

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ  
«БРЕСТСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра высшей математики

**Б. А. ГОДУНОВ**

# **СТАТИСТИКА**

**Часть 2**

**(Конспект лекций)**

Брест 2009

УДК 31  
ББК 60. 6я73  
Г 59

Рецензенты:

**В.Ф. Савчук**, доцент, зав. кафедрой информатики и прикладной математики УО «Брестский государственный университет им. А. С. Пушкина», к.ф.-м.н.

Кафедра высшей математики УО «БрГТУ»

Предлагаемая Вашему вниманию работа является продолжением пособия «Статистика, часть 1 (конспект лекций)», опубликованного в 2008 году. Она написана по материалам лекций по курсу «Статистика», читавшимся автором на протяжении десяти лет в Брестском государственном техническом университете. В «Статистике, ч.2» продолжена нумерация глав первой части.

В данном пособии три раздела курса. В первом из них рассматриваются основы корреляционного анализа. При этом основное внимание уделяется двухфакторному анализу. Второй раздел посвящен рядам динамики и основным показателям их, а также рассматриваются элементы выравнивания рядов. В третьем рассмотрены основы теории индексов.

Так же, как и часть 1, это учебное пособие не только поможет студентам самостоятельно освоить важные вопросы описательного плана, но и глубже разобраться в тех вопросах, где применяется математический аппарат статистики и статистический анализ данных и результатов их обработки.

Издается в 2-х частях. Часть 2.

**Годунов Б. А.**

Г 59 **Статистика, часть 2 (конспект лекций)**  
Брест, Издательство БрГТУ, 2009, с. 85.

**ISBN 978-985-493-139-5**

**ISBN 978-985-493-139-5**

© Годунов Б. А., 2009  
© Издательство БрГТУ, 2009

## ГЛАВА 6

### ОСНОВЫ КОРРЕЛЯЦИОННОГО АНАЛИЗА

#### 6.1. Различные формы связи между признаками

В этой главе будет изучен один из множества многочисленных видов связей между признаками.

Отметим сразу же два «крайних» вида. Первый – это отсутствие или практическое *отсутствие связи*. Например, разумно согласиться с тем, что между производительностью труда на молочном комбинате города  $N$  и числом библиотек в городе  $M$  нет никакой связи. Второй вид – *функциональные зависимости*, когда каждому допустимому значению переменной  $x$  ставится в соответствие одно определенное значение переменной  $y$ . Таковой является, например, зависимость пройденного пути от времени движения:  $S = v \cdot t$ , при постоянной скорости  $v$ . Заметим что, в некоторых случаях (многозначные функции) некоторым значениям аргумента может ставиться в соответствие несколько вполне определенных значений функции. Например, уравнение окружности  $x^2 + y^2 = R^2$  является функциональной зависимостью, в которой каждому значению  $x \in (-R; R)$  ставятся в соответствие два вполне определенных значения переменной  $y$  – одно с плюсом, другое с минусом.

В окружающем нас мире, и особенно в экономике, чаще случается так, что при заданном *значении*  $x$  одного признака  $X$  нельзя точно указать значение  $y$  другого признака  $Y$ , который может принять *одно из возможных значений*, то есть следует говорить о распределении признака  $Y$  при этом  $x$ . Причем с изменением  $x$  может меняться распределение признака  $Y$ . В этом случае речь идет о *стохастической связи*. Слово *стохастический* происходит от греческого «*stochasis*» – догадка и употребляется в смысле «случайный или вероятностный». Изучение стохастической связи является весьма трудоемкой задачей. На практике рассматривается частный случай таких связей, когда для каждого значения  $x$  одного признака рассматривается среднее значение другого признака. Если оно меняется при изменении  $x$ , то его называют *условным средним значением* и обозначают  $\bar{y}_x$  или  $\bar{y}(x)$ . Изучением зависимости между одним признаком и условным средним значением другого признака занимается *корреляционный анализ*. Слово «корреляция» является производным от латинского «*correlatio*» – соотношение, взаимосвязь предметов, явлений и понятий. Функция

$$\bar{y}_x = f(x),$$

выражающая эту связь, называется *уравнением регрессии*, а ее график – *кривой регрессии*  $Y$  по  $X$  (или  $Y$  на  $X$ ). «Регресс» происходит от

латинского «*regressus*» - возвращение, движение назад, и в данном случае означает тот факт, что изменение признака  $X$  возвращается нам как изменение условной средней признака  $Y$ .

Независимый признак  $X$  в статистике называют *признаком-фактором*, а зависящий от него признак  $Y$  называют *результативным* признаком или *признаком-следствием*. Если признак-фактор является одномерной величиной, то говорят об *однофакторном* корреляционном анализе, а если  $X$  – многомерный признак, то есть представляет собой совокупность нескольких факторов, то говорят о *многомерном* анализе. Нашей целью является изучение основных задач и методов однофакторного корреляционного анализа.

Уравнение регрессии позволяет сделать прогноз ожидаемого значения признака-следствия при планируемом значении признака-фактора. Возникает вопрос: почему с этой целью рассматривают условную среднюю? Оказывается, что функция  $\bar{y}(x)$  дает наименьшую среднюю погрешность оценки прогноза.

Отметим две основные задачи теории корреляции. *Первая* – установить форму корреляционной связи, то есть вид функции регрессии  $f(x)$  (линейная, квадратичная, показательная и т.д.), и рассчитать ее параметры. *Вторая* – оценить тесноту (силу) корреляционной связи между признаками, о которой судят по величине рассеяния эмпирических значений признака  $Y$  вокруг условной средней  $\bar{y}_x$ . При этом большое рассеяние говорит о *слабой* зависимости между признаками, а малое – о *сильной*, и даже, может быть, о функциональной зависимости.

И, наконец, отметим, что можно рассматривать «обратную» корреляционную связь признака  $X$  от признака  $Y$  в виде уравнения

$$\bar{x}_y = g(y),$$

называемого уравнением регрессии  $X$  на  $Y$ , где  $\bar{x}_y$  – условная средняя признака  $X$  при заданном значении  $y$  признака  $Y$ .

Применение термина *регрессия* (см.[1]) в статистике связывают с именами создателей корреляционного анализа Ф. Гальтона (1822 -1911) и К. Пирсона, которые интересовались связью между ростом отцов и сыновей. Ф. Гальтон изучил более 200 семей и обнаружил, что в группе семей с высокорослыми отцами сыновья в среднем ниже ростом, чем их отцы, а в группе семей с низкорослыми отцами сыновья в среднем выше отцов. Таким образом, отклонение роста от средней в следующем поколении уменьшается – *регрессирует*. Причина в том, что на рост сыновей влияет не только рост отцов, но и рост матерей и много других факторов развития ребенка. И эти факторы, случайно направленные как в сторону увеличения, так и снижения роста, приближают рост сыновей к среднему росту.

Отметим основные этапы корреляционного анализа.

1. Экономист, исходя из смысла признаков, ставит задачу о корреляционной связи, выбирая при этом признак-фактор и признак-следствие.
2. Статистик собирает статистические данные и, используя их, выясняет количественно вопрос о существовании связи между признаками.
3. Статистик устанавливает направление этой связи. Если при возрастании признака-фактора наблюдается тенденция к возрастанию признака-следствия, то говорят о *прямой связи*. И говорят об *обратной связи*, если признак-следствие при этом имеет тенденцию убывания.
4. Статистик выясняет вид связи, то есть вид уравнения регрессии, и рассчитывает его параметры. На этом этапе возможен другой подход: по одним и тем же данным можно рассчитать параметры нескольких наиболее подходящих уравнений, а затем, применяя известные методы, выбрать наилучшее из них.
5. Статистик рассчитывает тесноту связи между признаками с определенной степенью уверенности, то есть дает вероятностную оценку полученной тесноты.
6. Экономист делает окончательные выводы.

## 6.2. Предпосылки корреляционной связи

В дальнейшем мы будем рассматривать однофакторную корреляцию. В этом случае на начальном этапе располагают набором пар значений  $(x_i; y_i)$ ,  $i = 1 \div n$  признаков  $X$  и  $Y$ , полученных для  $n$  единиц некоторой статистической совокупности. Рассмотрим, к примеру, связь между объемом выполненных работ ( $Y$ , млн. руб.) и накладными расходами ( $X$ , млн. руб.). Выбор признака-фактора и признака-следствия определяется специалистами в каждом конкретном случае и зависит от цели исследования. По 30 предприятиям получены данные, которые затем были ранжированы по признаку  $X$ .

Таблица 6.1

<b>№</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
<b>X</b>	1,0	1,5	1,6	1,7	1,8	2,2	2,2	2,3	2,6	2,6
<b>Y</b>	12	16	23	13	19	27	36	14	33	27
<b>№</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>
<b>X</b>	2,8	2,9	2,9	3,2	3,3	3,4	3,5	3,6	3,6	3,8
<b>Y</b>	23	19	38	32	36	26	33	39	42	33
<b>№</b>	<b>21</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>28</b>	<b>29</b>	<b>30</b>
<b>X</b>	3,8	4,2	4,2	4,6	4,8	5,1	5,2	5,4	5,6	6,0
<b>Y</b>	36	36	46	39	42	48	42	49	43	47

Просматривая значения признака  $Y$  (по возрастанию признака  $X$ ), видим, что строгого возрастания их нет, но есть *тенденция к возрастанию*.

нию, т.е. в массе своей значения  $Y$  возрастают от начала ряда к его концу. Этот факт является *первой предпосылкой* существования *прямой связи* между рассматриваемыми признаками.

*Вторая предпосылка* связана с изображением каждой пары  $(x_i; y_i)$  точкой в системе координат, в которой на горизонтальной оси откладывают координаты признака-фактора, а на вертикальной – признака-следствия. Полезно предварительно разбить множества значений обоих признаков на интервалы (не обязательно в равных количествах) и отобразить это в виде сетки, что помогает дальнейшим подсчетам точек, попавших в те или иные интервалы. Такое изображение корреляционной зависимости называется *полем корреляции*.

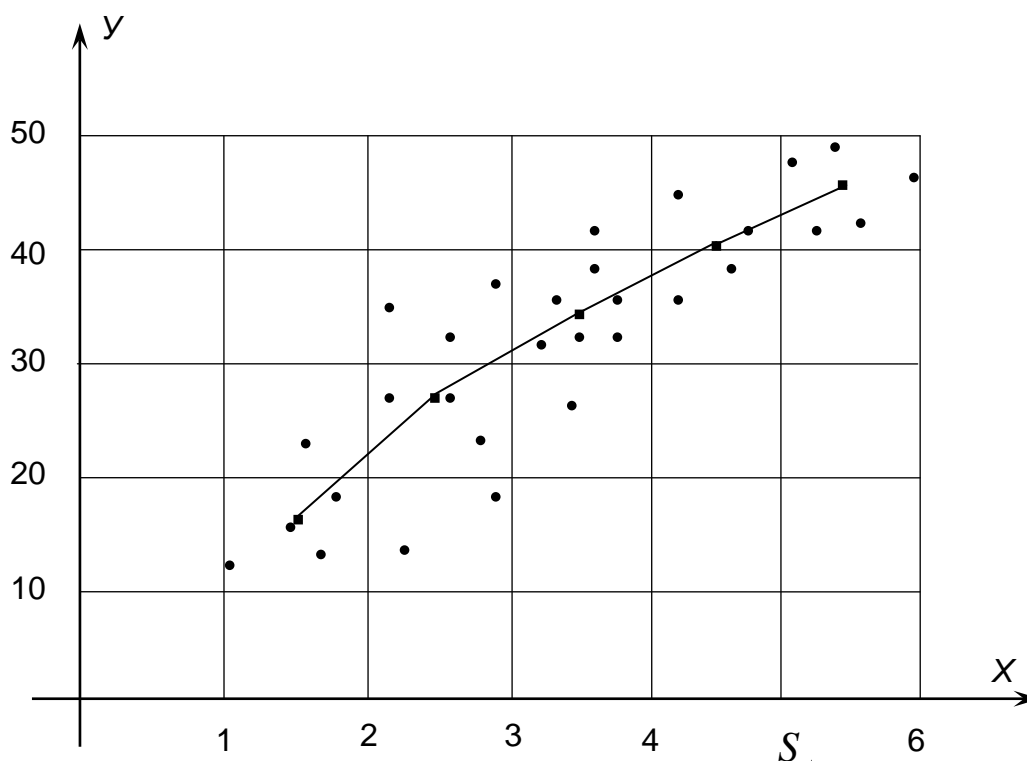


Рис. 1

По расположению точек на чертеже можно констатировать, что с увеличением  $X$  значения  $Y$  имеют *тенденцию* к возрастанию, то есть можно говорить о *прямой связи* между признаками. А несильный разброс точек вдоль направления возрастания дает повод предполагать достаточно тесную связь между ними.

*Третья предпосылка* предусматривает составление *групповой таблицы*, в которой вычисляются средние значения признака  $Y$  для точек, координаты  $x_i$  которых попадают в один интервал по признаку  $X$ .

В нашем случае весь интервал вариации признака-фактора разобьем на 5 группировочных интервалов: 1 – 2; 2 – 3; 3 – 4; 4 – 5; 5 – 6. Тогда для первого интервала получим среднее значение признака-следствия

$$\bar{y}_1 = \frac{12 + 16 + 23 + 13 + 19}{5} = 16,6, \text{ для второго получим аналогично}$$

$$\bar{y}_2 = \frac{27 + 36 + 14 + 33 + 27 + 23 + 19 + 38}{8} = 27,1 \text{ и так далее. В итоге со-}$$

ставляем *таблицу 6.2*. Тот факт, что при переходе к интервалам с большими значениями признака  $X$  *интервальные средние*  $\bar{y}_i$  ( $i = 1 \div 5$ ) строго возрастают, с еще большей уверенностью подтверждает наличие прямой связи между признаками.

*Четвертая предпосылка* заключается в том, данные полученной *групповой таблицы* можно использовать для прогноза вида

*Таблица 6.2*

№ п/п	Накладные расходы млн. руб. $a_i - a_{i+1}$	Число Предприятий $n_i$	Средний объем работ млн. руб. $\bar{y}_i$
1	1 – 2	5	16,6
2	2 – 3	8	27,1
3	3 – 4	8	34,6
4	4 – 5	4	40,8
5	5 – 6	5	45,8
Всего		30	

для уравнения регрессии. Для этого на корреляционном поле отложим точки с координатами  $(x_i; \bar{y}_i)$ , где  $x_i$  – середины интервалов, и последовательно соединим их отрезками прямых (на графике на стр. 6 эти точки изображены квадратиками). Как видим, полученная ломаная линия достаточно близка к прямой линии, поэтому можно предположить, что между признаками имеется линейная связь  $\bar{y}_x = a x + b$ .

И, наконец, *пятая предпосылка*. Составляется *корреляционная таблица*, которая является результатом группировки по обоим признакам с

*Таблица 6.3*

Объем выполненных работ, млн. руб.	Накладные расходы млн.руб.					$n_y$
	1 – 2 1,5	2 – 3 2,5	3 – 4 3,5	4 – 5 4,5	5 – 6 5,5	
10 – 20 15	4	2				6
20 – 30 25	1	3	1			5
30 – 40 35		3	6	2		11
40 – 50 45			1	2	5	8
$n_x$	5	8	8	4	5	30

одновременным выделением для них *группировочных интервалов* и подсчетом так называемых совместных частот  $n_{xy}$ , аналогично тому, как это делалось при построении корреляционного поля.

Во внутренней части таблицы находятся частоты  $n_{xy}$ . В частности,  $n_{xy} = 6$  означает, что в интервал 3 – 4 по  $X$  и одновременно в интервал 30 – 40 по  $Y$  попало 6 пар наблюдавшихся значений. Пустая клетка соответствует тому, что в пересечении соответствующих ей интервалов не наблюдалось ни одной точки.

Если частоты  $n_{xy}$  расположены в основном вдоль главной диагонали, как в нашем случае, то это говорит о наличии *прямой* корреляционной связи признаков, а если вдоль другой диагонали, то о наличии *обратной* связи. Если совместные частоты заполняют практически всю таблицу, то выделяют максимальные и по порядку их расположения судят о возможности существования и виде связи. В таких случаях говорят о заметных «помехах», то есть о влиянии неучтенных факторов.

### 6. 3. Различные уравнения регрессии

Изучение корреляционной связи сводится к установлению зависимости между признаком-фактором и условным средним значением признака-следствия в виде уравнения корреляции

$$\bar{y}_x = f(x).$$

Наиболее употребительными являются следующие зависимости:

- |                                |  |
|--------------------------------|--|
| 1. линейная зависимость        | $\bar{y}_x = a + bx;$                                    |
| 2. квадратичная зависимость    | $\bar{y}_x = a + bx + cx^2;$                             |
| 3. логарифмическая зависимость | $\bar{y}_x = a + b \ln x$ или $\bar{y}_x = a + b \lg x;$ |
| 4. показательная зависимость   | $\bar{y}_x = a \cdot b^x;$                               |
| 5. степенная зависимость       | $\bar{y}_x = a \cdot x^b;$                               |
| 6. гиперболическая зависимость | $\bar{y}_x = a + \frac{b}{x}.$                           |

Важным является вопрос выбора уравнения регрессии. Главным при этом является содержательный анализ природы изучаемой зависимости, механизм взаимодействия признаков. При этом можно учитывать теоретические исследования данных признаков, а также накопленный ранее опыт, если уже применявшиеся уравнения регрессии давали достаточно хороший результат.

Кроме этого, большую пользу дает анализ конкретных статистических данных  $(x_i; y_i)$ . А в случае, когда корреляционная связь ранее не изучалась и невозможно теоретически обосновать использование некоторого вида уравнения регрессии, эти данные сыграют основную роль.



Во-первых, по форме эмпирической линии регрессии (ломаная линия, рис.1) можно судить о том, какую зависимость выбрать. Во-вторых, можно по одним и тем же данным рассчитать параметры нескольких уравнений регрессии, и, используя некоторые критерии, выбрать из них лучший вариант.

В заключение отметим, что параметры уравнений регрессии определяются так называемым *методом наименьших квадратов* (в дальнейшем МНК), суть которого заключается в следующем. Для каждого значения  $x_i$  рассматривают теоретические значения признака следствия  $\bar{y}(x_i) = f(x_i)$  и требуют, чтобы сумма квадратов отклонений их от эмпирических значений  $y_i$  была наименьшей, то есть ищут

$$\min \Phi = \min \sum_{i=1}^n (\bar{y}(x_i) - y_i)^2 .$$

Функция  $\Phi$  является функцией параметров уравнения регрессии, поэтому при отыскании минимума используют равенство нулю ее частных производных.

Метод наименьших квадратов в начале XIX века независимо друг от друга открыли французский математик Адриен Мари Лежандр и немецкий ученый Карл Фридрих Гаусс, который, кроме того, дал теоретико-вероятностное обоснование этого метода как наилучшего из всех возможных.

#### 6. 4. Парная линейная корреляция

Для отыскания параметров  $a$  и  $b$  линейного уравнения регрессии  $\bar{y}_x = a + b x$  согласно МНК ищут минимум по  $a$  и  $b$ :

$$\min \Phi (a, b) = \min \sum_{i=1}^n (a + b x_i - y_i),$$

где  $(x_i; y_i)$ ,  $i = 1 \div n$  – эмпирические данные изучаемых признаков.

Для отыскания минимума приравняем к нулю частные производные функции  $\Phi(a; b)$ . Получим систему уравнений относительно  $a$  и  $b$ :

$$\begin{cases} \frac{\partial \Phi}{\partial b} = \sum 2(a + b x_i - y_i) \cdot x_i = 0, \\ \frac{\partial \Phi}{\partial a} = \sum 2(a + b x_i - y_i) \cdot 1 = 0. \end{cases}$$

В дальнейшем для упрощения записи и лучшего прочтения мы не пишем пределы суммирования по  $i$  от 1 до  $n$ .

Сокращая оба уравнения системы на 2 и суммируя их почленно, после вынесения постоянных  $a$  и  $b$  за знаки сумм, получим *систему нормальных уравнений*

$$\begin{cases} b \sum x_i^2 + a \sum x_i = \sum x_i y_i, \\ b \sum x_i + a \cdot n = \sum y_i. \end{cases}$$

Здесь искомыми являются  $a$  и  $b$ , а остальные члены предварительно вычисляются по исходной информации.

Решать эту систему можно, например, по правилу Крамера, используя определители

$$\Delta = \begin{vmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{vmatrix}, \quad \Delta_a = \begin{vmatrix} \sum x_i^2 & \sum x_i y_i \\ \sum x_i & \sum y_i \end{vmatrix}, \quad \Delta_b = \begin{vmatrix} \sum x_i y_i & \sum x_i \\ \sum y_i & n \end{vmatrix},$$

$$a = \frac{\Delta_a}{\Delta}, \quad b = \frac{\Delta_b}{\Delta}.$$

Например, по данным пункта 6. 2 будем иметь  $\sum x_i = 101,4$ ,  $\sum y_i = 969$ ,  $\sum x_i^2 = 394,91$ ,  $\sum x_i y_i = 3640,5$ . Получим систему нормальных уравнений

$$\begin{cases} 394,91 b + 101,4 a = 3640,5, \\ 101,4 b + 30 a = 969. \end{cases}$$

$$\text{Отсюда } \Delta = \begin{vmatrix} 394,91 & 101,4 \\ 101,4 & 30 \end{vmatrix} = 1565,34, \quad \Delta_a = \begin{vmatrix} 394,91 & 3640,5 \\ 101,4 & 969 \end{vmatrix} = 13521,09,$$

$$\Delta_b = \begin{vmatrix} 3640,5 & 101,4 \\ 969 & 30 \end{vmatrix} = 10958,4.$$

Тогда  $a = \frac{13521,09}{1565,34} = 8,64$ ,  $b = \frac{10958,4}{1565,34} = 7,00$ , и линейное уравнение регрессии будет иметь вид

$$\bar{y}_x = 8,64 + 7,00 x.$$

Начертим график теоретической прямой в корреляционном поле, проведя ее через две точки А (1; 15,64) и В (6; 50,64), вычисленные по полученному уравнению.

Как видим, полученная прямая регрессии проходит вдоль поля корреляции, и она вполне отражает корреляционную зависимость между признаками.

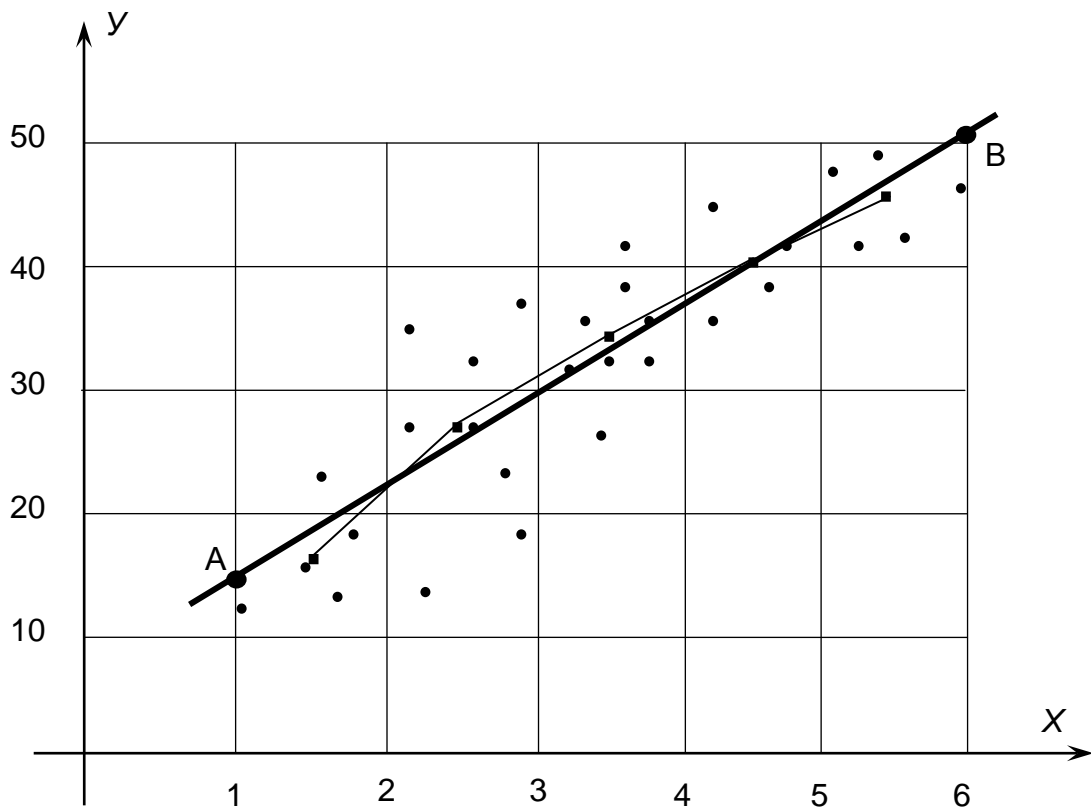


Рис. 2

Если пары значений данных  $(x_i, y_i)$  имеют частоты  $n_i$ , то система нормальных уравнений запишется в виде

$$\begin{cases} b \sum x_i^2 n_i + a \sum x_i n_i = \sum x_i y_i n_i, \\ b \sum x_i n_i + a \cdot n = \sum y_i n_i. \end{cases}$$

Например, если исходить из данных групповой *таблицы 2*, то в качестве  $x_i$  можно принять середины интервалов, в качестве  $y_i$  – интервальные средние  $\bar{y}_i$ , а в качестве частот – числа предприятий соответствующих интервалов. После этого производим предварительные расчеты и оформим их в *таблице 6.4*.

Таблица 6.4.

№ п/п	Накладные расходы млн. руб.	Середины интервалов $x_i$	Средний объем работ млн. руб. $y_i$	Число предприятий $n_i$	$x_i n_i$	$x_i^2 n_i$	$y_i n_i$	$x_i y_i n_i$
1	1 – 2	1,5	16,6	5	7,5	11,25	83,00	124,50
2	2 – 3	2,5	27,1	8	20,0	50,00	216,96	542,40
3	3 – 4	3,5	34,6	8	28,0	98,00	276,96	969,36
4	4 – 5	4,5	40,8	4	18,0	81,00	163,00	733,50
5	5 – 6	5,5	45,8	5	27,5	151,25	229,00	1259,50
Всего				30	101	391,50	968,92	3629,26

По данным итоговой строки составляем систему

$$\begin{cases} 391,50 b + 101 a = 3629,26, \\ 101 b + 30 a = 968,92, \end{cases}$$

решение которой приводит к уравнению регрессии  $\bar{y}_x = 8,275 + 7,135 x$ .

Эта прямая пройдет через точки (1; 15,410), (6; 51,085), практически совпадающие с аналогичными точками, полученными по исходным данным. Незначительное отклонение получилось за счет перехода к средним значениям признаков.

При использовании данных корреляционной таблицы, система нормальных уравнений запишется в виде

$$\begin{cases} b \sum x^2 n_x + a \sum x n_x = \sum x y n_{xy}, \\ b \sum x n_x + a \cdot n = \sum y n_y. \end{cases}$$

При этом в качестве значений признаков берутся середины соответствующих интервалов с частотами  $n_x$  и  $n_y$  в нижней строке и в правом столбце *таблицы* 6.3. Составим расчетную *таблицу* 6.5 (см. на следующей странице). Во внутренней части её под совместными частотами  $n_{xy}$ , выделенными жирным шрифтом, записаны произведения  $x y n_{xy}$ . Например,  $1,5 \cdot 15 \cdot 4 = 90$ ,  $1,5 \cdot 25 \cdot 1 = 37,5$ ,  $2,5 \cdot 15 \cdot 2 = 75$ ,  $2,5 \cdot 25 \cdot 3 = 187,5$  и т.д. В последней строке и в последнем столбце вычислены их суммы по столбцам и строкам соответственно. После этого они просуммированы. Обратите внимание – получен одинаковый результат 3590,0.

Используя итоговые данные, составляем систему

$$\begin{cases} 391,50 b + 101 a = 3590,0, \\ 101 b + 30 a = 960, \end{cases}$$

из которой получим уравнение корреляции  $\bar{y}_x = 8,582 + 6,956 x$ . Прямая регрессии пройдет через точки (1; 15,538) и (6; 50,318), и опять практически совпадающими с точками, полученными по исходным данным.

Таблица 6.5.

$x \backslash y$	1 – 2 1,5	2 – 3 2,5	3 – 4 3,5	4 – 5 4,5	5 – 6 5,5	$n_y$	$y n_y$	$x y n_{xy}$
10 – 20 15	4 90,0	2 75,0				6	90	165,0
20 – 30 25	1 37,5	3 187,5	1 87,5			5	125	312,5
30 – 40 35		3 262,5	6 735,0	2 315,0		11	385	1312,5
40 – 50 45			1 157,5	2 405	5 1237,5	8	360	1800,0
$n_x$	5	8	8	4	5	30	960	3590,0
$x n_x$	7,5	20,0	28,0	18,0	27,5	101		
$x^2 n_x$	11,25	50,00	98,00	81,00	151,25	391,50		
$x y n_{xy}$	127,5	525,0	980,0	720,0	1237,5	3590,0		

## 6. 5. Расчет параметров нелинейных уравнений регрессии

Для **квадратного уравнения** регрессии  $\bar{y}_x = a + b x + c x^2$  методом наименьших квадратов отыскивается минимум функции

$$\min \Phi(a, b, c) = \min \sum (a + b x_i + c x_i^2 - y_i)^2.$$

Находим частные производные и приравняем их к нулю:

$$\begin{cases} \frac{\partial \Phi}{\partial c} = \sum 2(a + b x_i + c x_i^2 - y_i) \cdot x_i^2 = 0, \\ \frac{\partial \Phi}{\partial b} = \sum 2(a + b x_i + c x_i^2 - y_i) \cdot x_i = 0, \\ \frac{\partial \Phi}{\partial a} = \sum 2(a + b x_i + c x_i^2 - y_i) \cdot 1 = 0. \end{cases}$$

Преобразуя эту систему аналогично тому, как это делалось в линейном случае, получим систему трех уравнений относительно трех неизвестных  $a$ ,  $b$  и  $c$ :

$$\begin{cases} c \sum x_i^4 + b \sum x_i^3 + a \sum x_i^2 = \sum x_i^2 y_i, \\ c \sum x_i^3 + b \sum x_i^2 + a \sum x_i = \sum x_i y_i, \\ c \sum x_i^2 + b \sum x_i + a \cdot n = \sum y_i. \end{cases}$$

Для всех остальных уравнений регрессии расчет их параметров сводится к линейному случаю.

**Логарифмическая** зависимость  $\bar{y}_x = a + b \ln x$ . С помощью замены переменной  $\ln x = t$  получаем линейную зависимость  $\bar{y}_t = a + bt$ . Для составления системы нормальных уравнений предварительно вычислим  $t_i = \ln x_i$ ,  $i = 1 \div n$ , после чего составляем систему уравнений

$$\begin{cases} b \sum t_i^2 + a \sum t_i = \sum y_i t_i, \\ b \sum t_i + a \cdot n = \sum y_i. \end{cases}$$

Найденные отсюда  $a$  и  $b$  подставляем в исходную функцию регрессии.

**Гиперболическая** зависимость  $\bar{y}_x = a + \frac{b}{x}$  заменой  $\frac{1}{x} = t$ ,  $t_i = \frac{1}{x_i}$ ,  $i = 1 \div n$  сводится к линейной зависимости  $\bar{y}_t = a + bt$  и системе

$$\begin{cases} b \sum t_i^2 + a \sum t_i = \sum y_i t_i, \\ b \sum t_i + a \cdot n = \sum y_i. \end{cases}$$

Уравнение **показательной** зависимости  $\bar{y}_x = a \cdot b^x$  прологарифмируем:  $\ln \bar{y}_x = \ln a + x \cdot \ln b$ . Далее заменим  $\ln \bar{y}_x = z$  и обозначим  $\ln a = a_1$  и  $\ln b = b_1$ . Получим линейное уравнение  $z = a_1 + b_1 x$ . Для отыскания его параметров решаем систему

$$\begin{cases} b_1 \sum x_i^2 + a_1 \sum x_i = \sum z_i x_i, \\ b_1 \sum x_i + a_1 \cdot n = \sum z_i, \end{cases}$$

где  $z_i = \ln y_i, i = 1 \div n$ . По найденному решению системы  $a_1$  и  $b_1$  находим исходные параметры  $a = e^{a_1}$  и  $b = e^{b_1}$  и тем самым определяем показательную зависимость.

**Степенное** уравнение регрессии  $\bar{y}_x = a \cdot x^b$  тоже логарифмируем:  $\ln \bar{y}_x = \ln a + b \cdot \ln x$ . Заменяя  $\ln \bar{y}_x = z, \ln x = t$  и обозначая  $\ln a = a_1$ , получим линейную зависимость  $z = a_1 + bt$ , для которой составляем систему

$$\begin{cases} b \sum t_i^2 + a_1 \sum t_i = \sum z_i t_i, \\ b \sum t_i + a_1 \cdot n = \sum z_i, \end{cases}$$

где  $z_i = \ln y_i, t_i = \ln x_i, i = 1 \div n$ . Решив систему, находим  $a = e^{a_1}$ , после чего получаем искомое уравнение регрессии.

Рассмотрим, к примеру, расчет параметров логарифмической корреляционной зависимости  $\bar{y}_x = a + b \ln x$  по данным *таблицы 6.4*. Учитывая частоты, система нормальных уравнений примет вид

$$\begin{cases} b \sum t_i^2 n_i + a \sum t_i n_i = \sum y_i t_i n_i, \\ b \sum t_i n_i + a \cdot n = \sum y_i n_i. \end{cases}$$

Составим расчетную таблицу.

Таблица 6.6

№ п/п	Сере- дины интер- валов $x_i$	Средний объем работ млн. руб. $y_i$	Число пред- при- ятий $n_i$	$t_i = \ln x_i$	$t_i n_i$	$t_i^2 n_i$	$y_i n_i$	$t_i y_i n_i$
1	1,5	16,6	5	0,405	2,025	0,820	83,00	33,615
2	2,5	27,1	8	0,916	7,328	6,712	216,96	198,735
3	3,5	34,6	8	1,253	10,024	12,560	276,96	347,031
4	4,5	40,8	4	1,504	6,016	9,048	163,00	245,152
5	5,5	45,8	5	1,705	8,525	14,535	229,00	390,445
Всего			30		33,918	43,675	968,92	1214,978

Подставляя данные итоговой строки, получим систему

$$\begin{cases} 43,675 b + 33,918 a = 1214,978, \\ 33,918 b + 30 a = 968,92. \end{cases}$$

Решением её является  $a = 22,435$ ,  $b = 6,933$ . Таким образом, логарифмическое уравнение регрессии имеет вид  $\bar{y}_x = 22,435 + 6,933 \ln x$ .

Наконец, отметим, что в ситуации, когда нет никаких предпосылок о виде уравнения корреляции, рассчитывают различные уравнения и для каждого из них вычисляют теоретические значения условных средних признака-следствия  $\bar{y}_{x_i} = \bar{y}(x_i)$ . Их сравнивают с эмпирическими значениями  $\bar{y}_i$ , вычисленными по статистическим данным при каждом  $x_i$ . Затем для каждой функции вычисляют среднеквадратическое отклонение теоретических и эмпирических значений условных средних

$$S = \sqrt{\frac{\sum (\bar{y}(x_i) - \bar{y}_i)^2 \cdot n_i}{n - m}},$$

где  $m$  - число параметров уравнения регрессии, которое равно трем для квадратичной зависимости и двум для остальных рассмотренных нами уравнений. После этого для использования выбирается функция с наименьшим значением  $S$ .

Например, для рассчитанных выше (см. таблицы 6.4 и 6.6) линейной и логарифмической функций получим:

Таблица 6.7

№ п/п	$x_i$	$\bar{y}_i = y_i$	$n_i$	Линейная функция		Логарифмическая функция	
				$\bar{y}(x_i)$	$(\bar{y}(x_i) - \bar{y}_i)^2 n_i$	$\bar{y}(x_i)$	$(\bar{y}(x_i) - \bar{y}_i)^2 n_i$
1	1,5	16,6	5	18,98	27,848	25,243	373,507
2	2,5	27,1	8	26,11	7,841	28,786	22,741
3	3,5	34,6	8	33,25	14,58	31,122	96,772
4	4,5	40,8	4	40,38	0,706	32,862	252,047
5	5,5	45,8	5	47,52	14,792	34,256	666,320
Всего			30		65,767		1411,387

Отсюда  $S_{\bar{y}\bar{y}} = \sqrt{\frac{65,767}{28}} = 1,48$ , а  $S_{\bar{y}\bar{a}} = \sqrt{\frac{1411,387}{28}} = 6,86$ . Таким образом, предпочтение следует отдать линейному уравнению регрессии.

### 6.6. Другая форма линейного уравнения регрессии. Коэффициент корреляции

Рассмотрим систему нормальных уравнений

$$\begin{cases} b \sum x_i^2 + a \sum x_i = \sum x_i y_i, \\ b \sum x_i + a \cdot n = \sum y_i \end{cases}$$



для расчета параметров линейного уравнения регрессии  $\bar{y}_x = a + b x$ . Разделив второе уравнение на  $n$ , получим  $b \frac{\sum x_i}{n} + a = \frac{\sum y_i}{n}$  или  $\bar{y} = b \bar{x} + a$ . Вычитая это равенство из уравнения регрессии, получим преобразованное уравнение

$$\bar{y}_x - \bar{y} = b(x - \bar{x}).$$

Отсюда в частности следует, что прямая линия регрессии проходит через точку со средними координатами признаков  $(\bar{x}, \bar{y})$ .

Далее из системы получаем  $b = \frac{n \cdot \sum x_i y_i - \sum x_i \cdot \sum y_i}{n \cdot \sum x_i^2 - (\sum x_i)^2}$ , а, учитывая, что  $\sum x_i = n \bar{x}$  и  $\sum y_i = n \cdot \bar{y}$ , получим

$$b = \frac{n \cdot \sum x_i y_i - n^2 \bar{x} \bar{y}}{n \cdot \sum x_i^2 - n^2 (\bar{x})^2} = \frac{n^2 \left( \frac{\sum x_i y_i}{n} - \bar{x} \bar{y} \right)}{n^2 \left( \frac{\sum x_i^2}{n} - (\bar{x})^2 \right)} = \frac{\overline{xy} - \bar{x} \bar{y}}{\sigma_x^2}.$$

Отсюда, умножая числитель и знаменатель на  $\sigma_y$ , получим

$$b = \frac{\sigma_y}{\sigma_x} \cdot \frac{\overline{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y} \quad \text{или} \quad b = \frac{\sigma_y}{\sigma_x} \cdot \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n \sigma_x \sigma_y}.$$

При этом

$$r_{xy} = \frac{\overline{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y} \quad \text{или} \quad r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n \sigma_x \sigma_y}$$

называется *коэффициентом корреляции* или *выборочным коэффициентом корреляции*, если он вычислен по выборочным данным. Он служит для оценки тесноты линейной корреляционной зависимости между признаками. Можно доказать, что  $|r| \leq 1$ . При этом близость  $|r|$  к единице говорит о достаточно тесной связи между рассматриваемыми признаками, близость к нулю – о слабой связи или об отсутствии таковой. Положительное значение  $r$  указывает на прямую связь, а отрицательное – на обратную связь между признаками.

Таким образом,  $b = r_{xy} \cdot \frac{\sigma_y}{\sigma_x}$ . Подставляя его в преобразованное уравнение, окончательно получим

$$\bar{y}_x - \bar{y} = r_{xy} \cdot \frac{\sigma_y}{\sigma_x} \cdot (x - \bar{x}) \quad \text{или} \quad \frac{\bar{y}_x - \bar{y}}{\sigma_y} = r_{xy} \cdot \frac{x - \bar{x}}{\sigma_x}.$$

Параметр  $b$  называется *коэффициентом регрессии*. Он показывает, что при изменении среднего значения признака-фактора на 1 измерения, признак-следствие изменится на  $b$  принятых для него единиц. Что же касается коэффициента корреляции, то из равенства

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y}$$

его называют еще *стандартизованным коэффициентом регрессии*. Он показывает, что при отклонении признака-фактора на величину  $\sigma_x$  от своего среднего значения признак-следствие в среднем по всей статистической совокупности отклонится от своей средней величины на  $r_{xy}\sigma_y$ . И, что существенно, в отличие от коэффициента регрессии  $b$  коэффициент корреляции не зависит от размерностей признаков. Следовательно, его можно использовать для сравнения тесноты связи для любых пар признаков. Это свойство применяется в теории множественной регрессии, в которой существенную роль играют *парные коэффициенты корреляции*, вычисляемые по таким формулам.

Нетрудно показать, что коэффициент корреляции можно представить в виде, предложенном английским статистиком и философом Карлом Пирсоном (1857 – 1936)

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

или в преобразованном виде

$$r_{xy} = \frac{\sum t_{x_i} \cdot t_{y_i}}{n},$$

где переменные  $t_{x_i} = \frac{x_i - \bar{x}}{\sigma_x}$ ,  $t_{y_i} = \frac{y_i - \bar{y}}{\sigma_y}$  не имеют размерности и в

теории множественной корреляции называются *стандартизованными или нормированными переменными*.

В заключение заметим, что уравнение регрессии  $X$  на  $Y$  можно записать в виде:

$$\frac{\bar{x}_y - \bar{x}}{\sigma_x} = r_{xy} \frac{y - \bar{y}}{\sigma_y}.$$

### 6.7. Коэффициент Фехнера оценки тесноты зависимости

Простейшим показателем тесноты связи между двумя статистическими признаками является *коэффициент корреляции знаков*  $K_\phi$ , предложенный немецким ученым Г. Фехнером (1801 – 1887). Для его вычисления находят средние значения признаков  $\bar{x}$  и  $\bar{y}$ , рассматривают знаки отклонений  $x_i - \bar{x}$  и  $y_i - \bar{y}$ . Затем подсчитывают  $n_a$  – число совпадений знаков соответствующих пар отклонений и  $n_b$  – число несовпадений (смотрите *таблицу 6.8* на следующей странице).

После этого вычисляют коэффициент Фехнера

$$K_\phi = \frac{n_a - n_b}{n_a + n_b}.$$

Очевидно, коэффициент Фехнера может принимать значения в пределах от  $-1$  до  $1$ . Если при этом все знаки совпадут, то  $n_b = 0$  и коэффициент будет равен  $1$ . В этом случае можно предполагать наличие прямой связи между признаками. Если же знаки всех пар будут несовпадающими ( $n_a = 0$ ), то коэффициент будет равен  $-1$ , что является свидетельством возможной обратной связи.

Рассчитаем  $K_\phi$  по данным *таблицы 6.1*. Среднее значение накладных расходов по всем предприятиям будет равно  $3,38$  млн. руб., а средний объем выполненных работ –  $32,3$  млн. руб. Распределение знаков отклонений показано в *таблице 6.8*. Из нее получаем  $n_a = 25$ , а  $n_b = 5$ . Тогда

$$K_\phi = \frac{25 - 5}{25 + 5} \approx 0,67.$$

Таблица 6.8.

№ п/п	$x_i$	$y_i$	Знаки отклонений		Совпадение ( a ) или несовпадение ( b )
			$x_i - \bar{x}$	$y_i - \bar{y}$	
1	2	3	4	5	6
1	1,0	12	-	-	a
2	1,5	16	-	-	a
3	1,6	23	-	-	a
4	1,7	13	-	-	a
5	1,8	19	-	-	a
6	2,2	27	-	-	a
7	2,2	36	-	+	b
8	2,3	14	-	-	a
9	2,6	33	-	+	b
10	2,6	27	-	-	a
11	2,8	23	-	-	a
12	2,9	19	-	-	a
13	2,9	38	-	+	b
14	3,2	32	-	-	a
15	3,3	36	-	+	b
16	3,4	26	+	-	b
17	3,5	33	+	+	a
18	3,6	39	+	+	a
19	3,6	42	+	+	a
20	3,8	33	+	+	a
21	3,8	36	+	+	a
22	4,2	36	+	+	a
23	4,2	46	+	+	a
24	4,6	39	+	+	a
25	4,8	42	+	+	a
26	5,1	48	+	+	a
27	5,2	42	+	+	a
28	5,4	49	+	+	a
29	5,6	43	+	+	a
30	6,0	47	+	+	a

Полученная величина коэффициента Фехнера 0,67 близка к 1, из чего можно предположить наличие достаточно тесной прямой корреляционной зависимости между исследуемыми признаками в данной статистической совокупности предприятий.

Заметим, что коэффициент Фехнера не учитывает числовые значения отклонений вариант от своих средних величин. Поэтому он может применяться только как еще один показатель наличия или отсутствия связи между признаками, а также направления этой связи. Если вспомнить, что аналогичные выводы можно получить из групповой или корреляционной таблицы, то станет ясно, что расчет коэффициента Фехнера лишней раз подтверждает уже имеющуюся *предварительную* информацию, которая в дальнейших исследованиях перекрывается более точными показателями.

## 6.8. Оценка тесноты линейной зависимости

Для оценки тесноты линейной корреляционной зависимости рассматривают отклонения  $x_i - \bar{x}$  и  $y_i - \bar{y}$ . Но поскольку природа признаков  $X$  и  $Y$ , как правило, различна, то сопоставление отклонений между собой оказывается некорректным. Поэтому рассматривают нормированные отклонения  $t_{x_i} = \frac{x_i - \bar{x}}{\sigma_x}$  и  $t_{y_i} = \frac{y_i - \bar{y}}{\sigma_y}$ , не имеющие размерности.

В качестве характеристики тесноты линейной связи рассматривают среднее произведение нормированных отклонений

$$r_{xy} = \frac{\sum t_{x_i} \cdot t_{y_i}}{n} = \frac{\sum \frac{x_i - \bar{x}}{\sigma_x} \cdot \frac{y_i - \bar{y}}{\sigma_y}}{n}.$$

Так как  $\sigma_x$  и  $\sigma_y$  являются постоянными величинами относительно индекса суммирования  $i$ , то их можно вынести за скобку (в нашем случае за знак суммы). Тогда получим

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y}.$$

Вычисление коэффициента корреляции по этой формуле достаточно громоздко. К тому же на его погрешность повлияют погрешности вычисления средних величин и среднеквадратических отклонений. Поэтому, проведем несложные преобразования

$$\begin{aligned} r_{xy} &= \frac{\sum x_i y_i - \bar{x} \cdot \sum y_i - \bar{y} \cdot \sum x_i + n \cdot \bar{x} \cdot \bar{y}}{n \cdot \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} \cdot \sqrt{\frac{\sum y_i^2}{n} - \left(\frac{\sum y_i}{n}\right)^2}} = \\ &= \frac{\sum x_i y_i - \frac{\sum x_i}{n} \cdot \sum y_i - \frac{\sum y_i}{n} \cdot \sum x_i + n \cdot \frac{\sum x_i}{n} \cdot \frac{\sum y_i}{n}}{n \cdot \frac{\sqrt{n \sum x_i^2 - (\sum x_i)^2}}{n} \cdot \frac{\sqrt{n \sum y_i^2 - (\sum y_i)^2}}{n}}, \end{aligned}$$

откуда легко получим

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \cdot \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n \sum y_i^2 - (\sum y_i)^2}}.$$

Воспользуемся данными *таблицы 6.1* и вычислим линейный коэффициент корреляции. Имеем

$$\begin{aligned} \sum_{i=1}^{30} x_i &= 101,4; & \sum_{i=1}^{30} y_i &= 969; & \sum_{i=1}^{30} x_i y_i &= 3640,5; \\ \left( \sum_{i=1}^{30} x_i \right)^2 &= 10281,96; & \left( \sum_{i=1}^{30} y_i \right)^2 &= 938961; \\ \sum_{i=1}^{30} x_i^2 &= 394,91; & \sum_{i=1}^{30} y_i^2 &= 34811. \end{aligned}$$

Тогда

$$r_{xy} = \frac{30 \cdot 3640,5 - 101,4 \cdot 969}{\sqrt{(30 \cdot 394,91 - 10281,96) \cdot (30 \cdot 34811 - 938961)}} = 0,8532.$$

Существенная близость к единице полученного линейного коэффициента корреляции говорит в пользу достаточно тесной прямой связи между рассматриваемыми признаками.

Величина  $r^2$  называется *коэффициентом детерминации*. В нашем случае  $r^2 = 0,7279$ . Умножив его на 100%, получим: 72,79% вариации объема выполненных работ (  $Y$  ) в рассматриваемой группе предприятий происходит за счет вариации накладных расходов (  $X$  ).

Если статистические данные представлены корреляционной таблицей, то в формуле коэффициента корреляции следует учесть частоты вариантов (  $n_x, n_y$  ) и совместные частоты (  $n_{xy}$  ):

$$r_{xy} = \frac{n \cdot \sum x y n_{xy} - \sum x n_x \cdot \sum y n_y}{\sqrt{n \sum x^2 n_x - (\sum x n_x)^2} \cdot \sqrt{n \sum y^2 n_y - (\sum y n_y)^2}}.$$

Например, чтобы вычислить коэффициент корреляции по сгруппированным данным *таблицы 6.5*, вычислим только  $\sum y^2 n_y = 34150$ . Тогда

$$r_{xy} = \frac{30 \cdot 3590,0 - 101 \cdot 960}{\sqrt{(30 \cdot 391,5 - 101^2) \cdot (30 \cdot 34150 - 960^2)}} = 0,8521.$$

Как видим, переход к сгруппированным данным, связанный с заменой конкретных вариантов серединами соответствующих интервалов с учетом частот, несущественно изменил значение коэффициента корреляции.

Отметим одно важное *свойство*: если сделать замену  $x = h_1 \cdot u + c_1$  и  $y = h_2 \cdot v + c_2$ , и вычислить коэффициент корреляции по формуле, аналогичной той, что была получена в п. 6.6,

$$r_{uv} = \frac{\sum u \cdot v - n \cdot \bar{u} \cdot \bar{v}}{n \cdot \sigma_u \cdot \sigma_v},$$

то окажется, что  $r_{uv} = r_{xy}$ . То есть линейная замена не меняет линейного коэффициента корреляции. Такой переход *удобно* использовать, когда исходные признаки имеют очень большие или очень малые значения. Кроме того, такая замена позволяет существенно сократить погрешности вычислений.

### 6.9. Существенность коэффициента корреляции

Как правило, при вычислении коэффициента корреляции получают величину, не равную нулю. Возникает вопрос: насколько существенно это отличие от нуля? В связи с этим требуется оценить надежность полученного коэффициента корреляции. Важную роль в решении этого вопроса играет объем выборочной совокупности.

1. Если *выборка большого объема* сделана из совокупности, имеющей нормальное распределение по обоим признакам, то распределение коэффициента корреляции можно считать близким к нормальному со средним значением, равным  $r$ , и дисперсией  $\sigma_r^2 = \frac{(1-r^2)^2}{n-1}$ . Тогда средняя ошибка коэффициента корреляции находится по формуле

$$s_r = \frac{1-r_e^2}{\sqrt{n-1}},$$

где  $n$  – объем выборочной совокупности, а  $r_e$  – выборочный коэффициент корреляции, полученный по выборке.

Задавая доверительную вероятность  $\gamma$ , равной 0,95 или 0,99, из равенства  $\Phi(t) = \gamma/2$  по таблице (**Приложение 2**) находим значение  $t$ , после чего доверительный интервал для коэффициента корреляции генеральной совокупности запишется в виде:

$$r_e - t \cdot s_r \leq r_{ген} \leq r_e + t \cdot s_r \quad \text{или} \quad r_e - t \cdot \frac{1-r_e^2}{\sqrt{n-1}} \leq r_{ген} \leq r_e + t \cdot \frac{1-r_e^2}{\sqrt{n-1}}.$$

Если окажется, что  $r_e - t \cdot s_r > 0$  в случае положительности  $r_e$  или  $r_e + t \cdot s_r < 0$  в случае отрицательности  $r_e$ , то с уверенностью в 95%

или в 99% можно говорить об отличии от нуля коэффициента корреляции и тем самым о существенно тесной линейной зависимости между рассматриваемыми признаками.

2. Если *объем выборки мал*, то приведенная выше формула доверительного интервала неприменима. В этом случае рассматривают статистику

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

и при уровне значимости  $\alpha = 0,05$  или  $\alpha = 0,01$  и проверяется гипотеза  $H_0: r_{ген} = 0$  при конкурирующей гипотезе  $H_1: r_{ген} \neq 0$ .

Если коэффициент корреляции генеральной совокупности равен нулю, то статистика  $t$  имеет распределение Стьюдента с  $n - 2$  степенями свободы.

Для проверки выдвинутой гипотезы по данным выборочного наблюдения вычисляется значение статистики  $t_{набл}$  и сравнивается с  $t_{крит} = t(\alpha; k = n - 2)$  – ее критическим значением, которое находят по таблице распределения Стьюдента (**Приложение 3**).

Если окажется, что  $t_{набл} > t_{крит}$ , то с уверенностью 95% или 99% можно считать существенной линейную корреляционную зависимость между признаками.

Если же  $t_{набл} < t_{крит}$ , то следует усомниться в линейной зависимости между признаками.

В нашем случае получим

$$t_{набл} = \frac{0,8521 \cdot \sqrt{30-2}}{\sqrt{1-0,8521^2}} = 8,615.$$

Теперь по таблице распределения Стьюдента находим  $t_{крит} = t(0,01; 28) = 2,763$ . Как видим,  $t_{набл} > t_{крит}$ . Следовательно, с уверенностью в 99% можно говорить о существенности линейной зависимости между признаками.

И, наконец, когда линейный коэффициент корреляции, полученный по данным малой выборки, оказался близок к единице ( $r \geq 0,8$ ), для проверки его значимости рекомендуется использовать так называемый *метод преобразованной корреляции*, предложенный Р. Фишером. Он показал, что функция

$$Z' = \frac{1}{2} \ln \frac{1+r}{1-r}$$



приближается к нормальной кривой даже для выборок очень небольшого объема. При этом среднеквадратическая ошибка зависит только от объема выборки и имеет вид

$$\sigma_{Z'} = \frac{1}{\sqrt{n-3}}.$$

Далее рассматривается отношение  $\frac{Z'}{\sigma_{Z'}}$ . Если оно окажется больше

трех, то значение выборочного коэффициента корреляции, близкое к единице, существенно и можно уверенно говорить о наличии тесной линейной корреляционной связи между признаками.

В рассмотренном выше примере

$$\sigma_{Z'} = \frac{1}{\sqrt{30-3}} = 0,1924, \quad Z' = \frac{1}{2} \ln \frac{1+0,8521}{1-0,8521} = 1,2638 \quad \text{и} \quad \frac{Z'}{\sigma_{Z'}} = 6,5686 > 3,$$

то есть значение линейного коэффициента корреляции существенно и можно быть уверенными в тесной связи между признаками.

### 6.10. Оценка тесноты связи в нелинейном случае

В случае, когда между признаками объективно существует нелинейная корреляционная связь, применение линейного коэффициента корреляции дает заниженный результат и поэтому не может использоваться как показатель тесноты связи между признаками. В этой ситуации используют эмпирическое *корреляционное отношение*

$$\eta = \pm \sqrt{\frac{S_{yx}^2}{S_y^2}} \quad \text{или} \quad \eta = \pm \sqrt{\frac{\sum_j (\bar{y}_j - \bar{y})^2 \cdot n_j}{\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2}},$$

достаточно подробно рассмотренное в п. 3.2 части 1 данного конспекта лекций. Напомним только, что величина  $\eta$ , близкая к единице, говорит о тесной связи между признаками, а близкая к нулю – о практическом отсутствии связи.

В рассмотренном выше примере о связи объема выполненных работ и накладных расходов средний объем работ по 30 предприятиям составил  $\bar{y} = 32,3$  млн. руб. Тогда, используя данные групповой таблицы 6.2,

$$\begin{aligned} \text{найдем} \quad \sum_{j=1}^5 (\bar{y}_j - \bar{y})^2 \cdot n_j &= (16,6 - 32,3)^2 \cdot 5 + (27,1 - 32,3)^2 \cdot 8 + \\ &+ (34,6 - 32,3)^2 \cdot 8 + (40,8 - 32,3)^2 \cdot 4 + (45,8 - 32,3)^2 \cdot 5 = 2691,34. \end{aligned}$$

Общую дисперсию признака-следствия найдем, используя исходные данные,

$$S_y^2 = \frac{\sum y^2}{n} - (\bar{y})^2 = \frac{34811}{30} - 32,3^2 = 117,68.$$

Отсюда  $\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 = S_y^2 \cdot n = 117,68 \cdot 30 = 3530,40$ , и величина корреляционного отношения получится равной

$$\eta = \sqrt{\frac{2691,34}{3530,40}} = 0,8731.$$

Перед корнем взят знак «+» в силу отмечавшейся ранее прямой связи между признаками, а близость величины  $\eta$  к единице говорит о тесной связи между ними.

Значимость  $\eta$  оценивается с помощью  $F$ -критерия, статистика которого, называемая *дисперсионным отношением*, имеет вид

$$F_{\text{набл}} = \frac{\sum_{j=1}^k (\bar{y}_j - \bar{y})^2 \cdot n_j}{k-1} : \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}{n-k},$$

где  $k$  – число групп, на которые разбиты исходные данные,

$n_j$  – объем  $j$ -ой группы,  $j = 1 \div k$ ,

$n$  – объем статистической совокупности,

$\bar{y}_j$  – групповая средняя,  $j = 1 \div k$ ,

$\bar{y}$  – общая средняя величина признака-следствия.

В таблицах  $F$ -распределения даны предельные значения  $F$ -критерия для различных сочетаний чисел степеней свободы  $k_1 = k - 1$  и  $k_2 = n - k$  при уровнях значимости  $0,05$  и  $0,01$ . Если окажется, что расчетное значение  $F$ -критерия больше табличного, то с уверенностью в  $95\%$  или  $99\%$  можно говорить о существенности значения корреляционного отношения, а в случае близости его к единице – еще и о существенности корреляционной связи между изучаемыми признаками.

В рассматриваемом примере сумма  $\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$ , связанная с

внутригрупповой дисперсией, можно вычислить непосредственно, используя данные признака-следствия из *таблицы 6.1* и групповые средние  $\bar{y}_j$  из *таблицы 6.2*. В нашем случае она равна  $825,48$ . Тогда расчетное значение  $F$ -критерия получится равным

$$F_{\text{набл}} = \frac{2691,34}{5-1} : \frac{825,48}{30-5} = 20,38.$$

По таблице  $F$ -распределения (приложение 5) при уровне значимости 0,99 находим критическое значение

$$F_{крит} = F(0,99; 4; 25) = 13,93.$$

Так как  $F_{набл} > F_{крит}$ , то с уверенностью (или вероятностью) в 99% можно говорить о существенности полученного корреляционного отношения и, следовательно, о существенности тесной корреляционной связи между рассматриваемыми признаками.

Отметим, что можно сопоставлять линейный коэффициент корреляции и эмпирическое корреляционное отношение при условии, что они вычислены по одним и тем же данным. При этом рассматривают величину  $\eta^2 - r^2$ . Доказано, что она положительна, а если к тому же  $\eta^2 - r^2 < 0,1$ , то считается предпочтительным использовать линейное уравнение корреляционной зависимости.

В нашем случае  $\eta^2 - r^2 = 0,8731^2 - 0,8521^2 = 0,036 < 0,1$ , что говорит в пользу практического применения полученного выше линейного уравнения корреляции.

И, наконец, отметим, что при малом количестве вариантов в выделенных группах рекомендуется сделать поправку корреляционного отношения, приняв к руководству его скорректированное значение

$$\eta_{скор.}^2 = 1 - (1 - \eta^2) \cdot \frac{n - 1}{n - k}$$

Для рассмотренного нами примера получим

$$\eta_{скор.}^2 = 1 - (1 - 0,8731^2) \cdot \frac{30 - 1}{30 - 5} = 0,7243, \text{ откуда } \eta_{скор.} = 0,8511. \text{ И его}$$

близость к единице еще раз говорит о тесной связи между признаками.

## ГЛАВА 7

### КОРРЕЛЯЦИОННЫЕ СВЯЗИ МЕЖДУ НЕКОЛИЧЕСТВЕННЫМИ ПРИЗНАКАМИ

Здесь рассматриваются так называемые *непараметрические методы*, которые применяются к признакам, не имеющим количественных значений. Приведем некоторые примеры.

#### 7.1 Корреляция рангов. Коэффициент Спирмена

Как известно, эксперты предварительно оценивают шансы каждого из  $n$  спортсменов, собирающихся принять участие в предстоящих соревнованиях. По итогам их работы каждому спортсмену прогнозируется номер места, которое он может занять, – *ранг*  $x_i$  факторного признака  $X$ ,  $i = 1 \div n$ . В результате соревнований определяются занятые места – *ранги*  $y_i$  результативного признака  $Y$ ,  $i = 1 \div n$ . Работу экспертов можно признать идеальной, если  $x_i = y_i$  для всех спортсменов. Но такое практически не случается. Возникает вопрос, как же оценить работу этих экспертов?

Английский психолог К. Спирмен (1863 – 1945) в качестве оценки тесноты связи предложил *коэффициент корреляции рангов*, основанный на рассмотрении разности рангов факторного и результативного признаков  $d_i = x_i - y_i$ ,  $i = 1 \div n$ . Очевидно,  $\bar{d} = \bar{x} - \bar{y}$ . Рассмотрим дисперсию  $\sigma_d^2$ :

$$\sigma_d^2 = \frac{\sum_1^n d_i^2}{n} - (\bar{d}^2) = \frac{\sum_1^n (x_i - y_i)^2}{n} - (\bar{x} - \bar{y})^2.$$

Раскроем скобки и перегруппируем слагаемые:

$$\sigma_d^2 = \left( \frac{\sum x_i^2}{n} - (\bar{x})^2 \right) + \left( \frac{\sum y_i^2}{n} - (\bar{y})^2 \right) - 2 \left( \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n} \right).$$

Первая и вторая скобки представляют собой дисперсии рассматриваемых признаков. Что касается третьей скобки, то вспомним формулу линейного коэффициента корреляции

$$r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n \sigma_x \sigma_y},$$

откуда

$$\frac{\sum x_i y_i - n \bar{x} \bar{y}}{n} = r_{xy} \sigma_x \sigma_y.$$

Таким образом,  $\sigma_d^2 = \sigma_x^2 + \sigma_y^2 - 2r_{xy} \sigma_x \sigma_y$ . Отсюда получим

$$r_{xy} = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_d^2}{2\sigma_x \sigma_y}.$$

Учитывая, что и факторный, и результативный признаки пробегают одни и те же целые значения от 1 до  $n$ , то их средние значения равны, и поэтому  $\bar{d} = 0$ . Дисперсии также равны и  $\sigma_x^2 = \sigma_y^2 = \frac{n^2 - 1}{12}$ .

Отсюда дисперсия разности рангов

$$\sigma_d^2 = \frac{\sum_1^n d_i^2}{n} - (\bar{d}^2) = \frac{\sum_1^n d_i^2}{n} \quad \text{и} \quad r_{xy} = \frac{2\sigma_x^2 - \sigma_d^2}{2\sigma_x^2} = 1 - \frac{\sigma_d^2}{2\sigma_x^2}.$$

Подставляя сюда выражения для дисперсии  $\sigma_x^2$  и  $\sigma_d^2$ , окончательно получим формулу *коэффициента корреляции рангов Спирмена*, который вместо  $r_{xy}$  обозначают буквой  $\rho$ :

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

Заметим, что коэффициент Спирмена, как всякий коэффициент корреляции, изменяется в пределах от  $-1$  до  $+1$  и близость его к единице говорит о хорошей согласованности рангов факторного и результативного признаков.

**Рассмотрим пример.** В *таблице 7.1* на следующей странице приведены экспертные оценки шансов десяти спортсменов и места, занятые ими на соревнованиях.

Данные графы 4 говорят о том, что в основном шансы спортсменов оценивались экспертами достаточно хорошо, не считая спортсменов под номерами 3, 4 и 9. Для ответа на вопрос, насколько точно эксперты предсказали результаты соревнований, вычислим коэффициент корреляции рангов Спирмена, используя итоговый результат графы 5:

$$\rho = 1 - \frac{6 \cdot 42}{10 \cdot (10^2 - 1)} = 0,745.$$

Близость  $\rho$  к единице указывает на то, что эксперты достаточно хорошо предсказали результаты соревнований.

Для совокупностей небольшого объема ( $n \leq 30$ ) ранговый коэффициент корреляции не подчиняется нормальному закону, и поэтому нет смысла применять упоминавшуюся ранее формулу доверительного интервала,

Таблица 7.1.

Порядковый номер спортсмена	Ранг спортсмена по оценке экспертов $x_i$	Ранг спортсмена по итогам соревнований $y_i$	Разность рангов $ d_i  =  x_i - y_i $	$d_i^2$
1	2	3	4	5
1	6	5	1	1
2	8	7	1	1
3	1	4	3	9
4	3	6	3	9
5	10	8	2	4
6	5	3	2	4
7	9	9	0	0
8	2	1	1	1
9	7	10	3	9
10	4	2	2	4
Итого				42

использующую значение  $t$ , получаемое по нормированной функции Лапласа  $\Phi(t)$ . В этом случае для оценки степени доверия полученному коэффициенту  $\rho$  можно воспользоваться таблицей «Значения коэффициента корреляции рангов Спирмена для двусторонних пределов при уровне значимости  $\alpha$ ». В ней приводятся предельные (минимальные) значения коэффициентов корреляции рангов Спирмена при заданном уровне значимости  $\alpha$  нулевой гипотезы об отсутствии корреляционной связи и определенном объеме выборочных данных.

Задавая уровень значимости 5% ( $\alpha = 0,05$ ), при объеме совокупности  $n = 10$  по таблице **Приложение 9** находим критическое значение  $\pm 0,6364$ . Поскольку полученное нами  $\rho = 0,745$  больше критического, то с 95-процентной уверенностью можно говорить о хорошей работе экспертов.

Однако при уровне значимости  $\alpha = 0,01$  критическое значение равно 0,7818, то есть наш коэффициент 0,745 оказался меньше критического. Это означает, что с вероятностью 99% нулевая гипотеза об отсутствии корреляционной связи не может быть отвергнута, а величина 0,745 является результатом случайных совпадений, в то время как связь между оценками экспертов и результатами соревнований может отсутствовать. Поэтому к работе экспертов следует относиться с достаточной осторожностью.

## 7. 2. Коэффициент корреляции рангов Кендэла

Еще один показатель связи между ранговыми признаками  $X$  и  $Y$  – коэффициент корреляции рангов Кендэла  $\tau$  – предложил М. Кендэл:

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

Таблица 7.2

Ранг спортсмена по оценке экспертов (x)	1	2	3	4	5	6	7	8	9	10
Ранг спортсмена по итогам соревнований (y)	4	1	6	2	3	5	10	7	9	8

Для вычисления величин  $P$  и  $Q$  совокупность ранжируют по признаку  $X$ , приводя его к ряду натуральных чисел. Затем рассматривают полученный при этом ряд рангов признака  $Y$ .

Для нахождения  $P$  каждому рангу признака  $Y$  ставят в соответствие количество рангов, находящихся справа и имеющих большее значение. Например, первому рангу 4 соответствуют 6 таких рангов (6, 5, 10, 7, 9, 8), второму рангу 1 соответствуют 8 рангов (6, 2, 3, 5, 10, 7, 9, 8), третьему рангу 6 соответствуют 4 ранга (10, 7, 9, 8) и т. д. Слагаемое  $P$  вычисляется как сумма полученных таким образом количеств рангов. В нашем случае  $P = 6+8+4+6+5+4+0+2+0=35$ . Полученное слагаемое отражает *соответствие* распределения рангов признака  $Y$  распределению рангов признака  $X$ .

Второе слагаемое  $Q$  характеризует *несоответствие* распределения рангов признака  $Y$  распределению рангов признака  $X$ . Оно вычисляется как сумма количеств рангов, стоящих опять же справа, но имеющих при этом меньшее значение. Эти величины берутся со знаком минус. В нашем примере получим  $Q = -3 - 0 - 3 - 0 - 0 - 0 - 3 - 0 - 1 = -10$ . Отсюда  $S = H + Q = 35 - 10 = 25$ . Тогда коэффициент корреляции рангов Кэндела будет равен:

$$\tau = \frac{2 \cdot 25}{10 \cdot (10 - 1)} = 0,556.$$

Величина  $\tau$  не превосходит единицы, а равенство его нулю говорит об отсутствии связи между признаками.

При достаточно большом числе рангов между коэффициентами рангов Спирмена и Кендэла выполняется соотношение:  $\rho \approx \frac{3}{2} \tau$ .

При больших значениях  $n$  и заданном уровне значимости  $\alpha$  существенными считаются значения коэффициента Кендэла, удовлетворяющие неравенству  $\tau > t_\alpha \cdot \sqrt{\frac{2(2n+5)}{9n(n-1)}}$ , где  $t_\alpha$  определяется по таблице

(Приложение 2).

### (Приложение 2).

И последнее замечание. Если эксперты присвоили один и тот же ранг двум единицам совокупности, то каждой из них присваивается ранг, равный средней арифметической занятых ими мест (он может оказаться дробным). Например, если дважды повторился ранг 5, то считается, что эти две единицы заняли пятое и шестое места. Тогда каждой из них приписывается ранг  $(5 + 6) : 2 = 5,5$ , а все последующие ранги увеличиваются на 1. Например, единице, имевшей ранг 6, приписывается ранг 7 и т.д. В такой ситуации говорят, что совершен переход к *перестроенным рангам*.

При другом подходе рассматривают *связанные ранги*, т.е. одинаковые средние номера. В таком случае коэффициент Спирмена вычисляется по формуле

$$\rho = 1 - \frac{6 \sum d^2 - A - B}{\sqrt{(n^3 - n - 12A) \cdot (n^3 - n - 12B)}},$$

где

$$A = \frac{1}{12} \sum_j (A_j^3 - A_j); \quad B = \frac{1}{12} \sum_k (B_k^3 - B_k);$$

$j$  – номера связок по порядку для признака  $X$ ;

$A_j$  – число одинаковых рангов в  $j$ -ой связке по  $X$ ;

$k$  – номера связок по порядку для признака  $Y$ .

### 7.3. Некоторые другие методы измерения тесноты связи

Рассматриваемые в этом пункте методы связаны с различиями в постановке задач.

Так, для оценки тесноты связи между несколькими признаками в случае ранговых оценок можно использовать *коэффициент конкордации*  $\omega$ , вычисляемый по формуле:

$$\omega = \frac{12 \cdot S}{m^2 \cdot (n^3 - n)},$$

где  $m$  – число факторов,  $n$  – число ранжируемых единиц,  $S$  – сумма квадратов отклонений рангов, вычисляемая по формуле:

$$S = \sum_1^n \left( \sum_1^m r_{ij} \right)^2 - \frac{1}{n} \left( \sum_1^n \sum_1^m r_{ij} \right)^2,$$

где  $r_{ij}$  – ранг  $i$ -го фактора  $j$ -ой единицы.

Например, восемь предприятий ранжированы экспертами по уровню качества выпускаемой продукции, спросу на продукцию и уровню рентабельности. Данные представлены в *таблице 7.3* на следующей странице.

По итоговым (расчетным) данным двух последних столбцов находим сумму квадратов отклонений:

$$S = 1788 - \frac{108^2}{8} = 330.$$



Тогда коэффициент конкордации получится равным 0,873:

$$\omega = \frac{12 \cdot 330}{3^2 \cdot (8^3 - 8)} = 0,873.$$

Таблица 7.3.

Порядковый номер предприятия	Ранг по показателям ( $r_{ij}$ )			$\sum_1^m r_{ij}$	$\left(\sum_1^m r_{ij}\right)^2$
	уровень рентабельности	уровень качества	уровень спроса		
1	4	4	3	11	121
2	1	3	1	5	25
3	3	1	2	6	36
4	7	6	5	18	324
5	5	5	7	17	289
6	6	8	6	20	400
7	2	2	4	8	64
8	8	7	8	23	529
Итого				108	1788

Близость полученного коэффициента конкордации к единице говорит о возможности существования достаточно тесной связи между рассматриваемыми признаками в этой группе предприятий.

При изучении степени тесноты связи между качественными альтернативными признаками ( да, нет ) можно воспользоваться *коэффициентом ассоциации* или *коэффициентом контингенции*, первый из которых был предложен английским статистиком Д. Юлом, а второй – К. Пирсоном. Например, желая оценить влияние существующих форм повышения квалификации преподавателей университета на уровень их профессионального мастерства, была проведена аттестация студентами 320 преподавателей, из которых 240 повысили квалификацию. Обработанные данные представлены в *таблице 7.4*.

Таблица 7.4.

Группы преподавателей	Средний балл по сравнению с предыдущим по результатам аттестации		Всего
	не изменился или вырос	снизился	
Повысившие квалификацию по одной из форм	163 (a)	77 (b)	240
Не проходившие повышение квалификации по принятым формам	46 (c)	34 (d)	80
Всего	209	111	320

Эта таблица называется «*таблицей четырех полей*», частоты которой обозначены буквами *a, b, c, d*.

Коэффициент ассоциации определяется по формуле

$$K_A = \frac{ad - bc}{ad + bc}.$$

В нашем примере получим

$$K_A = \frac{163 \cdot 34 - 77 \cdot 46}{163 \cdot 34 + 77 \cdot 46} = +0,22.$$

Близость его к нулю заставляет усомниться во влиянии повышения квалификации на уровень профессионального мастерства.

Из приведенной формулы видно, что величина коэффициента ассоциации будет равна единице, если одна из частот окажется равной нулю, что дает преувеличенную оценку степени тесноты связи между признаками. В этом случае предпочтение отдается коэффициенту контингенции:

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

Если же число градаций качественных признаков больше двух, можно воспользоваться *критерием «хи-квадрат»*. В *таблице 7.5* приводятся результаты группировки 53 предприятий по техническому и организационному уровням развития производства (наблюдаемые частоты  $f_{ij}$ ), каждый из которых представляется на трех уровнях: ниже среднего уровня, средний уровень и выше среднего уровня.

Таблица 7.5.

Группы предприятий по техническому уровню	Группы предприятий по организационному уровню			Итого
	ниже среднего уровня	среднего уровня	выше среднего уровня	
ниже среднего уровня	7 (7,1)	8 (9,4)	10 (8,5)	25
среднего уровня	4 (3,4)	7 (4,5)	1 (4,1)	12
выше среднего уровня	4 (4,5)	5 (6,0)	7 (5,5)	16
Итого	15	20	18	53

При исследовании вопроса о зависимости двух признаков исходят из следующего тезиса: если между признаками отсутствует связь, то частоты каждой строки должны быть пропорциональны элементам итоговой строки. Аналогично со столбцами. В предположении отсутствия связи между признаками такое распределение частот принимается за теоретическое. Исходя из этого, рассчитывают теоретические частоты  $f'_{ij}$ , пропорциональные итоговой строке с весами, равными элементам итогового столбца (в таблице они указаны в скобках):

$$f'_{11} = \frac{15 \cdot 25}{53} \approx 7,1; \quad f'_{12} = \frac{20 \cdot 25}{53} \approx 9,4; \quad f'_{13} = \frac{18 \cdot 25}{53} \approx 8,5;$$

$$f'_{21} = \frac{15 \cdot 12}{53} \approx 3,4; \quad f'_{22} = \frac{20 \cdot 12}{53} \approx 4,5; \quad f'_{23} = \frac{18 \cdot 12}{53} \approx 4,1;$$

$$f'_{31} = \frac{15 \cdot 16}{53} \approx 4,5; \quad f'_{32} = \frac{20 \cdot 16}{53} \approx 6,0; \quad f'_{33} = \frac{18 \cdot 16}{53} \approx 5,5.$$

Расчетное значение критерия «хи-квадрат» определяют по формуле:

$$\chi^2 = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \frac{(f_{ij} - f'_{ij})^2}{f'_{ij}},$$

где  $k_1$  и  $k_2$  – число групп соответственно в строках и столбцах таблицы. В нашем случае  $k_1 = k_2 = 3$ .

Получим расчетное значение критерия:

$$\begin{aligned} \chi_{\text{расч.}}^2 = & \frac{(7 - 7,1)^2}{7,1} + \frac{(8 - 9,4)^2}{9,4} + \frac{(10 - 8,5)^2}{8,5} + \frac{(4 - 3,4)^2}{3,4} + \frac{(7 - 4,5)^2}{4,5} + \\ & + \frac{(1 - 4,1)^2}{4,1} + \frac{(4 - 4,5)^2}{4,5} + \frac{(5 - 6,0)^2}{6,0} + \frac{(7 - 5,5)^2}{5,5} = 5,064. \end{aligned}$$

Теперь по числу степеней свободы  $\nu = (k_1 - 1) \cdot (k_2 - 1) = 4$  при заданном уровне значимости  $\alpha$ , равном 0,05 или 0,01, из таблицы (**приложение 6**) определяют критическое значение, которое в этом примере будет равно:

$$\chi_{\text{крит.}}^2 = \chi^2(0,05; 4) = 9,488,$$

то есть  $\chi^2_{расч.} < \chi^2_{крит.}$ . Это означает, что с 95%-ой уверенностью можно не отвергать гипотезу об отсутствии связи между техническим уровнем и организационным уровнем в группе данных предприятий. То есть следует считать, что влияния этих признаков друг на друга практически нет.

Зная расчетное значение критерия «хи-квадрат», можно оценить степень тесноты связи, исходя из *коэффициента сопряженности* К. Пирсона

$$P = \sqrt{\frac{\chi^2}{n + \chi^2}} = \sqrt{\frac{5,064}{53 + 5,064}} = 0,295$$

или из *коэффициента взаимной сопряженности* А. А. Чупрова, позволяющего учесть число групп по каждому из признаков:

$$C = \sqrt{\frac{\chi^2}{n \sqrt{(k_1 - 1) \cdot (k_2 - 1)}}} = \sqrt{\frac{5,064}{53 \cdot \sqrt{(3 - 1) \cdot (3 - 1)}}} = 0,219.$$

Отличие от нуля полученных коэффициентов указывает на наличие некоторой связи между признаками, но близость их к нулю говорит о том, что теснота этой связи невелика.

## Г Л А В А 8

### РЯДЫ ДИНАМИКИ

#### 8. 1. Определение и виды рядов динамики

Одной из важных задач изучения статистических признаков является анализ поведения их во времени, на основе которого можно пытаться предсказывать последующие значения.

*Ряд значений некоторого статистического признака, расположенных в хронологической последовательности, называют временным или динамическим рядом (рядом динамики).* Исходя из этого определения, заметим, что на самом деле ряд динамики представляет собой распределение двух признаков: *первый признак* – это время, задаваемое в виде временных показателей (дат или промежутков времени), *второй признак* – это непосредственно изучаемый признак, значения которого называют *уровнями (ряда)*. Таким образом, можно дать еще одно определение: *статистическое распределение, в котором одним из признаков является время, называется рядом динамики.* Например, при изучении средней урожайности пшеницы в течение нескольких лет были получены следующие данные:

Таблица 8.1.

Годы	1989	1990	1991	1992
Урожайность пшеницы ц / га	50,1	48,3	55,6	57,8

Первая строка содержит временные показатели, а во второй указаны соответствующие им уровни ряда – урожайность.

В зависимости от способа задания временного показателя различают два вида динамических рядов. Если время задано в виде промежутка, то ряд динамики называется *интервальным*, а если значения изучаемого признака связаны с определенными датами (моментами времени), то ряд называется *моментным*. Ряд, представленный *таблицей 7.1*, является интервальным, так как урожайность является результатом усилий, затраченных в течение промежутка времени, равного году. Моментными являются, например, ряды, в которых рассматривается численность населения некоторого региона на конец года на протяжении некоторого периода или состояние станочного парка некоторого предприятия на начало года в течение нескольких лет и т.д.

Определенный интерес представляет геометрическое изображение ряда динамики. Для интервального ряда по горизонтальной оси (оси времени) откладывают промежутки времени и на каждом из них строят прямоугольник с высотой, равной соответствующему уровню ряда. Получается график, аналогичный гистограмме. Для моментного ряда на

оси времени отмечают моменты времени и для каждого из них строят точки со второй координатой, равной соответствующему уровню ряда. Полученные таким образом точки последовательно соединяют отрезками прямых линий, то есть получают картину, аналогичную полигону. Кроме этих основных форм геометрической интерпретации рядов применяется большое количество других, с которыми можно познакомиться, обратившись, например, к программе Excel.

И, наконец, отметим важное требование: промежутки времени в интервальных рядах должны быть равными и равными должны быть расстояния между соседними датами в моментных рядах. Например, такие периоды как месяц или квартал не являются равными из-за разного количества дней в них. В таких случаях делаются поправки на уровни ряда, учитывающие эту разницу временных промежутков.

Рассмотрение рядов динамики позволяет изучить особенности и закономерности развития изучаемого объекта наблюдения. При этом к основным задачам относятся следующие:

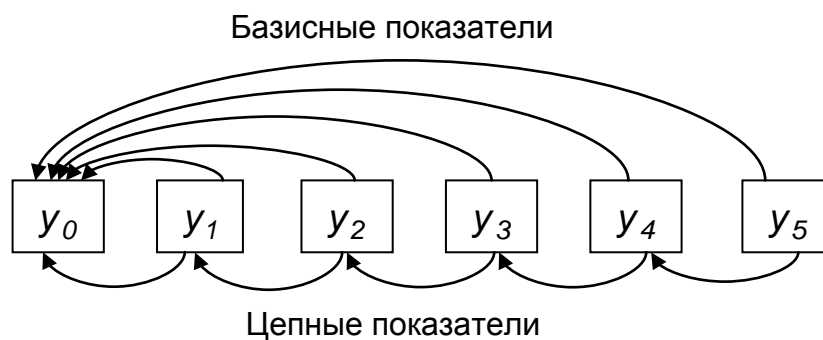
- характеристика интенсивности изменений уровней ряда от периода к периоду или от даты к дате;
- определение средних показателей ряда;
- выявление основных закономерностей динамики изучаемого признака на отдельных этапах и за весь период;
- выявление факторов, определяющих динамику уровней ряда;
- моделирование динамики уровней ряда и составление на его основе прогноза на будущее.

## 8. 2. Показатели рядов динамики

При сравнении уровней ряда динамики различают два вида показателей: 1) *абсолютные*, к числу которых относятся сами уровни ряда и абсолютный прирост, получаемый как разность уровней; 2) *относительные*, получаемые как отношение абсолютных или относительных показателей. К ним относятся темпы или коэффициенты роста или прироста, абсолютное значение одного процента прироста, пункты роста.

Последовательные во времени уровни ряда динамики признака  $Y$  обозначают  $y_0, y_1, y_2, \dots, y_{n-1}, y_n$ , при этом  $y_0$  называется *начальным* или *базисным* уровнем. Кстати, если возникает необходимость изучения динамики некоторого этапа данного ряда, то в качестве базисного берут ранний уровень этого этапа.

При построении показателей в качестве базы сравнения для уровня  $y_i$  выбирается либо начальный уровень  $y_0$ , либо непосредственно предшествующий ему уровень  $y_{i-1}$ . В первом случае говорят о *показателях с постоянной базой сравнения* и называют их *базисными показателями*, а во втором случае – о *показателях с переменной базой* и называют их *цепными показателями* (см. схему ниже).



### Схема построения базисных и цепных показателей

Базисные показатели характеризуют итоговый результат изменений за весь период, предшествующий уровню  $y_i$ , а цепные показатели отражают интенсивность изменения уровней от одного временного показателя к следующему за ним в пределах от начального до  $i$ -го периода. В статистике рядов динамики с уровнями  $y_0, y_1, y_2, \dots, y_i, \dots, y_n$  рассматривают следующие основные показатели.

1. *Абсолютные изменения уровня:*

базисные  $\Delta_{\text{б}} y_i = y_i - y_0$ ; цепные  $\Delta_{\text{ц}} y_i = y_i - y_{i-1}$ , ( $i = 1 \div n$ ).

2. *Темпы роста:*

базисные  $T_{\text{б}} y_i = \frac{y_i}{y_0}$  или в процентах  $T_{\text{б}} y_i = \frac{y_i}{y_0} \cdot 100\%$ ;

цепные  $T_{\text{ц}} y_i = \frac{y_i}{y_{i-1}}$  или в процентах  $T_{\text{ц}} y_i = \frac{y_i}{y_{i-1}} \cdot 100\%$ .

Заметим, что некоторые авторы темпы роста, вычисленные как отношение уровней, называют *коэффициентами роста* и обозначают буквой  $K$ , а вычисленные в процентах – собственно темпами роста.

3. *Темпы прироста:*

базисные  $\Delta T_{\text{б}} y_i = \frac{\Delta_{\text{б}} y_i}{y_0}$  или в процентах  $\Delta T_{\text{б}} y_i = \frac{\Delta_{\text{б}} y_i}{y_0} \cdot 100\%$ ;

цепные  $\Delta T_{\text{ц}} y_i = \frac{\Delta_{\text{ц}} y_i}{y_{i-1}}$  или в процентах  $\Delta T_{\text{ц}} y_i = \frac{\Delta_{\text{ц}} y_i}{y_{i-1}} \cdot 100\%$ .

Нетрудно увидеть связь между темпами роста и темпами прироста.

Действительно, например,  $\Delta T_{\text{ц}} y_i = \frac{\Delta_{\text{ц}} y_i}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} = \frac{y_i}{y_{i-1}} - 1$ , то есть

$\Delta T_{\text{ц}} y_i = T_{\text{ц}} y_i - 1$  или в процентах  $\Delta T_{\text{ц}} y_i(\%) = T_{\text{ц}} y_i(\%) - 100\%$ .

Аналогично  $\Delta T_{\text{б}} y_i = T_{\text{б}} y_i - 1$  и  $\Delta T_{\text{б}} y_i(\%) = T_{\text{б}} y_i(\%) - 100\%$  в процентах.

Если речь идет об одном и том же признаке, то в обозначениях рассмотренных показателей можно опустить обозначение признака и писать:  $\Delta_{\text{б}i}, \Delta_{\text{ц}i}, T_{\text{б}i}, T_{\text{ц}i}, \Delta T_{\text{б}i}, \Delta T_{\text{ц}i}$ .

4. Абсолютное значение одного процента прироста вычисляется по формуле

$$A_i = \frac{y_i - y_{i-1}}{\Delta T_{ц} y_i}$$

и показывает значимость цепного темпа прироста в сопоставлении его с абсолютным цепным приростом. Оказывается, замедление темпов прироста не всегда сопровождается уменьшением абсолютных приростов.

После простых преобразований получают удобную вычислительную формулу:

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

Отметим, что для практики большой интерес представляет параллельное рассмотрение и сравнение двух рядов динамики, технологически или экономически связанных между собой.

Рассмотрим пример вычисления введенных показателей для двух рядов динамики. Число перевезенных пассажиров за год (признак  $y$ ) – интервальный ряд, так как время здесь указано в виде промежутков, длиной в год. Число пассажирских вагонов на конец года (признак  $z$ ) – моментный ряд, поскольку время указано в виде даты – конец года.

Сначала рассмотрим ряд динамики для признака  $y$ . Данные и результаты расчетов внесем в таблицу.

Таблица 8.2.

$i$	Годы	Число пассажиров млрд. чел. $y_i$	Абсолютный прирост		Темп роста %		Темп прироста %		Абсолютное значение одного процента прироста млн. чел. $A_i$
			базисный $\Delta y_b$	цепной $\Delta y_c$	базисный $T_b$	цепной $T_c$	базисный $\Delta T_b$	цепной $\Delta T_c$	
1	2	3	4	5	6	7	8	9	10
0	1982	1,02	-	-	100,00	100,00	-	-	-
1	1983	1,03	0,01	0,01	100,98	100,98	0,98	0,98	10,2
2	1984	1,06	0,04	0,03	103,92	102,91	3,92	2,91	10,3
3	1985	1,11	0,09	0,05	108,82	104,72	8,82	4,72	10,6
4	1986	1,16	0,14	0,05	113,72	104,50	13,72	4,50	11,1
5	1987	1,20	0,18	0,04	117,65	103,45	17,65	3,45	11,6
6	1988	1,26	0,24	0,06	123,53	105,00	23,53	5,00	12,0

Поясним расчеты аналитических показателей ряда динамики.

Абсолютные приросты:

а) базисные:

$$\Delta y_{b1} = y_1 - y_0 = 1,03 - 1,02 = 0,01 \text{ млрд. чел.};$$

$$\Delta y_{b2} = y_2 - y_0 = 1,06 - 1,02 = 0,04 \text{ млрд. чел.};$$

$$\Delta y_{b3} = y_3 - y_0 = 1,11 - 1,02 = 0,09 \text{ млрд. чел. и т.д.};$$



б) цепные:

$$\Delta y_{ц1} = y_1 - y_0 = 0,01; \quad \Delta y_{ц2} = y_2 - y_1 = 1,06 - 1,03 = 0,03;$$

$$\Delta y_{ц3} = y_3 - y_2 = 1,11 - 1,06 = 0,05 \text{ млрд. чел. и т.д.}$$

Темпы роста:

а) базисные:

$$T_{б1} = \frac{y_1}{y_0} = \frac{1,03}{1,02} = 1,0098 \text{ или } 100,98\%, \quad T_{б2} = \frac{y_2}{y_0} = \frac{1,06}{1,02} = 1,0392 \text{ или}$$

$$103,92\%, \quad T_{б3} = \frac{y_3}{y_0} = \frac{1,11}{1,02} = 1,0882 \text{ или } 108,82\% \text{ и т.д.};$$

б) цепные:

$$T_{ц1} = \frac{y_1}{y_0} = 1,0098 \text{ или } 100,98\%, \quad T_{ц2} = \frac{y_2}{y_1} = \frac{1,06}{1,03} = 1,0291 \text{ или } 102,91\%,$$

$$T_{ц3} = \frac{y_3}{y_2} = \frac{1,11}{1,06} = 1,0472 \text{ или } 104,72\% \text{ и т.д.}$$

Темпы прироста:

а) базисные:

$$\Delta T_{б1} = \frac{y_1 - y_0}{y_0} = \frac{y_1}{y_0} - 1 = 1,0098 - 1 = 0,0098 \quad \text{или} \quad 0,98\%,$$

$$\Delta T_{б2} = 1,0392 - 1 = 0,0392 \text{ или } 3,92\%, \quad \Delta T_{б3} = 1,0882 - 1 = 0,0882 \text{ или } 8,82\% \text{ и т.д.};$$

б) цепные:

$$\Delta T_{ц1} = 1,0098 - 1 = 0,0098 \text{ или } 0,98\%, \quad \Delta T_{ц2} = 1,0291 - 1 = 0,0291 \text{ или } 2,91\%, \quad \Delta T_{ц3} = 1,0472 - 1 = 0,0472 \text{ или } 4,72\% \text{ и т.д..}$$

Абсолютное значение 1% прироста вычисляем по формуле

$$A_i = 0,01 y_{i-1}: \quad A_1 = 0,01 \cdot 1,02 = 0,0102 \text{ млрд. чел. или } 10,2 \text{ млн. чел.,}$$

$$A_2 = 0,01 \cdot 1,03 = 0,0103 \text{ млрд. чел. или } 10,3 \text{ млн. чел.,}$$

$$A_3 = 0,01 \cdot 1,06 = 0,0106 \text{ млрд. чел. или } 10,6 \text{ млн. чел. и т.д.}$$

Отметим связь между цепными и базисными показателями:

а) базисный абсолютный прирост  $\Delta y_{бi}$  равен сумме цепных абсолютных приростов за весь предыдущий период:  $\Delta y_{бi} = \Delta y_{ц1} + \Delta y_{ц2} + \dots + \Delta y_{цi}$ ,

например,  $\Delta y_{б5} = 0,18 = 0,01 + 0,03 + 0,05 + 0,05 + 0,04$ ;

б) базисный темп роста равен произведению цепных темпов роста за весь предшествующий период  $T_{бi} = T_{ц1} \cdot T_{ц2} \cdot \dots \cdot T_{цi}$ , например,

$$T_{б4} = 1,1372 = 1,0098 \cdot 1,0291 \cdot 1,0472 \cdot 1,0450.$$

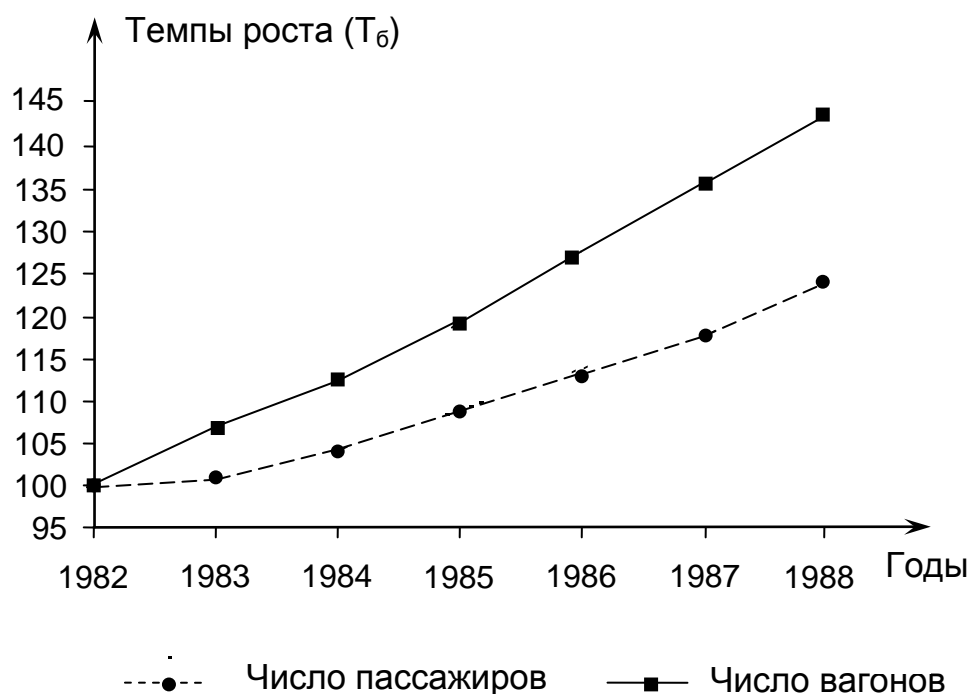
Аналогично рассчитывается таблица для признака z.

Таблица 8.3.

i	Годы	Число вагонов на конец года, тыс. шт. z	Абсолютный прирост		Темп роста %		Темп прироста, %		Абсолютное значение одного процента прироста, штук $A_i$
			базисный $\Delta y_b$	цепной $\Delta y_c$	базисный $T_b$	цепной $T_c$	базисный $\Delta T_b$	цепной $\Delta T_c$	
1	2	3	4	5	6	7	8	9	10
0	1982	1,20	-	-	100,0	100,0	-	-	-
1	1983	1,28	0,08	0,08	106,7	106,7	6,7	0,98	12,0
2	1984	1,35	0,15	0,07	112,5	102,91	12,5	2,91	12,8
3	1985	1,43	0,23	0,08	119,2	104,72	19,2	4,72	13,5
4	1986	1,53	0,33	0,10	127,5	104,50	27,5	4,50	14,3
5	1987	1,63	0,43	0,10	135,8	103,45	35,8	3,45	15,3
6	1988	1,72	0,52	0,09	143,3	105,52	43,3	5,52	16,3

Для наглядности некоторые показатели обоих рядов, например, темпы роста, можно изобразить геометрически. Для этого по данным второй и шестой граф таблицы 7.2 и таблицы 7.3 построим графики базисных темпов роста системы динамических рядов.

График базисных темпов роста системы рядов динамики



Рассматривая полученные показатели, следует отметить постоянный рост уровней обоих признаков с устойчивым увеличением абсолютных приростов, темпов роста и темпов прироста.

При этом по данным *таблиц 7.2, 7.3* и графика базисных темпов роста видно, что за весь период темпы роста числа вагонов опережают темпы роста числа пассажиров. Это говорит о том, что деятельность управления железной дороги направлена на лучшее обеспечение перевозок.

### 8. 3. Средние показатели ряда динамики

Средние показатели ряда динамики являются обобщающими характеристиками тенденции. К ним относятся средний уровень ряда, средний абсолютный прирост ряда, средние темпы роста и темпы прироста ряда и некоторые другие более специальные средние показатели. Если на протяжении рассматриваемого периода в ряде меняется тенденция, то может оказаться, что средний показатель не отражает динамики ряда. В таких случаях ряд разбивают на части с устойчивой тенденцией и рассматривают средние показатели для каждой из них. В случае больших колебаний уровней ряда построение средних показателей требует применения специальных методов статистики.

*Средний уровень интервального ряда с равными временными промежутками определяется как простая средняя арифметическая уровней ряда*

$$\bar{y} = \frac{\sum_{i=0}^n y_i}{n+1} \quad \text{или} \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n},$$

если нумерация начинается с 0 или с 1. Если же рассматривается ряд с неравными промежутками времени  $t_i$ , то средний уровень вычисляется как взвешенная средняя арифметическая

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

При вычислении среднего уровня *моментного* ряда условно переходят к интервальному ряду, полагая, что в течение периода между двумя соседними временными

рядов равен среднему арифметическому значению начального и конечного

уровней этого интервала:  $\frac{y_0 + y_1}{2}, \frac{y_1 + y_2}{2}, \dots, \frac{y_{n-1} + y_n}{2}$ . После этого

вычисляют среднюю арифметическую величину этих  $n$  показателей:

$$\bar{y} = \left( \frac{y_0 + y_1}{2} + \frac{y_1 + y_2}{2} + \dots + \frac{y_{n-1} + y_n}{2} \right) : n.$$

Преобразуя, окончательно получим формулу среднего уровня для моментного ряда динамики, которая называется *средней хронологической*:

$$\bar{y} = \left( \frac{y_0}{2} + y_1 + y_2 + \dots + y_{n-1} + \frac{y_n}{2} \right) : n.$$

Отметим, что эта формула получена в предположении *равенства промежутков* между соседними датами.

В примере предыдущего пункта для интервального динамического ряда ( $y$ ) получим средний уровень за 1982 -1988 годы

$$\bar{y} = \frac{1,02 + 1,03 + 1,06 + 1,11 + 1,16 + 1,20 + 1,26}{7} = \frac{7,84}{7} = 1,12,$$

а для моментного ряда динамики ( $z$ ) средний уровень будет равен:

$$\bar{z} = \left( \frac{1,20}{2} + 1,28 + 1,35 + 1,43 + 1,53 + 1,63 + \frac{1,72}{2} \right) : 6 = 8,68 : 6 = 1,45.$$

*Средний абсолютный прирост* или *средняя скорость роста* определяется как средняя арифметическая цепных абсолютных приростов:

$$\overline{\Delta y} = \frac{\sum_{i=1}^n \Delta_{цi}}{n},$$

где  $n+1$  – число уровней ряда. Учитывая свойство  $\sum_{i=1}^n \Delta_{цi} = \Delta_{бn} = y_n - y_0$ , получим формулу

$$\overline{\Delta y} = \frac{y_n - y_0}{n},$$

позволяющую получать средний абсолютный прирост без предварительного вычисления цепных абсолютных приростов.

Для нашего примера получим среднегодовой абсолютный прирост за 1982-1988 годы:

$$\overline{\Delta y} = \frac{y_6 - y_0}{6} = \frac{1,26 - 1,02}{6} = 0,040; \quad \overline{\Delta z} = \frac{z_6 - z_0}{6} = \frac{1,72 - 1,20}{6} = 0,087.$$

Исходя из того, что темп роста показывает, во сколько раз изменяется уровень ряда, *средний темп роста* находят как среднюю геометрическую цепных темпов роста:

$$\bar{T} = \sqrt[n]{T_{ц1} \cdot T_{ц2} \cdot \dots \cdot T_{цn}},$$

или  $\bar{T} = \sqrt[n]{\frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \cdot \dots \cdot \frac{y_{n-1}}{y_n}}$ , откуда

$$\bar{T} = \sqrt[n]{\frac{y_n}{y_0}} \text{ или в процентах } \bar{T}\% = \sqrt[n]{\frac{y_n}{y_0}} \cdot 100\%.$$

Средний темп прироста по определению считается равным

$$\overline{\Delta T} = \bar{T} - 1 \text{ или в процентах } \overline{\Delta T}\% = \bar{T}\% - 100\%.$$

За рассматриваемый период получим  $\bar{T}_y = \sqrt[6]{\frac{y_6}{y_0}} = \sqrt[6]{\frac{1,26}{1,02}} = 1,036$ ;

$\bar{T}_z = \sqrt[6]{\frac{z_6}{z_0}} = \sqrt[6]{\frac{1,72}{1,20}} = 1,062$  или в процентах  $\bar{T}_y = 103,6\%$ ,  $\bar{T}_z = 106,2\%$ .

Отсюда  $\overline{\Delta T}_y = 1,036 - 1 = 0,036$  или 3,6%, а  $\overline{\Delta T}_z = 1,062 - 1 = 0,062$  или в процентах 6,2%.

И в заключение еще раз заметим, что рассмотренные средние величины являются корректными обобщающими показателями в случае сохранения общей тенденции на протяжении всего периода времени. В противном случае можно прийти к казусным выводам. Если, к примеру, начальный и конечный уровни ряда динамики равны или практически равны между собой, то средний абсолютный прирост будет равен нулю, а средний темп роста – единице. То есть получается, что в рассматриваемый период в среднем не было никаких изменений. Например, такая ситуация сложилась бы, если в рассмотренном выше примере в силу некоторых случайных обстоятельств оказалось бы  $z_6 = 1,20$ . Тогда  $\bar{\Delta} = 0$  и  $\bar{T} = 1$ , что совсем не отражает положительного развития в предыдущие годы. Поэтому следует исключить из рассмотрения последний уровень и рассматривать средние показатели за предшествующий период, а затем провести подробный анализ причин падения уровня за последний год.

#### 8. 4. Типы тенденции динамики

Рассмотренные выше показатели и их средние значения позволяют проанализировать динамику ряда, исходя из того, что уже произошло. Следующей задачей является планирование на будущее с учетом ожидаемых значений уровней динамического ряда. Поскольку на признак влияют многие случайные факторы, вызывающие отклонения в сторону возрастания или убывания, то на основании уже полученных данных чаще всего можно говорить лишь об общей тенденции развития явления. В связи с этим возникает задача выявления основной тенденции развития ряда, выраженной в количественном виде.

*Выравнивание ряда – это выявление основной тенденции динамики в виде функции времени  $\hat{y} = f(t)$ , называемой трендом, ( $\hat{y}$  – значение уровня ряда на момент времени  $t$ ). Тренд позволяет учесть влияние всех основных факторов, влияющих на значения уровней ряда.*

Эта задача решается в два этапа. Первый – установление вида тенденции, второй – математическое моделирование тенденции, то есть расчет параметров тренда на основании данных ряда динамики.

Тренд	Название функции
1. $\hat{y} = a_0 + a_1 t$	Линейная
2. $\hat{y} = a_0 + a_1 t + a_2 t^2$	Парабола второго порядка
3. $\hat{y} = a_0 + a_1 t + a_2 t^2 + a_3 t^3$	Кубическая парабола
4. $\hat{y} = a_0 \cdot a_1^t$	Показательная
5. $\hat{y} = a_0 \cdot e^{a_1 t}$	Экспоненциальная
6. $\hat{y} = a_0 + a_1 \cdot a_2^t$	Модифицированная экспонента
7. $\hat{y} = a_0 \cdot a_1^{a_2^t}$	Кривая Гомперца
8. $\hat{y} = a_0 \cdot a_1^t \cdot a_2^{t^2}$	Логарифмическая парабола
9. $\hat{y} = a_0 + \frac{a_1}{t}$	Гиперболическая
10. $\hat{y} = \frac{a_0}{1 + a_1 \cdot e^{-a_2 t}}$	Логистическая кривая

Это наиболее часто употребляемые виды трендов.

Проблема выбора тренда начинает решаться специалистами в рассматриваемой области, опираясь на накопленный опыт. Рассмотрим некоторые применения трендов, оправдавшие себя на практике.

*Линейный тренд* хорошо отражает тенденцию в условиях действия большого числа случайных факторов, имеющих различные распределения и частично взаимно погашающих друг друга. Сюда можно отнести динамику урожайности области, региона и страны в целом.

*Параболический тренд* хорошо описывает ускорение или замедление уровней ряда с постоянным ускорением. Такая динамика наблюдается при действии факторов прогрессивного развития (прогрессирующее поступление высокотехнологичного оборудования, увеличение среднесуточного прироста живого веса поросят с возрастом). Такая же динамика наблюдается, когда снимаются преграды, ограничивающие развитие. Например, снятие ограничений в распределении дохода, в уровне оплаты труда, при повышении цены на дефицитную продукцию.

Параболический тренд с отрицательным ускорением ( $a_2 < 0$ ) может отражать не только приостановку роста уровня, но и его снижение с увеличивающейся скоростью. Такое характерно для производства устаревшей продукции или для ликвидируемого производства.

*Показательный тренд* с постоянным темпом роста  $a_1 > 1$  характеризует ускоряющееся возрастание уровней. Так, например, размножаются организмы при отсутствии ограничений со стороны среды – кормов, пространства обитания, хищников, болезней и т. д. По этому закону росло население Земли в период «демографического взрыва» в XX столетии. Сейчас этот период заканчивается, и темп роста населения стал уменьшаться. Если бы этого не случилось, то к 2500 г. на каждого человека приходилось бы около  $1 м^2$  суши. Понятно, что развитие по показательному тренду может продолжаться ограниченный период.

При  $a_1 < 1$  этот тренд описывает замедляющееся снижение уровней ряда динамики. Такова, к примеру, динамика трудоемкости продукции, удельных затрат, топлива, металла при технологическом прогрессе.

*Логарифмический тренд* характерен для тенденции замедляющегося изменения уровней ряда при отсутствии предельно возможного значения. При достаточно больших значениях  $t$  логарифмическая кривая становится по форме близкой к прямой, параллельной оси времени. Примерами являются рост спортивных достижений (высокие достижения трудно улучшить), рост производительности оборудования по мере его освоения.

*Гиперболический тренд* выражает замедляющееся изменение (убывание при  $a_1 > 0$ , возрастание при  $a_1 < 0$ ) уровня, имеющего предельное значение  $a_0$ . То есть речь идет о тенденциях, имеющих предельное значение – предельный КПД двигателя, 100%-ная грамотность и т. д.

*Логистический тренд* графически имеет форму латинской буквы S, положенной на бок и потому иногда называется *эсобразным* трендом. Он отражает, например, процесс насыщения потребителей новым товаром – сначала медленный, но ускоряющийся рост, затем рост равномерный, и, наконец, замедление роста до предельного насыщения.

## 8. 5. Аналитические методы выбора тренда

Прежде всего, следует ответить на три главных вопроса:

1. Были ли условия развития объекта достаточно однородными в изучаемый период?
2. Каков характер действия основных факторов развития?
3. Не было ли качественных, существенных изменений условий развития объекта внутри изучаемого периода времени?

Если, например, с некоторого момента были введены новые технологии, поточные линии, новые цеха, то единой тенденции за весь период не будет. В такой ситуации требуется дробление ряда на отдельные периоды: до введения этих изменений и после этого.

К простейшим методам обнаружения тенденции развития являются методы *усреднения* уровней ряда. При большом количестве уровней ряд динамики разбивается на группы одинакового объема и для каждой из них вычисляется среднее значение уровней, из которых образуется новый ряд. Так как при вычислении средних уровней действие случайных факторов взаимопогашается, то легче обнаруживается действие основных факторов, то есть тенденция.

Вторым является *метод скользящей средней*, особенно эффективный при небольшом числе уровней. В этом методе формируются укрупненные интервалы, и для каждого из них определяется средний уровень, который относят к середине укрупненного интервала. Если, например, укрупнение производится по три уровня, то первая средняя (арифметическая) определяется для уровней  $y_0, y_1, y_2$ :

$$\bar{y}_1 = \frac{y_0 + y_1 + y_2}{3};$$

вторая – для уровней  $y_1, y_2, y_3$ :

$$\bar{y}_2 = \frac{y_1 + y_2 + y_3}{3}$$

и так далее до

$$\bar{y}_{n-1} = \frac{y_{n-2} + y_{n-1} + y_n}{3}.$$

Аналогично проводится укрупнение по пяти уровням:

$$\bar{y}_2 = \frac{y_0 + y_1 + y_2 + y_3 + y_4}{5}, \quad \bar{y}_3 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5} \text{ и т. д.}$$

Рассмотрим этот метод на примере. Исходные и расчетные данные смотрите в *таблице 7.4*.

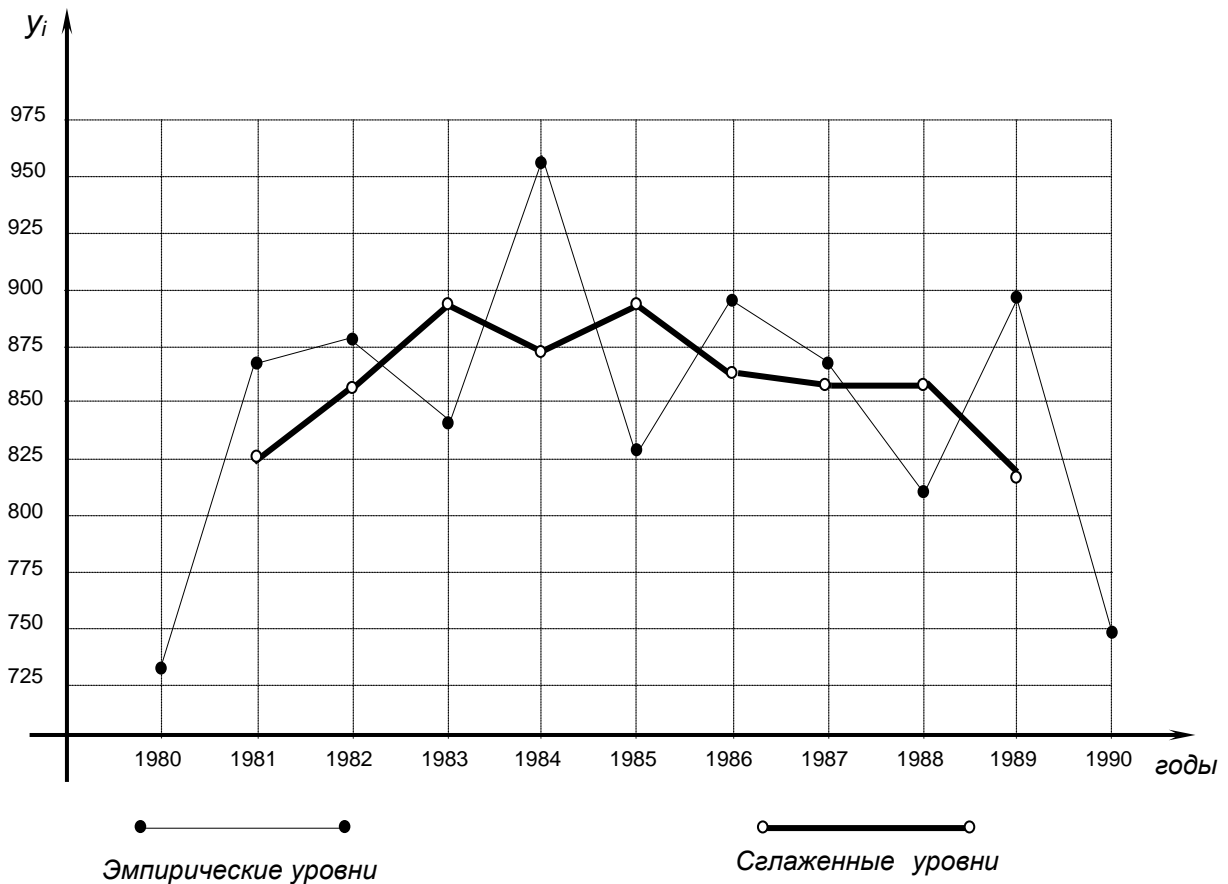
Для наглядности изобразим в системе координат совмещенные графики эмпирических и сглаженных уровней. Для этого по горизонтальной оси откладываются временные показатели (в нашем случае годы), а по вертикальной – соответствующие уровни. Затем полученные точки последовательно соединяются отрезками прямых линий. Для трехлетнего выравнивания получим график, на котором для сравнения построен и график эмпирических уровней (смотрите на следующей странице).

Таблица 8.4.

$i$	Год	Валовой сбор овощей тыс. т, $y_i$	Скользящие суммы		Скользящие средние	
			за 3 года	за 5 лет	за 3 года	за 5 лет
0	1980	733				
1	1981	868	2477		825,7	
2	1982	876	2583	4272	861,0	854,4
3	1983	839	2671	4367	890,3	873,4
4	1984	956	2623	4392	874,3	878,4
5	1985	828	2677	4384	892,3	876,8
6	1986	893	2589	4355	863,0	871,0
7	1987	868	2571	4293	857,0	858,6
8	1988	810	2572	4214	857,3	842,8
9	1989	894	2453		817,7	
10	1990	749				

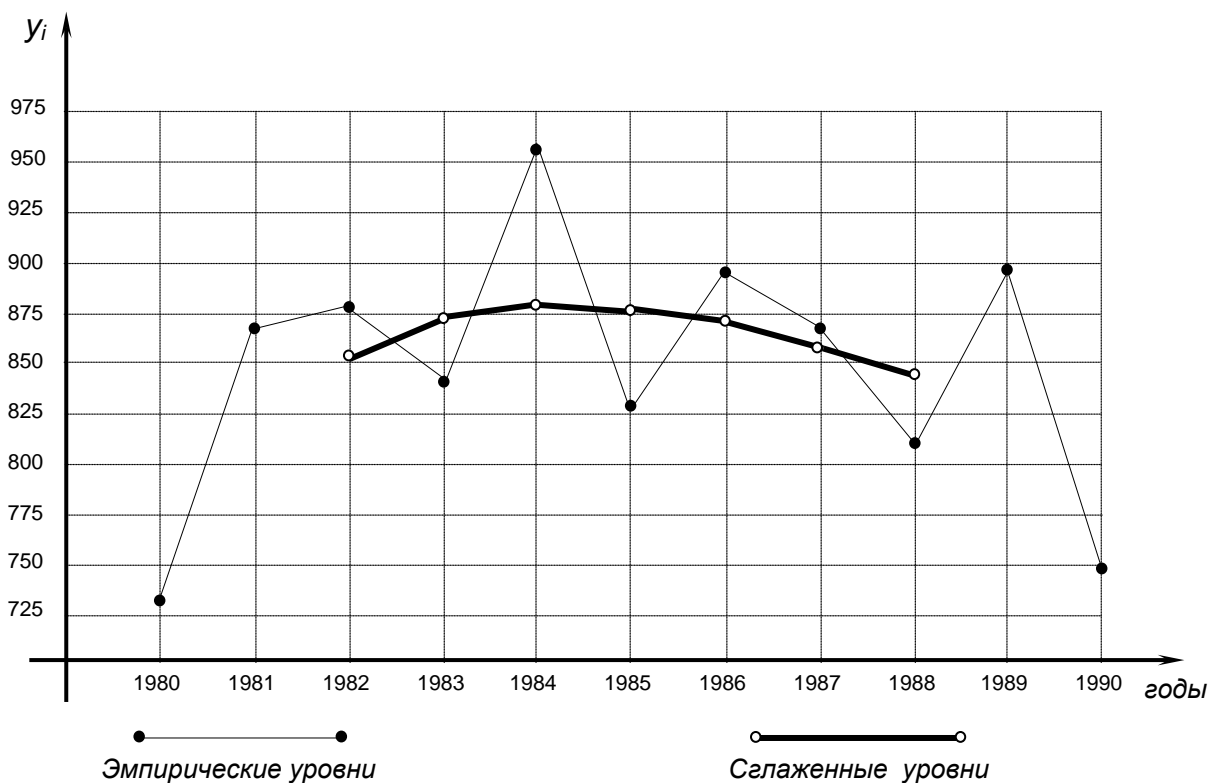


Графики эмпирических и сглаженных уровней по 3-летнему интервалу



Аналогично строим график для пятилетнего выравнивания.

Графики эмпирических и сглаженных уровней по 5-летнему интервалу



Сравнивая полученные графики, видим, что сглаживание по трех-летнему интервалу показывает меньшие колебания уровней, а по пяти-летнему интервалу явно выявляет тенденцию параболического типа.

Другой подход связан с изучением свойств теоретических уровней и последующей проверкой их для эмпирических уровней. Например, если тренд линейный или параболический, то рассматривают *конечные разности* первого, второго и т. д. порядков:  $\Delta_i^1 = y_i - y_{i-1}$ ,  $\Delta_i^2 = \Delta_i^1 - \Delta_{i-1}^1$ ,  $\Delta_i^3 = \Delta_i^2 - \Delta_{i-1}^2$  и т. д.

Если, к примеру, ряд динамики имеет линейную тенденцию с трендом  $\hat{y} = a_0 + a_1 t$ , то, введя условные временные показатели, получим:

$$\begin{aligned} t=0 & \quad y_0 = a_0 \\ t=1 & \quad y_1 = a_0 + a_1 \quad \Delta_1^1 = y_1 - y_0 = a_1 \\ t=2 & \quad y_2 = a_0 + 2a_1 \quad \Delta_2^1 = y_2 - y_1 = a_1 \quad \Delta_2^2 = \Delta_2^1 - \Delta_1^1 = 0 \\ t=3 & \quad y_3 = a_0 + 3a_1 \quad \Delta_3^1 = y_3 - y_2 = a_1 \quad \Delta_3^2 = \Delta_3^1 - \Delta_2^1 = 0 \\ & \quad \text{и так далее.} \end{aligned}$$

Как видим, в этом случае разности первого порядка получаются постоянными, а разности второго порядка – равными нулю. Исходя из этого свойства, для эмпирических уровней вычисляют конечные разности и, если окажется, что разности первого порядка *практически* постоянны, то считают, что в ряде динамики наблюдается линейная тенденция.

Аналогично для параболической тенденции второго порядка  $\hat{y} = a_0 + a_1 t + a_2 t^2$  получим:

$$\begin{aligned} t=0 & \quad y_0 = a_0 \\ t=1 & \quad y_1 = a_0 + a_1 \quad \Delta_1^1 = y_1 - y_0 = a_1 \\ t=2 & \quad y_2 = a_0 + 2a_1 \quad \Delta_2^1 = y_2 - y_1 = a_1 \quad \Delta_2^2 = \Delta_2^1 - \Delta_1^1 = 0 \\ t=3 & \quad y_3 = a_0 + 3a_1 \quad \Delta_3^1 = y_3 - y_2 = a_1 \quad \Delta_3^2 = \Delta_3^1 - \Delta_2^1 = 0 \\ & \quad \text{и так далее.} \end{aligned}$$

Отсюда, если для изучаемого ряда разности второго порядка *практически* постоянны, то можно говорить о квадратичной тенденции для этого ряда. Если же *практически* постоянными оказываются разности третьего порядка, то можно принимать кубическую тенденцию. Для показательного тренда  $\hat{y} = a_0 \cdot a_1^t$  со значениями  $y_0 = a_0$ ,  $y_1 = a_0 \cdot a_1$ ,

$y_2 = a_0 \cdot a_1^2$ ,  $y_3 = a_0 \cdot a_1^3$ ,  $y_4 = a_0 \cdot a_1^4$  и т.д. постоянными оказываются цепные темпы роста:  $T_{ci} = a_1$ . Отсюда, если для наблюдаемого ряда практически постоянны цепные темпы роста, то для выравнивания можно использовать показательный тренд.

Отметим еще некоторые признаки. Если первые конечные разности уменьшаются практически с постоянным темпом, то можно говорить о модифицированной экспоненте. Если средние уровни, нанесенные на полулогарифмическую сетку, близки к прямой линии, то можно принять к руководству простую экспоненту. Если первые разности обратных значений средних уровней изменяются на один и тот же процент, то следует остановиться на логистическом тренде.

В заключение отметим, что рассмотренные признаки «работают» при обязательном условии равенства временны  
седними уровнями ряда динамики.

□ x интер

## 8. 6. Расчет параметров тренда

Параметры тренда рассчитываются *методом наименьших квадратов* аналогично тому, как это делалось в корреляции. Только роль признака-фактора здесь играет временной параметр  $t$  – порядковый номер интервала или момента времени. Так, для линейного тренда  $\hat{y} = a_0 + a_1 t$  составляется система нормальных уравнений:

$$\begin{cases} a_1 \sum_1^n t_i^2 + a_0 \sum_1^n t_i = \sum_1^n y_i t_i, \\ a_1 \sum_1^n t_i + n a_0 = \sum_1^n y_i, \end{cases}$$

а для квадратичного тренда  $\hat{y} = a_0 + a_1 t + a_2 t^2$  система:

$$\begin{cases} a_2 \sum_1^n t_i^4 + a_1 \sum_1^n t_i^3 + a_0 \sum_1^n t_i^2 = \sum_1^n y_i t_i^2, \\ a_2 \sum_1^n t_i^3 + a_1 \sum_1^n t_i^2 + a_0 \sum_1^n t_i = \sum_1^n y_i t_i, \\ a_2 \sum_1^n t_i^2 + a_1 \sum_1^n t_i + n a_0 = \sum_1^n y_i. \end{cases}$$

Решение этих систем можно существенно упростить, удачным образом введя условные показатели времени. Если число уровней ряда нечетное,

то за начало отсчета принимается временной показатель уровня, находящегося в середине ряда. Ему придается значение 0, а датам, идущим вперед по времени, последовательно +1, +2, +3 и т. д. и идущим назад последовательно -1, -2, -3 и т. д.

Если число уровней четное, то за начало отсчета принимается середина между двумя средними показателями времени и временные показатели нумеруются ..., -5, -3, -1, +1, +3, +5, ... (-1 и +1 соответствуют двум средним показателям).

В обоих случаях получаются ряды показателей времени, симметричные относительно выбранных начал, и для них выполняются равенства  $\sum_1^n t_i = 0$  и  $\sum_1^n t_i^3 = 0$ . Это существенно упрощает системы нормальных уравнений. Действительно, и в линейном случае получим:

$$\left\{ \begin{array}{l} a_1 \sum_1^n t_i^2 = \sum_1^n y_i t_i \\ na_0 = \sum_1^n y_i \end{array} \right. , \text{ откуда получаем } a_1 = \frac{\sum_1^n y_i t_i}{\sum_1^n t_i^2} \text{ и } a_0 = \frac{\sum_1^n y_i}{n} = \bar{y}.$$

А в квадратичном случае система примет вид:

$$\left\{ \begin{array}{l} a_2 \sum_1^n t_i^4 + a_0 \sum_1^n t_i^2 = \sum_1^n y_i t_i^2 , \\ a_1 \sum_1^n t_i^2 = \sum_1^n y_i t_i , \\ a_2 \sum_1^n t_i^2 + na_0 = \sum_1^n y_i . \end{array} \right.$$

Из второго уравнения сразу определяется  $a_1 = \frac{\sum_1^n y_i t_i}{\sum_1^n t_i^2}$  и остается система только двух уравнений, из которой нетрудно получить решение

$$a_0 = \frac{\sum_1^n y_i \cdot \sum_1^n t_i^4 - \sum_1^n y_i t_i^2 \cdot \sum_1^n t_i^2}{n \cdot \sum_1^n t_i^4 - \left( \sum_1^n t_i^2 \right)^2} , \quad a_2 = \frac{n \cdot \sum_1^n y_i t_i^2 - \sum_1^n y_i \cdot \sum_1^n t_i^2}{n \cdot \sum_1^n t_i^4 - \left( \sum_1^n t_i^2 \right)^2} .$$

В качестве примера рассмотрим аналитическое выравнивание ряда динамики валового сбора овощей (таблица 8.4). Предварительные расчеты представим в таблице 8.5.

Таблица 8.5.

Годы	Валовой сбор овощей тыс. т, $y_i$	Условное время $t_i$	$t_i^2$	$t_i^4$	$y_i t_i$	$y_i t_i^2$	Выровненные уровни ряда динамики, тыс. т лин. тренд	Выровненные уровни ряда динамики, тыс. т квадр. тренд
1	2	3	4	5	6	7	8	9
1980	733	-5	25	625	-3655	18325	847,58	776,84
1981	868	-4	16	256	-3472	13888	847,41	819,05
1982	876	-3	9	81	-2628	7884	847,24	851,92
1983	839	-2	4	16	-1678	3356	847,07	875,35
1984	956	-1	1	1	-956	956	846,90	889,34
1985	828	0	0	0	0	0	846,73	893,89
1986	893	1	1	1	893	893	846,56	889,00
1987	868	2	4	16	1736	3472	846,39	874,67
1988	810	3	9	81	2430	7290	846,22	850,90
1989	894	4	16	256	3576	14304	846,05	817,69
1990	749	5	25	625	3745	18725	845,88	775,04
Всего	9314	0	110	1958	-19	89093	9314,03	9313,69

По итоговым данным граф 2, 4 и 6 найдем коэффициенты линейного тренда:

$$a_1 = \frac{-19}{110} = -0,17, \quad a_0 = \frac{9314}{11} = 846,73.$$

Получаем уравнение линейного тренда

$$\hat{y}_t = -0,17t + 846,73,$$

выражающего динамику валового сбора овощей в период с 1980 по 1990 годы. По этому уравнению вычисляем выровненные уровни ряда (графа 8):

для 1980 г.  $\hat{y}_{-5} = -0,17 \cdot (-5) + 846,73 = 847,58;$

для 1981 г.  $\hat{y}_{-4} = -0,17 \cdot (-4) + 846,73 = 847,41;$

.....

для 1990 г.  $\hat{y}_5 = -0,17 \cdot 5 + 846,73 = 845,88.$

Обратим внимание на то, что сумма выровненных уровней практически совпадает с суммой эмпирических уровней (итоги граф 2 и 8). Небольшое отклонение получилось за счет округлений.

Теперь по итоговым данным граф 2, 4 – 7 составим систему нормальных уравнений для определения коэффициентов квадратичного тренда:

$$\begin{cases} 1958 a_2 + 110 a_0 = 89093, \\ 110 a_1 = -19, \\ 110 a_2 + 11 a_0 = 9314. \end{cases}$$

Её решением являются  $a_0 = 893,89$ ,  $a_1 = -0,17$ ,  $a_2 = -4,72$ . Отсюда получаем квадратичный тренд:

$$\hat{y}_t = 893,89 - 0,17 \cdot t - 4,72 \cdot t^2.$$

Выровненные по нему уровни вносим в графу 9.

Еще один метод выбора тренда заключается в том, что по исходным данным рассчитывают параметры нескольких трендов и по некоторому признаку выбирают лучший. В качестве такого признака можно рассмотреть наименьшее среднеквадратическое отклонение эмпирических и теоретических уровней ряда с учетом числа параметров тренда:

$$S_{\hat{y}} = \sqrt{\frac{\sum (y_i - \hat{y}_{t_i})^2}{n - m}},$$

где  $m$  – количество параметров в уравнении тренда ( для линейного тренда  $m = 2$ , а для квадратичного  $m = 3$ ).

В нашем случае, используя данные граф 2, 8 и 9, для линейного случая получим

$$S_{\hat{y}_{\text{лин}}} = \sqrt{\frac{42378,9009}{11 - 2}} = 68,62,$$

а для квадратичного случая получим

$$S_{\hat{y}_{\text{квдр}}} = \sqrt{\frac{23238,8613}{11 - 3}} = 53,90.$$

Таким образом, предпочтение следует отдать квадратичному тренду, что вполне согласуется с выводом, сделанным в предыдущем пункте при расчете скользящих средних.

Величина относительной ошибки квадратичного тренда составит

$$6,26\% \left( \frac{S_{\hat{y}}^{у\text{квадр}}}{\bar{y}} \cdot 100\% = \frac{53,90}{846,73} \cdot 100\% = 6,26\% \right).$$

Одной из важных задач, решаемых с помощью тренда, является *экстраполирование* – прогноз на будущее. В нашем случае можно предсказать валовой сбор овощей в 1991 году, положив  $t = 6$ :

$$\hat{y}_{t=6} = 893,89 - 0,17 \cdot 6 - 4,72 \cdot 36 = 722,95,$$

то есть при сохранении тенденции следует ожидать валового сбора овощей в  $722,95$  тыс. т.

Однако более уверенным является интервальный прогноз при заданном уровне значимости. Например,

- при уровне значимости 5% и числе степеней свободы  $n - m = 8 - 3 = 5$  табличное значение критерия Стьюдента  $t_{\alpha} = 1,860$ ;

- предельная ошибка равна  $t_{\alpha} \cdot \frac{S_{\hat{y}}}{\sqrt{n}} = 1,860 \cdot \frac{53,90}{\sqrt{11}} = 30,2279$ ;

- доверительный интервал  $722,95 - 30,23 \leq y_{\text{прогноз}} \leq 722,95 + 30,23$ .

$$692,72 \leq y_{\text{прогноз}} \leq 753,18.$$

Это означает, что с уверенностью в 95% можно ожидать валовой сбор овощей не менее, чем  $692,72$  тыс. т, но не более  $753,18$  тыс. т.

В заключение отметим, что после несложных преобразований можно получить формулы для  $S_{\hat{y}}$ , дающие меньшую погрешность вычислений:

$$S_{\hat{y}} = \sqrt{\frac{\sum y_i^2 - a_0 \sum y_i - a_1 \sum y_i t_i}{n - 2}} \quad \text{для линейного тренда,}$$

$$S_{\hat{y}} = \sqrt{\frac{\sum y_i^2 - a_0 \sum y_i - a_1 \sum y_i t_i - a_2 \sum y_i t_i^2}{n - 3}} \quad \text{для квадратичного тренда.}$$

## Г Л А В А 9

### ИНДЕКСЫ

#### 9. 1. Понятие индекса и задачи, решаемые с их помощью

Слово *индекс* (от латинского *index* – указатель) в статистике употребляется как показатель изменений, как в простых, так и в сложных экономических процессах.

*Индекс – показатель сравнения двух состояний одного и того же явления, состоящего из соизмеримых или несоизмеримых элементов.*

С такими показателями мы уже встречались, когда рассматривали ряды динамики. Это темпы роста, прироста, абсолютный прирост, различные показатели средних изменений за весь период. Однако, понятие индекса, отражающее свойства упомянутых показателей, во многом отличается от них. В чем же эти отличия?

*Во-первых*, индекс – это *относительный* показатель, и в этом смысле темпы роста и темпы прироста можно было бы называть индексами роста и прироста.

*Во-вторых*, индексы позволяют сравнивать явления, состоящие из разнородных элементов. К примеру, как охарактеризовать общее изменение количества потребляемых населением продуктов, если из-за их разнообразия общее количество нельзя получить простым сложением (три килограмма картофеля плюс один килограмм мяса равно четыре..., но чего?). Таким образом, возникает необходимость соизмерения разнородных элементов. Здесь уместно вспомнить «Капитал» выдающегося экономиста Карла Маркса, который писал, что «различные виды становятся *количественно сравнимыми* лишь после того, как они сведены к одному и тому же единству. Только как выражения одного и того же единства они являются *одноименными*, а следовательно, соизмеримыми величинами». В качестве такой меры соизмерения разнородных продуктов в теории индексов используют цену, себестоимость или трудоемкость единицы продукции.

*В-третьих*, с помощью индексов можно характеризовать не только изменения во времени, но и так называемые территориальные изменения. Позаимствуем пример из [ 1 ]. Удой молока на одну корову в хозяйствах России в 1990 г. составил 2781 кг, а в 1989 – 2773 кг. Индекс равен  $2781 : 2773 = 1,0029$  или в процентах 100,29%. Увеличение на 0,29% или на 8 кг. на одну корову внушает хоть и умеренный, но оптимизм. Изменим направление сравнения, и посмотрим, что было в этот же период в странах Запада. В 1990 г. в Великобритании надой молока на одну корову составил 5213 кг, в Польше – 3234 кг, в Швеции – 6213 кг, а соответствующие индексы показали увеличение на 53,3%, 86%, 44,8%. Вспомните 0,29% в России, и оптимизм «испарится».



*В-четвертых*, индексы позволяют проводить сравнение с плановыми показателями. Например, на предприятиях некоторой отрасли, имеющих различные производственные мощности, или выпускающих различную продукцию, можно вычислить индексы выполнения плана (отношение объема выполненной работы к плановому объему) и по ним провести анализ успешности работы предприятий.

*В-пятых*, индексы позволяют произвести анализ изменений, то есть выяснить роль отдельных факторов. Например, определить, какую роль в изменении доходов от продаж сыграли изменения цен, и какую роль – изменения объемов продаж.

Пожалуй, это основные особенности и задачи, решаемые с помощью индексов. Другие свойства индексов будут рассматриваться в процессе изучения их.

## 9. 2. Индивидуальные индексы. Увязка в систему

*Индивидуальными* называют *индексы*, характеризующие изменение только *одного элемента* статистической совокупности. Их обозначают строчной латинской буквой  $i$ . При этом, как правило, для одного и того же элемента рассматриваются несколько признаков, каждому из которых в общей статистике присвоено свое символическое обозначение. Так, количество единиц рассматриваемого вида обозначается  $q$ , цена единицы –  $p$ , себестоимость –  $z$ , трудоемкость –  $t$  и т.д. Общая стоимость, общая себестоимость или общая трудоемкость, вычисляемые соответственно как произведение цены, себестоимости или трудоемкости одной единицы на количество единиц, обозначаются одной и той же буквой  $w$ . Если одновременно рассматриваются индивидуальные индексы для нескольких единиц совокупности, то признаки пишутся с подстрочным значком  $j$ :  $q_j, p_j, z_j, t_j, w_j, j = 1 \div n$ .

При построении индексов используются два вида данных: оцениваемые данные, которые называют *отчетными* и отмечают надстрочным значком «1»:  $q_j^1, p_j^1, z_j^1, t_j^1, w_j^1, j = 1 \div n$ , и *базисные* данные, которые используют в качестве базы сравнения и отмечают надстрочным знаком «0»:  $q_j^0, p_j^0, z_j^0, t_j^0, w_j^0, j = 1 \div n$ . Периоды соответственно называют отчетным и базисным.

*Индивидуальный индекс* вычисляется как *отношение отчетного значения признака к его базисному значению*. Для упомянутых выше признаков будем иметь:

$$i_q = \frac{q^1}{q^0}, \quad i_p = \frac{p^1}{p^0}, \quad i_z = \frac{z^1}{z^0}, \quad i_t = \frac{t^1}{t^0}, \quad i_w = \frac{w^1}{w^0}.$$

Например, в 1980 году на рынках страны было продано 387 тыс. кг сметаны по цене 19,7 ден. ед. за килограмм, а в 1990 – 241 тыс. кг по 24,8 ден. ед. Найти индивидуальные индексы цены, объема продаж и общей стоимости продаж сметаны.

В этом случае имеем базисные данные  $p^0 = 19,7$ ,  $q^0 = 387$ , а отчетные  $p^1 = 24,8$ ,  $q^1 = 241$ . Тогда получим индекс цены  $i_p = \frac{24,8}{19,7} = 1,26$

или 126% и индекс количества продаж  $i_q = \frac{241}{387} = 0,62$  или 62%. Для

отыскания индекса общей стоимости найдем ее в базисном году  $w^0 = p^0 q^0 = 19,7 \cdot 387 = 7623,9$  ден. ед. и в отчетном году  $w^1 = p^1 q^1 = 24,8 \cdot 241 = 5976,8$  ден. ед. Тогда получим индекс стоимости  $i_w = \frac{5976,8}{7623,9} = 0,78$  или 78%. Отсюда можно сделать вывод, что в 1990

году по сравнению с 1980 годом цена на сметану выросла 26% , а количество проданной сметаны уменьшилось на 38%. В результате общая стоимость продаж уменьшилась на 22%.

В рассмотренном примере три признака (количество единиц, цена единицы и общая стоимость) объединены равенством

$$w = p q.$$

В этом случае говорят, что признаки *увязаны в систему*.

Нетрудно увидеть, что

$$i_w = \frac{w^1}{w^0} = \frac{p^1 q^1}{p^0 q^0} = \frac{p^1}{p^0} \cdot \frac{q^1}{q^0} = i_p \cdot i_q,$$

$$i_w = i_p \cdot i_q,$$

то есть индивидуальные индексы «повторяют» связь между признаками. Это свойство называют *увязкой индексов в систему*.

В частности, это позволяет при неполных данных отыскивать индексы других признаков с меньшими усилиями. В задаче, рассмотренной выше, не обязательно было отыскивать общие стоимости в отчетном и базисном периодах. Для нахождения индекса общей стоимости достаточно было воспользоваться увязкой индексов в систему  $i_w = 1,26 \cdot 0,62 = 0,78$ .

Кроме того, это свойство позволяет контролировать правильность вычислений.

И, наконец, отметим некоторые другие примеры увязки индексов в систему:  $i_w = i_z \cdot i_q$  и  $i_w = i_t \cdot i_q$ , где  $i_w$  – индивидуальные индексы

общих затрат на производство в денежных единицах и общих трудовых затрат соответственно.

Кстати заметим, что обозначения признаков неслучайны. Они соответствуют начальным буквам английских слов price (цена), quantity (количество) и т. д. В статистике это закреплённые обозначения!

### 9.3. Агрегатные индексы

*Агрегатными (сводными или общими)* называют *индексы*, характеризующие изменения в целой статистической совокупности, и обозначают прописной латинской буквой *I*.

Пусть совокупность содержит  $n$  единиц, для каждой из которых имеются  $p_j^0, q_j^0, w_j^0$  – базисные данные и  $p_j^1, q_j^1, w_j^1$  – отчетные данные, где  $j$  – номер единицы совокупности,  $j = 1 \div n$ . Если общие стоимости выражены в одних и тех же денежных единицах, то *агрегатный индекс общей стоимости* имеет вид:

$$I_w = \frac{\sum_{j=1}^n w_j^1}{\sum_{j=1}^n w_j^0} = \frac{\sum_{j=1}^n p_j^1 q_j^1}{\sum_{j=1}^n p_j^0 q_j^0}.$$

В дальнейшем для простоты записи мы опускаем индекс суммирования, и будем писать

$$I_w = \frac{\sum p^1 q^1}{\sum p^0 q^0},$$

подразумевая, что суммирование осуществляется по всем единицам совокупности. Так как рассмотренный индекс отражает влияние изменения и цен и количеств, то можно употреблять еще одно обозначение индекса общей стоимости  $I_{pq}$ .

Проблемы возникают, когда речь заходит об агрегатных индексах цен. Причина заключается в их несоизмеримости. В самом деле, принято говорить, что цена единицы продукта выражается в некоторых денежных единицах, забывая при этом, что цена отнесена к одной единице его. Никто не подумает писать, что скорость равна 5 км, а укажет еще единицу времени: 5 км/час. То же самое и с ценой: правильнее было бы писать 5 д.е/шт, 5 д.е/кг, 5 д.е/м и так далее. Из-за этих знаменателей и возникает несоизмеримость и принципиальная невозможность сложить цены в отчетном и в базисном периоде и рассматривать индекс

цен, например, в виде:  $\sum p^1 / \sum p^0$ . Даже, если речь идет об одинаковой продукции, но изготовленной на разных предприятиях (например, сметана), то различия в технологических условиях, в закупочных ценах, в количестве производимой продукции и прочее делают цены качественно различными, то есть несоизмеримыми.

Аналогичное положение и с количествами, которые нельзя складывать для различных видов продукции (лук, арбузы, огурцы и т.д.) или из-за различия их цен.

Таким образом, требуются некоторые *коэффициенты соизмерения (веса)*, которые *элиминировали* бы влияние количеств при вычислении агрегатных индексов цен, делая их при этом соизмеримыми. В качестве таких коэффициентов можно выбрать количества, одинаковые для обоих периодов:

$$I_p = \frac{\sum p_j^1 q_j}{\sum p_j^0 q_j}.$$

В 1864 г. английский экономист Э. Ласпейрес предложил в качестве коэффициентов  $q_j$  принять количества базисного периода и рассматривать агрегатный индекс цен в виде

$$I_p^{\Pi} = \frac{\sum p_j^1 q_j^0}{\sum p_j^0 q_j^0},$$

а для агрегатного индекса количеств коэффициентами соизмерения взять цены базового периода:

$$I_q^{\Pi} = \frac{\sum q_j^1 p_j^0}{\sum q_j^0 p_j^0}.$$

Рассмотренные индексы называют *индексами Ласпейреса*. Как видим, в  $I_p^{\Pi}$  от числителя к знаменателю меняются только цены, следовательно, этот индекс характеризует совокупное изменение цен. А в  $I_q^{\Pi}$  меняются только количества, то есть этот индекс отражает совокупное изменение количеств.

Очевидно, эти агрегатные индексы Ласпейреса не увязываются в систему с агрегатным индексом общей стоимости, то есть в общем случае  $I_w \neq I_p^{\Pi} \cdot I_q^{\Pi}$ .

В 1874 г. английский экономист Г. Пааше вводит в рассмотрение два агрегатных индекса цен и количеств, в которых коэффициенты соизмерения берутся из отчетного периода:

$$I_p^\Pi = \frac{\sum p_j^1 q_j^1}{\sum p_j^0 q_j^1} \quad \text{и} \quad I_q^\Pi = \frac{\sum q_j^1 p_j^1}{\sum q_j^0 p_j^1}.$$

И эта пара *индексов Пааше* не увязывается в систему с агрегатным индексом общей стоимости, то есть  $I_p^\Pi \cdot I_q^\Pi \neq I_w$ . Однако, если умножить индекс цен Ласпейреса на индекс количеств Пааше, то получим индекс общей стоимости. Действительно,

$$I_p^\Pi \cdot I_q^\Pi = \frac{\sum p_j^1 q_j^0}{\sum p_j^0 q_j^0} \cdot \frac{\sum q_j^1 p_j^1}{\sum q_j^0 p_j^1} = \frac{\sum q_j^1 p_j^1}{\sum p_j^0 q_j^0} = I_w.$$

Аналогично выполняется равенство  $I_p^\Pi \cdot I_q^\Pi = I_w$ . Таким образом, агрегатные индексы Ласпейреса и Пааше в совокупности обеспечивают увязку в систему.

При построении агрегатных индексов и Ласпейреса и Пааше используются две суммы  $\sum p_j^1 q_j^0$  и  $\sum p_j^0 q_j^1$ , не отражавшиеся реально в деньгах. Однако их следует рассматривать как гипотетические общие стоимости количеств. В первой формуле это общая стоимость количеств базового периода в ценах отчетного периода, то есть сколько бы стоило общее количество прошлого периода, если бы продажи велись по ценам отчетного периода. Во второй формуле оценивается общая стоимость количеств отчетного периода в ценах базового периода. Это еще раз подтверждает практическую корректность рассмотренных нами агрегатных индексов.

#### 9.4. Абсолютные изменения общей стоимости

Очевидным образом можно определить *абсолютное изменение общей стоимости*

$$\sum \Delta w(pq) = \sum p^1 q^1 - \sum p^0 q^0,$$

выраженное в денежных единицах и представляющее собой разность между числителем и знаменателем в индексе общей стоимости. Если эта величина окажется положительной, то это значит, общая стоимость в отчетном периоде увеличилась по сравнению с общей стоимостью в базовом периоде. А ее отрицательность означает уменьшение общей стоимости.

Аналогичным образом, исходя из индексов Ласпейреса, получают *абсолютное изменение общей стоимости за счет изменения цен*

$$\sum^{\Pi} \Delta w(p) = \sum p^1 q^0 - \sum p^0 q^0,$$

отражающие влияние изменения цен на изменение общей стоимости, и *абсолютное изменение общей стоимости за счет изменения количества проданного*

$$\sum^{\Pi} \Delta w(q) = \sum q^1 p^0 - \sum q^0 p^0,$$

выделяющее долю общей стоимости, зависящую от изменения количества продаж. Заметим сразу же, что сумма этих изменений в общем случае не равна абсолютному изменению общей стоимости, т.е.

$$\sum \Delta w(pq) \neq \sum^{\Pi} \Delta w(p) + \sum^{\Pi} \Delta w(q),$$

что является отражением неувязки в систему агрегатных индексов Ласпейреса с агрегатным индексом общей стоимости.

Таким же образом, исходя из агрегатных индексов Паше, получают аналогичные абсолютные изменения общей стоимости за счет изменения цен

$$\sum^{\Pi} \Delta w(p) = \sum p^1 q^1 - \sum p^0 q^1$$

и за счет изменения количества продаж

$$\sum^{\Pi} \Delta w(q) = \sum q^1 p^1 - \sum q^0 p^1,$$

для которых в общем случае также выполняется неравенство

$$\sum \Delta w(pq) \neq \sum^{\Pi} \Delta w(p) + \sum^{\Pi} \Delta w(q).$$

Однако если сложить абсолютное изменение общей стоимости за счет изменения цен по Ласпейресу и абсолютное изменение общей стоимости за счет изменения количества по Паше, то получим абсолютное изменение общей стоимости. Действительно,

$$\begin{aligned} \sum^{\Pi} \Delta w(p) + \sum^{\Pi} \Delta w(q) &= \sum p^1 q^0 - \sum p^0 q^0 + \sum q^1 p^1 - \sum q^0 p^1 = \\ &= \sum p^1 q^1 - \sum p^0 q^0 = \sum \Delta w(pq). \end{aligned}$$

Аналогично получается равенство

$$\sum^{\Pi} \Delta w(p) + \sum^{\Pi} \Delta w(q) = \sum \Delta w(pq).$$

При проведении анализа экономической деятельности естественно исходить из принципа: изменение общей стоимости должно складываться из изменений за счет отдельных факторов (в нашем случае за счет изменения цен и количеств). Поэтому применяются два дополнительных правила, позволяющие обеспечить это условие:

1) при расчете *индексов количественных показателей* соизмерители берутся на уровне базисного периода, то есть расчет ведется по формуле Ласпейреса;

2) при расчете индексов качественных показателей (цен, себестоимостей, трудоемкостей ...) используется формула Пааше.

Рассмотрим применение агрегатных индексов к анализу условного рынка в целом.

Таблица 9.1.

Товары	Объем продаж, тыс. ед.		Цена единицы, ден. ед.		$q^0 p^0$	$q^1 p^1$	$q^1 p^0$	$q^0 p^1$
	1960	1970	1960	1960				
	$q^0$	$q^1$	$p^0$	$p^1$	тысяч денежных единиц			
Тушки кур, кг	1107	444	24,5	39,2	27121,5	17404,8	10878,0	43394,4
Яйца, шт.	21650	97016	1,8	1,5	38970,0	145524,0	174628,8	32475,0
Сметана, кг	387	241	19,7	24,8	7623,9	5976,8	4747,7	9597,6
Всего					73715,4	168905,6	190254,5	85467,0

Сводный индекс общей стоимости продукции

$$I_{pq} = \frac{\sum p_j^1 q_j^1}{\sum p_j^0 q_j^0} = \frac{168905,6}{73715,4} = 2,29.$$

Сводный индекс количества проданного  $I_q$  вычисляется как отношение взвешенных сумм  $q_j^1$  к взвешенным суммам  $q_j^0$ . В качестве весов выбираются цены отчетного периода (индекс Г. Пааше) либо цены базисного периода (индекс Ласпейреса).

Для вычисления сводного индекса цен  $I_p$  в качестве весов берутся количества проданного ( $q^1$  или  $q^0$ ).

Вычислим агрегатные индексы Г. Пааше:

$$I_q^{\Pi} = \frac{\sum q_j^1 p_j^1}{\sum q_j^0 p_j^1} = \frac{168905,6}{85467,0} = 1,98; \quad I_p^{\Pi} = \frac{\sum p_j^1 q_j^1}{\sum p_j^0 q_j^1} = \frac{168905,6}{190254,5} = 0,89$$

и индексы Ласпейреса:

$$I_q^{\Pi} = \frac{\sum q_j^1 p_j^0}{\sum q_j^0 p_j^0} = \frac{190264,5}{73715,4} = 2,58; \quad I_p^{\Pi} = \frac{\sum p_j^1 q_j^0}{\sum p_j^0 q_j^0} = \frac{85467,0}{73715,4} = 1,16.$$

Проверка согласования агрегатных индексов. Имеем

$$I_p^{\Pi} \cdot I_q^{\Pi} = 0,89 \cdot 2,58 = 2,296 = I_{pq}; \quad I_q^{\Pi} \cdot I_p^{\Pi} = 1,98 \cdot 1,16 = 2,2968 = I_{pq}.$$

(Небольшая погрешность получилась за счет округлений при счете).

**Выводы.** Общая стоимость продаж по ассортименту товаров в 1970 году увеличилась на 129% ( $2,29 \cdot 100\% - 100\% = 129\%$ ) по сравнению с 1960 годом. Количество проданного увеличилось на 98% в ценах 1970 года и 158% в ценах 1960 года. Цены в среднем уменьшились на 11% относительно объемов продаж 1970 года, и увеличились на 16% относительно объемов продаж 1960 года.

**4.** Абсолютное изменение общей стоимости проданного по всему ассортименту:

$$\sum \Delta w(pq) = \sum p_j^1 q_j^1 - \sum p_j^0 q_j^0 = 168905,6 - 73715,4 = 95190,2 \text{ тыс. д. ед.}$$

Абсолютные изменения общей стоимости за счет изменения количества проданного  $\sum \Delta w(q)$  и за счет изменения цен  $\sum \Delta w(p)$  рассчитываются как разность между числителями и знаменателями соответствующих агрегатных индексов.

Исходя из индексов Г. Пааше, получим

$$\sum^{\Pi} \Delta w(q) = 168905,6 - 85467,0 = 83438,6 \text{ тыс. д. ед.},$$

$$\sum^{\Pi} \Delta w(p) = 168905,6 - 190254,5 = -21348,9 \text{ тыс. д. ед.},$$

Исходя из индексов Ласпейреса, получим

$$\sum^{\Pi} \Delta w(q) = 190254,5 - 73715,4 = 116539,1 \text{ тыс. д. ед.},$$

$$\sum^{\Pi} \Delta w(p) = 85467,0 - 73715,4 = 11751,6 \text{ тыс. д. ед.}$$

Нетрудно заметить, что

$$\sum \Delta w(pq) = \sum^{\Pi} \Delta w(p) + \sum^{\Pi} \Delta w(q) = \sum^{\Pi} \Delta w(p) + \sum^{\Pi} \Delta w(q),$$

то есть абсолютное изменение общей стоимости разлагается на абсолютное изменение за счет изменения цен и на абсолютное изменение за счет изменения количеств проданного, вычисленных с весами разных периодов (базисного и отчетного).



### Логично сделать выводы:

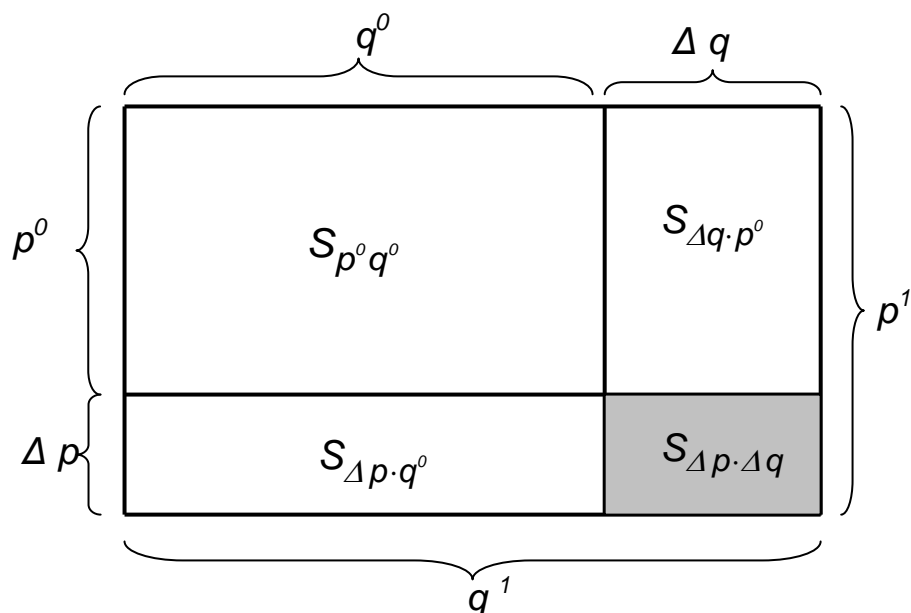
1. В 1970 году общая стоимость проданного увеличилась на  $95190.2$  тыс. д. ед., при этом за счет изменения количества проданного на  $83438.6$  тыс. д. ед. и на  $11751.6$  тыс. д. ед. за счет изменения цен, если исходить из цен отчетного периода и количеств проданного базисного периода. Если же исходить из количеств проданного отчетного периода и цен базисного, то за счет изменения количеств проданного произошло увеличение на  $116539.1$  тыс. д. ед., за счет изменения цен уменьшение на  $21348.9$  тыс. д. ед..

2. В итоге в 1970 году покупатели потратили на приобретение всего ассортимента на  $95190.2$  тыс. д. ед. больше, чем в 1960 году.

### 9.5. Знак Вазара. «Идеальная» формула Фишера

Рассмотрим теперь подробнее проблему увязки агрегатных индексов в систему. Как мы видели ранее, агрегатные индексы цен и количеств, в которых используются веса одного периода (отчетного или базового), не увязываются в систему с агрегатным индексом общей стоимости. Отсюда возникают две проблемы. Первая – попытаться объяснить, почему это происходит, если на общую стоимость влияют только цены и количества. Вторая – построить такие агрегатные индексы, которые удовлетворяли бы условию увязки в систему.

В первом случае наглядно это можно показать, например, с помощью плоскостной диаграммы – так называемого «знака Вазара» (по имени русского статистика В. Е. Вазара (1851-1940)). В этой схеме каждый



Знак Вазара

прямоугольник обозначен величиной его площади, каждая из которых соответствует некоторой стоимости. Так, площадь прямоугольника

$S_{p^0 q^0}$  равняется стоимости  $w^0 = p^0 q^0$ . Если при постоянной цене  $p^0$  изменится только количество на величину  $\Delta q$ , то стоимость изменится на величину, равную площади прямоугольника  $S_{\Delta q \cdot p^0}$ .

Если при постоянном количестве  $q^0$  изменится только цена на величину  $\Delta p$ , то стоимость изменится на величину, равную площади  $S_{\Delta p \cdot q^0}$ .

Эти изменения связаны с вычислением индексов Ласпейреса, в которых веса берутся на базовом уровне. Если же одновременно изменить и цены и количества, то получим стоимость  $w^1 = p^1 q^1$ , равную площади самого большого прямоугольника. По сравнению с суммой изменений по отдельным факторам получим дополнительное слагаемое, равное площади затушеванного прямоугольника:

$$S_{p^1 q^1} = S_{p^0 q^0} + S_{\Delta p \cdot q^0} + S_{\Delta q \cdot p^0} + S_{\Delta p \cdot \Delta q}.$$

Это объясняется как эффект от совместного изменения признаков. Удачное сравнение приведено в [ 1 ]. Допустим, что вы смотрите движущееся изображение без звука, а потом слышите звуковое сопровождение этого эпизода без изображения. Общее впечатление в итоге будет меньшим, чем в случае, когда бы вы и видели и слышали одновременно.

Следует отметить [ 1 ], что в статистической науке выведены формулы, объясняющие различие между индексами с разными весами, а также постоянно ведутся поиски индексов, удовлетворяющих требованию увязки в систему. Одним из таких вариантов являются «идеальные» индексы Фишера, вычисляемые как средние геометрические индексов Пааше и Ласпейреса:

$$I_p^\Phi = \sqrt{I_p^I \cdot I_p^P}, \quad I_q^\Phi = \sqrt{I_q^I \cdot I_q^P}.$$

Действительно,

$$I_p^\Phi \cdot I_q^\Phi = \sqrt{I_p^I \cdot I_p^P} \cdot \sqrt{I_q^I \cdot I_q^P} = \sqrt{(I_p^I \cdot I_q^I) \cdot (I_p^P \cdot I_q^P)} = \sqrt{I_w \cdot I_w} = I_w,$$

то есть эти индексы увязаны в систему с индексом общей стоимости.

## 9.6. Мультипликативная модель

В случае *мультипликативной модели*, когда результативный признак (например, общая стоимость) зависит от *двух и более факторов*, придерживаются концепции неравноправия факторов и строят индексы с разными весами в зависимости от принятой последовательности их в системе.

При этом соблюдается условие содержательной интерпретации при последовательном объединении признаков слева направо. Например, общие затраты на материалы для изготовления швейных изделий можно записать в виде  $\sum w = \sum q/p$ , где  $q$  – количество изделий,  $l$  – средний расход материала на одно изделие,  $p$  – цена материала. При такой последовательности произведение  $q/l$  имеет реальный экономический смысл – это расход материала на все производимые изделия, а  $lp$  – стоимость материалов, идущих на одно изделие. Если же признаки переставить в произведении, то этот смысл может быть потерян. Например, в последовательности  $lqr$  произведение  $qr$  не имеет экономического смысла.

В нашем случае первым стоит признак  $q$ , с которого и начинаются все изменения. Индексы будут иметь вид:

$$I_q = \frac{\sum q^1 l^0 p^0}{\sum q^0 l^0 p^0}, \quad I_l = \frac{\sum q^1 l^1 p^0}{\sum q^1 l^0 p^0}, \quad I_p = \frac{\sum q^1 l^1 p^1}{\sum q^1 l^1 p^0}.$$

Здесь использовано правило: признаки, предшествующие индексируемому признаку, берутся на отчетном уровне, а последующие признаки – на базовом уровне.

Перемножая рассмотренные индексы, очевидно получим

$$I_q \cdot I_l \cdot I_p = \frac{\sum q^1 l^1 p^1}{\sum q^0 l^0 p^0} = \frac{\sum w^1}{\sum w^0} = I_w,$$

то есть индексы увязаны в систему.

И отметим только, что при весах, взятых в одном периоде, это равенство нарушается. Если рассматривать индексы в виде

$$I_q = \frac{\sum q^1 l^0 p^0}{\sum q^0 l^0 p^0}, \quad I_l = \frac{\sum q^0 l^1 p^0}{\sum q^0 l^0 p^0}, \quad I_p = \frac{\sum q^0 l^0 p^1}{\sum q^0 l^0 p^0}, \text{ то } I_q \cdot I_l \cdot I_p \neq I_w.$$

В таком случае в левую часть неравенства вводят уравнивающий множитель, называемый *индексом совместных изменений*:

$$I_q \cdot I_l \cdot I_p \cdot I_{\text{совм. изменений}} = I_w, \text{ где } I_{\text{совм. изменений}} = \frac{I_w}{I_q \cdot I_l \cdot I_p}.$$

На практике показатель совместных изменений можно сохранить как самостоятельный член разложения или распределить его между индексами факторов. Как говорит И. И. Елисеева [ 1 ], «это зависит от поставленной задачи и от пристрастий исследователя».

## 9. 7. Агрегатные индексы как средние индивидуальных

Естественно ожидать, что между агрегатными и индивидуальными индексами, вычисленными для одной и той же статистической совокупности, имеется связь.

Рассмотрим индивидуальные индексы цен и количеств  $i_p = \frac{p^1}{p^0}$  и  $i_q = \frac{q^1}{q^0}$ , выразим отсюда их отчетные значения  $p^1 = p^0 i_p$  и  $q^1 = q^0 i_q$  и подставим их в агрегатные индексы Ласпейреса. Тогда получим  $I_p^{\Pi} = \frac{\sum p^1 q^0}{\sum p^0 q^0} = \frac{\sum p^0 q^0 i_p}{\sum p^0 q^0}$  и  $I_q^{\Pi} = \frac{\sum q^1 p^0}{\sum q^0 p^0} = \frac{\sum q^0 p^0 i_q}{\sum q^0 p^0}$ . Или окончательно

$$I_p^{\Pi} = \frac{\sum i_p \cdot (p^0 q^0)}{\sum p^0 q^0}, \quad I_q^{\Pi} = \frac{\sum i_q \cdot (p^0 q^0)}{\sum p^0 q^0}.$$

И как видим, агрегатный индекс Ласпейреса цен (количеств) является взвешенной средней арифметической индивидуальных индексов цен (количеств) с весами, равными индивидуальным стоимостям базового периода.

Аналогичным образом, выражая из индивидуальных индексов цены и количества базового периода  $p^0 = \frac{p^1}{i_p}$ ,  $q^0 = \frac{q^1}{i_q}$  и подставляя их в агрегатные индексы Пааше, получим окончательно

$$I_p^{\Pi} = \frac{\sum p^1 q^1}{\sum \frac{p^1 q^1}{i_p}}, \quad I_q^{\Pi} = \frac{\sum p^1 q^1}{\sum \frac{p^1 q^1}{i_q}}.$$

То есть, агрегатный индекс Пааше цен (количеств) является взвешенной средней гармонической индивидуальных индексов цен (количеств) с весами, равными индивидуальным стоимостям отчетного периода.

На практике полученные формулы позволяют, не имея информации о ценах и количествах, а, зная лишь их индексы и стоимости, находить агрегатные индексы этих факторов. К тому же, поскольку любая средняя величина заключена между максимальным и минимальным значениями признака, эти формулы дают возможность контролировать правильность вычислений: агрегатный индекс цен (количеств) заключен между минимальным и максимальным индивидуальными индексами цен (количеств).

И, наконец, учитывая роль весов при вычислении средних значений, следует ожидать близости агрегатного индекса цен (количества) к индивидуальному индексу цены ( $i_p$ ) (количества  $i_q$ ), имеющему наибольшую стоимость  $w_j^0 = p_j^0 q_j^0$  для индексов Ласпейреса и наибольшую стоимость  $w_j^1 = p_j^1 q_j^1$  для индексов Пааше.

*Замечание.* Применяя преобразования, аналогичные рассмотренным выше, можно выразить агрегатные индексы цен через индивидуальные индексы количеств и, наоборот, агрегатные индексы количеств через индивидуальные индексы цен. Действительно, нетрудно показать, что

$$I_p^L = \frac{\sum p^1 q^1}{\sum p^0 q^0}, \quad I_p^P = \frac{\sum p^1 q^1}{\sum i_q \cdot p^0 q^0}, \quad I_q^L = \frac{\sum p^1 q^1}{\sum p^0 q^0}, \quad I_q^P = \frac{\sum p^1 q^1}{\sum i_p \cdot p^0 q^0}.$$

Рассмотрим пример. По данным, приведенным в *таблице 9.2*, рассчитать для всей совокупности рудников агрегатные индексы общей добычи руды, производительности труда и количества работников.

*Таблица 9.2.*

Рудники	Добыча руды млн. т.		Индивидуальные индексы количе- ства работников $i_q$	Расчетная графа $p^0 q^0 i_q$
	1988 г. $w^0 = p^0 q^0$	1989 г. $w^1 = p^1 q^1$		
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>N1</b>	1,5	1,9	1,21	1,815
<b>N2</b>	7,4	7,1	0,93	6,882
<b>N3</b>	6,1	6,5	1,32	8,052
Итого	15,0	16,0		16,749

Обозначим:  $w$  - объем добычи руды,  $p$  - производительность труда,  $q$  - количество работников.

Тогда в силу полученных выше формул индекс Ласпейреса количества работников будет равен

$$I_q^L = \frac{\sum p^0 q^1}{\sum p^0 q^0} = \frac{\sum p^0 q^0 i_q}{\sum p^0 q^0} = \frac{16,749}{15,0} = 1,1166 \text{ или } 111,66 \%$$

Агрегатный индекс Паше производительности труда

$$I_p^\Pi = \frac{\sum q^1 p^1}{\sum q^1 p^0} = \frac{\sum q^1 p^1}{\sum q^0 p^0 i_q} = \frac{16,0}{16,749} = 0,9553 \text{ или } 95,53 \%$$

Общий индекс добычи руды

$$I_w = \frac{16,0}{15,0} = 1,0667 \text{ или } 106,67\%$$

Отметим, что  $I_p^\Pi \cdot I_q^\Pi = 0,9553 \cdot 1,1166 = 1,0667 = I_w$  (с погрешностью округлений), т.е. увязка индексов в систему имеется.

**Выводы.** В целом по рудоуправлению производительность труда снизилась на 4,47% на одного работающего при том, что численность работающих возросла на 11,66%. При этом общая добыча руды возросла на 6,67%, причем абсолютное увеличение ее на 1,749 млн.тонн произошло за счет изменения количества работников.

## 9.8. Индексы средней цены и структуры

Теория индексов позволяет анализировать не только факторы в совокупности, но и их средние величины. В основном интерес представляет средняя цена, которая в базовом и отчетном периодах представляется соответственно в виде:

$$\bar{p}^0 = \frac{\sum p^0 q^0}{\sum q^0}, \quad \bar{p}^1 = \frac{\sum p^1 q^1}{\sum q^1}.$$

Индекс средней цены определяется по формуле

$$I_{\bar{p}} = \frac{\bar{p}^1}{\bar{p}^0} = \frac{\sum p^0 q^0}{\sum q^0} : \frac{\sum p^1 q^1}{\sum q^1}.$$

Разделив почленно числители средних цен на знаменатели, они записываются в виде  $\bar{p}^0 = \sum p^0 d^0$ ,  $\bar{p}^1 = \sum p^1 d^1$ , где  $d^0 = \frac{q^0}{\sum q^0}$ ,  $d^1 = \frac{q^1}{\sum q^1}$

называются *долями* количеств и являются *показателями структуры* статистической совокупности. И поскольку в индексе средней цены меняется структура, то его называют еще *индексом переменного состава* и записывают в виде:

$$I_{\text{перем. состава}} = \frac{\sum p^1 d^1}{\sum p^0 d^0}.$$

Индекс переменного состава зависит от изменения цен и от изменения структуры статистической совокупности.

Желая элиминировать влияние структуры, в качестве весов рассматривают доли одного и того же периода, например, отчетного. Получаемый при этом индекс отражает изменение только средней цены и называется *индексом средней цены постоянного состава*:

$$I_{\text{пост. состава}} = \frac{\sum p^1 d^1}{\sum p^0 d^1} \quad \text{или} \quad I_{\text{пост. состава}} = \frac{\sum p^0 q^1}{\sum q^1} : \frac{\sum p^1 q^1}{\sum p^1}$$

Очевидно, после сокращения знаменателей, мы увидим, что этот индекс совпадает с агрегатным индексом цен Пааше.

С другой стороны, беря в качестве весов цены одного периода (базисного), мы имеем возможность изучить влияние структурных сдвигов с помощью индекса

$$I_{\text{струк. сдвигов}} = \frac{\sum p^0 d^1}{\sum p^0 d^0},$$

называемого *индексом структурных сдвигов*, или *индексом влияния структурных сдвигов* или просто *индексом структуры*.

Для введенных таким образом индексов выполняется равенство

$$I_{\text{струк. сдвигов}} = I_{\text{перем. состава}} : I_{\text{пост. состава}}$$

Для примера проведем анализ средней цены яблок по различным формам торговли, используя данные, приведенные ниже в *таблице 9.3* (графы 1 – 5).

Среднюю цену 1 кг яблок считаем как среднеарифметическую взвешенную (веса – количества продаж) по данным граф 2 и 3, 4 и 5:

$$\text{в феврале } \bar{p}^0 = \frac{\sum p_j^0 q_j^0}{\sum q_j^0} = \frac{8160}{270} = 30,222 \text{ ден. ед.},$$

$$\text{в марте } \bar{p}^1 = \frac{\sum p_j^1 q_j^1}{\sum q_j^1} = \frac{7640}{230} = 33,217 \text{ ден. ед.}$$

$$\text{Индекс средней цены } I_{\bar{p}} = 33,217 : 30,222 = 1,099.$$

Таблица 9.3.

Тип торговли	21 февраля 1992 г.		21 марта 1992 г.		Стоимость продаж, д.е.		Удельный вес продаж	
	цена за 1 кг д.е. $p_j^0$	прода-но кг $q_j^0$	цена за 1 кг д.е. $p_j^1$	прода-но кг $q_j^1$	базис-ный период $p_j^0 q_j^0$	отчет-ный период $p_j^1 q_j^1$	базис-ный период $d_q^0$	отчет-ный период $d_q^1$
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
Государственная	28	100	32	70	2800	2240	0,370	0,304
Кооперативная	28	120	30	100	3360	3000	0,445	0,435
Рынок	40	50	40	60	2000	2400	0,185	0,261
Итого		270		230	8160	7640	1,000	1,000

Удельные веса объемов продаж вычисляем для каждого периода по типам торговли по формуле  $d_j = \frac{q_j}{\sum q_j}$ . Например, для государственной

торговли в базисном периоде  $d_1^0 = \frac{q_1^0}{\sum q_j^0} = \frac{100}{270} = 0,370$ , а в отчетном

периоде  $d_1^1 = \frac{q_1^1}{\sum q_j^1} = \frac{70}{230} = 0,304$ ; остальные – аналогично.

Индекс цен переменного состава вычисляем по формуле

$$I_{\text{пер. сост.}} = \frac{\sum p_j^1 d_j^1}{\sum p_j^0 d_j^0} = \frac{32 \cdot 0,304 + 30 \cdot 0,435 + 40 \cdot 0,261}{28 \cdot 0,370 + 28 \cdot 0,445 + 40 \cdot 0,185} = \frac{33,218}{30,222} = 1,099.$$

Здесь веса (удельный вес продаж) являются переменными.

Индекс цен постоянного состава

$$I_{\text{пост. сост.}} = \frac{\sum p_j^1 d_{qj}^1}{\sum p_j^0 d_{qj}^1} = \frac{33,218}{28 \cdot 0,304 + 28 \cdot 0,435 + 40 \cdot 0,261} = \frac{33,218}{31,132} = 1,067$$

Индекс влияния структурных сдвигов:

$$I_{\text{стр. сдвигов}} = \frac{\sum p_j^0 d_j^1}{\sum p_j^0 d_j^0} = \frac{31,132}{30,222} = 1,030.$$

Найдем  $I_{\text{пер. сост.}} : I_{\text{пост. сост.}} = 1,099 : 1,067 = 1,030$ , т.е. найденные индексы увязаны в систему:  $I_{\text{стр. сдвигов}} = I_{\text{перем. сост.}} : I_{\text{пост. сост.}}$ .



### Анализ полученных показателей:

в 1992 году 21 марта по сравнению с 21 февраля:

- а) физический объем продаж в целом уменьшился (гр. 3 и 5, итог);
- б) средняя цена 1 кг яблок увеличилась, но общая сумма продаж уменьшилась, что является отражением снижения продаж в государственной торговле (особенно) и в кооперативной торговле (графы 6-9);
- в) индекс цен переменного состава указывает на то, что средняя цена 1 кг яблок повысилась на 9,9%, а индекс постоянного состава – на то, что среднее повышение цен произошло на 6,7% (против 9,9%). За счет изменения структуры продаж средняя цена 1 кг яблок повысилась на 3,0%.

Заканчивая этот конспект лекций, хочется заметить, что в данной работе рассмотрены только основные вопросы общей статистики и при том действительно на конспективном уровне. Задачей автора было показать основы аппарата статистической обработки данных. К сожалению, размеры издания и часы, планируемые на чтение курса статистики, не позволили практически во всех главах проводить достаточно широкий экономический анализ и показать, сколь широки рамки применения получаемых результатов. Однако автора успокаивает мысль, что те из читателей, кто заинтересуется в более глубоких знаниях статистики, могут обратиться к учебникам, упоминаемым в перечне литературы и ставших уже классическими, по крайней мере, на территориях, где в недавнем прошлом население было обязано знать русский язык.

# ПРИЛОЖЕНИЯ

## Статистико-математические таблицы

### Приложение 1.

Значения функции  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	С о т ы е д о л и									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
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	3726	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	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	2179	2155	2331	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	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
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	0208	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	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0001	0001	0001

При  $x \geq 4$  принимают  $\varphi(x) = 0$ .

Приложение 2.

Значения функции  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.0000	0.45	0.1736	0.90	0.3159	1.35	0.4115	1.80	0.4641	2.50	0.4938
0.01	0.0040	0.46	0.1772	0.91	0.3186	1.36	0.4131	1.81	0.4649	2.52	0.4941
0.02	0.0080	0.47	0.1808	0.92	0.3212	1.37	0.4147	1.82	0.4656	2.54	0.4945
0.03	0.0120	0.48	0.1844	0.93	0.3238	1.38	0.4162	1.83	0.4664	2.56	0.4948
0.04	0.0160	0.49	0.1879	0.94	0.3264	1.39	0.4177	1.84	0.4671	2.58	0.4951
0.05	0.0199	0.50	0.1915	0.95	0.3289	1.40	0.4192	1.85	0.4678	2.60	0.4953
0.06	0.0239	0.51	0.1950	0.96	0.3315	1.41	0.4207	1.86	0.4686	2.62	0.4956
0.07	0.0279	0.52	0.1985	0.97	0.3340	1.42	0.4222	1.87	0.4693	2.64	0.4959
0.08	0.0319	0.53	0.2019	0.98	0.3365	1.43	0.4236	1.88	0.4699	2.66	0.4961
0.09	0.0359	0.54	0.2054	0.99	0.3389	1.44	0.4251	1.89	0.4706	2.68	0.4963
0.10	0.0398	0.55	0.2088	1.00	0.3413	1.45	0.4265	1.90	0.4713	2.70	0.4965
0.11	0.0438	0.56	0.2123	1.01	0.3438	1.46	0.4279	1.91	0.4719	2.72	0.4967
0.12	0.0478	0.57	0.2157	1.02	0.3461	1.47	0.4292	1.92	0.4726	2.74	0.4969
0.13	0.0517	0.58	0.2190	1.03	0.3485	1.48	0.4306	1.93	0.4732	2.76	0.4971
0.14	0.0557	0.59	0.2224	1.04	0.3508	1.49	0.4319	1.94	0.4738	2.78	0.4973
0.15	0.0596	0.60	0.2257	1.05	0.3531	1.50	0.4332	1.95	0.4744	2.80	0.4974
0.16	0.0636	0.61	0.2291	1.06	0.3554	1.51	0.4345	1.96	0.4750	2.82	0.4976
0.17	0.0675	0.62	0.2324	1.07	0.3577	1.52	0.4357	1.97	0.4756	2.84	0.4977
0.18	0.0714	0.63	0.2357	1.08	0.3599	1.53	0.4370	1.98	0.4761	2.86	0.4979
0.19	0.0753	0.64	0.2389	1.09	0.3621	1.54	0.4382	1.99	0.4767	2.88	0.4980
0.20	0.0793	0.65	0.2422	1.10	0.3643	1.55	0.4394	2.00	0.4772	2.90	0.4981
0.21	0.0832	0.66	0.2454	1.11	0.3665	1.56	0.4406	2.02	0.4783	2.92	0.4982
0.22	0.0871	0.67	0.2486	1.12	0.3686	1.57	0.4418	2.04	0.4793	2.94	0.4984
0.23	0.0910	0.68	0.2517	1.13	0.3708	1.58	0.4429	2.06	0.4803	2.96	0.4985
0.24	0.0948	0.69	0.2549	1.14	0.3729	1.59	0.4441	2.08	0.4812	2.98	0.4986
0.25	0.0987	0.70	0.2580	1.15	0.3749	1.60	0.4452	2.10	0.4821	3.00	0.4987
0.26	0.1026	0.71	0.2611	1.16	0.3770	1.61	0.4463	2.12	0.4830	3.20	0.4993
0.27	0.1064	0.72	0.2642	1.17	0.3790	1.62	0.4474	2.14	0.4838	3.40	0.4997
0.28	0.1103	0.73	0.2673	1.18	0.3810	1.63	0.4484	2.16	0.4846	3.60	0.4998
0.29	0.1141	0.74	0.2703	1.19	0.3830	1.64	0.4495	2.18	0.4854	3.80	0.4999
0.30	0.1179	0.75	0.2734	1.20	0.3849	1.65	0.4505	2.20	0.4861	4.00	0.4999
0.31	0.1217	0.76	0.2764	1.21	0.3869	1.66	0.4515	2.22	0.4868	4.50	0.5000
0.32	0.1255	0.77	0.2794	1.22	0.3883	1.67	0.4525	2.24	0.4875	5.00	0.5000
0.33	0.1293	0.78	0.2823	1.23	0.3907	1.68	0.4535	2.26	0.4881		
0.34	0.1331	0.79	0.2852	1.24	0.3925	1.69	0.4545	2.28	0.4887	↓	↓
0.35	0.1368	0.80	0.2881	1.25	0.3944	1.70	0.4554	2.30	0.4893	+∞	0.5
0.36	0.1406	0.81	0.2910	1.26	0.3962	1.71	0.4564	2.32	0.4898		
0.37	0.1443	0.82	0.2939	1.27	0.3980	1.72	0.4573	2.34	0.4904		
0.38	0.1480	0.83	0.2967	1.28	0.3997	1.73	0.4582	2.36	0.4909		
0.39	0.1517	0.84	0.2995	1.29	0.4015	1.74	0.4591	2.38	0.4913		
0.40	0.1554	0.85	0.3023	1.30	0.4032	1.75	0.4599	2.40	0.4918		
0.41	0.1591	0.86	0.3051	1.31	0.4049	1.76	0.4608	2.42	0.4922		
0.42	0.1628	0.87	0.3078	1.32	0.4066	1.77	0.4616	2.44	0.4927		
0.43	0.1654	0.88	0.3106	1.33	0.4082	1.78	0.4625	2.46	0.4931		
0.44	0.1700	0.89	0.3133	1.34	0.4099	1.79	0.4633	2.48	0.4934		

При  $x > 5$  полагают  $\Phi(x) = 0,5000$ .

### Приложение 3.

Распределение Стьюдента (двусторонняя критическая область),  
 $\alpha$  – уровень значимости,  $\gamma = 1 - \alpha$  – доверительная вероятность,  $\nu$  – число степеней свободы,  $n = \nu + 1$  – объем выборки

$\alpha$ $\gamma$ $\nu \downarrow$	0,10 0,90	0,05 0,95	0,02 0,98	0,01 0,99	0,002 0,998	0,001 0,999
1	6,314	12,71	31,82	63,66	318,3	636,6
2	2,920	4,303	6,965	9,925	22,33	31,60
3	2,353	3,182	4,541	5,841	10,22	12,94
4	2,132	2,776	3,747	4,604	7,173	8,610
5	2,015	2,571	3,365	5,032	5,893	6,859
6	1,943	2,447	3,143	3,707	5,208	5,959
7	1,895	2,365	2,998	3,499	4,785	5,405
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,250	4,297	4,781
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,106	4,025	4,437
12	1,782	2,179	2,681	3,055	3,930	4,318
13	1,771	2,160	2,650	3,012	3,852	4,221
14	1,761	2,145	2,624	2,977	3,787	4,140
15	1,753	2,131	2,602	2,947	3,733	4,073
16	1,746	2,120	2,583	2,921	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,552	2,878	3,611	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,562	3,850
21	1,721	2,080	2,518	2,831	3,527	3,819
22	1,717	2,074	2,508	2,819	3,505	3,792
23	1,714	2,069	2,500	2,807	3,485	3,767
24	1,711	2,064	2,492	2,797	3,467	3,745
25	1,708	2,060	2,485	2,787	3,450	3,725
26	1,706	2,056	2,479	2,779	3,435	3,707
27	1,703	2,052	2,473	2,771	3,421	3,690
28	1,701	2,048	2,467	2,763	3,408	3,674
29	1,699	2,045	2,462	2,756	3,396	3,659
30	1,697	2,042	2,457	2,750	3,385	3,646
40	1,684	2,021	2,423	2,704	3,307	3,551
50	1,676	2,009	2,403	2,678	3,262	3,495
60	1,671	2,000	2,390	2,660	3,232	3,460
80	1,664	1,990	2,374	2,639	3,195	3,415
100	1,660	1,984	2,365	2,626	3,174	3,389
200	1,653	1,972	2,345	2,601	3,131	3,339
300	1,648	1,965	2,334	2,586	3,106	3,310
$\infty$	1,645	1,960	2,326	2,576	3,090	3,291

Приложение 4.

F- распределение (Фишера – Снедекора) при уровне значимости 0,01

$k_2$ ↓	$k_1$ – степени свободы для большей дисперсии											
	1	2	3	4	5	6	7	8	9	10	11	12
1	4 052	4 999	5 403	5 625	5 764	5 889	5 928	5 981	6 022	6 056	6 082	6 106
2	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,85	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45
20	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,30	3,23
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,09	3,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,90	2,84
	<b>14</b>	<b>16</b>	<b>20</b>	<b>24</b>	<b>30</b>	<b>40</b>	<b>50</b>	<b>75</b>	<b>100</b>	<b>200</b>	<b>500</b>	$\infty$
1	6 142	6 169	6 208	6 234	6 258	6 286	6 302	6 323	6 334	6 352	6 361	6 366
2	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,49	99,49	99,49	99,50	99,50
3	26,92	26,83	26,69	26,60	26,50	26,41	26,35	26,27	26,23	26,18	26,14	26,12
4	14,24	14,15	14,02	13,93	13,83	13,74	13,69	13,61	13,57	13,52	13,48	13,46
5	9,77	9,68	9,55	9,47	9,38	9,29	9,24	9,17	9,13	9,07	9,04	9,02
6	7,60	7,52	7,39	7,31	7,23	7,14	7,09	7,02	6,99	6,94	6,90	6,88
7	6,35	6,27	6,15	6,07	5,98	5,90	5,85	5,78	5,75	5,70	5,67	5,65
8	5,56	5,48	5,36	5,28	5,20	5,11	5,06	5,00	4,96	4,91	4,88	4,86
9	5,00	4,92	4,80	4,73	4,64	4,56	4,51	4,45	4,41	4,36	4,33	4,31
10	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91
11	4,29	4,21	4,10	4,02	3,94	3,86	3,80	3,74	3,70	3,66	3,62	3,60
12	4,05	3,98	3,86	3,78	3,70	3,61	3,56	3,49	3,46	3,41	3,38	3,36
13	3,85	3,78	3,67	3,59	3,51	3,42	3,37	3,30	3,27	3,21	3,18	3,16
14	3,70	3,62	3,51	3,43	3,34	3,26	3,21	3,14	3,11	3,06	3,02	3,00
15	3,56	3,48	3,36	3,29	3,20	3,12	3,07	3,00	2,97	2,92	2,89	2,87
16	3,45	3,37	3,25	3,18	3,10	3,01	2,96	2,89	2,86	2,80	2,77	2,75
17	3,35	3,27	3,16	3,08	3,00	2,92	2,86	2,79	2,76	2,70	2,67	2,65
20	3,13	3,05	2,94	2,86	2,77	2,69	2,63	2,56	2,53	2,47	2,44	2,42
24	2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,36	2,33	2,27	2,23	2,21
30	2,74	2,66	2,55	2,47	2,38	2,29	2,24	2,16	2,13	2,07	2,03	2,01

$k_2$  – степени свободы для меньшей дисперсии

Приложение 5.

F - распределение (Фишера – Снедекора) при уровне значимости 0,05.

$k_2$ ↓	$k_1$ – степени свободы для большей дисперсии											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
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
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
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	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,35	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
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
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
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
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
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
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
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
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
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
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
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
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
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
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
	<b>14</b>	<b>16</b>	<b>20</b>	<b>24</b>	<b>30</b>	<b>40</b>	<b>50</b>	<b>75</b>	<b>100</b>	<b>200</b>	<b>500</b>	$\infty$
1	245	246	248	249	250	251	252	253	253	254	254	254
2	19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,50	19,50
3	8,71	8,69	8,66	8,64	8,62	8,60	8,58	8,57	8,56	8,54	8,54	8,53
4	5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,66	5,65	5,64	5,63
5	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36
6	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67
7	3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23
8	3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93
9	3,02	2,98	2,93	2,90	2,86	2,82	2,80	2,77	2,76	2,73	2,72	2,71
10	2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54
11	2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40
12	2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,36	2,35	2,32	2,31	2,30
13	2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,26	2,24	2,22	2,21
14	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
15	2,43	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,12	2,10	2,08	2,07
16	2,37	2,33	2,28	2,24	2,20	2,16	2,13	2,09	2,07	2,04	2,02	2,01
17	2,33	2,29	2,23	2,19	2,15	2,11	2,08	2,04	2,02	1,99	1,97	1,96
20	2,23	2,18	2,12	2,08	2,04	1,99	1,96	1,92	1,90	1,87	1,85	1,84
24	2,13	2,09	2,02	1,98	1,94	1,89	1,86	1,82	1,80	1,76	1,74	1,73
30	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,69	1,66	1,64	1,62

Приложение 6.

$\chi^2$  – распределение

$\nu$  - число степеней свободы,  $\alpha$  - уровень значимости

$\alpha$ $\nu$	0,20	0,10	0,05	0,02	0,01	0,001
1	1,642	2,706	3,841	5,412	6,635	10,827
2	3,219	4,605	5,991	7,824	9,210	13,815
3	4,642	6,251	7,815	9,837	11,345	16,266
4	5,989	7,779	9,488	11,668	13,237	18,467
5	7,289	9,236	11,070	13,388	15,086	20,515
6	8,558	10,645	12,592	15,033	16,812	22,457
7	9,803	12,017	14,067	16,622	18,475	24,322
8	11,030	13,362	15,507	18,168	20,090	26,125
9	12,242	14,684	16,919	19,679	21,666	27,877
10	13,442	15,987	18,307	21,161	23,209	29,588
11	14,631	17,275	19,675	22,618	24,795	31,264
12	15,812	18,549	21,026	24,054	26,217	32,909
13	16,985	19,812	22,362	25,472	27,688	34,528
14	18,151	21,064	23,685	26,783	29,141	36,123
15	19,311	22,307	24,996	28,259	30,578	37,697
16	20,465	23,542	26,296	29,633	32,000	39,252
17	21,615	24,769	27,587	30,995	32,409	40,790
18	22,760	25,989	28,869	32,346	34,805	42,312
19	23,900	27,204	30,144	33,678	36,191	43,820
20	25,038	28,412	31,410	35,020	37,566	45,315
21	26,171	29,615	32,671	36,343	38,932	46,797
22	27,301	30,813	33,924	37,659	40,289	48,268
23	28,429	32,007	35,172	38,968	41,638	49,728
24	29,553	33,196	36,415	40,270	42,980	51,179
25	30,675	34,382	37,652	41,566	44,314	52,620
26	31,795	35,563	38,885	42,856	45,642	54,052
27	32,912	36,741	40,113	44,140	46,963	55,476
28	34,027	37,916	41,337	45,419	48,278	56,893
29	35,139	39,087	42,557	46,693	49,588	58,302
30	36,250	40,256	43,773	47,962	50,892	59,703

## Приложение 7.

Таблица значений  $q = q(\gamma, n)$ .

$(1 - q) s < \sigma < (1 + q) s$ , если  $q < 1$ ,  $0 < \sigma < (1 + q) s$ , если  $q > 1$ .

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

$n \backslash \gamma$	0,95	0,99	0,999	$n \backslash \gamma$	0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,211
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

## Приложение 8.

Критические значения  $\lambda_\alpha$ -распределения Колмогорова:

$$P(\lambda \geq \lambda_\alpha) = \alpha$$

Уровень значимости $\alpha$	0,20	0,10	0,05	0,02	0,01	0,001
$\lambda_\alpha$	1,073	1,224	1,358	1,520	1,627	1,950



## Приложение 9.

Значения коэффициента корреляции рангов Спирмэна  
для двусторонних пределов уровня значимости  $\alpha$

$\alpha$ $n$	0,20	0,10	0,05	0,02	0,01	0,001
4	0,8000	0,8000				
5	0,7000	0,8000	0,9000	0,9000		
6	0,6000	0,7714	0,8286	0,8857	0,9429	
7	0,5357	0,6786	0,7450	0,8571	0,8929	0,9643
8	0,5000	0,6190	0,7143	0,8095	0,8571	0,9286
9	0,4667	0,5833	0,6833	0,7667	0,8167	0,9000
10	0,4424	0,5515	0,6364	0,7333	0,7818	0,8667
11	0,4182	0,5273	0,6091	0,7000	0,7455	0,8364
12	0,3986	0,4965	0,5804	0,6713	0,7273	0,8182
13	0,3791	0,4780	0,5549	0,6429	0,6978	0,7912
14	0,3626	0,4593	0,5341	0,6220	0,6747	0,7670
15	0,3500	0,4429	0,5179	0,6000	0,6536	0,7464
16	0,3382	0,4265	0,5000	0,5824	0,6324	0,7265
17	0,3260	0,4118	0,4853	0,5637	0,6152	0,7083
18	0,3148	0,3994	0,4716	0,5480	0,5975	0,6904
19	0,3070	0,3895	0,4579	0,5333	0,5825	0,6737
20	0,2977	0,3789	0,4451	0,5203	0,5686	0,6586
21	0,2909	0,3688	0,4351	0,5078	0,5545	0,6455
22	0,2829	0,3597	0,4241	0,4963	0,5426	0,6318
23	0,2767	0,3518	0,4150	0,4852	0,5306	0,6186
24	0,2704	0,3435	0,4061	0,4748	0,5200	0,6070
25	0,2646	0,3362	0,3977	0,4654	0,5100	0,5962
26	0,2588	0,3299	0,3894	0,4564	0,5002	0,5856
27	0,2540	0,3236	0,3822	0,4481	0,4915	0,5757
28	0,2490	0,3175	0,3749	0,4401	0,4828	0,5660
29	0,2443	0,3113	0,3685	0,4320	0,4744	0,5567
30	0,2400	0,3059	0,3620	0,4251	0,4665	0,5479

## Греческий алфавит

Прописные		Строчные		Название	Прописные		Строчные		Название
1	2	1	2		3	1	2	1	
<i>A</i>	<b>A</b>	<i>α</i>	<b>α</b>	áльфа	<i>N</i>	<b>N</b>	<i>ν</i>	<b>ν</b>	ни (ню)
<i>B</i>	<b>B</b>	<i>β</i>	<b>β</b>	бéта	<i>Ξ</i>	<b>Ξ</b>	<i>ξ</i>	<b>ξ</b>	кси
<i>Γ</i>	<b>Γ</b>	<i>γ</i>	<b>γ</b>	гáμμα	<i>Ο</i>	<b>Ο</b>	<i>ο</i>	<b>ο</b>	ο микρόν
<i>Δ</i>	<b>Δ</b>	<i>δ</i>	<b>δ</b>	дéльта	<i>Π</i>	<b>Π</b>	<i>π</i>	<b>π</b>	пи
<i>E</i>	<b>E</b>	<i>ε</i>	<b>ε</b>	э псилóн	<i>Ρ</i>	<b>Ρ</b>	<i>ρ</i>	<b>ρ</b>	ро
<i>Z</i>	<b>Z</b>	<i>ζ</i>	<b>ζ</b>	дзéта	<i>Σ</i>	<b>Σ</b>	<i>σ ς</i>	<b>σ ς</b>	си <input type="checkbox"/>
<i>H</i>	<b>H</b>	<i>η</i>	<b>η</b>	э <input type="checkbox"/> та	<i>Τ</i>	<b>Τ</b>	<i>τ</i>	<b>τ</b>	тау
<i>Θ</i>	<b>Θ</b>	<i>θ</i>	<b>θ</b>	тéта	<i>Υ</i>	<b>Υ</b>	<i>υ</i>	<b>υ</b>	и псилóн
<i>I</i>	<b>I</b>	<i>ι</i>	<b>ι</b>	иόта	<i>Φ</i>	<b>Φ</b>	<i>φ</i>	<b>φ</b>	фи
<i>K</i>	<b>K</b>	<i>κ</i>	<b>κ</b>	кáппа	<i>Χ</i>	<b>Χ</b>	<i>χ</i>	<b>χ</b>	хи
<i>Λ</i>	<b>Λ</b>	<i>λ</i>	<b>λ</b>	лáмбда	<i>Ψ</i>	<b>Ψ</b>	<i>ψ</i>	<b>ψ</b>	пси
<i>M</i>	<b>M</b>	<i>μ</i>	<b>μ</b>	ми (мю)	<i>Ω</i>	<b>Ω</b>	<i>ω</i>	<b>ω</b>	ο méγα

В таблице приводится написание греческих букв курсивом двумя шрифтами: первые столбцы – «Times New Roman», вторые – «Arial».

## Рекомендуемая литература

1. Елисеева, И.И. Общая теория статистики: учебник / И.И. Елисеева, М.М. Юзбашев – М.: Финансы и статистика, 1996.
2. Ефимова, М.Р. Общая теория статистики: учебник / М.Р. Ефимова, Е.В. Петрова, В.Н. Румянцев – М.: ИНФРА–М, 1997.
3. Общая теория статистики / Под ред. А.А. Спирина, О.Э. Башиной – М.: Финансы и статистика, 1995.
4. Статистика: национальные счета, показатели и методы анализа: справочное пособие / Под общей ред. И.Э. Теслюка – Мн.: БГЭУ, 1995.
5. Общая теория статистики: практикум / Под общей ред. М.М. Новикова – Мн.: БГЭУ, 1996.
6. Теория статистики / Под ред. Р.А. Шмойловой – М.: Финансы и статистика, 1998.
7. Практикум по теории статистики: учебное пособие / Под ред. проф. Р.А. Шмойловой. – М.: Финансы и статистика, 1999.
8. Громыко, Г.М. Общая теория статистики: практикум / Г.М. Громыко – М.: ИНФРА-М, 1999.
9. Елисеева, И.И. Моя профессия – статистик / И.И. Елисеева – М.: Финансы и статистика, 1992.
10. Статистический словарь / Под ред. М.А. Королева. 2-е изд. – М.: Финансы и статистика, 1989.
11. Годунов, Б. А. Статистика, (Конспект лекций): учебное пособие / Б.А. Годунов – Брест: Издательство БрГТУ, 2008. – Ч.1.

## Оглавление

### ГЛАВА 6

<b>Основы корреляционного анализа .....</b>	<b>3</b>
6.1. Различные формы связи между признаками .....	3
6.2. Предпосылки корреляционной связи .....	5
6.3. Различные уравнения регрессии .....	8
6.4. Парная линейная корреляция .....	9
6.5. Расчет параметров нелинейных уравнений регрессии .....	13
6.6. Другая форма линейного уравнения регрессии. Коэффициент корреляции .....	16
6.7. Коэффициент Фехнера оценки тесноты зависимости .....	19
6.8. Оценка тесноты линейной зависимости .....	21
6.9. Существенность коэффициента корреляции .....	23
6.10. Оценка тесноты связи в нелинейном случае .....	25

### ГЛАВА 7

<b>Корреляционные связи между неколичественными признаками .....</b>	<b>28</b>
7.1. Корреляция рангов. Коэффициент Спирмена .....	28
7.2. Коэффициент корреляции рангов Кендэла .....	30
7.3. Некоторые другие методы измерения тесноты связи .....	32

### ГЛАВА 8

<b>Ряды динамики .....</b>	<b>37</b>
8.1. Определение и виды рядов динамики .....	37
8.2. Показатели рядов динамики .....	38
8.3. Средние показатели ряда динамики .....	43
8.4. Типы тенденции динамики .....	45
8.5. Аналитические методы выбора тренда .....	47
8.6. Расчет параметров тренда .....	51

## **Г Л А В А 9**

<b>Индексы .....</b>	<b>56</b>
9.1. Понятие индекса и задачи, решаемые с их помощью .....	56
9.2. Индивидуальные индексы. Увязка в систему .....	57
9.3. Агрегатные индексы .....	59
9.4. Абсолютные изменения общей стоимости .....	61
9.5. Знак Вазара. «Идеальная» формула Фишера .....	65
9.6. Мультипликативная модель .....	66
9.7. Агрегатные индексы как средние индивидуальных .....	68
9.8. Индексы средней цены и структуры .....	70
 <b>Приложения</b>	
<b>Статистико-математические таблицы .....</b>	<b>74</b>
<b>Греческий алфавит .....</b>	<b>82</b>
<b>Рекомендуемая литература .....</b>	<b>83</b>

УЧЕБНОЕ ИЗДАНИЕ

Годунов Борис Алексеевич

# **СТАТИСТИКА**

**Часть 2**

**(Конспект лекций)**

Ответственный за выпуск: Годунов Б. А.  
Редактор: Строкач Т. В.  
Компьютерная верстка: Боровикова Е. А.  
Корректор: Никитчик Е. В.

---

Подписано к печати 4.12.2009 г.  
Бумага «Снегурочка». Формат 60×84 <sup>1</sup>/<sub>16</sub>.  
Усл. п. л. 5,12. Уч.-изд. л. 5,5. Тираж 200 экз.  
Заказ № 1122. Отпечатано на ризографе  
учреждения образования  
«Брестский государственный  
технический университет».

224017, г. Брест, ул. Московская, 267.





