Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»

СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ

Учебное пособие

Владимир 2018

УДК 338.2
ББК 65.051
С78

Авторы: Е. М. Марченко, М. В. Рахова, А. К. Холодная, Т. Д. Белова

Рецензенты:
Доктор экономических наук, профессор
зав. кафедрой менеджмента и маркетинга
Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых
Н. М. Филимонова

Доктор экономических наук, профессор
зав. кафедрой экономики
Владимирского филиала Российской академии народного хозяйства
и государственных служащих при Президенте РФ
О. Б. Дигилина

Издается по решению редакционно-издательского совета ВлГУ

С78 **Статистическое** моделирование и прогнозирование : учеб. пособие / Е. М. Марченко [и др.] ; Владим. гос. ун-т им А. Г. и Н. Г. Столетовых. – Владимир : Изд-во ВлГУ, 2018. – 100 с.
ISBN 978-5-9984-0861-8

Посвящено методам моделирования и прогнозирования, в том числе методам многомерного анализа, методам выявления тенденции и сглаживания рядов динамики, построению прогнозов, выделению сезонной составляющей, расчету точности и адекватности моделей, понятиям адаптивных моделей и экспоненциального сглаживания.

Предназначено для студентов всех форм обучения по направлениям 38.03.01 «Экономика», 38.03.03 «Управление персоналом» и 38.05.01 «Экономическая безопасность».

Рекомендовано для формирования профессиональных компетенций в соответствии с ФГОС ВО.

Ил. 21. Табл. 48. Библиогр.: 16 назв.

УДК 338.2
ББК 65.051

ISBN 978-5-9984-0861-8

ВВЕДЕНИЕ

Прогнозирование – это взгляд в будущее, оценка возможных путей развития, последствий тех или иных решений. Поскольку любая будущая ситуация неопределенна, то не существует способов, позволяющих точно «угадать» ее развитие. Вместе с тем известны статистические методы, позволяющие обнаружить тенденцию, логику развития тех или иных процессов и на основе сопоставления с прошлым опытом предсказать характеристики наиболее вероятного развития событий. Процесс прогнозирования включает сбор и подготовку необходимых данных по рядам динамики, проверку гипотезы о наличии тенденции, выявление тренда и сезонной составляющей, проверку модели на точность и адекватность, построение точечного и интервального прогнозов. Представленный в пособии раздел об адаптивных моделях дает представление об экспоненциальном сглаживании рядов динамики.

Моделирование – это упрощение реальных взаимосвязей с помощью ряда допущений и ограничений, позволяющее выявить наиболее существенные закономерности и зависимости. В последние годы в экономических исследованиях важное место отводят многомерному анализу, который позволяет комплексно учесть влияние факторов, провести ранжирование пространственных или производственных систем.

Все эти вопросы представлены в пособии и раскрыты на конкретных примерах. В конце каждой темы предлагаются ряд контрольных вопросов для проверки уровня освоения теоретического материала и практические задания для закрепления изученного материала.

В качестве исходных данных для практических заданий и примеров использованы материалы официального сайта Федеральной службы государственной статистики Российской Федерации (http://www.gks.ru) и территориального органа Федеральной службы государственной статистики по Владимирской области (http://vladimirstat.gks.ru).

В целом пособие позволит студентам всех форм обучения по направлениям бакалавриата 38.03.01 «Экономика», 38.03.03 «Управление персоналом» и специальности 38.05.01 «Экономическая безопасность» овладеть компетенцией грамотной работы со статистическим материалом, что составляет основу подготовки любого экономиста и управленца.

1. РЯДЫ ДИНАМИКИ

Виды рядов динамики и их характеристика • Условия сопоставимости рядов динамики • Расчет среднего уровня ряда динамики • Показатели рядов динамики • Смыкание рядов динамики

1.1. Виды рядов динамики и их характеристика

Ряд динамики представляет собой ряд данных, характеризующих изменение социально-экономического явления во времени.

Основные характеристики рядов динамики – это уровень и время.

Уровень ряда динамики выражен абсолютными, относительными или средними величинами. Время может быть представлено моментными или интервальными показателями.

В зависимости от представления временной характеристики ряды динамики делят:

- на **моментные,** характеризующие состояние на конкретный момент, определенную дату. Моментные ряды динамики содержат результаты проверок на всех уровнях:

на уровне предприятия – остатки запасов на складе, средств на счетах предприятия, балансовые показатели на начало или конец отчётного периода;

на уровне города – наличие жилого фонда, посадочных мест в библиотеке;

на уровне страны – численность населения по результатам переписи;

- **интервальные** ряды динамики представлены показателями, характеризующими результаты деятельности за период времени (годовой объем производства, ежемесячные доходы, еженедельные затраты) и средними показателями (средней заработной платой, выработкой, затратами на единицу продукции).

Необходимо помнить, что уровни интервальных рядов динамики суммировать можно, а моментных рядов – нельзя.

1.2. Условия сопоставимости рядов динамики

Основное условие работы с рядами динамики – это сопоставимость и правильный выбор начального уровня ряда динамики. По умолчанию начальным уровнем считают первый уровень ряда динамики.

Ряды динамики должны быть сопоставлены по единицам измерения; масштабу, территории; структуре предприятия и продукции; методике исчисления показателей; уровню цен.

1.3. Расчет среднего уровня ряда динамики

Средний уровень интервального ряда рассчитывают по средней арифметической простой для равных интервалов и средней арифметической взвешенной – для неравных.

Средний уровень моментного ряда определяют по средней хронологической для равных интервалов:

$$\bar{x}_{\text{хр}} = \frac{\frac{1}{2}x_1 + x_2 + ... + x_{n-1} + \frac{1}{2}x_n}{n-1}$$, где n – количество дат.

Средний уровень моментного ряда с неравными интервалами между датами рассчитывают путем перехода от моментного ряда к интервальному с последующим расчетом по средней арифметической взвешенной.

Пример 1.1

По данным табл. 1.1 рассчитать средний размер активов ПАО «Сбербанк» за 2016 г.

Таблица 1.1. Размер активов ПАО «Сбербанк» в 2016 г., млрд руб.

Период	**1-й квартал 2016**	**2-й квартал 2016**	**3-й квартал 2016**	**4-й квартал 2016**	**1-й квартал 2017**
Активы на начало квартала	26 572	25 794	25 532	25 352	24 655

Решение.

Представленный ряд динамики является моментным, периоды между отчётными датами равны. Для расчёта среднего уровня применяем формулу средней хронологической

$$\bar{x}_{xp} = \frac{\frac{1}{2} \times 26572 + 25794 + 25532 + 25532 + \frac{1}{2} \times 24655}{5-1} = 25572,88 \text{ млрд руб.}$$

Ответ: средний размер активов ПАО «Сбербанк» в 2016 г. составил 25572,88 млрд руб.

Пример 1.2

По данным табл. 1.2 рассчитать среднегодовые остатки денежных средств клиентов ПАО «Сбербанк»

Таблица 1.2. Денежные средства клиентов ПАО «Сбербанк» за 2014 – 2016 гг., млрд руб.

Дата	**Денежные средства клиентов на отчетную дату, млрд руб.**	**Расчетные графы**	
		Длительность периода, мес.	**Средняя величина денежных средств за период, млрд руб.**
01.01.2014	12 566	6	(12 566+13 210) / 2 = 12 888,0
01.07.2014	13 210	3	14 386,5
01.10.2014	15 563	9	16 925,0
01.07.2015	18 287	3	19 042,5
01.10.2015	19 798	5	19 292,5
01.03.2016	18 787	7	18 736,0
01.10.2016	18 685	3	18 364,0
01.01.2017	18 043		

Решение.

Представленный ряд динамики является моментным, периоды между отчётными датами не равны. Для расчёта среднего уровня переходим к интервальному ряду путём расчёта средней величины денежных средств за период, далее применяем формулу средней арифметической взвешенной

$$\bar{x} = \frac{6\cdot 12888 + 3\cdot 14386,5 + 9\cdot 16925 + 3\cdot 19042,5 + 5\cdot 19292,5 + 7\cdot 18736 + 3\cdot 18364}{36} =$$

=17017,96 млрд руб.

Ответ: средний размер денежных средств клиентов ПАО «Сбербанк» составил 17 017,96 млрд руб.

1.4. Показатели рядов динамики

К показателям рядов динамики относят абсолютный прирост, темп роста, темп прироста, абсолютное значение одного процента прироста. Все показатели, за исключением абсолютного значения одного процента прироста, могут быть цепными, базисными, средними. Формулы расчета всех показателей представлены в табл. 1.3.

Цепные показатели характеризуют данный уровень по отношению к предыдущему.

Базисные показатели сравнивают данный уровень с уровнем, принятым за базу сравнения. Как правило, в качестве базисного принимают первый уровень ряда динамики.

Таблица 1.3. Формулы расчета основных показателей рядов динамики[1]

Показатель	Вид показателя		
	цепной	**базисный**	**средний**
Абсолютный прирост, ед.[2]	$\Delta x_i^{ц} = x_i - x_{i-1}$	$\Delta x_i^{б} = x_i - x_б$	$\overline{\Delta x} = \frac{\sum_{i=2}^{n} \Delta x_i}{n-1}$ или $\overline{\Delta x} = \frac{\Delta x_n^{б}}{n-1}$
Коэффициент роста[3]	$k_i^{ц} = \frac{x_i}{x_{i-1}}$	$k_i^{б} = \frac{x_i}{x_б}$	$\bar{k} = \sqrt[n-1]{\prod_{i=2}^{n} k_i^{ц}}$ или $\bar{k} = \sqrt[n-1]{k_n^{б}}$

[1] n в таблице соответствует числу дат, а $(n-1)$ – количеству периодов.

[2] Сумма цепных абсолютных приростов равна базисному абсолютному приросту последнего периода $\sum \Delta x_i^{ц} = x_n^{б}$.

[3] Произведение цепных коэффициентов роста равно базисному коэффициенту роста последнего периода $\prod_{i=2}^{n} k_i^{ц} = k_n^{б}$.

Окончание табл. 1.3

Показатель	Вид показателя		
	цепной	базисный	средний
Темп роста, %	$T_{i\%}{}^{ц} = \frac{x_i}{x_{i-1}} 100\%$	$T_{i\%}{}^{б} = \frac{x_i}{x_б} 100\%$	$\bar{T} = \bar{k}100\%$
Темп прироста, %	$\Delta T_{i\%}{}^{ц} = T_{i\%}{}^{ц} - 100\%$	$\Delta T_{i\%}^{б} = T_{i\%}^{б} - 100\%$	$\Delta\bar{T}_{i\%} = \bar{T}_{i\%} - 100\%$
Абсолютное значение одного процента прироста, ед.	$\overline{АЗП}_i = \frac{x_{i-1}}{100}$	–	$\overline{АЗП} = \frac{\overline{\Delta x}}{\overline{\overline{\Delta Т}}}$

Рассмотрим пример расчета показателей ряда динамики.

Пример 1.3

По данным о чистой прибыли ПАО «Сбербанк», представленным в табл. 1.4, рассчитать абсолютные и относительные показатели динамики.

Таблица 1.4. Ежеквартальная динамика чистой прибыли ПАО «Сбербанк» в 2016 г.

Показатель	1-й квартал 2016	2-й квартал 2016	3-й квартал 2016	4-й квартал 2016	Средний показатель
Исходные данные					
Чистая прибыль, млрд руб.	**117,7**	**145,4**	**137,0**	**141,8**	(117,7 + 145,4 + 137,0 + + 141,8) = 135,48
Решение					
Абсолютный прирост, млрд руб.					
Цепной	–	145,4 – – 117,7 = 27,7	137 – 145,4 = = – 8,4	4,8	(27,7 + (–8,4) + 4,8) / / (4 – 1) = = 8,03
Базисный	117,7 – – 117,7 = 0	145,4 – – 117,7 = 27,7	137 – – 117,7 = = 19,3	24,1	24,1 / (4 – 1) = 8,03
Коэффициент роста					
Цепной	–	145,4 / / 117,7 = = 1,235	137 / 145,4 = = 0,942	1,035	$\sqrt[4-1]{1,235 \cdot 0,942 \cdot 1,035} =$ = 1,0641
Базисный	117,7 / / 117,7 = 1	145,4 / / 117,7= = 1,235	137,0 / / 117,7 = = 1,164	1,205	$\sqrt[4-1]{1,205} = 1,0641$

Окончание табл. 1.3

Показатель	1-й квартал 2016	2-й квартал 2016	3-й квартал 2016	4-й квартал 2016	Средний показатель
Темп роста, %					
Цепной	–	1,235 · 100 = = 123,5	94,2	103,5	1,06 · 100 = 106,41
Базисный	100	123,5	116,4	120,5	106,41
Темп прироста, %					
Цепной	–	123,5 – 100 = = 23,5	–5,8	3,5	106,41 – 100 = 6,41
Базисный	0	23,5	16,4	20,5	6,41
Абсолютное значение 1 % прироста (АЗП)	–	117,7 / 100 = = 1,177	145,4 / 100 = = 1,454	1,370	8,03 / 6,41 = 1,25

Все показатели рассчитывают одинаково для моментных и интервальных рядов динамики.

1.5. Смыкание рядов динамики

Смыкание рядов динамики – это процесс приведения рядов динамики к сопоставимому виду по методике исчисления, ценам, структуре, единицам измерения, базисному уровню и т. д.

При этом возможен перевод базисных показателей в новый базисный уровень, цепных показателей – в базисные, базисных показателей – в цепные.

При переводе базисных темпов роста в новый базисный уровень необходимо все базисные показатели разделить на показатель нового базисного уровня, а его принять за 100 %. В примере 1.4 для расчёта базисного темпа роста объёма производства 2012 г. к 2013 г. необходимо разделить базисный темп роста 2012 года на тот же показатель 2013 г., т. е. $T^{б}_{2013Q} = \frac{100}{103}\ 100 = 97{,}1$ %.

При переводе цепных показателей в базисные следует использовать формулу взаимосвязи цепных и базисных коэффициентов роста (табл. 1.3), в соответствии с которой базисный коэффициент роста равен произведению цепных показателей, начиная со второго показателя от начала анализируемого периода. В примере 1.4 для расчета базисного темпа роста среднесписочной численности работников 2015 г. к 2013 г. необходимо перемножить цепные темпы роста 2014 и 2015 гг., т. е. $\text{Т}^{\text{б}}_{2015\ \text{ч}} = \frac{105 \cdot 103}{100} = 108{,}2\ \%$.

При переводе базисных темпов роста в цепные в качестве базы сравнения используют показатель предыдущего периода, т. е. для расчета цепного показателя 2014 г. необходимо базисный показатель 2014 г. разделить на базисный показатель 2013 г.

При совместном использовании моментных и интервальных рядов динамики необходимо моментный ряд приводить к интервальному виду путём расчёта средних показателей по периодам (годам, кварталам и т. д.). При этом полученный интервальный ряд укорачивается на одно значение в зависимости от исходных данных. Например, если данные в моментном ряду динамики приведены на начало года, то в итоговом ряду отсутствует последний период, если исходные данные представлены на конец года – первый период. В примере 1.5 отражен моментный ряд динамики численности работников по списку на начало года и интервальный ряд динамики объёма продукции. Чтобы рассчитать ежегодные показатели выработки, необходимо сопоставить среднегодовые показатели, т. е. преобразовать исходный ряд численности работников в интервальный ряд с данными за год. Так для расчёта среднегодовой численности работников за 2012 г. найдём полусумму значений на начало 2012 и 2013 гг. $\overline{\text{Ч}}_{2012} = \frac{\text{Ч}^{\text{н.г.}}_{2012} + \text{Ч}^{\text{н.г.}}_{2013}}{2} = \frac{30+40}{2} = 35$. Итоговый интервальный ряд среднегодовой численности работников сократился на одно значение в конце периода.

При смыкании рядов динамики, приведенных по двум методикам или представленных в разных ценах, необходимо иметь данные одного периода, рассчитанные при различных условиях, что даёт возможность получить коэффициент пересчёта из одних условий в другие. В примере 1.6 нужно привести объёмы работ, рассчитанные по разным методикам, к единому виду. Для этого следует вычислить коэффициент пересчета как отношение объемов работ 2015 г. к 2012 г.: $k_{2015/2012} = \frac{Q^{\text{нов}}_{2015}}{Q^{\text{стар}}_{2015}} = \frac{30}{28} = 1{,}07$. Далее необходимо умножить данные, полученные по методике 2012 г., на коэффициент.

Пример 1.4

По приведенным в табл. 1.5 данным составить ряд динамики базисных темпов роста выработки к 2013 г.

Решение.

$$\text{Т}_{\text{в}} = \frac{\text{Т}_{Q}}{\text{Т}_{\overline{\text{Ч}}}},$$

где Т_{Q} – темп роста объема производства; $\text{Т}_{\overline{\text{Ч}}}$ – темп роста среднесписочной численности работников.

Для построения динамического ряда темпов роста выработки по отношению к 2013 г. необходимо перевести базисные темпы роста объема производства к новой базе, а цепные темпы роста среднесписочной численности работников – к базисным показателям 2013 г.

Расчеты представлены в табл. 1.5.

Таблица 1.5. Расчёт динамики выработки по предприятию

Показатель	**Год**					
	2012	**2013**	**2014**	**2015**	**2016**	**2017**
Исходные данные						
Темп роста объема производства базисный к **2012** г., %	100	103	104	107	110	115
Темп роста среднесписочной численности работников цепной, %	101	102	105	103	101	102
Решение						
Темп роста объема производства базисный к **2013** г., %	$\frac{100}{103}100 = 97{,}1$	100,0	101,0	$\frac{107}{103}100 = 103{,}9$	106,8	111,7
Темп роста среднесписочной численности работников базисный к **2013** г., %	$\frac{100}{102}100 = 98{,}0$	100,0	105,0	$\frac{105 \cdot 103}{100} = 108{,}2$	109,2	111,4
Темп роста выработки базисный к **2013** г.	$\frac{97{,}1}{98{,}0}100 = 99{,}0$	100,0	96,2	$\frac{103{,}9}{108{,}2}100 = 96{,}1$	97,8	100,2

Пример 1.5

По данным табл. 1.6 рассчитать базисные темпы прироста выработки по годам к 2012 г.

Таблица 1.6. Расчёт базисных темпов прироста выработки

Показатель	**Год**					
	2012	**2013**	**2014**	**2015**	**2016**	**2017**
Исходные данные						
Объём продукции **за год**, тыс. шт.	150	200	220	280	300	310
Численность работников по списку **на начало года**, чел.	30	40	38	50	32	48
Решение						
Среднегодовая списочная численность работников $T_{\overline{ч}}$, чел.	(30 + 40) / 2 = **35**	(40 + 38) / 2 = **39**	44	41	40	–
Выработка на одного работника: $B = Q/T$	150 / 35 = 4,29	200 / 39 = 5,13	5,00	6,83	7,50	–
Темп роста выработки базисный к 2012 г., %	$\frac{4{,}29}{4{,}29} \cdot 100 = 100{,}00$	$\frac{5{,}71}{4{,}29} \cdot 100 = 119{,}66$	116,67	159,35	175,00	–
Темп прироста выработки базисный к 2012 г., %	100 – 100 = 0,00	119,66 – 100 = 19,66	16,67	59,35	75,00	–

Пример 1.6

По данным табл. 1.7 составить единый ряд динамики объемов работ предприятия с 2012 по 2017 гг.

Таблица 1.7. Пример смыкания рядов динамики

Показатель	**Год**					
	2012	**2013**	**2014**	**2015**	**2016**	**2017**
Объем работ, млн руб. по методике 2012 г.	15,0	20,0	22,0	28,0	–	–
Объем работ, млн руб. по методике 2015 г.	–	–	–	30,0	42,0	44,0
Коэффициент пересчета объемов 2015 г. по отношению к 2012 г.	30 / 28 = 1,07					
Объем работ, млн руб. по методике 2015 г.	15·1,07 = 16,1	21,4	23,5	30,0	42,0	44,0

Контрольные вопросы

1. Какие ряды динамики относятся к моментным?

2. Какие ряды динамики называются интервальными?

3. Как рассчитать средний уровень моментного ряда для равных и неравных интервалов?

4. Как определить средний уровень интервального ряда для равных и неравных интервалов?

5. Перечислите показатели рядов динамики.

Практические задания

Задача 1.1

По приведенным в табл. 1.8 данным рассчитать базисные темпы роста и прироста реальных располагаемых доходов населения РФ к 2008 г.

Таблица 1.8. Динамика реальных располагаемых доходов населения РФ за период с 2008 по 2017 гг.

Год	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017*
В процентах к предыдущему периоду	102,4	103	105,9	100,5	104,6	104	99,3	96,8	94,2	98,3

* Предварительные данные.

Задача 1.2

Рассчитать среднегодовой объем денежных накоплений населения РФ в 2016 г.

Таблица 1.9. Денежные накопления населения РФ в 2016 г.

Год	2016				2017
Квартал	I	II	III	IV	I
Всего накоплений на начало квартала, млрд руб.	24 125,8	24 209,5	25 319,7	25 759,9	27 423,0

Задача 1.3

Рассчитать среднегодовой объем денежных накоплений населения РФ в 2016 г.

Таблица 1.10. Денежные накопления населения РФ в 2016 г.

Год	2016					2017
Месяц	Январь	Февраль	Июль	Сентябрь	Октябрь	Январь
Всего накоплений на начало месяца, млрд руб.	24 125,8	23 547,9	25 319,7	25 638,3	25 759,9	27 423,0

Задача 1.4

Рассчитать базисный темп роста основных фондов в Российской Федерации 2017 г. к 2006 г. (табл. 1.11), имея в виду, что с 2006 г. по 2013 г. данные рассчитаны в ценах 2000 г., а с 2014 г. – в ценах 2010 г. Индекс пересчета 2010 г. по сравнению с 2000 г. принять равным 2,34.

Таблица 1.11. Динамика наличия основных фондов на конец года в Российской Федерации (в сопоставимых ценах), в % к предыдущему году

Год	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017*
Всего	102,4	103,1	103,6	103,2	103,0	104,0	104,3	104,1	103,7	103,2	103,9	103,9

* Предварительные данные.

Задача 1.5

По данным о производстве сельскохозяйственных машин в РФ, приведенным в табл. 1.12, рассчитать по каждому типу машин средние абсолютные и относительные приросты с 2010 по 2015 гг.

Таблица 1.12. Динамика производства сельскохозяйственных машин в РФ

Наименование	**Год**					
	2010	**2011**	**2012**	**2013**	**2014**	**2015**
Тракторы для сельского и лесного хозяйства прочие, тыс. шт.	6,9	13,2	13,6	7,6	6,7	5,5
Плуги общего назначения, шт.	1550	3719	4007	3369	2842	2993
Культиваторы для сплошной обработки почвы, тыс. шт.	25,6	28,8	24,2	16,5	14,2	13
Мотоблоки и мотокультиваторы, тыс. шт.	126	153	178	127	119	107
Сеялки тракторные (без туковых), шт.	2530	2351	2278	2303	2490	2052
Косилки тракторные (без косилок-измельчителей), шт.	3201	4187	4211	4021	3925	4441
Комбайны зерноуборочные, шт.	4295	6515	5798	5848	5547	4412
Комбайны силосоуборочные самоходные, шт.	268	305	890	431	240	379

2. ПРОГНОЗИРОВАНИЕ. ПРОВЕРКА ГИПОТЕЗЫ О СУЩЕСТВОВАНИИ ТРЕНДА

Прогнозирование. Классификация прогнозов • Требования, предъявляемые к исходной информации • Проверка гипотезы о существовании тенденции. Метод восходящих и нисходящих серий • Метод с использованием медианы выборки • Метод Фостера – Стюарта

2.1. Прогнозирование. Классификация прогнозов

Под **прогнозированием** принято понимать процесс разработки прогнозов, т. е. научно обоснованных описаний возможных будущих состояний объектов, а также альтернативных путей и сроков их достижения[4]. Различают поисковые и нормативные прогнозы.

Поисковые прогнозы отвечают на вопрос «к чему приведет сложившаяся тенденция развития событий?», **нормативные** – «как следует изменить исходные параметры, чтобы достичь ожидаемого результата в будущем?».

Все прогнозы можно дифференцировать по предметной области, масштабу, времени упреждения.

По предметной области различают экономические, социальные, политические, военные, научно-технические и т. д.

По масштабу прогнозирования существуют прогнозы:

- на микроуровне (уровне предприятия);
- макроуровне (уровне региона, страны);
- уровне глобальных прогнозов (в мировом масштабе).

По времени упреждения, т. е. по времени с момента последних статистических данных до момента прогнозирования выделяют прогнозы:

- оперативные (до 1 месяца);
- краткосрочные (от 1 месяца до 1 года);
- среднесрочные (от года до 5 лет);
- долгосрочные (более 5 лет).

Процесс прогнозирования включает:

▪ постановку задачи и сбор необходимой информации,

[4] Е. М. Четыркин в книге «Статистические методы прогнозирования» писал: «Прогнозирование – это научная деятельность, направленная на выявление и изучение возможных альтернатив будущего развития и структуры его вероятных траекторий» (Четыркин Е. М. Статистические методы прогнозирования. М. : Статистика,1975, С. 4)

- первичную обработку исходных данных,
- проверку гипотезы о наличии тенденции,
- выявление тренда,
- оценку параметров моделей,
- построение прогноза (точечный и интервальный),
- анализ полученного прогноза,
- корректировку полученного результата на циклическую и сезонную компоненты.

2.2. Требования, предъявляемые к исходной информации

Информация, на основе которой составляют прогноз, должна соответствовать ряду требований.

1. Необходимо обеспечить сопоставимость исходных данных по рядам динамики.

2. Рекомендуется принимать равные интервалы, чтобы признак распределился равномерно. Чем короче ретроспективный ряд, тем меньше принимаемая величина интервалов. При отсутствии промежуточной информации необходимо ее рассчитать.

3. Для изучения сезонных колебаний необходимо иметь информацию не менее чем за три года.

4. При использовании регрессионных методов анализа ретроспективный ряд должен быть в два-три раза больше количества независимых переменных.

5. Прогноз будет тем точнее, чем больше величина ретроспективного ряда, меньше период упреждения и более точно подобрана модель, описывающая тенденцию развития. Рекомендуется осуществлять прогноз на срок, который не превышает 1/3 длины анализируемого динамического ряда[5].

2.3. Проверка гипотезы о существовании тенденции. Метод восходящих и нисходящих серий

Уровни ряда динамики обычно рассматривают как результат воздействия четырех компонент, которые непосредственно не могут быть измерены: тренда, сезонной составляющей, циклической составляющей и случайной компоненты.

[5] Салин В. Н., Чурилова Э. Ю. Курс теории статистики для подготовки специалистов финансово-экономического профиля : учебник. М. : Финансы и статистика, 2007. 480 с. С. 394.

Тренд отражает тенденцию, под которой в статистической литературе понимают основное направление развития, выраженное более или менее гладкой траекторией.

Циклическая компонента отражает колебания с циклом более года.

Сезонная составляющая характеризует колебания в течение года.

Случайные компоненты могут быть результатом текущего воздействия ряда факторов или аномальных явлений.

Приведенное разделение составляющих уровня ряда динамики носит условный характер, но позволяет определить закономерности того или иного ряда динамики и может быть использовано при составлении прогноза.

К наиболее распространенным методам проверки гипотезы о наличии тренда можно отнести:

- метод нисходящих и восходящих серий;
- метод, основанный на медиане выборки;
- метод Фостера – Стюарта.

Метод восходящих и нисходящих серий

Метод восходящих и нисходящих серий включает несколько этапов:

1. Для временного ряда определяют последовательность знаков «+» и «–», исходя из следующих условий:

$$\begin{cases} \text{«+», если } y_t > y_{t-1} \\ \text{«–», если } y_t < y_{t-1} \\ \text{учитывается только одно значение, если } y_t = y_{t-1} \end{cases}$$

2. Определяют число серий в виде подряд идущих плюсов или минусов. Один плюс или один минус тоже может рассматриваться как серия.

3. Определяют протяженность самой длинной серии.

4. Проверяют гипотезу о наличии тренда, исходя из следующих условий:

$$\begin{cases} v(n) > \left[\frac{1}{3}(2n-1) - 1{,}96\sqrt{\frac{16n-29}{90}} \right] \\ \tau_{\max}(n) \le \tau_o(n) \end{cases}$$

где $v(n)$ – число серий n-го временного ряда; $\tau_{\max}$ – длина серий (протяженность самой длинной серии); n – количество членов временного ряда; τ_0 –

табличное значение, зависящее от длины временного ряда: $\tau_0 = 5$ при $n \leq 26$; $\tau_0 = 6$ при $26 < n \leq 153$; $\tau_0 = 7$ при $153 < n \leq 170$

Квадратные скобки указывают на то, что в расчет принимают только целую часть числа. Система неравенств рассчитана для уровня значимости 5 %.

Если нарушается хотя бы одно условие, то гипотеза о наличии тренда подтверждается для уровня значимости 5 %, т. е. с вероятностью 95 % можно говорить о наличии тенденции.

Пример 2.1

Проверим гипотезу о наличии тренда методом восходящих и нисходящих серий на примере динамики среднегодовой численности занятых в Центральном федеральном округе РФ за период с 2000 по 2017 гг. (табл. 2.1).

Таблица 2.1. Пример проверки гипотезы о наличии тренда методом восходящих и нисходящих серий

Год	**Среднегодовая численность занятых в ЦФО**	**Серии**
2000	18 014,4	
2001	18 209,9	+
2002	18 597,0	+
2003	18 609,9	+
2004	18 912,5	+
2005	19 159,4	+
2006	19 372,0	+
2007	19 945,8	+
2008	19 901,1	-
2009	19 471,2	-
2010	19 716,3	+
2011	20 056,9	+
2012	20 382,6	+
2013	20 309,6	-
2014	20 471,1	+
2015	20 363,3	-
2016	20 526,6	+
2017	20 591,0	+

Проверяем гипотезу о наличии тренда, исходя из следующих условий:

$$\begin{cases} \nu(n) > \left[\frac{1}{3}(2n-1) - 1{,}96\sqrt{\frac{16n-29}{90}}\right] \\ \tau_{max}(n) \le \tau_o(n) \end{cases}$$

По результатам расчётов получаем, что τ_{max} = 7 (самая длинная серия), а τ_0 (n < 26) = 5. Из этого следует, что второе условие не выполняется, так как $\tau_{max} > \tau_{0.}$

Проверим выполнение первого неравенства. В рассматриваемом ряду 18 периодов, при этом количество серий составляет v (18) = 7.

$$\left[\frac{1}{3}(2\cdot 18-1) - 1{,}96\sqrt{\frac{16\cdot 18-29}{90}}\right] = [8{,}342] = 8$$

Первое условие также не выполняется, так как v (18) < 8.

Таким образом, гипотеза о наличии тренда подтверждается, т. е. анализируемый ряд динамики содержит тенденцию с вероятностью 95 %.

2.4. Метод с использованием медианы выборки

Этапы расчета, основанного на медиане выборки:

1. Проводят ранжирование исходного временного ряда.

2. Определяют медиану этого ряда.

Для нечетного числа членов ряда как соответствующее значение $(2n + 1) / 2$ уровня, для четного – как полусумма двух серединных значений ряда динамики.

3. Формируют последовательность из плюсов и минусов по правилу

$$\begin{cases} \text{«+», если } y_t > M_e \\ \text{«–», если } y_t < M_e \\ \text{значение опускается, если } y_t = M_e \end{cases}$$

4. Определяют протяженность самой длинной серии $\tau_{max}(n)$ и общее число серий $v(n)$.

5. Проверяют гипотезу о наличии тренда.

Гипотезу о наличии тренда отвергают с вероятностью 95 %, если выполняются оба условия:

$$\begin{cases} v(n) > \left[\frac{1}{2}\left(n+1-1{,}96\sqrt{(n-1)}\right)\right] \\ \tau_{max}(n) < [3{,}3(\lg n+1)] \end{cases}$$

Если хотя бы одно неравенство нарушается, то можно говорить о наличии тенденции.

Пример 2.2

Проверим гипотезу о наличии тенденции по данным о среднегодовой численности занятых в ЦФО РФ методом с использованием медианы выборки. Расчёты серий приведены в табл. 2.2.

Так как в данном примере количество периодов в ряду чётное n = 18, то медиану рассчитывают как полусумму девятого и десятого значений ранжированного ряда: Me = (19 716,3 + 19 901,1) / 2 = 19 808,7 тыс. чел.

Сравнив с медианой значения **неранжированного** ряда, получаем совокупность знаков «+» и «–», которые можно объединить в две серии v (18). Продолжительность максимальной серии ($\tau_{max}(n)$) равна 9. Рассчитываем неравенства

$$\begin{cases} v(n) > \left[\frac{1}{2}\left(n+1-1{,}96\sqrt{(n-1)}\right)\right] \\ \tau_{max}(n) < [3{,}3(\lg n+1)] \end{cases} \qquad \begin{cases} 2 > \left[\frac{1}{2}\left(18+1-1{,}96\sqrt{(18-1)}\right)\right] \\ 9 < [3{,}3(\lg 18+1)] \end{cases}$$

$$\begin{cases} 2 > [5{,}459] \\ 9 < [7{,}442] \end{cases} \qquad \begin{cases} 2 > [5] \text{ – условие не выполняется} \\ 9 < [7] \text{ – условие не выполняется} \end{cases}$$

Таблица 2.2. Проверка гипотезы о наличии тенденции методом с использованием медианы выборки

Год	**Среднегодовая численность занятых в ЦФО**	**Ранжированный ряд**	**Серии**
2000	18 014,4	18 014,4	-
2001	18 209,9	18 209,9	-
2002	18 597,0	18 597,0	-
2003	18 609,9	18 609,9	-
2004	18 912,5	18 912,5	-
2005	19 159,4	19 159,4	-
2006	19 372,0	19 372,0	-
2007	19 945,8	19 471,2	-
2008	19 901,1	*19 716,3*	-

Окончание табл. 2.2.

Год	Среднегодовая численность занятых в ЦФО	Ранжированный ряд	Серии
2009	19 471,2	*19 901,1*	+
2010	19 716,3	19 945,8	+
2011	20 056,9	20 056,9	+
2012	20 382,6	20 309,6	+
2013	20 309,6	20 363,3	+
2014	20 471,1	20 382,6	+
2015	20 363,3	20 471,1	+
2016	20 526,6	20 526,6	+
2017	20 591,0	20 591,0	+

Результаты расчётов показали, что оба условия не выполняются, поэтому ряд изменения среднегодовой численности занятых в ЦФО имеет тенденцию, следовательно, исходные данные можно использовать для составления прогноза.

2.5. Метод Фостера – Стюарта

Данный метод был предложен Ф. Фостером и А. Стюартом и при относительной простоте, по мнению экономиста Е. М. Четыркина, дает более надежный результат, поскольку позволяет обнаружить тренд в значении дисперсии уровней[6].

В соответствии с данным методом определены следующие этапы:

1. Каждый уровень ряда сравнивают со всеми предшествующими, при этом определяют значения вспомогательных характеристик m_t и l_t:

$$m_t = \begin{cases} 1, \text{ } \textit{если } y_t > y_{t-1}, y_{t-2}, \ldots, y_1 \\ 0, \text{ } \textit{в противном случае} \end{cases}$$

$$l_t = \begin{cases} 1, \text{ } \textit{если } y_t > y_{t-1}, y_{t-2}, \ldots, y_1 \\ 0, \text{ } \textit{в противном случае} \end{cases}$$

2. Рассчитывают $d_t = m_t - l_t$, для всех $t = 2 \ldots n$
3. Определяют характеристику $D = \sum_{t=2}^{n} d_t$

[6] Четыркин Е. М. Указ. соч. С. 17.

4. Гипотезу о наличии тенденции проверяют с помощью критерия Стьюдента. Для этого определяют

$$t_{\text{набл}} = \frac{D}{\sigma_D},$$

где σ_D – средняя квадратическая ошибка величины D:

$$\sigma_D = \sqrt{2\sum_{t=2}^{n}\frac{1}{t}} \approx \sqrt{2\ln n - 0{,}8456}$$

Значения σ_D табулированы для n от 10 до 100 (табл. 2.3).

Таблица 2.3. Средняя квадратическая ошибка величины D

n	σ_D	n	σ_D	n	σ_D	n	σ_D
10	1,964	35	2,509	60	2,713	85	2,837
15	2,153	40	2,561	65	2,742	90	2,857
20	2,279	45	2,606	70	2,769	95	2,876
25	2,373	50	2,645	75	2,793	100	2,894
30	2,447	55	2,681	80	2,816	–	–

Расчетное значение $t_{\text{набл}}$ сравнивают с $t_{\text{кр}}$, взятым из таблицы t-распределения Стьюдента (прил. 1), для n – длины ряда, α – уровня значимости и $k = n - 1$ (для прямой) – степеней свободы.

5. Если $t_{\text{набл}} < t_{\text{кр}}$, то гипотеза о наличии тренда опровергается, т. е. тенденции нет. Если $t_{\text{набл}} > t_{\text{кр}}$, то гипотеза о наличии тренда подтверждается, т. е. тенденция присутствует.

Пример 2.3

Проверим гипотезу о наличии тренда в динамике среднегодовой численности занятых в ЦФО методом Фостера – Стюарта (табл. 2.4).

Таблица 2.4. Пример проверки гипотезы о наличии тенденции методом Фостера – Стюарта

Год	**Среднегодовая численность занятых в ЦФО**	m_t	l_t	d_t
2000	18 014,4	–	–	–
2001	18 209,9	1	0	1
2002	18 597,0	1	0	1
2003	18 609,9	1	0	1
2004	18 912,5	1	0	1

Окончание табл. 2.4

Год	**Среднегодовая численность занятых в ЦФО**	m_t	l_t	d_t
2005	19 159,4	1	0	1
2006	19 372,0	1	0	1
2007	19 945,8	1	0	1
2008	19 901,1	0	0	0
2009	19 471,2	0	0	0
2010	19 716,3	0	0	0
2011	20 056,9	1	0	1
2012	20 382,6	1	0	1
2013	20 309,6	0	0	0
2014	20 471,1	1	0	1
2015	20 363,3	0	0	0
2016	20 526,6	1	0	1
2017	20 591,0	1	0	1

Рассчитаем характеристику $D = \sum_{t=2}^{n} d_t = 12$

$$\sigma_D = \sqrt{2\sum_{t=2}^{n}\frac{1}{t}} \approx \sqrt{2\ln n - 0,8456}$$

$$\sigma_D = \sqrt{2\ln(18n) - 0,8456} = 2,2215$$

Рассчитаем $t_{\text{набл}}$, при $\sigma_D = 2,2215$:

$$t_{\text{набл}} = \frac{D}{\sigma D} = \frac{12}{2,2215} = 5,4017.$$

Определим $t_{\text{кр}}$ Стьюдента для $n = 18 t_{\text{кр}}$ ($\alpha = 0,05$, $k = 18 - 1$) = 2,1098 (прил. 1).

Таким образом, $t_{\text{набл}} > t_{\text{кр}}$ (5,4017 > 2,1098), т. е. гипотеза о наличии тренда подтверждается, и с вероятностью 95 % можно говорить о присутствии тенденции в динамике среднегодовой численности занятых в ЦФО.

Контрольные вопросы

1. Назовите виды прогнозов.

2. Чем поисковый прогноз отличается от нормативного?

3. Что представляет собой период упреждения?

4. От каких факторов зависит точность получаемых прогнозов?

5. В чем заключается метод восходящих и нисходящих серий и для чего он предназначен?

6. Раскройте содержание метода, основанного на медиане выборки.

7. *В чем суть метода Фостера – Стюарта?*
8. *Какие требования предъявляют к исходной информации при составлении прогнозов?*
9. *Какие компоненты формируют уровень ряда динамики?*
10. *Что представляет собой тенденция?*
11. *В чем отличие циклической составляющей от сезонной составляющей?*

Практические задания

Задача 2.1

По приведенным в табл. 2.5 данным проверить гипотезу о наличии тенденции в производстве сельскохозяйственных машин в РФ. Использовать метод восходящих и нисходящих серий.

Таблица 2.5. Производство сельскохозяйственных машин и оборудования в Российской Федерации

Наименование машин и оборудования	**Год**					
	2010	**2011**	**2012**	**2013**	**2014**	**2015**
Комбайны зерноуборочные, шт.	4295	6515	5798	5848	5547	4412
Комбайны силосоуборочные самоходные, шт.	268	305	890	431	240	379
Плуги общего назначения, шт.	1550	3719	4007	3369	2842	2993
Косилки тракторные (без косилок-измельчителей), шт.	3201	4187	4211	4021	3925	4441

Задача 2.2

По данным задачи 2.1 проверить гипотезу о наличии тенденции в производстве сельскохозяйственных машин в РФ методом с применением медианы.

Задача 2.3

По данным задачи 2.1 проверить гипотезу о наличии тенденции в производстве сельскохозяйственных машин в РФ методом Фостера – Стюарта.

Задача 2.4

По приведенным в табл. 2.6 данным проверить гипотезу о наличии тенденции в численности работников государственных органов и органов местного самоуправления на 10 000 человек постоянного населения по РФ. Использовать три метода проверки гипотезы.

Таблица 2.6. Динамика численности работников государственных органов и органов местного самоуправления на 10000 человек постоянного населения по РФ за период с 2000 по 2017 гг.

Год	2000	2001	2002	2003	2004	2005	2006	2007	2008
Численность, чел. на 10 000 чел. населения	79,4	78,3	86,4	90,2	91,9	102,4	110,9	114,4	117,7
Год	2009	2010	2011	2012	2013	2014*	2015	2016	2017
Численность, чел. на 10 000 чел. населения	118,0	115,4	112,1	109,7	107,8	153,6	148,5	146,2	147,9

* С 2014 г. в численность работников государственных органов включены сотрудники территориальных органов МВД России.

Задача 2.5

По приведенным в табл. 2.7 данным проверить гипотезу о наличии тенденции брачности в Российской Федерации за период с 2010 по 2016 гг. Использовать три метода проверки гипотезы.

Таблица 2.7. Динамика коэффициентов брачности на 1000 чел. населения Российской Федерации

Годы	2010	2011	2012	2013	2014	2015	2016
Коэффициент брачности на 1000 чел.	8,5	9,2	8,5	8,5	8,4	7,9	6,7

3. МЕТОДЫ ВЫЯВЛЕНИЯ ТЕНДЕНЦИИ С ПОМОЩЬЮ ВЫРАВНИВАНИЯ РЯДА

Метод укрупнения интервалов • Метод простой скользящей средней • Метод взвешенной скользящей средней

3.1. Метод укрупнения интервалов

Метод **укрупнения интервалов** подразумевает расчет средних значений, исходя из увеличения количества данных в каждом интервале, например в результате замены дневных интервалов недельными, декадными или месячными. При этом влияние отдельных значений уменьшается, кривая носит более сглаженный характер и тенденция проявляется в более явном виде. Рассчитанные средние записывают напротив серединного значения участка сглаживания.

Недостаток данного метода – значительное сокращение анализируемого ряда динамики. Например, при замене дневного интервала декадным или месячным ряд динамики сократится соответственно в 10 или 30 раз.

Пример 3.1

Сглаживание ряда динамики методом укрупнения интервалов

Проведём сглаживание ряда среднегодовой численности занятых в Центральном федеральном округе РФ, расчёты представим в табл. 3.1.

В рассмотренном примере полученный выровненный ряд динамики в явном виде отражает тенденцию роста показателя. В то же время при переходе от годового к трехлетнему периоду ряд динамики сократился с 18 до 6 значений.

Таблица 3.1. Среднегодовая численность занятых в ЦФО, чел.

Год	t	y_t	Расчёт	$\bar{y}_t$	Год	t	y_t	Расчёт	$\bar{y}_t$
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2000	1	18 014,4	(18 014,4+		2009	10	19 471,2	(19 471,2 +	
2001	2	18 209,9	+ 18 209,9+	18 273,8	2010	11	19 716,3	+ 19 716,3 +	19 748,1
2002	3	18 597,0	+ 18 597,0)/3 =		2011	12	20 056,9	+20 056,9)/3 =	
2003	4	18 609,9	(18 609,9 +		2012	13	20 382,6	(20 382,6 +	
2004	5	18 912,5	+18 912,5 +	18 893,9	2013	14	20 309,6	+ 20 309,6 +	20 387,8
2005	6	19 159,4	+ 19 159,4)/3 =		2014	15	20 471,1	+ 20 471,1)/3 =	
2006	7	19 372,0	(19 372,0+		2015	16	20 363,3	(20 363,3 +	
2007	8	19 945,8	+ 19 945,8 +	19739,6	2016	17	20 526,6	+ 20 526,6 +	20 493,6
2008	9	19 901,1	+ 19 901,1)/3 =		2017	18	20 591,0	+ 20 591,0)/3 =	

3.2. Метод простой скользящей средней

На практике чаще используют **метод скользящих средних**. В соответствии с данным методом ряд выравнивают с помощью средних величин, рассчитанных исходя из определенного числа звеньев ряда динамики, формирующих активный участок или участок сглаживания, который постепенно смещается по временному ряду на одно значение.

Если исходят из прямолинейной зависимости тенденции ряда динамики, то применяют сглаживание на основе **простой скользящей средней***,* например трех-, четырех- или пятизвенной, где каждый уровень ряда имеет весовой коэффициент, равный единице. Сглаженное значение записывают напротив серединного звена активного участка, поэтому при использовании скользящей средней с четным числом звеньев применяют дополнительную процедуру центрирования, т. е. нахождение средней из двух серединных сглаженных значений, либо сразу рассчитывают центрированные значения по формуле. Так для четырехзвенной простой скользящей средней центрированное значение для 3-го уровня ряда динамики можно рассчитать по формуле

$$\overline{y}_3 = 0{,}5 \times (\frac{y_1 + 2y_2 + 2y_3 + 2y_4 + y_5}{4}) .$$

При использовании активного участка с нечетным количеством звеньев ряд укорачивается на (k – 1) значение. Если активный участок включает четное число звеньев, то исходный ряд становится короче на k значений, где k – число звеньев активного участка.

Процедура **восстановления утраченных значений** включает расчет среднего абсолютного прироста активного участка и нахождение недостающих значений на основе корректировки крайних сглаженных значений. Если определяются первые члены сглаженного ряда динамики, то рассчитывают средний абсолютный прирост первого активного участка, который вычитают из первого сглаженного значения, а если восстанавливаются последние значения, то вычисляют средний абсолютный прирост последнего активного участка, который прибавляют к последнему сглаженному значению столько раз, сколько необходимо (табл. 3.2).

Следует помнить, что данный метод, как и метод укрупнения интервалов, устраняет лишь случайные колебания. Сезонные колебания сохраняются и после сглаживания. Данные методы считаются эмпирическими и не позволяют представить тенденцию в виде математической модели.

Пример 3.2

Сглаживание ряда динамики с помощью скользящей средней с нечётным числом звеньев

Выровняем ряд динамики численности занятых в ЦФО Российской Федерации с помощью трех- и пятизвенных простых скользящих средних. Результаты расчетов представлены в табл. 3.2.

При выравнивании ряда с помощью пятизвенной простой скользящей средней необходимо рассчитать среднюю величину значений первых пяти периодов и записать результат напротив третьего (серединного) периода:

$\bar{y}_3 =$ (18 014,4 + 18 209,9 + 18 597,0 + 18 609,9 + 18 912,5) / 5 = = 18 468,7 тыс. чел.

После этого рассчитываем среднее значение следующих пяти периодов с шагом в один год и т. д. (графа 3 в табл. 3.2).

После расчёта всех средних рассматриваемый исходный ряд динамики уменьшился на 4 значения (n – 1 = 5 – 1 = 4).

Восстановим первые утраченные уровни ряда. Для этого определим средний абсолютный прирост первого активного участка

$\overline{\Delta y}_{1\text{-}5} =$ (18 912,5 – 18 014,4) / 4 = 224,5 тыс. чел.

Утраченное значение второго периода

$\bar{y}_2 = \bar{y}_3 - \overline{\Delta y}_{1\text{-}5} =$ 18 468,7 – 224,5 = 18 244,2 тыс. чел.

Утраченное значение первого периода

$\bar{y}_1 = \bar{y}_2 - \overline{\Delta y}_{1\text{-}5} =$ 18 244,2 – 224,5 = 18 019,7 тыс. чел.

Восстановим последние утраченные значения. С этой целью определим средний абсолютный прирост последнего активного участка

$\overline{\Delta y}_{14\text{-}18} =$ (20 591,0 – 20 309,6) / 4 = 70,35 тыс. чел.

Утраченное значение 17-го периода

$\bar{y}_{17} = \bar{y}_{16} + \overline{\Delta y}_{14\text{-}18} =$ 20 452,3 + 70,35 = 20 522,7 тыс. чел.

Утраченное значение 18-го периода

$= \bar{y}_{17} + \overline{\Delta y}_{14\text{-}18} =$ 20 522,7 + 70,35 = 20 593,0 тыс. чел.

В графе 8 табл. 3.2 приведён пример сглаживания ряда динамики с помощью трехзвенной скользящей средней, из которого видно, что ряд сократился на два значения. Расчёт значений сглаженного ряда и восстановленных уровней аналогичен соответствующим расчётам с применением простой пятизвенной скользящей средней.

Данные таблицы свидетельствуют о том, что чем больше длина активного участка, тем меньше колебаний имеет сглаженный ряд динамики.

Таблица 3.2. Среднегодовая численность занятых в ЦФО, тыс. чел.

	y_t	**Динамика численности, сглаженная с помощью пятизвенной скользящей средней**	y_t	**Динамика численности, сглаженная с помощью четырехзвенной скользящей средней**		y_t	**Динамика численности, сглаженная с помощью трехзвенной скользящей средней**	y_t	**Динамика численности, сглаженная с помощью пятизвенной скользящей средней взвешенной**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2000	18 014,4	***18 019,7***	18 014,4		***18073,1***	18014,4	***17982,5***	18014,4	
2001	18 209,9	***18 244,2***	18 209,9		***18271,6***	18209,9	18273,8	18209,9	
2002	18 597,0	18 468,7	18 597,0	18 357,8	18470,1	18597,0	18472,3	18597,0	18491,6
2003	18 609,9	18 697,7	18 609,9	18 582,3	18701,0	18609,9	18706,5	18609,9	18696,4
2004	18 912,5	18 930,2	18 912,5	18 819,7	18916,6	18912,5	18893,9	18912,5	18881,1
2005	19 159,4	19 199,9	19 159,4	19 013,5	19180,4	19159,4	19148,0	19159,4	19127,3
2006	19 372,0	19 458,2	19 372,0	19 347,4	19471,0	19372,0	19492,4	19372,0	19489,9
2007	19 945,8	19 569,9	19 945,8	19 594,6	19633,6	19945,8	19739,6	19945,8	19841,8
2008	19 901,1	19 681,3	19 901,1	19 672,5	19715,6	19901,1	19772,7	19901,1	19830,2
2009	19 471,2	19 818,3	19 471,2	19 758,6	19772,5	19471,2	19696,2	19471,2	19611,7
2010	19 716,3	19 905,6	19 716,3	19 786,4	19846,6	19716,3	19748,1	19716,3	19676,1
2011	20 056,9	19 987,3	20 056,9	19 906,8	20011,6	20056,9	20051,9	20056,9	20080,3
2012	20 382,6	20 187,3	20 382,6	20 116,4	20210,7	20382,6	20249,7	20382,6	20295,4
2013	20 309,6	20 316,7	20 309,6	20 305,1	20343,4	20309,6	20387,8	20309,6	20407,1
2014	20 471,1	20 410,6	20 471,1	20 381,7	20399,7	20471,1	20381,3	20471,1	20381,6
2015	20 363,3	20 452,3	20 363,3	20 417,7	20452,8	20363,3	20453,7	20363,3	20441,3
2016	20 526,6	***20 522,7***	20 526,6	20 488,0	***20492,8***	20526,6	20493,6	20526,6	
2017	20 591,0	***20 593,0***	20 591,0		***20532,8***	20591,0	***20607,5***	20591,0	

Пример 3.3

Сглаживание ряда динамики с помощью простой скользящей средней с чётным числом звеньев

При сглаживании ряда по четырехзвенной простой скользящей средней на первом этапе рассчитывают средние значения для четырехзвенного активного участка и записывают между 2-м и 3-м периодами:

$$\bar{y}_{2-3} = (18\,014{,}4 + 18\,209{,}9 + 18\,597{,}0 + 18\,609{,}9) / 4 = 18\,357{,}8 \text{ тыс. чел.}$$

После этого рассчитывают среднее значение следующих четырёх периодов с шагом в один год и т. д. (графа 5 в табл. 3.2).

Для центрирования сглаженных значений находим среднее из двух соседних средних значений и записываем напротив 3-го звена активного участка:

$\bar{y}_3$ = (18357,8 + 18582,3) / 2 = 18470,1 тыс. чел.

Проверим правильность расчёта центрированного значения с помощью формулы

$$\bar{y}_3 = 0{,}5 \times \left(\frac{y_1 + 2y_2 + 2y_3 + 2y_4 + y_5}{4}\right) =$$
$$= 0{,}5 \times \left(\frac{18014{,}4 + 2*18209{,}9 + 2*18597{,}0 + 2*18609{,}9 + 18912{,}5}{4}\right) = 18470{,}1\,\textit{тыс.чел.}$$

Аналогично центрируют остальные сглаженные значения (графа 6 в табл. 3.2).

В результате преобразований исходный ряд динамики уменьшился на 4 звена (n = 4). Методика восстановления утраченных значений центрированного сглаженного ряда аналогична методике расчета, приведенной в предыдущем примере. Следует отметить, что при восстановлении утраченных значений с помощью четырехзвенной простой скользящей средней корректировке на величину среднего абсолютного прироста подвергаются центрированные значения сглаженного ряда.

3.3. Метод взвешенной скользящей средней

Простую скользящую среднюю используют, если предполагается прямолинейная зависимость. В случае криволинейной рекомендуется использовать **взвешенную скользящую среднюю**. Обычно сглаживание осуществляется на основе полиномов 2-й и 3-й степени. Каждому значению уровня ряда соответствует свой весовой коэффициент.

При сглаживании на основе параболы (полинома 2-й степени) следует учесть:

1) весовые коэффициенты симметричны относительно центрального уровня;

2) сумма всех весов с учетом общего множителя равна единице;

3) весовые коэффициенты зависят только от длины активного участка и степени полинома;

4) наличие положительных и отрицательных весов позволяет сглаженной кривой сохранять различные изгибы кривой тренда;

5) ряд динамики укорачивается на k значений и $(k - 1)$ значений соответственно для активного участка с четным и нечетным числом звеньев;

6) данный метод устраняет лишь случайные колебания, сезонные колебания сохраняются;

7) метод является эмпирическим и не позволяет представить тенденцию в аналитическом виде.

Таблица 3.3. Весовые коэффициенты сглаживания ряда динамики по полиномам 2-го и 3-го порядков для различной длины активных участков

Длина активного участка	**Весовой коэффициент**
5	$\frac{1}{35}$ [–3; +12; +17; +12; –3]
7	$\frac{1}{21}$ [–2; +3; +6; +7; +6; +3; –2]
9	$\frac{1}{231}$ [–21; +14; +39; +54; +59; +54; +39; +14; –21]

Пример 3.4

Рассчитаем среднюю взвешенную величину значений первых пяти периодов (графа 10 в табл. 3.2):

$$\bar{y}_3 = \frac{1}{35}(-3 \cdot 18\,014{,}4 + 12 \cdot 18\,209{,}9 + 17 \cdot 18\,597{,}0 + 12 \cdot 18\,609{,}9 + (-3)\ 18\,912{,}5) =$$

18 491,6 тыс. чел.

Остальные значения сглаженного ряда рассчитывают аналогично с шагом один год.

Контрольные вопросы

1. Перечислите методы выравнивания динамических рядов.

2. Каковы преимущества и недостатки метода укрупнения интервалов рядов динамики?

3. В чём заключается метод скользящей средней?

4. Каковы особенности метода с применением простой скользящей средней с четным числом звеньев?

5. В каком случае используют взвешенную скользящую среднюю и в чем особенности ее применения?

6. Как восстановить утраченные значения при выравнивании ряда динамики с помощью простой скользящей средней?

Практические задания

Задача 3.1

По приведенным в табл. 3.4 данным провести сглаживание ряда динамики плотности автомобильных дорог общего пользования с твердым покрытием по РФ и Центральному федеральному округу с помощью пятизвенной простой скользящей средней.

Таблица 3.4. Плотность автомобильных дорог общего пользования с твердым покрытием (на конец года; км путей на 1000 км^2 территории)

Год	2000	2001	2002	2003	2004	2005	2006	2007	2008
Российская Федерация	31	31	32	32	32	31	35	37	37
Центральный федеральный округ	184	185	186	187	188	179	206	214	215
Год	2009	2010	2011	2012	2013	2014	2015	2016	2017
Российская Федерация	38	39	43	54	58	60	61	62	62
Центральный федеральный округ	226	232	240	319	337	345	349	355	358

Задача 3.2

По приведенным в задаче 3.1 данным восстановить недостающие значения ряда динамики, предполагая в тенденции прямолинейную зависимость.

Задача 3.3

По приведенным в табл. 3.5 данным провести сглаживание ряда и восстановить недостающие значения ряда динамики ввода в действие жилых домов в Российской Федерации и Центральном федеральном округе с помощью трёхзвенной простой скользящей средней.

Таблица 3.5. Ввод в действие жилых домов, тыс. м^2 общей площади

Год	2000	2001	2002	2003	2004	2005	2006	2007	2008
Российская Федерация	30 296	31 703	33 832	36 449	41 040	43 560	50 552	60 989	64 058
Центральный федеральный округ	10 181	10 613	11 930	13 299	15 299	15 261	17 345	19 890	19 134
Год	2009	2010	2011	2012	2013	2014	2015	2016	2017
Российская Федерация	95 057	58 431	62 265	65 742	70 485	84 191	85 350	80 240	79 224
Центральный федеральный округ	30 387	17 461	18 064	18 179	20 255	24 547	25 550	23 982	24 285

Задача 3.4

По приведенным в табл. 3.6 данным провести сглаживание ряда динамики ввода в действие жилых домов в Российской Федерации и Центральном федеральном округе с помощью пятизвенной взвешенной скользящей средней.

Таблица 3.6. Ввод в действие жилых домов, построенных населением за счет собственных и привлеченных средств, млн $м^2$ общей площади жилых помещений

Год	2000	2001	2002	2003	2004	2005	2006	2007	2008
Российская Федерация	12,6	13,1	14,2	15,2	16,1	17,5	20	26,3	27,4
Год	2009	2010	2011	2012	2013	2014	2015	2016	2017
Российская Федерация	28,5	25,5	26,8	28,4	30,7	36,2	35,2	31,8	33,0

Задача 3.5

Провести сглаживание ряда ежеквартальной динамики ввода в действие жилых домов в Российской Федерации с помощью четырехзвенной простой скользящей средней, восстановить утраченные значения ряда.

Таблица 3.7. Ежеквартальная динамика ввода в действие жилых домов в Российской Федерации, млн $м^2$ общей площади

Квартал	**Ввод в действие жилых домов, млн $м^2$**		
	2014	**2015**	**2016**
I	14,0	18,6	15,6
II	15,7	16,1	15,9
III	18,9	17,7	18,0
IV	35,6	32,9	30,3

4. СГЛАЖИВАНИЕ РЯДОВ ДИНАМИКИ С ПРИМЕНЕНИЕМ ТЕОРЕТИЧЕСКИХ КРИВЫХ

Характеристика и методы выбора теоретических кривых • Использование метода «наименьших квадратов» для расчета параметров модели • Методы расчета прогнозных значений

4.1. Характеристика и методы выбора теоретических кривых

Аналитическое выравнивание считают наиболее совершенным методом обработки рядов динамики, поскольку позволяет подобрать аналитическую модель, наиболее точно описывающую основную линию движения показателя, и на ее основе рассчитать прогнозное значение. В соответствии с данным методом каждый уровень ряда представляют в виде суммы систематической составляющей, характеризующей тренд, и случайной компоненты, отражающей отклонения от линии тенденции. Систематическая составляющая находится на основании уравнения теоретической кривой $\hat{y}_t = f(t)$, выбранной в качестве математической модели тренда.

Аналитическое выравнивание включает:

- выбор типа теоретической кривой, которая наиболее адекватно отразит тенденцию;

- нахождение параметров уравнения выбранной теоретической кривой;

- расчет выровненных (теоретических) уровней на основе полученного уравнения.

Перед выбором конкретного типа кривой необходимо провести логический анализ анализируемого явления, т. е. ответить на следующие вопросы:

1. Возможен ли бесконечный рост уровня ряда (равномерный, равноускоренный)?

2. Есть ли точки насыщения?

3. Есть ли точки перегиба?

4. Есть ли асимптоты?

В качестве **теоретических кривых** обычно используют прямую или полином 1-й степени $\hat{y}_t = a_0 + a_1 t$,
полином 2-й степени, или параболу $\hat{y}_t = a_0 + a_1 t + a_2 t^2$,
полиномом 3-й степени, или кривую $\hat{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$,
экспоненциальную кривую $\hat{y}_t = a e^{bt}$,

логарифмическую кривую $\hat{y}_t = a_0 + a_1 \cdot \log(t)$.

При выборе теоретической кривой можно использовать результаты графического построения эмпирических данных – визуального метода отбора, метода последовательных разностей, расчета средней квадратической ошибки.

Визуальный метод в значительной мере субъективен, но наиболее прост и при относительно несложной конфигурации тенденции может дать вполне приемлемые результаты.

Метод **последовательных разностей** используют для определения степени полинома:

1) определяют разности уровней (абсолютные приросты);
2) находят разности абсолютных приростов;
3) вычисляют разности разностей абсолютных приростов и т. д.

Рассчитывают до тех пор, пока эти разности не будут примерно одинаковы. Порядок разностей определяет степень полинома.

Среднюю квадратическую ошибку находят по формуле

$$\sigma = \sqrt{\frac{\left(\sum\left(y_t - \hat{y}_t\right)^2\right)}{n}}$$

Чтобы просчитать ошибку, необходимо рассчитать фактические y_t и теоретические $\hat{y}_t$ значения. При прочих равных условиях выбирают теоретическую кривую с меньшим значением ошибки.

Использование данного метода тоже не исключает возможности ошибки, поскольку всегда можно найти кривую, которая пройдет через все точки, но вряд ли можно такую кривую считать моделью линии тренда и тем более строить на ее основе прогноз.

Экономисты предлагают следующий подход для выбора формы кривой. На первом этапе проводят содержательный анализ и отбирают типы кривых, которые наиболее близко описывают тенденции развития процесса, а на втором рассчитывают параметры теоретических кривых, величину ошибки и выбирают кривую, удовлетворяющую минимальному ее значению.

4.2. Использование метода «наименьших квадратов» для расчета параметров модели

В основе расчета параметров теоретической кривой лежит **метод «наименьших квадратов»**, обеспечивающий такое расположение теоретической кривой, при котором сумма квадратов отклонений всех фактических значений от теоретических будет минимальна:

$$\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \to \min$$

При этом для каждого вида теоретической кривой рассчитана своя **система нормальных уравнений**, на основе которых определяют коэффициенты регрессии и свободный член уравнений.

Так, для теоретической прямой $\hat{y}_t = a_0 + a_1 t$ система нормальных уравнений имеет следующий вид:

$$\begin{cases} \sum y_t = a_o + a_1 * \sum t \\ \sum y_t t = a_o \sum t + a_1 * \sum t^2 \\ t = 1,..., n \end{cases}$$

где n – количество членов ряда динамики; t – номер периода.

Расчеты можно упростить путем **переноса начала координат в середину ряда** динамики. При этом если число членов ряда нечетное, например семь периодов, то нумерация членов ряда имеет следующий вид: –3; –2; –1; 0; 1; 2; 3, а если число членов ряда четное, например шесть периодов, то нумерация членов ряда выглядит следующим образом: –5; –3; –1; 1; 3; 5. Это обстоятельство надо иметь в виду, когда выбирают номер прогнозного периода для расчета прогноза.

В любом случае $\sum t$ будет равна нулю, и значения a_0 и a_1 можно найти из уравнений

$$\begin{cases} a_0 = \sum y_t / n \\ a_1 = \sum y_t t / \sum t^2 \end{cases}.$$

Применительно к параболе 2-го порядка система нормальных уравнений имеет вид

$$\begin{cases} \sum y_t = a_0 n + a_1 \sum t + a_2 \sum t^2 \\ \sum y_t t = a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 \\ \sum y_t t^2 = a_0 \sum t^2 + a_1 \sum t^3 + a_2 \sum t^4 \end{cases},$$

$$\begin{cases} a_1 = \sum y_t t / \sum t^2 \\ a_2 = (n \sum y_t t - \sum t^2 - \sum y_t)/(n \sum t^4 - (\sum t^2)^2) \\ a_0 = \sum y_t t / n - \sum t^2 / n[(n \sum y_t t - \sum t^2 - \sum y_t) / (n \sum \mathrm{t}^4 - (\sum t^2)^2)] \end{cases}.$$

Исходя из уравнений находят параметры параболы и строят график.

Пример 4.1

Рассмотрим сглаживание ряда с помощью теоретической прямой и рассчитаем прогнозное значение среднегодовой численности занятых в ЦФО на 2020 г.

Для теоретической прямой вида $\hat{y}_t = a_0 + a_1 t$ система нормальных уравнений имеет следующий вид:

$$\begin{cases} \sum y_t = a_o * n + a_1 * \sum t \\ \sum y_t * t = a_o * \sum t + a_1 * \sum t^2 \end{cases}$$

$$\begin{cases} a_0 = \dfrac{\sum y_t - a_1 * \sum t}{n} \\ a_1 = \dfrac{n * \sum y_t t - \sum t * \sum y_t}{n * \sum t^2 - \left(\sum t\right)^2} \end{cases}$$

Для расчёта коэффициентов a_0 и a_1 используем вспомогательную табл. 4.1.

Таблица 4.1. Пример аналитического выравнивания ряда динамики на основе теоретической прямой

Год	**Среднегодовая численность занятых в ЦФО (y_t), чел.**	t	t^2	$y_t\, t$	$\hat{y}_{t\text{теор}}$
2000	18 014,4	1	1	18 014,4	18 313,8
2001	18 209,9	2	4	36 419,8	18 463,9
2002	18 597,0	3	9	55 791,0	18 613,9
2003	18 609,9	4	16	74 439,6	18 764,0
2004	18 912,5	5	25	94 562,5	18 914,1
2005	19 159,4	6	36	114 956,4	19 064,2
2006	19 372,0	7	49	135 604,0	19 214,3
2007	19 945,8	8	64	159 566,4	19 364,4
2008	19 901,1	9	81	179 109,9	19 514,4
2009	19 471,2	10	100	194 712,0	19 664,5
2010	19 716,3	11	121	216 879,3	19 814,6
2011	20 056,9	12	144	240 682,8	19 964,7
2012	20 382,6	13	169	264 973,8	20 114,8
2013	20 309,6	14	196	284 334,4	20 264,9
2014	20 471,1	15	225	307 066,5	20 414,9
2015	20 363,3	16	256	325 812,8	20 565,0
2016	20 526,6	17	289	348 952,2	20 715,1
2017	20 591,0	18	324	370 638,0	20 865,2
Итого	**352 610,6**	**171**	**2109**	**3 422 515,8**	–

Рассчитаем коэффициенты a_0 и a_1:

$$\begin{cases} 352\,610{,}6 = 18a_0 + 171a_1 \\ 3\,422\,515{,}8 = 171a_0 + 2109a_1 \end{cases}.$$

$$\begin{cases} a_0 = 18\,163{,}7 \text{ – начальный уровень выровненного ряда при } t = 0 \\ a_1 = 150{,}08 \text{ – средний ежегодный прирост численности занятых в ЦФО} \end{cases}.$$

Полученное уравнение теоретической прямой для расчёта динамики среднегодовой численности занятых в ЦФО имеет вид

$\hat{y}_t = 18\,163{,}7 + 150{,}08t$.

Подставляя в уравнение теоретической прямой значение t, рассчитываем теоретические уровни численности занятых для каждого года. Например, рассчитаем теоретические значения численности занятых для 2000 и 2017 гг.:

$\hat{y}_{2000}$ (t = 1) = 18 163,7 + 150,08 · 1 = 18 313,8 тыс. чел.

$\hat{y}_{2017}$ (t = 18) = 18 163,7 + 150,08 · 18 = 20 865,2 тыс. чел.

Пример расчета среднеквадратической ошибки для теоретической прямой приведен в табл. 4.2.

Таблица 4.2. Расчет среднеквадратической ошибки для теоретической прямой

Год	**Среднегодовая численность занятых в ЦФО (y_t), чел.**	t	$\hat{y}_t$	$(y_t - \hat{y}_t)$	$(y_t - \hat{y}_t)^2$
2000	18 014,4	1	18 313,8	–299,4	89 625,0
2001	18 209,9	2	18 463,9	–254,0	64 494,2
2002	18 597,0	3	18 613,9	–16,9	287,0
2003	18 609,9	4	18 764,0	–154,1	23 753,8
2004	18 912,5	5	18 914,1	–1,6	2,6
2005	19 159,4	6	19 064,2	95,2	9065,3
2006	19 372,0	7	19 214,3	157,7	24 878,5
2007	19 945,8	8	19 364,4	581,4	338 079,9
2008	19 901,1	9	19 514,4	386,7	149 508,7
2009	19 471,2	10	19 664,5	–193,3	37 372,3
2010	19 716,3	11	19 814,6	–98,3	9663,3
2011	20 056,9	12	19 964,7	92,2	8503,7
2012	20 382,6	13	20 114,8	267,8	71 734,3
2013	20 309,6	14	20 264,9	44,7	2002,5
2014	20 471,1	15	20 414,9	56,2	3154,7
2015	20 363,3	16	20 565,0	–201,7	40 689,2
2016	20 526,6	17	20 715,1	–188,5	35 531,7
2017	20 591,0	18	20 865,2	–274,2	75 175,4
Итого	**352 610,6**	**171**	–	–	**983 521,9**

Рассчитаем среднеквадратическую ошибку для теоретической прямой, характеризующей среднегодовую численность занятых в ЦФО:

$$\sigma = \sqrt{\frac{\left(\sum (y_t - \hat{y}_t)^2\right)}{n}} = \sqrt{\frac{983521{,}9}{18}} = 233{,}75$$

Для более точного выбора вида теоретической кривой необходимо также рассчитать её параметры и среднеквадратическую ошибку и выбрать ту кривую, у которой она ниже.

Рассчитаем прогноз численности занятых на 2020 г. ($t = 21$):

$\hat{y}_{2020} = 18\,163{,}7 + 150{,}08 \cdot 21 = 21\,315{,}4$ тыс. чел.

Графически прогнозное значение рассчитывают при построении **линии тренда,** как представлено на рис. 4.1.

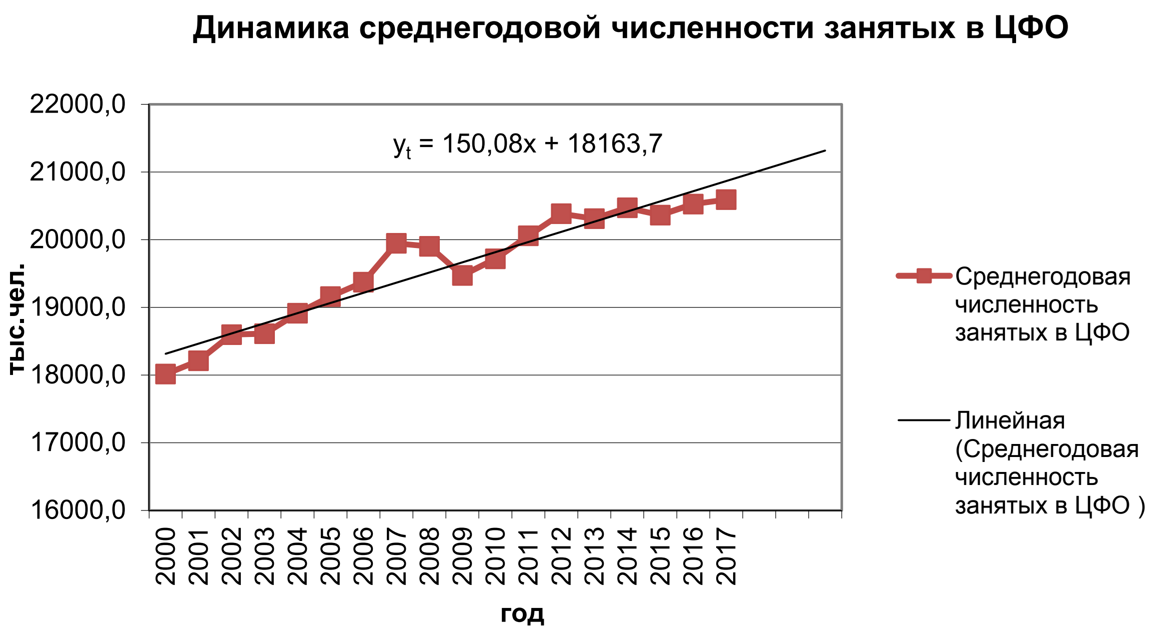


Рис. 4.1. Выравнивание ряда динамики на основе теоретической прямой

Пример 4.2

Вспомогательные расчёты для выравнивания ряда с помощью теоретической прямой методом переноса начала координат в середину временного ряда приведены в табл. 4.3.

Для упрощения расчета параметров a_0 и a_1 начало координат переносят в середину ряда динамики, т. е. между 2008 и 2009 гг., при этом коэффициенты рассчитывают по формулам

$$\begin{cases} a_0 = \sum y_t \,/\, n \\ a_1 = \sum y_t t \,/\, \sum t^2 \end{cases},$$

$$\begin{cases} a_0 = 352\ 610{,}6 \,/\, 18 \\ a_1 = 145\ 430{,}2 \,/\, 1938 \end{cases},$$

$$\begin{cases} a_0 = 19\ 589{,}5 \text{ – средний уровень выровненного ряда при } t = 0 \\ a_1 = 75{,}04 \text{ – половина среднего ежегодного прироста численности занятых} \end{cases}$$

Уравнение теоретической прямой для расчёта динамики среднегодовой численности занятых в ЦФО

$\hat{y}_t = 19\ 589{,}5 + 75{,}04t$.

Таблица 4.3. Выравнивание ряда динамики с помощью теоретической прямой методом переноса начала координат в середину временного ряда

Период	**Год**	**Среднегодовая численность занятых в ЦФО (y_t), чел.**	t	t^2	$y_t t$	$\hat{y}_{t\text{теор}}$
1	2000	18 014,4	–17	289	–306 244,8	18 313,8
2	2001	18 209,9	–15	225	–273 148,5	18 463,9
3	2002	18 597,0	–13	169	–241 761,0	18 613,9
4	2003	18 609,9	–11	121	–204 708,9	18 764,0
5	2004	18 912,5	–9	81	–170 212,5	18 914,1
6	2005	19 159,4	–7	49	–134 115,8	19 064,2
7	2006	19 372,0	–5	25	–96 860,0	19 214,3
8	2007	19 945,8	–3	9	–59 837,4	19 364,4
9	2008	19 901,1	–1	1	–19 901,1	19 514,4
10	2009	19 471,2	1	1	19 471,2	19 664,5
11	2010	19 716,3	3	9	59 148,9	19 814,6
12	2011	20 056,9	5	25	100 284,5	19 964,7
13	2012	20 382,6	7	49	142 678,2	20 114,8
14	2013	20 309,6	9	81	182 786,4	20 264,9
15	2014	20 471,1	11	121	225 182,1	20 414,9
16	2015	20 363,3	13	169	264 722,9	20 565,0
17	2016	20 526,6	15	225	307 899,0	20 715,1
18	2017	20 591,0	17	289	350 047,0	20 865,2
Итого		**352 610,6**	**0**	**1938**	**145 430,2**	–

Так же, как и в предыдущем примере, подставляя в уравнение теоретической прямой значение t, рассчитываем теоретические уровни численности занятых для каждого года. Рассчитаем теоретическое значение численности занятых в 2000 г. ($t = –17$):

$\hat{y}_{2000}$ = 19 589,5 + 75,04 (–17) = 18 313,8 тыс. чел.

Рассчитаем прогноз численности занятых на 2020 г. (t = 23):
$\hat{y}_{2020}$ = 19 589,5 + 75,04 · 23 = 21 315,4 тыс. чел.

4.3. Методы расчета прогнозных значений

Прогнозные значения можно рассчитать, продлевая линию тренда, сложившегося в анализируемый период, на период упреждения, т. е. экстраполируя линию тренда. Наиболее часто для этих целей используют средний абсолютный прирост и средний темп роста.

При составлении прогноза на основе среднего абсолютного прироста к данным последнего года анализируемого периода прибавляют абсолютный прирост показателя, умноженный на период упреждения:

$$\hat{y}_t = y_n + \bar{\Delta}\,(t - n),$$

где t – номер прогнозируемого периода; n – номер последнего периода.

При составлении прогноза на основе среднего темпа роста данные последнего года умножают на средний коэффициент роста в степени, соответствующей периоду упреждения:

$$\hat{y}_t = y_n\, \bar{k}^{(t-n)}.$$

Пример 4.1

Рассчитать прогнозное значение численности работников предприятия в 2019 г., если за последние 5 лет численность снижалась в среднем на 7 работников в год, а численность в 2017 г. составила 200 чел.

Решение: 200 чел. – 7 чел.·2 =186 чел.

Пример 4.2

Объем производства предприятия ежегодно увеличивается с примерно одинаковым темпом роста Т = 102 %. Рассчитать прогнозное значение объемов производства на 2020 г., если в 2017 г. объем производства составил 15 млн руб.

Решение: 15 млн руб.·$(1{,}02)^3$ = 15,918 млн. руб.

Контрольные вопросы

1. Какие типы кривых используют для аналитического выравнивания рядов динамики?

2. На чем основан метод «наименьших квадратов»?

3. Приведите методы отбора теоретических кривых.

4. Как нумеруют уровни ряда при переносе начала координат в середину ряда динамики?

5. Как рассчитать прогнозное значение на основе среднего абсолютного прироста?

6. Как определить прогнозное значение на основе среднего темпа роста?

Практические задания

Задача 4.1

По приведенным в табл. 4.4 данным провести сглаживание ряда динамики плотности автомобильных дорог общего пользования с твердым покрытием по одному из регионов Центрального федерального округа Российской Федерации с помощью аналитического выравнивания по теоретической прямой.

Таблица 4.4. Плотность автомобильных дорог общего пользования с твердым покрытием в ЦФО (на конец года; км путей на 1000 км^2 территории)

Регион	**Год**								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
Белгородская область	248	247	251	600	639	675	704	729	733
Брянская область	195	194	194	280	304	307	307	312	315
Владимирская область	194	216	314	315	330	332	336	340	343
Воронежская область	204	205	207	303	319	323	327	338	345
Ивановская область	183	217	248	313	333	340	333	335	334
Калужская область	301	304	307	312	318	319	321	321	323
Костромская область	92	92	92	112	127	130	133	136	136
Курская область	243	244	247	336	345	352	357	362	367
Липецкая область	253	256	264	482	507	514	520	528	533
Московская область	636	670	672	695	698	709	720	729	740
Орловская область	226	225	234	336	358	363	364	368	371
Рязанская область	185	197	203	228	258	261	264	269	269
Смоленская область	180	180	181	266	278	279	283	308	310
Тамбовская область	169	184	191	259	284	286	287	286	288
Тверская область	186	184	186	211	230	244	248	249	248
Тульская область	206	206	224	364	389	393	395	394	396
Ярославская область	195	193	196	256	262	264	265	269	270

Задача 4.2

На основе данных задачи 4.1 рассчитать прогнозное значение плотности автомобильных дорог общего пользования с твердым покрытием выбранного региона Центрального федерального округа на период упреждения, равный 5 годам. Построить графики исходного и сглаженного рядов динамики. Продлить линию тенденции на период упреждения, равный 5 годам.

Задача 4.3

По приведенным в табл. 4.5 данным провести сглаживание ряда динамики ввода в действие жилых домов по одному из регионов Центрального федерального округа Российской Федерации с помощью аналитического выравнивания по теоретической прямой. Начало координат перенести в середину ряда динамики.

Таблица 4.5. Ввод в действие жилых домов по регионам ЦФО, тыс. м^2 общей площади

Регион	**Год**								
	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Белгородская область	2718,2	2978,5	1100	1148	1215	1295	1469	1555	1350
Брянская область	458,6	520,9	391	421	453	527	551	644	665
Владимирская область	677,6	694,1	481	438	508	519	606	647	656
Воронежская область	1704,9	1392,4	1050	996	1110	1349	1573	1627	1679
Ивановская область	285,3	411,9	191	217	222	232	253	260	177
Калужская область	1010,1	800,2	501	598	613	657	808	796	737
Костромская область	228,1	259,1	151	156	205	228	328	322	309
Курская область	787,3	794,9	381	394	426	496	561	567	587
Липецкая область	1011,6	954,2	737	762	807	858	1009	1061	1078
Московская область	12119,2	11858,1	7939	8244	6620	7407	9945	9623	8914
Орловская область	612,5	508,6	249	334	359	379	469	479	360
Рязанская область	739,8	702,1	466	478	512	553	603	667	708
Смоленская область	434,2	505,3	348	371	261	411	449	514	630
Тамбовская область	769,0	747,0	569	603	636	704	771	826	833
Тверская область	535,4	700,1	452	422	410	505	538	556	496
Тульская область	511,2	564,1	395	263	310	503	580	771	623
Ярославская область	775,4	975	292	411	461	487	694	717	797

Задача 4.4

На основе данных задачи 4.3 рассчитать прогнозное значение общей площади ввода в действие жилых домов выбранного региона Центрального федерального округа на период упреждения, равный 3 годам. Построить графики исходного и сглаженного рядов динамики. Продлить линию тенденции на период упреждения, равный 3 годам.

Задача 4.5

На основе данных табл. 4.6 по одному из видов транспорта провести сглаживание ряда динамики грузооборота с помощью аналитического выравнивания, рассчитать среднеквадратическую ошибку для двух видов теоретических кривых. На основе уравнения теоретической кривой, наиболее точно отражающей тенденцию, построить график и прогноз до 2020 г.

Таблица 4.6. Грузооборот по видам транспорта РФ, млрд т · км

Вид транспорта	**Год**								
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Транспорт – всего	3 638	3 754	3 976	4 283	4 558	4 676	4 800	4 915	4 948
в том числе:									
железнодорожный	1 373	1 434	1 510	1 669	1 802	1 858	1 951	2 090	2 116
автомобильный	153	160	167	173	182	194	199	206	216
трубопроводный – всего	1 916	1 962	2 100	2 273	2 413	2 474	2 499	2 465	2 464
в том числе:									
газопроводный	1 171	1 164	1 203	1 270	1 297	1 317	1 345	1 324	1 351
нефтепроводный	718	769	867	971	1 083	1 123	1 119	1 106	1 077
нефтепродукто-проводный	27	28	30	32	33	33	35	35	36
морской	122	113	112	84	66	60	62	65	84
внутренний водный	71	83	84	81	92	87	87	86	64
воздушный	2,5	2,6	2,7	2,7	3,0	2,8	2,9	3,4	3,7
Вид транспорта	**Год**								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
Транспорт – всего	4 446	4 752	4 915	5 056	5 084	5 080	5 108	5 198	5 482
в том числе:									
железнодорожный	1 865	2 011	2 128	2 222	2 196	2 301	2 306	2 344	2 493
автомобильный	180	199	223	249	250	247	247	248	253
трубопроводный – всего	2 246	2 382	2 422	2 453	2 513	2 423	2 444	2 489	2 615
в том числе:									
газопроводный	1 123	1 259	1 302	1 265	1 289	1 203	1 176	1 181	1 300
нефтепроводный	1 087	1 084	1 083	1 152	1 182	1 178	1 226	1 262	1 265
нефтепродукто-проводный	36	39	38	36	42	42	42	46	50
морской	98	100	78	45	40	32	42	43	46
внутренний водный	53	54	59	81	80	72	64	67	67
воздушный	3,6	4,7	5,0	5,1	5,0	5,2	5,6	6,6	7,6

5. МЕТОДЫ ВЫЯВЛЕНИЯ СЕЗОННОСТИ В РЯДАХ ДИНАМИКИ

Расчеты сезонности и корректировка ряда динамики на основе индексов сезонности • Расчеты сезонности и корректировка ряда динамики на основе абсолютных приростов

5.1. Расчеты сезонности и корректировка ряда динамики на основе индексов сезонности

Сезонные колебания (сезонная неравномерность) представляют сравнительно устойчивые внутригодовые колебания. Они определяются специфическими условиями, влиянием ряда факторов, в том числе и природно-климатических.

Задача статистики состоит в выявлении и измерении колеблемости, или вариации признака. Наличие сезонных колебаний устанавливают с помощью графического метода. В этом случае применяют линейные диаграммы, на которые наносят значения показателей по месяцам не менее чем за три года.

Целесообразно для выявления сезонных колебаний использовать среднесуточные уровни за каждый месяц, что позволяет исключить влияние различной продолжительности месяцев. Эти уровни исчисляют путем деления общего объема явления за месяц на число календарных дней в месяце.

Измеряют сезонные колебания (сезонную волну) при помощи особых показателей, которые называют индексами сезонности. Обычно их применяют в мультипликативной модели сезонности, для которой характерно изменение амплитуды колебаний с течением времени.

Индексы сезонности – относительные показатели, которые отражают, во сколько раз фактический уровень ряда в момент или интервал времени t больше среднего уровня либо уровня, вычисляемого по уравнению тенденции $f(t)$.

Абсолютный прирост показывает, на сколько абсолютных единиц фактический уровень превышает или отстает от среднего уровня или уровня, вычисляемого по уравнению тенденции $f(t)$. Используют абсолютные приросты в моделях сезонности аддитивного вида, когда амплитуда колебаний показателя с течением времени не меняется.

В зависимости от характера динамики расчет индексов сезонности выполняют разными методами.

1-я ситуация. Анализируемые данные не содержат тенденции развития

В этом случае индексы сезонности определяют **по отношению к средней за весь анализируемый период** в следующей последовательности:

1. Рассчитывают средние уровни для каждого одноименного периода по данным за все годы наблюдения $\bar{y}_i$.

2. Определяют общую среднюю $\bar{y}_o$ за весь период наблюдения.

3. Вычисляют индексы сезонности по отношению к среднему уровню по формуле:

$$i_c = \frac{\bar{y}_i}{\bar{y}_o} * 100,$$

где $\bar{y}_i$ – средняя из фактических уровней одноименных месяцев; $\bar{y}_o$ – общая средняя за исследуемый период

2-я ситуация. Анализируемый ряд динамики имеет тенденцию в развитии

В данной ситуации индексы сезонности рассчитываются **по отношению к тренду.** Рекомендуется следующий порядок расчета:

1. Производят аналитическое выравнивание или сглаживание ряда динамики.

2. Исчисляют индексы сезонности для каждого периода: $i_c = \frac{\bar{y}_i}{\hat{y}_t} * 100$.

3. Рассчитывают среднюю арифметическую величину из полученных индексов:

$\bar{I}_c = \frac{\sum_{i=1}^{n} i_c}{n}$, где n – число одноименных моментов времени.

4. Рассчитывают выровненные уровни ряда с учётом сезонности.

В обоих методах для сопоставления величины сезонных колебаний по нескольким территориям или периодам может быть рассчитан **коэффициент сезонности**, исчисляемый как среднее квадратическое отклонение[7]:

$$k_S = \sigma = \sqrt{\frac{\sum (i_c - 1)^2}{n}},$$

где i_c – индекс сезонности для каждого месяца (недели, дня и т. д.); n – число месяцев (недель, дней и т.д.).

[7] Ефимова М. Р., Ганченко О. И., Петрова Е. В. Практикум по общей теории статистики : учеб. пособие. 2-е изд., перераб. и доп. М. : Финансы и статистика, 2006. 336 с. С. 229.

При этом чем меньше величина k_s, тем ниже величина сезонных колебаний.

Рассмотрим сезонные колебания в динамике ввода в действие жилых домов в Российской Федерации (табл. 5.1).

Пример 5.1

Расчет ведем, предполагая, что анализируемые данные **не имеют общей тенденции** развития. Рассчитаем средний уровень ввода в действие жилых домов по одноименным кварталам:

$\bar{y}_{I} = (14{,}0 + 18{,}6 + 15{,}6) / 3 = 16{,}07$ млн м^2, $\bar{y}_{II} = 15{,}90$ млн м^2, $\bar{y}_{III} = 18{,}20$ млн м^2, $\bar{y}_{IV} = 32{,}93$ млн м^2, $\bar{y}_{общ} = 20{,}78$ млн м^2

Индексы сезонности рассчитывают по отношению к среднему уровню за весь исследуемый период:

$i_{I} = \frac{\bar{y}_{I}}{\bar{y}_{общ}} = \frac{16{,}07}{20{,}78} = 0{,}773$, т. е. уровень ввода в действие жилых домов в I квартале по сравнению со среднегодовым уровнем ниже на 22,7 %;

$i_{II} = \frac{\bar{y}_{II}}{\bar{y}_{общ}} = \frac{15{,}90}{20{,}78} = 0{,}765$, уровень ввода в действие жилых домов во II квартале по сравнению со среднегодовым уровнем ниже на 23,5 %;

$i_{III} = \frac{\bar{y}_{III}}{\bar{y}_{общ}} = \frac{18{,}20}{20{,}78} = 0{,}876$, уровень ввода в действие жилых домов в III квартале по сравнению со среднегодовым уровнем ниже на 12,4 %;

$i_{IV} = \frac{\bar{y}_{IV}}{\bar{y}_{общ}} = \frac{32{,}93}{20{,}78} = 1{,}585$, в среднем уровень ввода в действие жилых домов в IV квартале по сравнению со среднегодовым уровнем выше на 58,5 % в период с 2014 по 2016 гг.

Таблица 5.1. Определение наличия сезонности в динамике ввода в действие жилых домов в Российской Федерации

Год	**Квартал**	**Введено общей площади, млн м^2**	**Средний уровень y_i**	**Индекс сезонности**	**$i_c - 1$**	**$(i_c - 1)^2$**
2014	I	14,0	16,07	0,773	–0,227	0,051
	II	15,7	15,90	0,765	–0,235	0,055
	III	18,9	18,20	0,876	–0,124	0,015
	IV	35,6	32,93	1,585	0,585	0,343
2015	I	18,6	16,07	0,773	–0,227	0,051
	II	16,1	15,90	0,765	–0,235	0,055

Окончание табл. 5.1

Год	**Квартал**	**Введено общей площади, млн м^2**	**Средний уровень y_i**	**Индекс сезонности**	$i_c - 1$	$(i_c - 1)^2$
	III	17,7	18,20	0,876	–0,124	0,015
	IV	32,9	32,93	1,585	0,585	0,343
2016	I	15,6	16,07	0,773	–0,227	0,051
	II	15,9	15,90	0,765	–0,235	0,055
	III	18,0	18,20	0,876	–0,124	0,015
	IV	30,3	32,93	1,585	0,585	0,343

Сезонные колебания представлены на рис. 5.1.

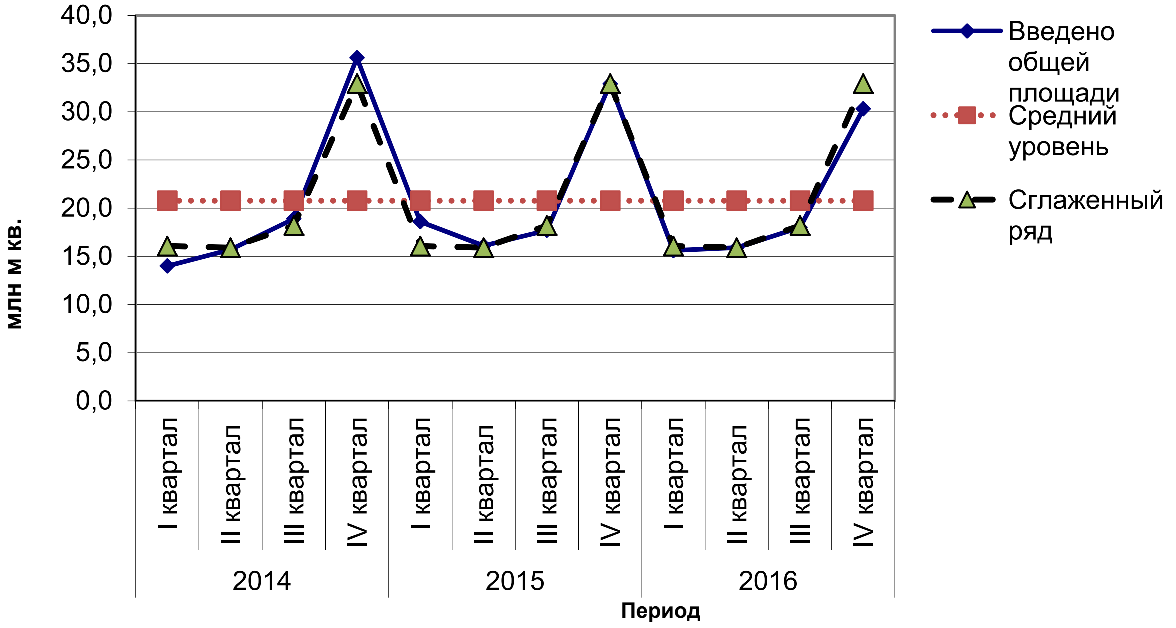


Рис. 5.1. Расчет индексов сезонности по отношению к среднегодовому уровню за период с 2014 по 2016 гг.

Рассчитаем средний коэффициент сезонности, исчисляемый как среднее квадратическое отклонение

$$k_S = \sigma = \sqrt{\frac{\sum (i_c - 1)^2}{n}} = \sqrt{\frac{1{,}393}{12}} = 0{,}341.$$

Пример 5.2

Рассчитаем индексы сезонности по средним ценам на куриные яйца по Владимирской области, предполагая **наличие тенденции в динамике** цен. Для расчётов построим вспомогательную табл. 5.2.

Для выравнивания примем линейную функцию $\hat{y}_t = a_0 + a_1 t$ и перенесём начало координат в середину ряда динамики. Параметры a_0 и a_1 определяем из системы нормальных уравнений

$$\begin{cases} a_0 = \sum y_t / n \\ a_1 = \sum y_t t / \sum t^2 \end{cases},$$

$$\begin{cases} a_0 = 65554{,}7 / 16 = 4097{,}17 \\ a_1 = 72356{,}0 / 1360 = 53{,}20 \end{cases}.$$

Отсюда уравнение теоретической прямой для расчёта принимает вид $\hat{y}_t = 4097{,}17 + 53{,}20t$.

Подставляя в данное уравнение значения t, находим выровненные значения временного ряда $\hat{y}_t$ (графа 7 табл. 5.2).

Таблица 5.2. Расчёт индексов сезонности по динамике средних цен на куриные яйца производителей Владимирской области[8]

Год	**Квартал**	**Средняя цена, руб./тыс. шт. (y_t)**	t	t^2	$y_t t$	$\hat{y}_t$	$I_{сез}$	$\overline{I_{сез}}$	**Выровненный уровень с учётом сезонности**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2013	I	3370,0	–15	225	–50 550,0	3299,1	1,021	1,086	3583,5
	II	3141,3	–13	169	–40 837,3	3405,5	0,922	0,936	3187,6
	III	3008,3	–11	121	–33 091,7	3511,9	0,857	0,873	3065,2
	IV	4429,0	–9	81	–39 861,0	3618,3	1,224	1,104	3993,7
2014	I	3609,0	–7	49	–25 263,0	3724,7	0,969	1,086	4045,8
	II	3285,7	–5	25	–16 428,3	3831,2	0,858	0,936	3586,0
	III	3675,3	–3	9	–11 026,0	3937,6	0,933	0,873	3436,7
	IV	4274,3	–1	1	–4274,3	4044,0	1,057	1,104	4463,5
2015	I	5064,7	1	1	5064,7	4150,4	1,220	1,086	4508,1
	II	4442,0	3	9	13 326,0	4256,8	1,044	0,936	3984,4
	III	3758,0	5	25	18 790,0	4363,2	0,861	0,873	3808,2
	IV	4941,0	7	49	34 587,0	4469,6	1,105	1,104	4933,3
2016	I	5189,3	9	81	46 704,0	4576,0	1,134	1,086	4970,4
	II	4310,0	11	121	47 410,0	4682,4	0,920	0,936	4382,8
	III	4022,0	13	169	52 286,0	4788,8	0,840	0,873	4179,6
	IV	5034,7	15	225	75 520,0	4895,2	1,028	1,104	5403,0
	Итого	**65554,7**	**0**	**1360**	**72 356,0**				

[8] Федеральная служба государственной статистики [Электронный ресурс]. URL: http://www.gks.ru/region/ind1117/IssWWW.exe/Stg/d100/tab3/tab3.2.htm (дата обращения: 15.01.2016).

Индексы сезонности рассчитываем как отношение уровня каждого квартала к выровненному за этот же квартал. Например, рассчитаем индекс сезонности I квартала 2013 г.

$I_I = \frac{y_I}{\hat{y}_I} = \frac{3370{,}0}{3299{,}1} = 1{,}021$, т. е. уровень средних цен на куриные яйца в I квартале 2013 г. по сравнению с теоретическим уровнем выше на 2,1 %.

Так как для четырёх лет поквартальные индексы сезонности различаются, рассчитаем среднее значение из четырехлетних данных (графа 9 табл. 5.2). Например, средний индекс сезонности I квартала равен

$$\bar{\mathrm{I}}_{\text{сез I}} = \frac{1{,}021 + 0{,}969 + 1{,}220 + 1{,}134}{4} = 1{,}086 .$$

Скорректируем выровненные уровни ряда с учётом сезонности. Для этого каждое значение $\hat{y}_t$ (графа 7 табл. 5.2) умножим на средний индекс сезонности (графа 9 таблица 5.2). Результаты расчетов приведены в графе 10 таблицы 5.2. График полученных данных представлен на рис. 5.2.

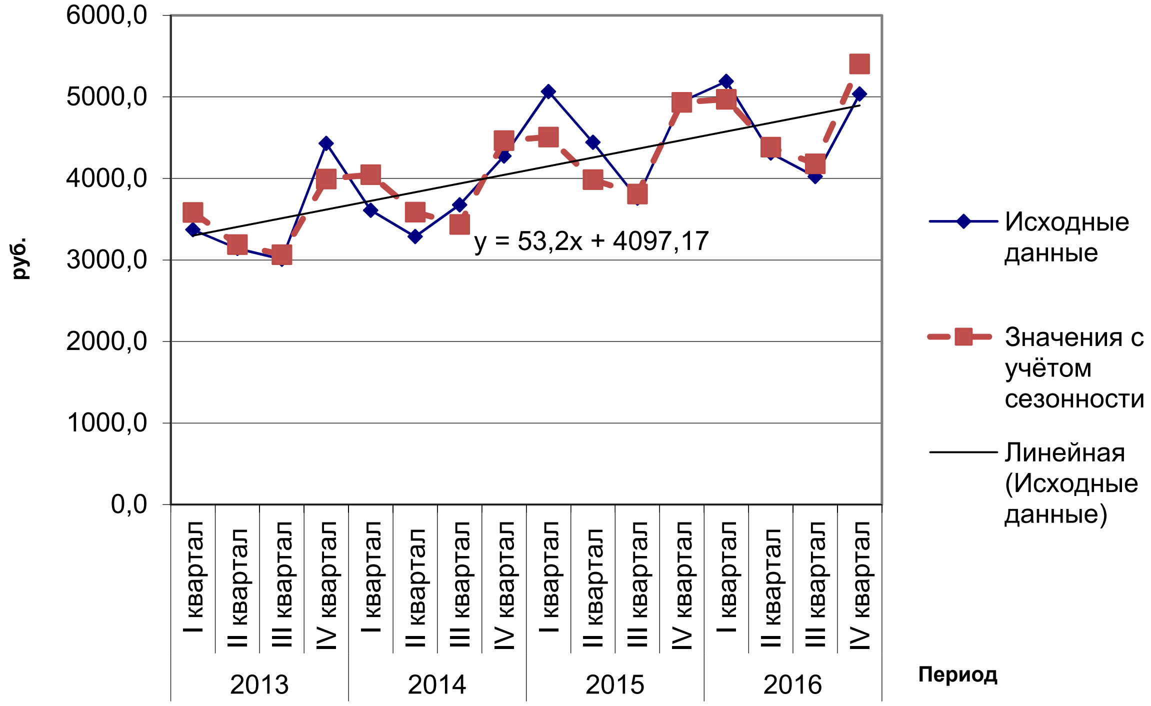


Рис. 5.2. Выявление сезонности по отношению к линии тренда

Прогнозирование в данном случае осуществляют также с учётом средних индексов сезонности.

Рассчитаем коэффициент сезонности для второго примера:

$$k_S = \sigma = \sqrt{\frac{\sum(i_c - 1)^2}{n}} = \sqrt{\frac{0,238}{16}} = 0,122\ .$$

Поскольку имеет место тенденция роста средних ежеквартальных цен на яйца во Владимирской области, то относительно тренда разброс индексов сезонности ниже, чем по отношению к среднему годовому уровню цен за четыре года.

5.2. Расчет сезонности и корректировка ряда динамики на основе абсолютных приростов

Если амплитуда сезонных колебаний с течением времени не изменяется, можно рассматривать модель с аддитивной сезонностью, определяемую на основе абсолютных отклонений.

С этой целью необходимо рассчитать средний уровень ряда за исследуемый период и абсолютные отклонения каждого уровня от рассчитанного среднего. Определяют среднее абсолютное отклонение по одноименным периодам. Выровненные значения получают суммированием среднего уровня со средним абсолютным отклонением за соответствующий период. Таким образом, корректируются прогнозные значения на сезонную составляющую при отсутствии тренда.

Если в ряду наблюдается тенденция, то определяют уравнение тренда, рассчитывают теоретические значения и определяют абсолютные отклонения фактических уровней от теоретических. Далее рассчитывают средние абсолютные отклонения по одноименным периодам и на них корректируют соответствующие прогнозные значения тренда.

Пример 5.3

Проведём корректировку на сезонную составляющую ряда динамики ввода в действие жилых домов в Российской Федерации в 2014 – 2016 гг., предполагая отсутствие тренда. Результаты расчётов представлены в таблице 5.3.

Рассчитаем абсолютные ежеквартальные отклонения от среднего уровня в графе 4 табл. 5.3. Среднеквартальный уровень введённой в действие площади жилых домов за весь период равен $\overline{у}_{общ}$ = 20,78 млн м^2.

Таблица 5.3. Расчёт сезонности динамики ввода в действие жилых домов в Российской Федерации на основе средних абсолютных отклонений

Год	**Квартал**	**Введено общей площади y_t, млн м^2**	**Абсолютное отклонение от среднего уровня, Δ**	**Среднее абсолютное отклонение от среднегодового уровня, $\overline{\Delta}$**	**Данные, выровненные с учётом сезонности, $\bar{y}_{сез}$**
1	**2**	**3**	**4**	**5**	**6**
2014	I	14,0	–6,78	–4,71	16,07
	II	15,7	–5,08	–4,88	15,90
	III	18,9	–1,88	–2,58	18,20
	IV	35,6	14,83	12,16	32,93
2015	I	18,6	–2,18	–4,71	16,07
	II	16,1	–4,68	–4,88	15,90
	III	17,7	–3,08	–2,58	18,20
	IV	32,9	12,13	12,16	32,93
2016	I	15,6	–5,18	–4,71	16,07
	II	15,9	–4,88	–4,88	15,90
	III	18,0	–2,78	–2,58	18,20
	IV	30,3	9,53	12,16	32,93

Так, для первого квартала 2013 г. абсолютное отклонение от среднего уровня составляет $\overline{\Delta_1} = 14{,}00 - 20{,}78 = -6{,}78$ млн м^2.

Рассчитаем средние отклонения среднеквартальных цен одноимённых периодов от среднегодового уровня:

$\overline{\Delta}_I = (-6{,}78 + (-2{,}18) + (-5{,}18)) / 3 = -4{,}71$ млн м^2

$\overline{\Delta}_{II} = -4{,}88$ млн м^2

$\overline{\Delta}_{III} = -2{,}58$ млн м^2

$\overline{\Delta}_{IV} = 12{,}16$ млн м^2

Рассчитаем данные, выровненные с учётом сезонности, в среднем за квартал:

$\bar{y}_{I} = 20{,}78 + (-4{,}71) = 16{,}07$ млн м^2

$\bar{y}_{II} = 20{,}78 + (-4{,}88) = 15{,}90$ млн м^2

$\bar{y}_{III} = 20{,}78 + (-2{,}58) = 18{,}20$ млн м^2

$\bar{y}_{IV} = 20{,}78 + 12{,}16 = 32{,}93$ млн м^2

Сезонные колебания представлены на рис. 5.3.

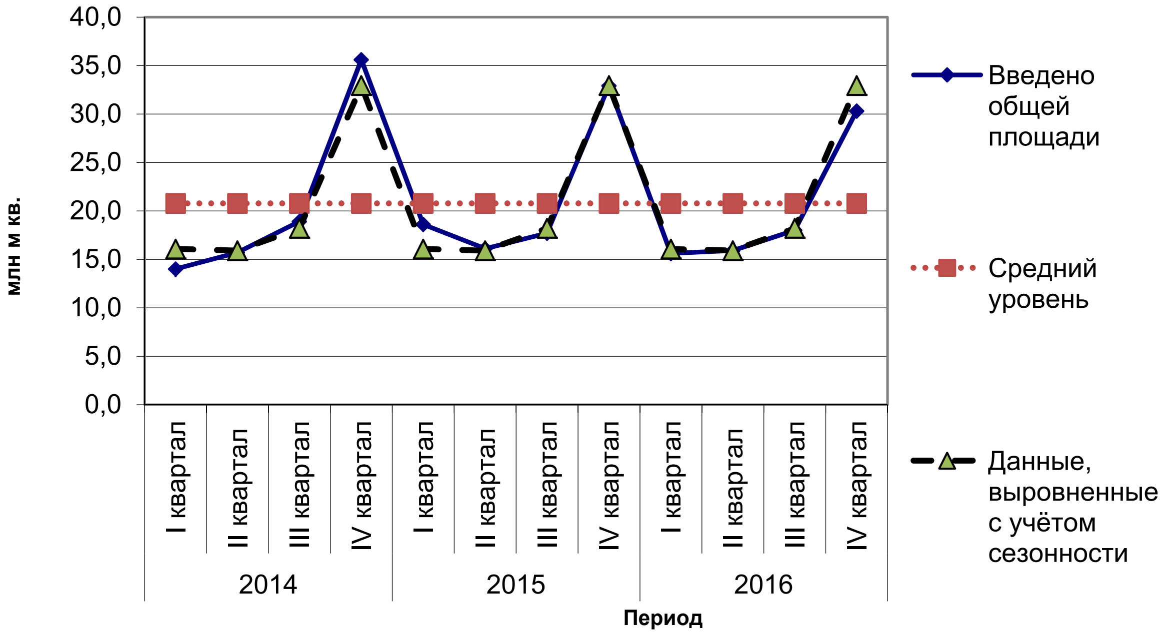


Рис. 5.3. Динамика сезонных колебаний ввода в действие жилых домов в Российской Федерации

Совершенствование подходов к изучению, выявлению и моделированию сезонных колебаний привело к изменению методики прогнозирования на основе теоретических кривых с учетом сезонной составляющей. Как отмечает Т. А. Дуброва, «сезонные и случайные колебания зачастую затрудняют анализ тенденции развития показателя, и их влияние должно быть элиминировано»[9].

С целью исключения влияния сезонности авторы предлагают:

1. Сгладить исходный ряд с помощью простой скользящей средней при четном числе звеньев на участке сглаживания для предварительной оценки тенденции развития.

2. Рассчитать отношения (разности) эмпирических значений уровней и сглаженных значений $I_{\text{сез. } t} = y_t / y_{t\,\text{сгл}}$.

3. Определить средние значения индексов (разностей) для одноименных периодов с целью исключения влияния случайной компоненты и расчёта предварительных значений сезонной составляющей:

$$\overline{I_{\text{сез}}} = (\sum I_{\text{сез. } t}) / n,$$

где n – количество одноименных периодов, например, I кварталов по всем годам.

[9] Дуброва Т. А. Прогнозирование социально-экономических процессов : учеб. пособие. 2-е изд., испр. и доп. М. : Маркет ДС, 2010. 192 с. С. 78.

4. Скорректировать исходные значения сезонной составляющей с целью нейтрализации ее влияния на динамику. Для этого необходимо рассчитать корректирующий коэффициент, так как считается, что сезонные воздействия за период взаимопогашаются, т. е. сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле. Коэффициент пересчета определяют путём деления теоретической суммы средних индексов (для четырех лет сумма индексов должна составить 4) на фактическую сумму средних индексов сезонности, полученную в результате расчетов по сглаженному ряду: $k_{кор} = \sum \overline{I_{сез}}^{теор} / \sum \overline{I_{сез}}^{факт}$

Уточненные индексы рассчитывают как произведение коэффициента пересчета на соответствующие средние индексы сезонности:

$\overline{I_{сез}}_{ут} = k_{кор} \overline{I_{сез}}$

5. Провести десезонализацию, для чего необходимо данные исходного ряда разделить на уточненный индекс сезонности:

$$y_{t\ ут} = y_t / \overline{I_{сез}}_{ут}$$

6. На основе данных десезонализированного ряда ($y_{tут}$) определить уравнение линии тренда $\widehat{y_t} = a_0 + a_1 t$ с учетом переноса начала координат в середину ряда динамики, при этом

$a_0 = (\sum y_{t\ ут}) / n$, где n – длина ряда динамики;

$a_1 = (\sum y_{t\ ут}\, t) / \sum t^2$

7. На основе уравнения линии тренда рассчитать $\widehat{y_t} = a_0 + a_1 t$.

8. Скорректировать линию тренда на уточненные индексы сезонности по формуле

$\widehat{y_t}_{сез} = \widehat{y_t} \bar{I}_{сез}$.

Пример 5.4

Рассмотрим предлагаемый алгоритм на примере построения прогнозных цен на куриные яйца во Владимирской области.

В соответствии с представленным выше расчетом сгладим исходный ряд динамики с помощью четырехзвенной простой скользящей средней (графа 4, табл. 5.4). Сглаженное значение для III квартала 2013 г. составит

$$y_{t\ сгл} = 0{,}5\left(\frac{3370 + 2 \cdot 3141{,}3 + 2 \cdot 3008{,}3 + 2 \cdot 4429 + 3609}{4}\right) = 3517{,}04.$$

Аналогично рассчитывают значения для остальных кварталов.

Определим индексы сезонности как отношение исходных значений к соответствующим сглаженным значениям.

Так, например, для III квартала 2013 г. индекс сезонности составит

$$I_{\text{сез}} = \frac{3008{,}3}{3517{,}04} = 0{,}8554.$$

Остальные значения по кварталам представлены в столбце 5 табл. 5.4.

Рассчитаем средние значения индексов по одноименным кварталам и уточним их значения.

Например, для III квартала 2013 г. средний индекс составит

$\bar{I}_{\text{сез}} = (0{,}8554 + 0{,}9441 + 0{,}8229) / 3 = 0{,}8741.$

Средние индексы по I, II и IV кварталам рассчитывают аналогично (графа 6 таблица 5.4).

Учитывая, что теоретическая сумма средних квартальных индексов за год должна оказаться равной 4, проведем корректировку средних квартальных индексов. Корректирующий коэффициент составит

$$k_{\text{кор}} = \frac{4}{0{,}8741 + 1{,}1125 + 1{,}0915 + 0{,}9355} = 0{,}9966.$$

Соответственно уточненные значения индексов сезонности по кварталам будут равны (графа 7 табл. 5.4):

для III квартала: $\bar{I}_{\text{сез. ут}}^{\,III} = 0{,}9966 \cdot 0{,}8741 = 0{,}8711$;

для IV квартала: $\bar{I}_{\text{сез. ут}}^{\,IV} = 0{,}9966 \cdot 1{,}1125 = 1{,}1087$;

для I квартала: $\bar{I}_{\text{сез. ут}}^{\,I} = 0{,}9966 \cdot 1{,}0915 = 1{,}0878$;

для II квартала: $\bar{I}_{\text{сез. ут}}^{\,II} = 0{,}9966 \cdot 0{,}9355 = 0{,}9323$.

Исключим влияние сезонности на исходный ряд динамики. С этой целью скорректируем исходный ряд на уточненные индексы сезонности. Так для I квартала 2013 г. скорректированное исходное значение на индекс сезонности составит $y^{1}_{t\,\text{ут}} = 3370 / 1{,}0878 = 3097{,}92$.

Аналогичные значения скорректированного исходного ряда представлены в графе 8 табл. 5.4.

Рассчитаем параметры уравнения линии тренда на основе скорректированных значений исходного ряда (графы 9, 10, 11 табл. 5.4).

$$a_0 = \frac{65\,573{,}21}{16} = 4098{,}33;$$

$$a_1 = \frac{72\,067{,}14}{1360} = 52{,}99.$$

Теоретические значения, рассчитанные на основе уравнения тренда, представлены в графе 13 табл. 5.4. Так, для I квартала 2013 г. теоретическое значение составит

$\widehat{y_1} = 4098{,}33 + 52{,}99 \cdot (-15) = 3303{,}47.$

Для получения значений линии тренда, скорректированных на индекс сезонности (графа 13), необходимо значения графы 12 умножить на соответствующий уточненный индекс сезонности (графа 7).

Например, скорректированное значение линии тренда для I квартала 2013 г. рассчитывают как

$\widehat{y}t_{\text{сез}} = 3303{,}47 \cdot 1{,}0878 = 3593{,}6.$

При необходимости рассчитать прогнозные значения, например для II квартала 2019 г., следует в уравнение линии тренда подставить прогнозный период и скорректировать полученное значение на уточненный индекс сезонности II квартала:

$\widehat{y}t_{\text{сез.прогноз}} = (4098{,}33 + 52{,}99 \cdot 35)\ 0{,}9323 = 5549{,}95$ руб.

Следовательно, при сохранении существующей тенденции во Владимирской области цена на тысячу штук яиц во II квартале 2019 г. составит 5549,95 руб.

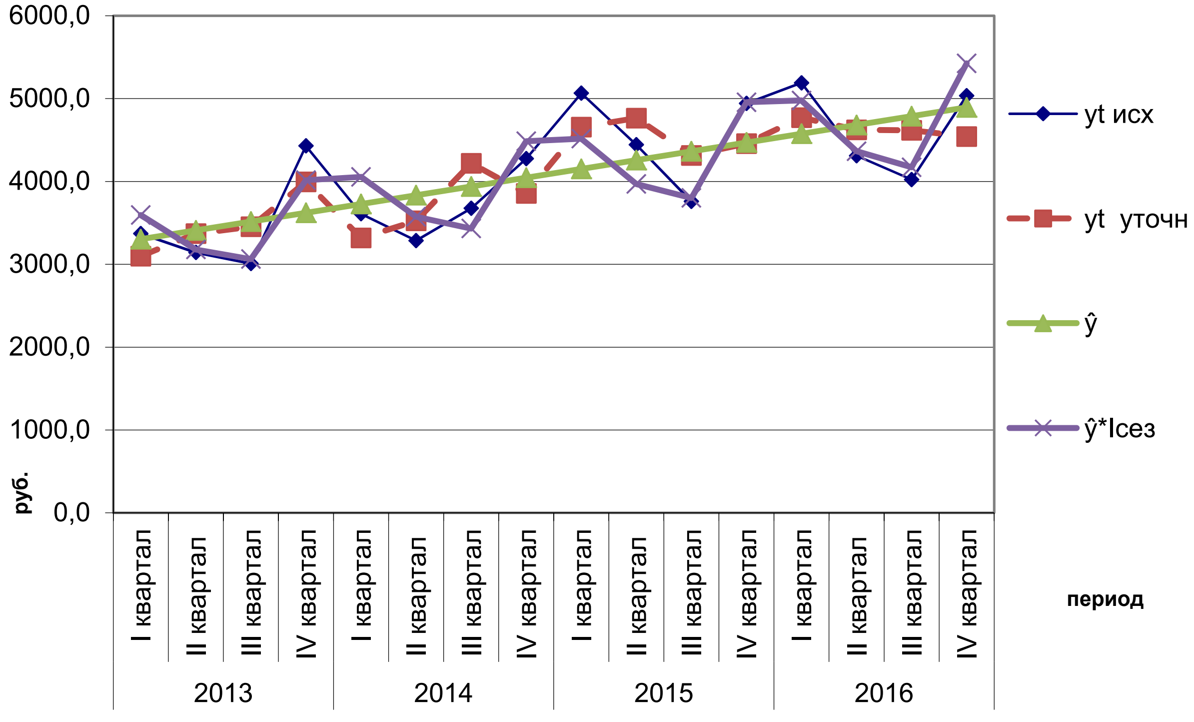


Рис. 5.4. Динамика цен на куриные яйца с выделением сезонной составляющей и линии тренда

Приведенные на рис. 5.4 графики наглядно иллюстрируют переход от исходного ряда к сглаженному, графику тренда и тренду, скорректированному на сезонную составляющую.

Таблица 5.4. Расчет прогнозных значений с учетом индексов сезонности на основе десезонализированного ряда динамики

Год	**Квартал**	**Средняя цена (y_t), руб./тыс. шт**	$y_{t\ \text{сгл}}$	$I_{\text{сез}}$	$\bar{I}_{\text{сез}}$	$\bar{I}_{\text{сез.ут}}$	$y_{t\ \text{ут}}$	t	t^2	$y_t t$	$\hat{y}_t$	$\hat{y}_{t\ \text{сез}}$
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
2013	I	3370,0	–	–	–	1,0878	3097,92	–15	225	–46 468,74	3303,47	3593,60
	II	3141,3	–	–	–	0,9323	3369,46	–13	169	–43 802,95	3409,45	3178,62
	III	3008,3	3517,04	0,8554	0,8741	0,8711	3453,31	–11	121	–37 986,37	3515,43	3062,45
	IV	4429,0	3564,96	1,2424	1,1125	1,1087	3994,66	–9	81	–35 951,95	3621,41	4015,17
2014	I	3609,0	3666,38	0,9844	1,0915	1,0878	3317,62	–7	49	–23 223,34	3727,39	4054,76
	II	3285,7	3730,42	0,8808	0,9355	0,9323	3524,27	–5	25	–17 621,36	3833,37	3573,84
	III	3675,3	3893,04	0,9441	0,8741	0,8711	4218,97	–3	9	–12 656,90	3939,35	3431,75
	IV	4274,3	4219,54	1,0130	1,1125	1,1087	3855,16	–1	1	–3855,16	4045,33	4485,18
2015	I	5064,7	4374,42	1,1578	1,0915	1,0878	4655,76	1	1	4655,76	4151,32	4515,92
	II	4442,0	4468,08	0,9942	0,9355	0,9323	4764,58	3	9	14 293,74	4257,30	3969,06
	III	3758,0	4567,00	0,8229	0,8741	0,8711	4313,86	5	25	21 569,30	4363,28	3801,05
	IV	4941,0	4566,08	1,0821	1,1125	1,1087	4456,45	7	49	31 195,15	4469,26	4955,20
2016	I	5189,3	4582,58	1,1324	1,0915	1,0878	4770,36	9	81	42 933,25	4575,24	4977,07
	II	4310,0	4627,29	0,9314	0,9355	0,9323	4622,99	11	121	50 852,93	4681,22	4364,29
	III	4022,0	–	–	–	0,8711	4616,91	13	169	60 019,81	4787,20	4170,35
	IV	5034,7	–	–	–	1,1087	4540,93	15	225	68 113,97	4893,18	5425,22
	Итого	**65 554,7**	–	–	–	–	**65 573,21**	–	**1360**	**72 067,14**	**–**	**–**

Контрольные вопросы

1. Как рассчитывают индексы сезонности при отсутствии тренда?

2. Как рассчитывают индексы сезонности при наличии тренда?

3. Каким образом корректируют точечный прогноз на основе индексов сезонности и абсолютных показателей?

4. Как рассчитывают коэффициент сезонности?

5. Сколько необходимо расчетных периодов для выявления сезонности?

Практические задания

Задача 5.1

По приведенным в табл. 5.5 данным сгладить ряд с помощью четырехзвенной простой скользящей средней и исключить сезонную составляющую из уровней исходного ряда 2012 – 2016 гг.

Таблица 5.5. Средние цены реализации сельскохозяйственной продукции (зерновые и зернобобовые культуры) сельхозпроизводителями всех категорий в Российской Федерации, руб.

Год	2003	2004	2005	2006	2007	2008	2009
I квартал	1599,06	3383,08	2953,88	2632,81	3501,46	5403,79	4509,00
II квартал	1906,24	4341,65	2783,35	3197,97	3847,08	6816,60	4624,66
III квартал	2294,28	2822,23	2371,17	2893,18	4867,80	4971,86	4482,12
IV квартал	2615,15	2864,37	2412,83	3189,47	4876,44	4444,85	4042,23
Год	2010	2011	2012	2013	2014	2015	2016
I квартал	3772,85	6445,71	5001,16	8579,53	6396,28	8588,4	9535,37
II квартал	3730,41	6237,48	5262,59	9074,17	7569,11	8685,2	10104,88
III квартал	3955,71	4944,87	6777,76	6265,82	6259,10	8250,89	8416,39
IV квартал	4714,21	5098,26	7625,55	6156,03	6674,95	9007,56	8438,69

Задача 5.2

По данным десезонализированного ряда цен (задача 5.1) рассчитать параметры уравнения линии тренда и скорректировать теоретические значения на индексы сезонности. На основе рассчитанных параметров уравнения линии тренда рассчитать прогнозное значение с учетом сезонной компоненты на IV квартал 2018 г.

Задача 5.3

По данным табл. 5.5 сгладить ряд с помощью четырехзвенной простой скользящей средней и исключить сезонную составляющую, носящую аддитивный характер из уровней исходного ряда с 2004 по 2009 гг.

Задача 5.4

По данным десезонализированного ряда цен (задача 5.1) рассчитать параметры уравнения линии тренда и скорректировать теоретические значения на показатели сезонности, носящие аддитивный характер.

Задача 5.5

По приведенным в табл. 5.6 данным сгладить ряд с помощью четырехзвенной простой скользящей средней и исключить сезонную составляющую из уровней исходного ряда 2014 – 2017 гг.

По данным десезонализированного ряда цен рассчитать параметры уравнения линии тренда и скорректировать теоретические значения на индексы сезонности. На основе рассчитанных параметров уравнения линии тренда рассчитать прогнозное значение с учетом сезонной компоненты на 3 месяца IV квартала 2018 г.

Таблица 5.6. Средние цены производителей Владимирской области на молоко сырое крупного рогатого скота, руб. за т

Год	**Месяц**					
	Январь	**Февраль**	**Март**	**Апрель**	**Май**	**Июнь**
2014	22 866	23 087	23 482	23 718	23 158	21 941
2015	23 714	24 075	24 324	24 081	23 463	22 356
2016	24 389	24 432	24 077	23 909	23 217	22 517
2017	26 755	26 991	27 377	26 910	26 369	25 611
Год	**Месяц**					
	Июль	**Август**	**Сентябрь**	**Октябрь**	**Ноябрь**	**Декабрь**
2014	20 762	20 446	21 043	22 177	23 525	23 907
2015	21 980	22 075	22 642	23 076	23 890	24 270
2016	22 069	22 043	23 248	24 575	25 282	26 215
2017	24 760	24 931	25 497	26 507	27 475	27 630

6. ОЦЕНКА ТОЧНОСТИ И АДЕКВАТНОСТИ ПРОГНОЗОВ

Расчет доверительных интервалов прогнозов • Оценка точности прогнозов • Оценка адекватности прогнозов

6.1. Расчет доверительных интервалов прогнозов

Для определения прогнозного значения в уравнение тренда подставляют значение t, соответствующее периоду упреждения продолжая, тем самым тенденцию, выявленную в анализируемом ряде динамики. Поскольку прогноз содержит только одно значение, он называется **точечным**.

В дополнение к точечному прогнозу рекомендуют рассчитывать **прогноз интервальный**, который определяет разброс возможных изменений прогнозируемого показателя. Несовпадения фактических значений с прогнозным могут быть связаны с неправильным выбором кривой, а также погрешностями при оценивании параметров кривой или отклонениями отдельных уровней от линии тренда. Погрешности могут быть отражены в виде **доверительного интервала** прогноза. Доверительный интервал, учитывающий неопределенность, связанную с положением тренда, и возможность отклонения от этого тренда, рассчитывают по формуле

$$\hat{y}_{n+L} \pm t_{\alpha} \cdot S_p,$$

где n – длина временного ряда; L – период упреждения; $\hat{y}_{n+L}$ – точечный прогноз на момент $n + L$; t_α – значение t – статистики Стьюдента; S_p – средняя квадратическая ошибка прогноза.

Рассмотрим расчет доверительного интервала для ряда, содержащего линейный тренд:

$$\hat{y}_t = a_0 + a_1 t .$$

Так как оценки параметров определяют по выборочной совокупности, представленной временным рядом, то они содержат погрешность. При этом погрешность параметра a_0 приводит к вертикальному сдвигу прямой, погрешность параметра a_1 – к изменению угла наклона прямой относительно оси абсцисс.

С учетом разброса конкретных реализаций относительно линии тренда, дисперсию S_P^2 можно представить в виде

$$S_p^2 = \frac{S_y^2}{n} + S_y^2 \frac{\left(t_1 - \bar{t}\right)^2}{\sum_{t=1}^{n}\left(t - \bar{t}\right)^2} + S_y^2,$$

где S_y^2 – дисперсия отклонений фактических наблюдений от расчетных; t_1 – время упреждения, для которого делается экстраполяция; $t_1 = n + L$; t – порядковый номер уровней ряда, $t = 1, 2, \ldots, n$; $\bar{t}$ – порядковый номер уровня, стоящего в середине ряда; $\bar{t} = (n+1):2$.

Тогда доверительный интервал можно представить в виде

$$\hat{y}_{n+L} \pm S_y * t_\alpha \sqrt{\frac{n+1}{n} + \frac{(t_1 - \bar{t})^2}{\sum_{t=1}^{n}(t - \bar{t})^2}} \quad .$$

Обозначим корень в данном выражении через K.

Значение K зависит только от длины ряда n и периода упреждения L, поэтому можно составить таблицы значений K или $K^* = t_\alpha K$. Тогда интервальная оценка будет иметь вид

$$\hat{y}_{n+L} \pm S_y \cdot K^*.$$

Аналогичное выражение можно получить для полинома 2-го порядка:

$$\hat{y}_{n+L} \pm t_\alpha \cdot S_y \sqrt{1 + \frac{t_1^2}{\sum t^2} + \frac{\sum t^4 - (2\sum t^2)t_1^2 + nt_1^4}{n\sum t^4 - (\sum t^2)^2}}$$

или $\hat{y}_{n+L} \pm S_y \cdot K^*$.

Дисперсию отклонений фактических наблюдений от расчетных можно представить формулой

$$S_y^2 = \frac{\sum_{t=1}^{n}(y_t - \bar{y}_t)^2}{n-k} \quad ,$$

где y_t – фактические значения уровней ряда; $\bar{y}_t$ – расчетные значение уровней ряда; n – длина временного ряда; k – число оцениваемых параметров выравнивающей кривой.

Следовательно, ширина доверительного интервала будет тем больше, чем меньше уровень значимости, больше период упреждения, среднее квадратическое отклонение от тренда и степень полинома.

В табл. 6.1 приведены значения K^* в зависимости от длины временного ряда n и периода упреждения L для прямой и параболы. Очевидно, что при увеличении длины рядов n значения K^* уменьшаются, с ростом периода упреждения L значения K^* увеличиваются. При этом влияние периода упреждения неодинаково для различных значений n: чем больше длина ряда, тем меньшее влияние оказывает период упреждения L.

Таблица 6.1. Значения K^* для оценки доверительных интервалов прогноза на основе линейного и параболического трендов при доверительной вероятности 0,95

Длина ряда (*n*)	**Линейный тренд**			**Длина ряда *n***	**Параболический тренд**		
	Период упреждения *L*				**Период упреждения *L***		
	1	**2**	**3**		**1**	**2**	**3**
7	2,6380	2,8748	3,1399	7	3,948	5,755	8,152
8	2,4631	2,6391	2,8361	8	3,459	4,754	6,461
9	2,3422	2,4786	2,6310	9	3,144	4,124	5,408
10	2,2524	2,3614	2,4827	10	2,926	3,695	4,698
11	2,1827	2,2718	2,3706	11	2,763	3,384	4,189
12	2,1274	2,2017	2,2836	12	2,636	3,148	3,808
13	2,0837	2,1463	2,2155	13	2,536	2,965	3,516
14	2,0462	2,1000	2,1590	14	2,455	2,830	3,286
15	2,0153	2,0621	2,1131	15	2,386	2,701	3,100
16	1,9883	2,0292	2,0735	16	2,330	2,604	2,950
17	1,9654	2,0015	2,0406	17	2,280	2,521	2,823
18	1,9455	1,9776	2,0124	18	2,238	2,451	2,717
19	1,9280	1,9568	1,9877	19	2,201	2,391	2,627
20	1,9117	1,9375	1,9654	20	2,169	2,339	2,549
21	1,8975	1,9210	1,9461	21	2,139	2,293	2,481
22	1,8854	1,9066	1,9294	22	2,113	2,252	2,422
23	1,8738	1,8932	1,9140	23	2,090	2,217	2,371
24	1,8631	1,8808	1,8998	24	2,069	2,185	2,325
25	1,8538	1,8701	1,8876	25	2,049	2,156	2,284

6.2. Оценка точности прогнозов

Качество выбранной модели можно оценить с помощью показателей точности и адекватности. Точность прогноза может характеризоваться величиной его ошибки, т. е. расхождением между фактическим и прогнозным значениями.

Для оценки используют абсолютную, относительную и среднюю ошибки.

Абсолютную ошибку прогноза можно рассчитать по формуле

$$\Delta t = \hat{y}_t - y_t,$$

где $\hat{y}_t$ – прогнозное значение показателя; y_t – фактическое значение.

Для определения относительной ошибки прогноза используют формулу

$$\delta_t = \frac{\hat{y}_t - y_t}{y_t} \cdot 100.$$

На основе абсолютных и относительных ошибок рассчитываются средние ошибки по модулю соответственно по формулам:

$$\left|\overline{\Delta}\right| = \frac{\sum_{t=1}^{n} \left|\hat{y}_t - y_t\right|}{n};$$

$$\left|\overline{\delta}\right| = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \cdot 100.$$

где n – число уровней ряда динамики, для которого рассчитывают прогноз.

Если абсолютная и относительная ошибки имеют положительные значения, то прогнозную оценку считают завышенной, если отрицательные, то говорят о заниженной прогнозной оценке.

Сравнительную точность моделей можно характеризовать величиной дисперсии (S^2) или среднеквадратической ошибки прогноза (S):

$$S^2 = \frac{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2}{n}; \; S = \sqrt{S^2}.$$

Точность модели тем выше, чем меньше полученные значения. Поскольку для проверки точности модели необходимы фактические значения прогнозных показателей, то поступают следующим образом.

Анализируемый ряд динамики делят на две части: по первой рассчитывают уравнение линии тренда, а по второй проверяют точность полученной модели.

Следует иметь в виду, что единичное совпадение точечного прогноза не дает основания для вывода о точности полученной модели. Лишь по совокупности сопоставлений прогнозных значений с их фактическими значениями можно делать такие выводы. Например, при расчете относительного числа случаев подтверждения прогнозов (μ) путем деления числа прогнозов, доказанных фактическими данными (*p*), на число прогнозов подтвержденных (*p*) и не доказанных фактическими данными (*q*). При этом необходимо соблюдать равенство доверительных вероятностей, если коэффициенты сопоставляются для разных моделей.

6.3. Оценка адекватности прогнозов

Считается, что **модель адекватно описывает процесс**, если несистематическая составляющая (e_t) уравнения удовлетворяет условию случайности, независимости и подчиняется закону нормального распределения. Если модель выбрана неудачно, то отклонения от тренда не могут носить случайный характер, в силу возможной корреляции между собой. В этом случае констатируют **автокорреляцию** ошибок, которая снижает надежность полученных результатов.

Из ряда методов обнаружения автокорреляции наибольшее распространение получил **предложенный Дарбиным и Уотсоном**. Данный метод связан с проверкой гипотезы о существовании автокорреляции между соседними остаточными членами ряда $(e_t - e_{t-1})$.

Этот критерий *d* рассчитывают по формуле

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

Приближенно величина *d* равна

$$d \approx 2(1 - r_1),$$

где r_1 – коэффициент автокорреляции первого порядка (т. е. парный коэффициент корреляции между двумя рядами $e_1, e_2, \dots, e_{n-1}$ и $e_2, e_3, \dots, e_n$).

Если имеет место полная корреляция, то есть r_1 равен 1, или –1, то *d* будет равен соответственно 0 или 4.

В случае отсутствия автокорреляции, когда r_1 равен 0, то *d* будет равен 2.

Для проверки гипотезы об отсутствии автокорреляции в остаточных величинах фактическое значение d сравнивают с табличными значениями критических границ критерия d_1 и d_2.

Таблица 6.2. Значения критерия Дарбина – Уотсона d_1 и d_2 при уровне значимости 5 %

n	**K' = 1**		**K' = 2**		**K' = 3**	
	d_1	d_2	d_1	d_2	d_1	d_2
15	1,08	1,36	0,95	1,54	0,82	1,75
16	1,1	1,37	0,98	1,54	2,86	1,73
17	1,13	1,38	1,02	1,54	0,9	1,71
18	1,16	1,39	1,05	1,53	0,93	1,69
19	1,18	1,4	1,08	1,53	0,97	1,68
20	1,2	1,41	1,1	1,54	1	1,68
21	1,22	1,42	1,13	1,54	1,03	1,67
22	1,24	1,43	1,15	1,54	1,05	1,66
23	1,26	1,44	1,17	1,54	1,08	1,66
24	1,27	1,45	1,19	1,55	1,1	1,66
25	1,29	1,45	1,21	1,55	1,12	1,66
26	1,3	1,46	1,22	1,55	1,14	1,65
27	1,32	1,47	1,24	1,56	1,16	1,65
28	1,33	1,48	1,26	1,56	1,18	1,65
29	1,34	1,48	1,27	1,56	1,2	1,65
30	1,35	1,49	1,28	1,57	1,21	1,65
31	1,36	1,5	1,3	1,57	1,23	1,65
32	1,37	1,5	1,31	1,57	1,24	1,65
33	1,38	1,51	1,32	1,58	1,26	1,65
34	1,49	1,51	1,33	1,58	1,27	1,65
35	1,4	1,52	1,34	1,58	1,28	1,65
36	1,41	1,52	1,35	1,59	1,29	1,65

В табл. 6.2. приведены критические значения для разного числа наблюдений и числа независимых переменных в уравнении регрессии и уровня значимости 5 %. В этой таблице d_1 и d_2 – соответственно нижняя и верхняя доверительные границы критерия Дарбина – Уотсона; K' – число переменных в модели; n – длина временного ряда. В табл. 6.3 приведены возможные соотношения показателей. При проверке отрицательной автокорреляции с критическими значениями сравнивается не сам коэффициент d, а $(4 - d)$.

Таблица 6.3. Варианты соотношений расчетного значения d
с критическими значениями доверительных границ d_1 и d_2

Соотношение показателей	**Вывод о принятии или отклонении гипотезы об отсутствии автокорреляции**
$d < d_1$	Гипотеза об отсутствии автокорреляции отвергается
$d > d_2$	Гипотеза об отсутствии автокорреляции принимается
$d_1 \geq d \leq d_2$	Нет достаточных оснований для принятия решений, величина находится в области неопределенности
$d > 2$	Существует отрицательная автокорреляция

Другое условие адекватности модели – **соответствие распределения остатков нормальному распределению**. Поскольку анализируемые ряды динамики экономических показателей, как правило, небольшие (< 50), то проверку распределения на нормальность обычно проводят лишь приближенно, например на основе исследования показателей асимметрии и эксцесса.

Показатели асимметрии (А) и эксцесса (Э) при нормальном распределении равны нулю. Предполагая, что отклонения от тренда представляют собой выборку из некоторой генеральной совокупности, можно рассчитать выборочные характеристики асимметрии

$$A = \frac{\frac{1}{n}\sum_{t=1}^{n} e_t^3}{\sqrt{\left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right)}}$$

и эксцесса

$$Э = \frac{\frac{1}{n}\sum_{t=1}^{n} e_t^4}{\sqrt{\left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right)}} - 3,$$

а также их среднеквадратические ошибки:

$$\sigma_A = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}};$$

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

где А – выборочная характеристика асимметрии; Э – выборочная характеристика эксцесса; σ_A – среднеквадратическая ошибка выборочной характеристики асимметрии; $\sigma_Э$ – среднеквадратическая ошибка выборочной характеристики эксцесса.

Если одновременно выполняются следующие неравенства:

$$\left.\begin{array}{l} |A| < 1{,}5\sigma_A; \\ \left|Э + \dfrac{6}{n+1}\right| < 1{,}5\sigma_Э \end{array}\right\}$$

, то гипотезу о нормальном характере распределения случайной компоненты не отвергают.

Если выполняется хотя бы одно из неравенств

$$\left.\begin{array}{l} |A| \geq 2\sigma_A \\ \left|Э + \dfrac{6}{n+1}\right| \geq 2\sigma_Э \end{array}\right\}$$

, то гипотезу о нормальном характере распределения отвергается.

Другие случаи требуют дополнительной проверки с помощью более точных критериев.

Пример 6.1

Проверим точность и адекватность линейной модели, заложенной в качестве линии тренда в примере о динамике численности занятых по Центральному федеральному округу, рассмотренном в табл. 4.3 раздела 4 (вспомогательная табл. 6.4).

Критерий Дарбина – Уотсона по данным таблицы составит

$$d = \frac{867541{,}65}{983521{,}9} = 0{,}882.$$

При сравнении расчетного значения d с табличными для $n = 18$, $\alpha = 0{,}05$, с объясняющей переменной в модели получаем, что $d < d_1 < d_2$ поскольку $0{,}882 < 1{,}16 < 1{,}39$, что с вероятностью 95 % подтверждает наличие положительной автокорреляции и свидетельствует о неадекватности линейной модели для данного ряда динамики.

Расчеты средних модулей абсолютных и относительных отклонений говорят о высокой точности расчетов. Так средняя абсолютная ошибка равна

$$|\bar{\Delta}| = \frac{3364{,}03}{18} = 186{,}9,$$

а средняя относительная ошибка составила

$|\bar{\partial}| = \frac{0{,}1715}{18} 100\ \% = 0{,}95\ \%$, что значительно меньше 10 % предельного значения для результатов высокой точности.

Таблица 6.4. Расчеты точности и адекватности модели ряда динамики

Год	**Среднегодовая численность y_t занятых в ЦФО, чел.**	t	$\hat{y}_t$	$e_t = y_t - \hat{y}_t$	e_t^2	$(e_t - e_{(t-1)})^2$	$\lvert e_t \rvert$	$\left\lvert \frac{e_t}{y_t} \right\rvert$
2000	18 014,4	–17	18 313,8	–299,4	89 625,0	–	299,37	0,0166
2001	18 209,9	–15	18 463,9	–254,0	64 494,2	2062,73	253,96	0,0139
2002	18 597,0	–13	18 613,9	–16,9	287,0	56 177,17	16,94	0,0009
2003	18 609,9	–11	18 764,0	–154,1	23 753,8	18 819,11	154,12	0,0083
2004	18 912,5	–9	18 914,1	–1,6	2,6	23 261,51	1,61	0,0001
2005	19 159,4	–7	19 064,2	95,2	9065,3	9373,58	95,21	0,0050
2006	19 372,0	–5	19 214,3	157,7	24 878,5	3908,40	157,73	0,0081
2007	19 945,8	–3	19 364,4	581,4	33 8079,9	179 536,29	581,45	0,0292
2008	19 901,1	–1	19 514,4	386,7	149 508,7	37 940,33	386,66	0,0194
2009	19 471,2	1	19 664,5	–193,3	37 372,3	336 380,01	193,32	0,0099
2010	19 716,3	3	19 814,6	–98,3	9663,3	9028,27	98,30	0,0050
2011	20 056,9	5	19 964,7	92,2	8503,7	36 296,82	92,22	0,0046
2012	20 382,6	7	20 114,8	267,8	71 734,3	30 841,41	267,83	0,0131
2013	20 309,6	9	20 264,9	44,7	2002,5	49 765,92	44,75	0,0022
2014	20 471,1	11	20 414,9	56,2	3154,7	130,35	56,17	0,0027
2015	20 363,3	13	20 565,0	–201,7	40 689,2	66503,52	201,72	0,0099
2016	20 526,6	15	20 715,1	–188,5	35 531,7	174,70	188,50	0,0092
2017	20 591,0	17	20 865,2	–274,2	75 175,4	7341,54	274,18	0,0133
Итого	**352 610,6**	**0**	**352 610,6**	**–**	**983 521,9**	**867 541,65**	**3364,03**	**0,1715**

Контрольные вопросы

1. Какие факторы влияют на ширину доверительного интервала?

2. Какая модель может считаться адекватной?

3. С помощью каких показателей оценивается точность полученного прогноза?

4. Что оценивается с помощью метода Дарбина – Уотсона?

5. Как можно оценить соответствие распределения остаточных величин нормальному распределению?

6. В чем особенность сравнения коэффициента d с табличными значениями при отрицательной автокорреляции?

Практические задания

Задача 6.1

По данным о плотности автомобильных дорог общего пользования с твердым покрытием (на конец года; км путей на 1000 км^2 территории) проверить адекватность выбранной линейной модели.

Таблица 6.5. Плотность автомобильных дорог общего пользования с твердым покрытием в ЦФО (на конец года; км путей на 1000 км^2 территории)

Регион	**Год**								
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
Белгородская область	248	247	251	600	639	675	704	729	733
Брянская область	195	194	194	280	304	307	307	312	315
Владимирская область	194	216	314	315	330	332	336	340	343
Воронежская область	204	205	207	303	319	323	327	338	345
Ивановская область	183	217	248	313	333	340	333	335	334
Калужская область	301	304	307	312	318	319	321	321	323
Костромская область	92	92	92	112	127	130	133	136	136
Курская область	243	244	247	336	345	352	357	362	367
Липецкая область	253	256	264	482	507	514	520	528	533
Московская область	636	670	672	695	698	709	720	729	740
Орловская область	226	225	234	336	358	363	364	368	371
Рязанская область	185	197	203	228	258	261	264	269	269
Смоленская область	180	180	181	266	278	279	283	308	310
Тамбовская область	169	184	191	259	284	286	287	286	288
Тверская область	186	184	186	211	230	244	248	249	248
Тульская область	206	206	224	364	389	393	395	394	396
Ярославская область	195	193	196	256	262	264	265	269	270

Задача 6.2

На основе указанного выше примера рассчитать показатели точности модели: среднюю абсолютную, среднюю относительную и среднеквадратическую ошибки. Сделать вывод о точности расчетов.

Задача 6.3

По приведенным в табл. 6.6 данным рассчитать адекватность предполагаемой линейной модели тренда.

Таблица 6.6. Число разработанных передовых производственных технологий по видам экономической деятельности в целом по Российской Федерации, ед.

Показатель	Год						
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Всего	864	1138	1323	1429	1409	1398	1534
из них по видам экономической деятельности:							
добыча полезных ископаемых	5	10	14	15	25	18	25
обрабатывающие производства	231	338	336	398	414	442	523
деятельность в области электросвязи	…	3	12	10	11	7	19
деятельность, связанная с использованием вычислительной техники и информационных технологий	13	15	23	27	28	44	79
научные исследования и разработки	443	531	562	619	546	529	502

Задача 6.3

По приведенным в табл. 6.6 данным оценить точность расчетов по выбранной линейной модели.

Задача 6.4

Оценить адекватность выбранной линейной модели по одному из видов деятельности из табл. 6.7.

Таблица 6.7. Образование отходов производства и потребления по видам экономической деятельности по Российской Федерации, млн т

Показатель	Год						
	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Всего	3505	3735	4303	5008	5153	5168	5060
из них по видам экономической деятельности: сельское хозяйство, охота и лесное хозяйство	77,4	24	27,5	26,1	40,3	43,1	45,8
добыча полезных ископаемых, в том числе:	3066,5	3335	3819	4629	4701	4807	4653
добыча топливно-энергетических полезных ископаемых	1984,9	2204	2528	3023	3011	3188	3107
добыча полезных ископаемых, кроме топливно-энергетических	1081,6	1130	1291	1607	1691	1620	1546

Окончание табл. 6.7

Показатель	2009	2010	2011	2012	2013	2014	2015
обрабатывающие производства, из них:	252,1	280,1	280,2	291	253,7	243,1	283
производство пищевых продуктов, включая напитки, и табака	25,1	20,2	16,2	19,8	20,5	19,1	19,5
обработка древесины и производство изделий из дерева	5	9,6	3,9	3,7	5,3	5	4,5
целлюлозно-бумажное производство; издательская и полиграфическая деятельность	5,3	5,7	6,1	6,1	8,9	6,2	6,8
производство кокса и нефтепродуктов	1,5	1,7	1,7	2,1	1,5	1,8	1,5
химическое производство	20,6	25,9	41,9	14,4	16,6	12,7	15,2
производство прочих неметаллических минеральных продуктов	12,1	15,4	15,5	16,8	18,3	19,2	13,4

Задача 6.5

Рассчитать показатели точности линейной модели по выбранному варианту экономической деятельности из табл. 6.7.

7. АДАПТИВНЫЕ МОДЕЛИ

Понятие адаптивных моделей • Расчеты прогнозов на основе адаптивных моделей

7.1. Понятие адаптивных моделей

Адаптивные методы – одни из перспективных методов исследования временных рядов.

Адаптивные методы позволяют учитывать различную временную ценность информации, которая устаревает со временем. Более поздние данные оказывают большее влияние на развитие тенденции, нежели предыдущий анализируемый ряд. Реализуется разнохарактерное влияние уровней с помощью различных весовых коэффициентов. Коэффициенты адаптивной модели оценивают на основе **рекуррентного метода**, позволяющего создать самокорректирующиеся модели, которые учитывают результат прогноза, полученного на предыдущем шаге.

Процесс адаптации заключается в следующем. На основе исходных данных ряда динамики составляют прогноз. Поступившие фактические данные сопоставляют с прогнозным значением, и разница между ними поступает в модель и влияет на расчет нового прогноза. Таким образом, адаптация осуществляется с каждым поступлением новых фактических данных, поэтому модель постоянно обновляется, приспосабливаясь к новой информации и отражая тенденцию развития, существующую в данный момент.

Скорость адаптации регулируется так называемым параметром адаптации, который можно получить на основе эмпирических данных или вывести аналитически. Адаптивные модели эффективно применяют в краткосрочном прогнозировании. Наиболее простой адаптивной моделью считают **экспоненциальная модель** вида

$$S_t = \alpha \cdot y_t + \beta \cdot S_{t-1},$$

где S_t – значение экспоненциальной средней для момента t; α – параметр сглаживания, постоянная величина, $0 < \alpha < 1$.

Если последовательно использовать предыдущее выражение, то значение экспоненциальной средней можно выразить через предшествующие значения анализируемого ряда динамики, а при $n \to \infty$ – формулой

$$S_t = \alpha \sum_{i=0}^{n} \beta^i \cdot y^{t-i} .$$

Из формулы следует, что экспоненциальная величина равна среднему взвешенному значению всех членов ряда. При этом весовые коэффициенты уровней ряда снижаются по мере их «устаревания». Так, последовательно подставляя в формулу значения 0; 1; 2 и далее при $\alpha = 0{,}3$, получаем весовые коэффициенты уровней соответственно 0,3; 0,21; 0,147 и т. д.

Английский математик Р. Браун доказал следующее соотношение дисперсии экспоненциальной средней и дисперсии ряда динамики (σ^2):

$$D[S_1] = \frac{\alpha}{2-\alpha}\sigma^2 .$$

В соответствии с вышеприведенным выражением при высоком значении α коэффициентом при σ^2 можно пренебречь.

С уменьшением α дисперсия экспоненциальной средней снижается, возрастает ее отличие от дисперсии временного ряда, она как бы сглаживает ряд, поглощая колебания ряда динамики. Так при $\alpha = 0{,}2$ дисперсия экспоненциальной средней составит 0,11 дисперсии временного ряда, в то время как при $\alpha = 0{,}5$ дисперсия экспоненциальной средней равна 0,33 дисперсии временного ряда, т. е. в 3 раза больше.

Выбор значения **параметра сглаживания** связан с разрешением противоречия, состоящего в том, что для увеличения веса современной информации следует увеличить α, а для повышения качества отсеивания случайных колебаний эту величину необходимо снижать. Решают данную проблему либо перебором данных, либо путем оптимизации. В качестве критерия выбирают минимум дисперсии ошибки.

7.2. Расчеты прогнозов на основе адаптивных моделей

При использовании экспоненциального сглаживания для составления краткосрочного прогноза на момент *t* величину прогноза можно представить выражением

$$S_t = S_{t-1} + \alpha(y_t - S_{t-1}),$$

где $(y_t - S_{t-1})$ – погрешность прогноза; S_{t-1} – прогноз предыдущего периода.

Таким образом, адаптивная модель позволяет получить новый прогноз в результате корректировки предыдущего прогноза с учетом его ошибки.

Проиллюстрируем экспоненциальное сглаживание на примере динамики среднемесячного курса доллара США к рублю за 2015 г.

В качестве исходного значения экспоненциальной средней примем среднее значение курса за полугодие, т. е.

S_{t-1} = 1/6 (68,93 + 61,27 + 58,46 + 51,7 + 52,97) = 58,14.

Значение экспоненциальной средней при α = 0,2 составит:

в январе $S_1 = 0{,}2 \cdot 68{,}93 + (1 - 0{,}2)\ 58{,}14\ 60{,}30$,

в феврале $S_2 = 0{,}2 \cdot 61{,}27 + (1 - 0{,}2)\ 60{,}30 = 60{,}49$,

в марте $S_3 = 0{,}2 \cdot 58{,}46 + (1 - 0{,}2)\ 60{,}49 = 60{,}09$.

Аналогично рассчитывают сглаженные значения курса доллара по другим месяцам.

Значения экспоненциальной средней при значениях параметра адаптации α = 0,5 составят

в январе $S_1 = 0{,}5 \cdot 68{,}93 + (1 - 0{,}5)\ 58{,}14 = 63{,}54$,

в феврале $S_2 = 0{,}5 \cdot 61{,}27 + (1 - 0{,}5)\ 63{,}54 = 62{,}01$,

в марте $S_3 = 0{,}5 \cdot 58{,}46 + (1 - 0{,}5)\ 60{,}49 = 61{,}05$.

Аналогично рассчитанные значения по остальным месяцам представлены в табл. 7.1.

Таблица 7.1. Ежемесячный курс доллара США к рублю в 2015 г.

Месяц	***t***	**Курс доллара США**	**Экспоненциальная средняя**	
			α = 0,2	**α = 0,5**
Январь	1	68,93	60,30	63,54
Февраль	2	61,27	60,49	62,01
Март	3	58,46	60,09	61,05
Апрель	4	51,7	58,41	59,73
Май	5	52,97	57,32	58,53
Июнь	6	55,52	56,96	57,74
Июль	7	58,99	57,37	57,56
Август	8	66,48	59,19	58,37
Сентябрь	9	66,24	60,60	59,49
Октябрь	10	64,37	61,35	60,42
Ноябрь	11	66,24	62,33	61,38
Декабрь	12	72,88	64,44	62,91

Сглаженные кривые и исходный график курса доллара представлены на рис. 7.1.

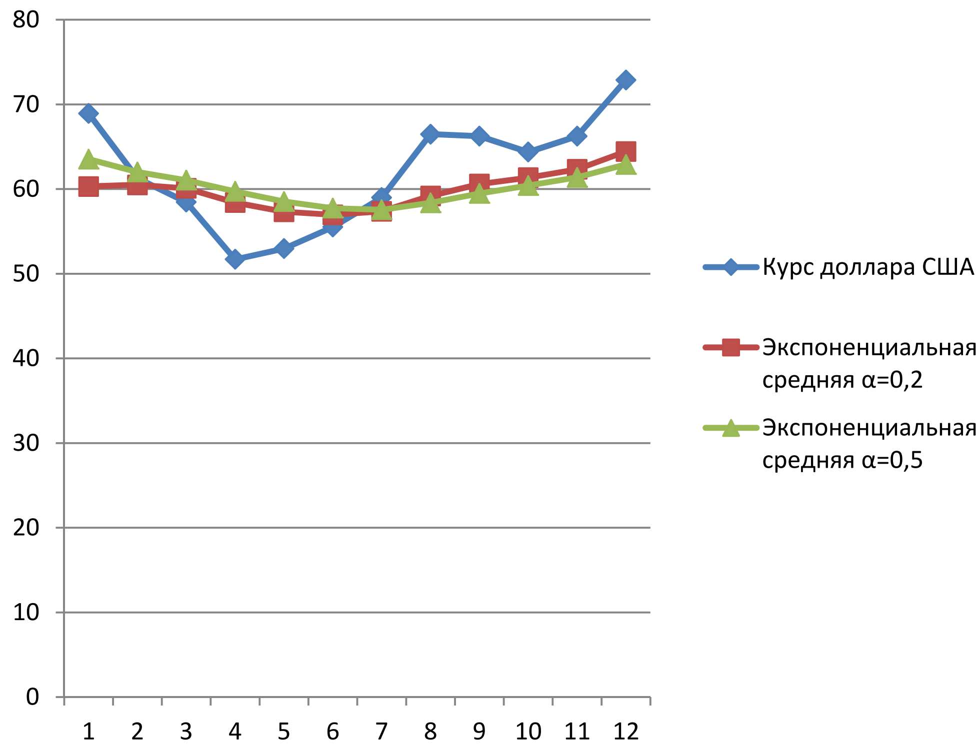


Рис. 7.1. Экспоненциальные средние курса доллара США в 2015 г.

Контрольные вопросы

1. В чем состоит цель применения адаптивных моделей?

2. В чем заключается принцип адаптации на основе рекуррентного метода?

3. Приведите аналитическое выражение рекуррентного метода.

4. В чем состоит противоречие выбора параметра сглаживания?

5. Какой критерий используют при выборе значения параметра сглаживания?

Практические задания

Задача 7.1

По приведенным ниже данным проведите экспоненциальное сглаживание курса евро за 2015 г. с помощью параметра адаптации, равного 0,2 и 0,5. Постройте графики курса евро и сглаженных рядов. Сделайте выводы.

Таблица 7.2. Динамика курса евро в 2015 – 2017 гг.

Месяц	***t***	**Курс евро**		
		2015	**2016**	**2017**
Январь	1	75,34	84,77	63,56
Февраль	2	73,30	85,95	62,40
Март	3	65,41	78,15	62,00
Апрель	4	57,36	75,61	60,46
Май	5	56,31	74,36	62,96
Июнь	6	61,10	73,29	64,97
Июль	7	62,99	71,20	68,62
Август	8	72,84	72,80	70,37
Сентябрь	9	75,04	72,36	68,79
Октябрь	10	71,04	69,13	67,86
Ноябрь	11	69,88	69,54	69,13
Декабрь	12	75,79	65,52	69,34

Задача 7.2

По приведенным данным проведите экспоненциальное сглаживание курса доллара США за 2016 и 2017 гг. с помощью параметра адаптации, равного 0,3 и 0,7. Постройте графики курса доллара и сглаженных рядов. Сделайте выводы.

Таблица 7.3. Курс доллара США в 2016 и 2017 гг.

Месяц	***t***	**Курс доллара США**	
		2016	**2017**
Январь	1	77,93	59,63
Февраль	2	77,33	58,54
Март	3	70,42	58,01
Апрель	4	66,68	56,44
Май	5	65,84	56,95
Июнь	6	65,22	57,89
Июль	7	64,34	59,69
Август	8	64,94	59,61
Сентябрь	9	64,56	57,75
Октябрь	10	62,62	57,69
Ноябрь	11	64,31	58,93
Декабрь	12	62,09	58,57

Задача 7.3

На основе приведенных данных проведите экспоненциальное сглаживание средних потребительских цен на отдельные виды продовольственных товаров с помощью $\alpha = 0,3$ и $\alpha = 0,5$.

Таблица 7.4. Средние потребительские цены на отдельные виды продовольственных товаров (на конец года; руб. за кг)

Вид товара \ Год	2010	2011	2012	2013	2014	2015	2016	2017
Говядина (кроме бескостного мяса)	197,64	234,49	248,47	244,55	272,28	314,94	315,02	320,34
Свинина (кроме бескостного мяса)	198,35	210,89	220,09	214,18	272,36	271,08	264,32	255,87
Куры охлажденные и мороженые	105,14	103,57	117,26	107,03	136,14	133,73	138,49	126,29
Колбаса вареная	235,96	270,28	288,23	302,94	310,54	344,81	351,27	360,88
Рыба мороженая неразделанная	79,22	86,79	85,67	90,79	110,65	138,16	147,68	153,03
Сливочное масло	239,55	256,48	260,84	308,92	357,54	397,75	477,13	528,83
Подсолнечное масло	72,60	76,79	78,51	75,47	78,09	107,62	110,10	100,16

Задача 7.4

По данным предыдущей задачи постройте графики цен на потребительские товары и графики сглаженных рядов.

Задача 7.5

На основе приведенных в табл. 7.6 данных проведите экспоненциальное сглаживание динамики численности населения Российской федерации с помощью $\alpha = 0{,}3$ и $\alpha = 0{,}5$.

Таблица 7.6. Численность населения Российской Федерации, млн чел.

Показатель \ Год	2010	2011	2012	2013	2014	2015	2016	2017	2018
Все население	142,9	142,9	143,0	143,3	143,7	146,3	146,5	146,8	146,9
в том числе мужчины	66,1	66,1	66,1	66,3	66,6	67,8	67,9	68,1	68,1
женщины	76,8	76,8	76,9	77,0	77,1	78,5	78,6	78,7	78,8
в том числе городское	105,3	105,4	105,7	106,1	106,6	108,3	108,6	109,0	109,3
сельское	37,6	37,5	37,3	37,2	37,1	38,0	37,9	37,8	37,6

8. МНОГОФАКТОРНЫЕ МОДЕЛИ

Кластерный анализ • Виды кластерного анализа • Правила иерархического объединения в кластер • Формирование кластеров с помощью программы STATISTICA

8.1. Кластерный анализ

Выделение типов в результате группировки или классификации данных приводит к получению однородных групп, которые различаются между собой. Это позволяет анализировать вариацию признака, рассчитывать сводные показатели. Кроме того, с ростом однородности обобщаемых данных повышается устойчивость всех статистических показателей.

Кластерный анализ – это метод классификационного анализа; его основное назначение – разбиение множества исследуемых объектов и признаков на однородные в некотором смысле группы, или кластеры. Это многомерный статистический метод, поэтому предполагается, что исходные данные могут быть значительного объема, т. е. существенно большим может быть как количество объектов исследования (наблюдений), так и признаков, характеризующих эти объекты.

Большое достоинство кластерного анализа в том, что он дает возможность разбивать объекты не по одному признаку, а по ряду признаков. Кроме того, кластерный анализ в отличие от большинства математико-статистических методов не накладывает никаких ограничений на вид рассматриваемых объектов и позволяет исследовать множество исходных данных практически произвольной природы. Так как кластеры – это группы однородности, то задача кластерного анализа заключается в том, чтобы на основании признаков объектов разбить их множество на m (m – целое) кластеров так, чтобы каждый объект принадлежал только одной группе разбиения. При этом объекты, принадлежащие одному кластеру, должны быть однородными (сходными), а объекты, принадлежащие разным кластерам, – разнородными. Если объекты кластеризации представить как точки в n-мерном пространстве признаков (n – количество признаков, характеризующих объекты), то сходство между объектами определяется через понятие расстояния между точками, так как понятно, что чем меньше расстояние между объектами, тем они более схожи.

Евклидово расстояние – наиболее популярная метрика, представляющая собой геометрическое расстояние в многомерном пространстве.

Данная метрика, как и большинство других, чувствительна к изменению единиц измерения осей. Например, если сантиметры перевести в миллиметры, то изменится и исчисляемое расстояние. Поэтому при использовании большинства метрик кластерный анализ предполагает предварительную стандартизацию исходных данных.

8.2. Виды кластерного анализа

Алгоритмов кластерного анализа достаточно много. Все их можно подразделить на **иерархические** и **неиерархические.**

Иерархические (древовидные) процедуры – наиболее распространённые алгоритмы кластерного анализа по их реализации на ЭВМ.

Различают **агломеративные** (от слова agglomerate – собирать) и **итеративные дивизивные** (от слова division – разделять) процедуры.

Принцип работы **иерархических агломеративных** процедур построен на последовательном объединении групп элементов, начиная с самых близких и переходя ко всё более отдалённым друг от друга.

Принцип работы **иерархических дивизивных** процедур, наоборот, предполагает последовательное разделение групп элементов сначала самых далёких, а затем всё более близких друг к другу. Большинство этих алгоритмов исходит из матрицы расстояний (сходства). К недостаткам иерархических процедур следует отнести громоздкость их вычислений. На каждом шаге алгоритмы требуют вычисления матрицы расстояний, а следовательно, значительного объема машинной памяти и большого количества времени. В этой связи реализация таких алгоритмов при числе наблюдений, превышающем несколько сотен, нецелесообразна, а в ряде случаев и невозможна.

Общий принцип работы агломеративного алгоритма следующий. На первом шаге каждое наблюдение рассматривают как отдельный кластер. В дальнейшем на каждом шаге работы алгоритма объединяют два самых близких кластера и с учётом принятого расстояния по формуле пересчитывают матрицу расстояний, размерность которой, очевидно, снижается на единицу. Работа алгоритма заканчивается, когда все наблюдения объединены в один класс. Большинство программ, применяющих алгоритм иерархической классификации, предусматривает графическое представление классификации в виде дендрограммы. В программе STATISTICA реализованы агломеративные методы минимальной дисперсии – древовидная кластеризация и двухвходовая кластеризация, а также дивизивный метод *k*-средних.

8.3. Правила иерархического объединения в кластер

В методе древовидной кластеризации предусмотрены различные правила иерархического объединения в кластеры:

1. **Правило одиночной связи**. Метод называют еще **методом ближайшего соседа**, так как расстояние между двумя кластерами определяется как между двумя наиболее близкими объектами в различных кластерах. На первом шаге объединяют два наиболее близких объекта, т. е. имеющие максимальную меру сходства. На следующем шаге к ним присоединяется объект с максимальной мерой сходства с одним из объектов кластера, т. е. для его включения в кластер требуется максимальное сходство лишь с одним членом кластера. Это правило «нанизывает» объекты для формирования кластеров. Недостаток данного метода – образование слишком больших продолговатых кластеров.

2. **Правило полных связей.** Метод позволяет устранить недостаток, присущий методу одиночной связи. В соответствии с правилом два объекта, принадлежащих одной и той же группе (кластеру), имеют коэффициент сходства, который превышает некоторое пороговое значение *S*. В терминах евклидова расстояния это означает, что расстояние между двумя точками (объектами) кластера не должно превышать некоторого порогового значения *d*. Таким образом, *d* определяет максимально допустимый диаметр подмножества, образующего кластер. Этот метод называют еще **методом наиболее удаленных соседей,** так как при достаточно большом пороговом значении *d* расстояние между кластерами определяется наибольшим расстоянием между любыми двумя объектами в различных кластерах.

3. **Правило невзвешенного попарного среднего**. Расстояние между двумя кластерами определяется как среднее расстояние между всеми парами объектов в них. Метод эффективен, когда объекты в действительности формируют различные группы, однако он работает одинаково хорошо и в случаях протяженных (цепочного типа) кластеров.

4. **Правило взвешенного попарного среднего**. Метод идентичен предыдущему, за исключением того, что при вычислении размер соответствующих кластеров используется в качестве весового коэффициента. Желательно этот метод применять, когда предполагаются неравные размеры кластеров.

5. **Невзвешенный центроидный метод**. Расстояние между двумя кластерами определяется как расстояние между их центрами тяжести.

6. **Взвешенный центроидный метод**. Идентичен предыдущему, за исключением того, что при вычислениях расстояния используют веса для

учета разности между размерами кластеров. Если имеются (или подозреваются) значительные отличия в размерах кластеров, этот метод оказывается предпочтительнее предыдущего.

7. **Правило Уорда (Варда).** В этом методе в качестве целевой функции применяют внутригрупповую сумму квадратов отклонений, которая представляет собой сумму квадратов расстояний между каждой точкой (объектом) и средней по кластеру, содержащему этот объект. На каждом шаге объединяются такие два кластера, которые приводят к минимальному увеличению целевой функции, т. е. внутригрупповой суммы квадратов отклонений. Этот метод направлен на объединение близко расположенных кластеров. Замечено, что метод Уорда приводит к образованию кластеров примерно равных размеров и имеющих форму гиперсфер.

Метод *k*-средних. Предположим, есть гипотезы относительно числа m кластеров (по переменным или наблюдениям). Тогда можно задать программе создать ровно m кластеров так, чтобы они были настолько различны, насколько это возможно. Именно для решения задач этого типа предназначен метод *k-means* (*k*-средних).

Гипотеза может основываться на теоретических соображениях, результатах предшествующих исследований или догадке. Выполняя последовательное разбиение на различное число кластеров, можно сравнивать качество получаемых решений. Программа начинает с m случайно выбранных кластеров, а затем изменяет принадлежность объектов к ним, чтобы минимизировать изменчивость внутри кластеров и максимизировать изменчивость между ними. Алгоритм случайным образом в пространстве назначает центры будущих кластеров. Затем вычисляет расстояние между центрами кластеров и каждым объектом, и объект приписывается к тому кластеру, к которому он ближе всего. Завершив приписывание, алгоритм вычисляет средние значения для каждого кластера. Этих средних будет столько, сколько используется переменных для проведения анализа, – k штук. Набор средних представляет собой координаты нового положения центра кластера. Алгоритм вновь вычисляет расстояние от каждого объекта до центров кластеров и приписывает объекты к ближайшему кластеру. Вновь вычисляются центры тяжести кластеров, и этот процесс повторяется до тех пор, пока центры тяжести не перестанут «мигрировать» в пространстве. Если в древовидной кластеризации можно использовать категориальные переменные, то так как в методе *k*-средних в качестве метрики используют евклидову метрику, то перед проведением кластеризации необходимо стандартизовать переменные. По этой же причине в методе предполагается, что переменные непрерывные и измерены как минимум в интервальной шкале.

Разбиение модельной базы предприятий на классы методом кластерного анализа сводится к разбиению множества элементов корреляционной матрицы признаков на *m* групп удовлетворяющих критерию оптимальности:

- каждое предприятие должно принадлежать одному и только одному подмножеству разбиения (кластеру);

- предприятия, принадлежащие одному и тому же кластеру, должны быть сходными;

- предприятия, принадлежащие разным кластерам, должны быть разнородными. Этот критерий может представлять собой целевую функцию – функционал, выражающий уровни желательности различных разбиений и группировок. Сходство и разнородность объектов будем определять по средствам *m*-мерного евклидового расстояния между векторами измерений.

Для проведения кластеризации необходимо обеспечить сопоставимость исходных показателей методом «нормированной разности» по формуле

$$d_i = \frac{x_i - \bar{x}}{\sigma},$$

где x_i – значение i-го признака; $\bar{x}$ – среднее значение признака; σ – среднее квадратическое отклонение признака x_i.

Кластеризация осуществляется методом Варда в пакете прикладного статистического анализа STATISTICA. В качестве целевой функции выступает евклидово расстояние, которое рассчитывается по формуле

$$d(x_i, y_j) = \sqrt{\sum_{i=1}^{n}(x_{il} - y_{jl})}^2,$$

где x_{il}, y_{jl} – значения l-го признака i-го и j-го объектов (l = 1, 2, …, n; I, j = 1, 2, … k).

Оптимальное число кластеров определяется на основании порогового расстояния в соответствии с графиком и таблицей объединения объектов, рассчитанных с помощью программного продукта STATISTICA.

Для анализа качества полученных групп необходимо провести **дискриминантный** анализ, который позволит оценить полученную классификацию с помощью статистических методов.

Дискриминация полученной классификации с помощью кластерного анализа оценивается по значению лямбды Уилкса.

Статистика лямбда Уилкса вычисляется как отношение детерминанта матрицы внутригрупповых дисперсий/ковариаций к детерминанту общей ковариационной матрицы и изменяется от 0 до 1.

Лямбда Уилкса может принимать значения от 0 (группы полностью однородны) до 1 (разделение объектов на группы не приводит к тому, что внутригрупповая изменчивость оказывается меньше общей). Чем меньшее значение имеет лямбда Уилкса, тем более эффективным оказывается разделение на группы. Значение лямбды Уилкса, лежащее около 1, свидетельствует о недостаточной дискриминации, что говорит о существенной неоднородности групп классификации, полученной с помощью кластерного анализа. В этом случае необходимо вернуться к п. 2 данной методики.

Оценка вклада в общую дискриминацию осуществляется на основании значения Partial Lambda (частной лямбды). Чем меньше статистика частной лямбды Уилкса, тем больше вклад в общую дискриминацию.

Успешность включения переменных в модель оценивается по показателю толерантности (Toler). Чем меньше значение толерантности, тем избыточнее переменная в модели, так как переменная несет малую дополнительную информацию. Причем нижняя граница толерантности составляет 0,01. Чем ниже значение Toler, тем сильнее данный признак связан со всеми остальными.

8.4. Формирование кластеров с помощью программы STATISTICA

Рассмотрим формирование региональных кластеров по отобранным показателям, характеризующим здоровый образ жизни населения РФ, с помощью программы STATISTICA.

Запускаем программу STATISTICA. Создаем новый документ с помощью комбинации Ctrl+N или выбираем New из меню (рис. 8.1).

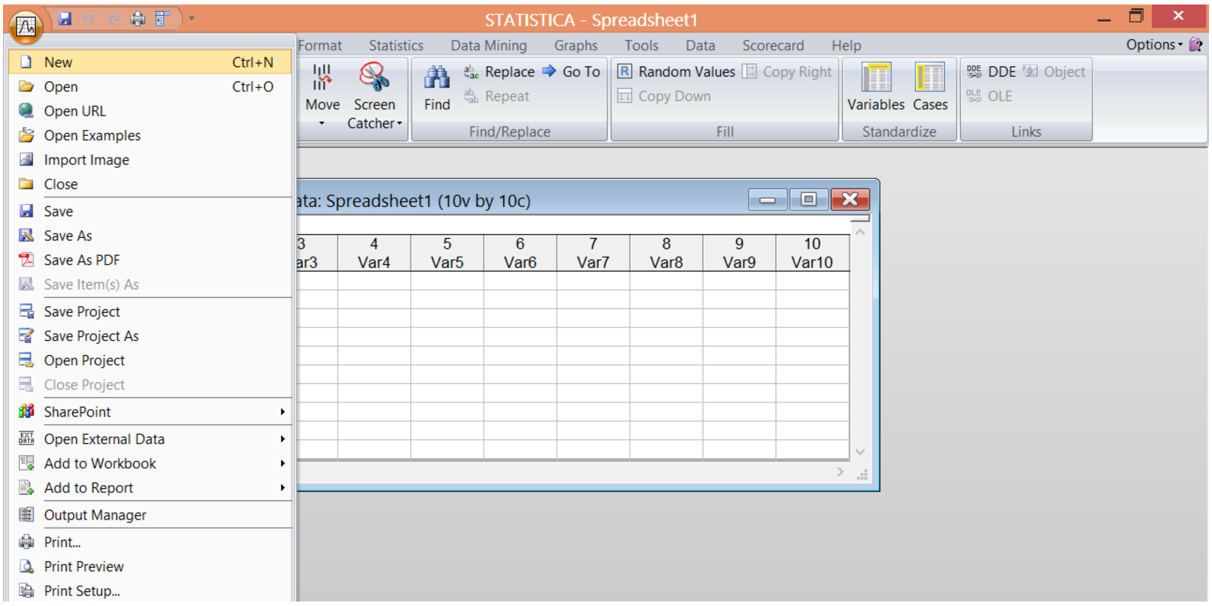


Рис. 8.1. Создание нового документа в программном комплексе STATISTICA

Задаем количественные характеристики исходных данных (рис. 8.2). В нашем случае – количество анализируемых показателей (variables) и количество регионов (cases).

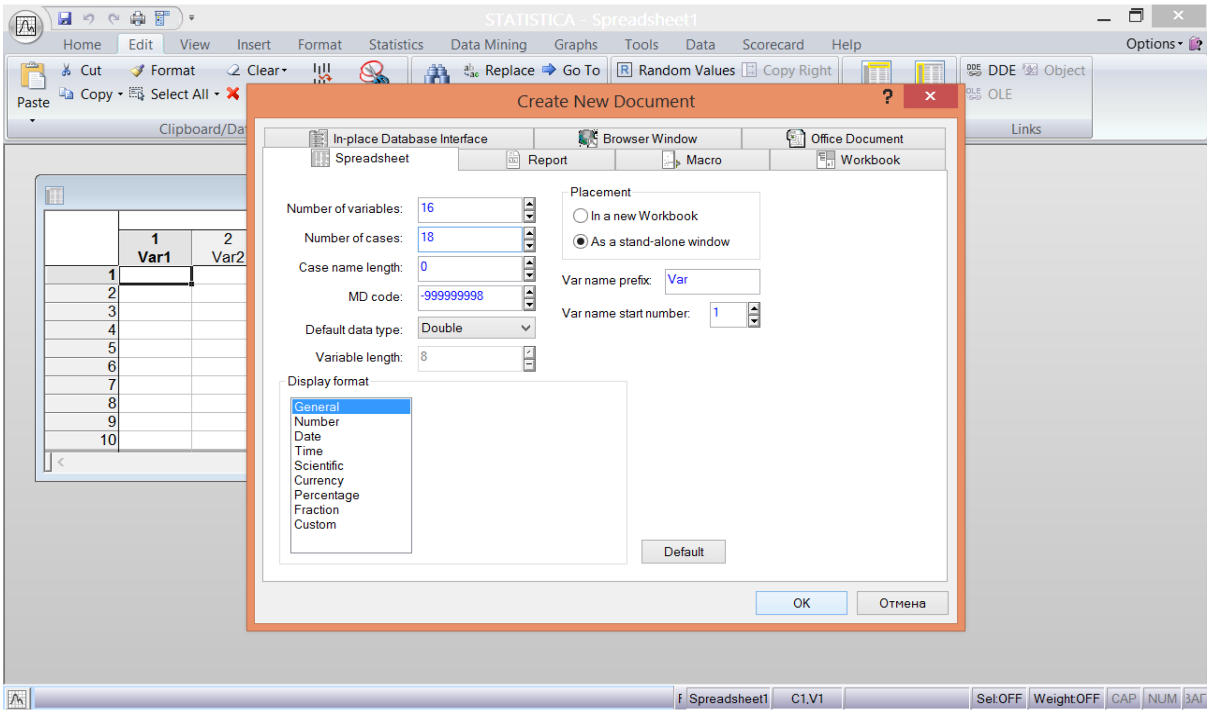


Рис. 8.2. Введение количественных характеристик исходных данных в программном комплексе STATISTICA

В качестве переменных (variables) на основе данных статистического сборника «Регионы России. Социально-экономические показатели» предлагается рассмотреть следующие показатели:

1. Среднедушевые денежные доходы (в месяц), руб.

2. Валовой региональный продукт на душу населения, руб.

3. Общие коэффициенты рождаемости (число родившихся на 1000 человек населения).

4. Общие коэффициенты смертности (число умерших на 1000 человек населения).

5. Ожидаемая продолжительность жизни при рождении (число лет).

6. Заболеваемость на 1000 человек населения (зарегистрировано заболеваний у пациентов с диагнозом, установленным впервые в жизни).

7. Уровень оснащенности стадионами с трибунами на 1500 мест и более (единиц на 1 млн чел.)*.

8. Уровень оснащенности плоскостными спортивными сооружениями (единиц на 1 млн чел.)*.

9. Уровень оснащенности спортивными залами (единиц на 1 млн чел.)*.

10. Уровень оснащенности плавательными бассейнами (единиц на 1 млн чел.)*.

11. Уровень оснащенности детскими оздоровительными лагерями (единиц на 1 млн чел.)*.

12. Удельное количество детей, отдохнувших в детских оздоровительных лагерях за лето (на 1000 чел.)*.

13. Выбросы загрязняющих веществ в атмосферный воздух, отходящих от стационарных источников (тыс. т).

14. Улавливание загрязняющих атмосферу веществ, отходящих от стационарных источников (тыс. т).

15. Использование свежей воды (млн $м^3$).

16. Сброс загрязненных сточных вод в поверхностные водные объекты (млн $м^3$).

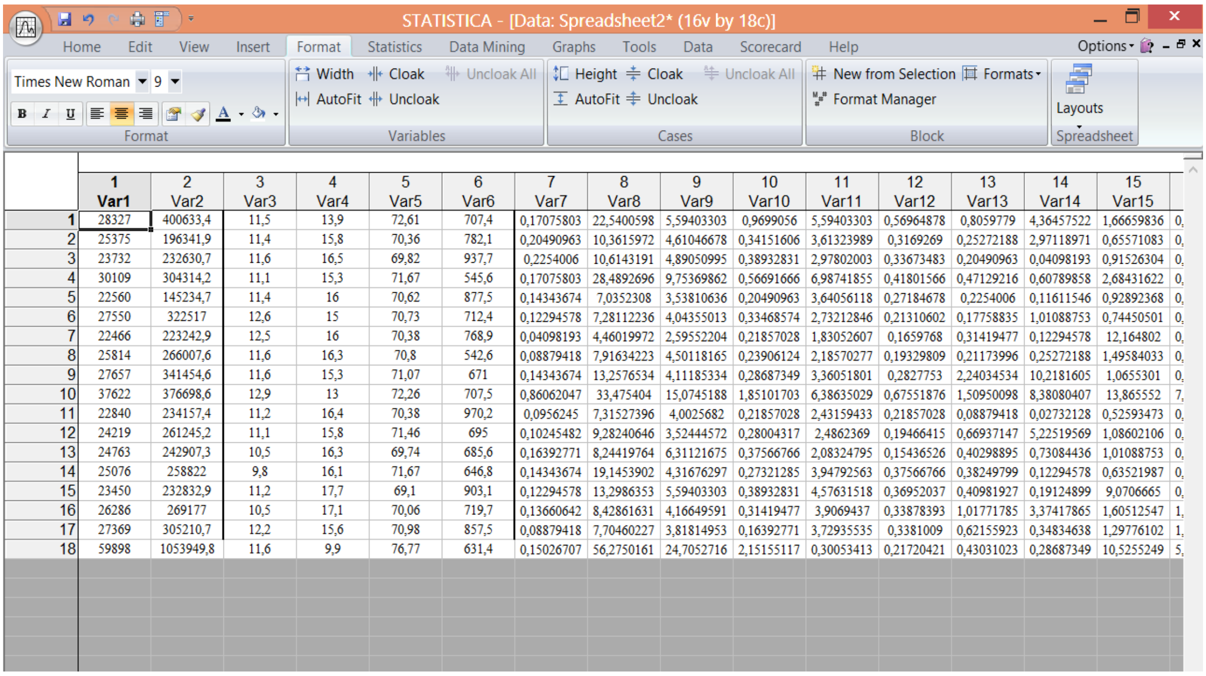
* Сопоставимость данных показателей по отдельным регионам обеспечивается их пересчетом на среднегодовую численность населения (тыс. чел.).

Приведенный перечень включает показатели, отражающие условия ведения здорового образа жизни населением (7 – 12); экологические факторы, влияющие на качество жизни (13 – 16); экономические факторы, обеспечивающие качество жизни (1 – 2); показатели результативности здорового образа жизни (3 – 6). Общепризнанный результирующий показатель – ожидаемая продолжительность жизни при рождении (число лет).

Анализ проводится в рамках Центрального федерального округа Российской Федерации (Cases), в состав которого входят:

1. Белгородская область.
2. Брянская область.
3. Владимирская область.
4. Воронежская область.
5. Ивановская область.
6. Калужская область.
7. Костромская область.
8. Курская область.
9. Липецкая область.
10. Московская область.
11. Орловская область.
12. Рязанская область.
13. Смоленская область.
14. Тамбовская область.
15. Тверская область.
16. Тульская область.
17. Ярославская область.
18. Г. Москва.

Далее необходимо осуществить ввод исходных данных (рис. 8.3). В программном комплексе STATISTICA допустимо копирование данных из расчетов, выполненных в Excel.

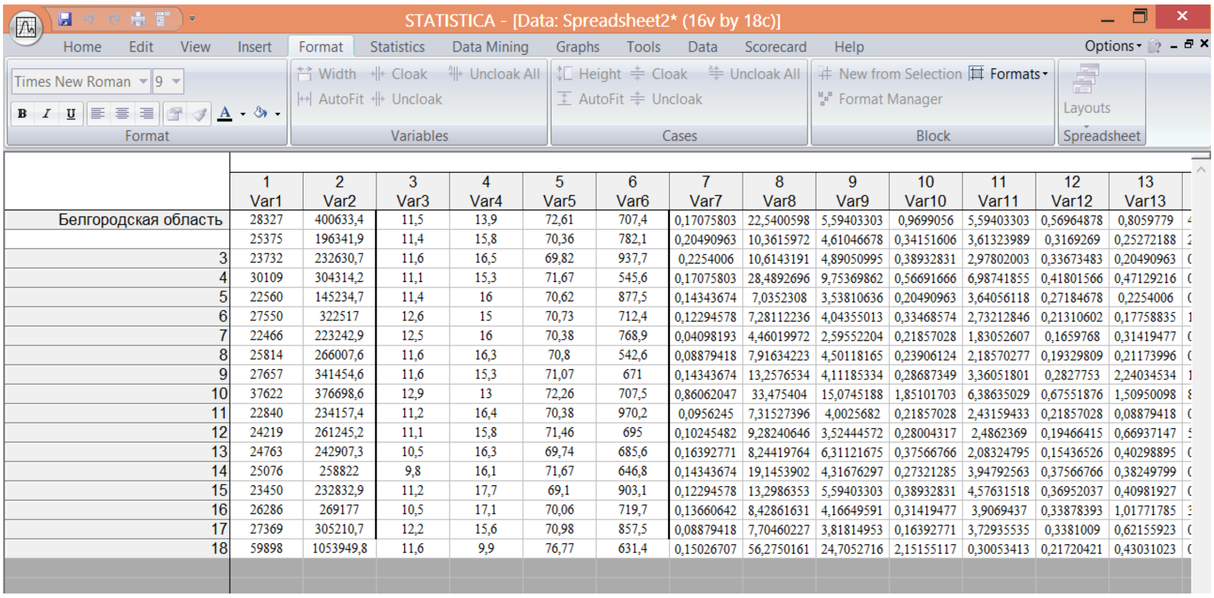


STATISTICA - [Data: Spreadsheet2* (16v by 18c)]

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10	11 Var11	12 Var12	13 Var13	14 Var14	15 Var15
1	28327	400633,4	11,5	13,9	72,61	707,4	0,17075803	22,5400598	5,59403303	0,9699056	5,59403303	0,56964878	0,8059779	4,36457522	1,66659836
2	25375	196341,9	11,4	15,8	70,36	782,1	0,20490963	10,3615972	4,61046678	0,34151606	3,61323989	0,3169269	0,25272188	2,97118971	0,65571083
3	23732	232630,7	11,6	16,5	69,82	937,7	0,2254006	10,6143191	4,89050995	0,38932831	2,97802003	0,33673483	0,20490963	0,04098193	0,91526304
4	30109	304314,2	11,1	15,3	71,67	545,6	0,17075803	28,4892696	9,75369862	0,56691666	6,98741855	0,41801566	0,47129216	0,60789858	2,68431622
5	22560	145234,7	11,4	16	70,62	877,5	0,14343674	7,0352308	3,53810636	0,20490963	3,64056118	0,27184678	0,2254006	0,11611546	0,92892368
6	27550	322517	12,6	15	70,73	712,4	0,12294578	7,28112236	4,04355013	0,33468574	2,73212846	0,21310602	0,17758835	1,01088753	0,74450501
7	22466	223242,9	12,5	16	70,38	768,9	0,04098193	4,46019972	2,59552204	0,21857028	1,83052607	0,1659768	0,31419477	0,12294578	12,164802
8	25814	266007,6	11,6	16,3	70,8	542,6	0,08879418	7,91634223	4,50118165	0,23906124	2,18570277	0,19329809	0,21173996	0,25272188	1,49584033
9	27657	341454,6	11,6	15,3	71,07	671	0,14343674	13,2576534	4,11185334	0,28687349	3,36051801	0,2827753	2,24034534	10,2181605	1,0655301
10	37622	376698,6	12,9	13	72,26	707,5	0,86062047	33,475404	15,0745188	1,85101703	6,38635029	0,67551876	1,50950098	8,38080407	13,865552
11	22840	234157,4	11,2	16,4	70,38	970,2	0,0956245	7,31527396	4,0025682	0,21857028	2,43159433	0,21857028	0,08879418	0,02732128	0,52593473
12	24219	261245,2	11,1	15,8	71,46	695	0,10245482	9,28240646	3,52444572	0,28004317	2,4862369	0,19466415	0,66937147	5,22519569	1,08602106
13	24763	242907,3	10,5	16,3	69,74	685,6	0,16392771	8,24419764	6,31121675	0,37566766	2,08324795	0,15436526	0,40298895	0,73084436	1,01088753
14	25076	258822	9,8	16,1	71,67	646,8	0,14343674	19,1453902	4,31676297	0,27321285	3,94792563	0,37566766	0,38249799	0,12294578	0,63521987
15	23450	232832,9	11,2	17,7	69,1	903,1	0,12294578	13,2986353	5,59403303	0,38932831	4,57631518	0,36952037	0,40981927	0,19124899	9,0706665
16	26286	269177	10,5	17,1	70,06	719,7	0,13660642	8,42861631	4,16649591	0,31419477	3,9069437	0,33878393	1,01771785	3,37417865	1,60512547
17	27369	305210,7	12,2	15,6	70,98	857,5	0,08879418	7,70460227	3,81814953	0,16392771	3,72935535	0,3381009	0,62155923	0,34834638	1,29776102
18	59898	1053949,8	11,6	9,9	76,77	631,4	0,15026707	56,2750161	24,7052716	2,15155117	0,30053413	0,21720421	0,43031023	0,28687349	10,5255249

Рис. 8.3. Ввод исходных данных в программном комплексе STATISTICA

Вводим наименования выбранных для наблюдения регионов. Переименование осуществляем двойным нажатием правой клавишей мыши и вводом текста (рис. 8.4).

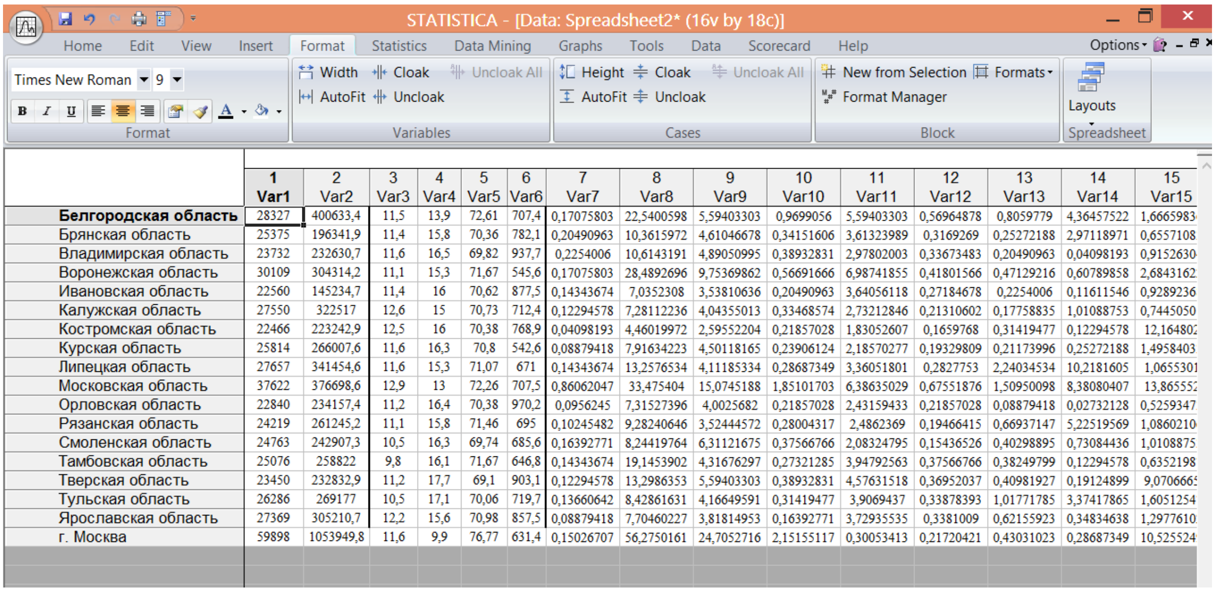


STATISTICA - [Data: Spreadsheet2* (16v by 18c)]

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10	11 Var11	12 Var12	13 Var13
Белгородская область	28327	400633,4	11,5	13,9	72,61	707,4	0,17075803	22,5400598	5,59403303	0,9699056	5,59403303	0,56964878	0,8059779
	25375	196341,9	11,4	15,8	70,36	782,1	0,20490963	10,3615972	4,61046678	0,34151606	3,61323989	0,3169269	0,25272188
3	23732	232630,7	11,6	16,5	69,82	937,7	0,2254006	10,6143191	4,89050995	0,38932831	2,97802003	0,33673483	0,20490963
4	30109	304314,2	11,1	15,3	71,67	545,6	0,17075803	28,4892696	9,75369862	0,56691666	6,98741855	0,41801566	0,47129216
5	22560	145234,7	11,4	16	70,62	877,5	0,14343674	7,0352308	3,53810636	0,20490963	3,64056118	0,27184678	0,2254006
6	27550	322517	12,6	15	70,73	712,4	0,12294578	7,28112236	4,04355013	0,33468574	2,73212846	0,21310602	0,17758835
7	22466	223242,9	12,5	16	70,38	768,9	0,04098193	4,46019972	2,59552204	0,21857028	1,83052607	0,1659768	0,31419477
8	25814	266007,6	11,6	16,3	70,8	542,6	0,08879418	7,91634223	4,50118165	0,23906124	2,18570277	0,19329809	0,21173996
9	27657	341454,6	11,6	15,3	71,07	671	0,14343674	13,2576534	4,11185334	0,28687349	3,36051801	0,2827753	2,24034534
10	37622	376698,6	12,9	13	72,26	707,5	0,86062047	33,475404	15,0745188	1,85101703	6,38635029	0,67551876	1,50950098
11	22840	234157,4	11,2	16,4	70,38	970,2	0,0956245	7,31527396	4,0025682	0,21857028	2,43159433	0,21857028	0,08879418
12	24219	261245,2	11,1	15,8	71,46	695	0,10245482	9,28240646	3,52444572	0,28004317	2,4862369	0,19466415	0,66937147
13	24763	242907,3	10,5	16,3	69,74	685,6	0,16392771	8,24419764	6,31121675	0,37566766	2,08324795	0,15436526	0,40298895
14	25076	258822	9,8	16,1	71,67	646,8	0,14343674	19,1453902	4,31676297	0,27321285	3,94792563	0,37566766	0,38249799
15	23450	232832,9	11,2	17,7	69,1	903,1	0,12294578	13,2986353	5,59403303	0,38932831	4,57631518	0,36952037	0,40981927
16	26286	269177	10,5	17,1	70,06	719,7	0,13660642	8,42861631	4,16649591	0,31419477	3,9069437	0,33878393	1,01771785
17	27369	305210,7	12,2	15,6	70,98	857,5	0,08879418	7,70460227	3,81814953	0,16392771	3,72935535	0,3381009	0,62155923
18	59898	1053949,8	11,6	9,9	76,77	631,4	0,15026707	56,2750161	24,7052716	2,15155117	0,30053413	0,21720421	0,43031023

Рис. 8.4. Переименование наблюдений в программном комплексе STATISTICA

Результат переименования наблюдений представлен на рис. 8.5.



	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10	11 Var11	12 Var12	13 Var13	14 Var14	15 Var15
Белгородская область	28327	400633,4	11,5	13,9	72,61	707,4	0,17075803	22,5400598	5,59403303	0,9699056	5,59403303	0,56964878	0,8059779	4,36457522	1,6665983
Брянская область	25375	196341,9	11,4	15,8	70,36	782,1	0,20490963	10,3615972	4,61046678	0,34151606	3,61323989	0,3169269	0,25272188	2,97118971	0,6557108
Владимирская область	23732	232630,7	11,6	16,5	69,82	937,7	0,2254006	10,6143191	4,89050995	0,38932831	2,97802003	0,33673483	0,20490963	0,04098193	0,9152630
Воронежская область	30109	304314,2	11,1	15,3	71,67	545,6	0,17075803	28,4892696	9,75369862	0,56691666	6,98741855	0,41801566	0,47129216	0,60789858	2,6843162
Ивановская область	22560	145234,7	11,4	16	70,62	877,5	0,14343674	7,0352308	3,53810636	0,20490963	3,64056118	0,27184678	0,2254006	0,11611546	0,9289236
Калужская область	27550	322517	12,6	15	70,73	712,4	0,12294578	7,28112236	4,04355013	0,33468574	2,73212846	0,21310602	0,17758835	1,01088753	0,7445050
Костромская область	22466	223242,9	12,5	16	70,38	768,9	0,04098193	4,46019972	2,59552204	0,21857028	1,83052607	0,1659768	0,31419477	0,12294578	12,164802
Курская область	25814	266007,6	11,6	16,3	70,8	542,6	0,08879418	7,91634223	4,50118165	0,23906124	2,18570277	0,19329809	0,21173996	0,25272188	1,4958403
Липецкая область	27657	341454,6	11,6	15,3	71,07	671	0,14343674	13,2576534	4,11185334	0,28687349	3,36051801	0,2827753	2,24034534	10,2181605	1,0655301
Московская область	37622	376698,6	12,9	13	72,26	707,5	0,86062047	33,475404	15,0745188	1,85101703	6,38635029	0,67551876	1,50950098	8,38080407	13,865552
Орловская область	22840	234157,4	11,2	16,4	70,38	970,2	0,0956245	7,31527396	4,0025682	0,21857028	2,43159433	0,21857028	0,08879418	0,02732128	0,5259347
Рязанская область	24219	261245,2	11,1	15,8	71,46	695	0,10245482	9,28240646	3,52444572	0,28004317	2,4862369	0,19466415	0,66937147	5,22519569	1,0860210
Смоленская область	24763	242907,3	10,5	16,3	69,74	685,6	0,16392771	8,24419764	6,31121675	0,37566766	2,08324795	0,15436526	0,40298895	0,73084436	1,0108875
Тамбовская область	25076	258822	9,8	16,1	71,67	646,8	0,14343674	19,1453902	4,31676297	0,27321285	3,94792563	0,37566766	0,38249799	0,12294578	0,6352198
Тверская область	23450	232832,9	11,2	17,7	69,1	903,1	0,12294578	13,2986353	5,59403303	0,38932831	4,57631518	0,36952037	0,40981927	0,19124899	9,070666
Тульская область	26286	269177	10,5	17,1	70,06	719,7	0,13660642	8,42861631	4,16649591	0,31419477	3,9069437	0,33878393	1,01771785	3,37417865	1,6051254
Ярославская область	27369	305210,7	12,2	15,6	70,98	857,5	0,08879418	7,70460227	3,81814953	0,16392771	3,72935535	0,3381009	0,62155923	0,34834638	1,2977610
г. Москва	59898	1053949,8	11,6	9,9	76,77	631,4	0,15026707	56,2750161	24,7052716	2,15155117	0,30053413	0,21720421	0,43031023	0,28687349	10,525524

Рис. 8.5. Результат переименования наблюдений в программном комплексе STATISTICA

Задаем переменные (в нашем случае обозначим их x_1, x_2, x_3, … и т. д.). Переименование осуществляем двойным нажатием правой клавиши мыши и вводом нужной переменной в поле Name в открывшемся диалоговом окне (рис. 8.6).

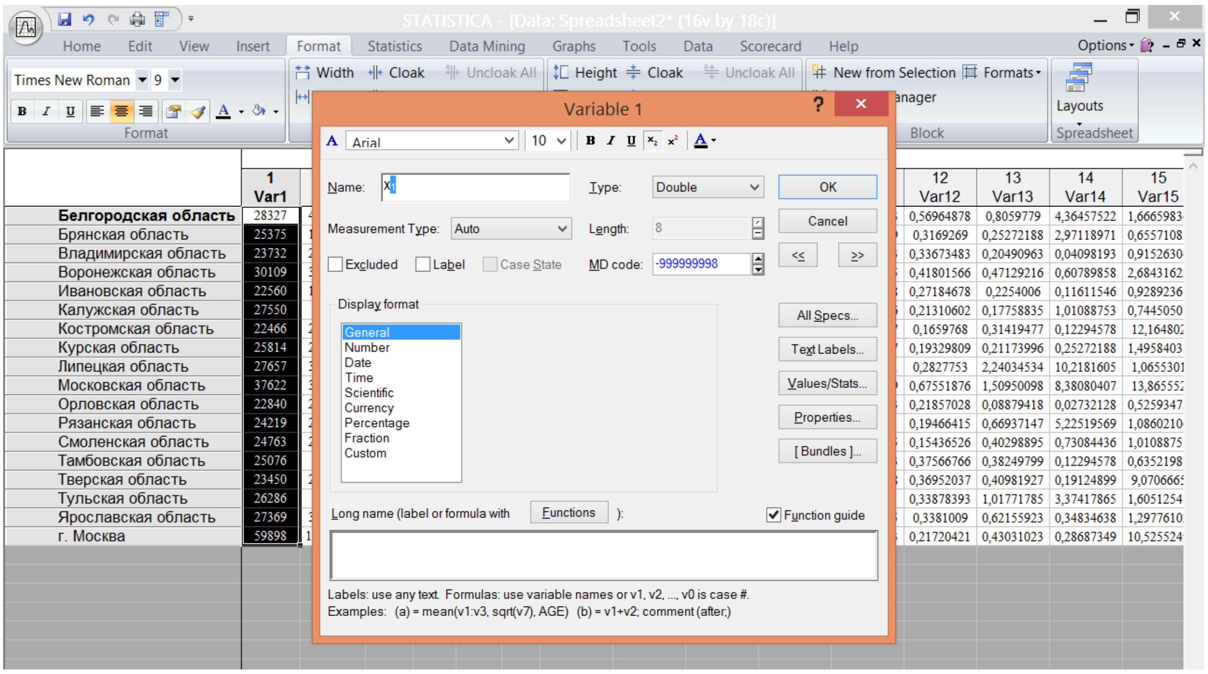
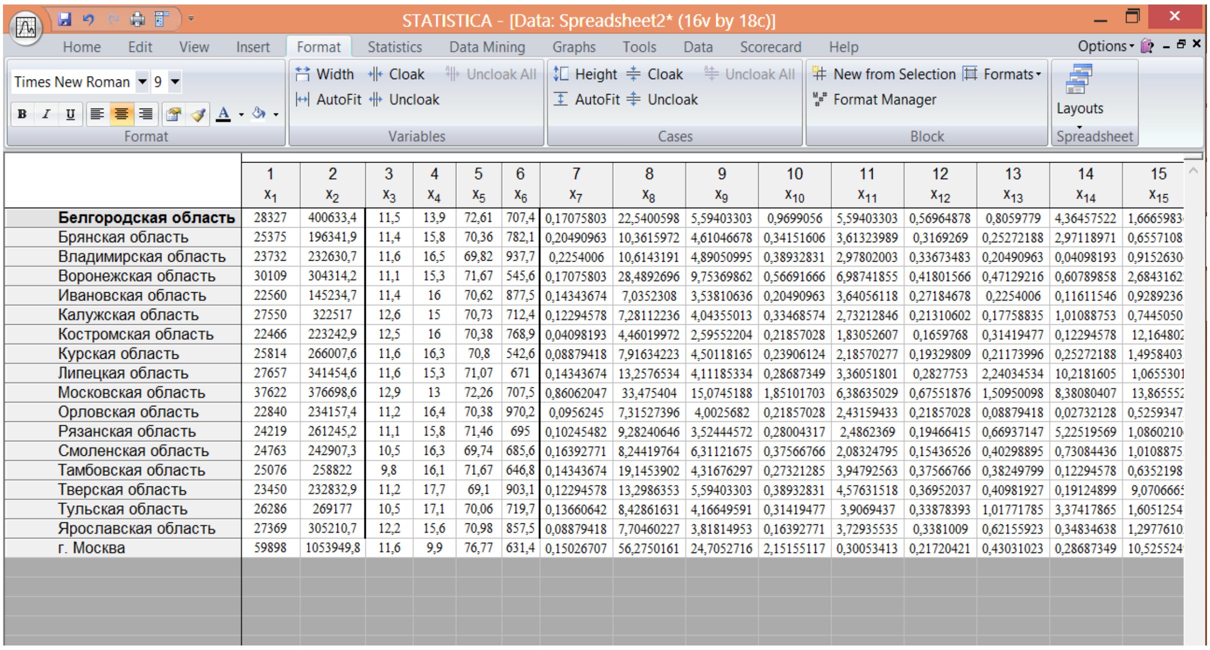


Рис. 8.6. Переименование переменных в программном комплексе STATISTICA

Результат переименования переменных представлен на рис. 8.7.



Рисy. 8.7. Результат переименования переменных в программном комплексе STATISTICA

Для выбора и начала кластерного анализа заходим на вкладку Statistics, выбираем Mult/Exploratory, затем выбираем Cluster (рис. 8.8).

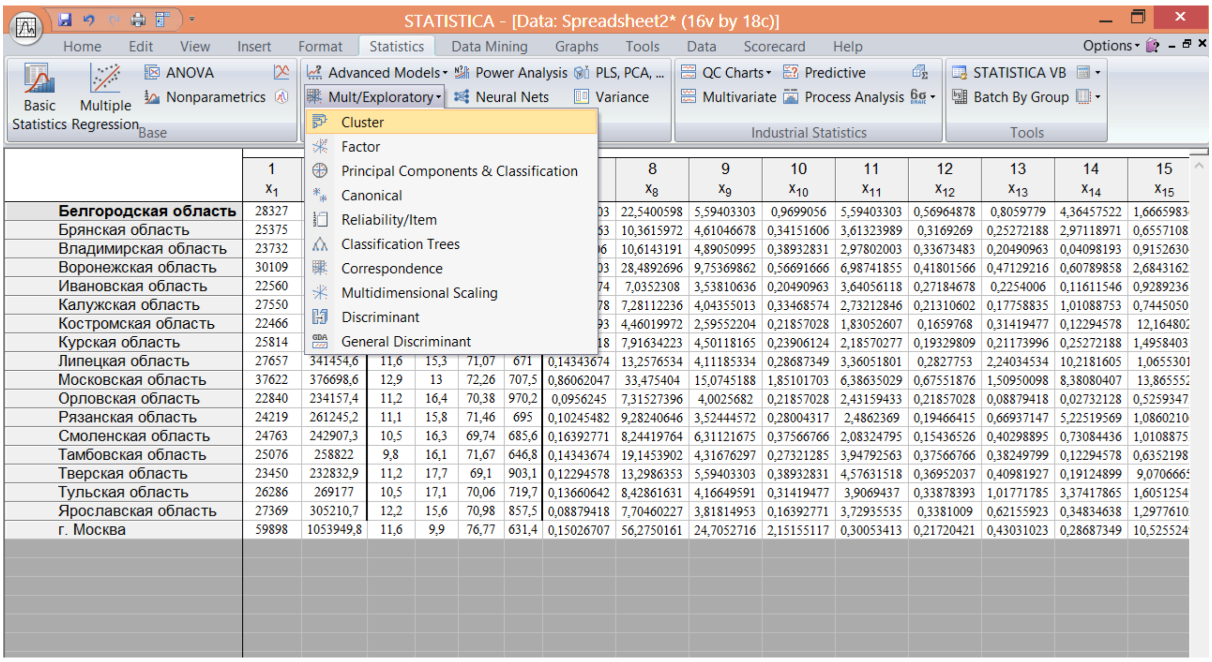


Рис. 8.8. Выбор кластерного анализа в программном комплексе STATISTICA

Далее выбираем Joining (tree clustering), т.е. иерархические агломеративные методы (рис. 8. 9).

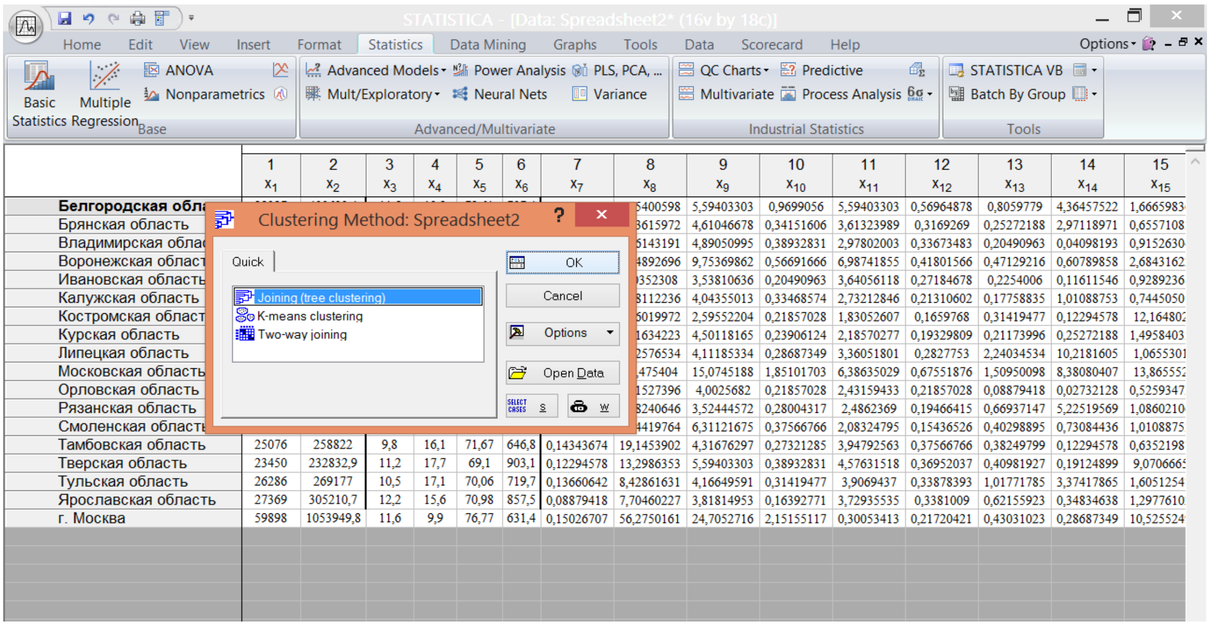


Рис. 8.9. Выбор иерархических агломеративных методов кластерного анализа в программном комплексе STATISTICA

В новом открывшемся диалоговом окне необходимо выбрать данные для анализа. Для этого нажимаем на Variables (рис. 8.10).

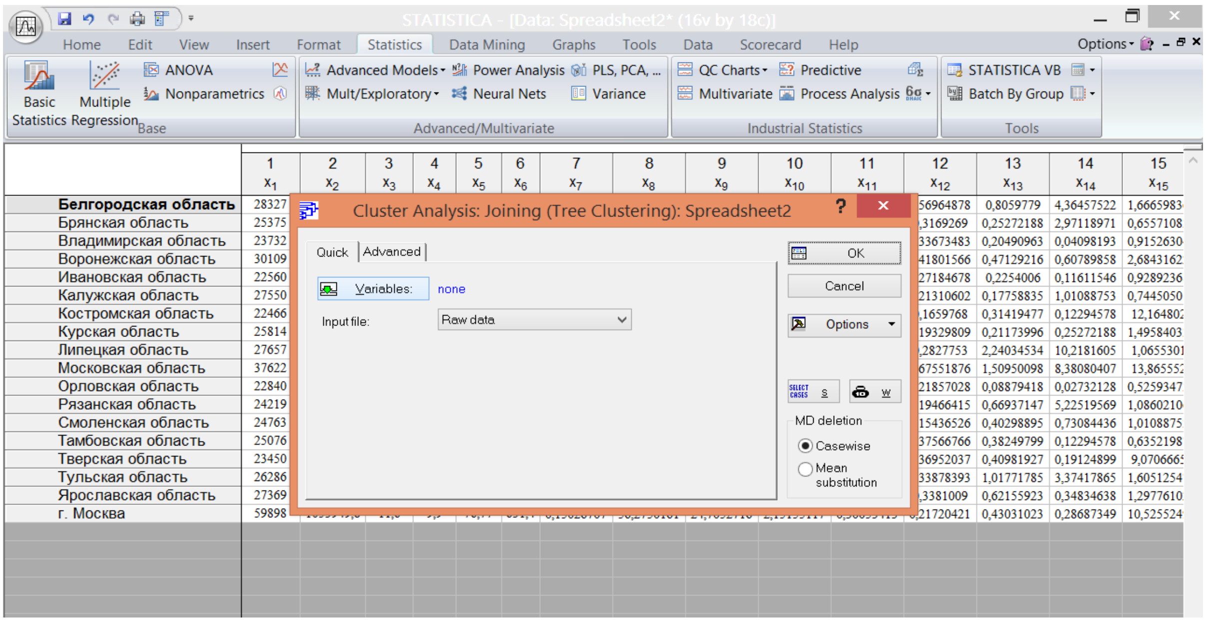


Рис. 8.10. Осуществление перехода в диалоговое окно по выбору переменных для кластерного анализа в программном комплексе STATISTICA

Поскольку предполагается осуществить кластерный анализ на основе всех 16 переменных, выбираем Select All, т. е. «Выбрать все» (рис. 8.11) и ОК.

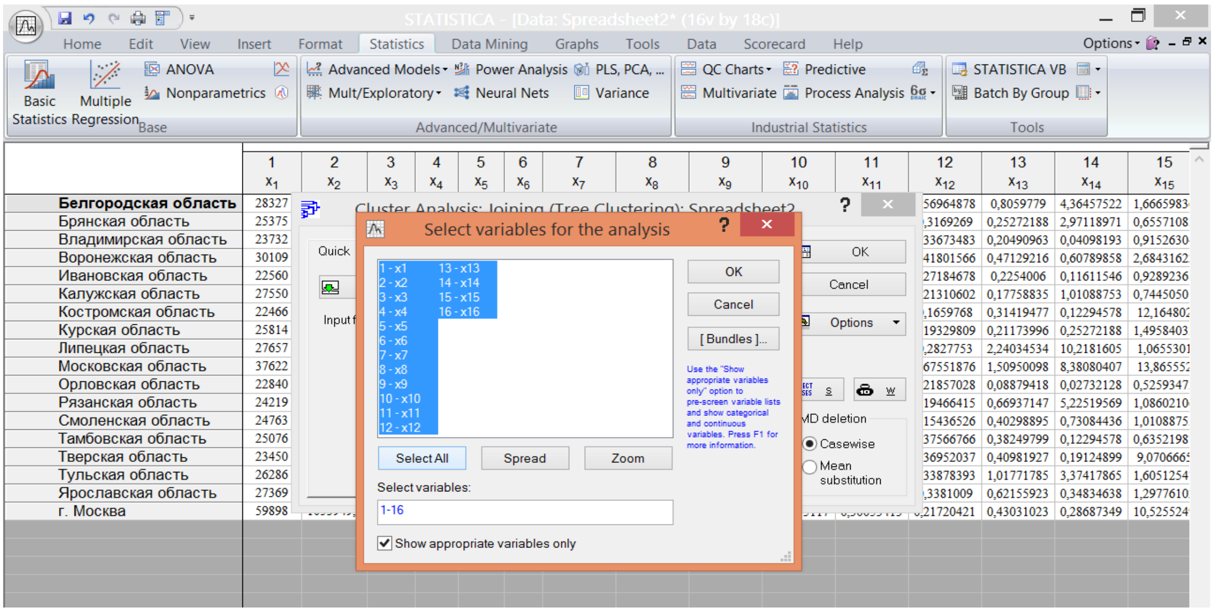


Рис. 8.11. Выбор переменных для кластерного анализа в программном комплексе STATISTICA

Цель анализа – упорядочение субъектов ЦФО РФ на основе 16 показателей, т. е. в результате необходима группировка регионов, для этого следует перейти во вкладку Advanced и в поле Cluster изменить Variables на Cases (рис. 8.12).

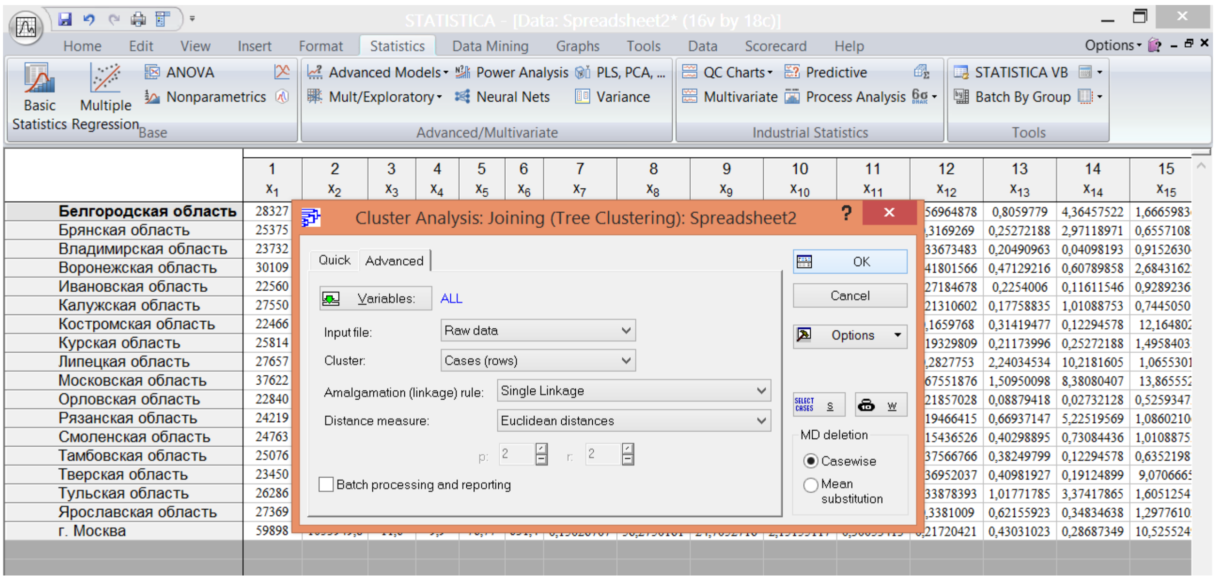


Рис. 8.12. Выбор наблюдений в качестве объектов для осуществления кластерного анализа в программном комплексе STATISTICA

В поле Amalgamation (linkage) rule (Правило объединения) выбирается Complete Linkage (метод полной связи) (рис. 8.13).

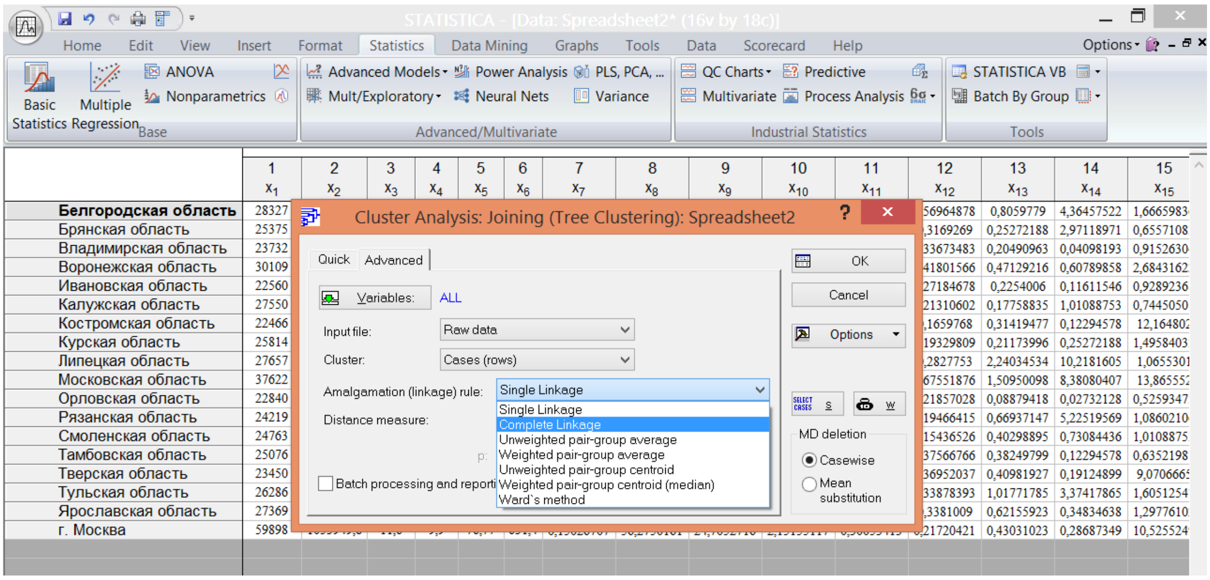


Рис. 8.13. Выбор метода полной связи в качестве правил объединения для осуществления кластерного анализа в программном комплексе STATISTICA

Далее необходимо осуществить выбор вида бинарного дерева кластеров для построения дендрограммы. В данном примере выбираем вертикальный вид – Vertical icicle plot (рис. 8.14).

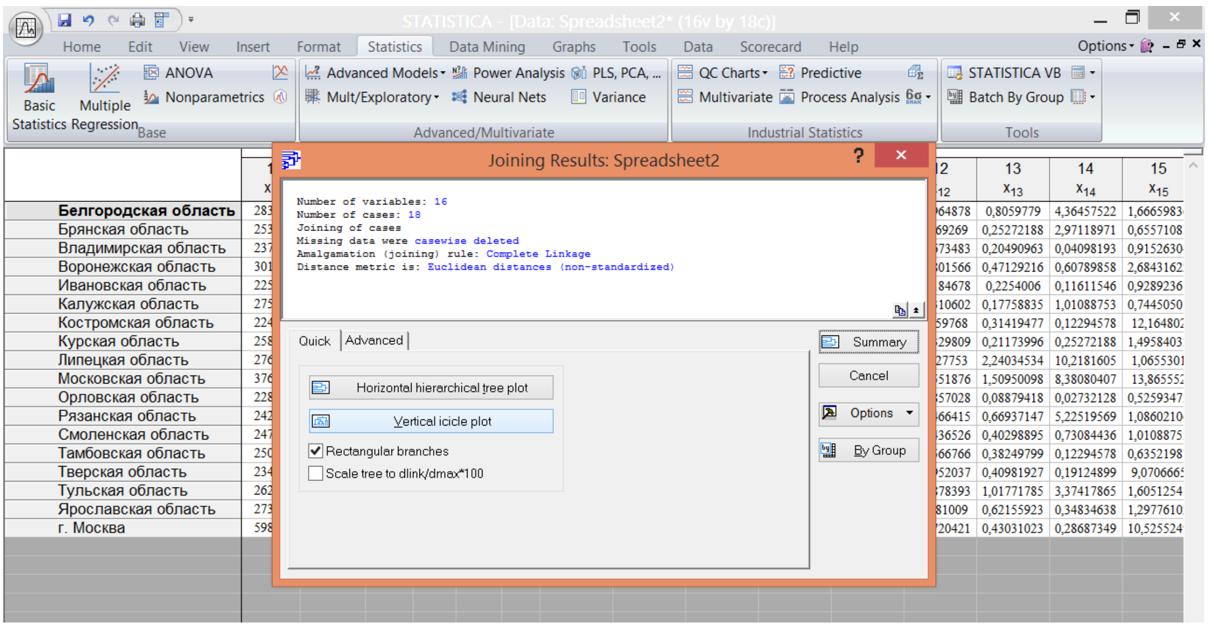


Рис. 8.14. Выбор вертикального вида бинарного дерева кластеров в программном комплексе STATISTICA

В результате на экран выводится дендрограмма (рис. 8.14).

На представленной дендрограмме можно выделить три укрупненные группы регионов:

1) Москва. Данный город федерального назначения имеет наиболее благоприятные значения анализируемых параметров;

2) Липецкая, Калужская, Ярославская, Воронежская, Московская и Белгородская области, которые имеют умеренные величины переменных;

3) Тамбовская, Рязанская, Тульская, Курская, Костромская, Смоленская, Орловская, Тверская, Владимирская, Ивановская и Брянская области, которые по анализируемым показателям уступают представителям предыдущих двух групп.

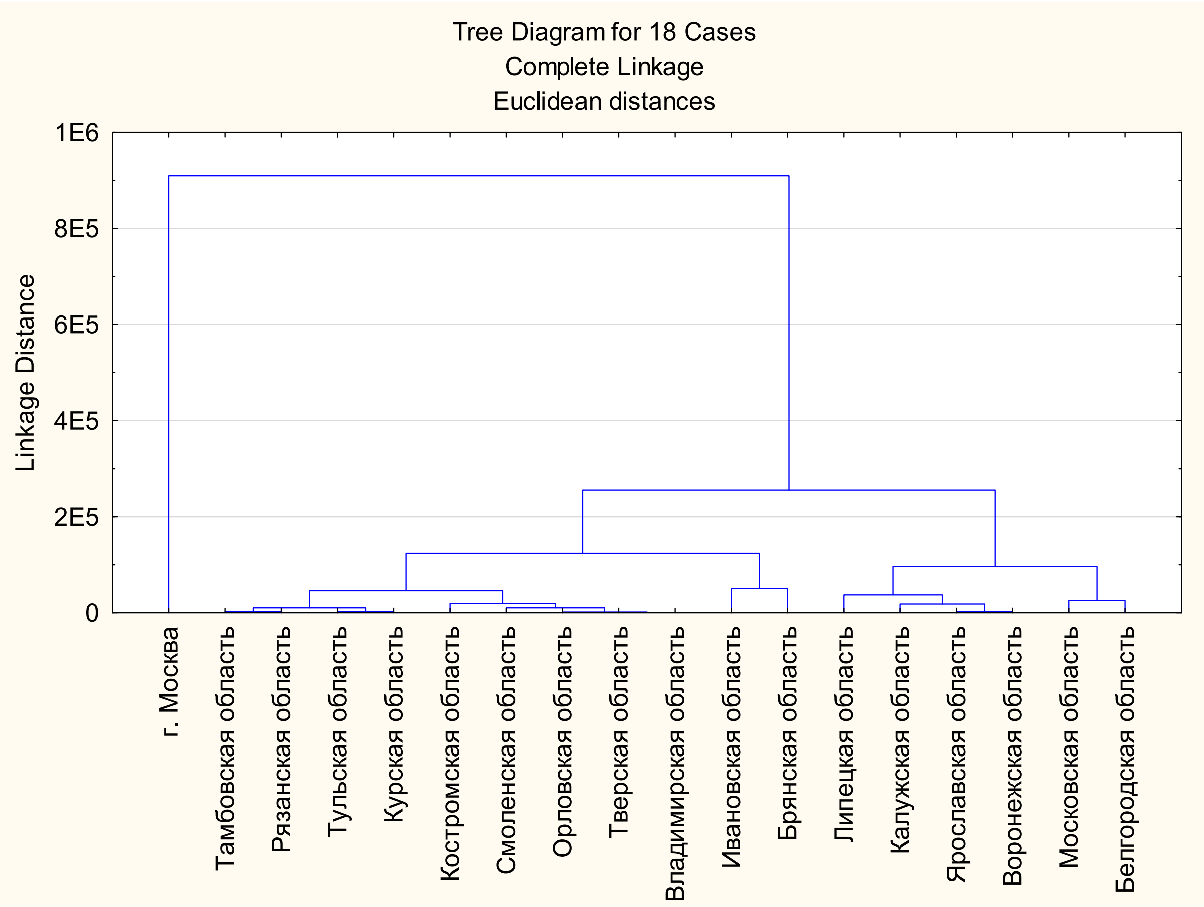


Рис. 8.15. Дендрограмма по результатам кластерного анализа в программном комплексе STATISTICA

Контрольные вопросы

1. Что такое кластерный анализ?

2. В чем заключается задача кластерного анализа?

3. Перечислите этапы кластерного анализа.

4. Назовите методы кластерного анализа.

5. В чем заключается подготовка данных для кластерного анализа?

6. Что такое дендрограмма?

7. В чем заключаются основные приемы чтения дендрограмм?

Практическое задание

Порядок выполнения работы

1. По табл. 8.1 определить федеральный округ, показатели регионов которого необходимо подвергнуть кластерному анализу.

2. Расчеты производить по данным статистического сборника «Регионы России. Социально-экономические показатели», для этого необходимо перейти на сайт http://www.gks.ru, выбрать вкладку «Официальная статистика», выбрать «Публикации», затем «Каталог публикаций». Далее переходим в раздел «Статистические сборники» и выбираем «Регионы России. Социально-экономические показатели».

3. В качестве переменных на основе данных статистического сборника «Регионы России. Социально-экономические показатели» необходимо использовать следующие:

1. Среднедушевые денежные доходы (в месяц), руб.
2. Валовой региональный продукт на душу населения, руб.*
3. Общие коэффициенты рождаемости (число родившихся на 1000 человек населения).
4. Общие коэффициенты смертности (число умерших на 1000 человек населения).
5. Ожидаемая продолжительность жизни при рождении (число лет).
6. Заболеваемость на 1000 человек населения (зарегистрировано заболеваний у пациентов с диагнозом, установленным впервые в жизни).
7. Уровень оснащенности стадионами с трибунами на 1500 мест и более (единиц на 1 млн чел.)**.
8. Уровень оснащенности плоскостными спортивными сооружениями (единиц на 1 млн чел.)**.
9. Уровень оснащенности спортивными залами (единиц на 1 млн. чел.)**.
10. Уровень оснащенности плавательными бассейнами (единиц на 1 млн чел.)**.

11. Уровень оснащенности детскими оздоровительными лагерями (единиц на 1 млн чел.)**.

12. Удельное количество детей, отдохнувших в детских оздоровительных лагерях за лето (на 1000 чел.)**.

13. Выбросы загрязняющих веществ в атмосферный воздух, отходящих от стационарных источников (тыс. т).

14. Улавливание загрязняющих атмосферу веществ, отходящих от стационарных источников (тыс. т).

15. Использование свежей воды (млн $м^3$).

16. Сброс загрязненных сточных вод в поверхностные водные объекты (млн $м^3$).

* Данный показатель рассматривается за предыдущий период, что обусловлено особенностями сбора статистических данных.

** Нахождение показателей необходимо осуществить за счет деления на среднегодовую численность населения (тыс. чел.).

Исходные данные представить в таблице исходных данных.

4. Открыть программный комплекс STATISTICA и осуществить кластерный анализ по алгоритму, представленному выше.

5. Построить дендрограмму по результатам кластерного анализа в программном комплексе STATISTICA.

6. Выделить группы, сформированные в результате кластерного анализа.

7. Сделать выводы.

Варианты заданий представлены в табл. 8. 1.

Таблица 8.1. Варианты заданий

Федеральный округ	**Номер варианта**
Северо-Западный	1
Южный	2
Северо-Кавказский	3
Приволжский	4
Уральский	5
Сибирский	6
Дальневосточный	7
Центральный	8

Исходные данные для кластерного анализа

Регионы	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
Регион 1																
…																
Регион *n*																

ЗАКЛЮЧЕНИЕ

В завершение следует отметить, что курс «Статистическое моделирование и прогнозирование» требует хорошей математической подготовки студентов, знания курса статистики, особенно раздела «Ряды динамики», владения навыками работы в Microsoft Excel.

В свою очередь, освоение программы курса «Статистическое моделирование и прогнозирование» позволит студентам выявлять тенденции и рассчитывать прогнозные значения показателей на уровне предприятий, регионов, отраслей.

Отметим, что для получения качественного прогноза необходимо не только знание методики его проведения, но и понимание сути того явления, по которому необходимо построить прогноз. Сейчас для построения прогнозных оценок применяют пакет электронных таблиц Excel, специализированные программы, например, STATISTICA или SPSS.

В последние годы на рынке труда возрос спрос на специалистов-аналитиков в разных областях. При этом особенно ценятся аналитики, владеющие навыками анализа ситуации как в статике, так и в динамике с расчетом прогнозных оценок. Успешное освоение данного курса позволит студентам овладеть соответствующими компетенциями и повысить свой профессиональный рейтинг.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Айвазян, С. А. Эконометрика. Краткий курс : учеб. пособие / С. А. Айвазян, С. С. Иванова. – М. : Маркет ДС, 2010. – 104 с. – (Университетская серия). – ISBN 978-5-94416-068-3.

2. Громыко, Г. Л. Теория статистики : практикум / Г. Л. Громыко. – 3-е изд., перераб. и доп. – М. : Инфра-М, 2004. – 160 с. – (Высшее образование). – ISBN 5-16-000583-8.

3. Дуброва, Т. А. Статистические методы прогнозирования : учеб. пособие / Т. А. Дуброва ; Моск. гос. ун-т экономики, статистики и информатики. – М. : МЭСИ, 1999. – 87 с.

4. *Она же.* Прогнозирование социально-экономических процессов : учеб. пособие / Т. А. Дуброва. – 2-е изд., испр. и доп. – М. : Маркет ДС, 2010. – 192 с. – (Университетская серия). – ISBN 978-5-94416-066-9.

5. Дубров, А. М. Многомерные статистические методы : учебник / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. – М. : Финансы и статистика, 2011. – 352 с. – ISBN 5-279-01945-3.

6. Ефимова, М. Р. Практикум по общей теории статистики : учеб. пособие / М. Р. Ефимова, О. И. Ганченко, Е. В. Петрова. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 2006. – 336. – ISBN 5-279-02555-0.

7. Ефимова, М. Р. Общая теория статистики : учеб. для вузов / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. – 2-е изд., испр. и доп. – М. : Инфра-М, 2006. – 413 с. – (Высшее образование). – ISBN 5-16-002179-5.

8. Салин, В. Н. Курс теории статистики для подготовки специалистов финансово-экономического профиля : учебник / В. Н. Салин, Э. Ю. Чурилова. – М. : Финансы и статистика, 2007. – 480 с. – ISBN 978-5-279-03063-7.

9. Статистика : учеб. для вузов / под. ред. И. И. Елисеева. – М. : Проспект, 2005. – 444 с. – ISBN 5-482-00031-1.

10. Теория статистики : учеб. для вузов / Г. Л. Громыко [и др.] ; под ред. Г. А. Громыко ; МГУ им. М. В. Ломоносова. – 2-е изд., перераб. и доп. – М. : Инфра-М, 2006. – 475 с. – (Классический университетский учебник). – ISBN 5-16-002158-2.

11. Теория статистики : учеб. для вузов / Р. А. Шмойлова [и др.] ; под ред. Р. А. Шмойловой ; МЭСИ. – 4-е изд., перераб. и доп. – М. : Финансы и статистика, 2006. – 655 с. – ISBN 5-279-02559-3.

12. Шмойлова, Р. А. Практикум по теории статистики : учеб. пособие для экон. специальностей вузов / Р. А. Шмойлова, В. Г. Минашкин, Н. А. Садовникова ; под ред. Р. А. Шмойловой. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 2006. – 415 с. – ISBN 5-279-02558-5.

13. Цены в России. 2016 : стат. сб. / Росстат. – М., 2016. – 151 с. – ISBN 978-5-89476-425-2.

14. Четыркин, Е. М. Статистические методы прогнозирования / Е. М. Четыркин. – М. : Статистика, 1975.

15. Федеральная служба государственной статистики [Электронный ресурс]. – URL: http://www.gks.ru (дата обращения: 12.01.2016).

ПРИЛОЖЕНИЯ

Приложение 1

Значения *t*-критерия Стьюдента при уровне значимости 0,10; 0,05; 0,01

df(v)	α			*df(v)*	α		
	0,10	**0,05**	**0,01**		**0,10**	**0,05**	**0,01**
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Приложение 2

Критические значения коэффициентов автокорреляции (r_a) при уровнях значимости α = 0,05 и α = 0,01

Объем выборки *n*	Положительное значение		Отрицательное значение	
	α = 0,05	α = 0,01	α = 0,05	α = 0,01
5	0,253	0,297	–0,753	–0,798
6	0,345	0,447	–0,708	–0,863
7	0,370	0,510	–0,674	–0,799
8	0,371	0,531	–0,625	–0,764
9	0,366	0,533	–0,593	–0,737
10	0,360	0,525	–0,564	–0,705
11	0,353	0,515	–0,539	–0,679
12	0,348	0,505	–0,516	–0,655
13	0,341	0,495	–0,497	–0,634
14	0,335	0,485	–0,479	–0,615
15	0,328	0,475	–0,462	–0,597
20	0,299	0,432	–0,399	–0,524

ОГЛАВЛЕНИЕ

Учебное издание

МАРЧЕНКО Елизавета Маратовна
РАХОВА Мария Владимировна
ХОЛОДНАЯ Анна Константиновна
и др.

СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ

Учебное пособие

Редактор Е. С. Глазкова
Технический редактор А. В. Родина
Корректор О. В. Балашова
Компьютерная верстка Е. А. Кузьминой
Выпускающий редактор А. А. Амирсейидова

Подписано в печать 06.11.18.
Формат 60×84/16 Усл. печ. л. 5,81. Тираж 50 экз.
Заказ
Издательство
Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых.
600000, Владимир, ул. Горького, 87.