Содержание

Аннотация ……3
1. Введение……3
2. Определение временного ряда……3
3. Генезис наблюдений, образующих временной ряд……4
4. Основные задачи анализа временных рядов……5
5. Методы сглаживания временного ряда……5
6. Метод сезонной декомпозиции……9
7. Практический пример. **Прогноз.**……14
8. Программа……15
Список использованной литературы……19

Аннотация. Временным рядом называют последовательность наблюдений, обычно упорядоченную во времени. Основной чертой, выделяющей анализ временных рядов среди других видов статистического анализа, является существенность порядка, в котором производятся наблюдения. Если во многих задачах наблюдения статистически независимы, то во временных рядах они, как правило, зависимы и характер этой зависимости может определяться положением наблюдений в последовательности. Почти в каждой области встречаются явления, которые важно изучать в их развитии и изменении во времени. В повседневной жизни могут представлять интерес, например, метеорологические условия, цены на тот или иной товар. С течением времени изменяются деловая активность, режим протекания того или иного производственного процесса. В тех случаях, когда приближенно выделяется периодическая составляющая у временного ряда (как правило, связанная с небольшими временными интервалами) можно выделить периодическую составляющую у него и сделать прогноз. Одним из методов является метод сезонной декомпозиции. В данной работе мы излагаем его теоретическое обоснование, а затем рассмотрим практический пример. Для этого примера мы проведем вычисления, используя пакет Maple.

1. Введение. Существуют две основные цели анализа временных рядов: (1) определение природы ряда и (2) прогнозирование (предсказание будущих значений временного ряда по настоящим и прошлым значениям). Обе эти цели требуют, чтобы модель ряда была идентифицирована и, более или менее, формально описана. Как только модель определена, вы можете с ее помощью интерпретировать рассматриваемые данные (например, использовать в вашей теории для понимания сезонного изменения цен на товары, если занимаетесь экономикой). Затем вы можете экстраполировать затем ряд на основе найденной модели, т.е. предсказать его будущие значения.

Пусть мы наблюдаем последовательность $x_1, x_2, \dots, x_n$. Достаточно общей статистической моделью служит модель вида

$$x_t = f(t) + u_t, t = 1, 2, \dots, T.$$

В этой модели наблюдаемый ряд рассматривается как сумма некоторой полностью детерминированной последовательности $\{f(t)\}$, которую можно назвать систематической составляющей, и случайной последовательности $\{u_t\}$, подчиняющейся некоторому вероятностному закону. Его компоненты ненаблюдаемы; они являются теоретическими величинами. Например, если производится измерение количества ежедневно выпадаемых осадков, то $f(t)$ может представлять собой климатическую норму, получающуюся долговременным усреднением за большой период, а u_t те капризы и нерегулярности в погоде, которые характеризуют отклонения от климатической нормы. Мы полагаем, что по крайней мере принципиально можно повторять ситуацию целиком, получая новые совокупности наблюдений. При таком повторении эксперимента функция $f(t)$ должна была бы оставаться одной и той же, а случайные составляющие оказывались бы различными как различные реализации случайного процесса. Случайные составляющие, кроме всего прочего, могут включать в себя и ошибки наблюдений.

2. Определение временного ряда. Говоря о расположенной в хронологическом порядке последовательности наблюденных значений

$$x_1, x_2, \dots, x_N \qquad (1)$$

какого-либо признака, мы, по существу, уже дали определение временного ряда. Однако в дальнейшем нас будет интересовать лишь некоторый подкласс подобных последовательностей, а именно тот, который связан с наблюдениями стохастических по своей природе признаков. Другими словами, мы исключаем из рассмотрения детерминированные схемы дина-

мических наблюдений, при которых элементы последовательности (1) могут быть в точности вычислены как значения некоторой неслучайной функции $f(t)$, т.е $x_t = f(t), t = 1, 2, \ldots$. Поэтому уточним понятие временного ряда.

Определение 1. *Ряд наблюдений* $x_{t_1}, x_{t_2}, \ldots, x_{t_N}$ *анализируемой случайной величины* $\xi(t)$, *произведенных в последовательные моменты времени* $t_1, t_2, \ldots, t_N$ *называется* ***временным рядом.***

Предметом нашего исследования будут временные ряды с равноотстоящими моментами наблюдений, а это позволяет представлять их в форме (1).

Определение 1 опирается на понятие случайной величины $\xi(t)$, зависящей от параметра *t,* интерпретируемого как время. То есть, по существу, речь идет об однопараметрическом семействе случайных величин $\{\xi(t)\}$. Это значит, что закон распределения вероятностей этих случайных величин, и в частности, их первые и вторые моменты, также, вообще говоря, могут зависеть от времени *t.*

В чем состоит принципиальное отличие временного ряда (1) от введенной последовательности наблюдений $x_1, x_2, \ldots, x_n$ образующих случайную выборку? Этих отличий два: а) во-первых, в отличие от элементов случайной выборки члены временного ряда не являются независимыми; б) во-вторых, члены ряда (1) не являются одинаково распределенными. Это значит, что мы не можем распространять свойства и правила статистического анализа случайной выборки на временные ряды. С другой стороны, взаимозависимость членов временного ряда создает свою специфическую базу для построения прогнозных значений анализируемого показателя (т.е. для построения оценок x_{N+k} для неизвестных значений x_{N+k}) по наблюдениям $x_1, x_2, \ldots, x_N$.

3. Генезис наблюдений, образующих временной ряд. Речь идет о структуре и классификации основных факторов, под воздействием которых формируются значения элементов временного ряда. Целесообразно выделить следующие 4 типа таких факторов.

(А) *Долговременные*, формирующие общую (в длительной перспективе) тенденцию в изменении анализируемого признака x_t. Обычно эта тенденция описывается с помощью той или иной неслучайной функции $m(t)$, медленно меняющейся во времени. Эту функцию называют *функцией тренда* или просто – *трендом.*

(Б) *Циклические*, формирующие изменения анализируемого признака, обусловленные действием долговременных циклов экономической, демографической или астрофизической природы (волны Кондратьева, демографические ямы, циклы солнечной активности, выборы и т.д.). Результат действия циклических факторов будем обозначать с помощью неслучайной функции $\psi(t)$. Ее период достаточно большой.

(В) *Сезонные*, формирующие периодически повторяющиеся в определенное время года колебания анализируемого признака. Условимся обозначать результат действия сезонных факторов с помощью неслучайной функции $\varphi(t)$. Поскольку эта функция должна быть периодической, в ее аналитическом выражении участвуют гармоники (тригонометрические функции), период которых определяется содержательной сущностью задачи. Ее период, как правило, небольшой.

(Г) *Случайные*, не поддающиеся учету и регистрации. Их воздействие на формирование значений временного ряда как раз и обусловливает стохастическую природу элементов x_t, а следовательно, и необходимость интерпретации $x_1, x_2, \ldots, x_N$ как наблюдений, произведенных над случайными величинами $\xi_1, \xi_2, \ldots, \xi_N$. Будем обозначать результат воздействия случайных факторов с помощью случайных величин (остатков, ошибок) $\varepsilon(t)$. В свою очередь этот остаток разлагают на авторегрессию (АР) и модель скользящего среднего (СС).

Конечно, вовсе не обязательно, чтобы в процессе формирования значений всякого временного ряда участвовали одновременно факторы всех четырех типов. Из приведенных ниже примеров мы увидим, что в одних случаях значения временного ряда формируются под воздействием факторов (А), (В), (Г) (пример 1), в других – под воздействием факторов (А), (В/Б), (Г) (пример 2) и, наконец, – исключительно под воздействием одних только случайных факторов (Г) (пример 3). Однако во всех случаях предполагается непременное участие случайных (эволюционных) факторов (Г). Кроме того, мы примем в качестве гипотезы аддитивную структуру схемы влияния факторов А, Б, В и Г на формирование значений x_t, которая означает правомерность представления значений членов временного ряда в виде разложения

$$x_t = a_1 m(t) + a_2 \psi(t) + a_3 \varphi(t) + \varepsilon(t), t = 1, 2, \ldots, N. \qquad (2)$$

Выводы о том, участвуют или нет факторы данного типа в формировании значений x_t, могут базироваться как на анализе содержательной сущности задачи, так и на специальном статистическом анализе исследуемого временного ряда.

4. Основные задачи анализа временных рядов. Отправляясь от проведенного аддитивного разложения (2) временного ряда X_t, можно дать общую формулировку базисной цели его статистического анализа[2]:

- *по имеющейся траектории* (1) *анализируемого временного ряда* x_t *требуется:*

(i) *определить, какие из неслучайных функций* $m(t), \varphi(t)$ *и* $\psi(t)$ *присутствуют в разложении* (2), *т.е. определить значения коэффициентов* $a_i, i = 1, 2, 3$ *в разложении* (2)*;*

(ii) *построить «хорошие» оценки для тех неслучайных функций, которые присутствуют в разложении* (2);

(iii) *подобрать модель, адекватно описывающую поведение «случайных остатков»* $\varepsilon(t)$ *и статистически оценить параметры этой модели.*

Успешное решение задач (i) – (iii), обусловленных базисной целью статистического анализа временного ряда, является основой для достижения конечных прикладных целей исследования и, в первую очередь, для решения задачи кратко- и среднесрочного прогноза значений временного ряда.

Одно из главных отличий последовательности наблюдений, образующих временной ряд, от случайной выборки заключается в том, что члены временного ряда являются, вообще говоря, статистически взаимосвязанными. Степень тесноты статистической связи между двумя случайными величинами может быть измерена парным коэффициентом корреляции.

5. Методы сглаживания временного ряда. Методы выделения неслучайной составляющей в траектории, отражающей поведение анализируемого временного ряда (1), можно условно разделить на два типа.

Методы *первого* типа (аналитические) основаны на допущении, что нам известен общий вид неслучайной составляющей в разложении (2)

$$f(t) = a_1 m(t) + a_2 \psi(t) + a_3 \varphi(t), \qquad (3)$$

Например, рассматривая график временного ряда из примера 2 (см. рис. 2), исследователь может предложить, что неслучайная составляющая динамики курса анализируемых акций описывается линейной функцией времени *t*, т.е.

$$f(t;\theta) = f(t) = \theta_0 + \theta_1 t,$$

где θ_0 и θ_1 — некоторые неизвестные (подлежащие статистическому оцениванию) параметры модели. Тогда задача выделения неслучайной составляющей сведется к задаче построения хороших оценок $\hat{\theta}_0$ и $\hat{\theta}_1$ для параметров θ_0 и θ_1.

Соответствующие методы мы назвали аналитическими, поскольку на выходе задачи мы будем иметь явное аналитическое задание (оценку) $\hat{f}(t)$ для искомой неслучайной функции $f(t)$. Другими словами, *f(t)* будет представлена в виде формулы $f(t;\hat{\theta})$ функции известного вида, в которой неизвестные параметры $\theta = (\theta_0, \theta_1, \ldots, \theta_m)$ заменены их статистическими оценками $\hat{\theta}$.

Методы *второго* типа (алгоритмические) не связаны ограничительным допущением о том, что общий аналитический вид искомой функции (3) известен исследователю. В этом смысле они являются более гибкими, более привлекательными. Однако на выходе задачи они доставляют исследователю лишь алгоритм расчета оценки $\hat{f}(t)$ для искомой функции $f(t)$ в любой наперед заданной точке *t* и не претендуют на аналитическое (т.е. заданное в виде явной формулы) представление функции (3).

Аналитические методы выделения (оценки) неслучайной составляющей временного ряда. Эти методы реализуются в рамках моделей регрессии, в которых в роли зависимой (объясняемой) переменной выступает переменная x_t, генерирующая анализируемый временной ряд (1), а в роли единственной объясняющей переменной – время *t*.

Таким образом, рассматривается модель регрессии вида

$$x_t = f(t;\theta) + \varepsilon_t, t = 1, 2, \dots, n,$$

в которой общий вид функции $f(t;\theta)$ известен, но неизвестны значения параметров $\theta = (\theta_0, \theta_1, \dots \theta_m)^{\mathrm{T}}$. Оценки $\hat{\theta}$ параметров θ строятся по наблюдениям $\{x_t\}_{1 \le t \le n}$. Выбор метода оценивания зависит от гипотетического вида функции $f(t;\theta)$ и стохастической природы случайных регрессионных остатков ε_t. Так, напри мер, если функция $f(t;\theta)$ имеет вид алгебраического полинома степени ***p***, т.е.

$$f(x;\theta) = \theta_0 + \theta_1 t + \dots + \theta_p t^p \qquad (4)$$

и при этом длина временного ряда *n* существенно превышает степень этого полином p, а регрессионные остатки $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ взаимно некоррелированны, то оценки $\hat{\theta}$ параметров θ могут быть получены с помощью обычного метода наименьших квадратов

$$\hat{\theta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}. \qquad (5)$$

В этой формуле $n \times (p-1)$ - матрица $\mathbf{X}$ наблюденных значений объясняющих переменных имеет в данном случае вид:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1^2 & \dots & 1^p \\ 1 & 2 & 2^2 & \dots & 2^p \\ 1 & 3 & 3^2 & \dots & 3^p \\ . & . & . & \dots & . \\ 1 & n & n^2 & \dots & n^p \end{pmatrix}$$

а вектор-столбец наблюденных значений зависимой переменной $\mathbf{Y} = (x_1, x_2, \dots, x_n)^{\mathrm{T}}$.

Отказ от взаимной некоррелированности регрессионных остатков в данной модели повлечет за собой необходимость использования практически реализуемого обобщенного МНК.

Алгоритмические методы выделения неслучайной составляющей временного ряда (методы скользящего среднего). В основе этих методов элиминирования случайных флуктуации в поведении анализируемого временного ряда лежит простая идея: если «индивидуальный» разброс значений членов временного ряда *x(t)* около своего среднего (сглаженного) значения *a* характеризуется дисперсией σ^2, то разброс среднего из *N* членов временного ряда $(x(1) + x(2) + \dots + x(N))/N$ около того же значения *a* будет характеризоваться гораздо меньшей величиной дисперсии, а именно дисперсией, равной σ^2/N. А уменьшение меры случайного разброса (дисперсии) и означает как раз *сглаживание* соответствующей траектории. Поэтому выбирают некоторую *«длину усреднения»* $N = 2m + 1$, измеренную в числе подряд идущих членов анализируемого временного ряда (конкретный выбор числа *m* зависит от специфики исходных данных, но, как правило, *m* выбирают таким образом, чтобы оно удовлетворяло неравенству $m < n/3$ обычно *m* не превышает трех; более подробно о выборе *m* см. ниже). А затем сглаженное значение $f(t)$ временного ряда $x(t)$ вычисляют по значениям $x(t-m), x(t-m+1), \dots, x(t), x(t+1), \dots, x(t+m)$ по формуле

$$\hat{f}(t) = \sum_{k=-m}^{m} w_k\, x(t+k),\ t = m+1, m+2, \ldots, n-m, \qquad (6)$$

где w_k $(k = -m, -m+1, \ldots, m-1, m)$ – некоторые положительные «весовые» коэффициенты (или просто «веса»), в сумме равные единице, т.е. $w_k > 0$ и $\sum_{k=-m}^{m} w_k = 1$. Поскольку, изменяя t от $m+1$ до $n-m$, *мы как бы «скользим» по оси времени* (так что при переходе от t к $t+1$ в составе слагаемых правой части (20) происходит замена только одного слагаемого $x(t-m)$ слагаемым $x(t+m+1)$, то и методы, основанные на формуле (6), принято называть *методами скользящей средней (МСС).* Очевидно, один МСС отличается о другого выбором параметров m и w_k $(k = -m, -m+1, \ldots, m)$. Остановимся на этом подробнее.

Определение параметров w_k основано на следующей процедуре. В соответствии известной теоремой Вейерштрасса любая гладкая функция $f(t)$ на ограниченном интервале может быть аппроксимирована алгебраическим полиномом подходящей степени p Поэтому возьмем первые $2m+1$ членов временного ряда $x(1), x(2), \ldots, x(2m+1)$, построим с помощью МНК полином $\hat{x}_{p1}(t)$ степени p, аппроксимирующий поведение этой начальной части траектории временного ряда, и используем этот полином для определение оценки $\hat{f}(t)$ *сглаженного* значения $f(t)$ временного ряда в *средней* (т.е. $(m+1)$-й точке этого отрезка ряда), т.е. полагаем $\hat{f}(m+1) = \hat{x}_{p1}(m+1)$. Затем «скользим» по оси времени на один такт и таким же способом подбираем полином $\hat{x}_{p2}(t)$ той же степени p к отрезку временного ряда $x(2), x(3), \ldots, x(2m+2)$ и определяем оценку сглаженного значения временного ряда в средней точке сдвинутого на единицу отрезка временного ряда, т.е. $\hat{f}(m+2) = \hat{x}_{p2}(m+2)$, и т.д. В результате мы найдем оценки для сглаженных значений $\hat{f}(t)$ анализируемого временного ряда при всех t, кроме $t = 1, 2, \ldots, m$ и $t = n-m+1, n-m+2, \ldots, n-1, n$.

Покажем, что подбор наилучшего (по методу наименьших квадратов) аппроксимирующего полинома к траектории анализируемого временного ряда (на заданном временном интервале) действительно приводит к формуле (6), причем результат (т.е. значения коэффициентов w_k $(k = -m, -m+1, \ldots, m)$ не зависит от того, для какого именно из «скользящих» временных интервалов был осуществлен этот подбор.

<u>Начнем с примера.</u> Будем полагать, что локальное поведение сглаженной функции $f(t)$ описывается алгебраическим полиномом 1-й степени ($p = 1$), т.е. на выбранном временном интервале (протяженностью $2m+1$ тактов) неслучайная составляющая $f(t)$ может быть аппроксимирована *линейной функцией времени:*

$$f(t) = \theta_0 + \theta_1 t,\ t = 1, 2, \ldots, 2m+1.$$

Без ограничения общности можно переобозначить моменты времени таким образом, ч т о б ы р а с с м а т р и в а т ь н а ш у м о д е л ь н а в р е м е н н о м о т р е з к е $t' = -m, -m+1, \ldots, -1, 0, 1, \ldots, m-1, m$ (тогда средняя точка будет соответствовать моменту $t' = 0$). «Новое» (условное) время t' связано со «старым» временем t соотношением $t' = t - (m+1)$. Согласно методу наименьших квадратов имеем систему уравнений:

$$\begin{cases} (2m+1)\theta_0 + \theta_1 \sum_{t=-m}^{m} t = \sum_{t=-m}^{m} x(t), \\ \theta_0 \sum_{t=-m}^{m} t + \theta_1 \sum_{t=-m}^{m} t^2 = \sum_{t=-m}^{m} t \cdot x(t). \end{cases}$$

Поскольку $\sum_{t=-m}^{m} t = 0$, а оценка сглаженного значения временного ряда определяется в средней точке рассматриваемого интервала, соответствующих значениям функции $\theta_0 + \theta_1 t$ при $t = 0$, то

$$\hat{f}(m+1) = \hat{\theta}_0 + \hat{\theta}_1 t\big|_{t=0} = \hat{\theta}_0 = \frac{1}{2m+1}\sum_{t=-m}^{m} x(t).$$

Проделав тот же анализ для любого другого временного интервала $t = k+1, k+2, \ldots, k+2m+1 \; (k = m, m+1, \ldots, n-3m-1)$, мы получим аналогичный результат. Таким образом, в общем случае имеем:

$$\hat{f}(t) = \frac{1}{2m+1}\sum_{t=-m}^{m} x(t+k), \qquad (7)$$

т.е. при линейном характере локальной аппроксимации траектории временного ряда в качестве его сглаженного значения в точке t следует брать среднее арифметическое из окаймляющих его $2m+1$ соседних значений $x(t-m), x(t-m+1), \ldots, x(t), \ldots, x(t+m)$ с весами $w_{-m} = w_{-m+1} = \ldots = w_m = 1/(2m+1)$.

Метод экспоненциально взвешенного скользящего среднего (метод Брауна).

Рассмотренный метод скользящего среднего построения оценок $\hat{f}(t)$ для неслучайной составляющей $f(t)$ анализируемого временного ряда $x(t)$ основывался на МНК, в соответствии с которым все исходные статистические данные $\{x(t)\}, t = 1, 2, \ldots, n$, имели равный вес. Однако в задачах прогноза, в которых сглаженная функция $\hat{f}(t)$ используется обычно для экстраполяции в будущее, кажется более естественным сравнительно *недавним* исходным данным придавать больший вес, чем наблюдениям, относящимся к далекому прошлому. Одним из наиболее распространенным методом, реализующим эту идею, является ***метод экспоненциально взвешенного скользящего среднего*** **(МЭВСС).**

В соответствии с этим методом оценка сглаженного значения $\hat{f}(t)$ в точке t определяется как решение оптимизационной задачи вида

$$Q(f) = \sum_{k=0}^{t-1} \lambda^k (x(t-k) - f)^2 \to \min_f, \qquad (8)$$

где λ - некоторое положительное число, меньшее единицы (т.е. $0 < \lambda < 1$). Веса λ^k в критерии $Q(f)$ уменьшаются экспоненциально по мере удаления $x(t-k)$ в прошлое, т.е. по мере роста k). Решение оптимизационной задачи (8), т.е. дифференцирование $Q(f)$ по f, приравнивание полученной производной к нулю и решение полученного уравнения относительно f дает:

$$\hat{f}(t) = \frac{1-\lambda}{1-\lambda^t}\sum_{k=0}^{t-1} \lambda^k x(t-k). \qquad (9)$$

Однако, в отличие от обычного МСС, скользит только *правый* конец интервала усреднения (левый его конец закреплен в точке $t = 1$), а, во-вторых веса при $x(t-k)$ экспоненциально уменьшаются по мере роста k. Кроме того, формула (9) дает оценку сглаженного значения не в средней, а в правой конечной точке интервала усреднения.

Посмотрим, как преобразовались взаимно некоррелированные случайные остатки ε_t из разложения (2) при переходе от исходных значений временного ряда $x(t)$ к сглаженным $\hat{f}(t)$ по формуле (9).

Поскольку преобразованные остатки $\hat{\varepsilon}(t)$ связаны с исходными остатками $\varepsilon(t)$ тем же соотношением, которым связаны $\hat{f}(t)$ и $x(t)$ (см. (9)), то

$$\mathbf{E}(\varepsilon(t)) = 0, \mathbf{D}(\varepsilon(t)) = \sigma^2 \frac{(1-\lambda)^2}{(1-\lambda^t)^2} \sum_{k=0}^{t-1} \lambda^{2k} = \sigma^2 \frac{(1-\lambda)(1+\lambda^t)}{(1+\lambda)(1-\lambda^t)}. \qquad (10)$$

Из (10) видно, что при значениях λ, не слишком близких к 1, и для достаточно удаленных от прошлого значениях t случайные остатки $\varepsilon(t)$, а, следовательно, и сглаженные значения $\hat{x}(t)$, подвержены существенно меньшему случайному разбросу, чем $x(t)$.

Замечание (**случай бесконечно удаленного прошлого**). Если анализируемый временной ряд достаточно длинный (т.е. n достаточно велико), то для представляющих главный интерес в прогнозе «свежих» значений временного ряда $x(t)$, т.е. при t, близких к n) можно приближенно допускать, что их прошлое «уходит в бесконечность». Тогда соотношения (8), (9) и (10) могут быть заменены, соответственно, следующими соотношениями:

$$Q(f) = \sum_{k=0}^{\infty} \lambda^k (x(t-k) - f)^2 \to \min_f, \qquad (8')$$

$$\hat{x}(t) = (1-\lambda) \sum_{k=0}^{\infty} \lambda^k x(t-k). \qquad (9')$$

$$\mathbf{D}(\varepsilon(t)) = \sigma^2 \frac{(1-\lambda)}{(1+\lambda)}. \qquad (10')$$

Кроме того, в этом случае имеют место соотношения:

$$\boxed{\hat{x}(t) = \lambda \hat{x}(t-1) + (1-\lambda) x(t)}$$

6. Метод сезонной декомпозиции.

Для оценки сезонной составляющей мы будем использовать ***метод сезонной декомпозиции.*** Сезонная компонента проявляется, когда временной ряд составляют квартальные или месячные наблюдения. Рассмотрим уровень ряда только как результирующую сезонности и тренда, т. е. представим его как произведение тренда и сезонной компоненты:

$$x(t) = m(t) + \psi(t) + \varepsilon(t).$$

Пример 1. В табл. 1 приведены квартальные данные о продажах фирмы за период 1990—1993 гг. (в млн долл.), где в скобках указаны обозначения соответствующих уровней временного ряда.

Таблица1 Квартальные данные об объеме продаж (млн долл.)

Год	Квартал 1	Квартал 2	Квартал 3	Квартал 4
1990	**20 (** $x(1)$ **)**	**12 (** $x(2)$ **)**	**47 (** $x(3)$ **)**	**60 (** $x(4)$ **)**
1991	**40 (** $x(5)$ **)**	**32 (** $x(6)$ **)**	**65 (** $x(7)$ **)**	**75 (** $x(8)$ **)**
1992	**56 (** $x(9)$ **)**	**50 (** $x(10)$ **)**	**85 (** $x(11)$ **)**	**100 (** $x(12)$ **)**
1993	**75 (** $x(13)$ **)**	**70 (** $x(14)$ **)**	**101 (** $x(15)$ **)**	**123 (** $x(16)$ **)**

Будем находить индекс сезонности для каждого квартала в течение года, т. е. вычислим четыре значения индекса. Идея метода отношения к центрированной скользящей средней состоит в том, что вначале на основе исходного временного ряда определяется новый временной ряд, не содержащий компоненту сезонности. Уровни нового ряда рассчитываются как центрированные скользящие средние.

Для вычисления центрированных скользящих средних определяются так называемые скользящие суммы. Для квартальной сезонности первая скользящая сумма будет включать значения первых четырех квартальных уровней исходного временного ряда:

$$y(1) = x(1) + x(2) + x(3) + x(4) = 20 + 12 + 47 + 60 = 139.$$

Во вторую скользящую сумму входят первые четыре уровня ряда, сдвинутого на один квартал вперед: $y(2) = x(2) + x(3) + x(4) + x(5) = 12 + 47 + 60 + 40 = y(1) - x(1) + x(5) = 139 - 20 + 40 = 159$.

Третья скользящая сумма определяется аналогично при сдвиге уровней на два квартала вперед:

$$y(3) = y(2) - x(2) + x(5) = 159 - 12 + 32 = 20 + 12 + 47 + 60 = 179.$$

Продолжая эту процедуру, получим 13 скользящих сумм (табл. 1), где последняя сумма вычисляется следующим образом:

$$y(13) = x(13) + x(14) + x(15) + x(16) = 75 + 70 + 101 + 123 = 369.$$

Таблица 2. **Вычисление центрированных скользящих средних**

Год	Квартал	*t*	*x(t)*	Скользящая сумма	Центрированная скользящая средняя	Отношение к центрированной скользящей средней
1990	1	1	20		–	–
	2	2	12		–	–
	3	3	47	139	37,25	1,2617
	4	4	60	159	42,25	1,42
1991	1	5	40	179	47,00	0,85
	2	6	32	197	51,25	0,62
	3	7	65	213	55,25	1,18
	4	8	76	229	59,50	1,28
1992	1	9	56	247	64,25	0,87
	2	10	50	267	69,75	0,72
	3	11	85	291	75,13	1,13
	4	12	100	310	80,00	1,25
1993	1	13	75	330	84,50	0,89
	2	14	70	346	89,38	0,78
	3	15	101	369	–	–
	4	16	123	–	–	–

Каждая скользящая сумма будет относиться к моменту времени, находящемуся посередине между периодами, на основании которых она была рассчитана. Если на основе каждой скользящей суммы вычислять квартальные средние (т. е. скользящие средние, полученные делением скользящих сумм на 4), то полученные величины будут относиться к моментам между кварталами. Например, первая скользящая средняя (или сумма) соответствует моменту между $t = 2$ и $t = 3$, т. е. $t = 2{,}5$ (конец июня – начало июля 1990 г.). Аналогично вторая скользящая сумма относится к $t = 3{,}5$ (конец сентября – начало октября 1990 г.) и т. д.

Для того чтобы скользящая средняя относилась непосредственно к периоду t (в данном случае к середине квартала t), следует вычислить двухлетние скользящие суммы (последовательно суммируя две соседние четырехквартальные скользящие суммы) и разделить их на 8. Например, вычислим первую скользящую квартальную среднюю, центрированную на середину квартала $t = 3$:

$$\frac{139 + 159}{8} = 37.25, \quad \frac{47}{37.25} = 1.2617,$$

Аналогично для $t = 4$ имеем

$$\frac{159 + 179}{8} = 42.25, \quad \frac{60}{42.75} = 1.42, \text{ и т. д.}$$

С помощью данной процедуры вычисляются 12 центрированных скользящих средних (табл. 2). Понятно, что их расчет для *t* = 1, 2, 15 и 16 невозможен. Заметим, что первые два значения из них соответствуют кварталам 1 и 2, а последние два – кварталам 3 и 4. В общем случае, если временной ряд содержит *T* наблюдений, то для определения квартальной сезонности можно вычислить $T - 4$ центрированных скользящих средних.

Скользящая сумма и центрированные скользящие средние определяются суммированием за четыре квартала (сезона). Поэтому они уже не содержат сезонной составляющей. В результате усреднения снижается также влияние нерегулярной компоненты

Процедура получения центрированных скользящих средних является, по существу, сглаживанием временного ряда. Она позволяет установить существование тренда, а также выявить его форму (прямая или кривая линия).

Центрированные скользящие средние, представленные в табл. 2, характеризуют стабильно возрастающий тренд. Ввиду того что разности между соседними средними почти одинаковы, тренд будет очень близок к линейному. Это можно наблюдать на рис. 5.

Для определения четырех квартальных сезонных индексов следует разделить значение каждого уровня $x(t)$ на соответствующую центрированную скользящую среднюю.

Как видно, вычисление отношения выявляет сезонный эффект в совокупности с нерегулярной составляющей. В табл. 2 представлены отношения $t = 3, 4, 5, \ldots, 14$. Они определялись следующим образом:

для $t = 3$ 47/37,25 = 1,26 (3-й квартал 1990 г.);

для $t = 4$ 60/42,25 = 1,42 (4-й квартал 1990 г.);

……

для $t = 14$ 70/89,38 = 0,78 (2-й квартал 1993 г.).

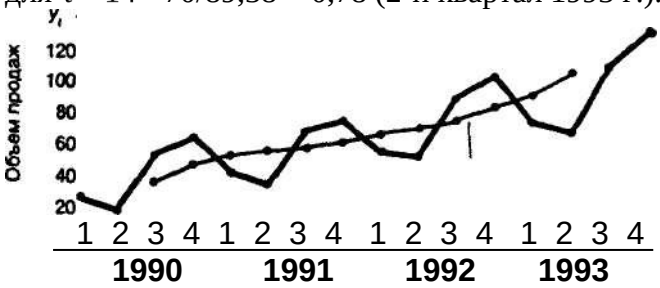


Рис. 5. Сглаживание временного ряда объемов продаж фирмы (ед.) на основе скользящих средних

Сведем все 14 отношений по кварталам и вычислим средние значения по каждому кварталу (табл. 3). Эти средние значения будем рассматривать в качестве соответствующих индексов.

Таблица .3 **Вычисление квартальных индексов сезонности**

Квартал 1	Квартал 2	Квартал 3	Квартал 4
–	–	1.26	1.42
0.85	0.62	1.18	1.28
0.87	0.72	1.13	1.25
0.89	0.78	–	–
Сумма 2.61	2.12	3.57	3.95
Средняя 0.870	0.707	1.190	1.317
= (2.16/3)	= (2.12/3)	= (3.57/3)	= (3.95/3)

Данная процедура позволяет значительно сократить эффект воздействия нерегулярной компоненты и получить практически в чистом виде квартальные индексы сезонности.

Они имеют следующую интерпретацию. Каждый индекс представляет собой отношение среднего значения уровня по данному кварталу к общему среднеквартальному уровню временного ряда. Если значение индекса меньше 1, то средний объем продаж в данном квартале меньше $^1/_4$ среднегодового объема продаж за все периоды временного ряда. Если индекс больше 1, то средний объем продаж в данном квартале превышает $^1/_4$ среднегодового объема продаж. В случае когда индекс равен 1, средний объем продаж по кварталу в точности равен 1/4 среднегодового объема продаж. Очевидно, средняя всех квартальных индексов есть 1, иначе сумма средних квартальных объемов продаж не будет равна среднегодовым продажам. Следовательно, сумма всех полученных квартальных индексов равна 4. Однако, как правило, возможны погрешности, связанные с округлениями результатов вычислений. Поэтому следует проверить точность расчетов и скорректировать полученные значения. Для этого определяется корректирующий множитель, который равен отношению 4 к сумме вычисленных индексов. На него

умножается значение каждого из четырех квартальных индексов. Сумма скорректированных индексов должна быть равна 4.

Определим сумму квартальных индексов, вычисленных в табл. 3:

0.87+0.707+1.19+1.317 = 4.084.

Найдем корректирующий множитель:

4/4.084 = 0.9794.

Скорректируем квартальные индексы:

Квартал	**Скорректированный индекс сезонности**
1	$0,870 \cdot 0,9794 = 0,852$
2	$0,707 \cdot 0,9794 = 0,692$
3	$1,190 \cdot 0,9794 = 1,166$
4	$1,317 \cdot 0,9794 = 1,290$

4

Индексы сезонности часто измеряют в процентах. Например, индекс первого квартала 85,2%. Это означает, что средний объем продаж по первому кварталу на 14,8% меньше 1/4 среднегодового объема продаж. Индекс третьего квартала, равный 116,6%, означает, что средний объем продаж по третьему кварталу на 16,6% больше '/4 среднегодового объема продаж. ■

Вычислим десезонализированные данные.

Год	***t***	**y_t**	**Индекс сезонности (S_t)**	**Десезонализированные данные ($d_t=y_t/S_t$)**
1990	1	20	0.852	23.47
	2	12	0.692	17.34
	3	47	1.166	40.31
	4	60	1.290	46.51
1991	5	40	0.852	46.95
	6	32	0.692	46.24
	7	65	1.166	55.75
	8	76	1.290	58.91
1992	9	56	0.852	65.73
	10	50	0.692	72.25
	11	85	1.166	72.90
	12	100	1.290	77.52
1993	13	75	0.852	88.03
	14	70	0.692	101.16
	15	101	1.166	86.62
	16	123	1.290	95.35

Приведем данные пакета Excel для данного примера.

Дисперс. статистика

Множеств 0,97262 R-квадрат 0,94599 Нормиров 0,942132

Стандарт 5,93174 Наблюден 16

Дисперсионный анализ

	df	SS	*MS*	*F*	*Значим F*
Регрессия	1	8627,877	8627,877	245,2109	2.88E-10
Остаток	14	492,5976	35,18554		
Итого	15	9120,475			

	Коэфф	*Станд ст*	*P-Значен.*	*Ниж 95%*	*Верх 95%;*	*Ниж 95%*	*Вер 95%*
Y-пересеч	**19,3715**	3,110631	6,227515	2.21E-05	12,69985	26,04315	12,69985	26,04315

Перемен **5,037471** 0,321694 15,65921 2.88E-10 4,347505 5,727436 4,347505 5,727436

Прогноз.

$x(17) = (19{,}372 + 5{,}0375 \cdot 17) \cdot 0{,}852 = 89{,}5$ $\quad x(19) = (19{,}372 + 5{,}0375 \cdot 19) \cdot 1{,}166 = 134{,}188$

$x(18) = (19{,}372 + 5{,}0375 \cdot 18) \cdot 0{,}692 = 76{,}2$ $\quad x(20) = (19{,}372 + 5{,}0375 \cdot 20) \cdot 1{,}29 = 154{,}96$

Мы можем осуществлять прогноз также вводя фиктивные переменные. Для этого введем переменные, принимающие значение 1 для соответствующего квартала и ноль – в остальных случаях и оценим коэффициенты при этих переменных. Ввиду того что сумма этих фиктивных переменных равна 1, то мы включим только 3 переменные. Именно, пусть

$$\boldsymbol{B}^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{X}^T = (20,12,47,60,40,32,65,76,56,50,85,100,75,70,101,123).$$

Тогда

$$\theta = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{X} = \begin{pmatrix} 41.75 \\ 4.8 \\ -27.6 \\ -39.15 \\ -10.45 \end{pmatrix}.$$

Однако построение линейной функции по данным (x_1, x_5, x_9, x_{13}), $(x_2, x_6, x_{10}, x_{14})$, $(x_3, x_7, x_{11}, x_{15})$, $(x_4, x_8, x_{12}, x_{16})$ дает коэффициенты $16.075 + 4.525t$, $2.6 + 4.8t$, $33.55 + 4.55t$, $36.5 + 5.325t$, т.е. эти оценки имеют разные коэффициенты (углы наклона), поэтому введем еще фиктивные переменные для наклона, т.е. положим

$$\boldsymbol{B}^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 9 & 0 & 0 & 0 & 13 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & 10 & 0 & 0 & 0 & 14 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 7 & 0 & 0 & 0 & 11 & 0 & 0 & 0 & 15 & 0 \end{pmatrix},$$

Тогда

$$\theta = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{X} = \begin{pmatrix} 36.5 \\ 5.325 \\ -20.425 \\ -33.9 \\ -2.95 \\ -0.8 \\ -0.525 \\ -0.775 \end{pmatrix}$$

Таким образом, прогноз можно осуществлять по формулам:

1) Для первого квартала

$$x_t = 36.5 - 20.425 + (5.325 - 0.8)t = 16.075 + 4.525\,t, \qquad x_{17} = 93.$$

1) Для второго квартала $x_t = 36.5 - 33.9 + (5.325 - 0.525)\,t = 2.6 + 4.8\,t$, $x_{18} = 89$.

3) Для третьего квартала $x_t = 36.5 - 2.95 + (5.325 - 0.775)t = 33.55 + 4.55\,t$, $x_{19} = 120$.

4) Для четвертого квартала (не вычитаем ничего) $x_t = 36.5 + 5.325t$, $x_{20} = 143$.

7. Практический пример.

Ниже представлены поквартальные объемы продаж автомобилей Ford Motor Company (по материалам этой компании). [4], с. 762, Таблица 14.2.4.

Таблица 1. Ford Motor Company

Год	Квартал	Объем продаж автомобилей, млн. долл.	Год	Квартал	Объем продаж автомобилей, млн. долл.
1991	1	17 115	1994	1	26 070
	2	19 833		2	28 375
	3	17 205		3	24 926
	4	17 898		4	27 766
1992	1	20 636	1995	1	28601
	2	22 903		2	29861
	3	19 370		3	24437
	4	21 498		4	27597
1993	1	22 264	1996	1	28297
	2	25 264		2	31762
	3	20 107		3	26459
	4	23 511		4	31505
			1997	1	30037
				2	32805
				3	28196
				4	31897

Проведем сначала сезонную декомпозицию для 1991-1194 и сделаем прогноз на следующий 1995 год.

Результаты вычислений проведены с помощью Maple-программы.

Таким образом, имеем следующий прогноз (1995): Реальные данные

$x(17) = 29012$	28601
$x(18) = 32008$	29861
$x(19) = 26800$	24437
$x(20) = 29298$	27597

8. Программа

```
X:=[17115.,19833.,17205.,17898.,20636.,22903.,19370.,21498.,22264.,2
5264.,20107.,23511.,26070.,28375.,24926.,27766.];
```

$$X := [17115., 19833., 17205., 17898., 20636., 22903., 19370., 21498., 22264., 25264., 20107., 23511., 26070., 28375., 24926., 27766.]$$

```
> x1:=stats[describe,mean](X);n:=stats[describe,count]
(X);n3:=n-3;n4:=n-4;
```

$$x1 := 22171.31250$$

$$n := 16$$

$$n3 := 13$$

$$n4 := 12$$

```
> A:=0;b:=array(1..n3);
```

$$A := 0$$

$$b := \text{array}(1 .. 13, [\ ])$$

```
> for j to n3 do b[j]:=0 od:
> print(b);
```

$$[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$$

```
> for i to n3 do for j to 4 do b[i]:=b[i]+X[j+i-1] od od;
> print(b);
```

$$[72051., 75572., 78642., 80807., 84407., 86035., 88396., 89133., 91146., 94952., 98063., 102882., 107137.]$$

```
> c:=array(1..12);
```

$$c := \text{array}(1 .. 12, [\ ])$$

```
> for j to n4 do c[j]:=(b[j]+b[j+1])/8 od;
```

$$c_1 := 18452.87500$$

$$c_2 := 19276.75000$$

$$c_3 := 19931.12500$$

$$c_4 := 20651.75000$$

$$c_5 := 21305.25000$$

$$c_6 := 21803.87500$$

$$c_7 := 22191.12500$$

$$c_8 := 22534.87500$$

$$c_9 := 23262.25000$$

$$c_{10} := 24126.87500$$

$$c_{11} := 25118.12500$$

$$c_{12} := 26252.37500$$

```
> c2:=array(1..n4);
```

$$c2 := \text{array}(1 .. 12, [\ ])$$

```
> for j to n4 do c2[j]:=X[j+2]/c[j] od;
```

$$c2_1 := 0.9323750364$$

$$c2_2 := 0.9284760138$$

$$c2_3 := 1.035365540$$

$$c2_4 := 1.109010132$$

$$c2_5 := 0.9091655813$$

$$c2_6 := 0.9859715303$$

$$c2_7 := 1.003283971$$

$$c2_8 := 1.121106729$$

$$c2_9 := 0.8643617879$$

$$c2_{10} := 0.9744734865$$

$$c2_{11} := 1.037895942$$

$$c2_{12} := 1.080854589$$

```
> c3:=array(1..4);c4:=array(1..4);
```

$$c3 := \text{array}(1 .. 4, [\ ])$$

$$c4 := \text{array}(1 .. 4, [\ ])$$

```
> for j to 4 do c3[j]:=c2[j]+c2[j+4]+c2[j+8] od;
```

$$c3_1 := 2.705902406$$

$$c3_2 := 2.888921030$$

$$c3_3 := 3.076545453$$

$$c3_4 := 3.310971450$$

```
> for j to 4 do c4[j]:=c3[j]/3 od;
```

$$c4_1 := 0.9019674687$$

$$c4_2 := 0.9629736767$$

$$c4_3 := 1.025515151$$

$$c4_4 := 1.103657150$$

```
> c5:=c4[1]+c4[2]+c4[3]+c4[4];
```

$$c5 := 3.994113446$$

```
> c6:=4/c5;
```

$$c6 := 1.001473808$$

```
> c7:=array(1..4);
```

$$c7 := \text{array}(1 .. 4, [\ ])$$

```
> for j to 4 do c7[j]:=c4[j]*c6 od;d1:=array(1..n);
```

$$c7_1 := 0.9032967956$$

$$c7_2 := 0.9643929150$$

$$c7_3 := 1.027026563$$

$$c7_4 := 1.105283729$$

$$d1 := \text{array}(1 .. 16, [\ ])$$

```
> for j to 4 do
d1[4*(j-1)+3]:=c7[1];d1[4*(j-1)+4]:=c7[2];d1[4*(j-1)+1]:=c7[3];d1[4*
(j-1)+2]:=c7[4] od:print(d1);
```

[1.027026563, 1.105283729, 0.9032967956, 0.9643929150, 1.027026563, 1.105283729, 0.9032967956, 0.9643929150, 1.027026563, 1.105283729, 0.9032967956, 0.9643929150, 1.027026563, 1.105283729, 0.9032967956, 0.9643929150]

```
> d:=array(1..n);for j to n do d[j]:=X[j]/d1[j] od; cc1:=0; for j to
n do cc1:=cc1+d[j]/n od:cc1;
```

$$d := \text{array}(1 .. 16, [\ ])$$

$$d_1 := 16664.61279$$

$$d_2 := 17943.80889$$

$$d_3 := 19046.89586$$

$$d_4 := 18558.82568$$

$$d_5 := 20092.95645$$

$$d_6 := 20721.37624$$

$$d_7 := 21443.67177$$

$$d_8 := 22291.74402$$

$$d_9 := 21678.11506$$

$$d_{10} := 22857.47934$$

$$d_{11} := 22259.57193$$

$$d_{12} := 24379.06753$$

$$d_{13} := 25383.95884$$

$$d_{14} := 25672.14124$$

$$d_{15} := 27594.47407$$

$d_{16} := 28791.16962$

$cc1 := 0$

22211.24184

```
> cc2:=0; for j to n do cc2:=cc2+(d[j])^2/n od:cc2;
```

$cc2 := 0$

$0.5045488907\ 10^{9}$

```
> cc3:=cc2-(cc1)^2;
```

$cc3 := 0.112096266\ 10^{8}$

```
> cc4:=0; for j to n do cc4:=cc4+(d[j]*j)/n od:cc4;
```

$cc4 := 0$

203888.7570

```
> m1:=(n+1)/2.;cc5:=cc4-cc1*m1;cc6:=(n+1)*(2*n+1)/6.-m1^2;
```

$m1 := 8.500000000$

$cc5 := 15093.2014$

$cc6 := 21.25000002$

```
> cc7:=cc5/cc6;
```

$cc7 := 710.2683005$

```
> cc8:=cc1-cc7*m1;
```

$cc8 := 16173.96129$

Прогноз `x17:=(cc7*(n+1)+cc8)*c7[3];`

$x17 := 29011.98287$

```
> x18:=(cc7*(n+2)+cc8)*c7[4];
```

$x18 := 32007.68017$

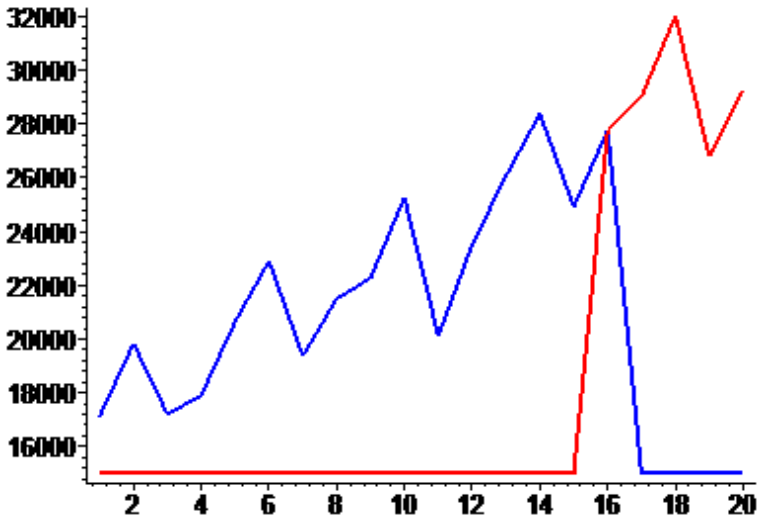
```
> x19:=(cc7*(n+3)+cc8)*c7[1];
```

$x19 := 26799.96592$

```
> x20:=(cc7*(n+4)+cc8)*c7[2];
```

$x20 := 29297.60801$

```
> plot({[[1,X[1]],[2,X[2]],[3,X[3]],[4,X[4]],[5,X[5]],[6,X[6]],
[7,X[7]],[8,X[8]],[9,X[9]],[10,X[10]],[11,X[11]],[12,X[12]],
[13,X[13]],[14,X[14]],[15,X[15]],[16,X[16]],[17,15000],[18,15000],
[19,15000],[20,15000]],[[1,15000],[2,15000],[3,15000],[4,15000],
[5,15000],[6,15000],[7,15000],[8,15000],[9,15000],[10,15000],
[11,15000],[12,15000],[13,15000],[14,15000],[15,15000],[16,X[16]],
[17,x17],[18,x18],[19,x19],[20,x20]]},color=[red,blue],thickness=2);
```



Список использованной литературы

1. Кендалл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976. - 736 с.

2. Андерсон Т. Статистический анализ временных рядов. М.: Мир, 1976 - 756 с.
(Anderson T. Statistical Analysis of Time Series. Hoboken: Wiley, 2011. - 358 p.)

3. Айвазян С.А. Прикладная статистика: основы эконометрики, т.2. М: ЮНИТИ-ДАНА, 2001. - 432 с.

4. Сигел Э.Ф. Практическая бизнес-статистика. М.: Издательский дом «Вильямс»,2002. - 1056с.
(Siegel A. Practical Business Statistics. NY: Academic Press, 2016. - 640 p. (7th ed.)