ГОСУДАРСТВЕННЫЙ КОМИТЕТ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПО ВЫСШЕМУ ОБРАЗОВАНИЮ

МАТИ им. К.Э.Циолковского - ГОСУДАРСТВЕННЫЙ АВИАЦИОННЫЙ
ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

Кафедра высшей математики

АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ

**Методические указания
к чтению лекций
и проведению практических занятий**

ГОРБАЦЕВИЧ В.В.

ЧАСТЬ II

МОСКВА 2000 год

§4. Спектральная теория случайных процессов

Для изучения функций в математическом анализе применяют разложение в тригонометрические ряды – ряды Фурье. При этом обычно используют только гармоники с частотами ω_k, кратными одной, основной частоте ω, т.е. берут $\omega_k=\omega\cdot k$, где $k=1,2,3...$ Более общее разложение в сумму гармоник с произвольными частотами используется реже, так как при этом получаются не привычные периодические функции, а так называемые почти периодические функции, которые еще не нашли широкого применения в инженерной практике. Разложение функции в ряд Фурье называется гармоническим анализом, а исследование амплитуд и фаз получившихся гармоник – спектральным анализом. Спектром функции называется совокупность частот ее разложения по гармоникам. Разложение в ряд Фурье используется для функций, заданных на конечном промежутке или для периодических функций, здесь получается дискретный спектр. Для более общих функций используется континуальный аналог ряда Фурье – интеграл Фурье. Здесь спектр уже не является дискретным, он "сплошной" и задается не дискретным набором частот, а спектральной функцией.

При исследовании случайных функций тоже часто удается воспользоваться методами спектрального анализа. Эта возможность основана на теореме, называемой теоремой о спектральном разложении.

Говорят, что случайный процесс $X(t)$ допускает спектральное разложение, если его можно представить в виде суммы (вообще говоря, бесконечной, т.е. в виде бесконечного функционального ряда) элементарных гармонических случайных процессов $C_k\sin(\omega_k t+\varphi_k)$, такое представление иногда называют полигармонической моделью временного ряда.

ТЕОРЕМА (О СПЕКТРАЛЬНОМ РАЗЛОЖЕНИИ). Предположим, что функция ковариации стационарного случайного процесса $X(t)$ может быть представлена в виде

$K_X(\tau)=\sum\limits_n C_n\cos\omega_n\tau$ (при этом можно доказать, что все коэффициенты C_n будут неотрицательны).

Тогда $X(t)$ допускает спектральное разложение вида

$$X(t)= MX+\sum_n A_n\cos\omega_n\tau +B_n\sin\omega_n\tau,$$

де A_n, B_n – некоторые случайные величины, независимые между собой, имеющие нулевое математическое ожидание и равные между собой (при каждом отдельном n) дисперсии, а MX – константа, равная математическому ожиданию процесса X.

Набор $\{\omega_n\}$ неслучайных частот, фигурирующих в этом разложении, называется спектром случайного процесса $X(t)$.

В некоторых случаях требуемое в этой теореме разложение функции K_X можно найти с помощью ряда Фурье. А именно, если K_X периодична, то она может быть разложена (при некоторых, довольно несущественных на практике, ограничениях на K_X) на отрезке $[-T,T]$

в ряд Фурье, причем только по косинусам (так как K_X –четная функция):

$$K_X(\tau)=\sum_{n=0}^{\infty} a_n\cos(\pi n/l).$$

Тут спектр имеет вид $\{\omega_n=\frac{pn}{l}\}$. Коэффициенты a_n вычисляются по обычным формулам Фурье:

$$a_n=\frac{1}{T}\int_{-T}^{T} K_X(t)\cos\frac{pn}{l}t\,dt\ .$$

На самом же деле периодической функция K_X бывает довольно редко (и практически никогда для стационарных процессов). Поэтому предположение о ее периодичности довольно ограничительно и обычно может рассматриваться как довольно грубое приближение к реальности. Но если имеются основания считать K_X периодической, то тут применима спектральная теорема и мы получаем разложение $X(t)$ в сумму элементарных гармоник, частоты которых имеют вид $\frac{pn}{l}$. При $l\to\infty$ спектр становится все более похож на непрерывный.

Для непериодических K_X используется не ряд Фурье, а преобразование Фурье. Преобразование Фурье для K_X (если оно существует, т.е. если сходится соответствующий несобственный интеграл), называется спектральной плотностью стационарного случайного процесса $X(t)$ и вычисляется по формуле:

$$S_X(\omega)=\frac{2}{p}\int_{0}^{\infty} K_X(\tau)\cos\omega\tau\,d\tau.$$

Спектральная функция всегда неотрицательна, это выводится из свойства положительной определенности функции автокорреляции. Ее график является важной характеристикой стационарного случайного процесса.

Сказанное здесь о случайных процессах относится и к временным рядам. Для них тоже используется спектральное разложение, которое широко применяется при практическом исследовании временных рядов.

ГЛАВА 3. Практические методы исследования временных рядов

§1. Тренд и его анализ.

Тренд или тенденция временного ряда – это несколько условное понятие. Под трендом понимают закономерную, неслучайную составляющую временного ряда (обычно монотонную), которая может быть вычислена по вполне определенному однозначному правилу. Тренд временного ряда часто связан с действием физических законов или каких-либо других объективных закономерностей. Однако, вообще говоря, нельзя однозначно разделить случайный процесс или временной ряд на регулярную часть (тренд) и колебательную часть (остаток). Поэтому обычно предполагают, что тренд - это некоторая функция простого вида (линейная, квадратичная и т.п.), описывающая "поведение в целом" ряда или процесса. Если выделение такого тренда упрощает исследование, то предположение о выбранной форме тренда считается допустимым.

Для временного ряда уравнение линейного тренда имеет вид

$$x-MX=r\frac{s_X}{s_T}(t-MT).$$

При $r>0$ говорят о положительном тренде (с течением времени значения временного ряда имеет тенденцию возрастать), при $r<0$ об отрицательном (тенденция убывания). При r, близких к нулю, иногда говорят о боковом тренде. Как было сказано выше, для случая, когда $t=1,2,3,...n$, имеем:

$MT=(n+1)/2$, $DT=n(n+2)/12$, а потому $\sigma_T=\sqrt{n(n+1)/12}$,

однако на практике не стоит отдельно вычислять r и σ_X и только потом подставлять их в уравнение тренда. Лучше прямо в формуле тренда произвести сокращения, после которых она примет вид:

$$x-MX=\frac{K_X}{DT}(t-(n+1)/2).$$

После выделения линейного тренда нужно выяснить, насколько он значим. Это делается с помощью анализа коэффициент корреляции. Дело в том, что отличие коэффициента корреляции от нуля и тем самым наличие реального тренда (положительного или отрицательного) может оказаться случайным, связанным со спецификой рассматриваемого отрезка временного ряда. Другими словами, при анализе другого набора экспериментальных данных (для того же временного ряда) может оказаться, что полученная при этом оценка намного ближе к нулю, чем исходная (и, возможно, даже имеет другой знак), и говорить о реальном тренде тут уже становится трудно.

Для проверки значимости тренда в математической статистике разработаны специальные методики. Одна из них основана на проверке равенства $r=0$ с помощью распределения Стьюдента.

Предположим, что имеется набор экспериментальных данных - значения $x_1, x_2, x_3,...x_N$ временного ряда в равноотстоящие моменты времени $t_1, t_2, t_3...t_N$.

Оценка для коэффициента автокорреляции, как уже было сказано, для временного ряда имеет вид:

$$r^*=K^*_X(\tau)/K^*_X(0),$$

где K^*_X - оценка для функции автоковариации, которую можно найти по следующей формуле:

$$K^*_X=\frac{1}{N+1}\sum_{k=0}^{N} x(k(T-\tau)/N+\tau)x(k(T-\tau)/N).$$

Назовем это значение r^* экспериментальным. Идея метода статистической проверки гипотез такова. Выдвигается некоторая гипотеза, в нашем случае это гипотеза о равенстве нулю коэффициента автокорреляции. Далее, задается некоторый уровень вероятности α. Смысл этой величины заключается в том, что она является мерой допустимой ошибки. А именно, мы допускаем, что сделанный нами вывод о справедливости или несправедливости (при более тонком исследовании эти две ситуации нужно различать) гипотезы на основании заданного массива экспериментальных данных может оказаться ошибочным, ибо абсолютно точного вывода на основании лишь частичной информации ждать, конечно, не стоит. Однако вероятность этой ошибки не должна превосходить выбранной величины α. Обычно берут это значением равным 0.05 (т.е.5%) или 0.10, иногда берут и 0.01.

Далее рассматриваем следующую величину:

$$U_{экс}=r^*\frac{\sqrt{N-2}}{\sqrt{1-r^2}}.$$

Можно доказать, что при достаточно больших значениях n эта случайная величина имеет распределение, мало зависящее от исходного временного ряда. Более того, это распределение близко к одному из стандартных распределений, используемых в математической статистике – к распределению Стьюдента (с числом степеней свободы k=N-2, где N – число экспериментальных данных).

Для распределения Стьюдента имеются подробные таблицы, в которых для заданного уровня вероятности α и числа степеней свободы k указывается критическое значение $U_{кр}$. Критическим или граничным оно называется потому, что ограничивает двустороннюю область, вне которой значения случайной величины могут оказаться только с вероятностью не большей, чем α (т.е. небольшой). Точнее, при условии r=0 имеет место равенство:

$$P[|U| \geq U_{кр}]=\alpha.$$

В настоящее время значение $U_{кр}$ можно находить не только из таблиц (где оно приведено только лишь для некоторых значений уровня вероятности). Любая современная статистическая программа для компьютера дает возможность вычислить $U_{кр}$ для произвольного заданного уровня вероятности. Как нетрудно понять, с ростом величины α значения $U_{кр}$ тоже растут.

Далее рассуждают следующим образом. Предположим, что число N достаточно велико. Тогда случайная величина $U_{экс}$ распределена приблизительно по закону Стьюдента. Если r=0, то с большой (т.е. близкой к 1) вероятностью, равной 1-α, значение $U_{экс}$ должно по модулю не превосходить $U_{кр}$. А вот выходить за пределы отрезка $[-U_{кр},U_{кр}]$ величина $U_{экс}$ может только с вероятностью α. Поэтому если $|U_{экс}| \geq U_{кр}$, то делают заключение о том, что гипотеза r=0 экспериментальными данными не подтверждается и потому тренд является выраженным. Ошибка такого заключения не превосходит заданного уровня α. Если же $|U_{экс}| \leq U_{кр}$, то говорят, что на заданном уровне вероятности α отвергнуть гипотезу r=0 нет оснований. В этом случае мы не имеем оснований говорить о выраженном тренде, а тем более использовать рост или убывание этого тренда при прогнозировании динамики временного ряда на будущее.

Например, пусть r*=0.20 и N=20. Тогда вычисление дает $U_{экс}$=0.87. Для уровня вероятности 5% находим из таблицы распределения Стьюдента $U_{кр}$=2.10. Сравнивая $U_{экс}$ и $U_{кр}$, видим, что тут гипотезу о равенстве нулю коэффициента автокорреляции отвергать нет основания. Тренд здесь не является выраженным.

Кроме линейного тренда, приходится рассматривать и тренды более сложной структуры. При этом заранее указать функцию, с помощью которой можно описать этот тренд, обычно не представляется возможным. Поэтому часто на практике просто перебирают несколько простых функциональных зависимостей (с параметрами) и для каждой из них оценивают, насколько успешно функцией того или иного вида можно описать тенденцию рассматриваемого временного ряда. При наличии компьютера эти вычисления не занимают много времени, а иногда могут проводиться даже в автоматическом режиме, выделяющем среди нескольких заданных видов трендов оптимальный. Однако далеко не всегда среди рассмотренных функций имеется та, которая достаточно эффективно описывает тенденцию развития заданного временного ряда. В этом случай приходится идти другими путями. В частности, часто в такой ситуации производят различные преобразования членов временного ряда (логарифмирование, дифференцирование - образование разностей соседних членов ряда, интегрирование – суммирование последовательных членов ряда и др.) для того, чтобы попытаться получить временной ряд с ясно выраженным линейным трендом. Если это удается сделать, то к полученному ряду применяют те методы, которые были описаны выше, а потом обратным преобразованием возвращаются к исходному ряду. Нужно еще раз подчеркнуть, что вид тренда не определяется однозначно самим рядом и является некоторым условным объектом, привлекаемым для более полного понимания особенностей рассматриваемого процесса.

§2. Кореллограмма и ее использование

Точное вычисление функции автокорреляции для заданного случайного процесса или временного ряда удается произвести очень редко. На практике единственно возможным остается нахождение оценок для K_X. Одна из оценок такого рода для центрированного временного ряда имеет, как уже отмечалось, следующий вид:

$$K^*_X(\tau)=\frac{1}{N+1}\sum_{k=0}^{N} x(k(T-\tau)/N+\tau)x(k(T-\tau)/N).$$

Однако вычисление и по такой приближенной формуле обычно бывает очень трудоемким. Только в последние два-три десятилетия вычисления оценок для функции автокорреляции перестали быть существенной проблемой. На современном компьютере такое вычисление занимает лишь доли секунды. Поэтому сейчас особый интерес начинают представлять уже не методы вычисления оценок, а методы адекватной интерпретации этих оценок.

Если временой ряд нестационарен, например, если он имеет явно выраженный тренд, чтобы более эффективно применять кореллограмму, предварительно нужно исключить нестационарную компоненту, т.е. тренд. Для исключения тренда обычно переходят к временному ряду, составленному из разностей последовательных значений ряда:

$x_n^*=x_n-x_{n-1}.$

Такого рода преобразование называют иногда дифференцированием. Название связано не только с тем, что слово "differentia" означает "разность". Тут есть и прямая связь с операцией дифференцирования, когда разность делится на приращение аргумента, которое для временного ряда равно 1 и потому в явном виде не выписывается. Если бы временной ряд описывался точно линейной функцией

$$X_n=an+b,$$

то продифференцированный указанным способом ряд описывался бы функцией, получаемой из исходной операцией дифференцирования (из дифференциального исчисления). Производная линейной функции постоянна, поэтому в нашем примере получился бы стационарный ряд без тренда. Так и в общем случае, переход к временному ряду из разностей - это попытка уничтожить линейный тренд или просто сделать ряд более похожим на стационарный. Но тренд далеко не всегда линеен и потому указанная операция не всегда приводит к цели. Если ряд стационарен, то его функция автокорреляции обычно довольно быстро стремится к нулю с ростом лага (хотя это свойство не является необходимым для стационарности). Поэтому об эффективности операции дифференцирования можно судить по поведению функции автокорреляции для преобразованного ряда. Если однократного применения операции дифференцирования оказалось недостаточно для получения ряда, похожего на стационарный, то эту операцию применяют еще раз, а потом еще раз... Если тренд полиномиальный, то в результате мы в конце концов сможем исключить влияние тренда, это связано с тем, что для полинома производные достаточно высокого порядка равны нулю. Но если тренд далек от полиномиального (например, логарифмический, экспоненциальный и др.), то таким методом мы ничуть не улучшим ситуацию. В дальнейшем мы будем считать, что нам тем или иным способом удалось от исходного ряда перейти к стационарному.

Кореллограммой называется график оценки для функции K_X или для функции $r_X(\tau)=K_X(\tau)/K_X(0)$. Большие значения оценки соответствуют наличию заметной корреляции между сечениями. Особо нагляден график для оценки функции автокорреляции, так как он изменяется в фиксированных пределах (от –1 до +1) и есть возможность наглядно соизмерять с 1 отдельные его значения. Если случайный процесс или временной ряд близок к белому шуму, то кореллограмма близка к горизонтальной оси, а ее значения близки к 0. Следует отметить, что при больших значениях τ приведенная выше оценка для коэффициента автокорреляции малонадежна, это связано с заменой полного суммирования частичным. Поэтому поведению кореллограммы при больших значениях τ не следует придавать большого значения.

Для заданного уровня вероятности α (по умолчанию при компьютерных вычислениях берется обычно α=5%) можно вычислить границы доверительного интервала, в котором нахождение значения функции автокорреляции при заданном лаге τ с вероятностью $1-\alpha$ не противоречит предположению об отсутствии корреляции сечений с этим лагом. При графическом изображении функции автокорреляции или ее оценки – кореллограммы – эти интервалы дают две граничные кривые (выше и ниже основного графика). Выход за эти граничные кривые рассматривается как указание на значимость корреляции с соответствующим лагом.

§3. Сглаживание временных рядов

Сглаживание временного ряда используется для удаления из него высокочастотных компонент (которые обычно являются несущественными, так как вызваны случайными факторами). Один из простейших методов сглаживания - метод скользящих или подвижных средних (MA в англоязычной нотации), он является одним из наиболее старых и широко известных. Этот метод основан на переходе от начальных значений временного ряда к их средним значениям на некотором заданном интервале времени (длина которого называется шириной окна). Этот интервал времени как бы скользит вдоль ряда, с чем и связано название метода. В каждый момент этого скольжения мы видим только часть ряда, чем и вызвана "оконная" терминология.

Полученный в результате такого сглаживания новый временной ряд обычно ведет себя более регулярно (гладко), что связано с удалением в процессе сглаживания резких случайных отклонений, попадающих в окно. Сглаживание полезно применять даже в самом начале исследования временного ряда, так как при этом часто удается прояснить вопрос о наличии и характере тренда, а также выявить сезонные колебания.

Несколько слов нужно сказать о сезонных колебаниях. Они проявляются во многих временных рядах, в частности, в экономике, метеорологии. Сезонными колебаниями называют все такие изменени, которые соответствуют определенному (почти) строго периодическому ритму (не обязательно равному одному году, как для обычных сезонов), присущему Вселенной, природе или человеческой деятельности. Такая периодичность может ярко проявляться в процессах человеческой деятельности, например, в изменениях объема перевозок местным транспортом в последние дни каждой недели или же утром и вечером в течение каждого дня, в росте ошибок при выполнении производственных операций по понедельникам и др. Но наиболее типичные сезонные колебания связаны именно со сменой сезонов года. Они затрагивают огромное число параметров жизни человека (как современного, так и в древности). Обычно при исследовании временных рядов стремятся выделить сезонные колебания для того, чтобы их изолировать и изучить другие, более сложные периодические компоненты.

Простейшее сглаживание методом MA с шириной окна $2m+1$ производится по следующим формулам:

$$x^*_k=(x_{k-m}+x_{k-m+1}+...+x_k+x_{k+1}+...+x_{k+m})/2m+1.$$

Выбор ширины окна диктуется содержательными соображениями, связанными с предполагаемым периодом сезонных колебаний или с желательным исключением определенного рода высокочастотных колебаний. На практике обычно при отсутствии сезонности ширину окна берут равной 3, 5 или 7. Не рекомендуется брать окно шире, чем в четверть числа анализируемых данных. Чем шире окно, тем больше колебательных компонент будет исключено и тем более гладкий вид полученного при сглаживании ряда. Однако при слишком больших окнах полученный ряд уже значительно отличается от исходного, теряются многие индивидуальные особенности и ряд все более приближается к постоянному. Если взять ширину окна максимально возможной (равной общему числу данных значений $x_1, x_2, \ldots$), то приходим просто к постоянной величине, равной среднему значению всех этих x_i.

Подвижные средние могут, к сожалению, искажать кратковременные колебания и порождать фиктивные гармонические компоненты при гармоническом анализе временных рядов.

Имеются различные модификации метода МА. В некоторых из них используются более сложные методы усреднения (с некоторыми весами и др.), которые подчеркивают большую или меньшую значимость отдельных слагаемых. Например, часто используемое экспоненциальное сглаживание основано на приписывании больших весов непосредственно предшествующим значениям. Этот подход очень широко распространен в социологии, экономике и других дисциплинах.

В настоящее время метод МА (с различными модификациями) реализован во всех статистических пакетах программ, а также в многих специализированных программах, предназначенных для обработки экономической и деловой информации.

Для случайных процессов тоже имеются разнообразные методы сглаживания. Здесь число методов чрезвычайно велико, это связано с тем, что усреднение может производиться с помощью интегрирования с некоторой весовой функцией, которую можно выбирать достаточно произвольно. Поэтому окно здесь задается не только своей шириной, а и видом усредняющей функции. Правильный выбор окна представляет собой весьма непростую задачу, этому посвящена обширная литература. Прямоугольное окно (используемое в классическом варианте метода МА) имеет целый ряд недостатков, которые в классической теории рядов Фурье связывают с явлением Гиббса, в технике именуемом вытеканием мощности. При исследовании случайных процессов часто говорят не о сглаживании, а о фильтрации (или о коррекции, очистке спектра), причем в области высоких частот говорят о применении фильтра высоких частот (ФВЧ), а в области низких частот – о фильтре низких частот (ФНЧ). Такого рода терминология принята, в частности, в теории распознавания сигналов и, вообще, в теории связи.

Другой (терминологически, но не по существу) подход к сглаживанию временных рядов и случайных процессов основан на модификации спектра. Если в спектре ряда просто полностью удалить высокочастотные компоненты, то получится новый ряд, который ведет себя более регулярно. Такого рода вычисление возможны только при наличии компьютера и специальной программы для работы с рядами и преобразованиями Фурье. Эти программы входят в состав большинства универсальных математических пакетов (Mathcad, Matlab, Maple, Mathematica) и многих статистических пакетов.

§4. Периодграмма и ее использование

Периодграмма – это графическое изображение оценки (точнее, одной из нескольких возможных оценок) модуля преобразования Фурье автокорреляционной функции стационарного временного ряда или случайного процесса. Другими словами, это – графическая форма одной из оценок спектра функции автокорреляции. Ее еще можно рассматривать как график зависимости мощности процесса (или, что то же, квадрата амплитуды) от частоты. Явные формулы для этих оценок довольно сложны и здесь приводиться не будут. Вычисления периодграммы основаны на БПФ (Быстром Преобразовании Фурье – см. ниже) и проводятся только при наличии компьютера с соответствующей программой. Если временной ряд имеет спектральное разложение вида:

$$X(t)= MX+\sum_n A_n\cos\omega_n\tau +B_n\sin\omega_n\tau,$$

то периодграмма состоит из точек $(\omega_n, a_n^2+b_n^2)$ или $(\omega_n, \sqrt{a_n^2+b_n^2})$. Величина $a_n^2+b_n^2$ - квадрат модуля - характеризует мощность соответствующей гармоники.

При анализе периодграммы нужно обращать особое внимание на ее пики. Большой пик в области некоторой частоты ω_0 указывает на то, что в спектральном разложении автокорреляционной функции присутствует соответствующая гармоническая компонента. Чем выше и резче выделен пик, тем большая часть мощности сосредоточена около частоты ω_0 и тем большую роль играет эта частота в описании соответствующего случайного процесса или временного ряда.

Более подробно использование периодграммы описывается в руководствах по работе с математическими пакетами (см. ниже список литературы). Дело в том, что методика этого исследования существенно зависит от специфики используемого статистического пакета и потому пытаться подробно описать эту методику вне связи с конкретным математическим пакетом нецелесообразно.

§5. Методика выделения скрытых периодичностей

Очень часто есть основания полагать, что временной ряд не является чисто случайным, т.е. в нем имеются некоторые, обычно не очень видные сразу (т.е. скрытые) закономерности. Разработано множество приемов для выделения этих закономерностей. Здесь основными современными инструментами являются кореллограмма и, особенно, периодграмма. По кореллограмме можно судить, как было описано выше, о наличии некоторых закономерностей в динамике случайного процесса или временного ряда. А вот периодграмма позволяется более конкретно указать регулярные компоненты колебательного типа.

Другой подход - прямое исследование временного ряда с помощью быстрого преобразования Фурье (БПФ). А именно, производится приближенное вычисление преобразования Фурье для заданного набора экспериментальных значений реализации временного ряда. При этом используются спектральная плотность (в технике называемая АЧХ – амплитудно-частотной характеристикой) и ФЧХ (фазово-частотная характеристика). АЧХ показывает, на каких частотах исходного временного ряда сосредоточена основная “энергия” и где находятся области резонанса. Фазо-частотная характеристика является более тонким инструментом, она описывает сдвиги фаз отдельных гармоник и не носит явного энергетического характера.

Следует отметить, что не все периоды, которые удается выделить указанными выше методами, являются реально существующими. Часть из них своим существованием обязана случайным колебаниям и погрешностям при вычислениях оценок. Более того, некоторые дополнительные периоды могут появиться и в результате сглаживания временного ряда, так как у сглаженного ряда корреляция близких сечений больше, чем у исходного.

В качестве примера появления ложных периодов часто приводится исследование урожайности пшеницы, проведенное в Великобритании. Там наряду с 11-летним периодом (который, как теперь хорошо известно, соотносится с периодом солнечной активности) было выделено еще несколько периодов. Были предприняты поиски естественных процессов с такими периодами, которые могли бы оказывать влияние на урожайность пшеницы. Однако потом в результате более подробного математического исследования выяснилось, что все эти периоды - ложные, а потому никакие естественные процессы им не соответствуют.

Более подробно методика выявления скрытых периодичностей описывается в руководствах по применению математических пакетов для исследования временых рядов (см. список литературы).

§6. Общая схема исследования временного ряда

Приводимая ниже схема исследования не претендует на роль единственно правильной, она приводится просто как пример естественной последовательности действий при исследовании временного ряда.

1. Общее описание временного ряда.

Здесь производятся вычисления различных числовых характеристик временного ряда, которые потом используются при более подробном его исследовании.

2. Выделение тренда и анализ его значимости.

Здесь производится определение вида тренда (линейный, полиномиальный, экспоненциальный и др.), вычисление параметров этого тренда и оценка значимости этих параметров.

3. Анализ остатков на случайность.

Здесь можно использовать несколько разных методов. В математических пакетах используются различные методы («тесты на регулярность»), основанные, например, на анализе числа значений, больших и меньших медианы (числа, разделяющего всю совокупность значений на две равновероятные части), числа участков понижений и повышений (пакеты STATGRAPHICS, SPSS, Statistica) и др. При этом может оказаться, что одни методы указывают на случайность ряда остатков, а другие – на наличие в нем определенных регулярностей. Это вовсе не является неразрешимым противоречием, так как разные тесты на регулярность понимают эту регулярность в разных смыслах. Поэтому в случае расхождений в результатах разных тестов исследователь должен отдать предпочтение либо выводу, за который имеется большее число показаний, или же отдать предпочтение результату того метода, который он считает более значимым (вообще или в данном конкретном исследовании).

В случае отсутствия компьютера можно применять более простые методы, не требующие значительных вычислений. Опишем здесь один из простейших методов анализа остатков ряда на случайность (или, говоря другими словами, на близость к белому шуму) – метод Юла-Кендалла.

Пусть имеется набор значений $x_1, x_2, \ldots x_n$ временного ряда. Назовем значение x_k максимумом, если выполняются условия $x_{k-1} < x_k > x_{k+1}$. Для граничных значений x_1 и x_n в силу невозможности выписать подобное условие принимается, что они никогда не могут быть максимумами. Другими словами, максимум (точнее, здесь стоило бы говорить о локальном максимуме) - это такое значение ряда, которое больше обоих соседних значений. Аналогично определяется понятие минимума – это такое значение, которое меньше обоих соседних, граничные значения никогда не могут быть и минимумами. Эстремумом называется минимум или максимум. На графике, изображающем точки временного ряда, экстремумы очень хорошо видны, они выделяются как точки поворота этого графика (поворота с убывания на возрастания и наоборот). Может оказаться, что несколько последовательных значений равны между собой, т.е. $x_k = x_{k+1} = \ldots = x_{k+p}$. В этом случае если соседние с ними значения меньше этого общего значения, то вся последовательность $x_k, \ldots x_{k+p}$ рассматривается как один единственный максимум. То же касается и случая минимума. Вычисление числа экстремумов особенно легко производить, если изобразить временной ряд графически.

Обозначим через ξ_n число точек поворота в некоторой случайной последовательности чисел $x_1, x_2, \ldots . x_n$. Величина ξ_n является случайной, ее математическое ожидание и дисперсия равны, как можно доказать

$M\xi_n = 2(n-2)/3$ и $D\xi_n = (16n-29)/90$,

независимо от распределения исходной совокупности значений ряда Приблизительно они равны $2n/3$ и $8n/45$ соответственно. При $n \to \infty$ величина ξ_n распределена приблизительно нормально. Для построения доверительного интервала можно использовать широко известное в математической статистике и е приложениях правило трех сигм (на уровне вероятности 99.73%) или же правило двух сигм (на уровне 95.44%).

Вычислим число ξ^*_n экстремумов в исследуемом временном ряде. Если этот ряд является чисто случайным, то с вероятность 99.73% величина ξ^*_n попадет в доверительный интервал, равный $M\xi \pm 3\sigma_\xi$, приблизительно это отрезок $\left(\frac{2n}{3} - \frac{2}{3}\sqrt{\frac{2n}{5}}, \frac{2n}{3} + \frac{2}{3}\sqrt{\frac{2n}{5}}\right)$. Если же ξ^*_n окажется вне этого интервала, то с вероятностью 99.73% можно утверждать, что изучаемый временной ряд нельзя рассматривать как чисто случайный. Можно использовать и правило двух сигм, тогда заключение делается на уровне 95.44%.

Например, пусть имеется остатков ряд из n=20 значений, среди которых мы насчитали 9 максимумов. При указанном значении n доверительный интервал равен приблизительно (11,19) (на уровне 99.73%). Но значение 9 не попадает в этот интервал, из чего можно сделать вывод о неслучайности ряда остатков.

Неслучайность означает, что ряд имеет некоторые (скрытые) периодичности, поиск которых производится с применением кореллограммы и периодграммы, а также других методов подобного сорта.

4. Выделение случайной компоненты остатка.

Из значений временного ряда вычитаются значения тренда и выделенных периодически компонент.

В результате указанных выше действий временной ряд разлагается на тренд (тенденцию, долговременное движение), циклические компоненты и случайные (несистематические) колебания. В свою очередь, циклические компоненты можно разделить на долговременные (годы, месяцы в зависимости от рассматриваемой задачи) и короткопериодические (недели, дни).

После получения указанного разложения можно заняться построением прогноза. При этом во внимание принимаются только тренд и регулярные компоненты, а случайные колебания отбрасываются. Возможное влияние случайных колебаний учитывается при прогнозе следующим образом. При выдаче прогноза на графике или в таблице указывают границы, в которых предсказанные значения содержатся с вероятностью не меньшей, чем $1-\alpha$, где α – заданный допустимый уровень ошибки (обычно это 5% или 10%). При компьютерной обработке результат прогноза выглядит примерно как отрезок ряда (или кривой), вокруг которого (сверху и снизу) изображаются две кривые, образующие доверительную полосу и соответствующие заданному уровню вероятности прогноза.

Методы прогноза основываются либо на чисто статистических оценках, либо используют спектральное разложение. Интересно отметить, что в разных прикладных областях предпочитают использовать разные методы прогноза. Так, например, в гуманитарных областях (социология, психология, лингвистика и др.) предпочитают чисто статистические методы а в инженерных и естественно-научных областях (исследование сигналов физического и технического происхождения: электроника, акустика, гидро- и аэродинамика, космические излучения, виброиспытания, механика конструкций и материалов) - спектральные. И у тех и у других методов есть свои приверженцы, которые усматривают значительные недостатки в других методах. На самом деле любые из этих методов прогнозирования имеют один существенный недостаток – они пытаются чисто формальными вычислениями восполнить недостаточную информацию о сущности изучаемых процессов. Именно с этим связаны частые неудачи, которые сопровождают все известные ныне формальные методы прогноза.

Надеяться на повышения точности прогноза можно будет только тогда, когда начнут использоваться более тонкие методы, основанные на использовании не просто математических моделей, а глубоких, поначалу хотя бы символических, аналогий между процессами разной природы.

ГЛАВА 4 Исследование временных рядов с помощью математических пакетов

В этой главе будут кратко описаны два основных вычислительных метода иследования временных рядов – метод ARIMA и быстрое преобразование Фурье (БПФ), а также некоторые основные математические пакеты, в которых эти методы реализованы.

§1. Метод ARIMA (АКПСС)

Так называется метод построения одного класса статистических моделей временного ряда. Название метода АКПСС расшифровывается (по русски) как Автокорреляция и Проинтегрированное Скользящее Среднее. Эти модели основаны на комбинации двух весьма эффективных методов построения статистических моделей. Первый метод – построение автокорреляционных моделей, он основан на представлении значения стационарного процесса в виде линейной комбинации предыдущих значений процесса. Второй метод основывается на представлении стационарного процесса в виде линейной комбинации последовательных значений белого шума. Если исходный ряд не стационарен (например, если он имеет выраженный тренд), то применяется метод Проинтегрированного Скользящего Среднего. Для стационарного ряда операция интегрирования (под которой тут понимается просто суммирование членов ряда) излишня и метод называется АКСС.

Само построение модели ARIMA требует предварительно выбора некоторого набора параметров (порядка модели авторегрессии - числа членов авторегрессии и порядка метода скользящих средних). Методика такого выбора неформальна, тут требуется использование специальных приемов, которые описаны в литературе. После выбора этих параметров необходимо произвести очень значительный объем вычислений. Без применения современного компьютера такие вычисления выполнить невозможно.

После построения модели на ее основе производится прогнозирование временного ряда. Нужно отметить, что точность такого прогноза далеко не всегда бывает достаточной для того, чтобы на основании только этого прогноза делать какие-либо далеко идущие практические выводы. Даже самые изощренные методы прогнозирования обычно могут дать только общее представление о динамике временного ряда за пределами изученного временного интервала.

§2. Быстрое преобразование Фурье

Компьютерное применение преобразования Фурье связано с использованием одного очень эффективного метода приближенного вычисления преобразования Фурье. Дело в том, что прямое применение компьютера для вычисления преобразования Фурье даже для несложной функции, заданной аналитически, наталкивалось некоторое время назад на большие трудности. Ведь для нахождения преобразования Фурье нужно вычислять несобственный интеграл от функции, зависящей от параметра, по всей числовой оси, что численно сделать удается далеко не всегда. Если же исходная функция задана не аналитически, а лишь своими значениями в ряде точек, то классические методы вычисления преобразования Фурье просто неприменимы. Для практического применения преобразования Фурье чрезвычайно полезным оказался метод БПФ – быстрого преобразования Фурье. На его основе строятся различные методы, имеющие общее название "спектральные".

§3. Специализированные математические пакеты

Здесь будут перечислены и кратко охарактеризованы несколько наиболее распространенных статистических пакетов программ (универсальных и специализированных), которые могут быть использованы при исследовании временных рядов. Более подробные сведения по использованию этих пакетов можно получить из прилагающихся к ним руководств, а также из дополнительной к этим описаниям литературы (см. ниже список литературы).

1. SPSS (для DOS и WINDOWS)

Один из самых обширных и надежных универсальных статистических пакетов. Для анализа временных рядов имеется отдельный модуль Trends. В нем можно производить выделение тренда (линейного и нелинейных девяти видов), анализ ряда остатков, выделение сезонной компоненты, сглаживание, регрессионный анализ, спектральный анализ, изучение функции авторегрессии, построение модели ARIMA. Объем обрабатываемой информации практически неограничен (лимитируется только величиной памяти компьютера).

Этот пакет очень широко используется психологами и социологами.

2. STATGRAPHICS (для DOS и WINDOWS)

Универсальный статистический пакет. Имеется специальный блок для анализа временных рядов. Этот блок содержит методы регрессии, сглаживания, сезонной декомпозиции и прогнозирования.

В пакете имеется очень полезный StatAdvisor – советчик, который комментирует результаты применных статистических методов, дает рекомендации по интерпретации и использованию этих результатов.

3. STATISTICA (только для WINDOWS)

Универсальный статистический пакет, содержит много современных статистических методик. Имеется отдельный модуль анализа временных рядов с широким набором возможностей.

4. STADIA (для DOS и WINDOWS)

Отечественный универсальный статистический пакет, имеет широкий набор методов для всех областей статистического анализа. Имеет очень хороший гипертекстовый справочник, а также экспертную систему по выбору и использованию статистических методов.

5. Отечественные специализированные статистические пакеты для анализа временных рядов:

   ЭВРИСТА, МЕЗОЗАВР (для DOS) – это хорошие и надежные пакеты . Есть и другие пакеты такого рода, но они автору не были доступны.

ГЛАВА 5. Задачи

Здесь будут приведены примеры постановки задач, которые для своего решения требуют исследования некоторого временного ряда. Эти задачи отличаются от большинства приводимых в опубликованых учебных руководствах тем, что они основаны на реальных проблемах, связанных с современностью. Основой для этих задач были темы, которые были представлены студентами МАТИ при выполнении курсовых работ по курсам лекций, читавшихся автором. Эти темы были в некоторой степени переработаны автором для того, чтобы получившиеся задачи носили более универсальный характер. Так как студенты специализировались по экологии, то многие задачи имеют достаточно выраженную экологическую направленность. Было немало задач, связанных со спортом, были и чисто бытовые задачи. Приведенные ниже задачи показывают, насколько широк спектр тем, в которых может быть использована методика обработки временных рядов.

Будет также приведен пример решения двух из этих задач (с использованием универсального пакета Statgraphics).

§1. Некоторые практические задачи исследования случайных величин и временных рядов

Во всех приведенных ниже 20 задачах нужно провести исследование соответствующего временного ряда: выделить тренд, проанализировать его значимость, проанализировать остатки на случайность и в случае их неслучайности выделить периодические компоненты. Вначале приводятся условия задач, а потом в таблицах числовые данные для них.

ЗАДАЧА 1. Исследовать динамику крупных землетрясений за год в 1940-1976 годах. Нет ли закономерностей, которые помогут предсказывать землетрясения?
ЗАДАЧА 2. Исследовать динамику числа попаданий в кольцо баскетболистом NBA Гленом Райсом в первых играх сезона 1996/1997 года. Стабилен ли он?
ЗАДАЧА 3. Исследовать коэффициент рождаемости (среднее число детей у одной семьи) в России в 1960-1995 годах. Что будет с нами дальше, не вымрем ли?
ЗАДАЧА 4. Исследовать число серьезных травм в футбольной команде в 1951-1983. Как меняется со временем футбольный травматизм?
ЗАДАЧА 5. Староста группы студентов 3 курса отмечала число студентов, отсутствующих на занятиях в течении 30 учебных дней (воскресенья не учитывались) весеннего семестра 1998/99 учебного года. Можно ли отсюда извлечь полезную информацию для деканата?
ЗАДАЧА 6. Исследовать динамику глобального потепления климата на Земле в 1910-1940 годах. За точку отсчета бралась средняя температура в 1914 году. Что происходит с нашим климатом?
ЗАДАЧА 7. Исследовать динамику цен на путевки на Кипр (в долларах США) в первые 28 недель 1999 года в одной из туристических фирм. Как купить путевку подешевле?
ЗАДАЧА 8. Исследовать динамику продолжительности залегания снежного покрова (в днях) в одном из регионов России в 1953-1979 годах. Не поможет ли это сельскому хозяйству?
ЗАДАЧА 9. Проанализировать динамику средней продолжительности жизни мужчин в России в 1959-1993 годах. Выживут ли наши мужчины?
ЗАДАЧА 10. Проанализировать динамику количества мячей, забитых футбольной командой командой “Спартак” в первых двух матчах каждого сезона в 1975-1999 годах. Играет ли команда сейчас лучше, чем прежде?
ЗАДАЧА 11. Исследовать динамику цен на бензин (в рублях за тонну) по месяцам с апреля 1996 года по март 1998 года. Случайны ли эти цены?
ЗАДАЧА 12. Исследовать помесячную динамику изменения площади осущенного дна одного водоема(в кв.км.). Насколько стабильно идет процесс осушение водоема?
ЗАДАЧА 13. Изучалась средняя длина перегонов (в км.) по 20 зонам подмосковной жедезной дороги (направление Москва-Владимир). Есть ли здесь закономерность?
ЗАДАЧА 14. Студентка изучала динамику своего артериального давления (в мм. рт. столба) с 1 по 30 апреля 1999 года. К каким выводам относительно своего здоровья она может прийти?

ЗАДАЧА 15. Исследовать динамику вылова трески в Баренцевом море в 1950-1994 годах (брались данные только за четные годы). Что посоветовать рыбакам?
ЗАДАЧА 16. Исследовать динамику числа пожаров на территории России (в тыс. случаев) в 1974-1994 годах. Как планировать число пожарных станций?
ЗАДАЧА 17. Проанализировать статистику по homerun (одно из трудновыполнимых действий в бейсболе) бейсболиста в 1974-1993 годах. Лучше ли он стал играть?
ЗАДАЧА 18. Проанализировать производительность рабочего, обучающегося на производстве трубок для газотурбинных двигателей в течении 20 дней. Выяснить, как идет процесс освоения производства.
ЗАДАЧА 19. Проанализировать динамику числа тех водоемов хозяйственно-питьевого назначения, которые не соответствовали требованиям российского стандарта (в процентах к общем числу водоемов) в 1974-1993 годах.
ЗАДАЧА 20. Проанализировать динамику реальной скорости (CPS) передачи данных (в бит/сек.) при первом подключении к серверу в локальеной сети в течении первых 25 дней в мае 1999 года.

Данные для задач 1-4 (в столбцах)

	1		2		3		4	
1	1940	1	1	21	1960	2.574	1951	10
2	1941	3	2	19	1961	2.522	1952	2
3	1942	4	3	18	1962	2.417	1953	6
4	1943	5	4	25	1963	2.322	1954	0
5	1944	1	5	20	1964	2.227	1955	13
6	1945	1	6	22	1965	2.176	1956	1
7	1946	2	7	23	1966	2.125	1957	4
8	1947	3	8	27	1967	2.062	1958	10
9	1948	3	9	21	1968	1.998	1959	4
10	1949	2	10	23	1969	1.985	1960	9
11	1950	5	11	26	1970	1.972	1961	11
12	1951	0	12	19	1971	2.013	1962	6
13	1952	3	13	28	1972	2.053	1963	7
14	1953	2	14	23	1973	2.026	1964	12
15	1954	3	15	30	1974	2.000	1965	14
16	1955	2	16	25	1975	1.985	1966	11
17	1956	2	17	27	1976	1.969	1967	9
18	1957	5	18	22	1977	1.934	1968	6
19	1958	2	19	24	1978	1.938	1969	5
20	1959	3	20	26	1979	1.913	1970	3
21	1960	4	21	28	1980	1.888	1971	6
22	1961	3	22	24	1981	1.920	1972	8
23	1962	1	23	31	1982	1.951	1973	3
24	1963	3	24	27	1983	2.017	1974	12
25	1964	2	25	23	1984	2.083	1975	10
26	1965	1	26	33	1985	2.097	1976	9

27	1966	1	27	25	1986	2.111	1977	15
28	1967	1	28	31	1987	2.194	1978	17
29	1968	5	29	26	1988	2.130	1979	5
30	1969	2	30	21	1989	2.007	1980	6
31	1970	3	31	25	1990	1.887	1981	9
32	1971	6	32	23	1991	1.732	1982	7
33	1972	3	33	29	1992	1.552	1983	11
34	1973	2	34	24	1993	1.385		
35	1974	2	35	28	1994	1.288		
36	1975	6	36	26	1995	1.198		
37	1976	4						

Данные для задач 5-8 (в столбцах)

	5		6		7		8	
1	1	6	1910	-0,24	1	300	1953	106
2	2	2	1911	-0,13	2	340	1954	150
3	3	4	1912	-0,22	3	360	1955	0
4	4	9	1913	-0,25	4	350	1956	159
5	5	8	1914	0	5	340	1957	144
6	6	2	1915	0,001	6	365	1958	131
7	7	6	1916	-0,013	7	370	1959	151
8	8	10	1917	-0,180	8	375	1960	141
9	9	2	1918	-0,26	9	370	1961	141
10	10	5	1919	-0,127	10	380	1962	148
11	11	4	1920	-0,129	11	360	1963	142
12	12	10	1921	-0,125	12	390	1964	127
13	13	7	1922	0,115	13	400	1965	143
14	14	7	1923	0,120	14	390	1966	144
15	15	4	1924	0,118	15	385	1967	133
16	16	7	1925	0,135	16	390	1968	150
17	17	3	1926	0,130	17	385	1969	136
18	18	2	1927	0,132	18	380	1970	135
19	19	11	1928	-0,130	19	370	1971	146
20	20	5	1929	0,175	20	390	1972	125
21	21	2	1930	0,175	21	380	1973	153
22	22	6	1931	0,125	22	375	1974	116
23	23	3	1932	0,178	23	395	1975	143
24	24	2	1933	0,150	24	400	1976	145
25	25	2	1934	0,152	25	390	1977	134
26	26	8	1935	0,320	26	420	1978	140
27	27	4	1936	0,400	27	400	1979	156
28	28	3	1937	0,26	28	430		
29	29	3	1938	0,150				
30	30	3	1939	0,152				
31			1940					

Данные для задач 9-12 (в столбцах)

	9		10		11		12	
1	1959	63,5	1975	8	1	792	1	37,2
2	1960	64	1976	1	2	799	2	42,5
3	1961	64	1977	3	3	814	3	33,6
4	1962	64	1978	6	4	813	4	45,8
5	1963	64,5	1979	6	5	823	5	58,9
6	1964	64,5	1980	9	6	835	6	55,2
7	1965	64,2	1981	5	7	865	7	56,9
8	1966	64,2	1982	0	8	888	8	70,1
9	1967	64	1983	2	9	912	9	69,4
10	1968	63	1984	3	10	934	10	74,7
11	1969	63	1985	5	11	902	11	71,1
12	1970	63	1986	7	12	922	12	86,3
13	1971	63	1987	2	13	933	13	70,3
14	1972	63	1988	4	14	886	14	92,5
15	1973	63	1989	5	15	907	15	68,7
16	1974	62,2	1990	8	16	932	16	93,1
17	1975	62	1991	0	17	949	17	81,4
18	1976	61,8	1992	4	18	968	18	89,7
19	1977	62	1993	7	19	968	19	94,8
20	1978	62	1994	2	20	978	20	103,9
21	1979	61,5	1995	1	21	1011	21	109,7
22	1980	61	1996	6	22	1007	22	99,3
23	1981	61,3	1997	5	23	1001	23	111,9
24	1982	62,5	1998	5	24	1054		
25	1983	62,1	1999	3				
26	1984	61,3						
27	1985	62,5						
28	1986	63,5						
29	1987	64,5						
30	1988	64,3						
31	1989	64						
32	1990	63,2						
33	1991	63,2						
34	1992	62						
35	1993	59						

Данные для задач 13-16 (в столбцах)

	13		14		15		16	
1	0	0	1	75	1950	510	1974	18,9
2	1	1,5	2	94	1952	726	1975	22,8
3	2	2,625	3	96	1954	1050	1976	15,1

4	3	2,2	4	95	1956	1300	1977	17
5	4	2,375	5	74	1958	890	1978	12,1
6	5	5,0	6	80	1960	380	1979	14,3
7	6	3,7	7	96	1962	400	1980	15,1
8	7	8,0	8	94	1964	410	1981	20
9	8	3,7	9	87	1966	430	1982	16
10	9	5,5	10	78	1968	850	1983	12
11	10	3,3	11	82	1970	1025	1984	15
12	11	3,5	12	94	1972	770	1985	12
13	12	4,25	13	86	1974	515	1986	16,1
14	13	4,0	14	74	1976	395	1987	13,5
15	14	7,0	15	87	1978	352	1988	19,3
16	15	6	16	92	1980	310	1989	22,8
17	16	7	17	76	1982	300	1990	18,2
18	17	9	18	89	1984	245	1991	17
19	18	9	19	80	1986	200	1992	23,3
20	19	5	20	82	1988	250	1993	18,2
21	20	12	21	78	1990	310	1994	20,1
22			22	94	1992	425		
23			23	75	1994	500		
24			24	92				
25			25	88				
26			26	90				
27			27	84				
28			28	92				
29			29	91				
30			30	96				

Данные для задач 17-20 (в столбцах)

	17		18		19		20	
1	1974	0	1	40	1974	10,24	1	1824
2	1975	10	2	43	1975	10,12	2	355
3	1976	5	3	44	1976	8,43	3	1747
4	1977	12	4	48	1977	13,02	4	1244
5	1978	0	5	38	1978	12,01	5	1193
6	1979	6	6	42	1979	14,9	6	1349
7	1980	13	7	48	1980	15,4	7	1497
8	1981	1	8	47	1981	15,7	8	1093
9	1982	10	9	49	1982	28,5	9	2284
10	1983	5	10	48	1983	27,6	10	659
11	1984	16	11	49	1984	27,5	11	491
12	1985	9	12	48	1985	25,9	12	555
13	1986	17	13	50	1986	26,9	13	1670
14	1987	15	14	50	1987	24,4	14	1690
15	1988	5	15	52	1988	26,8	15	95

16	1989	6	16	47	1989	28,3	16	240
17	1990	9	17	49	1990	30,2	17	2427
18	1991	11	18	53	1991	30,2	18	465
19	1992	13	19	55	1992	29,4	19	1343
20	1993	7	20	58	1993	20,4	20	2464
21							21	1045
22							22	1137
23							23	2029
24							24	1366
25							25	672

§2. Примеры решения задач

Рассмотрим задачу 1, условие и числовые данные для которой приведено выше. С помощью пакета STATGRAPHICS были найдены следующие значения статистических параметров:

Математическое ожидание числа X землетрясений =2.730

Среднеквадратичное отклонение X =1.503.

Коэффициент корреляции =0.178.

Уравнение регрессии имеет вид:

$$X=0.025T-45.55$$

Проведем анализ значимости коэффициента корреляции. Для этого вычислим величину Тэкс и сравним ее с найденной по таблице распределения Стьюдента величиной Ткр. Имеем

Тэкс=1.07

Ткр=2.02 на уровне вероятности 95% (число степеней свободы равно 35=37-2).

Ставнивая Ткр и Тэкс видим, что на уровне 95% коэффициент корреляции не является значимым, а тем самым тренд, определяемый уравнением регрессии, можно рассматривать как слабо выраженный.

Проанализируем теперь случайность ряда остатков. Для этого найдем число экстремумов – оно равно 14. Доверительный интервал на уровне 95% для числа экстремумов равен приблизительно (22.1,27.2). Так как 14 не попадает в этот интервал, то рассматривать ряд как белый шум оснований нет. Другими словами, имеет смысл искать закономерности в колебаниях ряда. Для этого используются кореллограмма и периодграмма. Кореллограмму и периодграмму можно найти с помощью той же программы STATGRAPHICS.

Кореллограмма, как оказывается, не выходит за пределы интервала, определяемого 95% уровнем, что означает отсутствие ярко выраженных корреляций между отдельными сечениями. Относительно большими являются значения кореллограммы для лагов 7 и 11. Для периодграммы имеются пики в районах частот 0.27 и 0.42. Однако в силу небольшого объема данных не следует в этой задаче придавать большого значения результатам исследования кореллограмма и периодграммы. Дело в том, что корреляционный анализ для небольших массивов данных дает очень ненадежные результаты. А преобразование Фурье и основанные на нем спектральные методы еще более капризны. Поэтому в этой задаче стоит остановить исследование на приведенных выше заключениях. Для более точного определения циклических компонент необходим более обширный экспериментальный материал.

Рассмотрим теперь решение задачи 18. Тут уравнение линейной регрессии имеет вид:

$$X=40.79+0.68T,$$

а коэффициент корреляции равен 0.83. Здесь и без всякого исследования ясно, что выявленный линейный тренд является значимым (стандартным образом проведенное исследование подтверждает, конечно, этот вывод). Тем самым можно считать установленным, что с течением времени рабочий осваивает производство и его производительность возрастает.

Что касается случайности ряда остатков, то исследование методом Юла-Кендалла показывает, что этот ряд нельзя рассматривать как белый шум. Другими словами, в ряде остатков имеются некоторые циклические компоненты. Однако применять кореллограмму и, тем более, периодграмму здесь не стоит, ибо, как и в задаче 1, объем данных недостаточно велик.

ГЛАВА 6 Циклические процессы в природе и общественной жизни

Если бы позволял объем издания, то в него бы вошли еще несколько параграфов. Ниже приводятся только их названия, которые могут подсказать читателю направление, в котором могло бы развиваться изложение в этой главе.

§1. Циклические процессы в природе.
§2. Циклические процессы в развитии человека, в его физиологии и психике.
§3. Циклические процессы в экономике.
§4. Циклические процессы в развитии общества.

ЗАКЛЮЧЕНИЕ

Нами были рассмотрены только некоторые, фундаментальные аспекты исследования временных рядов. Основной целью автора была демонстрация методики подхода к таким исследованиям и иллюстрация ее на конкретных практических задачах. Дальнейшее изучение затронутых здесь вопросов рекомендуется производить по более подробным руководствам, некоторые из которых указаны ниже в списке литературы. При выборе именно этих работ из большой массы сочинений по данной тематике основными критериями были фундаментальность подхода и современность изложения, а также ориентация на использование современных компьютерных программ.

ЛИТЕРАТУРА

1. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA. М., 1999.
2. Вентцель Е.С., Овчаров Л.А. Теория случайных процессов и ее инженерные приложения. М.,1991.
3. Дюк В. Обработка данных на ПК в примерах. СПб.,1997.
4. Кулаичев А.П. Методы и средства анализа данных. М.,1998.
5. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. М.,1998.
6. Яглом А.М. Корреляционная теория стационарных случайных функций. М.,1961.
7. Юл Дж., Кендалл М. Теория статистики. М.,1960.

СОДЕРЖАНИЕ