

МАТЕМАТИЧЕСКИЙ АППАРАТ ДЛЯ ИНВЕСТОРА.

Горчаков А.А., Аудиторская и консалтинговая фирма
«Росэкспертиза»

1. Методика статистического анализа и прогнозирования

При статистическом исследовании финансово-экономических показателей в ходе анализа вычисляют простейшие характеристики динамики их развития, выявляют закономерности прошлого развития и оценивают возможность их перенесения на будущее. Для успешного решения указанной задачи необходимо:

1. Иметь необходимый для проявления статистических закономерностей объем данных (для годовых наблюдений - не менее 5 уровней, для сезонных процессов - не менее трех периодов сезонности);

2. Обеспечить методологическую сопоставимость данных;

3. На основе содержательного анализа исследуемого показателя обосновать возможность переноса закономерностей прошлого на выбранный Вами период прогнозирования;

4. При помощи данной программы получить адекватную математическую модель и на ее основе построить точечные и интервальные прогнозы.

В случае невыполнения этапов (1-3) использовать математические методы бессмысленно!

Основной формой представления статистической информации являются временные ряды (ВР) наблюдений, т.е. ряды динамики, у которых в качестве признака упорядочения берется время. ВР, состоящий из N уровней $x(1), x(2), \dots, x(N)$, может быть записан в компактной форме: $X(t) \ t=1,2,\dots,N$, т.е. t - порядковый номер наблюдения.

Статистические методы исследования исходят из предположения о возможности представления уровней ряда в виде суммы нескольких компонент, отражающих закономерность и случайность развития, в частности, в виде суммы нескольких компонент:

$$X(t) = f(t) + S(t) + E(t) \quad (1.1)$$

где $f(t)$ - тренд (долговременная тенденция) развития;

$S(t)$ - сезонная компонента;

$E(t)$ - остаточная компонента.

Тренд представляет собой устойчивое изменение показателя в течение длительного времени. Он выражается аналитической функцией, которая используется для формирования прогнозных оценок.

Сезонная компонента характеризует устойчивые внутригодовые колебания уровней. Она проявляется в некоторых показателях, которые представлены квартальными или месячными данными. Наличие устойчивых колебаний в суточных или недельных данных может рассматриваться как циклическое и отображается сезонной компонентой.

Остаточная компонента представляет собой расхождение между фактическими и расчетными

значениями. Если построена адекватная (хорошая) модель, то $E(t)$ является близкой к 0 , случайной, независимой, подчиняющейся нормальному закону распределения компонентой. В противном случае модель является плохой.

Основной целью статистического анализа временных рядов является изучение соотношения между закономерностью и случайностью в формировании значений уровней ряда, оценка количественной меры их влияния. Закономерности, объясняющие динамику показателя в прошлом, могут быть использованы для прогнозирования его значений в будущем, а учет случайности позволяет определить вероятность отклонения от закономерного развития и их возможную величину.

Формирование уровней ряда определяется закономерностями трех основных типов: инерцией тенденции, инерцией взаимосвязи между последовательными уровнями ряда и инерцией взаимосвязи между исследуемым показателем и показателями-факторами, оказывающими на него причинное воздействие. Соответственно, различают задачи анализа и моделирования тенденций; взаимосвязи между последовательными уровнями ряда; причинных взаимодействий между исследуемым показателем и показателями-факторами. Первая из них решается с помощью методов компонентного анализа, вторая - с помощью адаптивных методов и моделей, а третья - на основе эконометрического моделирования, базирующегося на методах корреляционно-регрессионного анализа.

Статистический анализ выполняется в следующей последовательности:

1. Постановка задачи и подбор исходной информации.

2. Предварительный анализ исходных временных рядов и формирование набора моделей прогнозирования.

3. Численное оценивание параметров моделей.

4. Определение качества моделей (адекватности и точности).

5. Выбор одной лучшей или построение обобщенной модели.

6. Получение точечного и интервального прогнозов.

7. Содержательный комментарий полученного прогноза.

На первом этапе формулируется цель исследования, осуществляется содержательный (логический и экономический) анализ исследуемого процесса; решается вопрос о выборе показателя, характеризующего его наиболее полно; определяются показатели, оказывающие влияние на ход развития; определяется наиболее разумный период упреждения прогноза (горизонт прогнозирования, т.е. на сколько шагов вперед делается прогноз). Оптимальный **горизонт прогнозирования** определяется индивидуально для каждого показателя на основе содержательного суждения о его стабильности и с учетом статистической колеблемости данных. Он, как правило, не превышает 1/3 объема данных.

Предварительный анализ данных имеет целью определение соответствия имеющихся данных требованиям, предъявляемым к ним математическими методами (объективности, сопоставимости, полноты, однородности и устойчивости); строится график динамики, и рассчитываются основные динамические характеристики (приросты, темпы роста, темпы прироста, коэффициенты автокорреляции).

Набор моделей (исходная база моделей) формируется на основе интуитивных приемов (таких, например, как анализ графика динамики ряда), формализованных статистических процедур (исследование приростов уровней), исходя из целей исследования и качества имеющейся информации, а также содержательного анализа. Предпочтение отдается наиболее простым моделям, которые могут быть содержательно интерпретированы. При использовании мощных ПЭВМ эту проблему можно переложить на программы, поручив провести вычисления по всем доступным моделям и методам.

Метод наименьших квадратов (МНК) лежит в основе **численного оценивания** параметров моделей кривых роста. Параметры адаптивных методов оцениваются с использованием специальных процедур многомерной численной оптимизации. Во всех случаях основная идея оценки параметров заключается в наилучшем, т.е. максимальном приближении модели к исходным данным. Экстраполяционные методы прогнозирования строят модели кривых роста и адаптивные модели, которые используют лишь один фактор - "время". Этот фактор является условным представителем всей совокупности причинных факторов, влияющих на интересующий нас показатель. Кривые роста исходят из равноценности всех данных и отражают общую тенденцию развития, а адаптивные модели и методы исходят из большей значимости последних наблюдений и лучше отражают динамику изменения. Потенциально более мощным инструментом прогнозирования являются модели Бокса-Дженкинса и ОЛИМП. Поэтому именно они составляют основу рабочей базы моделей. Каждая построенная модель заносится в базу моделей. Максимальное количество моделей в базе моделей ограничено **20** (в текущей версии). Если рабочая база моделей заполнена (построено свыше двадцати моделей), то вновь построенная модель сравнивается с наихудшей моделью и вытесняет ее, если новая модель имеет лучшие характеристики качества. Внутренняя информация базы моделей включает в себя (для каждой модели): тип модели; количество и значения параметров построенной модели; вектор остатков; вектор прогнозов (включая границы) и ряд других.

Информация, содержащаяся в рабочей базе моделей, служит основой для построения прогноза как по лучшей модели, так и при формировании обобщенного прогноза. Методика измерения качества моделей в сочетании с высоким быстродействием современных вычислительных машин позволяет за короткое время просматривать большое количество моделей и оставлять из них наилучшие.

Качество модели с формально-статистической точки зрения оценивается на основе

ее адекватности и точности. Адекватность моделей оценивается путем исследования свойств остаточной компоненты, т.е. расхождений, рассчитанных по модели уровней и фактических наблюдений. Точность модели характеризует степень близости расчетных данных к фактическим. На основе характеристик точности и адекватности рассчитывается обобщенный показатель качества модели, который используется для определения лучшей модели.

В качестве прогнозной модели может быть выбрана лучшая модель из числа построенных, либо на основе нескольких моделей сформирована обобщенная модель (см. "Построение обобщенного прогноза").

При выборе лучшей модели следует учитывать не только формальные статистические характеристики, но и интерпретируемость их траектории развития с содержательной точки зрения. В случае несовпадения результатов выбора по статистическому и содержательному критериям предпочтение отдается последнему.

На основе построенной модели рассчитываются **точечный и интервальный** прогноз. Экстраполяция лежит в основе точечного прогноза. Он формируется путем подстановки в модель (уравнение тренда) соответствующего значения фактора "Время", т.е. $t=N+1, N+2...N+k$. Интервальные прогнозы строятся на основе точечных.

Доверительная вероятность прогноза характеризует степень уверенности в попадании прогнозируемой величины в построенный интервал прогнозирования. Она изменяется в пределах от 0 до 100% и задается пользователем. Следует помнить, что при ее увеличении интервальный прогноз расширяется, и потому полезность прогноза **обратно пропорциональна** доверительной вероятности. Можно построить прогноз, который свершится с вероятностью 99%, однако с практической точки зрения он будет бесполезен (например, прогноз погоды: ожидается температура воздуха от 5 до 25 градусов - не дает необходимой информации для принятия правильного решения о форме одежды). С математической точки зрения доверительной вероятностью для расчета прогноза можно пользоваться лишь при получении адекватной математической модели.

После получения прогнозных оценок необходимо убедиться в их разумности и непротиворечивости. Полученный прогноз должен быть подвергнут критическому рассмотрению с целью выявления возможных противоречий известным фактам и сложившимся к настоящему моменту представлениям о характере развития на периоде упреждения прогноза. В качестве средства оценки эффективности математического аппарата при исследовании конкретных процессов часто применяют ретропрогноз. При наличии данных о динамике других показателей можно построить модель их влияния на основной исследуемый показатель и в случае ее высокого качества получить прогнозные оценки. Для формирования набора факторов кроме содержательных аспектов необходимо учитывать формально - статистические, которые основываются на коэффициентах корреляции. Следовательно, перед регрессионным анализом

необходимо воспользоваться корреляционным анализом, а при необходимости получения прогнозов еще и экстраполяционными моделями.

2. Дескриптивная статистика

Исходные данные могут быть охарактеризованы простейшими средствами описательной статистики. Они позволяют получить представление об особенностях исследуемого показателя и перспективности использования более глубоких методов анализа.

Ниже приводятся формулы вычисления основных характеристик данных, в которых x_i - численные значения наблюдений переменной $X, i=1,2,...,n$.

Среднее значение:	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Среднеквадратическое отклонение (СКО):	
Дисперсия:	$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Несмещенная оценка дисперсии:	$s_H^2 = \frac{n}{n-1} s^2$
Среднеквадратическое отклонение для несмещенной оценки дисперсии:	$s_H = \sqrt{s_H^2}$
Среднее линейное отклонение:	$l_o = \frac{\sum_{i=1}^n x_i - \bar{x} }{n}$
Моменты начальные:	
2-го порядка:	$n^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$
3-го порядка:	$n^3 = \frac{1}{n} \sum_{i=1}^n x_i^3$
4-го порядка:	$n^4 = \frac{1}{n} \sum_{i=1}^n x_i^4$
Моменты центральные:	
3-го порядка:	$m^3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$
4-го порядка:	$m^4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4$
Коэффициент асимметрии:	$A_s = \frac{m^3}{s^3}$
Несмещенная оценка коэф. асимметрии:	$A'_s = \frac{\sqrt{(n-1)n}}{n-2} A_s$
Среднеквадратическое отклонение (СКО) коэффициента асимметрии	$S_A = \sqrt{\frac{6 \cdot n(n-1)}{(n-2)(n+1)(n+3)}}$
Показатель эксцесса:	$E = \frac{m^4}{s^4}$
Несмещенная оценка:	$E' = \frac{(n-1)}{(n-2)(n-3)} [(n+1)E + 6]$

Среднеквадратич. отклонение:	$S_E = \sqrt{\frac{24 \cdot n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}$
------------------------------	---

Коэффициенты асимметрии и эксцесса позволяют сделать предварительные заключения о близости изучаемого распределения к нормальному. Распределение принято считать нормальным, если выполняются условия: $A_s \leq 3S_A$ и $E \leq 5S_E$,

Коэффициенты вариации:	
$\sum$ по размаху:	$R / \bar{x}$
$\sum$ по среднему линейному отклонению:	$l_o / \bar{x}$
$\sum$ по среднеквадратическому отклонению:	$s / \bar{x}$
Медиана(исходный ряд считается отсортированным)	$X_{n/2}$
Мода -	Значение X , которое наблюдается наиболее часто
Минимальное значения ряда	X_{min}
Максимальное значения ряда	X_{max}
Размах	$R = X_{max} - X_{min}$

Для изучения пространственных данных используется технология их агрегирования путем построения интервального ряда. Ширина интервала для группировки (H) определяется следующим образом:

$$H = \frac{R}{L}, \quad L = 1 + 3.322 \cdot \lg(n),$$

где L - количество интервалов (округляется в большую сторону);

n - число членов ряда.

Если установлен соответствующий параметр, то изменяется значение H и пересчитывается L . Каждый j -й интервал ($j = 1, \dots, L$) характеризуется определенной частотой и частотой попадания в него соответствующих наблюдений заданного ряда.

Таблица интервального ряда распределений содержит разбивку данных на интервалы, числовую характеристику интервала (начало, середину и конец), а также частоту и частоту наблюдений.

В качестве характеристик интервального ряда используются:

- среднее значение
- дисперсия
- среднеквадратическое отклонение
- коэффициенты асимметрии и эксцесса
- мода и медиана

Смысл и назначение этих характеристик совпадает с вариационными характеристиками, а формулы вычисления содержат компоненту, учитывающую частоту попадания наблюдений в интервалы.

Бутстреп-оценки

Сущность метода сводится к дополнению данных фактических наблюдений данными численного моделирования. При этом моделирование произво-

дится только в рамках фактических данных. Входные параметры метода:

$\{X_1, X_2, \dots, X_n\}$ - исходная выборка;

k - количество моделируемых выборок ($k > 50$);

p - вероятностный уровень оценки математического ожидания

(рекомендуемые значения 0.7-0.9).

Решается задача оценки математического ожидания для малой выборки по следующему алгоритму:

1. Моделирование выборок с использованием датчика натуральных чисел, равномерно распределенных в интервале от 1 до n

$$v_j = \{X_{N_1^j}, X_{N_2^j}, \dots, X_{N_n^j}\}$$

2. Для каждой выборки V_j ищется оценка математического ожидания:

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{N_i^j}, \quad j = \overline{1, k}$$

3. Для вариационного ряда математических ожиданий выборки строится интервальный ряд, как описано в предыдущем разделе.

4. С хвостов построенного интервального ряда отсекаются интервалы таким образом, чтобы суммарная часть отброшенных интервалов не превосходила $(1-p)$. Оставшиеся интервалы определяют интервальную оценку математического ожидания.

3. Анализ временных рядов

Характеристика динамики

Динамика изменения исследуемого показателя может быть охарактеризована по отношению к какому-то базисному (обычно первому) наблюдению и величиной изменения соседних уровней. В этой связи вычисляются **базисные** и **цепные** характеристики. В качестве статистических характеристик временного ряда $Y_i, i = \overline{1, N}$ используются следующие величины:

абсолютный базисный прирост	$DY_i^b = Y_i - Y_1$
абсолютный цепной прирост	$DY_i^c = Y_i - Y_{i-1}$
базисный коэффициент роста	$Kr_i = Y_i / Y_1$
цепной коэффициент роста	$Kr_i^c = Y_i / Y_{i-1}$
базисный коэффициент прироста	$Kpr_i^b = (Y_i - Y_1) / Y_1$
цепной коэффициент прироста	$Kpr_i^c = (Y_i - Y_{i-1}) / Y_{i-1}$
темп роста	$Tr = Kr^b \cdot 100\%$
темп прироста	$Trp = Tr - 100\%$
средняя арифметическая	$y = \frac{1}{N} \sum_{j=1}^N Y_j$
средний абсолютный прирост	$\overline{DY} = (Y_n - Y_1) / (N - 1)$

средний темп роста	$\overline{Tr} = N \sqrt[N]{Y_N / Y_1} \cdot 100\%$
средний темп прироста	$\overline{Trp} = \overline{Tr} - 100\%$
	где N - число уровней ряда, Y_i - уровни ряда.

Примечание. Использование показателя средней арифметической величины для характеристики процессов, представленных временными рядами с ярко выраженной тенденцией, является некорректным.

Оценка наличия тренда

Оценка наличия тренда в исследуемом временном ряду осуществляется при помощи методов Фостера-Стюарта и средних в соответствии с методикой, которая изложена в работе Четыркина Е.М.[27]. В случае противоречивости их выводов предпочтение отдается первому методу.

В соответствии с методом проверки существенности разности средних исходный временной ряд разбивается на две равные (или почти равные) части, после чего проверяется гипотеза о существенности разности средних для этих частей. Недостаток метода состоит в невозможности правильно определить наличие тренда в том случае, когда временной ряд содержит точку изменения тенденции в районе середины ряда.

В методе Форстера-Стюарта гипотеза об отсутствии тренда проверяется с помощью вспомогательных функций:

$$L = \sum_{t=2}^T I_t,$$

$$I_t = U_t - V_t$$

$$u_t = \begin{cases} 1, & \text{если } y_t < y_{t-1}, \dots, y_1 \\ 0, & \text{иначе} \end{cases}$$

$$v_t = \begin{cases} 1, & \text{если } y_t > y_{t-1}, \dots, y_1 \\ 0, & \text{иначе} \end{cases}$$

Проверяется гипотеза о том, что $L=0$. Для проверки строится t -статистика:

$$t_L = \frac{L}{s_L}, \quad \text{где } s_L^2 = 2 \sum_{t=2}^t 1/t,$$

которая имеет распределение Стьюдента с $T-1$ степенями свободы. Гипотеза об отсутствии тенденции отклоняется, если расчетное t -значение больше табличного на выбранном уровне значимости 0.95.

Проверка однородности данных

Проверка однородности данных обычно проводится на основе критерия Ирвина, который основан на сравнении соседних значений ряда. В соответствии с ним рассчитывается характеристика I_t :

$$I_t = \frac{Y_t - Y_{t-1}}{s_y}.$$

Полученные значения сравниваются затем с табличными значениями. Однако критерий Ирвина недостаточно эффективен для выявления аномальности в динамических рядах, потому что величина

s_y характеризует отклонения значений показателя от среднего уровня по **всей** совокупности наблюдений, а значит, он не ловит выбросы внутри ряда наблюдений. В программе используется модифицированный метод, в соответствии с которым последовательно рассчитываются s_y не по всей совокупности, а по 3-4 наблюдениям, и рассчитанные с такими скользящими значениями s_y величины сравниваются с критическими значениями I^* для $n=3$.

Проверка не производится для временных рядов с периодом сезонности более единицы, а также для уровней на концах периода наблюдения.

Оценка автокорреляционных свойств

Оценка свойств сводится к исследованию автокорреляционной и частной автокорреляционной функции исходного и разностных рядов. Анализ автокорреляции выполняется с помощью графика и критических значений коэффициентов.

Автокорреляционная функция представляет собой совокупность коэффициентов автокорреляции, вычисленных для исследуемого показателя или разностного ряда.

Она используется для оценки тесноты взаимосвязи уровней ряда и подбора соответствующих авторегрессионных моделей. Анализ автокорреляции выполняется с помощью графика автокорреляции; коэффициенты автокорреляции для его построения вычисляются по формуле:

$$r_L = \frac{(N-L) \sum_{t=1}^{N-L} Y_t Y_{t+L} - \sum_{t=1}^{N-L} Y_t \sum_{t=1}^{N-L} Y_{t+L}}{\sqrt{[(N-L) \sum_{t=1}^{N-L} Y_t^2 - (\sum_{t=1}^{N-L} Y_t)^2] \cdot [(N-L) \sum_{t=1}^{N-L} Y_{t+L}^2 - (\sum_{t=1}^{N-L} Y_{t+L})^2]}}$$

$t = 1, N, L = 0, 1, 2, \dots$

Частная автокорреляционная функция вычисляется по формуле:

$$r_{kk} = \begin{cases} r_1, & k=1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-j} r_j}{1 - \sum_{j=1}^{k-1} r_j^2}, & k=2, 3, \dots, K \end{cases}$$

где K - максимальная задержка (лаг) функции (обычно $K \leq n/4$);

r - автокорреляционная функция (АКФ).

“Чистые” авторегрессионные процессы имеют плавно затухающую АКФ и резко прерывающуюся ЧАКФ. В этом случае в качестве порядка АР-модели выбирают лаг, после которого все ЧАКФ имеют незначительную величину.

4. Прогнозирование временных рядов

Для прогнозирования несезонных и сезонных процессов используется различный математический аппарат.

Динамика многих финансово-экономических показателей имеет устойчивую колебательную составляющую. При исследовании месячных и квартальных данных часто наблюдаются внутригодичные сезонные колебания соответственно с периодом 12 и 4. При использовании дневных наблюдений часто наблюдаются колебания с недельным (пятидневным) циклом. В этом случае для получения более точных прогнозных оценок необходимо не только правильно отобразить тренд, но и колебательную компоненту. Решение этой задачи возможно только при использовании специального класса методов и моделей.

В основе сезонных моделей лежат их несезонные аналоги, которые дополнены средствами отражения сезонных колебаний. Сезонные модели способны отражать как относительно постоянную сезонную волну, так и динамически изменяющуюся в зависимости от тренда. Первая форма относится к классу аддитивных, а вторая - к классу мультипликативных моделей. Большинство моделей имеет обе эти формы. Наиболее широко в практике используются модели Хольта-Уинтерса, авторегрессии, Бокса-Дженкинса.

Кривые роста

Для аналитического выравнивания временных рядов используются функции с одним параметром t , представляющим собой порядковый номер наблюдения ($t=1, 2, \dots, N$), который интерпретируется как “Время”. Модели этого класса получили название “кривые роста”. Оценка их параметров производится аналогично построению парной регрессии, в которой объясняющей переменной является время. Для кривых роста реализованы те же вычислительные процедуры, что и в парной регрессии. Как показывает практика, для целей краткосрочного и среднесрочного прогнозирования они являются надежным инструментом.

Метод наименьших квадратов является основным методом численной оценки параметров кривых роста. Оценка качества модели производится по критерию минимума средней квадратической ошибки. Аппроксимация наблюдений сложными функциями дает хорошее приближение к фактическим наблюдениям, но снижает устойчивость модели на периоде прогнозирования. Поэтому использовать для прогнозирования такие модели (например, полином выше второй степени) очень опасно. Особое место среди 18 задействованных в программе моделей занимают две функции, которые не сводятся к модели линейной регрессии. Это функции - Гомперца и Логистическая кривая. Для поиска их параметров используется метод многомерной численной оптимизации (в настоящей версии программы - метод деформируемого многогранника).

Экстраполяция траектории модели за период наблюдения, т.е. подстановка в модель очередного значения фактора “Время” $t=N+1, N+2, \dots$, является основой прогнозирования трендовых моделей. Интервальный прогноз в каждой прогнозной точке определяется по соотношениям регрессионного анализа с заданной пользователем доверительной вероятностью.

Адаптивные методы прогнозирования

При краткосрочном прогнозировании обычно более важна динамика развития исследуемого показателя на конце периода наблюдений, а не тенденция его развития, сложившаяся в среднем на всем периоде предыстории. Свойство динамичности развития финансово-экономических процессов часто преобладает над свойством инерционности. Поэтому более эффективными являются адаптивные методы, учитывающие информационную неравнозначность данных.

Адаптивные модели и методы имеют механизм автоматической настройки на изменение исследуемого показателя. Инструментом прогноза является модель (см. Базовые адаптивные модели), первоначальная оценка параметров которой производится по нескольким первым наблюдениям. На ее основе делается прогноз, который сравнивается с фактическими наблюдениями. Далее модель корректируется в соответствии с величиной ошибки прогноза и вновь используется для прогнозирования следующего уровня, вплоть до исчерпания всех наблюдений. Таким образом, модель постоянно "впитывает" новую информацию, приспосабливается к ней и к концу периода наблюдения отображает тенденцию, сложившуюся на текущий момент. Прогноз получается как экстраполяция последней тенденции. В различных методах прогнозирования процесс настройки (адаптации) модели осуществляется по-разному. Базовыми адаптивными моделями являются:

Модель Брауна;

Модель Хольта;

Модель авторегрессии.

Первые две модели относятся к схеме скользящего среднего, последняя - к схеме авторегрессии. Многочисленные адаптивные методы базируются на этих моделях и различаются между собой способом числовой оценки параметров, определения параметров адаптации и компоновкой.

Согласно **схеме скользящего среднего**, оценкой текущего уровня является взвешенное среднее всех предшествующих уровней, причем веса при наблюдениях убывают по мере удаления от последнего (текущего) уровня, т.е. информационная ценность наблюдений тем больше, чем ближе они к концу периода наблюдений.

Согласно **схеме авторегрессии**, оценкой текущего уровня является взвешенная сумма "p" предшествующих уровней (их количество называется порядком модели). Информационная ценность наблюдений определяется не их близостью к моделируемому уровню, а теснотой связи между ними.

Обе эти схемы имеют механизм отображения колебательного (сезонного или циклического) развития исследуемого процесса.

Модель Брауна

Пусть $X(t)$, $t=1, \dots, n$ - временной ряд наблюдений. Прогноз в момент времени t на τ шагов вперед может быть получен по формуле:

$$\hat{x}_t(t) = a_{1,t} + a_{2,t} \cdot t$$

где $a_{1,t}$ и $a_{2,t}$ - текущие оценки коэффициентов адаптивного полинома.

В модели Брауна модификация (адаптация) коэффициентов линейной модели осуществляется следующим образом:

$$a_{1,t} = a_{1,t-1} + a_{2,t-1} + (1-b^2)e_t$$

$$a_{2,t} = a_{2,t-1} + (1-b)^2 e_t$$

где b - коэффициент дисконтирования данных;

e_t - ошибка прогнозирования, $e_t = x_t - \hat{x}_{t-1}$

Начальные значения параметров модели определяются по МНК на основе нескольких первых наблюдений. Оптимальное значение параметра дисконтирования находится в пределах от нуля до единицы, определяется методом численной оптимизации и является постоянным для всего периода наблюдений.

Оператор B сдвигает всю последовательность на один шаг назад: $Bx(t) = x(t-1)$. Применение оператора B к наблюдениям и к коэффициентам адаптивного полинома позволяет выразить модель Брауна в виде:

$$(1-B)^2 x_t = (1-2Bb + B^2 b^2) e_t,$$

из чего следует, что модель Брауна можно трактовать как модель авторегрессии - скользящего среднего $APCC(p,d,q)$ с $p=0$, $d=2$, $q=2$ и коэффициентами скользящего среднего $-2b$ и b^2 .

В таблице "Параметров модели" для модели Брауна отображается оптимальное значение коэффициента b .

Модель Хольта

В модели Хольта коэффициенты линейной модели

$$\hat{x}_t(t) = a_{1,t} + a_{2,t} \cdot t$$

модифицируются по следующим соотношениям:

$$a_{1,t} = a_{1,t-1} + a_{2,t-1} + a_1 e_t$$

$$a_{2,t} = a_{2,t-1} + a_2 e_t$$

Начальные значения параметров модели находятся по МНК на основе нескольких первых наблюдений. Оптимальные значения параметров сглаживания a_1 и a_2 находятся в пределах от нуля до единицы. Они определяются методом многомерной численной оптимизации и являются постоянными для всего периода наблюдений.

Аналогично модели Брауна, модель Хольта в терминах АРС-моделей представима в виде:

$$(1-B)^2 x_t = (1-(2-(a_1+a_2)B+(1-a_1)B^2)) e_t$$

.Формулировка адаптивных моделей в терминах линейных параметрических моделей авторегрессии - скользящего среднего позволяет трактовать их как подмножество класса линейных параметрических

моделей. Таким образом, устанавливается соответствие между двумя, вообще говоря, различными подходами к моделированию временных рядов.

В таблице параметров модели для модели Хольта отображаются оптимальные значения коэффициентов a_1, a_2 .

Модель Хольта-Уинтерса

Модель для сезонных процессов существует в аддитивной форме и мультипликативной. Прогноз на t шагов вперед для аддитивной формы строится по формуле:

$$\hat{x}_t(t) = a_{1,t} + a_{2,t} \cdot t + g_{t-s+t},$$

а модификация параметров производится по соотношениям:

$$a_{1,t} = a_1(x_t - g_{t-s}) + (1 - a_1) \cdot (a_{1,t-1} + a_{2,t-1})$$

$$a_{2,t} = a_3(a_{1,t} - a_{1,t-1}) + (1 - a_3) \cdot a_{2,t-1}$$

где g - фактор сезонности,
 s - период сезонного цикла.

Для несезонных временных рядов вычислительные формулы упрощаются за счет исключения сезонной компоненты. При построении модели производится численная оптимизация параметров адаптации, значения которых изменяются от нуля до единицы.

Модель авторегрессии

В модели авторегрессии $AP(p)$ порядка " p " текущий уровень ряда представляется в виде взвешенной суммы " p " предыдущих наблюдений:

$$X(t) = a_1 \cdot X(t-1) + a_2 \cdot X(t-2) + \dots + a_p \cdot X(t-p)$$

Параметры модели могут быть оценены по МНК (простая авторегрессия) или иным методом (как в методе Бокса-Дженкинса). Порядок авторегрессии (величина " p ") определяется путем перебора, а его начальная оценка формируется на основе анализа автокорреляционной функции. Лучшей считается величина, при которой достигнута наименьшая дисперсия ошибок.

В сезонной модели авторегрессии $AP(p)$ порядок выбирается равной периоду сезонности (колебаний). Во многих случаях сезонная $AP(p)$ - модель с оценками по МНК оказывается "перегруженной" незначимыми коэффициентами, и вследствие этого она обычно уступает аналогичной модели Бокса-Дженкинса.

Для повышения устойчивости модели в большинстве случаев целесообразно строить ее для стационарного процесса, т.е. ряда с исключенной тенденцией. В программе удаление тенденции осуществляется на основе **разностного оператора**.

Метод Бокса - Дженкинса

Если временной ряд стационарный, что означает наличие статистического равновесия относительно постоянной средней c , он может быть представлен широким классом линейных моделей, называемых моделями авторегрессии-скользящего среднего (АРСС). Это значит, что

$$(z_t^* - c) = f_1(z_{t-1}^* - c) - \dots - f_p(z_{t-p}^* - c) + a_t - q_1 a_{t-1} - \dots - q_q a_{t-q}$$

где $z_t^* = z_t^{(1)}$ - значения предварительно

преобразованной переменной,

a_t - процесс "белого шума",

$f_1, \dots, f_p$ - параметры авторегрессии,

$q_1, \dots, q_q$ - параметры скользящего среднего.

Если использовать оператор сдвига назад B

$$Bz_t = z_{t-1},$$

то АРСС-модель можно записать в операторной форме:

$$z_t^* - c = \frac{q(B)}{f(B)} a_t$$

Параметры должны удовлетворять следующим условиям:

Для стационарности корни уравнения $f(B) = 0$ должны лежать вне единичного круга для оператора авторегрессии $f(B)$ (ряды находятся в статистическом равновесии относительно фиксированного среднего),

Для обеспечения обратимости корни уравнения $q(B) = 0$ должны лежать вне единичного круга для оператора скользящего среднего $q(B)$.

Чтобы добиться экономии параметров, в модель включают одновременно операторы авторегрессии и скользящего среднего.

В то время как авторегрессионные модели и модели скользящего среднего были известны относительно давно, их использование в моделировании временных рядов было затруднено по следующим причинам:

отсутствие соответствующих методов идентификации, оценивания и контроля этих моделей,

наличие неадекватных методов для описания нестационарных рядов.

При формализации нестационарных рядов используют такие классы моделей, которые пригодны для представления широкого диапазона практических ситуаций, т.е. используют конечные разности порядка d :

$$w_t = D^d z_t^*$$

(Конечная разность первого порядка

$$Dz_t^* = z_t^* - z_{t-1}^*).$$

Стационарный ряд можно затем представить в помощью АРСС модели

$$w_t - c = \frac{q(B)}{f(B)} a_t.$$

Определенная выше модель называется авторегрессионной интегрированной моделью скользящего среднего, или **АРИСС(p,d,q)**. Взаимосвязанная статистическая методика, включающая в себя:

• идентификацию временного ряда (т.е. определение размерностей операторов конечной разности, авторегрессии и скользящего среднего),

- оценивание параметров модели,
- проверку адекватности модели,
- получила название метода Бокса-Дженкинса по имени авторов.

Сезонная модель Бокса-Дженкинса содержит сезонные операторы конечной разности, авторегрессии и скользящего среднего. В операторном виде она приобретает вид:

$$D^d D_s^p f(B) F(B^S) Z_t^* = q(B) q(B^S) Q_t,$$

где S - период сезонности,

D_s^p - оператор сезонной конечной разности,

$$D_s Z_t^* = Z_t - Z_{t-s}$$

D - порядок сезонной конечной разности,

F - оператор сезонной авторегрессии порядка P ,

q - оператор сезонного скользящего среднего порядка Q ,

d, f, Q - определены выше.

Модель называется сезонной моделью авторегрессии-скользящего среднего $(p, d, q)x(P, D, Q)$. Основные этапы разработки сезонной модели такие же, как и для несезонной модели.

Метод ОЛИМП

Метод ОЛИМП является распространением моделей авторегрессии - скользящего среднего для моделирования нестационарных временных рядов. Нами теоретически доказано, что такое обобщение корректно для широкого класса временных рядов.

Формально соотношения модели ОЛИМП соответствуют модели $APCC(p, q)$, за исключением того, что на вход модели поступает нестационарный, вообще говоря, временной ряд. Так же как и для несезонных моделей, сезонная модель ОЛИМП отличается от $APCC$ -моделей тем, что на ее вход могут поступать нестационарные временные ряды, которые не приводятся к стационарным путем взятия конечных разностей. В операторном виде модель ОЛИМП $(p, q)x(P, Q)$ имеет вид:

$$f(B) F(B^S) Z_t^* = q(B) q(B^S) Q_t.$$

С точки зрения общих соображений размерности операторов авторегрессии для модели ОЛИМП должны быть несколько больше, чем для модели Бокса-Дженкинса при моделировании одинаковых временных рядов.

Если идентифицирована модель Бокса-Дженкинса с параметрами p, d, q , то соответствующая модель ОЛИМП должна иметь параметры: $p' = p + d, q' = q$.

В работе [3, 4], доказано следующее утверждение. Пусть процесс y_t удовлетворяет стохастическому разностному уравнению порядка p (авторегрессионный процесс)

$$\sum_{i=0}^p F_i y_{t-i} = u_t,$$

где F_i - коэффициенты оператора авторегрессии,

u_t - последовательность независимых одинаково распределенных случайных величин с диспер-

сией s_u^2 , $t=0, 1, \dots, T$, известны начальные значения $y_{-p}, y_{-p+1}, \dots, y_{-1}$. Тогда прогноз вида

$$E(y_t | y_{t-1}, \dots, y_0) = - \sum_{i=1}^p F_i y_{t-i},$$

где E - оператор математического ожидания, будет иметь наименьшую дисперсию вне зависимости от значения корней характеристического уравнения.

Сравнительные характеристики двух подходов к моделированию авторегрессионных процессов приведены в таблице.

Характеристики	Старая схема	Новая схема
Тип моделируемых процессов	Стационарные	Стационарные и нестационарные
Исходные предположения	Остатки независимы и одинаково распределены	Остатки независимы, одинаково распределены, заданы начальные условия
Базовое представление наблюдений	$y_t = \sum_{i=0}^{\infty} d_i u_{t-i}$	$y_t = \sum_{i=0}^t d_i u_{t-i} + \sum_{j=1}^p a_j y_{t-j}$
Ограничения	Корни характеристического уравнения вне единичного круга	Нет ограничений
Вид прогноза	$E(y_t y_1, \dots, y_p) = - \sum_{i=1}^p b_i y_{t-i}$	$E(y_t y_{t-1}, \dots, y_p) = - \sum_{i=1}^p b_i y_{t-i}$

Оказалось также, что статистические оценки модели являются состоятельными также вне зависимости от значения корней характеристического уравнения. В практическом плане свойства состоятельности оценок оказываются вполне достаточны для их использования.

Оценка качества моделей

Качество модели оценивается, как правило, двумя дополняющими друг друга характеристиками: точностью и адекватностью. Каждая из них, в свою очередь, имеет несколько критериев. Они с разных сторон и не всегда однозначно характеризуют исследуемый процесс. Поэтому существует необходимость в их интегрированной оценке. На основе отдельных критериев точности и адекватности, рассмотренных ниже, формируется обобщенный критерий.

Схема формирования интегрированных критериев точности и адекватности, а также общего критерия качества прогнозирования состоит в следующем. С помощью механизма параметров пакета формируется состав отдельных критериев, на основе которых рассчитывается интегрированный показатель. Так, точность может характеризоваться только коэффициентом детерминации, или дисперсией и средней ошибкой аппроксимации, или всеми тремя перечисленными выше критериями точности.

Предварительно для каждого отдельного критерия разрабатывается процедура его нормировки.

Нормированный критерий получается из исходной статистики критерия таким образом, чтобы выполнялись условия:

- нормированный критерий равен 100, если модель абсолютно точная (адекватная),
- нормированный критерий равен 0, если модель абсолютно неточная (неадекватная).

Проблема нормирования решается специальным образом для каждого из критериев качества модели прогнозирования. Числовое значение каждого показателя лежит в диапазоне от 0 до 100. То же самое относится к интегрированному критерию адекватности.

Обобщенный критерий качества модели формируется как взвешенная сумма обобщенного критерия точности и обобщенного критерия адекватности. Веса этих слагаемых составляют соответственно 0.75 и 0.25, т.е. точностным характеристикам придается больший вес. В качестве представителя характеристик точности используется нормированное значение средней относительной ошибки аппроксимации, а в качестве представителя критериев адекватности - нормированное значение критерия Дарбина-Уотсона и характеристики нормального закона распределения остаточной компоненты. Числовое значение обобщенного критерия качества лежит в диапазоне от 0 до 100. Минимальное значение соответствует абсолютно плохой модели, а максимальное - идеально отображающей развитие показателя. Обобщенный критерий качества модели сформирован в соответствии со схемой формирования интегрированных критериев. Наш опыт применения этого показателя показывает, что достаточно надежными являются модели, имеющие оценку качества не менее 75.

Формально-статистический выбор лучшей модели во многих случаях не дает полной уверенности в его правильности. Поэтому кроме указанной программой модели целесообразно просмотреть результаты прогнозирования других моделей, имеющих близкое значение критерия качества.

Адекватность моделей

АДЕКВАТНЫМИ моделями считаются такие, у которых остаточная компонента имеет свойства независимости, случайности и нормальности распределения.

Критерий **Дарбина-Уотсона** является наиболее распространенным критерием для проверки корреляции внутри ряда. Если величина

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2},$$

где e_i - расхождение между фактическими и расчетными уровнями, имеет значение, близкое к 2, то можно считать модель регрессии достаточно адекватной.

Для построения интервального прогноза необходимо выполнение свойства нормальности распределения остаточной компоненты. Оценка выполнения этого свойства осуществляется на основе коэффи-

циентов асимметрии и эксцесса, которые приведены в разделе дескриптивных статистик.

При оценке адекватности уравнения регрессии учитывается также корреляционное отношение, которое характеризует долю дисперсии зависимой переменной, объясняемой уравнением регрессии. Корреляционное отношение рассчитывается по формуле:

$$h = \sqrt{1 - \left(\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \right)},$$

где $\bar{y}_i$ - расчетные значения зависимой переменной,
 $\bar{y}$ - среднее значение.

Точность модели

Точность модели характеризует близость расчетных наблюдений к фактическим на периоде аппроксимации. Считается, что модели с меньшим расхождением между фактическими и расчетными значениями лучше отражают исследуемый процесс. Для характеристики степени близости используются:

- среднее квадратическое отклонение (или дисперсия), учитывающее сложность модели;
- коэффициент детерминации (чем ближе к 1, тем более точная модель);
- средняя относительная ошибка аппроксимации (чем ближе к 0, тем точнее модель);
- среднее значение (должно быть близко к нулю);
- максимальное отклонение.

Статистически **точность прогнозов** можно оценить только используя **ретропрогноз**, суть его состоит в построении модели по усеченному объему данных ($N-k$) точек с последующим сравнением прогнозных оценок с известными (фактическими), но умышленно "забытыми" к уровням ряда. По результатам сравнения вычисляются следующие показатели точности:

- среднее значение;
- среднеквадратическое отклонение;
- средний модуль ошибок прогнозирования (%);
- максимальное и минимальное отклонение.

Чем меньше значение этих величин, тем выше качество ретропрогноза. Этот подход дает хорошие результаты, если на периоде ретропрогноза не сойдутся принципиально новых закономерностей.

Построение обобщенного прогноза

На практике, часто встречается ситуация, когда среди построенных моделей несколько оказались адекватными, а различия между их характеристиками точности невелики. В этом случае целесообразно строить обобщенный прогноз. В программе он формируется как линейная комбинация частных прогнозов:

$$\hat{y}_0 = \sum_{j=1}^M p_j y_j,$$

где M - число объединяемых прогнозов;

p_j - весовые коэффициенты частных прогнозов;

y_j - частные прогнозы.

Весовые коэффициенты определяются из условия минимума дисперсии ошибок обобщающего прогноза - т.е. максимума его точности, которая находится как сумма всех элементов ковариационной матрицы ошибок частных прогнозов с соответствующими весами:

$$s_0^2 = \sum_{i=1}^M \sum_{j=1}^M k_{ij} p_i p_j,$$

$$k_{ij} = s_i s_j r_{ij},$$

где k_{ij} - корреляционный момент, характеризующий совместное распределение ошибок

i и j частных прогнозов;

s_i, s_j - средние квадратичные ошибки;

r_{ij} - коэффициент корреляции между рядами ошибок частных прогнозов y_i и y_j .

На весовые коэффициенты накладывается ограничение: их сумма должна давать единицу. Это необходимое условие того, чтобы дисперсия обобщающего прогноза не превышала дисперсии частных прогнозов. Тогда ковариационная матрица ошибок частных прогнозов будет иметь вид:

$$K = \begin{bmatrix} s_1^2 & k_{12} & \dots & k_{1M} \\ k_{21} & s_2^2 & \dots & k_{2M} \\ \dots & \dots & \dots & \dots \\ k_{M1} & k_{M2} & \dots & s_M^2 \end{bmatrix}$$

Дисперсия обобщающего прогноза будет равна сумме всех элементов матрицы:

$$\begin{bmatrix} s_1^2 p_1^2 & k_{12} p_1 p_2 & \dots & k_{1M} p_1 (1 - \sum_{j=1}^{M-1} p_j) \\ k_{21} p_1 p_2 & s_2^2 p_2^2 & \dots & k_{2M} p_2 (1 - \sum_{j=1}^{M-1} p_j) \\ \dots & \dots & \dots & \dots \\ k_{M1} p_1 (1 - \sum_{j=1}^{M-1} p_j) & k_{M2} p_2 (1 - \sum_{j=1}^{M-1} p_j) & \dots & s_M^2 (1 - \sum_{j=1}^{M-1} p_j) \end{bmatrix}$$

В точке минимума функции $s_0^2(p)$ все $(M-1)$ первые частные производные должны обращаться в нуль.

Приравняв к нулю все $(M-1)$ первые частные производные по переменным $p_1, p_2, \dots, p_{M-1}$ получаем систему $(M-1)$ линейных уравнений с $(M-1)$ неизвестными:

$$\begin{cases} p_1(s_1^2 - 2k_{1M} + s_M^2) + \sum_{i=2}^{M-1} p_i(k_{i1} - k_{1M} - k_{iM} + s_M^2) = s_M^2 - k_{1M} \\ p_{M-1}(s_{M-1}^2 - 2k_{(M-1)M} + s_M^2) + \sum_{i=1}^{M-2} p_i(k_{i(M-1)} - k_{(M-1)M} - k_{iM} + s_M^2) = s_M^2 - k_{(M-1)M} \end{cases}$$

Коэффициенты при переменных составят матрицу В, элементы которой определяются следующим образом:

$$b_{ij} = b_{ji} = \begin{cases} s_M^2 + k_{ij} - k_{iM} - k_{jM} & i \neq j \\ s_M^2 - 2k_{jM} + s_j^2 & i = j \end{cases}$$

Вектор свободных членов будет состоять из элементов:

$$c_j = s_M^2 - k_{jM}.$$

Такая система уравнений может быть решена одним из методов линейной алгебры.

Алгоритм объединения частных прогнозов можно представить в виде следующих последовательно выполняемых процедур:

1. Вычисляются дисперсии ошибок частных прогнозов и строится ковариационная матрица

$$s_j^2 = \frac{1}{n} \sum_{t=1}^n e_{jt}^2 \quad j = \overline{1, M}$$

где e_j - ошибки частных прогнозов

t - порядковый номер наблюдения ($t = \overline{1, n}$)

$$k_{ij} = \frac{1}{n} \sum_{t=1}^n e_{it} e_{jt}$$

2. Строятся матрица В и вектор С по формулам:

$$b_{ij} = b_{ji} = s_M^2 + k_{ij} - k_{iM} - k_{jM}$$

$$c_j = s_M^2 - k_{jM}$$

3. Из решения системы линейных уравнений определяется $(M-1)$ значение p_j , а последний весовой коэффициент p_M определяется по формуле:

$$p_M = 1 - \sum_{j=1}^{M-1} p_j$$

4. Проверка условия:

$$p_j > 0 \quad j = \overline{1, M}$$

Если условие не выполняется, прогнозы y_j исключаются и производится перерасчет весовых коэффициентов (к пункту 2).

5. Если все весовые коэффициенты положительны, вычисляется значение обобщающего прогноза:

$$y_0 = \sum_{j=1}^M p_j y_j$$

и коэффициент условной эффективности:

$$u_{y_1/y_0} = \frac{s_0^2}{s_1^2}$$

где s_0^2 - дисперсия ошибок комплексного прогноза;

s_1^2 - дисперсия ошибок наилучшего частного прогноза.

Так как в большинстве случаев точность прогнозов изменяется во времени, формулы оценки весовых коэффициентов модифицируются так, что более поздним ошибкам присваивается большее значение; таким образом происходит корректировка обобщающего прогноза путем изменения весовых

коэффициентов в сторону наилучшего частного прогноза:

$$y_{от} = \sum_{j=1}^M p_{jt} y_{jt}$$

где p_{jt} - весовой коэффициент частного прогноза в момент времени T ;

y_{jt} - частный прогноз в момент времени T ;

$y_{от}$ - обобщенный прогноз в момент времени T .

Для повышения стабильности динамики изменения весов в алгоритме их корректировки используется схема экспоненциального сглаживания.

Для проведения обобщения необходимо иметь не менее двух адекватных моделей. В целях повышения устойчивости результатов количество обобщаемых частных прогнозов не должно превышать пяти.

5. Корреляционный анализ

Основными **задачами** корреляционного анализа являются:

- измерение степени связи двух или более явлений;
- отбор факторов, оказывающих наиболее существенное влияние на результативный признак на основании измерения степени связности между явлениями;
- обнаружение ранее неизвестных причинных связей. Корреляция непосредственно не выявляет причинных связей между явлениями, но устанавливает численное значение этих связей и достоверность суждений об их наличии.

При проведении корреляционного анализа вся совокупность данных рассматривается как множество переменных (факторов), каждая из которых содержит n наблюдений; x_{ik} - наблюдение i переменной k ; $\bar{x}_k$ - среднее значение k -ой переменной; $i=1, \dots, n$.

Основными средствами анализа являются:

- парные коэффициенты корреляции;
- частные коэффициенты корреляции;
- множественные коэффициенты корреляции.

Парные коэффициенты корреляции опосредованно учитывают влияние других факторов. Для исключения этого влияния определяют частные коэффициенты корреляции.

Парные коэффициенты корреляции

Парный коэффициент корреляции между k -м и L -м факторами вычисляется по формуле:

$$r_{kL} = \frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)(x_{iL} - \bar{x}_L)}{\sqrt{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \sum_{i=1}^n (x_{iL} - \bar{x}_L)^2}}$$

Он служит показателем тесноты линейной статистической связи, но только в случае совместной нормальной распределенности случайных величин, выборками которых являются k -й и L -й факторы.

При этих же предпосылках для проверки гипотезы о равенстве нулю парного коэффициента корреляции используется t -статистика, распределенная по закону Стьюдента с $n-2$ степенями свободы. В программе для парного коэффициента корреляции сначала рассчитывается критическое значение t -статистики, а на его основе критическое значение коэффициента корреляции

$$r_{kp} = \frac{t_{kp}}{\sqrt{n-2+t^2}}$$

Если расчетное значение больше критического, то гипотеза о равенстве нулю данного коэффициента корреляции отвергается на соответствующем вероятностном уровне. Аналогичные выводы имеют место при проверке значимости частных коэффициентов корреляции.

Частные коэффициенты корреляции

Частный коэффициент корреляции первого порядка между k -м и L -м факторами характеризует тесноту их линейной связи при фиксированном значении j -го фактора. Он определяется как

$$r_{kL \cdot j} = \frac{r_{kL} - r_{kj}r_{Lj}}{\sqrt{(1-r_{kj}^2)(1-r_{Lj}^2)}}$$

Он распределен аналогично парному коэффициенту при тех же предпосылках, и для проверки его значимости используется t -статистика, в которой число степеней свободы равно $n-3$. В программе частный коэффициент корреляции рассчитывается в общем виде, т.е. при условии, что все остальные переменные - фиксированные:

$$r_{kL} (\text{частн.}) = \frac{-D_{kL}}{\sqrt{D_{kk}D_{LL}}}$$

Здесь D_{ij} - определитель матрицы, образованной из матрицы парных коэффициентов корреляции вычеркиванием i -й строки и j -го столбца. Для каждого частного коэффициента корреляции аналогично парному рассчитывается t -значение для проверки значимости коэффициента, а также доверительные интервалы. При этом дисперсия z -преобразованной величины будет равна $1/(n-L-3)$, где L - число фиксированных переменных (в программе $L=m-2$).

Множественные коэффициенты корреляции

Для определения тесноты связи между текущей k -й переменной и оставшимися (объясняющими) переменными, используется выборочный множественный коэффициент корреляции:

$$R_k = \sqrt{1 - \frac{D}{D_{kk}}},$$

где D - определитель матрицы парных коэффициентов корреляции.

Для проверки статистической значимости коэффициента множественной корреляции используется величина:

$$F = \frac{R^2 / L}{(1 - R^2) / (n - L - 2)},$$

имеющая F -распределение с L и $(n-L-2)$ степенями свободы соответственно.

Если рассчитанное F -значение больше значения F -распределения на соответствующем вероятностном уровне (0.9 и выше), то гипотеза о линейной связи между k -й переменной и остальными переменными не отвергается. В программе для каждого коэффициента множественной корреляции выводится F -значение и процентная точка F -распределения, которая ему соответствует.

6. Регрессионный анализ

В регрессионном анализе решаются следующие задачи:

- установление форм зависимости (положительная, отрицательная, линейная, нелинейная);

- определение функции регрессии. Важно не только указать общую тенденцию изменения зависимой переменной, но и выяснить, каково было бы действие на зависимую переменную главных факторов - причин, если бы прочие (второстепенные, побочные) факторы не изменялись бы (находились бы на одном и том же среднем уровне), и если были бы исключены случайные элементы;

- оценка неизвестных значений зависимой переменной.

Уравнение множественной линейной регрессии имеет вид:

$$y = a_0 + a_1 x_1 + \dots + a_m x_m$$

В каждом виде регрессионного анализа необходимо выбрать зависимую переменную Y (для которой строится уравнение регрессии) и одну или несколько независимых переменных x_i ($i=1, 2, \dots, m$).

Это уравнение позволяет установить статистическую взаимосвязь изучаемых показателей и, в случае ее устойчивости, давать аналитические и прогнозные оценки.

На базовом периоде времени строится уравнение регрессии зависимой переменной. Далее производится расчет прогнозных значений зависимой переменной по рассчитанному уравнению регрессии. При этом для всех регрессоров заранее должны быть получены их прогнозные оценки и дописаны в конец исходных данных. Для зависимой переменной в таблицу исходных данных на глубину периода прогнозирования необходимо дописать нулевые значения.

Линейная множественная регрессия

В линейном регрессионном анализе рассматривается зависимость случайной величины Y от ряда исходных факторов (регрессоров) $X_1, X_2, \dots, X_m$, которая в силу влияния неучтенных факторов будет стохастической. В матричной записи она имеет вид:

$$Y = Xb + e$$

где Y - вектор значений переменной,

X - матрица независимых переменных,

b - подлежащий определению вектор параметров,

e - вектор случайных отклонений.

В регрессионном анализе действуют следующие предположения:

$$M[e_j] = 0, \quad M[e_i \cdot e_j] = 0, \quad j \neq i,$$

$$M[e_i \cdot e_i] = s_e^2, \quad i = \overline{1, m}$$

матрица X детерминирована и ее столбцы линейно независимы.

МНК-оценки находятся из условия минимума функционала:

$$(Y - Xb)^T (Y - Xb)$$

Оценки параметров имеют вид:

$$\bar{b} = (X^T X)^{-1} X^T Y$$

и являются несмещенными и эффективными.

Пусть $\bar{y} = X \cdot \bar{b}$ - эмпирическая аппроксимирующая регрессия. Тогда элементы вектора

$$e = Y - \bar{Y}$$

называются остатками. Анализ остатков позволяет судить о качестве построенного уравнения регрессии.

Пошаговая регрессия

Пошаговая регрессия является одним из методов определения наилучшего подмножества регрессоров для объяснения Y . Реализуется пошаговая процедура с последовательным включением переменных в уравнение регрессии.

Пусть в уравнение регрессии включено L переменных, т.е. сделано L шагов алгоритма, и осуществляется $L+1$ шаг. Основным вопросом, который решается на каждой итерации - это вопрос о том, какую переменную включать в уравнение регрессии.

Для каждой переменной регрессии, за исключением тех переменных, которые уже включены в модель, рассчитывается величина C_j , равная отношению уменьшению суммы квадратов зависимой переменной. При включении переменной в уравнение регрессии она интерпретируется как доля оставшейся дисперсии независимой переменной, которую объясняет j -я переменная. Пусть k - номер переменной, имеющей максимальное значение j -го элемента. Тогда если $C_k < p$, где p - заранее определенная константа, то анализ переменных прекращается, и больше переменных не вводится в модель. В противном случае k -я переменная вводится в уравнение регрессии. Константа p является параметром метода и может быть изменена пользователем.

Гребневая регрессия

Гребневая регрессия основана на гребневых оценках, направленных на оценивание множественных линейных регрессий в условиях мультиколлинеарности, т.е. сильной корреляции независимых переменных. Как известно, следствием мультиколлинеарности является плохая обусловленность матрицы $X^T X$ и бесконечное возрастание по этой причине дисперсии оценок линейной регрессии.

Матрица $X'X$ регуляризуется путем добавления малого положительного числа к диагональным элементам. В программе реализован алгоритм построения однопараметрической гребневой оценки вида:

$$a(k) = (X'X + kD) X'Y, \quad k \geq 0,$$

где k - параметр регуляризации;

D - матрица регуляризации, в качестве которой может быть выбрана единичная матрица или диагональная матрица, составленная из диагональных элементов матрицы $X'X$.

Для автоматического расчета параметра k выбрана формула

$$k = ms/a'a,$$

где a - вектор оценок регрессии по МНК,

s - оценка остаточной дисперсии по МНК.

Тем не менее, пользователь имеет возможность произвольно изменять значения параметра регуляризации.

Парная регрессия

Парная регрессия устанавливает связь между откликом Y и функцией, зависящей от одной входной переменной X , т.е. регрессия имеет вид: $Y = f(X)$. Функции f , включенные в парную регрессию в настоящем пакете, удовлетворяют двум основным условиям: они распространены в практике экономических исследований, каждое из уравнений регрессии путем преобразований типа логарифмирования и возведения в степень сводится к линейной модели.

Для реализации функции парной регрессии необходимо выбрать переменную Y (зависимая переменная), переменную X (объясняющая переменная), а также сформировать список функций парной регрессии.

Основные функции парной регрессии и соответствующие преобразования приведены в таблице:

Модель	Преобразование	Матрицы	
		X	Y
$Y = a + bX$	нет	$\begin{pmatrix} 1 & x_1 \\ \dots & \dots \\ n & x_n \end{pmatrix}$	-
$Y = a + bX + cX^2$	нет	$\begin{pmatrix} 1 & x_1 & x_1^2 \\ \dots & \dots & \dots \\ n & x_n & x_n^2 \end{pmatrix}$	-
$Y = a + b/X$	нет	$\begin{pmatrix} 1 & 1/x_1 \\ 2 & 1/x_2 \\ \dots & \dots \\ n & 1/x_n \end{pmatrix}$	-
$Y = 1/(a + bX)$	возведение в степень (-1)	$\begin{pmatrix} 1 & x_1 \\ 2 & x_2 \\ \dots & \dots \\ n & x_n \end{pmatrix}$	$\begin{pmatrix} 1/y_1 \\ 1/y_2 \\ \dots \\ 1/y_n \end{pmatrix}$
$Y = 1/(a + b \cdot \exp(-X))$	возведение в степень (-1)	$\begin{pmatrix} 1 & \exp(-x_1) \\ 2 & \exp(-x_2) \\ \dots & \dots \\ n & \exp(-x_n) \end{pmatrix}$	$\begin{pmatrix} 1/y_1 \\ 1/y_2 \\ \dots \\ 1/y_n \end{pmatrix}$
$Y = a \cdot \exp(bX)$	логарифмирование	$\begin{pmatrix} 1 & x_1 \\ 2 & x_2 \\ \dots & \dots \\ n & x_n \end{pmatrix}$	$\begin{pmatrix} \ln(y_1) \\ \ln(y_2) \\ \dots \\ \ln(y_n) \end{pmatrix}$
$Y = a + b \cdot \lg(X)$	нет	$\begin{pmatrix} 1 & \ln(x_1) \\ 2 & \ln(x_2) \\ \dots & \dots \\ n & \ln(x_n) \end{pmatrix}$	-
$Y = a \cdot b^X \cdot c^{X^2}$	логарифмирование	$\begin{pmatrix} 1 & x_1 & x_1^2 \\ 2 & x_2 & x_2^2 \\ \dots & \dots & \dots \\ n & x_n & x_n^2 \end{pmatrix}$	$\begin{pmatrix} \ln(y_1) \\ \ln(y_2) \\ \dots \\ \ln(y_n) \end{pmatrix}$

Модель	Преобразование	Матрицы	
		X	Y
$Y = a b^x$	логарифмирование	$\begin{pmatrix} 1 & x_1 \\ 2 & x_2 \\ \dots & \dots \\ n & x_n \end{pmatrix}$	$\begin{pmatrix} \ln(y_1) \\ \ln(y_2) \\ \dots \\ \ln(y_n) \end{pmatrix}$
$Y = a + b/\ln(X)$	нет	$\begin{pmatrix} 1 & 1/\ln(x_1) \\ 2 & 1/\ln(x_2) \\ \dots & \dots \\ n & 1/\ln(x_n) \end{pmatrix}$	-
$Y = a X^b$	логарифмирование	$\begin{pmatrix} 1 & \ln(x_1) \\ 2 & \ln(x_2) \\ \dots & \dots \\ n & \ln(x_n) \end{pmatrix}$	$\begin{pmatrix} \ln(y_1) \\ \ln(y_2) \\ \dots \\ \ln(y_n) \end{pmatrix}$
$y = a + bX + c\sqrt{X}$	нет	$\begin{pmatrix} 1 & x_1 & \sqrt{x_1} \\ 2 & x_2 & \sqrt{x_2} \\ \dots & \dots & \dots \\ n & x_n & \sqrt{x_n} \end{pmatrix}$	-
$Y = X/(a+bX)$	нет	$\begin{pmatrix} 1 & x_1 \\ 2 & x_2 \\ \dots & \dots \\ n & x_n \end{pmatrix}$	$\begin{pmatrix} x_1/y_1 \\ x_2/y_2 \\ \dots \\ x_n/y_n \end{pmatrix}$
$Y = a \cdot \exp(b/X)$	логарифмирование	$\begin{pmatrix} 1 & 1/x_1 \\ 2 & 1/x_2 \\ \dots & \dots \\ n & 1/x_n \end{pmatrix}$	$\begin{pmatrix} \ln(y_1) \\ \ln(y_2) \\ \dots \\ \ln(y_n) \end{pmatrix}$
$y = a + bX^k$	нет	$\begin{pmatrix} 1 & x_1^k \\ 2 & x_2^k \\ \dots & \dots \\ n & x_n^k \end{pmatrix}$	-
$y = a + bX + cXX + \dots + dX^k$	нет	$\begin{pmatrix} 1 & x_1 & \dots & x_1^k \\ 2 & x_2 & \dots & x_2^k \\ \dots & \dots & \dots & \dots \\ n & x_n & \dots & x_n^k \end{pmatrix}$	-

Для каждой функции из списка будут найдены оценки регрессии по методу наименьших квадратов, а также рассчитан критерий. Критерием является величина:

$$\frac{1}{n-k} \sum_{i=1}^n (y_i - \bar{y})^2,$$

где k - число оцениваемых параметров функции.

Та функция, которой соответствует минимальное значение критерия, считается оптимальной. Для нее рассчитываются все параметры и результаты выводятся в протокол "Регрессионный анализ".

Экономическая интерпретация результатов

С помощью коэффициентов регрессии нельзя сопоставить факторы по степени их влияния на зави-

симую переменную из-за различий единиц измерения и степени колеблемости. Для устранения этого применяется:

- коэффициент эластичности;
- дельта-коэффициент;
- бета-коэффициент.

Как с помощью частных коэффициентов эластичности, так и с помощью бета-коэффициентов можно проранжировать факторы по степени их влияния на зависимую переменную, т.е. сопоставить их между собой по величине этого влияния. Вместе с тем нельзя непосредственно оценить долю влияния фактора в суммарном влиянии всех факторов. Для этой цели используют дельта-коэффициенты.

Коэффициент эластичности

Для экономической интерпретации нелинейных связей обычно пользуются коэффициентом эластичности, который характеризует относительное изменение зависимой переменной при изменении объясняющей переменной на 1%. Если уравнение регрессии имеет вид $y = f(x)$, то коэффициент эластичности рассчитывается как

$$\Theta = \frac{df}{dx} \cdot \frac{\bar{x}}{\bar{y}}$$

где $\bar{x}$ - среднее значение переменной x ,
 $\bar{y}$ - среднее значение переменной y .

Производная берется в точке $\bar{x}$.

Аналитические выражения для расчета коэффициента эластичности с точностью до знака приведены в таблице:

N	Функция	Формула коэффициента эластичности
1	$y = a + bx$	$\Theta = b \frac{x}{y}$
2	$y = a + bx + c x^2$	$\Theta = (b + 2cx) \frac{x}{y}$
3	$y = a + b/x$	$\Theta = b / (ax + b)$
4	$y = 1/(a + bx)$	$\Theta = bx / (a + bx)$
5	$y = 1/(a + b e^{-x})$	$\Theta = \frac{b x e^{-x}}{(a + b e^{-x})}$
6	$y = a e^{bx}$	$\Theta = bx$
7	$y = a + b \ln(x)$	$\Theta = b/y$
8	$y = a b^x c x^2$	$\Theta = \frac{a(\ln(b)e^x c x^2 + 2b^x x e^x \ln(c))}{y}$
9	$y = ab^x$	$\Theta = a \cdot \ln(b) e^x \frac{x}{y}$
10	$y = a + b/\ln(x)$	$\Theta = b / (\ln^2(x) \cdot y)$
11	$y = ax^b$	$\Theta = b$
12	$y = a + bx + c \sqrt{x}$	$\Theta = (b - c/\sqrt{x}) \frac{x}{y}$
13	$y = x/(a + bx)$	$\Theta = a/(a + bx)$
14	$y = a e^{(b/x)}$	$\Theta = b/x$
15	$y = a + bx^k$	$\Theta = b k x^k / (a + b x^k)$
16	$y = a_0 + a_1 x^1 + \dots + a_k x^k$	$\Theta = (\sum_{i=1}^k a_i i x^i) / y$

Дельта-коэффициент

Доля вклада каждого фактора в суммарное влияние всех факторов равна:

$$D_j = \frac{r_j \cdot b_j}{R^2}, \sum_k D_k = 1$$

$$R^2 = r_1 b_1 + r_2 b_2 + \dots + r_k b_k$$

R^2 - коэффициент множественной детерминации,

r_i - коэффициент парной корреляции между i -м фактором и зависимой переменной,

b_j^{CT} - b -коэффициент.

При корректно проводимом анализе величины дельта-коэффициентов положительны, т.е. все коэффициенты регрессии имеют тот же знак, что и соответствующие парные коэффициенты корреляции.

Тем не менее, в случаях сильной коррелированности объясняющих переменных, некоторые дельта-коэффициенты могут быть отрицательными вследствие того, что соответствующий коэффициент регрессии имеет знак, противоположный парному коэффициенту корреляции.

Бета-коэффициент

Для устранения различий в измерении и степени колеблемости факторов используется β -коэффициент, или коэффициент регрессии в стандартизованном виде:

$$b_j^{CT} = b_j \frac{S_j}{S_y}$$

где b_j - коэффициент регрессии при j -й переменной,

S_j - оценка среднеквадратического отклонения j -й переменной,

S_y - оценка среднеквадратического отклонения независимой переменной.

Он показывает, на какую часть величины среднего квадратического отклонения меняется среднее значение зависимой переменной с изменением соответствующей независимой переменной на одно среднеквадратическое отклонение при фиксированном на постоянном уровне значении остальных независимых переменных.

7. Факторный и компонентный анализ

Компонентный анализ является методом определения структурной зависимости между случайными переменными. В результате его использования получается сжатое описание малого объема, несущее почти всю информацию, содержащуюся в исходных данных. Главные компоненты $Y_1, Y_2, \dots, Y_m$ получаются из исходных переменных $X_1, X_2, \dots, X_m$ путем целенаправленного вращения, т.е. как линейные комбинации исходных переменных. Вращение производится таким образом, чтобы главные компоненты были ортогональны и имели максимальную дисперсию среди возможных линейных комбинаций исходных переменных X . При этом переменные $Y_1, Y_2, \dots, Y_m$ не коррелированы между собой и упорядочены по убыванию дисперсии (первая компонента имеет наибольшую дисперсию). Кроме того, общая дисперсия после преобразования

остаётся без изменений. Итак, i -я главная компонента Y_i :

$$Y_i = \sum_{j=1}^m a_{ij} x_j, \quad \sum_{j=1}^m a_{ij}^2 = 1, \quad i = \overline{1, m}$$

Пусть R - корреляционная матрица переменных X . Тогда a_{1j} - первый собственный вектор матрицы R , и т.д. Кроме того, дисперсия первой главной компоненты равна первому собственному числу матрицы R , дисперсия второй главной компоненты равна второму собственному числу матрицы R , и т.д.

Факторный анализ является более общим методом преобразования исходных переменных по сравнению с компонентным анализом. Модель факторного анализа имеет вид:

$$X_i = \sum_{j=1}^p l_{ij} F_j + e_i,$$

где l_{ij} - постоянные величины, называемые факторными нагрузками,

F_j - общие факторы, используемые для представления всех p исходных переменных,

e_i - специфические факторы, уникальные для каждой переменной, $p \leq m$.

Задачами факторного анализа являются: определение числа общих факторов, определение оценок l_{ij} , определение общих и специфических факторов.

Для получения оценок общностей и факторных нагрузок используется эмпирический итеративный алгоритм, который сходится к истинным оценкам параметров. Сущность **алгоритма** сводится к следующему.

Первоначальные оценки факторных нагрузок определяются с помощью метода главных факторов. На основании корреляционной матрицы R формально определяются оценки главных компонент:

$$Y_i = \sum_{j=1}^m a_{ij} x_j, \quad \sum_{j=1}^m a_{ij}^2 = 1, \quad i = \overline{1, m}$$

Оценки общих факторов ищутся в виде:

$$F_i = \frac{Y_i}{\sqrt{\lambda_i}}, \quad i = \overline{1, p}$$

где λ_i - соответствующее собственное значение матрицы R .

Оценками факторных нагрузок служат величины

$$L_{ij} = a_{ij} \sqrt{\lambda_j}, \quad i = \overline{1, m}, \quad j = \overline{1, p}$$

где a_{ij} - оценки a_{ij} ,

L_{ij} - оценки l_{ij} .

Оценки общностей получают как

$$h_i^2 = \sum_{j=1}^p L_{ji}^2$$

На следующей итерации модифицируется матрица R - вместо элементов главной диагонали подставляются оценки общностей, полученные на предыдущей итерации; на основании модифицированной матрицы R с помощью вычислительной схемы компонентного анализа повторяется расчет главных компонент (которые не являются таковыми с точки зрения компонентного анализа), ищутся оценки главных факторов, факторных нагрузок, общностей, специфичностей. Факторный анализ можно считать законченным, когда на двух соседних итерациях оценки общностей меняются слабо.

Примечание. Преобразования матрицы R могут нарушать положительную определенность матрицы R и, как следствие, некоторые собственные значения R могут быть отрицательными.

Для лучшей интерпретации полученных общих факторов к ним применяется процедура варимаксного вращения.

Если факторный анализ ведется в терминах главных компонент, то значения факторов могут быть вычислены непосредственно. Главные компоненты (без вращения) могут быть представлены в виде:

$$F_p = \sum_{j=1}^n \frac{a_{jp}}{l_p} x_j, \quad p = \overline{1, m}$$

где a_{jp} - коэффициенты при общих факторах,

l_p - собственные значения,

x_j - исходные данные (вектор-столбцы),

F_p - главные компоненты (вектор-столбцы).

В случае вращения главных компонент соотношения, связывающие исходные переменные и значения факторов, несколько усложняются. Ниже в матричном виде приведено соотношение, оптимальное по скорости вычисления, а также независимое от метода вращения факторов:

$$F = B^T A L_m^{-2} A^T x,$$

B^T - повернутая матрица A ,

A - матрица коэффициентов при общих факторах,

L_m - диагональная матрица m собственных членов,

x - матрица исходных данных,

F - матрица m повернутых факторов.

При определении числа **общих факторов** руководствуются следующими критериями: число существенных факторов можно оценить из содержательных соображений, в качестве p берется число собственных значений, больших либо равных единице (по умолчанию), выбирается число факторов, объясняющих определенную часть общей дисперсии или суммарной мощности.

8. Кластерный анализ

Классификация объектов по осмысленным группам, называемая кластеризацией, является важной процедурой в различных областях научных исследований. Кластерный анализ (**КА**) - это много-

мерная статистическая процедура, упорядочивающая исходные данные (объекты) в сравнительно однородные группы. Общим для всех исследований, использующих **КА**, являются пять основных шагов:

- отбор выборки для кластеризации;
- определение множества признаков, по которым будут оцениваться объекты в выборке;
- вычисление значений той или иной меры сходства между объектами;
- применение метода **КА** для создания групп исходных данных;
- проверка достоверности результатов кластерного решения.

Каждый из перечисленных шагов играет существенную роль при использовании кластерного анализа в прикладном анализе данных. При этом 1, 2 и 5 шаги целиком зависят от решаемой задачи и должны определяться пользователем. Шаги 3 и 4 выполняются программой кластерного анализа.

Сделаем несколько замечаний общего характера.

Многие методы **КА** - довольно простые процедуры, которые не имеют, как правило, строгого статистического обоснования. Другими словами, большинство методов **КА** являются эвристическими. Это позволяет повысить понимание метода и, таким образом, свести к минимуму вероятность допустить ошибку при трактовке результатов **КА**.

Разные кластерные методы могут порождать различные решения для одних и тех же данных. Это обычное явление в большинстве прикладных исследований. По-видимому, окончательным критерием является удовлетворенность исследователя результатами **КА**.

Разработанные кластерные методы образуют семь основных семейств:

- иерархические агломеративные методы;
- иерархические дивизимные методы;
- итеративные методы группировки;
- методы поиска модальных значений плотности;
- факторные методы;
- методы сгущений;
- методы, использующие теорию графов.

По данным некоторых исследований, приблизительно **2/3** приложений **КА** используют иерархические агломеративные методы. Рассмотрим его сущность на примере наиболее простого метода одиночной связи.

Процесс кластеризации начинается с поиска двух самых близких объектов в матрице расстояний. На последующих шагах к этой группе присоединяется объект, наиболее близкий к одному из уже находящихся в группе. По окончании кластеризации все объекты объединены в один кластер. Отметим несколько важных особенностей иерархических агломеративных методов. Во-первых, все эти методы просматривают матрицу расстояний размерностью **N*N** (где **N** - число объектов) и последовательно объединяют наиболее схожие объекты. Именно поэтому они называются агломеративными (объединяющими). Во-вторых, последовательность объединения

кластеров можно представить визуально в виде древовидной диаграммы, часто называемой дендрограммой. Наконец, для понимания этого класса методов не нужны обширные знания матричной алгебры или математической статистики. Вместо этого дается правило объединения объектов в кластеры.

Для "ОЛИМП:СтатЭксперт" разработана программа кластерного анализа, основанная на иерархической агломеративной процедуре и позволяющая пользователю управлять процессом кластеризации. Коротко поясним сущность предлагаемого метода.

Сначала ищутся два наиболее близких объекта (предположим, **A** и **B**). Предположим, что расстояние между объектами **A** и **B** равно **R**. В один кластер объединяются объекты, расстояние между которыми меньше, чем **(10-C)*R**, где **C** - четкость классификации, параметр управления процессом, принимающий значения от 1 до 10, который может меняться пользователем. При **C=10** на каждом шаге объединяются только два самых близких элемента, т.е. имеет место иерархическая агломеративная процедура в чистом виде. Однако, как показывает практика использования **КА**, пользователю важнее выделить в пространстве группы объектов с разной плотностью. В этом случае величину **C** необходимо уменьшать. Минимальное расстояние **R** пересчитывается на каждом шаге кластерного анализа.

Объединение. На каждом шаге кластерного анализа происходит объединение объектов, т.е. из нескольких объектов образуется один кластер. Процедура кластеризации заканчивается тогда, когда все первичные объекты исчерпаны. Допустим, на **k**-м шаге объединяются **n** объектов. Из этих объектов образуется один кластер как центр тяжести этих объектов (среднее арифметическое по каждой координате).

Размерность задачи уменьшается на величину **n-1** (**n** объектов удаляются, один добавляется). Далее производится пересчет матрицы расстояний.

В программе реализован кластерный анализ наблюдений, т.е. в результате вычислительной процедуры каждое наблюдение относится к той или иной группе. Кластеризация проводится на основе одной из двух метрик:

$$\text{Евклидово расстояние: } R = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

$$\text{Корреляционное расстояние: } R = |1 - r_{xy}|,$$

где $x = \{x_1, x_2, \dots, x_k\}$ и $y = \{y_1, y_2, \dots, y_k\}$ - две точки;

r_{xy} - парный коэффициент корреляции между x и y .

Графическая интерпретация

Для графической интерпретации результатов кластерного анализа приводится график расположения исходных объектов в пространстве первых двух главных компонент. При этом объекты, попавшие в один кластер, отображаются одним цветом.

Примечание. Иногда объекты из разных кластеров расположены столь близко, что может создаться иллюзия о неправильной классификации. Это

связано с тем, что классификация проводится по большому числу переменных, а график строится по двум координатам (хотя и отражающим основные особенности данных), поэтому некоторые расхождения между результатом классификации и графическим отображением неизбежны.

9. Частотный анализ

Вместе с долговременными изменениями во временных рядах часто появляются более или менее регулярные колебания. Эти изменения наблюдаемых значений могут быть строго периодическими или близкими к таковым и оцениваться в частотном аспекте. Для выявления наличия и устойчивости периода колебаний обычно используется следующий аппарат частотного анализа:

- гармонический анализ (
- спектральный анализ (
- частотная фильтрация (
- кросс-спектральный анализ.)

Этот аппарат позволяет с разных позиций анализировать исследуемый показатель, однако он эффективен лишь при наличии достаточно большого объема данных (по разным литературным источникам желательно иметь 200-300 наблюдений, но не менее 50 наблюдений), из которых предварительно исключена тенденция (за исключением методов частотной фильтрации).

Дадим определения основных терминов частотного анализа.

Интервал времени, необходимый для того, чтобы временной ряд начал повторяться, называется **периодом**. Он измеряется числом единиц времени за цикл и не является единственным. Если между пиками (высшими точками) или впадинами (низшими точками) проходит 10 месяцев, то период этого цикла равен 10 месяцам.

Величина, обратная периоду, называется **частотой ряда**. Она указывает число повторений цикла в единицу времени и поэтому измеряется числом циклов в единицу времени. Если между пиками (высшими точками) или впадинами (низшими точками) проходит 10 месяцев, то период этого цикла равен 10 месяцам, а частота 1/10.

Амплитуда периодического ряда - это отклонение от среднего значения до пика или впадины.

Фаза - представляет собой расстояние между началом отсчета времени и ближайшим пиковым значением.

Гармонический анализ

Временной ряд наблюдений может быть представлен с помощью линейных комбинаций функций времени - синусов и косинусов, на основании конечного преобразования Фурье. Гармонический анализ позволяет выявить наиболее существенные гармоники. Пусть $Y(t)$ - временной ряд $t=1, 2, \dots, T$. Тогда имеет место следующее представление ряда:

$$y_t = \bar{y} + \sum (a_j \cos(\frac{2\pi j}{T} t) + b_j \sin(\frac{2\pi j}{T} t)) + [a_{T/2}(-1)^t]$$

$$S = \lfloor \frac{T-1}{2} \rfloor,$$

где $\bar{y}$ - оценка математического ожидания ряда $Y(t)$. Последнее слагаемое добавляется в том случае, когда T - четное число. Коэффициенты вычисляются по соотношениям:

$$a_j = \frac{2}{T} \sum_{t=1}^S (y_t \cdot \cos(\frac{2\pi j}{T} \cdot t))$$

$$b_j = \frac{2}{T} \sum_{t=1}^S (y_t \cdot \sin(\frac{2\pi j}{T} \cdot t))$$

$$a_{T/2} = \frac{1}{T} \sum_{t=1}^S (y_t \cdot (-1)^t)$$

Таким образом, временной ряд представлен в виде суммы гармоник. Мощность каждой гармоники равна

$$R_k^2 = a_k^2 + b_k^2$$

k -я гармоника считается статистически значимой, если она вносит существенный вклад в дисперсию временного ряда, то есть если отвергается статистическая гипотеза о том, что $R_k = 0$. Для проверки гипотезы вычисляется критерий:

$$\frac{TR_k^2}{4s_k^2},$$

где s_k^2 - оценка дисперсии отклонения вычисляемых значений от фактических:

$$s_k^2 = \frac{1}{T-3} \left(\sum_{t=1}^T y_t^2 - T\bar{y}^2 - \frac{T}{2} R_k^2 \right)$$

Вычисляемая величина имеет **F**-распределение с $n_1 = 2$ и $n_2 = T-3$ степенями свободы. Гипотеза отвергается, то есть гармоника считается значимой, если вычисленная величина больше, чем 95% точка **F**-распределения с соответствующими степенями свободы.

Спектральный анализ

Рассмотрим алгоритм спектрального анализа. Пусть $x(t)$, $t = 0, 1, \dots, T$ - временной ряд. Тогда его **периодограмма** рассчитывается как

$$I_{xx}^{(T)}(f) = \frac{1}{T} \left| \sum_{t=0}^{T-1} x(t) \exp\{-i2\pi f t\} \right|^2$$

Предполагается, что исходные данные квантованы с интервалом 1 и, следовательно, частота Найквиста для них равна 0,5. Поэтому периодограмма и спектральная плотность рассчитывается на интервале от 0 до 0,5. в **точках** $f(j) = j/2M$, $j = 0, 1, \dots, M$.

Оценка **спектральной плотности**, реализованная в программе, основана на оценке Бартлетта, которая является усреднением периодограмм, вычисленных по непересекающимся отрезкам временных рядов. В программе спектральная плотность при $T = L \cdot V$ оценивается аналогично, только временные интервалы могут пересекаться. Пусть

$$I_{xx}^{(v)}(f, l) = \frac{1}{V} \left| \sum_{v=0}^{V-1} x(v + lV) \exp\{-i2\pi f(v + lV)\} \right|^2$$

, $l = 0, 1, \dots, L-1$,

где V - ширина временного интервала;

l - номер интервала;

S - смещение текущего временного интервала относительно предыдущего.

Тогда оценка спектральной плотности получается как

$$f_{xx}(f) = \frac{1}{L} \sum_{l=0}^{L-1} I_{xx}^{(v)}(f, l)$$

Спектральные оценки сглаживаются при помощи "окон", которые применяются с целью уменьшения дисперсии выборочной спектральной плотности. На практике из большого числа известных окон, используются следующие три:

- прямоугольное;
- окно Тьюки - Хеннинга;
- окно Парзена.

Окно	Формула
Прямоугольное	$l_k = 1, 0 \leq k \leq m$
Тьюки-Хеннинга	$l_k = \frac{1}{2} \left[1 + \cos \frac{\pi k}{m} \right]$
Парзена	$l_k = \begin{cases} 1 - \frac{6k^2}{m^2} \left(1 - \frac{k}{m} \right), & 0 \leq k \leq \frac{m}{2} \\ 2 \left(1 - \frac{k}{m} \right)^3, & \frac{m}{2} \leq k \leq m \end{cases}$

Параметры, необходимые для расчета спектра мощности, рассчитываются по следующему алгоритму:

$V = n/3$ (n - число наблюдений)

при $V < 10$ принимается $V = 10$;

при $V > 50$ принимается $V = 50$ $S = V/2$

Кросс-спектральный анализ

Кросс-спектральный анализ оценивает связь между частотными составляющими двух временных рядов при помощи параметров когерентности, фазового сдвига и коэффициента усиления. Рассчитываются оценки взаимных **ковариационных функций**:

$$C_{xy}(k) = \frac{1}{n-k} \left\{ \sum_{t=1}^{n-k} x_t y_{t+k} - \frac{1}{n-k} \sum_{t=k+1}^n y_t \sum_{t=1}^{n-k} x_t \right\}$$

$$C_{yx}(k) = \frac{1}{n-k} \left\{ \sum_{t=1}^{n-k} y_t x_{t+k} - \frac{1}{n-k} \sum_{t=k+1}^n x_t \sum_{t=1}^{n-k} y_t \right\}$$

Оценка **ко-спектра** (действительной части спектра):

$$\begin{aligned} \tilde{C}(w_j) &= \frac{1}{4\pi} (C_{xy}(0) + C_{yx}(0)) + \\ &+ \frac{1}{2\pi} \sum_{k=1}^m l_k (C_{xy}(k) + C_{yx}(k)) \cos w_j k \end{aligned}$$

Оценка квадратурного спектра (мнимой части):

$$\tilde{Q}(w_j) = \frac{1}{2\pi} \sum_{k=1}^m l_k (C_{xy}(k) - C_{yx}(k)) \sin w_j k$$

$$w_j = \frac{\pi j}{m}, j = 0, 1, \dots, m$$

Оценка когерентности:

$$\tilde{C}(w_j) = \frac{\tilde{C}(w_j)^2 + \tilde{Q}(w_j)^2}{\tilde{f}_x(w_j) \tilde{f}_y(w_j)}$$

Оценка фазового сдвига:

$$\tilde{\varphi}(w_j) = \arctg \left[\frac{\tilde{Q}(w_j)}{\tilde{C}(w_j)} \right]$$

Оценка коэффициента усиления:

$$\tilde{R}_{xy}^2(w_j) = \frac{\tilde{f}_x(w_j) \tilde{C}(w_j)}{\tilde{f}_y(w_j)}$$

Оценка спектра $\tilde{f}_x$ для ряда x в настоящем разделе имеет следующий вид:

$$\tilde{f}_x(w_j) = \frac{1}{2\pi} \left\{ l_0 C_x(0) + 2 \sum_{k=1}^m l_k C_x(k) \cos w_j k \right\}$$

$$C_x(k) = \frac{1}{n-k} \left\{ \sum_{t=1}^{n-k} x_t x_{t+k} - \frac{1}{n-k} \sum_{t=k+1}^n x_t \sum_{t=1}^{n-k} x_t \right\}$$

Аналогично получается оценка спектра $\tilde{f}_y$ для ряда y .

Интерпретация результатов кросс-спектрального анализа - довольно тонкий процесс. Отметим, что когерентность аналогична квадрату коэффициента корреляции на соответствующей частоте и интерпретируется таким же образом. Коэффициент усиления есть, по сути, коэффициент линейной регрессии процесса по процессу на соответствующей частоте. Фазовый сдвиг характеризует временное смещение между составляющими двух процессов.

Частотная фильтрация

Фильтрация осуществляется при помощи высокочастотного и низкочастотного фильтра, для каждого из которых рассчитывается соответствующая силовая и фазовая характеристики. Низкочастотный фильтр предназначен для устранения тренда (низкочастотной составляющей временного ряда наблюдений). Высокочастотный фильтр, наоборот, предназначен для выделения тренда из исходных данных.

Выход низкочастотного фильтра e_t получается из выражения:

$$x_t - 2x_{t-1} + x_{t-2} = a_0 e_t + a_1 e_{t-1} + a_2 e_{t-2},$$

где $a_0 = 1 + k^2 + k\sqrt{2}$,

$$a_1 = 2(k^2 - 1),$$

$$a_2 = 1 + k^2 - k\sqrt{2},$$

$$k = \operatorname{tg}\left(\frac{w}{2}\right),$$

W - частота отсечки,

e_t является оценкой высокочастотной составляющей. При оценке его теряются два первых

наблюдения. Оценкой тренда в этом случае является ряд $y_t = x_t - e_t$.

Выход высокочастотного фильтра e_t получается из выражения:

$$x_t - 2x_{t-1} + x_{t-2} = a_0e_t + a_1e_{t-1} + a_2e_{t-2},$$

где $a_0 = 1 + k^2 + k\sqrt{2}$,

$$a_1 = 2(k^2 + 1),$$

$$a_2 = 1 + k^2 - k\sqrt{2},$$

$$k = \operatorname{tg}\left(\frac{W}{2}\right),$$

W - частота отсечки,

e_t является оценкой низкочастотной составляющей. При оценке e_t теряются два первых наблюдения. Ряд может быть использован для прогнозирования.

10. Работа с математическим аппаратом на компьютере

Общие сведения о программном обеспечении

Программа "ОЛИМП:СтатЭксперт" предназначена для статистического анализа и прогнозирования развития финансово - экономических и ряда других процессов, представленных временными рядами наблюдений и пространственными данными.

Реализованный в ней математический аппарат позволяет решать широкий спектр практических задач: оценивать текущее состояние процесса, исследовать и прогнозировать динамику развития с учетом тенденции, а также сезонных и циклических колебаний, определять степень взаимосвязи исследуемых показателей и отражать их в форме математических моделей, проводить классификацию объектов и др.

Программа поставляется в двух версиях, отличающихся широтой реализованного в них математического аппарата: базовой и профессиональной. Базовая включает средства описательной статистики количественных данных, методы анализа и прогнозирования одномерных временных рядов, корреляционный и регрессионный анализ. Профессиональная версия включает базовую и имеет более широкие возможности обработки многомерных данных (факторный анализ, компонентный, гармонический, частотный, спектральный, кросс- спектральный и кластерный анализ).

Для независимого использования вычислительных библиотек "ОЛИМП:СтатЭксперт" отдельно поставляется руководство программиста.

Комплектность поставки

В комплект поставки входит одна (3.5-дюймовая) инсталляционная дискета, настоящее руководство, лицензионное соглашение и регистрационная карточка.

Варианты установки

Для работы "ОЛИМП:СтатЭксперт" необходимо:

Операционная система MS-DOS версии 4.0 или более поздней;

Одна из следующих операционных оболочек Windows:

русское издание Windows 3.1 или более поздней версии,

русское издание Windows для рабочих групп 3.11,

русифицированный Windows 3.1.

Windows'95

персональный компьютер с процессором 80386 или выше (рекомендуется Pentium);

8 Мбайт оперативной памяти;

3 Мбайта свободной памяти на жестком диске;

Дисковод для работы с 3.5 - дюймовыми дисками высокой плотности;

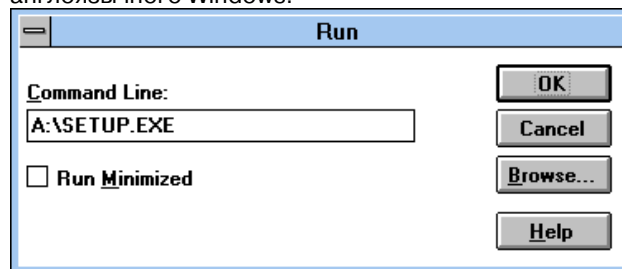
Графический адаптер VGA, совместимый с Microsoft Windows 3.1 (рекомендуется адаптер с разрешением не ниже 800 X 600);

Устройство "мышь";

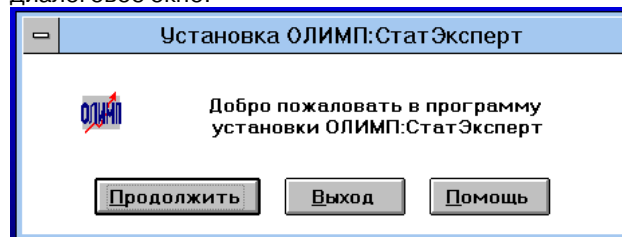
Установленная на компьютер программа Microsoft Excel 5.0 (Excel 7.0 для Windows'95).

При наличии указанных выше ресурсов можно начать инсталляцию программы. Для этого запустите программную оболочку Windows и вставьте установочную дискету в дисковод. Затем в Диспетчере программ в пункте "Файл" выполните функцию "Выполнить" как показано ниже.

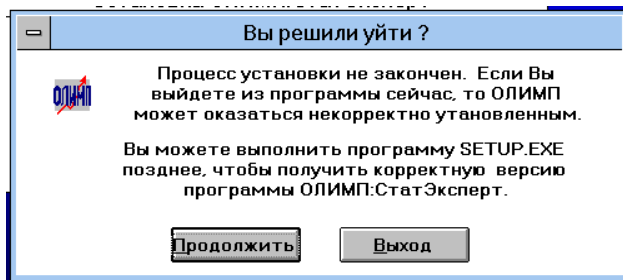
Примечание: все диалоги в данном разделе документации получены в русифицированной версии англоязычного Windows.



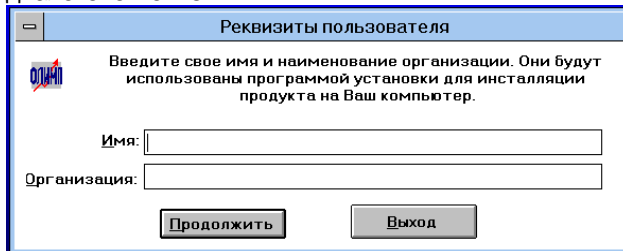
После того как Вы нажмете клавишу "OK", начинается работу программа установки. Она анализирует состояние системы, после чего отображается диалоговое окно.



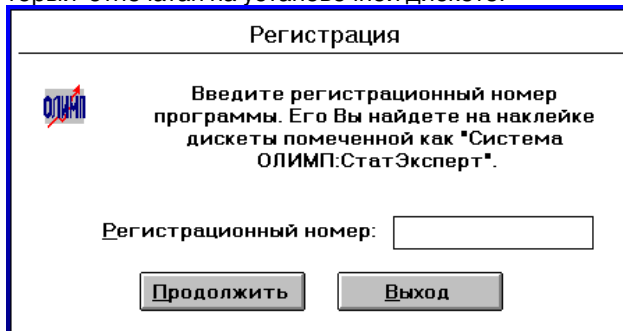
В том случае, когда Вы выбираете клавишу "Выход" (здесь или далее в процессе установки), отображается диалоговое окно, приведенное ниже.



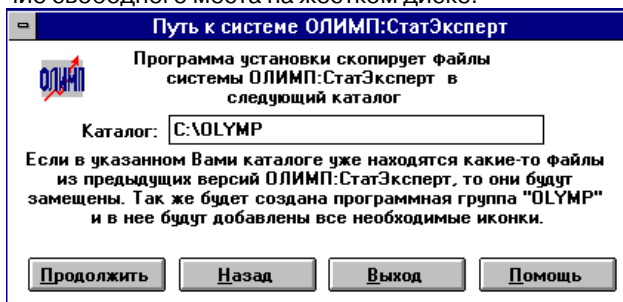
Если Вы выбираете "Продолжить", Вам необходимо заполнить свои координаты в следующем диалоговом окне.



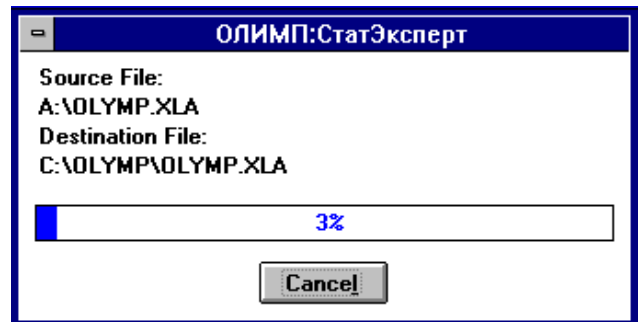
Далее Вы должны **правильно** (это очень важно) ввести регистрационный номер программы, который отпечатан на установочной дискете.



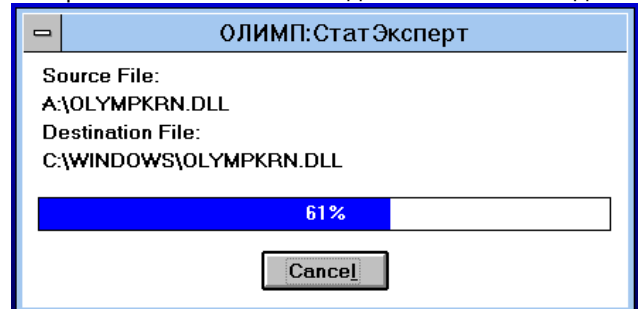
В следующем диалоге необходимо задать устройство и путь к директории, в которой будет размещена система. Следует учесть, что часть системных файлов будет размещена в системной директории Windows. Кроме того, текущая версия программы установки "ОЛИМП:СтатЭксперт" не проверяет наличие свободного места на жестком диске.



Процесс копирования и декомпрессирования файлов системы иллюстрируется в диалоговом окне следующего вида.



По окончании процесса установки система отображает заключительное диалоговое окно вида



после чего создается программная группа "ОЛИМП" с несколькими иконками.



По окончании работы программы инсталляции на Вашем компьютере появятся следующие модули, необходимые для нормальной работы "ОЛИМП:СтатЭксперт":

- OLYMP.XLA;**
- OLYMPSYS.XLS;**
- * OLYMP.HLP;**
- * OLYMPSTS.DLL;**
- * OLYMPSET.DLL (OLYMP32.DLL для Windows95);**
- * OLYMPRES.DLL (OLYMP32.DLL для Windows95);**
- * OLYMPKRN.DLL (OLYMPK32.DLL для Windows95);**

а также дополнительные текстовые и демонстрационные файлы:

- Read.Me;**
- ReadMe.wri;**
- Demo.XLS.**

Файлы, помеченные символом "*", размещаются в системной директории Windows, остальные файлы - в директории, указанной пользователем при установке программы.

Для того чтобы перенести систему на другой компьютер, необходимо сначала удалить ее с компьютера, на который она была установлена, т.е. выполнить деинсталляцию программы. Для этого следует вставить дистрибутивную дискету в устройство "а:"

и воспользоваться пиктограммой "Удаление Стат-Эксперта". После этого дистрибутивную дискету можно использовать для повторной инсталляции программы.

Технические ограничения

При практической работе с программой Вы вряд ли ощутите наличие ограничений, однако они, все же, существуют.

Для обработки пользователь должен выделить в электронной таблице блок данных. Величина этого блока не ограничена. Однако существует ограничение на объем непосредственно обрабатываемых данных вычислительными методами. В частности, поскольку вычисления, проводимые модулями Си++, осуществляются с числами с двойной точностью, а память под объекты выделяется сегментами, не превышающими 64K байт, объем обрабатываемых данных не должен превышать 8190 чисел. При проведении корреляционного анализа количество обрабатываемых переменных не должно превышать 120. При работе в Windows-95 эти ограничения отсутствуют!

Все результаты вычислений выдаются в таблицу "Report". Поэтому размер выходного протокола **каждой порции** результатов ограничен размерами электронной таблицы и не должен превышать 16000 строк.

Для нормальной работы программы в Excel необходимо проверить правильность установки некоторых параметров, связанных с работой модулей Visual Basic. Для этого войдите в меню "Сервис" ("Tools"), выберите пункт "Параметры" ("Parameters") и убедитесь в отсутствии символа выбора ("крестик") рядом с параметрами:

- "Останавливаться на всех ошибках" ("Break on all Errors");
- "Требовать описания всех переменных" ("Require variable Declaration").

Если символ выбора ("крестик") стоит слева от этих параметров - уберите его, сохраните новые установки, перезапустите Excel и запустите программу "ОЛИМП:СтатЭксперт".

Проверьте также системные установки. В приложении "Панель управления" в пункте "Стандарты" параметр "Разделитель списка" должен соответствовать символу "запятая", а параметр "Формат числа" должен использовать разделитель "точка".

В случае использования англоязычных версий Windows и MS Office, русифицированных программами Cyrwin, Parawin или другими, для корректного отображения текстовой информации, необходимо в секцию [Font Substitution] файла WIN.INI, поместить строку: ArialCyr=[имя русифицированного шрифта без засечек].

При этом шрифт Arial Cyr должен отсутствовать в списке установленных в Windows шрифтов. Имя шрифта различно для различных русификаторов. Например, для Cyrwin 4.x имя шрифта может быть "NT Harmonica". Подробнее смотрите документацию по конкретному русификатору. Программа "ОЛИМП:СтатЭксперт" не может нормально работать при двух одновременно работающих экземплярах программы Excel.

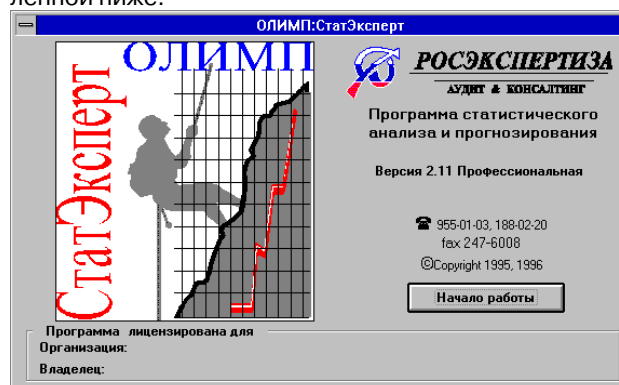
Перспективы развития программного обеспечения

Развитие данной версии программы, прежде всего, будет происходить за счет расширения круга задействованных моделей и методов, обеспечения более широкой возможности управления вычислительным процессом, путем изменения параметров моделей. В частности, предусматривается разработка модулей структурного анализа и обработки нечисловой информации, включающей в себя расчет ранговых корреляций и анализ таблиц сопряженности.

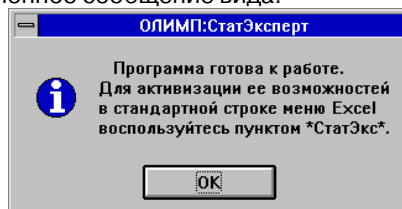
Запуск программы

После инсталляции программы создается программная группа "ОЛИМП" ("OLYMP"). Для запуска программы необходимо нажать на кнопку "СтатЭксперт". Кроме этого, программа может быть стартована непосредственно из Excel. В последнем случае в меню "Файл" - "Открыть" ("File" - "Open") необходимо найти файл OLYMP.XLA. После его запуска автоматически загружается файл OLYMPSYS.XLS, который открывает одноименную книгу. В ней содержится пустая таблица с именем "Data", предназначенная для размещения обрабатываемых данных, а для временного размещения результатов расчетов создается таблица "Report".

Признаком начала работы программы является модификация строки основного меню Excel и появления заставки с вашими реквизитами, представленной ниже.



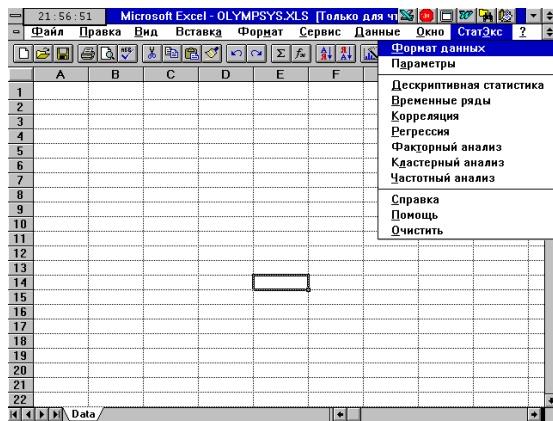
Для начала работы нажмите клавишу "Enter" или левую кнопку "мышки", установив ее указатель на кнопку "Начало работы". После этого появится информационное сообщение вида:



После нажатия кнопки "ОК" можно начинать работу с программой.

Меню программы

После запуска программы "ОЛИМП:СтатЭксперт" в строке основного меню Excel появляется пункт "СтатЭкс", при активизации которого ("мышкой" или горячей клавишей "Ctrl+ Э") раскрывается меню из 12 пунктов:



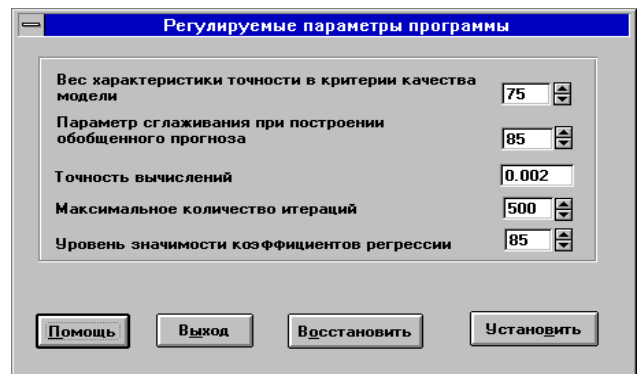
Это меню программы "ОЛИМП:СтатЭксперт" доступно из книги пользователя и из рабочей книги OLYMPSYS.XLS.

Семь пунктов меню "СтатЭкс" (Дескриптивная статистика, Временные ряды, Корреляция, Регрессия, Факторный анализ, Кластерный анализ, Частотный анализ) соответствуют вычислительным методам, назначение и способ реализации которых описаны в соответствующих разделах. Отметим, что перед тем как воспользоваться любым вычислительным методом, в Excel-таблицу следует ввести любым способом исходные данные. Техника их подготовки подробно описана в разделе 1.5.

Примечание. В базовой версии программы Факторный анализ, Кластерный и Частотный анализ отключены, а соответствующие им пункты меню заблокированы.

Пункт "**Формат данных**" определяет способ подготовки данных. При его запуске появляется диалоговое окно "Установки блока данных", вид которого представлен в разделе "Подготовка данных". Использование данного пункта целесообразно лишь в том случае, если Вы собираетесь длительное время работать с одним и тем же шаблоном данных и одинаковой ориентацией таблицы и хотите отключить постоянное высвечивание диалогового окна "Установка блока данных" перед запуском метода вычислений. Для этого достаточно убрать символ выбора ("крестик") перед элементом "Запрос шаблона данных на каждом шаге обработки". Когда появится необходимость в его появлении, в меню "СтатЭкс" снова выберите пункт "Формат данных" и сделайте новые установки.

Пункт "**Параметры**" позволяет переустановить числовые значения основных параметров программы. При активизации пункта "Параметры" меню "СтатЭкс" появляется диалоговое окно "Регулируемые параметры программы", позволяющее изменить значения ряда важных параметров.



Вес характеристики точности в критерии качества модели изменяется в диапазоне от 0 до 100% с шагом 1%. При минимальном значении этого параметра качество модели определяется только характеристикой ее адекватности, а при максимальном значении - только характеристикой точности. По умолчанию он задан величиной 75%, что свидетельствует о приоритете характеристики точности при формировании вывода о качестве построенной модели.

Параметр сглаживания при построении обобщенного прогноза изменяется в диапазоне от 0 до 100% с шагом - 1%. По умолчанию он задан величиной 85%, что свидетельствует о приоритете свойств инерционности над свойствами динамики исследуемого процесса.

Точность вычислений изменяется в диапазоне от 0.00001 до 0.5. По умолчанию она задана величиной 0.002.

Максимальное количество итераций по умолчанию задано величиной 500. При достижении указанного значения данного параметра процесс вычисления параметров модели завершается даже в том случае, когда заданная точность еще не достигнута. Диапазон изменений этого параметра - от 5 до 1000. Шаг изменения - 5. Параметры точности вычисления и максимального количества итераций используются при построении адаптивных и параметрических моделей прогнозирования, а также логистической кривой и кривой Гомперца.

Уровень значимости коэффициентов регрессии по умолчанию задан величиной 85. Диапазон изменений этого параметра - от 60 до 99. Шаг изменения - 1. Значение этого параметра используется для цветовой диагностики значимости коэффициентов линейной регрессионной модели.

Задание новых значений параметров (кроме точности) осуществляется при помощи спинера, а параметр точности устанавливается путем непосредственного ввода нового значения. Для фиксирования новых значений параметров необходимо нажать кнопку "Установить". Новые значения действуют в течение текущего сеанса работы и при перезапуске программы они принимают первоначальное значение, установленные программой.

Нажатие кнопки "Восстановить" приводит к восстановлению первоначально заданных значений параметров. Нажатие кнопки "Выход" не фиксирует новых значений параметров, если они были заданы.

Пункт **“Справка”** содержит краткую информацию о разработке и пользователе программы.

Пункт **“Помощь”** обеспечивает вызов описания программы. Техника использования “Помощи” в программе “ОЛИМП:СтатЭксперт” полностью совпадает с Excel. В частности, для того, чтобы подробнее узнать о каком-либо ключевом термине (в тексте “Помощи” он выделен подчеркиванием, а при сохранении системных установок - еще и зеленым цветом), необходимо установить на него курсор и нажать левую кнопку “мышки”. Кроме основного текста Help'a Вы сможете получать контекстную помощь на всех этапах работы. В диалоговых окнах, определяющих направление и параметры работы, она вызывается нажатием кнопки “Помощь” или функциональной клавиши “F1”, а при просмотре результатов - помощь придет к Вам при выборе пиктограммы “Помощь”, расположенной в левой части экрана (см. раздел 1.6.2).

Пункт **“Очистить”** закрывает книгу OLYMPSYS.XLS и удаляет все рабочие пиктограммы программы. После этой операции пункт “СтатЭкс” остается в основном меню, и все вычислительные возможности программы “ОЛИМП:СтатЭксперт”, по-прежнему, доступны. Полная выгрузка программы происходит лишь при выходе из Excel.

Технология обработки данных

Технология обработки данных при помощи программы сводится к выполнению следующих этапов:

- запуск программы;
- подготовка данных;
- проверка данных и восстановление пропущенных наблюдений;
- выполнение вычислений;
- просмотр, печать и сохранение результатов;
- завершение работы.

Очевидно, что первый и последний из указанных пунктов выполняются в течение одного сеанса работы в среде Excel однократно, в то время как остальные этапы могут быть реализованы многократно над одними и теми же или разными данными.

Этап 1. “Запуск программы” подробно описан в разделе 1.2.

Этап 2. “Подготовка данных” подробно описан в разделе 1.5. Отметим, что на вход вычислительным модулям могут быть поданы как исходные данные, так и преобразованные. Преобразование данных может осуществляться при помощи стандартных возможностей Excel, в частности, формульных соотношений, а также с использованием многочисленных функций (логарифмирования, квадратного корня, тригонометрических функций и др.).

Этап 3. “Проверка данных и восстановление пропущенных наблюдений”.

Выделенные пользователем данные могут содержать нечисловые и пропущенные наблюдения (клетки, не содержащие никакой информации). При обнаружении нечислового значения программа выдает соответствующее сообщение и прекращает работу. Пропущенные наблюдения должны быть предварительно устранены: либо удалены из информаци-

онного блока при помощи операций “Удалить строку”, “Удалить столбец”, либо заменены числовыми значениями. Восстановление пропущенных наблюдений можно осуществить самостоятельно (экспертным путем) или при помощи программы одним из трех способов, описанных в 1.7.1.

Этап 4. “Выполнение вычислений”. Запуск **вычислительных методов** осуществляется путем активизации внедренного в основное меню Excel пункта “СтатЭкс” и выбора в нем желаемого метода обработки. На этапе вычислений появляется различная по виду заставка, в нижней левой части которой отражается название этапа работы программы. В зависимости от этапа работы выдается сообщение:

“ОЛИМП:СтатЭксперт готовится” - на этапе удаления результатов предыдущих вычислений, переноса данных в рабочую таблицу;

“ОЛИМП: СтатЭксперт думает ” - на этапе выполнения вычислений;

“ОЛИМП: СтатЭксперт формирует отчет” - на этапе подготовки отчета в таблице Report;

“ОЛИМП: СтатЭксперт уже рисует графики!” - на заключительном этапе графического оформления результатов вычислений, если графики были заказаны.

Программа “ОЛИМП:СтатЭксперт” общается с пользователем при помощи диалоговых окон. Техника работы с ними является стандартной для Windows и одинаковой для всех этапов и методов вычислений. Выбор объектов в диалоговом окне можно осуществлять при помощи “мышки” или клавиатуры.

“Мышка” является наиболее удобным средством указания интересующего Вас объекта, а во многих случаях и изменения их состояний и значений. Ее можно быстро установить на любой объект окна (в отличие от клавиши табуляции, активизирующей их в строгой последовательности).

Многие параметры диалогового окна имеют только два состояния: “включен / не включен”. К их числу относятся, например, конкретные таблицы отчета. Графически состояние такого параметра отображается в виде маленьких квадрата или кружочка, располагающихся слева от параметра, к которому они относятся. Наличие символа “крестик” в квадрате или темной точки внутри кружочка означает выбор параметра. Для изменения состояния необходимо установить указатель “мышки” на квадрат или кружок и один раз нажать левую кнопку “мышки”.

Значение многих параметров меняется при помощи “спинера” (прямоугольник с двумя вертикальными стрелками), который располагается справа от параметра. При нажатии стрелки “вверх” значение показателя увеличивается, при нажатии стрелки “вниз” - уменьшается на единицу.

Если “мышка” не подключена к Вашей машине, для выбора параметров в диалоговых окнах придется использовать клавиатуру. В этом случае для обращения к объектам окна можно использовать клавишу табуляции “Tab” или “русскую букву-ускоритель”, которая в названии объекта помечена подчеркиванием. Клавиша табуляции активизирует объект, после чего нажатие клавиши “Enter” меняет его значение, а

буква-ускоритель сразу меняет значение параметра на противоположное.

Этап 5. "Просмотр, печать и сохранение результатов" подробно описан в разделе 1.6.

Этап 6. "Завершение работы". Для завершения работы можно просто выйти из Excel, либо воспользоваться пунктом **"Очистить"**, который закрывает книгу OLYMPSYS.XLS и удаляет все рабочие пиктограммы программы. После этой операции пункт **"СтатЭкс"** остается в основном меню, и все вычислительные возможности программы, по-прежнему, доступны.

Подготовка данных

Для всех вычислительных методов данные подготавливаются одинаково - в обычных Excel таблицах, однако перед выполнением вычислений **убедитесь**, что Ваши данные удовлетворяют всем требованиям, изложенным в разделе "Требования к данным". После этого в Excel - таблицу следует любым способом ввести данные, обозначить их и лишь затем запустить какой - либо метод анализа. Более подробно процесс подготовки данных описан ниже.

Источники и формат данных

Программа "ОЛИМП:СтатЭксперт" воспринимает данные только через Excel - таблицу. Первоначально данные в Excel - таблицу заносятся стандартным способом: курсор таблицы (светящийся прямоугольник) позиционируется на желаемую клетку и с клавиатуры вводится желаемое текстовое наименование или число. После нажатия клавиши "Enter" оно заносится в активную клетку. При вводе дробных чисел уточните используемый символ-разделитель целой и дробной части ("Сервис" (Tools) - "Параметры" (Options) - "Модуль" (Module General) - "Числовой формат" (Number Format)). Им может быть точка или запятая. При неправильном использовании этого символа набранные числовые значения заносятся как текст и воспринимаются нулевым значением.

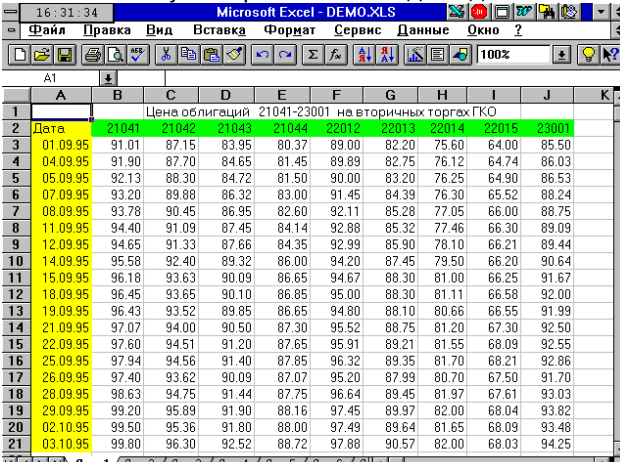
Если данные были однажды заведены и сохранены на диске, их можно прочесть. Проиллюстрируем этот процесс на демонстрационном примере. Находясь в Excel - таблице, выполните цепочку: "Файл" - "Открыть". Далее активизируйте каталог, в который при инсталляции записана программа, и, наконец, имя демонстрационного файла - DEMO.XLS. Через непродолжительное время перед вами появится указанная книга, открытая на первой таблице (см. ниже).

Как правило, для обработки непосредственно используются исходные данные, однако могут использоваться и производные от них - преобразованные данные. Преобразование данных может осуществляться при помощи стандартных возможностей Excel, в частности, формульных соотношений, а также с использованием многочисленных функций (логарифмирования, квадратного корня, тригонометрических функций и др.). Например, пусть в колонке "В" с переменной 21041 записан показатель "Объем производства", а в колонке "F" с именем переменной 22012 записан показатель "Численность работающих". На их основе можно сформировать

показатель производительности труда. Для этого в клетку "K2" занесите наименование переменной - "Производительность труда", а в клетку "K3" запишите формулу: =B3/F3. После нажатия клавиши Enter в этой клетке отобразится вычисленное по формуле значение. Чтобы не повторять эту операцию для всех других строк таблицы, воспользуйтесь командой копирования. Для этого установите курсор на клетку "K3", выделите клетку ("Правка", "Копировать"), выделите курсором диапазон клеток с K4 до K21 и занесите в них самонастраивающуюся формулу ("Правка", "Вставка"). После этого в указанном диапазоне появятся вычисленные значения.

Источником информации для вычислительных методов программы может быть активная таблица (Sheet) книги пользователя или рабочая книга OLYMPSYS.XLS. В первом случае программа автоматически переносит отмеченные данные в рабочую таблицу Data книги OLYMPSYS.XLS (делает их копию), причем расчетные показатели, заданные в таблице пользователя в виде формул, переносятся как числовые значения.

Если в момент запуска метода обработки активной была таблица "Data" или таблица "Report" (с невыделенными данными) рабочей книги OLYMPSYS.XLS, исходные данные берутся из таблицы "Data". В рабочую таблицу можно переносить свои данные "вручную" путем их копирования через Clipboard. Этот способ выбора позволяет при необходимости выбирать данные из нескольких таблиц и формировать новые, полученные на основе промежуточных вычислений. Такая технология работы обеспечивает полную сохранность исходных данных.



	A	B	C	D	E	F	G	H	I	J	K
1											
2	Дата	21041	21042	21043	21044	22012	22013	22014	22015	23001	
3	01.09.95	91.01	87.15	83.95	80.37	89.00	82.20	75.60	64.00	86.50	
4	04.09.95	91.90	87.70	84.65	81.45	89.89	82.75	76.12	64.74	86.03	
5	05.09.95	92.13	88.30	84.72	81.50	90.00	83.20	76.25	64.90	86.53	
6	07.09.95	93.20	89.88	86.32	83.00	91.45	84.39	76.30	65.52	88.24	
7	08.09.95	93.78	90.45	86.95	82.60	92.11	85.28	77.05	66.00	88.75	
8	11.09.95	94.40	91.09	87.45	84.14	92.88	85.32	77.46	66.30	89.09	
9	12.09.95	94.65	91.33	87.66	84.35	92.99	85.90	78.10	66.21	89.44	
10	14.09.95	95.58	92.40	89.32	86.00	94.20	87.45	79.50	66.20	90.64	
11	15.09.95	96.18	93.63	90.09	86.65	94.67	88.30	81.00	66.25	91.67	
12	18.09.95	96.45	93.65	90.10	86.85	95.00	88.30	81.11	66.58	92.00	
13	19.09.95	96.43	93.52	89.85	86.65	94.80	88.10	80.66	66.55	91.99	
14	21.09.95	97.07	94.00	90.50	87.30	95.52	88.75	81.20	67.30	92.50	
15	22.09.95	97.60	94.51	91.20	87.65	95.91	89.21	81.55	68.09	92.55	
16	25.09.95	97.94	94.56	91.40	87.85	96.32	89.35	81.70	68.21	92.86	
17	26.09.95	97.40	93.62	90.09	87.07	95.20	87.99	80.70	67.50	91.70	
18	28.09.95	98.63	94.75	91.44	87.75	96.64	89.45	81.97	67.61	93.03	
19	29.09.95	99.20	95.89	91.90	88.16	97.45	89.97	82.00	68.04	93.82	
20	02.10.95	99.50	95.36	91.80	88.00	97.49	89.64	81.65	68.09	93.48	
21	03.10.95	99.80	96.30	92.52	88.72	97.88	90.57	82.00	68.03	94.25	

Для определения данных, которые подлежат обработке, необходимо:

1. Сделать активной таблицу с данными;
2. Отметить в активной таблице блок данных, подлежащей обработке;
3. В стандартной строке меню выбрать внедренный пункт **"СтатЭкс"** и, в появившемся меню, выбрать метод обработки;
4. В диалоговом окне "Установки блока данных" указать ориентацию таблицы и шаблон данных.

Если в ходе работы Вы загрузили несколько книг, то для активизации нужной книги надо воспользоваться пунктом "Окно", выбрать имя желаемой кни-

ги, а затем в корешке ее таблиц выбрать нужную таблицу с данными.

В таблице пользователя данные могут располагаться в любом месте, но непременно должны быть выделены в виде **непрерывного блока**. Используя один из стандартных способов Excel, можно **выделить**:

блок клеток (левый верхний угол - правый нижний);

колонки (одну или несколько соседних);

строки (одну или несколько соседних).

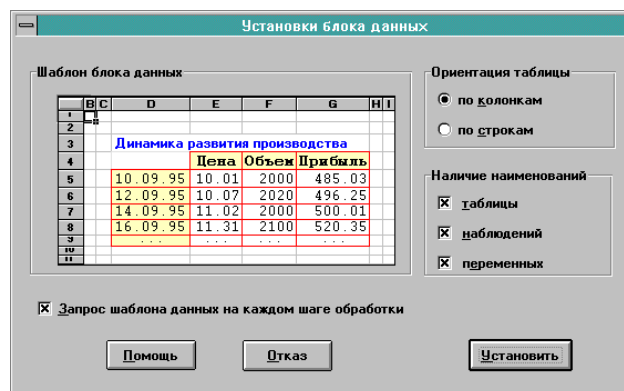
В случае отсутствия выделенного блока программа переносит все данные в таблицу Data и проводит их разбор начиная с клетки A1, учитывая установленный шаблон данных.

Явное выделение данных в виде фиксированного блока является наиболее предпочтительным. При этом способе независимо от установленного шаблона первая строка и первый столбец должны содержать информацию (наименования или числовые значения).

При неявном выделении данных (строк или столбцов) первая строка или первый столбец должны содержать информацию (наименования или числовые значения). Разбор таблицы производится до первой пустой клетки в строке. Количество строк блока определяется количеством непустых клеток в первом столбце выделенного блока, а количество столбцов блока определяется количеством непустых клеток в первой строке выделенного блока. Поэтому независимо от установленного шаблона первая строка и первый столбец должны содержать информацию (наименования или числовые значения). При любом способе определения блока данных Вы сможете указать конкретный показатель или набор обрабатываемых данных.

После выбора в стандартной строке меню пункта "СтатЭкс", а в нем - метода обработки, появится диалоговое окно "Установки блока данных".

Числовые значения показателей, содержащихся в выделенном блоке, могут быть записаны по колонкам или по столбцам. Выделенный блок кроме числовых данных может иметь текстовые атрибуты: наименование таблицы, боковик и шапку таблицы. При помощи левой кнопки "мышки" охарактеризуйте ориентацию таблицы, содержащую Ваши данные, и наличие текстовых наименований (см. подробное изложение в разделах 1.5.2 и 1.5.3). При изменении параметров в левой части диалогового окна, отображающей текущий шаблон условных данных, происходят соответствующие изменения.



Если Вы постоянно работаете с одним и тем же шаблоном данных и одинаковой ориентацией таблицы, целесообразно отключить высвечивание диалогового окна "Установка блока данных" перед запуском метода вычислений. Для этого достаточно убрать символ выбора ("крестик") перед элементом "Запрос шаблона данных на каждом шаге обработки". Когда появится необходимость в его появлении в меню "СтатЭкс", выберите пункт "Формат данных" и сделайте новые установки.

Для фиксации установок блока данных нажмите кнопку "**Установить**". Для отказа от продолжения вычислений (например, если Вы забыли структуру выделенного блока) - нажмите кнопку "Отказ".

Если программа неправильно воспринимает данные или выдает сообщение об ошибке, убедитесь, что:

- активной является непустая таблица;
- шаблон, объявленный Вами в диалоговом окне "Установки блока данных", соответствует шаблону выделенного блока (ориентация таблицы, наличие наименований таблицы, переменных, наблюдений);
- в выделенном блоке первая строка (при выделении по колонкам) и первый столбец (при выделении по строкам) являются непустыми;
- в активной, но невыделенной таблице пользователя информационный блок начинается с клетки A1;
- количество наблюдений в выделенном блоке не меньше 5;
- символ разделитель дробной и целой части в Ваших числовых данных совпадает с символом, который объявлен в Excel (им может быть запятая или точка, см. "Сервис"("Tools") - "Параметры"("Options") - "Модуль"("Module General") - "Числовой формат"("Number Format")).

Ориентация таблицы

Ориентация таблицы характеризует расположение **числовых значений** исследуемых показателей. Таблица может иметь две альтернативные ориентации:

- по колонке (столбцу);
- по строке.

Если каждое новое значение текущей переменной располагается в следующей строке, т.е. данные по каждому показателю располагаются сверху вниз в одной колонке (например, C4, C5, C6 и т.д.), то таблица имеет ориентацию "по колонке". Примером

такой ориентации являются приведенные выше данные первого листа демонстрационного файла.

Если каждое новое значение располагается в следующей колонке, т.е. данные по каждому показателю располагаются слева направо в одной строке (например, B4, C4, D4 и т.д.), то таблица имеет ориентацию "по строке". Примером такой ориентации являются данные второго листа демонстрационного файла.

Шаблон данных

Выделенный блок кроме числовых данных может иметь (или не иметь) текстовые стандартные атрибуты таблицы: наименование таблицы, боковик и шапку таблицы. Эти три элемента характеризуют **шаблон данных**.

Наименование таблицы (заголовок блока), когда оно присутствует, должно быть расположено в первой строке и записано в одной клетке.

При ориентации таблицы "по колонке" в ее боковике (то есть первой колонке) отображаются **наименования наблюдений** (обычно это даты или номера наблюдений по порядку); а в шапке таблицы, (то есть строке, непосредственно предшествующей числовым данным) - **наименования переменных** (например, "Курс доллара", "Объем продаж"). При ориентации таблицы "по строкам" - наименования переменных отображаются в боковике, а наименования переменных - в шапке таблицы. Для указания наличия:

- наименования таблицы;
- наименования переменных;
- наименования наблюдений,
- в диалоговом окне "Установки блока данных"

в блоке "Наименование" необходимо поставить символ выбора (крестик) рядом с их названием.

При изменении установок блока данных в диалоговом окне отображается их текущее состояние, что позволяет легко определить соответствие текущего шаблона данных и выделенного блока.

Совет. Для получения более компактного протокола результатов вычислений (в частности, при проведении корреляционного анализа) выделите блок данных без наименований показателей. В этом случае они будут идентифицироваться латинскими буквами, соответствующими их местоположению в рабочей таблице "Data".

Требования к данным

Формально математические методы анализа могут обрабатывать любые данные. Однако для правильного выявления закономерностей они должны удовлетворять определенным требованиям. Все математические методы обработки используют аппарат математической статистики, который предполагает, чтобы исходные данные были сопоставимы, достаточно полны, однородны, а временные ряды наблюдений должны быть еще и устойчивы. Невыполнение одного из этих требований делает **бессмысленным** применение любого математического аппарата исследования.

СОПОСТАВИМОСТЬ данных достигается в результате одинакового подхода к наблюдениям на

разных этапах формирования ряда динамики. Уровни во временных рядах должны измеряться в одних и тех же единицах, иметь одинаковый шаг наблюдений, рассчитываться для одного и того же интервала времени, по одной и той же методике, охватывать одни и те же элементы, относящиеся к неизменной совокупности.

Несопоставимость чаще всего проявляется в стоимостных показателях. Даже в тех случаях, когда значения этих показателей фиксируются в неизменных ценах (при наличии методики такого пересчета), их часто трудно сопоставить. Такого рода несопоставимость временных рядов не может быть устранена чисто формальными методами и может лишь учитываться при содержательной интерпретации результатов статистического анализа.

ПОЛНОТА данных связана с их количеством. Достаточное число наблюдений определяется в зависимости от цели проводимого исследования динамики. Если целью является описательный статистический анализ, то в качестве изучаемого можно выбрать очень короткий интервал времени (но **не менее 5 наблюдений**). Если же целью исследования является построение модели динамики, то число уровней исходного динамического ряда в этом случае должно не менее чем в три раза превышать период упреждения прогноза, и быть не менее 12. В случае использования квартальных или месячных данных для исследования и прогнозирования сезонных процессов, исходный временной ряд должен содержать данные не менее чем за четыре года, даже если требуется прогноз на 1-2 квартала (месяца). В противном случае эти данные будут обрабатываться без учета внутригодовых колебаний.

Отметим, что пороговое минимальное значение объема данных можно оценить по статистическим таблицам, например, таблица d-критерия Дарбина - Уотсона, имеет начальное значение 15. Следовательно, при меньшем объеме наблюдений использование этого ключевого критерия для оценки качества модели некорректно.

Исходные данные могут содержать пропущенные наблюдения. Программа имеет мощный математический блок их восстановления, который описан в разделе 1.7.1.

ОДНОРОДНОСТЬ данных означает отсутствие нетипичных, аномальных наблюдений, а также изломов тенденций. **Аномальность** приводит к смещению оценок и, следовательно, к искажению результатов анализа. Во временных рядах формально она проявляется как сильный скачок (спад) с последующим примерным восстановлением предыдущего уровня.

Для диагностики аномальных наблюдений предложено несколько критериев, которые, однако, не всегда надежны при исследовании временных рядов. Так, критерий Ирвина не "ловит" аномальность, если она проявляется в срединном уровне ряда с высокой динамикой, когда скачок велик, но не превышает уровней на конце периода наблюдений. В этой связи мы предлагаем модифицировать указанный метод следующим образом [12,13]. Для всех или только для подозрительных на аномальность

уровней вычисляем среднее значение и среднее квадратическое отклонение нескольких соседних с ними. Если рассчитанная величина превышает критическое значение, наблюдение считается аномальным. Аномальные наблюдения необходимо исключить из временного ряда и заменить расчетными значениями. Изломы тенденций свидетельствуют об изменении закономерностей протекания процесса или об изменении методики вычисления значений показателя. Если значение на конце временного ряда "выпадает" из общей тенденции, то без дополнительной содержательной информации о причинах "выброса" на конце ряда не представляется возможным определить, является ли это наблюдение аномальным или свидетельствует о наличии изменения тенденции. В этом случае также важно провести качественный анализ происходящих изменений или дождаться поступления нового наблюдения.

Если излом тенденции обусловлен изменением методики расчета показателя, то уровни, предшествующие излому тенденции, могут быть использованы в оценке характеристик динамики и для построения модели при условии их пересчета согласно новой методике. В случаях, когда такой пересчет невозможен, эти уровни ряда должны быть исключены из рассмотрения. Если излом тенденции отражает изменение закономерностей развития процесса, то в качестве информационной базы для статистического анализа могут быть использованы только значения, начиная с момента времени, соответствующего последнему излому тенденций.

УСТОЙЧИВОСТЬ временного ряда отражает преобладание закономерности над случайностью в изменении уровней ряда. На графиках устойчивых временных рядов даже визуально прослеживается закономерность. На графиках неустойчивых рядов изменения последовательных уровней представляются хаотичными, и поэтому поиск закономерностей в формировании значений уровней таких рядов лишен смысла.

Просмотр результатов

Во время формирования отчета на экране появляется заставка, характеризующая выполняемый этап. Результатом работы любого вычислительного метода является **отчет**, который всегда помещается в таблицу Report рабочей книги OLYMPSYS.XLS. При каждом запуске нового метода обработки предыдущее состояние таблицы "Report" **уничтожается!** Если имеющаяся таблица отчетов Вам нужно сохранить, распечатайте ее или перенесите ее стандартными средствами Excel в другую таблицу.

Содержание отчета

Содержание отчета определяется методом (т.е. выполненным этапом вычислений), а его структура - перечнем таблиц и графиков, заказанных в соответствующем диалоговом окне. На текущем шаге может быть получен один из следующих протоколов:

Протокол "Дескриптивная статистика";
Протокол "Предварительный анализ данных";
Протокол "Построение моделей и прогнозирование";

Протокол "Корреляционный анализ";
Протокол "Регрессионный анализ";
Протокол "Факторный и компонентный анализ";

Протокол "Кластерный анализ".

Протокол "Частотный анализ".

Подробное описание этих протоколов приведено в соответствующих разделах описания вычислительных методов обработки.

Техника просмотра протокола

Для просмотра таблиц можно воспользоваться стандартными клавишами перемещения курсора: на новую строку (стрелки Вниз-вверх), страницу (PgDn, PgUp) или в конец таблицы. Для просмотра графиков достаточно нажать "мышку" на пиктограмме желаемого графика. Однако более удобно воспользоваться меню графиков и таблиц текущей порции результатов. Эти меню появляются после нажатия на соответствующую пиктограмму "ОЛИМП:СтатЭксперт", расположенную с левой стороны экрана.



- список таблиц отчета;



- список графиков отчета;



- переход в рабочую таблицу исходных данных "Data";



- помощь.

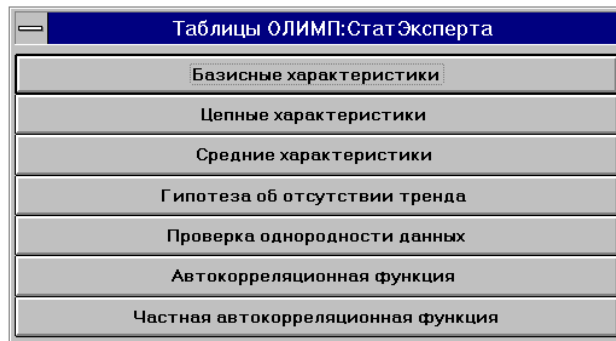
Для активизации любой пиктограммы программы "ОЛИМП:СтатЭксперт" необходимо установить на нее указатель "мышки" и нажать левую кнопку "мышки".

Ниже приведен фрагмент порции результатов обработки данных.

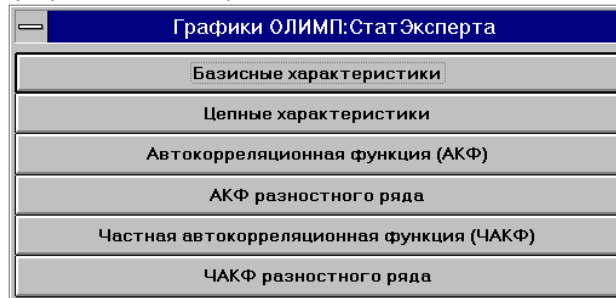
Наблюдение	Абс. прирост	Темп. роста	Темп. прироста
4 9.95	0.890	100.978	0.978
5 9.95	1.120	101.231	1.231
6 9.95	2.190	102.406	2.406
7 9.95	2.770	103.044	3.044
8 9.95	3.390	103.725	3.725
9 9.95	3.640	104.000	4.000
10 9.95	4.570	105.021	5.021
11 9.95	5.170	105.681	5.681
12 9.95	5.440	105.977	5.977
13 9.95	5.420	105.955	5.955
14 9.95	6.060	106.659	6.659
15 9.95	6.590	107.241	7.241
16 9.95	6.930	107.615	7.615
17 9.95	6.390	107.021	7.021
18 9.95	7.620	108.373	8.373
19 9.95	8.190	108.999	8.999
20 9.95	8.490	109.329	9.329
21 9.95	8.790	109.658	9.658

Воспользовавшись пиктограммой "Помощь", Вы сможете получить дополнительную информацию о содержании текущей порции результатов.

Нажатие "мышки", когда ее указатель установлен на пиктограмме "Список таблиц отчета" при просмотре отчета по предварительному анализу данных, приводит к появлению следующего меню:



Нажатие "мышки" на пиктограмму "Список графиков отчета" приводит к появлению меню вида:



Для появления таблицы (рисунка) установите указатель "мышки" на желаемый пункт и нажмите левую кнопку "мышки". При использовании разных вычислительных методов содержание этих меню будет иным, но техника их использования одинакова.

Пункты основного меню "СтатЭкс" доступны из таблицы "Report". Если Вы хотите продолжить вычисления с прежним выделенным блоком данных, можно сразу же запустить желаемый метод обработки. В противном случае воспользуйтесь пиктограммой перехода в рабочую таблицу "Data" или сделайте активной таблицу с данными в своей книге.

Источником данных может быть также таблица отчета. Для этого достаточно выделить в ней блок данных и запустить желаемый метод обработки. Например, если в режиме прогнозирования не удалось построить адекватную модель, т.е. в остаточной компоненте содержится невыявленная закономерность, можно в таблице остатков выделить в качестве информационного блока графу "Ошибка абсолютная" и вновь активизировать пункт "Временные ряды" - "Прогнозирование". В результате будет построена более точная модель, объединяющая первоначально построенную (функцию - пилот) и уточняющую функцию поведения остаточной компоненты.

Сохранение и печать протокола

Всю порцию результатов или ее часть (таблицы и графики) можно сохранить, переместить в другую программу или таблицу, переформатировать или выполнить какое-либо другое действие стандартными средствами Excel.

Таблицы отчета можно вывести на печать, используя стандартные возможности Excel, т.е. достаточно в основном меню активизировать пункт "Файл", далее в нем подпункт "Печать", подтвердив свое желание, нажатием кнопки "ОК". Если отчет по ширине не умещается на странице, для придания ему компактности следует воспользоваться возможностью

вертикальной печати, а также уменьшением шрифта, например, до 80% от обычного размера. Для этого следует выполнить следующую цепочку действий: "Файл"- "Печать"- "Страница"- "Ландшафт"- "Масштаб 80% от обычного размера".

Для печати графиков отчета их следует сначала активизировать (через пиктограмму или через список графиков), затем выделить и активизировать стандартное меню печати (Файл- Печать).

Примечание. Если Вы работаете с не русифицированной версией Excel, в распечатанном тексте или на графиках могут появиться неправильные символы.

Вычислительные методы обработки данных

В данном разделе представлено описание действий пользователя по реализации вычислительных методов. Если вы ознакомились с содержанием предыдущих разделов по технологии работы с программой, ее запуска, подготовки данных и просмотра результатов, у Вас не возникнет проблем на этом этапе работы.

Для иллюстрации действий пользователя воспользуемся данными, которые записаны в первых таблицах книги DEMO.XLS. Откройте этот файл самостоятельно. Для этого в строке меню Excel выберите пункт "Файл", а в ней команду "Открыть". Далее активизируйте каталог, в который при инсталляции записана программа, и, наконец, имя демонстрационного файла - DEMO.XLS.

Первая таблица содержит данные о динамике цен на вторичных торгах нескольких выпусков ГКО за период с 1 сентября по 3 октября 1995 года. Количество наблюдений - 19, шаг наблюдения - торговый день. Таблица с данными по ГКО расположена в клетках A1 - J21. Она имеет наименование, которое записано в клетке C1, а также наименования наблюдений (даты торгов), расположенные в первой колонке ("А"), и наименования переменных (номера выпусков ГКО), расположенные во второй строке блока. Числовые данные по каждому выпуску записываются сверху вниз, то есть таблица имеет ориентацию "по колонке". Предположим, что среди этих данных нас интересует ГКО 22014.

Вторая таблица содержит месячные данные о динамике изменения некоторых народнохозяйственных показателей РФ за период с 1 января 1993 г. по 31 декабря 1994 года. Количество наблюдений - 24, шаг наблюдения - месяц. Таблица ориентирована "по строке". Она имеет наименование, а также наименование переменных и наблюдений. Предположим, что среди этих данных нас интересует показатель "Индекс потребительских цен".

Третья таблица содержит данные о крупнейших банках РФ на 1 января 1996 года.

Числовые данные в этих таблицах являются фактическими, но мы будем акцентировать внимание на технической, а не на содержательной стороне анализа.

Примечание 1. В каждом вычислительном методе может быть получено графическое изображение результатов вычислений. Графики существенно

улучшают восприятие результатов вычислений, однако их формирование средствами Excel существенно (иногда почти в два раза) увеличивает общее время обработки данных. Если Вы хотите уменьшить время ожидания результатов, откажитесь от графиков и при установке параметров структуры отчета уберите символ “крестик” около пункта “Построение графиков”.

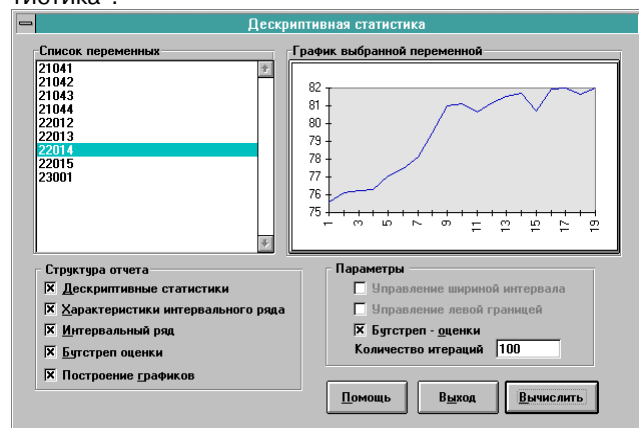
Примечание 2. При первоначальном запуске любого метода в блоке “Структура отчета” все таблицы помечены (символом “крестик”) для включения в отчет. В отличие от графиков, объем выдаваемых таблиц почти не влияет на общую продолжительность обработки данных.

Дескриптивная статистика

Дескриптивная (описательная) статистика является наиболее простым и распространенным средством характеристики исходных данных. Они позволяют получить представление об особенностях исследуемого показателя и перспективности использования более глубоких методов анализа.

Для запуска режима вычисления дескриптивной статистики активизируйте первую таблицу демонстрационного файла (см. раздел 1.5), установите курсор на пункт меню “СтатЭкс”, нажмите “мышку” и в появившемся меню выберите пункт “Дескриптивная статистика”.

Ввиду того, что обрабатываемый блок данных предварительно не был выделен, программа выдаст запрос о необходимости обработки данных с верхнего левого угла таблицы. В появившемся диалоговом окне “Установка блока данных” убедитесь, что все проставленные там установки соответствуют нашему блоку данных, и поэтому, оставив их без изменения, сразу воспользуйтесь кнопкой “Вычислить”. После этого появится диалоговое окно “Дескриптивная статистика”.



В “Списке переменных” отображаются наименования показателей, взятых из блока данных в соответствии с указанным ранее шаблоном. Текущей является переменная, которая выделена светящейся строкой курсора. В начальном состоянии всегда подсвечивается первая строка.

Примечание. Если бы у нас в списке переменных появились неправильные имена переменных или на графике отобразилась бы странная динамика, это свидетельствовало бы об ошибочном формате вы-

бранных данных. В этом случае следует выйти из окна диалога, нажав кнопку “Выход”, не начиная вычислений.

При перемещении курсора (движении указателя “мышки”) по списку переменных, справа отображается **график** их значений. Установите указатель “мышки” на строку с наименованием интересующего нас показателя (ГКО 22014) и нажмите один раз левую кнопку “мышки”. Справа появится график динамики этого показателя. Его можно увеличить до размера полного окна. Для этого достаточно нажать “мышку”, когда ее указатель установлен на графике. При желании график можно распечатать или перенести в другие программы (например, в текстовый редактор) стандартными средствами Excel. Возврат в диалоговое окно из режима просмотра укрупненного графика происходит при нажатии левой кнопки “мышки”, когда ее указатель стоит на графике, или клавиши Esc.

Для того чтобы заказать появление в отчете какой-либо таблицы, достаточно добиться появления символа выбора (крестик) слева от ее наименования. Предположим, что для нас представляют интерес все результаты вычислений. Поэтому в блоке “Структура отчета” оставим без изменения установки на выдачу всех таблиц. Для начала расчетов нажмите кнопку “Вычислить”, после чего появится заставка, внутри которой отображается сообщение о выполнении программой этапа работ.

При завершении вычислений в листе “Report” появляется протокол, содержащий 4 таблицы, полный список таблиц можно увидеть после нажатия на пиктограмму (см. также раздел 1.6. “Просмотр результатов”).

Структурно первые две таблицы состоят из двух граф: наименования характеристики и числового значения. Формулы их вычисления приведены в одноименном разделе второй части данного документа.

Две последние таблицы интервальных рядов имеют одинаковую структуру. Для них формируются графики-гистограммы, построенные по частости наблюдений. Для их появления на экране установите курсор на картинку, расположенную в таблице результатов, и нажмите “мышку” или воспользуйтесь второй слева пиктограммой (см. раздел “Просмотр результатов”).

Протокол режима “Дескриптивная статистика”

А	В
Описательные статистики переменной - 22014	
Характеристики вариационного ряда	
Характеристика	Значение
Число наблюдений	19,000
Среднее значение	79,575
Верхняя оценка среднего	80,500
Нижняя оценка среднего	78,649
Среднекв. отклонение (S)	2,320
Дисперсия	5,381
Дисперсия (несмещ. оценка)	5,680

Среднеkv. откл. (несмещ. оценка)	2,383				
Среднее линейное отклонение (L)	2,128				
Момент начальный 2-го порядка	6337,511				
Момент начальный 3-го порядка	505155,365				
Момент начальный 4-го порядка	40298300,756				
Момент центральный 3-го порядка	-6,454				
Момент центральный 4-го порядка	45,815				
Коэффициент асимметрии A	-0,517				
Несмещенная оценка A	-0,562				
Среднеkv. отклонение A	0,524				
Коэффициент эксцесса E	0,000				
Несмещенная оценка E	0,397				
Среднеkv. отклонение E	1,014				
Мода	82,000				
Минимальное значение	75,600				
Максимальное значение	82,000				
Размах (R)	6,400				
Коэффициент вариации по R	8,043				
Коэффициент вариации по L	2,674				
Коэффициент вариации по S	2,915				
Характеристики интервального ряда					
Характеристика	Значение				
Среднее значение	79,642				
Среднеkv. отклонение (S)	2,387				
Дисперсия	5,698				
Коэффициент асимметрии A	-0,525				
Коэффициент эксцесса E	-1,387				
Медиана	80,613				
Мода	81,619				
Коэффициент вариации по S	2,997				
A	B	C	D	E	F

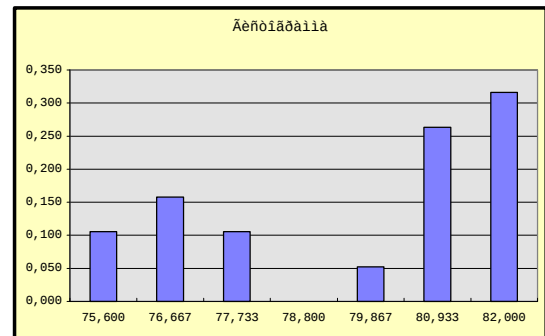
Интервальный ряд

Номер интервала	Начало	Середина	Конец	Частота	Частость
1	75,067	75,600	76,133	2,000	0,105
2	76,133	76,667	77,200	3,000	0,158
3	77,200	77,733	78,267	2,000	0,105
4	78,267	78,800	79,333	0,000	0,000
5	79,333	79,867	80,400	1,000	0,053
6	80,400	80,933	81,467	5,000	0,263
7	81,467	82,000	82,533	6,000	0,316

Интервальный ряд бутстреп-средних

Номер интервала	Начало	Середина	Конец	Частота	Частость
1	77,360	77,568	77,775	1,000	0,010
2	77,775	77,983	78,191	1,000	0,010
3	78,191	78,398	78,606	9,000	0,090
4	78,606	78,814	79,021	25,000	0,250
5	79,021	79,229	79,436	28,000	0,280

6	79,436	79,644	79,852	23,000	0,230
7	79,852	80,059	80,267	10,000	0,100
8	80,267	80,475	80,682	2,000	0,020
9	80,682	80,890	81,097	1,000	0,010



Предварительный анализ данных

Назначение этого этапа исследования состоит в формировании предварительных оценок об особенностях обрабатываемого показателя и перспективности применения отдельных методов их обработки.

Данный этап исследования реализуем на основе тех же данных (по цене ГКО 22014), что и при вычислении характеристик дескриптивной статистики, причем стартовать будем из рабочей таблицы Data книги OLYMSYS.XLS, куда они уже перенесены программой. Если Вы еще находитесь в таблице отчетов "Report", воспользуйтесь для перехода третьей сверху пиктограммой "Переход в рабочую таблицу данных (Data)". Для запуска режима вычисления, установите курсор на пункт меню "СтатЭкс", нажмите "мышку" и в появившемся меню выберите пункт "Временные ряды".

Для определения данных можно поступить тем же образом, что был использован в режиме "Дескриптивная статистика", но на этот раз выделить блок "по строке" с 1 по 21 строку. Выделенный блок кроме числовых данных имеет наименование таблицы, которое записано в клетке C1, а также наименования наблюдений (даты торгов), расположенные в первой колонке (A), и наименования переменных (номера выпусков ГКО), расположенные во второй строке блока. Числовые данные по каждому выпуску записываются сверху вниз, т.е. таб-

лица имеет ориентацию "по колонке". Исходя из этого, в появившемся диалоговом окне "Установки блока данных" (см. 1.5.1) убедимся, что все предоставленные там установки соответствуют нашему блоку данных, и поэтому, оставив их без изменения, сразу воспользуемся кнопкой "Установить". Программа сообщит, что последние 246 столбцов выделенных строк являются пустыми, и запросит разрешения не использовать их. Ответьте на этот запрос утвердительно, после чего появится диалоговое окно "Обработка временных рядов".



При перемещении курсора (движении "мышки") по списку переменных, справа отображается **график** их значений. Установите указатель "мышки" на строку с наименованием интересующего нас показателя (ГКО 22014) и нажмите один раз левую кнопку "мышки". Справа появится график динамики этого показателя. Его можно увеличить до размера полного окна. Для этого достаточно нажать "мышку", когда ее указатель установлен на графике. При желании график можно распечатать или перенести в другие программы (например, в текстовый редактор) стандартными средствами Excel. Возврат в диалоговое окно происходит при повторном нажатии мышки или клавиши Esc.

Параметр "**Шаг наблюдения**" означает периодичность наблюдений. С формально-статистической точки зрения все соседние значения должны отстоять друг от друга на одинаковом удалении (год, месяц, торговый день и т.д.). В нашем случае торговый день можно приравнять к календарному и оставить без изменения стоящее там значение "день". Если Вы не согласны с этим, установите указатель "мышки" на стоящую справа от этого параметра стрелку, нажмите "мышку" и в появившемся списке выберите иной шаг наблюдения, например, "прочее".

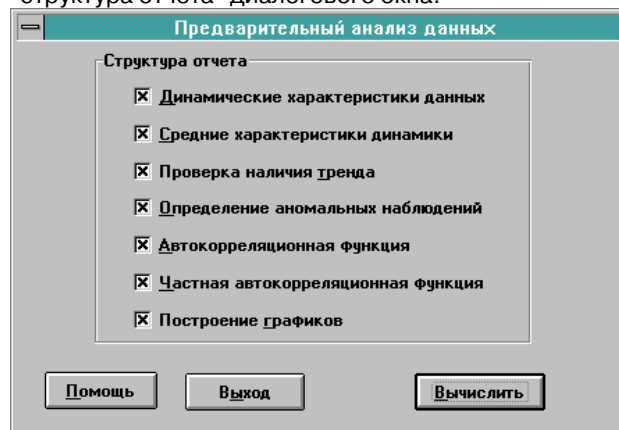
Параметр "**Период сезонности**" характеризует оценку периодичности колебаний уровней ряда. Он изменяется в пределах от 1 до 12. Значение "1" характеризует отсутствие колебаний, что характерно для годовых наблюдений. Значения "4" и "12" соответствуют периоду колебаний квартальных и месячных данных. Они устанавливаются программой автоматически при указании пользователем шага наблюдений "Квартал" и "Месяц". Указанные значения

можно изменить, но в большинстве случаев это делать нецелесообразно. При малом числе наблюдений (менее трех периодов сезонности), программа блокирует пользовательскую установку сезонности. Например, при 15 месячных наблюдениях нельзя установить период сезонности равный 12. При обработке дневных наблюдений, целесообразно провести вычисления с периодом сезонности равным "1" и "5". Последнее значение соответствует недельному колебанию многих финансовых показателей. В нашем примере установим период сезонности равный единице.

Примечание. Установка периода сезонности больше единицы предполагает использование более сложных математических моделей, что приведет, в свою очередь, к существенному увеличению времени расчетов.

Далее выберите этап работ – "Предварительный анализ", нажмите на кнопку "Вычислить", после чего появится диалоговое окно "Предварительный анализ".

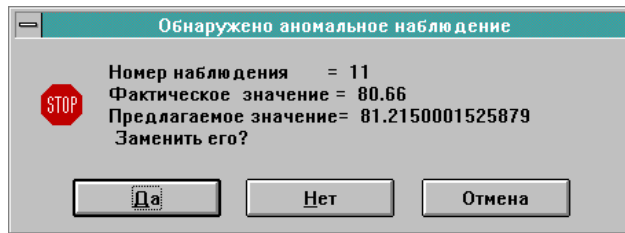
Полнота выдачи результатов вычислений определяется перечнем таблиц, заказанных в блоке "структура отчета" диалогового окна.



Результаты вычислений отражаются в протоколе, для получения которого необходимо заказать (добиться появления символа выбора "крестик") не менее одного элемента отчета (одной таблицы).

Среди всех предлагаемых направлений исследования особое место занимает пункт "Определение аномальных наблюдений", который доступен при обработке данных с периодом сезонности, не превышающем 1. Если он не заказан, проверка обрабатываемого показателя на наличие в нем резко выделяющихся наблюдений не производится.

После выбора полного отчета нажмите кнопку "Вычислить". Прежде всего, реализуется блок проверки статистической однородности наблюдений. При проверке данных на аномальность и наличие подозрительных наблюдений, программа выдает запрос на их устранение путем замены фактических наблюдений на расчетные. В нашем примере появится сообщение:



При получении подтверждения (нажатии кнопки "Да"), программа записывает новое расчетное значение в рабочую таблицу "ОЛИМП:СтатЭксперт", окрашивая его в зеленый цвет. Если на вопрос о замене аномального значения Вы ответите отрицательно ("Нет"), данные в рабочей таблице "Data" окрашиваются красным цветом, но не меняются, причем в этом случае результаты анализа последующих наблюдений могут быть искажены. Если нажать кнопку "Отмена" ("Cancel"), проверка прекращается. При получении сообщения об аномальности некоторых наблюдений, следует помнить, что их устранение может существенно изменить (как правило, улучшить) характеристики модели и результаты прогнозирования. В этой связи на заданный вопрос ответим "Да", так же как и на аналогичный запрос по поводу другого аномального наблюдения.

После этого появится заставка, внутри которой содержится сообщение о выполнении программой соответствующих этапов. Результаты обработки отображаются в Протоколе "Предварительный анализ данных", который содержит 7 таблиц (см. ниже).

Таблицы "**Базисные характеристики динамики**" и "**Цепные характеристики динамики**" содержат абсолютный прирост, темпы роста и темпы прироста, вычисленные по отношению к первому и предыдущему уровням ряда. Динамика показателя отчетливо видна на графиках. Список графиков появится при нажатии на вторую сверху пиктограмму отчета ОЛИМП:СтатЭксперт. На рисунке отражена динамика базисных темпов роста и прироста. Их траектории близки, но они существенно разнятся масштабом значений, представленных на левой и правой осях графика.

Примечание. На приведенных ниже графиках базисных и цепных темпов роста масштабы боковых осей не совпадают с теми, которые автоматически сформированы программой. Программа использует числовые данные соответствующих граф, которые всегда отображаются с тремя знаками в дробной части. Если Вас это не устраивает, можно, используя стандартные средства форматирования Excel (выделить диапазон данных и воспользоваться командой "Формат-Ячейки" ("Format-Cells")), установить желаемую точность представления данных. Например, сформатировав данные абсолютных приростов как целые значения, а темпов роста - с двумя знаками в дробной части, Вы получите их более удобное представление. Этим приемом можно пользоваться при графическом отображении и других характеристик.

Таблица "**Средние характеристики**" содержит усредненные показатели предыдущих таблиц, а также среднее арифметическое исходных данных. Учитывая ярко выраженный характер динамики ис-

следуемого показателя, ее значение (79,662) не целесообразно использовать для характеристики динамики.

Таблица "**Гипотеза об отсутствии тренда**" содержит результат проверки отсутствия тренда в исходном ряду наблюдений на основе двух критериев, причем в случае их противоречивости вывод, представленный в последней строке таблицы, формулируется на основе первого критерия. Если гипотеза принимается, использовать для поиска закономерностей изменения уровней ряда кривые роста и адаптивные методы нецелесообразно. В нашем примере оба критерия отвергают гипотезу об отсутствии тренда (в графе "Результат" стоит слово "Нет"), и, следовательно, весь реализованный в программе математический аппарат целесообразно использовать для последующего прогнозирования.

Таблица "**Проверка однородности данных**" содержит результат проверки исходных данных на аномальность. В таблице отображаются номера подозрительных наблюдений, фактические и расчетные значения, а также значение, присвоенное наблюдению после ответа пользователя на запрос о необходимости замены каждого аномального наблюдения. При положительном ответе на запрос о замене аномальных наблюдений новые значения заменяют старые в таблице "Data". В нашем примере аномальными признаны 11 и 15 наблюдения, которым присвоены новые значения. При отсутствии аномальных наблюдений выдается соответствующее сообщение.

В таблицах "**Автокорреляционная функция**" и "**Частная автокорреляционная функция**" содержатся результаты вычислений нескольких одноименных коэффициентов для исходного и разностного (порядка d) рядов. Разностный ряд идентифицируется таким значением порядка, который лучше других определяет стационарный процесс. На графиках АКФ и ЧАКФ критический уровень значений коэффициентов (представленный в последней строке таблиц) отображается пунктирной линией. Он помогает определить порядок авторегрессии и порядок скользящего среднего.

Протокол "Предварительный анализ данных"

А В С Д

Статистики временного ряда - 22014

Базисные характеристики			
Наблюдение	Абс. прирост	Темп роста	Темп прироста
4. 9. 95	0,520	100,688	0,688
5. 9. 95	0,650	100,860	0,860
7. 9. 95	0,700	100,926	0,926
8. 9. 95	1,450	101,918	1,918
11. 9. 95	1,860	102,460	2,460
12. 9. 95	2,500	103,307	3,307
14. 9. 95	3,900	105,159	5,159
15. 9. 95	5,400	107,143	7,143
18. 9. 95	5,510	107,288	7,288
19. 9. 95	5,615	107,427	7,427
21. 9. 95	5,600	107,407	7,407
22. 9. 95	5,950	107,870	7,870
25. 9. 95	6,100	108,069	8,069
26. 9. 95	6,205	108,208	8,208
28. 9. 95	6,370	108,426	8,426
29. 9. 95	6,400	108,466	8,466
2. 10. 95	6,050	108,003	8,003
3. 10. 95	6,400	108,466	8,466
Цепные характеристики			
Наблюдение	Абс. прирост	Темп роста	Темп прироста
4. 9. 95	0,520	100,688	0,688
5. 9. 95	0,130	100,171	0,171
7. 9. 95	0,050	100,066	0,066
8. 9. 95	0,750	100,983	0,983
11. 9. 95	0,410	100,532	0,532
12. 9. 95	0,640	100,826	0,826
14. 9. 95	1,400	101,793	1,793
15. 9. 95	1,500	101,887	1,887
18. 9. 95	0,110	100,136	0,136
19. 9. 95	0,105	100,129	0,129
21. 9. 95	-0,015	99,982	-0,018
22. 9. 95	0,350	100,431	0,431
25. 9. 95	0,150	100,184	0,184
26. 9. 95	0,105	100,129	0,129
28. 9. 95	0,165	100,202	0,202
29. 9. 95	0,030	100,037	0,037
2. 10. 95	-0,350	99,573	-0,427
3. 10. 95	0,350	100,429	0,429

А	В
Средние характеристики	
Характеристика	Значение
Среднее арифметическое	79,662
Средний темп роста	100,452
Средний темп прироста	0,452
Средний абсолютный прирост	0,356

Гипотеза об отсутствии тренда	
Метод проверки	Результат
Метод Форстера-Стюарта	Нет
Метод сравнения средних	Нет
Вывод: отвергается	

Проверка однородности данных

Номер на-блюдения	Факт	Расчет	Новое значение
11	80,660	81,215	81,215
15	80,700	81,805	81,805

Обнаружены аномальные наблюдения!

Автокорреляционная функция

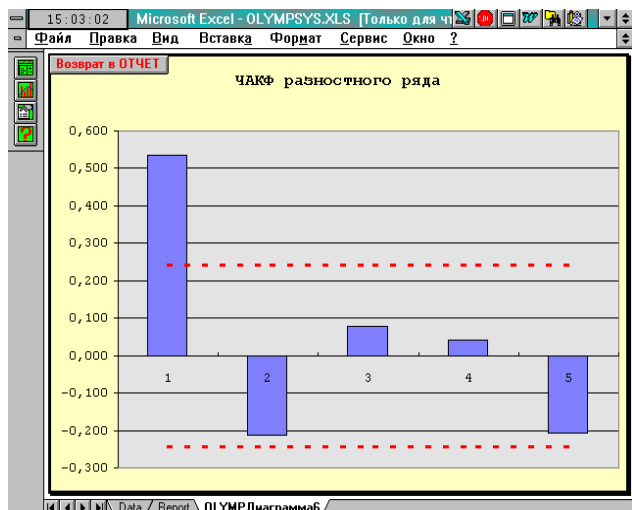
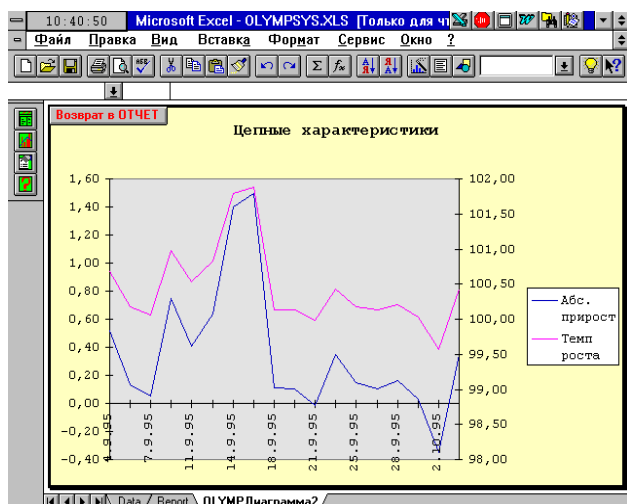
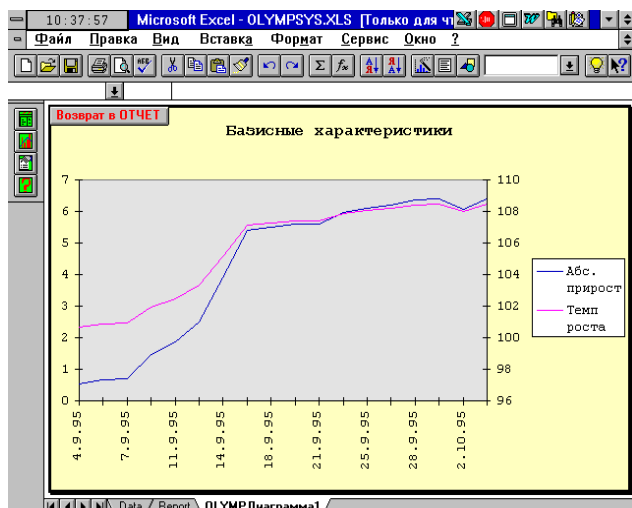
Лag	Исходный ряд	Разностный ряд (d=1)
1	0,869	0,449
2	0,730	0,058
3	0,565	0,020
4	0,380	-0,018
5	0,206	-0,199

Стандартное отклонение = +0.4916, +0.2800

Частная автокорреляционная функция

Лag	Исходный ряд	Разностный ряд (d=1)
1	0,890	0,536
2	0,091	-0,216
3	-0,002	0,079
4	-0,123	0,042
5	-0,074	-0,209

Стандартное отклонение = +0.2294, +0.2425



Прогнозирование временных рядов

Назначение статистических методов прогнозирования состоит в том, чтобы на основе математической модели сформировать точечный и интервальный прогнозы исследуемого показателя, охарактеризовав в доступной для пользователя форме степень доверия к полученным результатам.

Продолжим обработку показателя цены ГКО 22014, который был ранее исследован на этапе предварительного анализа, с целью выявить закономерности его изменения и получить прогнозные оценки на три шага (торговых дня) вперед. Доверительный интервал прогноза сформируем для уровня вероятности 80%. С учетом результатов предварительного анализа, подтвердившего, в частности, наличие тенденции, используем весь математический аппарат, эффективность которого предварительно проверим при помощи ретропрогноза на глубину в три торговых дня.

Для этого выполните все действия вплоть до появления диалога "Обработка временных рядов", таким же способом, как и на предыдущем этапе обработки данных (см. раздел 1.7.2). Выберите на этот раз этап "Построение моделей и прогнозирование", после чего появится соответствующее диалоговое окно.

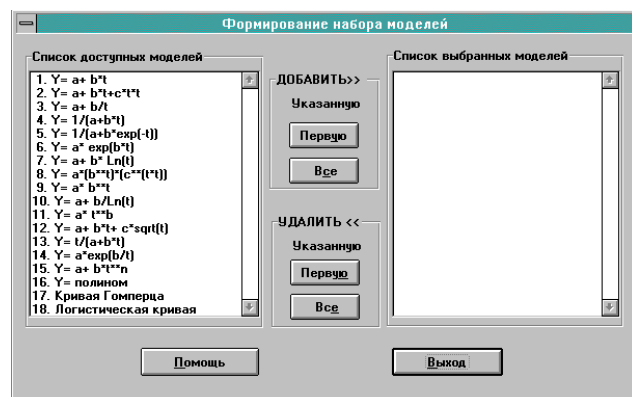
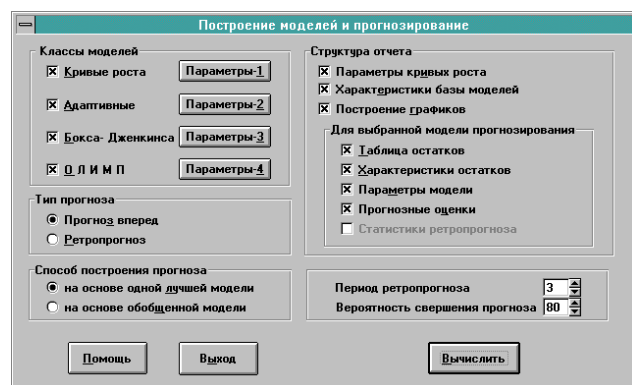
В этом окне во многих случаях можно сразу же нажать кнопку "Вычислить" и через непродолжительное время получить хорошие результаты. Однако программа в этом случае выполнит расчеты не в полную силу. Поэтому выполните следующие действия.

В блоке "Классы моделей", прежде всего, добьемся появления символа выбора "крестик" рядом со всеми наименованиями классов. Далее конкретизируем перечень моделей каждого класса. Для этого необходимо воспользоваться кнопкой "Параметры". Для моделей кривых роста и адаптивных методов установка параметров сводится к указанию состава моделей (по умолчанию программа использует не все модели), а для методов Бокса-Дженкинса и "ОЛИМП" их установка заключается в задании пределов изменения параметров модели.

Совет. При первоначальной обработке исследуемого показателя, целесообразно выбирать **все** классы моделей, а в них все имеющиеся модели. Это

несколько увеличит время вычислений, но позволит быстро сформировать круг достаточно надежных моделей.

Для задания **кривых роста** воспользуйтесь кнопкой "Параметры-1", после нажатия которой появляется диалоговое окно "Формирование набора моделей". Зададим все модели кривых роста, кроме полинома, который для целей прогнозирования подходит плохо. Для выбора моделей в группе "Добавить" воспользуйтесь кнопкой "Все" (все модели перейдут из списка доступных в список выбранных моделей), а затем в списке "Выбранные переменные" установите курсор на строку "Y = полином" и нажмите "мышку" (модель "полином" удалится из списка выбранных и вернется в список доступных переменных). Для выхода из диалога воспользуйтесь кнопкой "Вычислить".

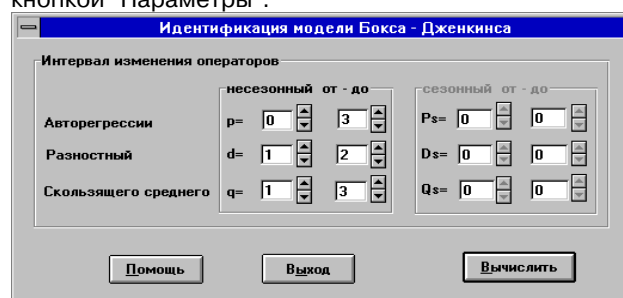


Примечание. Если Вы выйдете из диалога, не выбрав ни одной модели, программа воспользуется списком моделей, которые установлены по умолчанию.

Формирование набора **адаптивных методов** осуществляется аналогично, после нажатия соответствующей кнопки "Параметры-2", с той лишь разницей, что список доступных переменных содержит три модели.

Метод Бокса - Дженкинса для отражения особенностей развития показателя строит модель авторегрессии $AP(p)$ порядка "p" и (или) скользящего среднего $CC(q)$ порядка q для ряда наблюдений, полученного из исходного путем устранения тренда применением разностного оператора порядка d. Таким образом, в общем случае строится интегрированная модель АРСС (p, d, q). Основной сложностью

построения такой модели является ее идентификация, т. е. определение числовых значений трех указанных параметров. Как показывает практика, наиболее действенным путем решения этой задачи является перебор всех возможных моделей. Если Вы располагаете временем и достаточно быстродействующей ПЭВМ, можно поручить программе просмотреть весь диапазон возможных моделей в пределах от (0,0,1) до (5,2,5). Однако следует помнить, что более сложные модели обычно имеют лучшие аппроксимирующие свойства, но вместе с тем обладают меньшей устойчивостью, что для прогнозирования является определяющим отрицательным фактором. В большинстве случаев порядок параметров **p, d, q** не превышает **трех**. Среди моделей Бокса-Дженкинса построим модели, имеющие параметры авторегрессии в интервале от нуля до 3, разностный оператор – от 1 до 2, оператор скользящего среднего – от 1 до 3. Для задания этих установок воспользуемся кнопкой "Параметры".



Программа переспросит, действительно ли Вы хотите участвовать в процессе идентификации модели, и в случае положительного ответа появляется диалоговое окно "Идентификация модели Бокса - Дженкинса". При помощи спинеров, расположенных справа от параметров, выполните намеченные установки и нажмите кнопку "Вычислить".

Примечание 1. Для данных, содержащих циклическую или сезонную компоненту, дополнительно указываются три аналогичных параметра. Если период сезонности для обрабатываемых данных в диалоговом окне "Обработка временных рядов" был указан "1", то блок установки параметров сезонных операторов заблокирован.

Примечание 2. Если начальное значение операторов будет задано больше конечного, программа устранил это несоответствие, установив конечное значение параметра равным его начальному значению.

Формирование **набора моделей ОЛИМП** происходит аналогично после нажатия соответствующей кнопки "Параметры-4". Метод ОЛИМП для отражения особенностей развития показателя строит модель $APCC(p,q,d)$ порядка "p" и скользящего среднего порядка "q", но непосредственно для исходного ряда наблюдений. Поэтому порядок разностного оператора $d=0$. В большинстве случаев порядок параметров p и q не превышает **четырёх**. Установим параметры авторегрессии в интервале от нуля до 4, оператор скользящего среднего – от 1 до 4.

На этом завершается определение круга заказанных моделей.

Далее установим "Тип прогноза". С целью проверки эффективности математического аппарата выберите пункт "Ретропрогноз". Параметр "Период ретропрогнозирования" определяет количество точек проверочной последовательности (она должна быть не менее двух точек). Установим при помощи спинера период ретропрогноза 3 шага.

Доверительная вероятность прогноза первоначально задана уровнем 80%. Его можно изменить при помощи рядом стоящего спинера. Следует помнить, что при увеличении этой величины, ширина прогнозного интервала увеличивается, и слишком большое его значение (90% и выше) обычно приводит к потере информационной ценности прогноза.

Далее определим параметр в блоке "Способ построения прогноза". Прогнозные оценки можно получить двумя альтернативными способами: как прогноз по лучшей модели либо используя построение обобщенного прогноза на основе нескольких "хороших". Каждый из этих подходов имеет свои достоинства, однако не является заведомо лучше другого, и потому, в большинстве случаев, их целесообразно использовать совместно. Выберем режим получения прогноза "На основе лучшей модели".

Полнота выдачи результатов вычислений определяется перечнем таблиц, заказанных в блоке "Структура отчета". Количество заказанных таблиц практически не влияет на время формирования отчета. Поэтому имеет смысл заказать **все** перечисленные таблицы (поставить рядом с ними символ выбора - крестик). Это состояние уже зафиксировано.

Для начала расчетов следует нажать на кнопку "Вычислить".

Протокол ретропрогноза

А	В	С	Д
Модели временного ряда - 22014			
Таблица кривых роста			
Функция	Критерий	Эластичность	
$Y(t)=+75.058+0.493*t$	0,509	0,053	
$Y(t)=+74.122+0.805*t - 0.018*t*t$	0,401	0,053	
$Y(t)=+80.799-7.351/t$	3,046	-0,011	
$Y(t)=1./(+0.013-0.0000*t)$	0,562	-0,053	
$Y(t)=1./(+0.013+0.002*exp(-t))$	4,277	0,000	
$Y(t)=+75.117*exp(+0.006*t)$	0,534	0,053	
$Y(t)=+73.782+2.850*ln(t)$	0,984	0,036	
$Y(t)=(+74.190)*(+1.010)**t*(+1.000)**(t*t)$	0,392	0,000	
$Y(t)=(+75.117)*(+1.006)**t$	0,534	0,053	
$Y(t)=0+0/ln(t)$	Нет	0,000	
$Y(t)=(+73.900)*t**(+0.036)$	0,931	0,036	
$Y(t)=+73.438+0.233*t+1.376*sqr(t)$	0,501	0,050	
$Y(t)=t/(+0.004+0.012*t)$	11,120	0,034	

Функция	Критерий	Эластичность
$Y(t)=+80.793*exp(-0.094/t)$	2,981	-0,011
$Y(t)=+77.480+0.002*t**2$	1,236	0,003
$Y(t)=(+3.726)*((+1.735**(+1.127)**t))$	4199,821	0,000
$Y(t)=(+79.364)/(+22.466*exp(-22.420*t))$	6,914	0,000

Выбрана функция

$$Y(t)=(+74.190)*(+1.010)**t*(+1.000)**(t*t)$$

Характеристики базы моделей

Модель	Адекватность	Точность	Качество
ОЛИМП(4,1)	99,943	96,972	97,715
Метод Брауна(+0.106)	98,753	96,178	96,822
ОЛИМП(2,4)	99,994	97,600	98,199
ОЛИМП(2,3)	95,100	96,679	96,284
АРИСС(2,2,1)	99,843	95,910	96,893
АРИСС(0,1,1)	96,626	96,415	96,468
АРИСС(0,1,2)	99,600	96,314	97,136
АРИСС(0,1,3)	99,973	96,119	97,083
ОЛИМП(3,1)	98,828	96,538	97,110
АРИСС(0,2,2)	99,013	97,011	97,512
АРИСС(0,2,3)	98,333	96,198	96,731
АРИСС(1,1,1)	99,904	96,287	97,191
ОЛИМП(4,3)	94,689	97,235	96,598
АРИСС(1,1,3)	99,945	96,191	97,129
ОЛИМП(1,4)	98,335	97,420	97,649
ОЛИМП(2,2)	100,000	96,439	97,330
АРИСС(1,2,3)	99,477	96,183	97,007
АРИСС(2,1,1)	99,981	96,215	97,157
АРИСС(2,1,2)	99,635	96,196	97,056
АРИСС(2,1,3)	99,945	96,169	97,113
Лучшая модель ОЛИМП(2,4)			

Параметры моделей

Модель	a1	a2	a3	a4	a5	a6
ОЛИМП(2,4)	-0,863	-0,007	0,014	0,711	0,796	0,086

Таблица остатков

Номер	Факт	Расчет	Ошибка абс.	Ошибка относит.
8.9.95	77,050	76,818	0,232	0,300
11.9.95	77,460	77,044	0,416	0,538
12.9.95	78,100	77,871	0,229	0,293
14.9.95	79,500	79,836	-0,336	-0,422
15.9.95	81,000	81,496	-0,496	-0,613
18.9.95	81,110	80,807	0,303	0,373
19.9.95	81,215	81,530	-0,315	-0,388

Номер	Факт	Расчет	Ошибка абс.	Ошибка относит.
21.9.95	81,200	81,113	0,087	0,107
22.9.95	81,550	81,638	-0,088	-0,108
25.9.95	81,700	81,745	-0,045	-0,055
26.9.95	81,805	81,717	0,088	0,107
28.9.95	81,970	82,021	-0,051	-0,062

Характеристики остатков

Характеристика	Значение
Среднее значение	0,002
Оценка дисперсии	0,071
Приведенная дисперсия	0,086
Средний модуль остатков	0,224
Относительная ошибка	0,281
Критерий Дарбина-Уотсона	2,005
Коэффициент детерминации	1,000
F - значение (n1 = 1, n2 = 10)	100000,000
Критерий адекватности	99,994
Критерий точности	97,600
Критерий качества	98,199

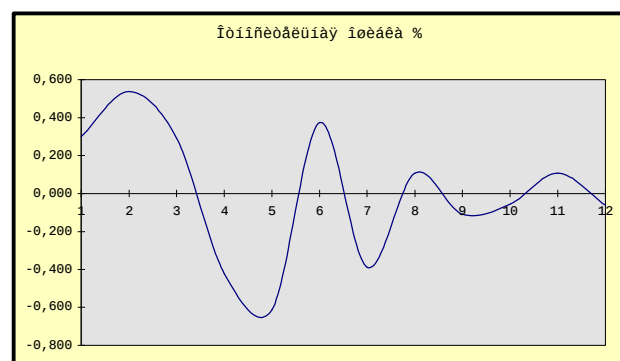
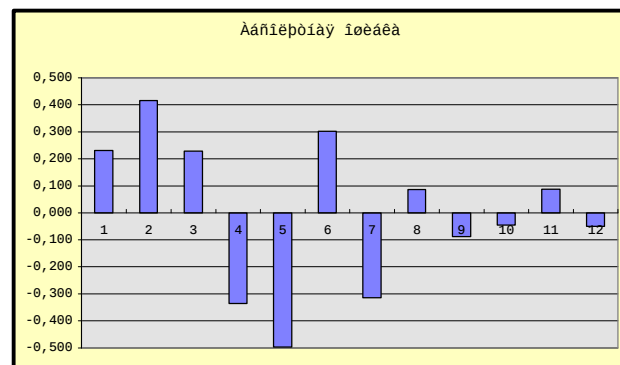
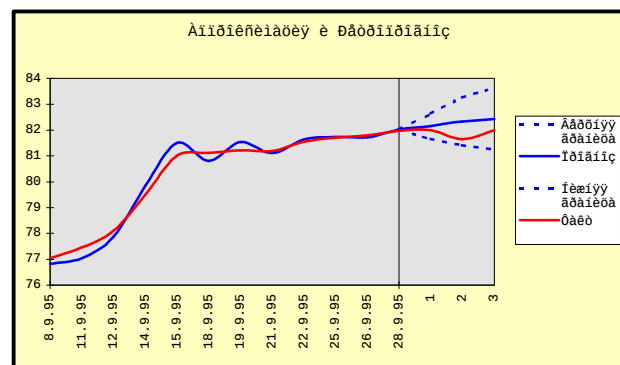
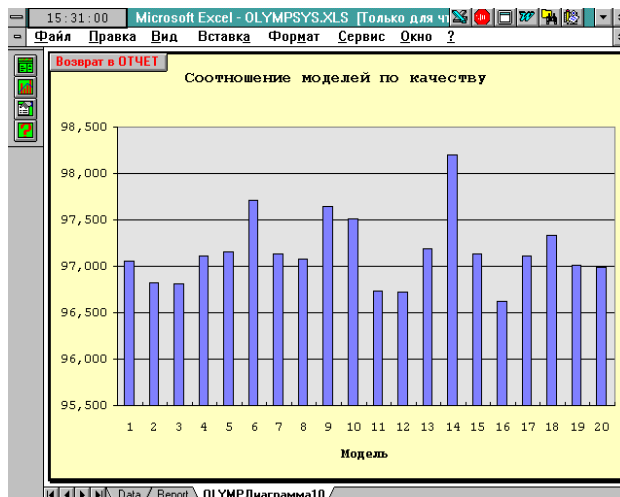
Уравнение значимо с вероятностью 0.95

Таблица ретро-прогнозов (p = 80%)

Упреждение	Факт	Прогноз	Нижняя граница	Верхняя граница	Абс. отклонение	Отн. отклонение
1	82,000	82,153	81,663	82,643	-0,153	-0,186
2	81,650	82,334	81,409	83,259	-0,684	-0,838
3	82,000	82,435	81,251	83,619	-0,435	-0,531

Таблица характеристик ретро-прогнозов

Характеристика	Абсолют. значение	Относит. значение(%)
Среднее значение	-0,424	-0,519
Среднеквадратическое откл.	0,217	0,266
Средний модуль ошибки	0,424	0,519
Максимальное отклонение	-0,153	-0,186
Минимальное отклонение	-0,684	-0,838



После этого появится заставка, внутри которой содержится сообщение о выполнении программой соответствующих этапов. Результаты обработки ото-

бражаются в протоколе "Модели временного ряда" (см. выше).

Таблица **"Моделей кривых роста"** содержит:

- математический вид функции с числовыми значениями параметров;
- значение критерия точности, в качестве которого выступает сумма квадратов отклонений расчетных данных от фактических;
- значение коэффициента эластичности.

В последней строке таблицы содержится модель, которая является лучшей (имеет минимальное значение критерия). В случае если количество заказанных моделей больше трех, на основе показателя "Критерий" строится график.

Таблица **"Характеристики базы моделей"** содержит:

- математический вид моделей с числовыми значениями параметров;
- значение интегрированного критерия адекватности;
- значение интегрированного критерия точности;
- значение интегрированного критерия качества.

База моделей, представленная в таблице, содержит не более 20 моделей, включающих в себя одну лучшую по критерию точности из построенных моделей кривых роста, а также заказанные адаптивные модели, модели Бокса-Дженкинса и модели ОЛИМП. Если рассчитываются более 20 моделей, то худшие из них вытесняются из базы моделей.

Для адаптивных моделей Брауна и Хольта в скобках приводятся значения параметров адаптации, найденных в результате оптимизации. Для авторегрессионной модели $AR(p,d)$ указывается порядок авторегрессии и разностного оператора.

Модели Бокса-Дженкинса идентифицированы как $ARIS(p,d,q)$ $((p,d,q)X(ps,ds,qs)$ для сезонных процессов), а модели ОЛИМП как (p,q) $((p,q)X(ps,qs)$ для сезонных процессов). Числовые значения параметров этих моделей можно увидеть в таблице параметров.

Методика оценки адекватности, точности и качества модели описана в разделе "Качество модели". В последней строке таблицы содержится модель, которая является лучшей по критерию качества. В нашем случае почти все модели имеют высокое качество, что особенно заметно на трех графиках к данной таблице, но чуть лучше других оказалась модель ОЛИМП (2,4).

В случае если количество заказанных моделей больше трех, на основе трех последних показателей строятся соответствующие графики.

Таблица **"Параметры моделей"** содержит:

- вид или наименование модели;
- числовые оценки параметров модели;
- веса моделей (только для обобщенного прогноза).

В таблицу из базы моделей включаются либо одна лучшая модель (по критерию качества), либо несколько (до четырех) моделей, если было заказано построение обобщенной модели. В каждой графе,

имеющей идентификатор "а" с порядковым номером, содержится значение коэффициентов моделей, причем для параметрических моделей сначала приводятся значения всех "р" коэффициентов авторегрессии, а затем "q" коэффициентов скользящего среднего. Если возникает необходимость узнать более точное значение коэффициентов модели, чем с тремя знаками в дробной части, достаточно установить курсор на клетку, содержащую значение интересующего параметра, и посмотреть в строку формул.

Примечание. В случае формирования обобщенного прогноза, после параметров моделей формируется графа "Вес моделей". Весовые коэффициенты в сумме дают единицу и показывают степень влияния каждой модели на обобщенный прогноз.

Таблица **"Характеристика остатков"** содержит наименование показателя и его числовое значение. В последней строке приведен вывод относительно значимости построенной модели. Если модель незначима или имеет низкое значение критерия качества (его значение в этом случае окрашивается в красный цвет), использовать ее для прогнозирования нельзя! В нашем примере построенная лучшая модель является значимой и имеет высокий уровень качества 98.2.

Таблица остатков содержит пять граф:

- имя переменной (в статистических таблицах она обычно выражается датой или порядковым номером наблюдений);
- фактическое значение (исходные данные в Вашей таблице);
- расчетное значение по модели;
- абсолютное отклонение расчетных уровней от фактических (факт минус расчет);
- относительное отклонение (факт минус расчет, деленное на фактическое значение и умноженное на 100%).

В динамических рядах наиболее важным является правильность отображения исследуемого показателя на последнем участке наблюдения. Поэтому, в качестве дополнительной информации для оценки точности модели, целесообразно привлекать величину ошибок аппроксимации на конце периода наблюдения. В нашем примере пять последних уровней относительных отклонений существенно меньше средней относительной ошибки аппроксимации, составляющей 0.28%. Для лучшего восприятия результатов, для наблюдений, которые отличаются от расчетных значений более чем на 10%, величина относительной ошибки выделяется красным цветом.

Для двух последних граф таблицы сформированы графики.

Таблица **"Ретропрогноз"** содержит:

- номер наблюдения на периоде прогнозирования;
- точечное прогнозное значение, полученное по модели;
- нижнюю и верхнюю оценку доверительного интервала прогноза;
- абсолютное отклонение расчетных данных от фактических значений;

•относительное отклонение (в %) расчетных данных от фактических значений.

Ретропрогноз вместе со значениями таблицы остатков отражаются на графике "Аппроксимация и ретропрогноз".

На графике ретропрогноза видно, что расчетные значения оказались достаточно близки к фактическим, причем все фактические данные попали в доверительный интервал прогноза. В этой связи воспользуемся построенной моделью для получения реального прогноза.

Таблица "Характеристики ретропрогноза" содержит наименование показателя и его числовое значение, вычисленное для абсолютных и относительных отклонений. Все характеристики (среднее значение, СКО, среднее по модулю, максимальное и минимальное отклонение) вычисляются аналогично характеристикам периода аппроксимации с той лишь разницей, что в качестве исходных данных берется отклонение прогнозных значений от фактических.

Итак, построенная модель имеет достаточно высокие статистические свойства на участке аппроксимации и ретропрогноза, и можно перейти к получению на ее основе прогнозных оценок на период после 9.09.95. Для этого повторим все действия описанные выше вплоть до выхода в диалоговое окно "Построение модели и прогнозирование". Далее в первых трех классах моделей уберем признак выбора "крестик" и нажмем кнопку "Параметры-4", соответствующую методу ОЛИМП. После этого появится диалоговое окно "Параметры ОЛИМП".

В нем сделаем установки параметров, которые соответствуют структуре построенной в режиме ретропрогнозирования лучшей модели: $p=2$, $q=4$ (нижняя и верхняя граница у этих параметров одинакова). В блоке "Тип прогноза" установим переключатель на "Прогноз вперед". Период прогнозирования определяется потребностями пользователя. Его первоначальное значение можно изменить при помощи рядом стоящего спинера. При задании этого параметра и использовании результатов следует помнить, что все экстраполяционные модели исходят из предположения о сохранении сложившихся закономерностей на некотором участке будущего, однако его справедливость при большом горизонте прогнозирования в большинстве случаев сомнительна. Поэтому не увлекайтесь дальними прогнозами!

Вероятность свершения прогноза зададим в 80 процентов, а остальные параметры оставим без изменения. После нажатия кнопки "Вычислить" через непродолжительное время появится протокол результатов вычислений, приведенный ниже. В этом

протоколе, в отличие от приведенного выше, нет таблиц характеристик ретропрогноза и параметров кривых роста, но есть таблица "Прогнозные оценки", которая содержит: номер наблюдения на периоде прогнозирования; точечное прогнозное значение, полученное по модели; нижнюю и верхнюю оценку доверительного интервала прогноза. Эти значения вместе со значениями таблицы остатков отражаются на графике "Аппроксимация и прогноз", вид которого идентичен аналогичному графику, полученному в режиме ретропрогноза.

Протокол прогнозирования по лучшей модели

A	B	C	D
---	---	---	---

Модели временного ряда - 22014

Характеристики базы моделей

Модель	Адекватность	Точность	Качество
ОЛИМП(2,4)	53,614	97,421	86,469

Лучшая модель ОЛИМП(2,4)

Параметры моделей

Модель	a1	a2	a3	a4	a5	a6
ОЛИМП(2,4)	-0,487	-0,412	-0,999	0,357	-0,119	-0,212

Таблица остатков

Номер	Факт	Расчет	Ошибка абс.	Ошибка относит.
8.9.95	77,050	76,881	0,169	0,220
11.9.95	77,460	77,491	-0,031	-0,040
12.9.95	78,100	78,006	0,094	0,120
14.9.95	79,500	80,536	-1,036	-1,304
15.9.95	81,000	81,443	-0,443	-0,547
18.9.95	81,110	81,293	-0,183	-0,225
19.9.95	81,215	80,793	0,422	0,520
21.9.95	81,200	81,029	0,171	0,211
22.9.95	81,550	81,390	0,160	0,196
25.9.95	81,700	81,716	-0,016	-0,019
26.9.95	81,805	81,628	0,177	0,217
28.9.95	81,970	82,044	-0,074	-0,090
29.9.95	82,000	81,647	0,353	0,431
2.10.95	81,650	81,358	0,292	0,358
3.10.95	82,000	82,024	-0,024	-0,029

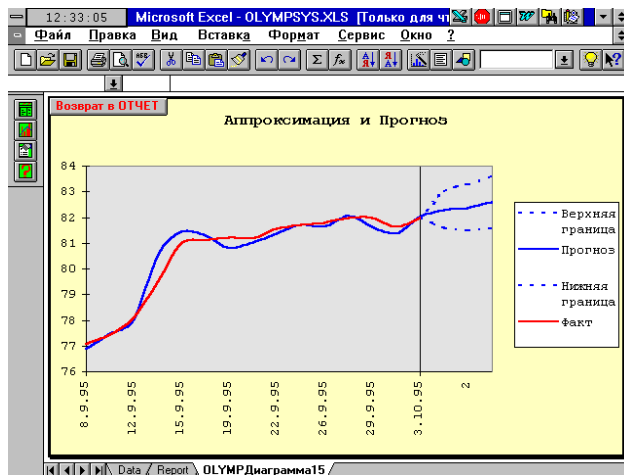
Характеристики остатков

Характеристика	Значение
Среднее значение	0,002
Оценка дисперсии	0,122
Оценка приведенной дисперсии	0,140
Средний модуль остатков	0,243
Относительная ошибка	0,302
Критерий Дарбина-Уотсона	1,442
Коэффициент детерминации	1,000
F - значение (n1 = 1, n2 = 13)	100000,000
Критерий адекватности	53,614
Критерий точности	97,421
Критерий качества	86,469

Уравнение значимо с вероятностью 0.95

Таблица прогнозов (p = 80%)

Упреждение	Прогноз	Нижняя граница	Верхняя граница
1	82,296	81,598	82,995
2	82,381	81,503	83,258
3	82,599	81,569	83,628



Корреляционный анализ

Корреляционный анализ предназначен для измерения степени связи двух или более переменных, отбор факторов, оказывающих наиболее существенное влияние на результирующий признак и обнаружения ранее неизвестных причинных связей.

Для иллюстрации возможностей корреляционного анализа исследуем взаимосвязь показателя "Индекс потребительских цен" от других приведенных во второй таблице показателей (см. фрагмент таблицы данных на рисунке). Активизируйте вторую таблицу демонстрационного файла Demo.xls и выделите с третьей по десятую строки. Таким образом, выделенный блок кроме числовых данных будет содержать наименования переменных, но не будет иметь наименования таблицы и наблюдений.

	A	B	C	D	E	F	G	H
1	Основные показатели развития							
2		01.93	02.93	03.93	04.93	05.93	06.93	07.93
3	Индекс потребительских цен	125.8	124.7	120.1	123.2	118.5	119.9	122.0
4	Индекс оптовых цен	132.0	130.0	123.0	124.0	119.0	117.0	129.0
5	Объем промышленного производства, млрд. руб.	3 000.0	3 860.0	5 340.0	5 900.0	6 200.0	7 200.0	9 100.0
6	M0, млрд. руб.	1 902.5	2 276.6	2 559.0	3 308.7	4 019.6	5 112.7	6 260.9
7	M2, млрд. руб.	8 447.3	9 299.5	10 889.5	13 400.2	15 960.8	18 960.0	22 203.8
8	Средневзвешенный обменный курс, руб./\$	488.3	569.5	663.7	766.1	911.7	1 077.7	1 020.0
9	Импорт (C.I.F.), млрд. руб.	700.3	1 059.3	1 259.1	1 580.4	2 319.2	3 147.0	1 727.8
10	Экспорт (F.O.B.), млрд. руб.	1 215.0	1 647.6	2 158.5	2 426.2	3 304.7	4 205.3	3 455.6

Установите курсор на пункт меню "СтатЭкс", нажмите "мышку" и в появившемся меню выберите пункт "Корреляция".

В появившемся диалоговом окне "Установка блока данных" в блоке "Ориентация таблицы" установите параметр "По строкам", отмените указатель наименований таблицы и наблюдений, а указатель наличия наименования переменных оставьте без изменения, (см. рисунок) после чего воспользуйтесь кнопкой "Установить".

Установка блока данных

Шаблон блока данных:

	10.01	10.07	11.02	11.31
Цена	10.01	10.07	11.02	11.31
Объем	2000.00	2020.00	2000.00	2100.00
Прибыль	485.03	496.25	500.01	520.35
%	98.01	98.91	98.07	97.90

Ориентация таблицы:

☐ по колонкам

☒ по строкам

Наличие наименований:

☐ таблицы

☐ наблюдений

☒ переменных

☒ Запрос шаблона данных на каждом шаге обработки

Помощь Отказ Установить

Программа сообщит, что последние столбцы выделенных строк являются пустыми, и запросит разрешения не использовать их. Ответьте на этот запрос утвердительно, после чего появится диалоговое окно "Корреляционный анализ".

Корреляционный анализ

Список доступных переменных:

- 4 M0, млрд. руб.
- 5 M2, млрд. руб.
- 6 Средневзвешенный обменный курс
- 8 Экспорт (F.O.B.), млрд. руб.

Добавить >>

Указанную Переу Все

Удалить <<

Указанную Переу Все

Список выбранных переменных:

- 1 Индекс потребительских цен
- 2 Индекс оптовых цен
- 3 Объем пром. производства, млрд. руб.
- 7 Импорт (C.I.F.), млрд. руб.

Структура отчета:

☒ Парные коэффициенты корреляции

☒ Частные коэффициенты корреляции

☒ Множественные коэффициенты корреляции

☒ Построение графиков

Вероятностный уровень значимости взаимосвязи: 90

Помощь Выход Вычислить

Для оценки взаимосвязи индекса оптовых цен с другими факторами необходимо определить набор не менее чем из двух переменных. Для этого в "Списке доступных переменных" установите курсор на строку "Индекс потребительских цен" и щелкните "мышкой". Указанная переменная будет перенесена в правое окно "Список выбранных переменных". Таким же способом закажите переменные "Индекс оптовых цен", "Объем промышленного производства", "Импорт".

Структура отчета зависит от сделанного выбора пунктов: в отчет попадут таблицы, рядом с которыми стоит символ "крестик". Суть всех коэффициентов и формулы расчета описаны во втором разделе.

Для начала расчетов нажмите на кнопку "Вычислить".

После этого появится заставка, внутри которой содержится сообщение о выполнении программой соответствующих этапов. Результаты обработки отображаются в Протоколе "Корреляционный анализ", который содержит 5 таблиц:

- Парные коэффициенты корреляции;
- Оптимальные лаги корреляции;
- Парные корреляции на оптимальных лагах;
- Частные коэффициенты корреляции;
- Множественные коэффициенты корреляции.

На основе этих коэффициентов с различных позиций можно судить о степени взаимосвязи переменных. При этом позиция коэффициентов (номер

строки и номер столбца) определяет между какими переменными оценивается связь. Номер строки характеризует первую переменную, а номер столбца - вторую.

В таблицах парных и частных коэффициентов корреляции красным цветом выделены незначимые коэффициенты, значение которых по абсолютной величине меньше порогового уровня. Пороговый уровень значимости, указанный в нижней строке этих таблиц, вычисляется исходя из заданного пользователем **уровня отбора**. Для первой переменной строится график парного и частного коэффициентов с остальными переменными.

Таблица "Оптимальные лаги корреляции" содержит оценку лага (сдвига, запаздывания) при котором связь между двумя переменными максимальна. Таблица "Парные корреляции на оптимальных лагах" отличается от таблицы парных корреляций, если между хотя бы одной из пар переменных оптимальный лаг больше нуля. Оценки этих таблиц важны при построении регрессионных моделей. В нашем примере все оптимальные корреляционные лаги равны нулю и, следовательно, построение регрессионных зависимостей следует проводить по исходным данным. Если провести корреляционный анализ по данным ГКО можно обнаружить, что наиболее сильная связь между многими выпусками проявляется на лаге 1, т.е. со сдвигом в один торговый день. Следовательно, в этом случае один выпуск ГКО по отношению к другому можно рассматривать как событие-предвестник, индикатор рынка.

В случае невозможности вычисления этих коэффициентов, например, из-за вырожденности матриц, в протоколе появляется соответствующее сообщение.

Примечание. Применение корреляционного анализа более эффективно при обработке не исходных временных рядов, а их остаточных компонент, полученных в результате исключения тренда из исходных данных. Поэтому перед запуском корреляционного анализа целесообразно для каждого исследуемого показателя воспользоваться режимом прогнозирования.

А	В	С	Д	Е
---	---	---	---	---

Протокол корреляционного анализа

Матрица парных корреляций

Переменная	Индекс потребительских цен	Индекс оптовых цен	Объем промышленного производства, млрд. руб.	Импорт (С.И.Ф.), млрд. руб.
Индекс потребительских цен	1.000	0.926	-0.562	-0.461
Индекс оптовых цен	0.926	1.000	-0.681	-0.627
Объем промышленного производства, млрд. руб.	-0.562	-0.681	1.000	0.965

млрд. руб.

Импорт (С.И.Ф.), млрд. руб.	-0.461	-0.627	0.965	1.000
-----------------------------	--------	--------	-------	-------

Критическое значение на уровне 90 при 2 степенях свободы = +0.2717

Матрица максимальных корреляций

Переменная	Индекс потребительских цен	Индекс оптовых цен	Объем промышленного производства, млрд. руб.	Импорт (С.И.Ф.), млрд. руб.
Индекс потребительских цен	1.000	0.926	-0.562	-0.461
Индекс оптовых цен	0.926	1.000	-0.681	-0.627
Объем промышленного производства, млрд. руб.	-0.562	-0.681	1.000	0.965
Импорт (С.И.Ф.), млрд. руб.	-0.461	-0.627	0.965	1.000

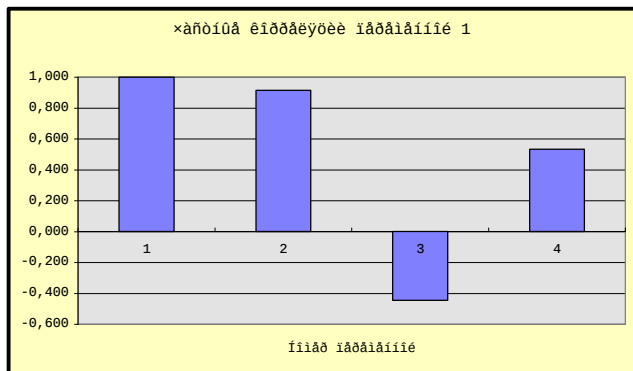
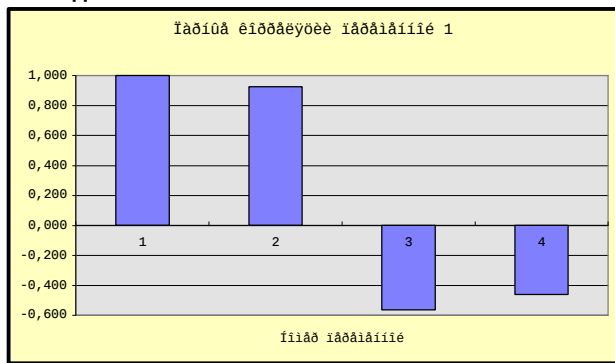
Матрица оптимальных лагов

Переменная	Индекс потребительских цен	Индекс оптовых цен	Объем промышленного производства, млрд. руб.	Импорт (С.И.Ф.), млрд. руб.
Индекс потребительских цен	0.000	0.000	0.000	0.000
Индекс оптовых цен	0.000	0.000	0.000	0.000
Объем промышленного производства, млрд. руб.	0.000	0.000	0.000	0.000
Импорт (С.И.Ф.), млрд. руб.	0.000	0.000	0.000	0.000

Матрица частных корреляций

Переменная	Индекс потребительских цен	Индекс оптовых цен	Объем промышленного производства, млрд. руб.	Импорт (С.И.Ф.), млрд. руб.
Индекс потребительских цен	1.000	0.915	-0.444	0.534
Индекс оптовых цен	0.915	1.000	0.272	-0.434
Объем промышленного производства, млрд. руб.	-0.444	0.272	1.000	0.952
Импорт (С.И.Ф.), млрд. руб.	0.534	-0.434	0.952	1.000

Критическое значение на уровне 90 при 4 степенях свободы = +0.2725



Регрессионный анализ

Регрессионный анализ предназначен для исследования зависимости исследуемой переменной от различных факторов и отображения их взаимосвязи в форме регрессионной модели.

Для иллюстрации возможностей регрессионного анализа исследуем взаимосвязь показателя "Индекс потребительских цен" от показателей, которые были задействованы в корреляционном анализе. Построим модель линейной, а так же парной регрессии и получим на их основе точечный и интервальный прогнозы с доверительной вероятностью 85%. Для оценки эффективности математического аппарата воспользуемся техникой ретропрогнозирования, оставив в качестве проверочного участка три последних наблюдения.

Сначала активизируйте вторую таблицу (Лист 2) демонстрационного файла Demo.XLS (см. рисунок в разделе 1.7.4). Выделите в ней строки со второй по десятую включительно, т.е. в блоке данных будут присутствовать наименования наблюдений (даты) и наименования переменных (показателей), но не будет наименования таблицы.

Далее установите курсор на пункт меню "СтатЭкс", нажмите кнопку "мышь" и в появившемся меню выберите пункт "Регрессия".

В диалоговом окне "Установка блока данных" (см. рисунок в разделе 1.7.4) в блоке "Ориентация таблицы" установите параметр "По строкам", отмените указатель наличия наименования таблицы, а указатель наличия наименования наблюдений и переменных оставьте без изменения. Далее нажмите кнопку "Установить". Программа сообщит, что последние столбцы (231) выделенных строк явля-

ются пустыми, и запросит разрешения не использовать их. Ответьте на этот запрос утвердительно, после чего появится диалоговое окно "Регрессионный анализ".

Модель регрессии включает зависимую и, по крайней мере, одну независимую переменную. Поэтому, прежде всего, необходимо определить набор не менее чем из двух переменных. Выберите те же переменные, что и при проведении корреляционного анализа. Для этого в "Списке доступных переменных" установите курсор на строку "Индекс потребительских цен" и щелкните "мышью". Указанная переменная будет перенесена в правое окно "Список выбранных переменных". Таким же способом закажите переменные "Индекс оптовых цен", "Объем промышленного производства", "Импорт".

Далее в списке выбранных переменных следует указать **зависимую переменную**. Для этого нажмите кнопку "Выбор", в появившемся окне установите курсор на "Индекс потребительских цен", а затем нажмите кнопку "Выход".

В поле "Зависимая переменная" вместо фразы "зависимая переменная не определена" появится наименование выбранной переменной. Ее наименование удаляется из списка выбранных переменных. Все остальные выбранные переменные считаются независимыми, т.е. **факторами**.

На основе регрессионной модели можно изучать взаимосвязь исследуемых показателей. Этот режим включен по умолчанию. На ее основе можно получать также прогнозные оценки. Для этого закажите пункт "Ретропрогнозирование по модели" и выберите пункт "ретропрогноз". Параметр "Период ретропрогнозирования" определяет количество точек проверочной последовательности, которая должна быть не менее двух точек. При помощи спинера установим цифру "3".

Примечание. При отключенном параметре "Ретропрогноз" программа находится в режиме прогнозирования. Чтобы получить правильный результат сначала сформируйте прогнозы всех включенных в нее факторов, потом запишите результаты в таблицу исходных данных (у зависимой переменной поставьте нулевые числовые значения). Затем снова войдите в режим регрессии и включите режим прогнозирования, установив символ "крестик" рядом с параметром "Прогнозирование по модели", и отключите параметр "Ретропрогноз".

Вероятность свершения прогноза первоначально задана уровнем 80%. Измените ее величину на 85% при помощи рядом стоящего спинера.

Полнота выдачи результатов вычислений определяется перечнем заказанных в блоке "Структура отчета" таблиц. Количество заказанных таблиц практически не влияет на время формирования отчета. Закажите **все** перечисленные таблицы (поставить рядом с ними символ выбора - крестик) и графики.

На последнем шаге следует определить **вид модели** регрессии и способ ее построения, т.е. выбрать одну из четырех возможностей:

- линейная множественная регрессия;
- пошаговая регрессия;
- гребневая регрессия;
- парная регрессия.

Примечание. При неопределенных или неправильно заданных параметрах, программа выдает соответствующее сообщение и ждет исправления ошибочного действия.

Выберите первый пункт - "Линейная регрессия" и нажмите на кнопку "Вычислить". Программа запросит разрешение подключить к выбранным факторам параметр "Время". Ответьте на этот запрос отрицательно.

Через непродолжительное время появится протокол регрессионного анализа, который представлен ниже. Структура отчета видна из приводимых ниже меню, доступ к которым осуществляется при помощи пиктограмм таблицы "Report".

Таблицы ОЛИМП:СтатЭксперта

Оценки коэффициентов линейной регрессии

Таблица остатков

Характеристики остатков

Таблица ретро-прогнозов (p = 85%)

Таблица характеристик ретро-прогнозов

Графики ОЛИМП:СтатЭксперта

Абсолютная ошибка

Относительная ошибка %

Аппроксимация и Ретропрогноз

Таблица "**Оценки коэффициентов линейной регрессии**" содержит:

- наименование переменной (для A0 имя "Св. член");
- коэффициент (параметры модели);

- среднеквадратическое отклонение;
- t-значение (статистика Стьюдента для оценки значимости коэффициентов модели);
- нижняя и верхняя оценка значений коэффициентов;

- коэффициенты эластичности, бета-коэффициенты и дельта-коэффициенты (см. раздел "Коэффициенты экономической интерпретации").

В последней строке приведено критическое значение t-критерия, ниже которого коэффициенты модели считаются незначимыми. В таблице они выделены красным цветом. Как правило, такую модель можно упростить, отбросив незначимые факторы.

"Таблица остатков", таблицы "Характеристика остатков", "Ретропрогноз", "Характеристики ретро-прогноза", а также графики расхождений фактических и расчетных значений полностью совпадают с аналогичными таблицами, которые описаны в разделе "Временные ряды".

На рисунке приведены результаты ретропрогнозирования, из которых следует, что, несмотря на заметное изменение поведения исследуемого показателя, его значение не вышло за пределы интервального прогноза.

Линейная регрессия:

Зависимая переменная - Индекс потребительских цен

Оценки коэффициентов линейной регрессии

Переменная	Коэффициент	Ср. кв. откл.	t-знач.	Нижняя оценка	Верхняя оценка	Эластичность	Бета-коэффициент	Дельта-коэффициент
Св. член	34,361	27,928	1,230	10,254	58,468	0,000	0,000	0,000
Индекс оптовых цен	0,713	0,208	3,434	0,534	0,893	0,720	0,736	-4,169
Объем пром. про-ва.	0,000	0,000	-2,277	-0,001	0,000	-0,063	-0,676	3,826
Импорт (С.И.Ф.)	0,001	0,001	0,877	0,000	0,002	0,039	-0,237	1,343

Критические значения t-распределения при 17 степенях свободы (p=80) = +0.863

Таблица остатков

Номер	Факт	Расчет	Ошибка абс	Ошибка относит.
1.1.93	125,800	128,140	-2,340	-1,860
1.2.93	124,700	126,790	-2,090	-1,676
1.3.93	120,100	121,434	-1,334	-1,110
1.4.93	123,200	122,301	0,899	0,730
1.5.93	118,500	119,489	-0,989	-0,834
1.6.93	119,900	118,638	1,262	1,052
1.7.93	122,000	124,745	-2,745	-2,250
1.8.93	126,000	122,806	3,194	2,535
1.9.93	123,000	118,716	4,284	3,483
1.10.93	120,000	116,734	3,266	2,722

Номер	Факт	Расчет	Ошибка абс	Ошибка относит.
1.11.93	116,000	114,708	1,292	1,114
1.12.93	113,000	111,988	1,012	0,896
1.1.94	117,900	115,796	2,104	1,785
1.2.94	110,700	113,753	-3,053	-2,758
1.3.94	107,400	109,240	-1,840	-1,714
1.4.94	108,800	108,532	0,268	0,247
1.5.94	108,100	107,307	0,793	0,734
1.6.94	105,000	108,157	-3,157	-3,007
1.7.94	105,000	107,010	-2,010	-1,914
1.8.94	104,000	104,483	-0,483	-0,465
1.9.94	107,200	106,032	1,168	1,090

Характеристики остатков

Характеристика	Значение
Среднее значение	-0,024
Оценка дисперсии	4,661
Оценка приведенной дисперсии	5,758
Средний модуль остатков	1,885
Относительная ошибка	1,618
Критерий Дарбина-Уотсона	1,305
Коэффициент детерминации	1,000
F - значение (n1 = 3, n2 = 17)	16288,395
Критерий адекватности	38,050
Критерий точности	86,545
Критерий качества	74,421

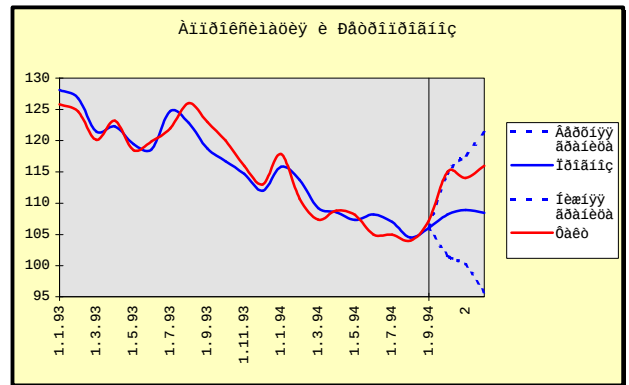
Уравнение значимо с вероятностью 0.95

Таблица ретро-прогнозов (p = 85%)

Упре- жде- ние	Факт	Прог- ноз	Ниж- няя границ а	Верх- няя границ а	Абс. Откло- нение	Отн. Откло- нение
1	115,00 0	108,18 2	101,64 9	114,71 5	6,818	5,929
2	114,00 0	108,90 5	100,14 5	117,66 5	5,095	4,469
3	116,00 0	108,44 8	95,649	121,24 8	7,552	6,510

Таблица характеристик ретро-прогнозов

Характеристика	Абсолют. значение	Отно- сит. значе- ние%
Среднее значение	6,488	5,636
Среднеквадратическое откло- нение	1,030	0,859
Средний модуль ошибки	6,488	5,636
Максимальное отклонение	7,552	6,510
Минимальное отклонение	5,095	4,469



Теперь построим модель парной регрессии, характеризующей зависимость индекса потребительских цен от индекса оптовых цен. Для этого повторите все действия, описанные в данном разделе выше, но в диалоговом окне “Регрессионный анализ” выберите лишь два указанных показателя.

Окажитесь от режима прогнозирования, а в блоке “Вид регрессии” выберите пункт “Парная регрессия”. Конкретный перечень моделей можно сформировать после нажатия кнопки “Вычислить”. Для **парной регрессии** появляется диалоговое окно выбора 16 регрессионных зависимостей.

Для выбора всех регрессионных моделей в окне “Список доступных переменных” в группе “Добавить” нажмите на кнопку “Все”. Для выхода из диалога “Формирование набора моделей” воспользуйтесь кнопкой “Выход”. (Если Вы выйдете из диалога, не выбрав ни одной модели, программа воспользуется списком моделей, которые установлены по умолчанию.).

Через непродолжительное время появится протокол, представленный ниже.

Для **парной регрессии** все таблицы, за исключением первой, аналогичны описанному выше. Вместо нее выдается таблица "Функции парной регрессии", которая содержит:

- вид регрессионной модели с оцененными параметрами;
- значение критерия точности, по минимальной величине которого определяется лучшая модель;
- значение коэффициента эластичности.

В последней строке приводится лучшая модель. Если было заказано построение более трех моделей, на основе графы "Критерий" строится соответствующий график (см. рисунок ниже).

Парная регрессия.

Y = Индекс потребительских цен

X = Индекс оптовых цен

Таблица функций парной регрессии

Функция	Критерий	Эластичность
$Y(X)=+18.621+0.831 \cdot X$	7,447	0,839
$Y(X)=-352.893+7.152 \cdot X - 0.027 \cdot X^2$	5,523	0,914
$Y(X)=0+0/X$	Нет	0,000
$Y(X)=1./(+0.016-0.0000 \cdot X)$	8,729	-0,839
$Y(X)=1/(0+0 \cdot \exp(-X))$	Нет	0,000
$Y(X)=+49.772 \cdot \exp(+0.007 \cdot X)$	8,020	0,840
$Y(X)= -353.172+98.545 \cdot \ln(X)$	6,871	0,852
$Y(X)= (+1.498) \cdot (+1.069) \cdot X^{(+1.000)} \cdot (X^X)$	7,790	0,000
$Y(X)= (+49.772) \cdot (+1.007) \cdot X$	8,020	0,840
$Y(X)= +588.319-2248.424/\ln(X)$	6,655	-0,858
$Y(X)= (+1.974) \cdot X^{(+0.855)}$	7,372	0,855
$Y(X)= -1424.050-11.314 \cdot X+266.385 \cdot \sqrt{X}$	299,465	0,898
$Y(X)= X/(+0.858+0.001 \cdot X)$	7,270	0,850
$Y(X)=0 \cdot \exp(0/X)$	Нет	0,000
$Y(X)= +84.364+0.0000 \cdot X^2$	8,102	0,006
$Y(X)= -352.893+7.152 \cdot X^2 - 0.027 \cdot X^3$	5,523	0,914

Выбрана функция $Y(X)=-352.893+7.152 \cdot X - 0.027 \cdot X^2$

Таблица остатков

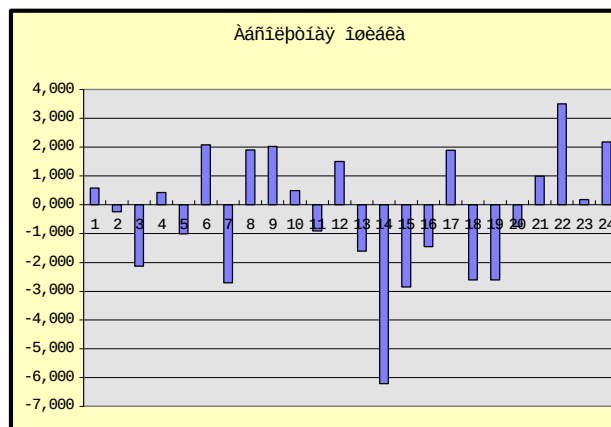
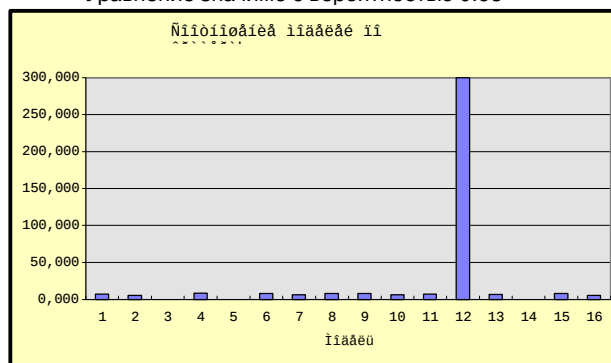
Номер	Факт	Расчет
1. 1. 93	125,800	125,224
1. 2. 93	124,700	124,933
1. 3. 93	120,100	122,230
1. 4. 93	123,200	122,777
1. 5. 93	118,500	119,509
1. 6. 93	119,900	117,827
1. 7. 93	122,000	124,707
1. 8. 93	126,000	124,096
1. 9. 93	123,000	120,976
1. 10. 93	120,000	119,509
1. 11. 93	116,000	116,906
1. 12. 93	113,000	111,499
1. 1. 94	117,900	119,509
1. 2. 94	110,700	116,906
1. 3. 94	107,400	110,257

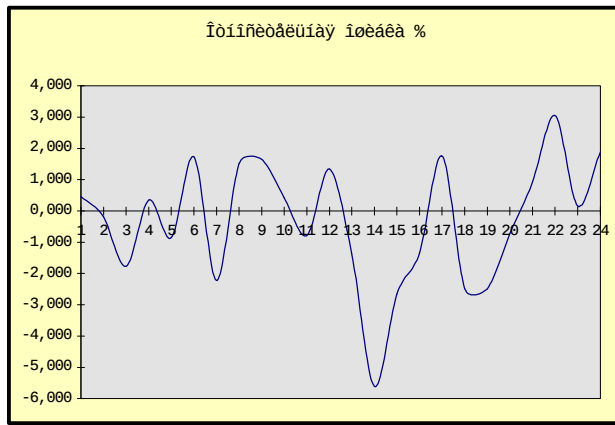
Номер	Факт	Расчет
1. 4. 94	108,800	110,257
1. 5. 94	108,100	106,210
1. 6. 94	105,000	107,612
1. 7. 94	105,000	107,612
1. 8. 94	104,000	104,754
1. 9. 94	107,200	106,210
1. 10. 94	115,000	111,499
1. 11. 94	114,000	113,822
1. 12. 94	116,000	113,822

Характеристики остатков

Характеристика	Значение
Среднее значение	-0,307
Оценка дисперсии	4,738
Оценка приведенной дисперсии	5,523
Средний модуль остатков	1,784
Относительная ошибка аппроксимации	1,569
Критерий Дарбина-Уотсона	1,551
Коэффициент детерминации	1,000
F - значение (n1 = 2, n2 = 21)	29066,275
Критерий адекватности	66,799
Критерий точности	86,937
Критерий качества	81,902

Уравнение значимо с вероятностью 0.95



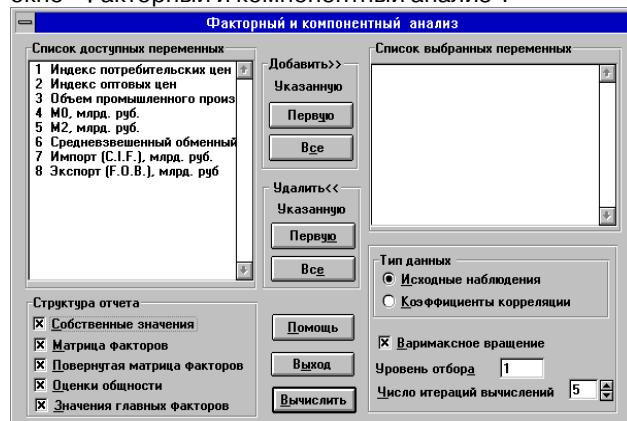


Факторный и компонентный анализ

Компонентный анализ является методом определения структурной зависимости между случайными переменными. В результате его использования получается сжатое описание малого объема, несущее почти всю информацию, содержащуюся в исходных данных. **Факторный анализ** является более общим методом преобразования исходных переменных по сравнению с компонентным анализом.

Проведем факторный и компонентный анализ всех показателей, содержащихся во второй таблице. Для этого активизируйте вторую таблицу демонстрационного файла Demo.xls и выделите весь блок данных, который кроме числовых значений будет содержать наименования таблицы, наблюдений и переменных. Установите курсор на пункт меню "СтатЭкс", нажмите "мышку" и в появившемся меню выберите пункт "Факторный анализ".

В диалоговом окне "Установка блока данных" в блоке "Ориентация таблицы" установите параметр "По строкам", а указатель наличия наименования переменных таблицы, наблюдений оставьте без изменения, далее воспользуйтесь кнопкой "Установить", после чего появится диалоговое окно "Факторный и компонентный анализ".



Выберите все доступные переменные. Для этого в группе "Добавить" воспользуйтесь кнопкой "Все", после чего наименования всех переменных из левого списка доступных переменных переместятся в правый список выбранных переменных. В поле "Тип данных" оставьте установку "Исходные наблюдения", а в поле "Варимаксное вращение" оставьте символ выбора (крестик). Установите **уровень отбора** рав-

ным 0.2, а **число итераций** - 5. Структура отчета зависит от сделанного выбора пунктов: в отчет попадут таблицы, рядом с которыми стоит символ "крестик". В блоке "Структура отчета" оставьте заказ всех таблиц отчета. Для начала вычислений нажмите кнопку "Вычислить".

После этого появится заставка, внутри которой содержится сообщение о выполнении программой соответствующих этапов. Результаты обработки отображаются в Протоколе "Факторный анализ", который может включать 5 таблиц. Суть всех коэффициентов, представленных в таблицах, и формулы расчета описаны во втором разделе данного документа.

А	В	С
---	---	---

Протокол факторного анализа

Оценки собственных значений

Фактор	Собств. значение	Накопленное отношение
1	6.759	0.875
2	0.962	1.000
3	0.070	1.009
4	0.007	1.010
5	0.000	1.010
6	-0.011	1.008
7	-0.028	1.005
8	-0.036	1.000

Отобрано факторов 2, количество итераций = 5

Матрица факторных нагрузок

Переменная	1_	2_
Индекс потребительских цен	0.700	0.700
Индекс оптовых цен	0.795	0.525
Объем промышленного производства.	-0.976	0.177
М0, млрд. руб.	-0.985	0.075
М2, млрд. руб.	-0.981	0.030
Средневзвешенный обменный курс, руб./\$	-0.986	0.128
Импорт (С.И.Ф.), млрд. руб.	-0.933	0.262
Экспорт (Ф.О.В.), млрд. руб.	-0.954	0.272

Матрица повернутых факторных нагрузок

Переменная	1_	2_
Индекс потребительских цен	0.000	0.958
Индекс оптовых цен	0.419	0.856
Объем промышленного производства,.	-0.931	-0.343
М0, млрд. руб.	-0.887	-0.435
М2, млрд. руб.	-0.860	-0.472
Средневзвешенный обменный курс, руб./\$	-0.914	-0.390
Импорт (С.И.Ф.), млрд. руб.	-0.937	0.000
Экспорт (Ф.О.В.), млрд. руб.	-0.960	0.000

Оценки общностей

Переменная	Общность	Специфичность
Индекс потребительских цен	0.979	0.093
Индекс оптовых цен	0.907	0.047

Переменная	Общность	Специфичность
Объем промышленного производства, .	0.985	0.017
M0, млрд. руб.	0.975	0.007
M2, млрд. руб.	0.962	0.004
Средневзвешенный обменный курс, руб./\$	0.988	0.012
Импорт (С.И.Ф.), млрд. руб.	0.939	0.029
Экспорт (F.O.B.), млрд. руб.	0.984	0.031

Значения факторов

Наблюдение	1_	2_
1.1.93	0.775	1.564
1.2.93	0.756	1.369
1.3.93	0.986	0.485
1.4.93	0.731	0.895
1.5.93	0.840	0.181
1.6.93	0.599	0.282
1.7.93	0.382	1.192
1.8.93	0.198	1.463
1.9.93	0.274	0.844
1.10.93	0.345	0.427
1.11.93	0.396	-0.083
1.12.93	0.391	-0.594
1.1.94	0.290	0.216
1.2.94	0.391	-0.586
1.3.94	0.339	-1.175
1.4.94	0.223	-1.070
1.5.94	0.160	-1.290
1.6.94	-0.053	-1.417
1.7.94	0.001	-1.513
1.8.94	-0.019	-1.750
1.9.94	-0.666	-1.160
1.10.94	-1.827	0.213
1.11.94	-2.279	0.466
1.12.94	-3.234	1.040

Кластерный анализ

Кластерный анализ предназначен для разбиения наблюдений на однородные группы (кластеры). Проиллюстрируем возможности его проведения на примере обработки данных о крупнейших банках РФ. Предположим, что нас интересует классификация банков, выполненная по всем 8 имеющимся показателям. Для этого активизируйте третью таблицу демонстрационного файла Demo.xls.

Крупнейшие банки России (на 1.01.96)						
	Собст. капитал	Прив. ресурсы (исключая бан.	Прив. ресурсы банков	Средняя задолжен. (рубля)	Средняя задолжен. (валюта)	Средств в банках
2	Название банка					
3	Сбербанк РФ	6099889	61067484	2693209	12245033	3506209
4	Внешторгбанк	3764350	9635474	11392991	31323	4629442
5	ОИЭКСИМбанк	1370596	8090357	4822784	3138452	7159184
6	Инкомбанк	1052618	7597614	1952306	1749462	3329294
7	Мосбизнесбанк	640478	3121342	874474	1177193	1793015
8	Роскредит	557032	1669324	2897602	809268	1360040
9	МФК	1120847	1800688	7513267	317719	4393383
10	Интерпирал	996003	1257359	2909163	772401	2055202
11	Моск. индустриальный	572385	2305445	11424	1234517	222695
12	Менатеп	625027	2155954	621010	3049381	2921867
13	Чинкомбанк	469496	6009677	155655	1381584	406261
14	Промстройбанк	487892	3617644	1056781	1009361	613389
15	Банк "С-Петербург"	139342	80577	60209	465825	142151
16	Межд. Моск. банк	731741	3578276	2079819	1883	1067004
17	Столичный банк сбереж.	867715	1810844	1148687	368887	1413719
18	Тверьичинверсалбанк	262228	1766648	2046477	760565	423275
19	Автобанк	615759	2337665	2262400	517422	798883

Выделите весь блок данных (A1: i29), который кроме числовых значений будет содержать наименования таблицы, наблюдений и переменных. Установите курсор на пункт меню "СтатЭкс", нажмите "мышку" и в появившемся меню выберите пункт "Кластерный анализ".

В диалоговом окне "Установка блока данных" в блоке "Ориентация таблицы" установите параметр "По колонкам", а указатель наличия наименования таблицы, переменных и наблюдений оставьте помеченными символом выбора (крестик). Далее воспользуйтесь кнопкой "Установить", после чего появится диалоговое окно "Кластерный анализ".

Выберите все доступные переменные. Для этого в группе "Добавить" воспользуйтесь кнопкой "Все", после чего наименования всех переменных из левого списка доступных переменных переместятся в правый список выбранных переменных. Выберите тип метрики "Евклидова", а в поле "Четкость классификации" при помощи спинера установите значение 3. В блоке "Структура отчета" оставьте заказ таблицы "Результаты классификации" и построение графика кластеризации.

Для начала вычислений нажмите кнопку "Вычислить". После этого появится заставка, внутри которой содержится сообщение о выполнении программой соответствующих этапов.

Результаты обработки отображаются в Протоколе "Кластерный анализ", который включает одну таблицу и один график. В таблице отражены наименование объектов классификации, номер группы, в которую они отнесены, расстояние объединения и координаты центра группы. Из приведенных результатов следует, что 27 банков разделены на 8 групп, причем банки Менатеп, Внешторгбанк и Сбербанк РФ по совокупности исследуемых показателей стоят особняком от остальных и каждый из них выделен в самостоятельную группу. В 5 остальных групп входят от 3 до 7 банков.

Для графической интерпретации результатов кластерного анализа приводится график расположения исходных объектов в пространстве первых двух главных компонент. При этом объекты, попавшие в один кластер, отображаются одним цветом.

Примечание. Иногда объекты из разных кластеров расположены столь близко, что может создаться иллюзия о неправильной классификации. Это связано с тем, что классификация проводится по большому числу переменных, а график строится по

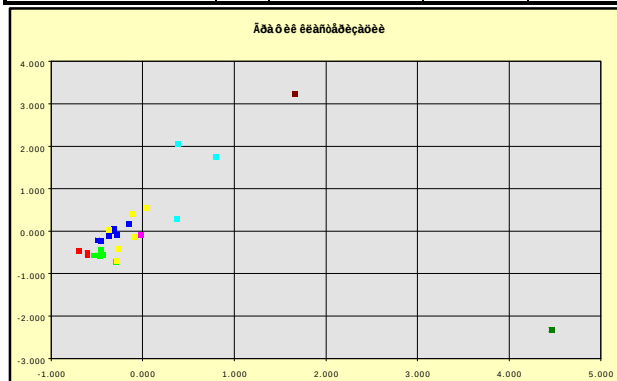
двум координатам (хотя и отражающим основные особенности данных), поэтому некоторые расхождения между результатом классификации и графическим отображением неизбежны.

A	B	C	D	E
---	---	---	---	---

Протокол кластерного анализа

Результат кластеризации

	Кластер	Расстояние объединения	Координата X	Координата Y
Банк "С-Петербург"	1	403105.594	-0.606	-0.504
Енисей	1	403105.594	-0.607	-0.541
Якиманка	1	403105.594	-0.699	-0.462
Башкредитбанк	2	658934.875	-0.442	-0.559
Возрождение	2	658934.875	-0.461	-0.433
Мосстройэкономбанк	2	658934.875	-0.533	-0.566
Промышл.-строительный	2	658934.875	-0.468	-0.577
Моск. индустриальный	2	658934.875	-0.300	-0.726
Альфа-банк	3	930044.938	-0.496	-0.208
Межкомбанк	3	930044.938	-0.454	-0.223
Столичный банк сбереж.	3	930044.938	-0.353	0.018
Автобанк	3	930044.938	-0.289	-0.074
Роскредит	3	930044.938	-0.152	0.181
Токобанк	3	930044.938	-0.316	0.051
Мостбанк	3	930044.938	-0.374	-0.106
Мосбизнесбанк	4	1750768.500	-0.092	-0.131
Промстройбанк	4	1750768.500	-0.268	-0.402
Уникомбанк	4	1750768.500	-0.284	-0.702
Тверь-универсалбанк	4	1750768.500	-0.375	0.035
Империял	4	1750768.500	0.044	0.563
Межд. Моск. банк	4	1750768.500	-0.119	0.416
Менатеп	5	3204395.750	-0.023	-0.079
ОНЭКСИМбанк	6	5386877.500	0.794	1.751
Инкомбанк	6	5386877.500	0.373	0.305
МФК	6	5386877.500	0.385	2.062
Внешторгбанк	7	13160116.000	1.656	3.233
Сбербанк РФ	8	61038228.000	4.458	-2.321



Частотный анализ

Вместе с долговременными изменениями, во временных рядах часто появляются более или менее регулярные колебания. Эти изменения наблюдаемых значений могут быть строго периодическими или близкими к таковым и оцениваться в частотном аспекте.

Для иллюстрации возможностей частотного анализа продолжим исследование показателя "Индекс потребительских цен", приведенного во второй таблице демонстрационного примера. Для этого активизируйте вторую таблицу файла Demo.xls и, не выделяя блок, активизируйте пункт "СтатЭкс", а в нем - "Частотный анализ". Таким образом, информационный блок, определенный "по умолчанию" кроме числовых данных будет содержать наименования таблицы, переменных и наблюдений. Поэтому, в появившемся диалоговом окне "Установка блока данных" в блоке "Ориентация таблицы" установите параметр "По строкам", а указатель наименований таблицы, наблюдений, и переменных оставьте без изменения, после чего воспользуйтесь кнопкой "Установить".

Для получения корректных результатов частотного анализа необходимо иметь достаточно большой объем данных (не менее 50 наблюдений), из которых предварительно исключена тенденция. Программа автоматически контролирует выполнение этих требований и в случае их нарушения выдает соответствующее сообщение, но не запрещает дальнейший анализ. В нашем примере объем данных меньше установленного критического уровня. Однако, на запрос о необходимости продолжения вычислений ответьте утвердительно.

После этого появится диалоговое окно частотного анализа.



Для выявления наличия и устойчивости периода колебаний обычно используется следующий аппарат частотного анализа:

- Гармонический анализ
- Спектральный анализ
- Частотная фильтрация
- Кросс-спектральный анализ.

В "Списке переменных" диалогового окна отображаются наименования показателей, взятых из базы данных в соответствии с указанным ранее шаблоном. Текущей является переменная, которая выделена светящейся строкой курсора.

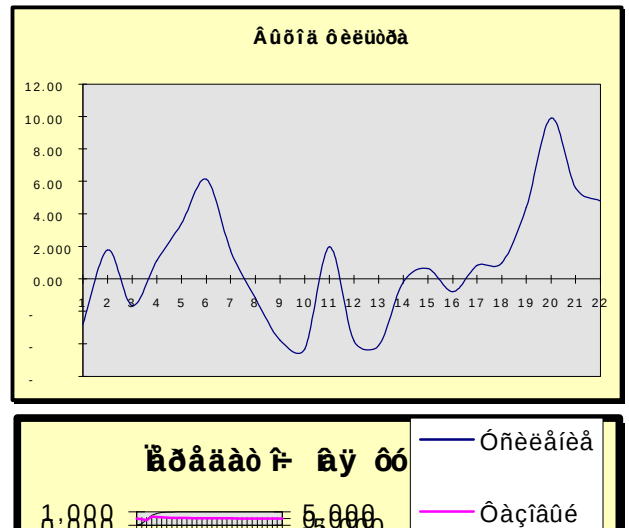
При перемещении курсора (движения "мышки") по списку переменных, справа отображается

Наблюдение	Значение
1.3.93	-2.802
1.4.93	1.790
1.5.93	-1.653
1.6.93	1.156
1.7.93	3.424
1.8.93	6.125
1.9.93	1.761
1.10.93	-1.179
1.11.93	-3.771
1.12.93	-4.329
1.1.94	1.985
1.2.94	-3.812
1.3.94	-4.101
1.4.94	-0.195
1.5.94	0.645
1.6.94	-0.789
1.7.94	0.836
1.8.94	1.010
1.9.94	4.404
1.10.94	9.909
1.11.94	5.599
1.12.94	4.786

Частота отсечки = +0.1000

Передаточная функция

Номер	Частота
1	0.000
2	0.033
3	0.067
4	0.100
5	0.133
6	0.167
7	0.200
8	0.233
9	0.267
10	0.300
11	0.333
12	0.367
13	0.400
14	0.433
15	0.467
16	0.500
17	0.533
18	0.567
19	0.600
20	0.633
21	0.667
22	0.700
23	0.733
24	0.767
25	0.800
26	0.833
27	0.867
28	0.900
29	0.933
30	0.967



Спектральный анализ

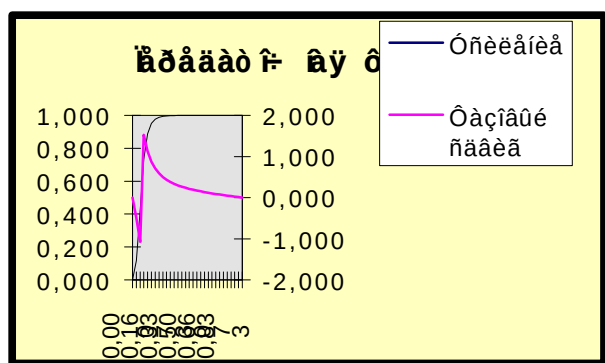
Для проведения спектрального анализа необходимо выполнить все действия, описанные в 1.7.8, а в блоке "Вид анализа" диалогового окна "Частотный анализ" следует выбрать пункт "Спектральный анализ". После этого появится диалоговое окно "Спектральный анализ", позволяющее выбрать один из трех типов окна и заказать графическое отображение результатов.

В блоке "Тип окна" выберите пункт "Окно Парзена" и закажите построение графиков.

Для начала расчетов следует нажать на кнопку "Вычислить", а для отказа от вычислений - нажать кнопку "Выход".

Протокол содержит одну таблицу, в которой для каждой частоты выдается соответствующее ей значение спектрограммы и спектральной плотности. В нижней строке таблицы представлено наименование типа использованного в расчетах окна.

Содержащиеся в таблице показатели отображаются на графике.



A	B	C	D
---	---	---	---

Спектральный анализ переменной Индекс потребительских цен

Характеристики спектра мощности

Номер	Частота	Спектрограмма	Спектральная плотность
1	0.000	1085.200	179615.531
2	0.033	1032.260	162821.531
3	0.067	882.427	119741.742
4	0.100	660.885	68112.641
5	0.133	404.045	26325.404
6	0.167	152.445	4287.958
7	0.200	57.104	334.900
8	0.233	197.845	5186.542
9	0.267	258.198	9317.914
10	0.300	242.416	8476.536
11	0.333	168.751	4273.881
12	0.367	65.008	719.639
13	0.400	43.723	186.333
14	0.433	122.648	1892.096
15	0.467	160.678	3384.813
16	0.500	151.994	3070.988
17	0.533	103.632	1428.615
18	0.567	32.107	134.919
19	0.600	43.174	277.147
20	0.633	100.281	1414.302
21	0.667	126.344	2213.405
22	0.700	116.204	1846.586
23	0.733	74.586	738.869
24	0.767	14.600	27.059
25	0.800	47.916	374.027
26	0.833	94.296	1340.624
27	0.867	112.978	1888.386
28	0.900	99.385	1451.974
29	0.933	57.797	493.540
30	0.967	1.800	6.675

Тип окна - Парзена

Кросс-спектральный анализ

Для проведения кросс-спектрального анализа (т.е. анализа взаимного спектра) необходимо выполнить все действия, описанные в 1.7.8, а в блоке "Вид анализа" диалогового окна "Частотный анализ" следует выбрать пункт "Кросс-спектральный анализ". После этого появится одноименное диалоговое окно.



Кросс-спектральный анализ проводится по двум переменным, первая из которых выбрана на предыдущем шаге в диалоге Частотного анализа. Ее наименование отображается в блоке "Первая выбранная переменная".

В "Списке переменных" диалогового окна отображаются наименования показателей, взятых из базы данных в соответствии с указанным ранее шаблоном. Текущей является переменная, которая выделена светящейся строкой курсора. Выберите в качестве второй переменной "Индекс оптовых цен".

При перемещении курсора (движения "мышки") по списку переменных, справа отображается график их значений, который можно увеличить до размера полного окна. Для этого достаточно нажать "мышку", когда ее указатель установлен на графике. При желании график можно распечатать или перенести в другие программы (например, в текстовый редактор) стандартными средствами Excel. Возврат в диалоговое окно происходит при повторном нажатии мышки или клавиши Esc.

В блоке "Тип окна" выберите пункт "Окно Тьюки".

Для начала расчетов следует нажать на кнопку "Вычислить", а для отказа от вычислений - нажать кнопку "Выход". Последней возможностью целесообразно воспользоваться в случае, если на графике или в списке переменных Вы заметили несоответствие со своими исходными данными.

Протокол содержит одну таблицу, в которой для каждой частоты выдается соответствующее ей значение когерентности, фазового сдвига и коэффициента усиления. В нижней строке таблицы представлено наименование использованного в расчетах окна. Содержащиеся в таблице показатели отражены на графике.

A	B	C	D	E
---	---	---	---	---

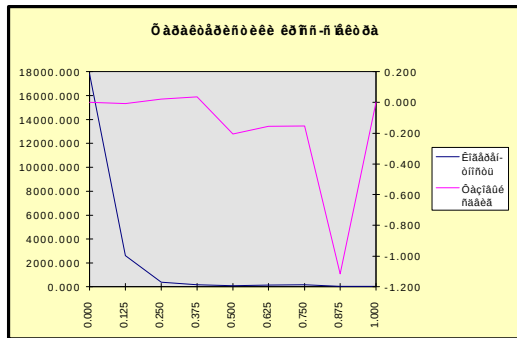
Кросс-спектральный анализ переменных
Индекс потребительских цен и Индекс оптовых цен

Характеристики кросс-спектра

Номер	Частота	Когерентность	Фазовый сдвиг	Усиление
-------	---------	---------------	---------------	----------

Номер	Частота	Когерентность	Фазовый сдвиг	Усиление
1	0.000	17813.813	0.000	85252.445
2	0.125	2607.673	-0.007	2135.664
3	0.250	387.910	0.022	110.815
4	0.375	160.673	0.037	33.410
5	0.500	93.293	-0.205	13.950
6	0.625	143.087	-0.155	25.818
7	0.750	166.281	-0.154	28.913
8	0.875	17.463	-1.117	1.047
9	1.000	52.800	0.000	2.207

Тип окна -Тьюки

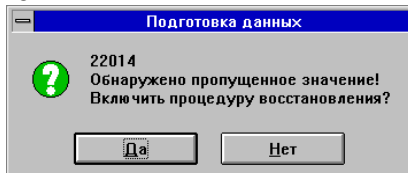


Восстановление пропущенных наблюдений

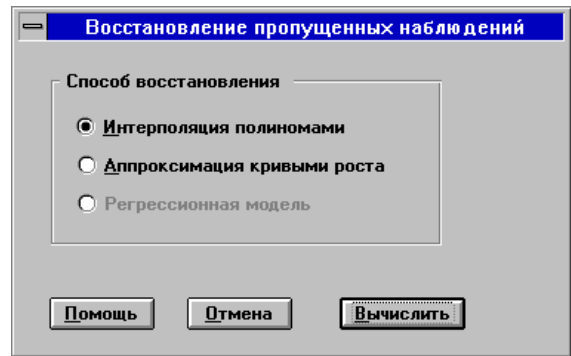
Выделенные пользователем данные могут содержать нечисловые и пропущенные наблюдения (клетки, не содержащие никакой информации). При обнаружении нечислового значения программа выдает соответствующее сообщение и прекращает работу.

Пропущенные наблюдения должны быть предварительно устранены: либо удалены из информационного блока при помощи операций "Удалить строку", "Удалить столбец", либо заменены числовыми значениями "вручную" или при помощи математических методов программным путем.

Предположим, что по каким-либо причинам нам не удалось получить сведения о цене на вторичных торгах по ГКО 22014 за 11.09.95. Следовательно, в таблице 1 книги Demo.xls в клетке Н9 вместо числа 78.10 должен стоять пробел. Для воссоздания этой ситуации сотрите указанное число (установите курсор на клетку Н9, активизируйте пункт "Правка", далее - "Очистить" - "Все"). Тогда в режиме "Обработка временных рядов" (см. раздел 1.7.2) после выбора ГКО 22014 и нажатия кнопки "Вычислить" появится сообщение:



Если на этот запрос ответить отрицательно, программа приостанавливает свою работу и возвратит пользователя в таблицу Data. В случае положительного ответа появляется диалоговое окно "Восстановление пропущенных наблюдений" с предложением выбрать способ восстановления.

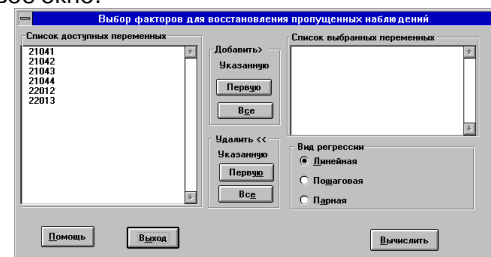


Для продолжения работы необходимо выбрать конкретный способ восстановления данных (последний способ доступен не всегда!) и нажать кнопку "Вычислить". Нажатие кнопки "Отмена" приведет к прекращению обработки данных.

В соответствии с первым способом пропущенные наблюдения восстанавливаются путем использования стандартных **интерполяционных** формул, в которых число точек интерполяционного полинома принимается равным 3. Восстановленное значение примерно равно 78.30.

Восстановление наблюдений при помощи **аппроксимации** сводится к подбору кривой роста, наилучшим образом отображающей особенности изменения исследуемого показателя. При этом пропущенные наблюдения заменяются на расчетные значения аппроксимирующей модели. Набор аппроксимирующих моделей определяется пользователем (см. режим "Обработка временных рядов. Прогнозирование"). Если выбрать все доступные модели, то программа, построив эти модели, автоматически выберет лучшую из них и выдаст восстановленное значение примерно равно 78.79.

Восстановление наблюдений при помощи **регрессионных зависимостей** заключается в построении модели взаимосвязи показателя, который содержит пропущенные наблюдения, с одним или несколькими показателями, не содержащими пропущенные наблюдения. Этот способ не доступен при обработке одномерных данных. Поэтому для его иллюстрации воспользуйтесь режимом корреляционного анализа, техника реализации которого описана в разделе 1.7.5. После выбора данных (вся таблица 1 книги Demo.xls), активизации режима "Корреляционный анализ", выбора в диалоговом окне "Корреляционный анализ" всех переменных и нажатия кнопки "Вычислить" появится следующее диалоговое окно:



Метод построения модели (парная регрессия, линейная, множественная) и круг показателей определяется пользователем таким же образом, как и в

режиме построения регрессионных моделей (см. раздел 1.7.5). Предположим, что у нас есть основания считать существенной зависимость между выпусками 22014 и, например, 21041, но форма этой зависимости не известна. Выберите режим "Парная регрессия", а в нем закажите все доступные модели. Программа, построив эти модели, автоматически выберет лучшую из них и выдаст восстановленное значение примерно равное 78.50.

Все описанные способы дают хорошие результаты при обработке временных рядов наблюдений. При обработке пространственных данных, таких, например, как данные о крупнейших банках РФ, приведенных в третьем листе демонстрационного файла, целесообразнее использовать последний способ восстановления пропущенных наблюдений.

Примечание 1. Восстановление пропущенных наблюдений при всех описанных выше способах осуществляется лишь в том случае, если по каждому показателю количество пропущенных наблюдений не превышает половины всего объема наблюдений, и выделенный блок данных содержит наименования наблюдений и переменных.

Примечание 2. Клетки рабочей таблицы "Data", содержащие пропущенные значения, выделяются красным цветом. Результаты восстановления пропущенных значений заносятся только в рабочую таблицу.

От авторов

ОЛИМП:СтатЭксперт (Windows 3.11, Windows 95) - это очередной шаг в развитии пакетов статистической обработки серии ОЛИМП. Пакет позволяет организовать полный цикл исследований по статистическому анализу и прогнозированию данных, начиная с их ввода, проверки, визуализации и заканчивая проведением расчетов и анализом результатов на основе широкого набора современных методов прикладной статистики, многие из которых реализованы по оригинальным алгоритмам.

Программа "ОЛИМП:СтатЭксперт" разработана сотрудниками Департамента информационных технологий ТОО "Росэкспертиза". Мы надеемся, что она существенно облегчит Вам решение задач анализа и прогнозирования финансово - экономических показателей, т.к. в ней реализованы наиболее эффективные вычислительные методы, хорошо зарекомендовавшие себя в практической работе, и оригинальные авторские алгоритмы решения указанных задач.

В отличие от своих прототипов "ОЛИМП" и "ОРАКУЛ-2", успешно функционирующих в среде MS DOS в течение многих лет, данная программа использует интерфейс Microsoft Excel. Использование этой популярной электронной таблицы в качестве оболочки вычислительных модулей, с одной стороны, полностью решает вопросы организации и ведения информационной базы, связи с другими прикладными программами и приложениями Windows, а с другой стороны, создает максимально удобную и привычную среду для многих пользователей. Для работы с программой достаточно иметь минимальный опыт работы с Windows и Excel. Для обеспечения высокого быстродействия все вычислительные модули реализованы на языке Си++ в виде библиотек динамической загрузки. Такая конфигурация программы принципиально отличает "ОЛИМП:СтатЭксперт" от аналогичных программ.

Документация по программе "ОЛИМП:СтатЭксперт" состоит из двух частей.

В первой части подробно описана установка программы на компьютер, основные понятия системы и техника решения задач анализа данных на примерах, являющихся элементами поставляемой системы. Вы можете воспроизвести расчеты, описанные в документации, на своем компьютере и убедиться в работоспособности своего экземпляра программы. При изучении примеров целесообразно также активно пользоваться встроенными в программу средствами помощи, что позволит Вам максимально быстро изучить и правильно использовать все вычислительные возможности программы.

Во второй части приведено описание математических методов, реализованных в программе. К ней имеет смысл обратиться в том случае, если Вас заинтересовала теоретическая основа реализованных в программе методов.

Список литературы

- 1.Андерсон Т. Статистический анализ временных рядов. - М.:Мир, 1976.
- 2.Бокс Дж., Дженкинс Г. Анализ временных рядов, прогноз и управление. - М.:Мир, 1974.
- 3.Богачев В.В. Моделирование нестационарных процессов авторегрессионными моделями. В сб.: Моделирование экономических процессов М.:МЭСИ, 1989.
- 4.Богачев В.В. Альтернативные представления адаптивных моделей прогнозирования. В сб.: Модели и методы экономической кибернетики М.:МЭСИ, 1988.
- 5.Богачев В.В. Моделирование нестационарных временных рядов с помощью несимметричного фильтра. - Материалы семинара по планированию эксперимента М.:МДНТП, 1985.
- 6.Богачев В.В. Метод прогнозирования одномерного временного ряда. В сб.: Применение методов вычислительной математики в экономических исследованиях М.:МЭСИ, 1985.
- 7.Горчаков А.А., Орлова И.В. Компьютерные экономико-математические модели. - М.: ЮНИТИ, 1995.
- 8.Горчаков А.А., Половников В.А. Тенденции развития кредитного рынка России. - Ж. "Банковское дело" N 3, 1995.
- 9.Горчаков А.А., Половников В.А. Обзор процентных ставок рынка ссудного капитала. - Ж. "Банковское дело" N 5, 1995.
- 10.Горчаков А.А., Половников В.А. Проблемные вопросы прогнозирования финансово-экономических показателей. - Бюллетень Информационного центра "Рейтинг" N 3, 1995.
- 11.Горчаков А.А., Половников В.А. Финансовая математика. - М.:ВЗФЭИ, 1995.
- 12.Горчаков А.А. и др. Методы экономико-математического моделирования и прогнозирования в новых условиях хозяйствования. - М.: ВЗФЭИ, 1991.
- 13.Горчаков А.А., Половников В.А. Одномерные методы и модели экономического прогнозирования. - Ташкент: Мехмат, 1990.
- 14.Горчаков А.А., Сенилов Д.А. Обработка временных рядов наблюдений с использованием ПЭВМ. - М.:МЭСИ, 1988.
- 15.Горчаков А.А. Использование статистических методов исследования временных рядов для оценки результативности новых условий хозяйствования. В кн.: Математико-статистический анализ в экономике и производстве. - М.:МЭСИ, 1988.
- 16.Горчаков А.А. Методические указания по работе на ПЭВМ с программой моделирования и прогнозирования социально-экономических процессов "ОРАКУЛ". - М.: АгроНИИТЭИПП, 1988.
- 17.Горчаков А.А., Половников В.А. Прогнозирование экономических процессов с сильной сезонностью. В кн.:

Моделирование экономических процессов. - М.: МЭСИ, 1987.

18. Горчаков А.А. Использование адаптивных методов в АСУ. В кн.: Организационно-методические аспекты автоматизированной обработки информации. - М.: МЭСИ, 1986.

19. Горчаков А.А. и др. Оценка точности и адекватности моделей экономического прогнозирования. В кн.: Математические методы моделирования экономических процессов. - М.: МЭСИ, 1986.

20. Горчаков А.А. Прогнозирование сезонных процессов на основе метода Тейла-Вейджа. В кн.: Проблемные вопросы конструирования АСУ. - М.: МЭСИ, 1985.

21. Горчаков А.А. Использование адаптивных моделей для краткосрочного экономического прогнозирования. В сб.: Актуальные вопросы экономико-статистических исследований. - М.: ЦСУ СССР, Городская конференция молодых ученых, 1982.

22. Горчаков А.А., Половников В.А. Использование метода адаптивной фильтрации в прогнозировании бытовых услуг. В сб.: Применение методов вычислительной математики в экономике. - М.: МЭСИ, 1981.

23. Горчаков А.А. Прогнозирование сезонных колебаний. В сб.: Вопросы эффективности и качества в системах управления народным хозяйством. - М.: МЭСИ, 1980.

24. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. - М.: Статистика, 1979.

25. Половников В.А., Горчаков А.А. Модели и методы экономического прогнозирования. - М.: МЭСИ, 1980.

26. Харман Г.Г. Современный факторный анализ. - М.: Статистика, 1972.

27. Четыркин Е.М. Статистические методы прогнозирования. - М.: Статистика, 1977.

МНОГОФАКТОРНАЯ МОДЕЛЬ ДЮПОНА АНАЛИЗА ЭФФЕКТИВНОСТИ ДЕЯТЕЛЬНОСТИ ПРЕДПРИЯТИЯ.

Аудиторская и консалтинговая фирма
«Росэкспертиза»

Современная деятельность любой компании направлена на получение прибыли в условиях острой конкурентной рыночной среды.

Для эффективного управления такой компанией необходимо внедрение и ведение на предприятии эффективной системы управленческого учета и финансового менеджмента.

В этих условиях финансовый анализ и управление рентабельностью и деловой активностью предприятия приобретает решающее значение.

Система или модель анализа корпорации Дюпон (в дальнейшем модель Дюпон) является согласованной методикой анализа способности предприятия эффективно генерировать прибыль, реинвестировать ее, наращивать обороты.

Эта модель позволяет исследовать динамику ключевых показателей эффективности деятельности предприятия:

- Оборачиваемости чистых активов
- Рентабельности чистых активов (RONA)
- Рентабельности собственного капитала (ROE)
- Коэффициента экономического роста компании (SG)
- Средневзвешенной стоимости капитала (WACC).

«Изюминкой» модели является расщепление этих комплексных показателей на факторы, их составляющие. Это позволяет определить первопричины и взаимосвязь изменения комплексных показателей эффективности деятельности предприятия. Схемы факторного анализа ключевых показателей позволяют наглядно проследить влияние первичных факторов на формирование комплексных показателей, причины их изменения.

Анализ динамики финансового состояния предприятия может проводиться в сопоставимых ценах с учетом индексов инфляции.

Эта система анализа полностью реализована в компьютерной программе финансового анализа Олимп: ФинЭксперт, созданной специалистами Департамента информационных технологий Росэкспертизы.

Схема проведения анализа по модели Дюпон, рассматриваемой в этой статье, такова.

Сначала формируется финансовая отчетность в форме GAAP.

Затем последовательно анализируются ключевые показатели модели Дюпон и факторы, их составляющие.

Динамика этих показателей исследуется, используя возможность их факторного расщепления. Это позволяет сделать содержательные выводы о причинах их изменения. И наконец, построение наглядных схем факторного анализа позволяет проследить первопричины изменения финального показателя.

1. Формирование финансовой отчетности в форме GAAP

Основные международные принципы формирования финансовой отчетности. Общие принципы и терминология

Структура отчета о прибылях и убытках (Income Statement or Profit and Losses Statement)

Данный документ отражает все доходы и расходы, понесенные компанией за отчетный период. Структурно это может быть представлено в следующем виде:

Объем продаж	Прямые произв. затраты
--------------	------------------------

Đã í â ñ ò đ. ĩ ð ò -
á û ë ü

ВСЕГО КАПИТАЛА

ЧИСТЫЕ АКТИВЫ

Подготовка российской финансовой отчетности к анализу в соответствии с международными принципами бухгалтерского учета (GAAP)

Одним из основополагающих принципов бухгалтерского учета в большинстве стран является принцип отражения учетных объектов по ценам приобретения. В условиях стабильных цен применение этого принципа вполне оправдано. Однако в периоды достаточно высокой инфляции отчетность, основанная на первоначальных стоимостных оценках, может давать искаженное представление о финансовом состоянии и результатах деятельности предприятия, более того, при анализе финансового состояния предприятия за несколько отчетных периодов, которые характеризуются высоким уровнем инфляции, для получения достоверной информации необходимо данные финансовой отчетности за все анализируемые периоды привести к уровню цен выбранного базового периода.

В системе анализа Дюпон в программе Фин-эксперт реализована возможность пересчета данных финансовой отчетности в цены последнего анализируемого временного периода. Для этого используются квартальные индексы инфляции.

Структура баланса GAAP

Актив	Пассив
Наименования позиций	Наименования позиций
ОБОРОТНЫЕ АКТИВЫ	КРАТКОСРОЧНЫЕ ПАССИВЫ
Денежные средства	Заемные средства
Краткосрочные финансовые вложения	Кредиторская задолженность
Дебиторская задолженность	Расчеты по дивидендам
Запасы	Прочие краткосрочные пассивы
Прочие текущие активы	ДОЛГОСРОЧНЫЕ ПАССИВЫ
ВНЕОБОРОТНЫЕ АКТИВЫ	Заемные средства
Нематериальные активы	Прочие долгосрочные пассивы
Основные средства	КАПИТАЛ И РЕЗЕРВЫ
Незавершенное строительство	Уставный капитал
Долгосрочные финансовые вложения	Накопленный капитал

Актив	Пассив
Выше вложения	
Прочие внеоборотные активы	Целевые финансирование и поступления
	Нераспределенная прибыль отчетного года- Убытки
БАЛАНС	БАЛАНС

Примечание: * Валюта баланса уменьшена на величину убытков.

Следующим этапом подготовки отчетности предприятия к анализу является подготовка трансформированного баланса, построенного в соответствии с принципами GAAP. Подготовка трансформированного баланса не является самоцелью. Благодаря укрупненной структуре статей трансформированного баланса во многом облегчается не только анализ финансовых показателей деятельности компании, но и сам баланс приобретает более удобный для анализа вид. Такие же или приблизительно такие же форматы балансового отчета используются на западных предприятиях. Основой для подготовки трансформированного баланса на российских предприятиях на сегодняшний день является обычный бухгалтерский баланс, введенный с первого полугодия 1998 года.