Министерство образования и науки Российской Федерации

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

Кафедра информационных систем и программной инженерии

Модели и методы планирования экспериментов, обработки экспериментальных данных

Методические указания к лабораторным работам

Составители:
Р. И. МАКАРОВ
Е. Р. ХОРОШЕВА

Владимир 2013

УДК 519.242 (076)
ББК 22.1я7
М74

Рецензент
Доктор технических наук, профессор
Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых
В. Н. Ланцов

Печатается по решению редакционно-издательского совета ВлГУ

М74 **Модели** и методы планирования экспериментов, обработки экспериментальных данных : метод. указания к лаб. работам / сост.: Р. И. Макаров, Е. Р. Хорошева ; Владим. гос. ун-т имени Александра Григорьевича и Николая Григорьевича Столетовых. – Владимир : Изд-во ВлГУ, 2013 – 61 с.

Знакомят магистрантов с методами построения моделей, планированием экспериментов и обработкой экспериментальных данных.

Лабораторные работы позволяют освоить методы построения моделей сложных систем с использованием регрессионного анализа, нейронных сетей и нечетких множеств; методы анализа временных рядов и их прогнозирование.

Предназначены для подготовки магистров по дисциплинам по выбору по направлениям 230400 «Информационные системы и технологии» и 231000 «Программная инженерия» по профилю «Информационные системы и технологии».

Рекомендованы для формирования профессиональных компетенций в соответствии с ФГОС 3-го поколения.

Ил. 11. Табл. 15. Библиогр.: 5 назв.

УДК 519.242 (076)
ББК 22.1я7

Лабораторная работа № 1
МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

1.1. Цель работы

Освоение методов построения регрессионных моделей на основе обработки статистических данных.

1.2. Теоретические сведения

Регрессионный анализ предназначен для исследования зависимости исследуемой переменной от различных факторов и отображения их взаимосвязи в форме регрессионной модели [1, 2].

В регрессионных моделях зависимая (объясняемая) переменная Y может быть представлена в виде функции $f\,(X_1,\ X_2,\ X_3,\ \dots Xm)$, где $X_1,\ X_2,\ X_3,\ \dots Xm$ – независимые (объясняющие) переменные, или факторы. В качестве зависимой переменной может выступать практически любой показатель, характеризующий, например, функционирование сложной системы, деятельность предприятия или курс ценной бумаги. В зависимости от вида функции $f\,(X_1,\ X_2,\ X_3,\ \dots Xm)$ модели делятся на линейные и нелинейные. В зависимости от количества включенных в модель факторов X модели делятся на однофакторные (парная модель регрессии) и многофакторные (модель множественной регрессии).

Связь между переменной Y и m независимыми факторами можно охарактеризовать функцией регрессии $Y= f\,(X_1,\ X_2,\ X_3,\ \dots Xm)$, которая показывает, каково будет в среднем значение переменной Y_i, если переменные X_i примут конкретные значения.

Данное обстоятельство позволяет использовать модель регрессии не только для анализа, но и для прогнозирования процессов и результатов деятельности предприятий.

Линейная модель множественной регрессии имеет вид:

$$Y_i = \alpha_0 + \alpha_1 x_{i\,1} + \alpha_2 x_{i\,2} + \dots + \alpha_m x_{i\,m} + \varepsilon_i\ ,\ i = 1,\dots n\ . \qquad (1.1)$$

Коэффициент регрессии α_j показывает, на какую величину в среднем изменится результативный признак Y, если переменную x_j увеличить на единицу измерения. Обычно предполагается, что случайная величина ε_i имеет нормальный закон распределения с математическим ожиданием, равным нулю, и с дисперсией σ^2.

Анализ уравнения (1.1) и методика определения параметров становятся более наглядными, а расчетные процедуры существенно упрощаются, если воспользоваться матричной формой записи уравнения

$$Y = X\alpha + \varepsilon, \tag{1.2}$$

где Y – вектор зависимой переменной размерности $n \times 1$, представляющий собой n наблюдений значений y_i; X – матрица n наблюдений независимых переменных X_1, X_2, $X_3, \dots X_m$, размерность матрицы X равна $n \times (m+1)$; α – подлежащий оцениванию вектор неизвестных параметров размерности $(m+1)\times 1$; ε – вектор случайных отклонений (возмущений) размерности $n \times 1$.

Уравнение (1.1) содержит значения неизвестных параметров α_0, α_1, $\alpha_2, \dots, \alpha_m$. Эти величины оцениваются на основе выборочных наблюдений, поэтому полученные расчетные показатели не являются истинными, а представляют собой лишь их статистические оценки. Модель линейной регрессии, в которой вместо истинных значений параметров подставлены их оценки (а именно такие регрессии и применяются на практике), имеет вид

$$Y = X\alpha + e = \hat{y} + e,$$

где a – вектор оценок параметров; e – вектор «оцененных» отклонений регрессии, остатки регрессии $e = Y - X\alpha$; $\hat{y}$ – оценка значений Y, равная $X\alpha$.

Параметры модели множественной регрессии можно оценить с помощью метода наименьших квадратов.

Формула для вычисления параметров регрессионного уравнения имеет вид

$$\alpha = (X^T X)^{-1} X^T Y.$$

Одним из условий регрессионной модели является предположение о линейной независимости объясняющих переменных, т.е. решение задачи возможно лишь тогда, когда столбцы и строки матрицы исходных данных линейно независимы. Это условие выполняется не всегда. Линейная или близкая к ней связь между факторами называется *мультиколлинеарностью* и приводит к линейной зависимости нормальных уравнений, что делает вычисление параметров либо невозможным, либо затрудняет содержательную интерпретацию параметров модели.

Мультиколлинеарность может возникать в силу разных причин. Например, несколько независимых переменных могут иметь общий временной тренд, относительно которого они совершают малые колебания. Считают явление мультиколлинеарности в исходных данных установленным, если коэффициент парной корреляции между двумя переменными больше 0,8. Чтобы избавиться от мультиколлинеарности, в модель включают лишь один из линейно связанных между собой факторов, причем тот, который в большей степени связан с зависимой переменной.

Качество модели регрессии оценивается проверкой:

1) качества всего уравнения регрессии;
2) значимости всего уравнения регрессии;
3) статистической значимости коэффициентов уравнения регрессии;
4) выполнения предпосылок метода наименьших квадратов (МНК).

Для оценки качества модели множественной регрессии вычисляют коэффициент множественной корреляции (индекс корреляции) R и коэффициент детерминации R^2:

$$R^2 = \frac{\textit{объясняемая сумма квадратов}}{\textit{общая сумма квадратов}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n}\varepsilon_i^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2},$$

где y – фактическое значение зависимой переменной; $\hat{y}$ – рассчитанное по уравнению регрессии значение зависимой переменной; $\overline{y}$ – среднее арифметическое значение переменной y.

Чем ближе к единице значение этих характеристик, тем выше качество модели.

В многофакторной регрессии добавление дополнительных объясняющих переменных увеличивает коэффициент детерминации. Следовательно, коэффициент детерминации должен быть скорректирован с учетом числа независимых переменных. Скорректированный R^2, или $\overline{\mathrm{R}}^2$, рассчитывается так:

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-k-1},$$

где n – число наблюдений; k – число независимых переменных.

Проверка значимости модели регрессии

Для проверки значимости модели регрессии используется F-критерий Фишера, вычисляемый по формуле:

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

Если расчетное значение с $\nu_1 = k$ и $\nu_2 = (n - k - 1)$ степенями свободы, где k – количество факторов, включенных в модель, больше табличного при заданном уровне значимости, то модель считается значимой.

Анализ статистической значимости параметров модели

Значимость отдельных коэффициентов регрессии проверяется по t-статистике путем проверки гипотезы о равенстве нулю j-го параметра уравнения (кроме свободного члена)

$$t_{aj} = a_j / S_{aj},$$

где S_{aj} – это стандартное (среднеквадратическое) отклонение коэффициента уравнения регрессии a_j. Величина S_{aj} представляет собой квадратный корень из произведения несмещенной оценки дисперсии S_ε^2 и j-го диагонального элемента матрицы, обратной матрице системы нормальных уравнений.

$$S_{aj} = S_\varepsilon \sqrt{b_{jj}},$$

где b_{jj} – диагональный элемент матрицы $(X^T X)^{-1}$.

Если расчетное значение t-критерия с $(n - k - 1)$ степенями свободы превосходит его табличное значение при заданном уровне значимости, коэффициент регрессии считается значимым. В противном случае фактор, соответствующий этому коэффициенту, следует исключить из модели, при этом оставшиеся в модели параметры должны быть пересчитаны.

Проверка выполнения предпосылок МНК

Проверка выполнения предпосылок МНК выполняется на основе анализа остаточной компоненты. Согласно общим предположениям регрессионного анализа, остатки должны вести себя как *независимые* (в действительности почти независимые) одинаково распределенные *случайные* величины. В классических методах регрессионного анализа предполагается также *нормальный закон распределения остатков.*

Исследование остатков полезно начинать с изучения их графика. Он может показать наличие какой-то зависимости, не учтенной в модели. График остатков может показать необходимость перехода к нелинейной модели (квадратичной, полиномиальной, экспоненциальной) или включения в модель периодических компонент.

График остатков показывает и резко отклоняющиеся от модели наблюдения – *выбросы.* Подобным аномальным наблюдениям надо уделять особо пристальное внимание, так как их присутствие может грубо искажать значения оценок. Устранение эффектов выбросов может проводиться либо с помощью удаления этих точек из анализируемых данных (эта процедура называется *цензурированием),* либо с помощью применения методов оценивания параметров, устойчивых к подобным грубым отклонениям.

Независимость остатков можно проверить расчетом первого коэффициента автокорреляции

$$r(1) = \left(\sum_{t=2}^{n} \varepsilon_i \varepsilon_{i-1} \right) / \sum_{i=1}^{n} \varepsilon_i^2 \ .$$

Для принятия решения о наличии или отсутствии автокорреляции в исследуемом ряду фактическое значение коэффициента автокорреляции $r(1)$ сопоставляется с табличным (критическим) значением для 5%-ного уровня значимости (вероятности допустить ошибку при принятии нулевой гипотезы о независимости уровней ряда). Если фактическое значение коэффициента автокорреляции меньше табличного, то гипотеза об отсутствии автокорреляции в ряду может быть принята, а если фактическое значение больше табличного – делают вывод о наличии автокорреляции в ряду динамики.

Обнаружение гетероскедастичности

Для обнаружения гетероскедастичности обычно используют три теста, в которых делаются различные предположения о зависимости между дисперсией случайного члена и объясняющей переменной: тест ранговой корреляции Спирмена, тест Голдфельда – Квандта и тест Глейзера [Доугерти].

При малом объеме выборки для оценки гетероскедастичности может использоваться метод Голдфельда – Квандта.

Данный тест используется для проверки такого типа гетероскедастичности, когда дисперсия остатков возрастает пропорционально квадрату фактора. При этом делается предположение, что случайная составляющая ε распределена нормально.

Чтобы оценить нарушение гомоскедастичности по тесту Голдфельда – Квандта, необходимо выполнить следующие шаги.

1) Упорядочение n наблюдений по мере возрастания переменной x.

2) Разделение совокупности на две группы (соответственно с малыми и большими значениями фактора x) и определение по каждой из групп уравнений регрессии.

3) Определение остаточной суммы квадратов для первой регрессии $S_{1\hat{y}} = \sum_{i=1}^{n1} (y_i - \hat{y}_{1i})^2$ и второй регрессии $S_{2\hat{y}} = \sum_{i=n-n1+1}^{n} (y_i - \hat{y}_{2i})^2$.

4) Вычисление отношений $\frac{S_{2\hat{y}}}{S_{1\hat{y}}}$ (или $\frac{S_{1\hat{y}}}{S_{2\hat{y}}}$). В числителе должна быть большая сумма квадратов.

Полученное отношение имеет F распределение со степенями свободы $k1=n1\text{-}m$ и $k2=n\text{-}n1\text{-}m$ (m – число оцениваемых параметров в уравнении регрессии).

Если $F_{\text{набл}} = \frac{S_{1\hat{y}}}{S_{2\hat{y}}} > F_{кр(\alpha;\, k1;\, k2)}$, то гетероскедастичность имеет место.

Чем больше величина F превышает табличное значение F-критерия, тем более нарушена предпосылка о равенстве дисперсий остаточных величин.

Оценка влияния отдельных факторов на зависимую переменную на основе модели (коэффициенты эластичности, β-коэффициенты)

Важную роль при оценке влияния факторов играют коэффициенты регрессионной модели. Однако непосредственно с их помощью нельзя сопоставить факторы по степени их влияния на зависимую переменную из-за различия единиц измерения и разной степени колеблемости. Для устранения таких различий при интерпретации применяются средние частные *коэффициенты эластичности* Э (j) и *бета-коэффициенты* β (j), которые рассчитываются по формулам

$$Э_j = \hat{a}_j \frac{\bar{x}_j}{\bar{y}};$$

$$\beta_j = \hat{a}_j \frac{S_{x_j}}{S_y};$$

где S_{xj} , S_y – среднеквадратическое отклонение соответственно фактора j и зависимой переменной y.

Коэффициент эластичности показывает, на сколько процентов изменяется зависимая переменная при изменении фактора j на один процент. Однако он не учитывает степень колеблемости факторов.

Бета-коэффициент показывает, на какую часть величины среднего квадратического отклонения S_y изменится зависимая переменная Y с изменением соответствующей независимой переменной X_j на величину своего среднеквадратического отклонения при фиксированном на постоянном уровне значении остальных независимых переменных.

Указанные коэффициенты позволяют упорядочить факторы по степени их влияния на зависимую переменную.

Долю влияния фактора i в суммарном влиянии всех факторов можно оценить по величине *дельта - коэффициентов* Δ (j)

$$\Delta_j = r_{y,x_j} \cdot \beta_j / R^2,$$

где r_{y,x_j} – коэффициент парной корреляции между фактором j (j = 1,...,m) и зависимой переменной.

Одна из важнейших целей моделирования заключается в *прогнозировании поведения исследуемого объекта*. При использовании построенной модели для прогнозирования делается предположение о сохранении в период прогнозирования существовавших ранее взаимосвязей переменных.

Для того чтобы определить область возможных значений результативного показателя, при рассчитанных значениях факторов следует учитывать два возможных источника ошибок: рассеивание наблюдений относительно линии регрессии и ошибки, обусловленные математическим аппаратом построения самой линии регрессии. Ошибки первого рода измеряются с помощью характеристик точности, в частности, величиной S_y. Ошибки второго рода обусловлены фиксацией численного значения коэффициентов регрессии, в то время как они в действительности являются случайными, нормально распределенными.

Для линейной модели регрессии доверительный интервал рассчитывается следующим образом. Оценивается величина отклонения от линии регрессии (обозначим ее U):

$$u = S_\varepsilon t_a \sqrt{V_{пр}} = S_\varepsilon t_a \sqrt{1 + X^T_{\text{прогн}} (X^T \cdot X)^{-1} X_{\text{прогн}}},$$

где $X^T_{\text{прогн}} = (1, X_{1\text{прогн}}, X_{2\text{прогн}}, \ldots, X_{k\,\text{прогн}})$.

Пример. Построить модель для предсказания объема реализации одного из продуктов фирмы. Объем реализации – это зависимая переменная Y. В качестве независимых объясняющих переменных выбраны время X_1, расходы на рекламу X_2, цена товара X_3, средняя цена товара у конкурентов X_4, индекс потребительских расходов X_5 [1].

Статистические данные по всем переменным приведены в табл. 1.1. В этом примере $n = 16$, $m = 5$.

Таблица 1.1. Данные объема реализации одного из продуктов фирмы

Объем реализации Y, млн руб.	Время X_1, мес.	Расходы на рекламу X_2, тыс. руб.	Цена X_3, руб.	Цена товара у конкурента X_4, руб.	Индекс потребительских расходов X_5, %
126	1	4	15	17	100
137	2	4,8	14,8	17,3	98,4
148	3	3,8	15,2	16,8	101,2
191	4	8,7	15,5	16,2	103,5
274	5	8,2	15,5	16	104,1
370	6	9,7	16	18	107
432	7	14,7	18,1	20,2	107,4
445	8	18,7	13	15,8	108,5
367	9	19,8	15,8	18,2	108,3
367	10	10,6	16,9	16,8	109,2
321	11	8,6	16,3	17	110,1
307	12	6,5	16,1	18,3	110,7
331	13	12,6	15,4	16,4	110,3
345	14	6,5	15,7	16,2	111,8
364	15	5,8	16	17,7	112,3
384	16	5,7	15,1	16,2	112,9

1) Осуществим выбор факторных признаков для построения двухфакторной регрессионной модели. Для этого проведем корреляционный анализ данных (табл. 1.2):

Таблица 1.2. Результат корреляционного анализа

Факторные признаки	Объем реализации	Время	Расходы на рекламу	Цена	Цена у конкурента	Индекс потребительских расходов
Объем реализации	1					
Время	0,678	1				
Реклама	0,646	0106	1			
Цена	0,233	0,174	-0,003	1		
Цена у конкурента	0,226	-0,051	0,204	0,698	1	
Индекс потребительских расходов	0,816	0,960	0,273	0,235	0,03	1

Анализ матрицы коэффициентов парной корреляции (см. табл.1.2) показывает, что зависимая переменная, т.е. объем реализации, имеет тесную связь с индексом потребительских расходов (r_{yx5} = 0,816), с расходами на рекламу (r_{yx2} = 0,646) и со временем (r_{yx1} = 0,678). Однако факторы X_2 и X_5 тесно связаны между собой ($r_{x\,1x5}$ = 0,96), что свидетельствует о наличии мультиколлинеарности. Из этих двух переменных оставим в модели X_5 – индекс потребительских расходов. После исключения незначимых факторов n = 16, k = 2.

2) Оценим параметры регрессии по методу наименьших квадратов. Расчеты произведем в программе *EXCEL* (табл. 1.3 – 1.6):

Таблица 1.3. Регрессионная статистика

Регрессионная статистика	
Множественный *R*	0.927
R-квадрат	0.859
Нормированный *R*-квадрат	0.837
Стандартная ошибка	41.473
Наблюдения	16.000

Таблица 1.4. Дисперсионный анализ

Дисперсионный анализ	*df*	*SS*	*MS*	*F*
Регрессия	2	136358.33	68179.167	39.639
Остаток	13	22360.104	1720.008	–
Итого	15	158718.44	–	–

Таблица 1.5. Коэффициенты модели

	Коэффициенты	Стандартная ошибка	t-статистика
Y-пересечение	-1471.314	259.766	-5.664
Реклама	9.568	2.266	4.223
Индекс потребительских расходов	15.753	2.467	6.386

Уравнение регрессии зависимости объема реализации от затрат на рекламу и индекса потребительских расходов можно записать в следующем виде:

$$y = -1471,314 + 9,568x_1 + 15,754x_2.$$

Таблица 1.6. Вывод остатка

Наблюдение	Предсказанное	Остатки
1	142,25	-16,25
2	124,70	12,30
3	159,24	-11,24
4	242,35	-51,35
5	247,02	26,98
6	307,06	62,94
7	361,20	70,80
8	416,80	28,20
9	424,18	-57,18
10	350,32	16,68
11	345,37	-24,37
12	334,72	-27,72
13	386,79	-55,79
14	352,05	-7,05
15	353,23	10,77
16	361,73	22,27

3) Оценим качество всего уравнения регрессии. Коэффициент детерминации $R^2 = 0{,}859$ показывает долю вариации результативного признака под воздействием изучаемых факторов. Следовательно, около 86 % вариации зависимой переменной учтено в модели и обусловлено влиянием включенных факторов.

Коэффициент множественной корреляции равен R = 0,927. Он показывает тесноту связи зависимой переменной Y с двумя включенными в модель объясняющими факторами.

4) Значимость уравнения регрессии проверим на основе вычисления $F_{рас}$-критерия Фишера. Расчетное значение критерия равно 39,639. Табличное значение $F_{табл}$ -критерия при доверительной вероятности 0,95 при $v_1 = k = 2$ и $v_2 =$ $= n - k$ -1 = 16 – 2 – 1 = 13 составляет 3,81. Поскольку $F_{рас} > F_{табл}$, уравнение регрессии следует признать адекватным.

5) Оценим с помощью t-критерия Стьюдента статистическую значимость коэффициентов уравнения множественной регрессии. Расчетные значения t-критерия приведены в табл. 1.5. Табличное значение t-критерия при 5%-ном уровне значимости и степенях свободы (16 – 2 – 1 = 13) составляет 2,16. Так как $|t_{рас}| > t_{табл}$, то коэффициенты a_1, a_2 существенны (значимы).

6) Проанализируем влияние факторов на зависимую переменную по модели (для каждого коэффициента регрессии вычисляем коэффициент эластичности, β-коэффициенты.

$Э_1 = 9{,}568 \cdot 9{,}294/306{,}813 = 0{,}2898$; $Э_2 = 15{,}7529 \cdot 107{,}231/306{,}813 = 5{,}506$;
$\beta_1 = 9{,}568 \cdot 4{,}913/102{,}865 = 0{,}457$; $\beta_2 = 15{,}7529 \cdot 4{,}5128/102{,}865 = 0{,}691$.

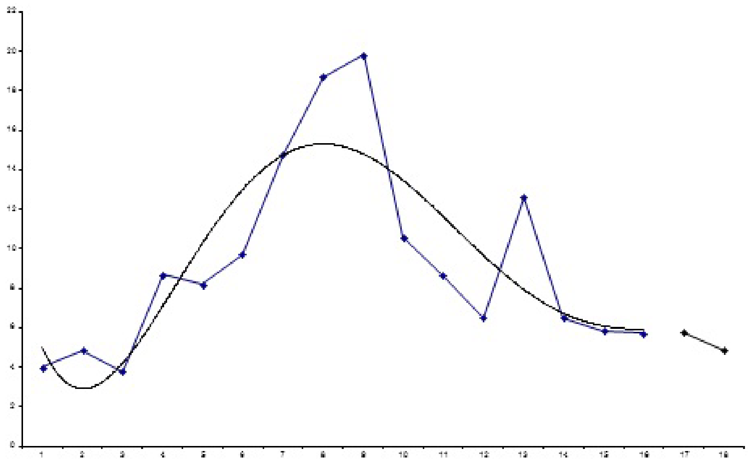
7) Определим точечную и интервальную прогнозные оценки объема реализации на два квартала вперед ($t_{0,7} = 1{,}12$). Исходные данные представлены временными рядами, поэтому прогнозные значения $X_{1,17}$, $X_{2,17}$ и $X_{1,18}$, $X_{2,18}$ можно определить с помощью методов экспертных оценок, с помощью средних абсолютных приростов или вычислить на основе экстраполяционных методов.

Для фактора X_2 «Затраты на рекламу» выбрана модель

$$X_2 = 12{,}83\text{-}11{,}616t + 4{,}319t^2 - 0{,}552t^3 + 0{,}020t^4 - 0{,}0006t^5,$$

по модели получаем прогноз на 2 месяца вперед. График модели временного ряда «Затраты на рекламу» приведен на рисунке.

Упреждение	Прогноз
1	5,75
2	4,85



Прогноз показателя «Затраты на рекламу»

Для временного ряда «Индекс потребительских расходов» в качестве аппроксимирующей функции выбираем полином второй степени (параболу), по которой строим прогноз на 2 шага вперед

Упреждение	Прогноз
1	112,468
2	112,488

$$X_5 = 97{,}008 + 1{,}739t - 0{,}0488t^2 .$$

Для получения прогнозных оценок зависимостей переменной Y по модели $Y = -1471{,}438 + 9{,}568X_2 + 15{,}754X_5$ подставим в нее найденные прогнозные значения факторов X_2 и X_5:

$$Y_{t=17} = -1471{,}438 + 9{,}568 \cdot 5{,}75 + 15{,}754 \cdot 112{,}468 = 355{,}399,$$
$$Y_{t=18} = -1471{,}438 + 9{,}568 \cdot 4{,}85 + 15{,}754 \cdot 112{,}488 = 344{,}179.$$

Результаты прогнозных оценок модели регрессии для выбранной вероятности 90 % с числом степеней свободы, равным 13 ($t_{кр}$ = 1,77), представлены в табл. 1.7.

Таблица 1.7. Таблица прогнозов (p = 90 %)

Упреждение	Прогноз	Нижняя граница	Верхняя граница
1	355,399	273,94	436,85
2	344,179	261,71	426,65

1.3. Задание к лабораторной работе

1. Осуществить выбор факторных признаков для построения двухфакторной регрессионной модели.
2. Рассчитать параметры модели.
3. Для оценки качества всего уравнения регрессии определить:
   - линейный коэффициент множественной корреляции,
   - коэффициент детерминации.
4. Осуществить оценку значимости уравнения регрессии.
5. Оценить с помощью t-критерия Стьюдента статистическую значимость коэффициентов уравнения множественной регрессии.
6. Оценить влияние факторов на зависимую переменную по модели.
7. Построить точечный и интервальный прогнозы результирующего показателя на два шага вперед $\alpha = 0{,}1$.

1.4. Порядок выполнения лабораторной работы

1. Ознакомиться с методикой проведения регрессионного анализа.
2. Получить от преподавателя задание на лабораторную работу.
3. Выбрать факторные признаки для построения двухфакторной регрессионной модели.
4. Оценить параметры регрессии по методу наименьших квадратов.
5. Оценить качество всего уравнения регрессии.
6. Проверить значимость уравнения регрессии на основе вычисления $F_{рас}$-критерия Фишера.
7. Оценить с помощью t-критерия Стьюдента статистическую значимость коэффициентов уравнения множественной регрессии.
8. Проанализировать влияние факторов на зависимую переменную по модели.

9. Определить точечные и интервальные прогнозные оценки зависимой переменной при заданных факторных переменных.

1.5. Содержание отчета

1. Индивидуальное задание.
2. Результаты корреляционного анализа и выбора значимых факторов.
3. Оценки параметров регрессии.
4. Расчеты точечных и интервальных прогнозных оценок.
5. Выводы по результатам моделирования.

1.6. Вопросы для самоконтроля

1. Назначение регрессионного анализа.
2. Как оцениваются параметры модели множественной регрессии?
3. Как оценивается качество модели регрессии, по каким направлениям?
4. Вычисление коэффициента множественной корреляции (индекс корреляции) R и коэффициента детерминации R^2 модели регрессии.
5. Каким образом проверяется значимость модели регрессии?
6. Как проводится анализ статистической значимости параметров модели регрессии?
7. Как выполняется проверка выполнения предпосылок МНК?
8. Для чего оценивается влияние отдельных факторов на зависимую переменную и как это производится?
9. Прогнозирование поведения исследуемого объекта с помощью регрессионной модели, построение точечного и интервального прогнозов.
10. От чего зависит точность прогнозирования по модели регрессии?

1.7. Список рекомендуемой литературы

1. Эконометрика : метод. указания по изучению дисциплины и выполнению контрольной работы и аудиторной работы на ПЭВМ. – М. : ВЗФЭИ, 2004. – 79 с.

2. Дубров, А. М. Многомерные статистические методы : учебник / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. – М. : Финансы и статистика, 2003. – 352 с. – ISBN 5-279-01945-3.

1.8. Варианты заданий

Вариант 1. Имеются данные о деятельности крупнейших компаний США в течение года. Исследовать зависимость переменной *y* от различных факторов и отобразить их взаимосвязь в форме регрессионной модели с двумя значимыми факторами.

№ п/п	*Y*	*X*1	*X*2	*X*3	*X*4
1	0,9	31,3	18,9	43	40,9
2	1,7	13,4	13,7	64,7	40,5
3	0,7	4,5	18,5	24	38,9
4	1,7	10	4,8	50,2	38,5
5	2,6	20	21,8	106	37,3
6	1,3	15	5,8	96,6	26,5
7	1,6	17,9	20,1	85,6	36,8
8	6,9	165,4	60,6	745	36,3
9	0,4	2	1,4	4,1	35,3
10	1,3	6,8	8	26,8	35,3
11	1,9	27,1	18,9	42,7	35
12	1,9	13,4	13,2	61,8	26,2
13	1,4	9,8	12,6	212	33,1
14	0,4	19,5	12,2	105	32,7
15	0,8	6,8	3,2	33,5	32,1
16	1,8	27	13	142	30,5
17	0,9	12,4	6,9	96	29,8
18	1,1	17,7	15	140	25,4
19	1,9	12,7	11,9	59,3	29,3
20	-0,9	21,4	1,6	131	29,2
21	1,3	13,5	8,6	70,7	29,2
22	2	13,4	11,5	65,4	29,1
23	0,6	4,2	1,9	23,1	27,9
24	0,7	15,5	5,8	80,8	27,2

Обозначения: *Y* – чистый доход, млрд дол. США; *X*1 – оборот капитала, млрд дол. США; *X*2 – использованный капитал, млрд дол.; *X*3 – численность служащих, тыс. чел; *X*4 – рыночная капитализация компании, млрд дол. США.

Вариант 2. Представлены данные о рынке строящегося жилья в Санкт-Петербурге (по состоянию на год). Исследовать зависимость переменной y от различных факторов и отобразить их взаимосвязь в форме регрессионной модели с двумя значимыми факторами.

№ п/п	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$	$X8$	Y
1	1	1	39	20	8,2	0	1	0	15,9
2	3	1	68,4	40,5	10,7	0	1	0	27
3	1	1	34,8	16	10,7	0	1	12	13,5
4	1	1	39	20	8,5	0	1	12	15,1
5	2	1	54,7	28	10,7	0	1	12	21,1
6	3	1	74,7	46,3	10,7	0	1	12	28,7
7	3	1	71,7	45,9	10,7	0	0	0	27,2
8	3	1	74,5	47,5	10,4	0	0	0	28,3
9	4	1	137,7	87,2	14,6	0	1	0	52,3
10	1	1	37	17,8	8,3	0	1	3	15,4
11	3	1	69	42,4	8,3	0	1	3	28,6
12	1	1	40	20	8,3	0	0	0	15,6
13	3	1	69,1	41,3	8,3	0	1	0	27,7
14	2	1	68,1	35,4	13	1	1	20	34,1
15	2	1	75,3	41,4	12,1	1	1	20	37,7
16	3	1	83,7	48,5	12,1	1	1	20	41,9
17	1	1	48,7	22,3	12,4	1	1	20	24,4
18	1	1	39,9	18	8,1	1	0	0	21,3
19	2	1	68,6	35,5	17	1	1	12	36,7
20	1	1	39	20	9,2	1	0	0	21,5
21	2	1	48,6	31	8	1	0	0	26,4

Обозначения: Y – цена квартиры, тыс. дол.; $X1$ – число комнат в квартире; $X2$ – район города (1 – Приморский, Шувалово - Озерки, 2 – Гражданка, 3 – Юго-запад, 4 – Красносельский); $X3$ – общая площадь квартиры, м^2; $X4$ – жилая площадь квартиры, м^2; $X5$ – площадь кухни, м^2; $X6$ – тип дома (1 – кирпичный, 0 – другой); $X7$ – наличие балкона (1 – есть, 0 – нет); $X8$ – число месяцев до окончания срока строительства.

Вариант 3. Построить модель с двумя наиболее значимыми факторами для предсказания объема реализации продукции фирмы по данным таблицы.

Y	$X1$	$X2$	$X3$	$X4$	$X5$
126	1	4	15	17	100
137	2	4,8	14,8	17,3	98,4

Окончание

Y	$X1$	$X2$	$X3$	$X4$	$X5$
148	3	3,8	15,2	16,8	101,2
191	4	8,7	15,5	16,2	103,5
274	5	8,2	15,5	16	104,1
370	6	9,7	16	18	107
432	7	14,7	18,1	20,2	107,4
445	8	18,7	13	15,8	108,5
367	9	19,8	15,8	18,2	108,3
367	10	10,6	16,9	16,8	109,2
321	11	8,6	16,3	17	110,1
307	12	6,5	16,1	18,3	110,7
331	13	12,6	15,4	16,4	110,3
345	14	6,5	15,7	16,2	111,8
364	15	5,8	16	17,7	112,3
384	16	5,7	15,1	16,2	112,9

Обозначения: Y – объем реализации, млн руб.; $X1$ – время, г.; $X2$ – расходы на рекламу, тыс. руб.; $X3$ – цена товара, руб.; $X4$ – средняя цена товара у конкурентов, руб.; $X5$ – индекс потребительских расходов, %.

Вариант 4. Установите направление и характер взаимосвязи между четырьмя факторами по 15 банкам зарубежной страны.

№ п/п	$X1$	$X2$	$X3$	$X4$
1	507,2	19,5	359,9	448,1
2	506,6	19,8	187,1	451,9
3	487,8	21,1	375,2	447,9
4	496	18,6	287,9	444,3
5	493,6	19,6	444	443,2
6	458,9	11,7	462,4	411,7
7	429,3	10,5	459,5	328,6
8	386,9	13,6	511,3	314,7
9	311,5	10,8	328,6	259,4
10	302,2	10,9	350	187,7
11	262	10,3	298,7	238,5
12	242,2	10,6	529,3	269,4
13	231,9	8,5	320	284
14	214,3	6,7	502	172,3
15	208,4	8,3	194,9	166,4

Обозначения: $X1$ – суммарный актив, млрд дол.; $X2$ – объем вложений акционеров, млрд дол.; $X3$ – чистый доход, млрд дол.; $X4$ – депозиты, млрд дол.

Вариант 5. Провести анализ деловой активности и прибыльности крупнейших банков России по данным за год. Определите факторы развития банковской системы.

№ п/п	*X*1	*X*2	*X*3	*X*4	*X*5
1	1370596	3138452	260727	600883	12913141
2	1052618	1749462	806316	722440	9549920
3	640478	1177193	482539	969496	3995816
4	557032	809268	400351	889704	4566926
5	1120847	317719	207889	753993	9393955
6	996003	772401	395220	626085	4166522
7	527385	1234517	609219	185066	2316869
8	625027	3049381	285677	191631	2776955
9	469296	1381584	463639	86559	6165342
10	487892	1009361	435813	587507	4674425
11	615759	517422	331008	535557	4600065
12	1032806	262494	120516	43653	1933402
13	413497	119884	187428	488837	1669520
14	246722	1115686	136567	12864	1129019
15	425144	191202	94535	392308	347461

Обозначения: *X*1 – собственный капитал; *X*2 – ссудная задолженность; *X*3 – балансовая прибыль; *X*4 – вложения в государственные бумаги; *X*5 – привлеченные ресурсы.

Вариант 6. Исследуйте зависимость курса доллара США по отношению к рублю по данным таблицы в зависимости от двух наиболее значимых факторов.

№ п/п	*X*1	*X*2	*X*3	*Y*
1	3051,08	16048	212,97	23,68
2	3051,08	16048	212,97	23,8
3	3031,1	16378	214,27	23,8
4	2935,16	14500	217,43	23,92
5	2952,09	16019	216,92	24,29
6	2962,22	15986	215,91	24,22
7	3009,52	16017	213,26	24,18
8	3001,4	16017	213,43	24,19
9	3001,4	16017	213,43	24,2
10	2993,27	16017	213,61	24,2

Окончание

№ п/п	$X1$	$X2$	$X3$	Y
11	3034,61	16009	213,02	24,2
12	3021,18	15836	212,72	24,18
13	2979,92	16327	214,71	24,16
14	2996,3	16290	219,49	24,29

Обозначения: Y – курс доллара, руб./1 дол.; $X1$ – *D*J индекс; $X2$ – *TN* индекс; $X3$ – цена золота, руб./г.

Вариант 7. По данным мониторинга о состоянии экологической защиты и охраны труда на промышленном предприятии за два года проанализируйте зависимость заболеваемости работников предприятия от содержания вредных веществ в производственно-ливневых водах после прохождения очистных сооружений.

Месяц	Случаи заболевания на 100 работающих y	Содержание вредных веществ в производственно-ливневых водах, мг/дм^3			
		Нефтепродукты	Железо	NO	NO_2
1	7	0,0625	0,7	39,375	5,625
2	7,9	0,225	0,68	29,375	6,9375
3	2,5	0,05	0,59	35	3,875
4	4,7	0,075	0,55	35	3,375
5	2,5	0,025	0,375	36,25	3,625
6	3,2	0,0125	0,6	26,875	1,375
7	3,6	0,325	0,5	34,375	2,3125
8	3,8	0,65	0,65	29,375	1,8125
9	3,3	0,75	0,21	27,5	1,625
10	5,4	0,55	0,64	26,873	1,5787
11	5,1	0,775	0,45	23,75	1,5625
12	3,2	0,3	0,43	30	2,75
13	3,4	0,2	0,54	22,9048	0,8415
14	3,9	0,2	0,675	21,2916	0,9442
15	6,3	0,39	0,7	19,5264	0,9331
16	5,5	0,175	0,625	23,4375	1
17	5,2	0,6	0,475	17	0,9375
18	4,6	0,75	0,25	18,125	0,40625
19	5,8	0,125	0,95	14	0,25
20	6	0,525	0,425	5,5625	0,375
21	5,3	0,35	0,475	6,5	0,25

Вариант 8. Имеются статистические данные функционирования многомерного объекта управления с семью входными переменными X=(x1, x2, ..., x7) и с одной выходной переменной y1. Исследовать зависимость переменной y1 от различных факторов и отобразить их взаимосвязь в форме регрессионной модели с двумя значимыми факторами.

№ п/п	x1	x2	x3	x4	x5	x6	x7	y1
1	1	655	0,1	1,22	0,3	0,35	5	245
2	1	679	0,2	1,2	0,35	0,35	7	253
3	1	644	0,21	1,22	0,3	0,35	7	246
4	1	644	0,17	1,11	0,2	0,4	7	230
5	1	635	0,2	1,21	0,2	0,4	6	245
6	1	655	0,2	1,21	0,2	0,4	7	215
7	1	655	0,2	1,21	0,2	0,4	7	210
8	1	655	0,2	1,21	0,2	0,4	7	208
9	1	650	0,22	1,24	0,3	0,35	15	189
10	1	672	0,22	1,2	0,3	0,35	16	150
11	1	679	0,22	1,2	0,3	0,35	15	198
12	1	635	0,22	1,2	0,3	0,35	15	157
13	1	644	0,21	1,22	0,3	0,35	15	190
14	1	644	0,17	1,11	0,2	0,4	15	150
15	1	575	0,17	1,11	0,2	0,4	15	139
16	1	635	0,18	1,11	0,2	0,4	15	184
17	1	635	0,2	1,21	0,2	0,4	14	193
18	1	655	0,2	1,21	0,2	0,4	14	200
19	2	650	0,2	1,21	0,3	0,35	17	145
20	2	650	0,2	1,21	0,3	0,35	17	137
21	2	650	0,2	1,21	0,3	0,35	17	124
22	2	650	0,2	1,21	0,3	0,35	18	132
23	2	650	0,2	1,21	0,4	0,4	18	126
24	2	650	0,2	1,21	0,4	0,4	18	117
25	2	650	0,2	1,21	0,4	0,4	18	113
26	2	650	0,2	1,21	0,4	0,4	18	119
27	2	650	0,2	1,21	0,4	0,4	18	127
28	2	650	0,2	1,21	0,4	0,4	18	118
29	2	650	0,2	1,21	0,4	0,4	18	125
30	2	650	0,2	1,21	0,4	0,4	18	131
31	2	650	0,2	1,21	0,4	0,4	18	120
32	2	650	0,2	1,21	0,4	0,4	18	121

Вариант 9. Имеются статистические данные функционирования многомерного объекта управления с семью входными переменными $X=(x1, x2, ..., x7)$ и с одной выходной переменной $y2$. Исследовать зависимость переменной $y2$ от различных факторов и отобразить их взаимосвязь в форме регрессионной модели с двумя значимыми факторами.

№ п/п	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$	$x7$	$y2$
1	1	655	0,1	1,22	0,3	0,35	5	89
2	1	679	0,2	1,2	0,35	0,35	7	91
3	1	644	0,21	1,22	0,3	0,35	7	83
4	1	644	0,17	1,11	0,2	0,4	7	81
5	1	635	0,2	1,21	0,2	0,4	6	78
6	1	655	0,2	1,21	0,2	0,4	7	75
7	1	655	0,2	1,21	0,2	0,4	7	70
8	1	655	0,2	1,21	0,2	0,4	7	68
9	1	650	0,22	1,24	0,3	0,35	15	61
10	1	672	0,22	1,2	0,3	0,35	16	59
11	1	679	0,22	1,2	0,3	0,35	15	62
12	1	635	0,22	1,2	0,3	0,35	15	60
13	1	644	0,21	1,22	0,3	0,35	15	71
14	1	644	0,17	1,11	0,2	0,4	15	59
15	1	575	0,17	1,11	0,2	0,4	15	55
16	1	635	0,18	1,11	0,2	0,4	15	67
17	1	635	0,2	1,21	0,2	0,4	14	68
18	1	655	0,2	1,21	0,2	0,4	14	71
19	2	650	0,2	1,21	0,3	0,35	17	59
20	2	650	0,2	1,21	0,3	0,35	17	55
21	2	650	0,2	1,21	0,3	0,35	17	51
22	2	650	0,2	1,21	0,3	0,35	18	55
23	2	650	0,2	1,21	0,4	0,4	18	54
24	2	650	0,2	1,21	0,4	0,4	18	51
25	2	650	0,2	1,21	0,4	0,4	18	49
26	2	650	0,2	1,21	0,4	0,4	18	52
27	2	650	0,2	1,21	0,4	0,4	18	61
28	2	650	0,2	1,21	0,4	0,4	18	53
29	2	650	0,2	1,21	0,4	0,4	18	51
30	2	650	0,2	1,21	0,4	0,4	18	53
31	2	650	0,2	1,21	0,4	0,4	18	54
32	2	650	0,2	1,21	0,4	0,4	18	49

Вариант 10. Имеются статистические данные функционирования многомерного объекта управления с семью входными переменными X=(x1, x2, ..., x7) и с одной выходной переменной y3. Исследовать зависимость переменной y3 от различных факторов и отобразить их взаимосвязь в форме регрессионной модели с двумя значимыми факторами.

№ п/п	x1	x2	x3	x4	x5	x6	x7	y3
1	1	655	0,1	1,22	0,3	0,35	5	40
2	1	679	0,2	1,2	0,35	0,35	7	35
3	1	644	0,21	1,22	0,3	0,35	7	38
4	1	644	0,17	1,11	0,2	0,4	7	39
5	1	635	0,2	1,21	0,2	0,4	6	37
6	1	655	0,2	1,21	0,2	0,4	7	42
7	1	655	0,2	1,21	0,2	0,4	7	45
8	1	655	0,2	1,21	0,2	0,4	7	51
9	1	650	0,22	1,24	0,3	0,35	15	55
10	1	672	0,22	1,2	0,3	0,35	16	57
11	1	679	0,22	1,2	0,3	0,35	15	52
12	1	635	0,22	1,2	0,3	0,35	15	53
13	1	644	0,21	1,22	0,3	0,35	15	48
14	1	644	0,17	1,11	0,2	0,4	15	59
15	1	575	0,17	1,11	0,2	0,4	15	60
16	1	635	0,18	1,11	0,2	0,4	15	54
17	1	635	0,2	1,21	0,2	0,4	14	49
18	1	655	0,2	1,21	0,2	0,4	14	37
19	2	650	0,2	1,21	0,3	0,35	17	46
20	2	650	0,2	1,21	0,3	0,35	17	53
21	2	650	0,2	1,21	0,3	0,35	17	58
22	2	650	0,2	1,21	0,3	0,35	18	64
23	2	650	0,2	1,21	0,4	0,4	18	62
24	2	650	0,2	1,21	0,4	0,4	18	67
25	2	650	0,2	1,21	0,4	0,4	18	65
26	2	650	0,2	1,21	0,4	0,4	18	66
27	2	650	0,2	1,21	0,4	0,4	18	67
28	2	650	0,2	1,21	0,4	0,4	18	64
29	2	650	0,2	1,21	0,4	0,4	18	68
30	2	650	0,2	1,21	0,4	0,4	18	69
31	2	650	0,2	1,21	0,4	0,4	18	65
32	2	650	0,2	1,21	0,4	0,4	18	69

Лабораторная работа № 2
КОМПОНЕНТНЫЙ АНАЛИЗ

2.1. Цель работы

Освоение методов построения регрессионных моделей на основе использования компонентного анализа.

2.2. Теоретические сведения

Компонентный анализ проводится с несколькими частными целями. Как метод снижения размерности он позволяет выявить закономерности, которые непосредственно не наблюдаются. Эта задача решается по матрице нагрузок, как и классификация признаков в пространстве главных компонент. А индивидуальные значения используются для классификации объектов (не по исходным признакам, а по главным компонентам) и для построения уравнения регрессии на эти обобщенные показатели. Кроме того, диаграмма рассеяния объектов, построенная в плоскости, образованной двумя первыми, наиболее весомыми главными компонентами, позволяет косвенно подтвердить или опровергнуть предположение о том, что исследуемые данные подчиняются многомерному нормальному закону распределения вероятностей.

Форма облака должна напоминать эллипс, более густо объекты расположены в его центре и разреженно по мере удаления от него. Интерпретируются главные компоненты, которым соответствуют дисперсии больше 1 и которые имеют хотя бы одну весомую нагрузку. Выбор критической величины, при превышении которой элемент матрицы нагрузок признается весовым и оказывает влияние на интерпретацию главной компоненты, определяется по смыслу решаемой задачи и может варьировать в пределах от 0,5 до 0,9 в зависимости от получаемых промежуточных результатов.

Идея метода заключается в замене сильно коррелированных переменных новыми переменными (главными компонентами), между которыми корреляция отсутствует. При этом главные компоненты $z_1 - z_n$ являются линейными комбинациями исходных переменных $x_1 - x_{\text{n}}$:

$$\begin{cases} z_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n , \\ z_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n , \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ z_n = a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n . \end{cases}$$

Главные компоненты подбираются так, чтобы z_1 имела наибольшую дисперсию. Для каждой следующей главной компоненты дисперсия убывает. Последняя компонента имеет наименьшую дисперсию.

Так как исходные переменные $x_1 - x_n$ измерены в несопоставимых величинах, то необходимо перейти к центрированным нормированным величинам. При этом все переменные будут иметь нулевое математическое ожидание и единичную дисперсию. Матрицу исходных центрированно-нормированных значений переменных находят из соотношения

$$X = \begin{vmatrix} \dfrac{X_{11} - \overline{X_1}}{\hat{S}_1} & \dots & \dfrac{X_{1N} - \overline{X_1}}{\hat{S}_1} \\ \dots & \dots & \dots \\ \dfrac{X_{n1} - \overline{X_n}}{\hat{S}_n} & \dots & \dfrac{X_{nN} - \overline{X_n}}{\hat{S}_n} \end{vmatrix}$$

где $\overline{X} = \dfrac{\sum_{i=1}^{N} X_i}{N}$ – несмещенная, состоятельная и эффективная оценка математического ожидания; N – количество наблюдений.

$\hat{S}^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2$ – несмещенная, состоятельная и эффективная оценка дисперсии.

Так как переменные центрированы и нормированы, то оценку корреляционной матрицы можно провести по формуле

$\hat{R} = \dfrac{1}{N-1} \cdot X \cdot X^T$ – размерность матрицы корреляций n х n.

Перед тем как проводить компонентный анализ, анализируется независимость исходных признаков. Проверяется значимость матрицы парных корреляций с помощью критерия Уилкса.

Выдвигается гипотеза: H_0: $\hat{R}$ незначима и альтернативная H_1: $\hat{R}$ значима.

Рассчитывается статистика $\gamma_н = -(N - \frac{1}{6}(2n+5))\ln|\hat{R}|$, которая распределена по закону χ^2 с $n\frac{n-1}{6}$ степенями свободы. Сравнивается расчетное значение с табличным значением $\chi^2_{кр}$ для уровня значимости α = 0,05.

Если расчетное значение критерия будет больше табличного значения $\gamma_н > \chi^2_{кр}$, то гипотеза H_0 отвергается и принимается альтернативная H_1: $\hat{R}$ значима, следовательно, имеет смысл проводить компонентный анализ.

Затем поверяется гипотеза о диагональности ковариационной матрицы. Выдвигается нулевая гипотеза H_0: $cov(x_i x_j)=0$, $\forall i \neq j$ и альтернативная H_1: $cov\ (x_i x_j) \neq 0$.

Рассчитывается статистика $\gamma_н = -(N - \frac{2n+11}{6})\ln|\hat{R}|$, которая распределяется по закону χ^2 с $n\frac{n-1}{2}$ степенями свободы.

Если расчетное значение критерия будет больше табличного значения $\gamma_н > \chi^2_{кр}$, то гипотеза H_0 отвергается и принимается альтернативная H_1: $\hat{R}$ значима, что подтверждает мультиколлениарность данных, следовательно, имеет смысл проводить компонентный анализ.

Для выделения главных компонент на уровне информативности 0,85 пользуются мерой информативности, которая показывает, какую часть или какую долю дисперсии исходных переменных составляют *k*-первых главных компонент. На заданном уровне информативности выделяются *k* главных компонент.

Для решения данной задачи необходимо использовать пакет прикладных программ статистического анализа, например *Statgraphics Plus*.

Программа выдает матрицу коэффициентов корреляции *A* между центрированно-нормированными исходными переменными и ненормированными главными компонентами размерностью (n x k)

$$A = \begin{vmatrix} a_{11} & a_{12} & .. & a_{1k} \\ a_{21} & a_{22} & .. & a_{2k} \\ a_{n1} & a_{n2} & .. & a_{nk} \end{vmatrix}$$

Коэффициенты показывают наличие, силу и направление линейной связи между соответствующими исходными переменными x_1 - x_n и соответствующими главными компонентами $z_1 - z_k$. Уравнение в матричной форме записи принимает вид $Z = X A$.

Программа выдает матрицу наблюдаемых значений главных компонент Z размерностью (n x k):

$$Z = \begin{vmatrix} z_{11} & z_{12} & \ldots & z_{1k} \\ z_{21} & z_{22} & \ldots & z_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ z_{N1} & z_{N2} & & z_{Nk} \end{vmatrix}.$$

Используя значения главных компонент, построим модель главных компонент:

$$x_{ji} = a_{1i} z_{j1} + a_{2i} z_{j2} + a_{3i} z_{j3} + \ldots\ldots + a_{ki} z_{jk}, \quad j=1, 2, ..N, \; i = 1, 2, ..n.$$

Для интерпретации используются, как правило, наиболее весомые главные компоненты. Уравнение регрессии на выделенные главные компоненты строится методом множественной регрессии. Оценивается значимость уравнения регрессии в целом и значимость коэффициентов регрессии при главных компонентах

$$y = b_0 + b_1 z_1 + b_2 z_2 + \ldots. + b_k z_k \;.$$

Подставляя в полученное уравнение значения главных компонент z_1 – – z_k, выраженные через центрированные переменные x_1 - x_n, получаем окончательное уравнение регрессии:

$$y = c_0 + c_1 \mathrm{x}_1 + c_2 x_2 + \ldots. + c_n x_n \, .$$

Это уравнение отличается более высокой точностью, чем получаемое с использованием классического метода наименьших квадратов.

Пример

Имеются данные, описывающие зависимость результирующей переменной y от факторных переменных $x1$ – $x3$ (табл. 2.1) с использованием метода главных компонент.

Таблица 2.1. Исходные данные

$x1$	$x2$	$x3$	y
1,1	1,1	1,2	26,2
1,4	1,5	1,1	25,9
1,7	1,8	2	32,5
1,7	1,7	1,8	31,7
1,8	1,9	1,8	31,7
1,8	1,8	1,9	33,6
1,9	1,8	2	34,2
2	2,1	2,1	34,4
2,3	2,4	2,5	35,5
2,5	2,5	2,4	36,5

1) Подготовим данные для использования пакета *Statgraphics Plus*. Данные вводим непосредственно в *Statgraphics Plus* путем копирования таблицы с данными.

Убираем с экрана лишние надписи удалением соответствующих строк и колонок таблицы с импортированными данными. Выполним форматирование данных по каждой колонке в отдельности. Для этого помечаем мышью редактируемую колонку с данными. Войти в пункт меню редактирования *Edit*. Выбрать режим *Modify Column*…Установить формат данных с фиксированной точкой с необходимым числом цифр после запятой, например с двумя цифрами после запятой *Fixed Decimal* 2 .

При этом необходимо запомнить размещение данных во вновь полученной табл. 2.2 по колонкам *Col_1 – Col_4*.

Таблица 2.2. Данные, размещенные в программе

*Col_* 1	*Col_* 2	*Col_* 3	*Col_* 4
1,10	1,10	1,20	26,20
1,40	1,50	1,10	25,90
1,70	1,80	2,00	32,50
1,70	1,70	1,80	31,70
1,80	1,90	1,80	31,70
1,80	1,80	1,90	33,60
1,90	1,80	2,00	34,20
2,00	2,10	2,10	34,40
2,30	2,40	2,50	35,50
2,50	2,50	2,40	36,50

2) Проверим мультиколлениарность факторов $x1 – x3$. Мультиколлениарность оцениваем по результатам анализа матрицы парных коэффициентов корреляции. Для расчета матрицы парных коэффициентов корреляции и выдачи ее на печать с исходными данными необходимо вызвать в главном меню программу *Summary stats*. В окно *Data* записать колонки *Col*_1, *Col*_2, *Col*_3, *нажать ОК***.** Вызвать подменю *Tabular options*. В окне табличных настроек поставить флажок напротив *Correlations*, нажать клавишу *OK*. При этом на экране появится матрица коэффициентов парной корреляции. Для записи матрицы в таблицу с данными необходимо вызвать пункт подменю *Save results,* в окне *Correlations* установить флажок и нажать *OK*. Файлу будет приписан идентификатор *CMAT.* Матрица коэффициентов парной корреляции будет продолжением таблицы с исходными данными с колонками *CMAT*_1, *CMAT*_2, *CMA*T_3. Матрица коэффициентов парной корреляции для рассматриваемого примера имеет вид табл. 2.3.

Таблица 2.3. Матрица парных коэффициентов корреляции

*CMAT*_1	*CMAT*_2	*CMAT*_3
1,0	0,985	0,931
0,985	1,0	0,915
0,931	0,915	1,0

Коэффициенты парной корреляции больше 0,8, что свидетельствует о коррелированности данных, следовательно, имеет смысл проводить компонентный анализ.

3) Выделим главные компоненты, построим уравнение главных компонент. Для выделения главных компонент воспользуемся специальной программой. Для этого в главном меню необходимо вызвать программу главных компонент: *Special \ Multivariate Methods \ Principal Components***.** В окно *Data* внесите имена колонок с исходными данными *Col*_1, *Col*_2, *Col*_3, нажать *OK*.

Для получения данных компонентного анализа вызываем подменю *Tabular options* и помечаем окно *Analysis Summaru*, нажимаем *OK*. При этом на экране отобразятся результаты анализа (табл. 2.4):

Таблица 2.4. Главные компоненты

Principal Components Analysis

Number	Component Eigenvalue	Percent of Variance	Cumulative Percentage
1	2,888	96,26	96,26
2	0,0985	3,28	99,54
3	0,0137	0,45	100,00

На уровне информативности 95 % и выше выделяется одна главная компонента. Она имеет наибольшую дисперсию, равную 96,26 %. Использование второй главной компоненты не приводит к существенному увеличению дисперсии (всего на 3,28 %). Главная компонента является линейной комбинацией исходных данных $x1 - x3$. Для выдачи на печать параметров модели необходимо пометить окно *Component Weights*. При этом на экране появятся параметры модели.

Table of Component Weights

	Component 1
Col_1	0,583
Col_2	0,580
Col_3	0,569

Имея параметры, записываем уравнение первой главной компоненты:

$$z\,1 = 0{,}583\,x1 + 0{,}58\,x2 + 0{,}569\,x3.$$

Программа рассчитывает значения главных компонент для всех опытных данных. Для выдачи данных необходимо пометить окно *Data Table* и нажать *OK***.** При этом на экране появятся значения главных компонент.

Row	Component 1
1	-2,983
2	-2,107
3	-0,107
4	-0,502
5	-0,073
6	-0,089
7	0,183
8	0,880
9	2,247
10	2,552

4) Построим уравнение регрессии на главных компонентах. Уравнение регрессии на выделенных главных компонентах строится методом множественной регрессии. Для чего воспользуемся программой *Multiple regression.* Результаты расчета уравнения регрессии приводим ниже:

Multiple Regression Analysis

Dependent variable: Col_4

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	32,22	0,394	81,782	0,00
PCOMP_1	2,000	0,244	8,187	0,00

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	104,039	1	104,039	67,03	0,00
Residual	12,417	8	1,552		
Total (Corr.)	116,456	9			

R-squared = 89,34 percent

R-squared (adjusted for d.f.) = 88,00 percent

Standard Error of Est. = 1,246

Mean absolute error = 0,914

Durbin-Watson statistic = 1,585

Полученное уравнение регрессии имеет вид:

$$y = 32{,}22 + 2{,}00\ z1.$$

Первая главная компонента $z1$ адекватно описывает зависимую переменную y. Подставляя в полученное уравнение выражение для первой главной компоненты, переходим к исходным переменным $x1 - x3$:

$$y = 16{,}542 + 2{,}822\ x1 + 2{,}808\ x2 + 2{,}755\ x3.$$

Полученное уравнение более точно описывает зависимость результирующей переменной от влияющих факторов по сравнению с уравнением множественной регрессии.

2.3. Задание к лабораторной работе

1. Получить задание от преподавателя на выполнение лабораторной работы.
2. Ввести исходные данные в ППП *Statgraphics Plus* или в другой пакет, например *STATISTICA*.
3. Оценить мультиколлениарность факторных переменных и сделать вывод о целесообразности построения модели на главных компонентах.
4. Рассчитать главные компоненты.
5. Построить уравнение в главных компонентах и оценить его адекватность.
6. Построить уравнение регрессии в исходных факторных переменных.
7. Сравнить точность регрессионной модели с моделью на главных компонентах. Сделать выводы по результатам исследований.

2.4. Порядок выполнения лабораторной работы

1. Подготовить данные для расчетов в ППП *Statgraphics Plus* либо в другой программе.
2. Проверить независимость (немультиколлениарность) факторных переменных.
3. Выделить главные компоненты, построить уравнения главных компонент.
4. Построить уравнение регрессии на выделенных главных компонентах методом множественной регрессии.

5. Оценить значимость уравнения регрессии в целом и коэффициентов регрессии при главных компонентах.

6. Сделать выводы по результатам исследований.

2.5. Содержание отчета

1. Исходные данные для исследования.
2. Матрица парных коэффициентов корреляции.
3. Таблица главных компонент.
4. Уравнения главных компонент.
5. Таблица со значениями главных компонент.
6. Результаты расчета уравнения регрессии на главных компонентах.
7. Уравнение регрессии с исходными факторными переменными.
8. Выводы по результатам выполненной лабораторной работы.

2.6. Вопросы для самоконтроля

1. С какими целями проводится компонентный анализ?

2. Как косвенно можно подтвердить или опровергнуть предположение о том, что исследуемые данные подчиняются многомерному нормальному закону распределения вероятностей.

3. В чем заключается идея метода главных компонент. Как подбираются главные компоненты?

4. Для чего проводится анализ независимости исходных факторных переменных?

5. Какую информацию содержат коэффициенты матрицы корреляций?

6. Как определяется целесообразность проведения компонентного анализа?

7. К чему приводит мультиколлениарность данных при регрессионном анализе?

8. Как оценивается точность модели на главных компонентах?

2.7. Список рекомендуемой литературы

1. Дубров, А. М. Многомерные статистические методы / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. – М. : Финансы и статистика, 2003. – 352 с. – ISBN 5-279-019450-3.

2. Яновский, Л. П. Введение в эконометрику : учеб. пособие / Л. П. Яновский, А. Г. Буховец; под. ред. Л. П. Яновского. – 2-е изд., доп. – М. : КНОРУС, 2007. – 256 с. – ISBN 5-85971-270-0.

2.8. Варианты заданий

Использовать данные из лабораторной работы № 1 с целью сравнения точности разработанных моделей на главных компонентах с регрессионными моделями, ранее разработанными в лабораторной работе № 1.

Лабораторная работа № 3
АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

3.1. Цель работы

Освоение методов построения моделей временных рядов на основе структурирования процессов.

3.2. Теоретические сведения

Модели, построенные по данным, характеризующим экономическую систему или процесс за ряд последовательных равноотстоящих моментов времени, называются моделями временных рядов, в дальнейшем – временными рядами. Простейшей является модель аддитивного случайного процесса, имеющая вид [1, 2]:

$$Y_t = U_t + V_t + e_t \,, \qquad (3.1)$$

где U_t – трендовая компонента; V_t – сезонная компонента; e_t – случайная компонента; t – уровни наблюдения, t =1, 2, 3,….

Для построения модели (3.1) необходимо получить оценки каждой компоненты. Для выделения составляющих компонент пользуются процедурами фильтрации, регрессионного и корреляционного анализов.

Относительно трендовой составляющей U_t предполагают, что она должна представлять некоторую гладкую функцию, описываемую полиномом минимальной степени. Для этого чаще всего используются следующие функции времени t:

- линейная $U_t = a+b\,t$;
- парабола второго и реже более высокого порядков

$$U_t = a+b_1\,t+b_2\,t^2+b_3\,t^3+\ldots+b_n\,t^n;$$

- экспонента $U_t = e^{a+bt}$ и др.

Параметры тренда определяются методом наименьших квадратов, в качестве независимой переменной выступает время t =1, 2, 3, ... , а в качестве зависимой переменной – уровни временного ряда Y_t. Критерием отбора наилучшей формы тренда является значение скорректированного коэффициента детерминации R^2.

Пример. Имеются данные о выработке продукции за 18 месяцев работы производственного участка (табл. 3.1). Требуется построить график динамики выработки продукции, подобрать наилучшую форму тренда, выделить сезонную компоненту и построить аддитивную модель.

Таблица 3.1. Выработка продукции

Месяцы	1	2	3	4	5	6
Выработка продукции, шт.	596488	615925	612846	634217	659835	615392
Месяцы	7	8	9	10	11	12
Выработка продукции, шт.	708291	580846	509008	ё	568649	420148
Месяцы	13	14	15	16	17	18
Выработка продукции, шт.	452529	447319	456579	505584	484261	453356

Решение проводим, используя ППП *MS EXCEL*. С использованием Мастера диаграмм строим график динамики выработки продукции (рис. 3.1).

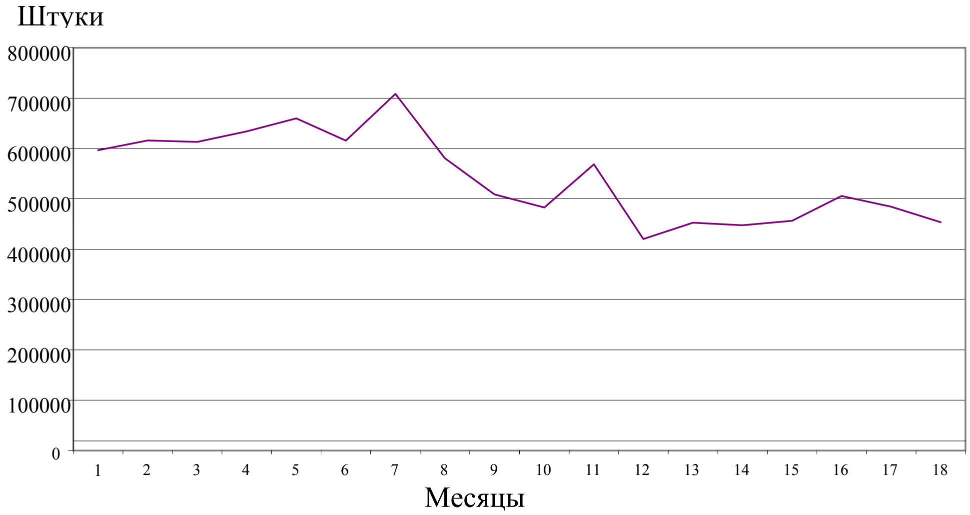


Рис. 3.1. График выработки продукции по месяцам

График характеризует убывающую тенденцию выработки продукции с периодическими колебаниями. Проведем подбор тренда путем добавления линий тренда. Одновременно установим режим отображения уравнения регрессии, описывающего тренд, и коэффициента детерминации. В табл. 3.2 приведены характеристики подбираемых линий тренда. Все три вида тренда адекватно описывают характер изменения выработки продукции во времени. Коэффициенты детерминации статистически значимы при уровне значимости 0,05, расчетные значения критерия Фишера превышают табличные данные.

Таблица 3.2. Подбор вида тренда

Вид тренда	Коэффициент детерминации, %	Уравнение тренда
Линейный	61	$U_t = 665390 - 12707\, t$
Парабола	61,5	$U_t = -50{,}31t^2 - 11751\, t + 662203$
Экспонента	62,4	$U_t = 672830e^{-0{,}0235\, t}$

Для математического описания тренда выбираем более простое линейное уравнение.

Для выделения сезонной компоненты совместно со случайной ($V_t + e_t$) из исходного ряда Y_t вычитаем трендовую компоненту U_t. При этом получаем центрированный временной ряд

$$(V_t + e_t) = Y_t - U_t\ .$$

График центрированного временного ряда отображен на рис. 3.2. Для определения периода циклической компоненты V_t вычисляем автокорреляционную функцию центрированного временного ряда (рис. 3.3). На графике просматривается периодическая составляющая с периодом (13-1) = = 12 месяцев и временным сдвигом (12 - 3) = 9 месяцев.

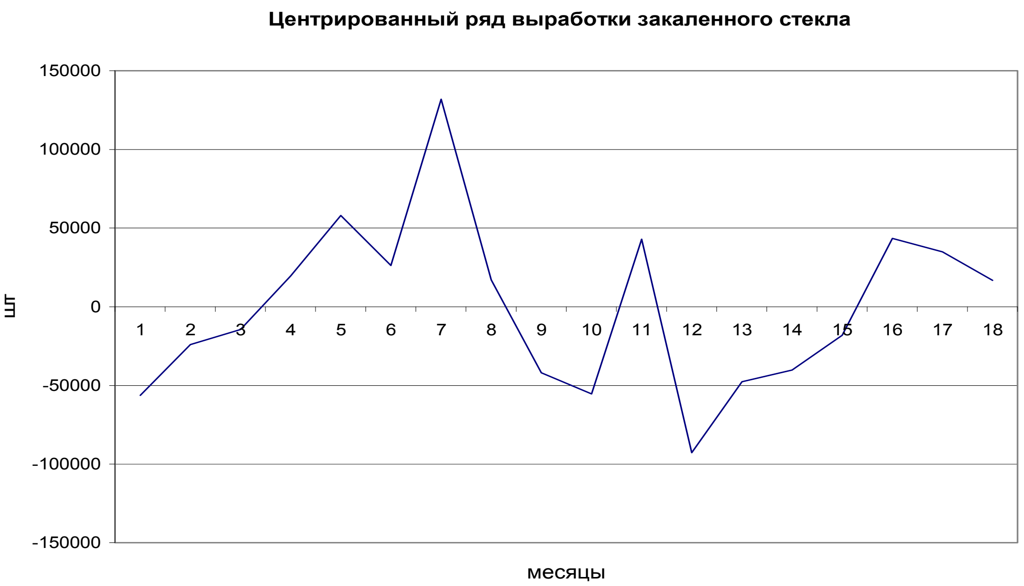


Рис. 3.2. График компонент ($V_t + e_t$) в динамическом ряду выработки продукции

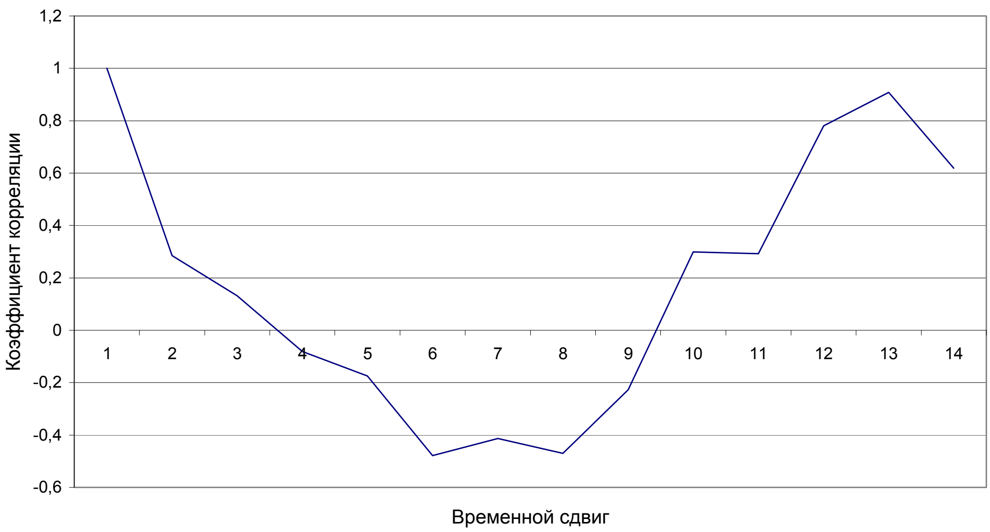


Рис. 3.3. Автокорреляционная функция центрированного временного ряда

Амплитуда гармоники может быть приближенно оценена через дисперсию центрированного временного ряда, т.к. из условия аддитивности модели вытекает баланс дисперсий центрированного ряда

$$S^2_{(Vt+et)} = S^2_{(Vt)} + S^2_{(et)},$$

где $S^2_{(Vt+et)}$ – оценка дисперсии центрированного временного ряда; $S^2_{(Vt)}$ – оценка дисперсии сезонной (гармонической) компоненты, равная квадрату амплитуды гармоники; $S^2_{(et)}$ – оценка дисперсии случайной компоненты.

Если пренебречь дисперсией случайной компоненты, то за амплитуду гармонической составляющей можно принять (оценка сверху) стандартное отклонение центрированного ряда. В рассматриваемом примере это будет

$$A_{Vt} = S_{(Vt)} = 53660.$$

Амплитуда гармоники может быть уточнена по критерию минимума случайной компоненты временного ряда. На графике (рис. 3.4) приведены совмещенные компоненты (V_t+e_t) и гармоническая компонента V_t с уточненной амплитудой, равной 50000:

$$V_t = 50000 \cdot Sin((2\pi/12)\,t + 2\pi \cdot 2{,}85/4).$$

Для выделения случайной компоненты e_t из центрированного временного ряда (V_t+e_t) вычитаем гармоническую компоненту V_t. График случайной компоненты приведен на рис. 3.5. Случайная компонента e_t имеет следующие параметры:

- среднее значение равно -226,3 (шт./мес.), что статистически незначимо при уровне значимости 0,05;

- оценка дисперсии равна 13,7 10^8 (шт./мес.)^2.

После подстановки в исходное уравнение (3.1) всех компонент, временной ряд выработки продукции, уровни которых представлены в табл. 3.1, описывается следующей аддитивной моделью:

$$Y_t = -12707 \cdot t + 665390 + 50000 \cdot Sin((2\pi/12)\ t + 2\pi \cdot 2{,}85/4) + e_t\ . \quad (3.2)$$

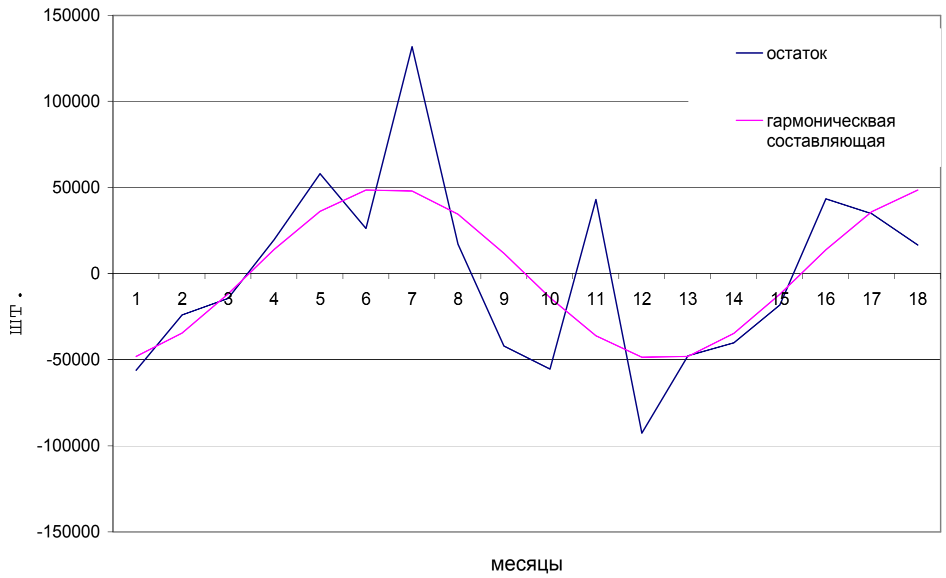


Рис. 3.4. График центрированного ряда ($V_t + e_t$) с наложением гармонической компоненты $V_t = 50000 \cdot Sin((2\pi/12)\ t + 2\pi \cdot 2{,}85/4)$

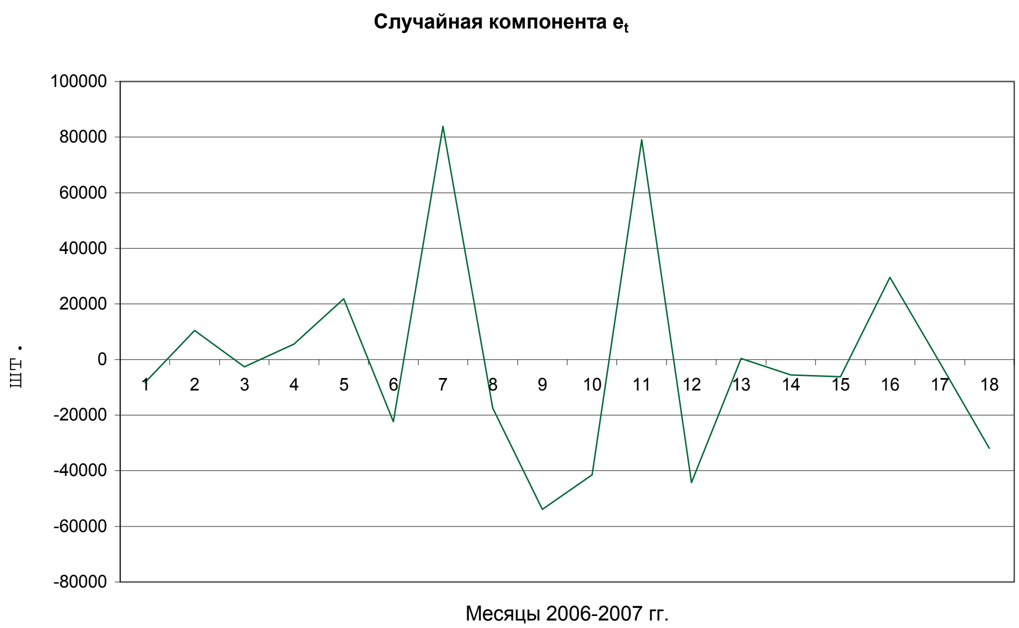


Рис. 3.5. График случайной компоненты временного ряда выработки продукции

Адекватность модели (3.2) оцениваем по результатам анализа случайной компоненты e_t,. Проверяем выполнение предпосылок МНК[1]:

- случайность остатков модели определяем по числу точек перегиба

$$p = 11 > p_к = 9;$$

- соответствие распределения нормальному закону определяем по *R/S* критерию:

расчетное значение *R/S* равно 3,69, находится в области критических границ для уровня значимости 0,05, равной $R/S = 3{,}3 – 4{,}21$;

- равенство нулю математического ожидания остатка определяем с помощью *t*-критерия Стьюдента:

$t_р = 0{,}48 \leq t_{кр} = 2{,}1$ (для уровня значимости 0,05);

- независимость значений уровней случайной компоненты определяем по *d*-критерию Дарбина – Уотсона:

$dw = 1{,}4 > d_2 = 1{,}39$ (для уровня значимости 0,05 имеем $d_1 = 1{,}16$; $d_2 = 1{,}39$).

Все предпоссылки м.н.к. выполняются, что подтверждает адекватность разработанной модели (3.2).

Оценим точность разработанной модели. Для этого вычисляем среднюю абсолютную и среднюю относительную ошибку. Расчеты показали следующие результаты:

- средняя абсолютная ошибка разработанной модели равна 25877,8 шт.;
- средняя относительная ошибка равна 4,7 %.

Приводим интерпретацию результатов исследований с учетом особенностей анализируемого производственного процесса. В рассматриваемом временном интервале работа производства характеризуется некоторой нестабильностью. Среднее абсолютное уменьшение выработки изделий в течение месяца составляет

$$\Delta y_{ср} = 12707 \text{ шт.}$$

Темп уменьшения выработки изделий в последнем месяце 2007 г. составил величину $12707/449371 \cdot 100 = 2{,}83$ %.

Сезонная компонента V_t отражает увеличение выработки изделий в зимние месяцы года (декабрь – январь) и уменьшение в летние месяцы (июнь – июль) на величину, примерно равную 50000 шт./мес. Причинами могут быть колебание спроса, а также влияние климатических условий на технологический процесс изготовления изделий.

3.3. Задание к лабораторной работе

1. Построить график анализируемого временного ряда и высказать гипотезу о возможности описания временного ряда моделью аддитивного случайного процесса.

2. Оценить составляющие аддитивной модели: трендовую, сезонную компоненты и случайную составляющую.

3. Оценить точность аппроксимации временного ряда моделью аддитивного случайного процесса.

4. Проверить выполнение предпосылок МНК по результатам анализа случайной компоненты.

5. Составить отчет по выполненным исследованиям.

3.4. Порядок выполнения лабораторной работы

1. Получить от преподавателя задание на лабораторную работу.

2. Оценить по графику возможность описания временного ряда моделью аддитивного случайного процесса.

3. Подобрать простейшую трендовую составляющую анализируемого временного ряда по критерию минимальной остаточной дисперсии.

4. Выделить периодическую составляющую во временном ряду.

5. Оценить параметры случайной компоненты.

6. Оценить точность описания моделью анализируемого временного ряда.

7. Проверить выполнение предпосылок МНК по результатам анализа случайной компоненты.

8. Сделать выводы по результатам исследований.

9. Составить отчет по выполненной лабораторной работе.

3.5. Содержание отчета

1. Задание на лабораторную работу.

2. График анализируемого временного ряда и выдвижение гипотезы о возможности описания временного ряда моделью аддитивного случайного процесса.

3. Таблица подбора вида трендовой составляющей временного ряда. Математическое описание тренда.

4. График центрированного временного ряда.

5. Автокорреляционная функция центрированного временного ряда и ее математическое описание.

6. График случайной компоненты временного ряда и ее оценки.

7. Аддитивная модель анализируемого временного ряда с оценками ее точности.

8. Результаты проверки выполнения предпосылок МНК.

9. Выводы по результатам исследований временного ряда.

3.6. Вопросы для самоконтроля

1. Модель аддитивного случайного процесса, интерпретация ее компонент.

2. Чем вызывается трендовая составляющая во временном ряду, ее аппроксимация?

3. Чем может вызываться периодическая составляющая во временном ряду, ее аппроксимация?

4. Как оценить случайную компоненту во временном ряду и чем она может вызываться?

5. Как оценивается точность разработанной модели временного ряда?

6. Для чего проверяют выполнение предпосылок МНК?

7. По каким пунктам проверяется выполнение предпосылок МНК?

3.7. Список рекомендуемой литературы

1. Эконометрика: учебник / под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 2005. – 576 с.

2. Яновский, Л. П. Введение в эконометрику: учеб. пособие / Л. П. Яновский, А. Г. Буховец; под ред. Л. П. Яновского. – 2-е изд. доп. – М. : КНОРУС, 2007. – 256 с. – ISBN 5-85971-279-0.

3.8. Варианты заданий

Вариант 1. Уровень дефектности PPM вырабатываемой продукции участка производства конкретного типа изделия по месяцам приведен в таблице.

Месяц	1	2	3	4	5	6	7	8	9
Выработка, шт.	4887	1148	1132	1883	1889	1371	452	1645	2797
Месяц	10	11	12	13	14	15	16	17	18
Выработка, шт.	1505	3071	0	738	567	2428	2061	3393	3882

Построить модель временного ряда и оценить ее точность.

Вариант 2. Выход годной продукции на производственном участке по месяцам приведен в таблице.

Месяц	1	2	3	4	5	6
Годная продукция, %	94,89	93,90	94,75	95,23	95,09	94,04
Месяц	7	8	9	10	11	12
Годная продукция, %	93,54	94,11	94,61	93,04	92,91	92,78
Месяц	13	14	15	16	17	18
Годная продукция, %	93,54	93,23	91,20	92,89	92,38	91,93

Построить модель временного ряда и оценить ее точность.

Вариант 3. Коэффициент использования оборудования на производственном участке по месяцам приведен в таблице.

Месяц	1	2	3	4	5	6
Использование оборудования	0,8895	0,8615	0,8492	0,8764	0,8563	0,8769
Месяц	7	8	9	10	11	12
Использование оборудования	0,8352	0,8612	0,8591	0,8532	0,8493	0,8479
Месяц	13	14	15	16	17	18
Использование оборудования	0,8524	0,8840	0,8485	0,8485	0,8203	0,8598

Построить модель временного ряда и оценить ее точность.

Вариант 4. Ритмичность процесса изготовления продукции на участке приведена в таблице, шт./ч.

Месяц	1	2	3	4	5	6
шт./ч.	1 190	1 142	1 132	1 176	1 194	1 150
Месяц	7	8	9	10	11	12
шт./ч.	1 164	1 164	1 252	1 198	1 195	1 147
Месяц	13	14	15	16	17	18
шт./ч.	1 116	1 155	1 092	1 094	1 156	973

Построить модель временного ряда и оценить ее точность.

Вариант 5. Среднее часовое потребление газа по дням промышленным предприятием приведено в таблице. Построить модель временного ряда и оценить ее точность.

Дата	Расход газа, м^3/ч	Дата	Расход газа, м^3/ч
13.09.2004 0:00	5000	21.09.2004 0:00	5150
14.09.2004 0:00	5100	22.09.2004 0:00	5056,4
15.09.2004 0:00	5174,7	23.09.2004 0:00	5184,2
16.09.2004 0:00	5183,5	24.09.2004 0:00	5179,7
17.09.2004 0:00	5179,1	25.09.2004 0:00	5082,8
18.09.2004 0:00	5159	26.09.2004 0:00	5164,4
19.09.2004 0:00	5150	27.09.2004 0:00	5149,5
20.09.2004 0:00	5150	28.09.2004 0:00	5167,1

Окончание

Дата	Расход газа, м³/ч	Дата	Расход газа, м³/ч
29.09.2004 0:00	5180,8	21.10.2004 0:00	5180
30.09.2004 0:00	5174,9	22.10.2004 0:00	5180
01.10.2004 0:00	5192,4	23.10.2004 0:00	5188,2
02.10.2004 0:00	5183,7	24.10.2004 0:00	5379,6
03.10.2004 0:00	5195	25.10.2004 0:00	5287,3
04.10.2004 0:00	5193,9	26.10.2004 0:00	5308,2
05.10.2004 0:00	5185,6	27.10.2004 0:00	5295,8
06.10.2004 0:00	5187,3	28.10.2004 0:00	5290
07.10.2004 0:00	5191,8	29.10.2004 0:00	5290
08.10.2004 0:00	5190	30.10.2004 0:00	5321
09.10.2004 0:00	5190	31.10.2004 0:00	5358
10.10.2004 0:00	5188,3	01.11.2004 0:00	5359,8
11.10.2004 0:00	5260,2	02.11.2004 0:00	5348,7
12.10.2004 0:00	5144,3	03.11.2004 0:00	5241,8
13.10.2004 0:00	5140	04.11.2004 0:00	5240
14.10.2004 0:00	5140	05.11.2004 0:00	5240
15.10.2004 0:00	5140	06.11.2004 0:00	5240
16.10.2004 0:00	5140	07.11.2004 0:00	5253,9
17.10.2004 0:00	5140	08.11.2004 0:00	5257
18.10.2004 0:00	5186,1	09.11.2004 0:00	5258,1
19.10.2004 0:00	5180	10.11.2004 0:00	5263,6
20.10.2004 0:00	5180	11.11.2004 0:00	5255,3

Вариант 6. В таблице приведены ежемесячные данные о случаях заболевания работников промышленного предприятия в пересчете на 100 работающих. Построить модель временного ряда и оценить ее точность.

Месяц	Случаи заболевания	Месяц	Случаи заболевания	Месяц	Случаи заболевания
1	7	9	3,3	17	2,8
2	7,9	10	5,4	18	9
3	2,5	11	5,1	19	5,2
4	4,7	12	3,2	20	4,6
5	2,5	13	3,4	21	5,8
6	3,2	14	3,9	22	6
7	3,6	15	6,3	23	5,3
8	3,8	16	5,5		

Вариант 7. Состояние охраны труда на промышленном предприятии оценивается числом нерабочих дней по больничным листам в пересчете на 100 работающих. Построить модель временного ряда и оценить ее точность.

Месяц	1	2	3	4	5	6	7	8	9	10	11
Больничные листы	43,6	49,1	21,9	30,4	37,7	40,8	30,9	90,5	49,4	31,6	59,2
Месяц	12	13	14	15	16	17	18	19	20	21	22
Больничные листы	33,9	63	71,1	30,5	76,5	59,2	51,1	59,1	80,8	60,3	

Вариант 8. Концентрация вредных выбросов промышленного предприятия в окружающую природную среду, мг/м^3, характеризуется данными, приведенными в таблице. Построить модель временного ряда и оценить ее точность.

Месяц	Концентрация оксида азота	Месяц	Концентрация оксида азота
1	39,37	13	22,90
2	29,37	14	21,29
3	35	15	19,53
4	35	16	23,44
5	36,25	17	15,94
6	26,87	18	25,62
7	34,37	19	17
8	29,37	20	18,12
9	27,5	21	14
10	26,87	22	5,56
11	23,75	23	6,5
12	30	24	9,13

Вариант 9. Выработка изделий на первом участке цеха по сменам, шт., характеризуется данными, приведенными в таблице. Построить модель временного ряда и оценить ее точность.

Смена	Выработка	Смена	Выработка	Смена	Выработка
1	6698	5	7384	9	8506
2	6740	6	7703	10	8737
3	6931	7	8005	11	8842
4	7089	8	8163	12	9022

Окончание

Смена	Выработка	Смена	Выработка	Смена	Выработка
13	9425	20	10867	27	12336
14	9752	21	10746	28	12568
15	9602	22	10770	29	12903
16	9711	23	10782	30	13027
17	10121	24	11179	31	13051
18	10425	25	11617	32	12889
19	10744	26	12015		

Вариант 10. Выработка изделий на втором участке цеха по сменам (шт.) характеризуется данными, приведенными в таблице. Построить модель временного ряда и оценить ее точность.

Смена	Выработка	Смена	Выработка	Смена	Выработка
1	7264	12	10111	23	12146
2	7382	13	10414	24	12349
3	7583	14	11013	25	13029
4	7718	15	10832	26	13258
5	8140	16	10906	27	13552
6	8508	17	11192	28	13545
7	8822	18	11406	29	13890
8	9114	19	11851	30	14030
9	9399	20	12039	31	14154
10	9606	21	12005	32	13987
11	9875	22	12156		

Лабораторная работа № 4
ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ

4.1. Цель работы

Изучение алгоритмов прогнозирования и оценка точности прогноза по ретроспективным данным показателей работы организации.

4.2. Теоретические сведения

Исследование динамики показателей работы организации, выявление и характеристика основных тенденций развития и моделей взаимосвязи дают основание для прогнозирования – определения ожидаемых показателей.

Важное место в прогнозировании занимают статистические методы прогноза. Применение прогнозирования предполагает, что закономерность развития, действующая в прошлом внутри ряда динамики, сохранится и в прогнозируемом будущем. Теоретической основой распространения тенденции на будущее является свойство инерционности, которое позволяет выявить сложившиеся взаимосвязи между уровнями динамического ряда, а также между группой взаимосвязанных рядов динамики. Надежность прогноза возрастает для сопоставимых рядов динамики, полученных на основе использования единой методологии. Точность прогноза зависит от периода упреждения: чем короче период упреждения, тем более надежные и точные результаты дает прогнозирование. За короткий период не успевают сильно измениться условия работы организации и характер ее динамики.

Наиболее часто используются простейшие алгоритмы прогнозирования:

- по среднему абсолютному приросту при линейной тенденции развития показателя во времени;
- по среднему темпу роста, когда тенденция ряда характеризуется показательной кривой;
- аналитическим описанием линии тренда, когда на показатель оказывает влияние множество факторов, и ее рассматривают в виде временной функции;
- по корреляционным связям между показателями ряда на ограниченном по времени интервале наблюдения;
- по среднему уровню ряда динамики в случае стационарного характера изменения во времени анализируемого показателя и др.

Прогнозирование по среднему абсолютному приросту проводится по формуле
$$y_{\text{пр}} = y + \Delta y \cdot t,$$
где y – последний уровень ряда динамики; t – период упреждения (прогноз); Δy – средний абсолютный прирост анализируемого показателя.

Прогнозирование по среднему темпу роста выполняется по формуле
$$y_{\text{пр}} = y \cdot T_{\text{р}}^{t},$$
где $T_{\text{р}}$ – средний темп роста показателя.

Простейшим считается прогноз средним значением уровня ряда
$$y_{\text{пр}} = y_{\text{ср}},$$
где $y_{\text{ср}}$ – среднее значение уровня анализируемого ряда динамики.

При подборе кривых для описания тренда многочленами стоит задача выбора степени многочлена. Для определения степени многочлена вычисляется последовательность разностей временного ряда

$$\Delta_i^{(1)} = y(t_i) - y\,(t_{i\text{-}1}), \quad i = 2.\ 3,\ \ldots,\ n$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\Delta_i^{(k)} = \Delta_i^{(k\text{-}1)} \text{-}\ \Delta_{i\text{-}1}^{(k\text{-}1)}, \qquad i = k,\ k{+}1,\ k{+}2,\ldots,\ n.$$

Порядок разностей, при котором они становятся примерно одинаковыми, берется в качестве степени аппроксимирующего многочлена. Так, если малыми оказываются вторые разности $\Delta_i^{(2)}$, i = 1, 2. 3, …, n, то для прогнозирования тренда берется многочлен второй степени

$$y\,(t) = a + bt + ct^2.$$

Прогнозирование стационарных случайных процессов по корреляционным связям между показателями ряда на ограниченном по времени интервале наблюдения выполняется по формуле

$$y_{\text{пр}} = (y_0\,(1 + y_1 \cdot y_4 + y_2 \cdot y_3 + y_3 \cdot y_2 + y_4 \cdot y_1\,))\ /\ (1 + y_4^2 + y_3^2 + y_2^2 + y_1^2\,), \quad (4.1)$$

где y_0, y_1, y_2, y_3, y_4 – уровни динамического ряда с показателями работы организации, y_4 соответствует последнему значению уровня ряда.

Для оценки точности прогноза с помощью вычислительного эксперимента анализируемая выборка временного ряда разбивается на две части: начало временного ряда используется для обучения и выбора алгоритма прогнозирования, а конец выборки используется для оценки точности прогнозирования. Абсолютная погрешность прогнозирования вычисляется по формуле

$$\Delta y_{\text{пр}} = y_{\text{пр}} - y, \qquad (4.2)$$

где $\Delta y_{\text{пр}}$ – абсолютная погрешность прогнозирования, отклонение прогнозного уровня относительно истинного значения; y – истинный уровень ряда; $y_{\text{пр}}$ – спрогнозированное значение уровня ряда.

Для анализа рядов динамики и их прогнозирования используются офисные информационные технологии, реализованные в электронной таблице *EXCEL*, а также ППП типа *STATISTICA, Matlab, STATISTICA Neural Networks* и др.

Пример. Имеются статистические данные за 24 месяца о негативном влиянии производства листового стекла на окружающую природную среду. Содержание концентрации оксида азота в дымовых газах отражено на рис. 4.1.

Выполнить прогнозирование временного ряда с использованием аналитического описания линии тренда: экспонентой, моделью на нейронных сетях и с использованием нечетких множеств. Выбрать наиболее точный алгоритм прогнозирования по величине абсолютной погрешности прогноза на интервале прогнозирования в четыре месяца.

Рис. 4.1. Содержание концентрации оксида азота в дымовых газах

Обучение алгоритма прогнозирования проводим по первым 20 данным временного ряда, а точность прогнозирования оценим по последним четырем данным.

1) Визуальный анализ графика на рис. 4.1 показывает возможность описания тренда многочленом. Оценим точность прогнозирования при аналитическом описании линии тренда простой экспонентой. Для этого воспользуемся ППП *STATISTICA*. Результаты математического описания линии тренда экспонентой и прогноза на четыре шага вперед отражены на рис. 4.2. Параметры алгоритма: Эксп. сглажив. $S0$ = 39,37 $T0$ = -1,12 (прогноз выбросов) Лин. тренд, нет сезон.; Альфа = 0,100 Гамма = 0,100 *VAR*1.

Вычислим погрешность прогнозирования по формуле (4.1) с 21-го по 24-й месяцы временного ряда (табл. 4.1).

Таблица 4.1. Погрешность прогнозирования

Месяц	21	22	23	24
Фактическое значение	14	5,56	6,5	9,12
Расчетный прогноз	15,3	14,24	13,15	12,1
Погрешность прогноза	1,3	8,68	6,65	2,98

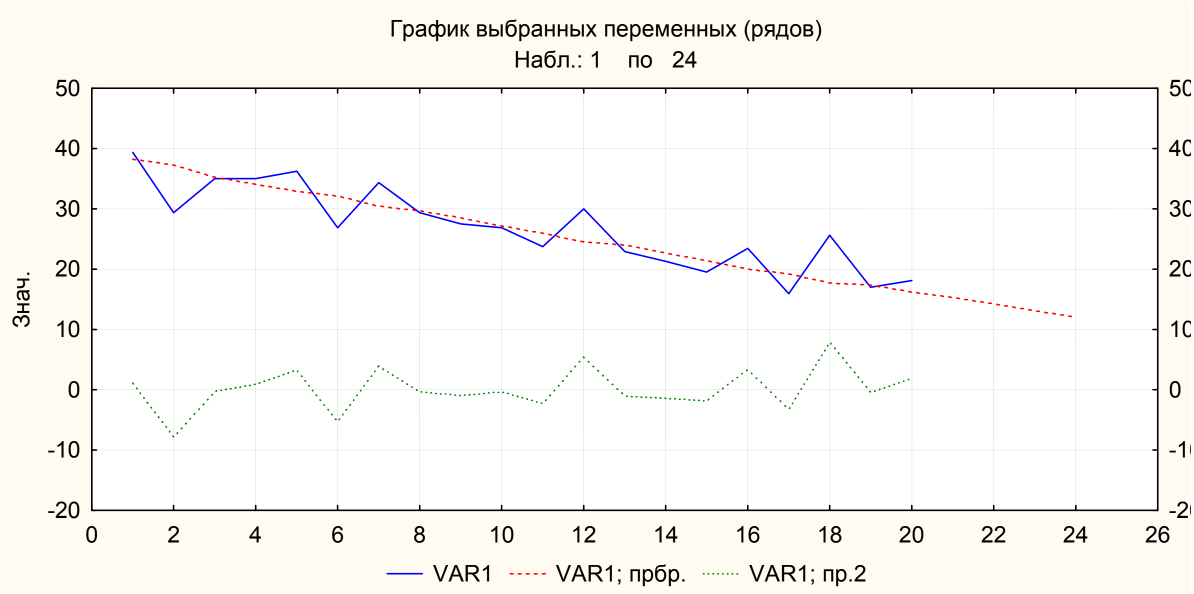


Рис. 4.2. Прогнозирование тренда

Наибольшая погрешность прогнозирования концентрации вредных веществ в выбросах линейным трендом равна $\Delta y_{пр}$ = 8,68 мг/дм^3.

2) Проведем анализ точности прогнозирования временного ряда с использованием моделей на нечетких множествах в ППП *Matlab*. Для этих целей составляем программу в виде *M*-файла (программа приведена ниже).

% ПОСТРОЕНИЕ МОДЕЛИ МВМВ:
***fis = genfis*2(*vremj, rjd*, 0.44);**
% ГДЕ:
%- ***vremj, rjd*** - МАТРИЦЫ ВХОДНЫХ/ВЫХОДНЫХ ИСХОДНЫХ ДАННЫХ
%- **0.44** - РАЗМЕР КЛАСТЕРА ДЛЯ ***genefis2***
% ПРОСМОТР МОДЕЛИ В РЕДАКТОРЕ МВМВ:
***fuzzy*(*fis*);**
% ВЫПОЛНИТЬ НА МОДЕЛИ МВМВ ПРОГНОЗ ДЛЯ НАБЛЮДЕНИЙ 21-22:
***chkfuzout = evalfis*(*progvremj, fis*);**
% где ***progvremj***-интервалы времени прогнозирования (21-24);
% ***chkfuzout***-прогнозное значение временного ряда;
% РАСЧЕТ ПО МОДЕЛИ МВМВ ВРЕМЕННОГО РЯДА ПРИ ОБУЧЕНИИ (1-20):
***model = evalfis*(*vremj, fis*);**
% ГДЕ: ***model***- СМОДЕЛИРОВАННЫЙ РЯД НА ОБУЧАЮЩЕЙ ВЫБОРКЕ (1-20);
% РАСЧЕТ ВРЕМЕННОГО РЯДА С ПРОГНОЗОМ (1-24):
graf=unnamed*1 + *unnamed;
% ГДЕ ***unnamed1***- ФАЙЛ ***model*** ДОПИСАН ЧЕТЫРМЯ НУЛЯМИ В КОНЦЕ;
% ***unnamed***-ФАЙЛ С 20-ю НУЛЯМИ В НАЧАЛЕ ДОПИСАН В КОНЦЕ ФАЙЛОМ ***chkfuzout*;**

```
% ПОСТРОЕНИЕ СОВМЕЩЕННОГО ГРАФИКА РАСЧЕТНОГО ВРЕМЕННОГО РЯДА
% С ИСХОДНЫМ РЯДОМ (1-24):
i=1;
plot([graf(:,i), prognoz(:,i)]);
   legend('show');
```

Результаты моделирования временного ряда отображены на рис. 4.3.

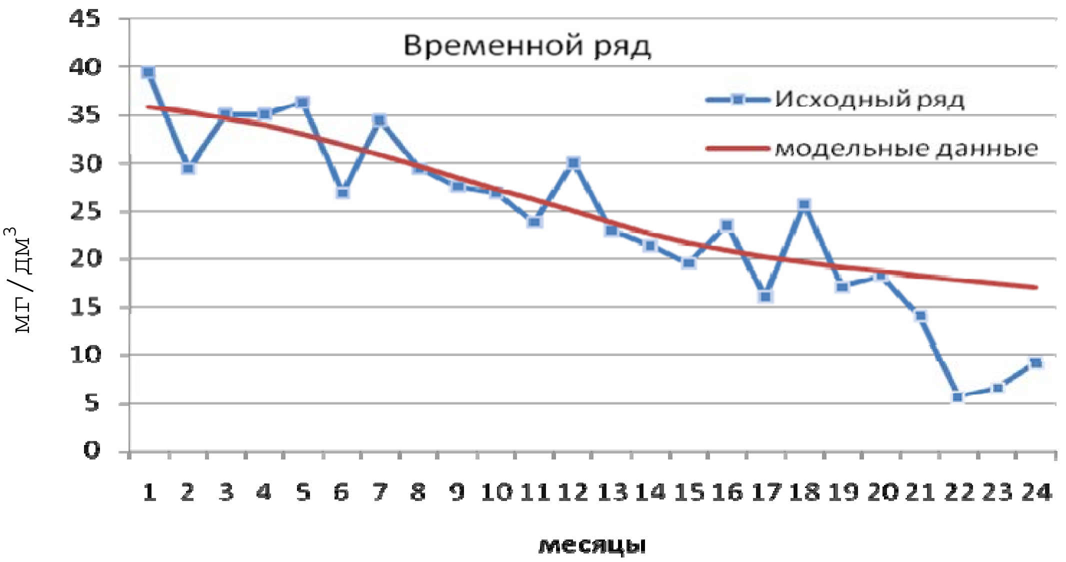


Рис. 4.3. Прогнозирование концентрации оксида азота в дымовых газах моделью на нечетких множествах

Расчеты погрешности прогнозирования представлены в табл. 4.2. Наибольшая погрешность прогнозирования концентрации вредных веществ в выбросах моделью на нечетких множествах составила $\Delta y_{пр}$ = =12,19 мг/дм^3.

Таблица 4.2. Погрешность прогнозирования нечеткими множествами

Месяц	Фактическое значение	Расчетный прогноз	Погрешность прогноза
21	14	18,18	4,18
22	5,56	17,75	12,19
23	6,5	17,34	10,84
24	9,12	16,92	7,8

3) Опишем временной ряд нейронными сетями с использованием ППП *STATISTICA Neural Networks*. Для этих целей была выбрана сеть *MLP* с одним входом и 13 нейронами в промежуточном слое (рис. 4.4).

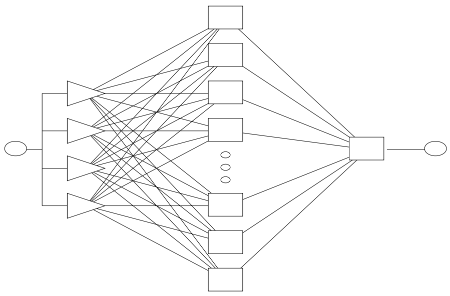


Рис. 4.4. Структура сети

Точность модели характеризуется следующими данными:

Регрессионная статистика

	Tr. VAR1	Ve. VAR1	Te. VAR1
Data Mean	21.69637	23.95	19.45725
Data S.D.	11.19525	5.918922	7.485738
Error Mean	1.956998	-0.08533	5.082375
Error S.D.	6.769452	4.853207	7.110437
Abs E. Mean	5.758195	3.875791	5.375157
S.D. Ratio	0.6046718	0.8199477	0.9498645
Correlation	0.9013178	0.577145	0.3153473

Ошибка обучения равна 6,628, ошибка контроля 4,43 и ошибка тестирования 8,244. Результаты моделирования временного ряда приведены на рис. 4.5.

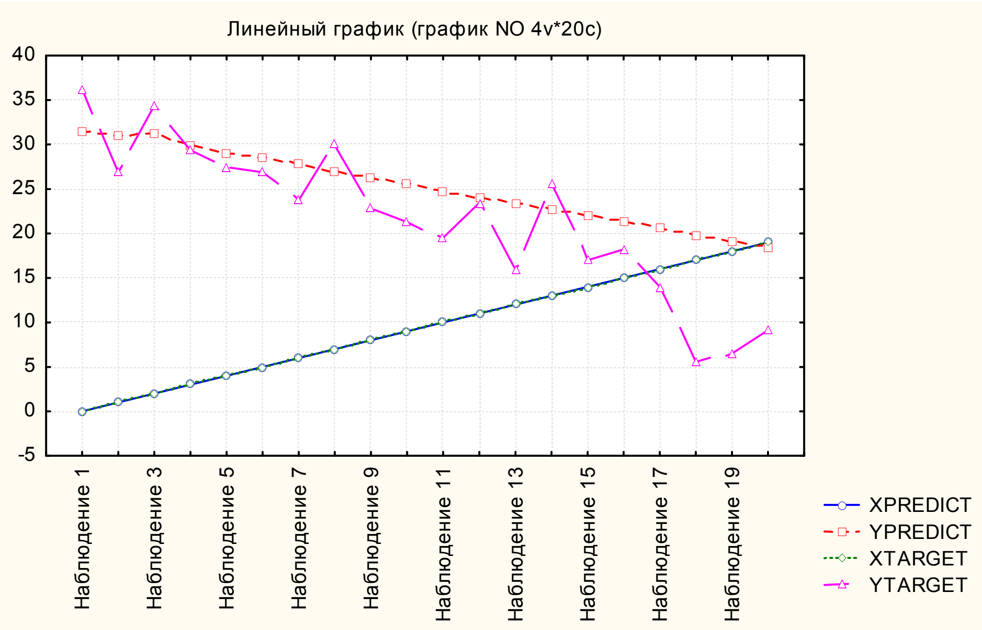


Рис. 4.5. Прогнозирование концентрации оксида азота в дымовых газах моделью на нейронных сетях

Проведенные исследования показали, что наибольшая точность прогнозирования достигается при аналитическом описании линии тренда простой экспонентой.

4.3. Задание к лабораторной работе

1. Выбрать тип кривой, соответствующей характеру изменения временного ряда.

2. Оценить точность прогнозирования временного ряда на интервале 3 – 5 шагов при описании тренда многочленом.

3. Оценить точность прогнозирования временного ряда с использованием модели на нечетких множествах.

4. Оценить точность прогнозирования временного ряда с использованием модели на нейронных сетях.

5. Выбрать наиболее точный алгоритм краткосрочного прогнозирования на интервале 3 – 5 шагов для исследуемого временного ряда.

6. Составить отчет по выполненным исследованиям.

4.4. Порядок выполнения лабораторной работы

1. Получить от преподавателя задание на проведение лабораторной работы.

2. На основе графического изображения временного ряда, формы его корреляционного поля выбрать тип кривой для описания временного тренда.

3. Выбрать степень многочлена для описания временного тренда по критерию минимальной погрешности прогноза на интервале 3 – 5 шагов.

4. Построить график временного ряда с наложением на него линии тренда.

5. Построить модель на нечетких множествах для описания временного ряда и рассчитать прогнозное значение ряда на интервале 3 – 5 шагов.

6. Построить график временного ряда с наложением линии тренда, рассчитанной по модели на нечетких множествах.

7. Построить модель на нейронных сетях для описания временного ряда и рассчитать прогнозное значение ряда на интервале 3 – 5 шагов.

8. Построить график временного ряда с наложением линии тренда, рассчитанной по модели на нейронных сетях.

9. Сравнить по точности алгоритмы краткосрочного прогнозирования временного ряда на интервале 3 – 5 шагов.

10. Оформить отчет по лабораторной работе.

4.5. Содержание отчета

1. Задание на лабораторную работу.

2. График анализируемого временного ряда и выдвижение гипотезы о возможности описания тренда временного ряда многочленом.

3. График временного ряда с наложением линии тренда, описываемого многочленом.

4. Таблица расчета погрешности прогнозирования на интервале 3 – 5 шагов.

5. *M*-файл программы прогнозирования в ППП *Matlab* с использованием модели на нечетких множествах.

6. График временного ряда с наложением линии тренда, описываемой моделью на нечетких множествах.

7. Таблица расчета погрешности прогнозирования на интервале 3 – 5 шагов с использованием модели на нечетких множествах.

8. Модель на нейронных сетях, ее структура и регрессионная статистика.

9. График временного ряда с наложением линии тренда, описываемой моделью на нейронных сетях.

10. Анализ результатов исследования по выбору алгоритма краткосрочного прогнозирования.

4.6. Вопросы для самоконтроля

1. Каковы основные принципы прогнозирования временных рядов?

2. Что такое метод и модель прогнозирования?

3. Какие подходы можно использовать для выделения тренда временного ряда?

4. Когда используется прогнозирование по среднему абсолютному изменению уровня временного ряда?

5. Когда используется прогнозирование по среднему темпу роста уровня временного ряда?

6. Когда используется прогнозирование с использованием аналитических временных функций?

7. Когда используется прогнозирование по корреляционным зависимостям?

8. Когда используется прогнозирование по среднему значению уровня временного ряда?

9. Что понимается под точностью прогнозирования временного ряда и как она оценивается?

10. На каких предпосылках основан прогноз временных рядов?

11. Как зависит точность прогнозирования от интервала упреждения и почему?

4.7. Список рекомендуемой литературы

1. Яновский, Л. П. Введение в эконометрику: учеб. пособие / Л. П. Яновский, А. Г. Буховец ; под ред. Л. П. Яновского. – 2-е изд. доп. – М. : КНОРУС, 2007. – 256 с. – ISBN 5-85971-270-7.

2. Осовский, С. Нейронные сети для обработки информации / С. Осовский ; пер. с польск. И. Д. Рудинского. – М. : Финансы и статистика, 2002. – 344 с. – ISBN 5-279-02567-4.

4.8. Варианты заданий

Вариант 1. Имеются следующие данные о розничном товарообороте во всех каналах реализации в регионе, млрд руб.

Месяц года	1998 г.	1999 г.	2000 г.
Январь	7,4	7,8	8,3
Февраль	7,9	8,2	8,6
Март	8,7	9,2	9,7
Апрель	8,2	8,6	9,1
Май	7,9	8,3	8,8
Июнь	8,2	8,7	9,1
Июль	8,3	8,8	9,3
Август	8,8	9,3	9,9
Сентябрь	8,7	8,9	9,3
Октябрь	8,8	8,2	9,9
Ноябрь	8,3	8,8	9,8
Декабрь	9,0	9,5	9,3

Изучите общую тенденцию изменения розничного товарооборота, выберите алгоритм и оцените точность прогнозирования на интервале в три месяца.

Вариант 2. Реализация скота и птицы на убой в живой массе в Российской Федерации, млн т.

Месяц года	1993 г.	1994 г.	1995 г.	1996 г.	1997 г.
Январь	510,8	414,7	372,3	286,1	222,5
Февраль	543,2	450,7	342,5	309,1	244,2
Март	555,1	476,7	349,2	333,1	253,5
Апрель	521,4	465,1	355,8	330,3	270,4
Май	503,5	434,9	341,6	299,2	244,1
Июнь	501,5	427,0	340,7	297,1	245,1
Июль	486,5	419,1	331,1	269,6	252,8

Окончание

Месяц года	1993 г.	1994 г.	1995 г.	1996 г.	1997 г.
Август	522,3	493,0	397,0	297,7	251,3
Сентябрь	636,5	501,1	392,2	332,2	263,8
Октябрь	641,5	523,4	396,4	303,2	269,7
Ноябрь	576,0	470,5	348,0	288,0	249,7
Декабрь	606,5	484,2	361,5	319,1	268,7

Изучите общую тенденцию изменения реализации скота и птицы на убой в живой массе, выберите алгоритм и оцените точность прогнозирования на интервале в три месяца.

Вариант 3. Производство электроэнергии в Российской Федерации, млрд кВт·ч.

Месяц года	1993 г.	1994 г.	1995 г.	1996 г.	1997 г.
Январь	96	91,3	90	91	88,1
Февраль	89	86,6	78,8	84,5	78
Март	92,3	87,2	82,6	82,5	76,5
Апрель	80,7	71,6	68,6	70,3	67,3
Май	70,1	62,6	62,7	59,8	58,2
Июнь	64,2	56,8	57,5	55	53,7
Июль	64	56,6	58,3	55,8	55,4
Август	64,9	58,3	59,6	56,2	56,6
Сентябрь	69,4	60,9	61,8	60,7	60,9
Октябрь	80,8	71,9	73	71,8	71,7
Ноябрь	91,7	81,3	79,6	75	78,6
Декабрь	94,4	91,2	89,3	84,7	88,7

Изучите общую тенденцию изменения производства электроэнергии в Российской Федерации, выберите алгоритм и оцените точность прогнозирования на интервале в три месяца.

Вариант 4. Производство газа в Российской Федерации, млн т.

Месяц года	1993 г.	1994 г.	1995 г.	1996 г.	1997 г.
Январь	57,2	57,2	57	56,8	57,4
Февраль	52,1	51,7	51,2	53,2	51,5
Март	57	56,1	55,2	56,3	54,2
Апрель	54,2	52,2	48,8	51,7	48,7
Май	51,8	49,5	48,9	46,9	45
Июнь	46,2	45,4	43,5	44,3	39,3
Июль	48,5	45,5	44	44	37,9

Окончание

Месяц года	1993 г.	1994 г.	1995 г.	1996 г.	1997 г.
Август	45,7	44,6	44	42,2	37,5
Сентябрь	42,5	44	43,2	44,2	40,7
Октябрь	52,1	50,4	50,8	52,5	48,6
Ноябрь	54,6	53,3	52,8	52,6	56,1
Декабрь	56,2	56,8	56	56	56,9

Изучите общую тенденцию изменения производства газа в Российской Федерации, выберите алгоритм и оцените точность прогнозирования на интервале в три месяца.

Вариант 5. Удельный расход газа в производстве листового стекла на технологические цели приведен ниже, м^3/кг.

Дата	Расход газа, м^3/кг	Дата	Расход газа, м^3/кг
18.05.2004 0:00	0,19453454	09.06.2004 0:00	0,195589645
19.05.2004 0:00	0,195360924	10.06.2004 0:00	0,194470925
20.05.2004 0:00	0,195007174	11.06.2004 0:00	0,194749403
21.05.2004 0:00	0,195952313	12.06.2004 0:00	0,195121951
22.05.2004 0:00	0,198176046	13.06.2004 0:00	0,19530876
23.05.2004 0:00	0,193468765	14.06.2004 0:00	0,194935499
24.05.2004 0:00	0,19402585	15.06.2004 0:00	0,179324427
25.05.2004 0:00	0,193804223	16.06.2004 0:00	0,198268827
26.05.2004 0:00	0,195154067	17.06.2004 0:00	0,191687916
27.05.2004 0:00	0,194964098	18.06.2004 0:00	0,191437321
28.05.2004 0:00	0,195488261	19.06.2004 0:00	0,195237369
29.05.2004 0:00	0,19378626	20.06.2004 0:00	0,194239847
30.05.2004 0:00	0,190760402	21.06.2004 0:00	0,194268258
31.05.2004 0:00	0,192065041	22.06.2004 0:00	0,19433652
01.06.2004 0:00	0,194137667	23.06.2004 0:00	0,195210728
02.06.2004 0:00	0,195022989	24.06.2004 0:00	0,172615091
03.06.2004 0:00	0,195769452	25.06.2004 0:00	0,172601432
04.06.2004 0:00	0,19504689	26.06.2004 0:00	0,172354623
05.06.2004 0:00	0,19526351	27.06.2004 0:00	0,172683859
06.06.2004 0:00	0,193901141	28.06.2004 0:00	0,172436814
07.06.2004 0:00	0,194579048	29.06.2004 0:00	0,172026641
08.06.2004 0:00	0,194825776	30.06.2004 0:00	0,195690831

Изучите общую тенденцию изменения удельного расхода газа, выберите алгоритм и оцените точность прогнозирования на интервале в 10 суток.

Вариант 6. Выработка стекла на производстве посменно в 1998 году составляла, кг/см.

Дата	Выработка стекла, кг/см	Дата	Выработка стекла, кг/см
00:00/21/07/1998	121000,0	00:00/13/08/1998	120400,0
00:00/22/07/1998	120000,0	00:00/14/08/1998	120600,0
00:00/23/07/1998	120000,0	00:00/15/08/1998	120000,0
00:00/24/07/1998	120300,0	00:00/16/08/1998	120000,0
00:00/25/07/1998	121200,0	00:00/17/08/1998	120600,0
00:00/26/07/1998	121000,0	00:00/18/08/1998	120800,0
00:00/27/07/1998	121300,0	00:00/19/08/1998	120000,0
00:00/28/07/1998	120000,0	00:00/20/08/1998	119500,0
00:00/29/07/1998	120300,0	00:00/21/08/1998	120000,0
00:00/30/07/1998	120100,0	00:00/22/08/1998	120500,0
00:00/31/07/1998	120120,0	00:00/23/08/1998	120500,0
00:00/01/08/1998	120000,0	00:00/24/08/1998	120500,0
00:00/02/08/1998	120200,0	00:00/25/08/1998	120800,0
00:00/03/08/1998	120000,0	00:00/26/08/1998	120400,0
00:00/04/08/1998	120100,0	00:00/27/08/1998	120000,0
00:00/05/08/1998	120400,0	00:00/28/08/1998	121100,0
00:00/06/08/1998	120700,0	00:00/29/08/1998	121200,0
00:00/07/08/1998	120300,0	00:00/30/08/1998	119600,0
00:00/08/08/1998	120400,0	00:00/31/08/1998	119200,0
00:00/09/08/1998	120000,0	00:00/01/09/1998	120000,0
00:00/10/08/1998	120100,0	00:00/02/09/1998	120000,0
00:00/11/08/1998	100000,0	00:00/03/09/1998	120300,0
00:00/12/08/1998	120700,0	00:00/04/09/1998	119500,0

Изучите общую тенденцию изменения выработки стекла в производстве, выберите алгоритм и оцените точность прогнозирования на интервале в 10 суток.

Вариант 7. Оптические искажения вырабатываемого листового стекла в проходящем свете составляли, угл. град.

Дата	Оптическое искажение	Дата	Оптическое искажение
00:00/01/08/1998	44,0	00:00/26/08/1998	47,0
00:00/02/08/1998	58,0	00:00/27/08/1998	50,0
00:00/03/08/1998	55,0	00:00/28/08/1998	53,0
00:00/04/08/1998	55,0	00:00/29/08/1998	54,0
00:00/05/08/1998	48,0	00:00/30/08/1998	55,0
00:00/06/08/1998	52,0	00:00/31/08/1998	43,0
00:00/07/08/1998	59,0	00:00/01/09/1998	46,0
00:00/08/08/1998	57,0	00:00/02/09/1998	46,0
00:00/09/08/1998	44,0	00:00/03/09/1998	59,0
00:00/10/08/1998	45,0	00:00/04/09/1998	55,0
00:00/11/08/1998	45,0	00:00/05/09/1998	47,0
00:00/12/08/1998	46,0	00:00/06/09/1998	51,0
00:00/13/08/1998	45,0	00:00/07/09/1998	45,0
00:00/14/08/1998	54,0	00:00/08/09/1998	45,0
00:00/15/08/1998	68,0	00:00/09/09/1998	44,0
00:00/16/08/1998	63,0	00:00/10/09/1998	58,0
00:00/17/08/1998	63,0	00:00/11/09/1998	53,0
00:00/18/08/1998	27,0	00:00/12/09/1998	50,0
00:00/19/08/1998	25,0	00:00/13/09/1998	59,0
00:00/20/08/1998	53,0	00:00/14/09/1998	58,0
00:00/21/08/1998	53,0	00:00/15/09/1998	55,0
00:00/22/08/1998	44,0	00:00/16/09/1998	47,0
00:00/23/08/1998	43,0	00:00/17/09/1998	55,0
00:00/24/08/1998	45,0	00:00/18/09/1998	55,0
00:00/25/08/1998	49,0	00:00/19/09/1998	54,0

Изучите общую тенденцию изменения оптических свойств вырабатываемого стекла, выберите алгоритм и оцените точность прогнозирования на интервале в 10 суток.

Вариант 8. В вырабатываемом листовом стекле обнаружены внешние дефекты, шт./м^2.

Дата	Оптическое искажение	Дата	Оптическое искажение
00:00/01/01/1999	2,3	00:00/26/01/1999	0,6
00:00/02/01/1999	0,7	00:00/27/01/1999	0,5
00:00/03/01/1999	0,9	00:00/28/01/1999	0,6
00:00/04/01/1999	0,4	00:00/29/01/1999	1,2
00:00/05/01/1999	0,3	00:00/30/01/1999	0,6
00:00/06/01/1999	0,6	00:00/31/01/1999	1,8
00:00/07/01/1999	0,8	00:00/01/02/1999	1,4
00:00/08/01/1999	0,9	00:00/02/02/1999	1,4
00:00/09/01/1999	0,8	00:00/03/02/1999	0,4
00:00/10/01/1999	0,4	00:00/04/02/1999	0,6
00:00/11/01/1999	0,6	00:00/05/02/1999	0,7
00:00/12/01/1999	0,4	00:00/06/02/1999	1,4
00:00/13/01/1999	0,5	00:00/07/02/1999	1,1
00:00/14/01/1999	0,5	00:00/08/02/1999	0,9
00:00/15/01/1999	0,7	00:00/09/02/1999	1,1
00:00/16/01/1999	0,5	00:00/10/02/1999	0,7
00:00/17/01/1999	0,2	00:00/11/02/1999	2,5
00:00/18/01/1999	0,1	00:00/12/02/1999	1,0
00:00/19/01/1999	0,0	00:00/13/02/1999	1,2
00:00/20/01/1999	0,2	00:00/14/02/1999	0,7
00:00/21/01/1999	0,6	00:00/15/02/1999	1,2
00:00/22/01/1999	0,2	00:00/16/02/1999	0,5
00:00/23/01/1999	0,1	00:00/17/02/1999	0,9
00:00/24/01/1999	2,8	00:00/18/02/1999	1,4

Изучите общую тенденцию изменения внешних дефектов в вырабатываемом стекле. Выберите алгоритм и оцените точность прогнозирования на интервале в 10 суток.

Вариант 9. Содержание оксида железа в листовом стекле колебалось в следующих пределах, %:

Дата	Содержание оксида железа	Дата	Содержание оксида железа
00:00/01/08/98	0,090	00:00/26/08/98	0,090
00:00/02/08/98	0,092	00:00/27/08/98	0,087
00:00/03/08/98	0,090	00:00/28/08/98	0,086
00:00/04/08/98	0,089	00:00/29/08/98	0,084
00:00/05/08/98	0,088	00:00/30/08/98	0,086
00:00/06/08/98	0,091	00:00/31/08/98	0,086
00:00/07/08/98	0,090	00:00/01/09/98	0,087
00:00/08/08/98	0,089	00:00/02/09/98	0,088
00:00/09/08/98	0,090	00:00/03/09/98	0,089
00:00/10/08/98	0,093	00:00/04/09/98	0,090
00:00/11/08/98	0,090	00:00/05/09/98	0,091
00:00/12/08/98	0,088	00:00/06/09/98	0,091
00:00/13/08/98	0,090	00:00/07/09/98	0,093
00:00/14/08/98	0,092	00:00/08/09/98	0,091
00:00/15/08/98	0,090	00:00/09/09/98	0,093
00:00/16/08/98	0,091	00:00/10/09/98	0,094
00:00/17/08/98	0,091	00:00/11/09/98	0,093
00:00/18/08/98	0,091	00:00/12/09/98	0,095
00:00/19/08/98	0,091	00:00/13/09/98	0,095
00:00/20/08/98	0,096	00:00/14/09/98	0,094
00:00/21/08/98	0,096	00:00/15/09/98	0,092
00:00/22/08/98	0,100	00:00/16/09/98	0,090
00:00/23/08/98	0,100	00:00/17/09/98	0,088
00:00/24/08/98	0,096	00:00/18/09/98	0,087
00:00/25/08/98	0,094	00:00/19/09/98	0,089

Изучите общую тенденцию изменения оксида железа в вырабатываемом стекле, выберите алгоритм и оцените точность прогнозирования на интервале в 10 суток.

Вариант 10. Расход шихты в производстве листового стекла в смену по суткам составлял, т/смена:

Дата	Расход шихты	Дата	Расход шихты
18.05.2004 0:00	513,7	11.06.2004 0:00	768,6
19.05.2004 0:00	752,1	12.06.2004 0:00	768,5
20.05.2004 0:00	727,5	13.06.2004 0:00	805,6
21.05.2004 0:00	682,3	14.06.2004 0:00	752,1
22.05.2004 0:00	719,2	15.06.2004 0:00	793,3
23.05.2004 0:00	715,2	16.06.2004 0:00	756,2
24.05.2004 0:00	628,8	17.06.2004 0:00	706,9
25.05.2004 0:00	669,9	18.06.2004 0:00	727,5
26.05.2004 0:00	690,5	19.06.2004 0:00	785
27.05.2004 0:00	711	20.06.2004 0:00	612,4
28.05.2004 0:00	698,7	21.06.2004 0:00	805,6
29.05.2004 0:00	649,4	22.06.2004 0:00	608,2
30.05.2004 0:00	641,2	23.06.2004 0:00	575,4
31.05.2004 0:00	612,4	24.06.2004 0:00	460,4
01.06.2004 0:00	550,7	25.06.2004 0:00	604,1
02.06.2004 0:00	604,2	26.06.2004 0:00	706,9
03.06.2004 0:00	686,3	27.06.2004 0:00	739,9
04.06.2004 0:00	645,3	28.06.2004 0:00	690,4
05.06.2004 0:00	608,3	29.06.2004 0:00	748,1
06.06.2004 0:00	583,6	30.06.2004 0:00	678,1
07.06.2004 0:00	637,1	01.07.2004 0:00	735,7
08.06.2004 0:00	649,3	02.07.2004 0:00	669,9
09.06.2004 0:00	674,1	03.07.2004 0:00	579,5
10.06.2004 0:00	711	04.07.2004 0:00	694,6

Изучите общую тенденцию изменения расхода шихты в производстве стекла, выберите алгоритм и оцените точность прогнозирования на интервале в 10 суток.

Оглавление

МОДЕЛИ И МЕТОДЫ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТОВ, ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Методические указания к лабораторным работам

Составители:
МАКАРОВ Руслан Ильич
ХОРОШЕВА Елена Руслановна

Ответственный за выпуск – зав. кафедрой профессор И. Е. Жигалов

Подписано в печать 17.04.13.
Формат 60х84/16. Усл. печ. л. 3,49. Тираж 75 экз.
Заказ
Издательство
Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых.
600000, Владимир, ул. Горького, 87.