УДК 519.237.5

С.Н. ЛАПАЧ, С.Г. РАДЧЕНКО

ОСНОВНЫЕ ПРОБЛЕМЫ ПОСТРОЕНИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ

Анотація. *Сформульовано проблеми побудови регресійних моделей на етапах формування плану експерименту, попередньої та кінцевої специфікації моделі, ідентифікації і оцінки якості отриманої моделі. Розроблено та приведено базові принципи використання регресійного аналізу і планування експерименту, які дозволяють вирішити описані проблеми.*
Ключові слова: *регресійний аналіз, планування експерименту, специфікація моделі, ідентифікація моделі.*

Аннотация. *Сформулированы проблемы построения регрессионных моделей на этапах формирования плана эксперимента, предварительной и окончательной спецификации модели, идентификации и оценки качества полученной модели. Разработаны и приведены базовые принципы использования регрессионного анализа и планирования эксперимента, позволяющие решить описанные проблемы.*
Ключевые слова: *регрессионный анализ, планирование эксперимента, спецификация модели, идентификация модели.*

Abstract. *The problems of regression models formation at the stages of experiment design, preliminary and final model specification, model identification stage and evaluation of result model quality were formulated. The fundamental principles of regression analysis and design of experiment were worked out and described. These principles allow solving the mentioned problems.*
Keywords: *regression analysis, design of experiment, model specification, model identification.*

1. Введение. Постановка проблемы в общем виде

Регрессионный анализ (РА) является одним из наиболее распространенных и мощных методов многомерного статистического анализа.

Широкое применение РА в настоящее время обуславливается следующими факторами: быстрая смена изучаемых областей (технологий, материалов, условий эксплуатации), которая не позволяет тратить много времени на исследования и получение требуемых зависимостей; расширение областей применения (социология, история и пр.); повсеместная автоматизация, требующая множества моделей, применимых в данном конкретном случае с заданной точностью. Тем не менее, следует признать, что как в теории, так и в практическом применении регрессионного анализа положение далеко не лучшее. А это приводит к непредвиденным последствиям.

Применение регрессионного анализа переросло свои разработанные теоретические основания, например, в [1] и [2]. Это солидные издания, но научного работника или инженера, для которого регрессионный анализ всего лишь один из инструментов, их большой объем может лишь отпугнуть.

Планирование эксперимента (ПЭ), которое должно было бы стать основанием и неотъемлемой частью РА, само находится в таком же состоянии и рассматривается как самостоятельная научная дисциплина с сильным уклоном от математической статистики и выборочного метода в сторону «чистой» математики. Кроме того, некоторые представители научных кругов считают ПЭ устаревшим и ненужным, т.е. наблюдается использование РА и ПЭ формально, без понимания сущности, ограничений и предпосылок.

Многие проблемы в построении эмпирических моделей есть следствие незавершенного исторического пути построения соответствующей теории. Регрессионный анализ и теория планирования экспериментов ни по отдельности, ни вместе не представляют собой единых теорий: это множество отдельных теорем и методов, собранных вместе постепенно и по необходимости решать реальные задачи, на что указывал еще Налимов В.В. Очень часто при решении конкретных задач происходит конфликт различных частей, входящих в эти теории. Для решения каждой проблемы разрабатывается много средств, ad hoc, т.е. для каждого конкретного случая. Такой подход только усугубляет ситуацию в связи с остающейся в целом несогласованностью.

Современное состояние дел в данной области требует построения новой теории, объединяющей регрессионный анализ и планирование эксперимента и разрешающей существующие противоречия.

При подходе к построению регрессионной модели следует исходить из цели исследования. Использование регрессионного анализа в терминах прикладной статистики возможно со следующими целями: выяснение наличия статистической связи между случайными переменными (раньше называлось нелинейной корреляцией); аппроксимация; построение математических моделей [3]. При построении моделей обычно требуется семантичность, т.е. возможность объяснения с помощью полученной модели происходящего процесса или явления. Это требует, чтобы структура уравнения регрессии и свойства коэффициентов в некотором смысле соответствовали исследуемому процессу. И в этом заключается главная проблема.

Теоретически известно, что результирующая ошибка при построении модели складывается из ошибки модели и из ошибки определения коэффициентов модели [4]. Тем не менее вся традиционная теория ПЭ и РА опирается на постулат априорной известности «истинной» структуры регрессионной модели. Сами свойства оценок сохраняются при соответствии структуры «истинной» [5]. Оптимальность плана, свойства регрессионных оценок, исследование уравнения на наличие гетероскедастичности, выбросов и прочее – все опирается на указанный постулат. Но при проведении прикладных исследований в подавляющем большинстве ситуаций в традиционной методологии указать эту структуру невозможно. Более того, исследователь во многих случаях как раз и желает получить эту структуру как результат исследования.

2. Основные этапы построения регрессионных моделей

Рассмотрим построение регрессионных моделей, считая структуру заранее неизвестной. Полагаем также, что имеющаяся в нашем распоряжении информация о процессе ограничена, т.е. мы имеем дело с «черным» или «серым» ящиком. Сразу скажем, что регрессионные модели не могут описывать любые процессы, а только те, которые можно представить в виде $\hat{y} = f(X) + \varepsilon$, где $f(X)$ – некоторая детерминированная функция от множества факторов, а ε – случайная величина.

В общем виде построение регрессионной модели проходит следующие этапы:

1. Формирование плана эксперимента (выборки).
2. Предварительная спецификация модели.
3. Окончательная спецификация модели.
4. Идентификация модели.
5. Оценка качества полученной модели.

Рассмотрим отражение проблемы с выбором структуры уравнения регрессии и ее решения на каждом из этих этапов.

Формирование плана эксперимента (выборки)

Проблемы.

• Игнорирование проблемы формирования спецификации модели. Она принимается известной до проведения эксперимента.

• Существование множества планов, оптимальных по разным критериям при заранее заданной структуре модели.

• Отсутствие в традиционной методологии методов работы с данными пассивного эксперимента.

• Отсутствие в традиционной методологии средств работы с нестандартными областями факторного пространства.

• Отсутствие методов (при наличии средств) работы с разрывными областями факторного пространства.

Имеется огромное количество видов планов с разными критериями качества: идентификации (коэффициентов регрессии; оценки поверхности отклика), спецификации (дискриминирующие; отсеивающие), оптимизирующие. Каждый из этих видов оптимальный для решения какой-либо конкретной специальной задачи и не является оптимальным для решения других задач. Кроме того, практически все планы, в частности, группы идентификации, для оптимальности требуют предварительного знания спецификации модели. Все планы рассчитаны на работу в некоторой стандартной области факторного пространства, которой является гиперпараллелепипед или гипершар, что далеко не всегда наблюдается в реальных задачах. Планирование эксперимента не рассматривает решение задач при пассивном или так называемом активно-пассивном эксперименте, который часто встречается на практике. Кроме того, достаточно часто встречаются задачи, в которых особенности протекающих процессов требуют для адекватного описания различные модели в различных областях факторного пространства. Теоретически проблема признана [6], но реального аппарата для формализованного решения подобных задач в настоящее время не существует, кроме случаев однофакторной регрессии.

Предварительная спецификация модели

Опираясь на теорему Вейерштрасса для неизвестной (непериодической) функции $f(X)$, принимается аппроксимация ее полиномами, т.е. общий вид модели задается полиномом определенного порядка. Предварительность связана с тем, что в большинстве реальных задач невозможно получить оценки сразу всех возможных коэффициентов модели и исключить из нее статистически незначимые в связи с тем, что число экспериментов значительно меньше количества возможных членов модели. Желательна линейная по параметрам регрессия, так как нелинейная теоретически не обоснована [7] и сопряжена с вычислительной неустойчивостью. Линеаризация также приводит к искажениям структуры статистических связей между откликом и независимыми переменными.

Проблемы.

• Плохая обусловленность матрицы, по которой выполняется идентификация, при увеличении степени полиномов.

• Невозможность удовлетворительной аппроксимации полиномами некоторых зависимостей при ограниченном числе уровней варьирования (быстроизменяющиеся и асимптотические функции).

Мультиколлинеарности в современной литературе уделяется достаточно много внимания. Но это мультиколлинеарность, которая имеет место в пассивном эксперименте при «неудачной» выборке. Однако мультиколлинеарность возникает в РА практически всегда даже при ортогональном плане эксперимента. Использование обычных полиномов при увеличении их степени даже при идеальных планах эксперимента (например, планах полного факторного эксперимента) и достаточно простых моделях приводит к плохо обусловленным матрицам. Это в свою очередь имеет результатом вычислительную и структурную неустойчивость полученной модели, что делает ее непригодной к использованию, о чем большинство исследователей даже не догадываются. А ведь совершенно непригодные для использования результаты получаются в достаточно простых ситуациях. Например, возьмем план полного факторного эксперимента $2^1 \times 4^1//8$. Такой план создает идеальные условия для устойчивости вычислительного процесса и использования статистических критериев. Посчитаем все коэффициенты для модели вида $\hat{y} = b_0x_0 + b_1x_1 + b_2x_2 + b_3x_2^2 + b_4x_2^3 + b_5x_1x_2 + b_6x_1x_2^2 + b_7x_1x_2^3$. В табл. 1 представлены результаты расчетов с использованием двух распространенных программных средств (ПС) Excel и MathCad. При этом расчеты в каждом ПС выполнялись двумя способами: средствами матричной алгебры и с помощью функций регрессионного анализа. Для возможности сравнения те же действия выполнены после преобразования исходных данных к ортогональным контрастам с помощью полиномов Чебышева.

Таблица 1. Значения коэффициентов регрессии

Название регрессора	Полиномы Чебышева				Обычные полиномы			
	Excel		MathCad		Excel		MathCad	
	Матричные вычисления	Регрессионный анализ	Матричные вычисления	Регрессионный анализ	Матричные вычисления	Регрессионный анализ	Матричные вычисления	Регрессионный анализ
x_0	162,56	162,56	162,56	162,56	7704	–6684	–5545	–4600
x_1	–7,688	–7,6875	–7,689	–7,689	–637,48	–15,375	–15,375	–15,375
x_2	–1,408	–1,408	–1,408	–1,408	–4184	0	913,584	610,308
x_2^2	–1251,3	–1251,3	–1251	–1251	580	425,887	–13,848	10,439
x_2^3	40,098	0	40,098	40,098	–21,375	–50,121	–0,058	–0,311
x_1x_2	–8,043	0	–8,043	–8,043	288	0	509,766	145,832
$x_1x_2^2$	–263,01	–263,01	–263,01	–263,01	8	1,56024	–48,766	–18,197
$x_1x_2^3$	–123,03	0	–123,03	–123,03	–4,125	0	–1,044	–1,61

В табл. 1 видно, что при использовании натуральных степеней полиномов даже для относительно простой модели и теоретически в идеальном плане полного факторного эксперимента не совпадают не только результаты, полученные в разных ПС, но и результаты, полученные в одном ПС разными способами. Т.е. мы имеем дело с потерей вычислительной устойчивости, которая делает результаты расчетов практически бесполезными.

Теорема Вейерштрасса выполняется только в предельном случае, т.е. при достаточном числе уровней варьирования данной переменной. При небольшом числе уровней варьирования, что имеет место в подавляющем числе экспериментальных исследований, для быстроизменяющихся функций при наличии хорошей аппроксимации в точках приближения имеют место значительные отклонения между ними [8–10]. Кроме того, полиномами невозможно описывать асимптотические процессы.

Окончательная спецификация модели

Формируется в процессе построения модели посредством алгоритмов выбора наилучшего подмножества регрессоров.

Проблемы.

• Разные алгоритмы дают разные структуры модели.

• Разные алгоритмы, по сути, имеют разные цели (разные показатели качества модели).

• Алгоритмы, как правило, неустойчивы при неортогональных матрицах, по которым выполняется поиск структуры модели.

• Даже для полного факторного эксперимента (полная ортогональность эффектов) существуют конфликты между отдельными элементами алгоритмов.

Существует множество алгоритмов формирования «наилучшего уравнения регрессии». Недостатком их есть то, что они дают различные «наилучшие уравнения». Особенно для неортогональных матриц эффектов. Кроме того, вопрос, что считать «наилучшим уравнением», также не разрешен. Конфликт существует даже для оптимальных матриц эксперимента. Например, для полного факторного эксперимента нередки ситуации, когда имеются статистически значимые коэффициенты регрессии, которые следует включить в модель, но их не следует включать, так как модель уже адекватна по критерию Фишера. Или же, для того чтобы модель была адекватной и информативной, в нее необходимо включать статистически незначимые коэффициенты.

Оценка качества модели

В большинстве случаев вопрос о показателях качества модели, их согласованности и соответствии цели моделирования даже не ставится.

Проблемы.

• Показатели качества модели не согласовывают с прикладной целью исследования.

• Не существует общепринятого набора показателей качества.

• Между отдельными показателями качества в большинстве случаев существуют противоречия.

• Разным показателям качества соответствует своя лучшая модель.

Статистические показатели, которые оценивают качество модели (значимость коэффициентов, информативность, адекватность, устойчивость), обычно используются по отдельности, между собой не согласованы. Например, модель может быть неадекватной, но информативной. Или информативной и адекватной может быть модель, состоящая из статистически незначимых коэффициентов. Адекватная и информативная модель может быть неустойчивой и прочее.

Опорные точки разработанной методологии

Как и требуется в математическом моделировании, сначала установим цель построения регрессионной модели. В данной работе таковой мы считаем построение математической модели явления или процесса, позволяющей проводить исследование моделируемого процесса. Такая цель является наиболее сложной.

Для достижения поставленной цели необходимо рассматривать ПЭ и РА как два этапа одного процесса, а не независимые методы. Для того чтобы конечный результат был оптимальным, необходимо, исходя из принципа динамического программирования, чтобы критерии оптимальности на каждом шаге были оптимальны с точки зрения достижения конечного результата, а не конкретного шага.

Для достижения поставленной цели необходимо обеспечить репрезентативность выборки, выделение истинной структуры модели, её информативность, структурную и вычислительную устойчивость модели. Решение об адекватности модели и соответствующие требования должны исходить из предметной области.

Поскольку данные, по которым мы будем получать коэффициенты модели, являются результатами эксперимента, то для последующего использования модели мы должны учитывать требования выборочного метода. Это означает требования как к свойствам выборки для обеспечения ее репрезентативности, так и требования к способам получения оценок для обеспечения соответствующих их свойств.

Теоретически в РА требования для обеспечения свойств оценок имеются, но они, во-первых, выполняются при известной структуре модели, а во-вторых, не учитывается влияние вычислительных методов на свойства этих оценок. В связи с тем, что структура модели заранее не известна, главным требованием становится обеспечение наилучших условий для определения структуры и обеспечение вычислительной устойчивости. Обосновывается это тем, что при неправильной структуре все остальные свойства не имеют места и бессмысленно за них бороться. Отсутствие вычислительной устойчивости делает получение модели бессмысленным. Т.е. ПЭ должно построить матрицу, которая обеспечивает наилучшие условия для 1) спецификации модели; и 2) устойчивых оценок коэффициентов модели; 3) матрица должна быть репрезентативной выборкой с точки зрения математической статистики.

Для обеспечения вышеуказанных требований матрица должна быть случайная, равномерно распределенная, число уровней достаточно для получения модели необходимой сложности, число опытов – достаточное для идентификации требуемого количества членов модели. Требуемое число уровней определяется при формализации. Если информация отсутствует, это число выбирается с «запасом».

Число опытов $N_{расч}$ для плана эксперимента по построению модели рассчитывается по следующей формуле (без учета дублирующих опытов):

$$N_{\text{расч}} = (1{,}5...2)(1+\sum_{i=1}^{k}(s_i - 1)),$$

где k – число независимых переменных (факторов), s_i – число уровней варьирования для каждой независимой переменной.

Эта формула базируется на допущении Саттерзвайта об экспоненциальном убывании силы влияния эффектов, ответственных за процесс [8].

Все эти требования наилучшим образом совместно удовлетворяются при использовании планов на основе многомерных равномерно распределенных псевдослучайных чисел, например, $ЛП_\tau$ равномерно распределенных последовательностей [8, 11].

Использование этих планов обеспечивает одновременно оптимальные условия для поиска неизвестной структуры уравнения регрессии и достаточно близкие к оптимальным условия получения устойчивых оценок коэффициентов регрессии. Кроме того, эти планы дополнительно устойчивы к отклонениям от самого плана: пропуски отдельных экспериментов и незначительные отклонения от значения уровней плана. Это свойство, а также возможность использовать такие планы как последовательные, представляет значительные удобства (и экономический выигрыш) для экспериментатора.

В табл. 2 показано место этих планов среди уже существующих.

Одного плана с хорошими свойствами недостаточно, требуется еще высокоэффективный алгоритм определения частной структуры уравнения регрессии. Был проведен сравнительный анализ эффективности разработанного алгоритма посредством вычислительного эксперимента на серии из 10 специально разработанных задач, подобных типовым техническим. Число регрессоров, которые подлежат рассмотрению в тестовых задачах – 180, количество значимых членов модели – 15, коррелированность между значимыми регрессорами – до 0,56. В результате вычислительного эксперимента установлено: традиционные методы выделяют не более 10…15 % (по количеству) истинных членов структуры. При этом выделенные элементы расположены по значимости случайным образом среди всего множества элементов, а коэффициенты регрессии в 60 % случаев имеют противоположный знак. В табл. 3 приведены результаты вычислительного эксперимента разработанного алгоритма в сравнении с часто используемыми алгоритмами.

Следует отметить, что при малом количестве регрессоров, которые подлежат анализу (1…15), и незначительном количестве членов модели (3…4) эффективность алгоритмов практически не различается.

Таблица 2. Место робастных планов в существующей классификации

<table>
<tr><th colspan="3">Классификация планов</th><th colspan="2">Категории робастных планов</th></tr>
<tr><td rowspan="12">Идентификация</td><td rowspan="6">Для оценки коэффициентов регрессии</td><td>D-оптимальность</td><td rowspan="16"></td><td rowspan="6">Робастные (на основе многофакторных регулярных планов)</td></tr>
<tr><td>A-оптимальность</td></tr>
<tr><td>E-оптимальность</td></tr>
<tr><td>Минимакс дисперсии оценки коэффициентов</td></tr>
<tr><td>Минимум суммы относительных ошибок оценок</td></tr>
<tr><td>Ортогональность</td></tr>
<tr><td rowspan="4">Для оценки поверхности отклика</td><td>G-оптимальность</td><td rowspan="10">Робастные (на основе псевдослучайных чисел)</td></tr>
<tr><td>Q-оптимальность</td></tr>
<tr><td>Ротатабельность</td></tr>
<tr><td>Униформность</td></tr>
<tr><td rowspan="2">Спецификация</td><td rowspan="2"></td><td>Дискриминирующие</td></tr>
<tr><td>Отсеивающие</td></tr>
<tr><td rowspan="2">Оптимизация</td><td rowspan="2"></td><td>Динамические</td></tr>
<tr><td>Статические</td></tr>
</table>

Таблица 3. Характеристики различных методов определения структуры уравнения регрессии

Показатель		Алгоритмы				
		АКМ	МГУА	ШРА	СП	ПВС
Доля выделенных истинных эффектов, %	MIN	0	0	0	0	47
	MAX	15	28	20	26,7	67
	Среднее	5,4	7,2	6,7	10,7	57,6
Доля рассеивания выделенных истинных эффектов, объясняемая уравнением, %	MIN	0	0	0	0	91
	MAX	9	15	8	11	98
	Среднее	1,7	4	2	3	93
Доля выделенных ложных эффектов, %	MIN	70	72	80	74,3	0
	MAX	100	100	100	100	15
	Среднее	92	81	94	89,3	11

Здесь АКМ – метод анализа корреляционной матрицы, МГУА – метод группового учета аргументов, ШРА – шаговый регрессионный анализ, СП – случайный поиск с адаптацией, ПВС – предлагаемый метод последовательного выделения структуры.

Поскольку целью является модель, то окончательное определение структуры должно выполняться по критерию Бокса-Веца γ [12], значимость коэффициентов регрессии является второстепенным показателем, поскольку возможна информативная и адекватная модель, в которой все коэффициенты статистически незначимы [13]. Информативность (значение критерия Фишера для значимости коэффициента множественной корреляции) имеет экстремум при увеличении числа членов в модели. Это отличает его от популярной остаточной дисперсии, которая неограниченно убывает, исключая случай, когда на ее величину начинают влиять вычислительные погрешности.

Среди требований к модели из предметной области наиболее часто встречаются следующие: плавный характер изменения зависимостей (отсутствие резких изменений, перегибов и прочее); прогностическая точность; наличие или отсутствие некоторых факторов; наличие или отсутствие взаимодействий (вариант: некоторых взаимодействий); характер поведения (особые точки, производная, общий характер изменения); точность во всей области определения, т.е. не среднее значение отклонения, а обеспечение непревышения ошибки некоторого значения во всей области описания.

Многие из этих требований могут быть учтены математически как требования сильной близости функций или гладкости аппроксимации [14]. Степень близости определяется следующим соотношением [14]:

$$\sum_{i=1}^{N}\left|\Delta^k f(x) - \Delta^k \varphi(x)\right| \to \min .$$

Здесь $f(x)$ и $\varphi(x)$ – функции, близость которых мы определяем, Δ^k – конечная разность порядка k как показатель близости функций. Согласно этому определению, требования метода наименьших квадратов представляют собой нулевую степень близости $(k=0)$.

3. Выводы

Разработанные базовые элементы технологии РА и ПЭ представляют собой единое целое и являются частью выборочного метода; построение регрессионной модели является частью математического моделирования. Ключевые моменты: 1) робастное планирование эксперимента [8]; 2) преобразование произвольной области факторного пространства в стандартную [11, 15]; 3) выделение или достройка выборки требуемого качества при пассивном эксперименте [16]; 4) разбиение на однородные области пространства (если удовлетворительное решение не получено, требует дальнейшей работы); 5) специальные преобразования отдельных факторов; 6) использование полиномов Чебышева [8, 17]; 7) алгоритм последовательного выделения структуры модели [8]; 8) совокупность показателей оценки качества [8]: наилучшая модель соответствует максимальному значению информативности (расчетное значение критерия Фишера F_R для значимости коэффициента множественной корреляции R; значение критерия Бокса-Веца γ); устойчивость структуры (анализ таблицы мультиколлинеарности) и вычислительная устойчивость $\mathrm{cond}(X^T X)$; адекватность – из предметной области или критерии гладкости [14, 18].

Разработанная методология и программное обеспечение [19], поддерживающее ее, успешно использовались для решения нескольких сотен прикладных задач [8, 11, 15].

Направление дальнейших работ

Разработка методов формального разбиения факторного пространства на однородные подобласти.

СПИСОК ЛИТЕРАТУРЫ

1. Грін Г.В. Економетричний аналіз / Грін Г.В. – К.: Основи, 2005. – 1198 с.
2. Дрейпер Н.Р. Прикладной регрессионный анализ / Н.Р. Дрейпер, Г. Смит; пер. с англ. – [3-е изд.] – М. – Санкт-Петербург – Киев: Диалектика, 2007. – 912 с.
3. Айвазян С.А. Прикладная статистика. Исследование зависимостей: справ. изд. / Айвазян С.А., Енюков И.С., Мешалкин Л.Д.; под. ред. С.А. Айвазяна. – М.: Финансы и статистика, 1985. – 487 с.
4. Математическая теория планирования эксперимента / Под ред. С.М. Ермакова. – М.: Наука. ГРФМЛ, 1983. – 392 с.
5. Демиденко Е.З. Линейная и нелинейная регрессия / Демиденко Е.З. – М.: Финансы и статистика, 1981. – 302 с.
6. Котюков В.И. Многофакторные кусочно-линейные модели / Котюков В.И. – М.: Финансы и статистика, 1984. – 216 с.
7. Швырков В.В. Тайна традиционной статистики Запада / Швырков В.В. – М.: Финансы и статистика, 1998. – 144 с.
8. Лапач С.Н. Статистические методы в фармакологии и маркетинге фармацевтического рынка / Лапач С.Н., Пасечник М.Ф., Чубенко А.В. – К.: ЗАТ «Укрспецмонтажпроект», 1999. – 312 с.
9. Лагутин М.В. Наглядная математическая статистика / Лагутин М.В. – М.: Бином. Лаборатория знаний, 2007. – 472 с.
10. Калиткин Н.Н. Численные методы / Калиткин Н.Н.; под ред. А.А. Самарского. – М.: Наука, ГРФМЛ, 1978. – 512 с.
11. Радченко С.Г. Методология регрессионного анализа / Радченко С.Г. – К.: Корнійчук, 2011. – 376 с.
12. Вучков И. Прикладной регрессионный анализ / Вучков И., Бояджиева Л., Солаков Е.; пер. с болг. и предисл. Ю.П. Адлера. – М.: Финансы и статистика, 1987. – 239 с.
13. Pardoux C. Sur la selection de variables en regression multiple / C. Pardoux // Cah. Bur. Univ. rech. oper. – 1982. – N 39–40. – P. 101 – 133.
14. Пухов Г.Е. Критерии и методы идентификации объектов / Г.Е. Пухов, Ц.С. Хатиашвили. – К.: Наукова думка, 1979. – 190 с.
15. Лапач С.М. Забезпечення необхідних властивостей вибірки для побудови регресійної моделі / С.М. Лапач // Труды 15-й Междунар. научн.-техн. конф. «Физические и компьютерные технологии», (Харьков, 2–3 декабря 2009 г.) – Харьков: ХНПК «ФЭД», 2009. – С. 179 – 182.
16. Радченко С.Г. Устойчивые методы оценивания статистических моделей / Радченко С.Г. – К.: ПП «Санспарель», 2005. – 504 с.
17. Пашковский С. Вычислительные применения многочленов и рядов Чебышева / Пашковский С. – М.: Наука, ГРФМЛ, 1983. – 384 с.
18. Компьютерный анализ и интерпретация эмпирических зависимостей / Под. ред. С.В. Поршнева. – М.: Бином, 2009. – 336 с.
19. Лапач С.Н. Планирование, регрессия и анализ моделей PRIAM (ПРИАМ) / С.Н. Лапач, С.Г. Радченко, П.Н. Бабич // Программные продукты Украины: каталог. – К., 1993. – С. 24 – 27.

Стаття надійшла до редакції 18.06.2012