

УДК 004.891.3.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И ПРОГРАММА ПРОГНОЗИРОВАНИЯ
УСПЕШНОСТИ ЛЕЧЕНИЯ НА ОСНОВЕ РЕГРЕССИОННОГО АНАЛИЗА© 2014 Б.А. Есипов¹, Е.С. Губанов², Е.А. Боряев²¹Самарский государственный аэрокосмический университет,²Самарский государственный медицинский университет

Поступила в редакцию 17.12.2013

В данной статье рассмотрена математическая модель для прогнозирования результатов хирургического лечения с использованием статистического анализа данных о пролеченных больных.

Ключевые слова: математическая модель, регрессионный многофакторный анализ, критериальное пространство.

На современном этапе развития медицины информационные нагрузки достигают пределов человеческих возможностей. Возникает дилемма: либо приходится жертвовать полнотой анализа информации, либо необходимо шире использовать различные методы компьютерной поддержки принятия решений. Медицинские экспертные системы позволяют врачу не только проверить собственные диагностические предположения, но и обратиться к компьютеру за консультацией в трудных диагностических случаях [2].

Рак предстательной железы является одной из актуальных проблем современной онкоурологии. Существует множество методов лечения данной патологии.

Например, высокоинтенсивный фокусированный ультразвук (HIFU) является новым методом лечения локализованного рака предстательной железы [3]. Эффективность, а также последствия его использования, малоизучены в России. Этот современный метод имеет как положительные моменты в лечении рака предстательной железы, так и отрицательные.

Целью настоящей работы является разработка математической модели и применение математических методов анализа результатов хирургических вмешательств на основе статистических данных о пролеченных пациентах и прогнозирования успешности будущих операций.

Прогнозирование успешности операций и возможных послеоперационных осложнений яв-

ляется основной функцией врача-хирурга. Так или иначе, при принятии решения о проведении хирургической операции он руководствуется большим числом индивидуальных факторов, присущих пациенту, а так же предшествующим опытом лечения большого числа больных. В этом смысле интуиция врача – это *знание*, базирующееся на опыте, т.е. статистическом материале. Поэтому естественно развивать индустрию получения медицинских знаний в том числе и на глубокой обработке статистического материала.

Математический аппарат многофакторного регрессионного анализа является эффективным приемом для решения разнообразных медицинских задач.

Будем определять успешность медицинской операции некоторыми количественными критериями y_j , j – номер критерия. В нашей работе мы применяли четыре критерия, измеряющих тяжесть послеоперационных осложнений: y_1 –держание мочи, y_2 – острая задержка мочи, y_3 – прогрессия заболевания, y_4 – стриктура. Идеальное значение всех критериев $y=0$.

С каждым пациентом связано большое количество индивидуальных факторов, которые по вероятности влияют на успешность операции A , обозначим их x_i , где i – номер фактора. Факторы могут быть как количественными, так и качественными. Дооперационные данные: возраст, стадия, предшествующие операции на предстательной железе, объем железы, наличие инфравезикальной обструкции (данные урофлоуметрии, объем остаточной мочи) и фактор, относящийся к техническим особенностям выполнения операции: объем предстательной железы подвергшейся воздействию высокоинтенсивного ультразвука – являются входными параметрами, влияющими на исход лечения.

Есипов Борис Алексеевич, кандидат технических наук, доцент кафедры информационных систем и технологий. E-mail: bobpereira@yandex.ru

Губанов Евгений Сергеевич, кандидат медицинских наук, ассистент кафедры урологии. E-mail: samara-urology@mail.ru

Боряев Евгений Александрович, ассистент кафедры урологии. E-mail: borjaev.evgeny@yandex.ru

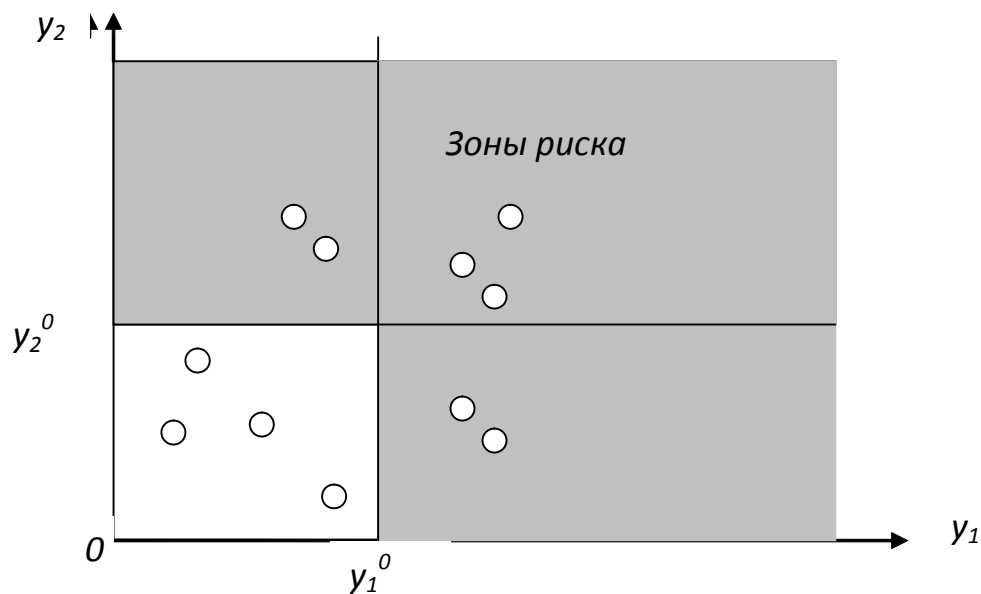


Рис. 1. Зоны риска в пространстве критериев

Первоначальные исследования использовали имеющиеся данные о 12 факторах: 1 – возраст, 2 – стадия заболевания по TNM, 3 – ПСА, 4 – сумма Глисона, 5 – объем простаты, 6 – объем простаты перед НIFU, 7 – длина простаты перед НIFU, 8 – ширина простаты перед НIFU, 9 – высота простаты перед НIFU, 10 – максимальная скорость потока мочи, 11 – средняя скорость потока мочи, 12 – остаточная моча.

В настоящее время имеется достаточно большой статистический материал, содержащий однородные по видам терапий (НIFU, НIFU+ТУР) группы данных для большой выборки больных: (y_j, x_i) , что дает возможность применить методику статистического многофакторного регрессионного анализа. Она заключается в получении уравнений регрессии $\hat{y}_j = f_A(x_1, x_2, \dots, x_m)$ – где \hat{y}_j – среднее ожидаемое значение критерия j . Технически задача получения такой зависимости не представляет трудности при применении стандартных компьютерных программ. Другое дело определение *значимости* полученной зависимости и сопоставление выводов статистических методов с медицинскими рекомендациями. С точки зрения создания экспертной системы такая зависимость представляет собой *знание*, на основе которого можно построить многие интеллектуальные процедуры вывода. Действительно, если на обучающей выборке получено уравнение регрессии, то можно рассмотреть задачу классификации пациентов по критериям успешности операции A .

Предварительно разбиваем пространство критериев на две части – допустимую и недопустимую. Для этого медики – эксперты указывают границы допустимых значений критериев y_j^0 . Тогда допустимым множеством критериев будет такое Y^0 , где по всем критериям выполняется не-

равенство $y_j < y_j^0$, для всех j . Все другие сочетания значений критериев образуют недопустимое множество Y^R (зона риска). На рис. 1 показан пример зон риска для двух критериев.

Тогда можно классифицировать нового пациента с точки зрения успешности операции A . Для этого необходимо получить результаты анализов этого пациента по стандартизованной схеме (x_1, x_2, \dots, x_m) , подставить эти значения в уравнение регрессии для предполагаемого вида терапии A и проверить полученные значения критериев на принадлежность к множеству Y^0 .

Такой подход, очевидно, может быть применен для прогнозирования успешности любых видов лечения.

На рис. 2 приведена структурная схема программной системы. Она состоит из взаимосвязанных подсистем, каждая из которых выполняет фиксированный набор функций.

Кратко охарактеризуем назначение подсистем.

1. *Подсистема управления* – представляет собой систему меню;
2. *Подсистема пациента* – отвечает за добавление/редактирование характеристик пациента;
3. *Подсистема обработки данных обследования* – отвечает за решение уравнений регрессии;
4. *Подсистема прогноза* – на основе статистических данных производит регрессионный анализ и выдает результирующие данные об уравнениях регрессии, а так же решение о прогнозе успешности того или иного вида лечения;
5. *Справочная подсистема* – дает необходимую информацию о программе и о том, как с ней работать;
6. *Подсистема визуализации* – обеспечивает удобную работу с программой, информирует пользователя в процессе работы о каких-либо исключительных ситуациях или ошибках.

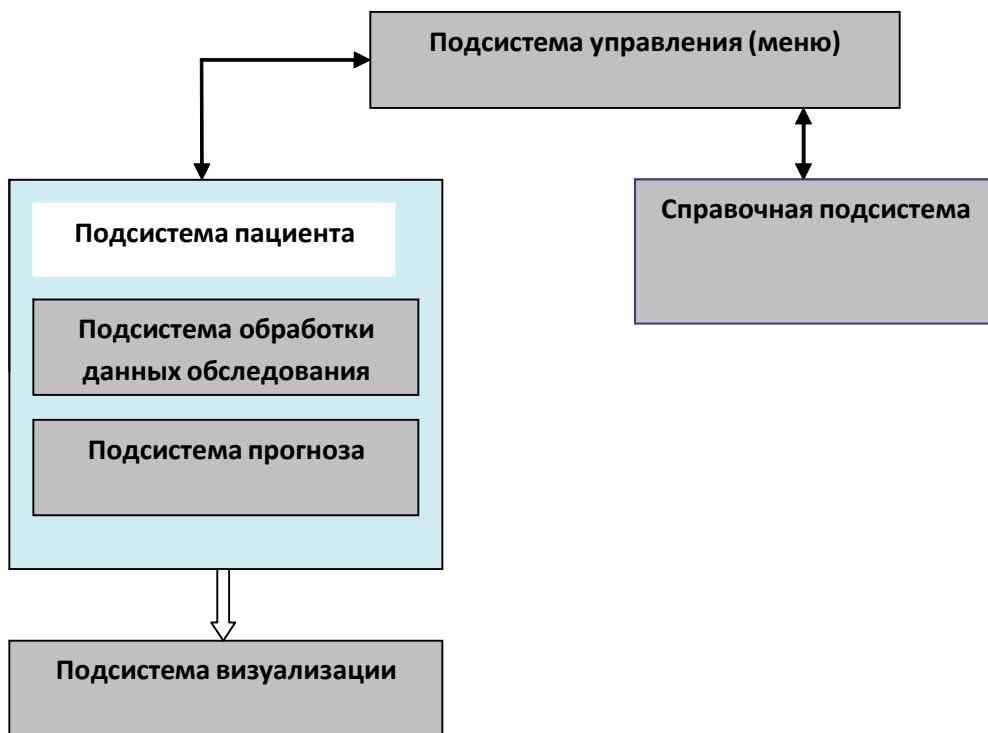


Рис. 2. Структура программной системы

В работе рассмотрен пример применения такого подхода для прогнозирования НIFU-терапии. В исследование вошли 40 больных раком предстательной железы, пролеченных методом НIFU-терапии. Все пациенты перед лечением прошли комплексное обследование для определения стадии заболевания и состояния мочевыделительной системы.

Прогнозирование исхода лечения проведено с помощью программных средств на основе многофакторного регрессионного анализа.

В основу была положена модель линейной множественной регрессии

$y^k = \beta_0 + \beta_1 x_1^k + \beta_2 x_2^k + \dots + \beta_m x_m^k + \epsilon^k$,
 где k - номер наблюдения ($k=1, n$) или в матричном виде

$$y = \beta X + \epsilon,$$

Здесь y – вектор объясняемых переменных; ϵ – вектор значений ошибки; β – вектор коэффициентов; X – матрица объясняющих переменных размером $n \cdot (m+1)$, в которой первый столбец содержит единичные элементы. Известно [1], что тогда оценки коэффициентов β , полученные по методу наименьших квадратов, определяются выражением

$$\beta = (X'X)^{-1}X'Y.$$

Это выражение справедливо и для других моделей регрессии, у которых неизвестными являются коэффициенты при любых степенях x .

На рис.3 представлена диаграмма прогнозирования результирующих параметров.



Рис. 3. Диаграмма прогнозирования результирующих параметров

Первые результаты показали работоспособность методики, хотя пока рано говорить о законченности системы прогнозирования и широкое ее использование. Выбранные 12 факторов для анализа показали хотя и заметную, но все-таки слабую корреляцию с критериями y_j .

Ф.И.О.	Возраст	стадия заболев...	ПСА	Глиссон(степен...
Васенков Нико...	69	2	21,2	7
Лисицин Васил...	82	3	29,8	6
Романов Влади...	78	2	15,2	7
Санников Рост...	71	2	13,9	5
Андреев Алекса...	76	2	28	6
Бочкарев Генна...	72	2	2,76	7
Колотилин Илья...	74	3	51,2	6
Лампеев Евге...	68	3	31	8
Ларионов Миха...	71	3	55	4
Ломака Вита...	70	2	1,47	5
Трунов Владим...	71	2	0,07	4

Рис. 4. Таблица пролеченных больных (обучающая выборка)

Коэффициент детерминации 35% – это значит, что только на одну треть критерии определяются выбранными показателями, а 65% этих величин зависят от чего-то другого (см. рис. 5). К счастью, разработанные программы дают возможность проэкспериментировать с другими факторами и найти то множество факторов, которое дает наибольшую корреляцию с критериями успешности. Это может привести к более широкому предоперационным анализам больных.

Проведен эксперимент с классификацией нового пациента. Получены значения критериев:

$y_1=1,37; y_2=1,43; y_3= 1,12; y_4= 0,98$. Оказалось, что этому пациенту показана HIFU+ТУР-терапия, что совпало с рекомендацией врачей. На рис. 4 и 5 показаны некоторые фрагменты выходных форм документов, реализованных в компьютерной программе.

В заключение следует отметить, что практическая эффективность предлагаемых средств зависит от дальнейших исследований информативности входных факторов, получаемых как в предоперационных исследованиях, так и в процессе проводимых операций. Созданная программа частично по-

Первое уравнение	Второе уравнение	Третье уравнение	Четвертое уравнение
$y_1 = -6.87 + 0.01 \cdot x_1 + 0.46 \cdot x_2 + 0.01 \cdot x_3 + 0.12 \cdot x_4 - 0.03 \cdot x_5 - 0.12 \cdot x_6 + 0.07 \cdot x_7 + 0.12 \cdot x_8 + 0.08 \cdot x_9 + 0.02 \cdot x_{10} + 0.14 \cdot x_{11} + 0.01 \cdot x_{12}$	$y_2 = 3.9 - 0.01 \cdot x_1 + 0.3 \cdot x_2 - 0.01 \cdot x_3 - 0.18 \cdot x_4 - 0.01 \cdot x_5 - 0.02 \cdot x_6 - 0.02 \cdot x_7 - 0.01 \cdot x_8 - 0.05 \cdot x_9 - 0.02 \cdot x_{10} + 0.05 \cdot x_{11} + 0.01 \cdot x_{12}$	$y_3 = 4.8 - 0.01 \cdot x_1 - 0.2 \cdot x_2 + 0.01 \cdot x_3 + 0.04 \cdot x_4 + 0.05 \cdot x_5 + 0.02 \cdot x_6 - 0.02 \cdot x_7 - 0.05 \cdot x_8 + 0.01 \cdot x_9 + 0.04 \cdot x_{10} - 0.11 \cdot x_{11} - 0.01 \cdot x_{12}$	$y_4 = 2.85 + 0.01 \cdot x_1 + 0.03 \cdot x_2 - 0.01 \cdot x_3 - 0.07 \cdot x_4 - 0.03 \cdot x_5 - 0.01 \cdot x_6 - 0.02 \cdot x_7 - 0.03 \cdot x_8 - 0.01 \cdot x_9 - 0.04 \cdot x_{10} + 0.15 \cdot x_{11} - 0.01 \cdot x_{12}$
R квадрат: 0,34	R квадрат: 0,33	R квадрат: 0,31	R квадрат: 0,35
Множественный R: 0,58	Множественный R: 0,54	Множественный R: 0,53	Множественный R: 0,59
Количество наблюдений: 44	Количество наблюдений: 44	Количество наблюдений: 44	Количество наблюдений: 44
Стандартная ошибка: 0,14	Стандартная ошибка: 0,08	Стандартная ошибка: 0,08	Стандартная ошибка: 0,07

Рис 5. Пример результатов регрессионного анализа

зволяет выявить состав факторов, наиболее влияющих на результаты лечения на основе вычисления коэффициентов множественной корреляции.

СПИСОК ЛИТЕРАТУРЫ

1. Приходько А.И. Практикум по эконометрике: регрессионный анализ средствами Excel. Ростов н/Д.: Феникс, 2007. 266 с.
2. Андрейчиков А.В. Андрейчикова О.Н. Интеллектуальные информационные системы. М.: Финансы и статистика, 2006. 424 с.
3. HIFU — новый метод лечения рака простаты [Электронный ресурс]. URL: <http://www.consmed.ru/news/view/526/> (дата обращения 24.11.2013).

MATHEMATIC MODEL AND PROGRAM FOR PREDICTION TREATMENT SUCCESS IS BASED ON REGRESSION ANALISIS

© 2014 В.А. Yesipov¹, Е.С. Gubanov², Е.А. Boriaev²

¹ Samara State Aerospace University named after Academician S.P. Korolyov
(National Research University)

² Samara State Medical University

In this article we described mathematic model for prediction treatment success is based in regression analysis

Key words: mathematic model, regression analysis, kriterium space.

Boris Yesipov, Candidate of Technics, Associate Professor at the Informatic Systems and Technology.

E-mail: bobpereira@yandex.ru

Evgeny Gubanov, Candidate of Medicine, Assistent at the Urology Department. E-mail: samara-urology@mail.ru

Evgeny Boriaev, Assistent at the Urology Department.

E-mail: boriaev.evgeny@yandex.ru