



Б.А. Есипов, Е.А. Боряев, Е.С. Губанов

МОДЕЛИРОВАНИЕ И РЕШЕНИЕ НЕКОТОРЫХ ПРАКТИЧЕСКИХ ЗАДАЧ МЕДИЦИНЫ НА ОСНОВЕ РЕГРЕССИОННОГО АНАЛИЗА

(Самарский государственный аэрокосмический университет,
Самарский государственный медицинский университет)

Медицинские экспертные системы позволяют врачу не только проверить собственные предположения, но и обратиться к компьютеру за консультацией в трудных диагностических случаях, а также в выборе метода лечения [2].

При принятии решения о проведении хирургической операции врач 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 на глубокой обработке статистического материала.

Целью настоящей работы является применение математических методов анализа результатов хирургических вмешательств на основе статистических данных о пролеченных пациентах и получение количественных моделей зависимости результатов медицинских операций от факторов, определяющих индивидуальные параметры больных и вида лечения.

Математический аппарат многофакторного регрессионного анализа является эффективным приемом для решения разнообразных медицинских задач. В частности отделение урологии клиник СГМУ располагает большим объемом статистических данных, однородных по видам терапий, с результатами лечения и послеоперационными наблюдениями. Глубокая статистическая обработка этой информации позволяет решать некоторые новые задачи медицинской практики.

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 – недержание 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. Знание этих значений (y_j, x_i) , полученных по результатам лечения большого количества больных 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 .

В основу была положена модель линейной множественной регрессии



$$y^k = \beta_0 + \beta_1 x_1^k + \beta_2 x_2^k + \dots + \beta_m x_m^k + \varepsilon^k,$$

где k - номер наблюдения ($k=1, n$)

или в матричном виде

$$y = \beta X + \varepsilon.$$

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 .

Технически задача получение такой зависимости не представляет трудности при применении 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. Перечислим некоторые практические задачи, которые можно решать с использованием найденного уравнения регрессии.

1. Если на обучающей выборке получено уравнение регрессии, то можно поставить задачу классификации пациентов по критериям успешности операции A . Для этого предварительно разбиваем пространство критериев на две части – допустимую и недопустимую. Медики - эксперты указывают границы допустимых значений критериев y_j^0 . Тогда допустимым множеством критериев будет такое Y^0 , где по всем критериям выполняется неравенство $y_j < y_j^0$, для всех j . Все другие сочетания значений критериев образуют недопустимое множество Y^R (зона риска).

2. 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 .

3. Уравнение регрессии можно использовать для установления границ изменения результирующего параметра y_j если известны некоторые границы изменения факторных величин x_i . Пусть $y = a_0 + a_1 x_1 + \dots + a_n x_n$ - полученное и статистически значимое уравнение регрессии. Нам нужно определить Δy , если известны Δx_i . Т.к. для линейной функции многих переменных полный дифференциал и полное приращение совпадают ($dy = \Delta y$), а частные производные по i -му аргументу равны a_i , то по формуле полного дифференциала получим $dy = \Delta y = a_1 \Delta x_1 + \dots + a_n \Delta x_n$. Здесь все величины a и Δx могут быть как >0 , так и <0 . Поэтому удобно использовать оценку

$$\Delta y \leq |a_1| |\Delta x_1| + \dots + |a_n| |\Delta x_n|. \quad (1)$$



4. Важным для практики является вопрос нахождения пределов изменения Δx_i при заданных границах изменения результирующего параметра Δy . То есть, если известны допустимые осложнения после операции A , то можно определить границы соответствующих факторов Δx_i .

Чтобы найти Δx_i , если задана величина Δy , используем (1) и получим

$$|\Delta x_i| \geq \frac{1}{|a_i|} (\Delta y - \sum_{k \neq i}^n |a_k| \cdot |\Delta x_k|) \quad (2)$$

Задача получения всех вариантов Δx_i ($i=1,2,3, \dots, n$) при заданном Δy может быть решена только перебором в компьютерной программе.

5. Важной практической задачей является определение факторов x_i наиболее коррелированных с результирующим параметром y . Эта задача решается специальными методами математической статистики на основе компьютерных экспериментов с массивом статистических данных.

Рассмотренный подход может быть применен для прогнозирования успешности любых видов лечения. В работе рассмотрены примеры применения регрессионного анализа для четырех видов терапий: *ультразвуковой абляции (HIFU)*, *радикальной простатэктомии (РПЭ)*, *дистанционной лучевой терапии (ДЛТ)*, *брахиотерапии (контактной лучевой терапии)*.

В исследование вошли от 50 до 80 больных раком предстательной железы пролеченных указанными методами. Все пациенты перед лечением прошли комплексное обследование для определения стадии заболевания и состояния мочевыделительной системы.

Разработана компьютерная программа получения уравнений множественной линейной регрессии. Результаты, проведенные для различных видов терапий и четырех результирующих параметров y_j , показали значимость полученных уравнений на уровне 0,05. Множественные коэффициенты детерминации оказались порядка 0,5, т.е коэффициенты корреляции около 0,7.

Литература

1. Приходько, А.И. Практикум по эконометрике: регрессионный анализ средствами Excel [Текст]/А.И. Приходько. – Ростов н/Д.: Феникс, 2007.-266 с.
2. Андрейчиков, А.В. Интеллектуальные информационные системы [Текст]/ А.В. Андрейчиков, О.Н. Андрейчикова – М.: Финансы и статистика, 2006. – 424 с.