

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ В МЕДИЦИНЕ. ПУТИ РЕШЕНИЯ ПРОБЛЕМЫ МУЛЬТИКОЛЛИНЕАРНОСТИ

Рассматриваются особенности применения метода наименьших квадратов при изучении патологических процессов. Предложены процедуры, позволяющие получать эффективные оценки в условиях мультиколлинеарности. Приведены результаты исследований состояния полости рта у детей с хроническими гепатитами и степени поражения организма при ожогах, полученные при помощи предлагаемых процедур.

Развитие медицинской науки на современном этапе позволяет проводить комплексное исследование различных систем организма. Это приводит к значительному росту объема разнообразной информации, что существенно усложняет ее анализ.

Так как организм описывается большим числом параметров и множеством внутренних взаимосвязей, то его успешное изучение невозможно без системного подхода. При этом одним из наиболее значимых, часто используемых методов является метод наименьших квадратов (МНК). Этот метод на сегодняшний день следует признать всесторонне изученным и имеющим несколько теоретических обоснований [2, 7, 8, 16]. Он является базовым методом математической статистики.

Главными преимуществами оценок МНК является то, что они обладают минимально возможной дисперсией в классе всех линейных несмещенных оценок и являются соответственно наилучшими линейными несмещенными оценками неизвестных параметров. Также следует отметить, что в случае некоррелированных показателей соответствующие весовые коэффициенты обратно пропорциональны своим дисперсиям. Это означает, что при неограниченном увеличении числа некоррелированных показателей ошибка в оценивании объясняемой переменной может быть сколь угодно малой.

По мере усложнения исследуемого процесса и вовлечения в рассмотрение новых компонентов оценки МНК можно использовать в качестве первого приближения. Все эти преимущества и являются причиной его широкого использования в медицине при исследовании различных патологических процессов.

Постановка задачи метода наименьших квадратов заключается в следующем. Пусть существует линейное соотношение между переменной Y , $k-1$, объясняющими переменными X_2, X_3, \dots, X_k и возмущением u . Если мы име-

ем выборку из n наблюдений над переменными Y и X_j , $j = 2, \dots, k$ то можно записать:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i, \quad (i = 1, \dots, n)$$

Коэффициенты β_k и параметры распределения u неизвестны. Задача состоит в получении эффективных оценок неизвестных коэффициентов β_k в соответствии с критерием минимизации: $E[Y_i - S]^2 \Rightarrow \min$.

Здесь E – оператор математического ожидания;

S – объясняемая переменная.

Для того чтобы оценить коэффициенты β_k , в методе наименьших квадратов принимаются предположения относительно того, как генерируются наблюдения. В связи с этим выдвигаются следующие основные гипотезы:

1) Предположение о нулевом среднем значении $E(u) = 0$.

При исследовании какого-либо процесса мы не можем охватить все параметры, имеющие влияние на исследуемый процесс, и включаем в рассмотрение, как правило, наиболее существенные параметры. Эта гипотеза требует, чтобы суммарное влияние исключенных переменных на исследуемый процесс взаимокомпенсировалось и сводилось к нулю.

2) Предположение о постоянстве условной дисперсии и об отсутствии автокорреляции $E(uu') = \Omega^2 I_n$.

Это предположение фактически требует, чтобы эксперимент был проведен при стабильных параметрах окружающей среды и чтобы влияние исключенных из рассмотрения параметров не преумножалось при их сочетании.

3) Предположение об отсутствии одновременности $E(X'u) = 0$ – фиксировано в повторяющихся выборках, либо распределено независимо от u .

Эта гипотеза требует, чтобы рассматриваемые нами показатели не имели функциональных зависимостей с исключенными. Выполне-

ние этого требования позволяет все случайные возмущения полученной модели отнести на счет влияния исключенных факторов, и тогда все свойства оценок β_k будут обусловлены качеством исходного вариационного ряда.

4) Предположение о ранге $\rho(X) = k < n$.

Это предположение требует, чтобы между исследуемыми параметрами не было функциональных связей, то есть исключается наличие линейной зависимости между ними.

На практике при исследовании патологических процессов не всегда представляется возможным удовлетворить всем этим требованиям. Это связано с тем, что живой организм характеризуется множеством внутренних связей различных систем, деятельности которых присуща согласованность и зависимость друг от друга. В результате большой размерности исходного материала и нарушения предположения о линейной независимости включенных в исследование показателей возникает проблема мультиколлинеарности.

Эта проблема, как правило, решается путем сужения признакового пространства за счет исключения из рассмотрения существенных, но менее значимых признаков. Возникает ситуация, когда необходимо выполнение двух противоположных требований. С одной стороны, исключение из регрессии существенного признака может привести к очень серьезным ошибкам. С другой – наличие линейной зависимости двух существенных признаков не позволяет получить эффективные оценки параметров модели при помощи метода наименьших квадратов.

Целью работы является разработка процедур, позволяющих получать эффективные оценки МНК в условиях мультиколлинеарности, которая обусловлена спецификой живого организма и систем, его составляющих.

Проблема мультиколлинеарности встречается достаточно часто. В частности, при решении задачи оценивания степени поражения организма при том или ином заболевании широкое распространение получили так называемые индексы тяжести [9, 10].

Практическое применение в клинической практике индексов тяжести позволило успешно решать задачи количественной оценки тяжести заболевания [14]. Кроме того, метод наименьших квадратов можно использовать для построения так называемых диагностических и прогностических решающих правил или критериев [1, 4, 15].

К настоящему времени разработано достаточно много методов идентификации параметров модели, оценивающей степень поражения организма при заболевании [12, 14]. Для учета нелинейности в поведении элементов вектора состояния предложены способы построения нелинейного обобщенного показателя [5, 17]. Так как интегральные характеристики дают возможность перейти от векторной характеристики состояния к скалярной, то это преимущество используется в модели темпа функционального восстановления [6, 12] для анализа динамики изменения состояния пациента и сравнения эффективности различных методов лечения.

Перейдем к рассмотрению процедур, которые были использованы нами при исследовании патологических процессов с целью решения проблемы мультиколлинеарности.

При построении интегральной характеристики состояния полости рта по показателям эмали у детей с хроническими гепатитами [3] у нас возникала мультиколлинеарность, которая была обусловлена достоверной корреляционной связью между показателями «электропроводность режущего края» и «электропроводность пришеечной поверхности второго моляра». Проблема была решена путем объединения этих показателей в виде $x_i = y_A - y_B$. В качестве обоснования такой процедуры было использовано то обстоятельство, что весовые коэффициенты показателей x_i обратно пропорциональны своим дисперсиям $D(x_i)$. Это позволяет применить свойство дисперсии, заключающееся в том, что если $x_i = y_A + y_B$, то

$$\overline{D}(x_i) = D(y_A) + D(y_B) + 2 * R(y_A, y_B),$$

где $R(y_A, y_B)$ – коэффициент корреляции между y_A и y_B .

Таким образом, вместо двух значимых показателей y_A и y_B в рассмотрение была введена их комбинация в виде $x_i = y_A - y_B$, причем дисперсия комбинированного показателя была уменьшена на величину $2 * R(y_A, y_B)$.

В результате была получена следующая интегральная характеристика, позволяющая оценить состояние полости рта у детей по показателям эмали:

$$S_{em} = 6,01 - 1,76 * y_1 - 0,21 * y_2 - 0,17 * y_3 + 0,20 * (y_4 - y_8) + 0,08 * y_5 + 0,15 * y_6 + 0,12 * y_7$$

где: y_1 – содержание кальция в смешанной слюне (мг/л);

y_2 – электропроводность режущего края резца (мкА);

y_3 – электропроводность экватора резца (мкА);

y_4 – электропроводность режущего края клыка (мкА);

y_5 – электропроводность центральной ямки первого моляра (мкА);

y_6 – электропроводность вестибулярной борозды первого моляра (мкА);

y_7 – электропроводность фиссуры второго моляра (мкА);

y_8 – электропроводность пришеечной поверхности второго моляра (мкА).

При исследовании ожоговой болезни исходная задача заключалась в построении интегральной характеристики степени поражения организма при данной патологии [13].

Как известно, в основе патогенеза ожоговой болезни лежат глубокие нарушения метаболических процессов в зоне ожога и в целом организме. Реакция организма на воздействие высокой температуры и термическое повреждение ткани складывается из болевого стресса, специфического токсинообразования и неспецифической воспалительной реакции.

Ожоги вызывают комплекс патологических изменений, охватывающих практически все жизненно важные системы организма. В итоге метаболических сдвигов наблюдаются нарушения функций различных органов – сердца, почек, печени и т. д. Это требовало комплексного подхода. С этой целью был организован сбор клинико-лабораторных данных, описывающих различные системы организма. В исходный вариационный ряд были включены значения иммунологических, биохимических показателей, липидов, показателей общего анализа крови, коагулограммы, а также кислотно-щелочного состава и эндотоксемии.

Проблема мультиколлинеарности была обусловлена взаимозависимостью систем организма, которая проявлялась в их ответной реакции на ожог. В частности, при ожоге отмечаются активация адренергической системы, нарушение эндокринной регуляции, угнетение работы щитовидной железы и эндогенной антиоксидантной активности, усиление секреции глюкагона, ускорение образования продуктов перекисного окисления липидов. Все это способствует нарушениям микроциркуляции.

В этом случае проблема мультиколлинеарности нами была решена при помощи подхода, основанного на идее исследования патологического

процесса с позиции каждой системы организма отдельно. Это позволило исходную задачу свести к задаче исследования реакции каждой системы организма в ответ на термическую травму.

В результате были построены следующие интегральные характеристики, оценивающие степень поражения организма при ожоге с точки зрения той или иной системы организма:

а) Иммунологический индекс тяжести:

$$J_{IM} = 2,78 - 0,003x_1 - 0,004x_2 + 0,004x_3 + 0,006x_4,$$

где x_1 – Т-лимфоциты (тыс.);

x_2 – Т-хелперы (тыс.);

x_3 – Т-супрессоры (тыс.);

x_4 – нулевые лимфоциты (%).

б) Липидный индекс тяжести:

$$J_L = 2,03 - 10,09y_1 + 1,34y_2 + 0,75y_3 - 0,66y_4 - 0,08y_5 + 2,90y_6$$

где y_1 – фосфолипиды (ммоль/л);

y_2 – моноглицериды (ммоль/л);

y_3 – ХГ (ммоль/л);

y_4 – желчные кислоты (мкмоль/л);

y_5 – диглицериды (ммоль/л);

y_6 – триглицериды (ммоль/л).

в) Индекс тяжести по общему анализу крови:

$$J_{KP} = 2,80 - 0,04z_1 + 0,01z_2 + 0,02z_3,$$

где z_1 – моноциты (10^9 /л);

z_2 – СОЭ (мм/ч); z_3 – Л.И.И.

г) Биохимический индекс тяжести:

$$J_{BX} = 3,86 - 0,01v_1 + 0,01v_2 - 0,03v_3 - 0,01v_4 - 0,01v_5 + \\ + 0,01v_6 + 0,001v_7 + 0,005v_8 - 0,26v_9 + 0,15v_{10} - \\ - 0,24v_{11} + 0,001v_{12} + 0,01v_{13} - 0,003v_{14} - 0,02v_{15} + \\ + 0,003v_{16} + 0,009v_{17} - 0,005v_{18},$$

где v_1 – билирубин общий (мкмоль/л);

v_2 – билирубин непрямой (мкмоль/л);

v_3 – общий белок (г/л);

v_4 – альбумин (г/л);

v_5 – глобулин (г/л);

v_6 – α_1 – глобулин (%);

v_7 – β – глобулин (%);

v_8 – γ – глобулин (%);

v_9 – А/Г коэффициент;

v_{10} – АЛТ (мкмоль/л);

v_{11} – АСТ (мкмоль/л);

v_{12} – диастаза (мкмоль/л);

v_{13} – креатинин (мкмоль/л);

v_{14} – мочевины (мкмоль/л);

v_{15} – холестерин (ммоль/л);

v_{16} – остаточный азот (ммоль/л);

v_{17} – сахар крови (ммоль/л);

v_{18} – щелочная фосфатаза (ммоль/л).

д) Индекс тяжести по показателям коагулограммы:

$$J_{\text{КОА}} = 0,19 + 0,01q_1 + 0,001q_2 + 0,02q_3 + 0,19q_4,$$

где q_1 – протромбиновый индекс (%);

q_2 – фибриноген (г/л);

q_3 – гематокрит;

q_4 – тромботест.

е) Индекс тяжести по показателям кислотно-щелочного состава

$$J_{\text{KSHC}} = 2,73 - 0,03g_1 + 0,001g_2 + 0,02g_3,$$

где g_1 – АВЕ; g_2 – SAT; g_3 – O_2 СТ.

ж) Эндотоксемический индекс тяжести:

$$J_{\text{ENDO}} = 0,33 + 0,04h_1 + 0,27h_2,$$

где h_1 – сорбционная способность эритроцитов;

h_2 – средне-молекулярные пептиды.

При построении уравнений по методу наименьших квадратов на коэффициенты модели накладывалось условие их значимости не ниже уровня $p < 0,05$ по критерию Фишера.

Полученные уравнения дали хорошее согласование с экспертными оценками тяжести патологического процесса, что говорит об адекватности этих интегральных показателей при описании тяжести поражения организма при ожогах.

Следует отметить, что полученные интегральные характеристики степени поражения организма при термической травме можно интерпретировать как нагрузку, которую испытывает та или иная система организма вследствие термической травмы. Это обстоятельство позволяет учесть индивидуальные особенности адаптационных возможностей пораженного организма и является основанием для выбора обоснованного терапевтического воздействия.

Таким образом, полученные результаты позволяют утверждать, что предложенные процедуры введения комбинированных показателей и разделения исходной задачи на подсистемы организма позволяют решать проблему мультиколлинеарности при изучении процессов, протекающих в живых системах.

Список использованной литературы:

1. Беккузин Р.Р., Кадиров Р.Х. Прогнозирование течения переломов нижней челюсти путем построения математической модели // Сб. Науч. труд. УзНПО «Кибернетика»: «Методы и модели систем обработки данных». – Ташкент, 1994. – С. 75-79.
2. Бурмистров Г.А. Основы способа наименьших квадратов. М., Госгеолтехиздат, 1963.
3. Даминов Т.А., Юлдашева А.С., Кадиров Р.Х., Исомов М.Д. Состояние полости рта у детей, больных хроническими гепатитами. Ташкент, изд-во А.А. ибн Сино, 1995. – 121 с.
4. Джаббарова Ю.К., Адильова Ф.Т., Ядгарова К.Т., Кадиров Р.Х. Математическая модель прогноза развития гипотрофии плода у беременных с хроническим пиелонефритом // Актуальные вопросы акушерства и гинекологии, вып. 2. Ташкент, 1993. – С. 25-29.
5. Зуев С.М., Кадиров Р.Х. Нелинейный обобщенный показатель тяжести заболевания // Сб. науч. труд. ОВМ АН СССР «Математическое моделирование в иммунологии и медицине» / под ред. Асаченкова А.Л., М., 1986. – С. 100-106.
6. Кадиров Р.Х., Ниязова И.А. Определение степени поражения сердечно-сосудистой системы и эффективности лечения больных артериальной гипертензией с помощью математического моделирования. // Докл. Акад. наук РУз. – 1998. – №11. – С. 50-52.
7. Крамер Г. Математические методы статистики. – М.: Мир, 1975. – 650 с.
8. Линник Ю.В. Метод наименьших квадратов и основы теории обработки наблюдений. ГИФМЛ, 1962.
9. Марчук Г.И., Бербенцова Э.Г. Результаты применения количественного метода для оценки тяжести и динамики острой пневмонии, хронического бронхита, бронхиальной астмы // Тер. архив – 1986. – Т. 3, С. 63-70.
10. Математический метод оценки тяжести течения инфаркта миокарда // Марчук Г.И., Захаров В.М., Суканова В.Н., Тимошкин Н.В. Препринт ОВМ АН СССР, М.: 1984. – 17 с.
11. Математическое моделирование в иммунологии и медицине. Под ред. Асаченкова А.Л. – М.: ОВМ АН СССР, 1987.
12. Математическое моделирование вирусного гепатита / Нисевич Н.И., Марчук Г.И., Зубикова И.И., Погожев И.Б. – М.: Наука, 1981. – 350 с.
13. Оценка эффективности применяемых методов детоксикации при лечении ожоговой болезни путем математического моделирования / Каримов Ш.Н., Саидов А.А., Исомов М.Д., Кадиров Р.Х., Пайзиев Д.Б. // Медицинский журнал Узбекистана, №1. – Ташкент. – 1994. – С. 37-42.
14. Погожев И.Б. Применение математических моделей заболеваний в клинической практике. – М.: Наука, 1988. – 192 с.
15. Прогнозирование течения острого одонтогенного остеомиелита нижней челюсти у детей / Адильова Ф.Т., Азимов М.И., Дусмухамедова Х.К., Кадиров Р.Х. // Сб. Алгоритмы, вып. 78. – Ташкент. – 1994. – С. 25-27.
16. Рао С.Р. Линейные статистические методы и их применение. М.: Наука, 1968. – С. 550.
17. Скалько Ю.И. Об одном подходе к построению и применению обобщенных индексов тяжести заболеваний // Сб. науч. труд. ОВМ АН СССР «Математическое моделирование в иммунологии и медицине» / Под ред. Асаченкова А.Л. М., 1986.