

РОЛЬ И МЕСТО ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ И ROC-АНАЛИЗА В РЕШЕНИИ МЕДИЦИНСКИХ ДИАГНОСТИЧЕСКИХ ЗАДАЧ

С.Г. Григорьев, Ю.В. Лобзин, Н.В. Скрипченко

Научно-исследовательский институт детских инфекций, Санкт-Петербург, Россия

The role and place of logistic regression and ROC analysis in solving medical diagnostic task

S.G. Grigoryev, Yu.V. Lobzin, N.V. Skripchenko

Science Research Institute of Children's Infections, Saint Petersburg, Russia

Резюме

Диагностика, наряду с профилактикой и лечением, является основой медицинской науки и практики. За свою историю медицина накопила множество диагностических методов различных заболеваний и патологических состояний. Тем не менее, и в настоящее время разрабатываются и рекомендуются к применению новые тесты, методы и инструменты. В качестве методов оценки их диагностической способности используются такие показатели, как чувствительность и специфичность, которые определяются на основе построения четырехпольных таблиц сопряженности или метода ROC-анализа с построением ROC-кривой (Receiver operating characteristic). Четырехпольная таблица используется в случае оценки метода, который подтверждает или отрицает диагноз — качественный показатель. ROC-кривая, являясь графиком, позволяет дать оценку качеству модели по разделению двух классов на основе определения точки отсечения непрерывного или дискретного количественного признака.

В статье представлен метод логистической регрессии в качестве инструмента по разработке математико-статистической модели прогноза вероятности наступления интересующего исследователя события при наличии двух возможных вариантов исхода. В качестве инструмента оценки качества модели избран и детально описан метод ROC-анализа. Возможности названных методов продемонстрированы на реальном примере создания и оценки эффективности (чувствительности и специфичности) модели прогноза вероятности развития у детей с atopическим дерматитом осложнения в форме пиодермита.

Ключевые слова: математико-статистическая модель, логистическая регрессия, ROC-анализ, чувствительность диагностического метода, специфичность диагностического метода, atopический дерматит.

Abstract

Diagnostics, equally with prevention and treatment, is a basis of medical science and practice. For its history the medicine has accumulated a great variety of diagnostic methods for different diseases and pathologic conditions. Nevertheless, new tests, methods and tools are being developed and recommended to application nowadays. Such indicators as sensitivity and specificity which are defined on the basis of fourfold contingency tables construction or ROC-analysis method with ROC — curve modelling (Receiver operating characteristic) are used as the methods to estimate the diagnostic capability. Fourfold table is used with the purpose to estimate the method which confirms or denies the diagnosis, i.e. a quality indicator. ROC-curve, being a graph, allows making the estimation of model quality by subdivision of two classes on the basis of identifying the point of cutting off a continuous or discrete quantitative attribute.

The method of logistic regression technique is introduced as a tool to develop some mathematical-statistical forecasting model of probability of the event the researcher is interested in if there are two possible variants of the outcome. The method of ROC-analysis is chosen and described in detail as a tool to estimate the model quality. The capabilities of the named methods are demonstrated by a real example of creation and efficiency estimation (sensitivity and specificity) of a forecasting model of probability of complication development in the form of pyodermitis in children with atopical dermatitis.

Key words: mathematical-statistical model, logistic regression, ROC-analysis, sensitivity of diagnostic method, specificity of diagnostic method, atopical dermatitis.

В медицинской науке и практике одним из решающих является вопрос диагностики. Решение диагностической задачи — повседневность любого практикующего врача. При этом, наряду с большим количеством проверенных методов диагностики, зарекомендовавших себя многолетней практикой, постоянно появляются новые, бази-

рующиеся на последних достижениях науки, техники и информационных технологий, которые требуют подтверждения своей диагностической способности [1, 2].

Настоящая статья посвящена оценке диагностической способности количественного показателя, являющегося непрерывной или дискретной

величиной. Предметом изучения стал метод диагностики, разработанный с помощью логистической регрессии, а в качестве инструмента оценки его диагностической способности — метод ROC-анализа. Логистическая регрессия применяется для решения задач отнесения исследуемого объекта к одной из двух ранее известных групп [2, 3, 4]. ROC-анализ тесно связан с логистической регрессией и, являясь, по сути, завершающим этапом разработки наиболее адекватной модели, применяется для оценки ее качества. Он позволяет проанализировать чувствительность и специфичность, подобрать порог отсечения и выбрать модель с наилучшей прогностической силой [5, 6, 7].

Построение моделей для показателей состояния многомерных медицинских систем в зависимости от воздействующих на них факторов является важной задачей статистического анализа, выполняемого исследователями с применением современных информационных технологий [3, 7, 8, 9]. С помощью модели решают основные задачи исследования, среди которых изучение характера изменения показателя при изменении действующих на систему факторов; оценка степени влияния факторов на величину показателя-отклика; прогнозирование показателя-отклика для заданных уровней факторов; определение оптимальных уровней факторов для получения требуемых или желаемых значений показателей состояния системы. Построение таких моделей проводится на основе выборочного наблюдения, по результатам которого формируется исходная база данных (обучающей информации), представляющая собой матрицу наблюдений с числом строк, равным числу наблюдавшихся объектов, и числом столбцов, равным числу контролируемых факторов и моделируемого показателя-отклика на воздействующие факторы.

Наряду с дисперсионным анализом, регрессионным анализом, анализом времени жизни, весомое место занимает логистическая регрессия как разновидность множественной регрессии, общее назначение которой состоит в анализе связи между несколькими независимыми переменными (называемыми также регрессорами или предикторами) и зависимой переменной (откликом). Логистическая регрессия применяется в случае, когда зависимая переменная является бинарной (т.е. может принимать только два значения). Прогнозируемым откликом в этом случае является вероятность наступления положительного события для конкретного испытуемого (больной/здоровый, ухудшение состояния/улучшение состояния, исход тяжелой травмы благополучный/летальный и т.д.). Возможный вариант решения является непрерывной величиной, значение которой может находиться в интервале от 0 до 1. Значение откли-

ка, близкое к 0, говорит о малой вероятности наступления интересующего нас исхода, и напротив — значение, близкое к 1, говорит в пользу высокой вероятности наступления этого исхода [3, 4].

Что является положительным событием, а что отрицательным, зависит от конкретной задачи. Например, в случае если прогнозируется вероятность наличия заболевания, то положительным исходом будет класс «Больной пациент», отрицательным — «Здоровый пациент». И наоборот, если необходимо определить вероятность того, что человек здоров, то положительным исходом будет класс «Здоровый пациент» и т.д.

Так как результат решения диагностической задачи по логистической регрессионной модели не дает однозначного ответа, а лишь предлагает вероятность одного из двух заданных состояний, то естественным является вопрос, какое же решение принять в каждой конкретной ситуации. В связи с этим принято, что прогноз положительного эффекта дается при его вероятности $>0,5$, отрицательного — при его вероятности $\leq 0,5$. Для более точного решения поставленной задачи, соответствующего конкретной ситуации, используется ROC-кривая (ROC — Receiver operating characteristic), которая, являясь графиком, позволяет дать оценку качеству модели по разделению двух классов [5–7]. По оси ординат обозначена частота истинно положительных результатов (чувствительность), по оси абсцисс — частота ложноположительных результатов (1 минус специфичность) по всему диапазону точек разделения. Значения по осям соответствуют вероятностям от 0 до 1 (т.е. от 0 до 100%). В результате вырисовывается некоторая кривая (рис. 1). График часто дополняют прямой $y = x$.

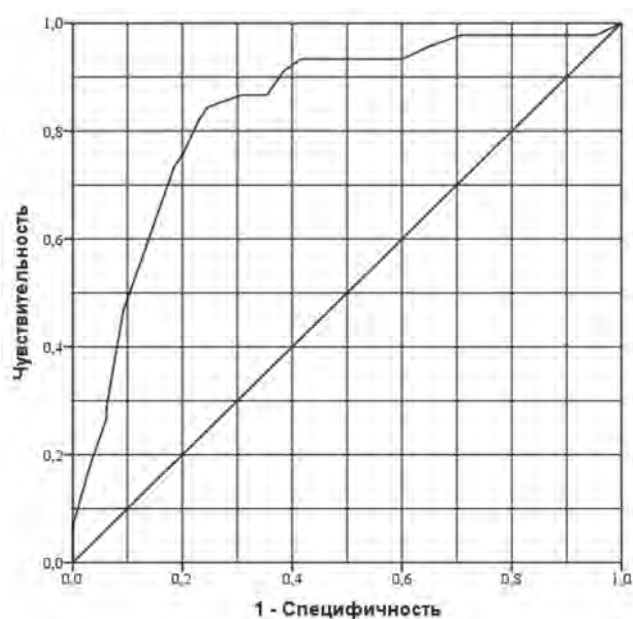


Рис. 1. Пример ROC-кривой

Поскольку классов два, один из них называется классом с положительными исходами, второй — с отрицательными исходами. ROC-кривая показывает зависимость количества верно классифицированных положительных примеров от количества неверно классифицированных отрицательных примеров. При этом предполагается, что у классификатора имеется некоторый параметр, варьируя который, возможно получить то или иное разбиение на два класса. Этот параметр часто называют порогом, или точкой отсечения (cut-off value), которая является наиболее адекватной точкой отсечения одной диагностируемой группы от другой. Как правило, этой точкой является точка, ближе всего расположенная к верхнему левому углу графика или максимально удаленная от диагональной прямой линии.

Для идеального классификатора график ROC-кривой проходит через верхний левый угол, где доля истинно положительных случаев составляет 1,0, или 100% (идеальная чувствительность), а доля ложноположительных примеров равна нулю. Поэтому чем ближе кривая к верхнему левому углу, тем выше предсказательная способность модели. Наоборот, чем меньше изгиб кривой и чем ближе она расположена к диагональной прямой, тем менее эффективна модель. Диагональная линия соответствует «бесполезному» классификатору, т.е. полной неразличимости двух классов.

На рисунке 1 такой точкой может быть точка для 0,8 (80%) чувствительности и 0,22 1-специфичности (соответствует специфичности, равной 78%).

Кроме графика ROC-кривой, для оценки качества модели используется характеристика площади под кривой ROC AUC (Area under ROC). С большими допущениями можно считать, что чем больше показатель AUC, тем лучшей прогностической силой обладает модель. Вся площадь в рамках осей графика равна 1, а площадь под кривой — в долях 1. Чем выше значение площади под ROC-кривой, тем лучше качество модели:

- 0,9–1,0 — отличное;
- 0,8–0,9 — очень хорошее,
- 0,7–0,8 — хорошее;
- 0,6–0,7 — среднее;
- 0,5–0,6 — неудовлетворительное.

Однако следует отметить, что показатель AUC предназначен скорее для сравнительного анализа нескольких моделей и не содержит никакой информации о чувствительности и специфичности модели. В нашем случае (см. рис. 1) площадь под кривой равна $0,836 \pm 0,04$ (95% доверительный интервал 0,758–0,914), что свидетельствует об очень хорошем качестве полученной модели.

Для единого понимания трактовки оценок качества тестируемого метода диагностики приняты

определенные характеристики (дефиниции). Исходные данные, на основе которых рассчитываются характеристики метода, задаются в виде четырехпольной таблицы (табл. 1).

Таблица 1

Четырехпольная таблица исходных данных, на основе которых рассчитываются характеристики метода

Решение по тестируемому методу	Фактическое состояние объектов	
	положительное	отрицательное
Положительное	TP	FP
Отрицательное	FN	TN

Определения характеристик следующие:

— TP (True Positives) — верно классифицированные положительные примеры (истинно положительные случаи);

— TN (True Negatives) — верно классифицированные отрицательные примеры (истинно отрицательные случаи);

— FN (False Negatives) — положительные примеры, классифицированные как отрицательные (ошибка I рода). Это так называемый «ложный пропуск» — когда интересующее нас событие ошибочно не обнаруживается (ложноотрицательные примеры);

— FP (False Positives) — отрицательные примеры, классифицированные как положительные (ошибка II рода). Это ложное обнаружение, т.к. при отсутствии события ошибочно выносится решение о его присутствии (ложноположительные случаи).

При анализе качества диагностического теста чаще оперируют не абсолютными показателями, а относительными — долями, выраженными в процентах, которые транслируются в такие определения, как: чувствительность и специфичность метода.

Чувствительность (Sensitivity) — это и есть доля истинно положительных случаев, которые были правильно идентифицированы тестируемым методом:

$$Se = \frac{TP}{TP + FN} \times 100\%.$$

Специфичность (Specificity) — доля истинно отрицательных случаев:

$$Sp = \frac{TN}{TN + FP} \times 100\%.$$

Модель с высокой чувствительностью часто дает истинный результат при наличии положительного исхода (обнаруживает положительные примеры). Наоборот, модель с высокой специфичностью чаще дает истинный результат при наличии отрицательного исхода (обнаруживает отри-

цательные примеры). Если рассуждать в терминах медицины — задачи диагностики заболевания, где модель классификации пациентов на больных и здоровых называется диагностическим тестом, то получится следующее:

1. Чувствительный диагностический тест проявляется в гипердиагностике — максимальном предотвращении пропуска больных. Допущение повышенной чувствительности приемлемо при разработке скрининговых методов диагностики, где гипердиагностика оправдана.

2. Специфичный диагностический тест диагностирует только подлинно больных. Это важно в случае, когда, например, лечение больного связано с серьезными побочными эффектами и гипердиагностика не желательна.

Не вызывает сомнения, что определение положительного и отрицательного события зависит от конкретной задачи. В одном случае положительным исходом будет класс «Больной пациент», отрицательным — «Здоровый пациент», в другом случае может быть и наоборот.

Идеальная модель обладает 100% чувствительностью и специфичностью. Однако на практике добиться этого невозможно, более того, невозможно одновременно повысить и чувствительность, и специфичность модели. Компромисс находится с помощью порога отсечения, т.к. пороговое значение определяется соотношением чувствительности и специфичности. В этих случаях принято говорить о задаче нахождения оптимального порога отсечения (optimal cut-off value). Порог отсечения нужен для того, чтобы применять модель на практике, а именно относить новые примеры к одному из двух классов. Для уточнения оптимального порога нужно задать критерий его определения, т.к. в разных задачах присутствует своя оптимальная стратегия. Критериями выбора порога отсечения могут выступать:

1. Требование минимальной величины чувствительности модели. Например, нужно обеспечить чувствительность теста не менее 80%. В этом случае оптимальным порогом будет максимальная специфичность, которая достигается при 80% чувствительности.

2. Требование максимальной суммарной чувствительности и специфичности модели, т.е.

$$\text{Cutt_off} = \max (\text{Se} + \text{Sp}).$$

3. Требование баланса между чувствительностью и специфичностью, т.е. когда $\text{Se} \approx \text{Sp}$:

$$\text{Cutt_off} = \min |\text{Se} - \text{Sp}|.$$

Последовательность и технику применения ROC-кривой продемонстрируем на конкретном примере.

Поставлена задача разработать математико-статистическую модель прогноза вероятности развития осложнения атопического дерматита в форме пиодермита у детей, пролеченных в стационаре и катamnестически наблюдавшихся в течение года после выписки [2]. На этапе одномерного анализа из общей совокупности первичных данных определены факторы (признаки), выявляемые при сборе анамнестических особенностей ребенка, наблюдавшиеся в возрасте до одного года и достоверно связанные с развитием осложнения атопического дерматита в форме пиодермита. Характерной особенностью этих признаков с позиций статистического анализа данных является их дихотомический характер — регистрация признака на двух уровнях: наличие его проявления или отсутствие. Дихотомическим является и признак-отклик — наличие или отсутствие осложнения атопического дерматита в форме пиодермита в течение годового катamnестического наблюдения после выписки из стационара. Именно преимущественно дихотомический характер предиктных признаков, а также признака отклика является весомым аргументом при выборе математико-статистического метода моделирования вероятности развития осложнения в пользу метода логистического регрессионного анализа. Селектированная с помощью этого метода модель предназначена для определения индивидуального прогноза вероятности возникновения осложнения атопического дерматита в форме пиодермита на первичном приеме. Признаки, включенные в модель, и их коэффициенты с уровнем значимости приведены в таблице 2.

Таблица 2

Признаки, включенные в модель прогноза вероятности развития осложнений атопического дерматита в форме пиодермита

Наименование признаков		Уровни признаков	Код признаков	Коэффициент (В)	Уровень значимости, р
Прием антибиотиков ребенком в возрасте до года	Да — 1, Нет — 0	X1	1,07	0,029	
Атопический дерматит в анамнезе	Да — 1, Нет — 0	X2	0,72	0,153	
В анамнезе заболевания гепато-билиарной системы	Да — 1, Нет — 0	X3	1,40	0,009	

Окончание таблицы 2

Наименование признаков		Уровни признаков	Код признаков	Коэффициент (В)	Уровень значимости, р
В анамнезе заболевания проявления аллергии	Да – 1, Нет – 0	X4	0,89	0,095	
Болезни кожи в возрасте до года	Да – 1, Нет – 0	X5	2,52	0,000	
Константа			-2,28	0,002	

Модель имеет вид:

$$y = \exp(-2,28 + X1 \times 1,07 + X2 \times 0,72 + X3 \times 1,40 + X4 \times 0,89 + X5 \times 2,52) / (1 + \exp(-2,28 + X1 \times 1,07 + X2 \times 0,72 + X3 \times 1,40 + X4 \times 0,89 + X5 \times 2,52)).$$

Подставляя в уравнение значения признаков, выявленных у конкретного больного, рассчитывается «у» — вероятность развития осложнения атопического дерматита в форме пиодермита. Если рассчитанное значение «у» равно или больше 0,5, то данного больного следует отнести в группу больных, у которых дерматит будет сопровождаться пиодермитом. Если же рассчитанное значение «у» меньше 0,5, то данного больного следует отнести в группу больных, у которых такого осложнения, вероятнее всего, не будет. Статистическая значимость модели оказалась вполне достаточной ($p < 0,001$). Классификационная способность модели определялась по данным обучающей выборки и составила 78,2%. Чувствительность модели оказалась равной 80,0%, а специфичность — 75,6% (табл. 3).

Таблица 3

Классификационная таблица обучающей информации по модели

Результаты прогноза	Результат наблюдения		Процент совпадения
	с осложнением	без осложнения	
С осложнением	52	13	80,0%
Без осложнения	11	34	75,6%

Проведена оценка качества созданной модели с помощью построения ROC-кривой. Предиктор — результат решения диагностической задачи с помощью логистической регрессионной модели для каждого ребенка, включенного в матрицу обучающей информации (Прогноз). Отклик — результат годового катамнестического наблюдения после выписки из стационара: наличие (1) или отсутствие (0) осложнения атопического дерматита в форме пиодермита (осложнение). Исходные данные представлены в таблице 4.

Таблица 4

Результаты катамнестического наблюдения и решения прогностической задачи с помощью математической модели

№	Осложнение	Прогноз	№	Осложнение	Прогноз	№	Осложнение	Прогноз
1	0	0,245	13	0	0,403	25	0	0,622
2	0	0,087	14	0	0,580	26	0	0,117
3	0	0,074	15	0	0,051	27	0	0,062
4	0	0,087	16	0	0,117	28	0	0,541
5	0	0,828	17	0	0,452	29	0	0,051
6	0	0,580	18	0	0,137	30	0	0,137
7	0	0,117	19	0	0,013	31	0	0,403
8	0	0,326	20	0	0,162	32	0	0,663
9	0	0,288	21	0	0,403	33	0	0,278
10	0	0,580	22	0	0,062	34	0	0,031
11	0	0,326	23	0	0,013	35	0	0,622
12	0	0,037	24	0	0,142	36	0	0,117

Окончание таблицы 4

№	Осложнение	Прогноз	№	Осложнение	Прогноз	№	Осложнение	Прогноз
37	0	0,828	62	0	0,051	87	1	0,663
38	0	0,142	63	0	0,051	88	1	0,801
39	0	0,213	64	0	0,403	89	1	0,663
40	0	0,051	65	0	0,142	90	1	0,622
41	0	0,580	66	1	0,801	91	1	0,663
42	0	0,142	67	1	0,452	92	1	0,663
43	0	0,213	68	1	0,541	93	1	0,622
44	0	0,142	69	1	0,908	94	1	0,117
45	0	0,663	70	1	0,801	95	1	0,663
46	0	0,253	71	1	0,908	96	1	0,663
47	0	0,051	72	1	0,580	97	1	0,663
48	0	0,213	73	1	0,580	98	1	0,403
49	0	0,213	74	1	0,828	99	1	0,288
50	0	0,497	75	1	0,707	100	1	0,908
51	0	0,100	76	1	0,497	101	1	0,013
52	0	0,801	77	1	0,828	102	1	0,137
53	0	0,326	78	1	0,580	103	1	0,580
54	0	0,142	79	1	0,622	104	1	0,253
55	0	0,497	80	1	0,580	105	1	0,801
56	0	0,801	81	1	0,497	106	1	0,663
57	0	0,137	82	1	0,580	107	1	0,828
58	0	0,051	83	1	0,828	108	1	0,278
59	0	0,253	84	1	0,828	109	1	0,622
60	0	0,013	85	1	0,497	110	1	0,580
61	0	0,100	86	1	0,580			

Задача построения ROC-кривой решается с помощью пакета по статистической обработке данных SPSS. Файл, созданный в формате Statistica или Excel, открывается в SPSS (рис. 2). Порядок и последовательность конвертации файла – типичные для Windows приложений.

Для решения задачи активируется пункт меню «Анализ» и выбирается процедура «ROC-кривые». В появившемся окне (рис. 3) в качестве проверяемых переменных задается переменная «Прогноз», в качестве переменной состояния – «Осложнение», значение переменной состояния – 1 (обозначение состояния – наличия осложнения).

Активируются ярлычки для вывода результатов решения:

1. ROC-кривые.
2. С диагональной опорной линией.
3. Стандартная ошибка и доверительный интервал.
4. Координатные точки ROC-кривой.

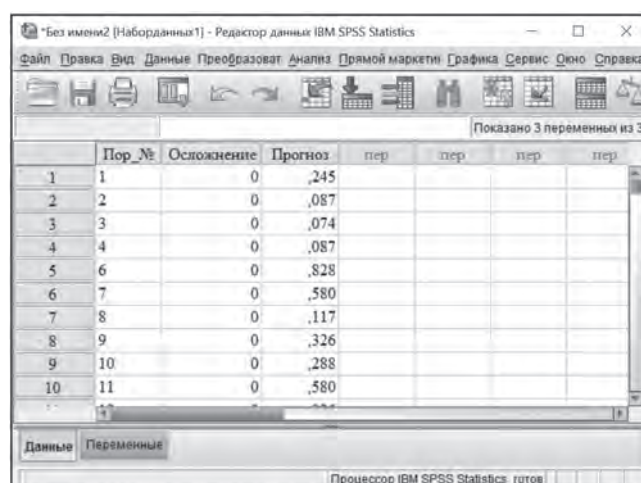


Рис. 2. Вид окна с открытым файлом исходных данных

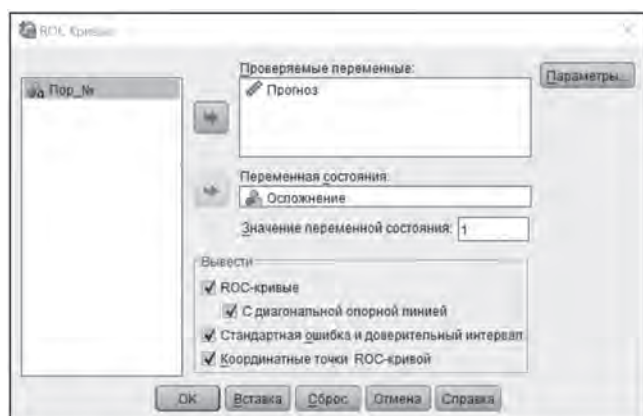


Рис. 3. Вид окна с заданными условиями решения

После нажатия кнопки «ОК» получается решение задачи.

Из анализа графического представления ROC-кривой оптимальной точкой отсечения являются показатель чувствительности, равный 0,85 (85%), и показатель 1-специфичности — 0,25 (специфичность 75%). Площадь под кривой (табл. 5), равная 0,84, говорит об очень хорошем качестве модели. Таблица координат ROC-кривой (табл. 6) позволяет более точно избрать сочетание чувствительности — 0,844 (84,4%) и 1-специфичности — 0,246 (специфичность 0,754 или 75,4%), определившие точку отсечения — 0,43.

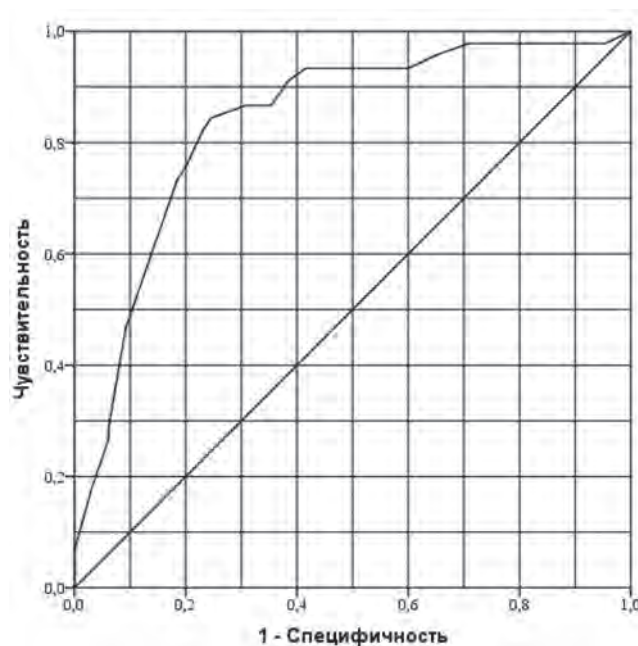


Рис. 4. ROC-кривая диагностической способности логистической регрессионной модели прогноза осложнения атопического дерматита

Таблица 5

Площадь под кривой

Тестовая переменная(ые): Прогноз

Площадь	Стандартная ошибка	Асимптотическая значимость	Асимптотический 95% доверительный интервал	
			Нижняя граница	Верхняя граница
,836	,040	,000	,758	,914

Таблица 6

Координаты кривой

Тестовая переменная(ые): Прогноз

Положительное, если больше или равно	Чувствительность	1-специфичность
0,000	1,000	1,000
0,022	,978	,954
0,034	,978	,938
0,044	,978	,923
0,057	,978	,815
0,068	,978	,785
0,080	,978	,769
0,093	,978	,738
0,108	,978	,708
0,127	,956	,646
0,139	,933	,600

Окончание таблицы 6

Положительное, если больше или равно	Чувствительность	1-специфичность
0,152	,933	,508
0,188	,933	,492
0,229	,933	,431
0,249	,933	,415
0,266	,911	,385
0,283	,889	,369
0,307	,867	,354
0,364	,867	,308
0,428	,844	,246
0,475	,822	,231
0,519	,756	,200
0,560	,733	,185
0,601	,556	,123
0,642	,467	,092
0,685	,289	,062
0,754	,267	,062
0,814	,178	,031
0,868	,067	0,000
1,000	0,000	0,000

Для определения точки отсечения возможно также воспользоваться двумя ранее описанными показателями: максимальной суммарной чувствительности и специфичности модели ($Cutt_off = \max (Se + Sp)$) и баланса между чувствительностью и специфичностью, т.е. когда $Se \approx Sp$ ($Cutt_off = \min |Se - Sp|$). Расчет этих показате-

телей представлен в таблице 7, из которой следует, что точка отсечения по первому показателю ($Se + Sp$) равна 0,428, которой соответствует чувствительность, равная 0,84 (84,4%), и специфичность — 0,754 ($1 - 0,246$) (75,4%). Близкие по значению данные определены по графику и координатам ROC-кривой.

Таблица 7

Координаты ROC-кривой с рассчитанными показателями определения точки отсечения

Положительное, если больше или равно	Чувствительность	специфичность	Специфичность	Se + Sp	Se - Sp
0,000	1,000	1,000	0,000	1,000	1,000
0,022	0,978	0,954	0,046	1,024	0,932
0,034	0,978	0,938	0,062	1,039	0,916
0,044	0,978	0,923	0,077	1,055	0,901
0,057	0,978	0,815	0,185	1,162	0,793
0,068	0,978	0,785	0,215	1,193	0,762
0,080	0,978	0,769	0,231	1,209	0,747
0,093	0,978	0,738	0,262	1,239	0,716
0,108	0,978	0,708	0,292	1,270	0,685
0,127	0,956	0,646	0,354	1,309	0,602

Окончание таблицы 7

Положительное, если больше или равно	Чувствительность	специфичность	Специфичность	Se + Sp	Se-Sp
0,139	0,933	0,600	0,400	1,333	0,533
0,152	0,933	0,508	0,492	1,426	0,441
0,188	0,933	0,492	0,508	1,441	0,426
0,229	0,933	0,431	0,569	1,503	0,364
0,249	0,933	0,415	0,585	1,518	0,349
0,266	0,911	0,385	0,615	1,526	0,296
0,283	0,889	0,369	0,631	1,520	0,258
0,307	0,867	0,354	0,646	1,513	0,221
0,364	0,867	0,308	0,692	1,559	0,174
0,428	0,844	0,246	0,754	1,598	0,091
0,475	0,822	0,231	0,769	1,591	0,053
0,519	0,756	0,200	0,800	1,556	0,044
0,560	0,733	0,185	0,815	1,549	0,082
0,601	0,556	0,123	0,877	1,432	0,321
0,642	0,467	0,092	0,908	1,374	0,441
0,685	0,289	0,062	0,938	1,227	0,650
0,754	0,267	0,062	0,938	1,205	0,672
0,814	0,178	0,031	0,969	1,147	0,791
0,868	0,067	0,000	1,000	1,067	0,933
1,000	0,000	0,000	1,000	1,000	1,000

Современный уровень развития и доступности пакетов прикладных программ по математико-статистической обработке данных научных исследований обеспечивает исследователя большим арсеналом анализа и моделирования сложных биомедицинских систем. Рассмотренный в статье метод математико-статистического моделирования прогноза вероятности развития отдаленных осложнений атопического дерматита в форме пиодермита, разработанный с использованием логистического регрессионного анализа, а также метод ROC-анализа в роли оценки его эффективности, являются лишь малой долей того потенциала, который представляют современные инновационные технологии.

Литература

1. Ермоленко, К.Д. Прогнозирование функциональных заболеваний органов пищеварения у реконвалесцентов ротавирусной инфекции / К.Д. Ермоленко, Н.В. Гончар, С.Г. Григорьев // Материалы всероссийского ежегодного конгресса «Инфекционные болезни у детей: диагностика, лечение и профилактика». СПб., 2016 // Журнал инфектологии. — 2016. — Т. 8, Приложение № 3. — С. 65–66.
2. Кострыкина, Л.Н. Роль инфекционного фактора при атопическом дерматите у детей / Л.Н. Кострыкина : дис.... канд.мед.наук. — СПб., 2007. — 154 с.
3. Юнкеров, В.И. Математико-статистическая обработка данных медицинских исследований / В.И. Юнкеров,

С.Г. Григорьев, М.В. Резванцев. — 3-е изд., доп. — СПб.: ВМедА, 2011. — 318 с.

4. Hosmer D.W. Applied Logistic Regression, 2nd ed. / D.W. Hosmer, S. Lemeshow — N.-Y.: Wiley, 2000. — 375 с.

5. Паклин Николай, <https://basegroup.ru/users/nikolay-paklin>

6. Файнзильберг, Л.С. Гарантированная оценка эффективности диагностических тестов на основе усиленного ROC-анализа / Л.С. Файнзильберг, Т.Н. Жук // Управляющие системы и машины. — 2009. — № 5. — С. 3–13.

7. Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers / T. Fawcett. — Kluwer Acad. Publ., 2004. — 38 p.

8. Каримов, Р.Н. Статистика для врачей, биологов и не только... : монография.-в 2 ч. / Р.Н. Каримов, Ю.Г. Шварц. — Саратов: Саратов. гос. мед. ун-т, 2007, 2010. — Ч.1. — 200 с. — Ч.2. — 204 с.

9. Себер, Дж. Линейный регрессионный анализ / Дж. Себер. — М.: Мир, 1980. — 456 с.

References

1. Ermolenko KD, Gonchar NV, Grigor'yev SG. Prognozirovanie funktsional'nykh zabolevaniy organov pishchevarenia u rekonvalescentov rotavirusnoy infektsii [Forecasting of Digestion Functional Diseases in Rotavirus Infection Reconvalescent Patients]. In: Infektsionnye bolezni u detey: diagnostika, lechenie i profilaktika [Infectious Diseases in Children: Diagnostics, Treatment and Prevention]. Materials of Russian annual congress. — Saint Petersburg (Russia), 2016. Zhurnal infektologii. 2016; V 8, Suppl. 3: 65-6. (in Russian).

2. Kostyrykina LN. Rol' infektsionnogo faktora pri atopicheskom dermatite u detey [Role of Infectious Factor in Case of Atopic Dermatitis in Children] [Dissertation]. Saint Petersburg; 2007. 154 p. (in Russian).
3. Yunkerov VI, Grigor'yev SG, Rezvantsev MV. Mathematical-Statistical Data Processing in Medical Research. 3-rd edition, supplemented. Saint Petersburg: Military Medical Academy; 2011. 318 p. (in Russian).
4. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. N.-Y.: Wiley; 2000. 375 p.
5. Paklin N. [cited 2016 oct16] Available from: <https://base-group.ru/users/nikolay-paklin>.
6. Faynzil'berg LS, Zhuk TN. Upravlyayushchie sistemy i mashiny. 2009; 5: 3-13 (in Russian).
7. Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers. Kluwer Acad. Publ; 2004. 38 p.
8. Karimov RN, Shvarts YuG. Statistics for Doctors, Biologists and Not Only for Them...: monograph work: in 2 parts. Saratov: Saratov State Medical University; 2007, 2010. (in Russian).
9. Seber Dzh. Linear Regression Analysis. Moscow; 1980. 456 p. (in Russian).

Авторский коллектив:

Григорьев Степан Григорьевич — старший научный сотрудник Научно-исследовательского института детских инфекций, д.м.н., профессор; тел.: 8(812)292-34-79, e-mail: gsg_rj@mail.ru

Лобзин Юрий Владимирович — директор Научно-исследовательского института детских инфекций, д.м.н., профессор, академик РАН, заслуженный деятель науки РФ; тел.: 8(812)234-60-04, e-mail: niidi@niidi.ru

Скрипченко Наталья Викторовна — заместитель директора по научной работе Научно-исследовательского института детских инфекций, д.м.н., профессор, заслуженный деятель науки РФ; тел.: 8(812)234-38-22, e-mail: snv@niidi.ru