

И.С. МОРОЗОВА, С.С. САДЫКОВ,
А.С. БЕЛЯКОВА

**Исследование математических
моделей диагностики и
прогнозирования ССЗ (на основе
стенокардии и инфаркта миокарда)**

УДК 004.67, 004.891.3

Муромский институт
(филиал) ФГБОУ ВПО
«Владимирский
государственный
университет имени
А.Г. и Н.Г. Столетовых»,
г.Муром

В статье рассматривается диагностика и прогнозирование ССЗ на основе математических моделей, описываются этапы их получения. После формирования математических моделей диагностики и прогнозирования ССЗ, производится их сравнение. Вычисляются коэффициенты информативности и погрешности, модели тестируются на контрольных выборках. В результате выявляется наиболее точная и надежная математическая модель диагностики и прогноза ССЗ, по которой будут рассчитываться диагнозы пациентов. Использование моделей предполагает упрощение и сокращение времени работы врачей.

Введение

В течение последних 40-50 лет, сердечно-сосудистые заболевания (ССЗ) взяли вверх над человечеством и являются основной причиной смертности большинства европейских популяций. Согласно статистике, в отличие от Западных стран, где отмечается тенденция к снижению смертности от ССЗ, в России наоборот, набирает масштабы, превышая в 6-8 раз аналогичные показатели в развитых странах Европы, Японии, США. Среди общей смертности в России, сердечно-сосудистые заболевания составляют 57% [1,2,3]. Именно поэтому диагностирование и прогнозирование кардиологических заболеваний является наиболее важной и актуальной задачей на сегодняшний день.

Установление диагноза врачом является тончайшим и хорошо обдуманым процессом. На преобразование мышления и анализа специалиста в последовательность математических методов, подталкивает время и сохранение жизни людей.

Не менее важным вопросом считается предварительная обработка исходной информации от пациента. Задача обнаружения взаимосвязанных данных имеет большое значение во многих сферах жизнедеятельности, в том числе и в кардиологии. Как правило, данные от пациента являются разрозненными, не приведенными к одной единице измерения, и на данный момент нет конкретных методов по выявлению связи между ними. Для установления корреляции между параметрами пациента применяются статистические методы обработки данных [4,5,6].

Они позволяют определять скрытые, неявные закономерности, объективно существующие в изучаемых медицинских явлениях, но не поддающиеся непосредственному измерению. Самыми перспективными в этом плане оказались факторный и регрессионный анализы, которые будем использовать для получения моделей диагностики и прогнозирования ССЗ[7,8].

Целью работы является разработка математической модели диагностики и прогнозирования ССЗ (на примере стенокардии и инфаркта миокарда). К задачам исследования относятся:

- выбор математической модели для диагностики и прогнозирования кардиологических заболеваний;
- разработка и исследование математической модели диагностики ССЗ;
- разработка и исследование индивидуальной математической модели прогноза ССЗ.

Разработка и исследование математической модели диагностики ССЗ

В данном случае исследуется достаточно распространенная в мире болезнь сердца – стенокардия. Для диагностики была взята выборка, состоящая из 200 пациентов (96 женщин и 104 мужчин, средний возраст 58.7 лет), для которых оценивались 25 параметров. Однако после предварительной обработки и приведения выборки к количественным единицам измерения, исходные данные сократились до 11 параметров и 200 пациентов. Параметры выборки: воз-

раст, среднее ЧСС, ЧСС, продолжительность боли в груди, Р продолжительность, QRS, QT, депрессия сегмента ST, САД, ДАД, диагноз.

Для построения адекватных математических моделей необходимо, чтобы исходная выборка данных соответствовала нормальному распределению. Для проверки такого соответствия необходимо рассчитать и проанализировать следующие характеристики: мода m_o , медиана m_e , среднее значение фактора \bar{x} .

Медиана определяется следующим образом[9]:

$$m_{ej} = \begin{cases} x_{(N+1)/2j}, & \text{если } N - \text{нечетное} \\ \frac{x_{N/2+1j} + x_{N/2j}}{2}, & \text{иначе} \end{cases}, \quad (1)$$

Среднее значение фактора имеет следующий вид:

$$\bar{x}_j = \frac{\sum_{i=1}^N x_{ij}}{N} \cdot \# \quad (2)$$

Мода m_{oj} j-го фактора, определяемая по выборке, равна наиболее часто встречающемуся значению этого параметра

Распределение является близким нормальным, если выполняется равенство:

$$m_{oj} \cong m_{ej} \cong \bar{x}_j.$$

Была произведена описательная статистика исходной выборки пациентов с диагнозом стенокардия. По результатам вычислений, необходимые параметры имели между собой практическое равенство, следовательно, обучающая выборка имеет нормальное распределение.

Основой для построения факторов служит корреляционная матрица. В виду ее громоздкости приводим только сильные корреляционные связи (таблица 1), которые рассчитываются по следующей формуле[10,11,12]:

$$r_{z_j z_j} = \frac{\sum_{i=1}^n (z_{ij} - \bar{z}_i)(z_{ji} - \bar{z}_j)}{\sqrt{\sum_{i=1}^n (z_{ij} - \bar{z}_i)^2 \sum_{i=1}^n (z_{ji} - \bar{z}_j)^2}} \# \quad (3)$$

где индекс i относится к объектам исследования, а индекс j – к переменным.

Таблица 1

Корреляционные сильные связи

$r_{x_1x_2}$	0.91
$r_{x_1x_3}$	0.69
$r_{x_2x_3}$	0.76
$r_{x_4x_3}$	0.72
$r_{x_4x_9}$	0.64
$r_{x_4x_{10}}$	0.60
$r_{x_4x_9}$	0.86
$r_{x_9x_3}$	0.62
$r_{x_3x_{10}}$	0.61
$r_{x_3x_{11}}$	0.82
$r_{x_9x_{10}}$	0.62
$r_{x_9x_{11}}$	0.78

По количеству наличия сильных (больше 0.7) и умеренных (больше 0.3) связей между параметрами, подводим итог, что дальнейший анализ имеет смысл.

Затем проводился собственно факторный анализ с вычислением факторных нагрузок по методу главных компонент. Жирным шрифтом в таблице 2 выделены факторообразующие факторные нагрузки показателей.

Таблица 2.

Факторные нагрузки восьми факторов

Параметр	$\Phi 1$	$\Phi 2$	$\Phi 3$	$\Phi 4$	$\Phi 5$	$\Phi 6$	$\Phi 7$	$\Phi 8$
Возраст	-0,03	0,93	0,04	0,07	0,10	0,12	-0,14	-0,21
ЧСС	0,01	0,87	0,04	0,06	0,08	0,10	-0,06	-0,12
Продолжительность боли в груди	-0,08	-0,02	0,01	-0,01	-0,02	-0,18	0,24	0,37
Продолжительность зубца Р	0,13	-0,03	0,56	-0,073	0,15	-0,27	-0,09	-0,05
QRS	-0,07	0,01	0,78	0,07	0,40	-0,01	0,01	0,03
QT	-0,09	0,24	-0,50	-0,64	-0,62	0,01	0,01	-0,05
Депрессия ST	-0,87	0,01	0,06	0,03	-0,52	0,01	-0,08	0,01
САД	-0,84	-0,01	-0,07	-0,08	-0,07	-0,23	-0,22	0,06
ДАД	-0,81	-0,1	0,06	0,02	-0,03	0,12	-0,13	-0,18

График собственных значений по Кеттелу («каменистая осыпь») выделенных факторов приведена на рисунке 1. Однако в расчет берем собственные значения факторов более 1.0. Таких факторов оказалось 3:

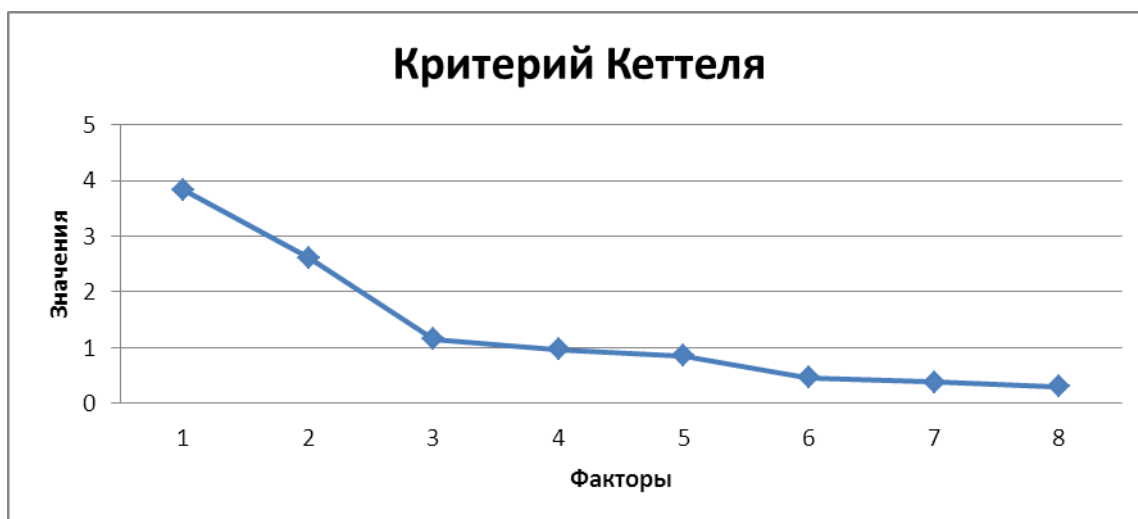


Рис. 1. График собственных значений факторов

Таким образом, обнаружено, что общее состояние организма, в основном, определяется следующими факторами: ЧСС, депрессия сегмента ST, QRS, САД, ДАД, продолжительность боли в груди и Р продолжительность.

Далее для построения математических моделей необходимо рассчитать коэффициенты регрессии. В общем случае уравнение характеризует зависимость между вариацией показателя и вариациями факторов. Функцию следует выбирать так, чтобы незначительное изменение аргументов приводило к незначительному изменению функции. Было рассмотрено 3 вида математических зависимостей: линейная полиномиальная и экспоненциальная.

Регрессионный анализ исследует влияние одной из нескольких независимых переменных x_i на зависимую переменную y_i . Связь между каждым значением y_i и $x_{i1}, x_{i2}, \dots, x_{in}$ описывается системой линейных уравнений.

$$\begin{cases} y_1 = a_0 + \sum_{j=1}^n a_j x_{1j} \\ y_2 = a_0 + \sum_{j=1}^n a_j x_{2j} \\ \dots \\ y_N = a_0 + \sum_{j=1}^n a_j x_{Nj} \end{cases}, \quad i=1, \dots, N; j=1, \dots, n, \quad (4)$$

где a_0 - свободный член;

a_1, a_2, \dots, a_n - коэффициенты регрессии - неизвестные параметры, которые необходимо оценить по результатам выборочных наблюдений; N - объем выборки; n - количество анализируемых параметров.

При проведении факторного анализа для случаев стенокардии критерий каменистой осыпи позволил выделить 3 основных фактора, имеющих наибольший процентный вклад в общую дисперсию:

- фактор 1 за продолжительность боли в груди в минутах, депрессию сегмента ST, САД и ДАД;
- фактор 2 за возраст, ЧСС и среднюю ЧСС;
- фактор 3 за QRS.

По итогу получения коэффициентов сформированы 3 вида функции диагностики ССЗ:

- линейная модель

$$\begin{aligned} \text{Результат} = & -0.001 * x_1 + 0.007 * x_2 - 0.001 * x_3 + 0.025 * x_4 + \\ & + 0.17 * x_5 - 0.24 * x_6 - 0.14 * x_7 + 2.47 * x_8 + 0.006 * x_9 + 0.007 * x_{10} - 0.79 \end{aligned} \quad (5)$$

где x_1 - возраст, x_2 - среднее ЧСС, x_3 - ЧСС, x_4 - продолжительность боли в груди, x_5 - Р продолжительность, x_6 - QRS, x_7 - QT, x_8 - депрессия сегмента ST, x_9 - САД, x_{10} - ДАД.

- экспоненциальная модель

$$\text{Результат} = e^x, \quad (6)$$

где

$$\begin{aligned} x = & -0.001 * x_1 + 0.007 * x_2 - 0.001 * x_3 + 0.025 * x_4 + 0.17 * x_5 - \\ & - 0.24 * x_6 - 0.14 * x_7 + 2.47 * x_8 + 0.006 * x_9 + 0.007 * x_{10} - 0.79 \end{aligned}$$

- полиномиальная модель

$$\begin{aligned}
\text{Результат} = & 0.01 * x_1 - 0.0002 * x_1^2 + 0.0002 * x_2^2 - 0.002 * x_3 + 0.00002 * x_3^2 + \\
& + 0.11 * x_4 - 0.003 * x_4^2 - 1.59 * x_5 + 18.57 * x_5^2 - 0.013 * x_6 - 2.42 * x_6^2 + 4.1 * x_7 - \\
& - 4.91 * x_7^2 + 1.78 * x_8 - 3.19 * x_8^2 + 0.002 * x_9 + 0.026 * x_{10} - 0.0001 * x_{10}^2 - 2.94
\end{aligned} \quad (7)$$

По результатам полученных моделей диагностики ССЗ был произведен анализ оценки информативности (коэффициент детерминации) моделей. Коэффициент показывает силу корреляционной связи между параметрами и откликом, а так же содержит ли построенная модель достаточную информацию об исследуемом объекте.

$$D = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (8)$$

Результаты критериев приведены в таблице 3:

Таблица 3

Вид модели	Линейная	Полиномиальная	Экспоненциальная
Оценка информативности	0.82	0.94	0.67

Следует отметить, что наиболее информативной по результатам коэффициента детерминации является полиномиальная модель. Однако, по результатам тестирования модели на исходной выборке, представленными на рисунке 2, наиболее предпочтительной и адекватной в практике оказалась линейная модель диагностики ССЗ.

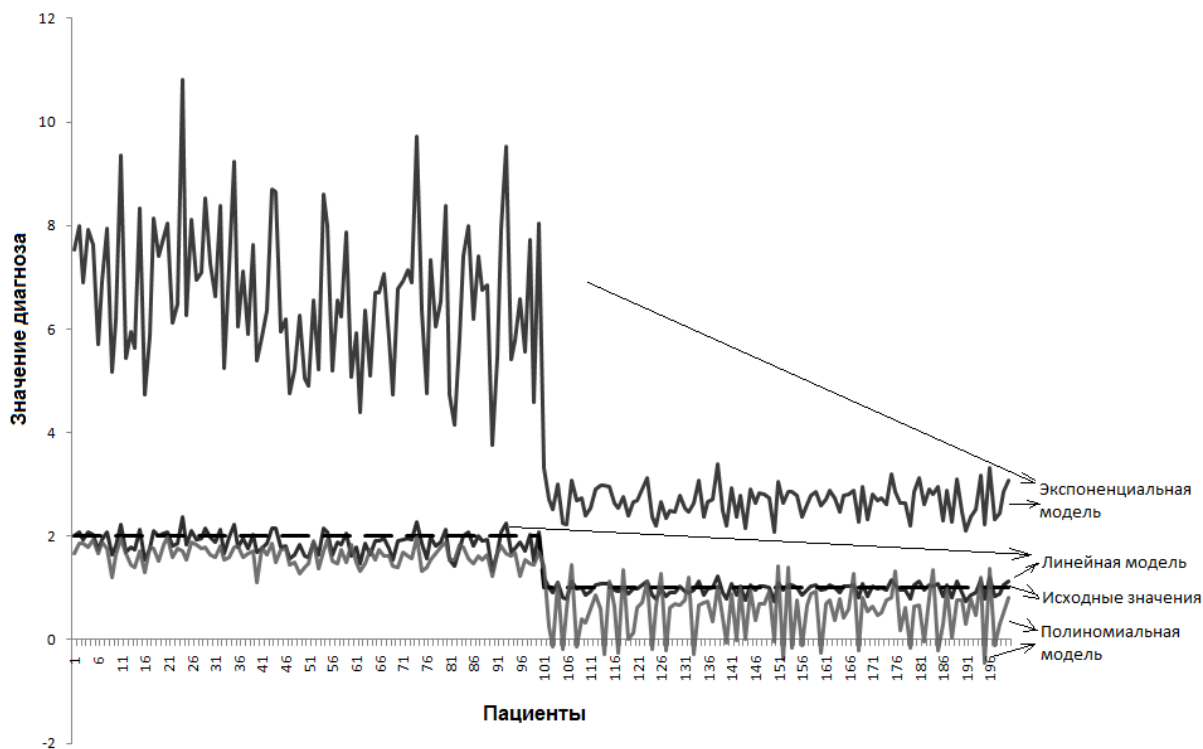


Рис. 2. Графическое сравнение результатов 3х моделей

Таким образом, при диагностировании ССЗ группы пациентов наибольшее внимание будет сконцентрировано на результатах линейной модели.

Разработка и исследование математической модели прогнозирования ССЗ

Построение модели прогнозирования возникновения инфаркта миокарда при стенокардии строится по аналогичному плану построения модели диагностирования стенокардии. Для вычисления математических моделей была взята выборка обследований одного пациента (13 просмотров по 12 параметрам). Зависимыми параметрами в данном случае являются SDR, SIT и Инд1.

Результаты пациента, в возрасте 45 лет, представлены в таблице 4.

Таблица 4

Результаты обследований пациента

Дата	ЧСС	СД	ДД	СГД	ПАД	РД	SDR	SIT	Инд1	МО	Амо	X
11.10.2011	82	117	82	93.67	35	5.68	174.66	31.38	79	625	25	173.08
13.10.2011	89	113	77	89	36	5.28	196.38	25.24	92	625	3	223.78
14.10.2011	84	111	69	83	42	4.28	214.26	39.55	100	714.29	36	117.06
15.10.2011	87	121	77	91.67	44	4.5	191.15	59.16	84	714.29	24	106.71
18.10.2011	87	119	70	86.33	49	3.86	195.49	36.27	90	681.82	31	150
19.10.2011	77	120	80	93.33	40	5	171.41	28.31	66	750	25	166.67

22.10.2011	82	115	69	84.33	46	4	143.59	23.8	64	652.17	21	210.81
23.10.2011	86	123	74	90.33	49	4.02	231.33	41.9	101	681.82	35	164.47
26.10.2011	87	123	71	88.33	52	3.73	196.21	22.88	88	652.17	29	210.81
27.10.2011	77	131	70	90.33	61	3.29	127.89	20.85	49	731.71	18	242.42
29.10.2011	73	133	74	93.67	59	3.51	121.91	30.27	43	833.33	18	333.33
30.10.2011	78	131	74	93	57	3.6	128.77	24.79	49	731.71	18	215.69
31.10.2011	76	122	78	92.67	44	4.54	176.29	28.34	67	789.47	27	194.81

Индивидуальная модель прогнозирования ССЗ, на примере параметра SDR, имеет следующий вид:

- линейная модель

$$\begin{aligned} \text{Результат} = & 7.73 * x_1 + 1194.75 * x_2 + 1292.63 * x_3 - 2487.98 * x_4 - \\ & - 364.61 * x_5 + 41.05 * x_6 + 0.37 * x_7 + 2.20 * x_8 + 0.15 * x_9 - 970.93 \end{aligned} \quad (9)$$

Где x_1 - ЧСС, x_2 - СД, x_3 - ДД, x_4 - СГД, x_5 - ПАД, x_6 - РД, x_7 - МО, x_8 - АМО, x_9 - Х.

- экспоненциальная модель

$$\text{Результат} = e^x, \quad (10)$$

где

$$\begin{aligned} x = & -970.93 + 7.73 * x_1 + 1194.75 * x_2 + 1292.63 * x_3 - 2487.98 * x_4 - \\ & - 364.61 * x_5 + 41.05 * x_6 + 0.37 * x_7 + 2.20 * x_8 + 0.15 * x_9 \end{aligned}$$

- полиномиальная модель

$$\begin{aligned} \text{Результат} = & 247583.8 - 2456.6 * x_1 + 14.8 * x_1^2 - 8598.6 * x_2 + 26.5 * x_2^2 - \\ & - 5232.5 * x_3 + 45.6 * x_3^2 + 15583.4 * x_4 - 77.5 * x_4^2 - 50085.9 * x_6 + \\ & + 3135.5 * x_6^2 + 70.3 * x_7 - 0.1 * x_7^2 \end{aligned} \quad (11)$$

По результатам тестирования модели на обучающей и контрольной выборке (таблица 5), наиболее точной в значениях и адекватной стала линейная модель.

Таблица 5

Сравнение результатов тестирования параметра SDR на обучающей выборке

SDR (исходный)	Результат SDR полученные с помощью линейной модели	Результат SDR полученные с помощью экспоненциальной модели	Результат SDR полученные с помощью полиномиальной модели
174.66	179.27	7245	-18.91

196.38	188.28	5880	-18.92
214.26	207.89	1930	-24.84
191.15	200.69	1443	-24.87
195.49	209.15	6823	-22.30
171.41	179.85	1291	-27.36
143.59	150.56	2442	-20.13
231.33	216.56	1125	-22.40
196.21	199.40	3995	-20.84
127.89	121.93	9046	-25.10
121.91	130.67	5616	-33.38
128.77	124.43	1101	-25.49
176.29	164.33	2344	-29.97

По результатам тестирования для расчета параметров SDR, SIT и Инд1, линейная модель прогнозирования ССЗ имеет наибольшее предпочтение.

Заключение

В результате анализа и выполнения поставленных задач было получено:

1. Полиномиальная, линейная, экспоненциальная математические модели диагностики ССЗ;
2. Наиболее значимой и информативной по отношению к получению результатов заключений пациентов является линейная модель. Именно на данный вид модели обращается наибольшее внимание.
3. Получены линейная, экспоненциальная и полиномиальная индивидуальные математические модели прогнозирования ССЗ.
4. Наиболее адекватной при тестировании и получении оценки эффективности оказалась линейная модель.

Использование математической модели диагностирования и прогнозирования ССЗ дает наиболее объективное установление диагноза и предназначено для помощи врачу-кардиологу при диагностике.

Литература

1. Садыков С.С., Белякова А.С. Математические модели некоторых сердечно-сосудистых заболеваний // Информационные технологии. №12, 2011, с. 59-63.
2. Садыков С.С., Белякова А.С. Регрессионные модели стенокардии и зависимость их информативности от количества параметров работы сердца // Системы управления и информационные технологии. №3.1(45), 2011, с.190-194.
3. Белякова А.С. Основные признаки оценки состояния сердечно-сосудистой системы // Алгоритмы, методы и системы обработки данных. 2009. №14. С. 14-31.
4. Садыков С.С., Белякова А.С., Пугин Е.А. Математическая модель диагностики инфекционного эндокардита // Алгоритмы, методы и системы обработки данных. 2012. №19 С. 150-154.
5. Евстигнеева О.И., Садыков С.С., Сулова Е.Е., Белякова А.С. Критерии выделения групп риска из лиц трудоспособного возраста при медицинских исследованиях системе АСПО // Алгоритмы, методы и системы обработки данных. 2012. №19 С. 33-39.
6. Садыков С.С., Орлов И.А., Белякова А.С.. Разработка и исследование математических моделей ссз: нарушения ритма и проводимости // Алгоритмы, методы и системы обработки данных. 2012. №19 С. 179-187.
7. Гарри Харман Современный факторный анализ. Перевод с английского В. Я. Лумельского. Москва: Статистика, 1972. – 484 с.;
8. Иберла К. Факторный анализ / Пер. с нем. В.М.Ивановой; Предисл. А.М.Дуброва. – М.: Статистика, 1980. – 398 с., ил. – (Математико-статистические методы за рубежом);
9. Новиков Д. А., Новочадов В.В. Статистические методы в медико-биологическом эксперименте (типовые случаи). – Волгоград: «ВолГМУ», 2005. – 84 с.;
10. Степанов В. Г. Эконометрика. – М.: МИЭМП, 2009.;
11. Халафян А. А. STATISTIC A 6. Статистический анализ данных. 3-е изд. Учебник — М: ООО «Бином-Пресс», 2007 г. — 512 с: ил.;
12. Юнкеров В. И., Григорьев С. Г. Математико-статистическая обработка данных медицинских исследований. – СПб.: ВМедА, 2002. – 266 с.

MAIL: IRINA@F5F5.RU