

ПРОГНОЗИРОВАНИЕ СТЕПЕНИ ТЯЖЕСТИ СОСТОЯНИЯ НЕДОНОШЕННЫХ НОВОРОЖДЕННЫХ СО СРОКОМ ГЕСТАЦИИ МЕНЕЕ 30 НЕДЕЛЬ

Тырсин А.Н.^{1,2}, Улезко Е.А.³, Остроушко Д.В.³, Свирская О.Я.³, Санковец Д.Н.³

¹ФГАОУ ВО Уральский федеральный университет имени первого Президента России Б.Н. Ельцина, Екатеринбург, e-mail: at2001@yandex.ru;

²ФГАОУ ВО Южно-Уральский государственный университет (национальный исследовательский университет), Челябинск, e-mail: at2001@yandex.ru;

³Государственное учреждение «Республиканский научно-практический центр «Мать и дитя», Минск, e-mail: sevenhos@mail.belpak.by

Целью исследования является разработка статистически достоверного метода оценки степени тяжести состояния недоношенных новорожденных со сроком гестации менее 30 недель при рождении с использованием многомерного статистического анализа данных. За основу была взята модель бинарной логистической регрессии. В качестве оценки степени тяжести новорожденного была использована вероятность его отнесения к соответствующей группе пациентов. Для решения поставленной задачи проведена оценка анамнестических данных и результатов лечебно-диагностических процедур 29 недоношенных новорожденных. Информативными признаками явились масса тела ребенка при рождении в граммах, метод родоразрешения, рН и уровень лактата крови КОС при рождении, минимальный уровень насыщения крови кислородом при проведении респираторной поддержки, зафиксированный за первые 24 часа жизни, гемодинамический статус, минимальное и максимальное значение общего содержания кислорода крови КОС первых 24 часов жизни, минимальное значение парциального давления кислорода крови КОС первых 24 часов жизни. По имеющейся выборке прецедентов разработано статистически достоверное решающее правило, на основе которого может быть создана методика распознавания степени тяжести состояния новых пациентов. Проведен сравнительный энтропийный анализ двух групп пациентов.

Ключевые слова: недоношенные новорожденные, прогноз, модель, логистическая регрессия, дискриминантный анализ, энтропия.

PREDICTING THE SEVERITY OF THE CONDITION OF PREMATURE NEWBORNS WITH A GESTATION PERIOD OF LESS THAN 30 WEEKS

Tyrsin A.N.^{1,2}, Ulezko E.A.³, Ostroushko D.V.³, Svirskaya O.Y.³, Sankovets D.N.³

¹Federal State Autonomous Educational Institution of Higher Education Ural Federal University named after the first President of Russia B.N. Yeltsin, Yekaterinburg, e-mail: at2001@yandex.ru;

²Federal State Autonomous Educational Institution of Higher Education South-Ural State University (National Research University), Chelyabinsk, e-mail: at2001@yandex.ru;

³Mother and Child National Research Center, Minsk, e-mail: darija12@yandex.ru

The aim of the study is to develop a statistically reliable method for assessing the severity of the condition of premature newborns with a gestation period of less than 30 weeks at birth using multivariate statistical data analysis. The model of binary logistic regression was used as a basis. As an assessment of the severity of the newborn, the probability of its assignment to the corresponding group of patients was used. To solve this problem, anamnestic data and results of medical and diagnostic procedures were evaluated in 29 premature newborns. Informative signs were the child's weight at birth in grams, delivery method, pH, and blood lactate level KOS at birth, the minimum level of blood oxygen saturation during respiratory support, fixed for the first 24 hours of life, hemodynamic status, minimum and maximum value of the total oxygen content of blood KOS first 24 hours of life, the minimum value of partial pressure of blood oxygen KOS first 24 hours of life. Based on the available sample of precedents, a statistically reliable decision rule has been developed, on the basis of which a method for recognizing the severity of the condition of new patients can be developed. A comparative entropy analysis of two groups of patients was performed.

Keywords: preterm infants, forecast, model, logistic regression, discriminant analysis, entropy.

В результате внедрения современных технологий интенсивной терапии и реанимации в неонатологии ежегодно увеличивается число выживших недоношенных детей со сроком

гестации менее 30 недель [1, 2]. Разработка новых подходов выхаживания и терапевтического ведения таких пациентов является весьма актуальной задачей, поскольку для недоношенных новорожденных характерен рост заболеваемости в неонатальном периоде. Для повышения эффективности выхаживания важно своевременно распознать тяжесть нарушений, происходящих в организме ребенка. Это позволит выбрать правильный режим лечения, снижающий риск возможных осложнений пациентов. Перинатальная патология у недоношенных детей характеризуется большей тяжестью и сочетанностью нарушений вследствие незрелости всех систем организма, развившейся на фоне отсутствия внутриутробного благополучия плода во время беременности [3, 4]. Это приводит к необходимости совершенствования методов оценивания состояния здоровья у таких детей.

В настоящее время накоплен огромный опыт в области моделирования биологических систем, в том числе в медицине. Такое моделирование основано на использовании различных направлений цифровых технологий, которые позволяют врачам независимо от уровня профессиональной подготовки принять правильное решение.

Многие модели успешно зарекомендовали себя и внедрены в лечебный процесс больниц, они являются дополнительным подспорьем для специалиста при принятии решения о выборе той или иной тактики лечения, а также позволяют оценивать имеющийся потенциал здоровья пациента [5]. Другие модели позволяют достоверно распознать возможность развития заболевания в будущем и привлечь дополнительное внимание лечащего врача к конкретному пациенту [6]. Поэтому для выбора оптимального метода выхаживания и реабилитации недоношенных детей важно дать объективный прогноз состояния их здоровья.

Целью исследования явилась разработка прогностического правила для методики прогнозирования в целях оценки степени тяжести состояния здоровья недоношенных новорожденных со сроком гестации менее 30 недель. В качестве инструментария для разработки методики были использованы методы многомерного статистического анализа.

Материалы и методы исследования

Для решения поставленной задачи проведена оценка анамнестических данных и результатов лечебно-диагностических процедур 29 недоношенных новорожденных со сроком гестации менее 30 недель при рождении. Наблюдение за пациентами и их обследование начинались непосредственно в родильном зале, а далее продолжались в отделении анестезиологии и реанимации (с палатами для новорожденных детей) ГУ «РНПЦ “Мать и дитя”», г. Минск. Средний гестационный возраст младенцев составил $28,25 \pm 1,03$ недели. Масса тела при рождении составила $1094,48 \pm 205,63$ г. Мальчиков было 15,00 (51,72%), девочек – 14 (48,27%). 20 (68,97%) новорожденных были от одноплодной беременности, 9 (31,03%) – от многоплодной. Через естественные родовые пути родился 1 (1/3,45%) ребенок,

при абдоминальном родоразрешении – 28 (96,55%) младенцев (20 (68,97%) – при экстренном и 8 (27,58%) – при плановом родоразрешении). Все дети (29 (100,00%)) в родильном зале получили сурфактантзаместительную терапию в дозе $227,86 \pm 72,47$ мг/кг. Среди основных клинических диагнозов отмечены следующие состояния: респираторный дистресс-синдром новорожденных, врожденная пневмония, инфекция, специфичная для перинатального периода неуточненная, врожденный сепсис, инфекция, специфичная для перинатального периода.

Проанализированы 20 показателей: анамнестические данные, кислотно-основное состояние (КОС) артериальной крови при рождении и первого дня жизни, спектроскопия в ближней инфракрасной области (Near-infrared spectroscopy, NIRS), респираторная поддержка новорожденного, гемодинамический статус и гидробаланс первых 24 часов жизни ребенка (табл. 1). Метод NIRS успешно применяется для оценки мозгового кровотока и объема крови в сосудах головного мозга как раннего предиктора перинатальных поражений центральной нервной системы у недоношенных новорожденных [7, 8]. Обследуемые пациенты нуждались в проведении респираторной поддержки. Время осмотра: для признаков P0 – P5 – в родильном зале, для признаков P6 – P19 – в первые 24 часа жизни.

Таблица 1

Анамнестические и лабораторные данные

Данные	Признак	Условное обозначение
Анамнез	Пол: ж – 0, м – 1	P0
	Масса при рождении (г.)	P1
	Беременность: 0 – одноплодная, 1 – многоплодная	P2
	Метод родоразрешения: 0 – через естественные родовые пути, 1 – плановое кесарево сечение, 2 – экстренное кесарево сечение	P3
КОС	pH	P4
КОС	Лас	P5
NIRS	ctSO ₂ (среднее)	P6
	ΔcrSO ₂ (макс-мин))	P7
Пульсоксиметрия	SpO ₂ MIN при проведении ИВЛ	P8
Клиническое наблюдение	Гемодинамический статус	P9
КОС (артериальная кровь)	pH MIN	P10
	pH MAX	P11
	pCO ₂ MIN	P12
	pCO ₂ MAX	P13
	pO ₂ MIN	P14
	pO ₂ MAX	P15
	ctO ₂ MIN	P16
	ctO ₂ MAX	P17
p50 MIN	P18	

Исследование включало в себя несколько этапов. Вначале необходимо у недоношенных детей найти число групп по степени тяжести состояния здоровья. Предварительная гипотеза состояла в том, что у недоношенных детей имеет место градация на две группы – «тяжелые» и «очень тяжелые». Необходимо было проверить справедливость предположения о двух уровнях степени тяжести. Здесь были использованы методы кластерного анализа данных [9]. На втором этапе была сформирована система информативных диагностических признаков. Эта процедура была выполнена с помощью дискриминантного анализа [10].

Важнейший этап – количественный прогноз степени тяжести состояния детей – осуществим с помощью логистической регрессии. Она удобна тем, что позволяет формировать хорошо интерпретируемый показатель степени тяжести в виде вероятности отнесения пациента к той или иной группе больных. Применительно к двум классам это выглядит следующим образом.

Дана обучающая выборка значений признаков X_1, X_2, \dots, X_m

$$(\mathbf{x}_i, y_i), i = 1, 2, \dots, n, \quad (1)$$

$$\text{где } \mathbf{x}_i = \begin{pmatrix} x_{i0} \\ x_{i1} \\ \dots \\ x_{im} \end{pmatrix} = \begin{pmatrix} 1 \\ x_{i1} \\ \dots \\ x_{im} \end{pmatrix} - \text{вектор значений } i\text{-го объекта, } \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix}; \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix},$$

$y_i \in \{0; 1\}$ – бинарная переменная, указывающая на принадлежность i -го объекта соответствующему классу. Например, для определенности первому классу при $y_i = 0$ и второму – при $y_i = 1$; m – число признаков у каждого объекта; n – количество наблюдений. Классификация выполняется с помощью логистической функции [11]:

$$h(\mathbf{x}) = \frac{1}{1 + \exp\{-\mathbf{b}^T \mathbf{x}\}}, \quad (2)$$

принимая значения в интервале $(0; 1)$. Пороговым значением является $h(\mathbf{x}) = 0,5$. Вектор $\mathbf{b}^T = (b_0 \ b_1 \ \dots \ b_m)$ в (2) задает разделяющую линейную границу, описываемую уравнением гиперплоскости Π : $\mathbf{b}^T \mathbf{x} = 0$.

Введем функцию $W(\mathbf{x}) = \mathbf{b}^T \mathbf{x}$. Зададим область D_1 возможных значений \mathbf{x} для первого класса как $D_1 = \{\mathbf{x} : W(\mathbf{x}) < 0\}$, а для второго класса – как $D_2 = \{\mathbf{x} : W(\mathbf{x}) > 0\}$. Тогда $\forall \mathbf{x} \in D_1$ $h(\mathbf{x}) < 0,5$ и $\forall \mathbf{x} \in D_2$ $h(\mathbf{x}) > 0,5$. Если \mathbf{x} принадлежит гиперплоскости Π , то $h(\mathbf{x}) = 0,5$.

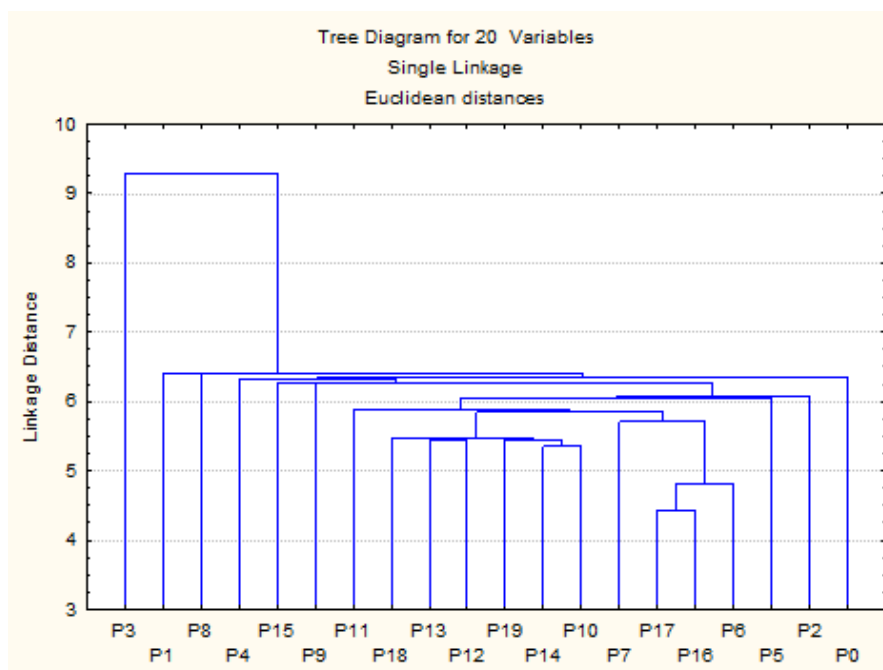
Следовательно, для произвольного наблюдения \mathbf{x}^* вероятность его отнесения к первому классу равна $P(\mathbf{x}^* \in D_1) = 1 - h(\mathbf{x}^*)$, а ко второму – $P(\mathbf{x}^* \in D_2) = h(\mathbf{x}^*)$.

Для оценивания вектора коэффициентов \mathbf{b} был использован алгоритм Ньютона–Рафсона [12]. Представляет интерес также исследование системных свойств каждой из групп пациентов с помощью энтропийного моделирования [6, 13].

Методы многомерного статистического анализа реализованы с помощью статистического пакета программ Statistica. Энтропийное моделирование выполнялось с помощью программы вычисления энтропии случайного вектора [14].

Результаты исследования и их обсуждение

Для уточнения числа групп недоношенных детей по степени тяжести состояния здоровья был использован традиционный в таких задачах кластерный анализ данных. Дендрограмма (рис.) наглядно показывает наличие двух выраженных кластеров.



Дендрограмма

Слева на рисунке приводится расстояние между объединяемыми на данном этапе кластерами (объектами), внизу указаны признаки. С учетом результатов кластерного анализа все пациенты ($n=29$) были разделены на две группы: к группе «тяжелые» (D_1) отнесены 17 пациентов, к группе «очень тяжелые» (D_2) отнесены 12 пациентов.

Решение задачи состоит в попытке различить между собой группы пациентов D_1 и D_2 и в случае успеха построить решающее правило классификации. Для этого нужно было решить задачу многомерной классификации (распознавания) двух групп (кластеров) по показателям анамнестических данных и результатов лечебно-диагностических процедур. Суть решения состояла в нахождении такой совокупности показателей из исходного множества,

которая позволила бы (если это в принципе возможно) статистически достоверно распознать различия в этих группах. Данная задача решалась в два этапа. На первом этапе с помощью дискриминантного анализа формировалась система информативных признаков, а на втором с помощью логистической регрессии строилось решающее правило классификации. Строилась дискриминантная функция с минимальным p -уровнем, который составил 0,0004. Информативными признаками явились следующие качественные и количественные показатели: X_1 – масса тела ребенка при рождении в граммах; X_2 – метод родоразрешения (0 – через естественные родовые пути, 1 – плановое кесарево сечение, 2 – экстренное кесарево сечение); X_3 – pH КОС при рождении, X_4 – уровень лактата крови КОС при рождении, X_5 – минимальная SpO_2 при проведении искусственной вентиляции легких, зафиксированная за первые 24 часа жизни, X_6 – гемодинамический статус (0 – нет терапии; 1 – допамин ≤ 5 мкг/кг мин; 2 – допамин > 5 и ≤ 10 мкг/кг мин; 3 – допамин > 10 мкг/кг мин или допамин+добутамин ≤ 10 мкг/кг мин; 4 – допамин+добутамин > 10 мкг/кг мин; 5 – эпинефрин или норэпинефрин и/или глюкокортикостероиды; X_7 – минимальное значение общего содержания кислорода (ctO_2) крови КОС первых 24 часов жизни; X_8 – максимальное значение ctO_2 крови КОС первых 24 часов жизни; X_9 – минимальное значение парциального давления кислорода ($p50$) крови КОС первых 24 часов жизни.

Распознавание двух групп (D_1) и (D_2) выполнено формально недостаточно достоверно, поскольку признаки pH крови КОС при рождении и $p50$ крови КОС первых 24 часов жизни имеют недостаточную статистическую надежность (менее 95%). Однако расчетное значение F-статистики статистически значимо со статистической надежностью 0,9996 (выше требуемой величины 0,95). Данные спектроскопии в ближней инфракрасной области (NIRS) оказались неинформативными показателями для решения поставленной задачи. Средняя вероятность правильного распознавания составила 96,552%, что выше требуемой величины 95%. Это означает, что сформированная система признаков в достаточной степени различает «тяжелых» и «очень тяжелых» пациентов.

Решающее правило построено по обучающему множеству признаков X_0 и Y_0 , которое образуют пациенты с известными диагнозами. Решающее правило предназначено для классификации состояний новых пациентов.

Учитывая существенные признаки, полученные ранее с помощью дискриминантного анализа, мы рассчитали интегративный прогностический индекс (ИПИ) по формуле:

$$Z = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9, \quad (4)$$

где $b_0 = -11,591$, $b_1 = 9,333$, $b_2 = 6,808$, $b_3 = -3,056$, $b_4 = -14,060$, $b_5 = 13,964$, $b_6 = -11,353$, $b_7 = 19,082$, $b_8 = -18,484$, $b_9 = -5,997$.

Если рассчитанный по прогностическому правилу (4) результат будет меньше нуля, то у ребенка прогнозируется «очень тяжелое» состояние с вероятностью $P_0 = e^{-Z}/1 + e^{-Z}$. Если результат Z будет больше нуля, то у ребенка прогнозируется «тяжелое состояние» с вероятностью $1 - P_0 = 1/1 + e^{-Z}$.

Все пациенты правильно классифицированы. В таблице 2 приведены результаты распознавания. Здесь обозначены: № – номер наблюдения (порядковый номер ребенка в общей выборке; IND – принадлежность наблюдений к одной из групп (IND=0 – группа «тяжелые», IND=1 – группа «очень тяжелые»); P_0 – вероятность отнесения наблюдения (ребенка) к группе «очень тяжелые» (IND=1); $1 - P_0$ – вероятность отнесения наблюдения (ребенка) к группе «тяжелые» (IND=0).

Таблица 2

Вероятности отнесения пациентов к группам «тяжелые» и «очень тяжелые»

№№	IND	Z	P_0	$1 - P_0$	№№	IND	Z	P_0	$1 - P_0$
1	1	-37,233	1,00000	0,00000	16	1	-9,655	0,99994	0,00006
2	1	-23,257	1,00000	0,00000	17	0	36,913	0,00000	1,00000
3	0	8,145	0,00029	0,99971	18	0	17,338	0,00000	1,00000
4	0	7,214	0,00074	0,99926	19	0	5,414	0,00443	0,99557
5	1	-27,895	1,00000	0,00000	20	0	16,257	0,00000	1,00000
6	1	-7,465	0,99943	0,00057	21	0	10,433	0,00003	0,99997
7	1	-5,686	0,99662	0,00338	22	0	32,437	0,00000	1,00000
8	0	6,101	0,00224	0,99776	23	0	21,573	0,00000	1,00000
9	0	25,104	0,00000	1,00000	24	1	-36,413	1,00000	0,00000
10	0	8,310	0,00025	0,99975	25	0	5,734	0,00322	0,99678
11	0	36,550	0,00000	1,00000	26	1	-30,228	1,00000	0,00000
12	1	-5,974	0,99746	0,00254	27	0	16,309	0,00000	1,00000
13	0	10,513	0,00003	0,99997	28	0	22,199	0,00000	1,00000
14	1	-7,951	0,99965	0,00035	29	1	-47,549	1,00000	0,00000
15	1	-30,438	1,00000	0,00000	–	–	–	–	–

Результаты энтропийного анализа по статистически значимым признакам приведены в таблице 3. Видно, что в группе «очень тяжелые» по сравнению с группой «тяжелые» энтропии хаотичности и самоорганизации меньше. Это означает, что в целом у «очень тяжелых» пациентов в целом вариация показателей, их изменчивость ниже, а теснота взаимосвязи между показателями выше.

Таблица 3

Результаты энтропийного анализа групп «тяжелые» и «очень тяжелые»

IND	Энтропия хаотичности H_V	Энтропия самоорганизации H_R	Общая энтропия H
0	11,108	-2,732	8,376
1	10,411	-3,803	6,609

Заключение

Использование методов многомерного статистического анализа позволило разработать алгоритм оценки степени тяжести недоношенных детей. Достоверность классификации выборки пациентов по полученному разделяющему правилу составила 100%. Это позволяет использовать данное решающее правило для создания методики в целях прогнозирования степени тяжести состояния недоношенных новорожденных со сроком гестации менее 30 недель. Использование доступных анамнестических, а также лечебно-диагностических данных дает возможность имплементировать разработанное правило после его апробации в практическую деятельность неонатальных отделений интенсивной терапии различных уровней оказания перинатальной помощи, что явится дополнительным критерием для врача при оценке степени тяжести новорожденного пациента и дальнейшем выборе персонализированной тактики терапевтических манипуляций и медикаментозного лечения.

Благодарности. Работа выполнена при поддержке совместного российско-белорусского проекта РФФИ (грант № 20-51-00001) и БРФФИ (грант № M20P-008).

Список литературы

1. Ely D.M., Driscoll A.K. Infant mortality in the United States, 2017: data from the period linked birth/infant deathfile. Natl Vital Stat Rep. 2019. Vol. 68 (10). P. 1-15.
2. Harrison W., Goodman D. Epidemiologic Trends in Neonatal Intensive Care 2007-2012. JAMA Pediatr. 2015. Vol. 169 (9). P. 855-862.
3. Chaudhari S. Neonatal intensive care practices harmful to the developing brain. Indian Pediatr. 2011. Vol. 48 (6). P. 437-440.
4. Pursley D.M., Zupancic J.A.F. Using Neonatal Intensive Care Units More Wisely for At-Risk Newborns and Their Families. JAMA Netw Open. 2020. Vol. 3 (6). P. 205693.
5. Белкин А.А., Кузнецов Ю.В., Тырсин А.Н., Помазкин Ю.В. Программа для ЭВМ «Автоматизированная информационно-аналитическая экспертная система (АИАЭС) для оценки прогноза больных с острой церебральной недостаточностью» // Свидетельство о государственной регистрации программы для ЭВМ РФ № 2016615536. Правообладатель Общество с ограниченной ответственностью «Клиника Института Мозга». Заявл. 07.04.2016; опубл. 20.06.2016, реестр программ для ЭВМ.
6. Тырсин А.Н., Шалькевич Л.В., Остроушко Д.В., Шалькевич О.В., Геворгян Г.Г. Исследование перинатального поражения центральной нервной системы у детей в

неонатальном периоде методами многомерного статистического анализа // Системный анализ и управление в биомедицинских системах. 2017. Т. 16. № 3. С. 595-605.

7. Sood B.J., McLaughlin K., Cortez J. Near-infrared spectroscopy. Applications in neonates. *Seminars Fetal Neonatal Med.* 2015. Vol. 20. P.164-172.

8. Kusaka T., Isobe K., Yasuda S. Evaluation of cerebral circulation and oxygen metabolism in infants using near-infrared light. *Brain Dev.* 2014. Vol. 36 (4). P. 277-283.

9. Мандель И.Д. Кластерный анализ. М.: Финансы и статистика. 1988. 176 с.

10. Hardle W.K., Simar L. *Applied Multivariate Statistical Analysis.* – 5th.ed. Springer. 2019. 558 p.

11. Hosmer D.W., Lemeshow S., Sturdivant R.X. *Applied Logistic Regression.* – 3d ed. Wiley. 2013. 397 p.

12. Васильев Н.П., Егоров А.А. Опыт расчета параметров логистической регрессии методом Ньютона–Рафсона для оценки зимостойкости растений // Математическая биология и биоинформатика. 2011. Том 6. № 2. С. 190-199. [Электронный ресурс]. URL: http://www.matbio.org/article_pdf.php?id=82 (дата обращения: 25.10.2020).

13. Тырсин А.Н. Энтропийное моделирование многомерных стохастических систем. Воронеж: Научная книга. 2016. 156 с.

14. Тырсин А.Н., Геворгян Г.Г. Программа вычисления энтропии случайного вектора // Свидетельство о государственной регистрации программы для ЭВМ РФ № 2017612851. Правообладатель Тырсин Александр Николаевич. Заявл. 31.10.2016; опубл. 03.03.2017, реестр программ для ЭВМ.