

Научная статья
УДК 519.218.28
<https://doi.org/10.24143/2072-9502-2024-3-95-106>
EDN FPWATQ

Аппроксимация статистических данных заболеваемости коронавирусной инфекцией с учетом расслоения по сопутствующим диагнозам

*Андрей Викторович Боровский^{1✉},
Андрей Леонидович Галкин², Светлана Сергеевна Козлова³*

^{1, 3}*Байкальский государственный университет,
Иркутск, Россия, andrei-borovskii@mail.ru*

²*Институт общей физики им. А. М. Прохорова Российской академии наук,
Москва, Россия*

Аннотация. Рассматривается расслоение по сопутствующим диагнозам статистики выздоровлений от Ковид-19 для города Иркутска за 2020–2021 гг. Предыдущее исследование было проведено без учета такого расслоения. Рассмотрены различные варианты аппроксимации реальной статистики гауссовой и лоренцевой функциями, гамма-распределением, кривыми Джонсона. Показано, что расслоение статистики выздоровлений улучшает аппроксимацию гауссовой и лоренцевой функциями по сравнению с интегральной статистикой, причем построение приближения на основе лоренцевой функции всегда лучше описывает реальную статистику. Оценки математического ожидания и дисперсии на основе статистических данных согласуются с оценками этих величин на основе гауссовой аппроксимации статистики методом наименьших квадратов, т. е. подходы являются эквивалентными. При этом расчеты критерия «Хи-квадрат» Пирсона отклоняют гипотезу о соответствии эмпирических данных предполагаемому теоретическому распределению. Поэтому нельзя говорить о нахождении функции распределения, а лишь об аппроксимации статистики некоторыми видами кривых. Подгонка эмпирических данных гауссовой и лоренцевой функциями осуществлялась по методу наименьших квадратов. В целом погрешность аппроксимации вследствие расслоения статистики по сопутствующим диагнозам уменьшается с 6 до 3 %.

Ключевые слова: коронавирусная инфекция, сопутствующие диагнозы, подгоночные кривые, аппроксимация, гауссова функция, лоренцева функция, гамма-распределение, кривая Джонсона

Для цитирования: Боровский А. В., Галкин А. Л., Козлова С. С. Аппроксимация статистических данных заболеваемости коронавирусной инфекцией с учетом расслоения по сопутствующим диагнозам // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2024. № 3. С. 95–106. <https://doi.org/10.24143/2072-9502-2024-3-95-106>. EDN FPWATQ.

Original article

Mathematical modeling of statistical data on the incidence of new coronavirus infection, taking into account the stratification by concomitant diagnoses

Andrey V. Borovsky^{1✉}, Andrey L. Galkin², Svetlana S. Kozlova³

^{1, 3}*Baikal State University,
Irkutsk, Russia, andrei-borovskii@mail.ru*

²*Prokhorov General Physics Institute of the Russian Academy of Sciences,
Moscow, Russia*

Abstract. The article considers the stratification of concomitant diagnoses of Covid-19 recovery statistics for the city of Irkutsk for 2020-2021. The previous study was conducted without taking into account such stratification. Various options for approximating real statistics by Gaussian and Lorentz functions, gamma distribution, and Johnson curves are considered. It is shown that the stratification of recovery statistics improves the approximation of Gaussian and Lorentz functions in comparison with integral statistics, and the construction of an approximation based on the Lorentz function always describes the real statistics better. Estimates of mathematical expectation and variance based on statistical data are consistent with estimates of these values based on the Gaussian approximation of statistics by the least squares method, i.e. the approaches are equivalent. At the same time, calculations of the Pearson Chi-squared criterion reject the hypothesis that empirical data correspond to the assumed theoretical distribution. Therefore, we cannot talk about finding the distribution function, but only about approximating statistics by certain types of curves. The fitting of empirical data by Gaussian and Lorentz functions was carried out using the least squares method. In general, the approximation error due to the stratification of statistics on concomitant diagnoses decreases from 6% to 3%.

Keywords: coronavirus infection, concomitant diagnoses, fitting curves, approximation, Gaussian function, Lorentz function, gamma distribution, Johnson curve

For citation: Borovsky A. V., Galkin A. L., Kozlova S. S. Mathematical modeling of statistical data on the incidence of new coronavirus infection, taking into account the stratification by concomitant diagnoses. *Vestnik of Astrakhan State Technical University. Series: Management, computer science and informatics.* 2024;3:95-106. (In Russ.). <https://doi.org/10.24143/2072-9502-2024-3-95-106>. EDN FPWATQ.

Введение

В настоящее время продолжается исследование различных особенностей коронавирусной инфекции [1–9]. В работе рассмотрены статистические данные по выздоровлению пациентов в стационарных и амбулаторных условиях в г. Иркутске в 2020–2021 гг., предоставленные в обезличенном виде ГБУЗ «Иркутская областная клиническая больница» (ИОКБ). Эти данные уже использовались авторами в работе [2], где применялись методы математической статистики и математического моделирования. Была изучена функция распределения (плотность вероятности) выздоровления больных, представлена аппроксимация функции распределения гауссовой и лоренцевой функциями с тремя подгоночными параметрами, найденными методом наименьших квадратов. Оказалось, что лоренцева функция лучше аппроксимирует статистические данные, чем гауссова функция. Точность аппроксимации в максимуме распределения вероятностей получилась 6,9 и 5,1 % для гауссовой и лоренцевой аппроксимаций соответственно. Первоначально авторы предполагали, что статистика будет описываться нормальным распределением вероятностей, т. е. гауссовой функцией. Однако практические расчеты не согласуются с таким предположением. Расчет критерия Пирсона для наилучшей аппроксимации статистических данных не подтверждает гипотезу о нормальном распределении выздоровлений.

Знание функции распределения выздоровлений пациентов позволяет упростить решение различных статистических задач теории эпидемий, например, позволяет получить формулу для среднего времени лечения в стационарных и амбулаторных условиях, а также вычислить средние затраты на лечение одного больного.

Ниже покажем, что основной причиной отклонения выписки выздоровевших пациентов от нормального распределения является наличие расслоения больных на группы с различными сопутствующими диагнозами. Учет расслоения улучшает аппроксимацию статистических данных до 3 %.

Построение сводных таблиц и графиков

В процессе работы в ГБУЗ ИОКБ Н. Е. Васильевой (врач-хирург) и С. С. Козловой (ИТ-специалист) была создана условная группировка сопутствующих диагнозов и осложнений, которую больница использовала при проведении внутренней экспертизы летальных случаев. Этот механизм обработки данных был введен в медицинскую информационную систему «Ариадна», в которой работает ГБУЗ ИОКБ на данный момент. Согласно этой классификации все сопутствующие диагнозы можно условно разбить на группы заболеваний в зависимости от поражения конкретной системы органов (табл.).

Укрупненная группировка по сопутствующим заболеваниям

An enlarged grouping of concomitant diseases

Сопутствующий диагноз	Медицинский классификатор	Выписка (число пациентов)	Всего койко-дней
Заболевания системы кровообращения	I00-I99	25 955	373 410
Злокачественные новообразования	C00-C48	696	10 595
Заболевания органов желудочно-кишечного тракта	K00-K93	6 133	89 932

Боровский А. В., Галкин А. Л., Козлова С. С. Аппроксимация статистических данных заболеваемости коронавирусной инфекцией с учетом расслоения по сопутствующим диагнозам

Окончание табл.
 Ending of the table

Сопутствующий диагноз	Медицинский классификатор	Выписка (число пациентов)	Всего койко-дней
Травмы, отравления	S00-T98	200	3 003
Инфекционные болезни	A00-B99	55 505	748 650
Заболевания органов дыхания	J00-J99	60 285	812 650
Заболевания мочевыделительной системы	N00-N99	4 558	67 346
Болезни обмена веществ	E00-E90	14 236	203 520
Гнойные заболевания мягких тканей	L00-99	94	1 327
Заболевания костно-мышечной системы и соединительной ткани	M00-M99	47	631
Болезни крови, нарушения свертываемости, иммунодефициты	D50-89	3 194	46 405
Заболевания нервной системы	G00-99	7	79
Неблагоприятные эффекты	W00-X59 Y40-84	15	147
Заболевания лор-органов	H65-H75	43	407
Алкогольная болезнь	XAII	8	91

Для анализа каждая группировка рассмотрена отдельно, независимо от того, были ли пересечения. На основе данных, приведенных в таблице, строим график среднего времени выздоровления

по каждой группе (рис. 1: ломаная линия – экспериментальные данные, гладкая сплошная – их аппроксимация параболой, выражение для которой приведено на поле графика).

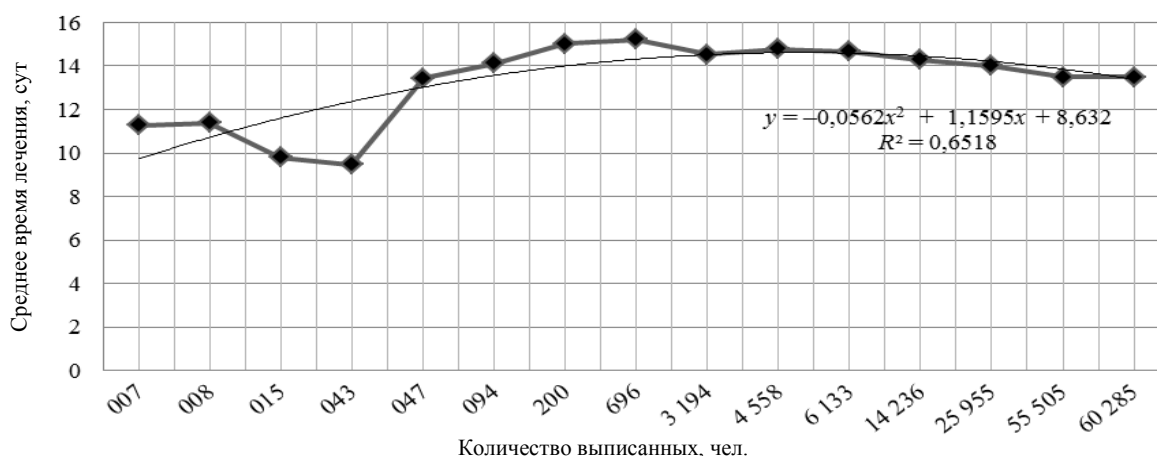


Рис. 1. Среднее время выздоровления по сопутствующим группам заболеваний

Fig. 1. Average recovery time for concomitant groups of diseases

В зависимости от сопутствующего диагноза среднее время лечения отмечается в диапазоне 9–15 дней.

Интегральная статистика выздоровлений

На рис. 2 представлена статистика выздоровлений пациентов по г. Иркутску за 2020–2021 гг. в зависимости от времени лечения.

На рис. 2, а приведен график выздоровления пациентов при амбулаторном лечении, на рис. 2, б – при госпитализации.

Локальные максимумы приходятся на 7, 14, 21, 28, 35 дни болезни и связаны с посещением врача в поликлинике амбулаторными больными 1 раз в неделю. Это один из факторов, заметно ухудшающих статистику амбулаторного лечения.

Bogovskiy A. V., Galkin A. L., Kozlova S. S. Mathematical modeling of statistical data on the incidence of new coronavirus infection, taking into account the stratification by concomitant diagnoses

Боровский А. В., Галкин А. Л., Козлова С. С. Аппроксимация статистических данных заболеваемости коронавирусной инфекцией с учетом расслоения по сопутствующим диагнозам

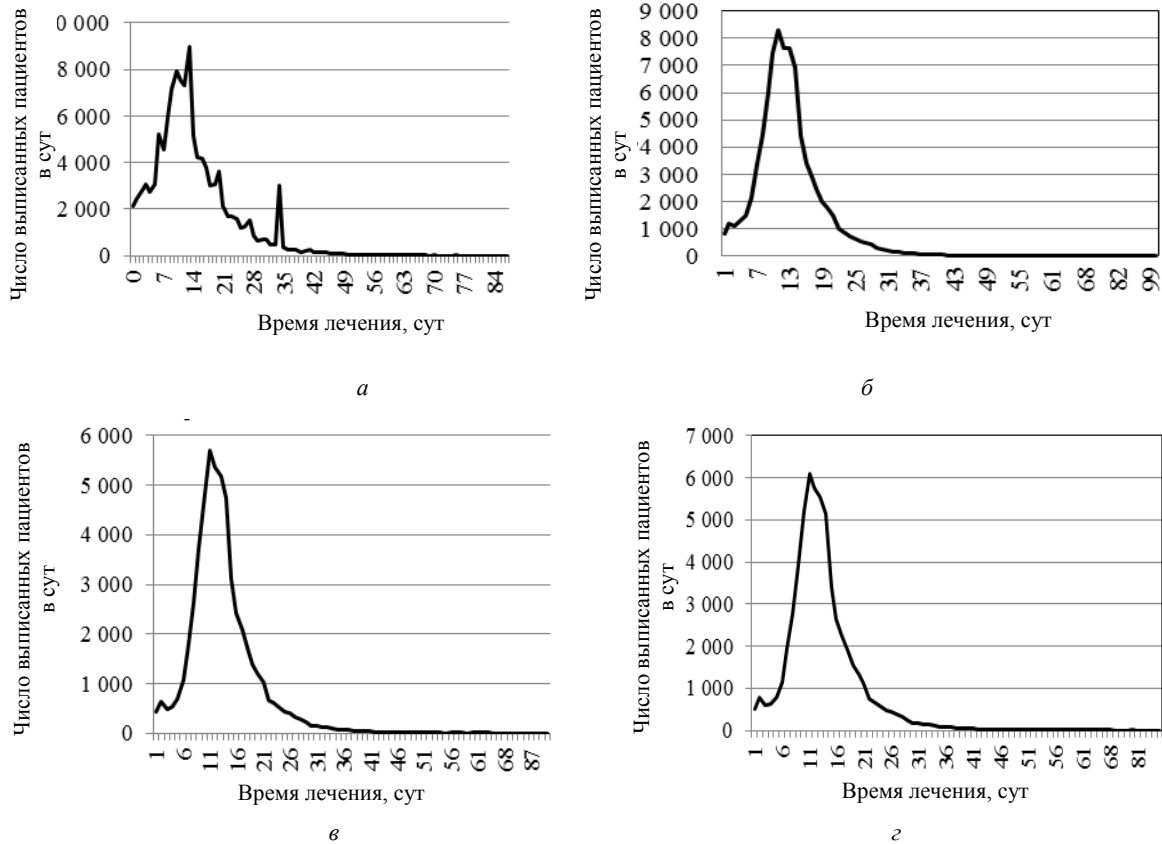


Рис. 2. Статистика выздоровлений по г. Иркутску за 2020–2021 гг. при амбулаторном лечении (а), при госпитализации интегральной (б), А00-В99 (в), J00-J99 (г)

Fig. 2. Statistics of recoveries in the city of Irkutsk for 2020-2021 for outpatient treatment (a), for integral hospitalization (б), A00-B99 (в), J00-J99 (г)

Поиск аппроксимации для функции распределения

Рассмотрим выздоровление при госпитализации. Данные для обработки соответствуют рис. 2, б:

$$\begin{pmatrix} t_i & 1, & 2, & \dots, & 182 \\ n_i & 2\,385, & 3\,065, & \dots, & 1 \end{pmatrix},$$

где t_i – нумерация суток, n_i – скорость выздоровления пациентов, т. е. количество пациентов, выздоровевших в течение суток t_i ,

$$n = \sum n_i = 98\,874,$$

где n – полное количество выздоровевших пациентов.

Математическое ожидание задачи, т. е. среднее время лечения одного пациента:

$$\bar{t} = \frac{\sum t_i n_i}{n} = 12,3835.$$

Дисперсия рассматриваемой выборки:

$$\sigma^2 = \frac{1}{(n-1)} \sum n_i (t_i - \bar{t})^2 = \frac{1}{(n-1)} \sum n_i t_i^2 - \frac{n\bar{t}^2}{(n-1)} = 46,4926;$$

$$\sigma = 6,81855.$$

Функция вероятности нормального распределения:

$$F_{N(c, \sigma^2)}(t) = \Phi\left(\frac{t-c}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{t-c}{\sigma}} e^{-\frac{x^2}{2}} dx.$$

Константы определяются ниже.

Отсюда плотность вероятности:

$$\rho(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-c}{\sigma}\right)^2}.$$

Максимум распределения достигается в точке $t = c$. Величина c определяет сдвиг функции распределения по оси времени вправо при условии, что ось времени направлена слева направо.

Рассмотрим аппроксимацию статистики гауссовой и лоренцевой функциями. Для гауссовой функции

$$f(t) = b_1 e^{-\left(\frac{t-c_1}{d_1}\right)^2}. \quad (1)$$

Согласование констант в (1) со статистическими моментами: $\bar{t} = c_1$, $\sigma\sqrt{2} = d_1$. Для лоренцевой функции

$$f(t) = \frac{b_2}{1 + \left(\frac{t-c_2}{d_2}\right)^2},$$

коэффициент b_2 соответствует максимальному значению функции, c_2 соответствует положению максимума функции на оси абсцисс, коэффициент d_2 равен полуширине распределения на полувысоте.

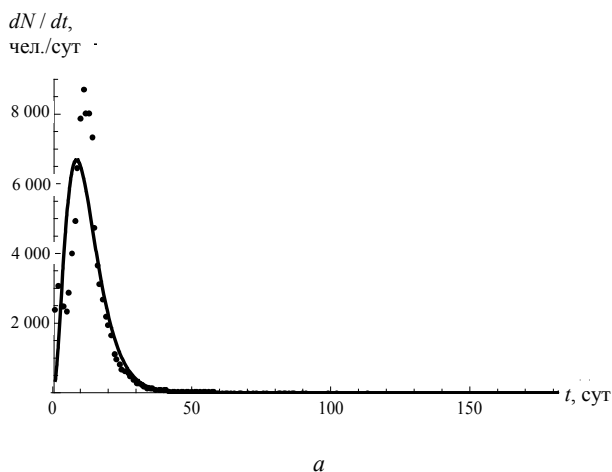
В случае малости величины

$$x(t) = \left(\frac{t-c_1}{d_1}\right)^2, \left(\frac{t-c_2}{d_2}\right)^2 \ll 1$$

обе функции дают одинаковый вид разложений

$$f(t) \cong b(1-x^2).$$

Это означает, что обе аппроксимации родственны, за исключением далекого хвоста распре-



деления. Гауссова аппроксимация спадает по экспоненте, а лоренцева как квадратичная гиперболоа.

Аппроксимация интегральной статистики выздоровлений рассматривалась в [2]. Применение метода наименьших квадратов привело к следующим значениям коэффициентов:

$$b_1 = 7618,61; \quad c_1 = 11,4359; \quad d_1 = 6,7879;$$

$$b_2 = 7726,68; \quad c_2 = 12,0259; \quad d_2 = 6,02823.$$

Гамма-распределение (непрерывное). Попробуем приблизить статистику гамма-распределением. Для него плотность вероятности

$$\rho(t) = \frac{t^{k-1} e^{-\frac{t}{\theta}}}{\theta^k \Gamma(k)}, \quad t \geq 0, \quad (2)$$

где $\Gamma(k)$ – гамма-функция.

Математическое ожидание в (2) $k\theta = \bar{t}$, дисперсия $k\theta^2 = \sigma^2$, поэтому

$$\theta = \frac{\sigma^2}{\bar{t}} = 3,754396; \quad k = \frac{(\bar{t})^2}{\sigma^2} = 3,298401.$$

На рис. 3, а представлены «экспериментальные данные» (точки) и плотность вероятности гамма-распределения с параметрами, выраженными через математическое ожидание и дисперсию статистики (сплошная кривая).

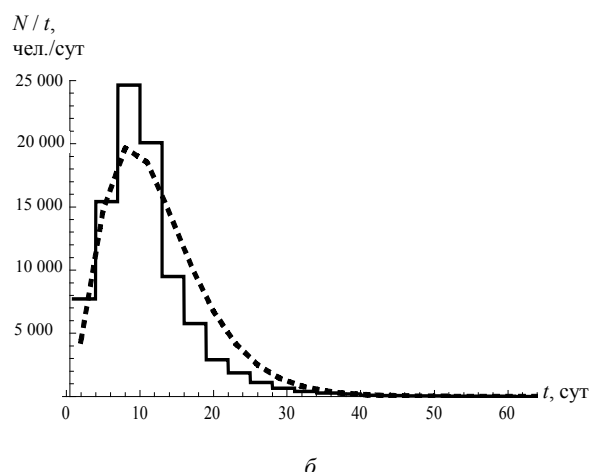


Рис. 3. Анализ гамма-распределения для амбулаторной статистики: «экспериментальные данные» и плотность вероятности гамма-распределения (а); группировка по 3 дня (б)

Fig. 3. Analysis of the gamma distribution for outpatient statistics: “experimental data” and probability density of the gamma distribution (a); grouping by 3 days (b)

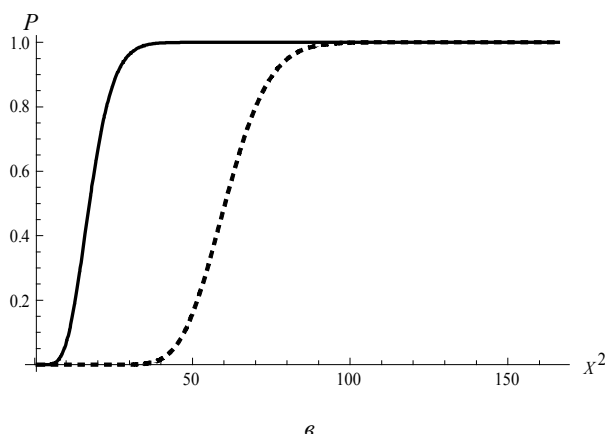


Рис. 3 (окончание). Анализ гамма-распределения для амбулаторной статистики: графики вероятности P для χ_{r-m-1}^2 на основе (3) (е)

Fig. 3 (ending). Analysis of the gamma distribution for outpatient statistics: probability graphs P for χ_{r-m-1}^2 based on (3) (e)

На рис. 3, б представлена группировка по 3 дня. Сплошной график гистограммы соответствует экспериментальным данным, пунктир – гамма-распределению. Улучшения укрупнение данных не дает. Видно, что выбор теоретического распределения не является удачным. Применим критерий Пирсона, для этого вычислим величину

$$X^2 = \sum_{i=1}^r \frac{n}{p_i} \left(\frac{n_i}{n} - p_i \right)^2,$$

где n_i – число событий из n с группировкой в r групп; p_i – теоретическая вероятность в группе. Проверка гипотезы по сложному критерию Пирсона, где 2 параметра оценены ($m = 2$) по выборкам объемом n , а плотность χ_{r-m-1}^2 с $r - m - 1$ степенями свободы имеет вид

$$P_{\chi_{r-m-1}^2}(x) = \begin{cases} \frac{x^{(r-m-1)/2-1} e^{-x/2}}{2^{(r-m-1)/2} \Gamma((r-m-1)/2)}, & x \geq 0; \end{cases} \quad (3)$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt,$$

где α – аргумент гамма-функции.

Рассмотрим 2 варианта в соответствии с рис. 2, а, б:

$$r - m - 1 = 65 - 2 - 1 = 62; \quad X^2 = 23\,732,1;$$

$$r - m - 1 = 22 - 2 - 1 = 19; \quad X^2 = 25\,33,2,$$

где X определяется по приведенной выше формуле.

Графики вероятности P для χ_{r-m-1}^2 на основе (3) (вместо таблиц) представлены на рис. 3, в: для первого варианта – сплошная линия, для второго варианта – пунктир. P -значения определяются по значениям X^2 на этих графиках, для обоих вариантов значение P приближается к 1, а для удовлетворения критерия надо попасть в левую ниспадающую часть графиков.

Отметим, что распределение Пуассона (дискретное) не подходит для рассматриваемой задачи, т. к. дисперсия для него равна математическому ожиданию, что не выполняется для статистики выздоровлений от Ковид-19.

Аппроксимация расслоенных данных

Данные по госпитальным выздоровлениям при сопутствующих диагнозах (инфекционные болезни А00-В99 и заболевания органов дыхания J00-J99) представлены на рис. 2, в, г. Имеем 3 варианта статистики:

$$1. \text{ Исходный суммарный вариант } \rightarrow n_1 = \sum_1^{182} N_1 = 98\,874.$$

$$2. \text{ A00-B99 } \rightarrow n_2 = \sum_1^{113} N_2 = 55\,505.$$

$$3. \text{ J00-J99 } \rightarrow n_3 = \sum_1^{91} N_3 = 60\,285.$$

Процедура расслоения может рассматриваться как фильтрация данных.

Аппроксимация данных второго и третьего вариантов представлена на рис. 4.

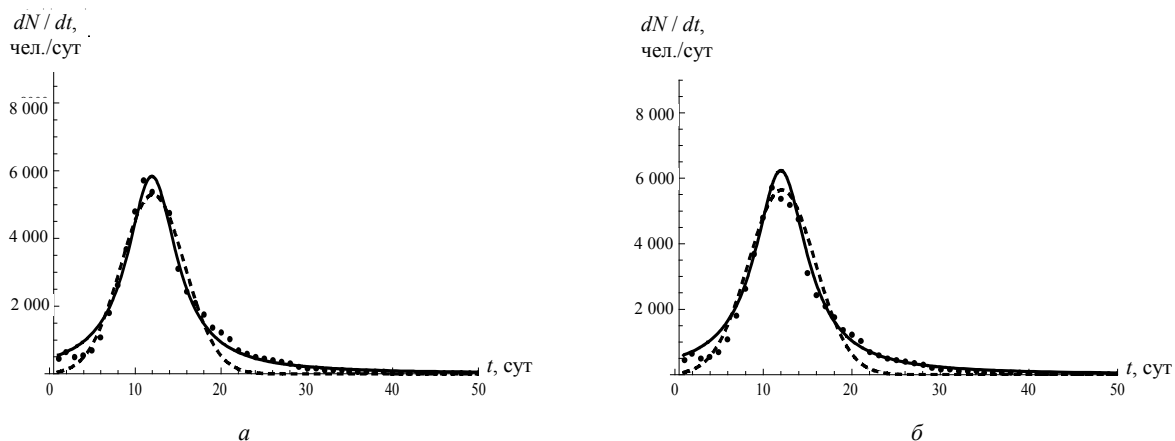


Рис. 4. Аппроксимация данных второго и третьего вариантов: точки – статистические данные, пунктир – гауссова функция; сплошная кривая – лоренцева функция:
 а – аппроксимация выздоровлений при сопутствующем диагнозе А00-В99; б – J00-Ј99

Fig. 4. Approximation of the data of the second and third variants: points – statistical data; dotted line is a Gaussian function; solid curve is a Lorentz function:
 а – approximation of recoveries with concomitant diagnosis А00-В99; б – J00-Ј99

Для рис. 4, а аппроксимация гауссовой функцией (пунктир) приводит к параметрам $b_1 = 5\,298,2$, $c_1 = 12,0542$, $d_1 = 5,25086$.

Аппроксимация лоренцевой функцией (сплошная кривая): $b_2 = 5\,835,56$, $c_2 = 11,9447$, $d_2 = 3,53038$. Расчет среднеквадратичного уклонения (число отсчетов $j = 74$) проводится по формуле

$$\rho = \sqrt{\frac{1}{j} \sum_{i=1}^j (f_{\text{эксп},i} - f_{\text{аппр},i})^2}. \quad (4)$$

Для аппроксимации гауссовой функцией $\rho = 257,952$, что по отношению к максимальному значению 5 712 составляет 0,0451596 (4,5 %). Наблюдаем уменьшение по сравнению с исходным вариантом без расслоения статистики, где было 0,0559 (5,6 %). Для аппроксимации лоренцевой функцией $\rho = 159,044$, что по отношению к максимальному значению 5 712 составляет 0,0278439 (2,7 %), – уменьшение по сравнению с основным вариантом, где было 0,0328 (3,3 %).

Для рис. 4, б аппроксимация гауссовой функцией (пунктир) приводит к следующим значениям параметров: $b_1 = 5\,647,93$, $c_1 = 12,0946$, $d_1 = 5,35026$. Аппроксимация лоренцевой функцией (сплошная кривая) дает: $b_2 = 6\,235,59$, $c_2 = 11,977$, $d_2 = 3,58749$. Расчет среднеквадратичного уклонения (число отсчетов $j = 75$) проводится по формуле (4).

$$\alpha_3 = \frac{1}{n\sigma^3} \sum_{i=1}^n (t_i - \bar{t})^3; \alpha_4 = \frac{1}{n\sigma^4} \sum_{i=1}^n (t_i - \bar{t})^4; \sigma^2 = \frac{1}{(n-1)} \sum_{i=1}^n (t_i - \bar{t})^2; \bar{t} = \frac{1}{n} \sum_{i=1}^n t_i.$$

В зависимости от соотношения выбирается класс семейства кривых Джонсона:

$$\alpha_4 > 3(1 + 0,641\alpha_3^2) \rightarrow S_U; \alpha_4 \approx 3(1 + 0,641\alpha_3^2) \rightarrow S_L; \alpha_4 < 3(1 + 0,641\alpha_3^2) \rightarrow S_B,$$

где S, L, B – обозначение классов.

Для аппроксимации гауссовой функцией $\rho = 282,875$, что по отношению к максимальному значению 5 712 составляет 0,049523 (4,95 %), – уменьшение по сравнению с основным вариантом, где было 0,0559 (5,6 %). Для аппроксимации лоренцевой функцией $\rho = 165,965$ или по отношению к максимальному значению 5 712 составляет 0,0290554 (2,9 %), – уменьшение по сравнению с основным вариантом, где было 0,0328 (3,3 %).

Таким образом, расслоение исходных данных по сопутствующим диагнозам улучшает аппроксимацию статистики выздоровлений как гауссовой, так и лоренцевой функциями, причем лоренцева функция лучше приближает статистические данные.

Поиск аппроксимации Джонсона

В статистической теории существует еще одна аппроксимация колоколообразной плотности вероятности, предложенная Джонсоном [10]. Процедура подбора кривой из системы Джонсона осуществляется на основе вычисления моментов высших порядков.

Наша выборка:

$$(1,1, \dots, 1,2,2, \dots, 2,3, \dots, 81,82,86,91) = \{t_i\}.$$

Для построения оценок удобно использовать упорядоченную выборку, где числа выстроены по возрастанию. Выражения для моментов:

1. Исходный суммарный вариант:

$$\bar{t} = 12,383; \sigma^2 = 46,4926; \alpha_3 = 1,73399; \alpha_4 = 14,2579; \alpha_4 - 3(1 + 0,641\alpha_3^2) = 5,47598 \rightarrow S_U.$$

2. A00-B99:

$$\bar{t} = 13,488; \sigma^2 = 39,6164; \alpha_3 = 1,94005; \alpha_4 = 12,5765; \alpha_4 - 3(1 + 0,641\alpha_3^2) = 2,33873 \rightarrow S_U.$$

3. J00-J99:

$$\bar{t} = 13,480; \sigma^2 = 39,8113; \alpha_3 = 1,80561; \alpha_4 = 11,1558; \alpha_4 - 3(1 + 0,641\alpha_3^2) = 1,88637 \rightarrow S_U.$$

Деление на классы условное [12], в данной работе используем семейство кривых Джонсона S_B , которое имеет вид

$$f(t) = \frac{\mu_1}{\sqrt{2\pi}} \frac{\mu_2}{(t-\varepsilon)(\mu_2-t+\varepsilon)} \exp \left\{ -\frac{1}{2} \left(\mu_3 + \mu_1 \ln \left\{ \frac{t-\varepsilon}{(\mu_2-t+\varepsilon)} \right\} \right)^2 \right\}, \quad \varepsilon \leq t \leq \varepsilon + \mu_2, \quad (5)$$

где μ и ε – подгоночные параметры Джонсона, которые определяются ниже.

Ввиду сложности (5) метод наименьших квадратов неприменим. В нашем случае $\varepsilon = 0$ (или даже 1),

константы предлагается определять с помощью квантилей [11].

Остановимся на варианте 2 (сплошная кривая на рис. 5, а).

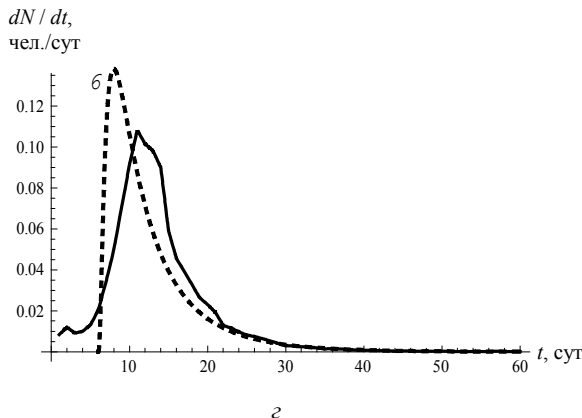
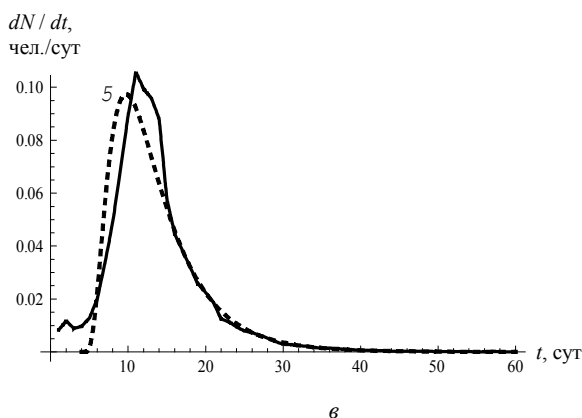
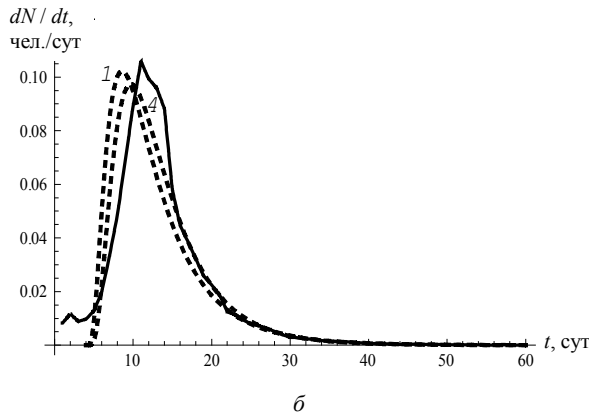
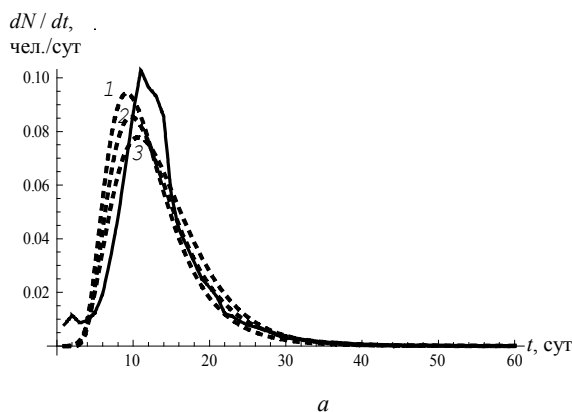


Рис. 5. Логико-эвристическая подгонка кривой Джонсона под экспериментальные данные: а-е – варианты кривых Джонсона 1-8 в сравнении друг с другом. Значения параметров для вариантов указаны в тексте статьи. Площади под всеми кривыми равны единице

Fig. 5. Logical-heuristic adjustment of the Johnson curve to experimental data: а-е – the variants of Johnson curves 1-8 are presented in comparison with each other. The parameter values for the variants are indicated in the text of the article. The areas under all curves are equal to one

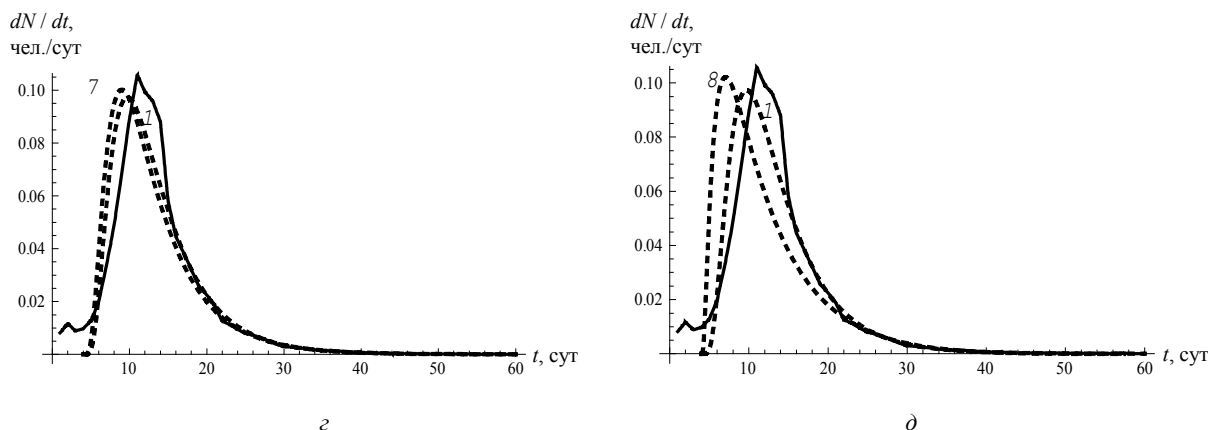


Рис. 5 (окончание). Логико-эвристическая подгонка кривой Джонсона под экспериментальные данные: *a-e* – варианты кривых Джонсона 1-8 в сравнении друг с другом. Значения параметров для вариантов указаны в тексте статьи. Площади под всеми кривыми равны единице

Fig. 5 (ending). Logical-heuristic adjustment of the Johnson curve to experimental data:

a-e – the variants of Johnson curves 1-8 are presented in comparison with each other.

The parameter values for the variants are indicated in the text of the article. The areas under all curves are equal to one

Будем считать, что известен только левый предел: $\varepsilon = 1$, тогда оценки параметров

$$\mu_1 = \frac{u_{\alpha_1} - u_{\alpha_2}}{\ln \left[\frac{(t_{\alpha_2} - \varepsilon)(\mu_2 - t_{\alpha_1} + \varepsilon)}{(t_{\alpha_1} - \alpha)(\mu_2 - t_{\alpha_2} + \varepsilon)} \right]}; \mu_3 = u_{\alpha_2} - \mu_1 \ln \left(\frac{(t_{\alpha_2} - \varepsilon)}{(\mu_2 - t_{\alpha_2} + \varepsilon)} \right);$$

$$\mu_2 = (\bar{t} - \varepsilon) \frac{(\bar{t} - \varepsilon)(t_{\alpha} - \varepsilon) + (\bar{t} - \varepsilon)(t_{1-\alpha} - \varepsilon) - 2(t_{\alpha} - \varepsilon)(t_{1-\alpha} - \varepsilon)}{(\bar{t} - \varepsilon)^2 - (t_{\alpha} - \varepsilon)(t_{1-\alpha} - \varepsilon)},$$

где u_{α} – квантиль стандартного распределения, рекомендуется $\alpha_1 = \alpha = 0,05$, $\alpha_2 = 1 - \alpha$; t_{α} – эмпирический квантиль $([\alpha(N_2 + 1)])$ -й упорядоченный по возрастанию член выборки):

$$\alpha = 0,05 \rightarrow u_{\alpha_1} = -u_{\alpha_2} = 1,645;$$

$$\alpha(55\,505 + 1) = 2\,775,3 \rightarrow 2\,775, t_{2\,775} = 6;$$

$$(1 - \alpha)(55\,505 + 1) = 52\,730,7 \rightarrow 52\,730, t_{52\,730} = 24.$$

В результате

$$2,1 \varepsilon = 1; \mu_1 = 2,11577; \mu_2 = 636,167; \mu_3 = 8,72904.$$

График кривой Джонсона S_B с этими значениями представлен на рис. 5, *a* (пунктир 1). Для сопоставления (пунктир 1) с экспериментальными данными их надо разделить на n_2 (сплошная кривая). Для улучшения можно фиксировать правый конец, сейчас он получился слишком большой. Добавляем вариант с $\varepsilon = 1$; $\mu_1 = 2,09088$; $\mu_2 = 3\,905,0$; $\mu_3 = 12,2788$ (пунктир 2) и вариант с фиксированным правым концом: $\varepsilon = 0$; $\mu_1 = 2,04182$; $\mu_2 = 113,0$; $\mu_3 = 4,21455$ (пунктир 3).

Анализ распределения рис. 5, *a* приводит к следующим выводам:

- правое крыло статистических данных может быть хорошо описано кривой Джонсона;
- на левой границе распределение резко выходит на 0 и не согласуется со статистическими данными;
- левая часть от максимума распределения и сам максимум сдвинуты влево относительно максимума статистических данных.

Формально этот сдвиг убирается, если в варианте 5, *a* изменить только ε , положив $\varepsilon = 4$ (пунктир 4, см. рис. 5, *b*). При отрицательных значениях ε распределение сдвигается влево.

На рис. 5, *b* представлен (пунктир 5) расчет с $\varepsilon = 4$ и оценкой параметров μ_1, μ_2, μ_3 по усеченной статистике:

$$n_2 = \sum_{\varepsilon}^{113} N_2 = 53\,953$$

в предположении неизвестного правого конца:

$$\alpha_1 = \alpha = 0,05; \alpha_2 = 1 - \alpha;$$

$$\alpha = 0,05 \rightarrow u_{\alpha_1} = -u_{\alpha_2} = -1,645$$

$$\alpha(53\,953 + 1) = 2\,697,7 \rightarrow 2\,697, t_{2\,697} = 7;$$

$$(1-\alpha)(53\,953 + 1) = 51\,256,3 \rightarrow 51\,256, t_{51\,256} = 25.$$

Медиана выборки:

$$\bar{t} = \frac{1}{2} \{t_{26\,976} + t_{26\,977}\} = \frac{1}{2} \{13 + 13\} = 13;$$

$$\mu_1 = 1,60484; \mu_2 = 185; \mu_3 = 4,9435.$$

$$\bar{t} = \frac{1}{2} \{t_{26\,369} + t_{26\,370}\} = \frac{1}{2} \{13 + 13\} = 13; \mu_1 = 1,03017; \mu_2 = 82,6; \mu_3 = 2,88962.$$

Таким образом (пунктир б), значение $\varepsilon = 6$ слишком велико. Методика определения коэффициентов допускает варьирование только коэффициентов α_1, α_2 (в нашем случае только α_1 для подстройки левой от максимума части статистики) плюс вариации $1 < \varepsilon < 4$.

Далее варьировем α_1 при базовом варианте $\varepsilon = 4, \alpha_1 = 0,05$.

$$\alpha_1(53\,953 + 1) = 4\,316,32 \rightarrow 4\,316; t_{4\,316} = 7; \alpha_2(53\,953 + 1) = 51\,256,3 \rightarrow 51\,256; t_{51\,256} = 25;$$

$$\mu_2 = 185; \mu_1 = 1,4878; \mu_3 = 4,70295.$$

Вариант $\varepsilon = 4, \alpha_1 = 0,02$ (пунктир в, см. рис. 5, в):

$$\alpha_1 = 0,02; \alpha_2 = 0,95; u_{\alpha_1} = -2,054; u_{\alpha_2} = 1,645;$$

$$\alpha_1(53\,953 + 1) = 1\,079,08 \rightarrow 1\,079; t_{1\,079} = 5; \alpha_2(53\,953 + 1) = 51\,256,3 \rightarrow 51\,256; t_{51\,256} = 25;$$

$$\mu_2 = 53,4; \mu_1 = 1,04928; \mu_3 = 2,10001.$$

Отклонение α_1 в обе стороны ухудшает положение кривой Джонсона слева от максимума, при этом правое крыло всегда ложится на статистику.

На рис. 5, г представлен (пунктир б) расчет с $\varepsilon = 6$ и оценкой параметров μ_1, μ_2, μ_3 по усеченной статистике:

$$n_2 = \sum_{\varepsilon}^{113} N_2 = 52\,739;$$

$$\alpha(52\,739 + 1) = 2\,637,0 \rightarrow 2\,637, t_{2\,637} = 7;$$

$$(1-\alpha)(52\,739 + 1) = 50\,103,0 \rightarrow 50\,103, t_{50\,103} = 25.$$

Рассмотрим вариант $\varepsilon = 4, \alpha_1 = 0,08$ (пунктир 7, см. рис. 5, д):

$$n_2 = \sum_{\varepsilon}^{113} N_2 = 53\,953;$$

$$\alpha_1 = 0,08; \alpha_2 = 0,95; u_{\alpha_1} = -1,40507; u_{\alpha_2} = 1,645$$

Приведем в заключение расчеты коэффициентов асимметрии для 8 кривых Джонсона и статистических данных:

$$M_1 = \int_{\varepsilon}^{\varepsilon+\mu_2} t f(t) dt; \sigma^2 = \int_{\varepsilon}^{\varepsilon+\mu_2} (t - M_1)^2 f(t) dt; M_3 = \int_{\varepsilon}^{\varepsilon+\mu_2} (t - M_1)^3 f(t) dt; A_s = \frac{M_3}{\sigma^3};$$

$$A_{s1} = 1,51634; A_{s2} = 1,63001; A_{s3} = 1,07073; A_{s4} = 2,07242;$$

$$A_{s5} = 1,79447; A_{s6} = 2,26922; A_{s7} = 1,96935; A_{s8} = 1,52575.$$

Коэффициенты асимметрии экспериментальных данных:

– исходный суммарный вариант: $A_s = 1,73399$;

– А00-В99: $A_s = 1,94005$

– J00-J99: $A_s = 1,80561$.

Кривые Джонсона асимметричны, что проявляется в наличии правого крыла и сдвиге максимумов кривых влево от максимума экспериментальных данных.

Заключение

Расслоение статистики выздоровлений по сопут-

ствующим диагнозам улучшает аппроксимацию эмпирических данных гауссовой и лоренцевой кривыми. Такое улучшение обусловлено тем, что расслоенные статистики имеют различные математические ожидания, т. е. времена выздоровления пациентов. Расчеты показали, что во всех случаях аппроксимация лоренцевой кривой дает лучшие результаты подгонки.

Показано, что статистика выздоровлений может быть аппроксимирована более сложной подгоночной кривой Джонсона (см., например, рис. 5, б, пунктир в). Строгое нахождение параметров под-

гонки кривой Джонсона под реальную статистику осложнено их количеством и сложными взаимосвязями между параметрами. По этой причине авторы применили логико-эвристический метод максимального правдоподобия, который позволил получить неплохие результаты.

Во всех вариантах применения кривой Джонсона

правая часть кривой хорошо ложится на реальную статистику. Кривая Джонсона плохо описывает реальную статистику вблизи нуля и во всех случаях дает отклонение подгоночной кривой влево. Максимальные значения совпадают в наилучших вариантах с точностью 5–6 %.

Список источников

1. Абдурахимов А. Х., Хегай Л. Н., Юсупова Ш. К. COVID-19 и его осложнения // *Re-health journal*. 2021. № 4 (12). URL: <https://cyberleninka.ru/article/n/covid-19-i-ego-oslozhneniya> (дата обращения: 01.02.2024).
2. Боровский А. В., Галкин А. Л., Козлова С. С. Анализ статистических данных амбулаторного лечения COVID-19 по городу Иркутску за 2020–2021 гг. // *System Analysis & Mathematical Modeling*. 2023. Т. 5. № 4. С. 494–504. DOI: 10.17150/2713-1734.2023.5(4).494-504.
3. Боровский А. В., Ильиных Н. Н., Козлова С. С. Математическая модель для скорости заражения SARS-COV-2 в неинфекционных больницах на примере города Иркутска // *System Analysis & Mathematical Modeling*. 2022. Т. 4. № 3. С. 187–199. DOI: 10.17150/2713-1734.2022.4(3).187-199.
4. Тамм М. В. Коронавирусная инфекция в Москве: прогнозы и сценарии // *Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология*. 2020. Т. 13. № 1. С. 43–51. DOI: 10.17749/2070-4909.2020.13.1.43-51.
5. Иванов М. В. Математическое моделирование процесса пандемии. Теория и практика // Институт развития стратегических инициатив. 2020. 30 апр. URL: <https://indsi.ru/2020/04/30> (дата обращения: 01.02.2024).
6. Головинский П. А. Математическое моделирование распространения вирусов с длинной инкубационной фазой в тесном мире // *Вестн. Воронеж. гос. ун-та. Сер.*

- Системный анализ и информационные технологии*. 2020. № 2. С. 5–14. DOI: 10.17308/sait.2020.2/2909.
7. Arenas A., Cota W., Gómez-Gardeñes J., Gomez S. Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions // *Physical Review X*. 2020. V. 10. Iss. 4. P. 041055.
8. Riyapan P., Shuaib S. E., Intarasit A. A Mathematical Model of COVID-19 Pandemic: A Case Study of Bangkok, Thailand // *Computational and Mathematical Methods in Medicine*. 2021. V. 9. P. 1–11. DOI: 10.1155/2021/6664483.
9. Xiao-Ping Li, Ye Wang, Muhammad Altaf Khan, Muhammad Y. Alshahrani, Taseer Muhammad. A Dynamical Study of SARS-COV-2: A Study of third Wave // *Results in Physics*. 2021. V. 29. P. 104705. DOI: 10.1016/j.rinp.2021.104705.
10. Johnson N. L. Systems of frequency curves generated by methods of translation // *Biometrika*. 1949. V. 36. N. 2. P. 149–176.
11. Кобзарь А. И. Прикладная математическая статистика. М.: Физматлит, 2006. 816 с.
12. Борбащ Н. М., Школина Т. В. Процедура подбора кривой из системы Джонсона методами процентилей и максимального правдоподобия – наименьших квадратов // *System Analysis & Mathematical Modeling*. 2023. Т. 5. № 4. С. 477–492. DOI: 10.17150/2713-1734.2023.5(4).476-493.

References

1. Abdurakhimov A. Kh., Khagai L. N., Iusupova Sh. K. COVID-19 i ego oslozhneniia [COVID-19 and its complications]. *Re-health journal*, 2021, no. 4 (12). Available at: <https://cyberleninka.ru/article/n/covid-19-i-ego-oslozhneniya> (accessed: 01.02.2024).
2. Borovskii A. V., Galkin A. L., Kozlova S. S. Analiz statisticheskikh dannykh ambulatornogo lecheniia COVID-19 po gorodu Irkutsku za 2020–2021 gg. [Analysis of statistical data on outpatient treatment of COVID-19 in the city of Irkutsk for 2020–2021]. *System Analysis & Mathematical Modeling*, 2023, vol. 5, no. 4, pp. 494–504. DOI: 10.17150/2713-1734.2023.5(4).494-504.
3. Borovskii A. V., Il'inykh N. N., Kozlova S. S. Matematicheskaia model' dlia skorosti zarazheniia SARS-COV-2 v neinfektsionnykh bol'nitsakh na primere goroda Irkutsk [A mathematical model for the rate of infection of SARS-COV-2 in non-infectious hospitals using the example of the city of Irkutsk]. *System Analysis & Mathematical Modeling*, 2022, vol. 4, no. 3, pp. 187–199. DOI: 10.17150/2713-1734.2022.4(3).187-199.
4. Tamm M. V. Koronavirusnaia infektsiia v Moskve: prognozy i stsenarii [Coronavirus infection in Moscow: forecasts and scenarios]. *Farmakoekonomika. Sovremennaiia farmakoekonomika i farmakoepidemiologia*, 2020, vol. 13, no. 1, pp. 43–51. DOI: 10.17749/2070-4909.2020.13.1.43-51.
5. Ivanov M. V. Matematicheskoe modelirovanie protsessa pandemii. Teoriia i praktika [Mathematical modeling of the pandemic process. Theory and practice]. *Institut razvitiia strategicheskikh initsiativ*, 2020, 30 apreliia. Available at: <https://indsi.ru/2020/04/30> (accessed: 01.02.2024).
6. Golovinskii P. A. Matematicheskoe modelirovanie rasprostraneniia virusov s dlinnoi inkubatsionnoi fazoi v tsnom mire [Mathematical modeling of the spread of viruses with a long incubation phase in a small world]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriia: Sistemnyi analiz i informatsionnye tekhnologii*, 2020, no. 2, pp. 5–14. DOI: 10.17308/sait.2020.2/2909.
7. Arenas A., Cota W., Gómez-Gardeñes J., Gomez S. Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions. *Physical Review X*, 2020, vol. 10, iss. 4, p. 041055.
8. Riyapan P., Shuaib S. E., Intarasit A. A Mathematical Model of COVID-19 Pandemic: A Case Study of Bangkok,

- casts and scenarios]. *Farmakoekonomika. Sovremennaiia farmakoekonomika i farmakoepidemiologia*, 2020, vol. 13, no. 1, pp. 43–51. DOI: 10.17749/2070-4909.2020.13.1.43-51.
5. Ivanov M. V. Matematicheskoe modelirovanie protsessa pandemii. Teoriia i praktika [Mathematical modeling of the pandemic process. Theory and practice]. *Institut razvitiia strategicheskikh initsiativ*, 2020, 30 apreliia. Available at: <https://indsi.ru/2020/04/30> (accessed: 01.02.2024).
6. Golovinskii P. A. Matematicheskoe modelirovanie rasprostraneniia virusov s dlinnoi inkubatsionnoi fazoi v tsnom mire [Mathematical modeling of the spread of viruses with a long incubation phase in a small world]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriia: Sistemnyi analiz i informatsionnye tekhnologii*, 2020, no. 2, pp. 5–14. DOI: 10.17308/sait.2020.2/2909.
7. Arenas A., Cota W., Gómez-Gardeñes J., Gomez S. Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions. *Physical Review X*, 2020, vol. 10, iss. 4, p. 041055.
8. Riyapan P., Shuaib S. E., Intarasit A. A Mathematical Model of COVID-19 Pandemic: A Case Study of Bangkok,

Thailand. *Computational and Mathematical Methods in Medicine*, 2021, vol. 9, pp. 1-11. DOI: 10.1155/2021/6664483.

9. Xiao-Ping Li, Ye Wang, Muhammad Altaf Khan, Muhammad Y. Alshahrani, Taseer Muhammad. A Dynamical Study of SARS-COV-2: A Study of third Wave. *Results in Physics*, 2021, vol. 29, p. 104705. DOI: 10.1016/j.rinp.2021.104705.

10. Johnson N. L. Systems of frequency curves generated by methods of translation. *Biometrika*, 1949, vol. 36, no. 1/2, pp. 149-176.

11. Kobzar' A. I. *Prikladnaia matematicheskaia statistika*

[Applied mathematical statistics]. Moscow, FIZMATLIT Publ., 2006. 816 p.

12. Borbats' N. M., Shkolina T. V. Protsedura podbora krivoi iz sistemy Dzhonsona metodami protsentilei i maksimal'nogo pravdopodobiiia – naimen'shikh kvadratov [The procedure for selecting a curve from the Johnson system using percentile and maximum likelihood methods – the name of the squares]. *System Analysis & Mathematical Modeling*, 2023, vol. 5, no. 4, pp. 477-492. DOI: 10.17150/2713-1734.2023.5(4).476-493.

Статья поступила в редакцию 16.05.2024; одобрена после рецензирования 06.06.2024; принята к публикации 19.07.2024
The article was submitted 16.05.2024; approved after reviewing 06.06.2024; accepted for publication 19.07.2024

Информация об авторах / Information about the authors

Андрей Викторович Боровский – доктор физико-математических наук; профессор кафедры математических методов и цифровых технологий; Байкальский государственный университет; andrei-borovskii@mail.ru

Andrey V. Borovsky – Doctor of Physico-Mathematical Sciences; Professor of the Department of Mathematical Methods and Digital Technologies; Baikal State University; andrei-borovskii@mail.ru

Андрей Леонидович Галкин – доктор физико-математических наук; старший научный сотрудник теоретической лаборатории, отдел ТИАМ; Институт общей физики им. А. М. Прохорова Российской академии наук; galkin@kapella.gpi.ru

Andrey L. Galkin – Doctor of Physico-Mathematical Sciences; Senior Researcher at the Theoretical Laboratory, EMAS Department; Prokhorov General Physics Institute of the Russian Academy of Sciences; galkin@kapella.gpi.ru

Светлана Сергеевна Козлова – аспирант кафедры математических методов и цифровых технологий; Байкальский государственный университет; svet201093@mail.ru

Svetlana S. Kozlova – Postgraduate Student of the Department of Mathematical Methods and Digital Technologies; Baikal State University; svet201093@mail.ru

