DOI 10.37882/2223-2966.2025.01.27

ЗАКОНОМЕРНОСТИ РАСПРЕДЕЛЕНИЯ ОНКОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ ПО РЕГИОНАМ РОССИЙСКОЙ ФЕДЕРАЦИИ

SIGNATURE VERIFICATION USING SIAMESE NEURAL NETWORKS

O. Mitina

Summary. Complete and reliable statistical information is a necessary basis for the management and organization of health care in all countries of the world. Information that has any medical significance is ultimately processed and analyzed using medical statistics. Statistical analysis is an integral part of clinical research. Statistics uses an extensive mathematical apparatus that helps to successfully analyze data and obtain the necessary information.
Nowadays, more and more people are being diagnosed with cancer. In 2021 alone, 283 thousand Russians died from malignant neoplasms. Now, there is a trend towards an increase in diseases associated with oncology in the world. The reasons are varied, but one thing is clear it is necessary to fight it, because cancer is one of the deadliest diseases.

Keywords: oncological diseases, integral law, Weibull distribution.

Митина Ольга Алексеевна
кандидат педагогических наук, МИРЭА — Российский технологический университет, г. Москва
alogmi@yandex.ru

Аннотация. Полная и достоверная статистическая информация является необходимым основанием для управления и организации здравоохранения во всех странах мира. Информация, имеющая какую-либо медицинскую значимость, в конечном счёте обрабатывают и анализируют с помощью медицинской статистики. Статистический анализ является интегральной частью клинических исследований. В статистике используется достаточно обширный математический аппарат, помогающий успешно анализировать данные, получать нужную информацию.
В наше время всё больше диагностируются онкологические заболевания. Только в 2021 году 283 тысячи россиян умерли от злокачественных новообразований. В данный момент в мире наблюдается тенденция к росту заболеваний, связанных с онкологией. Причины разнообразны, но понятно одно необходимо с этим бороться, ведь онкологические заболевания являются одними из самых смертельных.

Ключевые слова: онкологические заболевания, интегральный закон, распределение Вейбулла.

В современных реалиях проблемы онкологических заболеваний широко обсуждаются и остаются в центре внимания ввиду неуклонного роста заболеваемости, высокой инвалидизации и смертности пациентов. Подобные заболевания справедливо могут быть отнесены в группу социально-значимых болезней. [1] На первом году жизни с момента установления онкологического диагноза в РФ умирает каждый третий пациент. [2]

Заболевания, относящиеся к онкологическим, представляют собой обширный и разнородный класс заболеваний, включающий как доброкачественные, так и злокачественные новообразования. Злокачественные образования являются наиболее опасными.

Онкологические диагнозы зачастую воспринимаются как приговор, однако далеко не все из них приводят к смерти. Существуют доказанные методы по снижению смертности, такие как выявление рака на ранних стадиях, вовремя подобранное правильное лечение.

В России распространённость злокачественных новообразований ежегодно увеличивается на 1,5 % по данным за 2018 год, за последние 5 лет она выросла на 11,9 %. Сейчас смертность от онкологической заболеваемости находится на втором месте после заболеваний кровеносной системы. [3]

В 2021 году 933 тысячи россиян умерли от болезней сердца и сосудов, 283 тысячи — от злокачественных новообразований. Статистика же за 2020 год вызывает вопросы: официально от коронавируса в России умерло чуть больше 100 тысяч человек, но в сравнении с 2019 годом общая смертность выросла на 18 % — это больше на 324 тысячи смертей. Многие умершие от коронавируса имели различные хронические заболевания, в том числе онкологические, ковид же усугубил ситуацию. [4]

В России ситуация с онкологическими заболеваниями достаточно безрадостная, если сравнивать с другими странами. Российская Федерация находится на 49 месте в мире по заболеваемости онкологией: 234,4 диагноза на 100 тысяч жителей в 2020 году. По уровню смертности от рака — на 33 месте: 113,7 случая на 100 тысяч населения.

Данная статистика дана Всемирной организацией здравоохранения. Однако, к сожалению, она вызывает некоторые сомнения. Самые низкие показатели в этом рейтинге — у неразвитых государств Африки и Азии. Заболеваемость в Нигере составляет 78,4 на 100 тысяч жителей, в Гамбии — 79,5, в Непале — 80,9. Развитые же страны, наоборот, находятся на первых местах списка: Австралия — 452,4 диагноза на 100 тысяч, США — 362,2, Германия — 351,1, Франция — 313,7. [5]

В России тоже принимаются различные меры для улучшения ситуации с онкологическими заболеваниями. Так, согласно министру здравоохранения РФ М.А. Мурашко, введены новые требования к медицинским организациям, такие как лицензионные требования, требования к маршрутизации пациентов, утверждены новые требования по диспансерному наблюдению за пациентами с онкологическими заболеваниями, разработана и внедрена вертикально интегрированная информационная система, которая позволяет контролировать качество лечения и непосредственно прогнозировать последующие курсы химиотерапии и лучевой терапии. [6]

Сейчас в стране работает 86 онкологических диспансеров, ряд больниц, несколько профильных медучреждений находится на этапе строительства.

К сожалению, система здравоохранения РФ продолжает испытывать потребность в кадрах в сфере онкологии. На данный момент реализуется комплекс мер по обеспечению пациентов доступной медицинской помощью, в том числе в рамках развития кадрового потенциала.

На 2022 год количество онкологов возросло на 17 %, а также в связи с поставкой серьёзного радиологического оборудования, на 60 % увеличилось количество радиотерапевтов.

Несмотря на то, что в Российской Федерации ведётся работа по улучшению ситуации, связанной с онкологией, всё обстоит не так хорошо. Для улучшения этой ситуации важно проводить статистический анализ над данными по онкологической заболеваемости.

В статистике используются грубые и стандартизированные по возрасту показатели. Для характеристики закономерностей статистического распределения используют нормальное распределение, распределение Пуассона, распределение Вейбулла и многие другие. Соответствия теоретических и эмпирических частот получают при помощи критериев согласия.

При аппроксимации законов распределения, получаемых в результате обработки экспериментальных данных, иногда используется закон распределения Вейбулла. [7] Он назван по фамилии шведского инженера В. Вейбулла. Он впервые применил степенно-показательную функцию для аппроксимации экспериментальных данных о прочности стали на разрыв при усталостных испытаниях в 1939 году.

Вейбулл ввёл подобное распределение в практику анализа результатов усталостных испытаний. Впоследствии данное распределение использовалось для описания времён жизни разнообразных технических устройств, финансовых задач из страховой сферы, а также для описания социальных конфликтов.

Функция распределения трёхпараметрического распределения Вейбулла имеет вид:

$$F(x) = 1 - e^{-\left(\frac{x-\theta}{b}\right)^{c}},$$

где θ — параметр положения (минимальное значение случайной величины x);

b — параметр масштаба;

c — параметр формы. [8]

Функция выживаемости — это характеристика случайной величины, которая привязывает некоторое множество событий (обычно обозначающих смерть или поломку наблюдаемой системы) ко времени. Данная функция показывает вероятность того, что система не сломается к определённому времени. Сама же случайная величина показывает время поломки системы или время смерти. [9]

Распределение случайной величины можно характеризовать также функцией выживания, вероятностью противоположного события, состоящего в том, что случайная величина примет значение большее x. Обозначим функцию выживания через $W(x)$, тогда

$$W(x) = 1 - F(x) = e^{-\left(\frac{x-\theta}{b}\right)^{c}}.$$

У закона Вейбулла есть особое свойство: все кривые Вейбулла, различающиеся только параметром формы (c), пересекаются в одной точке.

В случае совпадения значений двух функций распределения Вейбулла, различающихся только параметром формы, будут совпадать и функции выживания, хотя бы для одного значения аргумента x^*:

$$W_1(x^*) = W_2(x^*),$$

тогда $e^{-\left(\frac{x^*-\theta}{b}\right)^{c_1}} = e^{-\left(\frac{x^*-\theta}{b}\right)^{c_2}}, \left(\frac{x^*-\theta}{b}\right)^{c_1-c_2} = 1.$

При $c_1 - c_2$ степенная функция принимает единичное значение тогда и только тогда, когда $x^* = \theta + b$.

В результате все кривые Вейбулла, различающиеся единственным параметром формы, пересекаются в одной точке:

— с координатами $(\theta + b, e^{-1})$ для функции выживания W;

— с координатами $\left(\theta + b, 1 - e^{-1}\right)$ для функции распределения F, (Рисунок 1).

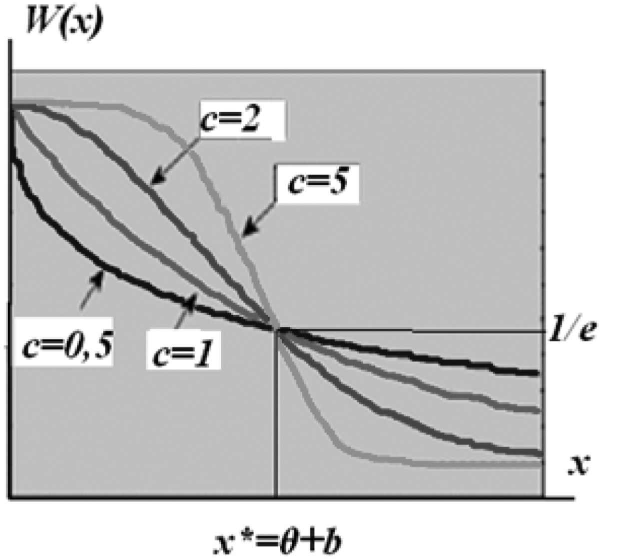

Рис. 1. Поведение функции выживания закона Вейбулла в зависимости от параметра формы c

В теории надёжности точка $e^* = \theta + b$ называется характеристическим временем жизни изделия. Оно определяет момент на временной шкале, до которого произойдет отказ 63,2 % изделий. В случае распределения онкологических заболеваний по регионам Российской Федерации — это характеристическая заболеваемость. Она определяет границу на шкале заболеваний, для которой вероятность события, состоящего в том, что число заболевших окажется менее граничного 0,632.

Для двухпараметрического распределения Вейбулла $\theta = 0$, поэтому абсцисса характеристической точки совпадает с параметром масштаба $x^* = b$.

Распределение Вейбулла спрямляется в координатах:

$$\left(\ln\left(x - \theta\right), \ln(-\ln(W\left(x\right)\right)).$$

Эти координаты называют координатами Вейбулла. Логарифмируя функцию выживания дважды, получим:

$$\ln\left(-\left(\ln\left(W\left(x\right)\right)\right)\right) = c \cdot \ln\left(\frac{x - \theta}{b}\right).$$

Ещё одним статистическим показателем анализа данных является функция риска (интенсивности). В общем случае функция риска определяется как величина относительного прироста функции выживания, взятая с противоположным знаком:

$$\lambda\left(x\right) = \frac{dF\left(x\right) / dx}{1 - F\left(x\right)} = -\frac{dW\left(x\right) / dx}{W\left(x\right)}.$$

Для закона Вейбулла, функция риска изменяется по степенному закону:

$$\lambda\left(x\right) = \frac{c\left(x - \theta\right)^{c-1}}{b^c}.$$

Поэтому спрямляющими координатами будут логарифмические координаты, $\left(\ln(x - \theta), \ln\lambda\right)$. [10]

При $c = 1$ распределение Вейбулла переходит в экспоненциальное распределение, функция риска равна константе. При $c < 1$ функция риска убывает, при $c > 1$ она возрастает (Рисунок 2).

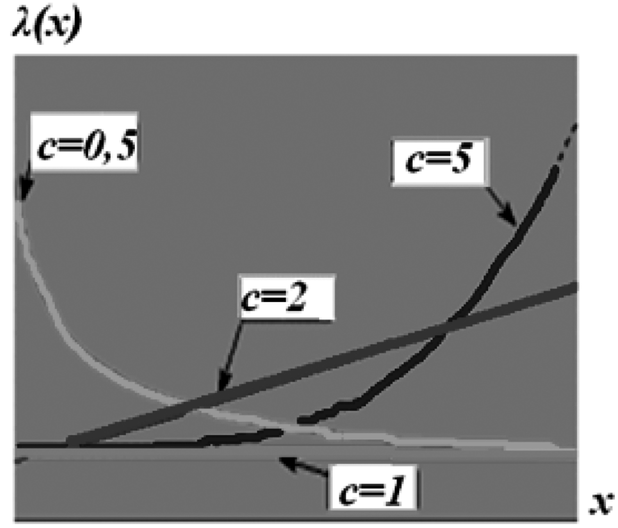

Рис. 2. Закон Вейбулла. Поведение функции риска при различных значениях параметра формы c

В результате, варьируя значения показателя c, становится возможным описать одним и тем же семейством функций разнообразные формы распределения вероятности.

Кривая Гомперца, названная в честь английского учёного Б. Гомперца, является сигмовидной функцией. Это тип математической модели для временных рядов, где рост медленнее в начале и в конце периода. Данная функция используется в актуарных расчётах для определения упрощённого закона о смертности. [11]

Кривую Вейбулла можно записать в форме Гомперца, если внести логарифмирование аргумента. Точка перегиба модифицированной таким образом кривой есть характеристическое значение аргумента кривой Вейбулла.

Введём логарифмический аргумент, $\ln\frac{x - \theta}{b} = -\tau$, тогда функция выживания, $W\left(\tau\right) = e^{-e^{-c\tau}}$. Полученная зависимость $W\left(\tau\right)$ есть кривая Гомперца, с точкой перегиба $\tau^* = 0$. В исходных переменных нулевому значению τ^* соответствует характеристическое значение распределения Вейбулла, $x^* = \theta + b$. [12], [13]

В практике распределение Вейбулла может применяться к широкому спектру данных.

Данные для анализа предоставлены московским научно-исследовательским онкологическим институтом имени П.А. Герцена. [14] Взяты стандартизованные данные по 85 регионам РФ, влияние возраста в этих данных исключено.

Упорядочим данные с Рисунка 3.1 по возрастанию, а затем построим график обратной функции распределения. [15] Сценарий модуля, выполняющего сортировку данных, а также построение графика на Рисунке 3.2 представлен в Приложении А. Значение оси ординат — это число заболевших на 100 000 человек, а значение по оси абсцисс — ранг. Полученный результат представлен на Рисунке 3.

Инверсия координатных осей на Рисунке 3 дает интегральный закон распределения числа онкологических заболеваний по регионам Российской Федерации.

Интегральный закон распределения, рассчитанный для стандартизованных данных, представлен на Рисунке 4. Накопленные частости получены путём нормировки рангов максимальным числом регионов.

График интегрального закона распределения допускает аппроксимацию кусочно-линейной зависимостью, ломаной. В пределах каждого линейного участка ломаной сохраняются параметры аппроксимирующей прямой, тогда как в граничных точках они изменяются скачком.

На Рисунке 4 выделяются четыре группы однородных по механизму формирования данных:
1. максимальная заболеваемость (более 300 заболевших на 100 000 человек): Сахалинская, Иркутская, Архангельская, Томская и Брянская области;
2. основная группа (диапазон от 242 до 300 заболевших на 100 000 человек): 55 регионов;
3. умеренная заболеваемость (диапазон от 197 до 242 случаев до 100 000 человек): 21 регион;
4. минимальная заболеваемость: Дагестан, Республика Алтай, Чукотская республика и Ингушетия.

Выделение однородных групп эмпирических данных позволяет построить классификацию регионов РФ по уровню заболеваемости:
— незначительная заболеваемость (до 197 заболеваний на 100 000 человек, частость до 0,035);
— заболеваемость среднего уровня (от 197 до 242 заболеваний на 100 000 человек, частость от 0,035 до 0,27);
— высокая заболеваемость (от 242 до 300 заболеваний на 100 000 человек, частость от 0,27 до 0,93);
— крайне высокая заболеваемость (выше 300 заболеваний на 100 000 человек, частость выше 0,93).

Закон распределения заболеваемости в каждой группе близок к равномерному, с постоянной плотностью.

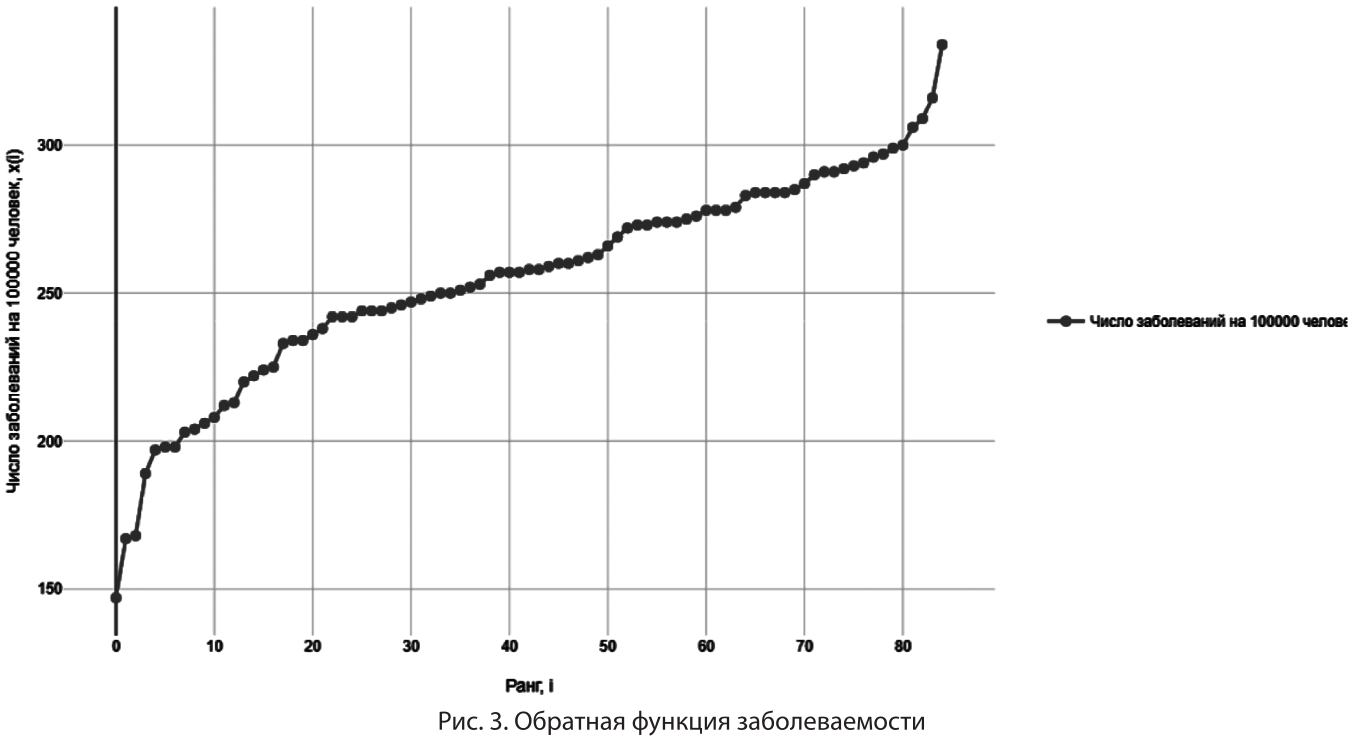

Рис. 3. Обратная функция заболеваемости

$t := 240..300 \qquad t1 := 180..240 \quad t2 := 145..200 \qquad t3 := 300..340$

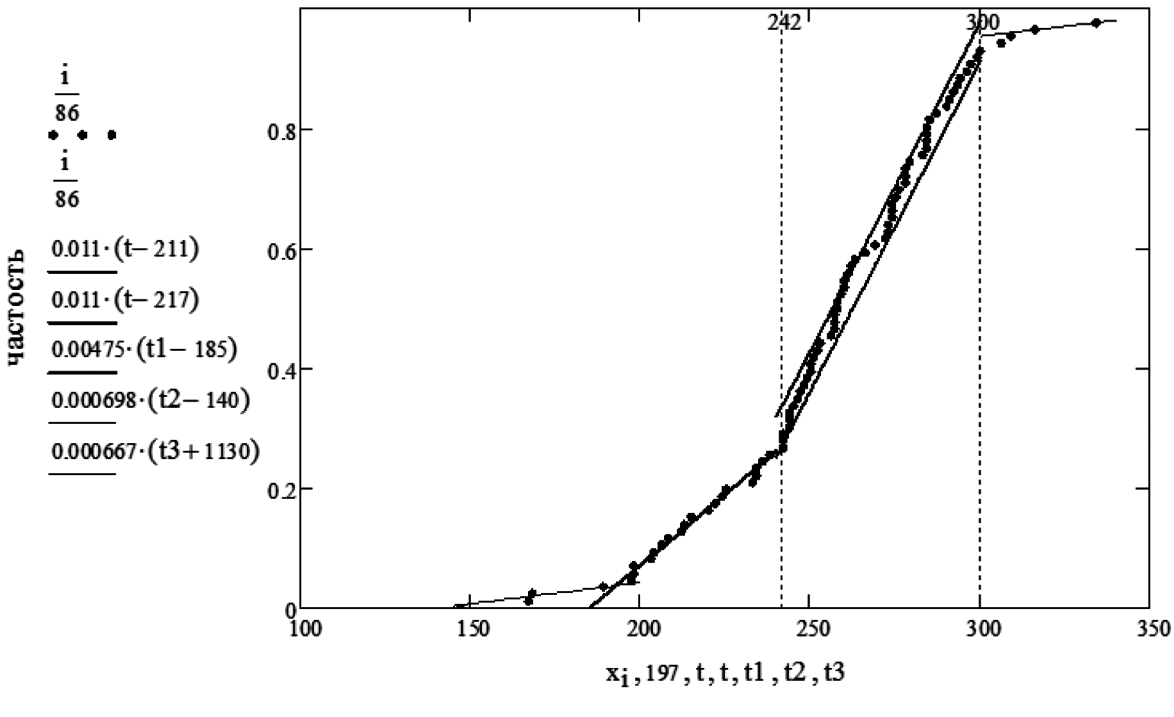

Рис. 4. Интегральный закон распределения заболеваемости в 2019 году

$$f(x) := \begin{cases} 0.0007 & \text{if } 147 \le x < 197 \\ 0.005 & \text{if } 197 \le x < 242 \\ 0.011 & \text{if } 242 \le x < 300 \\ 0.0007 & \text{if } 300 \le x < 335 \end{cases}$$

$x := 145..335$

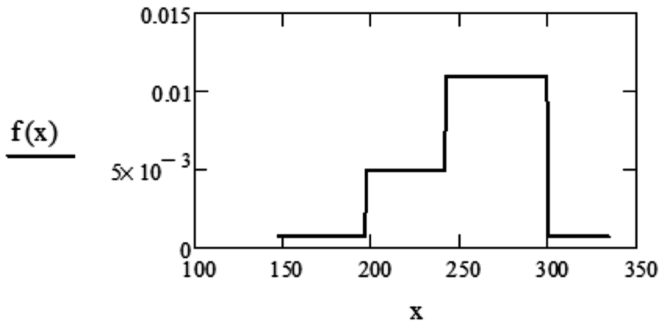

Рис. 5. Плотность равномерных распределений заболеваемости в диапазоне от 147 до 334 на 100 000 заболеваний

Идентификация угловых коэффициентов, аппроксимирующих прямых на Рисунке 4 позволяет определить изменение величин плотностей равномерных распределений с ростом числа заболеваний (Рисунок 5).

График Рисунка 5 имеет кусочно-постоянный, ступенчатый характер. В интервале от 147 до 300 заболеваний на 100 000 человек, величина плотности возрастает более чем на порядок, а затем резко сокращается. Такой характер изменения плотности распределения согласуется с интегральной функцией в том случае, когда последняя имеет точку перегиба.

Оценка положения точки перегиба на интегральной функции распределения приходится на интервал от 242 до 300 заболеваний на 100 000 человек, с арифметическим средним 271.

Подбор единой для всего диапазона заболеваемости унимодальной зависимости плотности распределения позволит также дать единое математической описание и для функции распределения.

Построим анаморфозы распределения Вейбулла по эмпирическим данным распределения онкологических заболеваний по регионам РФ в 2019 году. Полагая параметр положения $\theta = 0$, будем рассматривать двухпараметрическое распределение.

$$X := 5.1, 5.2 \,..\, 5.8 \qquad e^{5.6} = 270.426$$

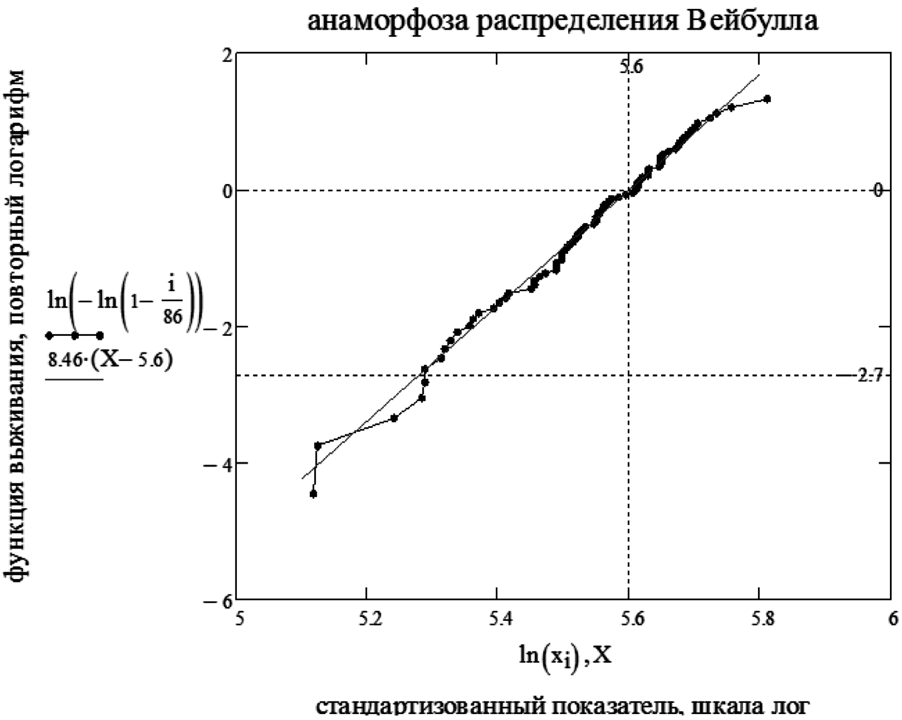

Рис. 6. Функция выживания в координатах Вейбулла, параметр формы $c \approx 8{,}46$, характеристическая заболеваемость $x^* \approx 270$ на 100 000 человек

$$y := 140 \,..\, 400$$

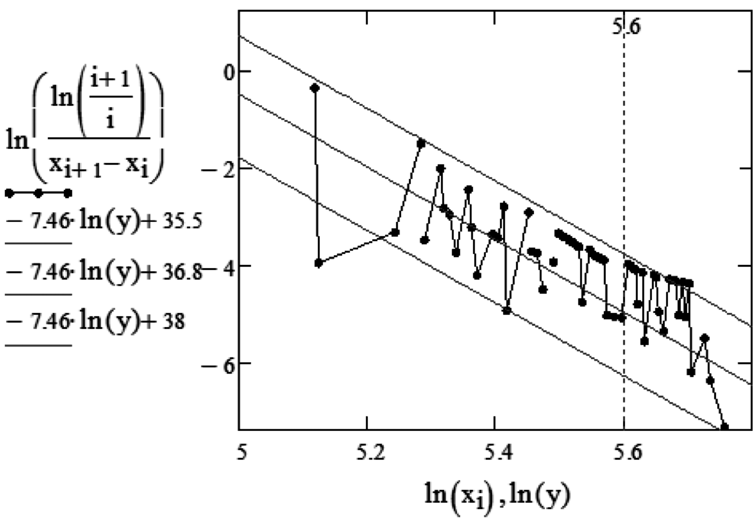

Рис. 7. Функция риска в логарифмической системе координат

На Рисунке 6 изображена функция выживания в координатах Вейбулла. Данный график демонстрирует хорошее спрямление почти во всём диапазоне измерения эмпирических данных, это подтверждает возможность применения распределения Вейбулла в качестве адекватного описания.

Графоаналитическое нахождение параметров распределения даёт значение $c \approx 8,46$ (угловой коэффициент аппроксимирующей прямой), а оценка характеристической заболеваемости $x^* \approx 270$ на 100 000 человек определяется положением точки пересечения аппроксимирующей прямой с осью абсцисс.

Построим график функции риска в логарифмических координатах (Рисунок 7). На нём вертикальной линией показана характеристическая заболеваемость $\ln\left(x^*\right) \approx 5,6$.

На Рисунке 7 наблюдается широкий линейный канал. Диапазон изменения заболеваемости укладывается в интервал, что говорит о степени полноты анализируемой эмпирической выборки.

Функция распределения заболеваемости по регионам РФ следует математической зависимости:

$$F\left(x\right) = 1 - e^{-\left(\frac{x}{270,4}\right)^{8,46}}.$$

Математическое ожидание:

$$M_x = \int_0^\infty e^{-\left(\frac{x-\theta}{b}\right)^c} dx = b \cdot \hat{}\left(\frac{1}{c} + 1\right) \approx 255.$$

Дисперсия:

$$D_x = \sigma^2 = b^2 \cdot \left(\Gamma\left(\frac{2}{c} + 1\right) - \Gamma^2\left(\frac{1}{c} + 1\right)\right) \approx 35,9^2.$$

Диапазон изменения заболеваемости укладывается в интервал $\left[M_x - 3\sigma, M_x + 2\sigma\right]$, что говорит о степени полноты анализируемой эмпирической выборки. [16]

Оценки математического ожидания и характеристической заболеваемости близки. Они близки также и к положению точки перегиба интегральной кривой, поэтому можно говорить о двух укрупнённых группах регионов — с заболеваемостью ниже и выше характеристической x^*.

Таким образом, приближение эмпирических данных кусочно-линейной зависимостью позволило получить однородные группы данных и построить на этой основе классификацию регионов по заболеваемости, в которой выделено четыре группы регионов. Показано, что единым математическим описанием для полного диапазона данных является закон Вейбулла. В качестве дальнейших исследований по данной теме можно предложить анализ динамики установленных параметров распределения, а также построение факторной модели.

ЛИТЕРАТУРА

1. Журавлев Е. Совершенствование диагностики злокачественных образований визуальных локализаций (на примере Челябинской области). [Текст]/ Журавлев Е. — Уфа: ГОУ ВПО «ЧелГМА Росздрава», 2011 — 28 с.
2. Чиссов В. Старинский В. Петрова Г. Состояние онкологической помощи населению в 2009 году. [Текст]/ В. Чиссов В. Старинский Г. Петрова М.: ФГУ «МНИОИ им. П.А. Герцена Росмедтехнологий», 2010. 188 с.
3. Каприн А., Старинский В., Петрова Г. Злокачественные новообразования в России в 2018 году (заболеваемость и смертность). [Текст]/ А. Каприн, В. Старинский, Г. Петрова М.: Филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2019. — 250 стр.
4. Ростовцев В.С. Искусственные нейронные сети / В.С. Ростовцев [Электронный ресурс] // Вятский Государственный Университет: — URL: http://iweb.vyatsu.ru/document/material/41/_Учебник%20ИНС_2014_Э4743.pdf (дата обращения: 18.10.2022) — 141 с.
5. Правительство России. Брифинг Татьяны Голиковой. [Электронный ресурс]. URL: http://government.ru/news/41486/ (дата обращения: 09.03.2023).
6. WHO. Estimated age-standardized incidence rates (World) in 2020, all cancers, both sexes, all ages. [Электронный ресурс]. URL: https://gco.iarc.fr/today/online-analysis (дата обращения: 12.03.2021).
7. ТАСС. Мурашко сообщил, что смертность от новообразований в России снизилась на 4,5 % за три года. [Электронный ресурс]. URL: https://tass.ru/obschestvo/15305967 (дата обращения: 10.03.2023).
8. Свешников А. Прикладные методы теории вероятностей: Учебник. [Текст]/ Под ред. О. Зайца. — СПб. Издательство «Лань», 2022. — 464 с.
9. ГОСТ 50779.27–2017. Статистические методы. Распределение Вейбулла. Анализ данных. [Текст] — М.: Стандартинформ, 2017. — 57 с.
10. MachineLearning.ru. [Электронный ресурс]. URL: http://www.machinelearning.ru/wiki/index.php?title=Функция_выживаемости (дата обращения: 01.04.2023).
11. Николаев Д. Антонов А. Чепурко В. Применение минимизации функции риска с учетом, усеченных слева и цензурированных справа данных для оценки параметров распределения Вейбулла. [Текст] — Надежность. №3, 2022. — 56 с.
12. Kenney J. Keeping E. Mathematics of Statistics, Pt. 1 [Текст] — Princeton N.J.: Van Nostrand, 1962. — p. 241.
13. Bowers N. Gerber H. Hickman J. Jones D. Nesbitt C. Acturial Mathematics. [Текст] — Itasca, I.L.: Society of Actuaries, 1997 — p. 71.
14. Gompertz B. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contigencies. [Текст] — London: Royal Society, 1832 — p. 585.
15. Каприн А. Старинский В. Шахзадова А. Злокачественные новообразования в России в 2019 году (заболеваемость и смертность). [Текст] — М.: МНИОИ им. П.А. Герцена — филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2020. — 252 с.
16. Шибинский В. Примеры и контрпримеры в курсе математического анализа. Учебное пособие. [Текст] — М.: Высшая школа, 2007. — 29 с.
17. Random. [Электронный ресурс]. URL: http://www.randomservices.org/random/special/Uniform.html (дата обращения: 12.04.2023).

Журнал «Современная наука: актуальные проблемы теории и практики»