

Опыт расчета параметров логистической регрессии методом Ньютона–Рафсона для оценки зимостойкости растений

Васильев Н.П.*, Егоров А.А.**

Санкт-Петербургская государственная лесотехническая академия им. С.М. Кирова,
Санкт-Петербург, 194021, Россия

Аннотация. При расчете параметров логистической регрессии, которая использовалась в качестве статистической модели для оценки зимостойкости растений, авторы столкнулись с проблемой неустойчивого счета. Как известно, задача построения логистической регрессии сводится к нелинейной системе уравнений, для решения которой использовался метод Ньютона–Рафсона. Для некоторых вариантов исходных данных отсутствует стабильная сходимость приближений. Рассматриваются возможные варианты решения этой проблемы. Окончательные результаты по ряду критериев сравниваются с результатами, полученными с помощью известной статистической программы SPSS.

Ключевые слова: статистическая обработка данных, логистическая регрессия, нелинейные системы уравнений, метод Ньютона–Рафсона, неустойчивость счета, зимостойкость растений.

ВВЕДЕНИЕ

В зимний период в условиях умеренной зоны растения испытывают влияние неблагоприятных факторов, и в первую очередь низких отрицательных температур. Устойчивость растений к низким отрицательным температурам называют морозостойкостью растений [1, 2]. Растения могут повреждаться в конце осени, когда начинают готовиться к переходу в состояние покоя и повреждаться в результате резких понижений температуры воздуха ниже 0°C, а также в начале весны, когда начинают выходить из состояния покоя, реагируя на температуры выше 0°C с последующим кратковременным понижением температуры ниже 0°C. В последнее время начали наблюдаться зимы с оттепелями, которые провоцируют растения к росту, что проявляется в набухании почек, а при последующем незначительном понижении температуры такие растения могут значительно повреждаться [3]. Устойчивость растений к целому комплексу неблагоприятных факторов конца осени – зимы – начала весны, когда растение еще находится в состоянии покоя, получило название зимостойкости [2].

При оценке морозостойкости и зимостойкости учитывают, какие органы растения повредились, степень и регулярность повреждения органов или растения в целом. Например, одна из первых шкал морозостойкости, разработанная Э.Л. Вольфом в 1917 г. [4], состоит из 5 основных баллов, где 1 балл означает, что растение не

* nikpv@mail.ru

** egorovfta@yandex.ru

повреждается морозами, а 5 балл, что растение погибает от морозов в первую или ближайшую зиму.

В двух крупнейших ботанических садах Санкт-Петербурга – Санкт-Петербургской государственной лесотехнической академии им. С.М. Кирова (ЛТА) и Ботанического института им. В.Л. Комарова РАН (БИН) – накоплены ряды наблюдений по зимостойкости древесных растений. В ЛТА эти ряды составлялись Н.Е. Булыгиным после зимы 1959–60 гг., и после его смерти продолжены А.А. Егоровым, а в БИН – Г.А. Фирсовым после зимы 1986–87 гг. Ряды наблюдений содержат информацию о годе наблюдения, названии вида и его зимостойкости. Зимостойкость приведена в семибальной шкале, предложенной для измерений Главным ботаническим садом в г. Москва [5] (характеристика шкалы приведена далее по тексту). Использование в ботанике при анализе зимостойкости статистических методов обычно ограничивается вычислением обычных статистик, иногда регрессии. Традиционные регрессионные модели, однако, в данном случае малоприспособлены, поскольку зависимая переменная (баллы) может принимать только ряд дискретных значений. С другой стороны, ряды наблюдений в БИН и ЛТА довольно объемны и содержат сведения о зимостойкости нескольких сотен древесных растений, поэтому их обработка должна носить массовый характер, что обычно удается с помощью собственных программных разработок.

Логистическая регрессия [6] является популярной статистической моделью для описания ситуаций, в которых зависимая переменная может принимать ряд дискретных значений, в простейшем случае, единицу или ноль, трактуемых как истина или ложь соответственно. Стандартный подход, который использует то или иное уравнение регрессии, как правило, линейное, в таких ситуациях является непригодным, поскольку не учитывает дискретность зависимой величины.

Логистическая регрессия в настоящее время является опцией практически любого серьезного программного продукта, посвященного статистическим расчетам, например, SPSS или STATGRAPHICS. Тем не менее, авторами была разработана программа для самостоятельного расчета параметров логит-регрессии, которая использовалась для оценки зимостойкости растений по значениям температур за зимний период. Причинами такого подхода являлись, прежде всего, возможность интеграции указанных расчетов в рамках разрабатываемого программного продукта – анализа данных метеостанций, с одной стороны, а с другой стороны, удобство интерфейса, как для ввода и выбора весьма объемных исходных данных, так и для вывода результатов в удобной форме.

Для проверки правильности расчетов, полученные результаты для некоторых выборок сравнивались с теми, которые дает программа SPSS. В большинстве случаев такого сопоставления получалось идеальное совпадение, однако для некоторых выборок наблюдался неустойчивый счет, и как следствие – расхождение результатов. Определение параметров логит-регрессии сводится к решению системы нелинейных уравнений. Существуют различные методы решения нелинейных систем, среди которых наиболее популярным и обеспечивающим наилучшую сходимость является метод Ньютона–Рафсона. Именно этот метод и был выбран. Опыт подобных расчетов излагается в настоящей статье.

ОСНОВЫ ТЕОРИИ

Логистическая регрессия, как уже было сказано, используется в ситуациях, когда отклик y является бинарным: единица – имеет место некоторое событие, ноль – если это событие не наступило. Множественный отклик, когда значений, а значит и исходов, может быть несколько, тоже возможен, но его оставим за рамками наших рассуждений. Нас будет интересовать вероятность появления события в зависимости от значений некоторых факторов (предикторов): (x_1, x_2, \dots, x_m) , а не сам отклик как таковой. Для

предсказания этой вероятности можно использовать, например, линейную регрессионную модель:

$$\text{Pr} = p_0 + p_1 x_1 + \dots + p_m x_m. \quad (1)$$

Здесь p_i – параметры модели. В принципе это возможно, но, очевидно, такой подход плох, поскольку не учитывает ограничений на значения вероятности Pr , которая в результате может выходить за рамки нуля и единицы.

Выходом из этого положения является использование специальной логит-функции, которая учитывает ограничения на значения вероятности. Модель связи вероятности с предикторами представляется в виде:

$$\text{Pr} = g(p_0 + p_1 x_1 + \dots + p_m x_m), \quad (2)$$

где $g(z)$ это так называемая логит-функция: $g(z) = 1/(1 + e^{-z})$. Логит-функция обладает тем замечательным свойством, что ее значения находятся в пределах от нуля до единицы при любых значениях аргумента, что, собственно, нам и нужно.

Предположим теперь, что у нас есть обучающая выборка, полученная в результате испытаний:

$$(x_1^i, x_2^i, \dots, x_m^i, y_i) \quad (i = 1, \dots, n). \quad (3)$$

Здесь y_i – значения отклика, n – количество испытаний. Требуется оценить параметры в уравнении (2), пользуясь полученной выборкой. Для оценки воспользуемся принципом максимума правдоподобия. Согласно этому принципу, за оценки параметров принимаются те значения, которые дают максимум для функции правдоподобия. Функция правдоподобия в нашем случае имеет следующий вид:

$$L(\bar{p}) = \prod_{i=1}^n g(\bar{p}\bar{x}^i)^{y_i} [1 - g(\bar{p}\bar{x}^i)]^{1-y_i}. \quad (4)$$

Здесь для краткости приняты следующие обозначения:

$$\begin{aligned} \bar{p} &= (p_0, p_1, \dots, p_m), \\ \bar{x}^i &= (1, x_1^i, \dots, x_m^i), \\ \bar{p}\bar{x}^i &= p_0 + p_1 x_1^i + \dots + p_m x_m^i. \end{aligned} \quad (5)$$

Обычно вместо функции (4) используется ее логарифм, что не меняет сути задачи, но позволяет избавиться от произведения:

$$l(\bar{p}) = \sum_{i=1}^n y_i \ln g(\bar{p}\bar{x}^i) + (1 - y_i) \ln(1 - g(\bar{p}\bar{x}^i)). \quad (6)$$

Очевидно, функция (6) имеет максимум. Для отыскания значений параметров, при которых достигается максимум функции (6), вычисляются частные производные по этим параметрам и приравниваются к нулю, в результате получается следующая система уравнений:

$$\begin{cases} f_0(\bar{p}) \equiv \sum_{i=1}^n g(\bar{p}\bar{x}^i) - \sum_{\{i: y_i=1\}} 1 = 0, \\ f_j(\bar{p}) \equiv \sum_{i=1}^n g(\bar{p}\bar{x}^i) x_j^i - \sum_{\{i: y_i=1\}} x_j^i = 0 \quad (j = 1, \dots, m). \end{cases} \quad (7)$$

Таким образом, задача свелась к решению системы уравнений (7) относительно неизвестных параметров \bar{p} . Решение этой системы представляет определенные сложности, поскольку она не является линейной.

ЧИСЛЕННОЕ РЕШЕНИЕ

Для решения воспользуемся методом Ньютона–Рафсона. Метод предполагает выбор некоторого начального приближения решения и последовательное его улучшение в ходе выполнения ряда вычислений, схему которых для нашего случая изложим здесь.

Для выбора начального приближения оценим вероятность появления единицы (успеха) по следующей формуле:

$$Pr_0 = \frac{\sum_{\{i: y_i=1\}} 1}{n}. \quad (8)$$

То есть количество удачных исходов поделили на общее количество испытаний. Наконец, начальное приближение для параметров модели получим, если не будем учитывать влияния факторов вообще, то есть:

$$\bar{p}^0 = (\ln \frac{Pr_0}{1 - Pr_0}, 0, \dots, 0). \quad (9)$$

В (9) использован тот факт, что для логит-функции выполняется следующее замечательное соотношение: $z \equiv \ln \frac{g(z)}{1 - g(z)}$.

Последовательные приближения будем вычислять согласно следующей формуле, которая связывает очередное $(q+1)$ - приближение с предыдущим q -приближением:

$$\bar{p}^{q+1} = \bar{p}^q - F(\bar{p}^q)J^{-1}(\bar{p}^q). \quad (10)$$

В соотношении (10) используется вектор $F(\bar{p}^q) = (f_0(\bar{p}^q), f_1(\bar{p}^q), \dots, f_m(\bar{p}^q))$, а также матрица Якоби для системы уравнений (7), которая в нашем случае имеет следующий вид:

$$J(\bar{p}) = \begin{pmatrix} \sum_{i=1}^n g(\bar{p}x^i)(1 - g(\bar{p}x^i)), \sum_{i=1}^n g(\bar{p}x^i)(1 - g(\bar{p}x^i))x_1^i, \dots, \sum_{i=1}^n g(\bar{p}x^i)(1 - g(\bar{p}x^i))x_m^i \\ \sum_{i=1}^n g(\bar{p}x^i)(1 - g(\bar{p}x^i))x_1^i, \sum_{i=1}^n g(\bar{p}x^i)(1 - g(\bar{p}x^i))x_1^i x_1^i, \dots, \sum_{i=1}^n g(\bar{p}x^i)(1 - g(\bar{p}x^i))x_m^i x_1^i \\ \dots \\ \sum_{i=1}^n g(\bar{p}x^i)(1 - g(\bar{p}x^i))x_m^i, \sum_{i=1}^n g(\bar{p}x^i)(1 - g(\bar{p}x^i))x_1^i x_m^i, \dots, \sum_{i=1}^n g(\bar{p}x^i)(1 - g(\bar{p}x^i))x_m^i x_m^i \end{pmatrix}. \quad (11)$$

Наконец, отметим, что вместо того чтобы вычислять обратную матрицу, которая фигурирует в соотношении (10), и которая требуется для вычисления величины $\Delta = F(\bar{p}^q)J^{-1}(\bar{p}^q)$, более эффективно методом исключений Гаусса решать линейную систему уравнений:

$$\Delta J(\bar{p}^q) = F(\bar{p}^q). \quad (12)$$

Для решения системы (12) использовался метод Гаусса с выбором главного элемента по всей матрице коэффициентов.

Вычисления прекращаются, когда на очередном шаге норма вектора Δ будет меньше некоторой заданной величины, характеризующей точность расчетов. Норма вектора Δ , очевидно, характеризует разницу между соседними приближениями.

Представленная схема вычислений была запрограммирована и результаты вычислений обсуждаются ниже.

РАСЧЕТ ВЕРОЯТНОСТИ ОБМЕРЗАНИЯ

В качестве примера рассмотрим зимостойкость пихты кавказской, или Нордмана (*Abies nordmanniana* (Stev.) Spach.), в ботаническом саду ЛТА. Ранее этот интродуцент с Кавказа значительно обмерзал (например, [4]).

Зимостойкость растений измерялась по 7-ми бальной шкале [5]:

- 1 – растения не обмерзают;
- 2 – у растения обмерзают не более 50 процентов однолетних побегов;
- 3 – у растения обмерзают от 50 до 100 процентов однолетних побегов;
- 4 – у растения обмерзают не только однолетние побеги, но и более старые;
- 5 – у растения обмерзает надземная часть до снегового покрова;
- 6 – у растения обмерзает вся надземная часть;
- 7 – растение вымерзает полностью.

Ранее нами было показано, что зимостойкость ряда хвойных и среднего модельного дерева показывает хорошие корреляции со среднемесячной температурой средних суточных за январь (T_{jm}) $r = -0,822$ в ботаническом саду БИН [7] и $r = -0,621$ в ботаническом саду ЛТА [8]. Поэтому дальнейший анализ проводится с учетом T_{jm} .

В таблице 1 приведена вероятность обмерзания пихты кавказской (т.е. появления баллов 2–7) с учетом зимостойкости по годам наблюдений, которые охарактеризованы T_{jm} .

Таблица 1. Вероятность обмерзания пихты кавказской (*Abies nordmanniana* (Stev.) Spach.) в ботаническом саду ЛТА за период с 1960 по 2000 гг.

Зима, год	T_{jm} , °C	Зимостойкость, балл	Вероятность обмерзания
1986/87	-17,94	3	0,98804
1965/66	-14,86	4	0,88524
1984/85	-13,68	2	0,75667
1962/63	-12,90	2	0,63040
1968/69	-12,23	1	0,50453
1971/72	-11,19	1	0,31375
1975/76	-11,15	1	0,30716
1981/82	-10,58	1	0,22231
1959/60	-10,01	1	0,15563
1979/80	-8,90	1	0,07271
1978/79	-8,85	2	0,07016
1998/99	-7,19	1	0,02059
1973/74	-7,11	1	0,01938
1977/78	-7,03	1	0,01824
1996/97	-5,60	1	0,00614
1989/90	-5,41	1	0,00531
1972/73	-5,33	1	0,00499
1964/65	-5,29	1	0,00484

Зима, год	T_{jm} , °C	Зимостойкость, балл	Вероятность обмерзания
1987/88	-4,70	1	0,00308
1999/00	-4,38	1	0,00241
1994/95	-4,13	1	0,00199
1963/64	-3,98	1	0,00177
1960/61	-3,44	1	0,00117
1961/62	-3,25	1	0,00101
1991/92	-2,72	1	0,00067
1970/71	-2,32	1	0,00049
1982/83	-2,25	1	0,00047
1974/75	-2,00	1	0,00039
1997/98	-1,71	1	0,00031
1988/89	-0,68	1	0,00014

Примечание: T_{jm} – среднемесячная температура средних суточных за январь.

Ранжирование данных по T_{jm} в табл. 1 наглядно показывает, что при температуре больше или равной $-12,23^{\circ}\text{C}$ растения в большинстве случаев не обмерзают (1 балл зимостойкости), и только в одном случае при $T_{jm} = -8,85^{\circ}\text{C}$ в 1978/79 г. наблюдается обмерзание не более 50% однолетних побегов (2 балл). Необходимо отметить, что редкие повреждения однолетних побегов у древесного растения не является серьезным повреждением и обычно такие растения восстанавливаются без последствий.

Принимая во внимание, что вероятность проявления события (обмерзания) при величине ниже 0,5 существенно падает, можно утверждать, что при T_{jm} равной или больше $-11,19^{\circ}\text{C}$ (см. табл. 1) пихта кавказская будет иметь высокую зимостойкость (балл 1).

АНАЛИЗ РЕЗУЛЬТАТОВ РАСЧЕТА

Представленная выше схема вычислений (гл. Численное решение) использовалась для оценки параметров логит-регрессии в качестве модели для прогнозирования влияния зимних температур на зимостойкость хвойных растений.

Чтобы воспользоваться бинарной моделью, баллы зимостойкости выше единицы объединялись в единый исход, при котором растения обмерзают в той или иной степени.

Для большинства вариантов исходных данных наблюдается устойчивый счет, который дает такие же результаты, как и пакет SPSS. Однако для некоторых ситуаций расчеты носят явно неустойчивый характер. Продемонстрируем это на примере использования следующих двух вариантов исходных данных, представленных в таблице 2.

Таблица 2. Пример исходных данных по лиственницам европейской (*Larix decidua* Mill.) и Кэмпфера (*Larix kaempferi* (Lamb.) Carriere) в ботаническом саду БИН за 1986–2009 гг.

Зима, год	T_{jm} , °C	Зимостойкость, балл	
		<i>Larix decidua</i>	<i>Larix kaempferi</i>
1986/87	-17,9	2	2
1987/88	-4,7	1	1
1988/89	-0,7	1	1
1989/90	-5,4	1	1
1990/91	-4,4	1	1
1991/92	-2,7	1	1

Зима, год	$T_{jm}, ^\circ\text{C}$	Зимостойкость, балл	
		<i>Larix decidua</i>	<i>Larix kaempferi</i>
1992/93	-2,3	1	1
1993/94	-3,3	1	1
1994/95	-4,1	1	1
1995/96	-7,0	1	1
1996/97	-5,6	1	1
1997/98	-1,7	1	1
1998/99	-7,2	1	1
1999/00	-4,4	1	1
2000/01	-2,5	1	1
2001/02	-4,4	1	1
2002/03	-9,5	1	1
2003/04	-8,2	1	1
2004/05	-1,6	1	1
2005/06	-5,9	1	1
2006/07	-2,5	1	1
2007/08	-1,8	1	1
2008/09	-3,6	2	1

Примечание: T_{jm} – среднемесячная температура средних суточных за январь.

Так, для *Larix decidua* (таблица 2), если предсказывать вероятность обмерзания, то есть считать балл 2 за истину (отклик равен единице), результаты расчета представлены в таблице 3.

Таблица 3. Результаты расчета параметров логит-регрессии для *Larix decidua* Mill.

Номер приближения	p_0	p_1	$f_0(\bar{p})$	$f_1(\bar{p})$
0	-2.351	0	-6.7e-16	11.8
1	-4.84	-0.514	1.31	-11.6
2	-3.53	-0.184	-0.135	6.22
3	-0.344	-0.31	0.0352	0.274
4	-4.504	-0.328	0.00129	0.0147
5	-4.512	-0.329	1.91e-6	4.41e-5
6	-4.512	-0.329	1.91e-6	4.41e-5

Примечание: p_0, p_1 – параметры модели (2); $f_0(\bar{p}), f_1(\bar{p})$ – значения левых частей уравнений (7), которые для искомым параметров p_0, p_1 должны обратиться в ноль.

Из таблицы видно, что уже шестая итерация дала желаемый результат. Пакет SPSS дает абсолютно такие же результаты. Кроме параметров логит-регрессии, в указанной таблице приводятся значения левых частей исходной системы уравнений (7), которая, собственно, и решается: эти значения позволяют оценить качество решения.

Однако самое интересное начинается для второго варианта исходных данных по *Larix kaempferi*. Несмотря на то, что исходные данные изменились всего лишь в одной строке, сходимость метода значительно ухудшилась, что демонстрирует таблица 4, в которой приведена лишь характерная часть результатов расчетов.

Таблица 4. Результаты расчета параметров логит-регрессии для *Larix kaempferi* (Lamb.) Carriere

Номер приближения	p_0	p_1	$f_0(\bar{p})$	$f_1(\bar{p})$
40	-120.753	-8.770	-1.67e-16	3.03e-15
41	-121.821	-8.873	5.04e-17	-4.79e-16
42	-122.821	-8.873	-2.03e-16	3.38e-15
43	-120.060	-8.762	1.02e-16	-9.71e-16
44	-121.060	-8.762	-1.84e-16	3.2e-15
45	-121.221	-8.814	-1.69e-16	3.05e-15
46	-122.192	-8.912	5.03e-17	-4.78e-16
47	-123.192	-8.912	-2.04e-16	3.38e-15
...
65	-121.420	-8.798	-1.85e-16	3.2e-15
66	-121.531	-8.848	5.29e-17	-5.03e-16

Примечание: сокращения см. в таблице 3.

Анализ данных таблицы 4 говорит о том, что явной сходимости параметров регрессии не наблюдается, не смотря на то, что качество решения исходной системы (7) довольно хорошее для всех итераций: $f_0(\bar{p})$ и $f_1(\bar{p})$ близки к нулю. Возникает вопрос выбора наилучшего приближения.

К ответу на этот вопрос можно подходить по-разному. Например, можно просто отбросить метод Ньютона–Рафсона, как несостоятельный для подобных ситуаций, или пытаться избавиться от неустойчивости счета. Однако, не смотря на то, что приближения плохо сходятся, они неплохо решают исходную систему уравнений (7).

Другой подход предполагает выбор наилучшего из приближений по некоторому критерию, отличному от оценки разности соседних приближений. В качестве такого критерия можно взять, например, показатель, характеризующий то, насколько хорошо полученное приближение удовлетворяет исходной системе (7), а именно, норму вектора левых частей этой системы, то есть:

$$\|F(\bar{p}^q)\| = \max(f_0(\bar{p}^q), f_1(\bar{p}^q), \dots, f_m(\bar{p}^q)). \quad (13)$$

С другой стороны, можно вспомнить, наконец, что решается задача определения максимума функции правдоподобия (4), и выбирать ту итерацию, для которой эта функция будет максимальной.

Результаты, которые дает каждый из указанных подходов, приведены в таблицах 5 и 6. Если использовать в качестве критерия выбора приближения минимум величины (13), то соответствующие результаты приведены в таблице 4.

Таблица 5. Приближение параметров логит-регрессии, наилучшим образом решающее систему (7)

Номер приближения	p_0	p_1	$f_0(\bar{p})$	$f_1(\bar{p})$	$L(\bar{p})$
38	-117.049	-8.576	1.32e-16	1.88e-16	0.9999999999999999445

Примечание: p_0, p_1 – искомые параметры модели (2); $f_0(\bar{p}), f_1(\bar{p})$ – значения левых частей уравнений (7); $L(\bar{p})$ – значение функции правдоподобия.

Если использовать в качестве критерия выбора приближения максимум функции правдоподобия, то соответствующие результаты приведены в таблице 5.

Таблица 6. Приближение, дающее максимальное значение функции правдоподобия $L(\bar{p})$

Номер приближения	p_0	p_1	$f_0(\bar{p})$	$f_1(\bar{p})$	$L(\bar{p})$
41	-121.821	-8.873	5.04e-17	-4.79e-16	1.00000000000000000000

Примечание: p_0, p_1 – искомые параметры модели (2); $f_0(\bar{p}), f_1(\bar{p})$ – значения левых частей уравнений (7); $L(\bar{p})$ – значение функции правдоподобия.

Если обратиться к программе SPSS, то соответствующие результаты приведены в таблице 7.

Таблица 7. Результаты расчета параметров логит-регрессии, которые дает пакет SPSS

p_0	p_1	$f_0(\bar{p})$	$f_1(\bar{p})$	$L(\bar{p})$
-57.084	-4.202	2.19e-8	-9.58e-8	0.99999995139628

Примечание: p_0, p_1 – искомые параметры модели (2); $f_0(\bar{p}), f_1(\bar{p})$ – значения левых частей уравнений (7); $L(\bar{p})$ – значение функции правдоподобия.

Представленные результаты в таблицах 5–7 говорят о том, что отдать явное предпочтение SPSS нельзя (таблица 7), если, конечно, использовать выбранные критерии.

Наконец, интересен следующий результат. Прежде чем его представить, заметим очевидный факт: если инвертировать в обучающей выборке отклик, то есть, заменить нули единицами и единицы нулями, то параметры логит-регрессии должны просто сменить знак. Для нашей выборки это означает, что теперь не двойка (2 балла зимостойкости) является истиной, а единица (1 балл). Оказалось, что в этом случае метод Ньютона–Рафсона хорошо сходится и соответствующие результаты приведены в таблице 8.

Таблица 8. Результаты расчета параметров логит-регрессии, которые получаются, если инвертировать отклик (заменить нули единицами и единицы нулями)

Номер приближения	p_0	p_1	$f_0(\bar{p})$	$f_1(\bar{p})$	$L(\bar{p})$
41	129.505	9.571	0	0	1.00000000000000000000

Примечание: p_0, p_1 – искомые параметры модели (2); $f_0(\bar{p}), f_1(\bar{p})$ – значения левых частей уравнений (7); $L(\bar{p})$ – значение функции правдоподобия.

С учетом смены знака эти результаты являются решением исходной задачи.

ВЫВОДЫ

Вычисление вероятности обмерзания с использованием ПЭВМ является необходимой процедурой, значительно снижающих трудоемкость обработки данных и приводящих к получению математически обоснованных результатов, а так же дает возможность построения модели прогнозирования – краткосрочного прогноза обмерзания по T_{jm} (среднемесячная температура средних суточных за январь), и восстановления выпавших (не измеренных) в прошлом данных.

Нелинейность задачи определения параметров логит-регрессии может привести к неустойчивому счету для некоторых исходных данных. Причиной этого, скорее всего, являются неустраняемые погрешности вычислений, избавиться от которых можно увеличив точность представления чисел.

Инвертирование отклика обучающей выборки в ряде случаев устраняет неустойчивость. Если учесть, что при этом меняется только знак рассчитываемых параметров логит-регрессии, то такой подход можно считать решением проблемы.

Для критических исходных данных, для которых наблюдаются описанные явления неустойчивого счета, пакет SPSS дает иные результаты, которые по критерию максимума правдоподобия не являются наилучшими.

СПИСОК ЛИТЕРАТУРЫ

1. Лир Х., Польстер Г., Фидлер Г.И. *Физиология древесных растений*. М.: Лесная пром-сть, 1974. 424 с.
2. Булыгин Н.Е., Ярмишко В.Т. *Дендрология*. СПб.: Наука, 2000. 528 с.
3. Фадеева И.В., Егоров А.А. Продолжительность зимнего покоя у *Tilia cordata* Mill. и *T. platyphyllos* Scop. в Санкт-Петербурге. В: *Фундаментальные и прикладные проблемы ботаники в начале XXI века: матер. всероссийской конф. в рамках XII съезда Русс. ботан. общ-ва*. Петрозаводск, 2008. Ч. 6. С. 350–353.
4. Вольф Э.Л. Наблюдения над морозостойкостью деревянистых растений. *Тр. бюро по прикл. ботан.* 1917. Т. 10. № 1. С. 1–146.
5. Лапин П.И., Сиднева С.В. Оценка перспективности интродукции древесных растений по данным визуальных наблюдений. В: *Опыт интродукции древесных растений*. М.: Гл. бот. Сад, 1973. С. 7–67.
6. Vapnik V.N. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 2000. 332 p.
7. *Реакция аборигенных и интродуцированных хвойных на изменение климата на Северо-Западе России. Научный отчет (заключительный)*. Рук. А.А. Егоров. СПб.: СПб ГЛТА, 2010. 102 с.
8. *Эколого-географический анализ коллекции голосеменных растений ботанического сада Санкт-Петербургской государственной лесотехнической академии. Научный отчет (заключительный)*. Рук. А.А. Егоров. СПб.: СПб ГЛТА, 2010. 66 с.

Материал поступил в редакцию 09.06.2011, опубликован 26.07.2011.