

ПОИСК ПРЕДСКАЗАТЕЛЬНЫХ ПРИЗНАКОВ ПРИ АНАЛИЗЕ ВЫЖИВАЕМОСТИ

Н. А. Бондарева

*Белорусский государственный университет, г. Минск;
n.bondareva1999@gmail.com;
науч. рук. – С. В. Гилевский, канд. техн. наук, доц.*

В работе рассмотрены методы поиска предсказательных признаков при анализе выживаемости. При помощи метода Каплана-Мейера проведен поиск предсказательных признаков для группы пациентов больных раком и был определен признак, оказавший наибольшее влияние на время жизни данной группы больных.

Ключевые слова: анализ выживаемости; таблица времен жизни; оценка Каплана-Мейера; оценка Нельсона-Аалена; цензурирование; функция выживаемости.

ВВЕДЕНИЕ

Анализ выживаемости – это общий термин для класса статистических методов, которые сосредоточены на вопросах, связанных с вероятностью наступления какого-либо события, и оцениванием времени до его наступления, с учетом того, что имеет место цензурение данных. Первоначально, анализ выживаемости появился в медицинских исследованиях, направленных на оценку продолжительности жизни при изучении эффективности методов лечения. Его появление было обусловлено необходимостью работать с так называемыми «неполными» или цензированными данными.

Анализ выживаемости применяется для работы с данными из совершенно разных научных и социальных сфер, таких как медицина, биология, маркетинг и реклама, менеджмент, государственное управление, страхование, электроника, техника.

Традиционно, события, время наступления которых нас интересует, называются терминальными событиями, также можно встретить термины «отказ» или «критическое событие». В медицине и биологии, терминальным событием мы считаем смерть наблюдаемого объекта.

МЕТОДЫ ОЦЕНКИ ФУНКЦИИ ВЫЖИВАЕМОСТИ

Функция выживаемости (в электронике и технике – функция надежности) для непрерывной случайной величины T с функцией распределения $F(t) = P\{T \leq t\}$ и плотностью распределения $f(t) = \frac{dF(t)}{dt}$ определяется следующим образом

$$S(t) = P\{T > t\} = 1 - F(t) = \int_t^{\infty} f(x)dx,$$

где t – некоторое время, в течение которого производилось наблюдение за совокупностью.

Функция $S(t)$ описывает вероятность того, что терминальное событие не наступит за все время наблюдения.

Важной функцией в анализе выживаемости является функция риска, или интенсивности отказов, которая аналитически задает вероятность того, что элемент, оставшийся в совокупности к началу некоторого временного интервала, покинет совокупность в течении этого интервала [2].

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)dt}{S(t)dt} = \frac{f(t)}{S(t)}$$

$$\lambda(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln S(t)$$

и, соответственно,

$$S(t) = \exp \left(- \int_0^t \lambda(x)dx \right)$$

Это соотношение легко объяснить тем, что чем больше риск умереть, тем меньше будет значение функции выживаемости.

Рассмотрены следующие непараметрические методы оценки функции выживаемости $S(t)$:

- таблица времен жизни;
- оценка Каплана-Майера;
- оценка Нельсона-Аалена.

Для построения таблицы времен жизни, необходимо всю область возможных времен жизни разбить на некоторое число интервалов n . Далее для каждого интервала рассчитывается число объектов, которые были живы к началу интервала r_j , число объектов, цензурированных или выбывших из интервала c_j , и число объектов, умерших в данном интервале d_j , $j = 1, 2, \dots, n$.

На основании этих величин вычисляются некоторые дополнительные статистики, такие, как число изучаемых объектов, доля умерших, доля выживших, функция выживания, плотность вероятности, медиана ожидаемого времени жизни.

При помощи метода Каплана-Майера можно провести оценку выживаемости двух или более, как правило не очень большого числа, групп [1].

Оценка Нельсона-Аалена – это оценка кумулятивной функции риска. Кривизна оценки Нельсона-Аалена дает представление о форме функции риска.

ПОИСК ПРЕДСКАЗАТЕЛЬНЫХ ПРИЗНАКОВ

Поиск предсказательных признаков проводился при помощи приведенных выше методов. Работа проводилась с базой данных Ovarian Cancer Survival Data 1979 года. В ней представлена информация о 26 пациентах больных раком яичников. Практическая часть была выполнена при помощи языка программирования R.

Пациенты были разбиты на несколько групп по какому-либо признаку, значимость которого необходимо было проверить. Например, вначале пациенты были разбиты на 2 группы (см. рис. 1). В каждой группе применялся свой метод лечения, и необходимо определить, какое лечение более эффективно.

rx=A							
time	n.risk	n.event	survival	std.err	lower	95% CI	upper
59	13	1	0.923	0.0739	0.789	1.000	
115	12	1	0.846	0.1001	0.671	1.000	
156	11	1	0.769	0.1169	0.571	1.000	
268	10	1	0.692	0.1280	0.482	0.995	
329	9	1	0.615	0.1349	0.400	0.946	
431	8	1	0.538	0.1383	0.326	0.891	
638	5	1	0.431	0.1467	0.221	0.840	

rx=B							
time	n.risk	n.event	survival	std.err	lower	95% CI	upper
353	13	1	0.923	0.0739	0.789	1.000	
365	12	1	0.846	0.1001	0.671	1.000	
464	9	1	0.752	0.1256	0.542	1.000	
475	8	1	0.658	0.1407	0.433	1.000	
563	7	1	0.564	0.1488	0.336	0.946	

Рис. 1. Таблицы времен жизни для двух групп с разными видами лечения

Далее выборка еще раз была разбита на две части как показано на рис. 2. За критерий разделения взято наличие остаточных болезней: у первой группы их не было, у второй были.

resid.ds=no							
time	n.risk	n.event	survival	std.err	lower	95% CI	upper
353	11	1	0.909	0.0867	0.754	1	
563	8	1	0.795	0.1306	0.577	1	
638	7	1	0.682	0.1536	0.438	1	

resid.ds=yes							
time	n.risk	n.event	survival	std.err	lower	95% CI	upper
59	15	1	0.933	0.0644	0.815	1.000	
115	14	1	0.867	0.0878	0.711	1.000	
156	13	1	0.800	0.1033	0.621	1.000	
268	12	1	0.733	0.1142	0.540	0.995	
329	11	1	0.667	0.1217	0.466	0.953	
365	10	1	0.600	0.1265	0.397	0.907	
431	8	1	0.525	0.1310	0.322	0.856	
464	7	1	0.450	0.1321	0.253	0.800	
475	6	1	0.375	0.1296	0.190	0.738	

Рис. 2. Таблицы времен жизни для групп с наличием остаточных болезней и без них

Построены кривые Каплана-Майера для каждой группы, приведенные на рис. 3 и рис. 4.

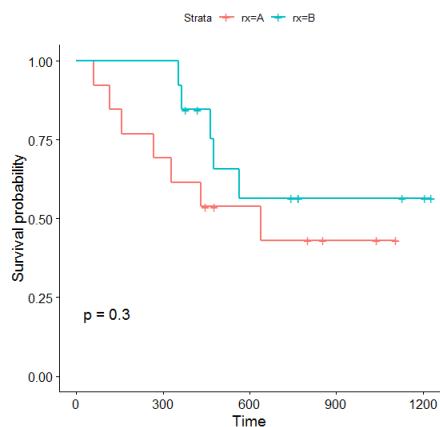


Рис. 3. Кривые Каплана-Мейера для двух групп с разными видами лечения (время жизни указано в днях)

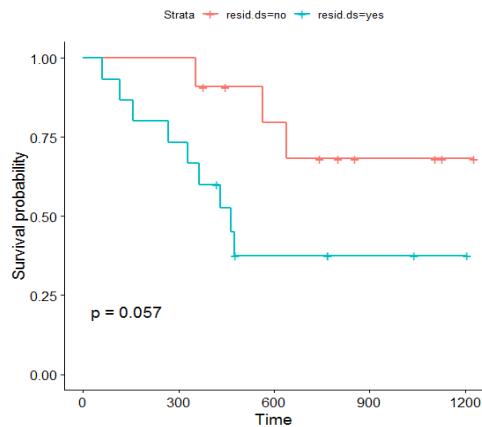


Рис. 4. Кривые Каплана-Мейера для групп с наличием остаточных болезней и без них (время жизни указано в днях)

Анализ графиков позволяет заключить, что метод лечения не может быть выбран как предсказательный признак продолжительности времени жизни, так как статистические различия в двух группах, разбитых по этому признаку, не слишком велики.

Однако разделение пациентов по группам, исходя из наличия остаточных болезней, дает значимое различие кривых рассмотренных выборок. Анализ кривых Каплана-Мейера позволяет сделать вывод, что у группы, где у пациентов не было остаточных болезней, продолжительность жизни была больше. Значит, наличие остаточных болезней мы можем считать значимым предсказательным признаком.

Библиографические ссылки

1. Mills M. Introducing survival and event history analysis. Groningen, 2011.
2. Singh R., Mukhopadhyay K. Survival analysis in clinical trials: Basics and must know areas // Perspect Clin Res. 2011. № 2 (4). P. 145-8.