

Яблонцева А.Д.

Хакасский государственный университет им. Н.Ф. Катанова, Россия, г. Абакан

РЕАЛИЗАЦИЯ МОДЕЛИ ПРОПОРЦИОНАЛЬНЫХ РИСКОВ КОКСА СРЕДСТВАМИ ЯЗЫКА ПРОГРАММИРОВАНИЯ R

Аннотация

В работе описывается регрессионная модель Кокса. Приводится пример реализации регрессии Кокса на языке программирования R в анализе времени до смерти на ковариатах постоянной времени. Визуализируется прогнозируемая доля выживаемости по времени для конкретной группы риска.

Ключевые слова: регрессия Кокса, анализ выживаемости, модели рисков

Keywords: Cox regression, survival analysis, risk models

Научный руководитель Голубничий А.А.
старший преподаватель кафедры ПОВТиАС,
ХГУ им. Н.Ф. Катанова, Россия, г. Абакан

Модели пропорциональных рисков Кокса [1] оценивают одновременно влияние нескольких факторов, влияющих на выживаемость. Другими словами, это позволяет исследовать, как указанные факторы влияют на частоту возникновения определенного события в определенный момент времени. Данный показатель обычно называют степенью опасности. В литературе по анализу выживаемости переменные-предикторы обычно называются ковариатами. Модель Кокса выражается функцией риска, также она может быть записана как множественная линейная регрессия логарифма риска по переменным, при этом базовая опасность является термином «intercept», который меняется со временем. Ключевое предположение модели Кокса состоит в том, что кривые риска для групп наблюдений (или пациентов) должны быть пропорциональными и не могут пересекаться. Следовательно, модель Кокса является моделью пропорциональных опасностей: опасность события в любой группе является постоянным числом, кратным опасности в любой другой.

Для реализации модели Кокса в языке программирования R существует множество решений, в рамках данного исследования выбор остановился на пакетах survival [2] и survminer [3]. Функция coxph(), может использоваться для вычисления модели регрессии пропорциональных рисков Кокс в R. Объект анализа выживаемости создается с помощью функции Surv().

Для анализа были использованы данные из датасета lung пакета survival, содержащие сведения о раке легких. На рисунке 1 приведены общий вид данных в датасете.

```
> head(lung)
  inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss
1    3  306     2  74  1     1         90      100     1175      NA
2    3  455     2  68  1     0         90       90     1225      15
3    3 1010     1  56  1     0         90       90       NA      15
4    5  210     2  57  1     1         90       60     1150      11
5    1  883     2  60  1     0        100       90       NA       0
6   12 1022     1  74  1     1         50       80       513       0
```

Рисунок 1 – Рендеринг исходного набора данных lung пакета survival

Для настройки регрессии Кокса, были использованы ковариаты: age, sex, ph.ecog и wt.loss. Для вычисления одномерного анализа Кокса используются всех переменные, затем необходимо провести многомерный анализ Кокса с использованием двух переменных, чтобы описать, как факторы совместно влияют на выживаемость (рисунок 2).

```
> res.cox <- coxph(Surv(time, status) ~ sex, data = lung)
> res.cox
Call:
coxph(formula = Surv(time, status) ~ sex, data = lung)

      coef exp(coef) se(coef)      z      p
sex -0.5310   0.5880   0.1672 -3.176 0.00149

Likelihood ratio test=10.63 on 1 df, p=0.001111
n= 228, number of events= 165
```

Рисунок 2 – Рендеринг одномерного анализа Кокса

Столбец с пометкой «z» дает статистическое значение Вальда. Он соответствует отношению каждого коэффициента регрессии к его стандартной ошибке. Вторая особенность, которую следует отметить в результатах модели Кокса – это знак коэффициентов регрессии (coef). Положительный знак означает, что риск смерти выше, а значит, и прогноз хуже для субъектов с более высокими значениями этой переменной. Переменная пол кодируется как числовой вектор 1 – мужчина, 2 – женщина. Итоговые данные также дают верхнюю и нижнюю оценку доверительного интервала для уровня 95% для отношения рисков (exp(coef)), нижняя граница 95% = 0,4237, верхняя граница 95% = 0,816. Результат регрессии Кокса приведен на рисунке 3.

```
> summary(res.cox)
Call:
coxph(formula = Surv(time, status) ~ sex, data = lung)

n= 228, number of events= 165

      coef exp(coef) se(coef)      z Pr(>|z|)
sex -0.5310   0.5880   0.1672 -3.176 0.00149 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
sex      0.588      1.701   0.4237   0.816

Concordance= 0.579 (se = 0.021 )
Likelihood ratio test= 10.63 on 1 df,  p=0.001
Wald test              = 10.09 on 1 df,  p=0.001
Score (logrank) test = 10.33 on 1 df,  p=0.001
```

Рисунок 3 – Результат анализа регрессии Кокса

Выходные данные указывают уровень значимости p-value для трех альтернативных тестов модели: тест отношения правдоподобия, тест Вальда и статистику логарифмического ранжирования. Эти три метода асимптотически эквивалентны и для достаточно большого N они дадут аналогичные результаты. Для малых N они могут несколько отличаться.

Результат применения одномерной функцию coxph к нескольким ковариатам приведен на рисунке 4.

	beta	HR (95% CI for HR)	wald.test	p.value
age	"0.019"	"1 (1-1)"	"4.1"	"0.042"
sex	"-0.53"	"0.59 (0.42-0.82)"	"10"	"0.0015"
ph.karno	"-0.016"	"0.98 (0.97-1)"	"7.9"	"0.005"
ph.ecog	"0.48"	"1.6 (1.3-2)"	"18"	"2.7e-05"
wt.loss	"0.0013"	"1 (0.99-1)"	"0.05"	"0.83"

Рисунок 4 – Результат применение одномерной функции coxph.

Age и ph.ecog имеют положительные бета-коэффициенты, а sex – отрицательный. Таким образом, пожилой возраст и более высокий показатель ph.ecog связаны с более низкой выживаемостью, тогда как принадлежность к женскому полу увеличивает вероятность выживаемости.

Регрессия Кокса по времени до смерти на ковариатах постоянной времени, показана на рисунке 5.

```
> res.cox <- coxph(Surv(time, status) ~ age + sex + ph.ecog, data = lung)
> summary(res.cox)
Call:
coxph(formula = Surv(time, status) ~ age + sex + ph.ecog, data = lung)

n= 227, number of events= 164
(1 observation deleted due to missingness)

      coef exp(coef) se(coef)      z Pr(>|z|)
age      0.011067  1.011128  0.009267  1.194 0.232416
sex     -0.552612  0.575445  0.167739 -3.294 0.000986 ***
ph.ecog  0.463728  1.589991  0.113577  4.083 4.45e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
age      1.0111      0.9890   0.9929   1.0297
sex      0.5754      1.7378   0.4142   0.7994
ph.ecog  1.5900      0.6289   1.2727   1.9864

Concordance= 0.637 (se = 0.025 )
Likelihood ratio test= 30.5 on 3 df,  p=1e-06
Wald test              = 29.93 on 3 df,  p=1e-06
Score (logrank) test = 30.5 on 3 df,  p=1e-06
```

Рисунок 5 – Регрессия Кокса по времени до смерти на ковариатах постоянной времени

Значение p-value для всех трех тестов значимы, указывая на то, что модель значима. Эти тесты оценивают общую нулевую гипотезу о том, что все бета-версии (β) равны 0. В приведенном выше примере статистические данные теста полностью совпадают, и общая нулевая гипотеза полностью отвергается. В многомерном анализе Кокса, ковариаты sex и rh.ecog остаются значимыми. Однако ковариационный возраст не может быть значимым. Отношения рисков ковариат интерпретируются как мультипликативный эффект на опасность.

Подбирая модель Кокса к данным, можно визуализировать прогнозируемую долю выживаемости в любой момент времени для конкретной группы риска. Функция `Survfit()` оценивает долю выживаемости по умолчанию при средних значениях ковариат. Реализация приведена на рисунке 6.

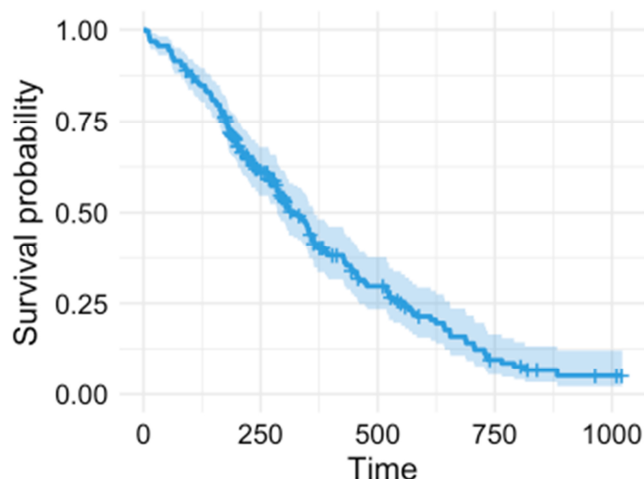


Рисунок 6 – Визуализация модели пропорциональных опасностей Кокса

Таким же образом можно проанализировать влияние пола на предполагаемую вероятность выживания. В этом случае, создается новый фрейм данных с двумя строками для каждого значения пола; другие ковариаты фиксируются на своих средних значениях или на самом низком уровне. Для фиктивной ковариаты среднее значение – это пропорция, закодированная в наборе данных. Набор данных передается в `Survfit()` через аргумент `newdata`. Результат приведен на рисунке 7.

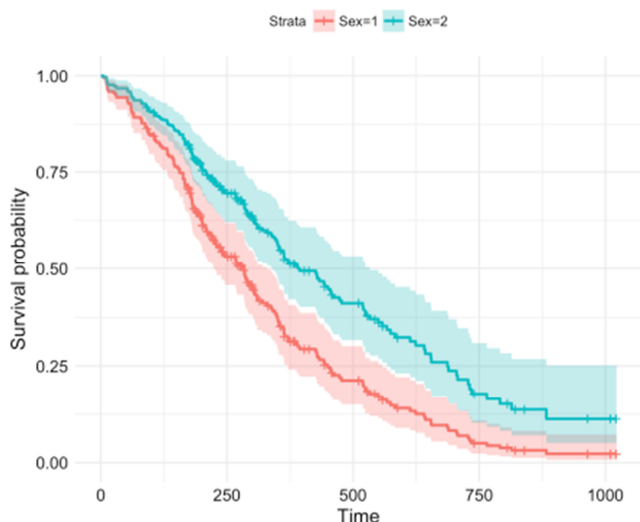


Рисунок 7 – Визуализация модели пропорциональных опасностей Кокса с учетом разделения по полу

Литература

1. Регрессия Кокса [Электронный ресурс] URL: <https://www.statmethods.ru/statistics-metody/regressiya-koksa/> (дата обращения: 11.08.2021).
2. CRAN – Package survival [Электронный ресурс] URL: <https://CRAN.R-project.org/package=survival> (дата обращения: 11.08.2021).
3. CRAN – Package survminer [Электронный ресурс] <https://CRAN.R-project.org/package=survminer> (дата обращения: 11.08.2021).