Applied Survival Analysis Using R
Chapter 5: Regression Analysis Using the Proportional Hazards Model

Qi Guo

Department of Mathematical Sciences
The University of Texas at Dallas

April, 12 2019

Covariates

Covariate

In general terms, *covariates* are characteristics (excluding the actual treatment) of the participants in an experiment. If you collect data on characteristics before you run an experiment, you could use that data to see how your treatment affects different groups or populations. It uses in *ANCOVA*

Characteristics:

- A covariate can be an *independent variable* or it can be an unwanted, confounding variable. Adding a covariate to a model can increase the accuracy of your results.
- Observed/measured (as opposed to a manipulated variable).
- A control variable.

The proportional hazards model

- Proportional hazard function:

$$h_1(t) = \psi h_0(t) \tag{1}$$

- Furthermore, we can extend the model to include covariate information in a vector z as follows:

$$\psi = e^{z\beta} \tag{2}$$

- *Partial likelihood* for assuming a form for $h_0(t)$.

The partial likelihood

- The partial likelihood will allow us to use an unspecified *baseline survival distribution* to define the survival distributions of subjects based on their covariates.
- Why partial likelihood?
  - It is a product of expressions, one for each failure time, while censoring times do not contribute any factors.
  - The factors of a partial likelihood are conditional probabilities.

Notation

- The hazard function for Subject i at failure time t_j is $h_i(t_j)$.
- j denotes the $j'th$ failure time (where the failure times are sorted from lowest to highest)
- Under the proportional hazards model,write this hazard function as:
  - $h_i(t_j) = h_0(t_j)\psi_i$
  - $\psi_i = e^{z_i\beta}$
  - Covariate z_i is either *1* (if the patient is in the experimental group) or *0* (of the patient is in the control group)
- Patients in the experimental group are expected less likely than control patients to experience the event(experiment work expected)
  - $\beta < 0$ and $\psi < 1$.
- $\psi_i = 1$ if a patient is in the control group or $\psi_i = \psi$ if a patient is in the experimental group.

Formula

- The first failure time t_1.The set of all subjects in the trial "**at risk**" for failure is denoted by R_1(Before the first failure, all of the patients in "**at risk**").
- Among them in R_1, the probability that Patient i fails in the hazard, $h_i(t_j) = h_0(t_j)\psi_i$, for that patient divided by the sum of the hazards of all of the patients:

$$p_1 = \frac{h_i(t_1)}{\sum_{k \in R_1} h_k(t_1)} = \frac{h_0(t_1)\psi_i}{\sum_{k \in R_1} h_0(t_1)\psi_k} \tag{3}$$

- where $h_0(t_1)$ is the hazard for a subject from the control group, and is that the *baseline hazard function*.
- Cancels out, we have:

$$p_1 = \frac{\psi_i}{\sum_{k \in R_1} \psi_k} \tag{4}$$

Example

- The partial likelihood:

$$L(\psi) = p_1 p_2 \cdots p_D \tag{5}$$

Patient	Survtime	Censor	Group
1	6	1	C
2	7	0	C
3	10	1	T
4	15	1	C
5	19	0	T
6	25	1	T

- We know $\psi_1 = \psi_2 = \psi_4 = 1$ and $\psi_3 = \psi_5 = \psi_6 = \psi$, substituting into (4), at time 6, a control patients failure, 3 control and 3 treated at risk.

$$p_1 = \frac{1 \cdot h_0(t_1)}{3 \cdot h_0(t_1)\psi + 3 \cdot h_0(t_1)} = \frac{1}{3\psi + 3} \tag{6}$$

Example

- The second failure time, $t = 10$:

$$p_2 = \frac{\psi}{3\psi + 1} \tag{7}$$

- The third failure time, $t = 15$:

$$p_3 = \frac{1}{2\psi + 1} \tag{8}$$

- The last failure time, $t = 25$ only one subject at risk and died, so it just 1:
- The partial likelihood:

$$L(\psi) = \frac{\psi}{(3\psi + 3)(3\psi + 1)(2\psi + 1)} \tag{9}$$

Transformation

- Let $\psi = e^{\beta}$,and take a derivative:

$$\ell(\beta) = \beta - \log(3e^{\beta} + 3) - \log(3e^{\beta} + 1) - \log(2e^{\beta} + 1) \qquad (10)$$

- The maximum partial likelihood estimate is the value of that *maximizes* this function.
- Defining the function $\ell(\beta)$:

```
1 >plsimple <- function(beta) {
2 psi <- exp(beta)
3 result <- log(psi)-log(3*psi+3)-log(3*psi+1)-log(2*psi+1) result}
```

- Find the m.p.l.e (maximum partial likelihood estimate) using the "optim" function(In my R for data science presentation).

```
1 >result <- optim(par=0, fn = plsimple, method = "L-BFGS-B",
2 control=list(fnscale = -1),lower = -3, upper = 1)
3 >result$par
4 [1] -1.326129
```

- Thus, the m.p.l.e is $\hat{\beta} = -1.326129$.

Recall

- Test of $H_0 : \beta = 0$
- The *score* function: $S(\beta) = \ell'(\beta)$
- The *information* function: $I(\beta) = -S'(\beta) = -\ell''(\beta)$, and the second derivative $-\ell''(\beta)$ is also known as the *Hessian*.
- Three forms of test:
  - *The Wald test*
  - The score test
  - *The likelihood ratio test*

The Wald Test

- *The Wald test* is perhaps the most commonly used test.
- The test form $Z = \hat{\beta}/s.e(\hat{\beta})$
  - $\hat{\beta}$ was the value of β that maximizes $\ell(\beta)$.
  - $s.e(\hat{\beta}) = 1/\sqrt{I(\hat{\beta})}$
  - Construct a normalized test statistic $Z_w = \hat{\beta}/s.e(\hat{\beta})$
- Reject $H_0 : \beta = 0$ if $|Z_w| > z_{\alpha/2}$
- Equivalently, Reject $H_0 : \beta = 0$ if $Z_w^2 > \chi^2_{\alpha,1}$
- Construct a $1 - \alpha$ confidence interval: $\hat{\beta} \pm z_{\alpha/2} \cdot s.e.(\hat{\beta})$.

The Score Test

- The test form $Z_s = S(\beta = 0)/\sqrt{I(\beta = 0)}$
  - Reject $H_0 : \beta = 0$ if $|Z_s| > z_{\alpha/2}$
  - Equivalently, Reject $H_0 : \beta = 0$ if $Z_s^2 > \chi^2_{\alpha,1}$
- *The score test* is equivalent to the log-rank test, as we saw in the previous section. This test can be carried out without finding the maximum likelihood estimate $\hat{\beta}$.

The Likelihood Ratio Test

- The likelihood ratio test uses the result $2[\ell(\beta = \hat{\beta}) - \ell(\beta = 0)]$ follows approximately a chi-square distribution with one df.
- The key advantage of this test over the other two is that it is invariant to monotonic transformations of β.

Example

We illustrate these three tests using the simple data last Example, begin by presenting the output from the "`coxph`" function. The result is put into a data structure called "`result.cox`", and a complete summary of the results we obtain using the "`summary`" function.

coxph function

```
>result.cox <- coxph(Surv(tt,status) ~ grp)
>summary(result.cox)
[1] Call:  coxph(Surv(tt,status) ~ grp)

      n = 6, number of events= 4

         coef       exp(coef)     se(coef)        z      Pr(>|z|)
grp    -1.3261       0.2655        1.2509       -1.06       0.289
         exp(coef)  exp(-coef)    lower.95    upper.95
grp       0.2655      3.766        0.02287      3.082
Concordance= 0.7 (se = 0.187)
Rsquare= 0.183 (max possible=0.76)
Likelihood ratio test= 1.21 on 1 df, p=0.2715
Wald test            = 1.12 on 1 df, p=0.2891
Score (logrank) test = 1.27 on 1 df, p=0.2591
```

Without `coxph` function

- The Score test

```
>library(numDeriv)
>grad(func=plsimple, x=0) #The score evaluated at beta = 0 numericall
the ``gradient'' function
[1] -0.917
>hessian(func=plsimple, x=0) #The information
     [,1] [1,] -0.660
```

- So the $Z_s^2 = (-0.917)^2/0.660 = 1.274$
- And the test *p-value* is given by the upper tail:

```
>pchisq(1.274, df=1, lower.tail=F)
[1] 0.259
```

Without `coxph` function

- The Wald test
- Before we get $\hat{\beta} = -1.326129$, and need information to get s.e:

```
>hessian(func=plsimple, x=result.cox$par)
      [,1]
[1,] -0.639
>sqrt(1/0.639) #s.e
[1] 1.251
```

- Finally, the Wald test statistic Z_W and two-sided p-value for the test are given by:

```
>-1.326/1.251
[1] -1.060
>2*pnorm(1.060, lower.tail=F)
[1] 0.289
```

Without `coxph` function

- The partial likelihood ratio test
- The likelihood ratio statistic is twice the difference in the log partial likelihoods evaluated at $\hat{\beta}$ and at 0:

```
>betahat <- result.cox$par
>2*(plsimple(betahat) - plsimple(0))
[1] 1.209
```

- Along with the `p-value` derived from the `chi-square` distribution:

```
>pchisq(1.209, 1, lower.tail=F)
[1] 0.271
```

- Besides, the statistic "`r-squared`" is an adaptation to survival analysis of the R^2 statistic from linear regression:

$$R^2 = 1 - \left(\frac{\ell(0)}{\ell(\hat{\beta})}\right)^{2/n} \tag{11}$$

Baseline Survival Function

- An estimate of the *baseline hazard function* is given by:

$$h_0(t_i) = \frac{d_i}{\sum_{j \in R_i} exp(z_j \hat{\beta})} \tag{12}$$

- An estimate of the *baseline survival function* is given by:

$$S_0(t) = exp[-H_0(t)] \tag{13}$$

- To find a survival curve for a particular covariate value z use:

$$S(t|z) = [S_0(t)]^{exp(z\hat{\beta})} \tag{14}$$

- In `R` the "`basehaz`" function will compute a cumulative baseline hazard function, use the option "`entered = F`" to cause it to estimate the cumulative hazard at $\beta = 0$

Example 1

- Enter the data as before(Chapter 3):

```
>tt <- c(6, 7, 10, 15, 19, 25)
>status <- c(1, 0, 1, 1, 0, 1)
>grp <- c(0, 0, 1, 0, 1, 1)
>backTime <- c(-3, -11, -3, -7, -10, -5)
```

- Without Left Truncation

```
>coxph(Surv(tt, status) ~ grp)
      coef    exp(coef)     se(coef)     z        p
grp -1.33        0.266         1.25    -1.06    0.29

Likelihood ratio test=1.21 on 1 df,p=0.271 n= 6,number
of events= 4
```

- Conclusion: The experimental group has a *lower hazard* than the control group, but this difference is *not statistically significant*(p-value = 0.271 based on the likelihood ratio test).

Example 1

- With Left Truncation

```
>tm.enter <- -backTime
>tm.exit <- tt - backTime
>coxph(Surv(tm.enter, tm.exit, status, type="counting") ~ grp)
        coef     exp(coef)    se(coef)      z        p
grp   -1.07         0.342        1.24    -0.869    0.39

Likelihood ratio test=0.81 on 1 df,p=0.368
```

- Conclusion: Similar non-significant treatment difference conclusion.

Example 2

- Another example is the "`Channing House data`" as before.
  - The "`start.time`" option we used previously is not available in the "`coxph`" function

```
>channing68 <- ChanningHouse[ChanningHouse$exitYears >= 68,]
>coxph(Surv(entryYears, exitYears, cens, type="counting") ~ sex,
data=channing68)
           coef     exp(coef)   se(coef)      z       p
sexMale   0.273         1.31      0.176     1.55   0.12

Likelihood ratio test=2.3 on 1 df,p=0.129
```

- Conclusion: Men have a *higher hazard* (and hence lower survival) than do women, but this difference is *not statistically significant*.