Regression models for time to event data

Shannon Pileggi

STAT 417

OUTLINE

Regression models for time to event data

- In the Veterans Administration lung cancer example, we investigated differences in the survival experiences for individuals with various types of lung cancer. Basically, we are addressing:

- In addition, are there other **explanatory variables** (i.e. **predictors**), that are associated with the risk of dying from lung cancer?
- To investigate the effects of predictor variables (**categorical** and **quantitative**) on the time to experience the event of interest (or some function of T) we can fit regression models to the time to event data.

Two approaches

1. The **Cox regression model** (also referred to as the **proportional hazards (PH) model**): The hazard function $h(t)$ depends on a set of p explanatory variables $X_1, X_2, \ldots, X_p$ through the model:

   where $h_0(t)$ is called a *baseline hazard function* that is assumed to be common for each subject.
2. The **accelerated failure time (AFT) model**. Expresses the (natural) log of the time-to-event random variable T as a linear combination of the p explanatory variables $X_1, X_2, \ldots, X_p$:

   where σ is a scale parameter and W represents a random error term with some specific distribution.

Why model hazard?

1.

2.

3.

Goals of a Cox regression model

1.

2.

3.

Overview

CR: one categorical predictor

Proportionality

Interpretations

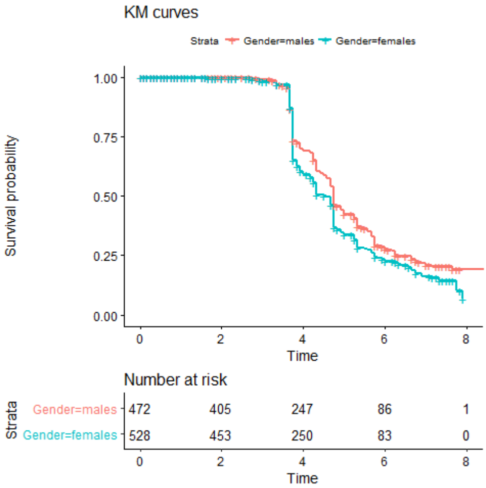
Estimation

College graduation data

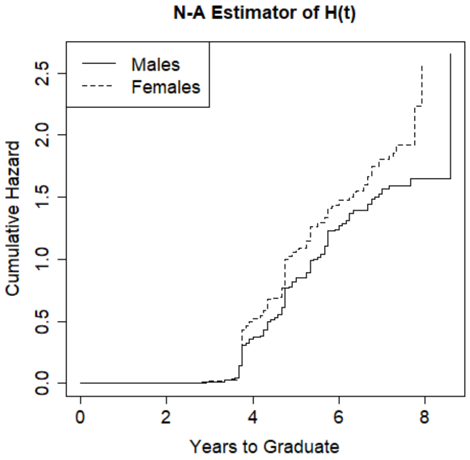
How long does it take for students to graduate from college? At what time(s) during their college career are they most likely to graduate? How do graduation experiences vary by gender?

- Using data on a sample of 1000 participants from the National Educational Longitudinal Survey (NELS) from 1988-2002, we will investigate whether student's gender has an effect on the time required to obtain a bachelor's degree.
- Let $T =$ time (in months) taken to complete the requirements for a bachelor's degree.
- If an individual had pursued and obtained a bachelor's degree by the interview date (in 2000), then her/his event time is complete. If the student dropped out or did not graduate by the date they were interviewed, then her/his time is right censored.

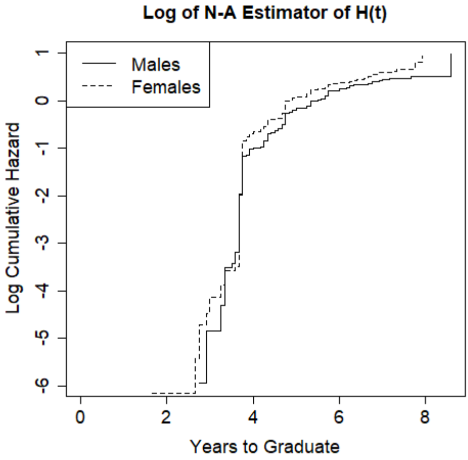
KM curves: years until graduation



$\tilde{H}(t)$: years until graduation



$\ln[\tilde{H}(t)]$: years until graduation



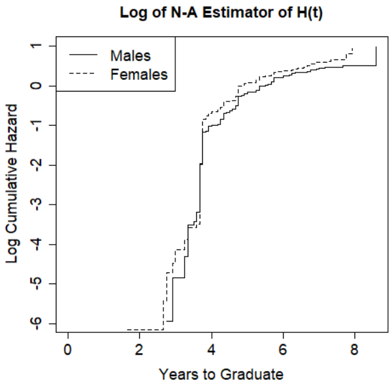
Cox regression model: one categorical predictor

Let's set up a CR model with one dichotomous (categorical) predictor, X, that only takes values 0 or 1. The form of the CR model is:

Cox regression model: one categorical predictor

Let `X = 1` for females and `X = 0` for males. Find the ratio of the hazard functions for females to males.

Back to the picture...



Example of proportionality

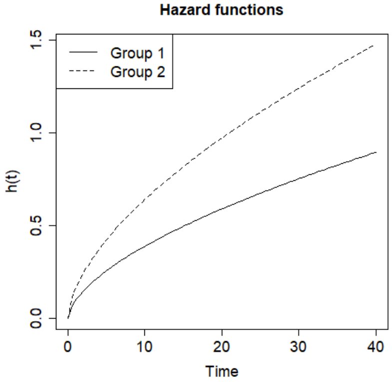
Suppose that the CR model is given by:

$$h(t|X) = \frac{(1.6)t^{.6}}{(5.74)^{1.6}} e^{.5X}$$

where $X = 1$ indicates membership to Group 1, and $X = 0$ indicates membership to Group 2.

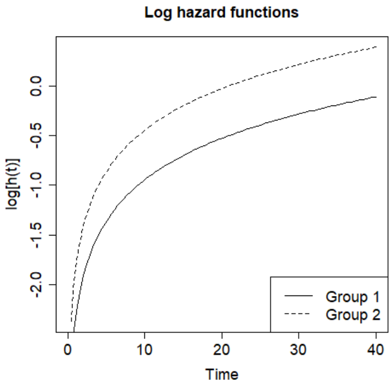
1. What is the baseline hazard function? What is the value of the parameter β?

Example of proportionality



2. Show that the ratio of the hazard functions (hazard ratio) of Group 1 to Group 2 is constant across time t.

Example of proportionality



3. Show that the difference in the log-hazard functions between Group 1 and Group 2 is constant across time t.

CR model: proportionality assumption

The main feature about the CR models observed in the above examples is known as the *proportionality assumption*. In general, for a set of p explanatory variables $X_1, X_2, \ldots, X_p$, the **hazard ratio (HR)** is given by:

CR model: one categorical predictor

Consider the CR model with one binary categorical predictor where X takes the value 0 or 1: $h(t|X) = h_0(t)e^{\beta X}$

- β represents:

- e^{β} is the **hazard ratio (HR)** and represents:

Interpretation of the hazard ratio e^{β}

If X is a categorical predictor taking values 0 or 1, then the hazard ratio (HR) of subjects with $X = 1$ to subjects with $X = 0$ is:

$$\boxed{HR = \frac{h(t|X=1)}{h(t|X=0)} = e^{\beta}}$$

- $e^{\beta} > 1$:
- $e^{\beta} < 1$:
- $e^{\beta} = 1$:

Example interpretation

Consider the CR model for the graduation data, and suppose that $\beta = .2$ (where $X = 1$ for females and $X = 0$ for males). Then the CR model would be given by:

1. Interpret the value of the coefficient $\beta = .2$.

Example interpretation, cont.

2. Interpret the value of the hazard ratio of females to males.

CR model: one quantitative predictor

- **Veterans Administration Lung Cancer Group Study**. Recall the lung cancer study to investigate the effects of two treatments on survival. Another explanatory variable available is the *Karnofsky performance score* which is used to quantify cancer patients' functional impairment. The Karnofsky performance score is measured on a 0-100 scale in increments of 10 with 0 indicating that the subject is dead, and 100 indicating that the subject shows no signs of the disease.
- Then $X =$ Karnofsky performance score is a *quantitative* variable.
- Suppose we want to investigate the effect of patient's Karnofsky score on the risk (hazard) of dying from lung cancer.

CR model: one quantitative predictor

Suppose X is quantitative and consider the CR model:

$$\boxed{h(t|X) = h_0(t)e^{\beta X}}$$

- The CR model for subjects with $X = a$ is:
- The CR model for subjects with $X = a + c$ is:
- Then the hazard ratio for subjects with $X = a + c$ to subjects with $X = a$ is:

CR model: one quantitative predictor interpretation of HR

The hazard ratio $e^{c\beta}$ is the multiplicative change in the hazard rate for a c unit increase in X, or alternatively, β is the change in log hazard for a c unit increase in X.

- $e^{\beta} > 1$:

- $e^{\beta} < 1$:

- $e^{\beta} = 1$:

Example interpretation with quantitative predictor

To investigate the effect of patient's Karnofsky score on the hazard of death from lung cancer, suppose we have the CR model:

$$h(t|X) = h_0(t)e^{-.03X}$$

1. Interpret $\beta = -.03$.

Example interpretation with quantitative predictor, cont.

2. Interpret $\beta = -.03$ in terms of a 10 point increase in Karnofsky score.

Example interpretation with quantitative predictor, cont.

3. Compute and interpret the hazard ratio for a 10 point increase in Karnofsky score.

Overview

CR: one categorical predictor

Proportionality

Interpretations

Estimation

Parameter estimation

- In the previous examples, values of the parameter β were given for the CR model. However, parameters values are typically unknown and must be estimated from data.
- Recall the form of the CR model with one explanatory variable:

$$h(t|X) = h_0(t)e^{\beta X}$$

- If we wanted to fully specify all the unknown quantities, what would we need?
- In general, the parameter β (or several β's) will be unknown, as well as the baseline hazard function $h_0(t)$.
- But since we treat $h_0(t)$ as a *nuisance parameter*, we are not interested in immediately estimating this quantity (however, we can estimate it after we estimate the β's).

Partial maximum likelihood

- A procedure called **partial maximum likelihood** that does not require any specification of the baseline hazard function is used to estimate the β parameters only.
- Skipping the mathematical details, but some notes:
  - Only subjects with complete event times will contribute terms to the (partial) likelihood function, and the likelihood function will only contain β terms (no $h_0(t)$ terms).
  - Using numerical methods, software obtains the β value(s) that maximize the likelihood function.
  - The estimates of $\beta_1, \beta_2, \ldots, \beta_p$ will be denoted by:
- Partial maximum likelihood is implemented in many statistical software packages, but not Minitab! CR models will be fit to time-to-event data in `R`.

Fitting CR model in R

R Code

```
CR_grad <- coxph(Surv(Years, Censor) ~ as.factor(Gender),
                 data = graduate)
summary(CR_grad)
```

R Code

R Output

```
  n= 1000, number of events= 614

                      coef exp(coef) se(coef)     z Pr(>|z|)
as.factor(Gender)1 0.20862   1.23197  0.08141 2.563   0.0104 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

R Output

Example interpretation of parameter estimates

1. State the *estimated* Cox regression model.

2. Interpret the value of the estimated coefficient $\hat{\beta}$.

3. Interpret the value of the estimated hazard ratio of females to males.

R Output

```
Call:
coxph(formula = Surv(Years, Censor) ~ as.factor(Gender), data = grad

  n= 1000, number of events= 614

                      coef exp(coef) se(coef)     z Pr(>|z|)
as.factor(Gender)1 0.20862   1.23197  0.08141 2.563   0.0104 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

                   exp(coef) exp(-coef) lower .95 upper .95
as.factor(Gender)1     1.232     0.8117      1.05     1.445

Concordance= 0.53  (se = 0.012 )
Rsquare= 0.007   (max possible= 0.999 )
Likelihood ratio test= 6.61  on 1 df,   p=0.01016
Wald test            = 6.57  on 1 df,   p=0.01039
Score (logrank) test = 6.59  on 1 df,   p=0.01025
```

R Output