Cox Proportional Hazards Regression Model

Mai Zhou
Department of Statistics
University of Kentucky

A search in *Journal of American Medical Association* yields similar results.

Some quotes from abstracts of NEJM:
..The time to cancer in the two groups was compared by Kaplan–Meier analysis and a **Cox proportional-hazards** model ….. (5/2002)

The presence of an interaction between sex and digoxin therapy with respect to the primary end point of death from any cause was evaluated with the use of Mantel–Haenszel tests of heterogeneity and a multivariable **Cox proportional-hazards** model, adjusted for demographic and clinical variables. (10/2002)

With the use of **Cox proportional-hazards** models, the body-mass index was evaluated ….. (8/2002)

We used **proportional-hazards** regression models to estimate the effect on mortality of combination therapy ….. (11/2001)

**Methods** We estimated graft survival using **proportional-hazards** techniques, adjusting for patient and donor characteristics, for a series of 30,564 Medicare patients receiving …. 

**Methods** In this prospective cohort study, we estimated the effects of air pollution on mortality, while controlling for individual risk factors. Survival analysis, including **Cox proportional-hazards** regression modeling, was conducted with data…….
Exponential Random Variable

- Two ways of describing an Exponential random variable.

1. Length of “life”… \( X \) CDF, pdf

2. Force of mortality, hazard, risk, intensity.
Hazard at time $t$

Imaging some evil force try to kill you: at time $t$ the intensity of the force is $h(t)$ [must $\geq 0$].

The (conditional) probability you die in the next small time interval $(t, t + \Delta t]$ (provided you still alive at time $t$) is

$$h(t) \Delta t$$
• If $h(t) = \text{constant}$, then we get Exponential distribution.

• Easiest in the language of hazard.
• But may not be appropriate for many cases.
  OK to describe an electric component under constant working condition.
  But my hazard goes up if I am going through a high stress time, downhill skiing……

• Average US population daily hazard based on 2000 census is bath-tub shaped.
Exponential Regression Model

• Every patient’s lifetime is an exponential r.v.
• The only difference is the $\lambda$ constant hazard.
  This patient is female, young and have no family history of heart problems --- her risk (constant) is low. (or 20% lower …)
• Constant for patient $i$, $\lambda_i$, depends on his/her covariates (gender, age, gene …..)
Exponential regression Model (cont.)

- The constant hazard for patient $i$ is

$$\lambda_i = \exp(\alpha + \beta_1 age_i + \beta_2 gender_i + \beta_3 trt_i)$$

The exp( ) is used to ensure the constants are always positive.

Other positive, monotone function can also be used. If the range of covariates is always positive then we may get by without function.
Cox Model is Exponential model under a variable (crazy) time clock

- “I went through two years worth of trouble in the past two months” (faster clock)
- “life in the fast lane”
- “One year for a dog is like 7 years for a human”.

- Cox model only uses the rank order of the data to estimate the risk ratios. Clocks do not change the rank ordering.

- We do not have to know how fast/slow the clock ---- (semiparametric).
• But every patient uses the same crazy clock!

• Still make sense to say patient A has 20% lower risk than average, but did not make sense to say the risk is 0.8 without specify the clock.

• Model assumption may not always be true. Solution----stratified Cox model, other models (AFT models, etc.)
• Censoring. The lifetime observations may be (right) censored. 34+, 55 +, etc
• We can estimate the crazy clock’s speed. (Kaplan-Meier estimator and its relatives.) (but this time we use more than just the orders of the data.)
• Easy to convert hazards to survival probability plots.