

# Applied Survival Analysis - January 2016

## Lab 4: Cox Proportional Hazards Model

In today's lab, we are going to review the basic interpretation of a Cox proportional hazards model. Then we are going to learn how to fit a Cox PH model using R and evaluate the implication of tied failure times.

- (a) **Interpretation of Cox Model: Prognosis with Breast Cancer.** A follow-up study of post-menopausal women diagnosed with breast cancer was performed to examine whether the estrogen receptor (ER) status of the tumor was related to prognosis, adjusting for stage of disease at diagnosis and age at diagnosis.

Let  $\lambda(t|\mathbf{X})$  be the hazard of death at month  $t$  after diagnosis for an individual with covariates  $\mathbf{X}$ ,  $\mathbf{X} = (X_E, X_A, X_S)$ , where

$$X_E = \begin{cases} 1, & \text{ER positive tumor} \\ 0, & \text{ER negative tumor} \end{cases}, X_A = \text{Age at diagnosis (years)}, X_S = \begin{cases} 1, & \text{is situ} \\ 2, & \text{local} \\ 3, & \text{regional} \\ 4, & \text{distant} \end{cases}$$

Suppose the following proportional hazards model was found to fit the data:

$$\lambda(t|\mathbf{X}) = \lambda_{0,X_S}(t) \exp \{-0.783X_E + 0.007X_A\},$$

where  $\lambda_{0,X_S}(t)$  is the baseline hazard function, specific for each stage (i.e., the value of  $X_S$ ). In other words, this is a stratified proportional hazards model, stratified by tumor stage. For the following questions, get an actual number if you can, showing how you got it. If you can't get an actual number, then write an expression for how it would be calculated if you had additional information

- (i) Based on this model, what is your best estimate of the hazard ratio (i.e. relative risk) of death for a woman with an ER positive tumor relative to a woman of the same age and with the same stage ER negative tumor, the same number of months beyond diagnosis? Do women with ER positive tumors have a more favorable or less favorable prognosis?

- (ii) Based on this model, what can you say about the hazard ratio of death for a woman 62 years old at diagnosis with a localized ER positive tumor, 24 months beyond diagnosis, relative to a woman 67 years old at diagnosis with a localized ER negative tumor, 24 months beyond diagnosis?
- (iii) Based on this model, what can you say about the hazard ratio of death for a woman 55 years old at diagnosis with an in situ ER positive tumor, 36 months beyond diagnosis, relative to a woman 55 years old at diagnosis with an in situ ER negative tumor, 24 months beyond diagnosis?
- (iv) Based on this model, what can you say about the hazard ratio of death for a woman 60 years old at diagnosis with an ER positive tumor with regional spread, 24 months beyond diagnosis, relative to a woman 60 years old at diagnosis with an ER negative in situ tumor, 24 months beyond diagnosis?

(b) **Fitting Cox Model and handling of ties: Nursing Home Data.** Now we are going to move into R considering the dataset *nurshome.dta*. The National Center for Health Services Research studied 36 for-profit nursing homes to assess the effects of different financial incentives on length of stay. “Treated” nursing homes received higher per diems for Medicaid patients, and bonuses for improving a patient’s health and sending them home. The study included 1601 patients admitted between May 1, 1981 and April 30, 1982.

Variables include:

***los*** - Length of stay of a resident (in days)

***age*** - Age of a resident

***rx*** - Nursing home assignment (1:bonuses, 0:no bonuses)

***gender*** - Gender (1:male, 0:female)

***married*** - Marital Status (1: married, 0:not married)

***health*** - health status (2:second best, 5:worst)

***fail*** - Failure/Censoring indicator (1:discharged, 0:censored)

- (i) Import the data in R. Use the **coxph** function to fit a Cox proportional hazards model focusing on the effect of marital status on the length of stay of a resident (type **?coxph** to see the syntax required). What is the estimated hazard ratio? Interpret the result.
- (ii) Which methods for tie handling are available in the **coxph** function? Refit the model of part (i) using all the available methods. How much impact do the different approaches have on the estimate of  $\beta$ ? Would you expect tied failure times to be a big issue in this dataset? Which approach would you recommend?

- (iii) Compute the *Logrank* and *Wilcoxon* test focusing on the effect of marital status on the length of stay, and compare the results with the test statistics you obtained in part (ii). Do any of them match?
- (c) **Optional:** Consider the case of a clinical trial in which a new treatment is compared with a standard treatment (control). Suppose that the hazard rate of death is 4 events/year in the control group, whereas it's 2 in the treatment group, assuming also that the rate is constant over time for both groups. Thus, the hazard ratio that compares the treatment group with the control group equals 0.5. To allow the occurrence of right censoring, suppose that there is an underlying censoring time that is exponentially distributed with its mean being 5 years. Assume that the number of patients in each treatment group is 5000.
- Simulate the above setting in R, using the `set.seed` function to get reproducible results. Then fit a cox model to those data. What hazard ratio did you get? Does the cox model seem appropriate for analysing data similar to the ones you simulated? Explain your opinion.