Applied Survival Analysis - January 2016

Lab 5: More on Cox Proportional Hazards Model

Today, we are going to see how to construct confidence intervals and tests for hazard ratios, and compare nested models using likelihood ratio tests. Then we are going to learn how to estimate the baseline survival function and predicted medians.

(a) **C.I., Wald test and Likelihood Ratio test: MAC Dataset.** The mac study was a randomized clinical trial to study the effects of combination regimens on prevention of MAC (*mycobacterium avium complex*), one of the most common OIs in AIDS patients. There were 3 regimens

- clarithromycin (new)
- rifabutin (standard)
- clarithromycin plus rifabutin

Here, we are interested in the time to MAC disease and not in time to death.

(i) Import the data into `R`. Have a look at the dataset. Fit a Cox PH model focusing on the effects of the Karnofsky score and treatment type on time to MAC disease. What is the hazard ratio of the Karnofsky score status? What is the interpretation of this hazard ratio?

(ii) Construct a 95%CI for the estimated hazard ratio in (i). Interpret your result.

(iii) Test the effect of the Karnofsky score using a Wald test. State your null and alternative hypothesis. What do you conclude?

(iv) Try to improve the model fit by adding the CD4 cell count as a covariate. Calculate the appropriate likelihood ratio test. What do you conclude from this result?

(v) Conduct an overall test of the treatment effect adjusted for the Karnofsky score and CD4 levels, using a multivariate Wald test.

(vi) Test whether there is a difference between the rifabutin and clarithromycin treatment arms after adjusting for the Karnofsky score and CD4 count.

(b) **Survival Function and Predicted Medians: Nursing Home Data.** We are going to consider again the dataset *nurshome.csv*.

(i) Import the data into `R`. Fit a Cox PH model focusing on the effects of marital and health status on length of stay using the `coxph` function.

(ii) Calculate the median length of stay for the following groups: (1) Single and healthy, (2) Single and unhealthy, (3) Married and healthy, and (4) Married and unhealthy, using a KM approach.

(iii) Calculate the medians again after taking advantage of the Cox model you fitted in (i). The `survfit` function would be very useful for doing this. Which options are available for estimating the baseline survival function? Which estimator did you actually use? Which subgroup has the longest length of stay?

(iv) **Optional:** Write your own `R` code to estimate the baseline survival function using the *Breslow* estimator of the baseline cumulative hazard. Verify the results using the function `survfit` with the appropriate option. Are the results identical?