

# Assignment 2

(Due: 2021/04/18, 11:59pm)

Note:

- No late assignment is accepted;
- Write your assignment in Chinese or English.

## Analysis of ACTG175 Data

Note: `cens` is binary variable  $\delta_i$ : 1 indicates event occurrence; 0 indicates censoring; `days` is the observed time  $T_i$ . We set significance level  $\alpha = 0.05$ .

### Questions:

1. (a) Run a Cox proportional hazards model (named Model 1), with `arms`, `cd40`, `cd80`, `age`, `wtkg`, `hemo`, `homo`, `drug`, `karnof`, `race`, `gender` and `symptom` as the explanatory variables (i.e., risk factors). In this Cox model, you need to treat `arms` as a factor, so you need to replace `arms` with `factor(arms)`, or deal with it using dummy variables directly. Which risk factors have statistically significant effects on survival?  
  
(b) What is the hazard ratio for a patient receiving the zidovudine and zalcitabine therapy compared to one receiving the zidovudine only therapy, adjusted for all other risk factors `cd40`, `cd80`, `age`, `wtkg`, `hemo`, `homo`, `drug`, `karnof`, `race`, `gender` and `symptom`? What is the 95% confidence interval of this hazard ratio? Based on the point estimate and the 95% confidence interval of this hazard ratio, is the the zidovudine and zalcitabine therapy better than the zidovudine only therapy on improving the patients' survival experience?  
  
(c) Check the proportional hazards assumptions of `cd40` and `gender` using the log-log plot, the goodness of fit test, and the Schoenfeld residual plot, respectively. Do these two variables violate the proportional hazard assumptions?

- (d) We suspect that **age** may have a nonlinear, rather than linear, effect on the log hazard ratio. Add  $\text{age}^2$  and  $\text{age}^3$  into the Cox model, and test whether these higher order terms are statistically significant.
  - (e) We suspect that there may be differential patterns of the treatment effects in the male and female groups. Add an interaction term of the treatments and **gender** to Model 1, and test whether the interaction is significant.
2. Suppose a newly HIV-infected patient is seeking for a therapy, who has the following characteristics (measured when he was admitted to the hospital): cd4 cell count equal to 400, cd8 cell count equal to 500, 25 years old, 70 kg, free of hemophilia, no history of homosexual activity, with a history of intravenous drug use, Karnofsky score equal to 90, non-white, male, and asymptomatic.

The doctor would like to know which out of the four treatments is most effective on prolonging the survival time for this patient, and will recommend this effective therapy to this patient. You may accomplish this task by drawing a Cox-adjusted survival plot using Model 1 for this subject if he had received each of the four treatments. Therefore, there will be four curves in this plot.

- (a) Based on this survival curve plot, which treatment do you recommend to the doctor if the goal is prolonging the survival experience?
- (b) What is the probability that this patient will survive more than two years if he had received each of these four treatments, respectively? Which therapy corresponds to the largest two-year survival for this patient?
- (c) Calculate the four risk scores for this patient if he had received each of the four treatments. For each of the four treatments, is his corresponding risk score high, low or normal compared to those in the same treatment arm in the ACTG175 Study (therefore, you need to make four separate comparisons)?