Tutorial 4: more on the Cox proportional hazards model in R

Download the datafile actg320.txt from the following url:

http://blog.nus.edu.sg/alexcook/files/2010/12/actg320.txt

This contains data on the occurrence of AIDS or death in HIV positive patients, from the AIDS clinical trials group (see Hosmer et al., 2008, for details). The columns in this dataset are, from left to right:

- patient identification
- time to AIDS or death (in days)
- censoring variable (1 if AIDS or death, 0 if right-censored)
- time of death
- censoring variable for death (1 if death, 0 if AIDS or right-censored)
- treatment (1 if IDV, 0 if control)
- treatment group (can be ignored: there were various sub-treatments that were applied)
- CD4 stratum at screening (0 if CD4 ≤ 50, 1 if CD4 > 50)
- sex (1 if male, 2 if female)
- “race” (1 if white and non-Hispanic, 2 if black and non-Hispanic, 3 if Hispanic, 4 if Asian or Pacific islander, 5 if native American, 6 if other)
- IV drug use history (1 if never, 2 if currently, 3 if previously)
- hæmophiliac (1 if yes, 0 if no)
- Karnofsky performance scale (100 indicates no evidence of disease, 90 minor symptoms, 80 some symptoms, 70 active work impossible)
• CD4 count (from multiple measurements) in cells ml\(^{-1}\)
• months of prior use of ZDV (another drug)
• age at enrollment.

For the following questions, consider death or onset of AIDS symptoms to be the event of interest.

[1] Use the Kaplan–Meier estimate to plot the estimated survival function against time for treatment (IDV versus control). Use a Cox proportional hazards model to do the same. Why are the two plots different? What does this tell us about the Cox PHM?

[2] Concentrating on “blacks” and “Hispanics”, test whether these two races have the same survival function using the Cox PHM framework. What is the maximum partial likelihood estimate of the hazard ratio comparing “blacks” to “Hispanics”? Construct a 95% confidence interval for this ratio.

[3] Do any of the following variables appear to influence survival individually: treatment, CD4 stratum or count, sex, IV drug use, haemophilia, severity of disease (indicated by the Karnofsky score), prior use of ZDV or age at enrollment? Using covariates that have coefficients significantly different from zero when fitted individually to construct a further model with multiple covariates. Test if there are interactions between these covariates.