SISCER Mod 12 Survival Analysis

Problem Set 2
Due: July 23, 2021

1. Consider the data from a two-arm clinical trial:

   - Control arm: 2, 3, 4+, 6, 12+, 18+, 20
   - Treatment arm: 7+, 12, 17+, 25, 29+, 30+

   Compute manually the log-rank test and provide a two-sided $p-$value. Do you reject the null hypothesis that there is no difference between these two arms?

2. In a two-arm randomized clinical trial, assume that failure times of the control group, T_0, follow an exponential distribution with density function of

$$f_0(t) = \theta e^{-\theta t}, \text{ for } t \geq 0.$$

   Let T_1 be the failure times of the treatment group. Consider two models to study a treatment effect between the treatment arms:

   - Cox proportional hazards model: $\lambda_1(t) = \lambda_0(t)e^{\beta}$, where $\lambda_1(t)$ and $\lambda_0(t)$ are the hazard functions for the treatment group and the control group, respectively, and β is the log hazards ratio;
   - Accelerate failure time model: $T_1 = T_0 e^{\alpha}$, where α is the log average time ratio.

   (a) Show that $\alpha + \beta = 0$ if both model are correct for T_0 and T_1.

   (b) In addition to the exponential distribution, are there any other distribution(s) for T_1 and T_0 under which both of these models are correct?

3. Analysis of the PBC data by the Cox proportional hazards model. Attach your computer outputs as appendix if you think they are helpful.

   (a) (Assessment of treatment efficacy) Apply the Cox proportional hazards model to estimate the hazards ratio between the D-penicillamine (DPCA) and the placebo. State your model and provide your model estimates. Based on your estimates, justify if DPCA significantly improves patients' survival.

(b) (Study of natural history) Apply the Cox proportional hazards model to study the association of patient's survival with following prognostic factors: Age, Albumin, Alkaline Phosphatase, Ascites, Bilirubin, Edema, Hepatomegaly, Platelets, Prothrombin Time, Sex and Spiders. Report your regression coefficients, standard errors and $p-$values. What would you conclude based on your estimates?