

Mei-Cheng Wang
Survival Analysis
Biostatistics 140.641

Problem Set 2

— Due Feb 23, 2006.

1. (Use of computing packages is not allowed) The following data are the remission times of 42 patients with acute leukemia: 6MP (mercaptopyurine) vs placebo. Reference: Freireich et al. (1963) Blood, p699.

Placebo: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

6MP: 6, 6, 6, 7, 10, 13, 16, 22, 23, 6⁺, 9⁺, 10⁺, 11⁺, 17⁺, 19⁺, 20⁺, 25⁺, 32⁺, 32⁺, 34⁺, 35⁺.

(a) Calculate both the Kaplan-Meier estimate and the empirical survival distribution on the basis of the placebo data. What is the relationship between the Kaplan-Meier estimate and the empirical survival distribution from your calculation?

(b) Do you observe a similar phenomenon which also holds for other uncensored survival data? Explain.

(c) Calculate the Kaplan-Meier estimate on the basis of the 6MP data.

(d) Use the variance estimates that you learn from Biostatistics 140.641 (or other courses) to produce a 95% confidence interval at each observed failure time for the empirical survival distribution on the basis of the placebo data. Comment on your results.

(e) Use the variance estimates that you learn from Biostatistics 140.641 to produce a 95% confidence interval, at each uncensored time, for the the Kaplan-Meier estimate on the basis of the 6MP data. Comment on

your results.

(f) Compare the survival function estimates from the placebo and 6MP groups. Interpret your results.

2. Answer this question based on the leukemia data given in Problem 1.

(a) Assume the remission times from the above two groups are associated through the proportional hazards model. Give the 2-sample proportional hazards model and your interpretation of the model. Give the constraints of the model, if any.

(b) Confirm the validity of the proportional hazards model by an exploratory method.

3. Assume the failure time T_0 has the exponential density $f_0(t) = \lambda e^{-\lambda t}$, $\lambda > 0$

(a) Identify the hazard function of T_0 .

(b) Verify that for a positive constant c , the random variable cT_0 has the hazard function λ/c . Interpret.

(c) Let $h(\cdot)$ denote a hazard function, and let T denote a failure time variable. Consider model I: $h(t; x) = \lambda e^{\beta x}$, and model II: $T = T_0 e^{\alpha x}$, where x is the covariate, and β and α are parameters. Identify the relationship between these two regression models.

(d) Interpret Models I and II. What do you conclude from (c)?

4. Familiarize yourself with a statistical package (SAS, STATA, S-plus, SPSS, etc.) for a coming problem set. You will be asked to use the Kaplan-Meier estimate, the two-sample test statistics, and the proportional hazards model for data analyses. The problem set will be due at the end of this quarter.