Survival Data Analysis (BIOS 7210)
Breheny

Assignment 4
Due Thursday, September 28

1. Patients with renal insufficiency require a procedure called dialysis, which is essentially an artificial replacement for their lost kidney function. One method of dialysis is known as peritoneal dialysis, which involves connecting a tube containing sterile solution into the peritoneum, thereby allowing diffusion to gradually remove waste products from the body.

   The tube can be placed using two methods, surgically or percutaneously (i.e., using a needle puncture). One complication that can arise with peritoneal dialysis is infection of site where the tube enters and exits the body (this known as the exit site). The course website contains data `Nahman1992.txt` from a study of whether the time until (first) exit-site infection depends on the method by which the dialysis tube was placed.

   (a) Fit Kaplan-Meier curves to each group and plot the resulting curves. Quietly ponder whether the difference looks significant or not.

   (b) Carry out a log-rank test of whether there is a significant difference between the two methods in terms of the time to exit-site infection. Comment on both the p-value and the difference(s) between observed and expected infections.

   (c) Would you expect a Peto-Prentice test to be more significant, or less significant, than the log-rank test in this scenario?

   (d) Carry out a Peto-Prentice test and comment on how it it differs (again, with respect to both the p-value and the observed/expected infections) from what you saw in (b).

   (e) Consider weighting the failure times by $\hat{S}(t)^{-1}$. Comment on the how this test would differ from the log-rank and Peto-Prentice tests.

   (f) Carry out the test in (e) for the kidney dialysis data; comment briefly on the result you obtain as well as how and why it differs from the results in (b) and (d).

2. Show that, in the absence of censoring due to other reasons, $\sum_i t_i = \sum_j u_j$, where t_i is the time on study for individual i and u_j is the jth normalized spacing: $u_j = (n-j+1)(T_{(j)} - T_{(j-1)})$. Here, $T_{(j)}$ is the jth ordered failure time, and we assume that all failure times are distinct.

3. Under the setup we introduced in class, where $Y(t)$ is the expected number of patient-years accumulated by time t and a denotes the average accrual rate per unit of time,

$$\mathbb{E}Y(t) = a \int_0^{t^*} \int_0^{t-v} S(u)\, du\, dv,$$

   where $t^* = \min(T, t)$. Show that for the exponential distribution,

$$\mathbb{E}Y(t) = \frac{at^*}{\lambda}\left\{1 - (\lambda t^*)^{-1} e^{-\lambda t}(e^{\lambda t^*} - 1)\right\}$$

4. Suppose that a random time-to-event T has the hazard function

$$\lambda(t) = \frac{\lambda}{\sigma} e^{t/\sigma}$$

for $t > 0$. This type of distribution is known as an *extreme value distribution.*

   (a) Derive the cumulative hazard, survival, and density functions for the extreme value distribution given above.

   (b) Let $X \sim \text{Exp}(\lambda)$. Consider the transformation $Y = \sigma \log(X + 1)$. Show that Y follows an extreme value distribution.

5. Suppose we are planning a randomized trial involving a new experimental treatment which the investigators believe will decrease the hazard of liver failure by 25% compared with the standard treatment.

   (a) Using the sample size formula we derived in class, how many subjects will the investigators need to enroll in each arm of the trial in order to achieve 80% power?

   (b) Our derivation of the formula in (a) assumed an exponential distribution, but I claimed that it was pretty accurate as long as the hazard ratio was constant. Suppose that the failure times in each group do not follow an exponential distribution, but instead follow an extreme value distribution with $\sigma = 1$ (note that the hazard ratio is still constant in this case). Carry out a simulation to estimate the empirical power of the log-rank test in this case: for 1,000 independent data sets, generate n observations (all of which are uncensored) for each group from the extreme value distribution and carry out a log-rank test, where n is the sample size from (a). What is the actual power? How accurate is the approximation we derived using the exponential distribution?

   (c) Repeat the simulation in b, but this time assume type 2 censoring. Specifically, suppose that $2n$ patients are enrolled in each arm, but that follow-up in each arm is discontinued after the first n subjects experience liver failure. Again, report the empirical power and comment on the accuracy of the sample size approach we worked out in class.

6. One of the earliest randomized clinical trials of a chemotherapuetic agent was conducted across 11 American hospitals, and involved children with acute leukemia that experienced a complete or partial remission (i.e, all/most of the signs of the disease had disappeared) of their leukemia following treatment with the drug prednisone. The trial was conducted by matching pairs of patients at a given hospital by remission status (complete or partial) and randomizing within the pair to either a 6-MP or placebo. Patients were followed until their leukemia returned (relapse) or until the end of the study. This data is available on the course website (`Freireich1963`).

   Another use of the stratified log-rank test we derived in class is that it can be applied to paired data like this. Analyze this data using a paired log-rank test. Provide the expected/observed counts, the p-value, and a brief conclusion regarding whether the drug is effective or not.