

Lab 15: Survival Analysis and Final Review

Survival Analysis

In lecture this week we got a brief overview of survival analysis. Today's lab will focus on Kaplan-Meier survival curves and the log-rank test. We will examine these using R on the aplastic anemia data set. The data set contains five variables:

- . Trt: Whether the patient received Methotrexate (MTX) or Methotrexate and cyclosporine (MTX+CSP).
- . Time_GVHD: Time until graft-versus-host disease. Measured in days.
- . Status_GVHD: What happened at the end of Time_GVHD. The patient was either censored (0) or got graft-versus-host disease (1).
- . Time: Time until death. Measured in days.
- . Status: What happened at the end of Time. The patient was either censored (0) or died (1).

Two common endpoints in survival analysis are Overall Survival and Progression-Free Survival. In this dataset, Overall Survival is time since randomization until death and Progression-Free Survival is time since randomization until GVHD.

Kaplan-Meier Estimates A **survival function** is a function of time, and is defined as the probability of the event in question not occurring by time t (i.e., the patient surviving until time t or later).

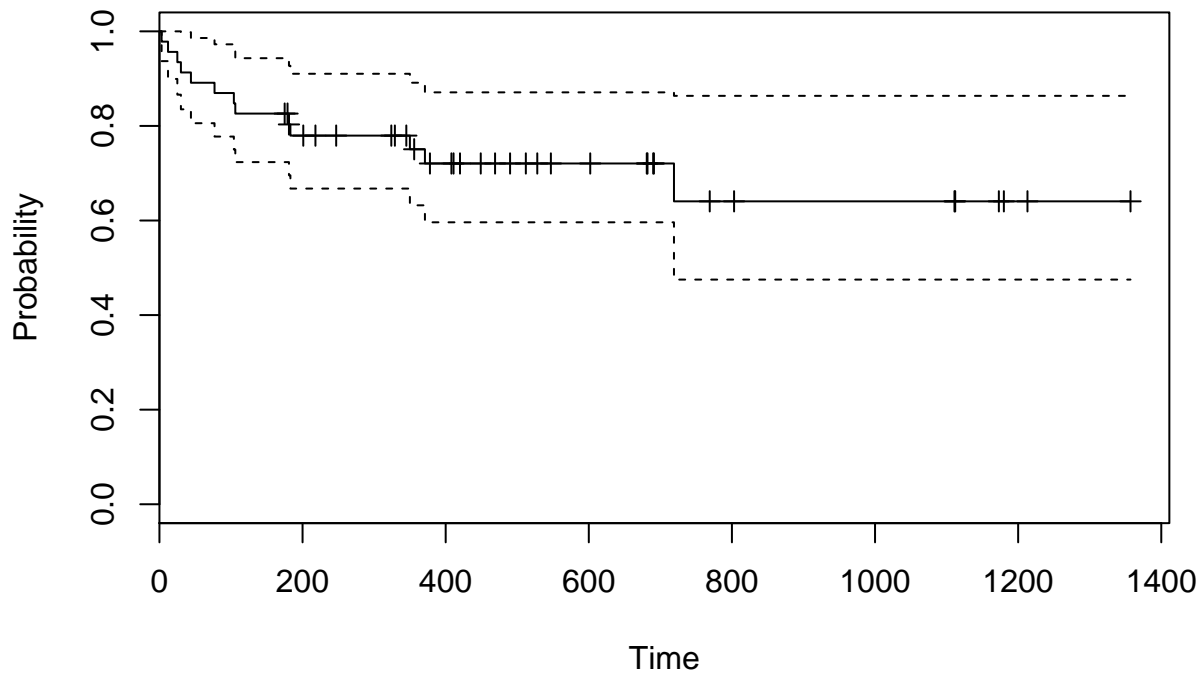
Ex: $S(10) = .95$ means that there is only a 5% chance of developing GVHD by day 10 (or equivalently, a 95% chance of surviving GVHD-free until day 10).

The most popular way to estimate survival functions is using Kaplan-Meier estimates. To estimate the Overall Survival function, do the following:

```
library(survival)
```

```
## Loading required package: splines
```

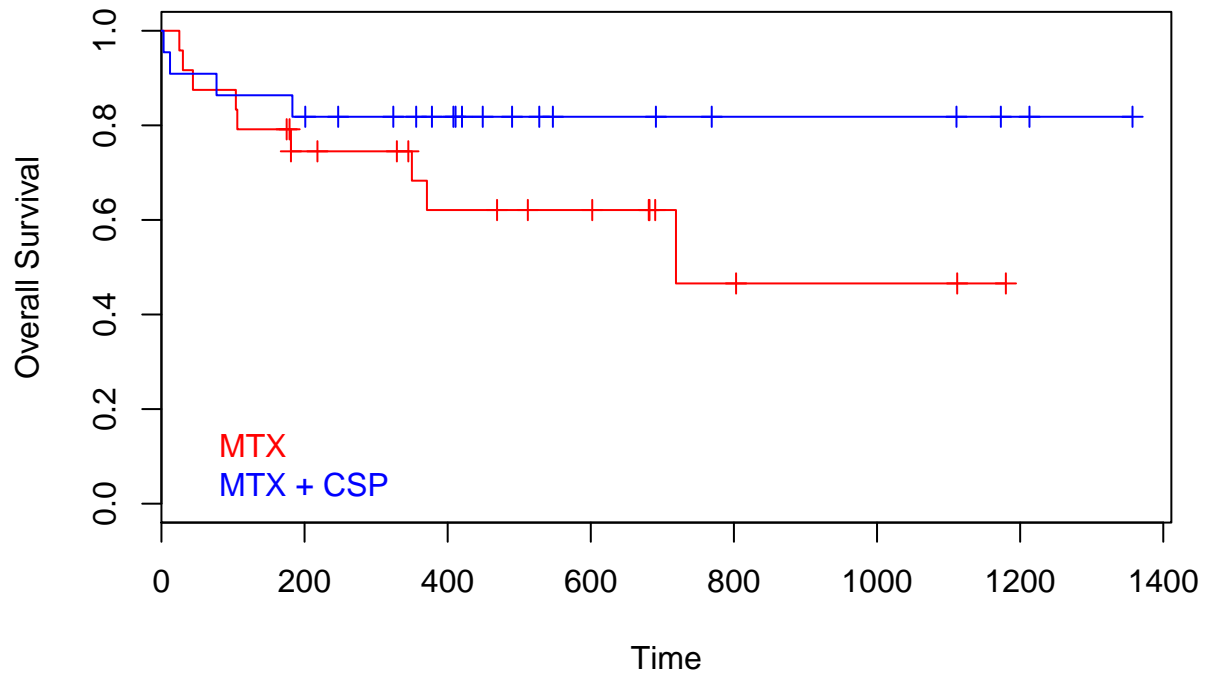
```
anemia <- read.delim("http://myweb.uiowa.edu/pbreheny/data/anemia.txt")
S <- with(anemia, Surv(Time, Status!=0))
fit <- survfit(S~1)
plot(fit, ylab = "Probability", xlab = "Time")
```



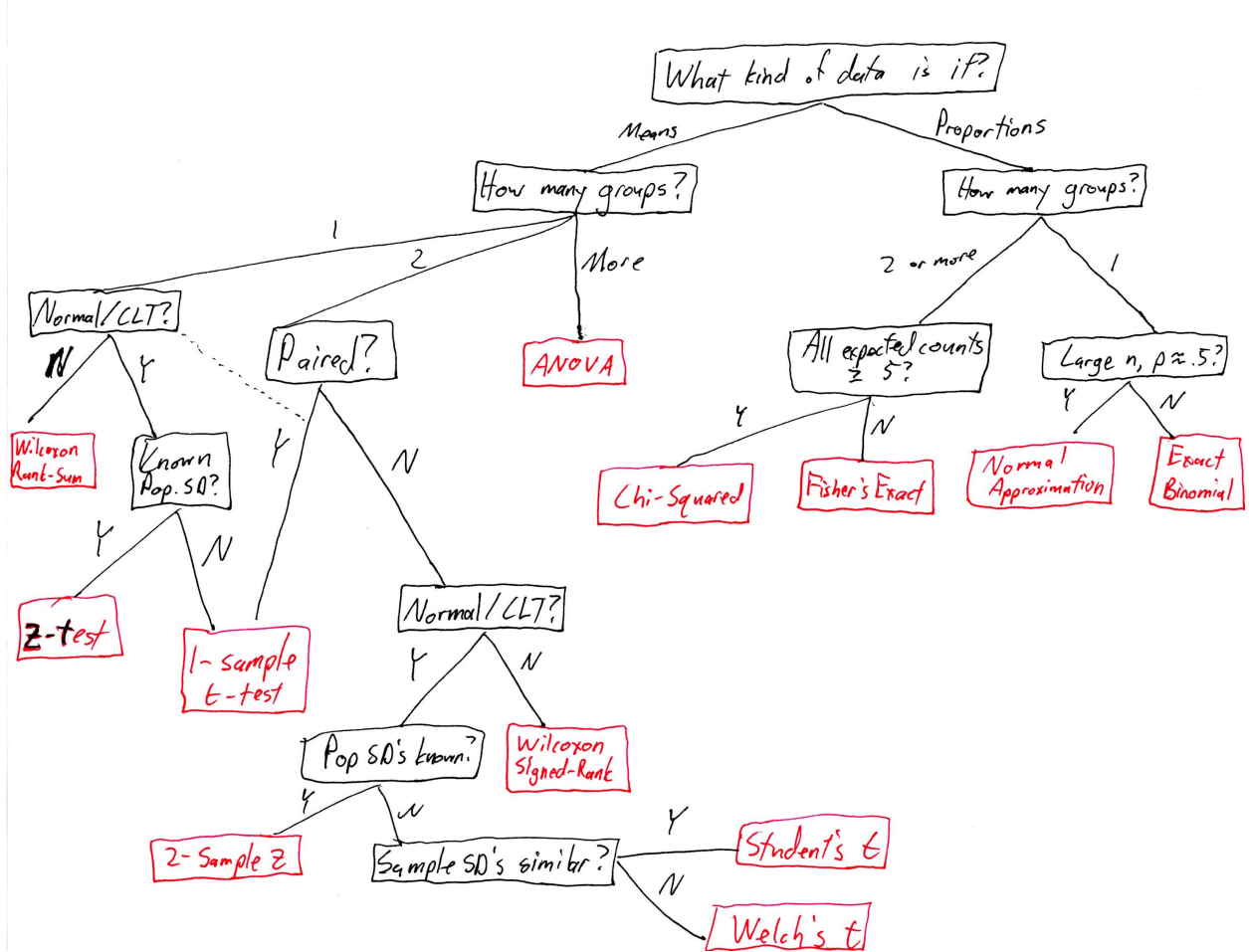
This is the Kaplan-Meier survival function estimate of the survival function, ignoring the different treatment groups. Pay particular attention to the confidence intervals. Why do they seem to get wider as time progresses? What do the tick marks represent? What is the median survival time?

We can also stratify by treatment group and examine both survival estimates.

```
fit2 <- with(anemia, survfit(S~Trt))
plot(fit2, ylab = "Overall Survival", xlab = "Time", col =
c("red", "blue"))
legend("bottomleft", c("MTX", "MTX + CSP"), text.col = c("red", "blue"),
bty = "n")
```



Final Review



For each problem, identify the type of study, what kind of test you would perform, and potential biases.

Example 1:

A team from Yale School of Medicine took a look at 1,433 people diagnosed with intracranial meningioma, the most commonly diagnosed brain tumor in the United States. Researchers compared these patients to a test group of 1,350 people without tumors. Participants offered self-reported lifetime dental X-ray histories. Researchers then analyzed the different types of X-rays these two groups had undergone. Patients with tumors were more than twice as likely to have had “bitewing” X-rays at least once per year. Bitewings, in which a patient bites down on X-ray film, take photos of the upper and lower back teeth.

Example 2:

In a study of 16 overweight young adults in India, participants were given, in turns, a dose of an extract made from unroasted coffee beans and a placebo, three times a day over 22 weeks. Their diet throughout the study was unchanged, and they were physically active. Between trials, the participants were given a two-week break for their bodies to reset. Though a few participants given the extract only lost 7 pounds, others lost as much as 26 pounds. On average, the subjects lost 17.5 pounds each, and reduced their body weight by 10.5 percent. Body fat also declined by 16 percent, even though the participants were eating an average of 2,400 calories and burning roughly 400.

Example 3:

Researchers at the University of College London surveyed nearly 8,000 participants over the age of 52. Using a fake aspirin bottle complete with instructions as the testing instrument, researchers asked participants to answer four basic questions, including “What is the maximum number of days you may take this medicine?” and “List three situations for which you should consult a doctor.” All the answers could be found on the label. One third of the adults failed to correctly answer all four questions, and one in eight got two or more wrong. Researchers then monitored the volunteers’ health for five years. During that time, 621 of the participants died, and people who missed two or more questions were more than twice as likely to have died than those who got the answers correct.

Example 4:

An experiment conducted by researchers at the Mayo Clinic took a close look at the sleeping and eating habits of 17 people between the ages of 18 and 40. The researchers put the volunteers in a Big Brother-style house, and monitored how much they slept, what they ate, and the physical activities they engaged in. For the first three nights the participants slept as much as they wanted. (They averaged about 6.5 hours per night.) Then they were split into two groups: Nine kept their normal sleep patterns, while the other eight were only allowed to sleep for 5 hours and 10 minutes a night. This schedule continued for eight nights. The leaders of the study found that with food readily available, those who slept less ate, on average, 549 more calories per day, or a little more than they’d get eating a McDonald’s Big Mac. And those who kept a normal sleep schedule actually ate 143 fewer calories per day, on average.

Solutions Example 1: Case-control (retrospective), chi-square or Fisher's Exact

Example 2: Crossover (prospective), paired t-test

Example 3: Prospective cohort, chi-square or Fisher's Exact

Example 4: Randomized (assumed) controlled clinical trial, 2-sample t-test (either Student or Welch, depending on data)