BIOS 4120: Introduction to Biostatistics
Breheny

Lab #15

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.

1. Kaplan-Meier Estimates

Survival function: *def.* 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, this 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, we will do:

```
require(survival)
S <- 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.

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

The log-rank test performs a test on the null hypothesis of "All the survival curves are not different from each other." The pvalue from the log-rank test is p = 0.16. What is the one year survival rate for both groups? What do we conclude?

2. Final Review

We have learned many methods of analyzing data of all types. But how do we know which methods are appropriate to use in certain situations? You will most likely come across scenarios where you are reviewing an analysis or performing one yourself. It is critical you choose an appropriate method and draw accurate conclusions. This will happen on the final as you will have to identify the type of problem and which analysis to perform. We will go over three examples to practice identifying the type of analysis, type of study, and what the outcome of interest is.

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 from Penn State found that increasing the amount of spices in your diet may lower the level of potentially harmful fat in your bloodstream. The experiment compared two groups of healthy, overweight men. One group ate meals seasoned with the special spice blend; the other ate the same meals prepared without the spices. Men who ate the spicy food saw a decrease of one-third in the level of triglycerides (a type of fat linked to heart disease) in their bloodstreams, and 20 percent lower insulin levels overall — even when the meals were high in fat and made with heavy oils.

****

Making a concise sheet for the final with formulas, about notes when to use them, and assumptions the formula makes about the underlying data, will be EXTREMELY useful. Reviewing all quizzes, summaries at the end of lectures, and procedure slides will also benefit you as well. The rest of the lab is open to questions until we do TA evaluations.