What is survival analysis and where can it be used?

Survival analysis is more than just death rates.

The methods for survival analysis were developed to handle the complexities of mortality studies, but they can be used for so much more.

You can study the “death” of mechanical devices, though the term “failure” is probably a better word to use for something that was never truly alive.

You can also study other health related events like relapse or rehospitalization. The events do not even need to be events that you’d like to avoid. Survival models are used to model the time to pregnancy for couples treated for fertility problems.

There are two features of survival models.

First is the process of measuring the time in a sample of people, animals, or machines until a specific event occurs. In fact, many people use the term “time to event analysis” or “event history analysis” instead of “survival analysis” to emphasize the broad range of areas where you can apply these techniques.

Second, is the recognition that not everyone/everything in your sample will experience the event. Those not experiencing the event, either because the study ended before they had the event or because they were lost to follow-up are classified as censored observations.

Censoring in time-to-event data and why you have to worry about it

One of the hallmarks of survival analysis is censoring. You are measuring the time until a certain event occurs in a sample of people, animals, or machines, and some of those in your sample never experience the event, at least not while you were studying them.

Consider a hypothetical experiment involving the survival times of a sample of 25 fruit flies. You wathc these flies daily and whenever a fly drops to the bottom of the cage, you give it a proper burial and record the number of days it was alive.

Suppose that you’ve done this for 15 of the flies, but on day 70 of the experiment, you carelessly leave the cage open and the 10 flies who are still alive bug out.

You might think that your experiment is ruined, but not so fast. You can still estimate the median survival time, because the median fly (#13) died before your gaffe occurred. Just be careful about this. The censored observations are not missing observations. You know something about these ten flies. They were the senior citizens in your sample and last longer than most of the other flies. You don’t ignore this information because ignoring information about the toughest ten flies in your sample seriously biases your outcome.

How do you handle censored observations? You use the censored observations to estimate survival probabilities up to the time of the great escape. Since these flies were alive through day seventy, they contribute to the denominator of the survival probabilty over time up to day seventy itself. After day seventy, they no longer contribute information about survival.

So censored observations represent data that is only partially missing and they can and should be incorporated into your statistical analysis.