2 Kaplan–Meier Survival Curves and the Log–Rank Test

■ Contents

Introduction

We begin with a brief review of the purposes of survival analysis, basic notation and terminology, and the basic data layout for the computer.

We then describe how to estimate and graph survival curves using the **Kaplan–Meier (KM)** method. The estimated survival probabilities are computed using a **product limit formula.**

Next, we describe how to compare two or more survival curves using the **log–rank test** of the null hypothesis of a common survival curve. For two groups, the log–rank statistic is based on the summed observed minus expected score for a given group and its variance estimate. For several groups, a computer should always be used because the log–rank formula is more complicated mathematically. The test statistic is approximately chi-square in large samples with G-1 degrees of freedom, where G denotes the number of groups being compared.

An alternative test is called the **Peto test,** which may be chosen if one wants to give more weight to the earlier part of the survival curves. This test is also a large sample chi-square test with G-1 degrees of freedom.

Abbreviated Outline

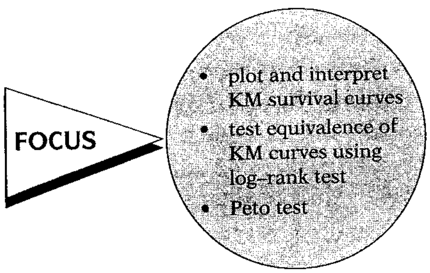
The outline below gives the user a preview of the material to be covered by the presentation. A detailed outline for review purposes follows the presentation.

Objectives

Upon completing the module, the learner should be able to:

1. Compute Kaplan–Meier (KM) probabilities of survival, given survival time and failure status information on a sample of subjects.
2. Interpret a graph of KM curves that compare two or more groups.
3. Draw conclusions as to whether or not two or more survival curves are the same based on computer results that provide a log–rank test.
4. Draw conclusions as to whether or not two or more survival curves are the same based on computer results that provide a Peto test.
5. Decide whether the log–rank test or the Peto test is more appropriate for a given set of survival data.

Presentation



This presentation describes how to plot and interpret survival data using Kaplan–Meier (KM) survival curves and how to test whether or not two or more KM curves are equivalent using the log–rank test. We also describe an alternative test called the Peto test.

I. Review

Start → TIME → Event

Event: death
disease
relapse

We begin by reviewing the basics of survival analysis. Generally, survival analysis is a collection of statistical procedures for the analysis of data in which the outcome variable of interest is **time until an event occurs.** By **event,** we mean death, disease incidence, relapse from remission, or any designated experience of interest that may happen to an individual.

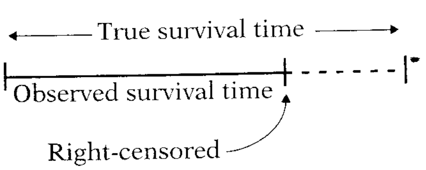
Time = survival time

Event = failure

When doing a survival analysis, we usually refer to the time variable as **survival time.** We also typically refer to the event as a **failure.**

Censoring: Don't know survival time exactly

Most survival analyses consider a key data analytical problem called **censoring.** In essence, censoring occurs when we have some information about individual survival time, but **we don't know the survival time exactly.**



Most survival time data is right-censored, because the true survival time interval, which we don't really know, has been cut off (i.e., censored) at the right side of the time interval, giving us the survival time actually observed. We want to use the observed survival time to draw implications about the true survival time.

Notation

T = survival time
↖ random variable
t = specific value for T

As notation, we denote by a **capital *T*** the random variable for a person's survival time. Next, we denote by a **small letter *t*** any specific value of interest for the variable T.