Paper PH09

KAPLAN-MEIER ANALYSIS: A PRACTICAL GUIDE FOR PROGRAMMERS

Madhi Saranadasa, Symbiance, Inc.

ABSTRACT

An important branch of statistics is survival analysis, which involves the modeling of time to event data. Within the context of clinical trials, this can represent the time between when a patient enrolls in a study and when a medically significant event occurs. Such analysis allows investigators to deduce, for example, the probability that an individual will survive past a certain time. A common problem in the analysis of clinical trials is how to appropriately consider censored data. The Kaplan-Meier (K-M) estimator (Kaplan, Meier 1985) of the survival function provides an elegant and robust method of survival analysis while properly handling censored data. Although it is common practice for SAS® programmers in the research community and pharmaceutical industry to implement the LIFETEST procedure to generate outputs, a comprehensive understanding of K-M survival analysis is required the appropriately interpret results. The objective of this presentation is to not only demonstrate the correct implementation of the LIFETEST procedure when studying survival data but also to describe the statistical fundamentals, the underlying calculations and the appropriate analytical tools so that the reader is well equipped to incorporate K-M analysis in their own research.

INTRODUCTION

Techniques that model the time until an event takes place are widely used in the medical, social and economic sciences. Known as survival analysis, this branch of statistics can answer questions such as: how long do people remain unemployed after job loss in different economic climates? Or, is the time it takes for a product to break different when originating from different manufacturing facilities? In the medical field a common application of survival analysis would be to determine whether or not a new drug aimed at treating a disease significantly improves the survival time of patients relative to a control. To answer such a question, it is initially tempting to employ standard parametric and non-parametric statistics for comparing the average survival time between the treatment and the control. However, how will patients who survive after the study ends be handled? Theoretically, they could survive an additional 20 days after the study or 20 years after the study, so it is not appropriate to assign any definite survival time to them to calculate an average. It is also very common in clinical trials to lose contact with patients during the study, obfuscating information about their survival. Especially when clinical trials are small, it is vital to incorporate any useful information about patients lost to follow-up in subsequent analysis. For example, while an investigator may not know how long a patient survived after losing contact, there is still useful information in the fact that the patient was enrolled in the study for six months before losing contact.

In this instance, rather than determining the average survival time, the appropriate analysis would be to calculate the probability that a patient will survive past a certain time. This is known as the survival function. As we will see, the Kaplan-Meier (K-M) estimator is an elegant method to compute the survival function while addressing the significant difficulties described above.

DERIVATION OF THE K-M ESTIMATE USING THE GEHAN DATA SET

The Gehan data set(Cox 1984) contains the length of remission in weeks for two groups of leukemia patients, treatment and control.Survival analysis traditionally considers death as the event of interest, however when using the Gehan data set we must consider the time at which remission ends. The goal of the K-M method is to derive a function that will tell us the likelihood that a patient will be “event-free” over a given period of time (this is typically called the “survival function”, regardless of application). In the case of the Gehan data set, that event will be the end of remission. The raw data for the amount of time each patient was in remission is presented below (note that * indicate patients who were lost due to follow-up, therefore we have no information regarding when or if remission ended, only that they continued to be in remission *at least* until the time denoted below):

Remission Time of each patient (weeks)

Treatment	Control
6	1
6	1
6	2
6*	2
7	3
9*	4
10	4
10*	5
11*	5
13	8
16	8
17*	8
19*	8
20*	11
22	11
23	12
25*	12
32*	15
32*	17
34*	22
35*	23

Let us first consider the control group to understand how the survival function is calculated without the need for censoring. Time to event data can be thought of as discrete events separated by intervals. In this case, the event would be whenever remission ends for a patient. The conditional survival probability, P_c, is the probability of surviving to a specific event, given that you survived all previous events. For a sample, this is simply the number patients in remission at an event divided by the number of patients surviving and available just prior to the event (otherwise known as the patients "at risk" of the event). This is often more easily calculated as 1 minus the conditional failure rate:

$$P_c = 1 - \frac{d_i}{n_i}$$

In this equation, d_i is the number of patients who ended remission at event *i*and n_i is the number of patients at risk (or in remission and enrolled in the study) just before event *i*. At the start of this study, the control group contains 21 at risk patients and the first event takes place at week 1. Here, remission ended for two patients. Therefore the conditional failure rate is 2/21 = 0.095 and the conditional survival rate is 1 – 0.095 = 0.905. The next event takes place at week 2, where again remission ended for 2 patients. However, because remission ended for two patients in the previous event they are no longer in the risk group. Thus the calculation of the conditional failure rate is 2 / (21 – 2) = 0.105 and the corresponding conditional survival rate is 1 – 0.105 = 0.895.

The calculations are similar for the treatment group, except patients who were lost to follow-up are now considered. In the K-M method, these patients do not represent end of remission events, but rather changes to the number of patients at risk. Consider the first event of the treatment group, which takes place at week 6. At this time, of the original 21 patients, remission ended for 3 of them and 1 was lost to follow-up. The conditional failure rate simply is 3/21 = 0.143. The next event takes place at week 7 where remission ended for 1 patient. In this instance, the number of patients at risk decreases not only by the patients for whom remission ended in the previous event, but also the patient lost to follow-up because they are no longer in the study. Therefore, the conditional failure rate is 1 / (21 – 3 – 1) = 0.058. In this way, the remission time of patients lost to follow-up is considered during their time in the study, but they are later censored after information about their clinical outcome is lost. The calculations for the first 4 events for the treatment group are shown below:

Week	# end remission	# lost	# at risk	Conditionalfailure	Conditional survival, P_c
6	3	1	21	3 / 21 = 0.142	0.8571
7	1	0	21 – 3 – 1 =17	1 / 17 = 0.058	0.942
9	0	1	17 – 1 – 0 = 16	0 / 16 = 0	1
10	1	1	16 – 0 – 1 = 15	1 / 15 = 0.067	0.9333

It is important to note that the previous calculations are for the *conditional* survival probability (P_c). This implicitly assumes that a patient has survived all events leading up to event i. In order to derive a survival function, we are interested in the probability that a patient cumulatively survived event i and all events prior to event i. This is known as the *unconditional* survival probability and, by the multiplication law of independent events, is calculated by multiply all conditional survival probabilities up to and including event *i*. The unconditional survival probability, P_u, is expressed as:

$$P_u = \prod_{j=1}^{i}\left(1 - \frac{d_j}{n_j}\right)$$

A summary of the calculations for the cumulative survival probability of the treatment group is provided below:

Week	# end remission	# lost	# at risk	Conditional Survival, P_c	Unconditional Survival, P_u
6	3	1	21	1-3/21 = .8571	.8571
7	1	0	17	1-1/17 = .9411	.8067
9	0	1	16	1-0 = 1	.8067
10	1	1	15	1-1/15 = .9333	.7529
11	0	1	13	1-0 = 1	.7529
13	1	0	12	1-1/12 = .9167	.6902
16	1	0	11	1-1/11 = .9090	.6275
17	0	1	10	1-0 = 1	.6275
19	0	1	9	1-0 = 1	.6275
20	0	1	8	1-0 = 1	.6275
22	1	0	7	1-1/7 = .8571	.5378
23	1	0	6	1-1/6 = .8333	.4482
25	0	1	5	1-0 = 1	.4482
32	0	2	4	1-0 = 1	.4482
34	0	1	2	1-0 = 1	.4482
35	0	1	1	1-0 = 1	.4482

P_u, also known as the K-M product limit estimate, is a function of time which gives the probability of survival from the start of the study to a specific time. It is this metric that we use to see if the survival difference between two groups is statistically significant.

THE K-M METHOD IN THE SAS ENVIRONMENT

While the previous hand calculations have been useful in understanding how K-M analysis calculates a survival function while also appropriately handling censored data, our next goal is to demonstrate how the SAS environment can programmatically perform such calculations. We will again be using the Gehan data set of remission times in leukemia patients. In its simplest form, using the LIFETEST procedure to perform K-M analysis requires essentially three pieces of information for each patient in a study. 1) Information regarding which treatment group the patient was placed in. 2) The amount of the time the patient was in the study. And finally, 3) whether or not the endpoint for the patient was due to the fact that the patient was lost to follow-up or that the event of interest occurred (in our Gehan example, this would signify the end of remission). This last element of the data is usually binary, in that '0' typically denotes occurrence of the event of interest while '1' denotes loss to follow-up. Below is an abbreviation of what K-M-formatted data typically looks like in SAS:

	GROUPID	TIME	CENSOR
1	1	6	0
2	1	6	0
3	1	6	0
4	1	6	1
5	2	1	0
6	2	1	0
7	2	2	0
8	2	2	0

Once survival data is appropriately formatted, using the LIFETEST procedure is relatively straightforward. Although a comprehensive overview of all the LIFETEST procedure options is beyond the scope of this discussion, there are key outputs and configurations that are fundamental to a statistical interpretation of the results. Presented below is sample code that will output the information required to correctly interpret results within the context of the previous statistical discussion:

```
proc lifetest data=GEHAN method=KM OUTSURV=SURV plots=(s);

   time TIME*CENSOR(1);

   ods output   productLimitEstimates = PLE
                CensoredSummary = CS
                homtests = STATS;

   strata GROUPID;

run;
```

In this example, GEHAN is the name of our appropriately formatted data set of survival information and our method will be KM for Kaplan-Meier analysis. Other key syntax includes the TIME statement, wherein the first argument denotes that variable that signifies time-until-event data and the second argument denotes the variable that signifies whether the record is censored or not. Importantly, one must set what value is defined as censored in the parentheses following the second argument. Here we are setting '1' to denote censored data. Finally, the STRATA statement is included to denote how the observations are grouped into different treatment strata.

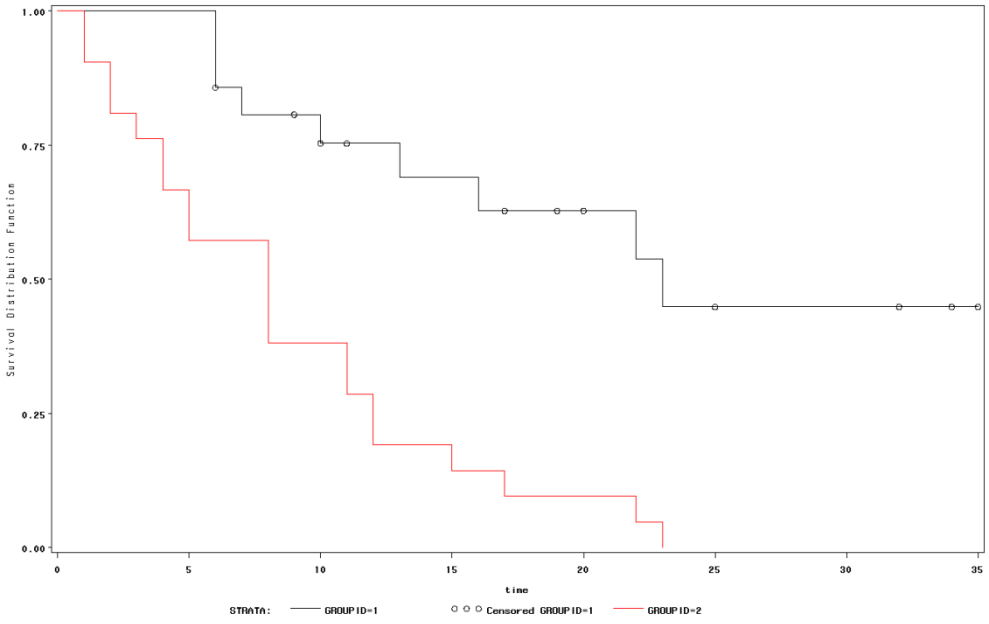
The remainder of the syntax presented are options that can be utilized for further statistical needs. The ODS output option CensoredSummary provides an at-a-glance summary of the number of patients that were censored and the number of patients that reached an endpoint due to the event of interest taking place (by default, SAS defines this event as 'failure'):

	Stratum Number	GROUPID	Total	Number Failed	Number Censored	Percent Censored
1	1	1	21	9	12	57.14
2	2	2	21	21	0	0.00
3	T	_	42	30	12	28.57

The OUTSURV statement and the ODS output option ProductLimitEstimates are two options to present the calculated survival function discussed earlier. While the ProductLimitEstimates output provides a patient-by-patient breakdown of the survival function, the OUTSURV output provides the time points where either endpoints or censoring occurred, and the survival function at these points as well as confidence intervals for the survival function. Presented below is the OUTSURV output for the Gehan data set, abbreviated to present only the treatment group:

	GROUPID	time	Censoring Flag: 0=Failed 1=Censored	Survival Distribution Function Estimate	SDF Lower 95.00% Confidence Limit	SDF Upper 95.00% Confidence Limit
1	1	0	.	1	1	1
2	1	6	0	0.8571428571	0.6197186381	0.9515516388
3	1	6	1	0.8571428571	.	.
4	1	7	0	0.8067226891	0.5631472111	0.922808871
5	1	9	1	0.8067226891	.	.
6	1	10	0	0.7529411765	0.5032001317	0.8893616478
7	1	10	1	0.7529411765	.	.
8	1	11	1	0.7529411765	.	.
9	1	13	0	0.6901960784	0.4316108256	0.8490657323
10	1	16	0	0.6274509804	0.3675114277	0.804911918
11	1	17	1	0.6274509804	.	.
12	1	19	1	0.6274509804	.	.
13	1	20	1	0.6274509804	.	.
14	1	22	0	0.5378151261	0.2677794772	0.7467903837
15	1	23	0	0.4481792717	0.1880524745	0.6801422376
16	1	25	1	.	.	.
17	1	32	1	.	.	.
18	1	32	1	.	.	.
19	1	34	1	.	.	.
20	1	35	1	.	.	.

Finally, the plots = (s) statement outputs the survival curves for the data set:



The survival curve is simply a depiction of how the survival function decreases over time. In the case of the Gehan data set, both treatment groups start at 100% survival because patient have yet to either reach an endpoint or be censored out of the study. Each 'step' in the survival curve represents a timepoint where a patient reached an endpoint in the study whereas each open circle is a point where a patient was censored from the study.Recall that the survival probability is 1 – (the number of patients lost divided by the number of patients at risk just before the event). Remember that at week six in treatment group, three patients reached the endpoint because their remission ended and one patient was censored from the study. This is reflected in the survival curve by a step-down in survival from 100% to 85.7% at week 6 and one open circle at the step.

The Log-Rank Test

A simple glance at the survival curves generated by the LIFETEST procedure suggests that there is a difference between the treatment and control groups of the Gehan data set. Judging by the separation in the two curves, patients appear to reach an endpoint (end of remission) relatively later and fewer patients seem to reach an endpoint at all in the treatment group relative to the control. From the CensoredSummary output, we know that a total of 21 patients were in each group. Therefore, are these apparent differences in the survival of patients between the groups due to the treatment introduced in the study or could it be chance variation due to the relatively small sample size? This question requires the introduction of a statistical test.

When censored observations are considered, a widely used method is the log-rank test. While a complete derivation of the log-rank test is beyond the scope of this discussion, the general motivations for the test are described here. This test equally weights the failure rates (this is also commonly referred to as the hazard rate) over time as the study progresses. For each 'step' in the survival curve, the expected number of failures is calculated assuming the populations are the same (in other words, the failure rates are the same). This expected value is compared to what was actually observed during the study. If the expected value varies significantly from the observed value, then we can say that the differences seen in the survival curve graph are not due to random chance, but rather because of the treatment given to the experimental group. For each step in the survival curve, that is, every time an event occurs, a conditional probability table is constructed to calculate the expected number of events. The table below represents how a typical conditional probability table would be constructed to calculate the expected number of failures assuming the failure rates are the same:

Treatment Group	Event Occurred	Event Did not Occur	Total patients
Treatment	d_1	$n_1 - d_1$	n_1
Control	d_2	$n_2 - d_2$	n_2
Overall	$d = d_1 + d_2$	$N - d$	$N = n_1 + n_2$

For each treatment group, *d* denotes the number of patients where an event occurred at this step and N is the total number of patients at risk at this step. Assuming a common failure rate between the two treatment groups, d_i is a hypergeometric random variable with an expected value $E(d_i)$ and variance $V(d_i)$ where:

$$E(d_i) = n_i \frac{d}{N}, \; V(d_i) = n_i \frac{d}{N} \frac{N-d}{N} \frac{N-n_i}{N-1}.$$

These tables are constructed and $E(d_i)$ and $V(d_i)$ are calculated for each step of the survival function and the overall sums are computed for one treatment group (you can choose any one of the groups). Specifically, $E = \Sigma E(d_i)$, $V = \Sigma V(d_i)$and the total observed events, O is the sum of all observed events over the course of the study, Σd_i.The log-rank Chi-squared test statistic is then calculated as:

$$X^2_{LR} = \frac{(O-E)^2}{V}.$$

Performing the calculations for each step in the Gehan data set and then calculating *O, E* and *V*, the chi-squared test statistic is calculated as

$$\frac{(9-19.251)^2}{6.257} = 16.79.$$

With 1 degree of freedom, this corresponds to a p-value of .00004. At an alpha level of 0.05 we can confidently reject the null hypothesis that the failure rates are the same between the treatment and control group and thus we can say that the difference seen in the survival curves is due to the treatment given in the study. Please note that the same test statistic is obtained if we instead considered other treatment group. The log-rank test can be easily implemented in the LIFETEST procedure using the ODS output option HomTests. The HomTests output for the Gehan data set is presented below:

	Test	Chi-Square	DF	Pr > Chi-Square
1	Log-Rank	16.7929	1	<.0001
2	Wilcoxon	13.4579	1	0.0002
3	-2Log(LR)	16.4852	1	<.0001

The first observation in the HomTests output displays the chi-squared value and p-value for the Log-Rank test as described above. As expected, the LIFETEST procedure correctly computes a chi-squared value of 16.79 and a p-value < .0001. Again, from this data we can reject the null hypothesis that the failure rate is the same between the treatment and control groups. We can therefore conclude that in the Gehan data set, the treatment given in the study significantly increases the survival time for patients.

CONCLUSION

Survival analysis methods are common in clinical trials and provide valuable information regarding the time it takes for death, relapse or treatment response. Because it is such an essential element of clinical analysis, programmers and investigators unfamiliar with statistically underpinnings of this analysis would still benefit greatly from a basic understanding of its methodology and implementation in SAS. Here, we have presented the computation behind the Kaplan-Meier Estimator and how the LIFETEST procedure can be used to generate the figures and statistics needed for Kaplan-Meier analysis. As demonstrated, the implementation is fairly straightforward and similarly the underlying methodology is fairly intuitive. Hopefully, the reader is left not only with confidence regarding using the LIFETEST procedure, but also with an understanding of the outputs and a rationale behind the calculations.

REFERENCES

Kaplan, E. L. and Meier, P. (1985) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association***53,** 457-481.

Cox, D. Roxbee, D. Oakes D. *Analysis of survival data*.Vol. 21.CRC Press, 1984.

CONTACT INFORMATION

Madhi Saranadasa
Symbiance, Inc.
231 Clarksville Road, Suite 1
Princeton, NJ 08550
E-Mail: msaranadasa@symbiance.com
www.symbiance.com