1. Introduction

Working alongside statisticians, statistical programmers can be required to produce or validate statistical outputs. Effective programmers understand the purpose of the outputs and the meaning of the results.

2. Kaplan-Meier

The Kaplan-Meier Estimate product limit is the most widely used method for representing time-to-event data and to estimate the survival over time:

\[
\hat{S}(t) = \frac{R(t)}{R_0(t)}
\]

where \( \hat{S}(t) \) = Survival Estimate

\( R(t) \) = number of subjects at the start

\( n_s \) = number of subjects experiencing event

**EXAMPLE by hand:** 10 subjects reviewed daily, all subjects followed up for 10 days (no censoring takes place, see (3) below):

<table>
<thead>
<tr>
<th>Time (days)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Censored?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

3. Censoring

Complete follow up of a subject is not always possible; a subject may withdraw from a trial early or before the event of interest has been observed. In survival analysis, this is called censoring. These subjects still provide useful information and should not be ignored as removing them could introduce a bias into the study data.

In clinical trials right censoring is most common. This is where we know that an event has not occurred by a certain time point (time of withdrawal), but we do not know what might have happened, had the subject been followed up for longer. We assume that censored subjects have the same risk of an event as those who continue to be followed up.

There could several reasons why a subject is censored:

- A study ends before the subject could experience the event, the subject withdrew from the study or lost to follow-up or the subject has another event that stops the event of interest from happening (e.g. dying before experiencing a different event of interest, known as a competing event).

**EXAMPLE by hand:** As above, but this time two subjects withdraw part way through at day 3 and 7.

4. Improvements

The plots option on PROC LIFETEST has limitations. Often, it is better to output the survival tables and then use PROC SGPLOT to customize the display. Using PROC LIFETEST in combination with ODS PRODUCTLIMITTESTESTIMATES gives a dataset of the Kaplan-Meier Estimates and from this we build the plot manually.

Using a PROC SGPlot and a SERIES statement, we can manually plot the Kaplan-Meier as survival vs. Time (Group).

**EXAMPLE in SAS:** 8,000 subjects randomized to Placebo or Treatment (trtp) for 3 years with a primary endpoint of death. Subjects can have event (die), complete (censored at 3 years) or withdraw (censor before 3 years).

**Survival Analysis**

Survival Analysis is commonly used statistical method in clinical trials that involves the investigation of time-to-event data. The key variable of interest is time and the event of interest must be a clearly measurable, predefined, discrete and recordable point, such as death, the first release of cancer or the failure of a component in engineering. This poster introduces the main concepts of survival analysis from a programming viewpoint.

**SAS® Proc Lifetest**

To produce Kaplan-Meier plots in SAS, a subject level dataset is needed containing a time variable (e.g. avail) and a censor variable (e.g. censored) which should be set to 0 for subjects who experienced an event and 1 for subjects who were censored. The PROC LIFETEST code below will output the plot to the right:

```sas
proc lifetest data=all output=call plots=(survival); time avail*censored(1); strata trtp; run;
```

Most clinical trials will compare treatments. To produce Kaplan-Meier plots comparing these, include a STRATA statement (see (4)).

**EXAMPLE in SAS:** As above plotted with PROC LIFETEST

**Covariates**

As with linear regression modelling, Cox PH models allows the use of covariates – variables that could also be associated with the outcome of interest. By including covariates our estimate of the treatment effect is improved.

In the example here we have two possible covariates: Age and Sex. However this data comes from a randomised clinical trial, so we would not expect to see much change in the hazard ratios.

**Covariates**

As with linear regression modelling, Cox PH models allows the use of covariates – variables that could also be associated with the outcome of interest.

5. Assumptions and Interpretation

- The Kaplan-Meier plot shows an estimate of the probability of surviving to a given time.
- A higher survival estimate suggests a longer time to death, not that the death is less likely to occur overall.
- Can represent as survival (going down) or time to event (going up) – however curve can only go in one direction.
- Assumes censored subjects have the same probability of survival as those who continue.
- Before any subjects are censored, the estimate is an exact representation of the data.
- More frequent follow-up of subjects increases accuracy of estimate.
- Due to censoring, the left hand side of the graph will have greater precision.
- Small subsets can be misleading due to high effect of a single event.
- Kaplan-Meier curves are limited in that multiple predictors cannot be taken into account. Theoretically, randomisation reduces this issue, however use of a Cox PH model (see (6)) with covariates may be preferred.

**Results**

Using the example from above, without Covariates, comparing Treatment against the reference group of Placebo, gives a hazard ratio of 0.441 (0.388, 0.502). This means that there is a 55% reduction in the risk of dying for subjects in the treatment group compared to the placebo group. With a p-value <0.0001 this can be considered significant.

When the covariates Age and Sex are included, however, there is little change in the hazard ratio. The reduction in risk is again 55% for subjects in the treatment group compared to the placebo group.

**SAS® Proc PHREG**

To compare Hazard Ratios between two groups, use:

```sas
proc phreg data=trial; class sex age; model time*status(0)=trtp / risklimits ri; output out=surv; run;
```

**EXAMPLE in SAS:** Placebo vs Treatment with covariates as sex and age.