

Generating Transition Probabilities from Your Survival Data

Question:

How can I derive Markov transition probabilities from Kaplan-Meier data?

Answer:

Kaplan-Meier data can be used as input data to a model. This example TreeAge Pro model shows two methods to derive transition probabilities from Kaplan Meier (KM) data.

Kaplan-Meier Data

- Kaplan-Meier data is typically collected from trial counts based on when events occur. In this example, we have collected the surviving proportion of the population at different times in months up to 46 months. The data can be reviewed in the Excel workbook - *Kaplan-Meier Survival Data.xlsx*.
- There are three columns:
 - Time in months
 - Percentage of cohort alive (Survival)
 - Percentage of cohort dead (= 1 – percentage alive).
- Data assumptions:
 - No duplicates – there is only one row for each time.
 - Equal time units – data is recorded at equally spaced time intervals. This may not be true of your real world data which may require some cleaning.
 - Time consistency – the same time units are used in the KM data and the model (months in this example).
 - Survival consistently decreases – survival in each month is always less than or equal to the previous month.

Methods to estimate Transition Probabilities

1. Represent the empirical data with a Parametric Distribution (**recommended**).
 - a. In an external software package, such as STATA, R or MatLab, analyze the empirical data to find a parametric distribution which represents the KM curve.
 - b. Create the appropriate parametric distribution type in TreeAge.
 - c. Use the TreeAge Pro function DistTransprob to calculate the transition probabilities based on the distribution.
2. Use Empirical data directly (**not recommended**).
 - a. Create a table in TreeAge representing the Cumulative Density Function (CDF).
 - b. Create a Table distribution using the CDF table.
 - c. Use the TreeAge Pro function DistTransprob to calculate the transition probabilities based on the distribution.

Generating Transition Probabilities from Your Survival Data

Method 1: Representing Survival with a Weibull Distribution

- In this method, we assume a regression analysis tool (such as STATA, SAS, R) identified a Weibull distribution (with parameters: shape 1.669511, scale 0.001457) as closely representing the empirical survival data.
 1. *This Weibull distribution is assumed to represent survival extrapolated beyond the data collected.*
 2. These estimations will have uncertainty, which is not discussed in this article.
- In a Markov model, we need the probability of an event occurring. The probability of death for each cycle is calculated from the underlying Weibull distribution using the DistTransProb function. In a DES model, the Weibull distribution could be sampled directly for time-to-death.
- Steps to implement in TreeAge, which are already complete in the TreeAge Pro model *Markov From Kaplan-Meier Data.trex*, are:
 1. Create parameter variables for the Weibull shape and scale.
 - `shape_Weibull = 1.669511`
 - `scale_Weibull = 0.001457`
 2. Create a Weibull distribution referencing the parameters above.
 - Distribution: Death_Weibull
 3. Create the probability of death from the Death_Weibull distribution using the DistTransProb function
 - `pDeath_Weibull = DistTransProb("Death_Weibull"; _stage; 1)`
 - The first argument references the distribution name in quotes.
 - The second argument, `_stage`, represents the start time in months for each Markov cycle.
 - The third argument, `1`, matches the time units in the distribution. In this case, the distribution and model have the same (monthly) units (i.e. `_stage` is equivalent to `1`).
 4. Enter the `pDeath_Weibull` variable as the probability of death in the Markov model.
- We will review the transition probabilities in the section Testing Results.

Generating Transition Probabilities from Your Survival Data

Method 2: Representing Survival with a Table Distribution

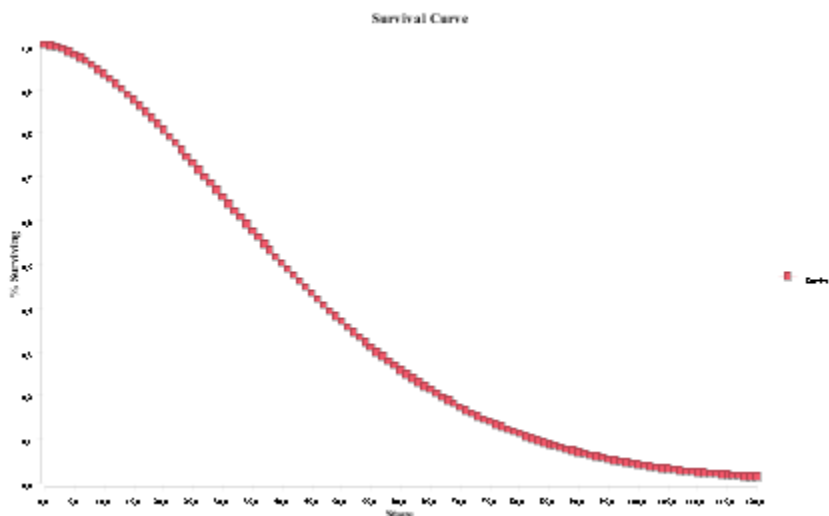
- In this method, we use the original survival data to generate transition probabilities via a table.
 1. The data can be reviewed in the Excel workbook - *Kaplan-Meier Survival Data.xlsx*. There are 3 columns with:
 - Time in months
 - Percentage of cohort alive (Survival)
 - Percentage of cohort dead (= 1 – percentage alive).
 2. *The data ends at 46 months, so to extend the survival beyond the data collected, we have assumed that everyone will be dead in 120 months.*
- The survival data is converted into a table which has the death percentage calculated as: $\text{Death} = 1 - \text{Survival}$.
- The death table is used to create a Table Distribution.
- The Table distribution is sampled for the probability of death for each cycle, calculated from the underlying distribution. The model uses DistTransProb function (like Method 1) to get the correct probability as `_stage` (the time) changes in the model.
- Steps to implement in TreeAge, which are already complete in the TreeAge Pro model: *Markov From Kaplan-Meier Data.trex*.
 1. Create a death table based on the cumulative percentage dead by month, as in the Excel worksheet.
 2. Add a row beneath the last observed entry of 46 months and probability 0.5089, with the data: 120 months and 100% dead (i.e. enter 1 in the table).
 3. Create a Table distribution referencing the death table.
 - Distribution: Death_Table
 4. Create the probability of death from the distribution Death_Table distribution using the DistTransProb function
 - `pDeath_Table = DistTransProb("Death_Table"; _stage; 1)`
 - The first argument references the distribution name in quotes.
 - The second argument, `_stage`, represents the start time in months for each Markov cycle.
 - The third argument, 1, represents the Markov cycle length in months.
 5. Enter the `pDeath_Table` variable as the probability of death in the Markov model.
- We will review the transition probabilities in the section Testing Results.

Generating Transition Probabilities from Your Survival Data

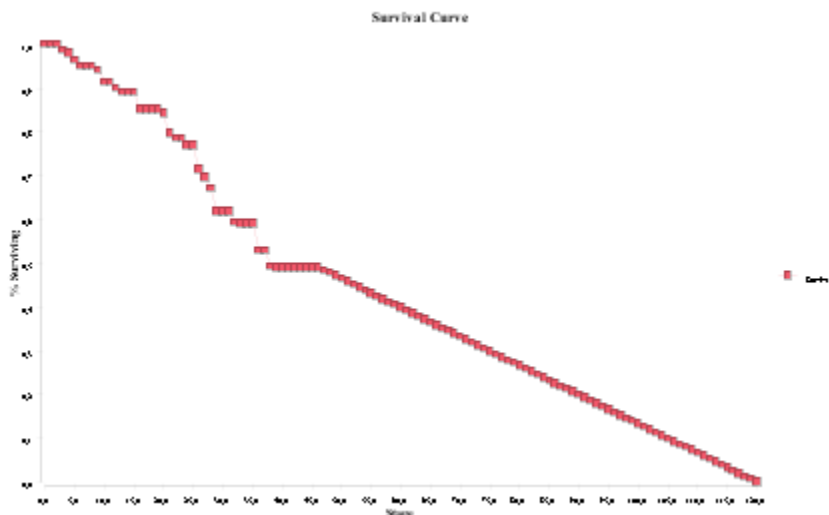
Testing Results

- It is critical to test your calculations against the data collected and the assumptions about that data.
 1. Assumption Method 1 – the survival data matches a Weibull Distribution
 2. Assumption Method 2 – the survival data is extended linearly from the last observed data to an “all dead” point in the future.
- To examine survival curves for each method:
 1. Select the Markov node.
 2. Choose Analysis > Markov Cohort > Extended Report from the menu.
 3. Click the survival curve link to the right of the data tables.

- Survival curve from Method 1:



- Survival curve from Method 2:



- The model accumulates Life Years in the Active Payoff. For each monthly cycle, it accumulates 1/12 LY. See the total LYs for the model at the bottom of the Extended Markov Cohort Reports.
 1. Life expectancy from Method 1: 3.75 years.
 2. Life expectancy from Method 2: 4.42 years.