



*Tutorials and worked examples for simulation,
curve fitting, statistical analysis, and plotting.*

<https://simfit.org.uk>

<https://simfit.silverfrost.com>

Often experimentalists prefer to fit a continuous parametric survivor function to survival data in order to quantify the data by best-fit parameters instead a nonparametric step-wise approximation, and the Weibull distribution is often used for this purpose.

From the main SIMFIT menu choose [Statistics], [Time series and survival], then [Kaplan-Meier], and study the default test file `survive.tf2` which has the following format.

| Time to relief | Censorship | Frequency |
|----------------|------------|-----------|
| 1.1 | 0 | 1 |
| 1.4 | 0 | 1 |
| 1.3 | 0 | 1 |
| 1.7 | 0 | 1 |
| 1.9 | 0 | 1 |
| 1.8 | 0 | 1 |
| 1.6 | 0 | 1 |
| 2.2 | 0 | 1 |
| 1.7 | 0 | 1 |
| 2.7 | 0 | 1 |
| 4.1 | 0 | 1 |
| 1.8 | 0 | 1 |
| 1.5 | 0 | 1 |
| 1.2 | 0 | 1 |
| 1.4 | 0 | 1 |
| 3.0 | 0 | 1 |
| 1.7 | 0 | 1 |
| 2.3 | 0 | 1 |
| 1.6 | 0 | 1 |
| 2.0 | 0 | 1 |

These data were for twenty patients taking an analgesic to relieve headache pain and the data have been formatted according to this scheme, where the critical event in this case is freedom from pain.

1. **First column**

Time in hours (not necessarily ordered)

2. **Second column**

Censoring code (0 = occurrence of the critical event, 1 = right-censored)

3. **Third column**

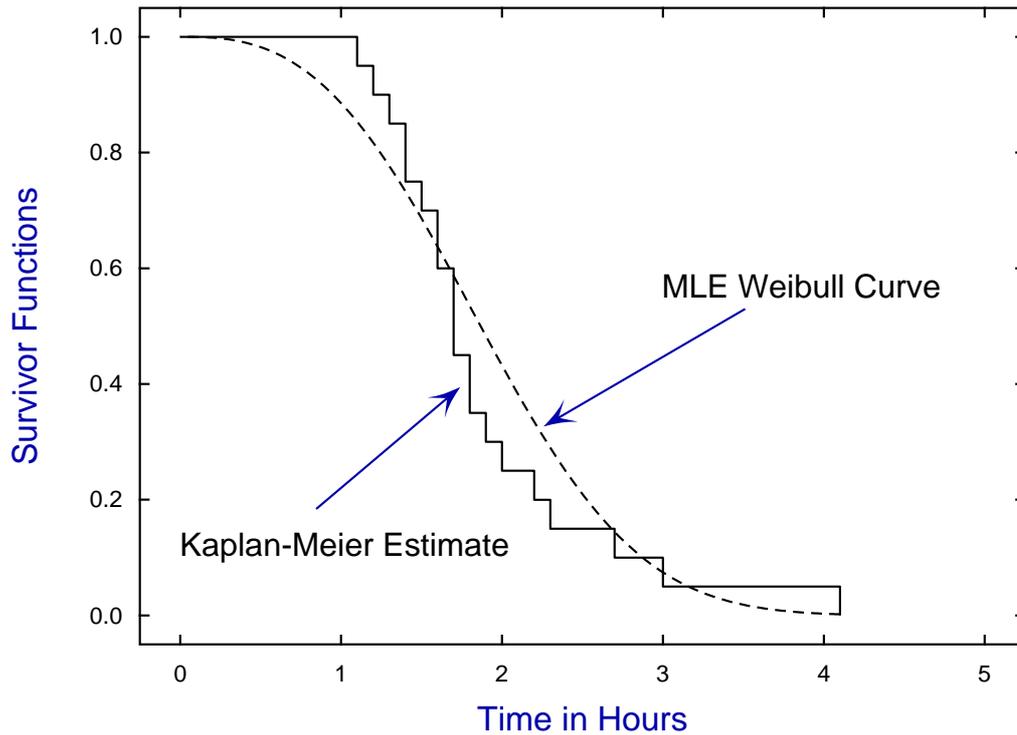
Frequency of the observation

4. **Note**

The starting sample size will be taken as the sum of all the frequencies in column 3. So subjects remaining at or after the last failure should be included as right-censored with the appropriate frequency. Failure to do this will lead to an underestimate of the starting size for the group and a biased estimator.

These data do not contain censored observations and all the subjects were observed until the headache ceased. The Kaplan-Meier menu then provides the option to fit a Weibull survivor function and to overlay it on the Kaplan-Meier step function as in the next figure.

Kaplan-Meier $\hat{S}(t)$ and Best Fit Weibull Curve



Of course the maximum likelihood Weibull estimate displayed is only one possible curve, and SIMFIT offers the opportunity to fit alternative curves by weighted least squares, or generalized linear interactive modeling (GLIM).

Tests for goodness of fit and for comparing parameter estimates require the actual parameter estimates and standard errors, i.e. estimates for their their standard errors, where the numerical values for these will depend upon the particular parameterization scheme used to define the Weibull model. So, as the next table shows, the Weibull model can be fitted with several parameterizations, and the one used by SIMFIT is designed to stabilize the calculations for the maximum likelihood estimate, as described subsequently.

Alternative MLE Weibull parameterizations

$$S(t) = \exp[-\exp(\beta)t^B]$$

$$S(t) = \exp[-\lambda t^B]$$

$$S(t) = \exp[-(At)^B]$$

| Parameter | Value | Standard error | Lower95%cl | Upper95%cl | <i>p</i> |
|-----------|---------|----------------|------------|------------|----------|
| <i>B</i> | 2.7870 | 0.4273 | 1.889 | 3.685 | 0.0000 |
| β | -2.1073 | 0.4627 | -3.079 | -1.135 | 0.0002 |
| λ | 0.1216 | 0.0563 | 0.003 | 0.240 | 0.0444 |
| <i>A</i> | 0.4695 | 0.0401 | 0.385 | 0.554 | 0.0000 |
| $t_{1/2}$ | 1.8675 | 0.1761 | 1.497 | 2.238 | 0.0000 |

Correlation coefficient(β, B) = -0.8755

Theory

To understand fitting the Weibull distribution, note that maximum likelihood parameter and standard error estimates are reported for three alternative parameterizations, namely

$$\begin{aligned} S(t) &= \exp(-\exp(\beta)t^B) \\ &= \exp(-\lambda t^B) \\ &= \exp(-(At)^B). \end{aligned}$$

Since the density and survivor function are

$$\begin{aligned} f(t) &= B\lambda t^{B-1} \exp(-\lambda t^B) \\ S(t) &= \exp(-\lambda t^B), \end{aligned}$$

and there are d failures and $n - d$ right censored observations, the likelihood function $l(B, \lambda)$ is proportional to the product of the d densities for the failures in the overall set of n observations and the survivor functions, that is

$$l(B, \lambda) \propto (B\lambda)^d \left(\prod_{i \in D} t_i^{B-1} \right) \exp\left(-\lambda \sum_{i=1}^n t_i^B\right)$$

where D is the set of failure times.

Actually, the log-likelihood function objective function

$$L(B, \beta) = d \log(B) + d\beta + (B - 1) \sum_{i \in D} \log(t_i) - \exp(\beta) \sum_{i=1}^n t_i^B$$

with $\lambda = \exp(\beta)$ is better conditioned, so it is maximized and the partial derivatives

$$\begin{aligned} L_1 &= \partial L / \partial \beta \\ L_2 &= \partial L / \partial B \\ L_{11} &= \partial^2 L / \partial \beta^2 \\ L_{12} &= \partial^2 L / \partial B \partial \beta \\ L_{22} &= \partial^2 L / \partial B^2 \end{aligned}$$

are used to form the standard errors and correlation coefficient according to

$$\begin{aligned} \text{se}(\hat{B}) &= \sqrt{-L_{11} / (L_{11}L_{22} - L_{12}^2)} \\ \text{se}(\hat{\beta}) &= \sqrt{-L_{22} / (L_{11}L_{22} - L_{12}^2)} \\ \text{corr}(\hat{B}, \hat{\beta}) &= L_{12} / \sqrt{L_{11}L_{22}}. \end{aligned}$$

Given the maximum likelihood estimates for B and β , the estimates for A and λ are calculated and the standard errors worked out by the propagation of errors technique, while the half-time $t_{1/2}$, which is the time to 50% survival, is calculated numerically using the survivor function with parameters equal to the maximum likelihood estimates.