

# Package: Sunclarco (via r-universe)

September 11, 2024

**Type** Package

**Title** Survival Analysis using Copulas

**Version** 1.0.0

**Date** 2017-10-04

**Author** Leen Preten, Roel Braekers, Luc Duchateau and Ewoud De Troyer

**Maintainer** Roel Braekers <roel.braekers@uhasselt.be>

**Description** Survival analysis for unbalanced clusters using Archimedean copulas (Preten et al. (2016) <[DOI:10.1111/rssb.12174](https://doi.org/10.1111/rssb.12174)>).

**Imports** survival, stats, utils

**License** GPL-3

**RoxygenNote** 6.0.1

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2017-10-04 17:15:48 UTC

## Contents

insem . . . . .	2
Sunclarco . . . . .	2
SunclarcoModel . . . . .	3

<b>Index</b>	<b>7</b>
--------------	----------

---

insem	<i>Insemination Data</i>
-------	--------------------------

---

**Description**

In dairy cattle, the calving interval (the time between two calvings) should be optimally between 12 and 13 months. One of the main factors determining the length of the calving interval is the time from parturition to the time of first insemination (Time variable in the data set). The data set includes 181 clusters (Herd variable in the data set) of different sizes. The parity of the cow (0 if multiparous; 1 if primiparous) is added as a covariate (Heifer in the data set).

**Format**

A dataframe with 10513 rows and 5 columns.

---

Sunclarco	<i>Survival analysis for unbalanced clusters using archimedes copula's.</i>
-----------	---

---

**Description**

Multivariate survival data occur in many different disciplines. If the correlation in the data is of interest by itself, two main modeling tools exist, the frailty model (Duchateau and Janssen, 2008) and the copula model. The use of copula modeling has been restricted due to the lack of software on the one hand. On the other hand, copula models mainly deal theoretically with small clusters, all of equal size, i.e., containing the same number of subjects. The Sunclarco package can handle large unbalanced clusters, i.e., of varying size. This allows the use of copula models to data sets that could previously not be handled, e.g., multicentre cancer clinical trials. Furthermore, the Sunclarco package is flexible in terms of the baseline hazard (Weibull, piecewise exponential, unspecified (using partial likelihood) and in terms of the copula function (Clayton and Gumbel-Hougaard).

**Author(s)**

Leen Prenen  
Ewoud De Troyer  
Roel Braekers  
Luc Duchateau

**References**

Prenen L, Braekers R, Duchateau L (2017). Extending the Archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. *Journal of the Royal Statistical Society*, 6, 1-24.

Duchateau L, Janssen P. (2008). *The frailty model*. Springer Verlag.

---

SunclarcoModel	<i>Sunclarco Model</i>
----------------	------------------------

---

**Description**

Model for Survival Analysis of Unbalanced Clusters using Archimedes Copula's.

**Usage**

```
SunclarcoModel(data, time, status, clusters, covariates, stage = 1,
  copula = "Clayton", marginal = "Weibull", n.piecewise = 20,
  init.values = NULL, baselevels = NULL, verbose = TRUE,
  summary.print = TRUE, optim.method = NULL, optim.bounds = NULL)
```

**Arguments**

<code>data</code>	Input dataframe containing all variables.
<code>time</code>	Which variable name is the time covariate?
<code>status</code>	The status indicator, 0=alive, 1=dead.
<code>clusters</code>	The variable name describing the clusters.
<code>covariates</code>	A vector of one or more covariates to be included in the model. Categorical covariates should be a factor in this data frame.
<code>stage</code>	Denotes whether the one-stage (stage=1, default) or the two-stage (stage=2) approach should be used. See Details for more information.
<code>copula</code>	Denotes which copula to use. Can be "Clayton" (default) or "GH" for Gumbel-Hougaard.
<code>marginal</code>	Denotes which marginal survival function to use. Can be "Weibull" (default), "PiecewiseExp" for Piecewise Exponential or "Cox" for non-parametric.
<code>n.piecewise</code>	For marginal="PiecewiseExp", denotes how many pieces the Piecewise Exponential should have (Default = 20).
<code>init.values</code>	A List object which contains the initial values for the parameters. This depends on the choice of the parameters stage, copula and marginal. See the <i>Initial Values</i> Section for more information. If no initial parameters are given, they will be chosen automatically (See Details for more information).
<code>baselevels</code>	Denotes the level of a categorical covariate in the covariates vector to be used as baseline. If not set, the first appearing level will be used as the baseline level. The specification should be done as a character vector and the names of this vector should coincide with the chosen factor variable (e.g. <code>c(disease='Other', region='Region1')</code> in which disease and region are factor covariates).
<code>verbose</code>	Print some in-between results as well as computation progress.
<code>summary.print</code>	Logical value to print a short summary at the end of the computation.

<code>optim.method</code>	Method used for optimization in one-stage estimation or in second stage of two-stage estimation. Can either be "Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN" or "Brent". Default in one-stage estimation is "Nelder-Mead" with Weibull margins and "BFGS" with piecewise exponential margins. Default in two-stage estimation is "Brent", except for the combination of Gumbel copula with Weibull margins, where the default is "BFGS".
<code>optim.bounds</code>	Lower and upper bounds on the variables for the "L-BFGS-B" method, or bounds in which to search for method "Brent". Should be a vector of length 2 in which the first element is the lower and the second the upper bound (e.g. <code>c(-Inf, Inf)</code> ). If <code>optim.method = NULL</code> and "Brent" is used, then default bounds will be chosen. Otherwise, <code>optim.bounds</code> is defaulted to <code>c(-Inf, Inf)</code> .

### Details

All copula models, regardless the choice of the marginal survival function, can be fitted with the two-stage approach. The one-stage approach, however, is only available for the "Weibull" and "PiecewiseExp" marginal survival functions choice. The one-stage approach is preferred as it leads to less biased estimates in the case of small sample sizes. When no initial values for the parameters are given, initial values for the optimisation procedure will be derived in the following way. Initial values for the marginal survival functions are obtained by estimating the parameters marginally, i.e., without taking into consideration the copula function. In the two-stage approach, these estimates are fixed, whereas in the one-stage approach, they are parameters in the optimisation. The association parameter is set arbitrarily to 0.5 for "Clayton" 0.55 for "GH". An initial value for the association parameter can be supplied as `c(theta=value)` Initial values for the marginal survival function parameters can only be supplied for the "Weibull" choice as `c(lambda=value, rho=value)` Initial values for the beta parameters from continuous covariates can be supplied as `c(beta_variablename=value)` Initial values for the beta parameters from categorical covariates can be supplied as `c(beta_variablename_level=value)`

### Value

S3 List object

- `Parameters`: Data frame containing estimates and standard errors of parameters.
- `Kendall_Tau`: Vector containing estimate and standard error of Kendall's Tau.
- `ParametersCov`: If available, covariance matrix of the parameters. For 2-stage approaches this is only available for the Weibull marginal.
- `logllh`: The log-likelihood value.
- `parameter.call`: A list containing all arguments given to the function, as well as the initial parameter values and the elapsed time.

### Initial Values

Initial values are provided in a `list()` object as following:

```
list(lambda=c(0.5), rho=0.5, theta=0.5 beta=c(0.5))
```

**Not all initial values need to be provided!** If only some of the initial values are provided, all initial parameters will be estimated (see Details), but the provided initial values will overwrite the generated ones.

Depending on the stage and marginal parameter, different initial values can be provided:

- *One-Stage:*
  - Weibull Marginal
    - \* lambda: Single initial value for marginal survival function.
    - \* rho: Single initial value for marginal survival function.
    - \* theta: Single initial value for the association parameter.
    - \* beta: Vector of multiple initial values for the continuous/categorical covariates.
  - Piecewise Exponential Marginal
    - \* lambda: Vector of multiple initial value for marginal survival function. The length of this vector should be the number of `n.piecewise` (see note down below).
    - \* theta: Single initial value for the association parameter.
    - \* beta: Vector of multiple initial values for the continuous/categorical covariates.
- *Two-Stage:*
  - Weibull or Cox Marginal
    - \* theta: Single initial value for the association parameter.
  - Piecewise Exponential Marginal
    - \* lambda: Vector of multiple initial value for marginal survival function. The length of this vector should be the number of `n.piecewise` (see note down below).
    - \* theta: Single initial value for the association parameter.
    - \* beta: Vector of multiple initial values for the continuous/categorical covariates.

### Initial Values Boundaries

- $\lambda > 0$
- $\rho > 0$
- $\theta$ :
  - GH Copula:  $\theta > 0$  &  $\theta < 1$
  - Clayton Copula:  $\theta > 0$

### Note on lambda and beta

For the Piecewise Exponential marginal, multiple  $\lambda$ 's should be provided in the lambda slot as a vector. This vector can have a maximum length of the number of pieces there were chosen (`n.piecewise`). In the scenario not all  $\lambda$ 's are provided, only the first few  $\lambda$ 's are overwritten.

In the beta slot, as many  $\beta$ 's should be provided as there are covariates (as well as in the same order of the covariates parameter). If one of the covariates is a categorical variable (factor), multiple  $\beta$ 's should be provided for a single covariate (namely the number of levels minus 1). In the scenario not all  $\beta$ 's are provided, only the first few  $\beta$ 's are overwritten.

### References

Prenen L, Braekers R, Duchateau L (2017). Extending the Archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. *Journal of the Royal Statistical Society*, 6, 1-24.

**Examples**

```
## Not run:
data("insem",package="Sunclarco")
result1 <- SunclarcoModel(data=insem,time="Time",status="Status",
                           clusters="Herd",covariates="Heifer",
                           stage=1,copula="Clayton",marginal="Weibull")

summary(result1)

result2 <- SunclarcoModel(data=insem,time="Time",status="Status",
                           clusters="Herd",covariates="Heifer",
                           stage=1,copula="GH",marginal="PiecewiseExp")

summary(result2)

result3 <- SunclarcoModel(data=kidney,time="time",status="status",
                           clusters="id",covariates="sex",
                           stage=2,copula="Clayton",marginal="Weibull")

summary(result3)

result4 <- SunclarcoModel(data=kidney,time="time",status="status",
                           clusters="id",covariates="sex",
                           stage=2,copula="Clayton",marginal="Cox")

summary(result4)

## End(Not run)
```

# Index

`insem`, [2](#)

`Sunclarco`, [2](#)

`Sunclarco-package (Sunclarco)`, [2](#)

`SunclarcoModel`, [3](#)