Package: survlab (via r-universe)

October 22, 2026

Type Package

Title Survival Model-Based Imputation for Laboratory Non-Detect Data

Version 0.2.0

Description Implements survival-model-based imputation for censored laboratory measurements, including Tobit-type models with several distribution options. Suitable for data with values below detection or quantification limits, the package identifies the best-fitting distribution and produces realistic imputations that respect the censoring thresholds.

License MIT + file LICENSE

Depends R (>= 4.1.0)

Imports data.table (>= 1.17.0), stats, survival (>= 3.0.0), truncnorm (>= 1.0.0)

Suggests ggplot2, knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Encoding UTF-8

Language en-US

LazyData true

URL https://lpereira-ue.github.io/survlab/, https://github.com/lpereira-ue/survlab

Config/roxygen2/version 8.1.1

Repository https://lpereira-ue.r-universe.dev

Date/Publication 2026-10-22 11:22:10 UTC

RemoteUrl https://github.com/lpereira-ue/survlab

RemoteRef HEAD

RemoteSha b7df91c0462e107f0a4183e0898ac454ddd8df5f

Contents

impute_nondetect 2
multi_censored_data 5
validate_imputation 6

Index **8**

`impute_nondetect` *Impute Non-Detect Values in Laboratory Data*

Description

This function imputes non-detect (censored) values in environmental laboratory analytical data using survival models with automatic distribution selection. It validates data quality requirements and fits multiple distributions to select the best model based on AIC. Each imputed value is guaranteed to be below its respective detection limit and above the specified minimum value.

Usage

```
impute_nondetect(
  dt,
  value_col = "value",
  cens_col = "censored",
  parameter_col = NULL,
  unit_col = NULL,
 dist = c("gaussian", "lognormal", "weibull", "exponential", "logistic", "loglogistic"),
  min_observations = 25,
  max_censored_pct = 75,
  min_value = 0,
  zero_value = 1e-05,
  control = survival::survreg.control(),
  verbose = FALSE
)
```

Arguments

`dt`	A data.frame or data.table containing laboratory analytical data
`value_col`	Character string specifying the column name containing values
`cens_col`	Character string specifying the column name containing censoring indicators (0 = non-detect/censored, 1 = detected/observed)
`parameter_col`	Character string specifying the column name containing parameter names (optional, for validation)
`unit_col`	Character string specifying the column name containing units (optional, for validation)
`dist`	Character vector of distributions to test. Options include: "gaussian", "lognormal", "weibull", "exponential", "logistic", "loglogistic"

`min_observations`
: Minimum number of observations required for modeling (default: 25)

`max_censored_pct`
: Maximum percentage of censored values allowed (default: 75)

`min_value`
: Minimum allowable value for imputed concentrations (default: 0, use 1e-10 for strictly positive)

`zero_value`
: Value that replaces exact zeros when fitting the distributions defined only for positive values ("lognormal", "weibull", "exponential", "loglogistic"), which otherwise cannot be fitted (default: 1e-5). The replacement is used for model fitting only; the zeros are kept in `[value_col]_final`. Use `NULL` to disable, in which case those distributions are skipped whenever the data contain a zero.

`control`
: A `survreg.control` object used to control the fitting algorithm, e.g. maximum number of iterations and convergence tolerance. Defaults to `survival::survreg.control()`. Increase `maxiter` (e.g. `survreg.control(maxiter = 200)`) when convergence warnings are raised for complex datasets.

`verbose`
: Logical indicating whether to display progress messages and distribution fitting information (default: FALSE)

Details

The function performs several validation checks: 1. Ensures sufficient sample size (>= min_observations) 2. Checks that censoring percentage is reasonable (<= max_censored_pct) 3. Validates that only one parameter and unit are present (if columns provided) 4. Tests multiple distributions and selects the best based on AIC among those that converged (if none converged, the best non-converged fit is used and a warning is raised) 5. Generates random imputed values below each observation's detection limit and above min_value

The imputed values are random draws from R's random number generator; call `set.seed` before `impute_nondetect()` to make them reproducible.

For non-detect observations (censored = 0), the value in value_col is treated as the detection limit for that specific analysis, allowing for different detection limits across samples or analytical methods.

IMPORTANT: This function should be applied to data containing only ONE parameter at a time. Different environmental parameters have different distributions and should not be modeled together.

When verbose = FALSE, the function operates silently except for critical errors, making it suitable for batch processing of multiple parameters.

Value

A data.table with additional columns:

[value_col _imputed] Imputed values for non-detect observations

[value_col _final] Final values combining original detected and imputed non-detect values

The returned object also has attributes containing model information:

best_model The fitted survival model object

best_distribution Name of the best-fitting distribution

detection_limits Vector of all detection limits found in the data

max_detection_limit The highest detection limit (for reference)

parameter Parameter name (if parameter_col provided)

unit Unit of measurement (if unit_col provided)

aic AIC value of the best model

sample_size Total number of observations

censored_pct Percentage of censored observations

best_converged FALSE if no distribution converged and the selected model is the best non-converged fit

fits A data.table with one row per distribution tried: AIC, convergence, and any error or warning message

n_zero_replaced Number of exact zeros replaced by zero_value when fitting the positive distributions

Examples

```
# Load example data
data(multi_censored_data)

# Basic imputation with default settings
set.seed(123)
result <- impute_nondetect(
  dt = multi_censored_data,
  value_col = "value",
  cens_col = "censored",
  verbose = FALSE
)

# View imputed values for non-detects
head(result[censored == 0, .(value, value_imputed, value_final)])

# Check best distribution selected
attr(result, "best_distribution")

# Increase max iterations for difficult datasets
result <- impute_nondetect(
  dt        = multi_censored_data,
  value_col = "value",
  cens_col  = "censored",
  control   = survival::survreg.control(maxiter = 200)
)

# With parameter and unit validation
result <- impute_nondetect(
  dt = multi_censored_data,
  value_col = "value",
  cens_col = "censored",
  parameter_col = "parameter",
  unit_col = "unit"
)
```

```
# For strictly positive values (avoiding exactly zero)
result <- impute_nondetect(
  dt = multi_censored_data,
  value_col = "value",
  cens_col = "censored",
  min_value = 1e-10,
  verbose = FALSE
)
```

`multi_censored_data` *Environmental Laboratory Nitrate Data with Non-Detects*

Description

A synthetic dataset containing environmental nitrate measurements with non-detect values, generated from a lognormal distribution. This dataset represents typical water quality monitoring data from an environmental laboratory, designed for demonstrating survival model-based imputation techniques.

Usage

```
multi_censored_data
```

Format

A data.table with 200 rows and 4 variables:

parameter Character string indicating the chemical parameter ("Nitrate")

unit Character string indicating the unit of measurement ("mg/l NO3")

value Numeric values representing either detected measurements or detection limits for non-detect observations

censored Integer indicator where 0 = non-detect (below detection limit), 1 = detected (above detection limit)

Details

This dataset simulates real-world environmental water quality data where nitrate measurements below certain detection limits are reported as non-detects. The data includes:

- Single parameter (Nitrate) with consistent units (mg/l NO3)
- Multiple detection limit levels reflecting different analytical conditions
- Realistic distribution of detected vs non-detect values (83.5
- Detection limits ranging from 5 to 25 mg/l NO3
- Lognormal distribution typical of environmental contaminant data

For non-detect observations (censored = 0), the 'value' column contains the detection limit for that specific analysis. For detected measurements (censored = 1), the 'value' column contains the actual measured nitrate concentration.

Source

Synthetic data generated for package demonstration, based on typical environmental water quality monitoring programs

Examples

```
data(multi_censored_data)

# Basic data exploration
multi_censored_data[, .(
  total_samples = .N,
  non_detects = sum(censored == 0),
  detects = sum(censored == 1)
)]

# View parameter and unit information
multi_censored_data[, .(
  parameter = unique(parameter),
  unit = unique(unit)
)]

# View detection limit levels
multi_censored_data[censored == 0, unique(value)]

# Apply survival model imputation
result <- impute_nondetect(multi_censored_data,
                           parameter_col = "parameter",
                           unit_col = "unit")
validate_imputation(result)
```

---

validate_imputation *Validate Laboratory Non-Detect Imputation Results*

Description

This function validates the quality of non-detect value imputation by checking that imputed values are below their respective limits of quantification and providing comprehensive summary statistics and model diagnostics.

Usage

```
validate_imputation(
  dt_imputed,
  value_col = "value",
  cens_col = "censored",
  verbose = TRUE
)
```

Arguments

dt_imputed	A data.table returned from impute_nondetect
value_col	Character string specifying the column name containing original values
cens_col	Character string specifying the column name containing censoring indicators
verbose	Logical indicating whether to print validation results to console (default: TRUE)

Details

The function checks:

- All imputed values are strictly below their respective limits of quantification
- Uniqueness of imputed values
- Summary statistics by limits of quantification level
- Model fit information including parameter and unit details
- Dataset characteristics (sample size, censoring percentage)

Value

Invisibly returns the input data.table. When verbose = TRUE, prints validation results to console including:

- Whether all imputed values are below their detection limits
- Number of duplicate imputed values (if any)
- Summary statistics by detection limit level
- Model fit information

Examples

```
data(multi_censored_data)
result <- impute_nondetect(multi_censored_data, verbose = FALSE)
validate_imputation(result)

# Silent validation for batch processing
validate_imputation(result, verbose = FALSE)
```

Index

∗ **datasets**
multi_censored_data, 5

impute_nondetect, 2, *7*

multi_censored_data, 5

set.seed, *3*
survreg.control, *3*

validate_imputation, 6