# Package 'reflimR'

January 10, 2024

Type Package

Title Reference Limit Estimation Using Routine Laboratory Data

Version 1.0.6

Date 2023-12-22

Maintainer Georg Hoffmann < georg.hoffmann@trillium.de>

Description

Uses an indirect method based on truncated quantile-quantile plots to estimate reference limits from routine laboratory data. The principle of the method was developed by Robert G Hoffmann (1963) <doi:10.1001/jama.1963.03060110068020> and modified by Georg Hoffmann and colleagues (2015) <doi:10.1515/labmed-2015-0104>, (2020) <doi:10.1515/labmed-2020-0005>, and (2022) <doi:10.1007/978-3-031-15509-3\_31>.

URL https://github.com/reflimR

License GPL-2

**Encoding** UTF-8

LazyData true

Imports stats, graphics, grDevices

**Depends** R (>= 3.5.0)

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

BugReports https://github.com/reflim/reflimR/issues

VignetteBuilder knitr

RoxygenNote 7.2.3

NeedsCompilation no

Author Georg Hoffmann [aut, cre],

Sandra Klawitter [aut], Frank Klawonn [aut]

Repository CRAN

**Date/Publication** 2024-01-10 13:10:02 UTC

2 adjust\_digits

# **R** topics documented:

adjust_digits	2
bowley	3
conf_int95	4
Dataset: livertests	5
Dataset: targetvalues	6
boxplot	7
Interpretation	8
lognorm	9
permissible_uncertainty	10
reflim	11
ri_hist	13
truncated_qqplot	15
	<b>17</b>

adjust\_digits

Plausible rounding

# Description

Rounds a quantitative laboratory result to a reasonable number of decimal places.

# Usage

**Index** 

```
adjust_digits(x)
```

# **Arguments**

Х

numeric

### Value

List containing a rounded value of x and the number of decimal places

```
adjust_digits(0.001234)
adjust_digits(0)
adjust_digits(-12.34)
adjust_digits(5.4321)$digits
```

bowley 3

bowley Bowley skewness
------------------------

# **Description**

Calculates a robust skewness measure for x based on the interquartile range (or any other quantile range).

# Usage

```
bowley(x, alpha = 0.25)
```

# **Arguments**

x numeric vector

alpha lower quantile of the range to be regarded (e. g. 0.25)

#### **Details**

Bowley's quantile skewness is calculated from (q[1] - 2 \* q[2] + q[3]) / (q[3] - q[1]), where q is a vector of quantiles alpha, 0.5, and 1 - alpha. The default value for alpha = 0.25 indicates an interval from the first to the third quartile.

# Value

Bowley's quantile skewness

### References

- 1. Bowley, AL (1920). Elements of Statistics. London: P.S. King & Son, Ltd.
- 2. Klawonn F, Hoffmann G, Orth M. Quantitative laboratory results: normal or lognormal distribution. J Lab Med 2020; 44: 143–50. doi:10.1515/labmed20200005.

```
bowley(1 : 100)
bowley(rnorm(1000, 3, 0.2))
bowley(rlnorm(1000, 3, 0.5))
```

4 conf\_int95

conf_int95	Confidence intervals of estimated reference limits

# **Description**

Calculates 95 percent confidence intervals for the lower and upper reference limits obtained with the reflim algorithm.

#### Usage

#### **Arguments**

n	number of observations
lower.limit	positive number indicating the lower limit of the reference interval
upper.limit	positive number indicating the upper limit of the reference interval
lognormal	Boolean indicating whether a lognormal distribution should be assumed
apply.rounding	Boolean indicating whether the confidence limits should be rounded

### **Details**

The confidence limits depend on the reference range (upper minus lower limit), and are proportional to 1/sqrt(n).

The coefficients used in this function have been determined by 100,000 Monte-Carlo simulations for sample sizes between n = 200 and n = 2,000 based on a standard normal distribution.

#### Value

95 percent confidence limits and total number of observations for the lower and the upper reference limit (ranging from lower.lim.low to lower.lim.upp for the lower reference limit, and from upper.lim.low to upper.lim.upp for the upper reference limit)

```
conf_int95(n = 250, lower.limit = 10, upper.limit = 50)
conf_int95(250, 135, 145, FALSE, FALSE)
```

Dataset: livertests 5

#### **Description**

Example data showing eight different biomarkers (laboratory tests), which are frequently measured in healthy controls and patients with liver diseases.

# Usage

livertests

#### **Format**

A data frame with 612 rows and 11 columns:

Category healthy reference individual or patient

Age age in years

**Sex** sex f = female or m = male

ALB albumin, g/L

ALT alanine aminotransferase, U/L

AST aspartate aminotransferase, U/L

BIL bilirubin, µmol/l

CHE choline esterase, kU/L

CREA creatinine, µmol/L

GGT gamma-glutamyl transferase, U/L

PROT total protein, mg/L

#### Source

<a href="https://archive.ics.uci.edu/ml/datasets/HCV+data">https://archive.ics.uci.edu/ml/datasets/HCV+data>

```
summary(livertests)
pie(table(livertests$Category), labels = c("patients", "controls"))
plot(livertests$Age, livertests$ALB, xlab = "Age [yr]", ylab = "ALB [g/L]")
grid()
abline(lm(livertests$ALB ~ livertests$Age))

che <- livertests$CHE
ref <- livertests$CHE[livertests$Category == "reference"]
pat <- livertests$CHE[livertests$Category == "patient"]

hist(che, breaks = 1 : 20, col = "white", main = "cholinesterase", xlab = "kU/L")
hist(ref, breaks = 1 : 20, col = rgb(0, 1, 0, 0.5), add = TRUE)</pre>
```

Dataset: targetvalues

```
hist(pat, breaks = 1 : 20, col = rgb(1, 0, 0, 0.5), add = TRUE)
legend("topright", fill = c(rgb(1,1,1,1), rgb(0,1,0,0.5), rgb(1,0,0,0.5)),
    legend = c("all", "controls", "patients"))

t.test(ref, pat)
var.test(ref, pat)

che.f <- livertests$CHE[livertests$Sex == "f"]
che.m <- livertests$CHE[livertests$Sex == "m"]
plot(density(che.f), xlim = c(0, 20), col = "red",
    main = "cholinesterase", xlab = "kU/L")
lines(density(che.m), col = "blue")
legend("topright", lty = 1, col = c("red", "blue"), legend = c("females", "males"))

reflim(che.m, main = "CHE (m)", xlab = "kU/L")
reflim(livertests$AST[livertests$Sex == "m"], main = "AST (m)", xlab = "U/L")</pre>
```

Dataset: targetvalues Dataset: target values

#### **Description**

Test names (analytes), units and reference limits from a textbook.

### Usage

targetvalues

#### **Format**

A data frame with 8 rows and 6 columns:

analyte short name of the analyte

unit measuring unit

**ll.female**, **ul.female** lower and upper reference limits for women

**Il.male**, **ul.male** lower and upper reference limits for men

#### **Details**

The table was created from the data in the textbook (web version 2023, https://www.clinical-laboratory-diagnostics.com). Missing data (i.e. the lower limits for ALT, AST and GGT) were supplemented from the product sheets of the respective tests.

#### Source

Thomas L. Clinical Laboratory Diagnostics, 2023

iboxplot 7

# **Examples**

```
targetvalues[, 1 : 4]
reflim(livertests$ALB[livertests$Sex == "m"],
main = targetvalues[1, 1], xlab = targetvalues[1, 2],
targets = targetvalues[1, 5 : 6])
```

iboxplot

Removal of pathological values

## Description

Iteratively truncates a vector of quantitative laboratory results until no more values outside the specified truncation interval are left.

# Usage

```
iboxplot(x, lognormal = NULL, perc.trunc = 2.5,
apply.rounding = TRUE,
plot.it = TRUE, main = "iBoxplot", xlab = "x")
```

#### **Arguments**

X	vector of positive numbers
lognormal	Boolean indicating whether a lognormal distribution should be assumed (NULL means that the distribution type is defined automatically)
perc.trunc	percentage of presumably normal values to be removed from each side
apply.rounding	Boolean indicating whether the estimated reference limits should be rounded
plot.it	Boolean indicating whether a graphic should be created
main, xlab	title and x label of the graphic

### **Details**

The truncated vector represents the estimated central 95 percent of values, which follow the assumed distribution (normal or lognormal). If the distribution of the reference values is unknown, medical laboratory results should be assumed to be lognormally distributed [2].

#### Value

\$trunc	truncated vector x
\$limits	truncation points, preliminary reference limits
\$lognormal	Boolean indicating whether a lognormal distribution has been assumed
<pre>\$perc.norm</pre>	proportion of the assumed non-pathological values
\$progress	results of the iterative truncation

8 interpretation

#### References

1. Klawonn F, Hoffmann G. Using fuzzy cluster analysis to find interesting clusters. In: L.A. Garcia-Escuderoet al. (eds.): Building bridges between soft and statistical methodologies for data science. Springer, Cham (2023), 231-239. doi:10.1007/9783031155093\_31.

2. Haeckel R, Wosniok W. Observed unknown distributions of clinical chemical quantities should be considered to be log-normal. Clin Chem Lab Med 2010; 48: 1393-6. doi:10.1515/CCLM.2010.273.

# **Examples**

```
set.seed(123)
iboxplot(rlnorm(n = 250, meanlog = 3, sdlog = 0.3))
iboxplot(rnorm(1000, 100, 10), apply.rounding = FALSE, plot.it = FALSE)$truncation.points
alb.trunc <- iboxplot(livertests$ALB, main = "ALB", xlab = "g/L")$trunc
summary(alb.trunc)</pre>
```

interpretation

Colors and text modules to interprete deviations from given target values

# **Description**

Creates traffic light colors green, yellow, and red as well as a textual description such as 'slightly increased'. This function is called by reflim(), if target values are available, and provides the required color information for ri\_hist().

# Usage

```
interpretation(limits, targets)
```

### **Arguments**

limits vector of two numbers indicating the reference limits that have been calculated

by the reflim function (or any other suitable algorithm)

targets vector of two numbers indicating target reference limits that may have been

obtained from external sources

#### **Details**

This algorithm compares the positions and tolerance intervals of the estimated upper and lower reference limits with the tolerance limits of the respective target values.

If the estimated reference limit is within the tolerance of the target value, the dev.lim text says 'within tolerance' and the color code #00FF0080 for semi-transparent green is returned.

If the position is outside and the two tolerance limits overlap, the dev.lim text says 'slightly increased' or 'slightly decreased' and the color code #FFFF0080 for semi-transparent yellow is returned.

lognorm 9

If the tolerance limits do not overlap, the dev.lim text says 'markedly increased' or 'markedly decreased' and the color code #FF000080 for semi transparent red is returned.

#### Value

```
$tol.lim and $tol.tar
tolerance limits for the estimated reference limits and the respective target values. If targets are not provided, the latter tolerance limits are returned as NA.
$col.lim and $col.tar
hexadecimal rgb values, indicating the traffic light colors green, yellow, and red
$dev.lim short text describing the deviations of the observed limit values from the target values
```

#### **Examples**

```
interpretation(limits = c(10, 50), targets = c(11, 49))
interpretation(limits = c(10, 50), targets = c(8, 60))$dev.lim
```

lognorm	Lognormal distribution model	

#### **Description**

Suggests lognormal modelling of a numeric vector x by comparing Bowley's quantile skewness for x and log(x). Lognormality is suggested if bowley(x) - bowley(log(x)) >= cutoff.

# Usage

#### **Arguments**

X	numeric vector of positive numbers
cutoff	a skewness threshold for the suggestion of a lognormal distribution
alpha	lower quantile of the range to be regarded (e. g. 0.25 for IQR)
digits	number of digits to be displayed for the Bowley skewness
plot.it	Boolean indicating whether a graphic should be created
plot.logtype	Boolean indicating whether the distribution type should be printed in the graphic
main, xlab	title and x label of the graphic

#### **Details**

If s lognorm is TRUE, a lognormal distribution is suggested for right-skewed density curves (bowley(x) > 0). The decision for a lognormal distribution is based on the difference between the skewness of the original and the logtransformed values (cut-off defaults to 0.05).

In the unusual case of a left-skewed distribution (bowley(x) < 0), a normal distribution is suggested (lognormal = FALSE), assuming that the left skew is caused by pathological low values rather than an unusual distribution of laboratory results.

The plot illustrates the skewness of x and log(x) showing density curves and boxplots with separate x-axes for the original values (bottom axis) and the log-transformed values (top axis). A skewness delta below the cut-off value means that both curves are quite symmetric. In this case, x can be approximated by a normal distribution. If the delta exceeds the cut-off, the density curve of the original values and the respective boxplot are right-skewed and become more symmetric after log-transformation.

The plot.logtype argument is used internally to suppress printing the automated definition of the distribution in case that the type has been set manually.

Extreme values are removed before plotting to improve the graphic.

#### Value

\$lognorm Boolean indicating whether a lognormal distribution should be assumed \$BowleySkewness

Bowley skewness of the original and the logtransformed values as well as the difference

# References

1. Klawonn F, Hoffmann G, Orth M. Quantitative laboratory results: normal or lognormal distribution? J Lab Med 2020; 44: 143-50. doi:10.1515/labmed20200005.

#### **Examples**

```
lognorm(rnorm(n = 1000, mean = 20, sd = 2))
lognorm(rlnorm(n = 1000, meanlog = 3, sdlog = 0.3))
lognorm(livertests$ALB, main = "albumin", xlab = "g/L")
lognorm(livertests$BIL, main = "bilirubin", xlab = "\u00c4mmol/L")
```

permissible\_uncertainty

Tolerance intervals of estimated and target limits

#### **Description**

Returns the permissible uncertainty of reference limits.

#### Usage

```
permissible_uncertainty(lower.limit, upper.limit, apply.rounding = TRUE)
```

reflim 11

#### **Arguments**

```
lower.limit positive number indicating the lower limit of the reference interval upper.limit positive number indicating the upper limit of the reference interval apply.rounding Boolean indicating whether the tolerance limits should be rounded
```

#### **Details**

The tolerance limits (also called equivalence limits) indicate the permissible uncertainty of a reference limit from a medical point of view (in contrast to the confidence interval, which reflects the statistical point of view). The calculation is based on a recommendation made by the DGKL [1, 2].

#### Value

Tolerance intervals for the lower and upper reference limits (ranging from lower.lim.low to lower.lim.upp for the lower reference limit and from upper.lim.low to upper.lim.upp for the upper reference limit)

#### References

- 1. Haeckel R et al. Permissible limits for uncertainty of measurement in laboratory medicine. Clin Chem Lab Med 2015;53:1161–71. doi:10.1515/cclm20140874.
- 2. Haeckel R et al. Equivalence limits of reference intervals for partitioning of population data. J Lab Med 2016; 40: 199-205. doi:10.1515/labmed20160002.

#### **Examples**

```
permissible_uncertainty(lower.limit = 10, upper.limit = 50)
permissible_uncertainty(10, 50, FALSE)
```

reflim

Reference limits (main function)

#### **Description**

Estimation of reference limits from mixed distributions of normal and pathological laboratory results. Estimates lower and upper reference limits and provides statistical characteristics and graphics to evaluate the results.

# Usage

12 reflim

#### Arguments

X	vector of positive numbers
lognormal	Boolean indicating whether a lognormal distribution should be assumed (NULL means that the distribution type is defined automatically)
targets	vector of two numbers indicating target reference limits that may have been obtained from external sources
perc.trunc	percentage of presumably normal values to be removed from each side
n.min	minimum number of observations needed for a robust estimate of reference intervals
apply.rounding	Boolean indicating whether the estimated reference limits should be rounded
plot.it	Boolean indicating whether graphics should be created
plot.all	Boolean indicating whether graphics of all process steps should be created
print.n	Boolean indicating whether the number of cases after truncation should be printed on the graph
main, xlab	title and x label of the graphic
,	

#### **Details**

The reflim function estimates reference limits from the linear part of a normal probability-probability or quantile-quantile plot [1, 2]. It combines several functions to determine the distribution type [3], to truncate the input vector [4], and to generate a truncated quantile-quantile plot [2, 4]. For details concerning the individual functions, which are called by reflim(), you may enter \*help(package = reflimR)\*.

The default value of perc.trunc is 2.5 meaning that 2.5 percent of the assumed non-pathological values are truncated on both sides of the quantile-quantile plot. By increasing perc.trunc (e.g. to 5 percent), a stronger cut can be applied to reduce the influence of potentially overlapping pathological values.

The argument plot.it is used to compare the observed and the theoretical distribution curves graphically. If target values have been specified, the tolerance intervals of the calculated reference limits will be drawn as colored vertical lines. Green means that the calculated limit is within the tolerance interval of the respective target, yellow means that the calculated limit falls outside but the two tolerance intervals overlap, and red means that there is no overlap between the tolerance intervals.

More detailed graphs of the three underlying steps can be generated by setting plot.all = TRUE. When plot.all is set to TRUE, plot.it is automatically set to TRUE as well.

#### Value

\$stats	mean and sd (or meanlog and sdlog) of the truncated vector, number of cases before and after truncation
\$lognormal	Boolean indicating whether a lognormal distribution has been assumed
\$limits	estimated reference limits with tolerance intervals
\$targets	target values with tolerance intervals
perc.norm	estimated percentage of non-pathological values

ri\_hist 13

\$confidence.int

95 percent confidence intervals for the estimated reference limits (depends on n)

\$interpretation

short text describing the deviation of observed limits from target values

remarks short text describing potential reasons why the reflim function could not be ex-

ecuted

#### References

- 1. Holmes D, Buhr K. Widespread incorrect implementation of the Hoffmann method, the correct approach, and modern alternatives. Am. J. Clin. Pathol. 2018; 151:328-36. doi:10.1093/ajcp/aqy149.
- 2. Hoffmann G, Lichtinghagen R, Wosniok W. Simple estimation of reference intervals from routine laboratory data. J Lab Med 2015; 39: 389-402. doi:10.1515/labmed20150104.
- 3. Klawonn F, Hoffmann G, Orth M. Quantitative laboratory results: normal or lognormal distribution? J Lab Med 2020; 44: 143-50. doi:10.1515/labmed20200005.
- 4. Klawonn F, Hoffmann G. Using fuzzy cluster analysis to find interesting clusters. In: L.A. Garcia-Escuderoet al. (eds.): Building bridges between soft and statistical methodologies for data science. Springer, Cham (2023), 231-239. doi:10.1007/9783031155093\_31.

#### **Examples**

```
x <- c(rnorm(800, 100, 10), rnorm(100, 70, 15), rnorm(100, 125, 15))
reflim(x, targets = qnorm(c(0.025, 0.975), 100, 10))

x.f <- subset(livertests, livertests$Sex == "f")
reflim(x.f$AST)
reflim(x.f$AST, targets = c(13, 40))
reflim(x.f$AST, targets = targetvalues[3, 3 : 4],
main = "AST/GOT", xlab = targetvalues[3, 2])
reflim(x.f$AST, plot.all = TRUE, main = "AST/GOT", xlab = "U/L")$limits
reflim(x.f$ALB, targets = targetvalues[1, 3 : 4],
    plot.all = TRUE, main = "ALB", xlab = "g/L")$limits</pre>
```

ri\_hist

Histogram with density plots and reference limits

#### **Description**

Creates a graphic illustrating the results of the reflim function.

#### **Usage**

ri\_hist

#### **Arguments**

x vector of positive numbers

lognormal Boolean indicating whether a lognormal distribution should be assumed

stats vector of mean and sd, or meanlog and sdlog, respectively

limits vector of reference limits calculated by the reflim function (or any other suitable

algorithm)

perc.norm estimated percentage of non-pathological values (usually provided by the ibox-

plot function)

targets vector of target reference limits obtained from external sources

remove.extremes

Boolean indicating whether extreme values should be removed to improve the

graphic

main, xlab title and x label of the graphic

#### **Details**

ri\_hist is called by the reflim function, but it can also be used to illustrate the results of other software packages (e. g. refineR), if the required arguments are available (see details).

It creates a graphic, which includes a histogram and a density curve of x, as well as a theoretical density curve of presumably non-pathological values (blue) and a calculated density curve of presumably pathological values (red). Calculated reference limits and target limits are shown as vertical lines, and their tolerance intervals (i. e., the permissible uncertainties) are represented by surrounding boxes. If target values are provided, traffic light colors indicate the deviation between target and actual results.

If the arguments lognormal or perc.norm are unknown, they can be set according to the user's expertise. For example, if the distribution type is unknown, a lognormal distribution can be assumed [1]. If the percentage of non-pathological values has not been provided by a foreign algorithm (e. g. refineR), it can be roughly estimated, if density curves of normal and pathological values are available (the argument perc.norm does not influence the result; its only effect is on the shape of the theoretical density curve).

# Value

\$lognormal assumed distribution model

\$percent\_normal

assumed percentage of non-pathological values

\$interpretation

text describing the deviation of observed limits from target values

#### References

1. Haeckel R, Wosniok W. Observed unknown distributions of clinical chemical quantities should be considered to be log-normal. Clin Chem Lab Med 2010; 48: 1393-6. doi:10.1515/CCLM.2010.273.

truncated\_qqplot 15

#### **Examples**

```
set.seed(123)
x1 <- rlnorm(800, 3, 0.3)
lim <- quantile(x1, c(0.025, 0.975))
ri_hist(x1, lognormal = TRUE, stats = c(3, 0.3), limits = lim, perc.norm = 100)

x2 <- rlnorm(200, 3.5, 0.4)
x <- c(x1, x2)
tar <- quantile(x, c(0.025, 0.975))
ri_hist(x, lognormal = TRUE, stats = c(3, 0.3), limits = lim, targets = tar, perc.norm = 80)</pre>
```

truncated\_qqplot

Quantile-Quantile plot of truncated laboratory results

#### **Description**

Generates and plots a quantile-quantile plot (q-q plot) with a theoretical normal distribution on the x-axis and the corresponding empirical distribution on the y-axis. Returns intercept, slope, and estimated quantiles 0.025 and 0.975 (i.e., reference limits).

# Usage

#### **Arguments**

x.trunc	truncated numeric vector of positive numbers, usually generated by iboxplot()
lognormal	Boolean indicating whether a lognormal distribution should be assumed (NULL means that the distribution type is defined automatically)
perc.trunc	percentage of values that has been removed from each side by truncation
n.min	minimal number of values in x.trunc for a robust estimate of reference limits
apply.rounding	Boolean indicating whether the reference limits should be rounded
plot.it	Boolean indicating whether a graphic should be created
main, xlab, ylab	
	title and labels of the graphic

#### **Details**

Intercept and slope of the q-q plot represent the robust mean and standard deviation of x.trunc. They serve as parameters to estimate the reference limits being represented by the quantiles 0.025 and 0.975 of a presumably normal subset of x.

16 truncated\_qqplot

# Value

\$stats intercept and slope of the q-q plot, lower and upper truncation points
\$lognormal Boolean indicating whether a lognormal distribution has been assumed

#### References

1. Hoffmann G et al. Simple estimation of reference intervals from routine laboratory data. J Lab Med 2016. doi:10.1515/labmed20150104.

```
set.seed(123)
x <- rlnorm(n = 250, meanlog = 3, sdlog = 0.3)
x.trunc <- iboxplot(x, plot.it = FALSE)$trunc
truncated_qqplot(x.trunc)

x.f <- subset(livertests, livertests$Sex == "f")
x.trunc <- iboxplot(x.f$ALT, plot.it = FALSE)$trunc
truncated_qqplot(x.trunc, n.min = length(x.trunc), main = "ALT")</pre>
```

# **Index**

```
adjust_digits, 2
bowley, 3
conf_int95, 4
Dataset: livertests, 5
Dataset: targetvalues, 6
iboxplot, 7
interpretation, 8
livertests (Dataset: livertests), 5
lognorm, 9
permissible_uncertainty, 10
reflim, 11
ri_hist, 13
targetvalues (Dataset: targetvalues), 6
truncated_qqplot, 15
```