Package 'labNorm'

January 15, 2023

Title Normalize Laboratory Measurements by Age and Sex

Version 1.0.1

Description
Provides functions for normalizing standard laboratory measurements (e.g. hemoglobin, cholesterol levels) according to age and sex, based on the algorithms described in ``Personalized lab test models to quantify disease potentials in healthy individuals'' (Netta Mendelson Cohen, Omer Schwartzman, Ram Jaschek, Aviezer Lifshitz, Michael Hoichman, Ran Balicer, Liran I. Shlush, Gabi Barbash & Amos Tanay, <doi:10.1038/s41591-021-01468-6>). Allows users to easily obtain normalized values for standard lab results, and to visualize their distributions. See more at <`https://tanaylab.weizmann.ac.il/labs/`>.

License MIT + file LICENSE

Suggests covr, mockery, spelling, testthat (>= 3.0.0)

Config/testthat/edition 3

Config/testthat/start-first download

Encoding UTF-8

Language en-US

RoxygenNote 7.2.3

Imports cli, dplyr, forcats, ggplot2, glue, purrr, rappdirs, scales,
stats, tibble, utils, withr, yesno

Depends R (>= 2.10)

LazyData true

NeedsCompilation no

Author Aviezer Lifshitz [aut, cre],
Netta Mendelson-Cohen [aut],
Weizmann Institute of Science [cph]

Maintainer Aviezer Lifshitz <`aviezer.lifshitz@weizmann.ac.il`>

Repository CRAN

Date/Publication 2023-01-15 14:30:05 UTC

R topics documented:

`hemoglobin_data` *Example values of Hemoglobin and Creatinine*

Description

Example datasets of Hemoglobin and Creatinine values for testing

Usage

```
hemoglobin_data

creatinine_data
```

Format

`hemoglobin_data` `creatinine_data`**:**

A data frame with 1000 rows and 3 columns:

age age of the patient

sex sex of the patient

value the lab value for the patient, in the default units for the lab

An object of class `data.frame` with 1000 rows and 3 columns.

Examples

```
head(hemoglobin_data)
head(creatinine_data)
```

`LAB_DETAILS` *Available lab names*

Description

Names of the labs available in the package.

Usage

```
LAB_DETAILS
```

Format

`LAB_DETAILS`**:**

A data frame with 95 rows and 4 columns:

short_name Short lab name

long_name Long lab name

units a list column with all the units available for the lab

default_units the default units for the lab

low_male,high_male,low_female,high_female the reference ranges for the lab, taken from the American Board of Internal Medicine. Can be NA if the lab does not have reference ranges.

Source

American Board of Internal Medicine. ABIM Laboratory Test Reference Ranges — July 2021. https://www.abim.org/~/media/ABIM%20Public/Files/pdf/exam/laboratory-reference-ranges.pdf (2021).

Examples

```
head(LAB_DETAILS)
```

`ln_convert_units` *Convert values to the default units for the lab*

Description

Convert values to the default units for the lab

Usage

```
ln_convert_units(values, units, lab)
```

Arguments

`values` a vector of lab values

`units` the units of the lab values. See `LAB_DETAILS$units` for a list of available units for each lab. If different values have different units then this should be a vector of the same length as `values`.

`lab` the lab name. See `LAB_DETAILS$short_name` for a list of available labs.

Value

the values converted to the default units for the lab

Examples

```
# emulate a dataset with different units

hemoglobin_diff_units <- hemoglobin_data

# first 50 values will be in mg/ML
hemoglobin_diff_units$value[1:50] <- hemoglobin_diff_units$value[1:50] * 10

# last 50 values will be in mmol/L
hemoglobin_diff_units$value[51:100] <- hemoglobin_diff_units$value[51:100] / 1.61


converted <- ln_convert_units(
    hemoglobin_diff_units$value[1:100],
    c(rep("mg/mL", 50), rep("mmol/L", 50)),
    "Hemoglobin"
)

head(converted)
head(hemoglobin_data$value)
```

`ln_download_data` *Download high-resolution reference distributions*

Description

The data is downloaded to the directory specified by the `dir` parameter. Note that if you specified a directory different from the default, you will need to set `options(labNorm.dir = dir)` in order for the package to use the downloaded data in future sessions.
Default directories are:

- Unix: ~/.local/share/LabNorm
- Mac OS X: `~/Library/Application Support/LabNorm`
- Win XP (not roaming): `C:\\Documents and Settings\\<username>\\Data\\<AppAuthor>\\LabNorm`

- Win XP (roaming): `C:\\Documents and Settings\\<username>\\Local Settings\\Data\\<AppAuthor>\\LabNo`
- Win 7 (not roaming): `C:\\Users\\<username>\\AppData\\Local\\<AppAuthor>\\LabNorm`
- Win 7 (roaming): `C:\\Users\\<username>\\AppData\\Roaming\\<AppAuthor>\\LabNorm`

Usage

```
ln_download_data(dir = NULL)

ln_data_downloaded()
```

Arguments

`dir` the directory to download the data to. If `NULL` and the user approves, the data will be downloaded to the package directory, using `rappdirs::user_data_dir("labnorm")`, otherwise - a temporary directory would be used.

Value

None.

True if the data was downloaded, false otherwise.

Examples

```

ln_download_data()


ln_data_downloaded()
```

`ln_lab_units` *Get available units for a lab*

Description

Get available units for a lab

Get the default units for a lab

Usage

```
ln_lab_units(lab)

ln_lab_default_units(lab)
```

Arguments

`lab` the lab name. See `LAB_DETAILS$short_name` for a list of available labs.

Value

a vector of available units for the lab

the default units for the lab

Examples

```
ln_lab_units("Hemoglobin")

ln_lab_default_units("Hemoglobin")
```

ln_normalize *Normalize lab values to age and sex*

Description

Normalize standard laboratory measurements (e.g. hemoglobin, cholesterol levels) according to age and sex, based on the algorithms described in "Personalized lab test models to quantify disease potentials in healthy individuals" doi:10.1038/s41591021014686.

The "Clalit" reference distributions are based on 2.1B lab measurements taken from 2.8M individuals between 2002-2019, filtered to exclude severe chronic diseases and medication effects. The resulting normalized value is a quantile between 0 and 1, representing the value's position in the reference distribution.

The "UKBB" reference distributions are based on the UK-Biobank, a large-scale population-based cohort study of 500K individuals, which underwent the same filtering process as the "Clalit" reference distributions.

The list of supported labs can be found below or by running `LAB_DETAILS$short_name`.

Usage

```
ln_normalize(
  values,
  age,
  sex,
  lab,
  units = NULL,
  reference = "Clalit",
  na.rm = FALSE
)

ln_normalize_multi(labs_df, reference = "Clalit", na.rm = FALSE)
```

Arguments

`values` a vector of lab values

`age` a vector of ages between 20-89 for "Clalit" reference and 35-80 for "UKBB". Can be a single value if all values are the same age.

`sex` a vector of either "male" or "female". Can be a single value if all values are the same sex.

`lab` the lab name. See `LAB_DETAILS$short_name` for a list of available labs.

`units` the units of the lab values. See `ln_lab_units(lab)` for a list of available units for each lab. If `NULL` then the default units (`ln_lab_default_units(lab)`) for the lab will be used. If different values have different units then this should be a vector of the same length as `values`.

`reference` the reference distribution to use. Can be either "Clalit" or "UKBB" or "Clalit-demo". Please download the Clalit and UKBB reference distributions using `ln_download_data()`.

`na.rm` if `TRUE`, then `NA` in age, sex or values will be ignored and 'NA' would be returned. Otherwise, an error will be thrown.

`labs_df` a data frame with the columns "value", "age", "sex", "units", and "lab". The "lab" column should be a vector with the lab name per row. See `ln_normalize` for details on the other columns.

Value

a vector of normalized values. If `ln_download_data()` was not run, a lower resolution reference distribution will be used, which can have an error of up to 5 quantiles (0.05). Otherwise, the full reference distribution will be used. You can check if the high resolution data was downloaded using `ln_data_downloaded()`.

You can force the function to use the lower resolution distribution by setting `options(labNorm.use_low_res = TRUE)`.

If the quantile information is not available (e.g. "Estradiol" for male patients, various labs which are not available in the UKBB data), then the function will return `NA`.

reference distribution

It is highly recommended to use `ln_download_data` to download the "Clalit" and "UKBB" reference distributions. If you choose not to download the data, the package will use the demo reference distributions included in the package ("Clalit-demo"), which doesn't include all the labs, and has a resolution of 20 quantile bins and therefore may have an error of up to 5 percentiles (0.05), particularly at the edges of the distribution.

labs

The following labs are supported in the "Clalit" reference (some labs are missing from the UKBB reference):

- WBC

- RBC
- Hemoglobin
- Hematocrit
- Platelets
- MCV
- MCH
- MCHC
- RDW
- MPV
- Large unstained cells, Abs
- Albumin
- Total Cholesterol
- Triglycerides
- BMI
- Iron
- Transferrin
- Ferritin
- PDW
- MPXI
- Total Globulin
- PCT
- HDW
- Fibrinogen
- CH
- Chloride
- Large unstained cells, %
- Macrocytic
- Microcytic
- Hyperchromic
- Hypochromic
- Lymphocytes, Abs
- Lymphocytes, %
- Neutrophils, Abs
- Neutrophils, %
- Monocytes, Abs
- Monocytes, %
- Eosinophils, Abs

- Eosinophils, %
- Basophils, Abs
- Basophils, %
- Microcytic:Hypochromic
- Glucose
- Urea
- Creatinine
- Uric Acid
- Calcium
- Phosphorus
- Total Protein
- HDL Cholesterol
- LDL Cholesterol
- Alk. Phosphatase
- AST
- ALT
- GGT
- LDH
- CPK
- Total Bilirubin
- Direct Bilirubin
- Hemoglobin A1c
- Sodium
- Potassium
- Vitamin D (25-OH)
- Microalbumin:Creatinine
- Urine Creatinine
- Urine Microalbumin
- Non-HDL
- TSH
- T3, Free
- T4, Free
- Blood Pressure, Systolic
- Blood Pressure, Diastolic
- Urine Specific Gravity
- Urine pH
- PT, INR

- PT, sec
- PT, %
- Vitamin B12
- PSA
- ESR
- aPTT, sec
- CRP
- Amylase
- Folic Acid
- Total:HDL
- Hematocrit:Hemoglobin
- Magnesium
- aPTT, ratio
- Indirect Bilirubin
- RDW-SD
- RDW-CV
- LH
- Estradiol

Examples

```
# Normalize Hemoglobin values to age and sex
hemoglobin_data$quantile <- ln_normalize(
    hemoglobin_data$value,
    hemoglobin_data$age,
    hemoglobin_data$sex,
    "Hemoglobin"
)

# plot the quantiles vs values for age 50-60
library(ggplot2)
library(dplyr)
hemoglobin_data %>%
    filter(age >= 50 & age <= 60) %>%
    ggplot(aes(x = value, y = quantile, color = sex)) +
    geom_point() +
    theme_classic()

# Different units
hemoglobin_diff_units <- hemoglobin_data
hemoglobin_diff_units$value <- hemoglobin_diff_units$value * 0.1
hemoglobin_diff_units$quantile <- ln_normalize(
    hemoglobin_data$value,
    hemoglobin_data$age,
    hemoglobin_data$sex,
    "Hemoglobin",
    "mg/mL"
)

# Multiple units
creatinine_diff_units <- creatinine_data
creatinine_diff_units$value <- c(
    creatinine_diff_units$value[1:500] * 0.011312,
    creatinine_diff_units$value[501:1000] * 11.312
)
creatinine_diff_units$quantile <- ln_normalize(
    creatinine_diff_units$value,
    creatinine_diff_units$age,
    creatinine_diff_units$sex,
    "Creatinine",
    c(rep("umol/L", 500), rep("mmol/L", 500))
)

# Use UKBB as reference
hemoglobin_data_ukbb <- hemoglobin_data %>% filter(age >= 35 & age <= 80)
hemoglobin_data_ukbb$quantile_ukbb <- ln_normalize(
    hemoglobin_data_ukbb$value,
    hemoglobin_data_ukbb$age,
    hemoglobin_data_ukbb$sex,
    "Hemoglobin",
    reference = "UKBB"
)

# plot UKBB vs Clalit
hemoglobin_data_ukbb %>%
    filter(age >= 50 & age <= 60) %>%
    ggplot(aes(x = quantile, y = quantile_ukbb, color = sex)) +
    geom_point() +
    geom_abline() +
    theme_classic()


# examples on the demo data


library(dplyr)
multi_labs_df <- bind_rows(
    hemoglobin_data %>% mutate(lab = "Hemoglobin"),
    creatinine_data %>% mutate(lab = "Creatinine")
)


multi_labs_df$quantile <- ln_normalize_multi(multi_labs_df)


# on the demo data
```

```
head(multi_labs_df)
```

ln_plot_dist *Plot age-sex distribution of a lab*

Description

Plot age-sex distribution of a lab

Usage

```
ln_plot_dist(
  lab,
  quantiles = c(0.03, 0.1, 0.15, 0.25, 0.35, 0.65, 0.75, 0.85, 0.9, 0.97),
  reference = "Clalit",
 pal = c("#D7DCE7", "#B0B9D0", "#8997B9", "#6274A2", "#3B528B", "#6274A2", "#8997B9",
    "#B0B9D0", "#D7DCE7"),
  sex = NULL,
  patients = NULL,
  patient_color = "yellow",
  patient_point_size = 2,
  ylim = NULL,
  show_reference = TRUE
)
```

Arguments

`lab` the lab name. See `LAB_DETAILS$short_name` for a list of available labs.

`quantiles` a vector of quantiles to plot, without 0 and 1. Default is `c(0.03, 0.1, 0.15, 0.25, 0.35, 0.5, 0.65, 0.75, 0.85, 0.9, 0.97)`. Note that if `reference="Clalit-demo"`, quantiles below 0.05 and above 0.95 would be rounded to 0.05 and 0.95 respectively, and the same would be done for quantiles below 0.01 and above 0.99 when the high-resolution version is available.

`reference` the reference distribution to use. Can be either "Clalit" or "UKBB" or "Clalit-demo". Please download the Clalit and UKBB reference distributions using `ln_download_data()`.

`pal` a vector of colors to use for the quantiles. Should be of length `length(quantiles) - 1`.

`sex` Plot only a single sex ("male" or "female"). If NULL - `ggplot2::facet_grid` would be used to plot both sexes. Default is NULL.

`patients` (optional) a data frame of patients to plot as dots over the distribution. See the `df` parameter of `ln_normalize_multi` for details.

`patient_color` (optional) the color of the patient dots. Default is "yellow".

patient_point_size
(optional) the size of the patient dots. Default is 2.

ylim
(optional) a vector of length 2 with the lower and upper limits of the plot. Default would be determined based on the values of the upper and lower percentiles of the lab in each age.

show_reference
(optional) if TRUE, plot two lines of the upper and lower reference ranges. Default is TRUE.

Value

a ggplot2 object

Examples

```
set.seed(60427)


ln_plot_dist("Hemoglobin")

# Plot only females
ln_plot_dist("Creatinine", sex = "female", ylim = c(0, 2))

# Set the ylim
ln_plot_dist("BMI", ylim = c(8, 50))

# Project the distribution of three Hemoglobin values
ln_plot_dist("Hemoglobin", patients = dplyr::sample_n(hemoglobin_data, 3))

# Change the quantiles
ln_plot_dist("Hemoglobin",
    quantiles = seq(0.05, 0.95, length.out = 10)
)

# Change the colors
ln_plot_dist(
    "Hemoglobin",
    quantiles = c(0.03, 0.1, 0.25, 0.5, 0.75, 0.9, 0.97),
    pal = c("red", "orange", "yellow", "green", "blue", "purple")
)

# Change the reference distribution
ln_plot_dist("Hemoglobin", reference = "UKBB")


# on the demo data
```

ln_quantile_value *Compute the lab value for a given quantile*

Description

The function ln_quantile_value calculates lab values at a specified quantile, using the default units for that lab. The function ln_patients_quantile_value does the same calculation for a specific group of patients.
Default units for a lab can be obtained using ln_lab_default_units.
If no quantile data is available for a particular lab, age, and sex, the function returns 'NA'.
It should be noted that the values of extreme quantiles (e.g. >0.95 or <0.05 on low resolution, >0.99 or <0.01 on high resolution) may not be reliable, as they may represent outliers in the data.

Note that ln_quantile_value returns values for all combinations of age, sex, and lab, while ln_patients_quantile_value returns values for a specific set of patients, similar to ln_normalize.

Usage

```
ln_quantile_value(
  quantiles,
  age,
  sex,
  lab,
  reference = "Clalit",
  allow_edge_quantiles = FALSE
)

ln_patients_quantile_value(
  quantiles,
  age,
  sex,
  lab,
  reference = "Clalit",
  allow_edge_quantiles = FALSE
)
```

Arguments

quantiles a vector of quantiles (in the range 0-1) to compute the lab value for, or a vector with a quantile for each patient when running ln_patients_quantile_value.

age a vector of ages to compute the lab values for or a vector with an age for each patient when running ln_patients_quantile_value. Note that the age should be in years, and would be floored to the nearest integer.

sex the sexes to compute the lab values for, or a vector with a sex for each patient when running ln_patients_quantile_value. Note that for ln_quantile_value this parameter can only be either: "male", "female" or c("male", "female") `lab` The lab name.

`reference` the reference distribution to use. Can be either "Clalit" or "UKBB" or "Clalit-demo". Please download the Clalit and UKBB reference distributions using `ln_download_data()`.

`allow_edge_quantiles`
If `TRUE` (default) then the function will return the value for the edge quantiles (<0.01 or >0.99) even though they are not reliable. If `FALSE` then the function will return `NA` for those quantiles. Note that for the "Clalit-demo" reference, the threshold would be <0.05 or >0.95.

Value

`ln_quantile_value` returns a data frame which contains the values for each combination of quantile, age and sex. The data frame has the the following columns:

- age: age in years
- sex: "male" or "female"
- quantile: he quantile
- value: the lab value
- unit: the lab unit
- lab: the lab name

`ln_patients_quantile_value` returns a vector of value per patient.

Examples

```
ln_quantile_value(c(0.05, 0.5, 0.95), 50, "male", "WBC")

ln_quantile_value(
    c(0, 0.05, 0.1, 0.4, 0.5, 0.6, 0.9, 1),
    c(50, 60),
    c("male", "female"),
    "Glucose"
)


# on the demo data



hemoglobin_data$quantile <- ln_normalize(
    hemoglobin_data$value,
    hemoglobin_data$age,
    hemoglobin_data$sex,
    "Hemoglobin"
)

hemoglobin_data$value1 <- ln_patients_quantile_value(
    hemoglobin_data$quantile,
    hemoglobin_data$age,
    hemoglobin_data$sex,
    "Hemoglobin"
)
head(hemoglobin_data)
```

Index