

Package ‘idarps’

May 11, 2026

Type Package

Title Datasets and Functions for the Class “Modelling and Data Analysis for Pharmaceutical Sciences”

Version 0.0.6

Description Provides datasets and functions for the class “Modelling and Data Analysis for Pharmaceutical Sciences”.

The datasets can be used to present various methods of data analysis and statistical modeling. Functions for data visualization are also implemented.

License AGPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Depends R (>= 3.5)

NeedsCompilation no

Author Lionel Voirol [aut, cre],
Stéphane Guerrier [aut],
Yuming Zhang [aut],
Luca Insolia [aut]

Maintainer Lionel Voirol <lionelvoiroi@hotmail.com>

Repository CRAN

Date/Publication 2026-05-11 17:10:03 UTC

Contents

boxplot_w_points	2
BreastCancer	3
bronchitis	4
centenarian	4
codex	5
cortisol	6
covid	7

data_covid_switzerland_spatial	7
diabetes	8
diet	9
fev	10
hist_compare_to_normal	10
HP13Cbicarbonate	11
kuwait_bp	12
PeruvianBP	13
pharmacy	13
pm_exposure	14
reading	14
snoring	15
students	15

Index	16
--------------	-----------

boxplot_w_points	<i>boxplot_w_points</i>
------------------	-------------------------

Description

boxplot_w_points

Usage

```
boxplot_w_points(
  ...,
  col_points = "#9033FF3F",
  col_boxplot = "#d2d2d2",
  horizontal = FALSE,
  main = "",
  names = NULL,
  las = 0,
  xlab = "",
  ylab = "",
  seed = 123,
  jitter_param = 0.25
)
```

Arguments

...	data vectors to be visualized.
col_points	color of the points to be added to the boxplot.
col_boxplot	color of the boxplot.
horizontal	logical indicating if the boxplots should be horizontal; default FALSE means vertical boxes.
main	string indicating the title of the plot.

names	vector of string indicating the group labels which will be printed under each boxplot.
las	a numeric value indicating the orientation of the tick mark labels and any other text added to a plot after its initialization. The options are as follows: always parallel to the axis (the default, 0), always horizontal (1), always perpendicular to the axis (2), and always vertical (3).
xlab	a string indicating the x label.
ylab	a string indicating the y label.
seed	an integer specifying a seed for the random jitter of the boxplot points.
jitter_param	a double specifying the amount of jittering applied on points.

Value

No return value. Plot a boxplot.

Examples

```
x <- rnorm(20, mean = 5)
y <- rnorm(20, mean = 10)
z <- rnorm(20, mean = 15)
boxplot_w_points(x, main = "test")
boxplot_w_points(x, y, names = c("x", "y"), las = 1, main = "Data")
boxplot_w_points(x, y, z, names = c("x", "y", "z"), horizontal = TRUE, las = 1, main = "Data")
boxplot_w_points(x, y, z, names = c("x", "y", "z"), horizontal = FALSE, las = 1, main = "Data")
```

BreastCancer

Breast Cancer

Description

This dataset consists of several clinical features observed or measured for 116 participants in a study of breast cancer.

Usage

```
BreastCancer
```

Format

Age Age in years

BMI Body mass index in kg/m^2

Glucose Glucose in mg/dL

Insulin Insulin in $\mu\text{U}/\text{mL}$

HOMA Homeostasis model assessment

Classification Presence of breast cancer (0 if no cancer, 1 if with cancer)

Source

<https://link.springer.com/article/10.1186/s12885-017-3877-1>

References

Patricio, Miguel, et al. "Using Resistin, glucose, age and BMI to predict the presence of breast cancer", BMC Cancer, (2018).

bronchitis	<i>Bronchitis</i>
------------	-------------------

Description

Data collected in a study to assess the effects of smoking and pollution on being diagnosed with bronchitis. This dataset is based on 212 subjects.

Usage

bronchitis

Format

bron Presence of bronchitis (0 for no and 1 for yes)

cigs Average daily number of smoked cigarettes

poll Pollution index

centenarian	<i>Centenarian Blood Pressure</i>
-------------	-----------------------------------

Description

This dataset consists of variables that are potentially related to blood pressure measurements and contains one group of patients aged between 52 and 89 years old who live in urban areas, and another group of 50 centenarian women aged between 101-121 who live in the island of Okinawa, which is known for its high number of centenarians. The dataset lists the following variables:

Usage

centenarian

Format

Age Age in years
Chin Chin skinfold in cm
Forearm Forearm skinfold in cm
Calf Calf skinfold in cm
Pulse Resting pulse rate
BMI The Body Mass Index (BMI) of the participant
Centenarian A dummy variable indicating if the participant is Centenarian
Cystol Systolic blood pressure

 codex

 codex

Description

This dataset is based on an observational study conducted at Geneva University Hospitals to assess the impact of weight on the pharmacokinetics of dexamethasone in normal-weight versus obese patients hospitalized for COVID-19.

Usage

codex

Format

id ID of the patient
gender Gender (0 for men and 1 for women)
age Age
bmi Body mass index
weight Weight in kg
number_doses Number of doses of the dexamethasone (DEX) drug
tmax The time it takes for the drug to reach the maximum concentration (i.e. C_{max}) after its administration in hours (h)
cmax The maximum concentration that achieves in the blood after the drug has been administered (ng/ml)
t1_2 t_{1_2} is the time required to decrease the drug concentration within the body by one-half during elimination in hours (h)
auc The integral (from 0 to 8 hours) of a curve that describes the variation of a drug concentration in the blood as a function of time it takes for a drug to reach the maximum concentration (C_{max}) after administration of a drug (ng.h/ml)
length_hospital Number of days the patient were hospitalized

length_intermed Number of days the patient were hospitalized at the intermediate and intensive care unit

crp crp

comor_e Presence of cormobidity type e

comor_p Presence of cormobidity type p

comor_v Presence of cormobidity type v

comor_c Presence of cormobidity type c

comor_r Presence of cormobidity type r

obese Indicator variable based on whether the subject is obese (i.e. with BMI > 30), 0 for no and 1 for yes.

cortisol

Biomarkers in pigs fed with various diets

Description

This dataset contains measured biomarkers in pigs fed with various diets.

Usage

cortisol

Format

A data frame with 61 rows and 9 variables:

id the id of the pig

group the diet fed to the pig (chipped diet or non-chipped diet)

gender the gender of the pig

cortisol urine cortisol in pg/ml

acth serum acth in pg/ml

crh serum crh in pg/ml

testosterone testosterone in ng/ml

lh LH in ng/ml

caloric daily caloric intake in kcal

`covid`*Intensive care admission of COVID-19 patients in Belgium*

Description

Data from Parisi, et al., (2021) which studies the applicability of predictive models for intensive care admission of COVID-19 patients in a secondary care hospital in Belgium. This study is based on data of patients admitted to an emergency department with a positive RT-PCR SARS-CoV-2 test.

Usage`covid`**Format**

A data frame with 64 rows and 5 variables:

icu admission to an Intensive Care Unit (0 for no, 1 for yes)

sex sex (men, women)

age age in years

ldh lactate dehydrogenase in U/L

spo2 oxygen saturation in percentage

Source

<https://jeccm.amegroups.org/article/view/6927/html>

References

Parisi, Nicolas, et al. "Non applicability of validated predictive models for intensive care admission and death of COVID-19 patients in a secondary care hospital in Belgium.", *Journal of Emergency and Critical Care Medicine*, (2021).

`data_covid_switzerland_spatial`*COVID-19 Spatial*

Description

Data from the COVID-19 Data Hub joined with spatial features for Switzerland.

Usage`data_covid_switzerland_spatial`

Format

admin Country
iso_alpha_3 3-letter code of the country according to the standard ISO 3166-1 Alpha-3
date Date
confirmed Cumulative number of confirmed cases
population Total population
tests Cumulative number of tests
diff_confirmed Daily number of confirmed cases
diff_test Daily number of tests
confirmed_per_pop Number of daily confirmed cases divided per the country population
confirmed_per_pop_ma Moving Average applied to confirmed_per_pop with a window of 7 days
geometry 'sf' geometry list of country

Source

<https://covid19datahub.io/>

diabetes

Diabetes study in Bangladesh

Description

This dataset contains reports of diabetes symptoms from 520 individuals, encompassing symptoms potentially associated with the condition. It was compiled through a questionnaire aimed at recently diagnosed diabetics or individuals displaying one or more symptoms. Data collection took place via direct questionnaire at Sylhet Diabetes Hospital in Bangladesh.

Usage

diabetes

Format

age Age of the patient in years
gender Gender of the patient (Male, Female)
polyuria Presence of polyuria (excessive urination) (Yes, No)
polydipsia Presence of polydipsia (excessive thirst) (Yes, No)
sudden_weight_loss Presence of sudden weight loss (Yes, No)
weakness Presence of weakness (Yes, No)
polyphagia Presence of polyphagia (excessive hunger) (Yes, No)
genital_thrush Presence of genital thrush (Yes, No)

visual_blurring Presence of visual blurring (Yes, No)
itching Presence of itching (Yes, No)
irritability Presence of irritability (Yes, No)
delayed_healing Presence of delayed healing (Yes, No)
partial_paresis Presence of partial paresis (Yes, No)
muscle_stiffness Presence of muscle stiffness (Yes, No)
alopecia Presence of alopecia (Yes, No)
obesity Presence of obesity (Yes, No)
class Diagnosis class (1 if presence of diabetes, 0 otherwise)

Source

https://link.springer.com/chapter/10.1007/978-981-13-8798-2_12

References

Islam, M. M. F., et al. "Likelihood prediction of diabetes at early stage using data mining techniques", Computer vision and machine intelligence in medical image analysis, (2020).

diet

Diet

Description

Diet

Usage

diet

Format

id ID
gender Gender (male or female)
age Age in years
height Height in m
diet.type Type of diet (A, B or C)
initial.weight Initial weight in kg
final.weight Final weight in kg

fev	<i>Forced Expiratory Volume</i>
-----	---------------------------------

Description

This dataset is based on a study conducted in suburban Boston in the late 1970s to investigate the relationship between forced expiratory volume and smoking behavior in 654 youths between the ages of 3 and 19.

Usage

fev

Format

fev forced expiratory volume or FEV, which measures the amount of air a person can exhale during a forced breath.

age age in years

sex gender of the person (0 for males and 1 for females)

height height in cm

smoke smoking behavior (0 for non-smokers and 1 for smokers)

hist_compare_to_normal	<i>hist_compare_to_normal</i>
------------------------	-------------------------------

Description

hist_compare_to_normal

Usage

```
hist_compare_to_normal(
  x,
  col = "lightgray",
  main = "",
  xlab = "",
  ylab = "",
  lwd_line = 1.5,
  col_line1 = "#ff160e",
  col_line2 = "#335bff",
  add_legend = TRUE,
  legend_position = "topleft",
  delta = 0.2,
  ...
)
```

Arguments

<code>x</code>	data vector to be visualized.
<code>col</code>	color of the histogram.
<code>main</code>	string indicating the title of the plot.
<code>xlab</code>	a string indicating the x label.
<code>ylab</code>	a string indicating the y label.
<code>lwd_line</code>	width of density lines.
<code>col_line1</code>	color of density line classic mle estimation.
<code>col_line2</code>	color of density line classic robust estimation.
<code>add_legend</code>	a Boolean if the estimated parameters of the Normal distribution should be plotted.
<code>legend_position</code>	a string specifying the position of the legend.
<code>delta</code>	graphic parameter to determine the shrinkage of the axis.
<code>...</code>	Extra graphical arguments.

Value

No return value. Plot a histogram.

Examples

```
n <- 1000
x <- rnorm(n = n)
hist_compare_to_normal(x)
x2 <- rexp(n, rate = 25)
hist_compare_to_normal(x2, legend_position = "topright")
```

HP13Cbicarbonate

HP13Cbicarbonate

Description

Data from an experiment made on rats which compares the HP13C bicarbonate signal intensities normalized to the total sum of metabolites and corresponding initial reaction rate as a function of the injected dose of HP1-13C pyruvate. Two groups of rats were compared (i.e. fed and overnight-fasted). Dataset from Can et al. 2022.

Usage

HP13Cbicarbonate

Format

signal HP13C bicarbonate signal intensities normalized to the total sum of metabolites

dose initial reaction rate as a function of the injected dose of HP13C pyruvate

group fed and overnight-fasted

Source

<https://www.nature.com/articles/s42003-021-02978-2>

kuwait_bp

Kuwait Blood Pressure

Description

This dataset contains a collection of variables believed to be potentially associated with the blood pressure measurements of 213 individuals from Kuwait. The dataset lists the following variables:

Usage

kuwait_bp

Format

age Age in years

weight Weight in kg

height Height in mm

chin Chin skinfold in cm

forearm Forearm skinfold in cm

calf Calf skinfold in cm

pulse Resting pulse rate

left_handed Whether or not the participant is left-handed

bmi The Body Mass Index (BMI) of the participant

systol Systolic blood pressure

PeruvianBP	<i>Peruvian Blood Pressure</i>
------------	--------------------------------

Description

This dataset consists of variables possibly relating to blood pressures of 39 Peruvians who have moved from rural high-altitude areas to urban lower-altitude areas.

Usage

PeruvianBP

Format

Age Age in years

Years Years in urban area

Weight Weight in kg

Height Height in mm

Chin Chin skinfold

Forearm Forearm skinfold

Calf Calf skinfold

Pulse Resting pulse rate

Systol Systolic blood pressure

pharmacy	<i>Customer attendance of a pharmacy in Geneva</i>
----------	--

Description

This dataset contains the number of clients in a pharmacy for each hour over two years.

Usage

pharmacy

Format

A data frame with 17520 rows and 4 variables:

date the date

hours the hour of the day

weekday the week day

attendance the recorded number of clients

pm_exposure	<i>Particulate Matter Exposure and Lung Function</i>
-------------	--

Description

This dataset contains demographic, environmental and respiratory data for 200 male adults (ages 40–75) to analyze factors affecting lung capacity. It contrasts populations across three distinct air-quality environments: a pristine high-altitude rural town, a suburban area with moderate air quality, and a highly industrialized urban region.

Usage

pm_exposure

Format

A data frame with 200 rows and 7 variables:

Age Age of the participant in years

Height Height of the participant in cm

Weight Weight of the participant in kg

PM2_5_Exposure Estimated 5-years average exposure to fine particulate matter (PM2.5), measured in micrograms per cubic meter

Rural A binary indicator of residential location (1 = Rural, 0 = Non-rural)

AreaType A factor indicating the specific type of residential area ("rural", "suburban", or "urban")

FEV1 Forced Expiratory Volume in 1 second (measured in Liters), representing lung function

reading	<i>Reading</i>
---------	----------------

Description

This dataset is based on the effectiveness of directed reading activities for elementary school students (6-12 years old).

Usage

reading

Format

id Student id

score Degree of Reading Power (DRP) test score

age Age of the students

group Binary variable indicating whether a student participated to the directed reading activities (Treatment if the student participated, Control otherwise)

 snoring

Snoring

Description

This dataset is based on a study on the physical and behavioral characteristics of snorers.

Usage

snoring

Format

sex gender of the person (0 for males and 1 for females)

age age in years

height height in cm

weight weight in kg

smoke smoking behavior (0 for non-smokers and 1 for smokers)

alcohol number of glasses drunk per day (in red wine equivalent)

snore snoring diagnosis (0 for not snoring, 1 for snoring)

 students

Students

Description

Students

Usage

students

Format

day day

case case

Index

* datasets

- BreastCancer, 3
 - bronchitis, 4
 - centenarian, 4
 - codex, 5
 - cortisol, 6
 - covid, 7
 - data_covid_switzerland_spatial, 7
 - diabetes, 8
 - diet, 9
 - fev, 10
 - HP13Cbicarbonate, 11
 - kuwait_bp, 12
 - PeruvianBP, 13
 - pharmacy, 13
 - pm_exposure, 14
 - reading, 14
 - snoring, 15
 - students, 15
-
- boxplot_w_points, 2
 - BreastCancer, 3
 - bronchitis, 4
-
- centenarian, 4
 - codex, 5
 - cortisol, 6
 - covid, 7
-
- data_covid_switzerland_spatial, 7
 - diabetes, 8
 - diet, 9
-
- fev, 10
-
- hist_compare_to_normal, 10
 - HP13Cbicarbonate, 11
-
- kuwait_bp, 12
-
- PeruvianBP, 13