

Package: med isr (via r-universe)

September 29, 2026

Type Package

Title Toolbox for Medical Statistics in R

Version 0.0.0.9000

Description This package serves as a toolbox for the course Medical Statistics.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 8.0.0

Depends R (>= 3.5)

LazyData true

URL <https://github.com/DanMazJen/med isr>,
<https://danmazjen.github.io/med isr/>

BugReports <https://github.com/DanMazJen/med isr/issues>

Suggests devtools, usethis

Imports cli, fs, withr

Config/pak/sysreqs cmake make libuv1-dev

Repository <https://danmazjen.r-universe.dev>

Date/Publication 2026-09-29 07:57:04 UTC

RemoteUrl <https://github.com/danmazjen/med isr>

RemoteRef HEAD

RemoteSha a7ee5bacdc7afde4e93e9a1dc0783c7276bc35b2

Contents

ad_data	2
ad_validation_logistic	3
beetle_dose	4

covid_data	5
dns	6
doctors	8
eq5d	9
hamilton_d17	10
hd_data	11
height_data	12
james_lind	12
mabp	13
medisr_init_project	15
medisr_install_packages	16
prostate_cancer	16
SUA_data	17
thrombectomy	18
Index	20

ad_data	<i>Alzheimer's Disease</i>
---------	----------------------------

Description

Clinical, demographic, and cerebrospinal fluid (CSF) measurements from subjects participating in a clinical study investigating whether biological measurements in CSF can be used to diagnose or predict Alzheimer’s disease.

Usage

ad_data

Format

ad_data:
A list containing:

diagnosis An R factor vector containing the Alzheimer’s disease diagnosis for 333 subjects.

predictors A data frame containing demographic and laboratory measurements for the study subjects, including the **Genotype** variable.

The **Genotype** predictor records the subject’s Apolipoprotein E (APOE) genotype. APOE has three variants (E2, E3, and E4), with six possible genotype combinations formed by inheriting one variant from each parent. APOE genotype is associated with Alzheimer’s disease risk.

These data are a modified version of the values used in the original publication.

Source

Craig-Schapiro, R., Kuhn, M., Xiong, C., Pickering, E. H., Liu, J., Misko, T. P., Perrin, R. J., et al. (2011). Multiplexed Immunoassay Panel Identifies Novel CSF Biomarkers for Alzheimer's Disease Diagnosis and Prognosis. *PLoS ONE*, 6(4), e18850.

ad_validation_logistic

Alzheimer's Disease Logistic Regression Predictions

Description

Predicted Alzheimer's disease probabilities and observed diagnoses from a logistic regression model fitted to a training subset of the `ad_data` dataset. The model uses age, sex, cerebrospinal fluid (CSF) tau and phosphorylated tau concentrations, amyloid beta 42 concentration, and Apolipoprotein E (APOE) genotype as predictors.

Usage

`ad_validation_logistic`

Format

ad_validation_logistic:

A data frame containing one row for each observation in the validation subset, with 3 columns:

.pred_impaired Predicted probability that the subject is cognitively impaired.

.pred_control Predicted probability that the subject is a cognitive control.

Class Observed diagnosis of the subject, with levels `impaired` and `control`.

The probabilities are generated from the logistic regression model:

`diagnosis ~ age + male + tau + p_tau + Ab_42 + Genotype`

Predictor variables include:

age Normalized patient age.

male Indicator of reported male sex.

tau Concentration of total tau protein in cerebrospinal fluid (CSF), measured by lumbar puncture.

p_tau Concentration of phosphorylated tau protein in CSF, measured by lumbar puncture.

Ab_42 Concentration of amyloid beta 42 (A 42) in CSF. A 42 is a 42-amino-acid peptide derived from amyloid precursor protein and is a component of the amyloid plaques associated with Alzheimer's disease.

Genotype APOE allele pair. Possible values are E2E2, E2E3, E3E3, E3E4, E2E4, and E4E4.

Details

The original `ad_data` observations were split by alternating rows, with even-numbered observations used for model training and odd-numbered observations used for validation. The logistic regression model was fitted to the training observations, and predicted probabilities were generated for the validation observations.

See Also

[ad_data](#)

`beetle_dose`

Bliss (1935) Beetle Mortality Data

Description

This dataset contains mortality data for adult flour beetles after five hours of exposure to gaseous carbon disulphide. The data were originally reported by Bliss (1935) and have subsequently been reanalyzed in studies of dose-response models and transformations for binary response data.

Usage

```
beetle_dose
```

Format

`beetle_dose`:

A data frame with 8 rows and 3 columns:

dose Numeric dose of gaseous carbon disulphide (CS), expressed as log10-transformed CS mg/L

died Integer number of adult flour beetles killed after five hours of exposure

n Integer total number of adult flour beetles exposed

Source

Bliss, C. I. (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology*, 22, 134–167.

References

Aranda-Ordaz, F. (1981). On two families of transformations to additivity for binary response data. *Biometrika*, 68, 357–363.

Hauck, W. (1990). Choice of scale and asymmetric logistic models. *Biometrical Journal*, 32, 79–86.

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics*, 38, 761–768.

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Journal of the Royal Statistical Society: Series C*, 29, 15–23.

covid_data	<i>COVID-19 Data</i>
------------	----------------------

Description

Real-world COVID-19 data for European countries.

Usage

```
covid_data
```

Format

A data frame containing population characteristics, COVID-19 cases and deaths, and other country- and week-specific variables:

country Country.

alpha2code Two-digit country code.

totalpop Total population (country-specific variable).

population_density Population density (country-specific variable).

human_development_index Human development index (country-specific variable).

male_smokers Proportion of male smokers (country-specific variable).

female_smokers Proportion of female smokers (country-specific variable).

diabetes_prevalence Diabetes prevalence (country-specific variable).

gdp_per_capita GDP per capita (country-specific variable).

aged_70_older Proportion of people aged 70 years or older (country-specific variable).

people_vaccinated_per_hundred Cumulative number of people vaccinated per hundred in the week of the baseline record.

temperature Average temperature in the week of the baseline record.

date Date of the Monday in the week of the baseline record.

cases Weekly number of cases four weeks after baseline.

cases_baseline Weekly number of cases in the baseline week.

cases_w1bb Weekly number of cases one week before baseline.

cases_w2bb Weekly number of cases two weeks before baseline.

cases_w3bb Weekly number of cases three weeks before baseline.

cases_w4bb Weekly number of cases four weeks before baseline.

cases_w5bb Weekly number of cases five weeks before baseline.

cases_w6bb Weekly number of cases six weeks before baseline.

cases_w7bb Weekly number of cases seven weeks before baseline.

deaths Weekly number of deaths four weeks after baseline.

deaths_baseline Weekly number of deaths in the baseline week.

deaths_w1bb Weekly number of COVID-19-related deaths one week before baseline.

deaths_w2bb Weekly number of COVID-19-related deaths two weeks before baseline.

deaths_w3bb Weekly number of COVID-19-related deaths three weeks before baseline.

deaths_w4bb Weekly number of COVID-19-related deaths four weeks before baseline.

deaths_w5bb Weekly number of COVID-19-related deaths five weeks before baseline.

deaths_w6bb Weekly number of COVID-19-related deaths six weeks before baseline.

deaths_w7bb Weekly number of COVID-19-related deaths seven weeks before baseline.

Details

The population consists of the populations of most European countries. Country-specific variables are combined with weekly COVID-19 data to describe the epidemiological situation around a baseline week.

Source

CIESIN (2018). Gridded Population of the World, Version 4 (GPWv4): Basic Characteristics, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), Columbia University Center for International Earth Science Information Network (CIESIN).

ECDC (2021). COVID-19 Situation Update Worldwide, as of Week 13, Updated 8 April 2021. European Centre for Disease Prevention and Control.

dns

Synthetic Danish National Health Survey Data

Description

Synthetic health survey data reconstructed from clustered and summarized data based on the Danish National Health Survey (Den Nationale Sundhedsprofil, DNS). The data preserve core patterns and correlations observed in the source data while not containing the original individual-level observations.

Usage

dns

Format

dns:

A tibble containing synthetic health survey data:

sex Sex: 1 = Male, 0 = Female

height Height in cm

weight Weight in kg

age Age in years

pcs Physical health rating derived from the 12-Item Short Form Survey (SF-12), with higher values representing better physical health

mcs Mental health rating derived from the 12-Item Short Form Survey (SF-12), with higher values representing better mental well-being

alcoholweek Number of alcohol units consumed per week, where one unit corresponds to 15 ml of alcohol

Details

The Danish National Health Survey is a large population-based survey concerning health, morbidity, and health-related behavior among the Danish population.

The data in this dataset are synthetic. A random subset of survey subjects from the source data was summarized into clusters while preserving core patterns and correlations among the variables. The synthetic observations were subsequently simulated from these clusters. Consequently, the dataset is intended for statistical analysis, teaching, and methodological demonstrations rather than as a replacement for the original survey data.

The pcs and mcs variables are derived from the physical and mental component scores of the 12-Item Short Form Survey (SF-12), respectively. Higher values indicate better health on the corresponding scale.

Source

Danish National Health Survey (Den Nationale Sundhedsprofil). Statens Institut for Folkesundhed, University of Southern Denmark. https://www.sdu.dk/da/sif/rapporter/2022/den_nationale_sundhedsprofil

References

Jensen, H. A. R., Davidsen, M., Ekholm, O., and Christensen, A. I. (2022). Danskernes sundhed: Den Nationale Sundhedsprofil 2021. Statens Institut for Folkesundhed, Syddansk Universitet.

doctors

*British Doctors Study Smoking Data***Description**

Data from the famous British Doctors Study of smoking conducted by Sir Richard Doll, Austin Bradford Hill, and colleagues. The study investigated the relationship between smoking and mortality among British doctors and became one of the landmark studies in the history of epidemiology.

Usage

doctors

Format**doctors:**

A tibble containing aggregated mortality data from the British Doctors Study:

age Age group: 1 = 35 to 44 years, 2 = 45 to 54 years, 3 = 55 to 64 years, 4 = 65 to 74 years, 5 = 75 to 84 years

agesq Age group squared

smoking Smoking status: smoker or non-smoker

deaths Number of deaths

personyears Person-years of observation at the time of the analysis

Details

The British Doctors Study was initiated in 1951 by Richard Doll and Austin Bradford Hill to investigate the relationship between smoking and mortality. Questionnaires concerning smoking habits were sent to British doctors, who were subsequently followed over many years.

The study provided important prospective evidence concerning the association between cigarette smoking and mortality, particularly from lung cancer. Its findings played a major role in the development of modern epidemiology and in the scientific debate over the causal effects of smoking.

The data in this dataset are aggregated by age group and smoking status. The **personyears** variable records the amount of observation contributed by each group and can be used as an exposure term in analyses of mortality rates, such as Poisson regression.

Source

Breslow, N. E. and N. E. Day (1987). Statistical Methods in Cancer Research, Volume 2: The Design and Analysis of Cohort Studies. Lyon: International Agency for Research on Cancer.

References

- Doll, R. and Hill, A. B. (1954). The mortality of doctors in relation to their smoking habits: A preliminary report. *British Medical Journal*, 1(4877), 1451-1455.
- Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research, Volume 2: The Design and Analysis of Cohort Studies*. Lyon: International Agency for Research on Cancer.

eq5d	<i>EQ-5D-5L and PEmb-QoL Data</i>
------	-----------------------------------

Description

Data for predicting quality of life measured by the EQ-5D-5L summary score from selected demographic and PEmb-QoL questionnaire variables in patients following pulmonary embolism.

Format

The **eq5d** dataset contains two data frames, **eq5d_train** and **eq5d_test**, which can be loaded together using `data(eq5d)`.

Both data frames contain the following variables:

eq5d Total EQ-5D-5L summary score.

AL Daily activity limitations domain of the PEmb-QoL questionnaire, reduced to a summary score of 7 items.

EC Emotional complaints domain of the PEmb-QoL questionnaire, reduced to a summary score of 9 items.

IC Intensity of complaints domain of the PEmb-QoL questionnaire.

age Age in years.

sex Sex (0 = Female, 1 = Male).

eq5d_train is the training set and **eq5d_test** is the test set. The data are split by time.

The dataset is a subset of the variables from the original study data. The values have been augmented with noise for use in methodological applications.

Source

Rolving, N., Lindegaard, S. F., Johnsen, S. P., et al. (2024). Effectiveness and cost-effectiveness of a structured integrated post-pulmonary embolism follow-up care model (Attend-PE): a rationale and protocol for a multicentre clinical pre-post study. *BMJ Open*, 14, e088533. doi:10.1136/bmjopen-2024-088533

Lindegaard, S. F., Valentin, J. B., Rolving, N., et al. (2025). Quality of life after pulmonary embolism: psychometric validation of the Danish electronic version of the PEmb-QoL questionnaire. *BMJ Open*, 15, e103473. doi:10.1136/bmjopen-2025-103473

hamilton_d17

*Synthetic Hamilton D17 Depression Data***Description**

Synthetic data reconstructed from clustered and summarized data from the Danish Depression Database (Dansk Depressionsdatabase, DDD). The source data contain patients admitted to psychiatric hospitals with depression in Denmark between 2011 and 2022.

Usage

hamilton_d17

Format**hamilton_d17:**

A tibble containing synthetic data on patients admitted with depression:

Sex Sex: 1 = Male, 0 = Female

readmitted90days Readmitted within 90 days after discharge: 1 = yes, 0 = no

age_at_admission Age in years at admission

admission_duration_days Duration of psychiatric admission in days

HAMD17_t0 Hamilton Depression Rating Scale (HAM-D17) score at admission

HAMD17_t1 Hamilton Depression Rating Scale (HAM-D17) score at discharge

HAMD17_diff Difference in Hamilton Depression Rating Scale (HAM-D17) score between admission and discharge

Details

The Hamilton Depression Rating Scale (HAM-D17) is a clinician-administered instrument used to assess the severity of depressive symptoms. In this dataset, the scale is recorded at admission and discharge, allowing the change in depressive symptom severity during the psychiatric admission to be examined.

The data in this dataset are synthetic. A random subset of observations from the source data was summarized into clusters while preserving core patterns and correlations among the variables. The synthetic observations were subsequently simulated from these clusters. The data are intended for statistical analysis, teaching, and methodological demonstrations and should not be interpreted as original patient-level registry data.

Source

Danish Depression Database (Dansk Depressionsdatabase, DDD). Danish Clinical Quality Program. <https://www.sundk.dk/kliniske-kvalitetsdatabaser/dansk-depressionsdatabase-ddd/om-databasen/>

hd_data

*Cleveland Heart Disease***Description**

Clinical and test measurements for 303 patients with chest pain. The outcome variable indicates whether angiographically confirmed heart disease is present.

Usage

hd_data

Format

hd_data:

A data frame with 303 rows and 14 columns:

HD Heart Disease Status. Confirmed coronary artery disease based on angiography (gold standard). Yes = heart disease present; No = no heart disease

Age Patient age in years

Sex Patient sex. 1 = male; 0 = female

ChestPain Type of chest pain. typical = typical angina; atypical = atypical angina; nonanginal = chest pain not due to heart disease; asymptomatic = no chest pain symptoms

RestBP Resting blood pressure, measured at rest, in mm Hg

Chol Serum cholesterol level in mg/dL

Fbs Fasting blood sugar. 1 = fasting glucose > 120 mg/dL; 0 = normal

RestECG Resting electrocardiogram result. 0 = normal; 1 = ST-T wave abnormality; 2 = left ventricular hypertrophy (LVH)

MaxHR Maximum heart rate achieved during exercise testing, in beats per minute

ExAng Exercise-induced angina. 1 = chest pain during exercise; 0 = no chest pain

Oldpeak ST depression, a numeric measure of ECG change during exercise compared with rest. Larger values suggest ischemia

Slope Slope of the ST segment during exercise. 1 = upsloping; 2 = flat; 3 = downsloping

Ca Number of major coronary vessels with visible disease on angiography. Values range from 0 to 3

Thal Thallium stress test result. normal = normal blood flow; fixed = fixed defect (old infarct or scar); reversible = reversible defect (ischemia under stress)

Source

UCI Machine Learning Repository, Cleveland Heart Disease dataset. <https://archive.ics.uci.edu/dataset/45/heart+>

height_data	<i>Howell !Kung Demography Data</i>
-------------	-------------------------------------

Description

Demographic data from Kalahari !Kung San people collected by Nancy Howell.

Usage

```
height_data
```

Format

height_data:

A data frame containing demographic measurements of !Kung San individuals:

height Height in cm

weight Weight in kg

age Age in years

male Gender indicator

age.at.death Age at death in years, if deceased

alive Indicator of whether the individual is still alive

Source

Downloaded from the University of Toronto TSpace repository: <https://tspace.library.utoronto.ca/handle/1807/10395>

References

Howell, N. Demographic data from the Kalahari !Kung San people.

james_lind	<i>James Lind Scurvy Trial Data</i>
------------	-------------------------------------

Description

Reconstructed data from the journals of James Lind describing his 1747 investigation of treatments for scurvy among sailors aboard HMS Salisbury. The data record the treatments administered to individual sailors, their vitamin C exposure, treatment duration, and outcomes.

Usage

```
james_lind
```

Format**james_lind:**

A data frame containing reconstructed observations from James Lind's scurvy trial:

Sailor_ID Identifier for the sailor

Group Treatment group

Treatment Description of the treatment administered

Vitamin_C Daily vitamin C intake in mg

Total_Vitamin_C Total vitamin C intake during the treatment period, in mg

Duration Treatment duration in days

Outcome_text Text description of the treatment outcome

Outcome Numeric outcome value

Diet Diet during the treatment period

Details

In 1747, James Lind conducted a controlled trial aboard HMS Salisbury to investigate treatments for scurvy. Twelve sailors suffering from scurvy were divided into pairs and given different dietary and medicinal treatments for a period of two weeks. The treatments included cider, elixir vitriol, vinegar, seawater, nutmeg, and citrus fruits.

The trial is widely recognized as an early example of a controlled clinical experiment and is frequently used in discussions of the development of clinical trial methodology and evidence-based medicine.

The data in this dataset are reconstructed from Lind's journals and historical accounts of the trial. They are intended for statistical analysis, teaching, and methodological demonstrations. The reconstructed data should not be interpreted as a verbatim transcription of the original historical records.

Source

Lind, J. (1753). A Treatise of the Scurvy. Edinburgh: Sands, Murray and Cochran.

References

Lind, J. (1753). A Treatise of the Scurvy. Edinburgh: Sands, Murray and Cochran.

mabp

Blood Pressure During Thrombectomy and Functional Outcomes

Description

Data on blood pressure during thrombectomy and functional outcomes at three months following treatment for acute ischemic stroke. The dataset was compiled from four clinical trials and contains a subset of the variables available in the original studies.

Usage

mabp

Format**mabp:**

A tibble containing data on patients undergoing thrombectomy for acute ischemic stroke:

sex Sex: 1 = Female, 0 = Male

age Age in years at admission

SBP__baseline Systolic blood pressure in mmHg at baseline

MAP__baseline Mean arterial blood pressure in mmHg at baseline

nihss__adm National Institutes of Health Stroke Scale (NIHSS) score at admission: 0 = no stroke, 1–4 = minor stroke, 5–15 = moderate stroke, 16–20 = moderate to severe stroke, 21–42 = severe stroke

Hyperton Hypertension: 1 = yes, 0 = no

Diabetes Diabetes: 1 = yes, 0 = no

mrs__3m Modified Rankin Scale (mRS) score at three months: 0 = no symptoms; 1 = no significant disability; 2 = slight disability; 3 = moderate disability; 4 = moderately severe disability; 5 = severe disability; 6 = dead

sbp__t Systolic blood pressure in mmHg at t minutes after groin puncture

map__t Mean arterial blood pressure in mmHg at t minutes after groin puncture

Details

The dataset concerns blood pressure measurements during thrombectomy and functional outcomes three months after treatment. The National Institutes of Health Stroke Scale (NIHSS) is used to quantify the severity of neurological impairment following stroke, while the Modified Rankin Scale (mRS) is used to assess functional disability after stroke.

The data were compiled from four distinct clinical trials. Only a subset of the variables from the original studies is included. Blood pressure measurements have been truncated and time-averaged, and noise has been added to the measurements. Consequently, the data do not represent the original trial observations and are intended for statistical analysis, teaching, and methodological demonstrations.

Source

Espelund, M., et al. (2023).

Rasmussen, M., et al. (2021).

References

Espelund, M., et al. (2023).

Rasmussen, M., et al. (2021).

medisr_init_project	<i>Initialize a Medisr R project</i>
---------------------	--------------------------------------

Description

Creates and initializes an R project directory with a standard folder structure and starter R scripts. If the requested directory does not exist, it is created recursively. Existing files and directories are preserved and skipped.

Usage

```
medisr_init_project(path = ".")
```

Arguments

path	A character string specifying the path to the project directory. Defaults to the current working directory.
------	---

Details

The project contains the following structure:

- `<project>.Rproj` – RStudio project file
- `raw-data/` – raw input data
- `data/` – processed data
- `R/` – R source code
- `docs/` – project documentation
- `figures/` – generated figures
- `R/learning.R` – learning or exploratory code
- `R/clean_script.R` – data-cleaning code

The function temporarily changes the working directory to the project directory while creating the project structure and restores the original working directory on exit.

Value

Invisibly returns the absolute path to the initialized project directory.

Examples

```
## Not run:
medisr_init_project("my-project")

# Initialize a project in the current directory
medisr_init_project()

## End(Not run)
```

```
medisr_install_packages
```

Install all packages used for the Medical Statistics course

Description

A simple high-level function to install all packages needed.

Usage

```
medisr_install_packages()
```

Examples

```
## Not run:
medisr_install_packages()

## End(Not run)
```

```
prostate_cancer
```

Prostate Cancer Data

Description

Data for predicting prostate cancer from blood sample measurements.

Format

The prostate_cancer dataset contains two data frames, `prostate_cancer_train` and `prostate_cancer_test`, which can be loaded together using `data(prostate_cancer)`.

Both data frames contain the following variables:

cancer Prostate cancer confirmed by biopsy (1 = yes, 0 = no).

symptomatic Indicator of whether the patient has trouble urinating and/or decreased force of urine stream.

biomarkerk Concentration of an unspecified biomarker extracted from a blood sample, where k denotes the biomarker number 1-12.

`prostate_cancer_train` is the training set and `prostate_cancer_test` is the test set.

Details

The population consists of patients screened for prostate cancer by biopsy. The data are sham data constructed by replicating the structure of the Messidor-2 dataset. Variables have been renamed and the data have been split into a training set and a test set.

Source

Decenci re, E., Cazuguel, G., Zhang, X., Thibault, G., Qu llec, G., Lamard, M., Conze, P.-H., Massin, P., Cochener, B., Trone, C., Gain, P., Ordonez, R., Abr moff, M. D., and others. (2014). Feedback on a publicly distributed database: the Messidor database. *Image Analysis & Stereology*, 33(3), 231–234.

Abr moff, M. D., Folk, J. C., Han, D. P., Walker, J. D., Williams, D. F., Russell, S. R., Massin, P., Cochener, B., Gain, P., Tang, L., Lamard, M., Moga, D. C., Qu llec, G., and Niemeijer, M. (2013). Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmology*, 131(3), 351–357.

References

Decenci re, E., Cazuguel, G., Zhang, X., Thibault, G., Qu llec, G., Lamard, M., Conze, P.-H., Massin, P., Cochener, B., Trone, C., Gain, P., Ordonez, R., Abr moff, M. D., and others. (2014). Feedback on a publicly distributed database: the Messidor database. *Image Analysis & Stereology*, 33(3), 231–234.

Abr moff, M. D., Folk, J. C., Han, D. P., Walker, J. D., Williams, D. F., Russell, S. R., Massin, P., Cochener, B., Gain, P., Tang, L., Lamard, M., Moga, D. C., Qu llec, G., and Niemeijer, M. (2013). Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmology*, 131(3), 351–357.

SUA_data

Serum Uric Acid and Cardiovascular Disease Risk Factors

Description

This dataset contains serum uric acid measurements, cardiovascular disease risk factors, and basic demographic information for participants in a retrospective cohort study conducted every two years from 2010 to 2018 in Hangzhou, Zhejiang Province, Southeastern China. A total of 6,119 participants aged 40 years and above who underwent at least three physical examinations were enrolled.

Usage

SUA_data

Format

SUA_data:

A data frame with 6,119 rows and 15 columns:

ID Unique identifier for each participant

Age Age of the participant in years

Sex Sex of the participant. 1 = male; 2 = female

BMI Body mass index of the participant

SBP Systolic blood pressure of the participant

DBP Diastolic blood pressure of the participant
FBG Fasting blood glucose level of the participant
TC Total cholesterol level of the participant
Cr Serum creatinine level of the participant
GFR Glomerular filtration rate of the participant
UA Serum uric acid level in the participant's blood
Times Number of medical follow-up visits for the participant
Hypertension Hypertension status. 1 = no; 2 = yes
Hyperglycemia Hyperglycemia status. 1 = no; 2 = yes
Dyslipidemia Dyslipidemia status. 1 = no; 2 = yes

Source

Luo, Y., Wu, Q., Meng, R., Lian, F., Jiang, C., Hu, M., Wang, Y., & Ma, H. (2023). Associations of serum uric acid with cardiovascular disease risk factors: a retrospective cohort study in Southeastern China Data set. Zenodo. <https://doi.org/10.5061/dryad.z08kprrk1>
 Zenodo record: <https://zenodo.org/records/8292712>
 Data file: https://zenodo.org/records/8292712/files/SUA_CVDs_risk_factors.csv

thrombectomy	<i>Synthetic Thrombectomy Data</i>
--------------	------------------------------------

Description

Synthetic data reconstructed from clustered and summarized data from the Danish Stroke Register (DanStroke) and the Danish National Patient Registry. The source data contain patients with acute ischemic stroke who were eligible for thrombectomy in Denmark between 2011 and 2016.

Usage

thrombectomy

Format

thrombectomy:

A tibble containing synthetic data on patients with acute ischemic stroke:

cost Hospital-associated costs in euros during the first year following stroke

LoS Length of hospital stay in days following stroke

mRS_3month Modified Rankin Scale (mRS) score three months after stroke: 0 = no symptoms; 1 = no significant disability; 2 = slight disability; 3 = moderate disability; 4 = moderately severe disability; 5 = severe disability; 6 = dead

age_at_diagnosis Age in years at the time of stroke

year Year of stroke

time_to_death Time to death or censoring, in days
death Death during follow-up: 1 = yes, 0 = no
sex Sex: 1 = Male, 0 = Female
prior_diabetes Diabetes prior to stroke: 1 = yes, 0 = no
prior_hypertention Hypertension prior to stroke: 1 = yes, 0 = no
thrombectomy Thrombectomy received: 1 = yes, 0 = no

Details

The source population consists of patients with acute ischemic stroke in Denmark who were eligible for thrombectomy between 2011 and 2016. Eligibility in the source data was restricted to patients older than 18 years of age with a National Institutes of Health Stroke Scale (NIHSS) score above 15.

The Modified Rankin Scale (mRS) is a commonly used measure of functional disability following stroke. Scores range from 0 (no symptoms) to 6 (death), with higher scores indicating greater disability.

The data in this dataset are synthetic. A random subset of observations from the source data was summarized into clusters while preserving core patterns and correlations among the variables. The synthetic observations were subsequently simulated from these clusters. The data are intended for statistical analysis, teaching, and methodological demonstrations and should not be interpreted as original patient-level registry data.

Source

Danish Stroke Register (DanStroke), Danish Clinical Quality Program. <https://www.sundk.dk/kliniske-kvalitetsdatabaser/dansk-stroke-register/om-databasen/>

Danish National Patient Registry.

References

Danish Stroke Register (Dansk Stroke Register, DanStroke). Danish Clinical Quality Program.

Index

- * datasets
 - ad_data, [2](#)
 - ad_validation_logistic, [3](#)
 - beetle_dose, [4](#)
 - covid_data, [5](#)
 - dns, [6](#)
 - doctors, [8](#)
 - hamilton_d17, [10](#)
 - hd_data, [11](#)
 - height_data, [12](#)
 - james_lind, [12](#)
 - mabp, [13](#)
 - SUA_data, [17](#)
 - thrombectomy, [18](#)
- ad_data, [2](#), [4](#)
- ad_validation_logistic, [3](#)
- beetle_dose, [4](#)
- covid_data, [5](#)
- dns, [6](#)
- doctors, [8](#)
- eq5d, [9](#)
- eq5d_test (*eq5d*), [9](#)
- eq5d_train (*eq5d*), [9](#)
- hamilton_d17, [10](#)
- hd_data, [11](#)
- height_data, [12](#)
- james_lind, [12](#)
- mabp, [13](#)
- medisr_init_project, [15](#)
- medisr_install_packages, [16](#)
- prostate_cancer, [16](#)
- prostate_cancer_test
 - (*prostate_cancer*), [16](#)
- prostate_cancer_train
 - (*prostate_cancer*), [16](#)
- SUA_data, [17](#)
- thrombectomy, [18](#)