Package 'limai'

April 11, 2025

Title Linear modeling and AI decision making

Version 0.0.1

Author Xu Liu [aut,cre]

Maintainer Xu Liu <liu.xu@sufe.edu.cn>

Description Data sets used in the book ``Linear modeling and AI decision making''

License GPL (>=2)

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Depends R (>= 3.5.0)

Imports Matrix, glmnet

LazyData true

R topics documented:

CHARLS	*China Health and Retirement Longitudinal Study (CHARLS) data of Hebei, Shandong, and Fujian provinces.*

Description

The dataset contains the CHARLS data collected in Hebei, Shandong, and Fujian provinces.

Usage

```
data(CHARLS)
```

Format

A list object containing the following 3 variables:

Name	Type	Description
hebei	matrix	The CHARLS data of Hebei province.
		A matrix with 257 rows and 50 columns including y and 49 covariates v1,...,v49.
shandong	matrix	The CHARLS data of Shandong province.
		A matrix with 413 rows and 50 columns including y and 49 covariates v1,...,v49.
fujian	matrix	The CHARLS data of Fujian province.
		A matrix with 167 rows and 50 columns including y and 49 covariates v1,...,v49.

Details

Each matrix containing y, v1-v49 variables:

Name	Description
y	Annual support income of elderly people.
v1	Gender.
v2	Age.
v3	Marital status.
v4	Live alone (indicator).
v5	Live with a spouse (indicator).
v6	Live with children (indicator).
v7	Live with other members (e.g., parents).
v8	Health status.
v9	Pension income.
v10	Whether to receive a pension.
v11	The number of surviving children.
v12	Wage income per household.
v13	Net operating income per household.
v14	Net transfer income per household.
v15	The number of children with a college degree or above.
v16	The number of children earning over 10000 CNY each year.
v17	Emotional comfort.
v18	The number of household members.
v19	The number of deceased biological children.
v20	The number of surviving adopted children.
v21	The number of surviving sons.
v22	Financial support provided for parents.
v23	Financial support provided for other relatives.
v24	Net financial support received from other relatives.
v25	The number of types of disability.
v26	Have a chronic illness (indicator).
v27	Whether to receive a retirement pension.
v28	Retirement pension income.
v29	New rural pension income.
v30	All other pension income.
v31	Pension income of elderly households.
v32	The total financial assets of the elderly and their spouses.
v33	The wage income of the main household members.
v34	Government subsidies for individual families in the past year.
v35	Government subsidies for the main household members.
v36	The wage income of other family members in the past year.
v37	Government subsidies for other family members.
v38	Total government subsidies for families.

	v39	Government transfer income for households.
	v40	Net household income excluding private transfer income.
	v41	Net household income.
	v42	Net household income per capita.
	v43	Other net private transfer income of the elderly.
	v44	The family shared income received by the elderly.
	v45	Whether to complete high school education (indicator).
	v46	Annual net income from other sources.
	v47	Financial support provided for children.
	v48	(Additional covariate if any).
	v49	(Additional covariate if any).

For a detailed explanation of this dataset, please refer to Ren et al. (2006).

Source

The CHARLS data from https://charls.charlsdata.com/pages/Data/2015-charls-wave4/zh-cn.html

References

Ren, P., Liu, X., Zhang, X., Zhan, P., & Qiu, T. (2024). Integrative analysis of high-dimensional quantile regression with contrasted penalization. *Journal of Applied Statistics*, 1-17.

Examples

```
library(glmnet)
data(CHARLS)
data_hebei = CHARLS$hebei
y = data_hebei$y
x = data_hebei[,-1]
x = matrix(unlist(x),nrow = nrow(x))
fit_lasso = cv.glmnet(x,y,alpha = 1)
coef(fit_lasso,s = "lambda.min")
```

---

pollution	*Air Quality Index (AQI) for Chinese Cities (2022)*

Description

A multidimensional dataset containing weekly Air Quality Index (AQI), meteorological parameters, and socioeconomic indicators for 173 Chinese cities in 2022.

Usage

```
data(pollution)
```

Arguments

A list object containing 173 city entries with the following 10 variables:

Name	Type	Description
AQI	matrix	Air Quality Index, a matrix with 173 rows (cities) and 51 columns (weekly AQI values). Higher values indicate poorer air quality.
city	character	City names vector (length 173).
temp	numeric	Annual mean air temperature in °C.
dew	numeric	Annual mean dew point temperature in °C.
windD	numeric	Wind direction in degrees (0-360).
windS	numeric	Annual mean wind speed in m/s.
pres	numeric	Annual mean atmospheric pressure in hPa.
pop	numeric	Household resident population (unit: 10,000).
green	numeric	Green Covered Area as percentage of Completed Area (0-100).
second	numeric	Secondary Industry as Percentage to GRP (0-100).

Details

The data provides AQI data for 173 Chinese cities for the 51 weeks of 2022 and economic and meteorological related annual average data.

Source

- Air Quality Index form China National Environmental Monitoring Center(https://air.cnemc.cn:18007/)

- Meteorological Data from NOAA National Centers for Environmental Information (https://www.ncei.noaa.gov/)

- Socioeconomic data from China City Statistical Yearbook (https://www.stats.gov.cn/)

References

Guan, X., Li, Y., Liu, X., & You, J. (2025). Subgroup learning in functional regression models under the RKHS framework. *arXiv preprint arXiv:2503.01515*.

Examples

```
data(pollution)

# Explore AQI distribution for Beijing
bj_aqi <- as.numeric(pollution$AQI[pollution$city == "Beijing", ])
plot(bj_aqi,
     type = "l",
     main = "Weekly AQI in Beijing (2022)",
     xlab = "Week",
     ylab = "AQI")

# Correlation analysis
cor(pollution$temp, rowMeans(pollution$AQI, na.rm = TRUE))
```

Index