

Package: dauR (via r-universe)

June 2, 2026

Version 1.0

Title Datasets for ``Sampling and Data Analysis Using R: Theory and Practice''

Maintainer Professor Dr. Mohammad Shahidul Islam
<shahed.stat@gmail.com>

Description Provides several datasets used throughout the book ``Sampling and Data Analysis Using R: Theory and Practice'' by Islam (2025, ISBN:978-984-35-8644-5). The datasets support teaching and learning of statistical concepts such as sampling methods, descriptive analysis, estimation and basic data handling. These curated data objects allow instructors, students and researchers to reproduce examples, practice data manipulation and perform hands-on analysis using R.

License CC BY 4.0

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

Depends R (>= 3.5)

NeedsCompilation no

Author Professor Dr. Mohammad Shahidul Islam [aut, cre]

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

Date/Publication 2025-12-04 15:10:08 UTC

RemoteUrl <https://github.com/cran/dauR>

RemoteRef HEAD

RemoteSha f25b55816e343313077dbbc592f50011e703174a

Contents

case_study	2
eng_data	3

Health	3
likert_data	4
reading	5
smoke_class	6
students_data	6
twoway	7
weight1	8
weight2	9
Index	10

case_study	<i>This dataset corresponds to a case study in the book "Sampling and Data Analysis Using R: Theory and Practice"</i>
------------	---

Description

A family library contains 1100 books. The owner is interested in exploring some features of the existing books through a short survey. Thirty books have been randomly selected, and four characteristics have been measured: the number of pages (Page), weight in grams (Weight), surface area in square inches (Surface), and type of each book (Type).

Usage

```
case_study
```

Format

A data frame with 30 rows and 4 variables:

Page Number of pages for each book.

Weight Weight of the book in grams.

Surface Surface area of each book in square inches.

Type Categorical variable defining whether this is a religious, science, or story book.

Source

Generated for the book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(case_study, package = "dauR")
head(case_study)
```

eng_data

This dataset is to demonstrate the concept of Survival Analysis

Description

This dataset is about an engineering firm which monitors industrial machines over time to understand how long they operate before experiencing a mechanical failure. Each machine is tracked from installation until either it fails or the study period ends. The record includes each machine's operating age (in years) and maintenance type of whether it receives regular maintenance or on-demand maintenance.

Usage

eng_data

Format

A dataframe with 4 variables and 223 observations:

time Time to malfunction (in days)

status Failure indicator (1 = still operational/censored, 2 = malfunctioned)

age Age of machine at installation (in years)

maintenance Maintenance type (1 = On-demand, 2 = Regular)

Source

Generated for book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(eng_data, package = "dauR")
head(eng_data)
```

Health

This dataset was used in different chapters of the book "Sampling and Data Analysis Using R: Theory and Practice"

Description

This dataset consists of a sample of eighty laborers from a large factory. Measurements on four variables, namely Gender, Age in years, Diastolic blood pressure and BMI are reported.

Usage

Health

Format

A data frame with 80 rows and 4 variables:

Gender Gender of individuals

Age Age of the respondents

Diastolic Diastolic blood pressure of the respondents

BMI BMI of the respondents

Source

Generated for book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(Health, package = "dauR")
head(Health)
```

likert_data

This dataset refers to the concept of ordinal data

Description

A survey was conducted on a group of 15 students about teaching-learning environment of an institution. One part of the survey (comprises five questions) was dedicated for opinions about one recently completed course.

Usage

```
likert_data
```

Format

A data frame with 15 rows and 5 variables, all in likert-scale:

good_exp Question about whether this was a good experience

disturb_free Question about whether enjoyed disturb-free internet connection

support Question about whether got support from my institution

instructors Question about whether instructors were good in distant teaching and new technology

better Question about whether it is better than in-class teaching

Source

Generated for book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(likert_data, package = "dauR")
head(likert_data)
```

reading

This dataset demonstrates the concept of Mediation, Moderation & Confounding

Description

The dataset contains 200 observations and five variables, namely reading_time, vocab, test_score, access_resources and read_motiv.

Usage

```
reading
```

Format

A data frame with 200 rows and five variables:

reading_time Time spent in reading

vocab Measure of richness of vocabulary

test_score Score from the

access_resources Access to Resources

read_motiv Measure of motivation for reading

Source

Generated for book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(reading, package = "dauR")
head(reading)
```

`smoke_class`*This dataset is used to illustrate the concept of contingency table*

Description

This dataset shows smoking habit of 198 drivers of three types of vehicles. The types of vehicles are Bus, Truck and Taxi. Here our aim is to find whether there is any association between smoking habit and the occupation type, represented here by the type of vehicle driven.

Usage`smoke_class`**Format**

A data frame with 198 rows and 2 variables:

Class Type of vehicle driven

SmokingStatus Categorical verible, showing whether smoker of non-smoker

Source

Generated for book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(smoke_class, package = "dauR")
head(smoke_class)
```

`students_data`*This dataset is to demonstrate the concept of PCA and Factor Analysis*

Description

The dataset contains nine variables and 20 recently graduates from university and college. For a high-school job, students sat for exam on 7 subjects and then appeared for oral (viva-voce) exam. The subjects are Math, Physics, Chemistry, Statistics, Bengali literature, English literature and History.

Usage`students_data`

Format

A dataframe with 9 variables and 20 observations:

ID Serial number

Group Whether a student is from science or humanities background

Math Score in Mathematics

Physics Scores in Physics

Chemistry Score in Chemistry

Statistics Score in Statistics

English Score in English

Bengali Score in Bengali

History Score in History

Viva Score in Viva-voce

Source

Generated for book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(students_data, package = "dauR")
head(students_data)
```

twoway

This dataset illustrates the application of two-way ANOVA

Description

A researcher seeks to investigate whether an individual's life satisfaction (happiness) is associated with gender and working status. The variable Gender includes two categories: Male and Female, while Working_Status comprises three categories: Self-employed, Student and Job. A random sample of ten participants was selected from each category. Life satisfaction was measured on a scale ranging from 0 to 100, with higher scores indicating greater happiness. Therefore, the dependent variable is life satisfaction (happiness), and the independent variables are gender and working status.

Usage

twoway

Format

A data frame with 60 rows and 3 variables:

Happiness Measurement of happiness

Gender Gender of individuals

Working_Status Working status of individuals. The classes are "Job", "Self_employed" and "Student"

Source

Generated for book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(twoway, package = "dauR")
head(twoway)
```

weight1 *The datasets weight1 and weight2 correspond to pair-wise observations*

Description

This dataset shows weight of 16 individuals before taking medication.

Usage

```
weight1
```

Format

A vector of 16 observations:

weight1 weight of individuals

Source

Generated for book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(weight1, package = "dauR")
head(weight1)
```

weight2

The datasets weight1 and weight2 correspond to pair-wise observations

Description

This dataset shows weight of 16 individuals after taking medication.

Usage

```
weight2
```

Format

A vector of 16 observations:

weight2 weight of individuals

Source

Generated for book "Sampling and Data Analysis Using R: Theory and Practice" by Dr. Mohammad Shahidul Islam

Examples

```
data(weight2, package = "dauR")  
head(weight2)
```

Index

* datasets

- case_study, 2
- eng_data, 3
- Health, 3
- likert_data, 4
- reading, 5
- smoke_class, 6
- students_data, 6
- twoway, 7
- weight1, 8
- weight2, 9

case_study, 2

eng_data, 3

Health, 3

likert_data, 4

reading, 5

smoke_class, 6

students_data, 6

twoway, 7

weight1, 8

weight2, 9