Package: sportsfeatures (via r-universe)

July 1, 2026

Type Package

Title Longitudinal Sports Analytics Asset and Workload Feature Processing

Version 0.1.0

Description A synthetic, longitudinal athletic dataset generated through a transparent, rule-based simulation engine. Captures individual activity sessions across multiple athletes, environmental conditions, and physiological responses. Specifically designed as an alternative to legacy teaching datasets by introducing realistic hierarchical repeated measures, complex two-way covariate interactions, and a deliberate Missing Not At Random (MNAR) tracking mechanism suitable for advanced imputation workflows. Methodologies implemented are based on van Buuren (2018) <doi:10.1201/9780429492259> and Bates et al. (2015) <doi:10.18637/jss.v067.i01>.

License MIT + file LICENSE

Depends R (>= 4.1.0)

Imports tibble, mice, modelsummary, lme4

Suggests tidyverse

Encoding UTF-8

LazyData true

Config/roxygen2/version 8.0.0

Config/Needs/editorial MNAR Buuren et al

NeedsCompilation no

Author Mohammad Abbas [aut, cre]

Maintainer Mohammad Abbas <ma.abbas3107@gmail.com>

Repository https://artisto123.r-universe.dev

Date/Publication 2026-06-30 21:11:13 UTC

RemoteUrl https://github.com/cran/sportsfeatures

RemoteRef HEAD

RemoteSha 504f7d775d0f21e44deb4da5962d91f8379adcd4

Contents

get_sportsdata 2
sports_features 2
sports_features_missing 4

Index **6**

`get_sportsdata` *Access Sports Feature Datasets*

Description

A convenient helper function to quickly load and return the package's internal sports features data assets directly into an active variable.

Usage

```
get_sportsdata(type = c("complete", "missing"))
```

Arguments

`type`	A character string specifying which dataset variant to load: `"complete"` (default) or `"missing"`.

Value

A tibble/data.frame containing the requested sports feature dataset.

Examples

```
# Get the clean complete dataset
clean_data <- get_sportsdata(type = "complete")

# Get the dataset containing systematic missingness
missing_data <- get_sportsdata(type = "missing")
```

`sports_features` *Comprehensive Sports Features Dataset*

Description

Comprehensive Sports Features Dataset

Usage

```
sports_features
```

Format

A tibble or data frame with 25 variables describing athlete sessions and performance metrics:

session_id Unique alphanumeric identifier for each training session.

athlete_id Unique alphanumeric identifier for each athlete.

datetime Timestamp of when the training session occurred.

activity_type Type of exercise performed (e.g., running, cycling, swimming).

region Geographical area where the session took place.

distance_km Total distance covered during the session in kilometers.

weather_type Weather condition during the session (e.g., sunny, rainy, cloudy).

temperature_c Ambient outdoor temperature in degrees Celsius.

personal_status Pre-activity physical or mental status reported by the athlete.

is_group_activity Logical indicator (TRUE/FALSE) if the session was done with a group.

gender Categorical gender of the athlete.

age Age of the athlete in years.

base_fitness Baseline fitness score of the athlete.

base_speed Baseline average speed capability of the athlete.

base_stamina Baseline stamina level of the athlete.

base_weight Baseline body weight of the athlete in kilograms.

resting_heart_rate Baseline resting heart rate in beats per minute (bpm).

device_type Type of tracking device used during the session.

speed_kmh Average speed maintained throughout the session in km/h.

duration_min Total duration of the training session in minutes.

heart_rate_avg Average heart rate monitored during the session in bpm.

calories_burned Estimated total energy expenditure in kilocalories (kcal).

exhaustion_level Subjective exhaustion level reported after the session.

hydration_status Hydration level (%) recorded during or after the session.

fatigue_score Calculated post-activity fatigue accumulation score.

Details

A rich, synthetic sports analytics dataset containing tracking metrics, environmental contexts, physiological markers, and performance data for athletes.

Source

Synthesized sports features analytics framework.

Examples

```
library(tidyverse)
library(lme4)

# Load the package data
data("sports_features")

# Downsample data for the example to ensure fast execution time (< 2.5s)
demo_data <- head(sports_features, 500)

# ------------------------------------------------------
# DEMO 1: Linear Regression (Fixed Effects)
# Predicting fatigue score based on workload metrics
# ------------------------------------------------------
lm_model <- lm(fatigue_score ~ distance_km + duration_min + speed_kmh + temperature_c,
               data = demo_data)

summary(lm_model)

# ------------------------------------------------------
# DEMO 2: Linear Mixed-Effects Model (Hierarchical MML)
# Controlling for variation across individual athletes (athlete_id)
# ------------------------------------------------------
mml_model <- lmer(fatigue_score ~ distance_km + duration_min + speed_kmh + temperature_c +
                    (1 | athlete_id),
                  data = demo_data)
summary(mml_model)
```

---

sports_features_missing

Comprehensive Sports Features Dataset (With Missing Values)

Description

Comprehensive Sports Features Dataset (With Missing Values)

Usage

```
sports_features_missing
```

Format

A tibble or data frame with 25 variables containing structured missing values.

Details

A variant of the core sports analytics dataset containing structured missingness (NA values) across performance tracking columns to demonstrate imputation workflows.

Source

Synthesized sports features analytics framework.

Index

∗ **datasets**
    sports_features, 2
    sports_features_missing, 4

get_sportsdata, 2

sports_features, 2
sports_features_missing, 4