

Package: lsda (via r-universe)

October 16, 2026

Type Package

Title Course Materials for Longitudinal and Spatial Data Analysis

Version 0.2.1

Description Datasets, teaching helper functions, interactive Shiny applications and 'learnr' tutorials supporting MATH48132 and MATH68132 Longitudinal and Spatial Data Analysis at the University of Manchester. Installing this package provides every dataset and dependency used across the eleven teaching weeks in a single step, so that tutorial code runs without students having to locate and read data files themselves.

License CC BY-NC-SA 4.0

Encoding UTF-8

LazyData true

Depends R (>= 4.1)

Imports graphics, grDevices, stats

Suggests CARBayes, geepack, ggplot2, gstat, knitr, learnr, MASS, nlme, rmarkdown, sf, shiny, spatstat, spdep, testthat (>= 3.0.0)

Config/testthat/edition 3

RoxygenNote 7.3.1

URL <https://github.com/olatunjijohnson/lsda>

BugReports <https://github.com/olatunjijohnson/lsda/issues>

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

Date/Publication 2026-10-16 20:28:11 UTC

RemoteUrl <https://github.com/olatunjijohnson/lsda>

RemoteRef HEAD

RemoteSha b54794893e1280343f4e907c27d533fd3ae2289c

Contents

cattle	2
cd4	3
cov_structure	4
emp_variogram_mc	5
fev1	5
galicia	6
galicia_grid	7
galicia_pred	7
gm_grid	8
gm_no2	8
ichs	9
liberia	10
liberia_grid	10
loaloa	11
nw_obesity	12
run_app	12
scotland_lip	13
sir	14
spaghetti_plot	14
tef	15
Index	17

cattle	<i>Cattle bodyweight data</i>
--------	-------------------------------

Description

Bodyweight of 60 cattle in two treatment groups, recorded at 11 equally spaced time points. A balanced design: every animal is measured at every time, which makes it the natural first example for exploring within-subject correlation.

Usage

cattle

Format

A data frame with 60 rows and 12 columns:

- group** treatment group, coded 0 and 1
- t1** bodyweight in kilograms at time 1
- t2** bodyweight in kilograms at time 2
- t3** bodyweight in kilograms at time 3
- t4** bodyweight in kilograms at time 4

- t5** bodyweight in kilograms at time 5
- t6** bodyweight in kilograms at time 6
- t7** bodyweight in kilograms at time 7
- t8** bodyweight in kilograms at time 8
- t9** bodyweight in kilograms at time 9
- t10** bodyweight in kilograms at time 10
- t11** bodyweight in kilograms at time 11

Source

Used throughout weeks 1-3 of MATH48132/68132. Originally distributed as a headerless text file; column names added here.

Examples

```
data(cattle)
spaghetti_plot(cattle, group = "group")
round(cor(cattle[, -1]), 2)
```

cd4

CD4+ cell counts in HIV seroconverters

Description

Longitudinal CD4+ counts with time measured in years relative to seroconversion, so negative times precede it. Highly unbalanced in both the number and the spacing of visits.

Usage

```
cd4
```

Format

A data frame with 2376 rows and 8 columns:

- time** years since seroconversion (negative before)
- cd4** CD4+ cell count
- age** age in years, centred
- cig** cigarette smoking
- drug** recreational drug use, 0/1
- partner** number of sexual partners
- cesd** depression score (CES-D scale)
- subject** subject identifier

Source

Multicenter AIDS Cohort Study.

cov_structure

*Build a within- or between-subject covariance matrix***Description**

Constructs the covariance structures met in week 2, so that slides, notes and tutorials all display the same matrices. Useful for showing students what a structure **is** before asking them to fit one.

Usage

```
cov_structure(
  structure = c("independent", "exchangeable", "ar1", "arma11", "unstructured"),
  n = 5,
  sigma2 = 1,
  rho = 0.5,
  theta = 0.5,
  values = NULL
)
```

Arguments

structure	One of "independent", "exchangeable" (compound symmetry), "ar1", "arma11", "unstructured".
n	Dimension of the matrix.
sigma2	Variance.
rho	Correlation parameter. For "arma11" this is the AR parameter ϕ .
theta	Moving-average parameter, "arma11" only.
values	Optional vector of correlations for "unstructured", filling the lower triangle by column.

Value

A symmetric n by n covariance matrix.

Examples

```
cov_structure("ar1", n = 5, rho = 0.8)
cov_structure("exchangeable", n = 4, rho = 0.5, sigma2 = 2)
```

emp_variogram_mc	<i>Monte Carlo envelope test for spatial correlation</i>
------------------	--

Description

Repeatedly permutes the values across locations and recomputes the empirical variogram. Under the null of no spatial correlation the observed variogram should sit inside the envelope of the permuted ones. This is the test used in the week 7 tutorial.

Usage

```
emp_variogram_mc(x, y, z, nsim = 999, nbins = 15, max_dist = NULL)
```

Arguments

x, y	Coordinates.
z	Values, typically residuals from a covariate-only fit.
nsim	Number of permutations.
nbins	Number of distance bins.
max_dist	Largest distance considered. Defaults to half the maximum separation.

Value

A list with the bin midpoints, the observed semivariance, the simulated envelope, and a permutation p-value.

Examples

```
data(galicia)
fit <- lm(loglead ~ pm10, data = galicia)
res <- emp_variogram_mc(galicia$long, galicia$lat, resid(fit), nsim = 99)
res$p_value
```

fev1	<i>Lung function growth in Topeka schoolgirls (FEV1)</i>
------	--

Description

Repeated spirometry measurements on 299 girls from Topeka, Kansas, enrolled in the Six Cities Study of Air Pollution and Health. Girls entered at around age 6-10 and were followed roughly annually. The number of observations per girl ranges from 1 to 12, so the design is markedly unbalanced – the counterpart to the balanced [cattle](#) data.

Usage

```
fev1
```

Format

A data frame with 1993 rows and 8 columns:

id girl identifier
ht height in metres at the visit
age age in years at the visit
baseht height in metres at the first visit
baseage age in years at the first visit
logfev1 log of FEV1
fev1 forced expiratory volume in one second, in litres
logfev1_ht $\log(\text{FEV1} / \text{height})$, the usual response

Note

The 2025 course tree also contained `air_pollution.csv`, which is this same data with `fev1` and `logfev1_ht` dropped. It is not reproduced here; use `fev1`.

Source

Six Cities Study of Air Pollution and Health.

Examples

```
data(fev1)
table(table(fev1$id)) # observations per girl: the design is unbalanced
```

galicia

Lead concentration in moss samples, Galicia

Description

Lead concentration measured in moss at 132 sampling stations across Galicia, north-west Spain, with a co-located PM10 measurement. The running example for the geostatistical weeks: variogram estimation in week 7, model fitting and kriging in week 8.

Usage

```
galicia
```

Format

A data frame with 132 rows and 5 columns:

lead lead concentration
loglead log lead concentration, the modelled response
long longitude in decimal degrees
lat latitude in decimal degrees
pm10 particulate matter under 10 micrometres

Name clash

RiskMap also ships a `galicia`, with three columns (`x`, `y`, `lead`). Both packages are attached in the week 7 and 8 tutorials, so load this one with `data(galicia, package = "lsda")`.

See Also

[galicia_grid](#) and [galicia_pred](#) for prediction.

<code>galicia_grid</code>	<i>Prediction grid for Galicia</i>
---------------------------	------------------------------------

Description

A regular grid covering the Galicia study region, for mapping kriging predictions in week 8.

Usage

```
galicia_grid
```

Format

A data frame with 1508 rows and 2 columns, `X` and `Y`, in kilometres.

See Also

[galicia](#)

<code>galicia_pred</code>	<i>Covariate surface for Galicia prediction</i>
---------------------------	---

Description

PM10 and coordinates at each prediction location, so that models including PM10 as a covariate can be used for prediction.

Usage

```
galicia_pred
```

Format

A data frame with 1508 rows and 3 columns: `pm10`, `long`, `lat`.

See Also

[galicia](#)

gm_grid	<i>Prediction grid for Greater Manchester</i>
---------	---

Description

A regular grid covering Greater Manchester, carrying the same covariates as [gm_no2](#), for mapping predictions in week 8.

Usage

```
gm_grid
```

Format

An sf data frame with 4996 features and 4 columns:

elevation elevation in metres

pop population count of the cell

road distance to the nearest major road, in metres

geometry grid point, EPSG:27700

Examples

```
data(gm_grid)
nrow(gm_grid)
```

gm_no2	<i>Nitrogen dioxide in Greater Manchester</i>
--------	---

Description

Nitrogen dioxide (NO₂) measured at 25 monitoring stations across Greater Manchester, the running geostatistical example in weeks 7 and 8: the variogram of the residuals in week 7, and model fitting and prediction in week 8.

Usage

```
gm_no2
```

Format

An sf data frame with 25 features and 8 columns:

station monitoring station name
NO2 nitrogen dioxide concentration
PM10, PM25 particulate matter concentrations
elevation elevation in metres
road distance to the nearest major road, in metres
pop population count of the surrounding cell
geometry station location, EPSG:27700

See Also

[gm_grid](#) for the prediction grid.

Examples

```
data(gm_no2)
summary(gm_no2$NO2)
```

ichs

Indonesian Children's Health Study (ICHS)

Description

250 preschool children examined at entry and then every three months for a year. The question of interest is whether vitamin A deficiency raises the risk of respiratory infection. The binary response and repeated measures make this the standard GEE and GLMM example.

Usage

```
ichs
```

Format

A data frame with 1500 rows and 6 columns:

id child identifier
infection respiratory infection present, 0 = absent, 1 = present
time months since the first visit
gender 0 = male, 1 = female
vitA vitamin A status, 0 = sufficient, 1 = deficient
age age in years

Source

Indonesian Children's Health Study. Used in week 5.

liberia	<i>Malaria incidence and child growth in Liberia</i>
---------	--

Description

Cluster-level malaria incidence with height-for-age z-scores, used as the week 8 challenge dataset.

Usage

```
liberia
```

Format

A data frame with 267 rows and 5 columns:

cluster cluster identifier

height_for_age_zscores mean height-for-age z-score

longitude, latitude decimal degrees

malaria_incidence_rate incidence rate

Name clash

RiskMap ships a different dataset also called `liberia` (onchocerciasis counts, used in the week 8 challenge). Name the package for whichever you want.

liberia_grid	<i>Prediction grid for Liberia</i>
--------------	------------------------------------

Description

Prediction grid for Liberia

Usage

```
liberia_grid
```

Format

A data frame with 3837 rows and 2 columns, X and Y.

See Also

[liberia](#)

loaloa	<i>Loa loa</i> prevalence in Cameroon and Nigeria
--------	---

Description

Village-level survey results for *Loa loa* (eyeworm) from 197 villages across Cameroon and southern Nigeria, with elevation and NDVI vegetation covariates. A binomial response, so it extends the Gaussian geostatistics of weeks 7-8 to the generalised case.

Usage

```
loaloa
```

Format

A data frame with 197 rows and 11 columns:

row row identifier

villcode village identifier

longitude, latitude decimal degrees

no_exam number of individuals tested

no_inf number testing positive

elevation metres above sea level

mean9901, max9901, min9901, stdev9901 NDVI summaries, 1999-2001

Name clash

RiskMap also ships a `loaloa` — the same survey, but with upper-case column names. Load this one with `data(loaloa, package = "lsda")`.

References

Diggle, P.J., Thomson, M.C., Christensen, O.F., Rowlingson, B., Obsomer, V., Gardon, J., Wanji, S., Takougang, I., Enyong, P., Kamgno, J., Remme, H., Boussinesq, M. and Molyneux, D.H. (2007). Spatial modelling and prediction of *Loa loa* risk: decision making under uncertainty.

nw_obesity

*Childhood obesity and deprivation in North West England***Description**

Childhood obesity and income deprivation for the 33 local authority districts of North West England, the areal example in weeks 9 and 10.

Usage

nw_obesity

Format

An sf data frame with 33 features and 8 columns:

lad_code, lad_name local authority district code and name

num_obese number of children measured as obese

total_sampled number of children measured

prop_obese num_obese / total_sampled

idaci_score Income Deprivation Affecting Children Index, average score

idaci_rank rank of the IDACI score (lower is more deprived)

geometry district polygon, EPSG:27700

Source

National Child Measurement Programme small-area data and the English Indices of Deprivation 2019, both published by gov.uk.

Examples

```
data(nw_obesity)
summary(nw_obesity$prop_obese)
```

run_app

*Launch a course Shiny application***Description**

Launch a course Shiny application

Usage

```
run_app(app = NULL, ...)
```

Arguments

app Application name. Omit to list what is available.
... Passed to `runApp`.

Value

Called for its side effect.

Examples

```
## Not run:
run_app()           # list the apps
run_app("variogram")

## End(Not run)
```

scotland_lip	<i>Male lip cancer in Scottish counties, 1975-1980</i>
--------------	--

Description

Observed and expected male lip cancer cases in each of the 56 Scottish counties, with the proportion of the workforce in agriculture, fishing and forestry (AFF) as a proxy for sunlight exposure. The running example for the areal weeks: Moran's I in week 9, CAR and ICAR models in week 10.

Usage

```
scotland_lip
```

Format

An sf data frame with 56 features and 6 columns:

county county name
cases observed cases
expected expected cases given the age-sex structure
aff proportion employed in agriculture, fishing or forestry
sir standardised incidence ratio, cases / expected
geometry county polygon

Coordinate reference system

The original shapefile declared no CRS and its coordinates were British National Grid divided by 1000. Here the geometry is rescaled to metres and tagged EPSG:27700, so basemaps and area calculations behave. Adjacency – the only property weeks 9 and 10 rely on – is unchanged by rescaling.

Source

Clayton and Kaldor (1987); Breslow and Clayton (1993).

Examples

```
data(scotland_lip)
head(scotland_lip$sir)
```

sir	<i>Standardised incidence ratio</i>
-----	-------------------------------------

Description

Standardised incidence ratio

Usage

```
sir(observed, expected)
```

Arguments

- observed Observed counts.
- expected Expected counts.

Value

Numeric vector of ratios.

Examples

```
data(scotland_lip)
head(sir(scotland_lip$cases, scotland_lip$expected))
```

spaghetti_plot	<i>Spaghetti plot for longitudinal data</i>
----------------	---

Description

One line per subject. Accepts either wide data (one row per subject, one column per time) or long data (one row per observation) via id, time and response.

Usage

```
spaghetti_plot(
  data,
  id = NULL,
  time = NULL,
  response = NULL,
  group = NULL,
  times = NULL,
  col = "grey30",
  alpha = 0.6,
  ...
)
```

Arguments

<code>data</code>	A data frame.
<code>id, time, response</code>	Column names, as strings, for long-format data. Leave NULL for wide format.
<code>group</code>	Optional grouping column; lines are coloured by it and a legend is drawn.
<code>times</code>	Numeric times for wide format. Defaults to <code>1:ncol</code> .
<code>col</code>	Line colour when group is NULL.
<code>alpha</code>	Line opacity between 0 and 1.
<code>...</code>	Passed to plot .

Value

Invisibly, the plotted data. Called for the side effect.

Examples

```
data(cattle)
spaghetti_plot(cattle, group = "group")

data(fev1)
spaghetti_plot(fev1, id = "id", time = "age", response = "logfev1_ht")
```

 tef

Teacher effectiveness survey summaries

Description

Summary statistics from a teacher effectiveness survey, used as a secondary example.

Usage

```
tef
```

Format

A data frame with 37 rows and 11 columns.

Index

* datasets

- cattle, [2](#)
- cd4, [3](#)
- fev1, [5](#)
- galicia, [6](#)
- galicia_grid, [7](#)
- galicia_pred, [7](#)
- gm_grid, [8](#)
- gm_no2, [8](#)
- ichs, [9](#)
- liberia, [10](#)
- liberia_grid, [10](#)
- loaloe, [11](#)
- nw_obesity, [12](#)
- scotland_lip, [13](#)
- tef, [15](#)

cattle, [2](#), [5](#)

cd4, [3](#)

cov_structure, [4](#)

emp_variogram_mc, [5](#)

fev1, [5](#)

galicia, [6](#), [7](#)

galicia_grid, [7](#), [7](#)

galicia_pred, [7](#), [7](#)

gm_grid, [8](#), [9](#)

gm_no2, [8](#), [8](#)

ichs, [9](#)

liberia, [10](#), [10](#)

liberia_grid, [10](#)

loaloe, [11](#)

nw_obesity, [12](#)

plot, [15](#)

run_app, [12](#)

runApp, [13](#)

scotland_lip, [13](#)

sir, [14](#)

spaghetti_plot, [14](#)

tef, [15](#)