

Package: incomeLato (via r-universe)

October 1, 2026

Title Relative Income Position for 'Latinobarometro' Respondents from a Frozen Imputation Model

Version 0.1.1

Description Applies a frozen multinomial model (version 'v1.5') that predicts each respondent's relative income position within their country and survey year -- the probabilities of belonging to the bottom 50 percent, the middle 40 percent and the top 10 percent -- to 'Latinobarometro' microdata that the user downloads from the official source. The model was trained on national household surveys and validated against 'LAPOP AmericasBarometer' data for 15 Latin American countries. It uses five predictors (sex, age, education, car ownership, washing machine ownership) plus a country effect, and the package harmonizes them across waves 1995-2024 through a manually curated crosswalk. The package ships only model parameters and variable-name metadata: it contains no survey microdata and no statistics computed from it. Probabilities are meant as continuous weights for aggregate analysis, not for classifying individuals.

License MIT + file LICENSE

URL <https://github.com/mrbpedro/incomeLato>,
<https://mrbpedro.r-universe.dev/incomeLato>

BugReports <https://github.com/mrbpedro/incomeLato/issues>

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Depends R (>= 3.5.0)

Imports dplyr, Matrix, glmnet, haven, stats, utils, tools

Suggests knitr, readstata13, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Config/pak/sysreqs make libx11-dev zlib1g-dev
Repository https://mrbpedro.r-universe.dev
Date/Publication 2026-10-01 15:34:31 UTC
RemoteUrl https://github.com/mrbpedro/incomeLato
RemoteRef HEAD
RemoteSha af2f999da69e4f4fa3f49211c02603a5205681e0

Contents

educ3_validated_years	2
harmonize_core_predictors	3
harmonize_educ3_latinobarometro	3
income_lato	4
predict_latinobarometro_income	6
prepare_latinobarometro_income_inputs	7
simulate_latinobarometro_wave	8
Index	10

educ3_validated_years	List the years with validated education logic
-----------------------	---

Description

Years outside this table are not processed by `harmonize_educ3_latinobarometro()` – the function refuses instead of guessing the scheme/variable for an unseen wave. Adding a new wave requires updating `inst/extdata/educ3_validated_years.csv` manually (see the package README).

Usage

```
educ3_validated_years()
```

Value

data.frame with `year`, `raw_variable_name`, `scheme`, `respondent_or_head`, `confidence_note` – one row per validated wave (1995-2024; 2024 with reduced confidence, see `confidence_note`).

Examples

```
anos <- educ3_validated_years()
anos[, c("year", "scheme", "confidence_note")]
```

```
harmonize_core_predictors
```

Harmonize the sex/age/car/washing machine/country/weight predictors for one wave

Description

Harmonize the sex/age/car/washing machine/country/weight predictors for one wave

Usage

```
harmonize_core_predictors(raw_data, year)
```

Arguments

raw_data	raw data.frame of one Latinobarometro wave.
year	survey year of the wave.

Value

data.frame with sexo_final, idade_final, owns_car, owns_washing_machine, country_raw, weight, and a column of flags for missing predictors.

Examples

```
# 100% synthetic data -- no real respondent value
sim <- simulate_latinobarometro_wave(2018, n = 50)
core <- harmonize_core_predictors(sim, 2018)
str(core)
```

```
harmonize_educ3_latinobarometro
```

Harmonize educ3_final for one Latinobarometro wave

Description

Builds the 3-category education variable (basic_or_less/secondary/higher_or_more) used by the income model, using the table of validated years (not the general crosswalk – see the note in the file header). Years outside the table of validated years are NOT processed – the function stops with an error and never guesses the scheme/variable for an unseen wave.

Usage

```
harmonize_educ3_latinobarometro(
  raw_data,
  year,
  on_unvalidated_year = c("error", "warning")
)
```

Arguments

`raw_data` raw data.frame of one Latinobarometro wave (columns with the original names of the file downloaded by the user).

`year` survey year of the wave (numeric).

`on_unvalidated_year` "error" (default, refuses) or "warning" (issues a strong warning and returns NA for the whole column – use with caution, never in production without manual review).

Details

The 2024 wave has reduced confidence – `readstata13` did not recover the value labels of this particular file, so the codes were inferred from the identical historical order confirmed in 9 other waves, not by direct confirmation from labels. Use with caution when `educ3_quality_flag` is "confianza_reduzida_2024_codigos_inferidos".

Value

data.frame with 1 row per respondent: `educ3_final`, `educ3_source_variable`, `educ3_source_scheme`, `educ3_quality_flag`.

Examples

```
# 100% synthetic data -- no real respondent value
sim <- simulate_latinobarometro_wave(2018, n = 50)
educ <- harmonize_educ3_latinobarometro(sim, 2018)
table(educ$educ3_final, useNA = "ifany")

# year outside the table of validated years -- refuses, does not guess
tryCatch(
  harmonize_educ3_latinobarometro(sim, 2099),
  error = function(e) message("Refused as expected: ", conditionMessage(e))
)
```

income_lato

Apply the v1.5 income model to Latinobarometro microdata (shortcut)

Description

Combines `prepare_latinobarometro_income_inputs()` and `predict_latinobarometro_income()` in a single step. Equivalent to: `predict_latinobarometro_income(prepare_latinobarometro_income_inputs(raw_data, years, countries), model_version, include_2024)`, plus a run summary emitted with `message()` at the end: waves detected, waves excluded by default (2024), waves skipped because they could not be read, and countries with no prediction at all, with the reason.

Usage

```
income_lato(
  raw_dir,
  years = NULL,
  countries = NULL,
  model_version = "v1_5",
  include_2024 = FALSE
)
```

Arguments

<code>raw_dir</code>	path to one wave's <code>.dta/</code> <code>.sav</code> file, or to a directory with one file per wave (downloaded by the user – not distributed with the package).
<code>years</code>	vector of years to process (default: every file found; see educ3_validated_years() for the years with validated education harmonization). For a single file without the year in its name, give the year here.
<code>countries</code>	vector of ISO3 codes to filter on (default: every country present in the file).
<code>model_version</code>	model version to apply (the only one available in this v0.1: "v1_5", the frozen model).
<code>include_2024</code>	if FALSE (default), drops the 2024 wave from the result – that wave's education codes were inferred from the historical order, not confirmed by labels (the 2024 <code>.dta</code> is only read by <code>readstata13</code> , without value labels). If TRUE, keeps 2024 with an additional warning explaining the specific reason for the reduced confidence.

Details

The package does NOT distribute Latinobarometro microdata – `raw_dir` must point to files the user downloaded from the official source (latinobarometro.org).

Value

`data.frame`: the harmonized data with `prob_bottom50`, `prob_middle40`, `prob_top10`, `elegivel`, `motivo_exclusao` and the per-country confidence flags APPENDED (not a separate object matched by position). The probabilities are continuous weights for describing the composition of groups (e.g., party support among the bottom 50%), within a wave and across waves. Do not classify individual respondents by a hard cutoff, and do not read `mean(prob_bottom50)` by country-wave as the size of the bottom half, nor its change over time as a trend (the aggregate margin drifts for reasons internal to the model; see `veredito_deriva_temporal` and `flag_margin`). A warning with this notice is issued on the first call of the session; see also `attr(result, "recommended_use") / attr(result, "prohibited_use")`. `motivo_exclusao` codes: `venezuela_architectural_quarantine`, `country_not_in_15`, `outside_country_valid_window`, `missing_or_invalid_predictor`.

Examples

```
# One wave, single file. 100% synthetic data written as a temporary
# .dta -- no real Latinobarometro respondent.
sim <- simulate_latinobarometro_wave(2011, n = 20)
arquivo <- file.path(tempdir(), "Latinobarometro_2011_sim.dta")
```

```

haven::write_dta(sim, arquivo)
res <- income_lato(arquivo)
head(res[, c("iso3", "prob_bottom50", "motivo_exclusao")])
unlink(arquivo)

## Not run:
# real use, with microdata downloaded by the user:
res2018 <- income_lato("~/downloads/Latinobarometro_2018_Eng_Stata_v20190303.dta")
todas <- income_lato("~/downloads/latinobarometro_dtas")

## End(Not run)

```

predict_latinobarometro_income

Apply the v1.5 income model to the prepared data

Description

Apply the v1.5 income model to the prepared data

Usage

```

predict_latinobarometro_income(
  data,
  model_version = "v1_5",
  include_2024 = FALSE
)

```

Arguments

data	data.frame returned by prepare_latinobarometro_income_inputs() .
model_version	for now only "v1_5" (the frozen model).
include_2024	if FALSE (default), drops the 2024 wave from the result (its education codes were inferred, not confirmed by labels). If TRUE, keeps it with an additional strong warning.

Value

data with the appended columns: prob_bottom50, prob_middle40, prob_top10, elegivel, motivo_exclusao, confianca_alvo_treino, flag_tier_reduzido, flag_piso_confianca, flag_fonte_censitaria, veredito_deriva_temporal, flag_margin. Attributes recommended_use/prohibited_use attached to the object. motivo_exclusao codes: venezuela_architectural_quarantine, country_not_in_15, outside_country_valid_window, missing_or_invalid_predictor.

Note

The probabilities are continuous weights for describing the composition of groups (e.g., party support among the bottom 50%), within a wave and across waves. Do NOT classify individual respondents into bottom50/middle40/top10 by hard cutoff. Do NOT read mean(prob_bottom50) by country-wave as the size of the bottom half, nor its change over time as a trend: the aggregate margin drifts for reasons internal to the model (see `veredito_deriva_temporal` and `flag_margin`). A warning with this notice is issued on the first call of the session.

Examples

```
# 100% synthetic data -- no real respondent value
sim <- simulate_latinobarometro_wave(2018, n = 30)
core <- harmonize_core_predictors(sim, 2018)
educ <- harmonize_educ3_latinobarometro(sim, 2018)
cw_pais <- read.csv(system.file("extdata", "country_crosswalk.csv", package = "incomelato"))
banco <- data.frame(year = 2018, country_raw = core$country_raw,
  respondent_row_id = seq_len(nrow(sim)), weight = core$weight,
  sexo_final = core$sexo_final, idade_final = core$idade_final,
  educ3_final = educ$educ3_final, owns_car = core$owns_car,
  owns_washing_machine = core$owns_washing_machine,
  source_file = "sintetico", harmonization_flags = NA_character_)
banco <- merge(banco, cw_pais, by.x = "country_raw", by.y = "country_numeric", all.x = TRUE)
resultado <- predict_latinobarometro_income(banco)
head(resultado[, c("iso3", "prob_bottom50", "elegivel")])
```

```
prepare_latinobarometro_income_inputs
```

Prepare the income predictors from the user's raw microdata

Description

Reads Latinobarometro .dta/.sav files (downloaded by the user from the official source – this package does NOT distribute microdata), detects the wave from the file name, and harmonizes the income model's predictors (sex/age/car/washing machine/country/weight via a minimal crosswalk; education via `harmonize_educ3_latinobarometro()`, restricted to validated years).

Usage

```
prepare_latinobarometro_income_inputs(raw_dir, years = NULL, countries = NULL)
```

Arguments

<code>raw_dir</code>	directory with the raw .dta or .sav files, or the path to a single file. One file per wave, with the year somewhere in the name (e.g. "Latinobarometro_2018_Eng_Stata_v20190303.dta"). If the same wave is present in both formats, keep only one in the directory. For a single file without the year in its name, give it in years.
<code>years</code>	vector of years to process (default: every file found in <code>raw_dir</code>).
<code>countries</code>	vector of ISO3 codes to filter on (default: all).

Value

data.frame with year, iso3, country_raw, respondent_row_id, weight, sexo_final, idade_final, educ3_final, Attributes ondas_detectadas (waves found and requested) and ondas_ignoradas (waves whose harmonization failed).

Examples

```
# runnable example with 100% synthetic data written as a temporary
# .dta -- no real Latinobarometro microdata is used or distributed. In
# real use, `raw_dir` points to .dta files downloaded by the user from
# latinobarometro.org. Year 2011 is chosen here because its raw variable
# names have no dot (e.g. "REEDUC1"), which haven::write_dta() accepts;
# some years (e.g. 2018, "REEDUC.1") have a dot in the original name and
# cannot be rewritten as .dta this way -- this does not affect reading
# files downloaded from the Latinobarometro, only the synthetic writing
# in this example.
sim <- simulate_latinobarometro_wave(2011, n = 30)
tmp_dir <- tempfile("lb_sim_"); dir.create(tmp_dir)
haven::write_dta(sim, file.path(tmp_dir, "Latinobarometro_2011_sim.dta"))
banco <- prepare_latinobarometro_income_inputs(tmp_dir, years = 2011)
str(banco)
unlink(tmp_dir, recursive = TRUE)
```

```
simulate_latinobarometro_wave
```

Generate a synthetic data.frame with the structure of a Latino-barometro wave

Description

No value comes from a real respondent – plausible distributions drawn at random, with deliberate test cases included (NA in car ownership, respondents from Venezuela, one country outside the 15 active ones).

Usage

```
simulate_latinobarometro_wave(year, n = 200, seed = 42)
```

Arguments

year	year of the synthetic wave (must be in <code>educ3_validated_years()</code> to be usable in the rest of the test pipeline).
n	number of synthetic respondents per country.
seed	random seed.

Value

data.frame with columns named IDENTICALLY to the raw Latinobarometro columns for the chosen year (found through the minimal crosswalk and the table of validated years), filled with synthetic values. Attribute `synthetic = TRUE` attached as a marker.

Examples

```
sim <- simulate_latinobarometro_wave(2018, n = 20)
dim(sim)
attr(sim, "synthetic")
```

Index

educ3_validated_years, [2](#)
educ3_validated_years(), [5](#), [8](#)

harmonize_core_predictors, [3](#)
harmonize_educ3_latinobarometro, [3](#)
harmonize_educ3_latinobarometro(), [2](#), [7](#)

income_lato, [4](#)

predict_latinobarometro_income, [6](#)
predict_latinobarometro_income(), [4](#)
prepare_latinobarometro_income_inputs,
 [7](#)
prepare_latinobarometro_income_inputs(),
 [4](#), [6](#)

simulate_latinobarometro_wave, [8](#)