

Package ‘DemogAnr’

July 26, 2026

Title Demographic Analysis

Version 0.2.0

Description Tool for demographic modeling and analysis, combining stochastic simulation-based projections with classic demographic methods. Provides utilities for disaggregating population data using Karup-King interpolation, fitting Brass relational logit models, calculating fertility and mortality metrics, building complete and abridged life tables with model-based separation factors, estimating child mortality indirectly from children ever born and surviving (Brass-Trussell method), decomposing and standardizing rates, and performing hierarchical subnational population projections. Relational life table estimation and interpolation methods are described in Brass (1975) ``Methods for Estimating Fertility and Mortality from Limited and Defective Data", Preston et al. (2001, ISBN:978-0631226161) ``Demography: Measuring and Modeling Population Processes", Siegel and Swanson (2004, ISBN:978-0126419559) ``The Methods and Materials of Demography", and Raftery et al. (2012) ``Bayesian probabilistic population projections for all countries" <[doi:10.1073/pnas.1211452109](https://doi.org/10.1073/pnas.1211452109)>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.3

Depends R (>= 4.1)

Imports dplyr, ggplot2, stats, tidyr

Suggests testthat (>= 3.0.0)

LazyData true

Config/testthat/edition 3

NeedsCompilation no

Author Edward Owusu Manu [aut, cre]

Maintainer Edward Owusu Manu <edward.manu@statsghana.gov.gh>

Repository CRAN

Date/Publication 2026-07-26 15:40:02 UTC

Contents

age_ratio	2
brass_logit	3
chm_brass	5
data	7
decompose_LE	8
decompose_rates	9
dem.cdr	10
dem.fert	12
dem.mmr	13
dm.chm	14
first_coef	15
ghmort2021	16
ghpop2021	17
gphc2010	17
interpolation	18
karup_king	19
last_coef	21
lifetable	21
lifetable_nqx	23
math_project	25
middle_coef	26
myers	27
project_population	28
pyramid	30
region2000	31
sex_ratio	32
standardize	33
standards	34
testdata	35
un_age_sex_accuracy	36
whipple	37
Index	39

age_ratio	<i>Age Ratios by Age Group</i>
-----------	--------------------------------

Description

Computes United Nations age ratios, $100 \times 2P_x / (P_{x-n} + P_{x+n})$, which measure how far each age group departs from the average of its neighbours. A ratio near 100 indicates a smooth age distribution.

Usage

```
age_ratio(data, age_col, pop_col, open_ended = TRUE)
```

Arguments

data	A data frame with one row per (equally spaced) age group.
age_col	Column name for the age group lower bound.
pop_col	Column name for the population count.
open_ended	Logical; if TRUE (default), the last age group is treated as open-ended, so no age ratio is computed for it or for the group below it (whose upper neighbour would be the open interval).

Value

A data frame with the age group, population, age ratio, and its deviation from 100 (NA where the ratio is not defined).

References

United Nations (1955). *Manual II: Methods of Appraisal of Quality of Basic Data for Population Estimates*. New York: United Nations.

Examples

```
d <- data.frame(age = seq(0, 20, 5), pop = c(1000, 900, 820, 760, 700))
age_ratio(d, "age", "pop", open_ended = FALSE)
```

brass_logit

Brass Relational Logit Model

Description

Fits the Brass relational logit model to an observed series of *cumulative* probabilities of dying $q(x) = 1 - l(x)/l(0)$ using a standard life table. The observed logits are regressed on the logits of the standard, and the fitted relationship (alpha, beta) is used to predict a complete, smoothed series of $q(x)$, including for ages where the observed value is missing (NA).

Usage

```
brass_logit(
  data,
  qx_col,
  age_col,
  standard,
  standards_data = NULL,
  verbose = FALSE
)
```

Arguments

data	A data frame containing the observed cumulative probabilities of dying $q(x)$ and ages.
qx_col	Column name for the cumulative probability of dying $q(x) = 1 - l(x)$ in the data frame. May contain NA for ages to be estimated.
age_col	Column name for age in the data frame.
standard	Standard to use ("African" or "GeneralUN").
standards_data	Standards dataset (default: included standards).
verbose	Logical. If TRUE, prints detailed status messages to the console.

Details

The Brass logit of a survivorship value is $Y(x) = 0.5 \log\{(1 - l(x))/l(x)\} = 0.5 \log\{q(x)/(1 - q(x))\}$, and predicted probabilities are recovered with the inverse transform $q(x) = \exp(2Y)/(1 + \exp(2Y))$.

Note that the input must be cumulative probabilities of dying by age x (i.e. $1 - l(x)$), which increase with age, and not the age-interval probabilities ${}_nq_x$. Values of exactly 0 or 1 (for example $q(0) = 0$) are excluded from the fit.

Value

A list containing the fitted model (model), the alpha/beta coefficients, and data: the input data with a predicted_qx column of smoothed/completed cumulative probabilities and a predicted_lx column of the implied survivorship $l(x) = 1 - q(x)$.

References

- Brass, W. (1971). On the scale of mortality. In W. Brass (Ed.), *Biological Aspects of Demography* (pp. 69-110). London: Taylor & Francis.
- Brass, W. (1975). *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill: Carolina Population Center, University of North Carolina.
- Moultrie, T., Dorrington, R., Hill, A., Hill, K., Timaeus, I., & Zaba, B. (Eds.). (2013). *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population (IUSSP). <https://demographicestimation.iussp.org/>
- Newell, C. (1988). *Methods and Models in Demography*. New York: Guilford Press. (Chapter on model life tables.)

Examples

```
# Observed cumulative probabilities of dying q(x) = 1 - l(x),
# with gaps (NA) to be filled by the model:
observed_data <- data.frame(
  age = c(1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50),
  qx = c(0.10, 0.15, NA, 0.19, 0.21, NA, 0.26, 0.30, NA, 0.38, 0.44)
)

model <- brass_logit(
```

```

    data = observed_data,
    qx_col = "qx",
    age_col = "age",
    standard = "African"
  )

  model$coefficients
  model$data

```

chm_brass

*Indirect Child Mortality Estimation (Brass-Trussell method)***Description**

Estimates child mortality indirectly from data on children ever born and children surviving (or dead), classified by five-year age group of mother, using the Trussell variant of the Brass method as set out in United Nations *Manual X* (1983). This is the standard technique for estimating child mortality from census or survey data in populations that lack complete death registration.

Usage

```

chm_brass(
  data,
  women_col,
  ceb_col,
  cd_col = NULL,
  cs_col = NULL,
  model = c("west", "north", "south", "east"),
  survey_year = NULL,
  verbose = FALSE
)

```

Arguments

data	A data frame with seven rows, one per five-year age group of mother from 15-19 to 45-49, in ascending order.
women_col	Column name for the number of women in each age group.
ceb_col	Column name for the number of children ever born.
cd_col	Column name for the number of children dead. Alternatively, supply cs_col (children surviving) instead.
cs_col	Optional column name for the number of children surviving; if given, children dead are computed as ceb - cs and cd_col is ignored.
model	Coale-Demeny model life table family: one of "west" (default), "north", "south", or "east".
survey_year	Optional numeric decimal year of the survey (e.g. 2010.4). If given, a reference date is computed for each estimate.
verbose	Logical; if TRUE, prints progress messages.

Details

For each five-year age group of mother i (15-19, ..., 45-49) the method computes the average parity $P(i) = CEB(i)/W(i)$ and the proportion of children dead $D(i) = CD(i)/CEB(i)$. Multipliers $k(i)$ that translate the reported proportions dead into life-table probabilities of dying $q(x)$ are obtained from

$$k(i) = a(i) + b(i) P(1)/P(2) + c(i) P(2)/P(3),$$

and $q(x) = k(i) D(i)$, where the age of the child x to which each estimate refers is 1, 2, 3, 5, 10, 15 and 20 for $i = 1, \dots, 7$. The coefficients $a(i), b(i), c(i)$ depend on the chosen Coale-Demeny model life table family (model).

Because a survey observes children born over a range of dates, each estimate refers to a point in the past. The number of years before the survey to which $q(x)$ applies, the reference period $t(x)$, is estimated with an analogous equation (Coale and Trussell 1977):

$$t(x) = a'(i) + b'(i) P(1)/P(2) + c'(i) P(2)/P(3).$$

If `survey_year` is supplied, the reference date of each estimate is returned as `survey_year - t(x)`, so that a series of estimates from one survey traces the trend in child mortality over the preceding two decades.

The estimate derived from the youngest mothers (15-19) is well known to be unreliable and is usually disregarded; see the references.

Value

A data frame (of class `dem_chm_indirect`) with one row per age group of mother, containing: `age_group`, index `i`, child age `x`, average parity `P`, proportion dead `D`, multiplier `k`, the probability of dying `qx = q(x)`, survivorship `lx = 1 - q(x)`, the reference period `t` (years before the survey) and, if `survey_year` was given, the `ref_date`.

References

- United Nations (1983). *Manual X: Indirect Techniques for Demographic Estimation*. Population Studies No. 81. New York: United Nations. (Chapter III; Tables 47 and 48.)
- Brass, W., & Coale, A. J. (1968). Methods of analysis and estimation. In W. Brass et al. (Eds.), *The Demography of Tropical Africa* (pp. 88-139). Princeton: Princeton University Press.
- Coale, A. J., & Trussell, T. J. (1977). Estimating the time to which Brass estimates apply. *Population Bulletin of the United Nations*, 10, 87-89.
- Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. (Chapter 11.)

Examples

```
# United Nations Manual X worked example: Panama, 1976 (both sexes)
panama <- data.frame(
  age_group = c("15-19", "20-24", "25-29", "30-34", "35-39", "40-44", "45-49"),
  women = c(2695, 2095, 1828, 1605, 1362, 1128, 930),
  ceb = c(557, 2633, 4757, 6085, 6722, 6367, 5276),
  cd = c(40, 130, 312, 435, 636, 686, 689)
```

```

)
est <- chm_brass(panama, women_col = "women", ceb_col = "ceb",
                 cd_col = "cd", model = "west", survey_year = 1976.7)
est

```

data

*Grouped population data by 5-year age groups***Description**

Example population counts in 5-year age groups (0-4 to 75-79) for the years 2021 to 2035, used to demonstrate [karup_king\(\)](#).

Usage

```
data
```

Format

A data frame with 16 rows (age groups) and 16 variables:

2021 Numeric. Population in the age group in 2021.

2022 Numeric. Population in the age group in 2022.

2023 Numeric. Population in the age group in 2023.

2024 Numeric. Population in the age group in 2024.

2025 Numeric. Population in the age group in 2025.

2026 Numeric. Population in the age group in 2026.

2027 Numeric. Population in the age group in 2027.

2028 Numeric. Population in the age group in 2028.

2029 Numeric. Population in the age group in 2029.

2030 Numeric. Population in the age group in 2030.

2031 Numeric. Population in the age group in 2031.

2032 Numeric. Population in the age group in 2032.

2033 Numeric. Population in the age group in 2033.

2034 Numeric. Population in the age group in 2034.

2035 Numeric. Population in the age group in 2035.

age_col Character. The 5-year age group, e.g. "0-4", "5-9", ..., "75-79".

Source

Simulated data based on population structures from the Ghana Statistical Service.

decompose_LE

*Arriaga Decomposition of a Difference in Life Expectancy***Description**

Decomposes the difference in life expectancy at birth between two life tables into the contributions of mortality differences in each age group, using Arriaga's (1984) discrete method (Preston et al. 2001, equations 3.11-3.12).

Usage

```
decompose_LE(
  lt1,
  lt2,
  age_col = "Age",
  lx_col = "lx",
  Lx_col = "Lx",
  Tx_col = "Tx",
  graph = TRUE
)
```

Arguments

lt1, lt2 Data frames for the two life tables, each with columns for age and the life-table functions lx, Lx and Tx (as returned in the `lifetable` component of `lifetable()`). Both must share the same age groups and radix.

age_col, lx_col, Lx_col, Tx_col Column names in lt1/lt2.

graph Logical. If TRUE (default), a **ggplot2** bar chart of the age-specific contributions is attached to the result as `$plot` and shown when the object is printed.

Details

For each closed age interval the contribution is

$${}_n\Delta_x = \frac{l_x^1}{l_0} \left(\frac{{}_nL_x^2}{l_x^2} - \frac{{}_nL_x^1}{l_x^1} \right) + \frac{T_{x+n}^2}{l_0} \left(\frac{l_x^1}{l_x^2} - \frac{l_{x+n}^1}{l_{x+n}^2} \right),$$

the first term being the direct effect and the second the indirect plus interaction effect. For the open-ended interval only the direct effect applies. The contributions sum to $e_0^2 - e_0^1$. Superscripts 1 and 2 refer to lt1 (baseline) and lt2 (comparison).

Value

An object of class `dem_le_decomp`: a list with the total difference in life expectancy at birth, its decomposition into direct and indirect effects, and a per-age table of contributions.

References

- Arriaga, E. E. (1984). Measuring and explaining the change in life expectancies. *Demography*, 21(1), 83-96. doi:10.2307/2061029
- Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography*. Oxford: Blackwell Publishers. (Equations 3.11-3.12.)

Examples

```
# Male-female gap in life expectancy, Ghana 2021
data(ghmort2021)
male <- subset(ghmort2021, Sex == "Male")
female <- subset(ghmort2021, Sex == "Female")
m <- lifetable(male, age = "Age", pop = "Population", Dx = "Deaths",
               sex = "male", nax_method = "keyfitz")
f <- lifetable(female, age = "Age", pop = "Population", Dx = "Deaths",
               sex = "female", nax_method = "keyfitz")
decompose_LE(m$lifetable, f$lifetable)
```

decompose_rates

Kitagawa Decomposition of a Difference Between Two Rates

Description

Decomposes the difference between two populations' crude rates into a component due to differences in age-specific rates and a component due to differences in age composition (Kitagawa 1955; Preston et al. 2001, Section 2.3).

Usage

```
decompose_rates(
  data,
  age_col,
  rate1_col,
  rate2_col,
  pop1_col,
  pop2_col,
  per = 1000
)
```

Arguments

data A data frame with one row per age group.

age_col Column name for age groups.

rate1_col, rate2_col Columns of age-specific rates for populations 1 and 2.

pop1_col, pop2_col Columns of population counts for populations 1 and 2.

per Scaling of the returned components (default 1000).

Details

With age composition $c_a = N_a / \sum N_a$ and age-specific rates M_a , the difference between the crude rates of populations 1 and 2 is split as

$$CDR_1 - CDR_2 = \sum_a (M_a^1 - M_a^2) \frac{c_a^1 + c_a^2}{2} + \sum_a (c_a^1 - c_a^2) \frac{M_a^1 + M_a^2}{2},$$

the first term being the rate (or effect) component and the second the age composition component.

Value

An object of class `dem_decomp`: a list with the total difference, the `rate_component`, the `composition_component`, and a per-age table.

References

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, 50(272), 1168-1194. doi:10.1080/01621459.1955.10501299

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography*. Oxford: Blackwell Publishers. (Section 2.3.)

Examples

```
d <- data.frame(
  age = c("0-14", "15-44", "45-64", "65+"),
  m1 = c(0.002, 0.001, 0.008, 0.060),
  m2 = c(0.003, 0.0015, 0.010, 0.065),
  n1 = c(300000, 400000, 200000, 100000),
  n2 = c(200000, 350000, 250000, 200000)
)
decompose_rates(d, "age", "m1", "m2", "n1", "n2")
```

dem.cdr

Calculate CDR and ASDR

Description

This function calculates the Crude Death Rate (CDR) and Age-Specific Death Rates (ASDR).

Usage

```
dem.cdr(
  data,
  type,
  age_col = NULL,
  population_col,
  deaths_col,
  graph = TRUE,
  verbose = FALSE
)
```

Arguments

<code>data</code>	A data frame containing demographic data.
<code>type</code>	A string or vector specifying the type(s) of death rate calculation. Options: "CDR", "ASDR". Use "all" to calculate both.
<code>age_col</code>	The column name for age groups (used for "ASDR").
<code>population_col</code>	The column name for total population.
<code>deaths_col</code>	The column name for total deaths.
<code>graph</code>	Logical. If TRUE (default), a ggplot2 plot of the age-specific death rates is attached to the result as <code>\$plot</code> and shown when the object is printed.
<code>verbose</code>	Logical. If TRUE, prints progress messages during execution.

Value

A list of named outputs for the requested death rate calculations.

References

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. ISBN 978-0631226161. (Chapter 2: Age-Specific Rates and Probabilities.)

Newell, C. (1988). *Methods and Models in Demography*. New York: Guilford Press.

Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The Methods and Materials of Demography* (2nd ed.). San Diego: Elsevier Academic Press. ISBN 978-0126419559. (Chapter on mortality measures.)

Examples

```
data <- data.frame(
  age = c("0-4", "5-9", "10-14"),
  population = c(10000, 12000, 11000),
  deaths = c(100, 50, 30)
)
dem.cdr(data, type = "all", age_col = "age",
        population_col = "population", deaths_col = "deaths")
```

dem.fert

*Calculate Fertility Metrics***Description**

This function calculates various fertility metrics (CBR, GFR, ASFR, TFR) based on user input and a dataframe with demographic data.

Usage

```
dem.fert(
  data,
  type,
  age_col = NULL,
  population_col,
  women_col = NULL,
  births_col,
  age_interval = 5,
  graph = TRUE,
  verbose = FALSE
)
```

Arguments

data	A dataframe containing demographic data.
type	A string or vector specifying the type(s) of fertility calculation. Options: "CBR", "GFR", "ASFR", "TFR". Use "all" to calculate all metrics.
age_col	The column name for age groups (used for "ASFR" and "TFR").
population_col	The column name for total population.
women_col	The column name for number of women (used for "GFR", "ASFR", and "TFR").
births_col	The column name for live births.
age_interval	Width (in years) of the age groups, used to compute TFR from the ASFRs (default 5, for standard 5-year age groups).
graph	Logical. If TRUE (default), a ggplot2 plot of the age-specific fertility rates is attached to the result as \$plot and shown when the object is printed.
verbose	Logical. If TRUE, prints progress messages during execution.

Value

A list of named outputs for the requested fertility calculations and the modified dataframe with ASFR.

References

- Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. ISBN 978-0631226161. (Chapter 5: Fertility and Reproduction.)
- Newell, C. (1988). *Methods and Models in Demography*. New York: Guilford Press.
- Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The Methods and Materials of Demography* (2nd ed.). San Diego: Elsevier Academic Press. ISBN 978-0126419559. (Chapter on fertility measures.)

Examples

```
demo_data <- data.frame(
  age = c("15-19", "20-24", "25-29"),
  population = c(10000, 12000, 11000),
  women = c(5000, 6000, 5500),
  live_births = c(200, 400, 300)
)
# Calculate ASFR
dem.fert(demo_data, type = "ASFR", age_col = "age",
         population_col = "population", women_col = "women", births_col = "live_births")
# Calculate all metrics
dem.fert(demo_data, type = "all", age_col = "age",
         population_col = "population", women_col = "women", births_col = "live_births")
```

dem.mmr

Calculate Maternal Mortality Metrics

Description

This function calculates the maternal mortality rate (maternal deaths per 100,000 women of reproductive age) and the maternal mortality ratio (maternal deaths per 100,000 live births).

Usage

```
dem.mmr(data, deaths_col, births_col, pop_women, verbose = FALSE)
```

Arguments

data	A dataframe containing demographic data.
deaths_col	The column name for maternal deaths.
births_col	The column name for live births.
pop_women	The column containing the number of women of reproductive age.
verbose	Logical. If TRUE, prints progress messages during execution.

Value

An object of class `dem_mmr`: a list with results (the maternal mortality rate MMR per 100,000 women, and the maternal mortality ratio MMR_Ratio per 100,000 live births) and `modified_data` (the input data with per-row `mm_rate` and `mm_ratio` columns).

References

World Health Organization (1992). *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)*. Geneva: World Health Organization. (Definitions of maternal death and maternal mortality measures.)

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. ISBN 978-0631226161.

Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The Methods and Materials of Demography* (2nd ed.). San Diego: Elsevier Academic Press. ISBN 978-0126419559. (Chapter on health demography and maternal mortality.)

Examples

```
demo_data <- data.frame(
  age=c("15-24", "25-34", "35-44"),
  maternal_deaths = c(5, 10, 15),
  live_births = c(5000, 7000, 6000),
  women =c(20000, 22000, 32000)
)
dem.mmr(demo_data, deaths_col = "maternal_deaths", births_col = "live_births", pop_women = "women")
```

dm.chm

Calculate Child Mortality Metrics by Age

Description

This function calculates child mortality metrics (neonatal, infant, child, under-5) based on single-age data. It requires columns for age, denominator (live births or population), and deaths.

Usage

```
dm.chm(
  data,
  age_col,
  live_births,
  deaths_col,
  type = c("neonatal", "infant", "child", "under5"),
  age_in_months = NULL,
  verbose = FALSE
)
```

Arguments

data	A dataframe containing demographic data.
age_col	The column name representing the age (in years or months for neonates).
live_births	The column name for the denominator at each age (live births, or mid-year population if births are unavailable).

deaths_col	The column name representing deaths at each age.
type	The type of mortality calculation to perform: "neonatal", "infant", "child", or "under5".
age_in_months	(Optional) The column name for age in months for neonatal mortality.
verbose	Logical. If TRUE, prints progress messages during execution.

Details

The result is expressed per 1,000 of the supplied denominator. If the denominator is live births (as is conventional for neonatal and infant mortality), the result is a rate per 1,000 live births. If a mid-year population is supplied instead (often the only option for the child and under-5 groups from census data), the result is a death *rate* per 1,000 population, which is not identical to the cohort probability of dying (e.g. 5q0) reported by survey programmes such as the DHS.

Value

A numeric value representing the mortality rate per 1,000 live births.

References

Pollard, A. H., Yusuf, F., & Pollard, G. N. (1990). *Demographic Techniques* (3rd ed.). Sydney: Pergamon Press.

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. ISBN 978-0631226161. (Chapters 1-2: rates versus probabilities of dying.)

Examples

```
demo_data <- data.frame(
  age = 0:10,
  population = c(1000, 950, 900, 850, 800, 750, 700, 650, 600, 550, 500),
  deaths = c(30, 20, 15, 10, 5, 2, 1, 1, 0, 0, 0)
)
dm.chm(demo_data, age_col = "age",
        live_births = "population",
        deaths_col = "deaths",
        type = "under5")
```

first_coef

Karup-King coefficients for the first age group

Description

The standard Karup-King third-difference osculatory interpolation multipliers used to split the *first* 5-year age group into single years. Rows are the five single-year ages within the group; columns are the weights applied to the first, second, and third 5-year groups.

Usage

```
first_coef
```

Format

A 5 x 3 data frame of numeric coefficients.

Source

Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The Methods and Materials of Demography* (2nd ed.), Appendix C.

ghmort2021

Ghana age-specific mortality by sex, 2021

Description

Age-specific population and deaths for Ghana in 2021, separately for males and females, taken as the base year of a national age-specific mortality projection. Because it is split by sex, this data set is suitable for demonstrating sex-specific life tables and the decomposition of the male-female gap in life expectancy with [decompose_LE\(\)](#).

Usage

```
ghmort2021
```

Format

A data frame with 36 rows (18 age groups x 2 sexes) and 5 variables:

Sex Character. "Male" or "Female".

Age Integer. Lower bound of the age group (0, 1, 5, 10, ..., 80).

AgeGroup Character. Age-group label ("<1", "1-4", ..., "80+").

Population Integer. Mid-year population (exposure) in the age group.

Deaths Integer. Deaths in the age group.

Source

Ghana Statistical Service, age-specific mortality projection (2021 base year).

ghpop2021

*Ghana single-year age distribution by sex, 2021***Description**

Population by single year of age (0 to 79, with 80 representing the open age group 80+), separately for males and females, Ghana 2021. Single-year age counts are the input required by the digit-preference indices [whipple\(\)](#) and [myers\(\)](#), and can be grouped into five-year age groups for [un_age_sex_accuracy\(\)](#).

Usage

ghpop2021

Format

A data frame with 162 rows (81 single ages x 2 sexes) and 3 variables:

Sex Character. "Male" or "Female".

Age Integer. Single year of age; 0 is age <1 and 80 is the open interval 80+.

Population Integer. Population count.

Source

Ghana Statistical Service, 2021 Population and Housing Census (national single-year age distribution).

gphc2010

*Ghana 2010 Population and Housing Census mortality data***Description**

Population and deaths by age group from the Ghana 2010 Population and Housing Census, used to demonstrate [lifetable\(\)](#).

Usage

gphc2010

Format

A data frame with 21 rows and 3 variables:

Age Numeric. Lower bound of the age group (0, 1, 5, 10, ...).

Pop Numeric. Population in the age group.

Deaths Numeric. Deaths in the age group.

Source

Ghana Statistical Service, 2010 Population and Housing Census.

interpolation

Smart Interpolation Function

Description

Performs interpolation on numeric data, supporting linear, quadratic, Lagrange, and cubic spline methods. Can handle both vector and dataframe inputs.

Usage

```
interpolation(
  x,
  y,
  x_new,
  data = NULL,
  y_cols = NULL,
  method = c("auto", "linear", "quadratic", "lagrange", "spline"),
  verbose = FALSE
)
```

Arguments

x	Numeric vector of x-values.
y	Numeric vector of y-values (ignored if using a dataframe).
x_new	Numeric vector of new x-values to interpolate.
data	Optional dataframe containing y-values to interpolate.
y_cols	Character vector of column names in data containing y-values.
method	Character. Interpolation method: "auto", "linear", "quadratic", "lagrange", or "spline". Default is "auto": linear for 2 points, quadratic for 3, and a natural cubic spline for more than 3 points. High-degree Lagrange polynomials are numerically unstable for many points (Runge phenomenon), so "lagrange" is only used when requested explicitly.
verbose	Logical. If TRUE, prints detailed status messages to the console during interpolation.

Value

If vectors x and y are provided, returns a named numeric vector of interpolated values. If a dataframe is provided, returns the original dataframe with additional interpolated columns.

References

- Shryock, H. S., Siegel, J. S., & Larmon, E. A. (1973). *The Methods and Materials of Demography*. Washington, DC: US Bureau of the Census. (Appendix C: Interpolation and graduation.)
- Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The Methods and Materials of Demography* (2nd ed.). San Diego: Elsevier Academic Press. ISBN 978-0126419559. (Appendix C: Selected General Methods.)
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing* (3rd ed.). Cambridge: Cambridge University Press. ISBN 978-0521880688. (Chapter 3: Interpolation and Extrapolation.)

Examples

```
# Example 1: Interpolating between points (vector input)
x <- c(1, 2, 3, 4, 5)
y <- c(2, 4, 6, 8, 10)
x_new <- c(2.5, 3.5, 4.5)
interpolation(x, y, x_new, method = "linear")

# Example 2: Interpolating for a dataframe
data <- data.frame(
  category = c("A", "B", "C"),
  y1 = c(2, 5, 9),
  y2 = c(3, 6, 10)
)
x_new <- c(2.5, 3.5)
interpolation(x = c(1, 2), data = data, y_cols = c("y1", "y2"), x_new = x_new, method = "spline")
```

karup_king

Split Age Groups into Single-Year Ages

Description

This function takes grouped population data (e.g., 5-year age groups) and distributes the population into single-year age groups using the Karup-King third-difference osculatory interpolation coefficients. The user can specify the column containing age groups, one or more population columns, and (optionally) custom coefficient matrices for the first, middle, and last groups.

Usage

```
karup_king(
  df,
  age_col = "age_group",
  pops = "population",
  first_coef = NULL,
  middle_coef = NULL,
  last_coef = NULL,
  verbose = FALSE
)
```

Arguments

<code>df</code>	A data frame containing grouped population data.
<code>age_col</code>	A string specifying the column in <code>df</code> that contains age group labels in the form "start-end" (e.g., "15-19"). Open-ended groups (e.g. "85+") are not supported and should be excluded or closed before calling.
<code>pops</code>	A string or vector of strings specifying one or more columns in <code>df</code> that contain population data.
<code>first_coef</code>	A 5x3 matrix of coefficients for splitting the first age group. Defaults to the packaged Karup-King multipliers (first_coef).
<code>middle_coef</code>	A 5x3 matrix of coefficients for splitting middle age groups. Defaults to the packaged Karup-King multipliers (middle_coef).
<code>last_coef</code>	A 5x3 matrix of coefficients for splitting the last age group. Defaults to the packaged Karup-King multipliers (last_coef).
<code>verbose</code>	Logical. If TRUE, prints detailed status messages to the console during interpolation.

Details

Each coefficient matrix has 5 rows (one per single year within the 5-year group) and 3 columns (weights applied to the neighbouring 5-year groups). The defaults are the standard Karup-King multipliers (Siegel & Swanson 2004, Appendix C): [first_coef](#) for the first group (columns: current, next, next+1), [middle_coef](#) for interior groups (columns: previous, current, next), and [last_coef](#) for the last group (columns: previous-1, previous, current).

Value

A data frame with single-year ages and corresponding population estimates for each specified population column. Ages beyond the range of the input age groups (overflow ages) are removed.

References

- Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The Methods and Materials of Demography* (2nd ed.). San Diego: Elsevier Academic Press. ISBN 978-0126419559. (Appendix C: Karup-King oscillatory interpolation multipliers.)
- Shryock, H. S., Siegel, J. S., & Larmon, E. A. (1973). *The Methods and Materials of Demography*. Washington, DC: US Bureau of the Census. (Appendix C: Interpolation and graduation.)

Examples

```
data(data)
head(data)

# Split the 5-year age groups into single years for the 2021 population,
# using the packaged Karup-King coefficients:
single <- karup_king(df = data, age_col = "age_col", pops = "2021")
head(single)
```

```
# Population totals are (approximately) preserved:
sum(data[["2021"]])
sum(single[["2021"]])
```

last_coef	<i>Karup-King coefficients for the last age group</i>
-----------	---

Description

The standard Karup-King third-difference osculatory interpolation multipliers used to split the *last* 5-year age group into single years. Rows are the five single-year ages within the group; columns are the weights applied to the antepenultimate, penultimate, and last 5-year groups.

Usage

```
last_coef
```

Format

A 5 x 3 data frame of numeric coefficients.

Source

Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The Methods and Materials of Demography* (2nd ed.), Appendix C.

lifetable	<i>lifetable</i>
-----------	------------------

Description

This function computes a complete (or abridged) life table from given mortality rates or from raw deaths and population counts. It returns the standard life-table columns using textbook notation.

Usage

```
lifetable(
  data,
  age = "Age",
  nMx = NULL,
  pop = NULL,
  Dx = NULL,
  sex = c("male", "female"),
  nax_method = c("cd", "keyfitz"),
  nax = NULL,
```

```

    radix = 1e+05,
    graph = TRUE,
    verbose = FALSE
  )

```

Arguments

<code>data</code>	A data frame containing age-specific mortality data.
<code>age</code>	String: name of the column with the lower bound of each age group, e.g., 0, 1, 5, 10, ...
<code>nMx</code>	String (optional): name of the column with observed mortality rates. If not provided, <code>pop</code> and <code>Dx</code> must be given to compute it.
<code>pop</code>	String (optional): name of the column with total population (exposure) in each age group. Used to compute <code>nMx</code> if not given.
<code>Dx</code>	String (optional): name of the column with deaths in each age group. Used to compute <code>nMx</code> if not given.
<code>sex</code>	Character: "male" (default) or "female". Determines the Coale-Demeny coefficients used for a_0 and ${}_4a_1$.
<code>nax_method</code>	Character: how to set ${}_na_x$ for interior age groups. Either "cd" (Coale-Demeny young ages, $n/2$ elsewhere; default) or "keyfitz" (iterative Keyfitz 1966 refinement for interior five-year intervals). Ignored if <code>nax</code> is supplied.
<code>nax</code>	Optional numeric vector of ${}_na_x$ values (one per age group) borrowed from an external model life table. Overrides <code>nax_method</code> .
<code>radix</code>	Numeric: the life table radix, i.e. survivors at exact age 0 (default 100,000).
<code>graph</code>	Logical. If TRUE (default), a ggplot2 survival curve (l_x against age) is attached to the result as <code>\$plot</code> and shown when the object is printed.
<code>verbose</code>	Logical. If TRUE, prints detailed status messages to the console during computation.

Details

Interval widths n are derived from the age column (e.g. ages 0, 1, 5, 10, ... give widths 1, 4, 5, ...). Probabilities of dying are obtained from the death rates with the standard conversion ${}_nq_x = n \cdot {}_nM_x / (1 + (n - {}_na_x) \cdot {}_nM_x)$, and person-years lived are ${}_nL_x = n \cdot l_{x+n} + {}_na_x \cdot {}_nd_x$ for closed intervals and l_x / M_x for the open-ended interval.

The average person-years lived by those dying in an interval, ${}_na_x$, controls the accuracy of the ${}_nM_x \rightarrow {}_nq_x$ conversion. `lifetable()` offers three ways to set it, chosen with `nax_method`:

"cd" (**default**) Coale-Demeny model values for ages 0 and 1-4 (sex-specific functions of ${}_1M_0$; Preston et al. 2001, Table 3.3), $n/2$ for other closed intervals, and $1/{}_nM_x$ for the open interval. This reproduces the behaviour of earlier versions.

"keyfitz" As "cd" for ages 0 and 1-4 and for the open interval, but for the interior five-year intervals ${}_na_x$ is refined with the iterative Keyfitz (1966) formula ${}_na_x = n/2 + (n/24)({}_nd_{x+n} - {}_nd_{x-n}) / {}_nd_x$, which borrows information on the slope of the death distribution from the neighbouring age groups. The table is solved iteratively (a few iterations) because ${}_nd_x$ itself depends on ${}_na_x$.

user-supplied Pass a numeric vector to `nax` (one value per age group) to borrow ${}_na_x$ values from any external model life table system, e.g. Coale-Demeny or the United Nations model life tables. This overrides `nax_method`.

Value

A list containing:

`metrics` A list with the life expectancy at birth and the total life table deaths.

`lifetable` A dataframe of the complete life table.

References

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. ISBN 978-0631226161. (Chapter 3: The Life Table and Single Decrement Processes; Table 3.3 for the Coale-Demeny separation factors.)

Keyfitz, N. (1966). A life table that agrees with the data. *Journal of the American Statistical Association*, 61(314), 305-312. doi:10.1080/01621459.1966.10480871

Chiang, C. L. (1984). *The Life Table and Its Applications*. Malabar, FL: Robert E. Krieger Publishing.

Coale, A. J., & Demeny, P. (1983). *Regional Model Life Tables and Stable Populations* (2nd ed.). New York: Academic Press. (Source of the `a0` and `4a1` separation factors.)

Examples

```
# Example computing rates from deaths and population
data(gphc2010)
res <- lifetable(data = gphc2010, age = "Age", pop = "Pop", Dx = "Deaths")
res$metrics
head(res$lifetable)

# Refine nax for adult ages with the iterative Keyfitz method
res2 <- lifetable(gphc2010, age = "Age", pop = "Pop", Dx = "Deaths",
                  nax_method = "keyfitz")
head(res2$lifetable[, c("Age", "nax", "nqx", "ex")])
```

lifetable_nqx

Compute Life Table Using Age and nqx

Description

This function computes a life table from a data frame containing age groups and age-interval probabilities of dying (`nqx`). It calculates key life table components such as `lx`, `dx`, `Lx`, `Tx`, and `ex`.

Usage

```
lifetable_nqx(data, age, nqx, l0 = 1e+05, verbose = FALSE)
```

Arguments

data	A data frame containing age and nqx columns.
age	A string specifying the column name for age groups (lower bounds, e.g. 0, 1, 5, 10, ...). Ages must be strictly increasing.
nqx	A string specifying the column name for the probabilities of dying between age x and $x+n$ (nqx). The last value should be 1 (open-ended group).
l0	Radix of the life table; survivors at age 0 (default = 100,000).
verbose	Logical. If TRUE, prints detailed status messages to the console during computation.

Details

Interval widths n are derived from the age column (e.g. ages 0, 1, 5, 10, ... give widths 1, 4, 5, ...). Person-years lived in each closed interval are approximated as ${}_nL_x = n(l_x - d_x/2)$, i.e. deaths are assumed to occur at the midpoint of the interval. For the open-ended last age group, person-years are approximated with the classical rule of thumb $L_{x+} = l_x \log_{10}(l_x)$ (equivalent to $e_x \approx \log_{10} l_x$), which is used when no death rate for the open interval is available. If death rates are available, [lifetable\(\)](#) provides a more precise treatment of the open interval.

Value

A data frame with the computed life table, including:

Age Age groups.
 n Width of each age interval (Inf for the last, open-ended group).
 nqx Probability of dying between age x and $x+n$.
 lx Number of survivors at the beginning of each age group.
 dx Number of deaths within each age group.
 Lx Person-years lived within each age group.
 Tx Cumulative person-years lived above a specific age group.
 ex Life expectancy at the beginning of each age group.

References

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. ISBN 978-0631226161. (Chapter 3: The Life Table and Single Decrement Processes.)
 Chiang, C. L. (1984). *The Life Table and Its Applications*. Malabar, FL: Robert E. Krieger Publishing.

Examples

```
# Example data
data <- data.frame(
  age = c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85),
  nqx = c(0.05, 0.01, 0.005, 0.002, 0.003, 0.004, 0.005, 0.007, 0.010, 0.015,
```



```

        0.020, 0.030, 0.050, 0.070, 0.100, 0.150, 0.200, 0.300, 1.000)
    )
    life_table <- lifetable_nqx(data, age = "age", nqx = "nqx")
    print(life_table)

```

math_project

*Hierarchical Subnational Population Projection***Description**

Projects lower-level geographies (e.g., Districts) while constraining them to higher-level control totals (e.g., Regions or National) provided within the dataset.

Usage

```

math_project(
  data,
  sub_group_col,
  parent_group_col = NULL,
  pop_cols,
  time_vals,
  target_times,
  method = c("linear", "exponential", "geometric", "constant_share", "shift_share",
    "share_of_growth"),
  parent_totals_cols = NULL,
  verbose = FALSE
)

```

Arguments

data	Dataframe containing the subnational population data.
sub_group_col	Character. Column name for the lower-level geography (e.g., "District").
parent_group_col	Character. Optional. Column name for the higher-level geography (e.g., "Region"). If NULL, assumes a single national parent group.
pop_cols	Character vector of length 2. Column names for Base Year and Launch Year populations.
time_vals	Numeric vector of length 2. The years corresponding to pop_cols.
target_times	Numeric vector. The future years to project.
method	Character. The projection method.
parent_totals_cols	Character vector. Optional. Column names in data containing the pre-calculated projected totals for the parent group at target_times. Must match the length of target_times.
verbose	Logical. If TRUE, prints detailed status messages to the console during projection.

Value

A dataframe appending the projected columns to the original data.

References

Smith, S. K., Tayman, J., & Swanson, D. A. (2013). *A Practitioner’s Guide to State and Local Population Projections*. Dordrecht: Springer. doi:10.1007/9789400775510 (Trend extrapolation, constant-share, shift-share, and share-of-growth methods.)

Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The Methods and Materials of Demography* (2nd ed.). San Diego: Elsevier Academic Press. ISBN 978-0126419559. (Chapter on population projections.)

Rowland, D. T. (2003). *Demographic Methods and Concepts*. Oxford: Oxford University Press. ISBN 978-0198752639. (Chapter on population projections.)

Examples

```
districts <- data.frame(
  Region = c("A", "A", "B", "B"),
  District = c("A1", "A2", "B1", "B2"),
  pop2010 = c(10000, 20000, 15000, 5000),
  pop2020 = c(12000, 26000, 16500, 5500)
)

# Exponential extrapolation of each district
math_project(districts,
  sub_group_col = "District", parent_group_col = "Region",
  pop_cols = c("pop2010", "pop2020"), time_vals = c(2010, 2020),
  target_times = c(2025, 2030), method = "exponential"
)
```

middle_coef	<i>Karup-King coefficients for middle age groups</i>
-------------	--

Description

The standard Karup-King third-difference osculatory interpolation multipliers used to split *interior* 5-year age groups into single years. Rows are the five single-year ages within the group; columns are the weights applied to the previous, current, and next 5-year groups.

Usage

```
middle_coef
```

Format

A 5 x 3 data frame of numeric coefficients.

Source

Siegel, J. S., & Swanson, D. A. (Eds.). (2004). *The Methods and Materials of Demography* (2nd ed.), Appendix C.

myers

Myers' Blended Index of Age Heaping

Description

Computes Myers' blended index, which measures preference for or avoidance of every terminal digit 0-9, from single-year-of-age population counts.

Usage

```
myers(data, age_col, pop_col, lower = 10, upper = 89, graph = TRUE)
```

Arguments

data	A data frame with single years of age.
age_col	Column name for single year of age.
pop_col	Column name for the population count.
lower, upper	The initial tabulation range (defaults 10 and 89). The blend extends nine years beyond upper.
graph	Logical. If TRUE (default), a ggplot2 bar chart of the deviation from 10% by terminal digit is attached to the result and shown on print.

Details

In the absence of digit preference, the population whose age ends in each digit should be one tenth of the total. Because a population normally declines with age, Myers "blends" the tabulation to remove that trend: each age a in the range $[lower, upper]$ receives weight 10, ages just below the range ($lower$ to $lower + 8$) receive weights 1 to 9, and ages just above ($upper + 1$ to $upper + 9$) receive weights 9 to 1 (Rodriguez 2015; equivalent to Myers 1940). The blended counts by terminal digit are expressed as percentages p_d , and the index is

$$M = \frac{1}{2} \sum_{d=0}^9 |p_d - 10|,$$

ranging from 0 (no heaping) to 90 (all ages at one digit). The data should extend up to age $upper + 9$ for the blending weights to apply.

Value

An object of class `dem_myers`: a list with the index value and a table giving, for each terminal digit, the reported and blended counts, the blended percentage, and its deviation from 10.

References

Myers, R. J. (1940). Errors and bias in the reporting of ages in census data. *Transactions of the Actuarial Society of America*, 41(2), 395-415.

Rodriguez, G. (2015). *Demographic Methods* (course notes). Princeton University. <https://grodrri.github.io/demography/>

Examples

```
# Ghana 2021 single-year ages, males (blend over 10-69)
data(ghpop2021)
male <- subset(ghpop2021, Sex == "Male")
myers(male, age_col = "Age", pop_col = "Population", lower = 10, upper = 69)
```

project_population	<i>Basic Probabilistic Population Projection</i>
--------------------	--

Description

This function projects population trajectories for multiple administrative regions using a simple stochastic (simulation-based) exponential growth model, inspired by the probabilistic projection literature (e.g. Raftery et al. 2012), but without a full Bayesian model.

For each region, `num_samples` growth-rate trajectories are simulated. The annual growth rate of each trajectory is the sum of a random baseline, a fertility effect (increasing in TFR), a mortality effect (decreasing in the death rate), a migration effect (with the sign of net migration), and random noise. Each trajectory is propagated independently through all years, so that uncertainty accumulates over the projection horizon, and the results are summarised per region and year.

Usage

```
project_population(  
  data,  
  future_year,  
  base_year,  
  region_var = "region",  
  subregion_var = "subregion",  
  base_pop_var = "base_pop",  
  TFR_var = "TFR",  
  death_rate_var = "death_rate",  
  net_migration_var = "net_migration",  
  num_samples = 2000,  
  random_seed = 42,  
  graph = TRUE,  
  verbose = FALSE  
)
```

Arguments

data	A data frame containing projection parameters.
future_year	Numeric. The target year for the projection.
base_year	Numeric. The base year for the projection.
region_var	Character. Column name for the region.
subregion_var	Character. Column name for the subregion.
base_pop_var	Character. Column name for the base population.
TFR_var	Character. Column name for the Total Fertility Rate (TFR).
death_rate_var	Character. Column name for the death rate.
net_migration_var	Character. Column name for net migration.
num_samples	Integer. Number of simulation samples (default: 2000).
random_seed	Integer. Random seed for reproducibility (default: 42).
graph	Logical. If TRUE (default), a ggplot2 plot of the projected trajectories (median and interquartile band, faceted by subregion) is attached to the result and shown when it is printed.
verbose	Logical. If TRUE, prints detailed status messages to the console during projection.

Value

A data frame with one row per region, subregion, and year, containing the summary statistics of the simulated population trajectories: lower (25th percentile), median, mean, and upper (75th percentile).

References

- Raftery, A. E., Li, N., Sevcikova, H., Gerland, P., & Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109(35), 13915-13921. doi:10.1073/pnas.1211452109 (Inspiration for the stochastic/probabilistic projection logic.)
- Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. ISBN 978-0631226161. (Chapter 6: Population Projections.)
- Rowland, D. T. (2003). *Demographic Methods and Concepts*. Oxford: Oxford University Press. ISBN 978-0198752639. (Chapter on population projections.)

Examples

```
data(region2000)
proj <- project_population(region2000,
  base_year = 2000, future_year = 2009,
  TFR_var = "TFR", base_pop_var = "base_pop",
  region_var = "Country", death_rate_var = "death_rate",
  net_migration_var = "net_migration",
```

```

    subregion_var = "Region",
    num_samples = 1000
)
head(proj)

```

pyramid

Population Pyramid

Description

Draws a population pyramid (back-to-back horizontal bar chart of population by age group and sex) using **ggplot2**.

Usage

```

pyramid(
  data,
  age_col,
  sex_col,
  count_col,
  male_label = "Male",
  female_label = "Female",
  male_color = .dem_male,
  female_color = .dem_female,
  title = NULL,
  base_family = ""
)

```

Arguments

<code>data</code>	A data frame with one row per age group and sex.
<code>age_col</code>	Column name for the age group.
<code>sex_col</code>	Column name for sex.
<code>count_col</code>	Column name for the population count.
<code>male_label, female_label</code>	The values in <code>sex_col</code> denoting males and females (defaults "Male" and "Female"). Males are drawn to the left.
<code>male_color, female_color</code>	Bar fill colours.
<code>title</code>	Optional plot title.
<code>base_family</code>	Font family passed to the theme; leave as "" to use the graphics device default, or set it to an installed font (e.g. "Century Gothic").

Value

A **ggplot2** object, which can be printed, saved with `ggplot2::ggsave()`, or further customised by adding layers.

References

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers.

Examples

```
# Ghana 2021: single-year ages grouped into five-year age groups, by sex
data(ghpop2021)
g <- within(ghpop2021, grp <- ifelse(Age >= 80, "80+",
  paste0((Age %/% 5) * 5, "-", (Age %/% 5) * 5 + 4)))
pyr <- aggregate(Population ~ grp + Sex, data = g, FUN = sum)
pyramid(pyr, age_col = "grp", sex_col = "Sex", count_col = "Population",
  title = "Ghana, 2021")
```

region2000

Regional projection parameters, Ghana 2000

Description

Base-year demographic parameters for the regions of Ghana in 2000, used to demonstrate `project_population()`.

Usage

```
region2000
```

Format

A data frame with 16 rows and 6 variables:

Country Character. Country name.

Region Character. Region name.

base_pop Numeric. Base-year population.

TFR Numeric. Total fertility rate.

death_rate Numeric. Crude death rate.

net_migration Numeric. Net migration (persons).

Source

Derived from Ghana Statistical Service census publications.

sex_ratio	<i>Sex Ratios by Age Group</i>
-----------	--------------------------------

Description

Computes the sex ratio (males per 100 females) for each age group.

Usage

```
sex_ratio(data, age_col, male_col, female_col)
```

Arguments

data	A data frame with one row per age group.
age_col	Column name for the age group.
male_col, female_col	Column names for the male and female counts.

Value

A data frame with the age group, male and female counts, and the sex ratio.

References

Shryock, H. S., & Siegel, J. S. (1976). *The Methods and Materials of Demography*. New York: Academic Press.

Examples

```
data(ghpop2021)
# collapse single years into five-year groups, by sex
g <- within(ghpop2021, grp <- ifelse(Age >= 80, 80, (Age %/% 5) * 5))
wide <- tapply(g$Population, list(g$grp, g$Sex), sum)
sr_data <- data.frame(age = rownames(wide),
                     m = wide[, "Male"], f = wide[, "Female"])
head(sex_ratio(sr_data, "age", "m", "f"))
```


Description

Removes the effect of differing age structures when comparing the crude rates of two populations, following Preston, Heuveline and Guillot (2001, Chapter 2). Each population's age-specific rates are re-weighted by a common standard age distribution – by default the average of the two populations' age compositions, their recommended choice for a two-population comparison.

Usage

```
standardize(
  data,
  age_col,
  rate1_col,
  rate2_col,
  pop1_col,
  pop2_col,
  standard = c("average", "first", "second"),
  std_pop_col = NULL,
  labels = c("1", "2"),
  per = 1000
)
```

Arguments

<code>data</code>	A data frame with one row per age group.
<code>age_col</code>	Column name for age groups.
<code>rate1_col, rate2_col</code>	Columns of age-specific rates for populations 1 and 2.
<code>pop1_col, pop2_col</code>	Columns of population counts (or proportions) for populations 1 and 2.
<code>standard</code>	How to choose the standard age distribution: "average" (default), "first", or "second". Ignored if <code>std_pop_col</code> is supplied.
<code>std_pop_col</code>	Optional column of an external standard population by age; overrides <code>standard</code> .
<code>labels</code>	Length-2 character vector naming the two populations.
<code>per</code>	Scaling of the returned rates (default 1000).

Details

For populations 1 and 2 with age-specific rates M_i^1, M_i^2 and age compositions $C_i^1 = N_i^1 / \sum N_i^1$ and C_i^2 , the crude rate of each is $CDR = \sum_i C_i M_i$. The age-standardized crude rate applies a common standard composition C_i^s :

$$ASCDR = \sum_i M_i C_i^s.$$

By default $C_i^s = (C_i^1 + C_i^2)/2$; set standard to "first" or "second" to use one population's composition, or supply an external standard with `std_pop_col`. The function also reports the comparative mortality ratio $CMR = \sum_i N_i^1 M_i^1 / \sum_i N_i^1 M_i^2$ (population 1's deaths relative to those expected under population 2's rates).

With the average standard, the difference between the two age-standardized rates equals the rate component of the Kitagawa decomposition (`decompose_rates()`) of the same difference.

Value

An object of class `dem_std`: a list with the crude rates (`crude`), the age-standardized rates (`standardized`), their ratio, the comparative mortality ratio (`CMR`), and a per-age table.

References

Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers. (Chapter 2; Box 2.1.)

Examples

```
# Preston et al. (2001) Box 2.1: Sweden vs Kazakhstan, females, 1992
box21 <- data.frame(
  age = c("0", "1-4", "5-9", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39",
    "40-44", "45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79",
    "80-84", "85+"),
  pop_Sw = c(0.0136, 0.0524, 0.0559, 0.0548, 0.0604, 0.0655, 0.0709, 0.0641, 0.0654,
    0.0703, 0.0730, 0.0552, 0.0481, 0.0493, 0.0512, 0.0508, 0.0420, 0.0321,
    0.0251),
  pop_K = c(0.0200, 0.0868, 0.1011, 0.0929, 0.0828, 0.0716, 0.0843, 0.0842, 0.0704,
    0.0561, 0.0327, 0.0579, 0.0347, 0.0430, 0.0295, 0.0178, 0.0172, 0.0102,
    0.0068),
  m_Sw = c(0.00467, 0.00008, 0.00013, 0.00014, 0.00023, 0.00030, 0.00032, 0.00050,
    0.00069, 0.00117, 0.00201, 0.00305, 0.00461, 0.00759, 0.01226, 0.02026,
    0.03664, 0.06815, 0.15729),
  m_K = c(0.02137, 0.00162, 0.00045, 0.00037, 0.00078, 0.00108, 0.00103, 0.00132,
    0.00182, 0.00288, 0.00430, 0.00571, 0.01082, 0.01392, 0.02679, 0.03998,
    0.05469, 0.10159, 0.18030)
)
standardize(box21, age_col = "age",
  rate1_col = "m_Sw", rate2_col = "m_K",
  pop1_col = "pop_Sw", pop2_col = "pop_K",
  labels = c("Sweden", "Kazakhstan"))
```

standards

Brass standard life tables

Description

Survivorship values $l(x)$ (radix 1) and their Brass logits for the African and General United Nations standards, used by `brass_logit()`.

Usage

```
standards
```

Format

A data frame with 24 rows and 5 variables:

Age Numeric. Exact age x .

GeneralUN Numeric. $l(x)$ of the General UN standard.

GeneralUN_logit Numeric. Brass logit of the General UN standard.

African Numeric. $l(x)$ of the African standard.

African_logit Numeric. Brass logit of the African standard.

Source

Brass, W. (1971). On the scale of mortality. In *Biological Aspects of Demography*. London: Taylor & Francis. United Nations model life table publications.

testdata

District-level demographic test data

Description

Simulated district-level demographic panel data (population, births, deaths, migration) by age group and sex, for testing subnational projection workflows such as [math_project\(\)](#).

Usage

```
testdata
```

Format

A data frame with 1224 rows and 10 variables:

region Character. Region name.

district Character. District name.

year Numeric. Calendar year.

age_group Character. Age group label.

sex Character. "Male" or "Female".

population Numeric. Population count.

births Numeric. Live births.

deaths Numeric. Deaths.

in_migration Numeric. In-migrants.

out_migration Numeric. Out-migrants.

Source

Simulated data.

un_age_sex_accuracy	<i>United Nations Age-Sex Accuracy Index</i>
---------------------	--

Description

Computes the United Nations joint age-sex accuracy index, a summary measure of the internal consistency of an age-sex distribution built from the sex ratio score and the male and female age ratio scores.

Usage

```
un_age_sex_accuracy(
  data,
  age_col,
  male_col,
  female_col,
  open_ended = TRUE,
  graph = TRUE
)
```

Arguments

data	A data frame with one row per five-year age group.
age_col	Column name for the age group.
male_col, female_col	Column names for the male and female counts.
open_ended	Logical; if TRUE (default), the last age group is treated as open-ended.
graph	Logical. If TRUE (default), a ggplot2 plot of the male and female age ratios by age is attached to the result and shown on print.

Details

Let the sex ratio score (SRS) be the mean absolute difference between successive sex ratios, and the age ratio scores (ARSM, ARSF) be the mean absolute deviation of the age ratios from 100 for males and females respectively. The joint index is

$$UN = 3 \times SRS + ARSM + ARSF.$$

Sex ratios use the closed age groups; age ratios exclude the first group, the open-ended group, and (when open_ended = TRUE) the group adjacent to it. United Nations guidance regards values under 20 as accurate, 20-40 as inaccurate, and over 40 as highly inaccurate.

Value

An object of class `dem_unasa`: a list with the joint index, its components (SRS, ARSM, ARSF), the quality band, and a full table with the male and female counts, their age ratios and deviations, the sex ratio and its successive differences.

References

United Nations (1955). *Manual II: Methods of Appraisal of Quality of Basic Data for Population Estimates*. New York: United Nations.

Kpedekpo, G. M. K. (1982). *Essentials of Demographic Analysis for Africa*. London: Heinemann. (Chapter 3.)

Examples

```
# Ghana 2021, five-year age groups by sex (from the single-year data)
data(ghpop2021)
g <- within(ghpop2021, grp <- ifelse(Age >= 80, 80, (Age %/% 5) * 5))
wide <- tapply(g$Population, list(g$grp, g$Sex), sum)
five <- data.frame(age = as.integer(rownames(wide)),
                  males = wide[, "Male"], females = wide[, "Female"])
un_age_sex_accuracy(five, "age", "males", "females")

# Reproduces the published 1960 Ghana census worked example (Kpedekpo 1982)
ghana1960 <- data.frame(
  age = c("0-4", "5-9", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39",
          "40-44", "45-49", "50-54", "55-59", "60-64", "65+"),
  males = c(643041, 515520, 357831, 275542, 268336, 278601, 242515,
            198231, 168937, 122756, 96775, 59307, 63467, 113185),
  females = c(654258, 503070, 323460, 265534, 322576, 306329, 245883,
             179182, 145572, 95590, 81715, 48412, 54572, 100392)
)
un_age_sex_accuracy(ghana1960, "age", "males", "females")
```

whipple

*Whipple's Index of Age Heaping***Description**

Computes Whipple's index, a measure of preference for ages ending in the digits 0 and 5, from single-year-of-age population counts.

Usage

```
whipple(data, age_col, pop_col, lower = 23, upper = 62, graph = TRUE)
```

Arguments

data	A data frame with single years of age.
age_col	Column name for single year of age.
pop_col	Column name for the population count.
lower, upper	Age range over which to evaluate the index (defaults 23 and 62, the UN standard).
graph	Logical. If TRUE (default), a ggplot2 bar chart of the distribution by terminal digit is attached to the result and shown on print.

Details

Over the standard age range 23 to 62, Whipple's index is

$$WI = 100 \times \frac{\sum_{a \in \{25, 30, \dots, 60\}} P_a}{\frac{1}{5} \sum_{a=23}^{62} P_a},$$

i.e. the population at ages ending in 0 or 5 relative to one fifth of the total over the range. A value of 100 indicates no preference; 500 means every age was reported as a multiple of 5. The United Nations quality bands are: under 105 highly accurate, 105-109.9 fairly accurate, 110-124.9 approximate, 125-174.9 rough, and 175 or more very rough.

Value

An object of class `dem_whipple`: a list with the index value, its quality band, the numerator/denominator used, and a table giving the population and share by terminal digit over the range.

References

Shryock, H. S., & Siegel, J. S. (1976). *The Methods and Materials of Demography*. New York: Academic Press.

United Nations (1955). *Manual II: Methods of Appraisal of Quality of Basic Data for Population Estimates*. New York: United Nations.

Examples

```
# Ghana 2021 single-year ages, males
data(ghpop2021)
male <- subset(ghpop2021, Sex == "Male")
whipple(male, age_col = "Age", pop_col = "Population")
```

Index

* datasets

- data, [7](#)
- first_coef, [15](#)
- ghmort2021, [16](#)
- ghpop2021, [17](#)
- gphc2010, [17](#)
- last_coef, [21](#)
- middle_coef, [26](#)
- region2000, [31](#)
- standards, [34](#)
- testdata, [35](#)

age_ratio, [2](#)

brass_logit, [3](#)
brass_logit(), [34](#)

chm_brass, [5](#)

data, [7](#)
decompose_LE, [8](#)
decompose_LE(), [16](#)
decompose_rates, [9](#)
decompose_rates(), [34](#)
dem.cdr, [10](#)
dem.fert, [12](#)
dem.mmr, [13](#)
dm.chm, [14](#)

first_coef, [15](#), [20](#)

ggplot2::ggsave(), [31](#)
ghmort2021, [16](#)
ghpop2021, [17](#)
gphc2010, [17](#)

interpolation, [18](#)

karup_king, [19](#)
karup_king(), [7](#)

last_coef, [20](#), [21](#)

lifetable, [21](#)
lifetable(), [8](#), [17](#), [24](#)
lifetable_nqx, [23](#)

math_project, [25](#)
math_project(), [35](#)
middle_coef, [20](#), [26](#)
myers, [27](#)
myers(), [17](#)

project_population, [28](#)
project_population(), [31](#)
pyramid, [30](#)

region2000, [31](#)

sex_ratio, [32](#)
standardize, [33](#)
standards, [4](#), [34](#)

testdata, [35](#)

un_age_sex_accuracy, [36](#)
un_age_sex_accuracy(), [17](#)

whipple, [37](#)
whipple(), [17](#)