Package 'educineq'

July 22, 2025

Type Package

Title Compute and Decompose Inequality in Education

Version 0.1.0

Maintainer Vanesa Jorda <jordav@unican.es>

Description Easily compute education inequality measures and the distribution
of educational attainments for any group of countries, using the data set
developed in Jorda, V. and Alonso, JM. (2017) <DOI:10.1016/j.worlddev.2016.10.005>.
The package offers the possibility to compute not only the Gini index, but
also generalized entropy measures for different values of the sensitivity
parameter. In particular, the package includes functions to compute the
mean log deviation, which is more sensitive to the bottom part of the
distribution; the Theil's entropy measure, equally sensitive to all parts
of the distribution; and finally, the GE measure when the sensitivity
parameter is set equal to 2, which gives more weight to differences in
higher education. The decomposition of these measures in the components
between-country and within-country inequality is also provided. Two
graphical tools are also provided, to analyse the evolution of the
distribution of educational attainments: The cumulative distribution
function and the Lorenz curve.

Depends R (>= 2.10)

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports ineq, flexsurv

RoxygenNote 5.0.1

NeedsCompilation no

Author Vanesa Jorda [aut, cre],
Jose Manuel Alonso [aut]

Repository CRAN

Date/Publication 2017-02-17 14:38:11

Contents

---

data_country	*This dataset contains information about the available countries, their corresponding country codes and the regions they belong to, which are used to with* educineq *functions.*

Description

- country. Country name

- code. World Bank country code

- region. Macro-region to which the country belongs

Usage

```
data(data_country)
```

Format

A data frame with 142 rows and 3 variables

---

edcdf   *Cumulative distribution function of time of schooling*

Description

edcdf is a function to graph the CDF of time of schooling for any group of countries using the set of estimates developed in Jorda and Alonso (2017).

Usage

```
edcdf(countries, init.y, final.y, database)
```

Arguments

countries	character vector with the country codes of the countries to be used. Some macro-regions are already defined and can be used instead of the country codes: South Asia, Europe and Central Asia,Middle East and North Africa, Latin America and the Caribbean, Advanced Economies, Sub-Saharan Africa, East Asia and the Pacific. (see data_country).
init.y	the first year in which the function is calculated. Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.
final.y	the last year in which the function is calculated Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.
database	population subgrup for which the function is calculated. The following options are available:

1. "total15": Total population aged over-15.
2. "total25": Total population aged over-25.
3. "male15": Male population aged over-15.
4. "male25": Male population aged over-25.
5. "female15": Female population aged over-15.
6. "female25": Female population aged over-25.

Details

We use the set of estimates developed in Jorda and Alonso (2017), where the generalized gamma distribution (Stacy, 1962) is used to model the time that individuals attend school until they complete the educational cycle or decide to drop out. The reason is twofold; first, the generalized gamma distribution is a parsimonious model that nests most of the parametric assumptions described in the literature (see, Marshall and Olkin, 2007). Second, the generalized gamma distribution is able to model one- and zero-mode distributions and to represent several types of hazard rates.The flexibility of this model to consider such heterogeneity, makes it an outstanding candidate to model the distribution of education. It is important to highlight that this parametric model includes as particular cases most of the distributions commonly used in survival analysis, including the Weibull, the exponential, and the gamma distributions, so it would converge to any of its special cases if needed.

To accommodate time and country varying parameters, the distribution of education of each country and year is estimated by non-linear least squares (see, Jorda and Alonso (2017) for further description on the estimation strategy).The distribution of education of a particular group or region of countries is simply defined as a mixture of the national distributions, weighted by their population shares.

Value

edcdf returns a graph of the evolution of the CDF of education over the specified period.

References

Jorda, V. and Alonso, J.M. (2017). New estimates on educational attainment using a continuous approach (1970-2010), World Development, 90, 281 - 293. http://www.sciencedirect.com/science/article/pii/S0305750X16305010

Marshall, A. W. and Olkin, I. (2007). Life distributions. Structure of nonparametric, semiparametric, and parametric families. New York: Springer.

Stacy, E. W. (1962). A generalization of the gamma distribution. Annals of Mathematical Statistics, 33, 1187 - 1192.

See Also

GenGamma.orig, data_country. Visit http://www.educationdata.unican.esfor more information on the constructoin of the dataset and the available countries.

Examples

```
edcdf(countries = "South Asia", init.y = 1980, final.y = 1990, database = "female25")
edcdf(countries = c("DNK", "FIN", "ISL", "NOR", "SWE"),init.y = 1995,
final.y = 2010, database = "male25")
```

---

ege2   *Generalized entropy measure of education*

Description

ege2 function computes the generalized entropy measure of education, with the sensitivity parameter set to 2, for any group of countries included in the dataset developed in Jorda and Alonso (2017). The function also provides a decomposition of this index in between-county and within- country inequality.

Usage

```
ege2(countries, init.y, final.y, database, plot = TRUE)
```

Arguments

`countries`	character vector with the country codes of the countries to be used. Some macroregions are already defined and can be used instead of the country codes: South Asia, Europe and Central Asia, Middle East and North Africa, Latin America and the Caribbean, Advanced Economies, Sub-Saharan Africa, East Asia and the Pacific and all for the 142 counrties included in the dataset.(see data_country).
`init.y`	the first year in which the function is calculated. Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.
`final.y`	the last year in which the function is calculated Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.
`database`	population subgrup for which the function is calculated. The following options are available:

1. `"total15"`: Total population aged over-15.
2. `"total25"`: Total population aged over-25.
3. `"male15"`: Male population aged over-15.
4. `"male25"`: Male population aged over-25.
5. `"female15"`: Female population aged over-15.
6. `"female25"`: Female population aged over-25.

`plot`	if TRUE (the default), displays a graph of the results.

Details

The estimates of the generalized entropy measure for the specified group of countries can be easily derived by taking advantage of the decomposition of this family. It is computed as the sum of the following terms, which correspond to within- country and between, country inequality respectively (see, e.g., Cowell, 2011):

$$GE(2)_W = \sum_{i=1}^{N} s_i^2 p_i^{-1} GE(2)_i;$$

$$GE(2)_B = 0.5 \sum_{i=1}^{N} p_i (\mu_i/\mu)^2 - 1,$$

where N is the number of countries, $GE(2)_i$ and p_i denote, respectively, the generalized entropy measure and the population weight of the country i, and s_i stands for the proportion of mean income of the country i in the overall mean of the group: $s_i = \lambda_i \mu_i / \sum_{i=1}^{N} \lambda_i \mu_i$.

Value

ege2 returns a list with the following objects:

1. `GE_2`: evolution of the generalized entropy measure of education from the initial to the last year, decomposed in between-country and within-country inequality.
2. `countries`: countries used to compute the generalized entropy measure.
3. If `plot = TRUE`, graphical representation of the numerical results.

References

Cowell, F. (2011). Measuring inequality. Oxford University Press.

Jorda, V. and Alonso, J.M. (2017). New estimates on educational attainment using a continuous approach (1970-2010), World Development, 90, 281 - 293. http://www.sciencedirect.com/science/article/pii/S0305750X16305010

See Also

data_country. Visit http://www.educationdata.unican.es for more information on the constructoin of the dataset and the available countries.

Examples

```
ege2(countries = "all", init.y = 1980, final.y = 2000,
     database = "total25")
ege2(countries = c("DNK", "FIN", "ISL", "NOR", "SWE"), init.y = 1980,
     final.y = 2000, database = "female15")
```

| egini | *Gini index of education* |

Description

egini is a function to compute the Gini index of education for any group of countries using the set of estimates developed in Jorda and Alonso (2017).

Usage

```
egini(countries, init.y, final.y, database, M = 5000, plot = TRUE)
```

Arguments

countries	character vector with the country codes of the countries to be used. Some macro-regions are already defined and can be used instead of the country codes: South Asia, Europe and Central Asia, Middle East and North Africa, Latin America and the Caribbean, Advanced Economies, Sub-Saharan Africa, East Asia and the Pacific and all for the 142 counrties included in the dataset (see data_country).
init.y	the first year in which the function is calculated. Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.
final.y	the last year in which the function is calculated Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.
database	population subgrup for which the function is calculated. The following options are available:

1. "total15": Total population aged over-15.
2. "total25": Total population aged over-25.

3. `"male15"`: Male population aged over-15.

4. `"male25"`: Male population aged over-25.

5. `"female15"`: Female population aged over-15.

6. `"female25"`: Female population aged over-25.

M   size of the simulated sample.

plot   if TRUE (the default), displays a graph of the results.

Details

We use the set of estimates developed in Jorda and Alonso (2017), where the generalized gamma distribution (Stacy, 1962) is used to model the time that individuals attend school until they complete the educational cycle or decide to drop out. The Gini index is computed from a synthetic sample of size M of the distribution of education of the specified group of countries. The sample is obtained by Monte Carlo simulation using the mixture of the national distributions, weighted by their population shares.

Value

`egini` returns a list with the following objects:

1. `Gini_index`: evolution of the Gini index of education from the initial to the last year.

2. `countries`: countries used to compute the Gini index.

3. If `plot = TRUE`, graphical representation of the numerical results.

References

Cowell, F. (2011). Measuring inequality. Oxford University Press.

Jorda, V. and Alonso, J.M. (2017). New estimates on educational attainment using a continuous approach (1970-2010), World Development, 90, 281 - 293. http://www.sciencedirect.com/science/article/pii/S0305750X16305010

Stacy, E. W. (1962). A generalization of the gamma distribution. Annals of Mathematical Statistics, 33, 1187 - 1192.

See Also

[GenGamma.orig](), [Gini](), [data_country](). Visit http://www.educationdata.unican.es for more information on the constructoin of the dataset and the available [countries]().

Examples

```
egini(countries = c("DNK", "FIN"), init.y = 1995, final.y = 1995,
  database = "male25", M = 100, plot = FALSE)
```

---

elc *Lorenz curve of education*

Description

elc is a function to graph the Lorenz curve of education for any group of countries using the set of estimates developed in Jorda and Alonso (2017).

Usage

```
elc(countries, init.y, final.y, database, M = 5000)
```

Arguments

countries   character vector with the country codes of the countries to be used. Some macro-regions are already defined and can be used instead of the country codes: South Asia, Europe and Central Asia, Middle East and North Africa, Latin America and the Caribbean, Advanced Economies, Sub-Saharan Africa, East Asia and the Pacific. (see data_country).

init.y   the first year in which the function is calculated. Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.

final.y   the last year in which the function is calculated Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.

database   population subgrup for which the function is calculated. The following options are available:

1. "total15": Total population aged over-15.
2. "total25": Total population aged over-25.
3. "male15": Male population aged over-15.
4. "male25": Male population aged over-25.
5. "female15": Female population aged over-15.
6. "female25": Female population aged over-25.

M   size of the simulated sample (default M = 5000).

Details

We use the set of estimates developed in Jorda and Alonso (2017), where the generalized gamma distribution (Stacy, 1962) is used to model the time that individuals attend school until they complete the educational cycle or decide to drop out. To accommodate time and country varying parameters, the distribution of education of each country and year is estimated by non-linear least squares (see, Jorda and Alonso (2017) for further description on the estimation strategy).The Lorenz curve is computed from a synthetic sample of size M of the distribution of education of the specified group of countries. The sample is obtained by Monte Carlo simulation using the mixture of the national distributions, weighted by their population shares.

Value

`elc` returns a graph of the evolution of the Lorenz curve of education over the specified period.

References

Cowell, F. (2011). Measuring inequality. Oxford University Press.

Jorda, V. and Alonso, J.M. (2017). New estimates on educational attainment using a continuous approach (1970-2010), World Development, 90, 281 - 293. http://www.sciencedirect.com/science/article/pii/S0305750X16305010

Stacy, E. W. (1962). A generalization of the gamma distribution. Annals of Mathematical Statistics, 33, 1187 - 1192.

See Also

GenGamma.orig, Lc, data_country. Visit http://www.educationdata.unican.es for more information on the constructoin of the dataset and the available countries.

Examples

```
elc(countries = c("CAN","USA"), init.y = 1985, final.y = 1985,
  database = "female25", M = 300)
```

---

emean   *Mean years of schooling*

Description

`emean` is a function to compute mean years of schooling for any group of countries included in the dataset developed in Jorda and Alonso (2017). It is computed as the average of the national years of schooling weighted by population weights.

Usage

```
emean(countries, init.y, final.y, database, plot = TRUE)
```

Arguments

`countries`	character vector with the country codes of the countries to be used. Some macro-regions are already defined and can be used instead of the country codes: `South Asia`, `Europe and Central Asia`,`Middle East and North Africa`, `Latin America and the Caribbean`, `Advanced Economies`, `Sub-Saharan Africa`, `East Asia and the Pacific` and `all` for the 142 counrties included in the dataset.(see `data_country`).
`init.y`	the first year in which the function is calculated. Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.

| final.y | the last year in which the function is calculated Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010. |
| database | population subgrup for which the function is calculated. The following options are available: |

1. "total15": Total population aged over-15.
2. "total25": Total population aged over-25.
3. "male15": Male population aged over-15.
4. "male25": Male population aged over-25.
5. "female15": Female population aged over-15.
6. "female25": Female population aged over-25.

| plot | if TRUE (the default), displays a graph of the results. |

Value

emean returns a list with the following objects:

1. mean_years_of_schooling: evolution of mean years of schooling from the initial to the last year.
2. countries: countries used to compute mean years of schooling.
3. If plot = TRUE, graphical representation of the numerical results.

References

Jorda, V. and Alonso, J.M. (2017). New estimates on educational attainment using a continuous approach (1970-2010), World Development, 90, 281 - 293. http://www.sciencedirect.com/science/article/pii/S0305750X16305010

See Also

[data_country](). Visit http://www.educationdata.unican.es for more information on the constructoin of the dataset and the available [countries]().

Examples

```
emean(countries = "Advanced Economies", init.y = 1980, final.y = 2000,
      database = "male25")
emean(countries = c("DNK", "FIN", "ISL", "NOR", "SWE"), init.y = 1980,
      final.y = 2000, database = "male25")
```

---

emld   *Mean log deviation (MLD) of education*

Description

emld function computes the MLD of education, with for any group of countries included in the dataset developed in Jorda and Alonso (2017). The function also provides a decomposition of this index in between-county and within-country inequality.

Usage

```
emld(countries, init.y, final.y, database, plot = TRUE)
```

Arguments

countries   character vector with the country codes of the countries to be used. Some macro-regions are already defined and can be used instead of the country codes: South Asia, Europe and Central Asia, Middle East and North Africa, Latin America and the Caribbean, Advanced Economies, Sub-Saharan Africa, East Asia and the Pacific and all for the 142 counrties included in the dataset.(see data_country).

init.y   the first year in which the function is calculated. Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.

final.y   the last year in which the function is calculated Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.

database   population subgrup for which the function is calculated. The following options are available:

1. "total15": Total population aged over-15.
2. "total25": Total population aged over-25.
3. "male15": Male population aged over-15.
4. "male25": Male population aged over-25.
5. "female15": Female population aged over-15.
6. "female25": Female population aged over-25.

plot   if TRUE (the default), displays a graph of the results.

Details

The estimates of the MLD for the specified group of countries can be easily derived by taking advantage of the decomposition of this family. It is computed as the sum of the following terms, which correspond to within- country and between, country inequality respectively (see, e.g., Cowell, 2011):

$MLD_W = \sum_{i=1}^{N} p_i MLD_i;$

$MLD_B = \sum_{i=1}^{N} p_i log(\mu/\mu_i),$

where N is the number of countries, MLD_i and p_i denote, respectively, the MDL and the population weight of the country i.

Value

emld returns a list with the following objects:

1. `MLD`: evolution of the MLD of education from the initial to the last year, decomposed in between-country and within-country inequality.

2. `countries`: countries used to compute the MLD.

3. If `plot = TRUE`, graphical representation of the numerical results.

References

Cowell, F. (2011). Measuring inequality. Oxford University Press.

Jorda, V. and Alonso, J.M. (2017). New estimates on educational attainment using a continuous approach (1970-2010), World Development, 90, 281 - 293. http://www.sciencedirect.com/science/article/pii/S0305750X16305010

See Also

[data_country](). Visit http://www.educationdata.unican.es for more information on the constructoin of the dataset and the available [countries]().

Examples

```
emld(countries = "East Asia and the Pacific", init.y = 1980,
     final.y = 2000, database = "female25")
emld(countries = c("DNK", "FIN", "ISL", "NOR", "SWE"), init.y = 1980,
     final.y = 2000, database = "total25")
```

---

etheil   *Theil index of education*

Description

`etheil` is a function to compute the Theil index of education for any group of countries included in the dataset developed in Jorda and Alonso (2017). The function also provides a decomposition of this index in between-county and within-country inequality.

Usage

```
etheil(countries, init.y, final.y, database, plot = TRUE)
```

Arguments

countries
: character vector with the country codes of the countries to be used. Some macro-regions are already defined and can be used instead of the country codes: South Asia, Europe and Central Asia, Middle East and North Africa, Latin America and the Caribbean, Advanced Economies, Sub-Saharan Africa, East Asia and the Pacific and all for the 142 counrties included in the dataset.(see data_country).

init.y
: the first year in which the function is calculated. Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.

final.y
: the last year in which the function is calculated Available years are 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010.

database
: population subgrup for which the function is calculated. The following options are available:

  1. "total15": Total population aged over-15.
  2. "total25": Total population aged over-25.
  3. "male15": Male population aged over-15.
  4. "male25": Male population aged over-25.
  5. "female15": Female population aged over-15.
  6. "female25": Female population aged over-25.

plot
: if TRUE (the default), displays a graph of the results.

Details

The estimates of the Theil index for the specified group of countries can be easily derived by taking advantage of the decomposition of this family. It is computed as the sum of the following terms, which correspond to within- country and between, country inequality respectively (see, e.g., Cowell, 2011):

$$T_W = \sum_{i=1}^{N} s_i T_i;$$

$$T_B = \sum_{i=1}^{N} s_i log(\mu_i/\mu),$$

where N is the number of countries, T_i denotes the Theil index of the country i and s_i stands for the proportion of mean income of the country i in the overall mean of the group: $s_i = \lambda_i \mu_i / \sum_{i=1}^{N} \lambda_i \mu_i$.

Value

etheil returns a list with the following objects:

1. Theli_index: evolution of the Theil index of education from the initial to the last year, decomposed in between-country and within-country inequality.

2. countries: countries used to compute the Theil index.

3. If plot = TRUE, graphical representation of the numerical results.

References

Cowell, F. (2011). Measuring inequality. Oxford University Press.

Jorda, V. and Alonso, J.M. (2017). New estimates on educational attainment using a continuous approach (1970-2010), World Development, 90, 281 - 293. http://www.sciencedirect.com/science/article/pii/S0305750X16305010

See Also

data_country. Visit http://www.educationdata.unican.es for more information on the constructoin of the dataset and the available countries.

Examples

```
etheil(countries = "Advanced Economies", init.y = 1980, final.y = 2000,
       database = "male25")
etheil(countries = c("DNK", "FIN", "ISL", "NOR", "SWE"), init.y = 1980,
       final.y = 2000, database = "female15")
```

ineq_female15a	*This dataset contains some statistics about the distribution of educational attainments for female population aged over 15, taken from* www.educationdata.unican.es. *The data on population come from Barro and Lee database available at* http://www.barrolee.com

Description

- country. Country name
- year
- code. World Bank country code
- region. Macro-region to which the country belongs
- mys. Mean years of schooling
- mld. Mean log deviation of education.
- theil. Theil index of education
- ge2. Generalized entropy measure of education
- pop. Total population, http://www.barrolee.com

Usage

```
data(ineq_female15a)
```

Format

A data frame with 1278 rows and 9 variables

ineq_female25a	*This dataset contains some statistics about the distribution of educational attainments for female population aged over 25, taken from* <www.educationdata.unican.es>. *The data on population come from Barro and Lee database available at* <http://www.barrolee.com>

Description

- country. Country name
- year
- code. World Bank country code
- region. Macro-region to which the country belongs
- mys. Mean years of schooling
- mld. Mean log deviation of education.
- theil. Theil index of education
- ge2. Generalized entropy measure of education
- pop. Total population, <http://www.barrolee.com>

Usage

```
data(ineq_female25a)
```

Format

A data frame with 1278 rows and 9 variables

ineq_male15a	*This dataset contains some statistics about the distribution of educational attainments for male population aged over 15, taken from* <www.educationdata.unican.es>. *The data on population come from Barro and Lee database available at* <http://www.barrolee.com>

Description

- country. Country name
- year
- code. World Bank country code
- region. Macro-region to which the country belongs
- mys. Mean years of schooling
- mld. Mean log deviation of education.
- theil. Theil index of education
- ge2. Generalized entropy measure of education
- pop. Total population, <http://www.barrolee.com>

Usage

```
data(ineq_male15a)
```

Format

A data frame with 1278 rows and 9 variables

---

ineq_male25a   *This dataset contains some statistics about the distribution of edu-
                      cational attainments for male population aged over 25, taken from
                      www.educationdata.unican.es. The data on population come from
                      Barro and Lee database available at* http://www.barrolee.com

---

Description

- country. Country name

- year

- code. World Bank country code

- region. Macro-region to which the country belongs

- mys. Mean years of schooling

- mld. Mean log deviation of education.

- theil. Theil index of education

- ge2. Generalized entropy measure of education

- pop. Total population, http://www.barrolee.com

Usage

```
data(ineq_male25a)
```

Format

A data frame with 1278 rows and 9 variables

ineq_total15a	*This dataset contains some statistics about the distribution of educational attainments for population aged over 15, taken from* www.educationdata.unican.es. *The data on population come from Barro and Lee database available at* http://www.barrolee.com

Description

- country. Country name
- year
- code. World Bank country code
- region. Macro-region to which the country belongs
- mys. Mean years of schooling
- mld. Mean log deviation of education.
- theil. Theil index of education
- ge2. Generalized entropy measure of education
- pop. Total population, http://www.barrolee.com

Usage

```
data(ineq_total15a)
```

Format

A data frame with 1278 rows and 9 variables

ineq_total25a	*This dataset contains some statistics about the distribution of educational attainments for population aged over 25, taken from* www.educationdata.unican.es. *The data on population come from Barro and Lee database available at* http://www.barrolee.com

Description

- country. Country name
- year
- code. World Bank country code
- region. Macro-region to which the country belongs
- mys. Mean years of schooling
- mld. Mean log deviation of education.
- theil. Theil index of education
- ge2. Generalized entropy measure of education
- pop. Total population, http://www.barrolee.com

Usage

```
data(ineq_total25a)
```

Format

A data frame with 1278 rows and 9 variables

Index