Package 'PHEindicatormethods'

July 30, 2018

Type Package

Version 1.0.8

Title Common Public Health Statistics and their Confidence Intervals

Description Functions to calculate commonly used public health statistics and their confidence intervals using methods approved for use in the production of Public Health England indicators such as those presented via Fingertips (<http://fingertips.phe.org.uk/>). It provides functions for the generation of proportions, crude rates, means, directly standardised rates, indirectly standardised rates and standardised mortality ratios. Statistical methods are referenced in the following publications.
Breslow NE, Day NE (1987) <doi:10.1002/sim.4780080614>.
Dobson et al (1991) <doi:10.1002/sim.4780100317>.
Armitage P, Berry G (2002) <doi:10.1002/9780470773666>.
Wilson EB. (1927) <doi:10.1080/01621459.1927.10502953>.
Altman DG et al (2000, ISBN: 978-0-727-91375-3).

BugReports https://github.com/PublicHealthEngland/PHEindicatormethods/issues

Depends R (>= 3.1.0)

License GPL-3

Encoding UTF-8

LazyData true

Imports dplyr, rlang

Suggests knitr, readxl, rmarkdown, RODBC, testthat,

RoxygenNote 6.0.1

VignetteBuilder knitr

NeedsCompilation no

Author Anderson Georgina [aut, cre],
Fox Sebastian [ctb],
Francis Matthew [ctb],
Fryers Paul [ctb]

Maintainer Anderson Georgina <georgina.anderson@phe.gov.uk>

Repository CRAN

Date/Publication 2018-07-30 10:50:13 UTC

R topics documented:

`byars_lower` *byars_lower*

Description

Calculates the lower confidence limits for observed numbers of events using Byar's method [1].

Usage

```
byars_lower(x, confidence = 0.95)
```

Arguments

`x`	the observed numbers of events; numeric vector; no default
`confidence`	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

Value

Returns lower confidence limits for observed numbers of events using Byar's method [1]

Notes

`byars_lower` and `byars_upper` together return symmetric confidence intervals around counts, therefore for a specified confidence level, α, the probability that, by chance, the lower limit returned will be above the true underlying value, is $\alpha/2$. If the confidence level is very close to 1 or the number of events is very small Byar's method is inaccurate and may return a negative number - in these cases an error is returned.

References

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.

See Also

Other PHEindicatormethods package functions: byars_upper, phe_dsr, phe_isr, phe_mean, phe_proportion, phe_rate, phe_smr, wilson_lower, wilson_upper

Examples

```
byars_lower(65)
byars_lower(65,99.8)
```

---

byars_upper *byars_upper*

Description

Calculates the upper confidence limits for observed numbers of events using Byar's method [1].

Usage

```
byars_upper(x, confidence = 0.95)
```

Arguments

x	the observed numbers of events; numeric vector; no default
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

Value

Returns upper confidence limits for observed numbers of events using Byar's method [1]

Notes

byars_lower and byars_upper together return symmetric confidence intervals around counts, therefore for a specified confidence level, α, the probability that, by chance, the upper limit returned will be below the true underlying value, is $\alpha/2$.

References

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.

See Also

Other PHEindicatormethods package functions: byars_lower, phe_dsr, phe_isr, phe_mean, phe_proportion, phe_rate, phe_smr, wilson_lower, wilson_upper

Examples

```
byars_upper(65)
byars_upper(65,99.8)
```

---

esp2013 *European Standard Population 2013*

Description

A numeric vector containing nineteen 5-year age band populations making up the 2013 European Standard Population ordered from age 0-4, 5-9, 10-14 ... to ... 85-89, 90+. Sorted by increasing age band.

Usage

```
esp2013
```

Format

A numeric vector with 19 elements

Value

5000 5500 5500 5500 6000 6000 6500 7000 7000 7000 7000 6500 6000 5500 5000 4000 2500 1500 1000

Notes

The 2013 European Standard Population is modelled and published by Eurostat [1] for use in the production of age-standardised rates. It uses the unweighted average 2010-based population projections of the European Union (x27) and European Free Trade Association (x4) countries for the period 2011-2030 broken down into 5-year age bands from age 0 - age 95+ with the 0-5 age band separated into age 0 and age 1-4. The version provided with this package combines the age 0 and age 1-4 populations into a single 0-4 age band and combines the 90-94 and 95+ populations into a single 90+ age band, giving 19 age bands in total.

References

[1] Eurostat Methodologies and Working Papers. Revision of the European Standard Population: Report of Eurostat's Taskforce, 2013.
http://ec.europa.eu/eurostat/documents/3859598/5926869/KS-RA-13-028-EN.PDF/e713fa79-1add-44e8-b23d-

Examples

```
esp2013
```

`PHEindicatormethods` *PHEindicatormethods: A package for performing standard statistics for public health indicators*

Description

PHEindicatormethods: A package for performing standard statistics for public health indicators

---

`phe_dsr` *phe_dsr*

Description

Calculates directly standardised rates with confidence limits using Byar's method [1] with Dobson method adjustment [2].

Usage

```
phe_dsr(data, x, n, stdpop = esp2013, stdpoptype = "vector",
  type = "standard", confidence = 0.95, multiplier = 1e+05)
```

Arguments

`data`	data.frame containing the data to be standardised, pre-grouped if multiple DSRs required; unquoted string; no default
`x`	field name from data containing the observed number of events for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
`n`	field name from data containing the populations for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
`stdpop`	the standard populations for each standardisation category (eg age band); unquoted string referencing a numeric vector or field name from data depending on value of stdpoptype; default = esp2013
`stdpoptype`	whether the stdpop has been specified as a vector or a field name from data; quoted string "field" or "vector"; default = "vector"
`type`	type of output; can be "value", "lower", "upper", "standard" (for all 3 previous fields) or "full"; quoted string; default standard
`confidence`	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
`multiplier`	the multiplier used to express the final values (eg 100,000 = rate per 100,000); numeric; default 100,000

Value

When type = "full", returns a tibble of total counts, total populations, directly standardised rates, lower confidence limits, upper confidence limits, confidence level, statistic and method for each grouping set

Notes

User MUST ensure that x, n and stdpop vectors are all ordered by the same standardisation category values as records will be matched by position.

For total counts >= 10 Byar's method [1] is applied using the byars_lower and byars_upper functions. When the total count is < 10 DSRs are not reliable and will therefore not be calculated.

References

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.

[2] Dobson A et al. Confidence intervals for weighted sums of Poisson parameters. Stat Med 1991;10:457-62.

See Also

Other PHEindicatormethods package functions: byars_lower, byars_upper, phe_isr, phe_mean, phe_proportion, phe_rate, phe_smr, wilson_lower, wilson_upper

Examples

```
library(dplyr)
df <- data.frame(indicatorid = rep(c(1234, 5678, 91011, 121314), each = 19 * 2 * 5),
                 year = rep(2006:2010, each = 19 * 2),
                 sex = rep(rep(c("Male", "Female"), each = 19), 5),
                 ageband = rep(c(0,5,10,15,20,25,30,35,40,45,
                                 50,55,60,65,70,75,80,85,90), times = 10),
                 obs = sample(200, 19 * 2 * 5 * 4, replace = TRUE),
                 pop = sample(10000:20000, 19 * 2 * 5 * 4, replace = TRUE))
df %>%
    group_by(indicatorid, year, sex) %>%
    phe_dsr(obs, pop)

## OR

df %>%
    group_by(indicatorid, year, sex) %>%
    phe_dsr(obs, pop, type = "full")
```

phe_isr *phe_isr*

Description

Calculates indirectly standardised rates with confidence limits using Byar's [1] or exact [2] CI method.

Usage

```
phe_isr(data, x, n, x_ref, n_ref, refpoptype = "vector", type = "standard",
  confidence = 0.95, multiplier = 1e+05)
```

Arguments

data	data.frame containing the data to be standardised, pre-grouped if multiple ISRs required; unquoted string; no default
x	field name from data containing the observed number of events for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
n	field name from data containing the populations for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
x_ref	the observed number of events in the reference population for each standardisation category (eg age band); unquoted string referencing a numeric vector or field name from data depending on value of refpoptype; no default
n_ref	the reference population for each standardisation category (eg age band); unquoted string referencing a numeric vector or field name from data depending on value of refpoptype; no default
refpoptype	whether x_ref and n_ref have been specified as vectors or a field name from data; quoted string "field" or "vector"; default = "vector"
type	type of output; can be "value", "lower", "upper", "standard" (for all 3 previous fields) or "full"; quoted string; default standard
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
multiplier	the multiplier used to express the final values (eg 100,000 = rate per 100,000); numeric; default 100,000

Value

When type = "full", returns a tibble of observed events, expected events, indirectly standardised rate, lower confidence limit, upper confidence limit, confidence level, statistic and method for each grouping set

Notes

User MUST ensure that x, n, x_ref and n_ref vectors are all ordered by the same standardisation category values as records will be matched by position.

For numerators >= 10 Byar's method [1] is applied using the byars_lower and byars_upper functions. For small numerators Byar's method is less accurate and so an exact method [2] based on the Poisson distribution is used.

References

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.

[2] Armitage P, Berry G. Statistical methods in medical research (4th edn). Oxford: Blackwell; 2002.

See Also

Other PHEindicatormethods package functions: byars_lower, byars_upper, phe_dsr, phe_mean, phe_proportion, phe_rate, phe_smr, wilson_lower, wilson_upper

Examples

```
library(dplyr)
df <- data.frame(indicatorid = rep(c(1234, 5678, 91011, 121314), each = 19 * 2 * 5),
                 year = rep(2006:2010, each = 19 * 2),
                 sex = rep(rep(c("Male", "Female"), each = 19), 5),
                 ageband = rep(c(0,5,10,15,20,25,30,35,40,45,
                                 50,55,60,65,70,75,80,85,90), times = 10),
                 obs = sample(200, 19 * 2 * 5 * 4, replace = TRUE),
                 pop = sample(10000:20000, 19 * 2 * 5 * 4, replace = TRUE))

refdf <- data.frame(refcount = sample(200, 19, replace = TRUE),
                    refpop = sample(10000:20000, 19, replace = TRUE))

df %>%
    group_by(indicatorid, year, sex) %>%
    phe_isr(obs, pop, refdf$refcount, refdf$refpop)

## OR

df %>%
    group_by(indicatorid, year, sex) %>%
    phe_isr(obs, pop, refdf$refcount, refdf$refpop, type="full", confidence=99.8)
```

phe_mean *phe_mean*

Description

Calculates means with confidence limits using Student's t-distribution method.

Usage

```
phe_mean(data, x, type = "standard", confidence = 0.95)
```

Arguments

data	a data.frame containing the data to calculate means for, pre-grouped if multiple means required; unquoted string; no default
x	field name from data containing the values to calculate the means for; unquoted string; no default
type	type of output; can be "value", "lower", "upper", "standard" (for all 3 previous fields) or "full"; quoted string; default standard
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

Value

When type = "full", returns a data.frame of value_sum, value_count, stdev, value, lowercl, uppercl, confidence, statistic and method for each grouping set

See Also

Other PHEindicatormethods package functions: byars_lower, byars_upper, phe_dsr, phe_isr, phe_proportion, phe_rate, phe_smr, wilson_lower, wilson_upper

Examples

```
library(dplyr)
df <- data.frame(values = c(30,40,50,60))
phe_mean(df, values)

## OR

df2 <- data.frame(area = rep(c("Area1", "Area2"),each=3),
                  values = c(20,30,40,200,300,400)) %>%
                  group_by(area)
phe_mean(df2,values)
phe_mean(df2,values,type="full", confidence=0.998)
```

phe_proportion *phe_proportion*

Description

Calculates proportions with confidence limits using Wilson Score method [1,2].

Usage

```
phe_proportion(data, x, n, type = "standard", confidence = 0.95,
  percentage = FALSE)
```

Arguments

data	a data.frame containing the data to calculate proportions for; unquoted string; no default
x	field name from data containing the observed numbers of cases in the sample meeting the required condition (the numerator for the proportion); unquoted string; no default
n	field name from data containing the number of cases in the sample (the denominator for the proportion); unquoted string; no default
type	type of output; can be "value", "lower", "upper", "standard" (for all 3 previous fields) or "full"; quoted string; default standard
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
percentage	whether the output should be returned as a percentage; logical; default FALSE

Value

When type = "full", returns the original data.frame with the following appended: proportion, lower confidence limit, upper confidence limit, confidence level, statistic and method

Notes

Wilson Score method [1,2] is applied using the wilson_lower and wilson_upper functions.

References

[1] Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc; 1927; 22. Pg 209 to 212.
[2] Newcombe RG, Altman DG. Proportions and their differences. In Altman DG et al. (eds). Statistics with confidence (2nd edn). London: BMJ Books; 2000. Pg 46 to 48.

See Also

Other PHEindicatormethods package functions: byars_lower, byars_upper, phe_dsr, phe_isr, phe_mean, phe_rate, phe_smr, wilson_lower, wilson_upper

Examples

```
df <- data.frame(area = c("Area1","Area2","Area3"),
                 numerator = c(65,82,100),
                 denominator = c(100,100,100))

phe_proportion(df, numerator, denominator)
phe_proportion(df, numerator, denominator, confidence=99.8)
phe_proportion(df, numerator, denominator, type="full")
```

---

phe_rate *phe_rate*

Description

Calculates rates with confidence limits using Byar's [1] or exact [2] CI method.

Usage

```
phe_rate(data, x, n, type = "standard", confidence = 0.95,
  multiplier = 1e+05)
```

Arguments

data	the data.frame containing the data to calculate rates for; unquoted string; no default
x	field name from data containing the rate numerators (eg observed number of events); unquoted string; no default
n	field name from data containing the rate denominators (eg populations); unquoted string; no default
type	type of output; can be "value", "lower", "upper", "standard" (for all 3 previous fields) or "full"; quoted string; default standard
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
multiplier	the multiplier used to express the final values (eg 100,000 = rate per 100,000); numeric; default 100,000

Value

When type = "full", returns the original data.frame with the following appended: rate, lower confidence limit, upper confidence limit, confidence level, statistic and method

Notes

For numerators >= 10 Byar's method [1] is applied using the byars_lower and byars_upper functions. For small numerators Byar's method is less accurate and so an exact method [2] based on the Poisson distribution is used.

References

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.
[2] Armitage P, Berry G. Statistical methods in medical research (4th edn). Oxford: Blackwell; 2002.

See Also

Other PHEindicatormethods package functions: byars_lower, byars_upper, phe_dsr, phe_isr, phe_mean, phe_proportion, phe_smr, wilson_lower, wilson_upper

Examples

```
df <- data.frame(area = rep(c("Area1","Area2","Area3","Area4"), 2),
                 year = rep(2015:2016, each = 4),
                 obs = sample(100, 2 * 4, replace = TRUE),
                 pop = sample(100:200, 2 * 4, replace = TRUE))
phe_rate(df, obs, pop)
phe_rate(df, obs, pop, type="full", confidence=99.8, multiplier=100)
```

---

phe_smr *phe_smr*

Description

Calculates standard mortality ratios (or indirectly standardised ratios) with confidence limits using Byar's [1] or exact [2] CI method.

Usage

```
phe_smr(data, x, n, x_ref, n_ref, refpoptype = "vector", type = "standard",
  confidence = 0.95, refvalue = 1)
```

Arguments

data	data.frame containing the data to be standardised, pre-grouped if multiple SMRs required; unquoted string; no default
x	field name from data containing the observed number of events for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
n	field name from data containing the populations for each standardisation category (eg ageband) within each grouping set (eg area); unquoted string; no default
x_ref	the observed number of events in the reference population for each standardisation category (eg age band); unquoted numeric vector or field name from data depending on value of refpoptype; no default

n_ref	the reference population for each standardisation category (eg age band); unquoted numeric vector or field name from data depending on value of refpoptype; no default
refpoptype	whether x_ref and n_ref have been specified as vectors or a field name from data; quoted string "field" or "vector"; default = "vector"
type	type of output; can be "value", "lower", "upper", "standard" (for all 3 previous fields) or "full"; quoted string; default standard
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95
refvalue	the standardised reference ratio, numeric, default = 1

Value

When type = "full", returns a tibble of observed events, expected events, standardised mortality ratios, lower confidence limits, upper confidence limits, confidence level, statistic and method for each grouping set

Notes

User MUST ensure that x, n, x_ref and n_ref vectors are all ordered by the same standardisation category values as records will be matched by position.

For numerators >= 10 Byar's method [1] is applied using the byars_lower and byars_upper functions. For small numerators Byar's method is less accurate and so an exact method [2] based on the Poisson distribution is used.

References

[1] Breslow NE, Day NE. Statistical methods in cancer research, volume II: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, World Health Organisation; 1987.

[2] Armitage P, Berry G. Statistical methods in medical research (4th edn). Oxford: Blackwell; 2002.

See Also

Other PHEindicatormethods package functions: byars_lower, byars_upper, phe_dsr, phe_isr, phe_mean, phe_proportion, phe_rate, wilson_lower, wilson_upper

Examples

```
library(dplyr)
df <- data.frame(indicatorid = rep(c(1234, 5678, 91011, 121314), each = 19 * 2 * 5),
                 year = rep(2006:2010, each = 19 * 2),
                 sex = rep(rep(c("Male", "Female"), each = 19), 5),
                 ageband = rep(c(0,5,10,15,20,25,30,35,40,45,
                                 50,55,60,65,70,75,80,85,90), times = 10),
                 obs = sample(200, 19 * 2 * 5 * 4, replace = TRUE),
```

```
                  pop = sample(10000:20000, 19 * 2 * 5 * 4, replace = TRUE))

refdf <- data.frame(refcount = sample(200, 19, replace = TRUE),
                    refpop = sample(10000:20000, 19, replace = TRUE))

df %>%
    group_by(indicatorid, year, sex) %>%
    phe_smr(obs, pop, refdf$refcount, refdf$refpop)

## OR

df %>%
    group_by(indicatorid, year, sex) %>%
   phe_smr(obs, pop, refdf$refcount, refdf$refpop, type="full", confidence=99.8, refvalue=100)
```

`wilson_lower` *wilson_lower*

Description

Calculates lower confidence limits for observed numbers of events using the Wilson Score method [1,2].

Usage

```
wilson_lower(x, n, confidence = 0.95)
```

Arguments

`x`	the observed numbers of cases in the samples meeting the required condition; numeric vector; no default
`n`	the numbers of cases in the samples; numeric vector; no default
`confidence`	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

Value

Returns lower confidence limits for observed numbers of events using the Wilson Score method [1,2]

Notes

`wilson_lower` and `wilson_upper` together return symmetric confidence intervals, therefore for a specified confidence level, α, the probability that, by chance, the lower limit returned will be above the true underlying value, is α/2.#'

References

[1] Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc; 1927; 22. Pg 209 to 212.
[2] Newcombe RG, Altman DG. Proportions and their differences. In Altman DG et al. (eds). Statistics with confidence (2nd edn). London: BMJ Books; 2000. Pg 46 to 48.

See Also

Other PHEindicatormethods package functions: byars_lower, byars_upper, phe_dsr, phe_isr, phe_mean, phe_proportion, phe_rate, phe_smr, wilson_upper

Examples

```
wilson_lower(65,100)
wilson_lower(65,100,99.8)
```

---

wilson_upper *wilson_upper*

Description

Calculates upper confidence limits for observed numbers of events using the Wilson Score method [1,2].

Usage

```
wilson_upper(x, n, confidence = 0.95)
```

Arguments

x	the observed numbers of cases in the samples meeting the required condition; numeric vector; no default
n	the numbers of cases in the samples; numeric vector; no default
confidence	the required level of confidence expressed as a number between 0.9 and 1 or 90 and 100; numeric; default 0.95

Value

Returns upper confidence limits for observed numbers of events using the Wilson Score method [1,2]

Notes

wilson_lower and wilson_upper together return symmetric confidence intervals, therefore for a specified confidence level, α, the probability that, by chance, the upper limit returned will be below the true underlying value, is α/2.#'

References

[1] Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc; 1927; 22. Pg 209 to 212.
[2] Newcombe RG, Altman DG. Proportions and their differences. In Altman DG et al. (eds). Statistics with confidence (2nd edn). London: BMJ Books; 2000. Pg 46 to 48.

See Also

Other PHEindicatormethods package functions: byars_lower, byars_upper, phe_dsr, phe_isr, phe_mean, phe_proportion, phe_rate, phe_smr, wilson_lower

Examples

```
wilson_upper(65,100)
wilson_upper(65,100,99.8)
```

Index