

Package ‘sampleone’

October 13, 2026

Type Package

Title Survey Sampling Methods for Teaching and Practice

Version 0.1.0

Description Planning, selection, and estimation for surveys under the main probability sampling designs taught in introductory sampling courses: simple random sampling with and without replacement, stratified sampling (proportional, Neyman, and cost-based allocation), systematic sampling, one- and two-stage cluster sampling, and ratio and regression estimators. Functions accept a superset of the inputs a given formula requires and silently ignore arguments that are not needed for the selected computation. Each function returns a structured object with the point estimate, the associated variance and confidence interval, a step-by-step breakdown of the calculation, and a short pedagogical note suggesting related methods to explore, making the package suitable both for applied survey work and for classroom use. Methods follow Cochran (1977, ISBN:0-471-16240-X) and Bolfarine & Bussab (2005, ISBN:8521203675), among others referenced in each function's documentation.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 4.1.0)

Imports stats, utils, withr

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

Config/testthat/edition 3

VignetteBuilder knitr

URL <https://github.com/mdelapi/sampleone>

BugReports <https://github.com/mdelapi/sampleone/issues>

Language en-US

Config/roxygen2/version 8.1.0

NeedsCompilation no

Author Miguel Luiz de Almeida Pinto [aut, cre],
 Mariano Martinez Espinosa [aut] (Author of the source course material
 (course notes) this package implements)

Maintainer Miguel Luiz de Almeida Pinto <phd20x@gmail.com>

Repository CRAN

Date/Publication 2026-10-13 15:00:02 UTC

Contents

aas_ajustar_perdas	3
aas_comparar_reposicao	4
aas_estima_media	4
aas_estima_proporcao	5
aas_estima_total	6
aas_tamanho_amostra	7
aas_tamanho_media	8
aas_tamanho_proporcao	9
amostragem_ajustar_deff_resposta	10
amostragem_ajustar_resposta	11
amostragem_comparar_planos	12
amostra_estimadores	13
congl_ajustar_deff	14
congl_estima_media	15
congl_estima_proporcao	16
congl_tamanho_media	17
congl_tamanho_proporcao	18
cong2_estima_media	19
cong2_estima_proporcao	20
cong2_estima_razao_media	21
cong_estima	22
estr_alocacao_custo	24
estr_alocacao_neyman	25
estr_alocacao_proporcional	26
estr_estima_media	26
estr_estima_proporcao	27
estr_tamanho_amostra	28
estr_tamanho_media	29
estr_tamanho_proporcao	30
pop_media	31
pop_proporcao	32
pop_total	32
print.sampling_result	33
razao_estima_media	34
razao_estima_R	35
razao_estima_total	36
razao_tamanho_media	37
razao_tamanho_R	38

razao_tamanho_total	39
razao_variancia_piloto	40
razao_vs_regressao_decidir	41
reg_ajustar	42
reg_estima_media	42
reg_teste_origem	43
sist_intervalo	44
sist_selecionar_amostra	45
summary.sampling_result	46
Index	47

aas_ajustar_perdas	<i>Adjust a sample size by a flat percentage buffer ("acrescimo")</i>
--------------------	---

Description

Inflates a sample size by a flat percentage buffer, following Example 3.14 of the course notes: $n_{ajustado} = n * (1 + taxa_acrescimo)$ (there illustrated as a 25% buffer added to protect against losses, refusals, or other contingencies during fieldwork). This is a distinct, simpler convention from the formal non-response adjustment of Section 4.7.2 (dividing by an expected response rate, see the course notes' Eq. 4.25 and the package's stratified-sampling module) – do not confuse the two: dividing by $(1 - taxa)$ targets a *given number of completed responses* after losses, while multiplying by $(1 + taxa)$ is a simpler flat safety margin, as used here.

Usage

```
aas_ajustar_perdas(n, taxa_acrescimo, arredondar = c("cima", "proximo"))
```

Arguments

n	Base (unadjusted) sample size.
taxa_acrescimo	Flat buffer to add, as a proportion (e.g. 0.25 for a 25% increase).
arredondar	Either "cima" (default, always round up) or "proximo" (round to the nearest integer, reproducing Example 3.14 of the course notes exactly: $357 * 1.25 = 446.25$, rounded there to 446, not 447).

Value

A `sampling_result` object with the adjusted sample size.

Examples

```
aas_ajustar_perdas(357, 0.25) # arredondar="cima" (padrao) -> 447
aas_ajustar_perdas(357, 0.25, arredondar = "proximo") # Exemplo 3.14 -> 446
```

aas_comparar_reposicao

Compare the "with replacement" and "without replacement" designs (EPA/deff)

Description

Computes the sampling-plan effect (EPA, "efeito do planejamento amostral", a.k.a. design effect) comparing AASs to AASc for the same N and n, following Eq. (3.46)-(3.47) of the course notes.

Usage

```
aas_comparar_reposicao(N, n)
```

Arguments

N	Population size.
n	Sample size.

Value

A `sampling_result` object with the EPA value and its interpretation.

References

Bolfarine, H.; Bussab, W.O. (2005). *Elementos de Amostragem*. Edgard Blucher.

Examples

```
aas_comparar_reposicao(N = 2695, n = 336)
```

aas_estima_media

Estimate the population mean under simple random sampling (AAS)

Description

Estimates the population mean from a simple random sample, with or without replacement, together with its variance and a confidence interval. Corresponds to Eqs. (3.4)-(3.7) (with replacement, AASc) and (3.25)-(3.28.1) (without replacement, AASs) of the course notes.

Usage

```
aas_estima_media(x, N = NULL, reposicao = c("sem", "com"), conf = 0.95, ...)
```

Arguments

x	Numeric vector with the sample data.
N	Optional population size. If NULL or Inf, the population is treated as unknown/very large (the finite-population correction $1 - n/N$ is dropped, i.e. $f = 0$), which also makes the "com" ("with replacement") and "sem" ("without replacement") results coincide.
reposicao	Either "sem" (without replacement, AASs – the default and the usual case in practice) or "com" (with replacement, AASc).
conf	Confidence level, default 0.95.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated mean, its variance, coefficient of variation, and confidence interval.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley. Bolfarine, H.; Bussab, W.O. (2005). *Elementos de Amostragem*. Edgard Blucher.

Examples

```
aas_estima_media(c(12, 30, 18), N = 3, reposicao = "com")
```

aas_estima_proporcao	<i>Estimate a population proportion under simple random sampling (AAS)</i>
----------------------	--

Description

Estimates a population proportion from a simple random sample, with or without replacement. Corresponds to Eqs. (3.16)-(3.21) (AASc) and (3.39)-(3.44) (AASs) of the course notes.

Usage

```
aas_estima_proporcao(
  x,
  N = NULL,
  reposicao = c("sem", "com"),
  conf = 0.95,
  ...
)
```

Arguments

x	Numeric or logical vector coded 0/1 (or FALSE/TRUE).
N	Optional population size. If NULL or Inf, the population is treated as unknown/very large (the finite-population correction $1 - n/N$ is dropped, i.e. $f = 0$), which also makes the "com" ("with replacement") and "sem" ("without replacement") results coincide.
reposicao	Either "sem" (without replacement, AASs – the default and the usual case in practice) or "com" (with replacement, AASc).
conf	Confidence level, default 0.95.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated proportion, its variance, and confidence interval.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley.

Examples

```
aas_estima_proporcao(c(1, 0, 1, 1, 0, 1, 0, 0), N = 500)
```

aas_estima_total	<i>Estimate the population total under simple random sampling (AAS)</i>
------------------	---

Description

Estimates the population total $\hat{\tau} = N\bar{x}$ from a simple random sample without replacement (Eq. 3.28.2-3.28.3 of the course notes). For the "with replacement" plan, the same expressions are used with $f = 0$.

Usage

```
aas_estima_total(x, N, reposicao = c("sem", "com"), conf = 0.95, ...)
```

Arguments

x	Numeric vector with the sample data.
N	Population size (required to scale the total).
reposicao	Either "sem" (without replacement, AASs – the default and the usual case in practice) or "com" (with replacement, AASc).
conf	Confidence level, default 0.95.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated total, its variance, and confidence interval.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 2 (extension not explicit in the course notes for AAS, but consistent with Eq. 3.28.3).

Examples

```
aas_estima_total(c(12, 30, 18), N = 300)
```

aas_tamanho_amostra	<i>Sample size for AAS: automatic formula selection (umbrella function)</i>
---------------------	---

Description

Convenience "umbrella" function that accepts every input that might be relevant to determine an AAS sample size (population variance or standard deviation, expected proportion, population size, expected losses) and automatically dispatches to [aas_tamanho_media\(\)](#) or [aas_tamanho_proporcao\(\)](#) depending on `parametro`, silently ignoring whichever inputs are not needed for the chosen calculation.

Usage

```
aas_tamanho_amostra(
  parametro = c("media", "proporcao"),
  N = NULL,
  sigma2 = NULL,
  sigma = NULL,
  p = NULL,
  d,
  conf = 0.95,
  perda = 0,
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

<code>parametro</code>	Either "media" or "proporcao": which population parameter the sample size should be computed for.
<code>N</code>	Optional population size.
<code>sigma2</code>	Population variance (used only when <code>parametro = "media"</code>).
<code>sigma</code>	Population standard deviation (alternative to <code>sigma2</code> , used only when <code>parametro = "media"</code>).

p	Expected proportion (used only when parametro = "proporcao").
d	Tolerable margin of error.
conf	Confidence level, default 0.95.
perda	Optional expected non-response/loss rate in $[0, 1)$.
arredondar	Either "cima" (default) or "proximo" – see aas_tamanho_media() .
...	Ignored. Any other variables from a broader problem statement (e.g. sample data, stratum information) may be passed here without causing an error.

Value

A `sampling_result` object, as produced by [aas_tamanho_media\(\)](#) or [aas_tamanho_proporcao\(\)](#).

Examples

```
aas_tamanho_amostra(parametro = "media", sigma = 7, d = 2)
aas_tamanho_amostra(parametro = "proporcao", p = 0.6, d = 0.03, N = 5000)
```

aas_tamanho_media	<i>Sample size to estimate a population mean (AAS)</i>
-------------------	--

Description

Computes the sample size needed to estimate a population mean with a given margin of error, choosing automatically between the infinite/unknown-population formula (Eq. 3.15) and the finite-population formula (Eq. 3.37/3.38) depending on whether N is supplied. Extra arguments not needed for the chosen formula are ignored.

Usage

```
aas_tamanho_media(
  sigma2 = NULL,
  sigma = NULL,
  d,
  conf = 0.95,
  N = NULL,
  perda = 0,
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

sigma2	Population variance (or its estimate). Alternatively supply sigma (standard deviation); if both are given, sigma2 takes precedence.
sigma	Optional population standard deviation, used only if sigma2 is not supplied.
d	Tolerable margin of error (same units as the variable).

conf	Confidence level, default 0.95.
N	Optional population size. If NULL or Inf, the infinite/unknown-population formula (3.15) is used; otherwise the finite-population formula (3.37) is used (cross-checked internally against the algebraically equivalent Eq. 3.38).
perda	Optional flat safety-margin proportion (e.g. 0.25 for a 25% buffer), applied via aas_ajustar_perdas() to inflate the raw sample size ($n * (1 + perda)$), following Example 3.14 of the course notes). Not to be confused with a formal response-rate adjustment (dividing by an expected response rate, Section 4.7.2).
arredondar	Either "cima" (the default – always round up, the conventional, conservative practice in sampling theory, e.g. Cochran 1977) or "proximo" (round to the nearest integer). The course notes round to the nearest integer using a 2-decimal critical value ($z = 1.96/1.64$); use arredondar = "proximo" together with a matching conf to reproduce the worked examples exactly.
...	Ignored (see package overview on tolerated extra arguments).

Value

A sampling_result object with the required sample size.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley.

Examples

```
aas_tamanho_media(sigma = 7, d = 2, conf = 0.95)
aas_tamanho_media(sigma2 = 1240, d = 7, conf = 0.95, N = 10000)
```

aas_tamanho_proporcao *Sample size to estimate a population proportion (AAS)*

Description

Computes the sample size needed to estimate a population proportion with a given margin of error, using the infinite/unknown-population formula (Eq. 3.23, or the conservative Eq. 3.24 when p is unknown) or the finite-population formula (Eq. 3.45) when N is supplied.

Usage

```
aas_tamanho_proporcao(
  p = NULL,
  d,
  conf = 0.95,
  N = NULL,
  perda = 0,
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

p	Expected proportion. If NULL, the conservative value $p = 0.5$ is used (maximises $p(1-p)$), following Eq. (3.24).
d	Tolerable margin of error (same units as the variable).
conf	Confidence level, default 0.95.
N	Optional population size. If NULL or Inf, the infinite/unknown-population formula (3.15) is used; otherwise the finite-population formula (3.37) is used (cross-checked internally against the algebraically equivalent Eq. 3.38).
perda	Optional flat safety-margin proportion (e.g. 0.25 for a 25% buffer), applied via <code>aas_ajustar_perdas()</code> to inflate the raw sample size ($n * (1 + perda)$), following Example 3.14 of the course notes). Not to be confused with a formal response-rate adjustment (dividing by an expected response rate, Section 4.7.2).
arredondar	Either "cima" (the default – always round up, the conventional, conservative practice in sampling theory, e.g. Cochran 1977) or "proximo" (round to the nearest integer). The course notes round to the nearest integer using a 2-decimal critical value ($z = 1.96/1.64$); use <code>arredondar = "proximo"</code> together with a matching <code>conf</code> to reproduce the worked examples exactly.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the required sample size.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley.

Examples

```
aas_tamanho_proporcao(p = 0.6, d = 0.03, conf = 0.95)
aas_tamanho_proporcao(d = 0.05, conf = 0.95) # p desconhecido -> conservador
```

`amostragem_ajustar_deff_resposta`

Adjust a sample size for design effect and expected response rate

Description

Inflates a base sample size (computed under a simple-random-sampling assumption) by a design effect (`deff`, accounting for the extra variability introduced by a more complex design such as clustering or multi-stage sampling) and by an expected response rate, following Eq. (6.6) of the course notes (Espinosa et al., 2019).

Usage

```
amostragem_ajustar_deff_resposta(n_prime, deff = 1.5, taxa_resposta = 0.85)
```

Arguments

n_prime	Base sample size (e.g. from <code>aas_tamanho_media()</code> , <code>aas_tamanho_proporcao()</code> , <code>estr_tamanho_media()</code> , or <code>cong1_tamanho_media()</code>), ignoring the design effect and non-response.
deff	Design effect (≥ 1), default 1.5 as commonly assumed in the absence of better information. <code>deff = 1</code> means no loss of efficiency relative to simple random sampling.
taxa_resposta	Expected response rate, default 0.85.

Details

Although this formula is first introduced in the course notes' cluster sampling chapter, it is a **general-purpose adjustment**: it only needs a base sample size, a design effect, and a response rate – nothing specific to clusters. Course exercises apply it to plain AAS and stratified problems just as often as to cluster problems (e.g. to scale up a prevalence-survey sample size for an anticipated design effect and non-response), so it is exposed here as a cross-cutting utility rather than tucked away under the `cong1_` prefix. `cong1_ajustar_deff()` remains available as an alias, for continuity with the cluster-sampling chapter and its documentation.

Value

A `sampling_result` object with the adjusted sample size.

References

Espinosa et al. (2019), as cited in the course notes, Eq. (6.6).

Examples

```
# Cap. 3 (AAS) + deff + resposta -- ex.: planejamento de prevalencia
n_prime <- aas_tamanho_proporcao(p = 0.20, d = 0.05)$estimativa$n
amostragem_ajustar_deff_resposta(n_prime, deff = 1.2, taxa_resposta = 0.85)
```

amostragem_ajustar_resposta

Adjust a sample size for a minimum expected response rate

Description

Inflates a sample size to guarantee a target number of completed responses given an expected response rate, following Eqs. (4.24)-(4.25) of the course notes: $n^* = n / \text{taxa_resposta}$. This targets *completed responses after losses* and is conceptually distinct from the flat safety-margin buffer of `aas_ajustar_perdas()` (which multiplies by $1 + \text{taxa}$, see Example 3.14) – do not confuse the two.

Usage

```
amostragem_ajustar_resposta(n, taxa_resposta)
```

Arguments

<code>n</code>	Base sample size computed ignoring non-response.
<code>taxa_resposta</code>	Expected response rate, in $(0, 1]$ (e.g. 0.85 for an expected 85% response rate).

Value

A `sampling_result` object with the adjusted sample size.

Examples

```
amostragem_ajustar_resposta(1525, 0.85)
```

```
amostragem_comparar_planos
```

Compare the variance of different sampling plans for the same problem

Description

Convenience function that lines up the estimated variance obtained under different sampling designs (as returned by `sampling_result` objects from this package) so they can be compared directly, computing the sampling-plan effect (EPA / design effect) of each one relative to a chosen baseline. This operationalises the comparisons illustrated throughout the course notes (Section 3.7.7, Examples 4.4-4.5, and the design-effect discussion in Section 6.1.5).

Usage

```
amostragem_comparar_planos(..., referencia = NULL)
```

Arguments

<code>...</code>	Two or more named <code>sampling_result</code> objects (e.g. <code>aas = aas_estima_media(...)</code> , <code>estratificada = estr_estima_media(...)</code>), each with a <code>variancia</code> element containing <code>var_media</code> or <code>var_p</code> .
<code>referencia</code>	Name of the plan to use as the baseline (denominator of the EPA ratio). Defaults to the first plan supplied.

Value

A `sampling_result` object with a comparison table of variances and EPA (design effect) ratios relative to `referencia`.

Examples

```
x <- rnorm(30, 50, 5)
plano_aas <- aas_estima_media(x, N = 300)
# (em um caso real, plano_estrat viria de estr_estima_media() com os
# mesmos dados organizados em estratos)
amostragem_comparar_planos(aas = plano_aas)
```

amostra_estimadores	<i>Sample-based estimators (mean, variance, proportion)</i>
---------------------	---

Description

Computes the sample mean (Eq. 2.8), the unbiased sample variance (Eq. 2.10, divisor $n-1$), and, when the data are binary, the sample proportion (Eq. 2.9) – all from a single call. This function is used internally as a building block by most other functions in the package.

Usage

```
amostra_estimadores(x, ...)
```

Arguments

x	Numeric vector with the sample data.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the requested sample estimators.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley.

Examples

```
amostra_estimadores(c(12, 30, 18))
amostra_estimadores(c(1, 0, 1, 1, 0))
```

cong1_ajustar_deff	<i>Adjust a cluster/element sample size for design effect and response rate</i>
--------------------	---

Description

Alias for [amostragem_ajustar_deff_resposta\(\)](#), kept for continuity with the cluster-sampling chapter, since that is where the course notes first introduce Eq. (6.6). The underlying formula is general-purpose (not specific to clusters) – see [amostragem_ajustar_deff_resposta\(\)](#) for the full documentation and for use with non-cluster problems (AAS, stratified, etc.).

Usage

```
cong1_ajustar_deff(n_prime, deff = 1.5, taxa_resposta = 0.85)
```

Arguments

n_prime	Base sample size (e.g. from aas_tamanho_media() , aas_tamanho_proporcao() , estr_tamanho_media() , or cong1_tamanho_media()), ignoring the design effect and non-response.
deff	Design effect (≥ 1), default 1.5 as commonly assumed in the absence of better information. $deff = 1$ means no loss of efficiency relative to simple random sampling.
taxa_resposta	Expected response rate, default 0.85.

Value

A `sampling_result` object with the adjusted sample size.

References

Espinosa et al. (2019), as cited in the course notes, Eq. (6.6).

Examples

```
cong1_ajustar_deff(n_prime = 518, deff = 1.5, taxa_resposta = 0.85)
```

cong1_estima_media *Estimate the population mean under one-stage cluster sampling*

Description

Estimates the population mean from a sample of whole clusters, all of whose elements were observed (one-stage cluster sampling), following Eqs. (6.1)–(6.4) of the course notes.

Usage

```
cong1_estima_media(yi, Mi, Nc, M = NULL, conf = 0.95, ...)
```

Arguments

yi	Numeric vector with the cluster totals (y_i) for each sampled cluster.
Mi	Numeric vector with the number of elements in each sampled cluster.
Nc	Total number of clusters in the population.
M	Optional total number of elements in the population (used to compute the population average cluster size $M_{\text{barra}} = M/Nc$, as required by Eq. 6.2). When M is unknown, M_{barra} falls back to the sample average of M_i , which is only an approximation and is flagged as such in the result.
conf	Confidence level, default 0.95.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated mean, variance, and confidence interval.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 9.

Examples

```
# Exemplo 6.1 (turmas da UFMT)
cong1_estima_media(
  yi = c(1209, 921, 1116, 1433, 1037, 2112, 1737, 2206, 2112, 1870),
  Mi = c(20, 16, 17, 21, 16, 25, 21, 26, 25, 23),
  Nc = 373, M = 8018
)
```

 congl_estima_proporcao

Estimate the population proportion under one-stage cluster sampling

Description

Analogous to `congl_estima_media()` but for a population proportion, following Eqs. (6.7)-(6.10) of the course notes.

Usage

```
congl_estima_proporcao(yi, Mi, Nc, M = NULL, conf = 0.95, ...)
```

Arguments

<code>yi</code>	Numeric vector with the number of "successes" (elements with the characteristic of interest) in each sampled cluster.
<code>Mi</code>	Numeric vector with the number of elements in each sampled cluster.
<code>Nc</code>	Total number of clusters in the population.
<code>M</code>	Optional total number of elements in the population (see <code>congl_estima_media()</code> for details on $M_{\text{barra}} = M/Nc$).
<code>conf</code>	Confidence level, default 0.95.
<code>...</code>	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated proportion, variance, and confidence interval.

Examples

```
# Exemplo 6.3 (proporcao de alunos com peso acima da mediana)
congl_estima_proporcao(
  yi = c(7, 6, 5, 6, 5, 7, 5, 6, 6, 6),
  Mi = c(20, 16, 17, 21, 16, 25, 21, 26, 25, 23),
  Nc = 373, M = 8018
)
```

cong1_tamanho_media	<i>Sample size (number of clusters) to estimate a mean (one-stage cluster sampling)</i>
---------------------	---

Description

Computes the number of clusters needed to estimate a population mean with a given margin of error under one-stage cluster sampling (Eq. 6.5 of the course notes).

Usage

```
cong1_tamanho_media(  
  Nc,  
  sigmac2,  
  M_barra,  
  d,  
  conf = 0.95,  
  arredondar = c("cima", "proximo"),  
  ...  
)
```

Arguments

Nc	Total number of clusters in the population.
sigmac2	Between-cluster population variance (or its estimate, from a pilot study).
M_barra	Average number of elements per cluster.
d	Tolerable margin of error (in the original units of the variable, not per-cluster).
conf	Confidence level, default 0.95.
arredondar	Either "cima" (default) or "proximo" – see aas_tamanho_media() . Example 6.1(c) of the course notes uses "proximo" (41.87 rounds to 42, same result either way here).
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the required number of clusters.

Examples

```
cong1_tamanho_media(Nc = 373, sigmac2 = 51054.3854, M_barra = 8018/373, d = 3)
```

cong1_tamanho_proporcao

Sample size (number of clusters) to estimate a proportion (one-stage cluster sampling)

Description

Analogous to [cong1_tamanho_media\(\)](#) but for a population proportion (Eq. 6.11 of the course notes).

Usage

```
cong1_tamanho_proporcao(
  Nc,
  sigmacp2,
  M_barra,
  d,
  conf = 0.95,
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

Nc	Total number of clusters in the population.
sigmacp2	Between-cluster population variance for the proportion (or its estimate).
M_barra	Average number of elements per cluster.
d	Tolerable margin of error (in the original units of the variable, not per-cluster).
conf	Confidence level, default 0.95.
arredondar	Either "cima" (default) or "proximo" – see aas_tamanho_media() . Example 6.1(c) of the course notes uses "proximo" (41.87 rounds to 42, same result either way here).
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the required number of clusters.

Examples

```
cong1_tamanho_proporcao(Nc = 373, sigmacp2 = 0.91785840, M_barra = 8018/373, d = 0.013)
```

cong2_estima_media *Estimate the population mean under two-stage cluster sampling*

Description

Estimates the population mean from a two-stage cluster sample (a sample of clusters – UPA – followed by a sample of elements within each sampled cluster – USA), when the total population size M is known, following Eqs. (6.12)-(6.18) of the course notes.

Usage

```
cong2_estima_media(
  dados = NULL,
  Mi = NULL,
  mi = NULL,
  yi_bar = NULL,
  si2 = NULL,
  Nc,
  M,
  conf = 0.95,
  ...
)
```

Arguments

<code>dados</code>	A <code>data.frame</code> (or list of data frames / a list with one element per cluster) with, for every sampled element, at least the columns <code>conglomerado</code> (cluster id), <code>Mi</code> (cluster population size), and <code>y</code> (the measured value for that element). Alternatively, supply <code>Mi</code> , <code>mi</code> , <code>yi_bar</code> , and <code>si2</code> directly (pre-summarised per cluster) – see Details.
<code>Mi</code>	Optional numeric vector with the population size of each sampled cluster (used instead of <code>dados</code> when the data are already summarised per cluster).
<code>mi</code>	Optional numeric vector with the number of sampled elements in each cluster.
<code>yi_bar</code>	Optional numeric vector with the sample mean of each cluster.
<code>si2</code>	Optional numeric vector with the sample variance within each cluster (Eq. 6.15).
<code>Nc</code>	Total number of clusters in the population.
<code>M</code>	Total number of elements in the population.
<code>conf</code>	Confidence level, default 0.95.
<code>...</code>	Ignored (see package overview on tolerated extra arguments).

Details

Two calling styles are supported: (1) pass raw per-element data via `dados` (a `data.frame` with one row per sampled element), or (2) pass already-summarised per-cluster vectors `Mi`, `mi`, `yi_bar`, `si2`

directly. The second style is convenient when a problem statement (as in a textbook exercise) already gives per-cluster summaries instead of raw data – any of `dados/Mi/mi/yi_bar/si2` not needed for the chosen style are simply ignored.

Value

A `sampling_result` object with the estimated mean, its two-part variance decomposition, and confidence interval.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 10.

Examples

```
Mi <- c(50, 65, 45, 48, 52, 58, 42, 66, 40, 56)
mi <- c(10, 13, 9, 10, 10, 12, 8, 13, 8, 11)
yi_bar <- c(5.4, 4.0, 5.6667, 4.8, 4.3, 3.8333, 5.0, 3.8462, 4.875, 5.0)
si2 <- c(11.3778, 10.6667, 16.75, 13.2889, 11.1222, 14.8788, 5.1429,
        4.3077, 6.125, 11.8)
cong2_estima_media(Mi = Mi, mi = mi, yi_bar = yi_bar, si2 = si2,
                  Nc = 90, M = 4500)
```

`cong2_estima_proporcao`

Estimate the population proportion under two-stage cluster sampling

Description

Analogous to `cong2_estima_media()` but for a population proportion, following Eqs. (6.24)-(6.29) of the course notes. Note that, unlike `cong2_estima_media()`, the proportion estimator is inherently a ratio-type estimator (structurally analogous to `cong2_estima_razao_media()`) and therefore always uses the *sample* average cluster size $m_{\text{barra}} = \text{mean}(Mi)$ in its variance formula, even when the population totals Nc/M are known – this is exactly how Example 6.6 of the course notes computes it (using $m_{\text{barra}} = 52.2$ rather than $M/Nc = 50$, even though both M and Nc are known in that example).

Usage

```
cong2_estima_proporcao(Mi, mi, pi_hat, Nc, M = NULL, conf = 0.95, ...)
```

Arguments

<code>Mi</code>	Numeric vector with the population size of each sampled cluster.
<code>mi</code>	Numeric vector with the number of sampled elements in each cluster.
<code>pi_hat</code>	Numeric vector with the sample proportion of each cluster (Eq. 6.24).
<code>Nc</code>	Total number of clusters in the population.

M	Ignored (kept for signature symmetry with <code>cong2_estima_media()</code> and so that extra arguments from a broader problem statement can be passed without error) – see Details.
conf	Confidence level, default 0.95.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated proportion, its variance, and confidence interval.

Examples

```
Mi <- c(50, 65, 45, 48, 52, 58, 42, 66, 40, 56)
mi <- c(10, 13, 9, 10, 10, 12, 8, 13, 8, 11)
pi_hat <- c(0.4, 0.3077, 0.3333, 0.3, 0.3, 0.25, 0.375, 0.3077, 0.375, 0.3636)
cong2_estima_proporcao(Mi = Mi, mi = mi, pi_hat = pi_hat, Nc = 90)
```

cong2_estima_razao_media

Estimate the population mean via ratio estimator (two-stage cluster sampling, M unknown)

Description

Analogous to `cong2_estima_media()` but for the case where the total population size M is unknown, using the ratio estimator of Eqs. (6.19)-(6.23) of the course notes, with m_{barra} (the sample average cluster size) substituted for $M_{\text{barra}} = M/Nc$.

Usage

```
cong2_estima_razao_media(
  dados = NULL,
  Mi = NULL,
  mi = NULL,
  yi_bar = NULL,
  si2 = NULL,
  Nc,
  M = NULL,
  conf = 0.95,
  ...
)
```

Arguments

<code>dados</code>	A <code>data.frame</code> (or list of data frames / a list with one element per cluster) with, for every sampled element, at least the columns <code>conglomerado</code> (cluster id), <code>Mi</code> (cluster population size), and <code>y</code> (the measured value for that element). Alternatively, supply <code>Mi</code> , <code>mi</code> , <code>yi_bar</code> , and <code>si2</code> directly (pre-summarised per cluster) – see Details.
<code>Mi</code>	Optional numeric vector with the population size of each sampled cluster (used instead of <code>dados</code> when the data are already summarised per cluster).
<code>mi</code>	Optional numeric vector with the number of sampled elements in each cluster.
<code>yi_bar</code>	Optional numeric vector with the sample mean of each cluster.
<code>si2</code>	Optional numeric vector with the sample variance within each cluster (Eq. 6.15).
<code>Nc</code>	Total number of clusters in the population.
<code>M</code>	Ignored for this estimator (kept for signature symmetry with <code>cong2_estima_media()</code> so extra arguments from a broader problem can be passed without error); <code>m_barra</code> is estimated from <code>Mi</code> instead.
<code>conf</code>	Confidence level, default 0.95.
<code>...</code>	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the ratio-estimated mean, its variance, and confidence interval.

Examples

```
Mi <- c(50, 65, 45, 48, 52, 58, 42, 66, 40, 56)
mi <- c(10, 13, 9, 10, 10, 12, 8, 13, 8, 11)
yi_bar <- c(5.4, 4.0, 5.6667, 4.8, 4.3, 3.8333, 5.0, 3.8462, 4.875, 5.0)
si2 <- c(11.3778, 10.6667, 16.75, 13.2889, 11.1222, 14.8788, 5.1429,
        4.3077, 6.125, 11.8)
cong2_estima_razao_media(Mi = Mi, mi = mi, yi_bar = yi_bar, si2 = si2, Nc = 90)
```

cong_estima

Umbrella estimator for two-stage cluster sampling

Description

Convenience "umbrella" function analogous to `aas_tamanho_amostra()`: dispatches automatically to `cong2_estima_media()` (when `M` is supplied) or `cong2_estima_razao_media()` (when `M` is `NULL`), ignoring whichever inputs are not needed.

Usage

```
cong_estima(
  dados = NULL,
  Mi = NULL,
  mi = NULL,
  yi_bar = NULL,
  si2 = NULL,
  Nc,
  M = NULL,
  conf = 0.95,
  ...
)
```

Arguments

<code>dados</code>	A data.frame (or list of data frames / a list with one element per cluster) with, for every sampled element, at least the columns conglomerado (cluster id), Mi (cluster population size), and y (the measured value for that element). Alternatively, supply Mi, mi, yi_bar, and si2 directly (pre-summarised per cluster) – see Details.
<code>Mi</code>	Optional numeric vector with the population size of each sampled cluster (used instead of dados when the data are already summarised per cluster).
<code>mi</code>	Optional numeric vector with the number of sampled elements in each cluster.
<code>yi_bar</code>	Optional numeric vector with the sample mean of each cluster.
<code>si2</code>	Optional numeric vector with the sample variance within each cluster (Eq. 6.15).
<code>Nc</code>	Total number of clusters in the population.
<code>M</code>	Total number of elements in the population.
<code>conf</code>	Confidence level, default 0.95.
<code>...</code>	Ignored. Any other variables from a broader problem statement may be passed here without causing an error.

Value

A `sampling_result` object, as produced by `cong2_estima_media()` or `cong2_estima_razao_media()`.

Examples

```
Mi <- c(50, 65, 45, 48, 52, 58, 42, 66, 40, 56)
mi <- c(10, 13, 9, 10, 10, 12, 8, 13, 8, 11)
yi_bar <- c(5.4, 4.0, 5.6667, 4.8, 4.3, 3.8333, 5.0, 3.8462, 4.875, 5.0)
si2 <- c(11.3778, 10.6667, 16.75, 13.2889, 11.1222, 14.8788, 5.1429,
        4.3077, 6.125, 11.8)
cong_estima(Mi = Mi, mi = mi, yi_bar = yi_bar, si2 = si2, Nc = 90) # M ausente -> razao
```

estr_alocacao_custo *Cost-based optimum allocation of a stratified sample*

Description

Allocates strata sample sizes to minimise variance subject to a fixed total budget, when per-unit sampling costs differ across strata (Eqs. 4.7-4.8 for means, Eq. 4.13 for proportions).

Usage

```
estr_alocacao_custo(n, Nk, sigmak = NULL, pk = NULL, ck, ...)
```

Arguments

n	Total sample size to allocate.
Nk	Numeric vector with the population size of each stratum.
sigmak	Numeric vector with the population standard deviation of each stratum (for a mean problem). Ignored if pk is supplied.
pk	Numeric vector with the population proportion of each stratum (for a proportion problem); if supplied, sigmak is derived as $\sqrt{pk*(1-pk)}$ and any sigmak argument is ignored.
ck	Numeric vector with the sampling cost per unit in each stratum.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the per-stratum sample sizes under cost-based optimum allocation.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 5.

Examples

```
estr_alocacao_custo(n = 491, Nk = c(4000, 3000, 2000), sigmak = c(3, 3, 2),
  ck = c(10, 10, 20))
```

estr_alocacao_neyman *Neyman (optimum) allocation of a stratified sample*

Description

Allocates a total sample size across strata to minimise variance for a fixed total sample size, assuming equal per-unit cost across strata (Eqs. 4.18-4.20 of the course notes). Works for both mean (σ_k known) and proportion (p_k known, substituting $\sigma_k = \sqrt{p_k(1-p_k)}$) problems.

Usage

```
estr_alocacao_neyman(n, Nk, sigma_k = NULL, pk = NULL, ...)
```

Arguments

n	Total sample size to allocate.
Nk	Numeric vector with the population size of each stratum.
sigma_k	Numeric vector with the population standard deviation of each stratum (for a mean problem). Ignored if pk is supplied.
pk	Numeric vector with the population proportion of each stratum (for a proportion problem); if supplied, sigma_k is derived as $\sqrt{p_k(1-p_k)}$ and any sigma_k argument is ignored.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the per-stratum sample sizes under Neyman allocation.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 5.

Examples

```
estr_alocacao_neyman(n = 261, Nk = c(1540, 770, 385), sigma_k = c(2.6, 3, 3.3))
```

estr_alocacao_proporcional

Proportional allocation of a stratified sample

Description

Allocates a total sample size across strata proportionally to stratum size, $n_k = n * W_k$ (Eq. 4.6, or Eq. 4.14-4.15 for proportions).

Usage

```
estr_alocacao_proporcional(n, Nk, ...)
```

Arguments

n	Total sample size to allocate.
Nk	Numeric vector with the population size of each stratum.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the per-stratum sample sizes.

Examples

```
estr_alocacao_proporcional(n = 120, Nk = c(1540, 770, 385))
```

estr_estima_media

Estimate the population mean under stratified sampling

Description

Estimates the population mean from data collected under stratified random sampling, given per-stratum sample means and sizes (Eqs. 4.1-4.4 of the course notes).

Usage

```
estr_estima_media(Nk, xk_bar, sk2, nk, conf = 0.95, ...)
```

Arguments

Nk	Numeric vector with the population size of each stratum.
xk_bar	Numeric vector with the sample mean of each stratum.
sk2	Numeric vector with the sample variance of each stratum.
nk	Numeric vector with the sample size of each stratum.
conf	Confidence level, default 0.95.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated stratified mean, its variance, coefficient of variation, and confidence interval.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 5.

Examples

```
estr_estima_media(Nk = c(1540, 770, 385), xk_bar = c(120, 118, 125),
                 sk2 = c(7, 9, 11), nk = c(69, 34, 17))
```

`estr_estima_proporcao` *Estimate the population proportion under stratified sampling*

Description

Estimates a population proportion from data collected under stratified random sampling, given per-stratum sample proportions and sizes (Eqs. 4.9-4.11 of the course notes).

Usage

```
estr_estima_proporcao(Nk, pk, nk, conf = 0.95, ...)
```

Arguments

<code>Nk</code>	Numeric vector with the population size of each stratum.
<code>pk</code>	Numeric vector with the sample proportion of each stratum.
<code>nk</code>	Numeric vector with the sample size of each stratum.
<code>conf</code>	Confidence level, default 0.95.
<code>...</code>	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated stratified proportion, its variance, and confidence interval.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 5.

Examples

```
estr_estima_proporcao(Nk = c(1540, 770, 385), pk = c(0.75, 0.50, 0.25),
                     nk = c(161, 81, 40))
```

estr_tamanho_amostra *Sample size for stratified sampling: automatic formula selection (umbrella)*

Description

Convenience "umbrella" function analogous to [aas_tamanho_amostra\(\)](#): accepts every input potentially relevant to a stratified sample-size problem and dispatches to [estr_tamanho_media\(\)](#) or [estr_tamanho_proporcao\(\)](#) depending on parametro, ignoring whichever inputs are not needed.

Usage

```
estr_tamanho_amostra(
  parametro = c("media", "proporcao"),
  Nk,
  sigmak2 = NULL,
  pk = NULL,
  d,
  conf = 0.95,
  ck = NULL,
  alocacao = c("proporcional", "neyman", "custo", "nenhuma"),
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

parametro	Either "media" or "proporcao".
Nk	Numeric vector with the population size of each stratum.
sigmak2	Numeric vector with the population variance of each stratum.
pk	Numeric vector with the population proportion of each stratum (used only when parametro = "proporcao").
d	Tolerable margin of error.
conf	Confidence level, default 0.95.
ck	Optional numeric vector with the sampling cost per unit in each stratum. If supplied, the cost-based formula (Eq. 4.8) is used; otherwise the standard formula (Eq. 4.5) is used.
alocacao	How to allocate n across strata once computed: "proporcional" (default), "neyman", "custo" (requires ck), or "nenhuma" (only the total n is returned).
arredondar	Either "cima" (default, always round up – the conventional, conservative practice) or "proximo" (round to the nearest integer, reproducing the course notes' worked examples, which use a 2-decimal critical value and round to the nearest integer).
...	Ignored. Any other variables from a broader problem statement may be passed here without causing an error.

Value

A `sampling_result` object, as produced by `estr_tamanho_media()` or `estr_tamanho_proporcao()`.

Examples

```
estr_tamanho_amostra(parametro = "media", Nk = c(1540, 770, 385),
  sigmak2 = c(7, 9, 11), d = 0.5)
```

<code>estr_tamanho_media</code>	<i>Sample size for stratified sampling: mean (proportional / Neyman / cost)</i>
---------------------------------	---

Description

Computes the total sample size and, when `alocacao` is not `"nenhuma"`, the per-stratum allocation, to estimate a population mean under stratified sampling, dispatching automatically between the proportional/Neyman formula (Eq. 4.5) and the cost-based formula (Eq. 4.8) depending on whether `ck` is supplied. Extra arguments not needed for the chosen formula are ignored.

Usage

```
estr_tamanho_media(
  Nk,
  sigmak2,
  d,
  conf = 0.95,
  ck = NULL,
  alocacao = c("proporcional", "neyman", "custo", "nenhuma"),
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

<code>Nk</code>	Numeric vector with the population size of each stratum.
<code>sigmak2</code>	Numeric vector with the population variance of each stratum.
<code>d</code>	Tolerable margin of error.
<code>conf</code>	Confidence level, default 0.95.
<code>ck</code>	Optional numeric vector with the sampling cost per unit in each stratum. If supplied, the cost-based formula (Eq. 4.8) is used; otherwise the standard formula (Eq. 4.5) is used.
<code>alocacao</code>	How to allocate <code>n</code> across strata once computed: <code>"proporcional"</code> (default), <code>"neyman"</code> , <code>"custo"</code> (requires <code>ck</code>), or <code>"nenhuma"</code> (only the total <code>n</code> is returned).

arredondar	Either "cima" (default, always round up – the conventional, conservative practice) or "proximo" (round to the nearest integer, reproducing the course notes' worked examples, which use a 2-decimal critical value and round to the nearest integer).
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the total sample size and, unless `alocacao = "nenhuma"`, the per-stratum allocation.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 5.

Examples

```
estr_tamanho_media(Nk = c(1540, 770, 385), sigmak2 = c(7, 9, 11), d = 0.5, conf = 0.95)
```

`estr_tamanho_proporcao`

Sample size for stratified sampling: proportion (proportional / Neyman / cost)

Description

Analogous to `estr_tamanho_media()` but for a population proportion, using Eq. (4.12), with per-stratum proportions substituting for variances ($\text{sigmak2} = \text{pk} \times (1 - \text{pk})$).

Usage

```
estr_tamanho_proporcao(
  Nk,
  pk,
  d,
  conf = 0.95,
  ck = NULL,
  alocacao = c("proporcional", "neyman", "custo", "nenhuma"),
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

Nk	Numeric vector with the population size of each stratum.
pk	Numeric vector with the population proportion of each stratum.
d	Tolerable margin of error.
conf	Confidence level, default 0.95.
ck	Optional numeric vector with the sampling cost per unit in each stratum (Eq. 4.13).
alocacao	How to allocate n across strata: "proporcional" (default), "neyman", "custo" (requires ck), or "nenhuma".
arredondar	Either "cima" (default) or "proximo" – see estr_tamanho_media() .
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the total sample size and, unless `alocacao = "nenhuma"`, the per-stratum allocation.

Examples

```
estr_tamanho_proporcao(Nk = c(1540, 770, 385), pk = c(0.75, 0.50, 0.25), d = 0.05)
```

pop_media	<i>Population mean</i>
-----------	------------------------

Description

Computes the population mean $\mu = \sum X_i/N$ (Eq. 2.3) and, when requested, the population variance (Eq. 2.5).

Usage

```
pop_media(x, ...)
```

Arguments

x	Numeric vector with the full population values.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the population mean and variance.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley.

Examples

```
pop_media(c(44, 45, 48, 42, 46))
```

pop_proporcao	<i>Population proportion</i>
---------------	------------------------------

Description

Computes the population proportion $P = \sum X_i/N$ for a binary (0/1) characteristic (Eq. 2.4) and its variance $P(1 - P)$ (Eq. 2.5.1).

Usage

```
pop_proporcao(x, ...)
```

Arguments

- x Numeric or logical vector coded as 0/1 (or FALSE/TRUE).
- ... Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the population proportion.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley.

Examples

```
pop_proporcao(c(1, 0, 0, 1, 1))
```

pop_total	<i>Population total</i>
-----------	-------------------------

Description

Computes the population total $X_T = \sum_{i=1}^N X_i$ (course notes, Eq. 2.2).

Usage

```
pop_total(x, ...)
```

Arguments

- | | |
|-----|--|
| x | Numeric vector with the full population values. |
| ... | Ignored. Present so that extra, unrelated arguments can be passed without causing an error (see package overview). |

Value

A `sampling_result` object with the population total.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley.

Examples

```
pop_total(c(1, 2, 3, 4, 5))
```

`print.sampling_result` *Print method for sampling_result objects*

Description

Print method for `sampling_result` objects

Usage

```
## S3 method for class 'sampling_result'  
print(x, ...)
```

Arguments

- | | |
|-----|--|
| x | A <code>sampling_result</code> object. |
| ... | Further arguments (unused, kept for S3 consistency). |

Value

x, invisibly.

razao_estima_media	<i>Ratio estimator of the population mean</i>
--------------------	---

Description

Estimates the population mean using the ratio estimator (Eq. 7.36-7.38 of the course notes). Unlike `razao_estima_R()`, this does not require knowing X (the population total of the auxiliary variable) – only N .

Usage

```
razao_estima_media(
  x = NULL,
  y = NULL,
  n = NULL,
  sum_x = NULL,
  sum_y = NULL,
  sum_x2 = NULL,
  sum_y2 = NULL,
  sum_xy = NULL,
  N,
  X = NULL,
  conf = 0.95,
  ...
)
```

Arguments

<code>x, y</code>	Optional raw numeric vectors (sample data).
<code>n, sum_x, sum_y, sum_x2, sum_y2, sum_xy</code>	Optional pre-aggregated summary statistics, used when raw data are not available (e.g. a problem statement that only gives totals).
<code>N</code>	Population size.
<code>X</code>	Optional population total of the auxiliary variable; not needed to compute the mean itself (only N is required), but accepted here so that a full problem statement (which usually supplies X) can be passed without error. Ignored if <code>NULL</code> .
<code>conf</code>	Confidence level, default 0.95.
<code>...</code>	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated mean, its variance, and confidence interval.

Examples

```
x <- c(1, 30, 44, 20, 0, 10, 15, 5, 2, 50, 35, 25)
y <- c(2, 35, 50, 27, 1, 15, 17, 7, 0, 53, 35, 30)
razao_estima_media(x = x, y = y, N = 1500)
```

razao_estima_R

Ratio estimator of the population ratio R

Description

Estimates the population ratio $R = Y/X$ (Eq. 7.27-7.30 of the course notes) and its variance (Eq. 7.31), from either raw sample data (x, y) or pre-aggregated sums – see [.razao_resumo\(\)](#).

Usage

```
razao_estima_R(
  x = NULL,
  y = NULL,
  n = NULL,
  sum_x = NULL,
  sum_y = NULL,
  sum_x2 = NULL,
  sum_y2 = NULL,
  sum_xy = NULL,
  N,
  X,
  conf = 0.95,
  ...
)
```

Arguments

x, y	Optional raw numeric vectors (sample data).
$n, \text{sum}_x, \text{sum}_y, \text{sum}_{x2}, \text{sum}_{y2}, \text{sum}_{xy}$	Optional pre-aggregated summary statistics, used when raw data are not available (e.g. a problem statement that only gives totals).
N	Population size.
X	Population total of the auxiliary variable (used to derive $\bar{X} = X/N$, required by Eq. 7.31).
conf	Confidence level, default 0.95.
$\dots$	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated ratio, its variance, and confidence interval.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 6.

Examples

```
x <- c(1, 30, 44, 20, 0, 10, 15, 5, 2, 50, 35, 25)
y <- c(2, 35, 50, 27, 1, 15, 17, 7, 0, 53, 35, 30)
razao_estima_R(x = x, y = y, N = 1500, X = 15000)
```

razao_estima_total	<i>Ratio estimator of the population total</i>
--------------------	--

Description

Estimates the population total using the ratio estimator (Eq. 7.33-7.35 of the course notes).

Usage

```
razao_estima_total(
  x = NULL,
  y = NULL,
  n = NULL,
  sum_x = NULL,
  sum_y = NULL,
  sum_x2 = NULL,
  sum_y2 = NULL,
  sum_xy = NULL,
  N,
  X,
  conf = 0.95,
  ...
)
```

Arguments

x, y	Optional raw numeric vectors (sample data).
n, sum_x, sum_y, sum_x2, sum_y2, sum_xy	Optional pre-aggregated summary statistics, used when raw data are not available (e.g. a problem statement that only gives totals).
N	Population size.
X	Population total of the auxiliary variable (used to derive $\bar{X} = X/N$, required by Eq. 7.31).
conf	Confidence level, default 0.95.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the estimated total, its variance, and confidence interval.

Examples

```
x <- c(1, 30, 44, 20, 0, 10, 15, 5, 2, 50, 35, 25)
y <- c(2, 35, 50, 27, 1, 15, 17, 7, 0, 53, 35, 30)
razao_estima_total(x = x, y = y, N = 1500, X = 15000)
```

razao_tamanho_media	<i>Sample size to estimate the population mean via the ratio estimator</i>
---------------------	--

Description

Analogous to `razao_tamanho_R()` but for the population mean (Eq. 7.44 of the course notes). Unlike `razao_tamanho_R()`, this does not require $\bar{X}$.

Usage

```
razao_tamanho_media(
  sr2,
  N,
  d,
  conf = 0.95,
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

<code>sr2</code>	Variance of the ratio estimator (population value, or a pilot-sample estimate from <code>razao_variancia_piloto()</code>).
<code>N</code>	Population size.
<code>d</code>	Tolerable margin of error.
<code>conf</code>	Confidence level, default 0.95.
<code>arredondar</code>	Either "cima" (default) or "proximo".
<code>...</code>	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the required sample size.

Examples

```
razao_tamanho_media(sr2 = 2.2462, N = 1000, d = 0.75)
```

razao_tamanho_R	<i>Sample size to estimate the population ratio R</i>
-----------------	---

Description

Computes the sample size needed to estimate the population ratio R with a given margin of error (Eq. 7.41 of the course notes).

Usage

```
razao_tamanho_R(
  sr2,
  N,
  X_barra,
  d,
  conf = 0.95,
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

sr2	Variance of the ratio estimator (population value, or a pilot-sample estimate from razao_variancia_piloto()).
N	Population size.
X_barra	Population mean of the auxiliary variable (X/N).
d	Tolerable margin of error.
conf	Confidence level, default 0.95.
arredondar	Either "cima" (default) or "proximo" – see aas_tamanho_media() .
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the required sample size.

Examples

```
razao_tamanho_R(sr2 = 7.7853, N = 1500, X_barra = 10, d = 0.05)
```

razao_tamanho_total	<i>Sample size to estimate the population total via the ratio estimator</i>
---------------------	---

Description

Analogous to [razao_tamanho_R\(\)](#) but for the population total (Eq. 7.43 of the course notes).

Usage

```
razao_tamanho_total(
  sr2,
  N,
  d,
  conf = 0.95,
  arredondar = c("cima", "proximo"),
  ...
)
```

Arguments

sr2	Variance of the ratio estimator (population value, or a pilot-sample estimate from razao_variancia_piloto()).
N	Population size.
d	Tolerable margin of error.
conf	Confidence level, default 0.95.
arredondar	Either "cima" (default) or "proximo" – see aas_tamanho_media() .
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the required sample size.

Examples

```
razao_tamanho_total(sr2 = 7.7853, N = 1500, d = 500)
```

razao_variancia_piloto

Pilot-sample variance estimate for the ratio estimator

Description

Computes s_r^2 , the pilot-sample estimate of the population variance of the ratio estimator (Eq. 7.42 of the course notes), used as an input to [razao_tamanho_R\(\)](#), [razao_tamanho_total\(\)](#), and [razao_tamanho_media\(\)](#) when no prior estimate of the variance is available.

Usage

```
razao_variancia_piloto(
  x = NULL,
  y = NULL,
  n = NULL,
  sum_x = NULL,
  sum_y = NULL,
  sum_x2 = NULL,
  sum_y2 = NULL,
  sum_xy = NULL,
  N = NULL,
  X = NULL,
  ...
)
```

Arguments

x, y	Optional raw numeric vectors (sample data).
n, sum_x, sum_y, sum_x2, sum_y2, sum_xy	Optional pre-aggregated summary statistics, used when raw data are not available (e.g. a problem statement that only gives totals).
N, X	Ignored (kept for signature symmetry with razao_estima_R() so a full problem statement can be passed without error) – s_r^2 itself does not require them.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with s_r^2 .

Examples

```
x <- c(1, 30, 44, 20, 0, 10, 15, 5, 2, 50, 35, 25)
y <- c(2, 35, 50, 27, 1, 15, 17, 7, 0, 53, 35, 30)
razao_variancia_piloto(x = x, y = y)
```

razao_vs_regressao_decidir

Ratio vs. regression: full side-by-side comparison (umbrella function)

Description

Convenience "umbrella" function that fits the regression model, tests whether the line passes through the origin, and computes *both* the ratio and the regression estimators of the population mean side by side, clearly flagging which one the origin test recommends – following the full workflow of Examples 7.1-7.2 of the course notes. This lets students and instructors compare both estimators even when only one is formally "correct" for the data at hand.

Usage

```
razao_vs_regressao_decidir(x, y, X_barra, N, conf = 0.95, ...)
```

Arguments

x	Numeric vector with the auxiliary variable.
y	Numeric vector with the variable of interest.
X_barra	Known population mean of the auxiliary variable (required for the regression estimator; not required for the ratio estimator, whose mean form only needs N).
N	Population size.
conf	Confidence level, default 0.95.
...	Ignored. Any other variables from a broader problem statement may be passed here without causing an error.

Details

The function itself does not write anything to the console; print the returned object (or just type its name) to see a side-by-side summary.

Value

An object of class `razao_vs_regressao`: a named list with elements `teste` (from `reg_teste_origem()`), `razao` (from `razao_estima_media()`), and `regressao` (from `reg_estima_media()`), plus a `recomendado` character scalar. It has a `print()` method.

Examples

```
x <- c(1, 30, 44, 20, 0, 10, 15, 5, 2, 50, 35, 25)
y <- c(2, 35, 50, 27, 1, 15, 17, 7, 0, 53, 35, 30)
res <- razao_vs_regressao_decidir(x, y, X_barra = 10, N = 1500)
res$recomendado
print(res)
```

reg_ajustar	<i>Fit a simple linear regression (ratio/regression estimator groundwork)</i>
-------------	---

Description

Fits the simple linear regression of y on x by least squares (Eqs. 7.9-7.10 of the course notes), and computes the residual variance, the standard error of the intercept, and the t-statistic used to test whether the line passes through the origin (Eqs. 7.18-7.20) – the basis for deciding between the ratio and the regression estimator (see [reg_teste_origem\(\)](#)).

Usage

```
reg_ajustar(x, y, ...)
```

Arguments

x	Numeric vector with the auxiliary variable.
y	Numeric vector with the variable of interest.
$\dots$	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the fitted coefficients, residual variance, and the intercept's standard error and t-statistic.

References

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. Wiley, cap. 7.

Examples

```
x <- c(39, 43, 21, 64, 57, 47, 28, 75, 34, 52)
y <- c(65, 78, 52, 82, 92, 89, 73, 98, 56, 75)
reg_ajustar(x, y)
```

reg_estima_media	<i>Regression estimator of the population mean</i>
------------------	--

Description

Estimates the population mean using the regression estimator (Eq. 7.22-7.25 of the course notes), which uses a known population mean $\bar{X}$ of an auxiliary variable correlated with y .

Usage

```
reg_estima_media(x, y, X_barra, N = NULL, conf = 0.95, ...)
```

Arguments

x	Numeric vector with the auxiliary variable.
y	Numeric vector with the variable of interest.
X_barra	Known population mean of the auxiliary variable.
N	Optional population size, used for the finite-population correction $f = n/N$. If NULL, $f = 0$ is assumed.
conf	Confidence level, default 0.95.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the regression-estimated mean, its variance, and confidence interval.

Examples

```
x <- c(39, 43, 21, 64, 57, 47, 28, 75, 34, 52)
y <- c(65, 78, 52, 82, 92, 89, 73, 98, 56, 75)
reg_estima_media(x, y, X_barra = 52, N = 486)
```

reg_teste_origem

Decide between the ratio and regression estimators

Description

Tests whether the regression line of y on x passes through the origin ($H_0: a_0 = 0$), using both the t -statistic (Eq. 7.20) and the confidence interval for a_0 (Eq. 7.21), and recommends the ratio estimator if the line passes through the origin (fails to reject H_0) or the regression estimator otherwise – following the logic of Examples 7.1-7.2 of the course notes.

Usage

```
reg_teste_origem(x, y, conf = 0.95, ...)
```

Arguments

x	Numeric vector with the auxiliary variable.
y	Numeric vector with the variable of interest.
conf	Confidence level, default 0.95.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object stating the recommended estimator, the test statistic, the confidence interval for a_0 , and both estimators computed side by side for comparison.

Examples

```
x <- c(1, 30, 44, 20, 0, 10, 15, 5, 2, 50, 35, 25)
y <- c(2, 35, 50, 27, 1, 15, 17, 7, 0, 53, 35, 30)
reg_teste_origem(x, y)
```

sist_intervalo	<i>Sampling interval for systematic sampling</i>
----------------	--

Description

Computes the sampling interval $k = N/n$ used to select a systematic sample, following Section 5.1-5.2 of the course notes. When N/n is not an integer, k is first rounded up; if $k*n > N$, k is then decremented by 1, and the correction constant $k^* = N - k*n$ is computed (Example 5.2 of the course notes).

Usage

```
sist_intervalo(N, n, ...)
```

Arguments

N	Population size.
n	Desired sample size.
...	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the sampling interval (and, when N/n is not an integer, the correction constant k^* needed by [sist_selecionar_amostra\(\)](#)).

Examples

```
sist_intervalo(N = 2000, n = 80)
sist_intervalo(N = 2000, n = 75)
```

`sist_seleccionar_amostra`*Select a systematic sample*

Description

Selects a systematic sample of size n from a population of size N , handling both the integer-interval case and the non-integer-interval case (with the correction constant k^*), following the selection rules of Section 5.2 of the course notes (Examples 5.1-5.2). In the non-integer case, the first selected element is r ; from the second element onward, k^* is added on top of the regular step k (i.e. element i for $i \geq 2$ equals $(i-1)*k + k^*$ plus r), exactly as illustrated in Example 5.2.

Usage

```
sist_seleccionar_amostra(N, n, r = NULL, semente = NULL, ...)
```

Arguments

<code>N</code>	Population size.
<code>n</code>	Desired sample size.
<code>r</code>	Optional random start (an integer between 1 and k , inclusive). If <code>NULL</code> , one is drawn uniformly at random.
<code>semente</code>	Optional integer seed for reproducibility. The user's global random state is saved and restored, so calling this function does not affect subsequent unrelated random draws.
<code>...</code>	Ignored (see package overview on tolerated extra arguments).

Value

A `sampling_result` object with the selected element positions (1:N).

Examples

```
sist_seleccionar_amostra(N = 2000, n = 80, r = 8)
sist_seleccionar_amostra(N = 2000, n = 75, r = 26)
```

`summary.sampling_result`*Summary method for sampling_result objects*

Description

Prints the print method output plus the full step-by-step calculation trace and bibliographic references.

Usage

```
## S3 method for class 'sampling_result'  
summary(object, ...)
```

Arguments

<code>object</code>	A <code>sampling_result</code> object.
<code>...</code>	Further arguments (unused, kept for S3 consistency).

Value

object, invisibly.

Index

.razao_resumo(), 35

aas_ajustar_perdas, 3

aas_ajustar_perdas(), 11

aas_comparar_reposicao, 4

aas_estima_media, 4

aas_estima_proporcao, 5

aas_estima_total, 6

aas_tamanho_amostra, 7

aas_tamanho_amostra(), 22, 28

aas_tamanho_media, 8

aas_tamanho_media(), 7, 8, 11, 14, 17, 18, 38, 39

aas_tamanho_proporcao, 9

aas_tamanho_proporcao(), 7, 8, 11, 14

amostra_estimadores, 13

amostragem_ajustar_deff_resposta, 10

amostragem_ajustar_deff_resposta(), 14

amostragem_ajustar_resposta, 11

amostragem_comparar_planos, 12

cong1_ajustar_deff, 14

cong1_ajustar_deff(), 11

cong1_estima_media, 15

cong1_estima_media(), 16

cong1_estima_proporcao, 16

cong1_tamanho_media, 17

cong1_tamanho_media(), 11, 14, 18

cong1_tamanho_proporcao, 18

cong2_estima_media, 19

cong2_estima_media(), 20–23

cong2_estima_proporcao, 20

cong2_estima_razao_media, 21

cong2_estima_razao_media(), 20, 22, 23

cong_estima, 22

estr_alocacao_custo, 24

estr_alocacao_neyman, 25

estr_alocacao_proporcional, 26

estr_estima_media, 26

estr_estima_proporcao, 27

estr_tamanho_amostra, 28

estr_tamanho_media, 29

estr_tamanho_media(), 11, 14, 28–31

estr_tamanho_proporcao, 30

estr_tamanho_proporcao(), 28, 29

pop_media, 31

pop_proporcao, 32

pop_total, 32

print.sampling_result, 33

razao_estima_media, 34

razao_estima_media(), 41

razao_estima_R, 35

razao_estima_R(), 34, 40

razao_estima_total, 36

razao_tamanho_media, 37

razao_tamanho_media(), 40

razao_tamanho_R, 38

razao_tamanho_R(), 37, 39, 40

razao_tamanho_total, 39

razao_tamanho_total(), 40

razao_variancia_piloto, 40

razao_variancia_piloto(), 37–39

razao_vs_regressao_decidir, 41

reg_ajustar, 42

reg_estima_media, 42

reg_estima_media(), 41

reg_teste_origem, 43

reg_teste_origem(), 41, 42

sist_intervalo, 44

sist_selecionar_amostra, 45

sist_selecionar_amostra(), 44

summary.sampling_result, 46