Package 'ltgsmd'

September 2, 2026

Title Latent True-Score and Target-Population Anchored Geometric SMD

Version 0.2.2

Description Implements the latent true-score and target-population anchored geometric standardized mean difference (LTG-SMD) framework for two-group effect-size analysis. Provides plug-in estimation, analytic delta-method confidence intervals using a sample fourth-moment plug-in, bias-corrected and bias-corrected-accelerated nonparametric bootstrap confidence intervals with study x group stratification, denominator-sensitivity profiles, and multi-site meta-analytic wrappers. Includes denominator-diagnostic reporting, a pluggable reliability estimator interface, and an explicit interface for specifying the target reference distribution. Companion software to the methodological paper ``The Denominator Chooses the Estimand: A Target-Population True-Score Framework for Standardized Mean Differences" (Nakamura, in press, Psychological Methods).

License MIT + file LICENSE

URL https://doi.org/10.17605/OSF.IO/KW9R6

Encoding UTF-8

Depends R (>= 4.1.0)

Imports stats

Suggests boot, metafor, psych, testthat (>= 3.0.0), knitr, rmarkdown

Config/testthat/edition 3

VignetteBuilder knitr

Config/roxygen2/version 8.1.0

NeedsCompilation no

Author Daiki Nakamura [aut, cre] (ORCID: <https://orcid.org/0000-0002-0625-6918>)

Maintainer Daiki Nakamura <nakamura@miyazaki-u.ac.jp>

Repository CRAN

Date/Publication 2026-09-02 11:30:09 UTC

Contents

`ltgsmd-package` *ltgsmd: Latent True-Score and Target-Population Anchored Geometric SMD*

Description

Implements the LTG-SMD framework for two-group effect-size analysis. The LTG-SMD is a standardized mean difference whose denominator is the geometric mean of group-specific true-score standard deviations in an explicitly chosen target reference population. The package supports single-study analyses with internal holdout references, analyses with external norm samples, and multi-site coordinated-replication analyses with cross-site pooled references.

Core functions

`compute_ltg_smd()` Compute the LTG-SMD and four conventional comparators (Hedges's g, Welch-type SMD, observed geometric SMD, external observed SMD) from study and reference data.

`ltg_smd_ci()` Analytic delta-method and bias-corrected (BC) or BCa nonparametric bootstrap confidence intervals.

`denominator_diagnostics()` Table F.1 diagnostics for reporting (Section 8.3 of the paper).

`denominator_sensitivity()` Profile across six symmetric denominators (Section 7 Simulation D).

`sensitivity_reference()` Sensitivity to the choice of target reference.

`multisite_ltg_smd()` Multi-site analysis with cross-site pooled reference and meta-analytic comparison.

`export_supplementary()` Bundle diagnostics, CIs, and sensitivity profiles for a supplementary appendix.

Companion paper

"The Denominator Chooses the Estimand: A Target-Population True-Score Framework for Standardized Mean Differences" (Author, submitted to Psychological Methods). See the package vignette `vignette("getting-started", package = "ltgsmd")` for worked examples corresponding to Sections 8.1, 8.2, and 8.3 of the paper.

Author(s)

Maintainer: Daiki Nakamura <nakamura@miyazaki-u.ac.jp> (ORCID)

Authors:

- Daiki Nakamura <nakamura@miyazaki-u.ac.jp> (ORCID)

See Also

Useful links:

- doi:10.17605/OSF.IO/KW9R6

---

coef_alpha *Compute coefficient alpha (Cronbach's alpha)*

Description

Computes Cronbach's coefficient alpha from item-level data. Uses listwise-complete cases by default. The implementation follows the standard formula

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_i s_i^2}{s_T^2}\right)$$

where k is the number of items, s_i^2 are item variances, and s_T^2 is the variance of the total score.

Usage

```
coef_alpha(items, na.rm = TRUE)
```

Arguments

items	A data frame, matrix, or list-like object whose columns are item-level responses (one row per respondent).
na.rm	Logical. If TRUE (default), use complete cases only.

Value

A numeric scalar, the value of coefficient alpha.

Examples

```
set.seed(1)
items <- matrix(rnorm(400), ncol = 4)
coef_alpha(items)
```

compute_ltg_smd *Compute the LTG-SMD and related effect-size estimates*

Description

Computes the latent true-score and target-population anchored geometric standardized mean difference (LTG-SMD) along with conventional comparators: Hedges's g, Welch-type SMD, observed geometric SMD (within study), and external observed geometric SMD (in the reference sample).

Usage

```
compute_ltg_smd(
  study_data,
  reference_data,
  group_var,
  score_var = NULL,
  items = NULL,
  group_levels = NULL,
  reliability_estimator = "alpha",
  rho_external = NULL,
  se_method = c("fourth_moment", "normal"),
  verbose = FALSE
)
```

Arguments

study_data	A data frame containing the study sample, with one row per participant.
reference_data	A data frame containing the reference sample. May be the same as study_data if an internal (holdout) reference is used; see Section 8.2 of the paper.
group_var	Character. Name of the grouping variable in both data frames.
score_var	Character. Name of the precomputed composite score variable. If NULL, the composite is built by averaging the items in items.
items	Optional character vector of item names. Required if score_var is NULL, or if reliability is to be estimated from item- level data.
group_levels	Named character vector with names "reference" (group 0) and "focal" (group 1) specifying the levels of group_var. If NULL, the two unique values are sorted alphabetically with the first taken as reference.
reliability_estimator	Either the string "alpha" (default; uses coef_alpha()) or a function taking a matrix/data frame of items and returning a numeric scalar in (0, 1]. Custom estimators allow plugging in psych::omega(), CFA-based reliability, or any user-supplied procedure.
rho_external	Optional named numeric vector with names "focal" and "reference" specifying externally supplied reliability estimates. If provided, these override the internal estimator. Each value must lie in (0, 1].

`se_method`	Character. "fourth_moment" (default; uses sample fourth moments per Section 4 of the paper) or "normal" (uses the simpler 2*sigma^4/(m-1) approximation that is valid under normality).
`verbose`	Logical. If TRUE, prints intermediate quantities.

Details

The LTG-SMD estimand (Section 4 of the companion paper) is

$$\delta_{\mathrm{LTG}} = \frac{\mu_{T1} - \mu_{T0}}{(\sigma^2_{T1,R}\sigma^2_{T0,R})^{1/4}}$$

where μ_{Tg} are true-score means and $\sigma^2_{Tg,R}$ are group-specific true-score variances in the target reference population R. The plug-in estimator substitutes observed study-sample means for the true-score means (assuming $\mu_{Yg} = \mu_{Tg}$) and reference-sample reliabilities times reference-sample observed variances for the true-score variances.

Hedges's small-sample correction factor $J(N) = 1 - 3/(4N - 9)$ is applied to the conventional Hedges's g estimate but not to the LTG-SMD or to the observed geometric SMDs, consistent with conventional practice.

The analytic SE under `se_method = "fourth_moment"` uses $\widehat{\mathrm{Var}}(s^2) = (\hat{\mu}_4 - \frac{m-3}{m-1}(s^2)^2)/m$, which is consistent under arbitrary distributions of the score with finite fourth moments and reduces to the normal-theory approximation $2\sigma^4/(m-1)$ when the score is normal. Reliability is treated as fixed at its point estimate (the reliability-fixed approximation of Section 6); a reliability-propagated alternative is implemented by the bootstrap in `ltg_smd_ci()`.

Value

An object of class "ltg_smd" containing:

point Numeric. The LTG-SMD point estimate.

se_analytic Numeric. The analytic delta-method standard error (reliability-fixed approximation; see Section 6).

se_method Character. Which SE method was used.

estimates Numeric vector with elements hedges_g, welch_smd, observed_geometric, external_observed, ltg_smd.

study Data frame of group-level study-sample statistics.

reference Data frame of group-level reference-sample statistics.

c_g Numeric vector of length 2: study-to-reference SD ratios (c_focal, c_reference).

rho_g Numeric vector of length 2: reliability estimates (rho_focal, rho_reference) in the reference sample.

factors Named numeric vector with reliability_factor, study_to_target_factor, predicted_ratio_observed_to_LTG.

call The matched call.

Examples

```
set.seed(2026)
gen_items <- function(n, shift = 0) {
  true <- rnorm(n)
  data.frame(
    item1 = true + rnorm(n, sd = 0.6) + shift,
    item2 = true + rnorm(n, sd = 0.6) + shift,
    item3 = true + rnorm(n, sd = 0.6) + shift,
    item4 = true + rnorm(n, sd = 0.6) + shift
  )
}
study_df <- rbind(
  cbind(condition = "control",   gen_items(15)),
  cbind(condition = "treatment", gen_items(15, shift = 0.5))
)
reference_df <- rbind(
  cbind(condition = "control",   gen_items(30)),
  cbind(condition = "treatment", gen_items(30, shift = 0.5))
)

result <- compute_ltg_smd(
  study_data     = study_df,
  reference_data = reference_df,
  group_var      = "condition",
  items          = c("item1", "item2", "item3", "item4"),
  group_levels   = c(reference = "control", focal = "treatment")
)
print(result)

# Custom reliability estimator: use psych::omega instead of alpha
# (illustrated on a larger synthetic sample, since omega needs more
# data than alpha to fit stably)
if (requireNamespace("psych", quietly = TRUE)) {
  gen_items6 <- function(n, shift = 0) {
    true <- rnorm(n)
    items <- as.data.frame(replicate(6, true + rnorm(n, sd = 0.6),
                                     simplify = FALSE))
    names(items) <- paste0("item", 1:6)
    items[] <- lapply(items, `+`, shift)
    items
  }
  study_big <- rbind(
    cbind(condition = "control",   gen_items6(60)),
    cbind(condition = "treatment", gen_items6(60, shift = 0.5))
  )
  reference_big <- rbind(
    cbind(condition = "control",   gen_items6(100)),
    cbind(condition = "treatment", gen_items6(100, shift = 0.5))
  )
  my_omega <- function(items) psych::omega(items, plot = FALSE)$omega.tot
  result_omega <- compute_ltg_smd(
    study_big, reference_big,
```

```
    group_var = "condition",
    items = paste0("item", 1:6),
    reliability_estimator = my_omega,
    group_levels = c(reference = "control", focal = "treatment")
  )
  print(result_omega)
}
```

`denominator_diagnostics` *Denominator diagnostics for an LTG-SMD analysis*

Description

Produces the diagnostic table recommended in Section 8.3 of the companion paper (Table F.1 format). The table reports group-level sample sizes, observed standard deviations, reference reliabilities, true-score standard deviations, study-to-reference ratios c_g, Hedges's g, the LTG-SMD, the reliability factor $(\rho_1\rho_0)^{1/4}$, the study-to-target factor $(c_1c_0)^{-1/2}$, the predicted ratio of the within-study observed geometric SMD to the LTG-SMD (exact under the decomposition), and the observed ratio of Hedges's g to the LTG-SMD (which approximately matches the predicted ratio with small deviations due to the pooled-versus-geometric denominator and Hedges's J correction).

Usage

```
denominator_diagnostics(x)
```

Arguments

x	An "ltg_smd" object.

Value

A data frame in a long format suitable for direct reporting. The data frame has class "denominator_diagnostics".

Examples

```
set.seed(2026)
gen_items <- function(n, shift = 0) {
  true <- rnorm(n)
  data.frame(
    item1 = true + rnorm(n, sd = 0.6) + shift,
    item2 = true + rnorm(n, sd = 0.6) + shift,
    item3 = true + rnorm(n, sd = 0.6) + shift,
    item4 = true + rnorm(n, sd = 0.6) + shift
  )
}
study_df <- rbind(
  cbind(condition = "control",   gen_items(15)),
```

```
  cbind(condition = "treatment", gen_items(15, shift = 0.5))
)
reference_df <- rbind(
  cbind(condition = "control",   gen_items(30)),
  cbind(condition = "treatment", gen_items(30, shift = 0.5))
)
result <- compute_ltg_smd(study_df, reference_df,
  group_var = "condition", items = c("item1", "item2", "item3", "item4"),
  group_levels = c(reference = "control", focal = "treatment"))
diag <- denominator_diagnostics(result)
print(diag)
```

`denominator_sensitivity` *Denominator-sensitivity profile across six symmetric denominators*

Description

Computes the LTG-SMD estimate under each of six symmetric denominator rules applied to the group-specific reference-sample true-score standard deviations: geometric (default), arithmetic, harmonic, root mean square, minimum, and maximum. This profile reproduces the six rows of Simulation D in Section 7 of the companion paper, and is recommended as a supplementary report when group-specific reference true-score variances differ markedly (Section 8.3).

Usage

```
denominator_sensitivity(x)
```

Arguments

`x`	An "ltg_smd" object.

Value

A data frame with one row per denominator rule, columns `denominator`, `D`, `estimate`. Class "denominator_sensitivity".

Examples

```
set.seed(2026)
gen_items <- function(n, shift = 0) {
  true <- rnorm(n)
  data.frame(
    item1 = true + rnorm(n, sd = 0.6) + shift,
    item2 = true + rnorm(n, sd = 0.6) + shift,
    item3 = true + rnorm(n, sd = 0.6) + shift,
    item4 = true + rnorm(n, sd = 0.6) + shift
  )
}
```

```
study_df <- rbind(
  cbind(condition = "control",   gen_items(15)),
  cbind(condition = "treatment", gen_items(15, shift = 0.5))
)
reference_df <- rbind(
  cbind(condition = "control",   gen_items(30)),
  cbind(condition = "treatment", gen_items(30, shift = 0.5))
)
result <- compute_ltg_smd(study_df, reference_df,
  group_var = "condition", items = c("item1", "item2", "item3", "item4"),
  group_levels = c(reference = "control", focal = "treatment"))
sens <- denominator_sensitivity(result)
print(sens)
```

`export_supplementary` *Export an LTG-SMD analysis to a supplementary appendix template*

Description

Bundles the point estimates, confidence intervals, denominator diagnostics, and denominator-sensitivity profile into a list and (optionally) writes them to a Markdown file for inclusion in a supplementary appendix.

Usage

```
export_supplementary(object, ci = NULL, file = NULL)
```

Arguments

`object`	An "ltg_smd" object.
`ci`	An optional "ltg_smd_ci" object from `ltg_smd_ci()`.
`file`	Optional path. If supplied, the supplementary appendix is written there as Markdown. If NULL, only the list is returned.

Value

Invisibly, a list with elements diagnostics, sensitivity, estimates, and ci. If `file` is supplied, the Markdown is also written to disk.

`ltg_smd_ci` *Compute confidence intervals for the LTG-SMD*

Description

Computes one or more confidence intervals for the LTG-SMD, including the analytic delta-method interval (reliability-fixed, Section 6 of the paper) and the bias-corrected (BC) or bias-corrected-and-accelerated (BCa) nonparametric bootstrap interval.

Usage

```
ltg_smd_ci(
  object,
  method = c("analytic", "bootstrap"),
  boot_type = c("bc", "bca", "percentile"),
  B = 2000,
  level = 0.95,
  seed = NULL,
  parallel = c("no", "multicore", "snow"),
  ncpus = 1L,
  study_data = NULL,
  reference_data = NULL,
  group_var = NULL,
  score_var = NULL,
  items = NULL,
  group_levels = NULL,
  reliability_estimator = "alpha",
  rho_external = NULL
)
```

Arguments

`object`	An object of class "ltg_smd" returned by `compute_ltg_smd()`.
`method`	Character vector with elements from c("analytic", "bootstrap"). Both can be requested.
`boot_type`	Character. Type of bootstrap interval: "bc" (bias-corrected, default), "bca" (bias-corrected and accelerated), or "percentile" (uncorrected percentile interval).
`B`	Integer. Number of bootstrap replicates. Default 2000.
`level`	Numeric. Confidence level, default 0.95.
`seed`	Optional integer for reproducibility of the bootstrap.
`parallel`	Character; passed to `boot::boot()`. Default "no".
`ncpus`	Integer; passed to `boot::boot()` if parallel != "no".
`study_data`	Data frame. Required for bootstrap (the original study sample passed to compute_ltg_smd).

`reference_data` Data frame. Required for bootstrap.

`group_var`, `score_var`, `items`, `group_levels`, `reliability_estimator`
Same as in `compute_ltg_smd()`; required when bootstrap is requested.

`rho_external` Optional; see `compute_ltg_smd()`.

Details

Stratification. The bootstrap resamples within the four strata defined by the cross-classification of source (study vs reference) and experimental group (focal vs reference). This preserves the sample sizes of all four study x group cells across replicates, which is the appropriate stratification for a two-group LTG-SMD analysis. Earlier versions of this package stratified only on source; that design allowed the group-1/group-0 ratio within each source to vary across replicates, producing unstable confidence intervals under small samples or unequal group sizes.

Reliability propagation. Because the bootstrap recomputes the entire LTG-SMD plug-in within each replicate, item-level reliability is re-estimated each time (when `items` is supplied). This propagates uncertainty in the reliability estimator through to the interval. By contrast, the analytic interval treats reliability as fixed at its point estimate.

Value

An object of class "ltg_smd_ci" containing:

analytic Named numeric vector with elements estimate, lower, upper, se, level.

bootstrap Named numeric vector with bootstrap interval, or NULL if bootstrap not requested.

boot_details If bootstrap requested, list with B, type, seed, z0, acceleration, strata_design, and the bootstrap distribution.

Examples

```
set.seed(2026)
gen_items <- function(n, shift = 0) {
  true <- rnorm(n)
  data.frame(
    item1 = true + rnorm(n, sd = 0.6) + shift,
    item2 = true + rnorm(n, sd = 0.6) + shift,
    item3 = true + rnorm(n, sd = 0.6) + shift,
    item4 = true + rnorm(n, sd = 0.6) + shift
  )
}
study_df <- rbind(
  cbind(condition = "control",   gen_items(15)),
  cbind(condition = "treatment", gen_items(15, shift = 0.5))
)
reference_df <- rbind(
  cbind(condition = "control",   gen_items(30)),
  cbind(condition = "treatment", gen_items(30, shift = 0.5))
)
result <- compute_ltg_smd(study_df, reference_df,
  group_var = "condition", items = c("item1", "item2", "item3", "item4"),
  group_levels = c(reference = "control", focal = "treatment"))
```

```
# B is kept small here so the example runs quickly; in practice use a
# larger B (e.g. 2000, the default) for stable bootstrap intervals.
ci <- ltg_smd_ci(result,
  method = c("analytic", "bootstrap"),
  B = 200, seed = 20260814,
  study_data = study_df, reference_data = reference_df,
  group_var = "condition", items = c("item1", "item2", "item3", "item4"),
  group_levels = c(reference = "control", focal = "treatment"))
print(ci)
```

`multisite_ltg_smd` *Multi-site LTG-SMD analysis with cross-site reference anchoring*

Description

Implements the multi-site workflow described in Section 7.3 and illustrated in Example 3 of the paper (Many Labs 2 Anderson.1). Each site is treated as a study sample, the cross-site pooled distribution (or an external sample) is treated as the target reference, and site-level Hedges's g and LTG-SMD estimates are returned along with a meta-analytic comparison from `metafor::rma()`.

Usage

```
multisite_ltg_smd(
  data,
  site_var,
  group_var,
  score_var = NULL,
  items = NULL,
  group_levels = NULL,
  reliability_estimator = "alpha",
  reference_strategy = c("pooled_across_sites", "external"),
  reference_data = NULL,
  min_n_per_group = 50L,
  rho_external = NULL
)
```

Arguments

`data`	A data frame with one row per participant containing group, site, item-level, and (optionally) composite-score columns.
`site_var`	Character. Name of the site grouping variable.
`group_var`	Character. Name of the experimental-condition variable.
`score_var`	Optional character. Name of a precomputed composite score column. If NULL, built from `items` (in both `data` and, if supplied, `reference_data`).
`items`	Optional character vector of item column names.

`group_levels` Optional named character vector with names "reference" and "focal".

`reliability_estimator`
Either "alpha" (default) or a function; see `compute_ltg_smd()`.

`reference_strategy`
One of "pooled_across_sites" (default; uses all retained sites as the reference) or "external" (requires `reference_data`).

`reference_data` Optional data frame; required when `reference_strategy = "external"`.

`min_n_per_group`
Integer. Minimum per-group sample size for a site to be retained. Defaults to 50.

`rho_external` Optional named numeric vector with names "focal" and "reference" to override item-level reliability estimation.

Details

The site-level standard error for the LTG-SMD is approximated as

$$\widehat{\mathrm{SE}}_{\mathrm{LTG,site}} = \sqrt{s^2_{1,\mathrm{site}}/n_{1,\mathrm{site}} + s^2_{0,\mathrm{site}}/n_{0,\mathrm{site}}} \,/\, D,$$

treating the cross-site (or external) reference denominator D as fixed. This approximation is reasonable when the reference sample is large relative to each site, which is typical in coordinated multi-site projects, but it understates uncertainty when the reference is small. For a full reliability-propagated interval at the site level, use `ltg_smd_ci()` with `method = "bootstrap"` on each site separately.

Value

An object of class "multisite_ltg_smd" containing:

site_table Data frame of site-level estimates: site, n_focal, n_reference, hedges_g, ltg_smd, se_hedges, se_ltg.

reference_summary List of pooled reference quantities (variances, reliabilities, true-score geometric SD).

meta_hedges Output of `metafor::rma()` for Hedges's g.

meta_ltg Output of `metafor::rma()` for the LTG-SMD.

meta_comparison Data frame summarizing the two meta-analytic pools side by side.

se_ltg_method Character. Notes that the site-level SE for the LTG-SMD treats the reference denominator as fixed.

Examples

```
set.seed(2026)
gen_site <- function(site_name, n_per_cond = 15, shift = 0.3) {
  true_ctrl <- rnorm(n_per_cond)
  true_trt  <- rnorm(n_per_cond)
  ctrl <- data.frame(
    site = site_name, condition = "control",
```

```
    item1 = true_ctrl + rnorm(n_per_cond, sd = 0.6),
    item2 = true_ctrl + rnorm(n_per_cond, sd = 0.6),
    item3 = true_ctrl + rnorm(n_per_cond, sd = 0.6),
    item4 = true_ctrl + rnorm(n_per_cond, sd = 0.6)
  )
  trt <- data.frame(
    site = site_name, condition = "treatment",
    item1 = true_trt + rnorm(n_per_cond, sd = 0.6) + shift,
    item2 = true_trt + rnorm(n_per_cond, sd = 0.6) + shift,
    item3 = true_trt + rnorm(n_per_cond, sd = 0.6) + shift,
    item4 = true_trt + rnorm(n_per_cond, sd = 0.6) + shift
  )
  rbind(ctrl, trt)
}
ml_data <- do.call(rbind, lapply(
  c("SiteA", "SiteB", "SiteC", "SiteD"), gen_site
))

result <- multisite_ltg_smd(
  data = ml_data,
  site_var = "site",
  group_var = "condition",
  items = paste0("item", 1:4),
  reference_strategy = "pooled_across_sites",
  min_n_per_group = 10
)
print(result)
```

`sensitivity_reference` *Sensitivity of the LTG-SMD to the choice of target reference*

Description

Recomputes the LTG-SMD under each of several user-supplied alternative reference data frames. Useful when more than one target reference is defensible (Section 8.3 of the companion paper).

Usage

```
sensitivity_reference(
  object,
  alternative_references,
  study_data,
  group_var,
  score_var = NULL,
  items = NULL,
  group_levels = NULL,
  rho_external = NULL
)
```

Arguments

`object`	An "ltg_smd" object computed with a primary reference.
`alternative_references`	A named list of reference data frames. Names are used as labels in the output.
`study_data`	Original study data (required to recompute).
`group_var`, `score_var`, `items`, `group_levels`, `rho_external`	Arguments to pass through to `compute_ltg_smd()`.

Value

A data frame with columns `reference`, `ltg_smd`, `c_focal`, `c_reference`, `rho_focal`, `rho_reference`, including the primary reference in the first row.

Index