

Package ‘dsFedAte’

October 15, 2026

Title Server-Side Treatment Effect Computation for Federated Studies

Version 0.1.0

Description Server-side 'DataSHIELD' package that computes local average treatment effect (ATE) statistics from arm-specific regression coefficients provided by the analyst. Implements the server-side component of the G-formula estimator from 'Federated Causal Inference: Multi-Study ATE Estimation beyond Meta-Analysis' (R. Khellaf, A. Bellet and J. Josse, 2025), enforcing 'DataSHIELD' disclosure controls throughout.

License AGPL (>= 3)

Encoding UTF-8

RoxygenNote 8.0.0

Imports cli, dsBase, stats

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Francisco Andrade [aut, cre],
Rémi Khellaf [aut],
Julie Josse [aut]

Maintainer Francisco Andrade <francisco.de-lima-andrade@inria.fr>

Repository CRAN

Date/Publication 2026-10-15 17:00:12 UTC

Contents

addStudyIdDS	2
localAteDS	3

Index	6
--------------	----------

addStudyIdDS

Add study membership dummy columns to a server-side data frame

Description

Server-side assign function called once per server before the study-effects arm-specific regression models are fitted via `ds.glm()`. Called automatically when `study_effects = TRUE` is passed to `dsFedAteClient::ds.EstimateAte()`. Appends $K-1$ explicit 0/1 integer columns named `study_id2`, ..., `study_idK` to the server-side data frame.

Usage

```
addStudyIdDS(base_name, study_index, K_studies)
```

Arguments

<code>base_name</code>	A character string naming the server-side data frame.
<code>study_index</code>	A character string giving this server's 1-based study index (its position in the client's <code>datasources</code> list).
<code>K_studies</code>	A character string giving the total number of studies K .

Details

These are dummy columns: column `study_idk` equals 1 on the server whose study index is k and 0 on every other server. Study 1 is the reference level (all dummies zero).

Because the columns are plain integers, `ds.glm()` includes them in the model formula without invoking R's factor-contrast machinery. This avoids the *"contrasts can be applied only to factors with 2 or more levels"* error that arises when a factor column carries only one level locally (every server holds observations from exactly one study).

The column values are entirely determined by the study index supplied by the client — they are constants, not individual-level data — so appending them carries no disclosure risk beyond the existing `nfilter.subset` check on the dataset size.

Value

The augmented data frame with $K-1$ integer dummy columns (`study_id2`, ..., `study_idK`) appended. The return value is assigned to a server-side symbol by `DSI::datashield.assign()` on the client.

Privacy control level

- Permitted: permissive, banana.
- Blocked: carrot, avocado, non-permissive.

References

Khellaf, R., Bellet, A., and Josse, J. (2025). Federated Causal Inference: Multi-Study ATE Estimation beyond Meta-Analysis. *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS 2025)*.

See Also

[localAteDS](#), `dsFedAteClient::ds.EstimateAte()`.

Examples

```
## Not run:
# Called automatically by dsFedAteClient::ds.EstimateAte(..., study_effects = TRUE).
# Direct invocation (per server, from the client):
cally <- call("addStudyIdDS", "study_data", "2", "3")
DSI::datashield.assign(conns["site2"], "study_data_adj", cally)

## End(Not run)
```

localAteDS	<i>Compute local statistics for the federated average treatment effect (ATE) and (optionally) its variance</i>
------------	--

Description

Server-side aggregate function called after the federated GLM has converged. Always returns the quantities needed for the ATE point estimate. When `compute_variance = TRUE` (the default) it also returns the statistics required for the delta-method variance. Set `compute_variance = FALSE` to limit disclosure to the minimum needed for the point estimate only.

Usage

```
localAteDS(
  base_name,
  outcome_name,
  covariate_names_str,
  treatment_name,
  alpha0_str,
  beta0_str,
  alpha1_str,
  beta1_str,
  compute_variance = TRUE,
  heterogeneous = TRUE
)
```

Arguments

<code>base_name</code>	A character string naming the server-side data frame.
<code>outcome_name</code>	A character string naming the outcome variable inside <code>base_name</code> .
<code>covariate_names_str</code>	A character string of comma-separated covariate names.
<code>treatment_name</code>	A character string naming the binary treatment indicator inside <code>base_name</code> .
<code>alpha0_str</code>	Character. Intercept of the control-arm model.
<code>beta0_str</code>	Character. Comma-separated slope coefficients of the control-arm model.
<code>alpha1_str</code>	Character. Intercept of the treated-arm model.
<code>beta1_str</code>	Character. Comma-separated slope coefficients of the treated-arm model.
<code>compute_variance</code>	Logical. If TRUE (default), also return the within-site sum of squared deviations of the individual treatment-effect projections from their site mean, and — when <code>heterogeneous = TRUE</code> — the column sums of the augmented design matrix. Set to FALSE to limit disclosure to the three scalars strictly needed for the point estimate.
<code>heterogeneous</code>	Logical. Controls which variance statistics are returned. TRUE (default) makes no assumption about the participating RCTs: each site may have a different covariate distribution and a different treatment probability. The column sums of the augmented design matrix are returned so the client can assemble the global covariate mean needed for the leverage-corrected variance. FALSE assumes all RCTs share the same covariate distribution; these column sums are then unnecessary and are not returned, limiting disclosure. Has no effect when <code>compute_variance = FALSE</code> .

Value

A named list. The following three elements are always present:

- n_k** Total sample size on this server.
- n_k_1** Number of treated observations on this server.
- tau_k** Local average treatment effect estimate $\hat{\tau}_k$.

When `compute_variance = TRUE`, one further element is added:

- s_k** Within-site sum of squared deviations of the individual treatment-effect projections from their site mean: $\sum_{i \in k} (X_i^\top \hat{\gamma} - \bar{X}_k^\top \hat{\gamma})^2$, where $\hat{\gamma} = \hat{\beta}_1 - \hat{\beta}_0$ is the difference of the arm-specific slope vectors.

When `compute_variance = TRUE` and `heterogeneous = TRUE`, two additional elements:

- g_raw_k** Column sums of the augmented design matrix $(1, X)$, i.e. the vector

$$(n_k, \sum_i X_{i1}, \dots, \sum_i X_{id}).$$

Accumulated across servers by the client to obtain the global covariate mean used for both arm gradients.

Privacy control level

- Permitted: permissive, banana.
- Blocked: carrot, avocado, non-permissive.

References

Khellaf, R., Bellet, A., and Josse, J. (2025). Federated Causal Inference: Multi-Study ATE Estimation beyond Meta-Analysis. *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS 2025)*.

See Also

[addStudyIdDS](#), `dsFedAteClient::ds.EstimateAte()`.

Examples

```
## Not run:
df <- data.frame(
  X1 = rnorm(30), X2 = rnorm(30),
  W = rbinom(30, 1, 0.5)
)
df$Y <- 1 + 0.5 * df$W + 0.3 * df$X1 + rnorm(30, sd = 0.5)
options(datashield.privacyControlLevel = "banana", nfilter.subset = 3)
localAteDS("df", "Y", "X1,X2", "W", "1.0", "0.3,0.0", "1.5", "0.3,0.0")

## End(Not run)
```

Index

addStudyIdDS, [2](#), [5](#)

localAteDS, [3](#), [3](#)