

Package: robustT2 (via r-universe)

May 9, 2026

Title Robust Hotelling-Type T^2 Control Chart Based on the Dual STATIS Approach

Version 0.1.0

Description Implements a robust multivariate control-chart methodology for batch-based industrial processes with multiple correlated variables using the Dual STATIS (Structuration des Tableaux A Trois Indices de la Statistique) framework. A robust compromise covariance matrix is constructed from Phase I batches with the Minimum Covariance Determinant (MCD) estimator, and a Hotelling-type T^2 statistic is applied for anomaly detection in Phase II. The package includes functions to simulate clean and contaminated batches, to compute both robust and classical Hotelling T^2 control charts, to visualize results via robust biplots, and to launch an interactive 'shiny' dashboard. An internal dataset (pharma_data) is provided for reproducibility. See Lavit, Escoufier, Sabatier and Traissac (1994) [doi:10.1016/0167-9473\(94\)90134-1](https://doi.org/10.1016/0167-9473(94)90134-1) for the original STATIS methodology, and Rousseeuw and Van Driessen (1999) [doi:10.1080/00401706.1999.10485670](https://doi.org/10.1080/00401706.1999.10485670) for the MCD estimator.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Imports dplyr, MASS, ggplot2, rrcov, shiny, ggrepel, forcats, Matrix

Depends R (>= 3.5)

URL <https://github.com/SergioDanielFG/robustT2>

BugReports <https://github.com/SergioDanielFG/robustT2/issues>

Suggests spelling

Language en-US

Config/pak/sysreqs cmake make libuv1-dev zlib1g-dev

Repository <https://sergiodanielfg.r-universe.dev>
Date/Publication 2025-09-05 15:46:36 UTC
RemoteUrl <https://github.com/sergiodanielfg/robustt2>
RemoteRef HEAD
RemoteSha c2e5c29b4b59e25b5e169bb209204bc20ebc8db0

Contents

pharma_data	2
plot_classical_hotelling_t2_chart	3
plot_classical_hotelling_t2_phase2_chart	4
plot_statis_biplot_projection	5
plot_statis_hj_biplot	6
plot_statis_phase1_chart	7
plot_statis_phase2_chart	8
robust_statis_phase1	9
run_statis_dashboard	10
simulate_pharma_batches	11

Index	12
--------------	-----------

pharma_data	<i>Simulated Pharmaceutical Manufacturing Data</i>
-------------	--

Description

This dataset contains simulated pharmaceutical manufacturing data generated by `simulate_pharma_batches()` with `seed = 780` and `obs_per_batch = 30`.

Usage

```
data("pharma_data")
```

Format

A data frame with 450 rows and 7 variables:

Batch Batch identifier (factor)
Phase Phase indicator: "Phase 1" or "Phase 2" (factor)
Status Batch status: "Under Control" or "Out of Control" (factor)
Concentration Concentration of active ingredient (mg/mL)
Humidity Humidity percentage (% w/w)
Dissolution Dissolution percentage (% released)
Density Density (g/cm³)

Details

Phase 1 includes 10 under-control batches with natural variability in mean and covariance, without contamination.

Phase 2 includes 2 additional under-control batches and 3 out-of-control batches. The out-of-control batches exhibit shifts in both mean and variability, along with moderate contamination in a portion of their observations.

Each batch contains 30 observations measured across four quantitative quality-control variables.

Source

Simulated using [simulate_pharma_batches](#) with seed = 780 and obs_per_batch = 30.

plot_classical_hotelling_t2_chart

Plot Classical Hotelling T2 Control Chart

Description

Plots the classical Hotelling T2 statistics per batch with a uniform color line. Batches are evaluated against a control threshold obtained from the chi-squared distribution with degrees of freedom equal to the number of variables.

Usage

```
plot_classical_hotelling_t2_chart(  
  t2_statistics,  
  num_vars,  
  title = "Classical Hotelling T2 Control Chart"  
)
```

Arguments

t2_statistics	A data frame with columns Batch and T2_Stat.
num_vars	Integer. Number of variables used in the multivariate analysis (to compute the Chi ² threshold).
title	Optional string. Plot title.

Value

A ggplot2 object representing the control chart.

Examples

```
# Simulate pharmaceutical manufacturing batches
sim_batches <- simulate_pharma_batches()

# Phase 1 analysis: use Phase 1 data
phase1_data <- subset(sim_batches, Phase == "Phase 1")

# Apply classical Hotelling T2 methodology
t2_result <- hotelling_t2_phase1(
  data = phase1_data,
  variables = c("Concentration", "Humidity", "Dissolution", "Density")
)

# Plot classical Hotelling T2 control chart
plot_classical_hotelling_t2_chart(
  t2_statistics = t2_result$batch_statistics,
  num_vars = 4
)
```

```
plot_classical_hotelling_t2_phase2_chart
```

Plot Classical Hotelling T2 Control Chart - Phase 2

Description

Plots the classical Hotelling T^2 statistics per batch for Phase 2 data, using the reference mean and covariance matrix estimated from Phase 1. Batches are color-coded by control status ("Under Control" = blue, "Out of Control" = red).

Usage

```
plot_classical_hotelling_t2_phase2_chart(
  t2_statistics,
  num_vars,
  title = "Classical Hotelling T2 Control Chart (Phase 2)"
)
```

Arguments

t2_statistics	A data frame with columns Batch, T2_Stat, and Status.
num_vars	Integer. Number of variables used in the multivariate analysis (degrees of freedom for Chi ²).
title	Optional string. Plot title.

Value

A ggplot2 object with the Phase 2 control chart.

Examples

```

# Simulate pharmaceutical manufacturing batches
sim_batches <- simulate_pharma_batches()

# Split by phase
phase1_data <- subset(sim_batches, Phase == "Phase 1")
phase2_data <- subset(sim_batches, Phase == "Phase 2")

# Fit Phase 1 classical estimators
t2_phase1 <- hotelling_t2_phase1(
  data = phase1_data,
  variables = c("Concentration", "Humidity", "Dissolution", "Density")
)

# Evaluate Phase 2 batches
t2_phase2 <- hotelling_t2_phase2(
  new_data = phase2_data,
  variables = c("Concentration", "Humidity", "Dissolution", "Density"),
  center = t2_phase1$center,
  covariance = t2_phase1$covariance
)

# Combine with status for plotting
status_info <- phase2_data[!duplicated(phase2_data$Batch), "Status"]
t2_phase2_plot <- cbind(t2_phase2$batch_statistics, Status = status_info)

# Plot Phase 2 control chart
plot_classical_hotelling_t2_phase2_chart(
  t2_statistics = t2_phase2_plot,
  num_vars = 4
)

```

plot_statis_biplot_projection

HJ-Biplot Projection - Robust STATIS Dual (Phase 2)

Description

Projects new batches from Phase 2 into the HJ-Biplot space defined by the robust compromise matrix and eigen decomposition from Phase 1.

Usage

```
plot_statis_biplot_projection(phase1_result, phase2_result, dims = c(1, 2))
```

Arguments

phase1_result Result from robust_statis_phase1().

phase2_result Result from robust_statis_phase2() (must include standardized_data, t2_stats_by_batch and threshold).

dims Dimensions to plot (default: c(1, 2)).

Details

This implementation follows the HJ-Biplot formulation of Galindo-Villardón (1986). The compromise matrix C , being symmetric and positive semidefinite, is decomposed via an eigen decomposition (not a rectangular SVD). The square roots of eigenvalues are used to build the biplot scaling, consistent with robust STATIS Dual.

Value

A ggplot2 object with the projected HJ-Biplot for Phase 2 batches.

Examples

```
sim_batches <- simulate_pharma_batches()
phase1_data <- subset(sim_batches, Phase == "Phase 1" & Status == "Under Control")
phase2_data <- subset(sim_batches, Phase == "Phase 2")

phase1 <- robust_statis_phase1(
  data = phase1_data,
  variables = c("Concentration", "Humidity", "Dissolution", "Density")
)

phase2 <- robust_statis_phase2(
  new_data = phase2_data,
  variables = c("Concentration", "Humidity", "Dissolution", "Density"),
  medians = phase1$global_medians,
  mads = phase1$global_mads,
  compromise_matrix = phase1$compromise_matrix,
  global_center = phase1$global_center
)

plot_statis_biplot_projection(phase1, phase2)
```

plot_statis_hj_biplot *HJ-Biplot of Robust STATIS Dual Compromise (Galindo-Villardón)*

Description

Generates an HJ-Biplot using the compromise matrix obtained from robust STATIS Dual. Individuals (batch centers) are projected as $G = U D$, and variables as $H = V D$, where D is the diagonal matrix of square roots of eigenvalues.

Usage

```
plot_statis_hj_biplot(  
  phase1_result,  
  dims = c(1, 2),  
  color_by = c("none", "weight", "distance"),  
  highlight_batches = NULL  
)
```

Arguments

`phase1_result` Result from `robust_statis_phase1()`.

`dims` Dimensions to plot (default: `c(1, 2)`).

`color_by` One of "none", "weight", or "distance" for coloring batches.

`highlight_batches` Optional vector of batch names to emphasize.

Value

ggplot2 object with HJ-Biplot.

Examples

```
sim_batches <- simulate_pharma_batches()  
phase1 <- robust_statis_phase1(  
  data = subset(sim_batches, Phase == "Phase 1" & Status == "Under Control"),  
  variables = c("Concentration", "Humidity", "Dissolution", "Density")  
)  
plot_statis_hj_biplot(phase1)
```

plot_statis_phase1_chart

Plot Control Chart - Robust STATIS Dual (Phase 1)

Description

Plots the Hotelling T^2 statistic per batch using the robust center and compromise matrix estimated in `robust_statis_phase1()`. The control limit is based on a Chi-squared distribution with degrees of freedom equal to the number of variables.

Usage

```
plot_statis_phase1_chart(  
  batch_statistics,  
  num_vars,  
  title = "Robust STATIS Dual Control Chart - Phase 1"  
)
```

Arguments

<code>batch_statistics</code>	A data frame with columns <code>Batch</code> and <code>T2_Stat</code> , typically from <code>phase1_result\$batch_statistics</code> .
<code>num_vars</code>	Integer. Number of variables used in the multivariate analysis (to compute the Chi^2 threshold).
<code>title</code>	Optional string. Plot title.

Value

A `ggplot2` object.

Examples

```
sim_batches <- simulate_pharma_batches()

# Phase 1 analysis: select under control batches from Phase 1
phase1_result <- robust_statis_phase1(
  data = subset(sim_batches, Phase == "Phase 1" & Status == "Under Control"),
  variables = c("Concentration", "Humidity", "Dissolution", "Density")
)

# Plot the Phase 1 robust control chart
plot_statis_phase1_chart(
  batch_statistics = phase1_result$batch_statistics,
  num_vars = 4
)
```

`plot_statis_phase2_chart`

Plot STATIS Dual Robust Control Chart - Phase 2 Only

Description

Plots the robust Hotelling T^2 statistics for Phase 2 batches only, using the results from the robust STATIS Dual method.

Usage

```
plot_statis_phase2_chart(
  phase2_result,
  title = "Robust STATIS Dual Control Chart - Phase 2"
)
```

Arguments

<code>phase2_result</code>	A list returned by <code>robust_statis_phase2()</code> , including <code>t2_stats_by_batch</code> with Hotelling T^2 values and a control threshold.
<code>title</code>	Optional string. Plot title.

Value

A ggplot2 object representing the control chart for Phase 2 batches.

Examples

```
sim_batches <- simulate_pharma_batches()
phase1 <- robust_statis_phase1(
  data = subset(sim_batches, Phase == "Phase 1" & Status == "Under Control"),
  variables = c("Concentration", "Humidity", "Dissolution", "Density")
)
phase2 <- robust_statis_phase2(
  new_data = subset(sim_batches, Phase == "Phase 2"),
  variables = c("Concentration", "Humidity", "Dissolution", "Density"),
  medians = phase1$global_medians,
  mads = phase1$global_mads,
  compromise_matrix = phase1$compromise_matrix,
  global_center = phase1$global_center
)
plot_statis_phase2_chart(phase2_result = phase2)
```

robust_statis_phase1 *Robust STATIS Dual - Phase 1 (Under Control Batches)*

Description

Applies the Robust STATIS Dual methodology to Phase 1 data (under control batches), using robust batch-wise standardization (median and MAD). Covariance matrices are robustly estimated using the MCD method and used directly (without trace normalization) to construct the compromise matrix.

Usage

```
robust_statis_phase1(data, variables)
```

Arguments

data A data frame containing the process data with batch information.
variables Character vector with the names of the variables to be used in the analysis.

Value

A list containing:

compromise_matrix Robust compromise matrix (without trace normalization)

global_center Global robust center of the batches

batch_statistics Data frame with Batch, T2_Stat (Hotelling-type robust statistic), and Weight

batch_medians List of medians per batch and variable

batch_mads List of MADs per batch and variable
global_medians Global medians per variable (for use in Phase 2)
global_mads Global MADs per variable
robust_means List of robust centers of each batch (estimated by MCD)
standardized_data Data set standardized batch by batch
robust_covariances List of robust covariance matrices per batch
similarity_matrix Hilbert-Schmidt similarity matrix between batches
statis_weights Weights obtained from the first eigenvector of the similarity matrix
first_eigenvector First eigenvector of the similarity matrix (unnormalized)

Examples

```
# Simulate new pharmaceutical manufacturing batches
sim_batches <- simulate_pharma_batches()

# Select only Phase 1 under control batches
phase1_data <- subset(sim_batches, Phase == "Phase 1" & Status == "Under Control")

# Apply robust STATIS Dual methodology
result <- robust_statis_phase1(
  data = phase1_data,
  variables = c("Concentration", "Humidity", "Dissolution", "Density")
)

# View main outputs
result$compromise_matrix
result$batch_statistics
result$robust_covariances
result$similarity_matrix
result$statis_weights
result$robust_means
```

run_statis_dashboard *Launch STATIS Dual Robust Dashboard (Shiny)*

Description

Launches an interactive Shiny dashboard that includes:

- Phase 1 control chart (sum of robust Mahalanobis distances)
- Phase 2 control chart (for new batches)
- HJ-Biplot visualization

Usage

```
run_statis_dashboard()
```

Value

No return value, called for side effects (launches a Shiny application).

Examples

```
if (interactive()) {  
  run_statis_dashboard()  
}
```

simulate_pharma_batches

Simulate Pharmaceutical Manufacturing Batches (Realistic Variability)

Description

Simulates pharmaceutical manufacturing batches across two phases. Phase 1 includes 10 under-control batches, each with natural variability in mean and covariance. Phase 2 includes 2 clean under-control batches and 3 out-of-control batches with shifted mean, increased dispersion, and moderate contamination.

Usage

```
simulate_pharma_batches(obs_per_batch = 30, seed = 780)
```

Arguments

obs_per_batch Integer. Number of observations per batch. Default is 30.
seed Optional integer. If provided, sets a random seed for reproducibility.

Details

The simulated data includes four quality control variables: Concentration, Humidity, Dissolution, and Density.

Value

A data frame with 450 observations and the following columns:

Batch Factor. Batch identifier (Batch_1 to Batch_15).

Phase Factor. Phase of the process: "Phase 1" or "Phase 2".

Status Factor. Control status: "Under Control" or "Out of Control".

Concentration, Humidity, Dissolution, Density Numeric quality control variables.

Index

* datasets

- pharma_data, [2](#)

- pharma_data, [2](#)
- plot_classical_hotelling_t2_chart, [3](#)
- plot_classical_hotelling_t2_phase2_chart,
[4](#)
- plot_statis_biplot_projection, [5](#)
- plot_statis_hj_biplot, [6](#)
- plot_statis_phase1_chart, [7](#)
- plot_statis_phase2_chart, [8](#)

- robust_statis_phase1, [9](#)
- run_statis_dashboard, [10](#)

- simulate_pharma_batches, [3](#), [11](#)