

# Package: DTSR (via r-universe)

December 9, 2024

**Version** 0.1.0

**Date** 2024-10-22

**Type** Package

**Title** Distributed Trimmed Scores Regression for Handling Missing Data

**Author** Guangbao Guo [aut, cre, cph]  
(<https://orcid.org/0000-0002-4115-6218>), Ruiling Niu [aut]

**Maintainer** Guangbao Guo <ggb11111111@163.com>

**Description** Provides functions for handling missing data using Distributed Trimmed Scores Regression and other imputation methods. It includes facilities for data imputation, evaluation metrics, and clustering analysis. It is designed to work in distributed computing environments to handle large datasets efficiently. The philosophy of the package is described in Guo G. (2024) <[doi:10.1080/03610918.2022.2091779](https://doi.org/10.1080/03610918.2022.2091779)>.

**License** GPL-3

**Encoding** UTF-8

**Depends** R (>= 3.5.0)

**Repository** CRAN

**Imports** stats, mvdalab, DMwR2, cluster, MASS

**RoxygenNote** 7.3.2

**NeedsCompilation** no

**Date/Publication** 2024-11-08 14:30:06 UTC

**Config/pak/sysreqs** cmake make libicu-dev libssl-dev libx11-dev

## Contents

DEM	2
DRPCA	3
DTSR	4
EM	6
IndexCPP	7

KNN . . . . .	8
mean . . . . .	9
MLPCA . . . . .	10
NIPALS . . . . .	12
RPCA . . . . .	13
SVD . . . . .	14
SVDImpute . . . . .	15
TSR . . . . .	16

<b>Index</b>	<b>19</b>
--------------	-----------

DEM

*Distributed EM Imputation (DEM) for Handling Missing Data***Description**

This function performs DEM to handle missing data by dividing the dataset into D blocks, applying the EM imputation method within each block, and then combining the results. It calculates various evaluation metrics including RMSE, MMAE, RRE, and Consistency Proportion Index (CPP) using different hierarchical clustering methods.

**Usage**

```
DEM(data0, data.sample, data.copy, mr, km, D)
```

**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.
<code>D</code>	The number of blocks to divide the data into.

**Value**

A list containing:

<code>XDEM</code>	The imputed dataset.
<code>RMSEDEM</code>	The Root Mean Squared Error.
<code>MAEDEM</code>	The Mean Absolute Error.
<code>REDEM</code>	The Relative Eelative Error.
<code>GCVDEM</code>	The DEM Imputation for Generalized Cross-Validation.
<code>timeDEM</code>	The DEM algorithm execution time.

**See Also**

[EM](#) for the original EM function.

**Examples**

```
# Create a sample dataset with missing values
set.seed(123)
n <- 100
p <- 5
data.sample <- matrix(rnorm(n * p), nrow = n)
data.sample[sample(1:(n*p), 20)] <- NA
data.copy <- data.sample
data0 <- data.frame(data.sample, response = rnorm(n))
mr <- sample(1:n, 10) # Sample rows for evaluation
km <- 3 # Number of clusters
D <- 2 # Number of blocks
# Perform DEM imputation
result <- DEM(data0, data.sample, data.copy, mr, km, D)
# Print the results
print(result$XDEM)
```

---

DRPCA

*Distributed Robust Principal Component Analysis (DRPCA) for Handling Missing Data*

---

**Description**

This function performs DRPCA to handle missing data by dividing the dataset into D blocks, applying the Robust Principal Component Analysis (RPCA) method to each block, and then combining the results. It calculates various evaluation metrics including RMSE, MMAE, RRE, and Generalized Cross-Validation (GCV) using different hierarchical clustering methods.

**Usage**

```
DRPCA(data0, data.sample, data.copy, mr, km, D)
```

**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.
<code>D</code>	The number of blocks to divide the data into.

**Value**

A list containing:

XDRPCA	The imputed dataset.
RMSEDRPCA	The Root Mean Squared Error.
MAEDRPCA	The Mean Absolute Error.
REDRPCA	The Relative Eelative Error.
GCVDRPCA	Distributed DRPCA Imputation for Generalized Cross-Validation.
timeDRPCA	The DRPCA algorithm execution time.

**See Also**

[RPCA](#) for the original RPCA function.

**Examples**

```
# Create a sample dataset with missing values
set.seed(123)
n <- 100
p <- 10
D <- 2
data.sample <- matrix(rnorm(n * p), nrow = n)
data.sample[sample(1:(n-10), (p-2))] <- NA
data.copy <- data.sample
data0 <- data.frame(data.sample, response = rnorm(n))
mr <- sample(1:n, 10) # Sample rows for evaluation
km <- 3 # Number of clusters
result <- DRPCA(data0, data.sample, data.copy, mr, km, D)
#Print the results
print(result$XDRPCA)
```

---

DTSR

---

*Distributed Trimmed Scores Regression (DTSR) for Handling Missing Data*


---

**Description**

This function performs DTSR to handle missing data by dividing the dataset into  $D$  blocks, applying the Trimmed Scores Regression (TSR) method to each block, and then combining the results. It calculates various evaluation metrics including RMSE, MMAE, RRE, and Consistency Proportion Index (CPP) using different hierarchical clustering methods.

**Usage**

```
DTSR(data0, data.sample, data.copy, mr, km, D)
```

**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.
<code>D</code>	The number of blocks to divide the data into.

**Value**

A list containing:

<code>XDTSR</code>	The imputed dataset.
<code>RMSDTSR</code>	The Root Mean Squared Error.
<code>MAEDTSR</code>	The Mean Absolute Error.
<code>REDTSR</code>	The Relative Eelative Error.
<code>GCVDTSR</code>	The DTSR for Generalized Cross-Validation.
<code>timeDTSR</code>	The DTSR algorithm execution time.

**See Also**

[TSR](#) for the original TSR function.

**Examples**

```
# Create a sample matrix with random values and introduce missing values
set.seed(123)
n <- 100
p <- 10
D <- 2
data.sample <- matrix(rnorm(n * p), nrow = n)
data.sample[sample(1:(n-10), (p-2))] <- NA
data.copy <- data.sample
data0 <- data.frame(data.sample, response = rnorm(n))
mr <- sample(1:n, 10) # Sample rows for evaluation
km <- 3 # Number of clusters
# Perform DTSR imputation
result <- DTSR(data0, data.sample, data.copy, mr, km,D)
# Print the results
print(result$XDTSR)
```

**Description**

This function performs Expectation-Maximization (EM) imputation on a dataset with missing values. It uses the ‘imputeEM’ function from the ‘mvdalab’ package to estimate the missing values. The function also calculates various evaluation metrics including RMSE, MMAE, and RRE. Additionally, it performs k-means and hierarchical clustering to assess the quality of the imputation.

**Usage**

```
EM(data0, data.sample, data.copy, mr, km)
```

**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.

**Value**

A list containing:

<code>Xnew</code>	The imputed dataset.
<code>RMSE</code>	The Root Mean Squared Error.
<code>MMAE</code>	The Mean Absolute Error.
<code>RRE</code>	The Relative Erelative Error.
<code>CPP1</code>	The K-means clustering Consistency Proportion Index.
<code>CPP2</code>	The Hierarchical Clustering Complete Linkage Consistency Proportion Index.
<code>CPP3</code>	The Hierarchical Clustering Single Linkage Consistency Proportion Index.
<code>CPP4</code>	The Hierarchical Clustering Average Linkage Consistency Proportion Index.
<code>CPP5</code>	The Hierarchical Clustering Centroid linkage Consistency Proportion Index.
<code>CPP6</code>	The Hierarchical Clustering Median Linkage Consistency Proportion Index.
<code>CPP7</code>	The Hierarchical Clustering Ward’s Method Consistency Proportion Index.
<code>timeEM</code>	The EM algorithm execution time.

## Examples

```
# Create a sample matrix with random values and introduce missing values
set.seed(123)
n <- 100
p <- 5
data.sample <- matrix(rnorm(n * p), nrow = n)
data.sample[sample(1:(n*p), 20)] <- NA
data.copy <- data.sample
data0 <- data.frame(data.sample, response = rnorm(n))
mr <- sample(1:n, 10) # Sample rows for evaluation
km <- 3 # Number of clusters
# Perform EM imputation
result <- EM(data0, data.sample, data.copy, mr, km)
# Print the results
print(result$RMSE)
print(result$MMAE)
print(result$RRE)
print(result$CPP1)
print(result$Xnew)
```

---

IndexCPP

*Calculate the Consistency Proportion Index (CPP)*

---

## Description

This function calculates the Consistency Proportion Index (CPP), a measure of the consistency of clustering results. The CPP is calculated by determining the most common cluster assignment for each group and then computing the proportion of cases that are assigned to these clusters.

## Usage

```
IndexCPP(I)
```

## Arguments

**I** A matrix where each row represents a case and each column represents a cluster assignment. The last column should indicate the group membership (1, 2, or 3).

## Value

A list containing:

**ICPP** The Consistency Proportion Index.

**Examples**

```
# Example usage
set.seed(123)
n <- 100
values1 <- sample(1:3, 30, replace = TRUE)
values2 <- sample(1:3, 30, replace = TRUE) + 1
values3 <- sample(1:3, 40, replace = TRUE) + 2
values <- c(values1, values2, values3)
categories <- c(rep(1, 30), rep(2, 30), rep(3, 40))
I <- cbind(1:n, values, categories)
CPP <- IndexCPP(I)
print(CPP)
```

KNN

*This function performs imputation using the K-Nearest Neighbors (KNN) algorithm and calculates various evaluation metrics including RMSE, MMAE, RRE, and Consistency Proportion Index (CPP) using different hierarchical clustering methods. It also records the execution time of the process.*

**Description**

This function performs imputation using the K-Nearest Neighbors (KNN) algorithm and calculates various evaluation metrics including RMSE, MMAE, RRE, and Consistency Proportion Index (CPP) using different hierarchical clustering methods. It also records the execution time of the process.

**Usage**

```
KNN(data0, data.sample, data.copy, mr, km)
```

**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.

**Value**

A list containing:

<code>Xnew</code>	The imputed dataset.
<code>RMSE</code>	The Root Mean Squared Error.



MMAE	The Mean Absolute Error.
RRE	The Relative Erelative Error.
CPP1	The K-means clustering Consistency Proportion Index.
CPP2	The Hierarchical Clustering Complete Linkage Consistency Proportion Index.
CPP3	The Hierarchical Clustering Single Linkage Consistency Proportion Index.
CPP4	The Hierarchical Clustering Average Linkage Consistency Proportion Index.
CPP5	The Hierarchical Clustering Centroid linkage Consistency Proportion Index.
CPP6	The Hierarchical Clustering Median Linkage Consistency Proportion Index.
CPP7	The Hierarchical Clustering Ward's Method Consistency Proportion Index.
timeKNN	The KNN algorithm execution time.

---

 mean

*Mean Imputation with Evaluation Metrics*


---

### Description

This function performs mean imputation on a dataset with missing values. It replaces missing values with the column means and calculates various evaluation metrics including RMSE, MMAE, and RRE. Additionally, it performs k-means and hierarchical clustering to assess the quality of the imputation.

### Usage

```
mean(data0, data.sample, data.copy, mr, km)
```

### Arguments

data0	The original dataset containing the response variable and features.
data.sample	The dataset used for sampling, which may contain missing values.
data.copy	A copy of the original dataset, used for comparison or validation.
mr	Indices of the rows with missing values that need to be predicted.
km	The number of clusters for k-means clustering.

### Value

A list containing:

Xnew	The imputed dataset.
RMSE	The Root Mean Squared Error.
MMAE	The Mean Absolute Error.
RRE	The Relative Erelative Error.
CPP1	The K-means clustering Consistency Proportion Index.

CPP2	The Hierarchical Clustering Complete Linkage Consistency Proportion Index.
CPP3	The Hierarchical Clustering Single Linkage Consistency Proportion Index.
CPP4	The Hierarchical Clustering Average Linkage Consistency Proportion Index.
CPP5	The Hierarchical Clustering Centroid linkage Consistency Proportion Index.
CPP6	The Hierarchical Clustering Median Linkage Consistency Proportion Index.
CPP7	The Hierarchical Clustering Ward's Method Consistency Proportion Index.
timemean	The mean algorithm execution time.

**See Also**

[kmeans](#) in the stats package for more information on k-means clustering.

[hclust](#) in the stats package for more information on hierarchical clustering.

**Examples**

```
# Create a sample matrix with random values and introduce missing values
set.seed(123)
n <- 100
p <- 5
data.sample <- matrix(rnorm(n * p), nrow = n)
data.sample[sample(1:(n*p), 20)] <- NA
data.copy <- data.sample
data0 <- data.frame(data.sample, response = rnorm(n))
mr <- sample(1:n, 10) # Sample rows for evaluation
km <- 3 # Number of clusters
# Perform mean imputation
result <- mean(data0, data.sample, data.copy, mr, km)
# Print the results
print(result$RMSE)
print(result$MMAE)
print(result$RRE)
print(result$CPP1)
print(result$Xnew)
```

**Description**

This function performs Multilinear Principal Component Analysis (MLPCA) to handle missing data by imputing the missing values based on the correlation structure within the data. It also calculates the RMSE and Consistency Proportion Index (CPP) using different hierarchical clustering methods.

**Usage**

```
MLPCA(data0, data.sample, data.copy, mr, km)
```

**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.

**Value**

A list containing:

<code>Xnew</code>	The imputed dataset.
<code>RMSE</code>	The Root Mean Squared Error.
<code>CPP1</code>	The K-means clustering Consistency Proportion Index.
<code>CPP2</code>	The Hierarchical Clustering Complete Linkage Consistency Proportion Index.
<code>CPP3</code>	The Hierarchical Clustering Single Linkage Consistency Proportion Index.
<code>CPP4</code>	The Hierarchical Clustering Average Linkage Consistency Proportion Index.
<code>CPP5</code>	The Hierarchical Clustering Centroid linkage Consistency Proportion Index.
<code>CPP6</code>	The Hierarchical Clustering Median Linkage Consistency Proportion Index.
<code>CPP7</code>	The Hierarchical Clustering Ward's Method Consistency Proportion Index.
<code>timeKNN</code>	The MLPCA algorithm execution time.

**See Also**

[princomp](#) and [svd](#) for more information on PCA and SVD.

**Examples**

```
# Create a sample matrix with random values and introduce missing values
set.seed(123)
n <- 100
p <- 5
data.sample <- matrix(rnorm(n * p), nrow = n)
data.sample[sample(1:(n*p), 20)] <- NA
data.copy <- data.sample
data0 <- data.frame(data.sample, response = rnorm(n))
mr <- sample(1:n, 10) # Sample rows for evaluation
km <- 3 # Number of clusters
# Perform MLPCA imputation
result <- MLPCA(data0, data.sample, data.copy, mr, km)
# Print the results
print(result$RMSE)
print(result$CPP1)
print(result$Xnew)
```

**Description**

This function performs the NIPALS (Nonlinear Iterative Partial Least Squares) algorithm to handle missing data by imputing the missing values based on the correlation structure within the data. It also calculates the RMSE and Consistency Proportion Index (CPP) using different hierarchical clustering methods.

**Usage**

```
NIPALS(data0, data.sample, data.copy, mr, km)
```

**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.

**Value**

A list containing:

<code>Xnew</code>	The imputed dataset.
<code>RMSE</code>	The Root Mean Squared Error.
<code>CPP1</code>	The K-means clustering Consistency Proportion Index.
<code>CPP2</code>	The Hierarchical Clustering Complete Linkage Consistency Proportion Index.
<code>CPP3</code>	The Hierarchical Clustering Single Linkage Consistency Proportion Index.
<code>CPP4</code>	The Hierarchical Clustering Average Linkage Consistency Proportion Index.
<code>CPP5</code>	The Hierarchical Clustering Centroid linkage Consistency Proportion Index.
<code>CPP6</code>	The Hierarchical Clustering Median Linkage Consistency Proportion Index.
<code>CPP7</code>	The Hierarchical Clustering Ward's Method Consistency Proportion Index.
<code>timeNIPALS</code>	The NIPALS algorithm execution time.

**See Also**

[princomp](#) and [svd](#) for more information on PCA and SVD.

**Examples**

```

# Create a sample matrix with random values and introduce missing values
set.seed(123)
n <- 100
p <- 5
data.sample <- matrix(rnorm(n * p), nrow = n)
data.sample[sample(1:(n*p), 20)] <- NA
data.copy <- data.sample
data0 <- data.frame(data.sample, response = rnorm(n))
mr <- sample(1:n, 10) # Sample rows for evaluation
km <- 3 # Number of clusters
# Perform NIPALS imputation
result <- NIPALS(data0, data.sample, data.copy, mr, km)
# Print the results
print(result$RMSE)
print(result$CPP1)
print(result$Xnew)

```

RPCA

*Robust Principal Component Analysis with Missing Data***Description**

This function performs Robust Principal Component Analysis (RPCA) to handle missing data by imputing the missing values based on the correlation structure within the data. It also calculates various evaluation metrics including RMSE, MMAE, RRE, and Consistency Proportion Index (CPP) using different hierarchical clustering methods.

**Usage**

```
RPCA(data0, data.sample, data.copy, mr, km)
```

**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.

**Value**

A list containing:

<code>Xnew</code>	The imputed dataset.
<code>RMSE</code>	The Root Mean Squared Error.

MMAE	The Mean Absolute Error.
RRE	The Relative Relative Error.
CPP1	The K-means clustering Consistency Proportion Index.
CPP2	The Hierarchical Clustering Complete Linkage Consistency Proportion Index.
CPP3	The Hierarchical Clustering Single Linkage Consistency Proportion Index.
CPP4	The Hierarchical Clustering Average Linkage Consistency Proportion Index.
CPP5	The Hierarchical Clustering Centroid linkage Consistency Proportion Index.
CPP6	The Hierarchical Clustering Median Linkage Consistency Proportion Index.
CPP7	The Hierarchical Clustering Ward's Method Consistency Proportion Index.
timeRPCA	The RPCA algorithm execution time.

**See Also**

[princomp](#) and [svd](#) for more information on PCA and SVD.

**Examples**

```
# Create a sample matrix with random values and introduce missing values
set.seed(123)
n <- 100
p <- 5
data.sample <- matrix(rnorm(n * p), nrow = n)
data.sample[sample(1:(n*p), 20)] <- NA
data.copy <- data.sample
data0 <- data.frame(data.sample, response = rnorm(n))
mr <- sample(1:n, 10) # Sample rows for evaluation
km <- 3 # Number of clusters
# Perform RPCA imputation
result <- RPCA(data0, data.sample, data.copy, mr, km)
# Print the results
print(result$RMSE)
print(result$MMAE)
print(result$RRE)
print(result$CPP1)
print(result$Xnew)
```

---

SVD

*This function performs imputation using Singular Value Decomposition (SVD) and calculates various evaluation metrics including RMSE, MMAE, RRE, and Consistency Proportion Index (CPP) using different hierarchical clustering methods.*

---

**Description**

This function performs imputation using Singular Value Decomposition (SVD) and calculates various evaluation metrics including RMSE, MMAE, RRE, and Consistency Proportion Index (CPP) using different hierarchical clustering methods.

**Usage**

```
SVD(data0, data.sample, data.copy, mr, km)
```

**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.

**Value**

A list containing:

<code>Xnew</code>	The imputed dataset.
<code>RMSE</code>	The Root Mean Squared Error.
<code>MMAE</code>	The Mean Absolute Error.
<code>RRE</code>	The Relative Erelative Error.
<code>CPP1</code>	The K-means clustering Consistency Proportion Index.
<code>CPP2</code>	The Hierarchical Clustering Complete Linkage Consistency Proportion Index.
<code>CPP3</code>	The Hierarchical Clustering Single Linkage Consistency Proportion Index.
<code>CPP4</code>	The Hierarchical Clustering Average Linkage Consistency Proportion Index.
<code>CPP5</code>	The Hierarchical Clustering Centroid linkage Consistency Proportion Index.
<code>CPP6</code>	The Hierarchical Clustering Median Linkage Consistency Proportion Index.
<code>CPP7</code>	The Hierarchical Clustering Ward's Method Consistency Proportion Index.
<code>timeSVD</code>	The SVD algorithm execution time.

**See Also**

[princomp](#) and [svd](#) for more information on PCA and SVD.

---

SVDImpute

*Improved SVD Imputation*

---

**Description**

This function performs imputation using Singular Value Decomposition (SVD) with iterative refinement. It begins by filling missing values with the mean of their respective columns. Then, it computes a low-rank ( $k$ ) approximation of the data matrix. Using this approximation, it refills the missing values. This process of recomputing the rank- $k$  approximation with the newly imputed values and refilling the missing data is repeated for a specified number of iterations, 'num.iters'.

**Usage**

```
SVDImpute(x, k, num.its = 10, verbose = TRUE)
```

**Arguments**

<code>x</code>	A data frame or matrix where each row represents a different record.
<code>k</code>	The rank-k approximation to use for the data matrix.
<code>num.its</code>	The number of times to compute the rank-k approximation and impute the missing data.
<code>verbose</code>	If TRUE, print status updates during the process.

**Value**

A list containing:

<code>data.matrix</code>	The imputed matrix with missing values filled.
--------------------------	--

**Examples**

```
# Create a sample matrix with random values and introduce missing values
x = matrix(rnorm(100), 10, 10)
x[x > 1] = NA

# Perform SVD imputation
imputed_x = SVDImpute(x, 3)

# Print the imputed matrix
print(imputed_x)
```

---

TSR

*Trimmed Scores Regression with Missing Data*

---

**Description**

This function performs Trimmed Scores Regression (TSR) to handle missing data by imputing the missing values based on the correlation structure within the data. It also calculates various evaluation metrics including RMSE, MMAE, RRE, and Consistency Proportion Index (CPP) using different hierarchical clustering methods.

**Usage**

```
TSR(data0, data.sample, data.copy, mr, km)
```



**Arguments**

<code>data0</code>	The original dataset containing the response variable and features.
<code>data.sample</code>	The dataset used for sampling, which may contain missing values.
<code>data.copy</code>	A copy of the original dataset, used for comparison or validation.
<code>mr</code>	Indices of the rows with missing values that need to be predicted.
<code>km</code>	The number of clusters for k-means clustering.

**Value**

A list containing:

<code>Xnew</code>	The imputed dataset.
<code>RMSE</code>	The Root Mean Squared Error.
<code>MMAE</code>	The Mean Absolute Error.
<code>RRE</code>	The Relative Relative Error.
<code>CPP1</code>	The K-means clustering Consistency Proportion Index.
<code>CPP2</code>	The Hierarchical Clustering Complete Linkage Consistency Proportion Index.
<code>CPP3</code>	The Hierarchical Clustering Single Linkage Consistency Proportion Index.
<code>CPP4</code>	The Hierarchical Clustering Average Linkage Consistency Proportion Index.
<code>CPP5</code>	The Hierarchical Clustering Centroid linkage Consistency Proportion Index.
<code>CPP6</code>	The Hierarchical Clustering Median Linkage Consistency Proportion Index.
<code>CPP7</code>	The Hierarchical Clustering Ward's Method Consistency Proportion Index.
<code>timeTSR</code>	The TSR algorithm execution time.

**See Also**

[princomp](#) and [svd](#) for more information on PCA and SVD.

**Examples**

```
# Create a sample matrix with random values and introduce missing values
set.seed(123)
n <- 100
p <- 5
data.sample <- matrix(rnorm(n * p), nrow = n)
data.sample[sample(1:(n*p), 20)] <- NA
data.copy <- data.sample
data0 <- data.frame(data.sample, response = rnorm(n))
mr <- sample(1:n, 10) # Sample rows for evaluation
km <- 3 # Number of clusters
# Perform TSR imputation
result <- TSR(data0, data.sample, data.copy, mr, km)
# Print the results
print(result$RMSE)
print(result$MMAE)
print(result$RRE)
```

```
print(result$CPP1)
print(result$Xnew)
```

# Index

- \* **DEM**
  - DEM, [2](#)
- \* **DRPCA**
  - DRPCA, [3](#)
- \* **DTSR**
  - DTSR, [4](#)
- \* **EM**
  - DEM, [2](#)
- \* **KNN**
  - KNN, [8](#)
- \* **MLPCA**
  - MLPCA, [10](#)
- \* **NIPALS**
  - NIPALS, [12](#)
- \* **PCA**
  - DRPCA, [3](#)
  - DTSR, [4](#)
  - MLPCA, [10](#)
  - NIPALS, [12](#)
  - RPCA, [13](#)
  - SVD, [14](#)
  - TSR, [16](#)
- \* **RPCA**
  - DRPCA, [3](#)
  - NIPALS, [12](#)
  - RPCA, [13](#)
- \* **SVD**
  - DRPCA, [3](#)
  - DTSR, [4](#)
  - MLPCA, [10](#)
  - NIPALS, [12](#)
  - RPCA, [13](#)
  - SVD, [14](#)
  - TSR, [16](#)
- \* **TSR**
  - DTSR, [4](#)
  - TSR, [16](#)
- \* **clustering**
  - DEM, [2](#)
  - IndexCPP, [7](#)
- \* **consistency**
  - IndexCPP, [7](#)
- \* **evaluation**
  - mean, [9](#)
- \* **imputation**
  - DEM, [2](#)
  - DRPCA, [3](#)
  - DTSR, [4](#)
  - KNN, [8](#)
  - mean, [9](#)
  - MLPCA, [10](#)
  - NIPALS, [12](#)
  - RPCA, [13](#)
  - SVD, [14](#)
  - TSR, [16](#)
- \* **k-means**
  - DEM, [2](#)
  - DEM, [2](#)
  - DRPCA, [3](#)
  - DTSR, [4](#)
  - EM, [3, 6](#)
  - hclust, [10](#)
  - IndexCPP, [7](#)
  - kmeans, [10](#)
  - KNN, [8](#)
  - mean, [9](#)
  - MLPCA, [10](#)
  - NIPALS, [12](#)
  - princomp, [11, 12, 14, 15, 17](#)
  - RPCA, [4, 13](#)
  - SVD, [14](#)

svd, [11](#), [12](#), [14](#), [15](#), [17](#)

SVDImpute, [15](#)

TSR, [5](#), [16](#)