

# Package ‘xtsum’

December 7, 2023

**Type** Package

**Title** Summary Statistics for Panel Data

**Version** 0.1.0

**Depends** R (>= 3.2.0), knitr, magrittr, rlang, plm

**Imports** dplyr, kableExtra, sampleSelection

**Author** Joao Claudio Macosso

**Maintainer** Joao Claudio Macosso <joaoclaudiomacosso@gmail.com>

**URL** <https://github.com/Macosso/xtsum>

**BugReports** <https://github.com/Macosso/xtsum/issues>

**VignetteBuilder** knitr

**Description** Based on 'STATA' xtsum command, it is used to compute summary statistics for a panel data set. It generates overall, between-group, and within-group statistics for specified variables in a panel data set, as presented in S. Porter (2023) <[https://stephenporter.org/files/xtsum\\_handout.pdf](https://stephenporter.org/files/xtsum_handout.pdf)>, Stat-aCorp (2023) <<https://www.stata.com/manuals/xtxtsum.pdf>>.

**License** GPL-3

**Encoding** UTF-8

**RoxygenNote** 7.2.3

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2023-12-07 16:20:02 UTC

## R topics documented:

between_max	2
between_min	3
between_sd	3

within_max	4
within_min	5
within_sd	6
xtsum	6

<b>Index</b>	<b>8</b>
--------------	----------

---

between_max	<i>Compute the maximum between-group</i>
-------------	--

---

### Description

This function calculates the maximum between-group in a panel data.

### Usage

```
between_max(data, variable, id = NULL, t = NULL, na.rm = FALSE)
```

### Arguments

data	A data.frame or pdata.frame object containing the panel data.
variable	The variable for which the maximum between-group effect is calculated.
id	(Optional) Name of the individual identifier variable.
t	(Optional) Name of the time identifier variable.
na.rm	Logical. Should missing values be removed? Default is FALSE.

### Value

The maximum between-group effect.

### Examples

```
# Example using pdata.frame
data("Gasoline", package = "plm")
Gas <- pdata.frame(Gasoline, index = c("country", "year"), drop.index = TRUE)
between_max(Gas, variable = "lgaspcar")

# Using regular data.frame with id and t specified
data("Crime", package = "plm")
between_max(Crime, variable = "crmrte", id = "county", t = "year")
```

---

between_min	<i>Compute the minimum between-group</i>
-------------	--

---

**Description**

This function calculates the minimum between-group of a panel data.

**Usage**

```
between_min(data, variable, id = NULL, t = NULL, na.rm = FALSE)
```

**Arguments**

data	A data.frame or pdata.frame object containing the panel data.
variable	The variable for which the minimum between-group effect is calculated.
id	(Optional) Name of the individual identifier variable.
t	(Optional) Name of the time identifier variable.
na.rm	Logical. Should missing values be removed? Default is FALSE.

**Value**

The minimum between-group effect.

**Examples**

```
# Example using pdata.frame
data("Gasoline", package = "plm")
Gas <- pdata.frame(Gasoline, index = c("country", "year"), drop.index = TRUE)
between_min(Gas, variable = "lgaspcar")

# Using regular data.frame with id and t specified
data("Crime", package = "plm")
between_min(Crime, variable = "crmrt", id = "county", t = "year")
```

---

between_sd	<i>Compute the standard deviation of between-group</i>
------------	--

---

**Description**

This function calculates the standard deviation of between-group in a panel data.

**Usage**

```
between_sd(data, variable, id = NULL, t = NULL, na.rm = FALSE)
```

**Arguments**

data	A data.frame or pdata.frame object containing the panel data.
variable	The variable for which the standard deviation of between-group effects is calculated.
id	(Optional) Name of the individual identifier variable.
t	(Optional) Name of the time identifier variable.
na.rm	Logical. Should missing values be removed? Default is FALSE.

**Value**

The standard deviation of between-group effects.

**Examples**

```
# Example using pdata.frame
data("Gasoline", package = "plm")
Gas <- pdata.frame(Gasoline, index = c("country", "year"), drop.index = TRUE)
between_sd(Gas, variable = "lgaspcar")

# Using regular data.frame with id and t specified
data("Crime", package = "plm")
between_sd(Crime, variable = "crrmrte", id = "county", t = "year")
```

---

within\_max

*Compute the maximum within-group for a panel data*

---

**Description**

This function computes the maximum within-group for a panel data.

**Usage**

```
within_max(data, variable, id = NULL, t = NULL, na.rm = FALSE)
```

**Arguments**

data	A data.frame or pdata.frame object containing the panel data.
variable	The variable for which the maximum within-group effect is calculated.
id	(Optional) Name of the individual identifier variable.
t	(Optional) Name of the time identifier variable.
na.rm	Logical. Should missing values be removed? Default is FALSE.

**Value**

The maximum within-group effect.

**Examples**

```
# Example using pdata.frame
data("Gasoline", package = "plm")
Gas <- pdata.frame(Gasoline, index = c("country", "year"), drop.index = TRUE)
within_max(Gas, variable = "lgaspcar")

# Using regular data.frame with id and t specified
data("Crime", package = "plm")
within_max(Crime, variable = "crmrte", id = "county", t = "year")
```

within\_min

*Compute the minimum within-group for panel data***Description**

This function computes the minimum within-group for a panel data.

**Usage**

```
within_min(data, variable, id = NULL, t = NULL, na.rm = FALSE)
```

**Arguments**

data	A data.frame or pdata.frame object containing the panel data.
variable	The variable for which the minimum within-group effect is calculated.
id	(Optional) Name of the individual identifier variable.
t	(Optional) Name of the time identifier variable.
na.rm	Logical. Should missing values be removed? Default is FALSE.

**Value**

The minimum within-group effect.

**Examples**

```
# Example using pdata.frame
data("Gasoline", package = "plm")
Gas <- pdata.frame(Gasoline, index = c("country", "year"), drop.index = TRUE)
within_min(Gas, variable = "lgaspcar")

# Using regular data.frame with id and t specified
data("Crime", package = "plm")
within_min(Crime, variable = "crmrte", id = "county", t = "year")
```

---

within_sd	<i>Compute the standard deviation of within-group for a panel data</i>
-----------	--

---

**Description**

This function computes the standard deviation of within-group for a panel data.

**Usage**

```
within_sd(data, variable, id = NULL, t = NULL, na.rm = FALSE)
```

**Arguments**

data	A data.frame or pdata.frame object containing the panel data.
variable	The variable for which the standard deviation of within-group effects is calculated.
id	(Optional) Name of the individual identifier variable.
t	(Optional) Name of the time identifier variable.
na.rm	Logical. Should missing values be removed? Default is FALSE.

**Value**

The standard deviation of within-group effects.

**Examples**

```
# Example using pdata.frame
data("Gasoline", package = "plm")
Gas <- pdata.frame(Gasoline, index = c("country", "year"), drop.index = TRUE)
within_sd(Gas, variable = "lgaspcar")

# Using regular data.frame with id and t specified
data("Crime", package = "plm")
within_sd(Crime, variable = "crm rte", id = "county", t = "year")
```

---

xtsum	<i>Calculate summary statistics for panel data</i>
-------	--

---

**Description**

This function computes summary statistics for panel data, including overall statistics, between-group statistics, and within-group statistics.

**Usage**

```
xtsum(  
  data,  
  variables = NULL,  
  id = NULL,  
  t = NULL,  
  na.rm = FALSE,  
  return.data.frame = FALSE,  
  dec = 3  
)
```

**Arguments**

data	A data.frame or pdata.frame object representing panel data.
variables	(Optional) Vector of variable names for which to calculate statistics. If not provided, all numeric variables in the data will be used.
id	(Optional) Name of the individual identifier variable.
t	(Optional) Name of the time identifier variable.
na.rm	Logical indicating whether to remove NAs when calculating statistics.
return.data.frame	If the return object should be a dataframe
dec	Number of significant digits to report

**Value**

A table summarizing statistics for each variable, including Mean, SD, Min, and Max, broken down into Overall, Between, and Within dimensions.

**Examples**

```
# Using a data.frame and specifying variables, id, it, na.rm, dec  
data("nlswork", package = "sampleSelection")  
xtsum(nlswork, "hours", id = "idcode", t = "year", na.rm = TRUE, dec = 6)  
  
# Using pdata.frame object without specifying a variable  
data("Gasoline", package = "plm")  
Gas <- pdata.frame(Gasoline, index = c("country", "year"), drop.index = TRUE)  
xtsum(Gas)  
  
# Using regular data.frame with id and t specified  
data("Crime", package = "plm")  
xtsum(Crime, variables = c("crmrte", "prbarr"), id = "county", t = "year")  
  
# Specifying variables to include in the summary  
xtsum(Gas, variables = c("lincomep", "lgaspcar"))
```

# Index

between\_max, 2  
between\_min, 3  
between\_sd, 3

within\_max, 4  
within\_min, 5  
within\_sd, 6

xtsum, 6