Gráfico

Descripción generada automáticamente

Guide

**Guía de Usuario del paquete de R #rstats {MATdatatools}, Versión 0.1.0**

**Miguel Ángel Tarancón Morán**

Catedrático de Economía Aplicada. Universidad de Castilla – La Mancha

Dibujo en blanco y negro

Descripción generada automáticamente con confianza media

MATdatatools © 2025 by [Miguel-Ángel Tarancón](http://www.miguelangeltarancon.com) is licensed under [Creative Commons Attribution-NonCommercial 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/?ref=chooser-v1)

MATdatatools Guide

Miguel-Ángel Tarancón

2025-01-15

Tabla de contenidos

[1 Introducción 3](#_Toc190449977)

[2 Funciones de MATdatatools 3](#_Toc190449978)

[3 Descarga e Instalación 3](#_Toc190449979)

[3.1 Desde GitHub 3](#_Toc190449980)

[3.2 En RStudio 4](#_Toc190449981)

[4 Detalle de cada Función 4](#_Toc190449982)

[4.1 MATfexcel 4](#_Toc190449983)

[4.1.1 Uso 4](#_Toc190449984)

[4.1.2 Objetivo 4](#_Toc190449985)

[4.1.3 Argumentos 4](#_Toc190449987)

[4.1.4 Salida 4](#_Toc190449992)

[4.1.5 Ejemplo 5](#_Toc190449993)

[4.1.6 Resultados Esperados 5](#_Toc190449994)

[4.2 MATmv 6](#_Toc190449995)

[4.2.1 Uso 6](#_Toc190449996)

[4.2.2 Objetivo 6](#_Toc190449997)

[4.2.3 Argumentos 6](#_Toc190449998)

[4.2.4 Salida 6](#_Toc190449999)

[4.2.5 Ejemplo 6](#_Toc190450003)

[4.2.6 Resultados Esperados 6](#_Toc190450004)

[4.3 MATout 8](#_Toc190450005)

[4.3.1 Uso 8](#_Toc190450006)

[4.3.2 Objetivo 8](#_Toc190450007)

[4.3.3 Argumentos 8](#_Toc190450008)

[4.3.4 Salida 8](#_Toc190450009)

[4.3.5 Ejemplo 8](#_Toc190450010)

[4.3.6 Resultados Esperados 8](#_Toc190450011)

[4.4 MATout\_Mahalanobis 11](#_Toc190450012)

[4.4.1 Uso 11](#_Toc190450013)

[4.4.2 Objetivo 11](#_Toc190450014)

[4.4.3 Argumentos 11](#_Toc190450016)

[4.4.4 Salida 11](#_Toc190450017)

[4.4.5 Ejemplo 11](#_Toc190450022)

[4.4.6 Resultados Esperados 11](#_Toc190450023)

[4.5 MATdescribe 13](#_Toc190450024)

[4.5.1 Uso 13](#_Toc190450025)

[4.5.2 Objetivo 13](#_Toc190450026)

[4.5.3 Argumentos 13](#_Toc190450028)

[4.5.4 Salida 14](#_Toc190450029)

[4.5.5 Ejemplo 14](#_Toc190450034)

[4.5.6 Resultados Esperados 14](#_Toc190450035)

[4.6 MATtaf 15](#_Toc190450036)

[4.6.1 Uso 15](#_Toc190450037)

[4.6.2 Objetivo 15](#_Toc190450038)

[4.6.3 Argumentos 15](#_Toc190450039)

[4.6.4 Salida 15](#_Toc190450040)

[4.6.5 Ejemplo 15](#_Toc190450041)

[4.6.6 Resultados Esperados 15](#_Toc190450042)

[5 Documentación del Paquete 16](#_Toc190450043)

[6 ¡Empieza a explorar tus datos con MATdatatools! 17](#_Toc190450044)

# 

# 1 Introducción

El paquete **MATdatatools** ha sido diseñado para facilitar el análisis de datos y la generación de resultados gráficos y estadísticos en R. Este paquete es una herramienta ágil y accesible, especialmente útil para estudiantes y profesionales de Administración y Dirección de Empresas que buscan una manera intuitiva de explorar, analizar y visualizar sus datos sin necesidad de profundos conocimientos de programación.

**¿Qué ofrece MATdatatools?**

* Automatización de procesos estadísticos comunes.
* Gráficos de alta calidad con interpretaciones claras.
* Análisis descriptivo univariante.

# 2 Funciones de MATdatatools

El paquete incluye las siguientes funciones principales:

* **MATfexcel:** Importa datos desde hojas de Excel, convirtiendo fácilmente las primeras columnas en nombres de filas.
* **MATmv:** Filtra datos con casos completos, muestra gráficamente los datos faltantes y genera un resumen.
* **MATout:** Detecta y elimina valores atípicos en una variable, proporcionando gráficos y tablas explicativas.
* **MATout\_Mahalanobis:** Identifica valores atípicos multidimensionales basándose en la distancia de Mahalanobis.
* **MATdescribe:** Realiza un análisis descriptivo completo, incluyendo gráficos y estadísticos.
* **MATtaf:** Genera tablas de frecuencias agrupadas en intervalos y crea histogramas relacionados.

Cada función está diseñada para ser intuitiva y generar resultados listos para interpretar. El archivo de Microsoft® Excel® de los ejemplos, eolica\_100\_mv.xls, puede descargarse [aquí](https://docs.google.com/spreadsheets/d/1f9KV5epNpBRM6v9sKPtgFDkzMlOo75Y2/edit?usp=sharing&ouid=115375878280465826079&rtpof=true&sd=true).

# 3 Descarga e Instalación

## 3.1 Desde GitHub

1. Asegúrate de tener instalado el paquete devtools:

* install.packages("devtools")

1. Descarga e instala el paquete desde GitHub:

* devtools::install\_github("teckel71/R\_for\_Economics/packages/MATdatatools")

1. Carga el paquete:

* library(MATdatatools)

## 3.2 En RStudio

1. Abre RStudio y ve a **Tools > Install Packages**.
2. Selecciona **Install from GitHub** y escribe:

* teckel71/R\_for\_Economics/packages/MATdatatools

1. Haz clic en **Install**.

# 4 Detalle de cada Función

## 4.1 MATfexcel

### 4.1.1 Uso

MATfexcel(file\_path, sheet\_name, na\_values = NULL, viz = FALSE)

### 4.1.2 Objetivo

### Esta función carga datos desde un archivo Excel, convirtiendo la primera columna en nombres de fila y generando un resumen.

### 4.1.3 Argumentos

### file\_path: Ruta al archivo Excel.

### sheet\_name: Nombre o índice de la hoja a importar.

### na\_values: Vector de caracteres que representan valores NA.

### viz: Si es TRUE, genera una visualización automática de las variables. Es necesario que no existan NAs.

### 4.1.4 Salida

* Dataframe con los datos importados.
* Si viz = TRUE, genera un objeto con visualización de las variables (<nombre\_dataframe>\_viz). Es necesario que no existan NAs.

### 4.1.5 Ejemplo

rm(list = ls())

**library**(MATdatatools)

eolica\_100 <- MATfexcel("eolica\_100\_mv.xlsx", "Datos",

na\_values = c("n.d.", "s.d."),

viz = FALSE)

### 4.1.6 Resultados Esperados

* Un dataframe llamado eolica\_100

## RES ACTIVO FPIOS   
## Min. : -5661.5 Min. : 24944 Min. : -77533   
## 1st Qu.: 669.5 1st Qu.: 34547 1st Qu.: 2305   
## Median : 2084.5 Median : 46950 Median : 11936   
## Mean : 11529.8 Mean : 277270 Mean : 123743   
## 3rd Qu.: 3806.7 3rd Qu.: 85610 3rd Qu.: 28292   
## Max. :727548.0 Max. :13492812 Max. :6904824   
## NA's :1 NA's :1   
##   
## RENECO RENFIN LIQUIDEZ   
## Min. :-2.813 Min. :-359.773 Min. : 0.0140   
## 1st Qu.: 1.558 1st Qu.: 2.556 1st Qu.: 0.6567   
## Median : 4.236 Median : 15.326 Median : 1.0650   
## Mean : 5.416 Mean : 17.243 Mean : 2.7214   
## 3rd Qu.: 7.970 3rd Qu.: 31.307 3rd Qu.: 1.6078   
## Max. :35.262 Max. : 588.190 Max. :128.4330   
## NA's :2   
##   
## ENDEUDA MARGEN SOLVENCIA   
## Min. : 0.917 Min. :-2248.157 Min. :-40.74   
## 1st Qu.: 50.852 1st Qu.: 12.316 1st Qu.: 4.71   
## Median : 83.346 Median : 26.618 Median : 16.65   
## Mean : 72.227 Mean : 3.228 Mean : 27.57   
## 3rd Qu.: 95.388 3rd Qu.: 39.590 3rd Qu.: 45.59   
## Max. :140.745 Max. : 400.899 Max. : 99.08   
## NA's :2 NA's :2   
##   
## APALANCA MATRIZ DIMENSION   
## Min. :-8254.11 Length:100 Length:100   
## 1st Qu.: 16.13 Class :character Class :character   
## Median : 161.97 Mode :character Mode :character   
## Mean : 345.03   
## 3rd Qu.: 623.13   
## Max. :12244.35

## 4.2 MATmv

### 4.2.1 Uso

MATmv(dataframe, columnas)

### 4.2.2 Objetivo

Selecciona variables de un dataframe, identifica casos completos y genera un resumen de valores faltantes.

### 4.2.3 Argumentos

* dataframe: El dataframe que se desea analizar.
* columnas: Las columnas (variables) que se desean evaluar, separadas por comas (sin comillas). Si no se especifican las variables, se procesarán todas las variables del dataframe.

### 4.2.4 Salida

* Un nuevo dataframe con los casos completos de las columnas seleccionadas.

### Una lista con:

### Tabla de valores faltantes.

### Gráfico de valores ausentes.

### 4.2.5 Ejemplo

MATmv(eolica\_100, c(RENECO, ACTIVO, RES, RENFIN, FPIOS, MARGEN))

### 4.2.6 Resultados Esperados

1. Un dataframe llamado eolica\_100\_sm con los casos completos en las variables seleccionadas.
2. Una lista eolica\_100\_sm\_info que contiene:
   * Gráfico que visualiza los valores faltantes.
   * Tabla con los casos faltantes.

## RES ACTIVO FPIOS   
## Min. : -5661.5 Min. : 24944 Min. : -77533   
## 1st Qu.: 739.6 1st Qu.: 34249 1st Qu.: 2380   
## Median : 2114.7 Median : 46653 Median : 11936   
## Mean : 12080.6 Mean : 285091 Mean : 128818   
## 3rd Qu.: 3844.2 3rd Qu.: 83091 3rd Qu.: 27991   
## Max. :727548.0 Max. :13492812 Max. :6904824   
##   
## RENECO RENFIN LIQUIDEZ   
## Min. :-2.813 Min. :-359.773 Min. : 0.0140   
## 1st Qu.: 1.558 1st Qu.: 2.212 1st Qu.: 0.6675   
## Median : 4.236 Median : 15.924 Median : 1.0795   
## Mean : 5.478 Mean : 17.906 Mean : 2.8401   
## 3rd Qu.: 8.107 3rd Qu.: 34.167 3rd Qu.: 1.6350   
## Max. :35.262 Max. : 588.190 Max. :128.4330   
##   
## ENDEUDA MARGEN SOLVENCIA   
## Min. : 0.917 Min. :-2248.157 Min. :-40.745   
## 1st Qu.: 54.406 1st Qu.: 12.793 1st Qu.: 4.779   
## Median : 83.346 Median : 27.638 Median : 16.653   
## Mean : 72.002 Mean : 9.393 Mean : 27.773   
## 3rd Qu.: 95.289 3rd Qu.: 41.264 3rd Qu.: 43.812   
## Max. :140.745 Max. : 400.899 Max. : 99.082   
## NA's :2   
##   
## APALANCA MATRIZ DIMENSION   
## Min. :-8254.11 Length:94 Length:94   
## 1st Qu.: 25.71 Class :character Class :character   
## Median : 223.21 Mode :character Mode :character   
## Mean : 378.36   
## 3rd Qu.: 670.30   
## Max. :12244.35

eolica\_100\_sm\_info$grafico\_vis\_miss

Texto

Descripción generada automáticamente

eolica\_100\_sm\_info$tabla\_na

Tabla

Descripción generada automáticamente

## 4.3 MATout

### 4.3.1 Uso

MATout(data, variable)

### 4.3.2 Objetivo

Detectar valores atípicos unidimensionales basados en el rango intercuartílico (IQR).

### 4.3.3 Argumentos

* data: El dataframe que contiene los datos.
* variable: La variable numérica a analizar (sin comillas).

### 4.3.4 Salida

* Identifica valores atípicos y los muestra en un boxplot.
* Filtra los datos sin valores atípicos.

### 4.3.5 Ejemplo

MATout(data = eolica\_100\_sm, variable = RENECO)

### 4.3.6 Resultados Esperados

1. Un dataframe llamado eolica\_100\_sm\_so con los datos filtrados.
2. Una lista eolica\_100\_sm\_so\_info que contiene:
   * Box-Plot de la variable analizada.
   * Tabla de valores atípicos.

## RES ACTIVO FPIOS   
## Min. : -5661.5 Min. : 24944 Min. : -77533   
## 1st Qu.: 718.8 1st Qu.: 34437 1st Qu.: 2305   
## Median : 2016.7 Median : 46653 Median : 10870   
## Mean : 12109.0 Mean : 290312 Mean : 131112   
## 3rd Qu.: 3666.9 3rd Qu.: 84061 3rd Qu.: 28292   
## Max. :727548.0 Max. :13492812 Max. :6904824   
##   
## RENECO RENFIN LIQUIDEZ   
## Min. :-2.813 Min. :-359.773 Min. : 0.0140   
## 1st Qu.: 1.421 1st Qu.: 1.951 1st Qu.: 0.6915   
## Median : 4.144 Median : 15.460 Median : 1.1115   
## Mean : 4.977 Mean : 17.078 Mean : 2.8952   
## 3rd Qu.: 7.904 3rd Qu.: 33.163 3rd Qu.: 1.6567   
## Max. :15.882 Max. : 588.190 Max. :128.4330   
##

## ENDEUDA MARGEN SOLVENCIA   
## Min. : 0.917 Min. :-2248.157 Min. :-40.74   
## 1st Qu.: 56.187 1st Qu.: 12.493 1st Qu.: 4.71   
## Median : 83.648 Median : 27.409 Median : 16.35   
## Mean : 72.646 Mean : 8.631 Mean : 27.14   
## 3rd Qu.: 95.388 3rd Qu.: 39.580 3rd Qu.: 43.80   
## Max. :140.745 Max. : 400.899 Max. : 99.08   
## NA's :2   
##   
## APALANCA MATRIZ DIMENSION   
## Min. :-8254.11 Length:92 Length:92   
## 1st Qu.: 26.17 Class :character Class :character   
## Median : 233.75 Mode :character Mode :character   
## Mean : 385.76   
## 3rd Qu.: 693.49   
## Max. :12244.35

eolica\_100\_sm\_so\_info$Boxplot

Gráfico

Descripción generada automáticamente

eolica\_100\_sm\_so\_info$Outliers\_Table

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente con confianza media

### 

## 4.4 MATout\_Mahalanobis

### 4.4.1 Uso

MATout\_Mahalanobis(data, variables)

### 4.4.2 Objetivo

### Usa la distancia de Mahalanobis para detectar y eliminar outliers multivariados en un conjunto de variables.

### 4.4.3 Argumentos

* data: El dataframe que contiene los datos.
* variables: Variables a incluir en el análisis (sin comillas, separadas por comas).

### 4.4.4 Salida

### Un dataframe sin outliers (<nombre\_original>\_so).

### Una lista con:

### Tabla de outliers detectados.

### Boxplot de distancias de Mahalanobis.

### 4.4.5 Ejemplo

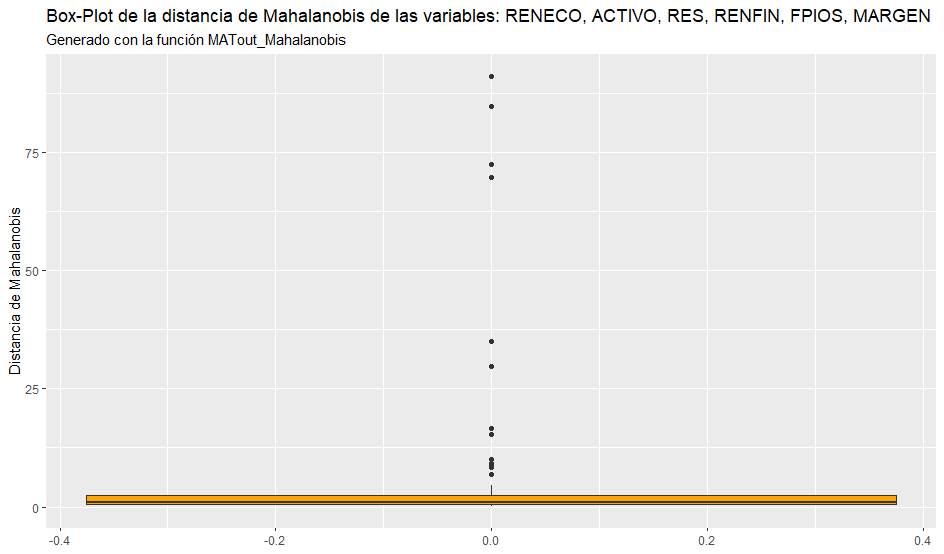
MATout\_Mahalanobis(data = eolica\_100\_sm,  
 variables = c(RENECO, ACTIVO, RES, RENFIN, FPIOS, MARGEN))

### 4.4.6 Resultados Esperados

1. Un dataframe llamado eolica\_100\_sm\_so con los datos filtrados.
2. Una lista eolica\_100\_sm\_so\_info que contiene:
   * Tabla de valores atípicos.
   * Gráficos de las distancias de Mahalanobis.

## RES ACTIVO FPIOS   
## Min. :-5661.5 Min. : 24944 Min. :-77533   
## 1st Qu.: 650.7 1st Qu.: 33607 1st Qu.: 2842   
## Median : 1949.0 Median : 43998 Median : 10980   
## Mean : 3253.1 Mean : 89534 Mean : 28365   
## 3rd Qu.: 3497.5 3rd Qu.: 79873 3rd Qu.: 25995   
## Max. :67033.0 Max. :1275939 Max. :726783   
##   
## RENECO RENFIN LIQUIDEZ   
## Min. :-2.813 Min. :-165.348 Min. : 0.029   
## 1st Qu.: 1.424 1st Qu.: 6.904 1st Qu.: 0.690   
## Median : 4.237 Median : 16.684 Median : 1.177   
## Mean : 5.236 Mean : 19.599 Mean : 3.129   
## 3rd Qu.: 8.560 3rd Qu.: 34.669 3rd Qu.: 1.648   
## Max. :15.882 Max. : 207.801 Max. :128.433   
##   
## ENDEUDA MARGEN SOLVENCIA   
## Min. : 0.917 Min. :-302.03 Min. :-40.745   
## 1st Qu.: 56.193 1st Qu.: 11.66 1st Qu.: 5.981   
## Median : 83.098 Median : 28.52 Median : 16.901   
## Mean : 72.071 Mean : 34.10 Mean : 27.670   
## 3rd Qu.: 94.059 3rd Qu.: 42.02 3rd Qu.: 43.796   
## Max. :140.745 Max. : 400.90 Max. : 99.082   
## NA's :2   
##   
## APALANCA MATRIZ DIMENSION   
## Min. :-7770.00 Length:81 Length:81   
## 1st Qu.: 31.93 Class :character Class :character   
## Median : 239.37 Mode :character Mode :character   
## Mean : 435.09   
## 3rd Qu.: 684.22   
## Max. : 8049.39

eolica\_100\_sm\_so\_info$Boxplot



eolica\_100\_sm\_so\_info$Outliers\_Table

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

## 4.5 MATdescribe

### 4.5.1 Uso

MATdescribe(data, variable)

### 4.5.2 Objetivo

### Esta función realiza un análisis descriptivo y gráfico detallado de una variable numérica en un dataframe. Incluye histogramas, densidad, boxplots y pruebas de normalidad.

### 4.5.3 Argumentos

* data: El dataframe que contiene los datos.
* variable: La variable numérica a analizar (sin comillas).
* bins: Número de bins para el histograma. Si es 0, o no se especifica, se usa el método *Freedman-Diaconis* para calcular el número de intervalos de valores en el histograma.

### 4.5.4 Salida

### Genera un objeto en el entorno global con el nombre <variable>\_describe\_info, que incluye:

### **Resumen gráfico** (histograma, densidad, boxplot, QQ-Plot).

### **Estadísticos descriptivos** (media, mediana, asimetría, curtosis, etc.).

### **Pruebas de normalidad** (Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling).

### 4.5.5 Ejemplo

MATdescribe(eolica\_100\_sm\_so, RENECO, bins = 0)

### 4.5.6 Resultados Esperados

1. Una lista llamada RENECO\_describe\_info que contiene:
   * Gráfico resumen con varios paneles.
   * Tabla de estadísticos descriptivos.
   * Resultado de la prueba de normalidad de Shapiro-Wilk, Kolmogorov-Smirnov y Anderson-Darling  (para significación de 0.05).

RENECO\_describe\_info$grafico\_resumen

Gráfico

El contenido generado por IA puede ser incorrecto.

RENECO\_describe\_info$estadisticos

Imagen que contiene Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

RENECO\_describe\_info$normalidad

Interfaz de usuario gráfica, Texto

El contenido generado por IA puede ser incorrecto.

## 4.6 MATtaf

### 4.6.1 Uso

MATtaf(data, variable, breaks = NULL)

### 4.6.2 Objetivo

Generar una tabla de frecuencias agrupadas en intervalos y un histograma relacionado.

### 4.6.3 Argumentos

* data: El dataframe que contiene los datos.
* variable: La variable numérica para analizar (sin comillas).
* breaks: Número de intervalos (opcional, se calcula automáticamente si no se especifica, mediante el método de Freedman-Diaconis).

### 4.6.4 Salida

* Una tabla de frecuencias.
* Un histograma con los mismos intervalos.

### 4.6.5 Ejemplo

MATtaf(eolica\_100\_sm\_so, RENECO)

### 4.6.6 Resultados Esperados

1. Una lista llamada RENECO\_intervalos\_frecuencia que contiene:
   * Histograma con los mismos intervalos que la tabla de frecuencias.
   * Tabla de frecuencias agrupadas en intervalos.

RENECO\_intervalos\_frecuencia$histograma

Gráfico, Histograma

El contenido generado por IA puede ser incorrecto.

RENECO\_intervalos\_frecuencia$tabla

Tabla

El contenido generado por IA puede ser incorrecto.

# 5 Documentación del Paquete

* **Autor:** Miguel Ángel Tarancón miguelangel.tarancon@uclm.es
* **Versión:** 0.1.0
* **Fecha:** 2025-02-14
* **Licencia:** MIT

# 6 ¡Empieza a explorar tus datos con MATdatatools!

Este paquete está diseñado para facilitarte el trabajo y ayudarte a generar análisis rápidos y visuales. ¡Pruébalo hoy y optimiza tus estudios de datos!

Logotipo

Descripción generada automáticamente