



ANÁLISIS EXPLORATORIO DE 100 EMPRESAS EÓLICAS ESPAÑOLAS

Ramírez Guijarro, Marta

ÍNDICE

1. INTRODUCCIÓN	3
2. PROCEDIMIENTOS PREVIOS AL ANÁLISIS	3
3. ANÁLISIS DE LA VARIABLE <i>ENDEUDA</i>	5
3.1. TRATAMIENTO DE <i>MISSING VALUES</i>	5
3.2. TRATAMIENTO DE <i>OUTLIERS</i>	7
3.3. DESCRIPTIVOS BÁSICOS DE LA VARIABLE <i>ENDEUDA</i>	8
3.4. ¿SIGUE <i>ENDEUDA</i> UNA DISTRIBUCIÓN NORMAL?: GRÁFICO QQ	9
4. ANÁLISIS DE LAS VARIABLES <i>ENDEUDA</i> Y <i>RENECO</i>	10
4.1. TRATAMIENTO DE <i>MISSING VALUES</i>	11
4.2. TRATAMIENTO DE <i>OUTLIERS</i>	11
5. ANÁLISIS DE LAS VARIABLES <i>ENDEUDA</i> , <i>RES</i> , <i>FPIOS</i> , Y <i>ACTIVO</i>	14
5.1. TRATAMIENTO DE <i>MISSING VALUES</i> Y <i>OUTLIERS</i>	14
5.2. ANÁLISIS DE CORRELACIONES	16
6. CONCLUSIONES	18
ANEXO 1. CÓDIGO	19

List of Figures

1	Valores perdidos	5
2	ENDEUDAMIENTO	6
3	ENDEUDAMIENTO (%)	7
4	ENDEUDAMIENTO	9
5	ENDEUDAMIENTO: QQ-PLOT	10
6	DISTANCIA DE MAHALANOBIS	15
7	Correlaciones	17

List of Tables

1	Relación de variables y descripción	3
2	Estadísticos básicos	4
3	Valores perdidos de la variable ENDEUDA.	6
4	Valores atípicos	8
5	Estadísticos básicos de ENDEUDA	8
6	Valores perdidos en las variables ENDEUDA y RENECO	11
7	Outliers de las variables ENDEUDA y RENECO	13
8	Valores perdidos de las variables ENDEUDA, RES, FPIOS y ACTIVO	14
9	Outliers de las variables ENDEUDA, RES, FPIOS y ACTIVO (incluye distancias de Mahalanobis)	16

1. INTRODUCCIÓN

El presente proyecto se basa en el análisis de una base de datos extraída de la plataforma SABI, la cual contiene información económica correspondiente a empresas eólicas de España, dicha información está recogida en las siguientes variables (Table 1).

Table 1: Relación de variables y descripción

VARIABLE (atributo)	DESCRIPCIÓN	OBSERVACIONES
RENECO	Rentabilidad económica (%)	Numérica: continua
RENFIN	Rentabilidad financiera (%)	Numérica: continua
LIQUIDEZ	Liquidez general (%)	Numérica: continua
ENDEUDA	Endeudamiento (%)	Numérica: continua
ACTIVO	Total Activo (mil EUR)	Numérica: continua
FPIOS	Fondos propios (mil EUR)	Numérica: continua
RES	Resultado del Ejercicio (mil EUR)	Numérica: continua
MARGEN	Margen de beneficio (%)	Numérica: continua
SOLVENCIA	Coefficiente de solvencia (%)	Numérica: continua
APALANCA	Apalancamiento (%)	Numérica: continua
DIMENSIÓN	Grande, Medio, Pequeño	Numérica: discreta
MATRIZ	GUO - Nombre	Cualitativa: nominal

2. PROCEDIMIENTOS PREVIOS AL ANÁLISIS

En primer lugar, se importará la base de datos anteriormente mencionada al software Rstudio.

Una vez importada la base de datos, se muestran los estadísticos básicos de todas las variables anteriormente mencionadas (Table 2).

Table 2: Estadísticos básicos

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

Así, en la Table 2 se pueden observar los estadísticos básicos de las variables como, por ejemplo, la rentabilidad económica media, la rentabilidad financiera media, la liquidez media mantenida, el endeudamiento medio, el número de empleados medio, los activos medios, los fondos propios medios, el resultado medio, los ingresos medios, el margen medio, la solvencia media, el grado de apalancamiento medio y el número de accionistas medio, entre otros. Cabe destacar que para las variables nominales estos estadísticos básicos no son válidos.

Dado que para realizar muchas de las Técnicas de Análisis Multivariantes (en adelante TAM) de tipo paramétrico es necesario que las variables implicadas sigan una distribución normal, se procederá en epígrafes posteriores a la comprobación de la normalidad de las variables analizadas. Además, para que los resultados de las TAM sean plenamente confiables y no se vean alterados o invalidados, es necesario comprobar la existencia de observaciones “raras” o *outliers*. Pues estos suelen alterar la normalidad de las variables o afectar a los resultados de los análisis, por lo

que en caso de existir es recomendable evaluar su eliminación de la muestra o su conservación. Paralelamente, también se comprobará la existencia de *valores perdidos* o *missing values* para, en caso de existir, proceder a su eliminación o imputación de un valor discrecional, pues estos impedirán la correcta realización de los análisis pertinentes.

3. ANÁLISIS DE LA VARIABLE ENDEUDA

La primera variable que se va a analizar es ENDEUDA, para ello, en primer lugar, se va a proceder a la comprobación de la existencia de *missing values*.

3.1. TRATAMIENTO DE MISSING VALUES

Como ya se ha comentado, la existencia de valores perdidos dificulta e incluso impide el correcto análisis de las variables, por ello, a continuación, se va a proceder a comprobar la existencia de valores perdidos (en la Table 3 se puede observar como la mayoría de las variables los presentan) a través de un gráfico.

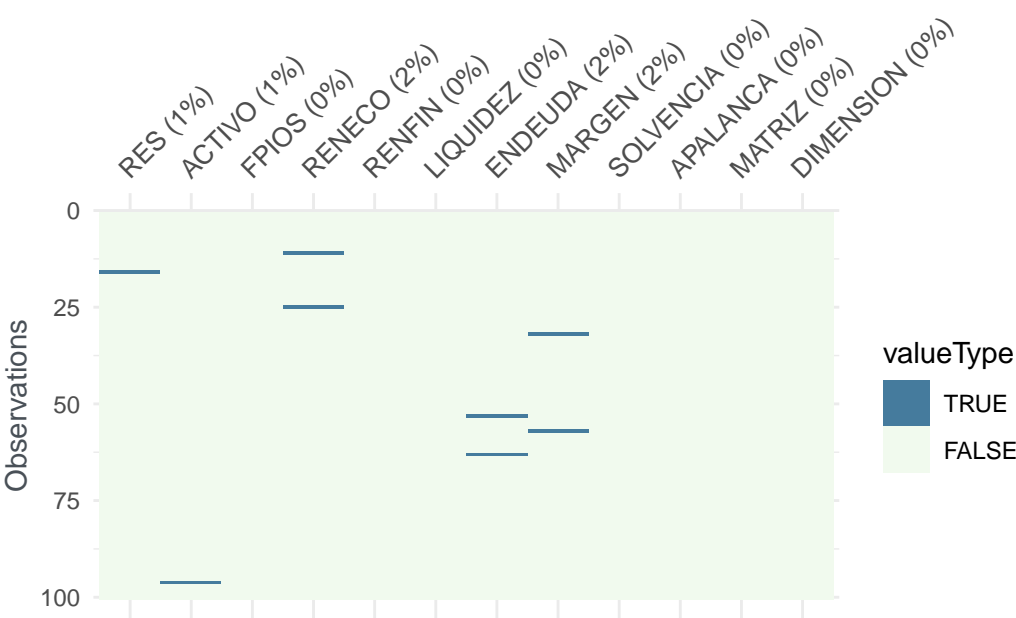


Figure 1: Valores perdidos

A continuación, en la Table 3, se muestran las observaciones que presentan valores perdidos en la variable analizada ENDEUDA.

Table 3: Valores perdidos de la variable ENDEUDA.

	ENDEUDA
Acampo Arias SL	NA
Parque Eolico Marmellar Sociedad Limitada.	NA

Así, existen dos empresas eólicas que no tienen datos de endeudamiento, por lo que, en este caso se va a proceder a la eliminación de dichas observaciones. Por ello, a partir de ahora se trabajará con 98 observaciones, en vez de con 100.

Una vez eliminados los *missing values* de la variable ENDEUDA se presenta su representación gráfica a través de un histograma, lo que permitirá observar la distribución de dicha variable.

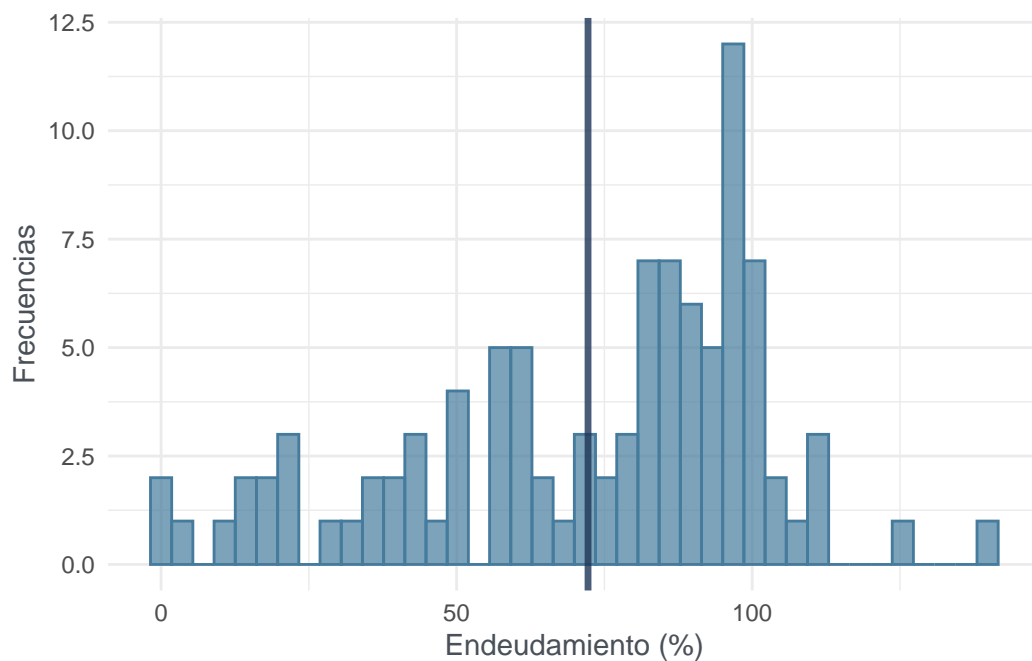


Figure 2: ENDEUDAMIENTO

A priori, se puede observar como aparentemente la variable ENDEUDANO sigue una distribución normal, pues su histograma dista de asemejarse a la Campana de Gauss correspondiente. Pero dicha observación no es concluyente a través de simplemente un análisis gráfico como este, por ello más adelante se comprobará de forma más exhaustiva la normalidad de la variable.

3.2. TRATAMIENTO DE *OUTLIERS*

Tras eliminar los valores perdidos de la variable objeto de análisis resulta necesario determinar si existen observaciones “raras” o *outliers* que alteren la normalidad de esta o interfieran en los resultados de los análisis pertinentes. La forma más sencilla de comprobar si existen *outliers* a primera vista en una variable es a través de un gráfico de cajas como el mostrado en la Figure 3.

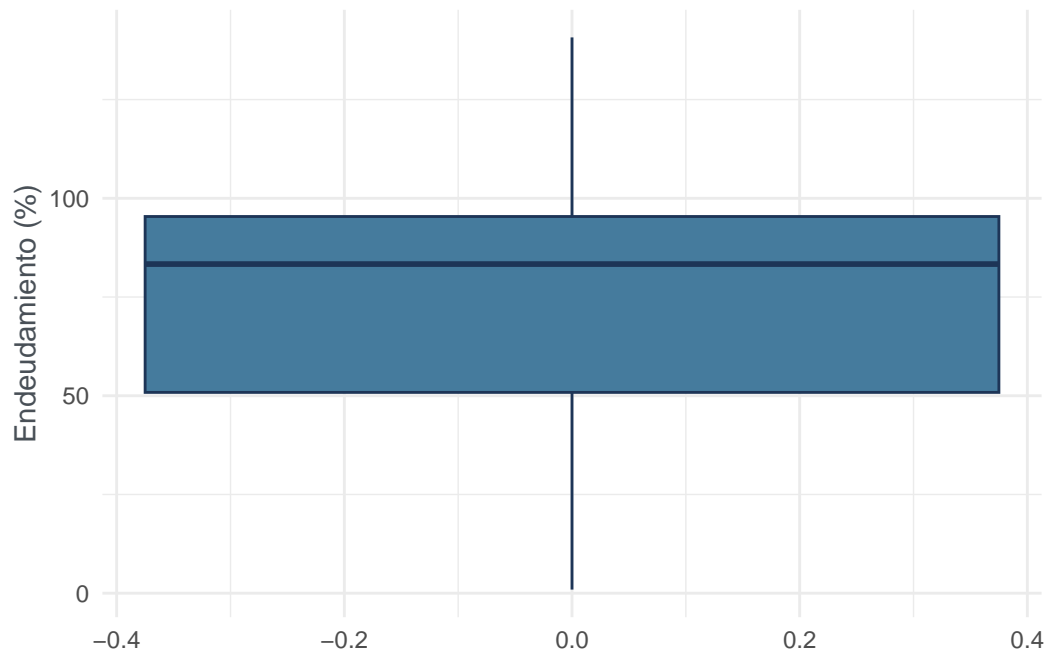


Figure 3: ENDEUDAMIENTO (%)

Así, se puede observar como la variable ENDEUDA no presenta valores atípicos a simple vista. Para asegurar dicho resultado a continuación se muestra el valor del primer cuantil y del tercero.

Creando el objeto Q1 en el que van a estar el 25% de los datos de la variable ENDEUDA (que está expresada en %) y en el objeto Q3 el 75% de los datos.

Obteniéndose que el 25% de las empresas eólicas tienen un endeudamiento del 50.85% y el 75% cuentan con un endeudamiento del 95.40%.

Endeudamiento	
25%	50.85250
75%	95.38825

A continuación, se mostrarán los *outliers* existentes, si es que los hay. Para ello, mostramos las

observaciones mayores a $Q3 + 1.5 \cdot IQR$ y todos los valores menores a $Q1 - 1.5 \cdot IQR$.

Table 4: Valores atípicos

ENDEUDA

Como se puede comprobar, la variable ENDEUDA no presenta valores atípicos.

3.3. DESCRIPTIVOS BÁSICOS DE LA VARIABLE *ENDEUDA*

Una vez eliminados los valores perdidos y comprobado que no existen *outliers* se presentan los estadísticos básicos de la variable ENDEUDA.

Table 5: Estadísticos básicos de ENDEUDA

	value
Mean	72.2268265
Std.Dev	30.2707472
Min	0.9170000
Q1	49.0760000
Median	83.3460000
Q3	95.4380000
Max	140.7450000
IQR	44.5357500
CV	0.4191067

Donde se puede observar como el endeudamiento medio de las empresas eólicas analizadas es del 72.23% , siendo el menor endeudamiento presentado por las empresas eólicas analizadas un 0.92% y el máximo del 140.75%. A continuación, se muestra de nuevo la distribución de la variable ENDEUDA a través de su histograma, mostrándose también en este el endeudamiento medio.

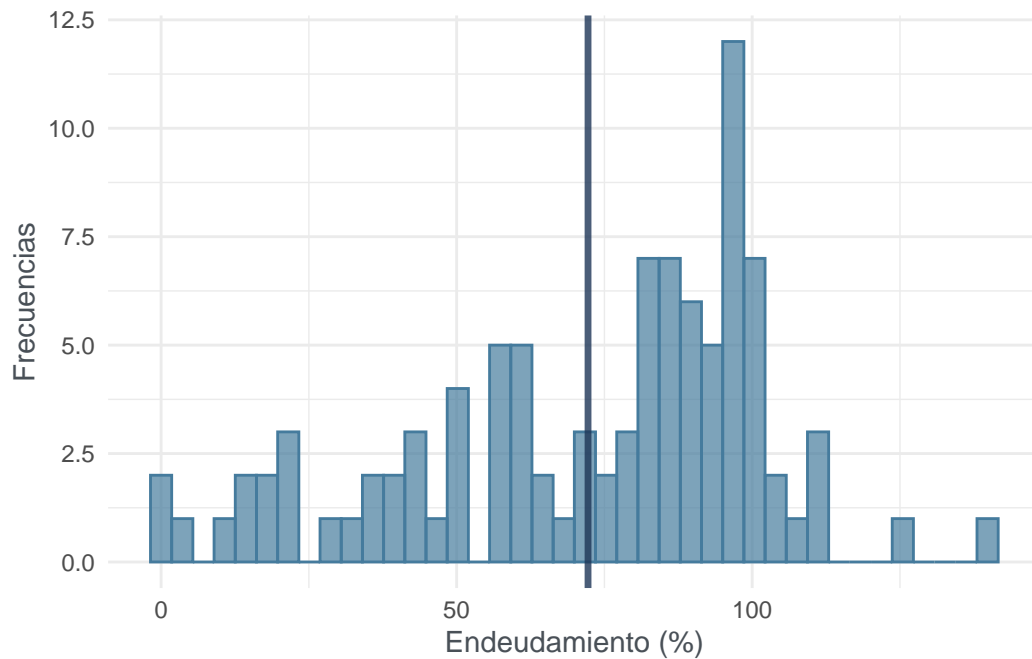


Figure 4: ENDEUDAMIENTO

En el histograma se aprecia como, aparentemente, ENDEUDA no sigue una distribución normal, tal y como ya se señaló anteriormente, pero de forma gráfica no se puede determinar con total seguridad, por lo que a continuación se presenta un análisis más detallado de la normalidad de ENDEUDA. También se puede observar como muchas de las empresas eólicas presentan ratios de endeudamiento entre la media y el 100%.

3.4. ¿SIGUE ENDEUDA UNA DISTRIBUCIÓN NORMAL?: GRÁFICO QQ

El primer análisis para comprobar la normalidad es gráfico, al igual que el histograma, este es el gráfico cuantil-cuantil. Donde se comparan los cuantiles, sin *outliers*, de la muestra con los cuantiles teóricos de una distribución normal.

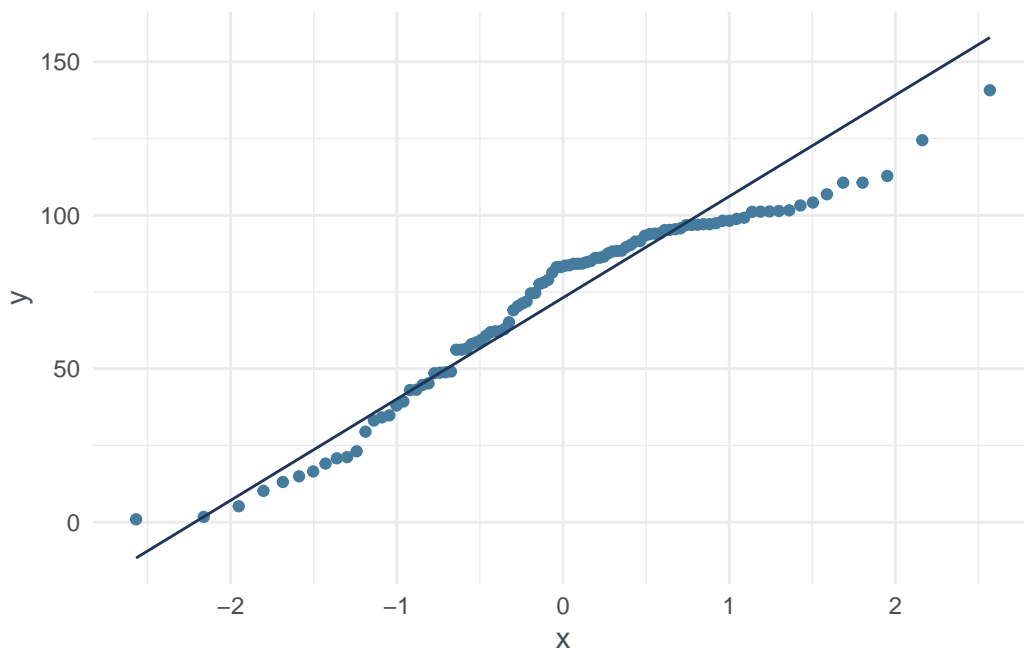


Figure 5: ENDEUDAMIENTO: QQ-PLOT

Con este análisis se observa como la variable ENDEUDA no parece presentar normalidad, pues en el extremo superior y en el centro la muestra se distancia mucho de la línea teórica. Como no se puede dar una conclusión sólida se realizará un contraste paramétrico.

3.4.1. PRUEBA DE SHAPIRO-WILK

El contraste paramétrico para determinar la normalidad de una variable es el Test de Shapiro-Wilk, el cual utiliza como hipótesis nula (H_0) que la muestra presenta normalidad.

Así, realizando dicho test para la variable ENDEUDA se determina que esta no sigue una distribución normal, pues el p-valor resultante del test es inferior a 0.05, por lo que se rechaza H_0 .

Shapiro-Wilk normality test

```
data: muestra$ENDEUDA
```

```
W = 0.93863, p-value = 0.0001865
```

4. ANÁLISIS DE LAS VARIABLES *ENDEUDA* Y *RENECO*

A continuación se analizarán conjuntamente las variables ENDEUDA y RENECO, es decir, el endeudamiento y la rentabilidad económica.

4.1. TRATAMIENTO DE *MISSING VALUES*

En primer lugar, se comprobará la existencia de valores perdidos en dichas variables que nos dificulten el análisis.

Table 6: Valores perdidos en las variables ENDEUDA y RENECO

	ENDEUDA	RENECO
Viesgo Renovables SL.	34.116	NA
Sargon Energias SLU	112.811	NA
Acampo Arias SL	NA	6.591
Parque Eolico Marmellar Sociedad Limitada.	NA	2.323

Tal y como se puede comprobar en la Table 6 ambas variables presentan valores perdidos. Por lo que estos van a ser eliminados, pasándose a tener 96 observaciones (en el `data frame muestra2`) en vez de 100.

4.2. TRATAMIENTO DE *OUTLIERS*

Una vez eliminados los valores perdidos, se procederá a la localización y eliminación de los *outliers*, para poder realizar correctamente los análisis pertinentes.

Dado que en este caso se está trabajando con dos variables, para detectar los posibles *outliers*, se realizará un gráfico de dispersión, donde se representaran ENDEUDA y RENECO al mismo tiempo. Además, también se presenta un gráfico de cajas de cada una de las variables objeto de análisis.



Tal y como se puede observar en los gráficos superiores, existen varias observaciones de las que cabe pensar que se tratan de *outliers* (por su distancia con la mayor densidad de la nube de puntos). Así, en el gráfico se puede comprobar la posible existencia de al menos 2 observaciones atípicas en la variable RENEKO.

Por lo que, se van a crear los objetos Q1 (donde se mostrará el valor medio de cada variable en el primer cuantil, donde estarán situadas el 25% de las empresas analizadas) y Q3 (donde se mostrará el valor medio de cada variable en el tercer cuantil, donde estarán situadas el 75% de las empresas analizadas) para cada una de las variables objeto de análisis.

A continuación, se muestran los *outliers* de ambas variables, es decir, las observaciones que presentan valores, para cada una de las variables analizadas, por encima del tercer cuantil más 1.5 veces el rango intercuantílico ($Q3 + 1.5 \cdot IQR$) y por debajo del primer cuantil menos 1.5 veces el rango intercuantílico ($Q1 - 1.5 \cdot IQR$).

Table 7: Outliers de las variables ENDEUDA y RENECO

	ENDEUDA	RENECO
Molinos Del Ebro SA	56.546	35.262
Sierra De Selva SL	29.475	21.761

Así, el resultado obtenido es que existen dos *outliers* (tal y como se puede ver en la Table 7) que van a ser eliminados. Quedándose el data frame muestra2_so con 94 observaciones.

De esta forma, volviéndose a realizar el gráfico de dispersión se observa como la muestra ya no presenta tantos valores extremos, aunque es cierto que se pueden apreciar observaciones que se encuentran alejadas de la mayor densidad de observaciones, pero estas ya no serán tratadas como *outliers*.



5. ANÁLISIS DE LAS VARIABLES ENDEUDA, RES, FPIOS, Y ACTIVO

A continuación, se analizarán las variables ENDEUDA, RES, FPIOS y ACTIVO, de la misma forma que anteriormente se han analizado otras, es decir, se analizará el endeudamiento, resultado, fondos propios y el activo de las empresas eólicas de la muestra.

5.1. TRATAMIENTO DE *MISSING VALUES* Y *OUTLIERS*

Así, tal y como se ha hecho anteriormente, se va a determinar si la nueva muestra de cuatro variables presenta valores perdidos. Los análisis pertinentes indican, tal y como se muestra en la Table 8, que hay varias observaciones que presentan *missing values* en las distintas variables objeto de análisis, por lo que se procederá a la eliminación de estas observaciones, por lo que a partir de ahora se trabajará con una muestra de 96 observaciones, en vez de con 100.

Table 8: Valores perdidos de las variables ENDEUDA, RES, FPIOS y ACTIVO

	ENDEUDA	RES	FPIOS	ACTIVO
Biovent Energia SA	61.917	NA	70033.000	183899.00
Acampo Arias SL	NA	2186.314	5621.975	44222.76
Parque Eolico Marmellar Sociedad Limitada.	NA	631.352	8654.645	38911.08
La Caldera Energia Burgos SL	110.636	511.304	-2752.605	NA

Como ya se ha hecho en apartados anteriores, tras la eliminación de los *missing values* es pertinente analizar si existen *outliers* que distorsionen el análisis.

Dado que, en esta ocasión, se está trabajando con más de dos variables, no es posible realizar un gráfico de dispersión de las variables analizadas para observar de forma gráfica la existencia de valores atípicos. Por ello, dicho análisis se realizará a través de las distancias de Mahalanobis, que son las que se muestran en el Figure 6. De esta forma, a priori, se pueden identificar varias observaciones atípicas.

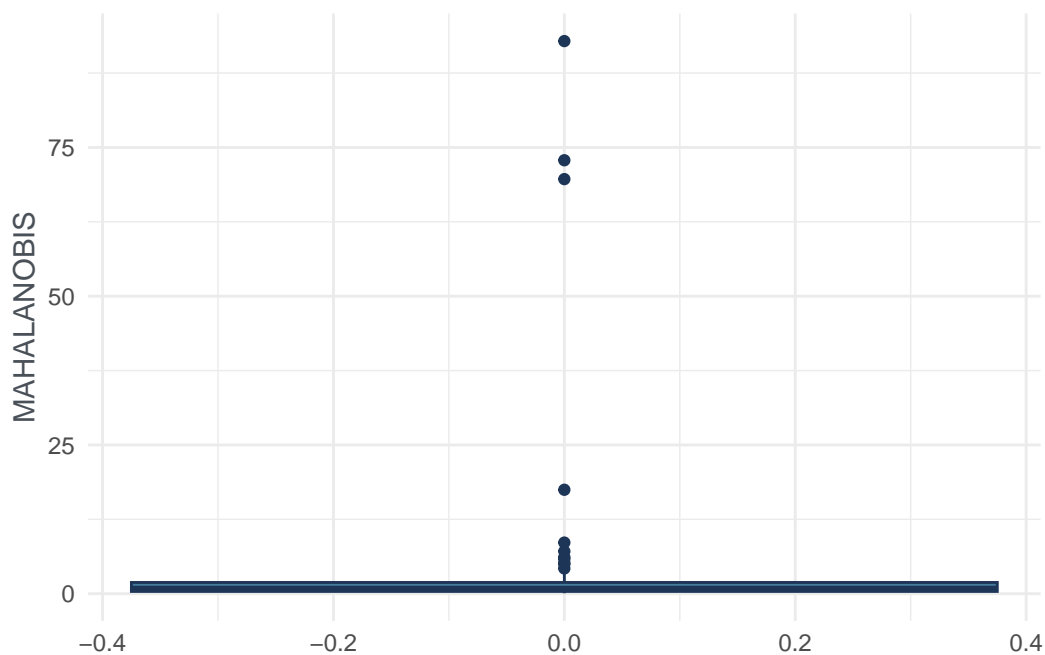


Figure 6: DISTANCIA DE MAHALANOBIS

Tal y como se ha realizado anteriormente, a continuación se va a determinar paramétricamente la existencia de *outliers*. Para ello se van a crear los objetos Q1 (donde se mostrará el valor medio de cada variable en el primer cuantil, donde estarán situadas el 25% de las empresas analizadas) y Q3 (donde se mostrará el valor medio de cada variable en el tercer cuantil, donde estarán situadas el 75% de las empresas analizadas) para las distancias de Mahalanobis.

A continuación, se muestran los *outliers* de las variables analizadas, es decir, las observaciones que presentan valores, para cada una de las variables analizadas, por encima del tercer cuantil más 1.5 veces el rango intercuantílico ($Q3 + 1.5 \cdot IQR$) y por debajo del primer cuantil menos 1.5 veces el rango intercuantílico ($Q1 - 1.5 \cdot IQR$). Así, tal y como se observa en la Table 9, existen once observaciones que presentan valores atípicos, por lo que estas serán eliminadas para que no influyeran los resultados de los análisis posteriores.

Table 9: Outliers de las variables ENDEUDA, RES, FPIOS y ACTIVO (incluye distancias de Mahalanobis)

	MAHALANOBIS	ENDEUDA	RES	FPIOS	ACTIVO
Holding De Negocios De GAS SL.	92.870789	48.825	727548.000	6904824.00	13492812.00
Global Power Generation SA.	72.842172	13.082	39995.000	1740487.00	2002458.00
Naturgy Renovables SLU	69.673916	83.725	42737.000	318475.00	1956869.00
Corporacion Acciona Eolica SL	8.593143	84.262	29592.000	136064.00	864606.00
Saeta Yield SA.	17.470320	16.510	2084.476	665319.56	796886.38
Guzman Energia SL	7.123109	140.745	-5661.463	-77532.70	190286.98
WPD Wind Investment SL.	5.836603	0.917	-850.068	108023.83	109023.80
Renovalia Reserve SL.	5.194619	5.169	2485.164	88563.23	93391.14
Disa Duna SL.	6.150857	1.696	2594.401	42283.14	43012.79
Parques Eolicos San Lorenzo SL	4.951056	10.193	4311.980	37396.17	41640.82
Parques Eolicos Ciudad Real Sociedad Limitada	4.242076	14.934	2543.000	21810.00	25639.00

5.2. ANÁLISIS DE CORRELACIONES

El último análisis que se va a realizar a las variables elegidas es determinar si estas presentan correlación entre ellas, un análisis que es útil para posteriores procedimientos como puede ser la creación de componentes principales, o para evitar la multicolinealidad en un modelo regresivo.

Generalmente, se dice que dos variables están correlacionadas cuando presentan un índice de Pearson por encima de 0.7 (en valor absoluto), teniendo en cuenta que el índice de correlación toma valores desde -1 hasta 1. Significando una correlación negativa entre dos variables que están tienen un comportamiento opuesto y una correlación positiva que ambas variables aumentan o disminuyen al mismo tiempo.

Así, tal y como muestra el Figure 7, la variable ENDEUDA no presenta correlación elevada con las otras tres, RES presenta una elevada correlación positiva con los FPIOS y ACTIVO (cuando una de las tres aumenta las otras dos también lo hacen, aunque con este estudio no se puede determinar el sentido de la causalidad) y, por último, FPIOS, y ACTIVO también tiene una elevada correlación positiva, por lo que, por ejemplo, podría resultar apropiado utilizar en ellas técnicas de componentes principales u otros análisis.

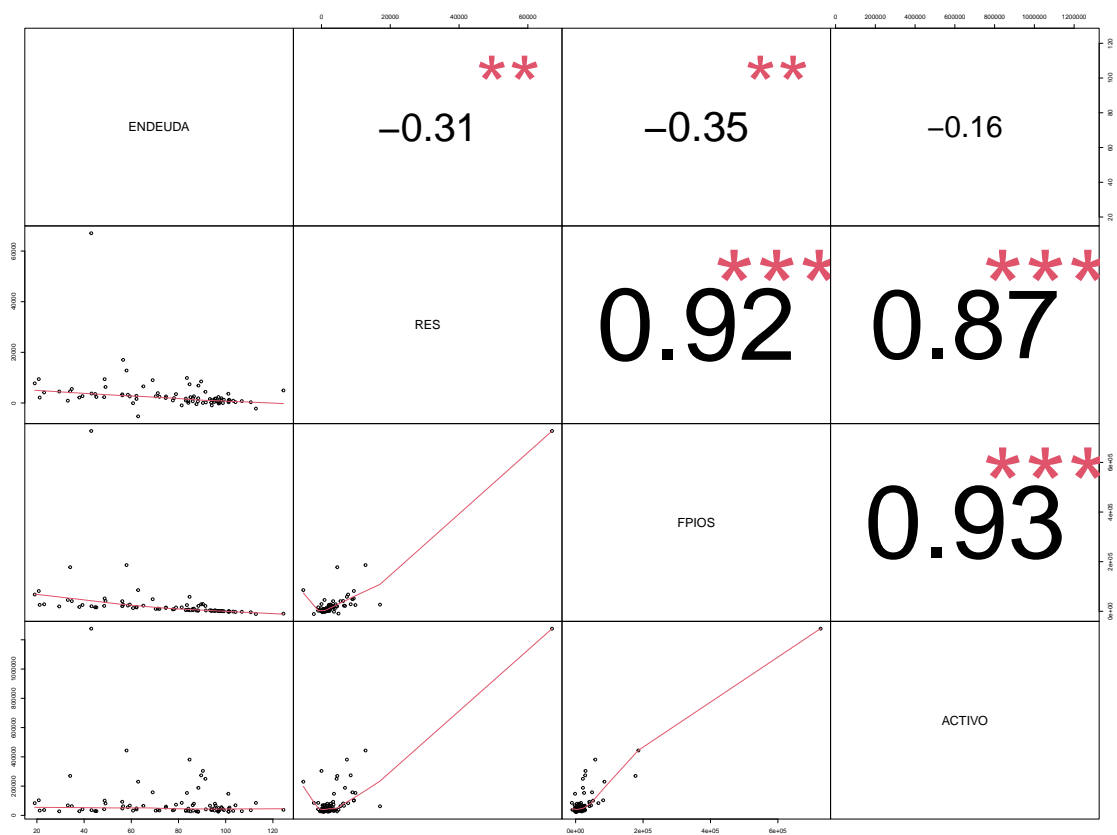


Figure 7: Correlaciones

6. CONCLUSIONES

De este análisis exploratorio se puede concluir que las empresas eólicas españolas analizadas presentan un elevado ratio de endeudamiento, no existiendo entre ellas ninguna que presente valores atípicamente altos o bajos de esta variable. Además, la distribución de estos ratios de endeudamiento no sigue una distribución normal, sino que presenta una asimetría positiva.

Por el contrario, en la variable rentabilidad económica si que existen empresas que presentan valores atípicamente altos, respecto al resto. Por último, se puede decir que la variable resultado, fondos propios y activo presentan una alta correlación positiva entre las tres. Por el contrario, el ratio de endeudamiento no presenta correlación con ninguna de las otras tres.

ANEXO 1. CÓDIGO

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)

rm(list = ls())

library(readxl)

eolicas <- read_excel("eolica_100_mv.xlsx", sheet = "Datos")

library(printr)
library(kableExtra)

#Asignar la primera columna como nombre de fila

eolicas <- data.frame(eolicas, row.names = 1)

knitr::kables(list(kable(summary (eolicas[1:5])),
                    kable(summary (eolicas[6:10])),
                    kable(summary (eolicas[11:12]))),
               caption = "Estadísticos básicos") %>%
  kableExtra::kable_classic() %>%
  kableExtra::row_spec(0, bold = T, color = "black",
                       background = "#F1FaEE", align = "c") %>%
  kableExtra::row_spec(8, bold = T, color = "black",
                       background = "#F1FaEE", align = "c") %>%
  kableExtra::row_spec(16, bold = T, color = "black",
                       background = "#F1FaEE", align = "c") %>%
  kableExtra::kable_styling(full_width = F,
                             latex_options = c("basic", "HOLD_position"),
                             position = "center", font_size = 10)
```

```

library(visdat)
library(ggplot2)
library(ggthemes)

vis_miss(eolicas) +
  scale_color_manual(aesthetics = "fill", values = c("#F1FaEE", "#457b9d")) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, size = 10, hjust = 0, vjust = 1),
        axis.title = element_text(color = "#495057"),
        plot.title = element_text(color = "#495057"))
library (dplyr)

muestra<- select(eolicas, everything())

knitr::kable(muestra %>%
              filter(is.na(ENDEUDA)) %>%
              select(ENDEUDA),
              caption = "Valores perdidos de la variable ENDEUDA.") %>%
kableExtra::kable_classic() %>%
kable_styling(full_width = F, latex_options = c("basic", "HOLD_position"),
              "condensed", position = "center", font_size = 10)
muestra <- muestra %>% filter(! is.na(ENDEUDA))
ggplot(data = muestra, map = aes(x = ENDEUDA)) +
  geom_histogram(bins = 40, colour = "#457b9d", fill = "#457b9d", alpha = 0.7) +
  geom_vline(xintercept = mean(muestra$ENDEUDA),
            color = "#1d3557", size = 1.2, alpha = 0.8) +
  xlab("Endeudamiento (%)") +
  ylab("Frecuencias") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057"),
        plot.title = element_text(color = "#495057"))
library (ggplot2)

```

```

ggplot(data = muestra, map = (aes(y = ENDEUDA))) +
  geom_boxplot(fill = "#457b9d", colour = "#1d3557") +
  ylab("Endeudamiento (%)") +
  scale_color_manual(aesthetics = "fill", values = c("#457b9d", "#1d3557")) +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057"),
        plot.title = element_text(color = "#495057"))

Q1 <- quantile (muestra$ENDEUDA, c(0.25))
Q3 <- quantile (muestra$ENDEUDA, c(0.75))
table <- c(Q1, Q3)

knitr::kable(table, col.names = c("Endeudamiento")) %>%
  kableExtra::kable_classic() %>%
  kableExtra::row_spec(0, bold = T, color = "black",
                      background = "white", align = "c") %>%
  kableExtra::kable_styling(full_width = F,
                          latex_options = c("basic", "HOLD_position"),
                          position = "center", font_size = 10)

table2 <- muestra %>% filter(ENDEUDA > Q3 + 1.5*IQR(ENDEUDA) |
                          ENDEUDA < Q1 - 1.5*IQR(ENDEUDA),
                          na.rm = TRUE) %>%
  select(ENDEUDA)

knitr::kable(table2, caption = "Valores atípicos") %>%
  kableExtra::kable_classic() %>%
  kable_styling(full_width = F, latex_options = c("basic", "HOLD_position"),
                "condensed", position = "center", font_size = 10)

library (summarytools)

knitr::kable(descr(muestra$ENDEUDA,
                  stats = c("mean", "sd", "min", "q1", "med", "q3", "max", "iqr", "cv"),
                  transpose = FALSE,

```

```

    style = "simple",
    justify = "center",
    headings = T), caption = "Estadísticos básicos de ENDEUDA") %>%
kableExtra::kable_classic() %>%
kable_styling(full_width = F, latex_options = c("basic", "HOLD_position"),
               "condensed", position = "center", font_size = 10)

ggplot(data = muestra, map = aes(x = ENDEUDA)) +
  geom_histogram(bins = 40, colour = "#457b9d", fill = "#457b9d", alpha = 0.7) +
  geom_vline(xintercept = mean(muestra$ENDEUDA),
             color = "#1d3557", size = 1.2, alpha = 0.8) +
  xlab("Endeudamiento (%)") +
  ylab("Frecuencias") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057"),
        plot.title = element_text(color = "#495057"))

ggplot(data = muestra, aes(sample = ENDEUDA)) +
  stat_qq(colour = "#457b9d") +
  stat_qq_line(colour = "#1d3557") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057"),
        plot.title = element_text(color = "#495057"))

shapiro.test(x = muestra$ENDEUDA)

muestra2 <- select(eolicas, everything())

knitr::kable(muestra2 %>% filter(is.na(ENDEUDA) | is.na(RENECO)) %>%
             select(ENDEUDA, RENECO),
             caption = "Valores perdidos en las variables ENDEUDA y RENECO") %>%
kableExtra::kable_classic() %>%
kable_styling(full_width = F, latex_options = c("basic", "HOLD_position"),
               "condensed", position = "center", font_size = 10)

```

```

muestra2 <- muestra2 %>% filter(! is.na(ENDEUDA) & ! is.na(RENECO) )

dispersion <- ggplot(data = muestra2, map = (aes(x = ENDEUDA, y = RENECO))) +
  geom_point(colour = "#457b9d", size = 2, alpha = 0.5) +
  ggtitle("ENDEUDAMIENTO vs RENTABILIDAD ECONÓMICA") +
  xlab("Endeudamiento (%)") +
  ylab("Rentabilidad Económica (%)") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057", size = 25),
        plot.title = element_text(color = "#495057", size = 25))
caja_ENDEUDA <- ggplot(data = muestra2, map = (aes(y = ENDEUDA))) +
  geom_boxplot(fill = "#457b9d", colour = "#1d3557") +
  ggtitle("ENDEUDAMIENTO") +
  ylab("Endeudamiento (%)") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057", size = 25),
        plot.title = element_text(color = "#495057", size = 25))
caja_RENECO <- ggplot(data = muestra2, map = (aes(y = RENECO))) +
  geom_boxplot(fill = "#457b9d", colour = "#1d3557") +
  ggtitle("RENTABILIDAD ECONÓMICA") +
  ylab("Rentabilidad Económica (%)") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057", size = 25),
        plot.title = element_text(color = "#495057", size = 25))
library(patchwork)

dispersion / (caja_ENDEUDA | caja_RENECO)
Q1_ENDEUDA <- quantile (muestra2$ENDEUDA, c(0.25))
Q3_ENDEUDA <- quantile (muestra2$ENDEUDA, c(0.75))
Q1_RENECO <- quantile (muestra2$RENECO, c(0.25))
Q3_RENECO <- quantile (muestra2$RENECO, c(0.75))

```

```

knitr::kable(muestra2 %>%
  filter(ENDEUDA > Q3_ENDEUDA + 1.5*IQR(ENDEUDA) |
         ENDEUDA < Q1_ENDEUDA - 1.5*IQR(ENDEUDA) |
         RENECO > Q3_RENECO + 1.5*IQR(RENECO) |
         RENECO < Q1_RENECO - 1.5*IQR(RENECO)) %>%
  select(ENDEUDA, RENECO),
  caption = "Outliers de las variables ENDEUDA y RENECO") %>%
kableExtra::kable_classic() %>%
kable_styling(full_width = F, latex_options = c("basic", "HOLD_position"),
              "condensed", position = "center", font_size = 10)

muestra2_so <- muestra2 %>%
  filter(ENDEUDA <= Q3_ENDEUDA + 1.5*IQR(ENDEUDA) &
         ENDEUDA >= Q1_ENDEUDA - 1.5*IQR(ENDEUDA) &
         RENECO <= Q3_RENECO + 1.5*IQR(RENECO) &
         RENECO >= Q1_RENECO - 1.5*IQR(RENECO))

dispersion_so <- ggplot(data = muestra2_so, map = (aes(x = ENDEUDA, y = RENECO))) +
  geom_point(colour = "#457b9d", size = 2, alpha = 0.5) +
  ggtitle("ENDEUDAMIENTO vs RENTABILIDAD ECONÓMICA") +
  xlab("Endeudamiento (%)") +
  ylab("Rentabilidad Económica (%)") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057", size = 25),
        plot.title = element_text(color = "#495057", size = 25))

caja_ENDEUDA_so <- ggplot(data = muestra2_so, map = (aes(y = ENDEUDA))) +
  geom_boxplot(fill = "#457b9d", colour = "#1d3557") +
  ggtitle("ENDEUDAMIENTO") +
  ylab("ENDEUDAMIENTO (%)") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057", size = 25),

```



```

    plot.title = element_text(color = "#495057", size = 25))

caja_RENECO_so <- ggplot(data = muestra2_so, map = (aes(y = RENECO))) +
  geom_boxplot(fill = "#457b9d", colour = "#1d3557") +
  ggtitle("RENECO") +
  ylab("Rentabilidad económica (%)") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057", size = 25),
        plot.title = element_text(color = "#495057", size = 25))

dispersion_so / (caja_ENDEUDA_so | caja_RENECO_so) +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057"),
        plot.title = element_text(color = "#495057"))

muestra3<- select(eolicas, everything())
knitr::kable(muestra3 %>% filter(is.na(ENDEUDA) | is.na(RES) |
                                is.na(FPIOS) | is.na(ACTIVO)) %>%
              select(ENDEUDA, RES, FPIOS, ACTIVO),
              caption = "Valores perdidos de las variables ENDEUDA, RES, FPIOS y ACTIVO") %>%
kableExtra::kable_classic() %>%
kable_styling(full_width = F, latex_options = c("basic", "HOLD_position"),
              "condensed", position = "center", font_size = 10)

muestra3 <- muestra3 %>% filter(! is.na(ENDEUDA) &
                                ! is.na(RES) & ! is.na(FPIOS) &
                                ! is.na(ACTIVO))

muestra3.variables <- muestra3 %>% select(ENDEUDA, RES, FPIOS, ACTIVO)

muestra3.maha <-mahalanobis(muestra3.variables,
                             center = colMeans(muestra3.variables),

```

```

cov = cov(muestra3.variables))

muestra3 <- cbind(muestra3, muestra3.maha)

muestra3 <- rename(muestra3, MAHALANOBIS = muestra3.maha)

ggplot(data = muestra3, map = (aes(y = MAHALANOBIS))) +
  geom_boxplot(fill = "#457b9d", colour = "#1d3557") +
  ylab("MAHALANOBIS") +
  theme_minimal() +
  theme(axis.title = element_text(color = "#495057"),
        plot.title = element_text(color = "#495057"))

Q1M <- quantile (muestra3$MAHALANOBIS, c(0.25))
Q3M <- quantile (muestra3$MAHALANOBIS, c(0.75))

knitr::kable(muestra3 %>%
  filter(MAHALANOBIS > Q3M + 1.5*IQR(MAHALANOBIS) |
         MAHALANOBIS < Q1M - 1.5*IQR(MAHALANOBIS)) %>%
  select(MAHALANOBIS, ENDEUDA, RES, FPIOS, ACTIVO),
  caption = "Outliers de las variables ENDEUDA, RES, FPIOS y ACTIVO (incluye dis
kableExtra::kable_classic()) %>%
kable_styling(full_width = F, latex_options = c("basic", "HOLD_position"),
              "condensed", position = "center", font_size = 8)

muestra3_so <- muestra3 %>% filter(MAHALANOBIS <= Q3M +
                                1.5*IQR(MAHALANOBIS) &
                                MAHALANOBIS >= Q1M - 1.5*IQR(MAHALANOBIS))

muestra3_so_variables <- muestra3_so %>%
  select(ENDEUDA, RES, FPIOS, ACTIVO)

```

```
library(PerformanceAnalytics)

chart.Correlation(muestra3_so_variables, histogram = F, pch = 18)

tidytidy.opts = list(width.cutoff = 30)
```