

Ejercicio Evaluable Tema IV. Modelo de Regresión Múltiple (corte transversal)

Bermann, M.A. & Pérez, R.S.

31 octubre, 2021

Índice

Introducción	3
1. Análisis de posibles datos perdidos y casos atípicos en la muestra	5
1.1. Detección y tratamiento de <i>missing values</i>	5
1.2. Detección y tratamiento de <i>outliers</i>	5
2. Análisis de problemas de multicolinealidad	9
3. Análisis del modelo de regresión inicial de partida y mejora final	10
3.1. Modelo inicial de partida	10
3.2. Mejora del modelo inicial de partida	11
4. Análisis gráfico	14
5. Contraste: forma funcional	15
6. Contraste: normalidad del vector de perturbaciones aleatorias	16
6.1. Análisis gráfico de la normalidad	16
6.2. Contraste con <i>Shapiro-Wilk</i>	17

7. Contraste: homocedasticidad del vector de perturbaciones aleatorias	18
8. Sentido económico del modelo final	19
Referencias bibliográficas	20
Anexos	21
Anexo 1. Valores perdidos en la muestra inicial	21
Anexo 2. Código (<i>script</i>) utilizado	23
Anexo 3. Datos de la sesión	37

Introducción

En este informe¹ se va a proceder a desarrollar las cuestiones planteadas en el ejercicio evaluable del Tema 4 correspondiente al programa de la asignatura Técnicas Multivariantes Aplicadas al Análisis Sectorial del Máster Universitario en Modelización y Análisis de Datos Económicos (MUMADE). Para ello se va a utilizar información sobre **empresas comercializadoras de textil** españolas con el objetivo final de poder responder a las cuestiones mencionadas.

En un paso previo a comenzar el desarrollo de este informe es preciso **definir las variables**² que forman parte de la base de datos de las 319 empresas comercializadoras textiles iniciales con las que vamos a trabajar.

Cuadro 1: Definición de variables

Variable	Descripción
ACTIVO	Total Activo (mil EUR) Últ. año disp.
APALANCA	Apalancamiento (%) Últ. año disp.
CCAA	Comunidad Autónoma
EMP	Número de empleados
ENDEUDA	Endeudamiento (%) Últ. año disp.
FJUR	Forma jurídica.
FPIOS	Fondos propios (mil EUR Últ.) año disp.
ING	Ingresos de explotación (mil euros) Últ.año disp.
LIQUIDEZ	Liquidez general (%) Últ. año disp.
LOCAL	Localidad
MARGEN	Margen de beneficio (%) Últ. año disp.
NCORP	Número de empresas en el grupo corporativo.
RENECO	Rentabilidad económica (%) Últ. año disp.
RENFIN	Rentabilidad financiera (%) Últ. año disp.
RES	Resultado del Ejercicio (mil EUR) Últ. año disp.
SOLVENCIA	Coficiente de solvencia (%) Últ. año. disp.

Los datos a utilizar en este informe se basan en información que puede ser extraída de la base de datos Sabi, la cual contiene datos sobre empresas de España y Portugal (BVD 2021), habiéndose personalizado dichos datos en la

¹Para la elaboración de este informe se ha utilizado el software R, a través de su entorno RStudio y generándose la maquetación vía R Markdown. Se han utilizado numerosas fuentes para el maquetado a partir de ayudas de Allaire et al. (2021), Cano (2021), CRAN R-Project (2021), DataCamp (2021), Hlavac (2018), Keyes (2019), Kobi (2010), Luque (2019b), Luque (2019a), Van Hespén (2016), Xie, Dervieux, y Riederer (2021), Xie, Allaire, y Golemund (2021) y Zhu (2020).

²Cabe decir que en el archivo original Microsoft® Excel® viene definida la variable **GCORP** en la hoja de “Variables”, pero en la hoja de datos “GRUPO_03” no aparece como variable. En cambio sí aparece en dicha hoja “GRUPO_03” la variable **SOLVENCIA**. Es por ello por lo que se ha eliminado de la tabla de definición de variables **GCORP** y se ha incorporado **SOLVENCIA**. Por otro lado se ha procedido a transformar el formato de los valores para la variable **NCORP** en numéricos ya que la importación original los identifica como **character**.

hoja de Excel para el GRUPO_03³ y habiendo para este informe, tal y como se ha mencionado anteriormente, un total de 319 empresas como muestra inicial a estudiar.

A continuación se ha procedido a seleccionar la variable del resultado del ejercicio (**RES**) como variable métrica dependiente y total activo (**ACTIVO**), el endeudamiento (**ENDEUDA**), el número de empleados (**EMP**) y la forma jurídica (**FJUR**) como variables explicativas, donde una de ellas tiene carácter de factor (**FJUR**), con el **objetivo final de construir un modelo** donde el conjunto de explicativas sean un reflejo del comportamiento de la dependiente. Así, el modelo inicial planteado para trabajar es el siguiente:

$$RES = f(ACTIVO, ENDEUDA, EMP, FJUR)$$

Los principales datos estadísticos de las variables explicativas seleccionadas son los siguientes:

Cuadro 2: Datos estadísticos de las variables explicativas seleccionadas

ACTIVO	ENDEUDA	EMP	FJUR
Min. : 0.689	Min. : 0.00	Min. : 1.00	Length:319
1st Qu.: 50.799	1st Qu.: 57.12	1st Qu.: 1.00	Class :character
Median : 152.862	Median : 87.66	Median : 3.00	Mode :character
Mean : 446.233	Mean : 130.60	Mean : 5.15	NA
3rd Qu.: 372.114	3rd Qu.: 129.32	3rd Qu.: 6.00	NA
Max. :13451.004	Max. :2244.76	Max. :41.00	NA
NA	NA	NA's :79	NA

Dichos datos nos revelan, de forma preliminar, que **van a existir *missing values*** para la variable del número de empleados (**EMP**), así como que la variable de la forma jurídica (**FJUR**) es de tipo factor al no arrojar datos estadísticos a diferencia de las demás variables seleccionadas.

³GRUPO_03 es el nombre de la hoja del libro de Excel asignada para el informe.

1. Análisis de posibles datos perdidos y casos atípicos en la muestra

En este primer apartado se procederá a detectar y eliminar, en su caso, posibles *missing values* y *outliers* que pudieran distorsionar los resultados del análisis.

1.1. Detección y tratamiento de *missing values*

En primer lugar, el análisis de valores perdidos, tal y como habíamos adelantado en la introducción, nos ha arrojado 79 *missing values*. Debido a la cantidad elevada que son se ha trasladado la información sobre los casos que representan dichos valores perdidos en el Anexo I.

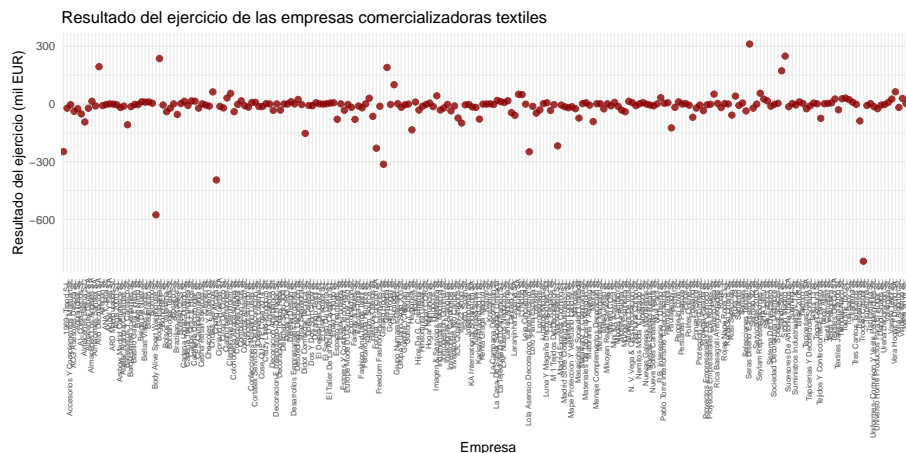
Una vez eliminados los *missing values* la muestra ha pasado de 319 observaciones a 240.

1.2. Detección y tratamiento de *outliers*

Tras el análisis y eliminación de *missing values* se va a proceder a determinar si existen outliers en la muestra. Siguiendo a Tarancón y Fernández (2021), distinguiremos este análisis de datos atípicos según si se analiza la variable dependiente o las explicativas.

1.2.1. *Outliers* en la variable dependiente

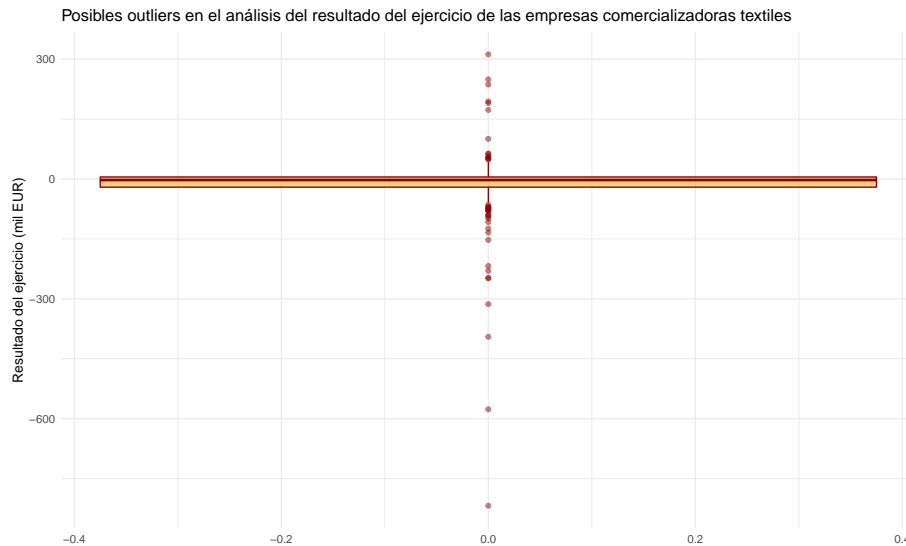
En primer lugar, para detectar si existen datos atípicos en la muestra para la variable dependiente, en este caso el resultado del ejercicio (RES), se recurrirá a un gráfico de dispersión que se muestra a continuación.



1. ANÁLISIS DE POSIBLES DATOS PERDIDOS Y CASOS ATÍPICOS EN LA MUESTRA

1.2. Detección y tratamiento de outliers

El gráfico de dispersión nos ha mostrado síntomas claros de que en la muestra existen posibles casos atípicos, por lo que se va a recurrir también al análisis gráfico *boxplot* para corroborar las “sospechas” y ayudarnos a definir a partir de qué valor vamos a considerar estos casos como verdaderos *outliers*.



El análisis *boxplot*, junto al de la dispersión de frecuencias, nos confirman la **existencia clara de outliers**. Así, siguiendo un criterio gráfico de observación, se ha procedido a eliminar estos datos atípicos considerando aquellos que su Resultado del Ejercicio (RES) tiene un valor superior e inferior a 150 (mil EUR) y que se reflejan en el Cuadro 3.

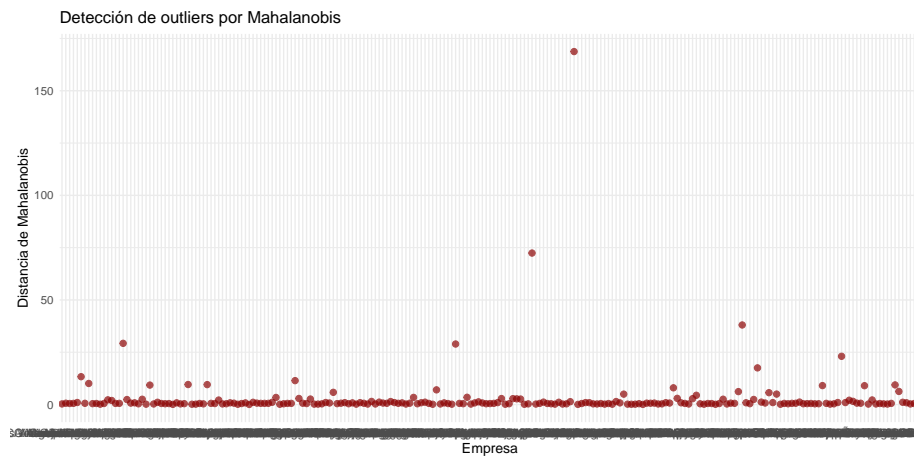
1.2.2. Outliers en las variables explicativas

Tras el análisis de la variable dependiente, es también relevante determinar si existen casos atípicos en las variables explicativas, para lo que, al haber varias variables, se recurrirá al análisis de la *distancia de Mahalanobis* que, siguiendo a Tarancón y Fernández (2021), es como un “*resumen del comportamiento de cada caso en todas las variables explicativas conjuntamente*”. Recordemos que este análisis, por el momento, solo procede a las variables cuantitativas (FJUR es una variable de tipo factor), por lo que se aplicará a las variables explicativas ACTIVO, ENDEUDA y EMP.

1. ANÁLISIS DE POSIBLES DATOS PERDIDOS Y CASOS ATÍPICOS EN
1.2. Detección y tratamiento de outliers LA MUESTRA

Cuadro 3: Casos atípicos omitidos en la muestra para la variable RES

	RES
Loro Piana España SL	-248.31
Serman 92 SL	312.00
Madrid Express Clothing SL.	-217.62
Gofer Hogar SL	190.31
Sucesores De Antonio Ubillos SA	249.46
Boggi Hispania SL	-576.31
Tropezina SL	-817.78
Boho BEF SL.	236.49
Altecon SL	194.28
Cipriano G Huertas SA	-394.91
Stoun Madrid SL.	172.85
Gemespa SL	-313.32
1969-Trend S.L.	-247.57
Doña Coletas SL	-152.62
Freedom Fashion Investments SL.	-229.91



El análisis nos arroja la posibilidad de que existan diversos *outliers* para las 3 variables explicativas analizadas, hecho que se confirma, a continuación, con el análisis gráfico *boxplot*.

1. ANÁLISIS DE POSIBLES DATOS PERDIDOS Y CASOS ATÍPICOS EN LA MUESTRA

1.2. Detección y tratamiento de outliers



Siguiendo un criterio gráfico de observación, se ha procedido a **eliminar estos datos atípicos** considerando aquellos que tienen una distancia de *Mahalanobis* superior a 20 y que se reflejan en la siguiente tabla.

Cuadro 4: Casos atípicos omitidos en la muestra para las variables ACTIVO, ENDEUDA y EMP

	ACTIVO	ENDEUDA	EMP
Sabamin SL	1554.60	5.79	34
Matarranz Y Compañia SL	13451.00	2.65	13
Barquisimeto Inversiones SL.	268.22	147.28	29
Isalto Spain SL	202.07	815.29	5
Lozueposeis SL	3.36	1230.00	1
Tiendas Naco SL	30.58	737.15	6

1.2.3. Variable explicativa factor (FJUR)

Previamente a seguir trabajando con el informe es necesario considerar que en la especificación del modelo de regresión múltiple se pretende incorporar la forma jurídica de la empresa (**FJUR**), que es una variable de tipo factor, como variable explicativa. Tal y como señala Tarancón y Fernández (2021), es necesario que **R reconozca dicha variable como un “verdadero factor”**, el cual queda definido con las siguientes 2 categorías o niveles:

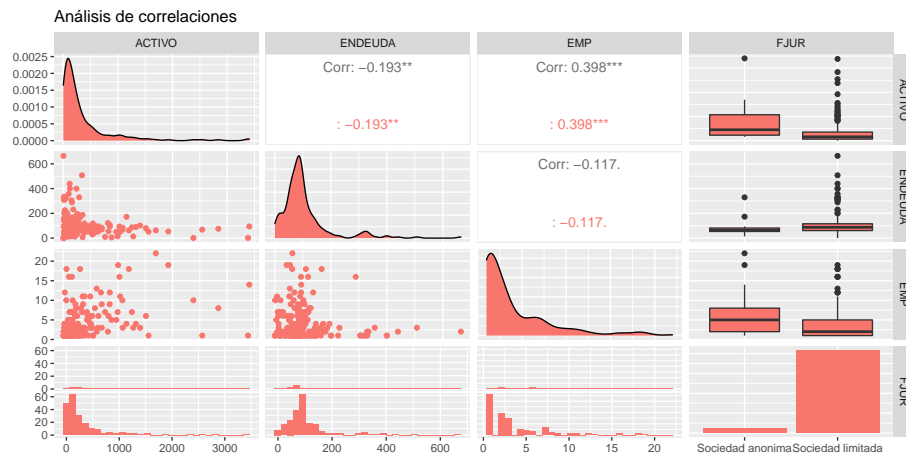
Cuadro 5: Tipos de formas jurídicas

FJUR
Sociedad anonima
Sociedad limitada

2. Análisis de problemas de multicolinealidad

En este segundo apartado, a través de las correlaciones entre las variables explicativas, se procederá a explicar razonadamente si hay alguna(s) variable(s) explicativa(s) que pudieran originar un problema de multicolinealidad en la estimación por mínimos cuadrados ordinarios (MCO) del modelo planteado.

Hay que tener en cuenta que el análisis de posibles problemas de multicolinealidad es fundamental como paso previo a la especificación y estimación del modelo. Para ello se analizará si las correlaciones son altas.



Este análisis de correlaciones arroja que podría haber cierta correlación entre la variable del total de activo (**ACTIVO**) y la variable del número de empleado (**EMP**), aunque como no es superior o 0,5 (en valor absoluto), se considera mantenerla. Parece evidente que a mayor tamaño de la empresa, representada por un mayor número de empleados, el total del activo también lo sea y por tanto, exista este problema de colinealidad. También cabe mencionar los resultados del análisis gráfico de la variable de la forma jurídica (**FJUR**), los cuales no arrojan valores de correlación al ser una variable de tipo factor. Sin embargo, vemos gráficamente que la muestra representada por las *Sociedades anónimas* es muy limitada como para trazar conclusiones y que, además, no presenta diferencias considerables respecto a las *Sociedades limitadas*.

3. Análisis del modelo de regresión inicial de partida y mejora final

En esta sección se realizará, si es posible, un proceso de mejora del modelo inicial y se comentará razonadamente el por qué se ha elegido el modelo final del proceso de mejora.

3.1. Modelo inicial de partida

Recordemos que el modelo inicial de partida planteado era el siguiente:

$$RES = f(ACTIVO, ENDEUDA, EMP, FJUR)$$

Así, una vez especificado el modelo, su estimación, a través de **mínimos cuadrados ordinarios (MCO)**, arroja los siguientes resultados:

Cuadro 6: Estimación del modelo de partida (ecua1)

	Dependent variable:
	RES
ACTIVO	0.009** (0.004)
ENDEUDA	-0.071*** (0.022)
EMP	-1.369*** (0.503)
FJUR Sociedad limitada	15.241* (8.537)
Constant	-12.287 (9.337)
Observations	219
R ²	0.104
Adjusted R ²	0.087
Residual Std. Error	27.827 (df = 214)
F Statistic	6.182*** (df = 4; 214)
Note:	*p<0.1; **p<0.05; ***p<0.01

Cabe comentar diversas cuestiones sobre dichos resultados:

- El valor del estadístico R^2 ajustado (corregido) es muy bajo (0.086).
- Los parámetros estimados son estadísticamente significativos (**ACTIVO**, **ENDEUDA**, **EMP**) (no se rechaza la hipótesis nula de significación de los parámetros individuales) excepto los niveles del factor **FJUR**, aunque tiene significación a partir de un nivel de confianza entre el 5 % y el 10 %.
- Existe significación conjunta de los parámetros estimados, la cual se analiza observando el p-valor de la prueba de la F (no se rechaza la hipótesis nula de significación conjunta de los parámetros individuales).
- Los índices de inflación de la varianza (*vif*) arrojan resultados que no superan el umbral de 4: 1.2544775, 1.040838, 1.2064335, 1.0676902.
- El criterio de información de Akaike (AIC) arroja el siguiente resultado: 2085.2216366.
- Como el factor **FJUR** se incluye en la especificación, mediante tantas variables dicotómicas o *dummies* como niveles o categorías existen, menos uno (en este caso menos la categoría *Sociedad anonima*), su resultado, siguiendo a Tarancón y Fernández (2021), debe ser interpretado en relación con el nivel o categoría que falta (*Sociedad anonima*). La interpretación, por tanto, muestra que, el hecho de ser una empresa comercializadora textil con una forma jurídica de *Sociedad limitada* significa que su resultado del ejercicio aumenta en 15.2410811 puntos respecto a una que tome como forma jurídica *Sociedad anónima*.

3.2. Mejora del modelo inicial de partida

Ante los resultados del modelo inicial, es de interés realizar una **especificación y estimación alternativa** que busque mejorar dicho modelo de partida. Para ello, considerando la posible correlación que podía existir entre las variables **ACTIVO** y **EMP**, se ha decidido suprimir dicha variable del total de activo (**ACTIVO**) e incorporar la variable del margen de beneficio (**MARGEN**)⁴.

Este modelo ha mejorado en diferentes sentidos que se recogen a continuación:

- El valor del estadístico R^2 ajustado (corregido) (0.151) sigue siendo bajo pero ha mejorado respecto al modelo anterior.
- Los parámetros estimados siguen estadísticamente significativos (**MARGEN**, **ENDEUDA**, **EMP**) (no se rechaza la hipótesis nula de significación de los parámetros individuales) excepto los niveles del factor **FJUR**.
- Sigue existiendo significación conjunta de los parámetros estimados, la cual se analiza observando el p-valor de la prueba de la F (no se rechaza la hipótesis nula de significación conjunta de los parámetros individuales).

⁴Cabe considerar que se ha procedido a eliminar los 2 *missing values* existentes para la variable **MARGEN**.

3. ANÁLISIS DEL MODELO DE REGRESIÓN INICIAL DE PARTIDA Y
3.2. Mejora del modelo inicial de partida MEJORA FINAL

Cuadro 7: Estimación del modelo corregido (ecua2)

	<i>Dependent variable:</i>
	RES
MARGEN	0.136*** (0.029)
ENDEUDA	-0.073*** (0.023)
EMP	-1.136** (0.453)
FJUR Sociedad limitada	10.799 (8.106)
Constant	-3.049 (8.754)
Observations	217
R ²	0.167
Adjusted R ²	0.151
Residual Std. Error	26.780 (df = 212)
F Statistic	10.594*** (df = 4; 212)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

3. ANÁLISIS DEL MODELO DE REGRESIÓN INICIAL DE PARTIDA Y
3.2. Mejora del modelo inicial de partida MEJORA FINAL

- Los índices de inflación de la varianza (*vif*) siguen arrojando resultados que no superan el umbral de 4: 1.0291418, 1.0327298, 1.0568666, 1.0386324.
- El criterio de información de Akaike (AIC) ha disminuido respecto al del anterior caso, arrojando ahora el siguiente resultado: 2049.6071458.
- El hecho de ser una empresa comercializadora textil con una forma jurídica de *Sociedad limitada* significa que su resultado del ejercicio aumenta en 10.7985401 puntos respecto a una que tome como forma jurídica *Sociedad anónima*.

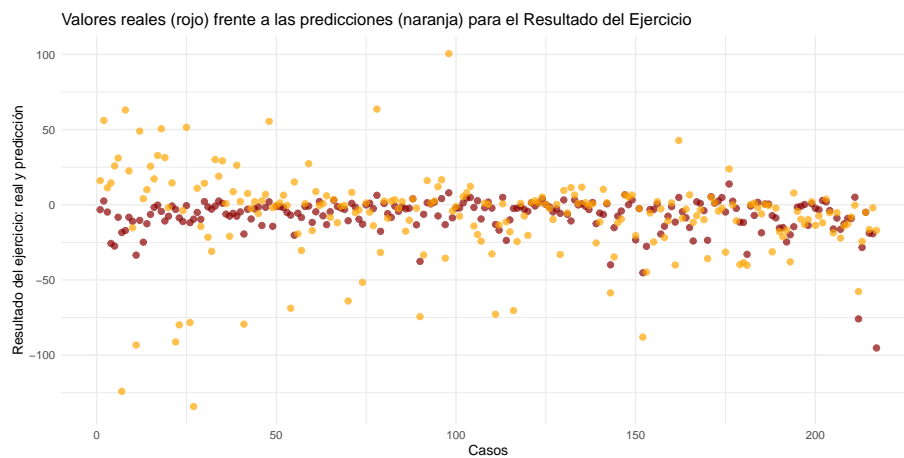
Así, el modelo final con el que se trabajará queda definido de la siguiente forma:

$$RES = f(MARGEN, ENDEUDA, EMP, FJUR)$$

4. Análisis gráfico

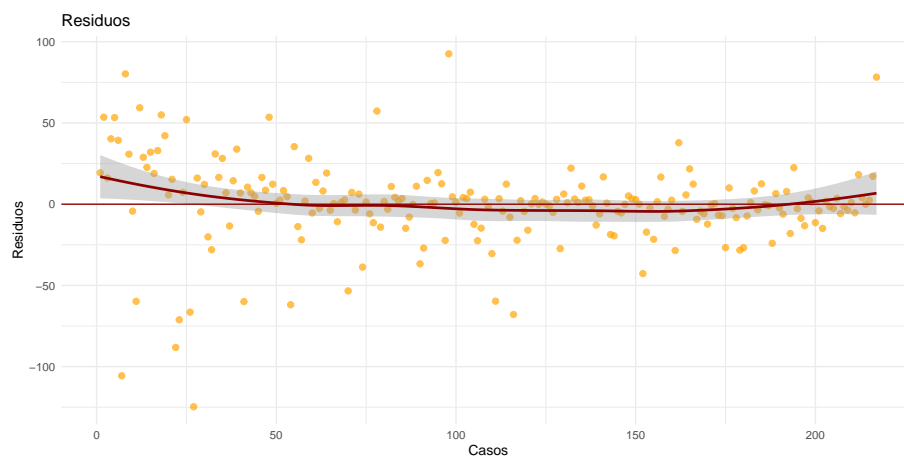
A continuación, se va a construir el gráfico variable dependiente real/estimada y el gráfico de los residuos del modelo final estimado, comentándolos sucintamente.

En primer lugar, el gráfico variable dependiente real/estimada permite comparar las predicciones del modelo estimado (`ecua2`) sobre la variable dependiente (`RES`) respecto a los valores reales.



Como se puede observar **las predicciones frente a los valores reales tienen cierta similitud**, aunque existen muchas predicciones dispersas respecto a la realidad, lo que puede deberse a las diferentes circunstancias señaladas anteriormente (por ejemplo, recordemos que el R^2 ajustado es de tan solo 0.151).

Respecto al gráfico de los residuos este se puede observar a continuación.



5. Contraste: forma funcional

En este quinto apartado se procederá al contraste, en el modelo final (`ecua2`), del cumplimiento de la **hipótesis de forma funcional** correcta de las relaciones planteadas en el modelo.

Para ello se recurrirá a la prueba de *Ramsey-RESET*, cuyo resultado rechaza la hipótesis nula de forma funcional correcta (p-valor menor que 0.05).

RESET test

```
data: .  
RESET = 22.525, df1 = 2, df2 = 210, p-value = 1.372e-09
```

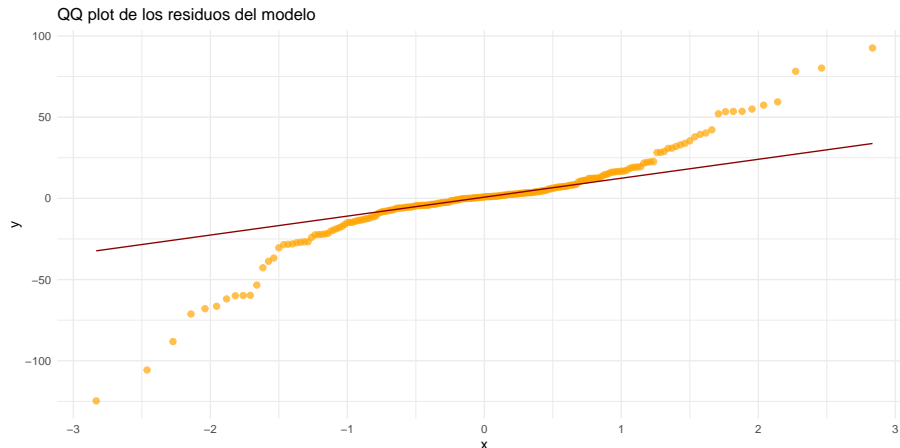
6. Contraste: normalidad del vector de perturbaciones aleatorias

A continuación se procederá a contrastar, en el modelo final (**ecua2**), el cumplimiento de la hipótesis de normalidad del vector de perturbaciones aleatorias.

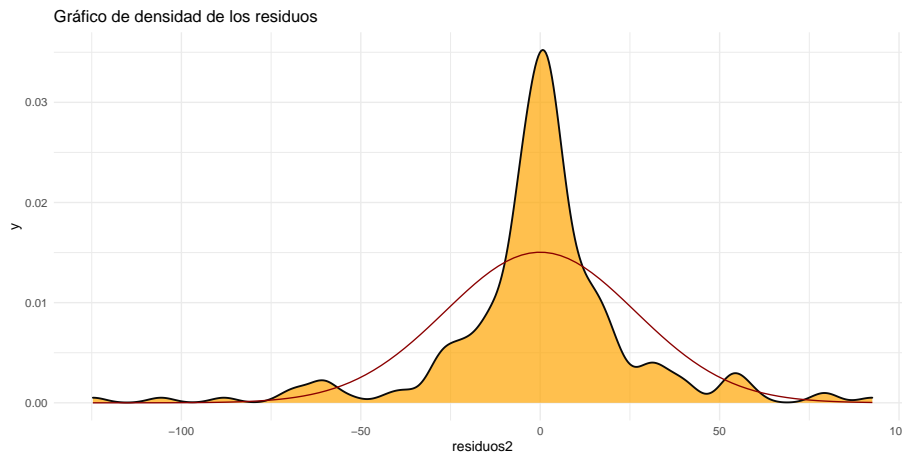
Para realizar este contraste, y siguiendo a Tarancón y Fernández (2021), se va a proceder a trabajar con los residuos del modelo (**ecua2**), considerados como estimaciones de dichas perturbaciones.

6.1. Análisis gráfico de la normalidad

Para determinar si la distribución de los residuos sigue una *ley Normal*, se recurre, en primer lugar, al gráfico *qq* (*cuantil-cuantil*), el cual compara los cuantiles teóricos que tiene una distribución normal con los que tienen la distribución de nuestra variable.



El gráfico *qq* nos muestra una considerable separación de los puntos que se encuentran en los extremos respecto a los de la línea diagonal, lo que es síntoma de “no normalidad”. De forma complementaria, un gráfico analítico de la normalidad es el de las densidades, donde se compara la función de densidad empírica de los residuos con la de la distribución normal con la misma media y desviación típica.



Parece que el gráfico muestra que la función de densidad empírica de los residuos es similar a la distribución normal (campana centrada en 0), pero junto al análisis anterior y a la excesiva función apuntada que observamos será necesario recurrir a un contraste más específico.

6.2. Contraste con *Shapiro-Wilk*

El contraste de *Shapiro-Wilk* permite contrastar estadísticamente la hipótesis de normalidad. Este, como a continuación se observa, nos confirma la *no existencia de normalidad* (p-valor menor que 0.05), al rechazar la hipótesis nula de normalidad en las perturbaciones aleatorias.

Shapiro-Wilk normality test

```
data: resultados2$residuos2
W = 0.88663, p-value = 1.054e-11
```

Siguiendo a Tarancón y Fernández (2021), la no existencia de un comportamiento normal afecta a la eficiencia de los estimadores MCO, es decir, son menos precisos (mayor varianza) (es necesario que contemos con estimadores insesgados, eficientes y consistentes, es decir, ELIO)

7. Contraste: homocedasticidad del vector de perturbaciones aleatorias

Como último análisis de contraste, en este séptimo apartado, se va a contrastar, en el modelo final, el cumplimiento de la hipótesis de homoscedasticidad del vector de perturbaciones aleatorias.

Para ello, se va a recurrir al contraste de *Breush-Pagan* y de *Goldfeld-Quandt*.

studentized Breusch-Pagan test

```
data: ecua2
BP = 43.115, df = 4, p-value = 9.793e-09
```

Goldfeld-Quandt test

```
data: ecua2
GQ = 0.47719, df1 = 104, df2 = 103, p-value = 0.9999
alternative hypothesis: variance increases from segment 1 to 2
```

En primer lugar, el test de *Breush-Pagan*, nos lleva a no rechazar la hipótesis nula de homoscedasticidad, cuestión que es totalmente contraria al resultado del test de *Goldfeld-Quandt*, el cual arroja un p-valor inferior a 0.05 y que, por tanto, rechaza la hipótesis nula de homocedasticidad, y está indicando, con ello, heterocedasticidad. Este resultado contradictorio requiere que, de forma adicional, se siguiera mejorando el modelo y se pudiera trabajar finalmente con estimadores ELIO y con un modelo mucho más cercano a la realidad del que finalmente se ha desarrollado en este informe.

8. Sentido económico del modelo final

En este octavo, y último apartado, se procederá a interpretar el significado económico de los coeficientes obtenidos en el modelo final.

- El margen de beneficio (**MARGEN**) tendrá un estimador positivo, es decir, cuanto más margen de beneficio tenga la empresa mejor resultado obtendrá. Esto es corroborado por nuestro modelo, en el que el estimador es de 0.1356469.
- El endeudamiento de una empresa (**ENDEUDA**) al conllevar un pago de intereses, afectará de manera negativa al resultado del ejercicio. Esta relación seguramente no tendría relevancia si trabajásemos con el resultado de operaciones. El estimador es -0.0730974.
- El número de empleados (**EMP**) tiene dos efectos opuestos: por un lado, es un indicador del tamaño de la empresa y por tanto de su capacidad para obtener beneficios, y por otro, al presentar gastos de salarios mayores (costes salariales), reducen el resultado. En nuestro caso, al ser empresas pequeñas, las que tengan pocos empleados será debido a que los socios propietarios se encargan de buena parte del trabajo y, por lo tanto, obtengan su retribución como beneficios de la empresa, no como salario. Es la explicación más cercana al resultado del estimador, que es -1.1360969.
- Respecto a la forma jurídica (**FJUR**), esperábamos que las *Sociedades anónimas* pudiesen conseguir un crecimiento mayor debido a la flexibilidad que ofrecen, pero ya que a penas hay, el resultado no es relevante ni siquiera al 10 %, por lo que no podemos confirmar ni desmentir el razonamiento económico que pueda existir detrás.

Referencias bibliográficas

- Allaire, J. J., Rich Iannone, Alison P. Hill, y Yihui Xie. 2021. «Distill: R Markdown Format for Scientific and Technical Writing».
- BVD. 2021. «Sabi». <https://www.bvdinfo.com/es-es/nuestros-productos/datos/nacional/sabi>.
- Cano, Emilio. 2021. «Introducción al software estadístico R». https://www.lcano.com/b/iser/%7B/_%7Dbook/index.html.
- CRAN R-Project. 2021. «The YAML Fieldguide». <https://cran.r-project.org/web/packages/ymlthis/vignettes/yaml-fieldguide.html>.
- DataCamp. 2021. «RDocumentation». <https://www.rdocumentation.org/>.
- Hlavac, Marek. 2018. «beautiful LaTeX, HTML and ASCII tables from R statistical output». <https://cran.r-project.org/web/packages/stargazer/vignettes/stargazer.pdf>.
- Keyes, David. 2019. «How to make beautiful tables in R». <https://rfortherestofus.com/2019/11/how-to-make-beautiful-tables-in-r/>.
- Kobi. 2010. «Display a (R) or (TM) symbol on SE sites». <https://meta.stackexchange.com/a/68202>.
- Luque, Pedro L. 2019a. «Cómo crear tablas de información en R Markdown». Universidad de Sevilla. http://destio.us.es/calvo/ficheros/ComoCrearTablasRMarkdown%7B/_%7DPedroLuque%7B/_%7D2019Sep%7B/_%7Dlibrodigital.pdf.
- . 2019b. «Construcción de tablas con knitr-kableExtra».
- Tarancón, Miguel Á., y Gema Fernández. 2021. «Técnicas Multivariantes Aplicadas al Análisis Sectorial». En *Máster Universitario en Modelización y Análisis de Datos Económicos (MUMADE)*.
- Van Hespén, Rossana. 2016. «Writing your thesis with R Markdown (2) – Text, citations and equations». <https://rosannavanhespen.nl/rmarkdown/writing-your-thesis-with-r-markdown-2-text-citations-and-equations/>.
- Xie, Yihui, J. J. Allaire, y Garrett Golemund. 2021. «R Markdown: The Definitive Guide». <https://bookdown.org/yihui/rmarkdown/>.
- Xie, Yihui, Christophe Dervieux, y Emily Riederer. 2021. «R Markdown Cookbook». <https://bookdown.org/yihui/rmarkdown-cookbook/>.
- Zhu, Hao. 2020. «Create awesome LaTeX table with knitr::kable and kableExtra». https://haozhu233.github.io/kableExtra/awesome%7B/_%7Dtable%7B/_%7Din%7B/_%7Dpdf.pdf.

Anexos

Anexo 1. Valores perdidos en la muestra inicial

A continuación se presentan los datos de las empresas que contienen *missing values* del total de la muestra de las 319 empresas comercializadoras textiles inicialmente importadas para su análisis.

Cuadro 8: Localización de valores perdidos en las variables seleccionadas para el modelo de regresión múltiple

	RES	ACTIVO	ENDEUDA	EMP	FJUR
Juan Sanchez Ruiz SL	1.87	3588.04	42.65	NA	Sociedad limitada
Madrid Sistemas Confort SL	-28.11	655.30	96.13	NA	Sociedad limitada
Sport&Cheap SL.	17.15	182.70	14.48	NA	Sociedad limitada
Bellido Decoracion S.L.	1.83	652.41	85.40	NA	Sociedad limitada
Cero Negativo SL	-10.00	1001.38	89.91	NA	Sociedad limitada
Stichbee SL	60.42	157.68	24.61	NA	Sociedad limitada
Mariaeugenia Amores SL	13.70	26.84	1015.38	NA	Sociedad limitada
Prosol Tejido Tecnico SL.	35.46	64.21	40.10	NA	Sociedad limitada
Next Bid SL	70.64	224.37	50.42	NA	Sociedad limitada
Seka Soluciones Integrales SL	-2.22	61.95	104.19	NA	Sociedad limitada
Gold-Wagen Styl SL	-0.63	1830.49	70.73	NA	Sociedad limitada
Millenium Home Market S.L.	-2.16	651.67	10.73	NA	Sociedad limitada
Hipertextil Mari Javi SL	2.09	188.58	118.99	NA	Sociedad limitada
Moda AI ER TE SL	-17.34	31.98	142.49	NA	Sociedad limitada
Entrecolores 130698 SL	1.52	104.75	143.74	NA	Sociedad limitada
Bleis Madrid Fashion SL.	-46.71	154.15	87.55	NA	Sociedad limitada
Sister Sierte SL.	-19.64	89.42	184.77	NA	Sociedad limitada
Textil Aranjuez SL.	1.14	16.85	102.18	NA	Sociedad limitada
Decorazona Home SL.	5.26	58.29	31.59	NA	Sociedad limitada
La Llave Hueca SL	-3.88	64.75	78.58	NA	Sociedad limitada
DPH Capital Partners, SL.	6.05	16.47	94.51	NA	Sociedad limitada
Gespostas Gestion SL.	-29.26	1592.87	92.15	NA	Sociedad limitada
Masse Textil Y Decoracion SL.	0.32	19.55	81.50	NA	Sociedad limitada
Serrano 21 SA	13.78	311.63	25.34	NA	Sociedad anonima
Comercial Carlos Gutierrez Arias SL	0.64	11.81	38.40	NA	Sociedad limitada
Luras De Amanecer S.L.	-3.03	34.24	913.73	NA	Sociedad limitada
Mundouniforme Ropa Laboral Y Complementos De Seguridad SL.	-3.36	21.02	63.31	NA	Sociedad limitada
Leader One SL.	3.27	82.06	0.61	NA	Sociedad limitada
Foupar Labores SL	2.14	176.20	138.93	NA	Sociedad limitada
S D C Carlos S.L.	-22.53	79.65	155.44	NA	Sociedad limitada
KTF Decoracion SL	-25.50	50.83	66.04	NA	Sociedad limitada
Modogar Textil SL	-9.73	182.59	62.40	NA	Sociedad limitada
Casa Añil Decoracion SL	8.22	164.82	93.13	NA	Sociedad limitada
Tandem Perfecto SL.	7.86	27.36	15.19	NA	Sociedad limitada
Noisno Brand SL.	-11.27	11.53	583.42	NA	Sociedad limitada
Andyans Decoracion Reformas Y Mantenimiento SL	3.83	59.18	2.32	NA	Sociedad limitada
Lorena Y Maria Boutique Alta Costura SLL	-17.95	70.18	45.83	NA	Sociedad limitada
Concrete Skateboards SL	3.01	19.90	16.54	NA	Sociedad limitada
Bozo&Bozo SL.	-4.64	39.20	156.07	NA	Sociedad limitada
Moda Global 2000 D.Q. SL.	-3.22	91.87	98.33	NA	Sociedad limitada
Caricia Feliz SL	-6.98	73.23	336.79	NA	Sociedad limitada
Sares Decoracion S.L.	1.49	486.08	98.42	NA	Sociedad limitada
Alondra Petite SL.	7.08	453.16	85.15	NA	Sociedad limitada
Closet Club La Latina SL.	-31.44	1.88	2244.76	NA	Sociedad limitada
Moreno Lancheros SL.	-15.22	22.28	198.29	NA	Sociedad limitada
Maria Y Carlos Decoracion SL	3.33	21.14	893.01	NA	Sociedad limitada
El Encanto De Los Bebés SL.	-7.39	66.55	87.92	NA	Sociedad limitada
DUO M Y A Coleccion SL.	1.41	94.83	132.23	NA	Sociedad limitada
Apney Spearfishing SL.	-16.53	7.55	279.62	NA	Sociedad limitada
Bustaexport SL	-38.11	31.05	326.17	NA	Sociedad limitada

continúa en la siguiente página

Cuadro 8: Localización de valores perdidos en las variables seleccionadas para el modelo de regresión múltiple continuación

	RES	ACTIVO	ENDEUDA	EMP	FJUR
Decoracion De Interiores Luz De Luna SL	-19.84	188.59	187.14	NA	Sociedad limitada
Eureka Haus SL.	-1.84	2.30	197.99	NA	Sociedad limitada
Artecrochet Online SL.	-3.20	14.86	89.00	NA	Sociedad limitada
Propuestas Laborales SL.	-13.11	12.74	193.83	NA	Sociedad limitada
Look At ME Europe SL.	1.46	69.67	134.65	NA	Sociedad limitada
Marcus Brand SL.	-11.41	24.50	1.73	NA	Sociedad limitada
Decoracion E Interiorismo Mandala SL.	-0.05	5.91	3.50	NA	Sociedad limitada
Guillecruz S.L.	-2.36	7.60	91.56	NA	Sociedad limitada
Meridien Clothing SL.	-3.46	25.05	124.51	NA	Sociedad limitada
PJ3 Comercio Agil SL.	-28.94	35.87	553.38	NA	Sociedad limitada
Paquiesma SL	6.68	160.45	106.21	NA	Sociedad limitada
Armana 3 SL.	-6.88	20.46	76.66	NA	Sociedad limitada
Badribel SL	-5.23	37.01	326.50	NA	Sociedad limitada
Webconta Aplicaciones SL.	-4.02	6.55	84.50	NA	Sociedad limitada
Guapissimissima SL.	1.79	7.79	37.21	NA	Sociedad limitada
Dibo Dreams SL.	-6.04	5.45	272.12	NA	Sociedad limitada
Inversiones Castaño Y Romeral SL.	0.14	20.14	83.49	NA	Sociedad limitada
Demas Tallas SL.	-0.41	5.76	163.31	NA	Sociedad limitada
Yadawee España SL.	-3.15	61.81	191.75	NA	Sociedad limitada
Cascia Style SL.	NA	3.00	0.00	NA	Sociedad limitada
Claumor Capital S.L.	NA	3.01	0.00	NA	Sociedad limitada
Click-Couture SL	-1.08	81.03	25.67	NA	Sociedad limitada
Creaciones Usera SL	NA	61.36	420.97	NA	Sociedad limitada
Gonibarri SL	NA	127.26	320.21	NA	Sociedad limitada
International Threads & Buttons SL.	-7.63	14.89	131.11	NA	Sociedad limitada
La Tienda Somnium On Line SL	NA	45.00	1.97	NA	Sociedad limitada
Lipizanta S.L.	-2.95	81.97	229.85	NA	Sociedad limitada
Los 3 Evils SL.	-4.76	18.58	37.67	NA	Sociedad limitada
Umami Global SL	21.53	78.60	351.70	NA	Sociedad limitada

Anexo 2. Código (*script*) utilizado

A continuación se presenta el *script* utilizado para desarrollar el informe.

```
[1] "---"
[2] "title: \"Ejercicio Evaluable Tema IV. Modelo de Regresión Múltiple (corte transversal)\"
[3] "author: \"Bermann, M.A. & Pérez, R.S.\"
[4] "lang: es"
[5] "date: \"r format(Sys.time(), '%d %B, %Y')\"
[6] "header-includes:"
[7] "- \\usepackage{fancyhdr}"
[8] "- \\pagestyle{fancy}"
[9] "- \\fancyfoot[CO,CE]{Grupo 03 - TMAAS - MUMADE}"
[10] "- \\fancyfoot[LE,RO]{\\thepage}"
[11] "- \\usepackage{titling}"
[12] "- \\pretitle{\\begin{center}}
[13] "    \\includegraphics[width=2in,height=2in]{logo_color.png}\\LARGE\\\\"
[14] "- \\posttitle{\\end{center}}
[15] "documentclass: article"
[16] "bibliography: library.bib"
[17] "output:"
[18] "  pdf_document:"
[19] "    toc: yes"
[20] "----"
[21] ""
[22] ""
[23] ""
[24] ""
[25] ""
[26] ""
[27] ""
[28] ""
[29] ""
[30] ""
[31] ""
[32] ""
[33] ""
[34] ""
[35] ""
[36] ""
[37] ""
[38] ""
[39] ""
[40] ""
[41] ""
[42] ""
```

```

[43] ""
[44] "\\newpage"
[45] ""
[46] "# Introducci3n"
[47] ""
[48] "En este informe[~1] se va a proceder a desarrollar las cuestiones planteadas en el e
[49] ""
[50] "En un paso previo a comenzar el desarrollo de este informe es preciso **definir las v
[51] ""
[52] "''{r}"
[53] "# Creaci3n de la tabla que define las variables"
[54] ""
[55] "definiciones <- read_excel(\"tmaas_evalua_04.xlsx\", sheet = \"Variables\")"
[56] ""
[57] "# Eliminaci3n de filas sin datos o que deben corregirse"
[58] "definiciones <- definiciones[-c(8,17,18), ]"
[59] ""
[60] "# Creaci3n del data.frame que incorporar3 la nueva fila"
[61] "solvencia <- data.frame(\"SOLVENCIA\", \"Coeficiente de solvencia (%) 3lt. a3o. dis
[62] ""
[63] "# Descripci3n de los datos incorporados en la nueva fila"
[64] "names(solvencia) <- c(\"Variable\", \"Descripci3n\")"
[65] ""
[66] "# Incorporaci3n de la nueva fila"
[67] "definiciones <- rbind(definiciones, solvencia)"
[68] ""
[69] "# Creaci3n de una tabla"
[70] "definiciones %>% "
[71] "  kable(booktabs = TRUE, "
[72] "        format = \"latex\","
[73] "        caption = \"Definici3n de variables\") %>% "
[74] "  kable_styling(font_size = 8,"
[75] "                latex_options = c(\"striped\", \"condensed\", \"hold_position\"), "
[76] "                position = \"center\", "
[77] "                full_width = FALSE) %>% "
[78] "  column_spec(1, bold = T, color = \"black\") %>% "
[79] "  column_spec(2, width = \"30em\")"
[80] "'''"
[81] ""
[82] "Los datos a utilizar en este informe se basan en informaci3n que puede ser extra3-
da de la base de datos Sabi, la cual contiene datos sobre empresas de Espa3a y Portugal [O
[83] ""
[84] "[~1]: Para la elaboraci3n de este informe se ha utilizado el software R, a trav3s d
a R Markdown. Se han utilizado numerosas fuentes para el maquetado a partir de ayudas de @A
[85] ""
[86] "[~2]: Cabe decir que en el archivo original Microsoft&reg; Excel&reg; viene definida

```


aparece en dicha hoja \"GRUPO_03\" la variable 'SOLVENCIA'. Es por ello por lo que se ha el

```

[87] ""
[88] "[^3]: GRUPO_03 es el nombre de la hoja del libro de Excel asignada para el informe."
[89] ""
[90] ""{r}"
[91] "# NCORP como variable numérica y no cualitativa"
[92] ""
[93] "comercializadoras$NCORP <- as.numeric(comercializadoras$NCORP)"
[94] ""
[95] ""
[96] "A continuación se ha procedido a seleccionar la variable del resultado del ejercicio
dica ('FJUR') como variables explicativas, donde una de ellas tiene carácter de factor ('FJUR'),
, el modelo inicial planteado para trabajar es el siguiente:"
[97] ""
[98] "\\begin{center}"
[99] "$RES = f(ACTIVO, ENDEUDA, EMP, FJUR)$"
[100] "\\end{center}"
[101] ""
[102] "Los principales datos estadísticos de las variables explicativas seleccionadas son los
[103] ""
[104] ""{r}"
[105] "# Selección de las variables explicativas"
[106] "explicativas <- select(comercializadoras, ACTIVO, ENDEUDA, EMP, FJUR)"
[107] ""
[108] "# Creación de una tabla con las variables explicativas"
[109] "explicativas %>% "
[110] "  summary() %>% "
[111] "  kable(booktabs = TRUE, "
[112] "        format = \"latex\", "
[113] "        caption = \"Datos estadísticos de las variables explicativas seleccionadas\", "
[114] "        digits = 2) %>% "
[115] "  kable_styling(font_size = 8, "
[116] "                latex_options = c(\"striped\", \"condensed\", \"hold_position\"), "
[117] "                position = \"center\", "
[118] "                full_width = F) %>% "
[119] "  row_spec(0, bold = T, color = \"black\")"
[120] ""
[121] ""
[122] "Dichos datos nos revelan, de forma preliminar, que **van a existir _missing values_**
como que la variable de la forma jurídica ('FJUR') es de tipo factor al no arrojar datos estadísticos
a diferencia de las demás variables seleccionadas."
[123] ""
[124] "\\newpage"
[125] ""
[126] "# 1. Análisis de posibles datos perdidos y casos atípicos en la muestra"

```

```

[127] ""
[128] "En este primer apartado se procederá a detectar y eliminar, en su caso, posibles _m
[129] ""
[130] "## 1.1. Detección y tratamiento de _missing values_"
[131] ""
[132] "```{r, include = FALSE}"
[133] "# Nuevo vector para conservar el original de la muestra"
[134] "muestra_1 <- select(comercializadoras, everything())"
[135] ""
[136] "# Detección de missing values"
[137] "muestra_1 %>% "
[138] "  filter(is.na(RES) | is.na(ACTIVO) | is.na(ENDEUDA) | is.na(EMP) | is.na(FJUR))"
[139] "```"
[140] ""
[141] "En primer lugar, el análisis de valores perdidos, tal y como habíamos
[142] amos adelantado en la introducción, nos ha arrojado 'r muestra_1 %>% filter(is.na(RES) | is
[143] [142] ""
[143] "```{r, include = FALSE}"
[144] "# Eliminación de los _missing values_"
[145] "muestra_1_1 <- muestra_1 %>% filter(! is.na(RES) & ! is.na(ACTIVO) & ! is.na(ENDEUDA)
[146] [145] ""
[146] "```"
[147] ""
[148] "Una vez eliminados los _missing values_ **la muestra ha pasado de 'r count(comercial
[149] [148] ""
[150] "## 1.2. Detección y tratamiento de _outliers_"
[151] ""
[152] "Tras el análisis y eliminación de _missing values_ se va a proceder a determinar si
[153] picos según si se analiza la variable dependiente o las explicativas."
[154] [152] ""
[154] "### 1.2.1. _Outliers_ en la variable dependiente"
[155] [153] ""
[155] ""
[156] "En primer lugar, para detectar si existen datos atípicos
[157] en la muestra para la variable dependiente, en este caso el resultado del ejercicio (
[158] [156] ""
[158] "```{r, fig.align = 'center', fig.height = 5, fig.width = 10}"
[159] "# Gráfico de dispersión para la detección de outliers"
[160] [157] ""
[160] "muestra_1_1 %>% "
[161] [158] "  ggplot(aes(x = row.names(muestra_1_1), y = RES)) +"
[162] [159] "  geom_point(size = 2, alpha = 0.8, colour = 'red4') +"
[163] [160] "  xlab('Empresa') +"
[164] [161] "  ylab('Resultado del ejercicio (mil EUR)') +"
[165] [162] "  ggtitle('Resultado del ejercicio de las empresas comercializadoras textiles') +"
[166] [163] "  theme_minimal() +"
[167] [164] "  theme(axis.text.x = element_text(angle = 90, size = 6, hjust = 1, vjust = 1))"
[168] [165] "```"
[169] [166] ""

```

```

[170] "El gráfico de dispersión nos ha mostrado síntomas claros de que en la muestra exist
[171] ""
[172] ""{r, fig.align = 'center', fig.height = 6, fig.width = 10}"
[173] "# Gráfico boxplot para la detección de outliers"
[174] "muestra_1_1 %>% "
[175] "  ggplot(aes(y = RES)) +"
[176] "  geom_boxplot(alpha = 0.5, fill = \"orange\", color = \"red4\") +"
[177] "  ylab('Resultado del ejercicio (mil EUR)') +"
[178] "  ggtitle('Posibles outliers en el análisis del resultado del ejercicio de las empre
[179] "  theme_minimal()"
[180] ""
[181] ""
[182] "El análisis_boxplot_, junto al de la dispersión de frecuencias, nos confirman la *
, siguiendo un criterio gráfico de observación, se ha procedido a eliminar estos datos at
picos considerando aquellos que su Resultado del Ejercicio ('RES') tiene un valor superior e
[183] ""
[184] ""{r}"
[185] "# Detección de outliers en la variable RES"
[186] "muestra_1_1 %>% "
[187] "  filter(RES >= 150 | RES <= -150) %>% "
[188] "  select(RES) %>% "
[189] "  kable(booktabs = TRUE, "
[190] "        format = \"latex\", "
[191] "        caption = \"Casos atípicos omitidos en la muestra para la variable RES\", "
[192] "        digits = 2) %>% "
[193] "  kable_styling(font_size = 7, "
[194] "                latex_options = c(\"striped\", \"condensed\", \"hold_position\"), "
[195] "                position = \"center\", "
[196] "                full_width = F) %>% "
[197] "  row_spec(0, bold = T, color = \"black\")"
[198] ""
[199] "# Eliminación de outliers en la variable RES"
[200] "muestra_1_2 <- muestra_1_1 %>% "
[201] "  filter(RES < 150 & RES > -150)"
[202] ""
[203] ""
[204] "### 1.2.2. _Outliers_ en las variables explicativas"
[205] ""
[206] "Tras el análisis de la variable dependiente, es también relevante determinar si ex
picos en las variables explicativas, para lo que, al haber varias variables, se recurrirá a
[207] ""
[208] ""{r, fig.align = 'center', fig.height = 5, fig.width = 10}"
[209] "# Creación del vector de la distancia de Mahalanobis"
[210] "muestra_1_2_distancias_maha <- mahalanobis(muestra_1_2[,c(9,14,15)],"
[211] "                                             center = colMeans(muestra_1_2[,c(9,14,15)])"

```

```

[212] "                                cov = cov(muestra_1_2[,c(9,14,15)]))"
[213] ""
[214] "# Gráfica de la distancia de Mahalanobis"
[215] "muestra_1_2 %>% "
[216] "  ggplot(aes(x = row.names(muestra_1_2), y = muestra_1_2_distancias_maha)) +"
[217] "  geom_point(size = 2, alpha = 0.7, color = 'red4') +"
[218] "  xlab('Empresa') +"
[219] "  ylab('Distancia de Mahalanobis') +"
[220] "  ggtitle('Detección de outliers por Mahalanobis') +"
[221] "  theme(axis.text.x = element_text(angle = 90, size = 6,hjust = 1, vjust = 1)) +"
[222] "  theme_minimal()"
[223] "'''"
[224] ""
[225] "El análisis nos arroja la posibilidad de que existan diversos _outliers_ para las 3"
[226] ""
[227] "'''{r, fig.align = 'center', fig.height = 4, fig.width = 10}"
[228] "# Detectando outliers"
[229] "muestra_1_2 %>% "
[230] "  ggplot(aes(y = muestra_1_2_distancias_maha)) +"
[231] "  geom_boxplot(alpha = 0.5, fill = \"orange\", color = \"red4\") +"
[232] "  ylab('Distancia de Mahalanobis') +"
[233] "  ggtitle('Posibles outliers para el análisis conjunto de variables explicativas') +"
[234] "  theme_minimal()"
[235] "'''"
[236] ""
[237] "Siguiendo un criterio gráfico de observación, se ha procedido a **eliminar estos d"
[238] "picos** considerando aquellos que tienen una distancia de _Mahalanobis_ superior a 20 y que"
[239] "'''{r}"
[240] "muestra_1_2 %>% "
[241] "  filter(muestra_1_2_distancias_maha >= 20) %>% "
[242] "  select(ACTIVO, ENDEUDA, EMP) %>% "
[243] "  kable(booktabs = TRUE, "
[244] "        format = \"latex\", "
[245] "        caption = \"Casos atípicos omitidos en la muestra para las variables ACTIVO,"
[246] "        digits = 2) %>% "
[247] "  kable_styling(font_size = 8, "
[248] "               latex_options = c(\"striped\", \"condensed\", \"hold_position\"), "
[249] "               position = \"center\", "
[250] "               full_width = F) %>% "
[251] "  row_spec(0, bold = T, color = \"black\")"
[252] ""
[253] "# Eliminando los outliers"
[254] "muestra_1_3 <- muestra_1_2 %>% "
[255] "  filter(muestra_1_2_distancias_maha < 20)"
[256] "'''"

```

```

[257] ""
[258] "### 1.2.3. Variable explicativa factor ('FJUR')"
[259] ""
[260] "Previamente a seguir trabajando con el informe es necesario considerar que en la espe
dica de la empresa ('FJUR'), que es una variable de tipo factor, como variable explicativa.
as o niveles:"
[261] ""
[262] ""
[263] "# Transformando el data.frame en factor"
[264] "muestra_1_3$FJUR <- as.factor(muestra_1_3$FJUR)"
[265] ""
[266] "# Mostrando los niveles de la variable FJUR"
[267] "muestra_1_3$FJUR %>% "
[268] "  levels() %>% "
[269] "  kable(col.names = 'FJUR',"
[270] "         booktabs = TRUE, "
[271] "         format = \"latex\", "
[272] "         caption = 'Tipos de formas jurÁdicas') %>% "
[273] "  kable_styling(font_size = 8,"
[274] "                latex_options = c(\"striped\", \"condensed\", \"hold_position\"), "
[275] "                position = \"center\", "
[276] "                full_width = T) %>% "
[277] "  row_spec(0, bold = T, color = \"black\")"
[278] ""
[279] ""
[280] "\\newpage"
[281] ""
[282] "# 2. Análisis de problemas de multicolinealidad"
[283] ""
[284] "En este segundo apartado, a través de las correlaciones entre las variables explicat
nimos cuadrados ordinarios (MCO) del modelo planteado."
[285] ""
[286] "Hay que tener en cuenta que el análisis de posibles problemas de multicolinealidad e
[287] ""
[288] ""
[289] "# Selección de las variables del análisis de correlaci³n"
[290] "muestra_1_3_cor <- muestra_1_3 %>% "
[291] "  select(ACTIVO, ENDEUDA, EMP, FJUR)"
[292] ""
[293] "# Análisis de correlaci³n"
[294] "ggpairs(muestra_1_3_cor,"
[295] "        title = \"Análisis de correlaciones\", "
[296] "        mapping = aes(colour = ''))"
[297] ""
[298] ""
[299] "Este análisis de correlaciones arroja que podrÁa haber cierta correlaci³n entre la

```

```

dica ('FJUR'), los cuales no arrojan valores de correlación al ser una variable de tipo fa
[300] ""
[301] "\\newpage"
[302] ""
[303] "# 3. Análisis del modelo de regresión inicial de partida y mejora final"
[304] ""
[305] "En esta sección se realizará, si es posible, un proceso de mejora del modelo inicial
[306] ""
[307] "## 3.1. Modelo inicial de partida"
[308] ""
[309] "Recordemos que el modelo inicial de partida planteado era el siguiente:"
[310] ""
[311] "\\begin{center}"
[312] "$RES = f(ACTIVO, ENDEUDA, EMP, FJUR)$"
[313] "\\end{center}"
[314] ""
[315] "Así, una vez especificado el modelo, su estimación, a través de mínimos
cuadrados ordinarios (MCO)**, arroja los siguientes resultados:"
[316] ""
[317] "```{r, comment = '', results= 'asis'}"
[318] "# Estimación del modelo inicial (ecua1)"
[319] "ecua1 <- lm (data = muestra_1_3, RES ~ ACTIVO + ENDEUDA + EMP + FJUR)"
[320] "ecua1 %>% "
[321] "  stargazer(header = F,"
[322] "            title = \"Estimación del modelo de partida (ecua1)\")"
[323] "```"
[324] ""
[325] "Cabe comentar diversas cuestiones sobre dichos resultados:"
[326] ""
[327] " + El valor del estadístico  $R^2$  ajustado (corregido) es muy bajo (0.086)."
[328] " + Los parámetros estimados son estadísticamente significativos ('ACTIVO', 'ENDEUDA'
[329] " + Existe significación conjunta de los parámetros estimados, la cual se analiza c
[330] " + Los Índices de inflación de la varianza (_vif_) arrojan resultados que no supera
[331] " + El criterio de información de Akaike (AIC) arroja el siguiente resultado: 'r ec
[332] " + Como el factor 'FJUR' se incluye en la especificación, mediante tantas variables
as existen, menos uno (en este caso menos la categoría _Sociedad anonima_), su resultado, si
a que falta (*Sociedad anonima*). La interpretación, por tanto, muestra que, el hecho de se
dica de _Sociedad limitada_ significa que su resultado del ejercicio aumenta en 'r ecua1$cc
dica _Sociedad anónima_."
[333] ""
[334] "## 3.2. Mejora del modelo inicial de partida"
[335] ""
[336] "Ante los resultados del modelo inicial, es de interés realizar una **especificación
a existir entre las variables 'ACTIVO' y 'EMP', se ha decidido suprimir dicha variable del t
[337] ""
[338] "```{r, comment = '', results= 'asis'}"

```

```

[339] "# Eliminaci3n de missing values en la nueva variable del modelo"
[340] "muestra_1_4 <- muestra_1_3 %>% filter(! is.na(MARGEN))"
[341] ""
[342] "# Estimaci3n del segundo modelo"
[343] "ecua2 <- lm (data = muestra_1_4, RES ~ MARGEN + ENDEUDA + EMP + FJUR)"
[344] "ecua2 %>% "
[345] "  stargazer(header = F,"
[346] "            title = \"Estimaci3n del modelo corregido (ecua2)\")"
[347] "'''"
[348] ""
[349] "[^4]: Cabe considerar que se ha procedido a eliminar los 'r muestra_1_3 %>% filter(
[350] ""
[351] "***Este modelo ha mejorado** en diferentes sentidos que se recogen a continuaci3n:"
[352] ""
[353] "  + El valor del estadístico R^2 ajustado (corregido) (0.151) sigue siendo bajo pero
[354] "  + Los parámetros estimados siguen estadísticamente significativos ('MARGEN', 'ENDE
[355] "  + Sigue existiendo significaci3n conjunta de los parámetros estimados, la cual se
[356] "  + Los Índices de inflaci3n de la varianza (_vif_) siguen arrojando resultados que
[357] "  + El criterio de informaci3n de Akaike (AIC) ha disminuido respecto al del anterior
[358] "  + El hecho de ser una empresa comercializadora textil con una forma jurÁ-
dica de _Sociedad limitada_ significa que su resultado del ejercicio aumenta en 'r ecua2$co
dica _Sociedad anónima_."
[359] ""
[360] "Así, el modelo final con el que se trabajarÁ; queda definido de la siguiente forma:"
[361] ""
[362] "\\begin{center}"
[363] "$RES = f(MARGEN, ENDEUDA, EMP, FJUR)$"
[364] "\\end{center}"
[365] ""
[366] "\\newpage"
[367] ""
[368] "# 4. Análisis gráfico"
[369] ""
[370] "A continuaci3n, se va a construir el gráfico variable dependiente real/estimada y e
[371] ""
[372] "En primer lugar, el gráfico variable dependiente real/estimada permite comparar las
[373] ""
[374] "'''{r, fig.align = 'center', fig.height = 5, fig.width = 10}"
[375] "# Creaci3n del vector con las predicciones"
[376] "numcasos <- nrow(muestra_1_4)"
[377] "ORDEN <- (c(1:217))"
[378] "resultados2 <-data.frame(RES_predic2 = ecua2$fitted.values,"
[379] "                          residuos2 = ecua2$residuals,"
[380] "                          RES = muestra_1_4$RES,"
[381] "                          ORDEN = c(1:217))"
[382] ""

```

```

[383] "# Gráfico realidad vs. predicción"
[384] "muestra_1_4 %>% "
[385] "  ggplot() +"
[386] "    geom_point(aes(x = ORDEN, y = ecua2$fitted.values),"
[387] "                size = 2, alpha = 0.7, color = \"red4\") + "
[388] "    geom_point(aes(x = ORDEN, y = muestra_1_4$RES),"
[389] "                size = 2, alpha = 0.7, color = \"orange\") + "
[390] "    ggtitle(\"Valores reales (rojo) frente a las predicciones (naranja) para el Resulta")
[391] "    xlab('Casos') + "
[392] "    ylab('Resultado del ejercicio: real y predicción') + "
[393] "    theme(axis.text.x = element_text(angle = 90, "
[394] "                                          size = 6, "
[395] "                                          hjust = 1, "
[396] "                                          vjust = 1)) + "
[397] "    theme_minimal()"
[398] "'''"
[399] ""
[400] "Como se puede observar **las predicciones frente a los valores reales tienen cierta s"
[401] ""
[402] "Respecto al gráfico de los residuos este se puede observar a continuaci3n."
[403] ""
[404] "'''{r, fig.align = 'center', fig.height = 5, fig.width = 10}"
[405] "# Gráfico de residuos"
[406] "muestra_1_4 %>% "
[407] "  ggplot(aes(x = ORDEN, y = ecua2$residuals)) +"
[408] "    geom_point(size = 2, "
[409] "               alpha= 0.7, "
[410] "               color = \"orange\") +"
[411] "    geom_smooth(color = \"red4\") +"
[412] "    geom_hline(yintercept = 0, "
[413] "               color = \"red4\")+"
[414] "    ggtitle(\"Residuos\")+"
[415] "    xlab('Casos') +"
[416] "    ylab('Residuos') +"
[417] "    theme_minimal()"
[418] "'''"
[419] ""
[420] "\\newpage"
[421] ""
[422] "# 5. Contraste: forma funcional"
[423] ""
[424] "En este quinto apartado se procederá al contraste, en el modelo final ('ecua2'), del"
[425] ""
[426] "Para ello se recurrirá a la prueba de _Ramsey-RESET_, cuyo resultado rechaza la hip3"
[427] ""
[428] "'''{r, comment = ''}"

```



```

[429] "# Prueba de Ramsey-RESET"
[430] "ecua2 %>% "
[431] "  resettest(data = muestra_1_4)"
[432] "'''"
[433] ""
[434] "\\newpage"
[435] ""
[436] "# 6. Contraste: normalidad del vector de perturbaciones aleatorias"
[437] ""
[438] "A continuaci3n se proceder3 a contrastar, en el modelo final ('ecua2'), el cumplimiento
[439] ""
[440] "Para realizar este contraste, y siguiendo a @Tarancon2021, se va a proceder a trabajar
[441] ""
[442] "## 6.1. An3lisis gr3fico de la normalidad"
[443] ""
[444] "Para determinar si la distribuci3n de los residuos sigue una _ley Normal_, se recurrir3
[445] ""
[446] "'''{r, fig.align = 'center', fig.height = 5, fig.width = 10}"
[447] "# Contraste de normalidad: gr3fico qq"
[448] "resultados2 %>% "
[449] "  ggplot(aes(sample = residuos2)) +"
[450] "    stat_qq(size = 2,"
[451] "           alpha= 0.7, "
[452] "           color = \"orange\") + "
[453] "    stat_qq_line( color = \"red4\") + "
[454] "    ggtitle('QQ plot de los residuos del modelo') + "
[455] "    theme_minimal()"
[456] "'''"
[457] ""
[458] "El gr3fico _qq_ nos muestra una considerable separaci3n de los puntos que se encuentran
[459] ""
[460] "'''{r, fig.align = 'center', fig.height = 5, fig.width = 10}"
[461] "# Contraste de normalidad: gr3fico de densidad"
[462] "resultados2 %>% "
[463] "  ggplot(aes(residuos2)) +"
[464] "    geom_density(alpha = 0.7, "
[465] "                fill = \"orange\", "
[466] "                size = 0.7, "
[467] "                color = \"grey4\") + "
[468] "    ggtitle('Gr3fico de densidad de los residuos') + "
[469] "    stat_function(fun = dnorm, "
[470] "                 color= \"red4\", "

```

```

[471] "                args = list(mean = mean(resultados2$residuos2), "
[472] "                        sd = sd(resultados2$residuos2))) +"
[473] "  theme_minimal()"
[474] "'''"
[475] ""
[476] "Parece que el gráfico muestra que la función de densidad empí-
rica de los residuos es similar a la distribución normal (campana centrada en 0), pero jun-
fico."
[477] ""
[478] "## 6.2. Contraste con _Shapiro-Wilk_"
[479] ""
[480] "El contraste de _Shapiro-Wilk_ permite contrastar estadísticamente la hipótesis de normalidad. Este, como a continuación se observa, nos confirma
[481] ""
[482] "'''{r, comment = ''}"
[483] "# Test de Shapiro-Wilk"
[484] "shapiro.test(x = resultados2$residuos2)"
[485] "'''"
[486] ""
[487] "Siguiendo a @Tarancon2021, la no existencia de un comportamiento normal afecta a la e
[488] ""
[489] "\\newpage"
[490] ""
[491] "# 7. Contraste: homocedasticidad del vector de perturbaciones aleatorias"
[492] ""
[493] "Como último análisis de contraste, en este último apartado, se va a contrastar, e
[494] ""
[495] "Para ello, se va a recurrir al contraste de _Breush-Pagan_ y de _Goldfeld-Quandt_."
[496] ""
[497] "'''{r, comment = ''}"
[498] "# Test de Breush-Pagan y Goldfeld-Quandt"
[499] "bptest(ecua2)"
[500] "gqtest(ecua2)"
[501] "'''"
[502] ""
[503] "En primer lugar, el test de _Breush-Pagan_, nos lleva a no rechazar la hipótesis nul
[504] ""
[505] "\\newpage"
[506] ""
[507] "# 8. Sentido económico del modelo final"
[508] ""
[509] "En este octavo, y último apartado, se procederá a interpretar el significado econó
[510] ""
[511] " + El margen de beneficio ('MARGEN') tendrá un estimador positivo, es decir, cuanto
[512] " + El endeudamiento de una empresa ('ENDEUDA') al conllevar un pago de intereses, af
a relevancia si trabajamos con el resultado de operaciones. El estimador es 'r ecua2$coe

```

```

[513] " + El número de empleados ('EMP') tiene dos efectos opuestos: por un lado, es un in
[514] " + Respecto a la forma jurídica ('FJUR'), esperamos que las _Sociedades anónimas
[515] ""
[516] "\\newpage"
[517] ""
[518] "# Referencias bibliográficas"
[519] ""
[520] "<div id=\"refs\"></div>"
[521] ""
[522] "\\newpage"
[523] ""
[524] "# Anexos"
[525] ""
[526] "## Anexo 1. Valores perdidos en la muestra inicial"
[527] ""
[528] "A continuación se presentan los datos de las empresas que contienen _missing values.
[529] ""
[530] "{'r}"
[531] "#Valores perdidos de la muestra inicial"
[532] "muestra_1 %>% "
[533] "  filter(is.na(RES) | is.na(ACTIVO) | is.na(ENDEUDA) | is.na(EMP) | is.na(FJUR)) %>%
[534] "  select(RES, ACTIVO, ENDEUDA, EMP, FJUR) %>% "
[535] "  kable(booktabs = TRUE,"
[536] "        format = \"latex\", "
[537] "        digits = 2,"
[538] "        longtable=TRUE,"
[539] "        caption = \"Localización de valores perdidos en las variables seleccionadas
[540] "  kable_styling(font_size = 5,"
[541] "               latex_options = c(\"striped\", "
[542] "                                \"hold_position\", "
[543] "                                \"repeat_header\", "
[544] "                                \"consensed\"), "
[545] "               full_width = F,"
[546] "               position = \"center\", "
[547] "               repeat_header_text = \"continuación\\'on\", "
[548] "               repeat_header_continued = \"continúa en la siguiente página\\'agina
[549] "{''"
[550] ""
[551] "\\newpage"
[552] ""
[553] "## Anexo 2. Código (_script_) utilizado"
[554] ""
[555] "A continuación se presenta el _script_ utilizado para desarrollar el informe."
[556] ""
[557] "{'{r, echo = FALSE, comment= '}"
[558] "script <- readLines(\"TMAAS_04.Rmd\")"

```

```
[559] "print(script)"
[560] "'''"
[561] ""
[562] "\\newpage"
[563] ""
[564] "## Anexo 3. Datos de la sesión"
[565] ""
[566] "En esta sección se recogen los datos de la sesión utilizada para elaborar este informe
como las versiones de los paquetes bajo los cuales se ha ejecutado el código o _script_."
[567] ""
[568] "'''{r, echo = FALSE, comment = ''}"
[569] "sessionInfo()"
[570] "'''"
```

Anexo 3. Datos de la sesión

En esta sección se recogen los datos de la sesión utilizada para elaborar este informe. Siguiendo a Cano (2021), es fundamental observar la versión de R, así como las versiones de los paquetes bajo los cuales se ha ejecutado el código o *script*.

```
R version 4.1.1 (2021-08-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19042)
```

```
Matrix products: default
```

```
locale:
```

```
[1] LC_COLLATE=Spanish_Spain.1252 LC_CTYPE=Spanish_Spain.1252
[3] LC_MONETARY=Spanish_Spain.1252 LC_NUMERIC=C
[5] LC_TIME=Spanish_Spain.1252
```

```
attached base packages:
```

```
[1] stats      graphics  grDevices  utils      datasets  methods    base
```

```
other attached packages:
```

```
[1] lmtree_0.9-38      zoo_1.8-9          stargazer_5.2.2    car_3.0-11
[5] carData_3.0-4      GGally_2.1.2       ggplot2_3.3.5      knitr_1.36
[9] kableExtra_1.3.4   dplyr_1.0.7        tidyr_1.1.4        readxl_1.3.1
```

```
loaded via a namespace (and not attached):
```

```
[1] Rcpp_1.0.7          svglite_2.0.0       lattice_0.20-44     assertthat_0.2.1
[5] digest_0.6.28       utf8_1.2.2          R6_2.5.1            cellranger_1.1.0
[9] plyr_1.8.6          evaluate_0.14       httr_1.4.2          pillar_1.6.3
[13] rlang_0.4.11        curl_4.3.2          rstudioapi_0.13     data.table_1.14.2
[17] Matrix_1.3-4        rmarkdown_2.11      splines_4.1.1       labeling_0.4.2
[21] webshot_0.5.2       stringr_1.4.0       foreign_0.8-81      munsell_0.5.0
[25] compiler_4.1.1      xfun_0.26           pkgconfig_2.0.3     systemfonts_1.0.3
[29] mgcv_1.8-36         htmltools_0.5.2     tidyselect_1.1.1    tibble_3.1.4
[33] rio_0.5.27          reshape_0.8.8       fansi_0.5.0         viridisLite_0.4.0
[37] crayon_1.4.1        withr_2.4.2         grid_4.1.1          nlme_3.1-152
[41] gtable_0.3.0        lifecycle_1.0.1     DBI_1.1.1           magrittr_2.0.1
[45] scales_1.1.1        zip_2.2.0           stringi_1.7.4       farver_2.1.0
[49] xml2_1.3.2          ellipsis_0.3.2      generics_0.1.0      vctrs_0.3.8
[53] openxlsx_4.2.4      RColorBrewer_1.1-2  tools_4.1.1         forcats_0.5.1
[57] glue_1.4.2          purrr_0.3.4         hms_1.1.1           abind_1.4-5
[61] fastmap_1.1.0       yaml_2.2.1          colorspace_2.0-2    rvest_1.0.1
[65] haven_2.4.3
```