

**Estudio econométrico del método de cálculo óptimo para
la transformación de unidades de calzado en metros
cúbicos.**

**TRABAJO FIN DE GRADO
CURSO 2013-14**



Autor: Francisco David García Agulló

Tutor: Juan Aparicio Baeza

Fecha de entrega: 18 de junio de 2014

Grado de Administración y Dirección de Empresa

Facultad CC Sociales y Jurídicas de Orihuela

Universidad Miguel Hernández

ÍNDICE GENERAL

1. RESUMEN	3
2. INTRODUCCIÓN	4
1.1. Para quien se va a realizar este estudio	4
3. OBJETIVOS	5
4. METODOLOGÍA	5
5. ESTUDIO ECONOMÉTRICO	6
a. ESTUDIO ECONOMÉTRICO PTR	7
b. ESTUDIO ECONOMÉTRICO MOD	12
c. ESTUDIO ECONOMÉTRICO DENS	16
d. ESTUDIO ECONOMÉTRICO BLT	21
e. ESTUDIO ECONOMÉTRICO DE TODOS LOS MÉTODOS	25
6. PREDICCIÓN DEL VOLUMEN CON DATOS DE 2014	31
7. CONCLUSIONES	31
8. BIBLIOGRAFIA	33

1. RESUMEN

El objetivo de este estudio es determinar cuál es el método de cálculo que más se aproxima a la realidad para la transformación de Uds. de calzado en m^3 . La fuente de información utilizada es obtenida de las bases de datos de Tempe, S.A. Grupo INDITEX.

Una vez analizadas las diferentes formas de cálculo propuestas y la utilizada actualmente podemos confirmar que el modelo de cálculo más óptimo es el método llamado DENS. Este método es el calculado a partir de las uds. facturadas entre la densidad (Uds./ m^3) enviada a la misma tienda en el mismo día de salida de la semana anterior. A parte de ser el método que más se aproxima a la realidad, es uno de los métodos más sencillos de calcular y que no depende de demasiadas variables, que en el caso de otros modelos, como es PTR o MOD, algunas de las variables necesarias para el cálculo no se pueden obtener por falta de información en los sistemas.



2. INTRODUCCIÓN

Hemos centrado este estudio en el área de logística con el fin de optimizar al máximo los recursos de transporte de calzado y complementos en una gran empresa especializada en este segmento. Partiendo de la necesidad de transportar diariamente cantidades muy importantes de calzado y complementos a todo el mundo con multitud de operativas y diferentes medios de transporte donde el coste cada vez resulta ser un factor más importante a tener en cuenta en la cadena de costes, intentaremos aportar nuevos conocimientos que permitan la optimización de los recursos en determinadas fases de todo el proceso de envío de mercancías.

La gran cantidad de referencias con las que se trabaja, con multitud de posibles combinaciones y a una gran cantidad de destinos hace que se pueda generar un número enorme de combinaciones.

En este estudio vamos a evaluar cómo se transforma un conjunto de unidades dadas de calzado y complementos en m^3 , ya que este volumen en m^3 es la unidad de medida utilizada para la contratación de los diferentes medios de transporte.

Un exceso en el cálculo del volumen nos genera gastos de recursos no utilizados y un cálculo a la baja del volumen genera que tengamos que buscar combinaciones de transporte totalmente desoptimizadas a un coste superior al regular, dado que se trabaja con la máxima de que todo el calzado y complementos deben ser enviados, nada puede quedarse sin expedir.

1.1. Para quien se va a realizar este estudio

Este estudio se va a realizar basándose en la forma de trabajo de Tempe, S. A. Grupo INDITEX.

“Tempe es la empresa especializada en calzado y complementos del Grupo Inditex. Su papel es el de concebir todas las colecciones de calzado de las ocho cadenas comerciales del Grupo (Zara, Pull&Bear, MassimoDutti, Bershka, Stradivarius, Oysho, Zara Home y Uterqüe), controlar su producción y llevar a cabo la distribución.

Los servicios centrales de Tempe, con una superficie de 180.000 m², están ubicados en uno de los principales núcleos de producción de calzado de España, Elche (Alicante). Sólo desde este centro se envían más de 50 millones de pares de zapatos al año a los más diversos destinos del mundo. Actualmente la plantilla de Tempe cuenta con más de 1.500 empleados.

Tempe asume el modelo de gestión de Inditex basado en la innovación, la flexibilidad y en una forma propia de entender la moda –creatividad y diseño de calidad, junto con una rápida respuesta a las demandas del mercado-. Para seguir el ritmo de expansión del Grupo, quien ya cuenta con más de 6.000 tiendas repartidas por más de 86 países en todo el mundo, destina recursos y esfuerzos a la búsqueda continua de la excelencia en el diseño, la calidad de sus productos y la innovación en sus procesos. El éxito de sus colecciones de calzado es el resultado del conocimiento de las necesidades de cada mercado y de su capacidad de adaptación a las demandas de los clientes.

La importancia de acortar plazos

Tempe cuenta con una estructura flexible y una fuerte orientación al cliente en todas sus áreas de actividad: diseño, fabricación, comercial, logística y distribución. La clave de este modelo de negocio, en el que se controlan todos los procesos, es la capacidad de adaptar la oferta a los deseos de los clientes en el menor tiempo posible. Un elevado grado de integración vertical permite disponer de una gran flexibilidad para acortar plazos y controlar stocks. De un modo dinámico, ágil y eficiente, Tempe se adapta al entorno cambiante de la moda atendiendo a las peticiones de los clientes en todo el mundo.

Funcionamiento de la organización Tempe

La actividad de Tempe empieza y termina en las tiendas de Inditex, siendo éstas la pieza clave en el desarrollo de las colecciones. De ellas proviene la información que genera el cliente y que permite modular la oferta en función de la demanda, y a ellas se destina el producto final. Los equipos de diseño se nutren de todo tipo de tendencias a través de sus viajes, de la observación de lo que ocurre en la calle y de la información que reciben desde las tiendas, para la elaboración de las colecciones de calzado y complementos. Al mismo tiempo, coordinan las tendencias con los equipos de ropa y accesorios, para obtener una colección completa y coordinada. Por su parte, los departamentos de compras y fabricación tienen la labor de materializar las ideas de los diseñadores y transformarlas en el calzado y los complementos que llevarán nuestros clientes de todo el mundo.

*Los avanzados sistemas informáticos y las alianzas establecidas con proveedores de múltiples nacionalidades son imprescindibles para llevar a cabo esa labor. **En cuanto a la distribución y la logística, el objetivo es cubrir la demanda de las tiendas acortando en la medida de lo posible los tiempos de tránsito y buscando el medio de transporte más eficiente. Tempe dispone de una cadena de suministro global, con múltiples orígenes y destinos de sus productos lo que convierte la logística en un elemento clave y estratégico para la evolución y el crecimiento de la compañía. El departamento comercial está especializado por mercados y su responsabilidad se centra en conocer las necesidades de cada uno de ellos, dar atención a las tiendas y asegurar el cumplimiento de los objetivos financieros establecidos.**¹*

3. OBJETIVOS

El objetivo de este estudio es definir el método de cálculo que más se aproxime al m^3 real comparando el método actual utilizado y junto con otros tres métodos nuevos propuestos.

4. METODOLOGÍA

Para la determinación de relaciones lineales entre variables haremos uso de gráficos de dispersión y cálculo de correlaciones (simples y parciales), junto a tests de la t de Student. Para el establecimiento del mejor modelo lineal (regresión lineal múltiple), utilizaremos el estimador máximo verosímil de los parámetros del modelo, junto con el método Stepwise ligado al criterio AIK (Akaike Information Criteria). Para analizar la bondad del ajuste haremos uso del error estándar residual, la tabla de ANOVA (Análisis de la Varianza) y el coeficiente de determinación. Finalmente, las hipótesis del modelo de regresión lineal múltiple (linealidad, normalidad, Homocedasticidad, media cero e incorrelación) serán analizadas mediante el test de Shapiro-Wilk, el test de Breusch-Pagan, el test de Durbin-Watson, el histograma de los residuos, el gráfico qq-plot y otra serie de gráficos habituales. Para todos nuestros análisis asumiremos un error no superior al 5%.

El programa utilizado para todos y cada uno de los pasos seguidos en este estudio estadístico es R (www.r-project.org). R es un software libre (GNU), cuyo uso se ha generalizado en los últimos años entre estadísticos y usuarios de estas técnicas en todo el mundo.

Para la realización del análisis econométrico someteremos a estudio los cuatro métodos de cálculo basados en datos de 2013 asegurándonos de que tenemos información detalle a detalle para confirmar cual es el mejor método siempre y cuando dispongamos siempre de todos los datos.

¹<http://www.tempe.es/documents/11643/54052/Tempe-Grupo-INDITEX-Press-ESP.pdf/7293490b-7962-4c39-a64d-7b12765beed7>

Primero realizaremos el estudio de forma independiente método a método y por último realizaremos un estudio completo con todos los métodos.

Por último se realizará un estudio comparativo de todos los métodos basados en los datos de 2014 y sin depurar los datos, es decir, sin eliminar los valores calculados a cero por cualquiera de los métodos. Esto ocurre cuando por problemas de los sistemas no se dispone de toda la información necesaria para realizar alguno de los cálculos. Por otra parte, para cada método se explicará con detalle cómo se calcula y posibles errores que se pueden cometer.

5. ESTUDIO ECONOMÉTRICO

A partir de unas unidades (uds.) dadas en los repartos de las cadenas, vamos a utilizar cuatro métodos de cálculo diferentes para hallar el volumen (vol.) de envío que será el utilizado para la contratación del transporte.

Este proyecto estudiará y comparará uno a uno los cuatro métodos de cálculo con su vol. de envío real y confirmará cuál de ellos es el más óptimo que será el que más se aproxime al vol. real del envío.

Lo métodos de cálculo a estudiar serán los siguientes:

- **PTR** método utilizado actualmente
- **MOD** = uds facturadas * dimensiones ref. + %aire misma tda. sem. anterior y mismo ciclo
- **DENS** =
$$\frac{\text{uds facturadas}}{\left(\frac{\text{uds}}{\text{vol}}\right) \text{ misma tda., sem. anterior y mismo ciclo}}$$
- **BLT** =
$$\left(\frac{\text{uds facturadas}}{\left(\frac{\text{uds}}{\text{bulto}}\right) \text{ misma tda., sem. ant., mismo ciclo}}\right) \text{ Prom. vol. bulto misma tda., sem - 1 y mismo ciclo}$$

Considerando Ciclo 1 días de salida (Martes, Miércoles y Jueves) y Ciclo 2 (Viernes, Lunes)

Para hacer la comparativa método a método se partirá de los envíos realizados cogiendo la información del facturado y a partir de las uds. facturadas se calculará el vol. de envío esperado según los métodos descritos. A priori se presupondrá que las uds. repartidas coinciden perfectamente con las uds. facturadas para que la variable uds. sea igual en todos los casos y la desviación en el cálculo del vol. provenga únicamente por el método y no por la diferencia de uds.

El estudio de los cuatro métodos comparados con el envío real se realizará basado en un estudio econométrico utilizando un programa estadístico llamado R a través del cual y siguiendo unos pasos establecidos nos confirmará cual será el método más óptimo de cálculo y con qué % de error.

La muestra que se tomará para realizar la comparativa será de 30 días elegidos aleatoriamente entre 5 periodos de tiempo durante el año 2013 seleccionados estratégicamente en los cambios de temporada y envíos de una única temporada, posteriormente según los resultados obtenidos se aplicará el método confirmado como el más óptimo en otros días para validar la calidad del mismo.

Detalles:

- No se tendrá en cuenta para el cálculo los envíos que se han grabado sin volumen o sin unidades.

Conociendo cuando se están enviando modelos de una o varias campañas se seleccionan 5 periodos durante el año de los cuales en 3 de ellos solo se envían uds. de una única campaña y 2 periodos en los que se están enviando uds. de dos campañas a la vez.

De cada uno de estos periodos se selecciona aleatoriamente una muestra de 30 días para el cálculo de los diferentes métodos tomando un total de 150 días y llegando al nivel para el cálculo de fecha y tienda. En el caso concreto del modelo por volumen se llega al nivel de fecha, tienda y referencia de modelo. La selección de estos periodos podemos verla representada en el Gráfico 1.

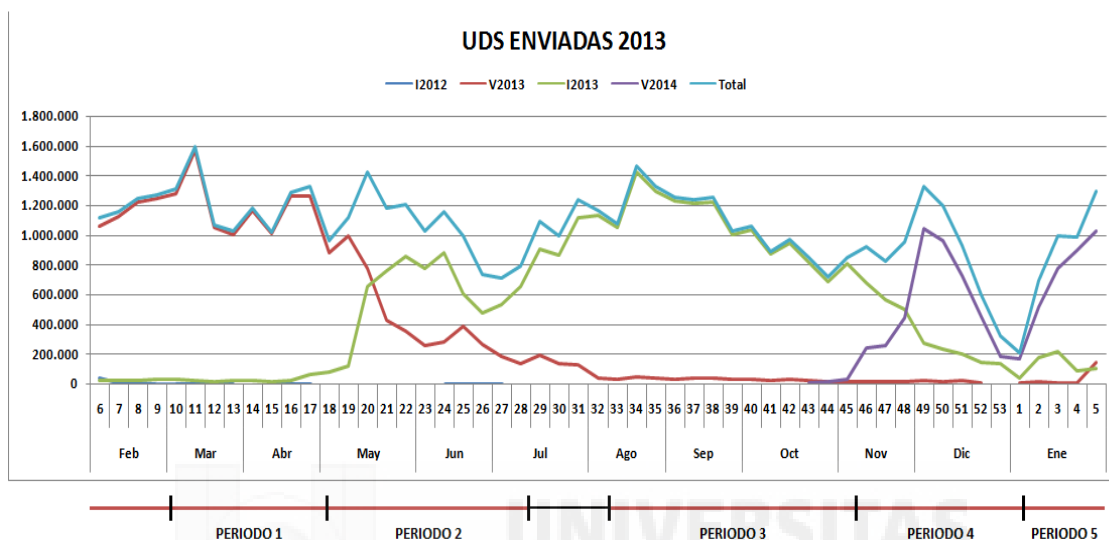


Gráfico 1 UDS. ENVIADAS POR TEMPORADA

a. ESTUDIO ECONOMÉTRICO PTR

En primer lugar analizamos el método de cálculo utilizado actualmente, llamado PTR, el cual a partir de las uds. que se reparten a tienda y mediante la conversión a vol. en función de la combinación de cálculos como por ej. la densidad (Uds./m³) de la entrada de la mercancía, el vol. de entrada de los bultos de surtido (bultos de 4 a 6 uds. aprox. del mismo modelo con distintas tallas), un %Aire parametrizable en la propia aplicación y algunas otras variables, se realiza la conversión de uds. en m³. Hay que tener en cuenta que puede haber diferencias entre las uds. que se quiere enviar a tienda y las uds. que realmente se envían producidas por ejemplo por diferencias de stock, previsiones de entrada de modelos no recibidos, etc. aunque estas diferencias suelen ser alrededor del 1%.

El nivel de detalle que utilizamos para realizar el cálculo es fecha, cadena, país (en el resto de modelos el detalle al que podemos llegar es mucho mayor) por lo que este nivel de agrupación puede hacer que se compensen algunas diferencias de cálculo.

Para este método no hemos podido disponer de información del primer periodo a diferencia de para el resto de métodos y la muestra de datos total utilizada para el estudio es de 78 días con un total de 9.475 registros.

La formulación del modelo para nuestro análisis es:

$$VOL = \beta_0 + \beta_1 PTR$$

Definición de los parámetros:

β_0 : término de interceptación o intercepto.

β_1 es el coeficiente de regresión, la pendiente de su respectiva variable, por tanto el valor de β_1 me indica que ceteris paribus, si aumento en una unidad la variable PTR implica que VOL aumentará en el valor de β_1 .

El modelo de regresión múltiple se va a analizar en cuatro fases:

Análisis preliminar
Ajuste del modelo
Bondad del ajuste
Diagnóstico del modelo

ANALISIS PRELIMINAR

GRÁFICA

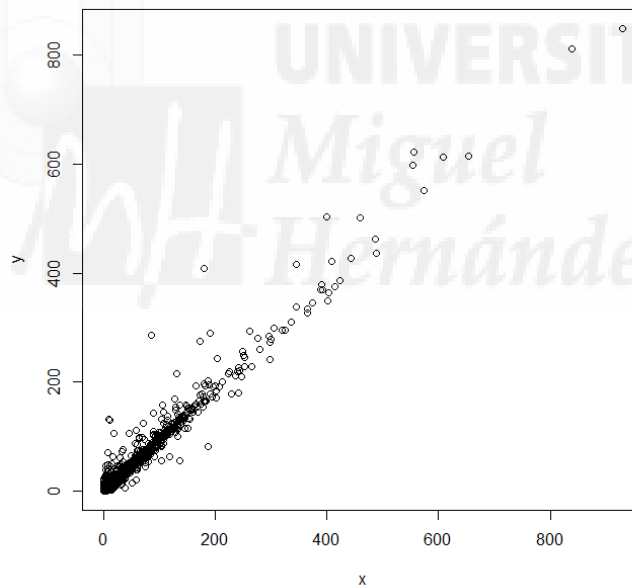


Gráfico 2 PTR

En el Gráfico 2 podemos observar como sí existe una clara relación lineal positiva entre las dos variables a estudio.

NUMÉRICA

El Coeficiente de Correlación Lineal es 0.9847303, redondeando 0.98, muy alto y por lo tanto muy bueno ya que es prácticamente igual a 1.

El p-valor asociado al contraste de hipótesis de si la correlación es cero o no es inferior a $2.2 \cdot 10^{-16}$, menor por tanto que 0.05, confirmándonos así que se rechaza la hipótesis de correlación cero. Por

otra parte, el intervalo de confianza obtenido está entre 0.9841077 y 0.9853286, siendo éste un intervalo muy reducido.

AJUSTE DEL MODELO

La recta estimada es:

$$y = \beta_0 + \beta_1 * x$$

$$y = 0.003380 + 0.967444 * x$$

Esta recta estimada indica que para cada unidad que se incremente x el valor de y aumentará en 0.967444 unidades.

El intervalo de confianza al 95% que tenemos para β_0 está comprendido entre -0.1261007 y 0.1328608 vemos como no se puede descartar que este coeficiente sea nulo.

El intervalo de confianza al 95% para el coeficiente β_1 está entre 0.9639994 y 0.9708886, un rango bastante reducido, sin embargo, no tiene contenido el valor 1. De haberse contenido el valor 1 podríamos haber sugerido que $y = x$, es decir, el método PTR proporcionaría el ajuste perfecto.

BONDAD DEL AJUSTE

Error estándar residual

El error estándar residual es igual a 6.107 y para saber si el dato obtenido es grande o pequeño calculamos el coeficiente de variación:

$$100 * (6.107 / \text{mean}(y)) = 53.62812 = 54\% > 10\%$$

Al comprobar el dato obtenido observamos que el valor es bastante alto por lo que no lo podemos tomar como bueno.

El coeficiente de determinación o R^2

Se obtiene un valor de 0.9697 > 0.65. Este valor es muy superior al umbral del 65% y casi igual a la unidad por lo que sí podemos tomarlo como bueno. Podemos, de hecho, afirmar que nuestro modelo de regresión explica el 97% de la variabilidad total de los datos.

El contraste asociado a la tabla de ANOVA

Con este criterio cuantificamos cuánta de la variabilidad contenida en los datos ha conseguido ser explicada por el modelo. Es bueno cuando la variabilidad explicada es alta, es decir, si las diferencias entre los datos y las predicciones según el modelo son pequeñas. Una vez descompuesta la variabilidad en la parte que es explicada por el modelo y la parte que deja sin explicar, los residuos, las comparamos para ver si la variabilidad explicada por el modelo ajustado es suficientemente grande.

$H_0: \beta_1 = 0$ (El modelo es malo) vs $H_1: \beta_1 \neq 0$ (El modelo es bueno)

El p-valor obtenido es inferior a $2.2 * 10^{-16}$, valor muy pequeño e inferior a 0,05. Por lo tanto, podemos rechazar H_0 y confirmar que el modelo lineal explica bien la variable respuesta (y).

DIAGNÓSTICO DEL MODELO O DE RESIDUOS

Después de ajustado el modelo ahora en la cuarta fase vamos a verificar si el modelo satisface las hipótesis básicas del modelo de regresión lineal para los errores:

- Normalidad
- Varianza constante (Homocedasticidad)
- Incorrelación
- Media cero

El análisis de los residuos nos permitirá detectar deficiencias en la verificación de estas hipótesis, así como descubrir observaciones anómalas o especialmente influyentes en el ajuste. Una vez encontradas las deficiencias, si existen, debemos analizar si replanteamos el modelo empleando alguna transformación sobre las variables. La fase del diagnóstico del modelo suele ser gráfica, aunque también existen varios test estadísticos que mostraremos y resolveremos. Los residuos de un modelo lineal los definimos, como habitualmente se hace en la literatura, como las desviaciones entre las observaciones y los valores ajustados.

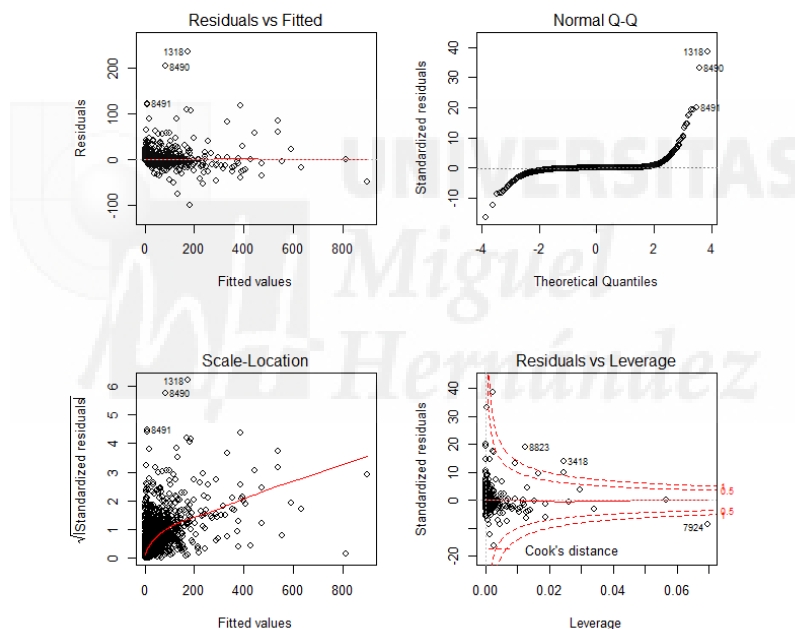


Gráfico 3 RESIDUOS Y QQPLOT

Normalidad: En el gráfico qqplot (superior derecho) observamos que existen colas muy pronunciadas y por lo tanto no podemos confirmar la hipótesis de normalidad.

Homocedasticidad (varianza cte.): El gráfico superior izquierdo muestra una forma de embudo por lo que tampoco podemos confirmar esta hipótesis.

Ahora calcularemos los residuos para poder realizar unos gráficos adicionales.

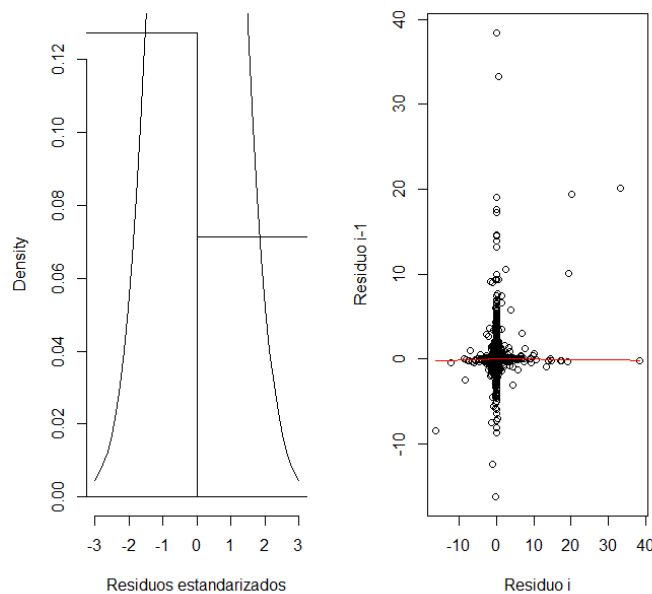


Gráfico 4

En el Gráfico 4 de Residuos estandarizados confirmamos que la normalidad no se cumple porque las columnas no quedan representadas dentro de la curva en forma de campana.

Por otra parte, en el gráfico de Residuo i vs $i-1$ observamos que la línea roja representada es bastante horizontal por lo que parece indicar que se verifica la hipótesis de incorrelación.

CONTRASTES DE HIPOTESIS

$H_0: e \sim \text{Normal}$
 $H_1: \text{Lo contrario}$

El p-valor es inferior a $2.2 \cdot 10^{-16} < 0.05$, por lo que rechazamos la hipótesis de normalidad definitivamente.

Homocedasticidad (hipótesis de varianza constante)

$H_0: \text{Homocedasticidad}$
 $H_1: \text{Heterocedasticidad}$

El test de Breusch-Pagan nos devuelve un p-valor de $2.2 \cdot 10^{-16}$, inferior a 0.05 por lo que podemos rechazar la hipótesis de varianza constante (Homocedasticidad).

Incorrelación

Para salir de dudas sobre el cumplimiento de esta hipótesis hallamos el siguiente contraste:

$H_0: \rho = 0$
 $H_1: \rho \neq 0$

El test de Durbin-Watson nos devuelve un p-valor inferior a $2.2 \cdot 10^{-16} < 0.05$ por lo que rechazamos la hipótesis de incorrelación.

Hipótesis de media cero de los errores:

Siempre se verifica por cómo ajustamos el modelo (Máxima verosimilitud).

En conclusión:

- Aunque en la bondad del ajuste del error estándar residual es elevado, un 54%, la tabla de ANOVA y R cuadrado se aprueban perfectamente por lo que podemos dar por buena la bondad del ajuste.
- No se cumplen los diagnósticos del modelo de Normalidad, Homocedasticidad e Incorrelación, este último incluso en el caso de tomar una muestra de días al azar.

b. ESTUDIO ECONOMÉTRICO MOD

A continuación analizaremos uno de los métodos propuestos llamado MOD, el cual a partir de las uds. que se facturan a tienda y mediante la conversión a vol. en función de las medidas de cada modelo por referencia (largo*ancho*alto) obtenidas a la entrada de la mercancía más un porcentaje de aire tomado del envío de la semana anterior a la misma tienda y en el mismo día de salida, realiza la conversión de uds. en m^3 .

El nivel de detalle que utilizamos para realizar el cálculo es fecha, tienda, cadena, país.

La muestra de datos total utilizada para el estudio es de 150 días con un total de 16.519 registros.

La formulación del modelo para nuestro análisis es:

$$VOL = \beta_0 + \beta_1 MOD$$

Definición de los parámetros:

β_0 : término de interceptación o intercepto.

β_1 es el coeficiente de regresión, la pendiente de su respectiva variable, por tanto el valor de β_1 me indica que ceteris paribus, si aumento en una unidad la variable PTR implica que VOL aumentará en el valor de β_1 .

El modelo de regresión múltiple se va a analizar en cuatro fases:

Análisis preliminar
Ajuste del modelo
Bondad del ajuste
Diagnóstico del modelo

ANÁLISIS PRELIMINAR

GRÁFICA

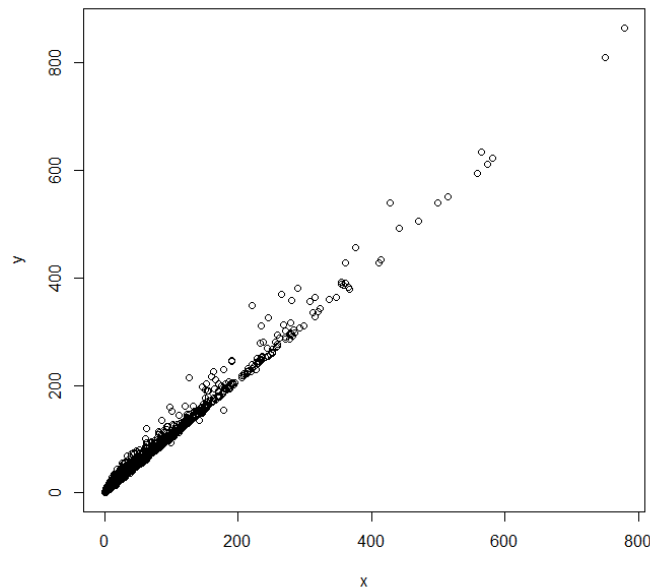


Gráfico 5

En el Gráfico 5 observamos una clara tendencia positiva.

NUMÉRICA

El Coeficiente de Correlación Lineal es 0.9962612 redondeando 99,6%, el % es bastante alto y por lo tanto bastante bueno ya que es prácticamente igual a la unidad. El $p\text{-value} < 2.2 \cdot 10^{-16}$ es inferior a 0.05 confirmándonos que es un buen valor, el intervalo de confianza obtenido está entre 0.9961456 y 0.9963733 nótese que la amplitud de este intervalo es prácticamente nula.

AJUSTE DEL MODELO

La recta estimada es

$$y = \beta_0 + \beta_1 \cdot x$$

$$y = 0.3154086 + 1.0944828 \cdot x$$

Esta recta explica que cada unidad que se incremente x el valor de y aumentará en 1.0944828 unidades. El intervalo de confianza que tenemos para β_0 está entre 0.2681266 y 0.3626906, por lo que descartamos que su valor sea igual a cero.

El intervalo de confianza de β_1 está entre 1.0930353 y 1.0959304, donde vemos que tiene un rango bastante ajustado.

BONDAD DEL AJUSTE

Error estándar residual

El error estándar residual es igual a 2.945 y para saber si el dato obtenido es grande o pequeño calculamos la siguiente fórmula:

$$100 * (2.945 / \text{mean}(y)) = 25.59783 = 26\% > 10\%$$

Al comprobar el dato obtenido observamos que el valor es relativamente alto por lo que no lo podemos tomar como bueno.

a. El coeficiente de determinación o R^2

Multiple R-squared: 0.9925 > 0.65

Este valor es muy superior al 65% y casi igual a la unidad por lo que sí podemos tomarlo como bueno.

b. El contraste asociado a la tabla de ANOVA

$H_0: \beta_1=0$ (El modelo es malo) vs $H_1: \beta_1 \neq 0$ (El modelo es bueno)

El p-valor obtenido es $< 2.2 * 10^{-16}$. Podemos observar que es muy pequeño e inferior a 0,05, por lo tanto podemos rechazar H_0 y confirmar que el modelo lineal explica bien la variable respuesta (y).

DIAGNÓSTICO DEL MODELO O DE RESIDUOS

Después de ajustado el modelo ahora en la cuarta fase vamos a verificar si el modelo satisface las hipótesis básicas del modelo de regresión lineal para los errores:

- Normalidad
- Varianza constante (Homocedasticidad)
- Incorrelación
- Media cero

El análisis de los residuos nos permitirá detectar deficiencias en la verificación de estas hipótesis, así como descubrir observaciones anómalas o especialmente influyentes en el ajuste. Una vez encontradas las deficiencias, si existen, debemos analizar si replanteamos el modelo empleando alguna transformación sobre las variables. La fase del diagnóstico del modelo suele ser gráfica, aunque también existen varios test estadísticos que mostraremos y resolveremos. Los residuos de un modelo lineal los definimos, como habitualmente se hace en la literatura, como las desviaciones entre las observaciones y los valores ajustados.

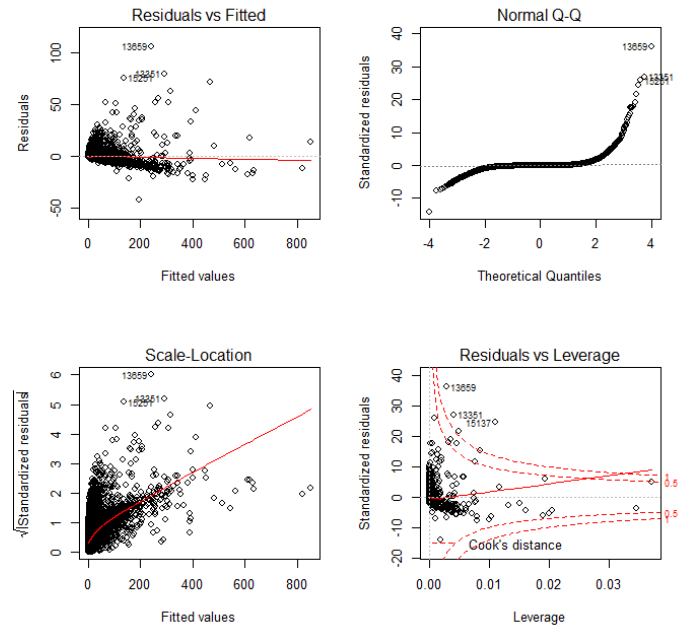


Gráfico 6

Normalidad: En el gráfico qqplot (superior derecho) observamos que existen colas muy pronunciadas y por lo tanto no podemos confirmar que existe una distribución normal.

Homocedasticidad (varianza cte.): El gráfico superior izquierdo muestra una forma de embudo por lo que tampoco podemos confirmar este diagnóstico.

Ahora calcularemos los residuos para poder realizar unos gráficos adicionales.

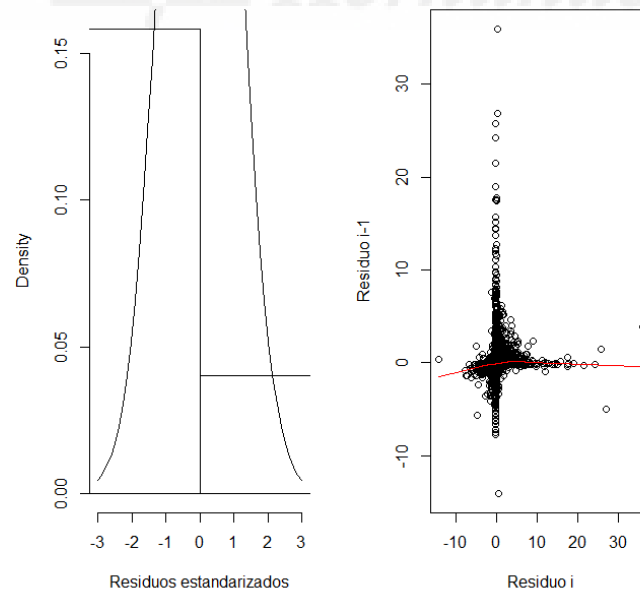


Gráfico 7

En el gráfico de Residuos estandarizados confirmamos que la **normalidad** no se cumple porque las columnas no quedan representadas dentro de la línea que dibujamos de campana.

En el gráfico de Residuo i observamos que la línea roja representada es bastante horizontal por lo que podemos indicar que parece existir autocorrelación en los residuos, que tendrá que ser confirmada o no con el test correspondiente.

CONTRASTE DE HIPOTESIS

$H_0: e \sim N$

H_1 : Lo contrario

El p -value $< 2.2 \cdot 10^{-16} < 0.05$ donde también rechazamos la hipótesis de normalidad definitivamente.

Homocedasticidad (hipótesis de varianza constante)

H_0 : Homocedasticidad

H_1 : Heterocedasticidad

El test de Breusch-Pagan se obtiene con

El p -valor obtenido $2.2 \cdot 10^{-16} < 0.05$ por lo que podemos rechazar la hipótesis de varianza constante de Homocedasticidad.

Para salir de dudas hallamos el siguiente contraste:

$H_0: \rho = 0$

$H_1: \rho \neq 0$

El test de Durbin-Watson:

p -value $< 2.2 \cdot 10^{-16} < 0.05$ por lo que rechazamos la hipótesis de incorrelación.

Hipótesis de media cero:

Siempre se verifica por cómo ajustamos el modelo (EMV).

En conclusión podemos dar por bueno el modelo por varios motivos como son:

- Aunque en la bondad del ajuste el error estándar residual es algo elevado, un 26%, la tabla de ANOVA y R cuadrado se aprueban perfectamente por lo que podemos dar por buena la bondad del ajuste.
- No se cumplen los diagnósticos del modelo de Normalidad, Homocedasticidad e Incorrelación.

c. ESTUDIO ECONOMÉTRICO DENS

A continuación analizaremos uno de los métodos propuestos, llamado DENS, el cual a partir de las uds. que se facturan a tienda y mediante la conversión a vol. en función de las uds. y vol. real enviado por tienda el mismo día de salida de la semana anterior, realiza la conversión de uds. en m^3 .

El nivel de detalle que utilizamos para realizar el cálculo es fecha, tienda, cadena, país.

La muestra de datos total utilizada para el estudio es de 150 días con un total de 211.548 registros.

La formulación del modelo para nuestro análisis es:

$$VOL = \beta_0 + \beta_1 DENS$$

Definición de los parámetros:

β_0 : término de interceptación o intercepto.

β_1 es el coeficiente de regresión, la pendiente de su respectiva variable, por tanto el valor de β_1 me indica que ceteris paribus, si aumento en una unidad la variable DENS implica que VOL aumentará en el valor de β_1 .

El modelo de regresión múltiple se va a analizar en cuatro fases:

Análisis preliminar
Ajuste del modelo
Bondad del ajuste
Diagnóstico del modelo

ANALISIS PRELIMINAR

GRÁFICA

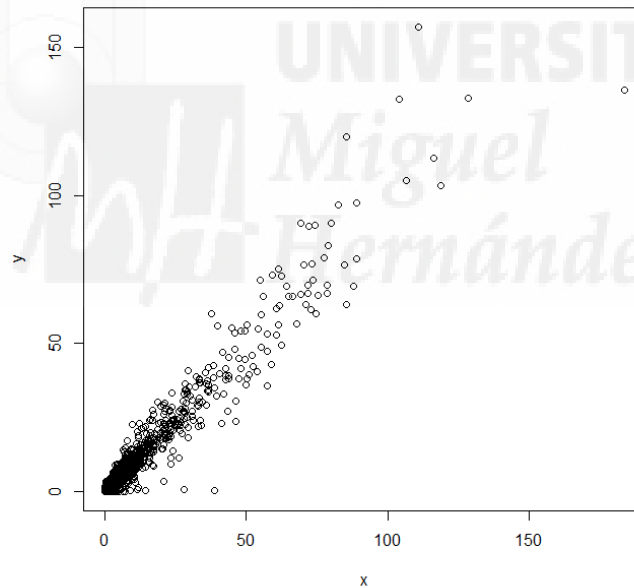


Gráfico 8

NUMÉRICA

El Coeficiente de Correlación Lineal es 0.9757682 redondeando 98%, el % es muy alto y por lo tanto muy bueno ya que se aproxima bastante a la unidad.

El p-value $< 2.2 \cdot 10^{-16}$ es inferior a 0.05 confirmándonos que es un buen valor, el intervalo de confianza obtenido está entre 0.9755633 y 0.9759713, mostrando una amplitud muy reducida.

AJUSTE DEL MODELO

La recta estimada es

$$y = \beta_0 + \beta_1 * x$$

$$y = 0.0185565 + 0.9674069 * x$$

Esta recta explica que cada unidad que se incremente x el valor de y aumentará en **0.9674069** unidades.

El intervalo de confianza que tenemos para β_0 está entre 0.01668657 y 0.02042641, por lo que podemos descartar que el valor sea igual a cero.

El intervalo de confianza de β_1 está entre 0.96648248 y 0.96833133 donde vemos que se aproxima bastante a la unidad y con un rango bastante acotado.

Estos intervalos obtenidos para $\beta_0 \approx 0$ y $\beta_1 \approx 1$ nos confirman que el valor de $y \approx x$ ajustándose casi perfectamente el valor obtenido mediante este método en x con el valor real de y .

BONDAD DEL AJUSTE

Error estándar residual

El error estándar residual es igual a 0.3974 y para saber si el dato obtenido es grande o pequeño calculamos la siguiente fórmula.

$$100 * (0.3974 / \text{mean}(y)) = 46.82109 = 47\% > 10\%$$

Al comprobar el dato obtenido observamos que el valor es muy alto por lo que no lo podemos tomar como bueno.

c. El coeficiente de determinación o R^2

Multiple R-squared: 0.9521 > 0.65

Este valor es muy superior al 65% por lo que sí podemos tomarlo como bueno.

El contraste asociado a la tabla de ANOVA

H0: $\beta_1=0$ (El modelo es malo) vs **H1: $\beta_1 \neq 0$** (El modelo es bueno)

El p-valor obtenido es $2.2 * 10^{-16}$. Podemos observar que es muy pequeño e inferior a 0,05, por lo tanto podemos rechazar **H0** y confirmar que el modelo lineal explica bien la variable respuesta (y).

DIAGNÓSTICO DEL MODELO O DE RESIDUOS

Después de ajustado el modelo ahora en la cuarta fase vamos a verificar si el modelo satisface las hipótesis básicas del modelo de regresión lineal para los errores:

- Normalidad
- Varianza constante (Homocedasticidad)
- Incorrelación
- Media cero

El análisis de los residuos nos permitirá detectar deficiencias en la verificación de estas hipótesis, así como descubrir observaciones anómalas o especialmente influyentes en el ajuste. Una vez encontradas las deficiencias, si existen, debemos analizar si replanteamos el modelo empleando alguna transformación sobre las variables. La fase del diagnóstico del modelo suele ser gráfica, aunque también existen varios test estadísticos que mostraremos y resolveremos. Los residuos de un modelo lineal los definimos, como habitualmente se hace en la literatura, como las desviaciones entre las observaciones y los valores ajustados.

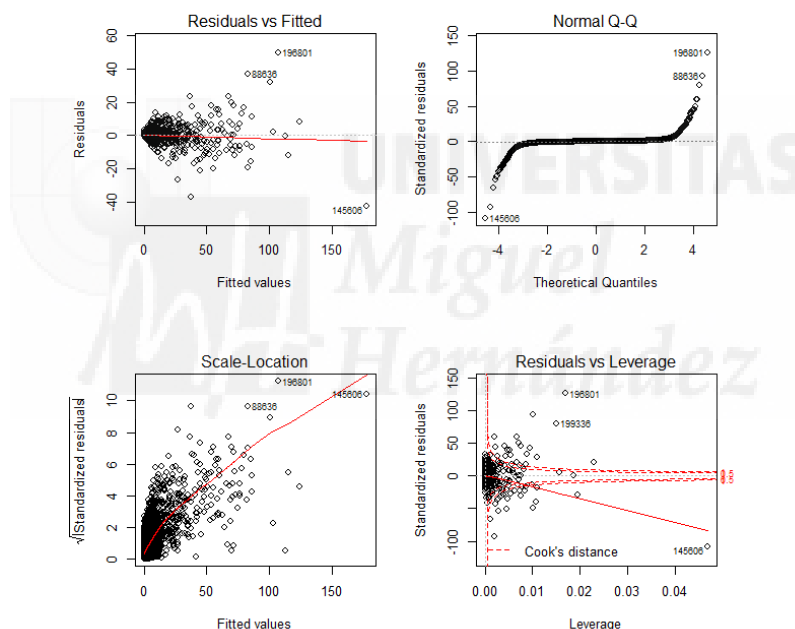


Gráfico 9

Normalidad: En el gráfico qqplot (superior derecho) observamos que existen colas y por lo tanto no podemos confirmar que existe una distribución normal.

Homocedasticidad (varianza cte.): El gráfico superior izquierdo muestra una forma de embudo por lo que tampoco podemos confirmar este diagnóstico.

Ahora calcularemos los residuos para poder realizar unos gráficos adicionales.

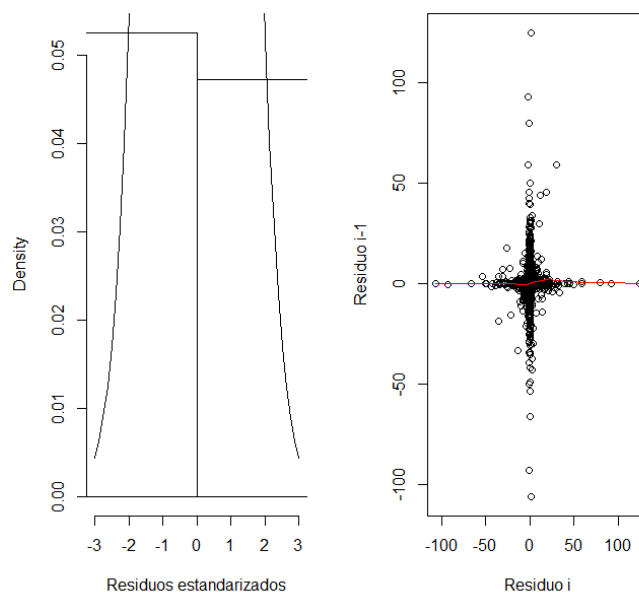


Gráfico 10

En el gráfico de Residuos estandarizados confirmamos que la **normalidad** no se cumple porque las columnas no quedan representadas dentro de la línea que dibujamos de campana.

En el gráfico de Residuo i vs i-1 observamos que la línea roja representada es bastante horizontal por lo que parece podemos indicar que existe **autoincorrelación** en el modelo.

CONTRASTE DE HIPOTESIS

$H_0: e \sim N$

$H_1: \text{Lo contrario}$

El $p\text{-value} < 2.2 \cdot 10^{-16} < 0.05$ donde también rechazamos la hipótesis de **normalidad** definitivamente.

Homocedasticidad (hipótesis de varianza constante)

$H_0: \text{Homocedasticidad}$

$H_1: \text{Heterocedasticidad}$

El test de Breusch-Pagan se obtiene con:

El $p\text{-valor}$ obtenido $2.2 \cdot 10^{-16} < 0.05$ por lo que podemos rechazar la hipótesis de varianza constante de Homocedasticidad.

Incorrelación:

Para salir de dudas hallamos el siguiente contraste:

$H_0: \rho = 0$

$H_1: \rho \neq 0$

El test de Durbin-Watson:

$p\text{-value} < 2.2 \cdot 10^{-16} < 0.05$ por lo que rechazamos la hipótesis de incorrelación.

Hipótesis de media cero:

Siempre se verifica por cómo ajustamos el modelo (EMV)

En conclusión podemos dar por bueno el modelo por varios motivos como son:

- Aunque en la bondad del ajuste del error estándar residual es elevado, un 47%, la tabla de ANOVA y R cuadrado se aprueban perfectamente por lo que podemos dar por buena la bondad del ajuste.
- No se cumplen los diagnósticos del modelo de Normalidad, Homocedasticidad e Incorrelación.

d. ESTUDIO ECONOMÉTRICO BLT

A continuación analizaremos el último de los métodos propuestos llamado BLT, el cual a partir de las uds. que se facturan a tienda y mediante la conversión a vol. en función de las uds. y la media de uds. por bulto multiplicado por el vol. medio de los bultos enviados por tienda el mismo día de salida de la semana anterior, realiza la conversión de uds. en m^3 .

El nivel de detalle que utilizamos para realizar el cálculo es fecha, tienda, cadena, país.

La muestra de datos total utilizada para el estudio es de 150 días con un total de 211.548 registros.

La formulación del modelo para nuestro análisis es:

$$VOL = \beta_0 + \beta_1 BLT$$

Definición de los parámetros:

β_0 : término de interceptación o intercepto.

β_1 es el coeficiente de regresión, la pendiente de su respectiva variable, por tanto el valor de β_1 me indica que ceteris paribus, si aumento en una unidad la variable BLT implica que VOL aumentará en el valor de β_1 .

El modelo de regresión múltiple se va a analizar en cuatro fases:

Análisis preliminar
Ajuste del modelo
Bondad del ajuste
Diagnóstico del modelo

ANÁLISIS PRELIMINAR

GRÁFICA

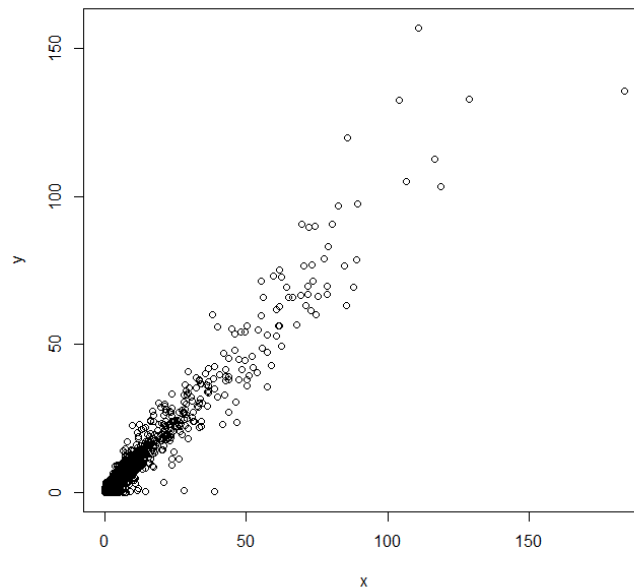


Gráfico 11

NUMÉRICA

El Coeficiente de Correlación Lineal es 0.9755372, redondeando 98%, el % es muy alto y por lo tanto muy bueno ya que se aproxima bastante a la unidad.

El $p\text{-value} < 2.2 \cdot 10^{-16}$ es inferior a 0.05 confirmándonos que es un buen valor, el intervalo de confianza obtenido está entre 0.9753304 y 0.9757423.

AJUSTE DEL MODELO

La recta estimada es

$$y = \beta_0 + \beta_1 \cdot x$$

$$y = -0.0644229 + 0.9661751 \cdot x$$

Esta recta explica que cada unidad que se incremente x el valor de y aumentará en 0.9661751 unidades.

El intervalo de confianza que tenemos para β_0 está entre -0.06633687 y -0.06250897.

El intervalo de confianza de β_1 está entre 0.96524732 y 0.96710293 donde vemos que se aproxima bastante a la unidad y con un rango bastante acotado.

BONDAD DEL AJUSTE

Error estándar residual

El error estándar residual es igual a **0.3992** y para saber si el dato obtenido es grande o pequeño calculamos la siguiente fórmula:

$$100 * (0.3992 / \text{mean}(y)) = 47.03316 = 47\% > 10\%$$

Al comprobar el dato obtenido observamos que el valor es muy alto por lo que no lo podemos tomar como bueno.

d. El coeficiente de determinación o R^2

Multiple R-squared: 0.9517 > 0.65.

Este valor es muy superior al 65% por lo que si podemos tomarlo como bueno.

e. El contraste asociado a la tabla de ANOVA

H0: $\beta_1=0$ (El modelo es malo) vs **H1: $\beta_1 \neq 0$** (El modelo es bueno)

El p-valor obtenido es $2.2 * 10^{-16}$. Podemos observar que es muy pequeño e inferior a 0,05, por lo tanto podemos rechazar **H0** y confirmar que el modelo lineal explica bien la variable respuesta (y).

DIAGNÓSTICO DEL MODELO O DE RESIDUOS

Después de ajustado el modelo ahora en la cuarta fase vamos a verificar si el modelo satisface las hipótesis básicas del modelo de regresión lineal para los errores:

- Normalidad
- Varianza constante (Homocedasticidad)
- Incorrelación
- Media cero

El análisis de los residuos nos permitirá detectar deficiencias en la verificación de estas hipótesis, así como descubrir observaciones anómalas o especialmente influyentes en el ajuste. Una vez encontradas las deficiencias, si existen, debemos analizar si replanteamos el modelo empleando alguna transformación sobre las variables. La fase del diagnóstico del modelo suele ser gráfica, aunque también existen varios test estadísticos que mostraremos y resolveremos. Los residuos de un modelo lineal los definimos, como habitualmente se hace en la literatura, como las desviaciones entre las observaciones y los valores ajustados.

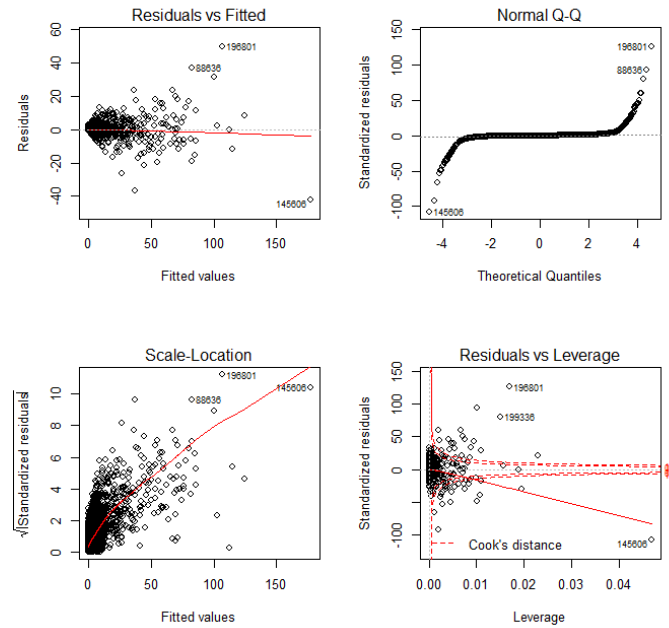


Gráfico 12

Normalidad: En el gráfico qqplot (superior derecho) observamos que existen colas y por lo tanto no podemos confirmar que existe una distribución normal.

Homocedasticidad (varianza cte.): El gráfico superior izquierdo muestra una forma de embudo por lo que tampoco podemos confirmar este diagnóstico.

Ahora calcularemos los residuos para poder realizar unos gráficos adicionales.

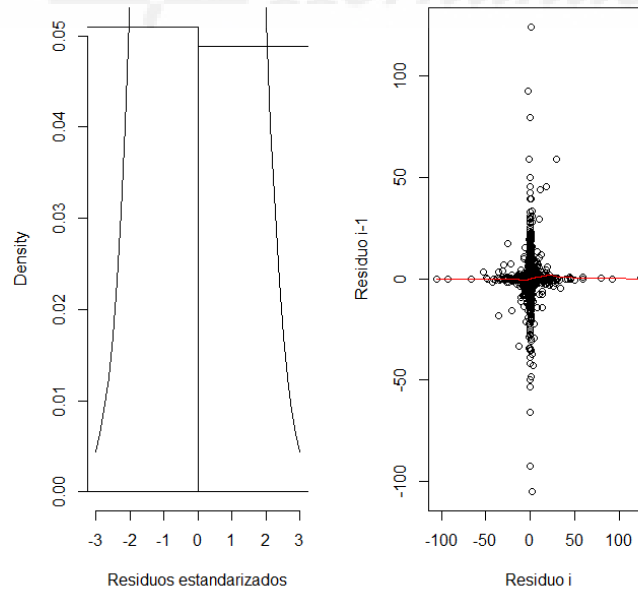


Gráfico 13

En el gráfico de Residuos estandarizados confirmamos que la **normalidad** no se cumple porque las columnas no quedan representadas dentro de la línea que dibujamos de campana.

En el gráfico de Residuo i observamos que la línea roja representada es bastante horizontal por lo que parece podemos indicar que existe **autoincorrelación** en el modelo.

CONTRASTE DE HIPOTESIS

$H_0: e \sim N$

H_1 : Lo contrario

El p -value $< 2.2 \cdot 10^{-16} < 0.05$ donde también rechazamos la hipótesis de **normalidad** definitivamente.

Homocedasticidad (hipótesis de varianza constante)

H_0 : Homocedasticidad

H_1 : Heterocedasticidad

El test de Breusch-Pagan se obtiene con

El p -valor obtenido $2.2 \cdot 10^{-16} < 0.05$ por lo que podemos rechazar la hipótesis de varianza constante de Homocedasticidad.

Para salir de dudas hallamos el siguiente contraste:

$H_0: \rho = 0$

$H_1: \rho \neq 0$

El test de Durbin-Watson:

p -value $< 2.2 \cdot 10^{-16} < 0.05$ por lo que rechazamos la hipótesis de incorrelación.

Hipótesis de media cero:

Siempre se verifica por cómo ajustamos el modelo (EMV).

En conclusión podemos dar por bueno el modelo por varios motivos como son:

- Aunque en la bondad del ajuste del error estándar residual es elevado, un 47%, la tabla de ANOVA y R cuadrado se aprueban perfectamente por lo que podemos dar por buena la bondad del ajuste.
- No se cumplen los diagnósticos del modelo de Normalidad, Homocedasticidad e Incorrelación.

e. ESTUDIO ECONOMÉTRICO DE TODOS LOS MÉTODOS

En esta sección analizamos los cuatro métodos juntos agrupados por cadena y país con un total de datos por método a analizar de 7.686 registros. El modelo llamado PTR es el utilizado actualmente y el resto (MOD, DENS y BLT) son los nuevos modelos propuestos que comparamos con el volumen real llamado VOL.

La formulación del modelo para nuestro análisis es:

$$VOL = \beta_0 + \beta_1 PTR + \beta_2 MOD + \beta_3 DENS + \beta_4 BLT + \varepsilon$$

Definición de los parámetros:

β_0 : término de interceptación o intercepto.

$\beta_1, \beta_2, \dots, \beta_4$ son los coeficientes de regresión, son las pendientes de sus respectivas variables, por tanto el valor de β_1 me indica que ceteris paribus, si aumento en una unidad la variable PTR implica que VOL aumentará en el valor de β_1 , etc....

El modelo de regresión múltiple se va a analizar en cuatro fases:

- Análisis preliminar
- Ajuste del modelo
- Bondad del ajuste
- Diagnóstico del modelo

ANÁLISIS PRELIMINAR

GRÁFICA

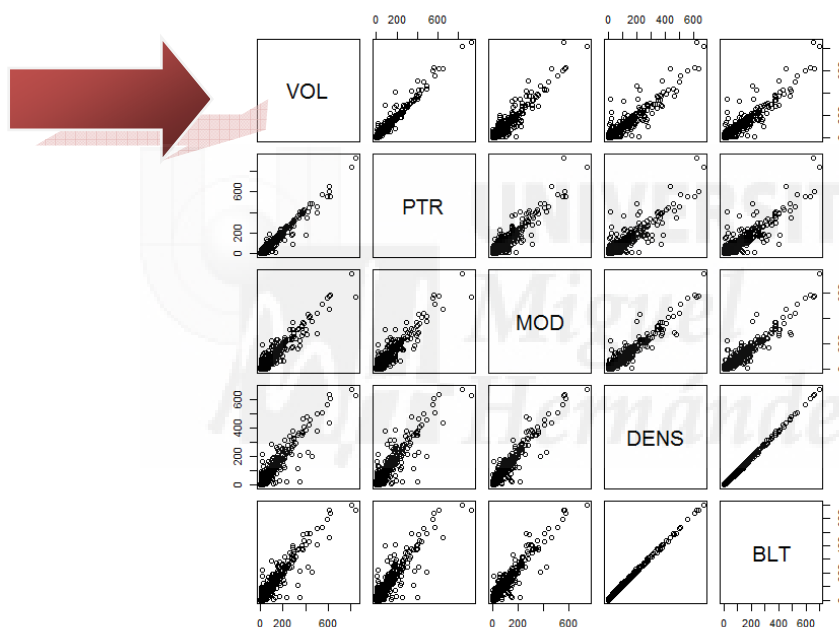


Gráfico 14

NUMÉRICA

Las correlaciones simples calculadas con respecto a VOL son:

VOL	PTR	MOD	DENS	BLT
1.0000000	0.9855866	0.9660818	0.9566094	0.9576074

Todas ellas son correlaciones muy altas y positivas.

Mostramos a continuación las correlaciones parciales:

VOL	PTR	MOD	DENS	BLT
0.0000000	0.81288458	0.2133637	-0.04277576	0.10011983

Las correlaciones parciales nos confirman que el método PTR es la variable más importante, en presencia de las demás, al igual que se observa tanto en el gráfico como en las correlaciones simples con una diferencia muy importante con respecto al resto de variables.

AJUSTE DEL MODELO

El modelo ajustado (estimado) queda como:

$$\text{VOL} = -0.326203 + 0.656498 \cdot \text{PTR} + 0.164666 \cdot \text{MOD} - 0.152968 \cdot \text{DENS} + 0.329632 \cdot \text{BLT}$$

El intervalo de confianza al 95% para los coeficientes es:

	2.5 %	97.5 %
(Intercept)	-0.4554672	-0.19693821
PTR	0.6459771	0.66701884
MOD	0.1478017	0.18153074
DENS	-0.2328796	-0.07305568
BLT	0.2563613	0.40290216

Estudiamos si cada coeficiente puede o no ser cero (contrates de hipótesis).

$H_0: \beta_1=0$	$H_0: \beta_2=0$	$H_0: \beta_3=0$	$H_0: \beta_4=0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$	$H_1: \beta_3 \neq 0$	$H_1: \beta_4 \neq 0$

p-valor PTR $< 2 \cdot 10^{-16} < 0.05$	→	Se rechaza H_0 y tomamos H_1
p-valor MOD $< 2 \cdot 10^{-16} < 0.05$	→	Se rechaza H_0 y tomamos H_1
p-valor DENS = 0.000176 < 0.05	→	Se rechaza H_0 y tomamos H_1
p-valor BLT $< 2 \cdot 10^{-16} < 0.05$	→	Se rechaza H_0 y tomamos H_1

En cualquiera de los casos descartamos que el coeficiente sea cero y, por lo tanto, no descartamos ninguna variable ya que todas ellas son importantes a la hora de explicar volumen facturado.

BONDAD DEL AJUSTE

Error estándar residual

El error estándar residual es igual a **5.209** y para saber si el dato obtenido es grande o pequeño calculamos la siguiente fórmula:

$$100 \cdot (5.209 / \text{mean}(\text{VOL})) = 40.33186 = 40\% > 10\%$$

Al comprobar el dato obtenido observamos que el valor es bastante alto por lo que no lo podemos tomar como bueno.

El coeficiente de determinación o R^2

Otro estadístico útil para chequear la bondad del ajuste se define como la proporción de la varianza que es explicada por la recta de regresión. Obtenemos un número que nos indicará el valor en tanto por ciento de la variabilidad total de los datos. Cuando el valor sea cercano a 1 (100%) implicará que buena parte de la varianza es explicada por la recta de regresión y cuando R^2 tenga un valor cercano a cero significará que prácticamente toda la variabilidad de los datos queda sin explicar por el modelo de regresión.

Compararemos, además, el valor del coeficiente de R^2 con el valor de coeficiente R^2 ajustado para ver si son parecidos, ya que cuando hay varias variables explicativas, el mero hecho de tener muchas variables puede distorsionar el R^2 original, por eso lo comparamos con el ajustado. Si el ajustado fuera muy diferente, tomaríamos este como bueno y no el coeficiente de R^2 .

Para nuestros datos, observamos un valor del coeficiente $R^2 = 0.9815$ (>65%), por lo que el modelo es bastante bueno, explica el 98% de la variabilidad total de la variable VOL. Por otra parte, el valor de coeficiente R^2 ajustado = 0.9815, es igual al coeficiente de R^2 . Por lo tanto, damos por válido el valor del coeficiente R^2 original, 0.9815, que como hemos visto es elevado.

Multiple R-squared: 0.9815, Adjusted R-squared: 0.9815

Multiple R-squared: 0.9815 > 0.65.

El contraste asociado a la tabla de ANOVA

En este criterio cuantificamos cuánta de la variabilidad contenida en los datos ha conseguido ser explicada por dicho modelo. Un modelo es bueno si la variabilidad explicada es mucha, o lo que es lo mismo, si las diferencias entre los datos y las predicciones según el modelo son pequeñas. Construir la tabla de Anova consiste en descomponer la variabilidad de los datos en la parte que es explicada por el modelo y la parte que se deja sin explicar, es decir, la variabilidad de los residuos, y compararlas y valorarlas estadísticamente para ver si la variabilidad explicada por el modelo ajustado es suficientemente grande. En particular, damos respuesta al contraste:

H0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (El modelo es malo)

H1: Lo contrario (El modelo es bueno)

A este respecto, si obtenemos un p-valor inferior a 0,05 rechazaremos H0 y el modelo será bueno, mientras que, por el contrario, si el p-valor es superior al 0,05 el modelo será malo; contrastamos si las variables explicativas explican suficientemente bien la variable respuesta a través del modelo lineal propuesto.

El p-valor asociado al contraste de la tabla de ANOVA de nuestro modelo es $< 2 \cdot 10^{-16}$ por lo tanto inferior a 0,05, y rechazamos H0. En definitiva, el modelo es bueno según este criterio.

DIAGNÓSTICO DEL MODELO O DE RESIDUOS

Después de ajustado el modelo ahora en la cuarta fase vamos a verificar si el modelo satisface las hipótesis básicas del modelo de regresión lineal para los errores:

- Normalidad
- Varianza constante (Homocedasticidad)
- Incorrelación
- Media cero

El análisis de los residuos nos permitirá detectar deficiencias en la verificación de estas hipótesis, así como descubrir observaciones anómalas o especialmente influyentes en el ajuste. Una vez encontradas las deficiencias, si existen, debemos analizar si replanteamos el modelo empleando alguna transformación sobre las variables. La fase del diagnóstico del modelo suele ser gráfica, aunque también existen varios test estadísticos que mostraremos y resolveremos. Los residuos de un modelo lineal los definimos, como habitualmente se hace en la literatura, como las desviaciones entre las observaciones y los valores ajustados.

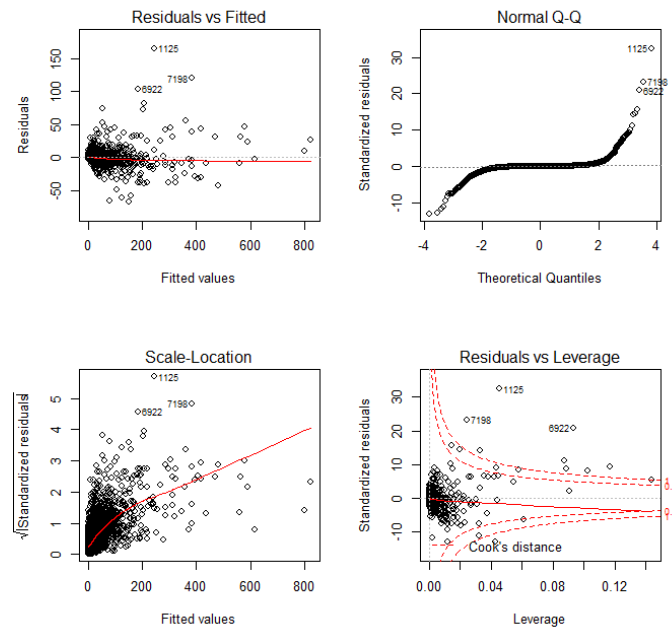


Gráfico 15

Normalidad: En el gráfico qqplot (superior derecho) observamos que existen colas muy pronunciadas y por lo tanto no podemos confirmar que existe una distribución normal.

Homocedasticidad (varianza cte.): El gráfico superior izquierdo muestra una forma de embudo por lo que tampoco podemos confirmar este diagnóstico.

Ahora calcularemos los residuos para poder realizar unos gráficos adicionales.

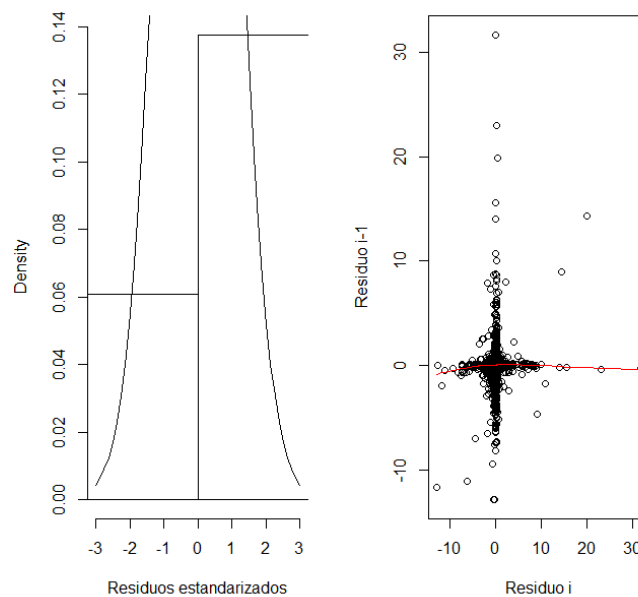


Gráfico 16

En el gráfico de Residuos estandarizados confirmamos que la **normalidad** no se cumple porque las columnas no quedan representadas dentro de la línea que dibujamos de campana.

En el gráfico de Residuo i vs $i-1$ observamos que la línea roja representada es bastante horizontal por lo que podemos indicar que existe autocorrelación en el modelo.

CONTRASTE DE HIPOTESIS

$H_0: e \sim N$

H_1 : Lo contrario

El p -value $< 2.2 \cdot 10^{-16} < 0.05$ donde también rechazamos la hipótesis de **normalidad** definitivamente.

Homocedasticidad (hipótesis de varianza constante)

H_0 : Homocedasticidad

H_1 : Heterocedasticidad

El test de Breusch-Pagan se obtiene con

El p -valor obtenido $< 2.2 \cdot 10^{-16} < 0.05$ por lo que podemos rechazar la hipótesis de varianza constante de Homocedasticidad.

Para salir de dudas hallamos el siguiente contraste:

$H_0: \rho = 0$

$H_1: \rho \neq 0$

El test de Durbin-Watson:

p -value $< 2 \cdot 10^{-16} < 0.05$ por lo que rechazamos la hipótesis de incorrelación.

Hipótesis de media cero:

Siempre se verifica por cómo ajustamos el modelo (EMV)

En conclusión podemos dar por bueno el modelo por varios motivos como son:

- Aunque en la bondad del ajuste del error estándar residual es elevado, un 40%, la tabla de ANOVA y R cuadrado se aprueban perfectamente por lo que podemos dar por buena la bondad del ajuste.
- No se cumplen los diagnósticos del modelo de Normalidad, Homocedasticidad e Incorrelación.
- En definitiva podemos confirmar que la variable que más explica VOL es PTR, que resulta ser el modelo de cálculo utilizado en la actualidad, por lo tanto con los modelos propuestos no podemos mejorar el método de cálculo del VOL, salvo probablemente con el modelo completo.

TRANSFORMACIÓN DEL MODELO

Intentamos transformar el modelo con BOX-COX pero $\lambda=0.8$ lo que significa que no podemos mejorar el modelo ya que el valor de λ es muy próximo a 1.

CRITERIO AIC

Con el criterio AIC intentamos reducir el número de variables explicativas del modelo original, pero al aplicar este criterio confirmamos que no podemos reducir ninguna variable (x).

6. PREDICCIÓN DEL VOLUMEN CON DATOS DE 2014

Tomando una muestra de 31 días elegidos al azar desde febrero a abril de 2014 con un total de 4.389 registros, calculamos el volumen según cada uno de los cuatro métodos analizados. Para cada valor obtenido de los diferentes métodos aplicamos la recta estimada y hallamos registro a registro el error cometido con la estimación.

En la Tabla 1 podemos observar el error promedio para cada uno de los métodos de forma individual.

Tabla 1

	ERRORES PROMEDIO
PTR	19.66%
MOD	-43.79%
DENS	-1.77%
BLT	-3.57%
TODOS	43.02%

Según podemos observar en la Tabla 1 los métodos DENS y BLT son los que más se aproximan a explicar el VOL final con un menor error promedio cometido.

7. CONCLUSIONES

En este apartado expondremos las principales conclusiones a las que hemos ido llegando con cada uno de los análisis realizados.

Hemos ido analizando cuatro métodos diferentes de cálculo realizando para cada uno de ellos un Análisis Lineal Simple y un Análisis Lineal Múltiple con todos los métodos propuestos.

En el análisis preliminar hemos observado tanto gráfica como numéricamente que todos los modelos propuestos tienen una relación bastante elevada, ninguno de los modelos es inferior al 95% (correlación simple) como podemos ver en el resumen de la Tabla 2.

Para analizar la bondad del ajuste hemos tenido en cuenta tres criterios como son el error estándar residual, donde en todos los casos el valor obtenido es superior al 10%, y por tanto demasiado elevado, R-squared o R^2 donde los valores obtenidos son muy buenos ya que son superiores al 65% y muy cercanos a la unidad, y por último el criterio de ANOVA donde los valores son inferiores a 0,05 y por tanto hemos podido descartar que el valor pueda ser igual a cero.

Desafortunadamente no se cumplen los principios de normalidad, Homocedasticidad e incorrelación.

Se ha intentado realizar una transformación de las diferentes variables respuesta con BOX-COX pero el valor obtenido para lambda que es el valor que transforma la variable (y), en este caso VOL, en este caso es un valor igual o muy cercano a uno lo cual nos indica que el modelo no se puede mejorar mediante este método de transformación.

Tabla 2

	<u>PRT</u>	<u>MOD</u>	<u>DENS</u>	<u>BLT</u>
Preliminar	98,47% ✓	99,63% ✓	97,58% ✓	97,55% ✓
Error st. Res.	53,63% ✗	25,59% ✗	46,82% ✗	47,03% ✗
R-squared	96,97% ✓	99,25% ✓	95,21% ✓	95,17% ✓
ANOVA	$<2.2*10^{-16}$ ✓	$<2.2*10^{-16}$ ✓	$<2.2*10^{-16}$ ✓	$<2.2*10^{-16}$ ✓
Normalidad	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗
Homocedasticidad	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗
Incorrelación	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗
Media Cero	siempre se cumple ✓	siempre se cumple ✓	siempre se cumple ✓	siempre se cumple ✓
RESULTADO	OK ✓	OK ✓	OK ✓	OK ✓

	<u>TODOS</u>			
	<u>PTR</u>	<u>MOD</u>	<u>DENS</u>	<u>BLT</u>
Preliminar Simp.	98,55% ✓	96,61% ✓	95,66% ✓	95,76% ✓
Preliminar Parc.	81,28% ✓	21,33% ✗	-4,28% ✗	10,01% ✗
Error st. Res.	40,33% ✗	40,33% ✗	40,33% ✗	40,33% ✗
Multiple R-squared	98,15% ✓	98,15% ✓	98,15% ✓	98,15% ✓
Adjusted R-squared	98,15% ✓	98,15% ✓	98,15% ✓	98,15% ✓
ANOVA	$<2*10^{-16}$ ✓	$<2*10^{-16}$ ✓	$<2*10^{-16}$ ✓	$<2*10^{-16}$ ✓
Normalidad	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗
Homocedasticidad	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗
Incorrelación	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗	$<2.2*10^{-16}$ ✗
Media Cero	siempre se cumple ✓	siempre se cumple ✓	siempre se cumple ✓	siempre se cumple ✓

8. BIBLIOGRAFIA

Aparicio J, Martínez Mayoral M.A. y Morales J. Modelos Lineales Aplicados en R (consultado en mayo 2014) disponible en <http://umh3067.edu.umh.es/wpcontent/uploads/sites/240/2013/02/Modelos-Lineales-Aplicados-en-R.pdf>

Dosier de prensa de Tempe, S.A. Grupo INDITEX (consultado en mayo 2014) <http://www.tempe.es/documents/11643/54052/Tempe-Grupo-INDITEX-Press-ESP.pdf/7293490b-7962-4c39-a64d-7b12765beed7>

Información de las bases de datos de Tempe, S.A. Grupo INDITEX con su autorización.

Beatriz Asunción Ramón Puente, Estudio de la relación existente entre la Q de Tobin y el modelo CAMELS en un conjunto de bancos europeos, año 2013: Trabajo Final de Grado CCSSJJ de Orihuela, Universidad Miguel Hernández de Elche.

