



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CAMPUS D'ALCOI

Obtención de un modelo econométrico para explicar la producción de energía en Europa

MEMORIA PRESENTADA POR:

Ana Teresa Olcina Climent

GRADO DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Convocatoria de defensa: Marzo 2016

ÍNDICE

1. INTRODUCCIÓN Y OBJETIVOS	1
1.1 Motivación	2
1.2 Hipótesis.....	5
1.3 Objetivos	6
2. METODOLOGÍA Y TÉCNICAS.....	7
2.1 Metodología.....	8
2.2 Técnicas.....	9
3. CÁLCULOS, RESULTADOS Y EXPLOTACIÓN	30
3.1 Univariante.....	32
3.2 Bivariante	45
3.3 Multivariante.....	63
3.4 Predicciones	82
4. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN.....	84
4.1 CONCLUSIONES	85
4.2 FUTURAS LÍNEAS DE INVESTIGACIÓN	86
5. BIBLIOGRAFÍA	87

1. INTRODUCCIÓN Y OBJETIVOS

1.1 Motivación

La energía, una capacidad para generar un movimiento, que está presente todos los días en nuestras vidas de diferentes maneras. Como puede afectar en la actualidad a nivel económico y social en el día a día de una persona.

La decisión de realizar un proyecto sobre la energía, ya que no es un tema relacionado con mis estudios de Administración y Dirección de Empresas, ha sido porque hoy en día, es un tema que nos puede afectar fácilmente, ya que cada vez existen más tipos de energía que son más económicos, que consumen menores recursos naturales y con mejores finalidades para mejorar nuestro planeta.

La estructura del sector de la energía ha sufrido una gran transformación desde el año 1998 hasta la norma del año 2013 que en la actualidad regula la estructura y el funcionamiento del sector. En el año 1998, la actividad del sector estaba estructurada de forma vertical y que ejercían un monopolio en las distintas regiones españolas, en el 2007 se aprobó otra norma en la que permitió la apertura de redes con terceros, el establecimiento de un mercado organizado de negociación de la energía y la reducción de la intervención pública en la gestión del sistema. La norma básica de la actualidad mantiene la distinción entre las actividades reguladas y no reguladas, al mismo tiempo se impulsa la competencia del sector, mejorando la posición del consumidor en cuanto a información y facilitando los procesos de cambio de suministrador [1].

El suministro de energía eléctrica se define como la entrega de energía a través de las redes de transporte y distribución mediante contraprestación económica en las condiciones de regularidad y calidad que resulten exigibles. Las actividades destinadas al suministro de energía eléctrica son: generación, transporte, servicios de recarga energética, comercialización e intercambios intracomunitarios e internacionales, así como la gestión económica y técnica del sistema eléctrico [1].

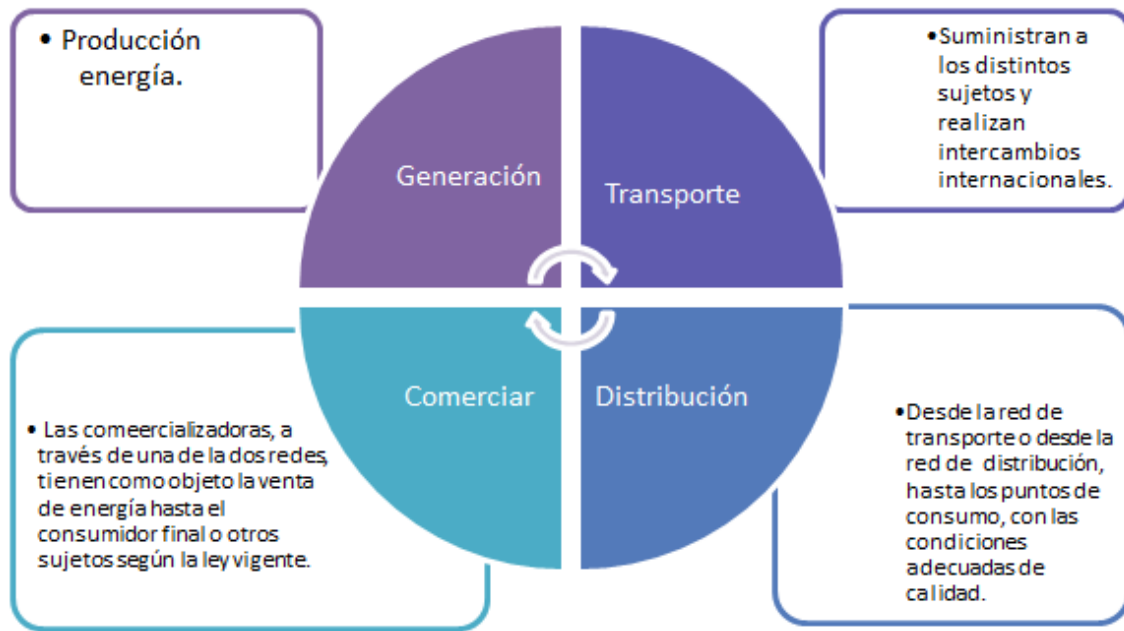


Ilustración 1. 1 Proceso para suministrar la energía.

Las energías convencionales frente a las energías renovables, recursos limpios e inagotables que proporciona la naturaleza tienen un impacto nulo y reversible, por lo cual, el Ministerio de Industria, Energía y Turismo está trabajando en energías como:

1. Energías Renovables, que contribuyen a disminuir la dependencia de nuestro país de los suministros externos, disminuyen el riesgo de un aprovisionamiento poco diversificado y favorecen el desarrollo tecnológico y la creación de empleo. Hay una colección de seis manuales monográficos de las tecnologías de energías renovables, donde se explica cada una de las siguientes: eólica, solar térmica, mini centrales hidroeléctricas, biomasa, geotérmica, biocarburantes del transporte. Esta colección es una consecuencia de un convenio de colaboración firmado por el instituto para la Diversificación y Ahorro de Energía (IDEA) y la Asociación de Periodistas de Información Ambiental (APIA) [2].
2. Energía e I+D+i, por lo que conlleva a Investigación, desarrollo e internalización, colaboran con la Unión Europea, con proyectos de Horizonte 2020 [3].
3. La eficiencia energética, permite el ahorro, además el gobierno impulsa auditorías energéticas para las grandes empresas para mejorar el consumo [4].

Cuando se empezó este proyecto se buscaron diferentes variables que tuvieran relación con la energía, se utilizó la base de datos Eurostat, se decidió por esta base de datos, ya que al ser a nivel Europeo, se podía extraer más información y hacer un análisis más amplio. Se buscaron variables que tuvieran relación con los indicadores principales, precios y cantidades. En un principio se escogieron 12 variables explicativas o independientes para explicar, la variable dependiente que en este caso, era interpretar la producción de cada uno de los países de Europa. Se decidió, por la PRODUCCIÓN DE ENERGÍA, de las 12 elegidas se eliminaron dos que se pensó, que podían explicar todo el modelo económico, finalmente se trabajó con 10 variables independientes más la variable dependiente.

Por último, las asignaturas elegidas para analizar este tema, han sido Introducción a la Estadística y Econometría, por el motivo en que, en la primera de ellas se ha podido comprobar, mediante un análisis univariante, el comportamiento de cada una de estas variables elegidas y en la segunda ofrece una herramienta que recibe el nombre de Programa Statgraphics,[5] el cual, es una potente herramienta para analizar gran variedad de datos y en el que también proporciona diferentes tipos de gráficos, donde, se pueden observar soluciones rápidas y poder llegar a unos objetivos claros y concisos.

1.2 Hipótesis

- Explicar la variable Producción de Energía a partir de otras variables explicativas.
- Comprobar si la variable Población (POB) afecta de manera directa al modelo económico, ya que cuando más habitantes, probablemente se tenga que producir más energía.
- Comprobar si la variable Número de Empresas (NºEMP) y Población (POB), tienen influencia directa, en el caso de que, mayor número de habitantes, más empresas en un determinado país.
- Comprobar si países con un gran mayor número de población (POB) se produce más energía.
- Comprobar si depende de la riqueza del país que se produzca más energía o no.

1.3 Objetivos

El objetivo principal del presente trabajo consiste en analizar y estudiar el comportamiento de la producción de energía a nivel europeo para obtener un modelo matemático que explique la producción a partir de unas variables explicativas. Para conseguir esto se plantean los siguientes objetivos secundarios:

- Realizar un análisis descriptivo de las variables que puedan explicar la producción de energía.
- Comprobar la existencia de relación lineal entre las variables independientes y la Producción de Energía.
- Determinar si existe relación entre las variables explicativas mediante el estudio de multicolinealidad.
- Realizar el ajuste y validación de un modelo econométrico con las variables explicativas resultantes del análisis de multicolinealidad frente a la Producción de Energía.
- Identificar si algún país tiene un comportamiento distinto al resto.
- Realizar predicciones con el modelo definitivo para ver que cabría esperar de un país que modificara su comportamiento.

2. METODOLOGÍA Y TÉCNICAS

2.1 Metodología

A continuación, en la (Ilustración 2.1) se desglosa cada uno de los aspectos del algoritmo de procesos que se va a seguir durante el trabajo, mostrando los análisis en concreto que se han realizado.

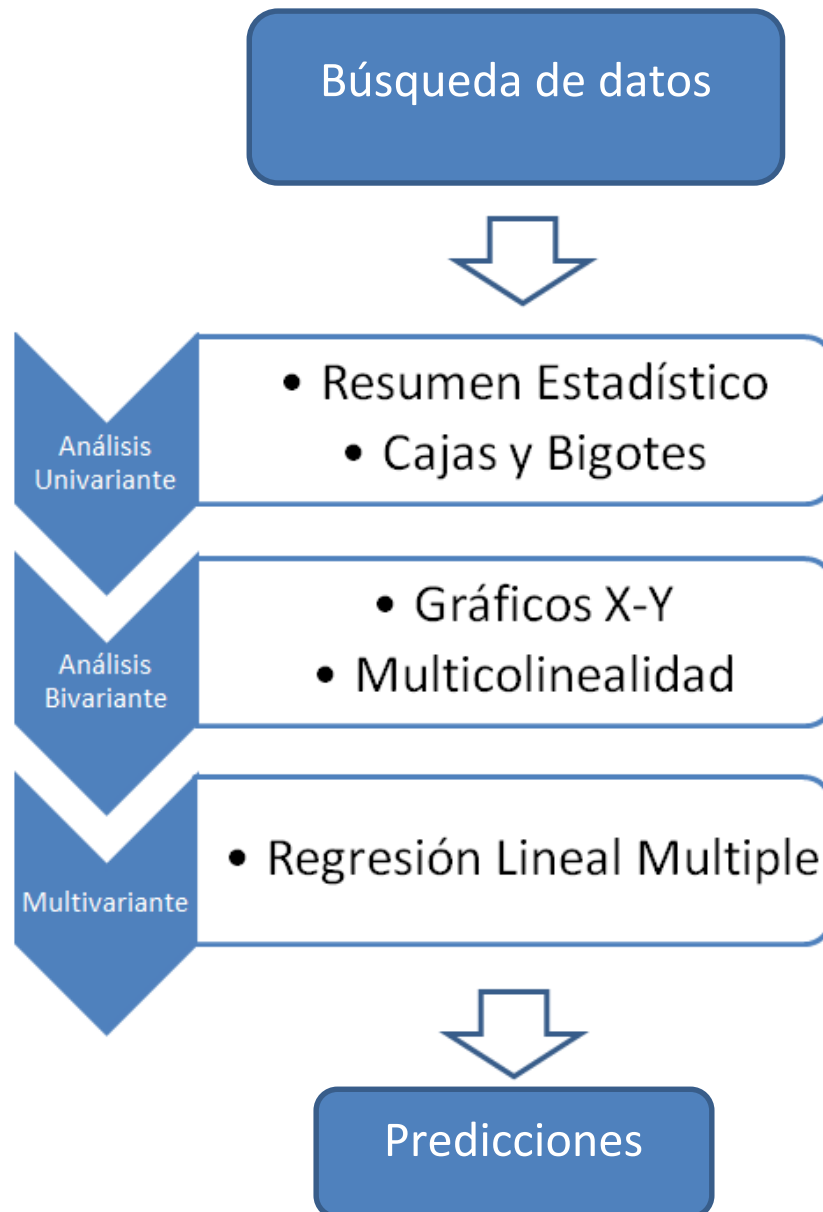


Ilustración 2. 1 Algoritmo de Procesos.

2.2 Técnicas

2.2.1 Análisis Univariante

El análisis Univariante estudia el comportamiento de cada una de las variables que se va a analizar por separado, por lo cual, se analizará el resumen de estadísticos que incluye las medidas de centralización, dispersión, y asimetría. Finalmente, también se estudiará el gráfico de cajas y bigotes que resume parte de estas medidas.

2.2.1.1 Resumen de Estadísticos

El Resumen de Estadísticos de una variable resumirá la información más relevante. Existen dos tipos de medidas:

- Medidas de centralización o de tendencia central que dan información sobre la posición de una variable, es decir, sobre los valores en torno a los cuales se sitúan los datos. Las medidas de centralización son:

- i) Media: La media de una variable es la suma total de los valores observados dividida entre el número total de datos, es la medida de centralización más usual, pero es poco informativa en el caso de que existan datos atípicos (outliers). Es obvio que la media siempre estará entre el mínimo y el máximo de la muestra y por lo tanto da una idea clara sobre la posición de los datos.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

- ii) Mediana: La mediana de una muestra es el valor central de los datos observados, es decir, es el valor que deja tanto a su izquierda y a su derecha el mismo número de datos cuando están ordenados de menor a mayor, también se puede definir como el segundo cuartil. Esta medida no se ve afectada por la presencia de datos extremos. Ambas son usadas en variables cuantitativas.

$$\text{Si } n \text{ es impar: } Me = X\left(\frac{n+1}{2}\right)$$

$$\text{Si } n \text{ es par: } Me = \frac{x_{\frac{n}{2}} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

- Medidas de dispersión o variación.
 - i) Rango, es la diferencia entre el mayor y el menor valor de la muestra. Se puede decir que es la medida de variación más simple, pero es muy sensible a la presencia de datos atípicos.

$$R = x_{Max} - x_{Min}$$

- ii) 1^{er} Cuartil, C1, es el valor que deja por debajo el 25% de los datos.

iii) 3^{er} Cuartil, C3, es el valor que deja por debajo de 75% de los datos.

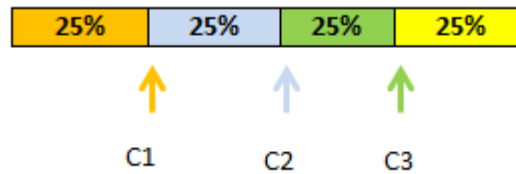


Ilustración 2. 2 Los cuartiles dividen la muestra ordenada en cuatro segmentos con igual número de datos.

iv) Varianza, la varianza de una muestra de valores $x_1, x_2, x_3, \dots, x_n$ de una variable estadística es, la medida de la distancia de los valores a la media. Se toman las distancias al cuadrado para que no se anulen unas “distancias positivas” con otras “distancias negativas”. Por ello, la varianza viene expresada en las unidades de la variable al cuadrado.

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

v) Desviación Típica, la desviación típica de una muestra de valores $x_1, x_2, x_3, \dots, x_n$ de una variable estadística es la raíz cuadrada de la varianza. Esta misma, viene expresada en las mismas unidades que la variable.

$$s = \sqrt{s^2}$$

vi) Mínimo, es el valor mínimo de la variable.

vii) Máximo, es el valor máximo de la variable.

viii) Asimetría, en esta medida se puede conocer si los datos están dispuestos de forma simétrica en torno a un valor central de posición, que en realidad es la media aritmética [6].

$$CA = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{(N - 1) \cdot S^3}$$

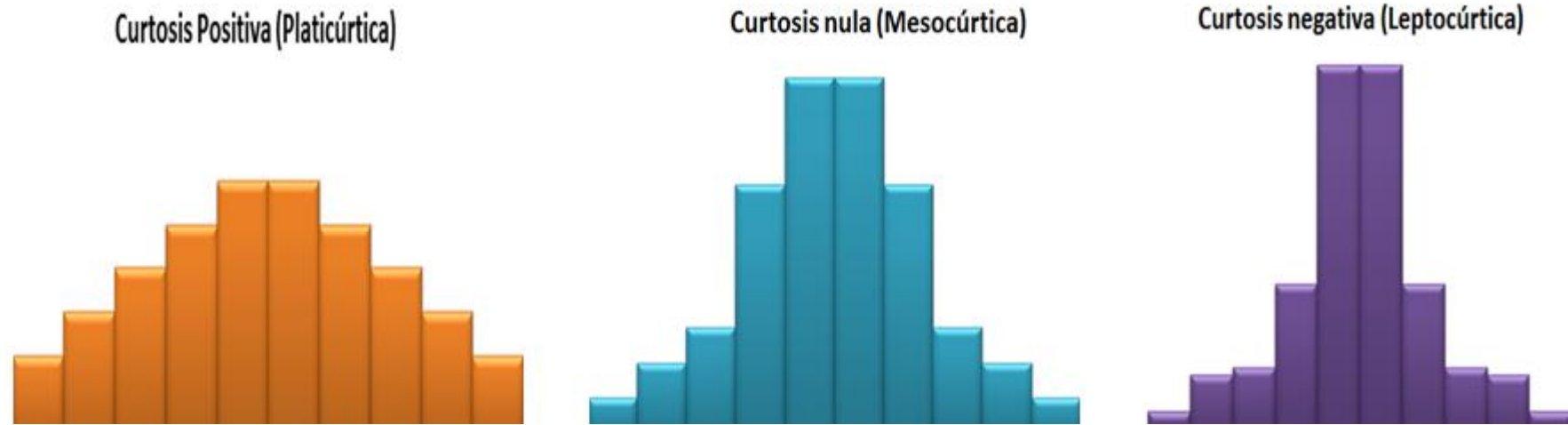


Ilustración 2. 3 Tipos de curtosis.

ix) Curtosis Tipificada, la curtosis permite conocer si unos datos están más cerca o lejos de lo que sería lo esperado con unos datos normales, es decir, lo que se llega a medir realmente es la mayor o menor concentración de la frecuencia alrededor de la media [6].

$$CC = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{(N-1) \cdot S^4} - 3$$

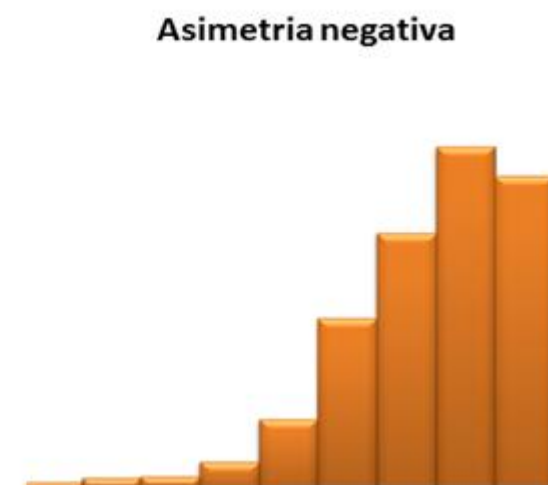
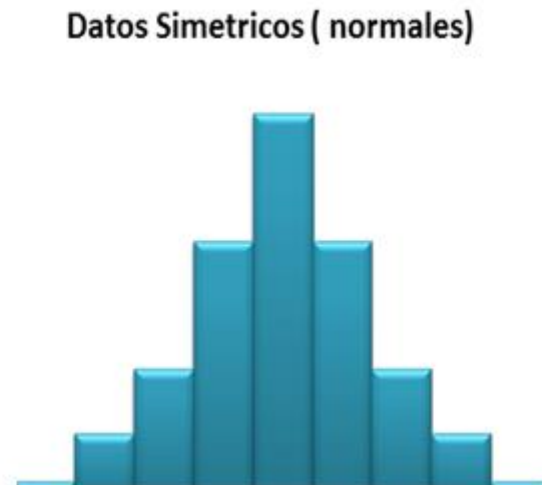
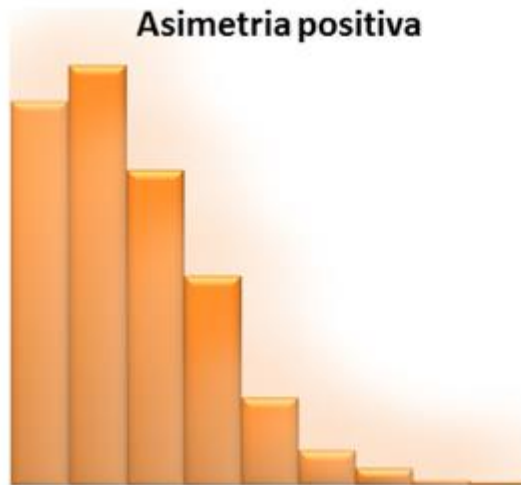


Ilustración 2. 4 Tipos de asimetría.

2.2.1.2 Gráfico de Cajas y Bigotes

El gráfico de cajas y bigotes es capaz de resumir al mismo tiempo seis estadísticos a la vez, tanto de dispersión y posición, como se representa en la (Ilustración 2. 5), la cual representa el C1 (cuartil 1), C2 (cuartil 2) a lo que también se llama mediana, C3 (cuartil 3), los valores máximo y mínimo, y la media. A la vez también puede mostrar valores atípicos tanto inferiores y superiores, que se encuentran fuera de la distribución del diagrama y la simetría de la distribución [7].

Está compuesta por un rectángulo (caja) y dos brazos (bigotes).

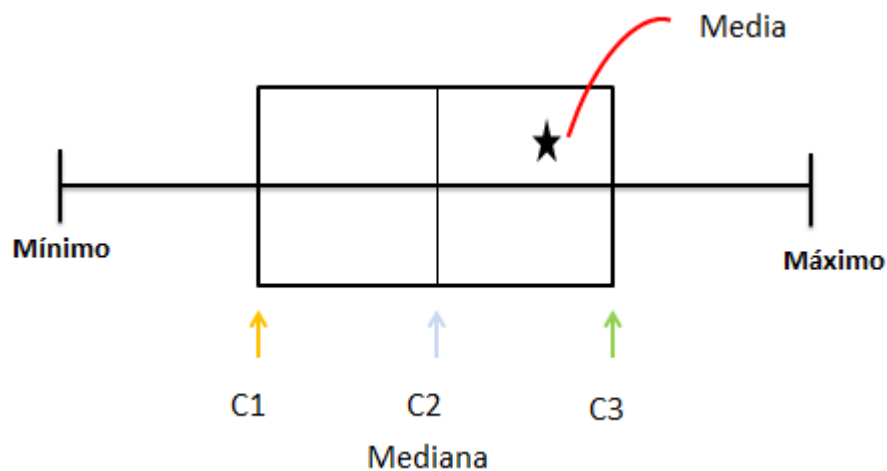


Ilustración 2. 5 Gráfico de Cajas y Bigotes.

2.2.2 Análisis Bivariante

En el apartado anterior se ha explicado las características que pertenecen a un único carácter. Pero en el modelo matemático a explicar, es necesario estudiar el comportamiento de dos o más caracteres y la relación que existe entre ellos.

2.2.2.1 Gráfico X-Y

Es una representación de dos variables en un plano de dos dimensiones, el cual, sirve para ver si existe algún tipo de relación entre la variable Y con la variable X.

2.2.2.2 Multicolinealidad

Se entiende por concepto de multicolinealidad cuando las variables explicativas están relacionadas entre sí, de manera perfecta o casi perfecta. La existencia de multicolinealidad puede provocar problemas en el ajuste del modelo econométrico, debido a que existe una misma información explicada por dos o más variables a la vez.

Ejemplo:

Se plantea un modelo económico que es el siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

Y= Ingresos

X₁: Renta de las familias

X₂: Gastos de las familias

Donde X₁ está relacionada con X₂, y como el modelo económico en el que se va a trabajar existen más de dos variables explicativas, se comprobará mediante la (Ilustración 2. 7) que ocurre con ellas y como afecta cada variable al explicar la variable Y. Pero antes en la (Ilustración 2. 6) se explicará la relación existente entre dos variables para que todo esto se entienda mejor.

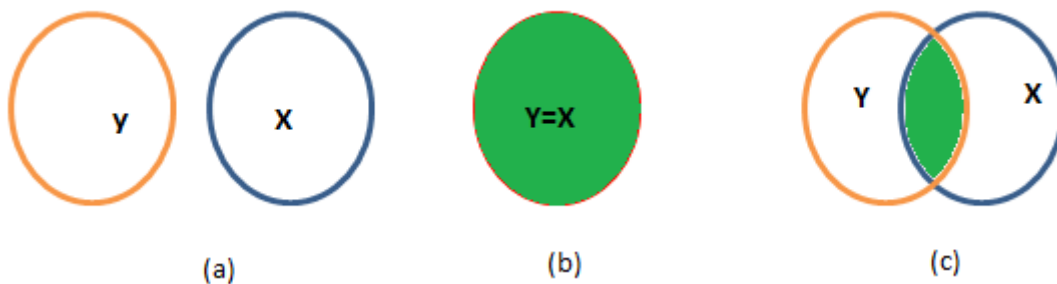


Ilustración 2. 6 Relación entre dos variables.

En la (Ilustración 2. 6.a) significa que ambas no están relacionadas entre sí, lo contrario que en la (Ilustración 2. 6.b) y en el caso de la (Ilustración 2. 6.c) corresponde a la situación más habitual que se puede encontrar que la variable Y este relacionada con la variable X, pero esa relación no es perfecta total, y la variable X solo permite explicar una parte una parte de la variable Y, la cual aparece en color verde.

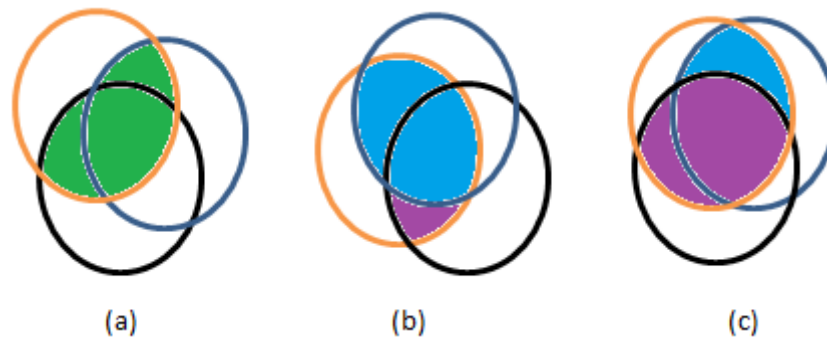


Ilustración 2. 7 Comportamiento de las variables explicativas al explicar Y.

En la (Ilustración 2. 7.a) las variables explicativas X_1 Renta (esfera azul) y X_2 Gastos (esfera negra) ambas relacionadas entre sí, contribuyen de la misma manera a explicar la variable ingresos (esfera naranja). En el caso de la (Ilustración 2. 7.b) presentaría cuando se trata de descomponer la suma de cuadrados explicada en la colaboración de cada una de las dos variables. Por tanto, si a la hora de realizar el cálculo la primera variable explicativa a introducir, que en este caso es la X_1 Renta (color azul), resulta que le corresponde una suma de cuadrados mayor (trazo pintado de color azul) que a la variable X_2 Gastos (color violeta), ya que esta presentación es tan pequeña al hacer la prueba F , para comprobar si es explicativa la respuesta sería que no. En definitiva X_1 sería explicativa y X_2 no.

Por último en la (Ilustración 2. 7.c) en esta se ha descompuesto la suma de cuadrados también, pero en primer lugar se introduce la variable X_2 gastos (esfera negra) trazo pintado de color violeta es bastante grande se aceptaría como significativa y después la variable X_1 (esfera azul) trazo pintado de color azul, ya que ambos trazos son bastante semejantes se podrían aceptar las dos variables como significativas.

¿Cómo identificamos si existe o no multicolinealidad?

Existen varios métodos pero en el modelo económico, se van utilizar dos métodos para identificarla:

1. Matriz de correlación R y su inversa R^{-1} .

Lo que explica la matriz de correlación es que cuantifica la relación entre una variable explicativa y otra, es decir, entre parejas, por lo tanto si existen relaciones lineales entre las variables explicativas, el coeficiente de correlación lineal es un método ideal para cuantificar dicha relación. La matriz de correlación R de las variables explicativas es una matriz simétrica de orden K cuyo término general ij es el coeficiente de correlación lineal simple entre las variables X_j y Y_j .

$$R_{ij} = \frac{cov(X_i, X_j)}{\sqrt{var(X_i)var(X_j)}}$$

Por lo cual, la matriz de correlación R tiene una diagonal principal de 1 y el resto de los elementos son los coeficientes de correlación de cada pareja de variables explicativas. En conclusión, en el caso de que por debajo de la diagonal principal exista algún valor de $|R_{ij}| \geq 0,7$, habrá problema de multicolinealidad.

La inversa de la matriz de correlación R^{-1} , si tiene en cuenta el resto de variables, por lo cual, se compara una variable explicativa con el resto de variables.

$$R_{ii}^{-1} = \frac{1}{1 - R_{i-resto}^2}$$

En este caso si los valores de $R_{ii}^{-1} \geq 10$, se interpreta que X_i está muy relacionada con el resto de las variables, entonces existe multicolinealidad.

2. Método de Belsley, kug, Welsch

Este método se basa en los autovalores de la matriz de correlación R . Belsley, kug, Welsch propusieron el índice de acondicionamiento (I.C) [9].

$$I.C = \sqrt{\frac{\text{Autovalor}_{Max}}{\text{Autovalor}_{Min}}}$$

Si el valor del índice de acondicionamiento esta:

$1 \leq I.C \leq 10 \rightarrow$ No existe multicolinealidad.

$10 \leq I.C \leq 20 \rightarrow$ La multicolinealidad es moderada.

$20 \leq I.C \rightarrow$ La multicolinealidad es elevada.

Para solucionar este tipo de problemas en el que las variables explicativas están altamente relacionadas, la mejor manera es eliminar una variable, para saber que variable se debe eliminar, la matriz de correlación inversa R^{-1} será la que proporciona la ayuda para saber cuál es la variable a eliminar, de hecho, los valores que superen el valor de 10 serían los candidatos, pero se debe eliminar el más grande, en el caso de que los valores sean muy cercanos (del orden de una diferencia menor a la unidad) se plantearían dos modelos del cual se haría una regresión múltiple omitiendo una de las variables candidatas a eliminar. El modelo con un R^2 mayor será con el que se deba continuar. La variable problemática perteneciente al otro modelo sería la candidata a eliminar.

Pongamos el caso que de nuestro modelo ejemplo y le introducimos dos variables más;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + U$$

Existe el problema en la variable X_1 y X_4 , por lo tanto planteamos dos modelos para saber que R_2 es más grande.

$$Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + U \rightarrow R^2 = 87,9\%$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U \rightarrow R^2 = 60,8\%$$

Por lo tanto la variable a eliminar sería la X_1 .

Finalmente ya se tendría el modelo económico definitivo para realizar la regresión lineal múltiple y comprobar cada uno de los siguientes métodos que se van a explicar a continuación, para que el modelo esté libre de otros problemas o si se tienen que solucionar.

2.2.3 Análisis Multivariante

La Regresión Lineal Múltiple (RLM) pretende conseguir un modelo matemático, en el que, de unas variables explicativas (X) se consiga explicar una variable dependiente (Y).

Para que el modelo sea válido se deben cumplir una serie de características como la no presencia de heteroscedasticidad, y autocorrelación, y la normalidad de residuos, etc..., cada una de estas características se puede comprobar mediante contraste de hipótesis para saber si existe algún problema, todo esto será explicado a continuación.

También se comprobará el coeficiente de determinación R^2 ayuda a explicar mejor si un modelo es adecuado o no, es decir, explica lo bastante bien la variable a estudiar o para comparar diferentes estructuras del modelo.

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

El coeficiente de determinación siempre toma valores entre cero y uno.

$$0 < R^2 \leq 1$$

El modelo explica mejor los datos cuanto más se acerca R^2 a 1.

$$-1 < R \leq 1$$

El signo indica si la pendiente es positiva o negativa. En valor absoluto explica lo mismo que R^2 .

2.2.3.1 Contraste de Significatividad

Una vez obtenido el modelo matemático, a partir de que se hayan eliminado las variables correspondientes después de haber realizado el análisis de multicolinealidad, se realiza la regresión múltiple para saber que variables son significativas o no.

Hay diversas maneras de identificar si el modelo y los parámetros son significativos o no, mediante:

- Se compara el estadístico F_{calc} frente $T_{teórico}$, este se utiliza para el modelo.
- Se compara el estadístico T_{calc} frente $T_{teórico}$, este se utiliza para los parámetros.
- P-valor para ambos.

En este proyecto se utiliza P-Valor.

$$P - Value = P(F_{k,n-k-1}^{\alpha} \geq F_{calc})$$

Primero se realiza un contraste del modelo que acompaña a la variable explicativa y luego de cada uno de los parámetros que acompañan al modelo.

$$\left. \begin{array}{l} H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0 \\ H_1 : \text{Algún } \beta_i \neq 0 \end{array} \right\} \text{Contraste modelo}$$

$$\left. \begin{array}{l} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{array} \right\} \text{Contraste de uno de los parámetros}$$

Si $P\text{-valor} \geq \alpha$ se acepta H_0 , donde $\alpha=0,05$ y por tanto concluirá que el modelo o/y los parámetros no son significativos.

2.2.3.2 Normalidad de residuos

En este apartado se explicará, el papel probabilístico normal y el histograma acompañado de su contraste de hipótesis, ya que ambos son útiles para comprobar la normalidad de residuos.

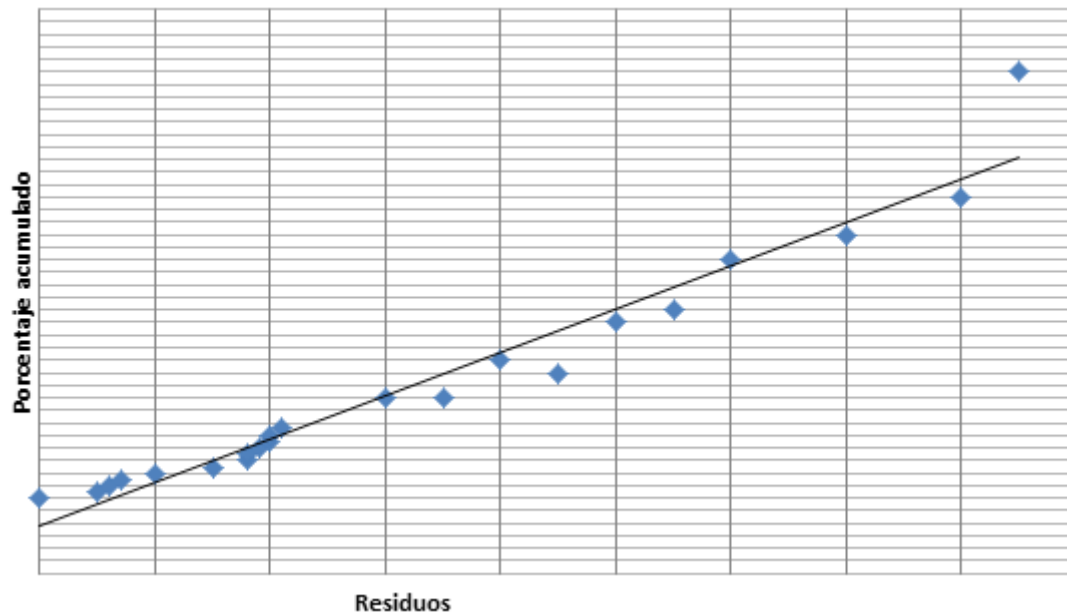


Ilustración 2. 8 Papel probabilístico normal.

La (Ilustración 2. 8) representa la alineación de los residuos (puntos azules) que indicará la normalidad de los mismos.

El papel probabilístico normal muestra una mayor cantidad de información sobre la distribución de los residuos que cualquier otro gráfico o prueba numérica, ya que, permite que se tomen decisiones importantes como si se tiene que modificar el modelo para conseguir la normalidad.

Por lo tanto se analizan los (puntos azules) se dice que existe normalidad de residuos, cuando más cercanos estén a la línea de tendencia, si están lejos no son normales, los cuales, podrían ser puntos anómalos o influyentes, que, se explicaran más adelante.

Otra forma para analizar la normalidad de residuos es mediante el histograma que se explicara a continuación:



Ilustración 2. 9 Histograma.

Si el histograma representa una forma de campana perfecta de Gauss o casi perfecta, indicara que la mayoría de los residuos están en la media, por tanto si existiría la normalidad de residuos.

EL contraste de hipótesis también se determina mediante el P-valor, que mediante la herramienta statgraphics se realiza un test para la normalidad, el cual, calcula tres test que tienen el nombre de:

1. Chi-cuadrado
2. Shapiro Wilks
3. Z para asimetría

Se escoge el más restrictivo de los tres

H_0 : Los residuos son normales

H_1 : Los residuos no son normales

P-valor $\geq \alpha$, donde, $\alpha=0,05$.

Se acepta la hipótesis nula (H_0) y la conclusión sería que los residuos se distribuyen normalmente.

2.2.3.3 Heteroscedasticidad

Se dice que aparece problema de heteroscedasticidad cuando la varianza de la perturbación no es constante, el motivo de que una de las hipótesis del modelo sea que la varianza de U sea constante es porque es más fácil estimar algo que es constante que algo que no lo es. También la heteroscedasticidad surge por la presencia de datos atípico, una observación atípica es la que es muy diferente (muy pequeña o muy grande) en relación con las demás observaciones de la muestra [6].

Existen dos formas para detectar la presencia de heteroscedasticidad en un modelo econométrico:

- Mediante contraste de hipótesis.
- Mediante análisis gráfico.

Para el modelo econométrico que se está analizando se ha utilizado el método de contraste de hipótesis el cuál se explicará a continuación:

- Contraste de Hipótesis.

Se plantea un modelo lineal que relacione los residuos del modelo al cuadrado con nuestras variables explicativas, es decir, de nuestro modelo inicial, se guardaran los residuos en el programa que se está utilizando para llevar a cabo todos los análisis y cálculos, el cual tiene el nombre de Statgraphics. Por ejemplo, nuestro modelo lineal inicial es el siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + U$$

Y una vez guardado los residuos se podrá comprobar el contraste de hipótesis mediante el siguiente modelo:

$$e^2 = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + U$$

Dónde:

$H_0 : \sigma^2 = cte \rightarrow$ No existe heteroscedasticidad

$H_1 : \sigma^2 \neq cte \rightarrow$ Si existe heteroscedasticidad



De manera que si alguno de los parámetros beta que no sea la constante, fuera significativo, se aceptaría que hay heteroscedasticidad. Se sabe si es significativo cuando el P-valor de alguna variable es menor que $\alpha=0,05$.

2.2.3.4 Autocorrelación

En este proyecto para determinar la existencia de autocorrelación, se ha utilizado:

- Contraste de autocorrelación de 1^{er} orden
- Los gráficos FAS y FAP

Autocorrelación de 1^{er} orden.

H ₀ : $\rho_1 = 0$	}	No hay autocorrelación de 1 ^{er} orden
H ₁ : $\rho_1 \neq 0$		Si hay autocorrelación de 1 ^{er} orden

Este contraste se resuelve mediante el P-valor del Estadístico de Durbin – Watson.

Si el P-valor es mayor que $\alpha=0,05$, se acepta la hipótesis nula, por tanto, no existe autocorrelación de 1^{er} orden.

La prueba de Durbin Watson, es una de las pruebas más utilizada y que es calculada, para solucionar el problema detectado de autocorrelación, mediante los dos casos explicados anteriormente, el estadístico de Durbin Watson se define de la siguiente manera:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=2}^n e_t^2}$$

Donde e_t es el residuo estimado para el periodo t.

$$d \approx 2(1 - \hat{\rho})$$

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}$$

En la siguiente imagen se representará como se construyen unos intervalos para la aceptación de que existe o no autocorrelación, y en su caso, de si es positiva o negativa. Como el estadístico d depende en gran manera de la matriz de datos X , no es posible calcular valores críticos de d para todos los casos posibles, por lo que Durbin Watson [9] establecieron las cotas superior d_U e inferior d_L para los valores críticos.

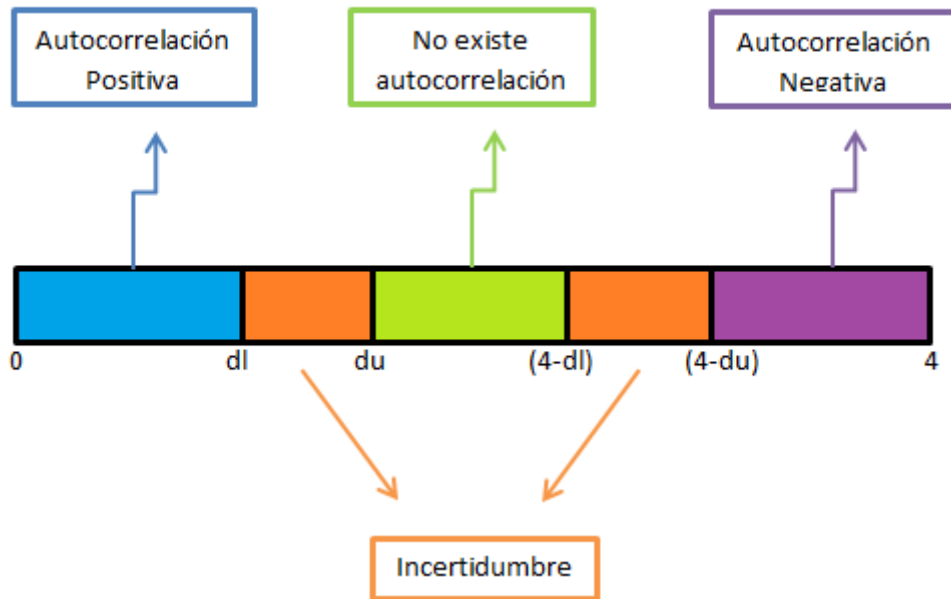


Ilustración 2. 10 Zona de aceptación/rechazo de autocorrelación en la prueba de Durbin Watson.

Representación de la FAS y el FAP, otra mejor forma para asegurar la existencia de autocorrelación, es mediante, la obtención y representación de las Funciones de Autocorrelación Simple y Parcial de los residuos que se muestran a continuación.

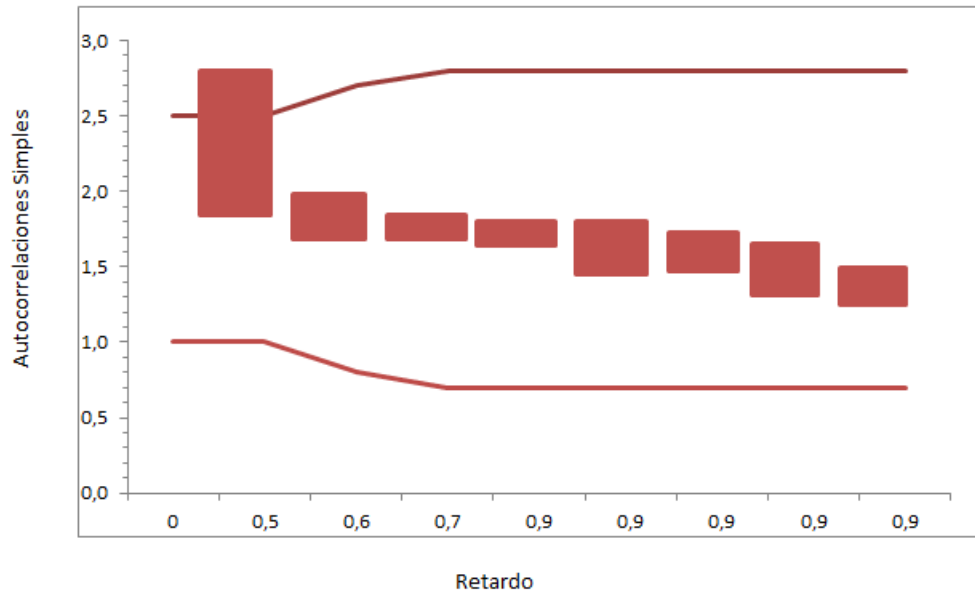


Ilustración 2.11 Función de autocorrelación simple.

La función de autocorrelación simple (FAS) es todo el conjunto de los coeficientes de autocorrelación simple ρ_s , (con $s= 1,2,3,4\dots$) y mide la relación lineal que hay entre el residuo en el instante de tiempo t , e_t , y el residuo en el instante $t-s$, e_{t-s} [9].

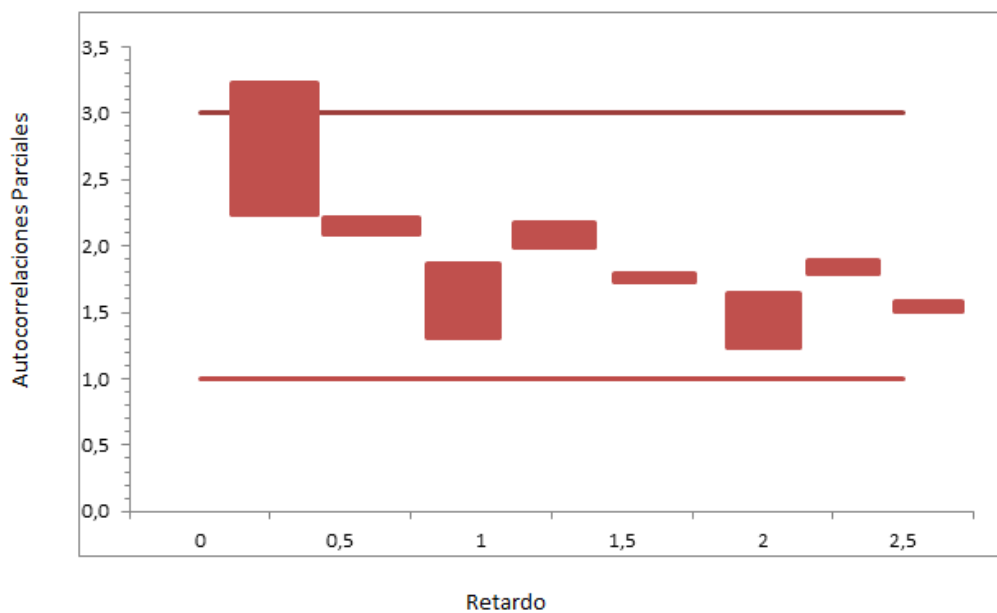


Ilustración 2.12 Función autocorrelación parciales.

La función de autocorrelación parcial (FAP) es el conjunto de coeficiente de autocorrelación parcial (con $s=1, 2, 3, 4, \dots$) y mide la relación directa que existe entre el residuo de tiempo, t , e_t , y el residuo en el instante $t-s$, e_{t-s} , eliminando el posible efecto de los residuos intermedios o ambos [9].

2.2.3.5 Puntos Influyentes y Residuos Atípicos o anómalo.

Los puntos influyentes (outliers), se detectan cuando al realizar una regresión simple, hay un dato que no llega a formar parte esa relación lineal. Una forma de detectarlos es calculando los residuales, como se explicó, anteriormente en la (Ilustración 2. 13).

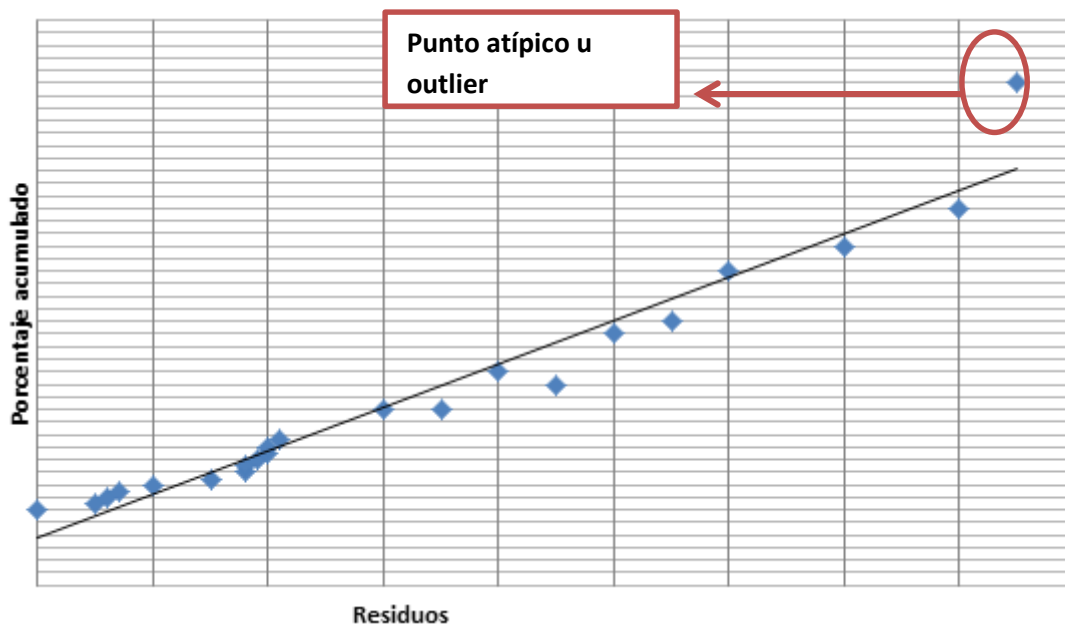


Ilustración 2. 13 Representación de un punto influyente (outlier).

Existen puntos influyentes a priori y posterior. Mediante la herramienta Statgraphics se calcularán dichos puntos, los cuál, el programa muestra la siguiente información, que se tiene que tener en cuenta para poder saber si son puntos influyentes a priori o a posteriori o ambos.

- La influencia(ni)
- DFITS
- La influencia media de un punto (\bar{n})

$$2 \times \bar{n}$$

Si la influencia es mayor que $2 \times \bar{n} \rightarrow$ Es un punto influyente a priori

Si la influencia es menor que $2 \times \bar{n} \rightarrow$ No es un punto influyente a priori

$$2\sqrt{n}$$

Si $|DFITS| > 2\sqrt{n} \rightarrow$ Es un punto influyente a posteriori

Si $|DFITS| < 2\sqrt{n} \rightarrow$ No es un punto influyente a posteriori

Los residuos atípicos o anormales son puntos mal estimados para el modelo. El programa Statgraphics mediante los residuos, calcula un valor estudentizado, el cual, mediante un gráfico se comprueba en que situación, se encuentra el punto. En la (Ilustración 2. 14) se puede observar, donde, se situaría cada valor estudentizado en valor absoluto.

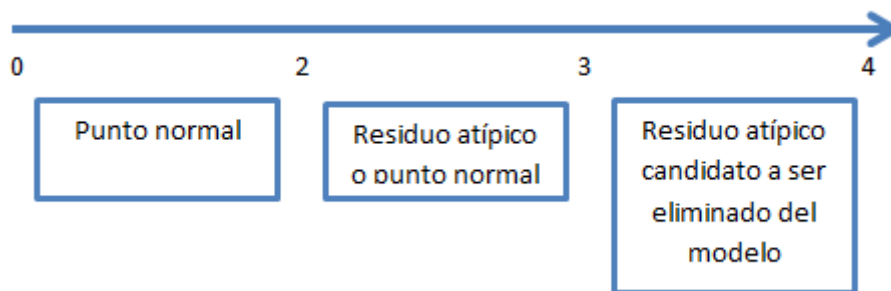


Ilustración 2. 14 Ilustración para localizar los puntos atípicos o anómalos.

2.2.4 Predicciones

Uno de los objetivos que se quiere lograr al hacer modelos económicos, es el poder realizar predicciones de los valores de la variable, por lo cual, a partir del modelo ajustado, se pretende predecir un valor medio ajustado, tanto en el caso si la variable aumenta, disminuye o es el mismo valor dado en los datos iniciales.

Existen dos formas de realizar predicciones:

- Predicción puntual
- Predicción por intervalos

Ambas se ven representadas en la (Ilustración 2. 15).

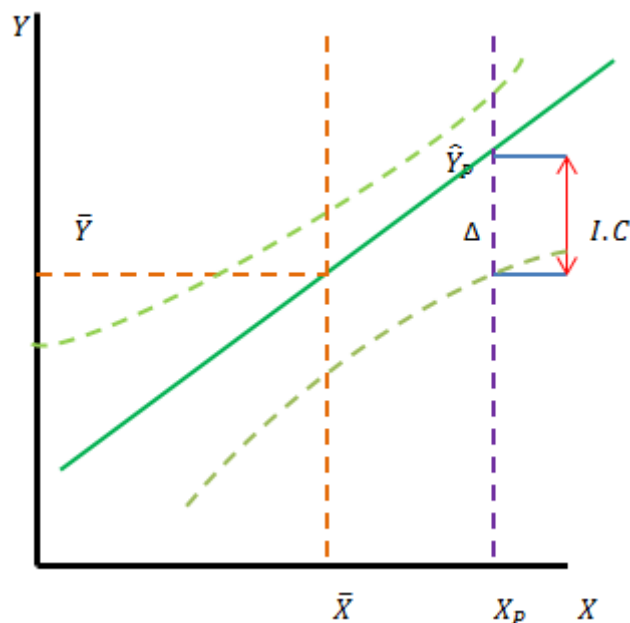


Ilustración 2. 15 Predicción puntual de Y , dado X_p , \hat{Y}_p , y por intervalos de confianza ($I.C$).

La predicción puntual para X_p de la variable explicativa es el valor que se obtiene con el corte de la vertical con la recta \hat{Y}_p . La predicción es el valor esperado de la variable explicada para los valores dados a las variables explicativas.

El intervalo de confianza para X_p , es el intervalo ($I.C$) que se encuentra entre los trazos de las dos curvas de la recta de regresión. Es un intervalo que contiene con una probabilidad determinada el valor verdadero y desconocido de la variable explicada para los valores dados a las variables explicativas. Efectivamente, el valor esperado se encuentra dentro del intervalo de confianza ($I.C$).



3. CÁLCULOS, RESULTADOS Y EXPLOTACIÓN

A continuación se observa la (Tabla 2. 1) donde se muestran todas las variables que vamos a analizar en el modelo teórico, datos extraídos en el Eurostat que es una base de datos donde se han buscado estas variables que aparecen en la tabla a nivel Europeo para el año 2013. No se han podido analizar para todos los países de la Unión Europea, ya que se encontró el problema de que no había datos en todos los países, pues la mejor opción fue eliminar dichos países. En la (Tabla 2. 1) los datos están ordenados alfabéticamente y en el idioma Español.

Tabla 2. 1 Tabla de la base de datos del Eurostat de las variables con las que se va a trabajar a continuación.

Pais	Producción Energía	Consumo energía final	Con. Electrico Hogar	Con. Energetico Hogar Combustible	Con. energia transprte/PIB	Participacion Energias Renov.	Tasa Impositiva	Precio Gas	Precio electricidad	Población (millones)	Nº empresas por pais
Alemania	120566,3	217251,4	11693,9	23,8	81,6	6,3	205,93	18,35	0,2919	80620000	2824
Austria	12104	27950	1520,8	18,5	105,2	7,5	168,98	21,32	0,2082	8474000	851
Belgica	14633,5	34802,3	1703,1	31,5	85,6	4,3	127,74	18,32	0,2173	11200000	363
Bulgaria	10538	8769,6	906	1,1	89,6	5,6	111,06	14,2397	0,0924	7265000	60
Croacia	3624,5	5812,5	535,6	9	107	2,1	148,44	12,9191	0,1372	4253000	64
Dinamarca	16622,5	14194,7	886,2	7,4	92,8	5,7	417,68	27,5854	0,3	5614000	328
Eslovaquia	6408,2	10864,4	423,9	0,4	95,6	5,3	92,36	13,85	0,1698	5414000	8
Eslovenia	3550,5	4798,3	277,6	15,8	117,1	3,4	216,58	18,57	0,161	2060000	8
España	34338,8	81137,6	6235	18,4	80,8	0,4	170,22	20,34	0,2228	47130000	862
Estonia	5653,3	2870,4	160,4	0,8	78,9	0,2	128,98	14,57	0,1351	1325000	29
Francia	135087,1	152056,5	14437,1	16,5	85,2	7,2	214,26	18,83	0,1524	66030000	1654
Grecia	9312,4	15339,6	1500,1	26,3	85,5	1,1	257,19	21,44	0,1563	11030000	68
Irlanda	2268,8	10738,9	683,4	35,7	79,4	5	228,18	18,15	0,2295	4595000	32
Italia	36868,3	118696,2	5759,5	7,9	91,2	5	363,12	23,17	0,2292	59830000	5550
Letonia	2143,5	3855,1	153,3	4,2	86,2	3,1	101,55	14,0521	0,1378	2013000	7
Lituania	1414,2	4736,4	222,8	3,3	85,4	4,6	104,76	16,7458	0,137	2956000	16
Luxemburgo	139,7	4129	75,4	38	99,1	3,9	214,63	17,34	0,1665	543202	39
Países Bajos	69652,2	51155,5	2161	0,8	90,2	5	233,3	22,56	0,1916	16800000	527
Polonia	70578	63412,7	2445,6	2,9	103,4	6	125,82	13,0548	0,148	38530000	389
Portugal	5764,9	15846,7	1058,8	19,5	95,4	0,7	133,13	23,23	0,2081	10460000	288
Reino Unido	109520,4	136432,4	9755,2	6,5	78,6	4,4	247,21	14,7385	0,1741	64100000	757
Republica Checa	29947,7	23860,4	1265,3	0,1	99,2	5,7	127,41	17,8029	0,1525	10520000	65
Rumania	26110,9	21757,8	1022,9	2,6	95,4	4,6	108,61	7,9114	0,1323	19960000	42
Suecia	34683,2	31593,7	3284,9	0,6	77,9	16,7	230,55	34,088	0,2101	9593000	13

3.1 Univariante

En primer lugar se realiza el análisis univariante en el cual se calcula el resumen de estadísticos para cada una de las variables y su respectiva imagen de cajas y bigotes para el año 2013 y los países de la Unión Europea como Alemania, Austria, Bélgica, Bulgaria, Croacia, Dinamarca, Eslovaquia, Eslovenia, España, Estonia, Francia, Grecia, Irlanda, Italia, Letonia, Lituania, Luxemburgo, Países Bajos, Polonia, Portugal, Reino Unido, República Checa, Rumania y Suecia.

La variable dependiente que se utiliza es la Producción de energía primaria por parte de los recursos (1000 toneladas de petróleo equivalente) es cualquier tipo de extracción de productos energéticos a partir de fuentes naturales de una forma utilizable se llama producción primaria, que se lleva a cabo cuando se explotan las fuentes naturales, por ejemplo, en las minas de carbón, yacimientos de petróleo crudo, plantas de energía hidroeléctrica o la fabricación de biocombustibles [10].

Se calcula el resumen de estadísticos (Tabla 3.1) la producción de energía medida por 1000 toneladas de petróleo equivalente en el año 2013 para los países citados anteriormente.

Tabla 3. 1 Resumen de estadísticos para la producción de energía primaria por parte de los recursos (1000 toneladas de petróleo equivalente) año 2013.

Estadístico	Valor
Media	31730,5
Mediana	13368,8
Varianza	1,60E+09
Desviación Típica	39971,1
Mínimo	139,7
Máximo	135087
Rango	134947
1ºCuartil	4638,9
3ºCuartil	35775,8
Asimetría	1,59339
Curtosis Tipificada	1,52097

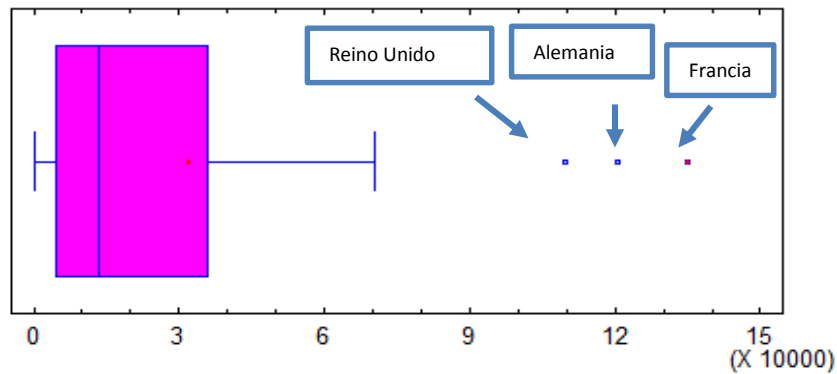


Ilustración 3. 1 Gráfico de Caja y Bigotes de la producción de energía (1000 toneladas por petróleo equivalente) en el año 2013.

Como se observa en la (Ilustración 3. 1) hay tres puntos fuera de la imagen llamados puntos anómalos o atípicos que explican que hay tres países que no se comportan como el resto, esto quiere decir que tienen una producción más elevada.

Hay una gran distancia entre el máximo y el mínimo esto puede ser causa por la elevada población de algunos países como Alemania y la poca población de Luxemburgo.

Luego también tenemos una gran diferencia entre la media y la mediana, lo cual luego puede traer problemas a la hora de realizar el modelo de regresión.

A continuación se realizan los siguientes cálculos de las variables dependientes que se hará una breve explicación sobre cada una de ellas y en el cuál se mostraran los resúmenes estadísticos y su imagen correspondiente de cajas y bigotes.

COENG: Consumo de energía final (1000 tonelada de petróleo equivalente) incluye toda la energía entregada a la puerta del consumidor (en la industria, el transporte, los hogares y otros sectores) para todos los usos de la energía. Los datos de la base del Eurostat representan todos los productos y el consumo de energía de un determinado número de producto o grupos de productos [11].

Se calcula la siguiente (Tabla 3.2) de resumen de estadísticos para el consumo de energía final medida en 1000 toneladas de petróleo equivalente para el año 2013 en cada uno de los países citados anteriormente de Unión Europea.

Tabla 3. 2 Resumen de estadísticos para el consumo de energía final (1000 toneladas de petróleo equivalente) año 2013.

Estadístico	Valor
Media	44252,6
Mediana	18802,3
Varianza	3,23E+09
Desviación Típica	56873,3
Mínimo	2870,4
Máximo	217251
Rango	214381
1ºCuartil	7291,05
3ºCuartil	57284,1
Asimetría	1,82588
Curtosis Tipificada	2,81301

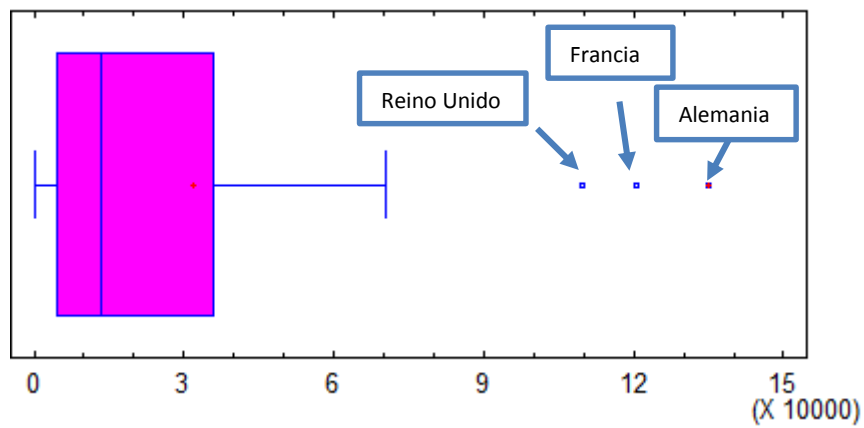


Ilustración 3. 2 Gráfico de cajas y bigotes para el consumo de energía final en el año 2013.

Se observa en la (Ilustración 3. 2) que Alemania es el país que más consume siguiéndole Francia y el Reino Unido.

COELECHOG: El consumo de electricidad de los hogares (1000 toneladas de petróleo equivalente) este indicador se define como la cantidad de electricidad consumida por los hogares. Consumo de los hogares cubre todo el uso de la electricidad para el espacio y el calentamiento de agua y todos los aparatos eléctricos [12].

Se realiza el cálculo de resumen de estadísticos en la (Tabla 3.3) para el consumo de electricidad de los hogares medida en 1000 toneladas de petróleo equivalente en el año 2013 para cada uno de los países citados anteriormente.

Tabla 3. 3 Resumen de estadísticos para el consume de electricidad de los hogares (1000 toneladas de petróleo equivalente) año 2013.

Estadístico	Valor
Media	2840,33
Mediana	1162,05
Varianza	1,54E+07
Desviación Típica	3924,09
Mínimo	75,4
Máximo	14437,1
Rango	14361,7
1ºCuartil	479,75
3ºCuartil	2865,25
Asimetría	1,94713
Curtosis Tipificada	3,0238

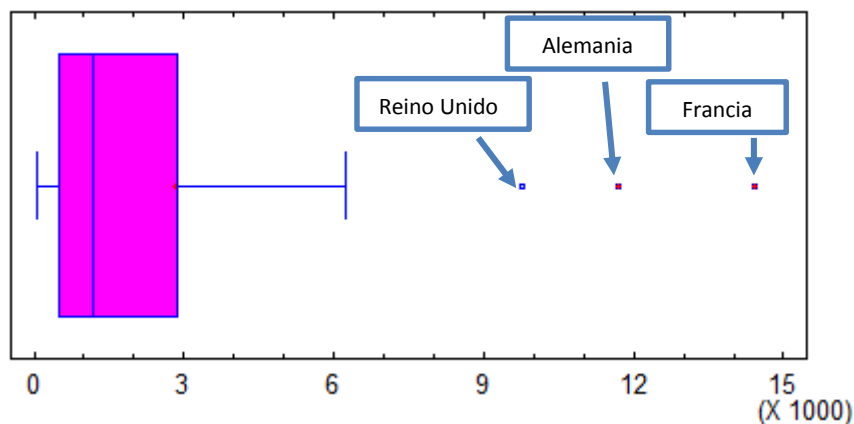


Ilustración 3. 3 Gráfico de cajas y bigotes para el consumo eléctrico de los hogares (1000 toneladas de petróleo equivalente) año 2013.

Como se observa en la (Ilustración 3. 3) volvemos a tener puntos anómalos o atípicos en los países como Francia, Alemania, Reino Unido que es donde se consume más electricidad en los hogares, el nombre de estos tres países se repite en las tres imágenes analizadas (Ilustración 3. 1) (Ilustración 3. 2) y (Ilustración 3. 3) esto es debido a la variable población ya que son países que destacan por su población y es normal que se consuma más que el resto.

COENGHOGCOMB: Consumo de energía final en los hogares por combustible (expresada en tanto por ciento) este indicador presenta la proporción de seis tipos de combustibles para el consumo final de energía residencial: combustibles sólidos, combustibles totales de petróleo, gas, energía eléctrica, calor derivada y energías renovables [13].

Se calcula en la (Tabla 3.4) el resumen de estadísticos para el consumo final de energía en los hogares por combustible se expresa en tanto por ciento (%) del consumo total para el año 2013 en cada uno de los países citados anteriormente.

Tabla 3. 4 Resumen de estadísticos para el consumo final de energía en los hogares por combustible (%) año 2013.

Estadístico	Valor
Media	12,15
Mediana	7,65
Varianza	1,43E+02
Desviación Típica	11,971
Mínimo	0,1
Máximo	38
Rango	37,9
1ºCuartil	1,85
3ºCuartil	19
Asimetría	0,845544
Curtosis Tipificada	-0,429357

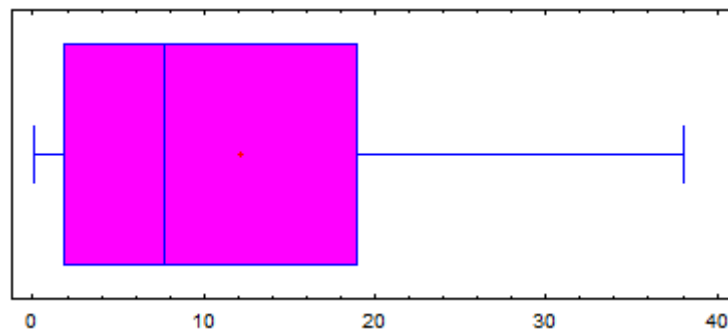


Ilustración 3. 4 Gráfico de cajas y bigotes para el consumo de energía en los hogares por combustible (%) año 2013.

Como se observa en la (Ilustración 3. 4) la mayoría de los países no tiene esta participación a la hora de consumir este tipo de energía en sus hogares ya que la caja se acerca al valor más mínimo y vuelve a ver una gran diferencia en el máximo y el mínimo.

Esta vez no existen puntos anómalos o atípicos por lo tanto, se puede decir que no hay nada que se comporte con una gran diferencia del resto.

COENGTRANSP: El consumo de energía del transporte en relación con el PIB –Índice (2000=100) este indicador se define como la relación entre el consumo de energía del transporte y el PIB (volúmenes encadenados, a tipos de cambio de 2000). La energía consumida por todos los tipos de transporte (carretera, ferrocarril, navegación interior, aviación, transporte comercial, individual y público) con la excepción del transporte marítimo y tubería [14].

Se calcula en la (Tabla 3.5) el resumen de estadísticos de la relación entre el consumo de energía de transporte y el precio interior bruto (PIB) en volúmenes encadenados a tipos de cambio de 2000 para los países citados anteriormente en el año 2013.

Tabla 3. 5 Resumen de estadísticos para el consume de energía del transporte en relación con el precio interior bruto (volúmenes encadenados a tipos de cambio 2000) año 2013.

Estadístico	Valor
Media	91,0958
Mediana	89,9
Varianza	1,06E+02
Desviación Típica	10,2939
Mínimo	77,9
Máximo	117,1
Rango	39,2
1ºCuartil	83,4
3ºCuartil	97,35
Asimetría	0,744156
Curtosis Tipificada	0,149942

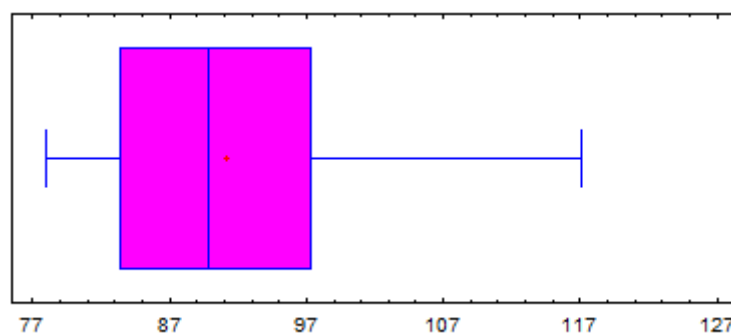


Ilustración 3. 5 Gráfico de caja y bigotes para el consumo de energía del transporte en relación con el precio interior bruto (volúmenes encadenados a tipos de cambio de 2000) para el año 2013.

Como se observa en la (Ilustración 3. 5) los medios de transporte consumen este tipo de energía.

Y según los datos de la media y la mediana, el mínimo y el máximo no existe gran diferencia entre ellos.

PARTENRENOV: Participación de las energías renovables en el consumo de combustible del transporte (%) este indicador se calcula sobre la base de los datos incluidos en el reglamento (CE) nº 1099/2008 relativo a las estadísticas de energía. Países que presentaron informes proporcionan información adicional sobre la fuente renovable no cubiertos por el Reglamento [15].

Se calcula la (Tabla 3.6) de resumen de estadísticos para la participación de las energías renovables en el consumo de combustible de transporte se expresa en tanto por ciento (%) para los países citados anteriormente en el año 2013.

Tabla 3. 6 Resumen de estadísticos para la participación de energías renovables en el consumo de combustible de transporte (%) año 2013.

Estadístico	Valor
Media	4,74167
Mediana	4,8
Varianza	1,07E+01
Desviación Típica	3,26615
Mínimo	0,2
Máximo	16,7
Rango	16,5
1ºCuartil	3,25
3ºCuartil	5,7
Asimetría	1,95817
Curtosis Tipificada	7,40037

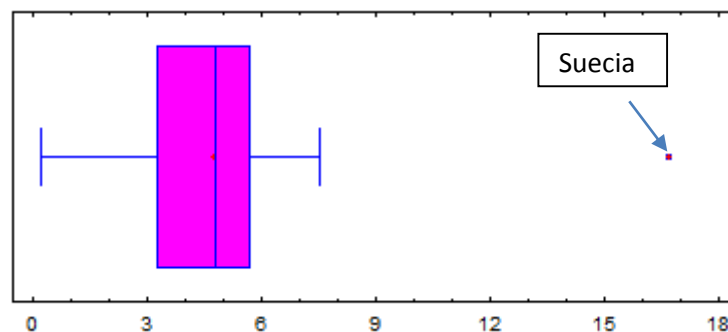


Ilustración 3. 6 Gráfico de cajas y bigotes para la participación de energías renovables en consumo de combustible (%) para el año 2013.

Como se observa en la (Ilustración 3. 6) hay un punto anómalo o atípico perteneciente a Suecia, por lo tanto este país tiene un exceso de consumo en diferencia con el resto de países citados anteriormente en el año 2013.

Entre el máximo y el mínimo si existe gran diferencia que será debido al número de población perteneciente a cada país.

Pero la mediana y la media son prácticamente iguales, esto quiere decir que esta variable no nos perjudicará en el modelo de regresión.

TAIMP: Tasa impositiva implícita sobre la energía (Euros por tonelada de equivalente de petróleo) este indicador se define como la relación entre los ingresos fiscales de energía (se miden en €) y el consumo final se mide en forma de dedo (toneladas equivalentes de petróleo) [16].

Se calcula para la (Tabla 3.7) el resumen de estadísticos para la tasa impositiva expresada en Euros por tonelada de petróleo para el año 2013 en cada uno de los países citados anteriormente.

Tabla 3. 7 Resumen estadísticos de la tasa impositiva (€/tonelada de petróleo) año 2013.

Estadístico	Valor
Media	186,57
Mediana	169,6
Varianza	6,75E+03
Desviación Típica	82,1864
Mínimo	92,36
Máximo	417,68
Rango	325,32
1ºCuartil	126,615
3ºCuartil	229,365
Asimetría	1,26558
Curtosis Tipificada	1,73933

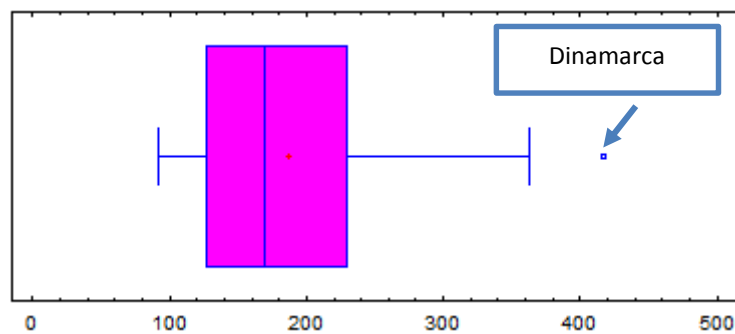


Ilustración 3. 7 Gráfico de caja y bigotes para la tasa impositiva (€/tonelada de petróleo) año 2013.

Como se observa en la (Ilustración 3. 7) hay otro punto anómalo o atípico que pertenece al país de Dinamarca, esto quiere decir, que la tasa impositiva es mucho más alta que el resto de países citados anteriormente de la Unión Europea.

La gran diferencia entre el máximo y el mínimo es obvia en todas las imágenes.

PGAS: Precios del gas por tipo de usuario (expresado en Euros per KWh) este indicador presenta los precios del gas natural que se cobran a los consumidores finales. Precios del gas natural para los consumidores industriales se define como: el precio nacional promedio en Euro por Giga Joule (GJ), sin impuestos aplicables para el primer semestre de cada año para el tamaño medio de los consumidores industriales y para el precio del gas natural para los consumidores domésticos se definen como el precio nacional promedio en Euro por GJ impuestos y gravámenes aplicables para el primer semestre de cada año para los medianos consumidores domésticos [17].

Se calcula la (Tabla 3.8) de resumen de estadísticos para los precios del gas por tipo de usuario que se expresa en Euro per KWh para los países citados anteriormente en el año 2013.

Tabla 3. 8 Resumen estadísticos para los precios del gas por tipo de usuario (€ per KWh) año 2013.

Estadístico	Valor
Media	18,4657
Mediana	18,235
Varianza	2,93E+01
Desviación Típica	5,411
Mínimo	7,9114
Máximo	34,088
Rango	26,1766
1ºCuartil	14,4049
3ºCuartil	21,38
Asimetría	0,905919
Curtosis Tipificada	2,07947

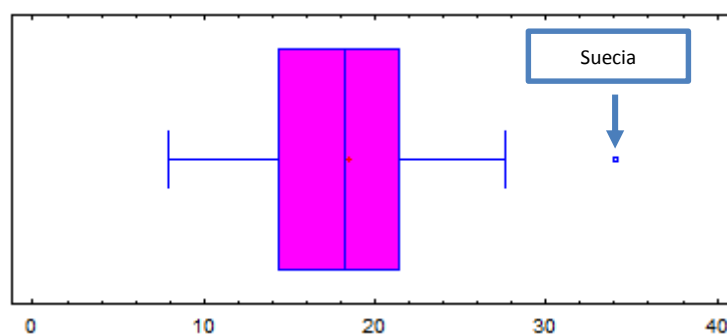


Ilustración 3. 8 Gráfico de caja y bigotes de los precios del gas por tipo de usuario (€ per KWh) año 2013.

Como se observa en la (Ilustración 3. 8) hay un punto anómalo o atípico perteneciente al país de Suecia, por lo cual quiere expresarnos que este país tiene el precio del gas por tipo de usuario más elevado que el resto de países citados anteriormente.

También se observa en la (Ilustración 3. 8) que la caja está más centrada entre el máximo y el mínimo. Y la mediana y la media son prácticamente iguales por lo tanto esta variable tampoco expresará problemas en el modelo de regresión.

PELEC: Precios de la electricidad por tipo de usuario expresado en Euros per KWh este indicador presenta precios de la electricidad se cobran a los consumidores finales. Precios de electricidad para los consumidores industriales se define como el precio nacional promedio en euros por KWh sin impuestos aplicables para el primer semestre de cada año para los medianos consumidores industriales. Para los precios de la electricidad para los consumidores domésticos se definen como el precio nacional promedio en euros por KWh incluyendo impuestos y gravámenes aplicables para el primer semestre de cada año para los medianos consumidores domésticos [18].

Se calcula la (Tabla 3.9) para el resumen de datos estadísticos de los precios de la electricidad por tipo de usuario expresado en Euros para cada uno de los países citados anteriormente en el año 2013.

Tabla 3. 9 Resumen de Estadísticos de los precios de la electricidad por tipo de usuario (€ por KWh) año 2013.

Estadístico	Valor
Media	0,181713
Mediana	0,16815
Varianza	2,56E-03
Desviación Típica	0,0505548
Mínimo	0,0924
Máximo	0,3
Rango	0,2076
1ºCuartil	0,1429
3ºCuartil	0,2137
Asimetría	0,759272
Curtosis Tipificada	0,445911

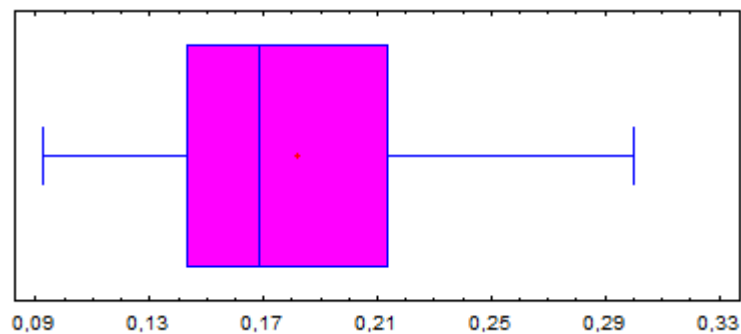


Ilustración 3. 9 Gráfico de cajas y bigotes de los precios de la electricidad por tipo de usuario (€ por KWh) año 2013.

Como se observa en la (Ilustración 3. 9) no existen puntos anómalos o atípicos, esto quiere decir que los precios de la electricidad por tipo de usuario se comportan dentro de lo normal.

La media y la mediana son muy parecidas también a la (Ilustración 3. 9), por lo cual esta variable tampoco expresará problema en el modelo de regresión.

POB: Población (expresada en millones) esta variable hace referencia al grupo formado por las personas que viven en un determinado lugar.

Ahora se realiza la (Tabla 3.10) número de población expresada en millones en el año 2013 para cada uno de los países citados anteriormente donde se realiza el resumen de estadísticos.

Tabla 3. 10 Resumen de estadísticos para el número de población (millones) año 2013.

Estadístico	Valor
Media	20429800
Mediana	10026500
Varianza	5,97E+14
Desviación Típica	24424100
Mínimo	543202
Máximo	80620000
Rango	80076800
1ºCuartil	4424000
3ºCuartil	29245000
Asimetría	1,38661
Curtosis Tipificada	0,540157

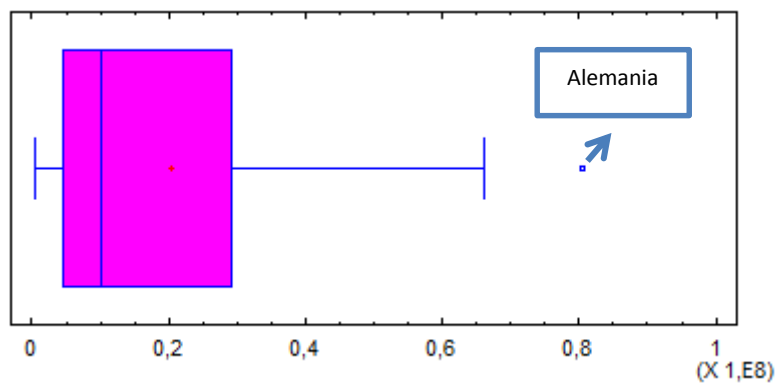


Ilustración 3. 10 Gráfico de cajas y bigotes de la población (millones) año 2013.

Como se observa en la (Ilustración 3. 10) hay un punto atípico o anómalo el cual corresponde a Alemania por su gran número de población por ser un país tan extenso.

Entre la media y mediana existe una gran diferencia por lo cual es probable que esta variable también de problemas en el modelo de regresión.

Nº EMP: Nº de Empresas por país, en esta variable se refleja el número de empresas que proporcionan los servicios de energía, ya sean productoras, proveedoras o que presten el servicio de cada uno de los tipos de energía.

Se realiza por última vez el cálculo de resumen de estadísticos (Tabla 3.11) número de empresas por cada uno de los países citados para el año 2013.

Tabla 3. 11 Resumen de Estadísticos para el número de empresas año 2013.

Estadístico	Valor
Media	618,5
Mediana	66,5
Varianza	1,53E+06
Desviación Típica	1236,16
Mínimo	7
Máximo	5550
Rango	5543
1ºCuartil	30,5
3ºCuartil	642
Asimetría	3,24193
Curtosis Tipificada	11,5046

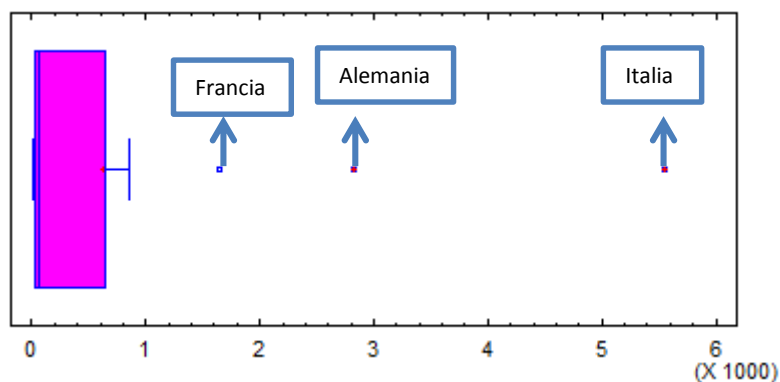


Ilustración 3. 11 Gráfico de cajas y bigotes para el número de empresas por país (unidades) año 2013.

Como se observa en la (**Ilustración 3. 11**) se aprecian tres puntos anómalos o atípicos los cuales pertenecen a Italia, Alemania, y Francia, esto explica que estos tres países resaltando a Italia, tienen más empresas que proporcionan el servicios de energía, ya sean productoras, proveedoras o que presten el servicio de cada uno de los tipos de energía.

Entre la media y la mediana también existe una diferencia muy grande por lo cual esta variable es probable que de problemas en el modelo de regresión.

En la (**Ilustración 3. 11**) apenas se observa dónde está la media y la mediana.

3.2 Bivariante

3.2.1 Gráficos X-Y

Se realiza la (Ilustración 3. 12) X-Y donde el eje de la “X” se representa el consumo de energía final por producto expresada en 1000 toneladas de petróleo equivalente y en el eje de la “Y” la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Donde también podremos observar su modelo correspondiente y el coeficiente de correlación también llamado “R”. Todo esto realizado en el año 2013 y para los países de la Unión Europea como Alemania, Austria, Bélgica, Bulgaria, Croacia, Dinamarca, Eslovaquia, Eslovenia, España, Estonia, Francia, Grecia, Irlanda, Italia, Letonia, Lituania, Luxemburgo, Países Bajos, Polonia, Portugal, Reino Unido, República Checa, Rumania y Suecia.

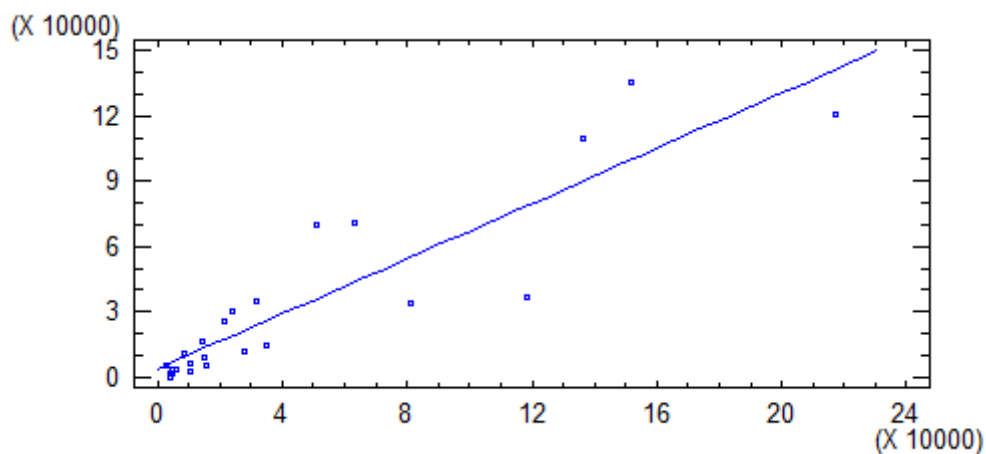


Ilustración 3. 12 Ilustración X-Y para el consumo de energía final (1000 toneladas por petróleo equivalente) año 2013.

En la (Ilustración 3. 12) se puede decir que existe relación directa.

$$\text{Producción Energía} = 3696,12 + 0,633507 * \text{COENG}$$

$$\text{Coeficiente de Correlación} = 0,901392$$

Se realiza la (Ilustración 3. 13) X-Y donde el eje de la “X” se representa el consumo de electricidad de los hogares expresada en 1000 toneladas de petróleo equivalente y en el eje de la “Y” la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Donde también podremos observar su modelo correspondiente y el coeficiente de correlación también llamado “R”. Todo esto realizado en el año 2013 y para los países de la Unión Europea citados anteriormente en la (Tabla 2. 1)

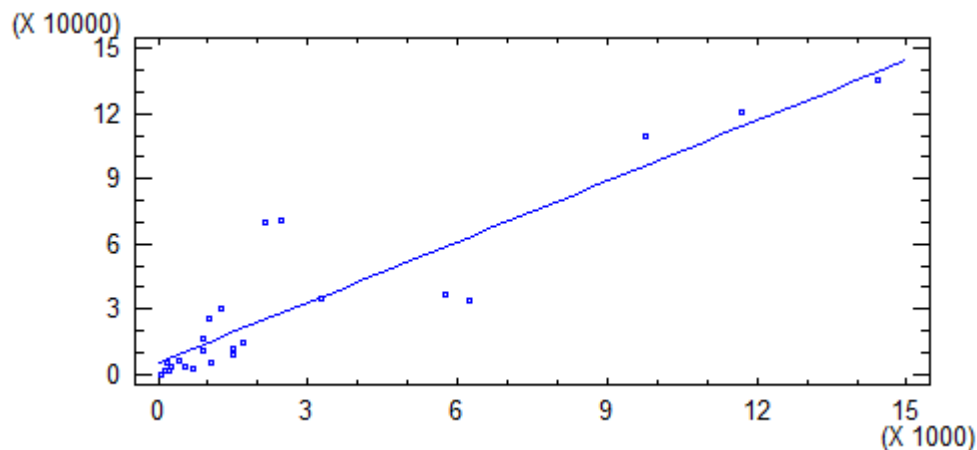


Ilustración 3. 13 Ilustración X-Y para el consumo de electricidad de los hogares (1000 toneladas de petróleo equivalente) año 2013.

En la (Ilustración 3. 13) se observa que la relación es directa porque tiende a ser positiva, a la vez que el Beta es positivo.

$$\text{Producción Energía} = 5344,59 + 9,28974 * \text{COELECHOG}$$

$$\text{Coeficiente de Correlación} = 0,912001$$

Se realiza la (Ilustración 3. 14) X-Y donde el eje de la “X” se representa el consumo final de energía en los hogares por combustible expresada en tanto por ciento del consumo final y en el eje de la “Y” la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Donde también podremos observar su modelo correspondiente y el coeficiente de correlación también llamado “R”. Todo esto realizado en el año 2013 y para los países de la Unión Europea citados anteriormente en la (Tabla 2. 1).

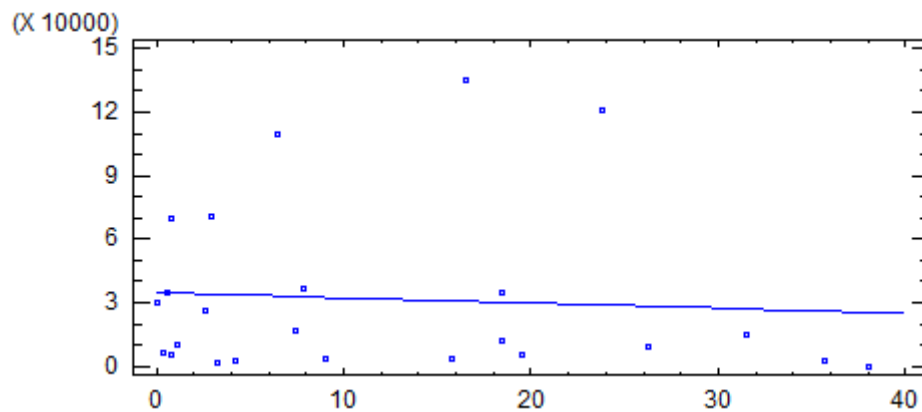


Ilustración 3. 14 Ilustración X-Y para el consumo final de energía en los hogares por combustible (%) año 2013.

Como se observa en la (Ilustración 3. 14) existe una relación indirecta prácticamente inexistente.

$$\text{Producción Energía} = 34659,3 - 241,058\text{COENGHOGCOMB}$$

$$\text{Coeficiente de Correlación} = -0,0721947$$

Se realiza la (Ilustración 3. 15) X-Y donde el eje de la “X” se representa el consumo de energía del transporte en relación con el precio interior bruto expresada en volúmenes encadenados a tipos de cambio de 2000 y en el eje de la “Y” la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Donde también podremos observar su modelo correspondiente y el coeficiente de correlación también llamado “R”. Todo esto realizado en el año 2013 y para los países de la Unión Europea citados anteriormente en la (Tabla 2. 1).

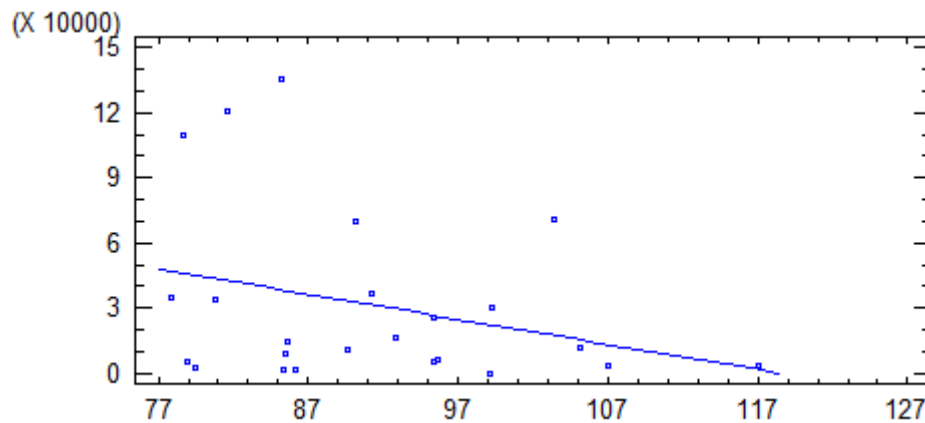


Ilustración 3. 15 Ilustración X-Y para el consumo de energía del transporte en relación con el precio interior bruto (volúmenes encadenados a tipos de cambio de 2000) año 2013.

Como se observa en la (Ilustración 3. 15) existe una relación indirecta muy pequeña.

$$\text{Producción Energía} = 136936,0 - 1154,89 * \text{COENGTRANSP}$$

$$\text{Coeficiente de Correlación} = -0,297423$$

Se realiza la (Ilustración 3. 16) X-Y donde el eje de la “X” representa la participación de las energías renovables en el consumo de combustible de transporte expresada en tanto por ciento y en el eje de la “Y” la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Donde también podremos observar su modelo correspondiente y el coeficiente de correlación también llamado “R”. Todo esto realizado en el año 2013 y para los países de la Unión Europea citados anteriormente en la (Tabla 2. 1).

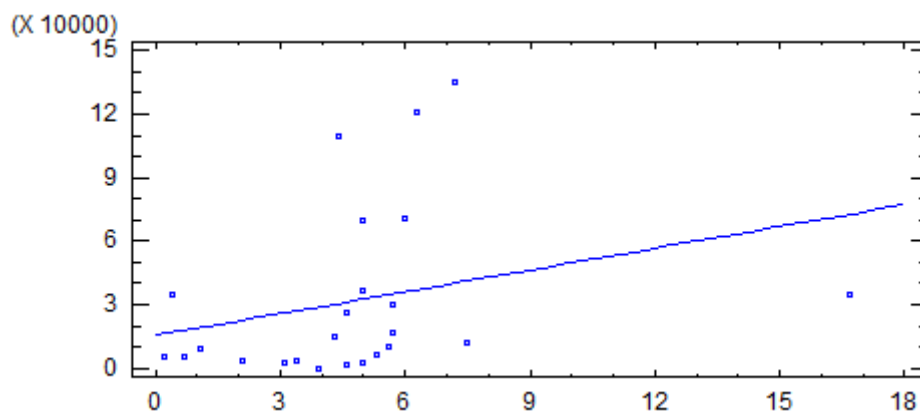


Ilustración 3. 16 Ilustración X-Y para la participación de las energías renovables en el consumo de combustible de transporte (%) año 2013.

Como se observa en la (Ilustración 3. 16) existe una relación directa bastante pequeña.

$$\text{Producción Energía} = 15409,2 + 3442,1 * \text{PARTENGRE}$$

$$\text{Coeficiente de Correlación} = 0,281263$$

Se realiza la (Ilustración 3. 17) X-Y donde el eje de la “X” se representa la tasa impositiva expresada en Euros por tonelada de petróleo y en el eje de la “Y” la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Donde también podremos observar su modelo correspondiente y el coeficiente de correlación también llamado “R”. Todo esto realizado en el año 2013 y para los países de la Unión Europea citados anteriormente en la (Tabla 2. 1).

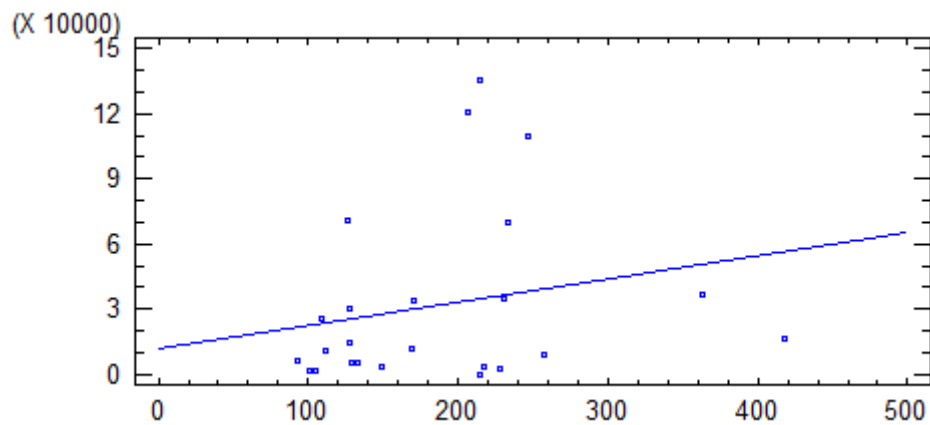


Ilustración 3. 17 Ilustración X-Y para la tasa impositiva (€/tonelada de petróleo) año 2013.

Como se observa en la (Ilustración 3. 17) existe una relación directa relativamente pequeña.

$$\text{Producción Energía} = 11903,5 + 106,271 * \text{TAIMP}$$

$$\text{Coeficiente de Correlación} = 0,218508$$

Se realiza la (Ilustración 3. 18) X-Y donde el eje de la “X” se representa por los precios del gas por tipo de usuario expresado en Euros y en el eje de la “Y” la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Donde también podremos observar su modelo correspondiente y el coeficiente de correlación también llamado “R”. Todo esto realizado en el año 2013 y para los países de la Unión Europea citados anteriormente en la (Tabla 2. 1).

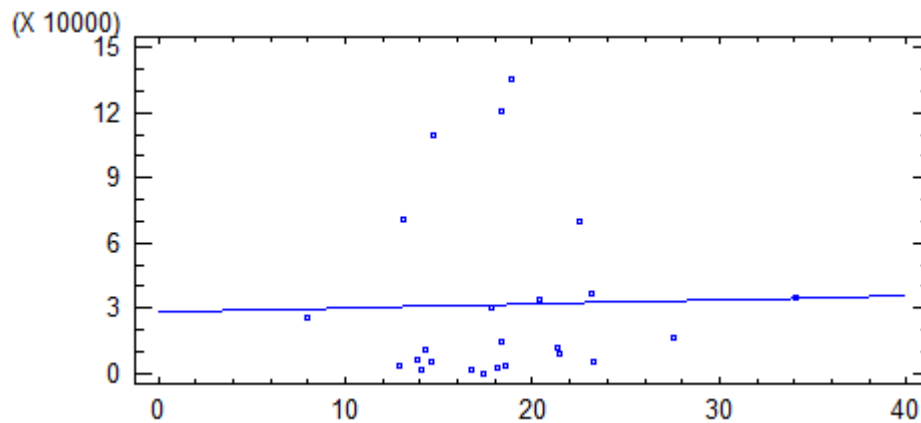


Ilustración 3. 18 Ilustración X-Y para los precios del gas por tipo de usuario (€) año 2013.

Como se observa en la (Ilustración 3. 18) se aprecia una relación directa casi inexistente.

$$\text{Producción Energía} = 28301,3 + 185,701 * PGAS$$

$$\text{Coeficiente de Correlación} = 0,0251388$$

Se realiza la (Ilustración 3. 19) X-Y donde el eje de la “X” se representa por los precios de la electricidad por tipo de usuario expresado en Euros y en el eje de la “Y” la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Donde también podremos observar su modelo correspondiente y el coeficiente de correlación también llamado “R”. Todo esto realizado en el año 2013 y para los países de la Unión Europea citados anteriormente en la (Tabla 2. 1).

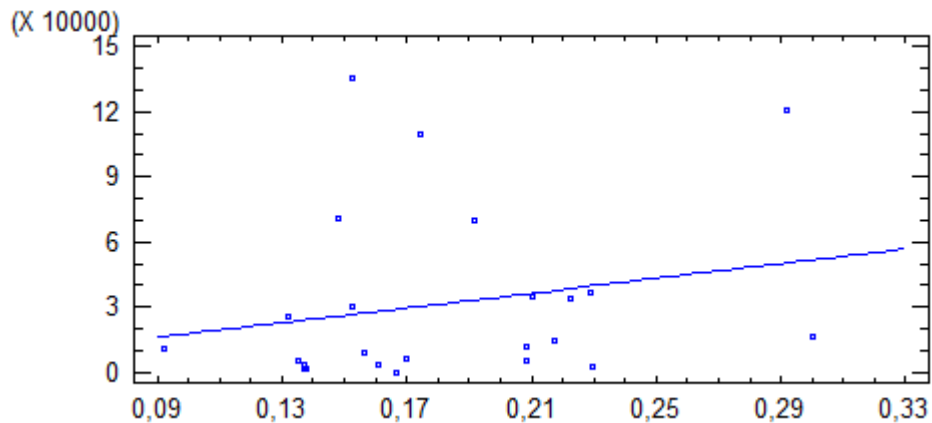


Ilustración 3. 19 Ilustración X-Y para los precios de la electricidad por tipo de usuario (€) | año 2013.

Como se observa en la (Ilustración 3. 19) existe una leve relación directa.

$$\text{Producción Energía} = 1009,59 + 169063,0 * \text{PELEC}$$

$$\text{Coeficiente de Correlación} = 0,213828$$

Se realiza la (**Ilustración 3. 20**) X-Y donde el eje de la “X” se representa el número de población expresada en millones por cada país de la Unión Europea como Alemania, Austria, Bélgica, Bulgaria, Croacia, Dinamarca, Eslovaquia, Eslovenia, España, Estonia, Francia, Grecia, Irlanda, Italia, Letonia, Lituania, Luxemburgo, Países Bajos, Polonia, Portugal, Reino Unido, República Checa, Rumania y Suecia y en el eje de la “Y” la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Para el año 2013.

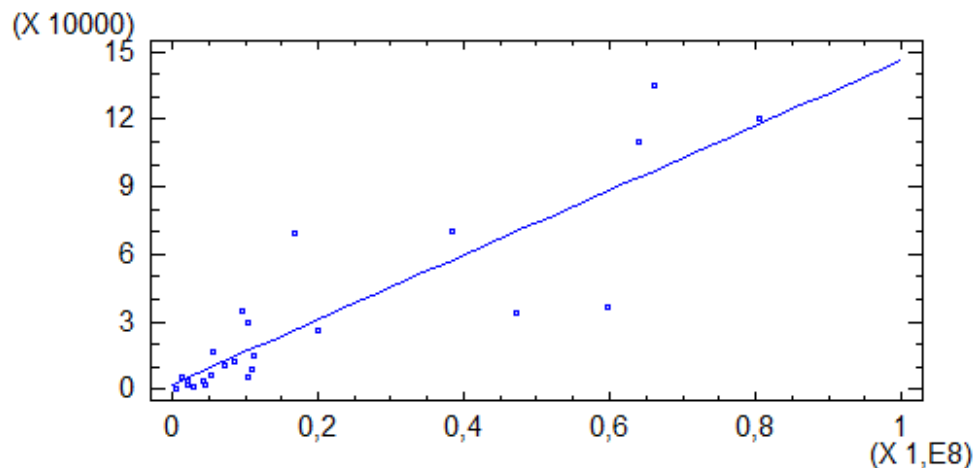


Ilustración 3. 20 Ilustración X-Y número de población (millones) para cada país de la U.E citado anteriormente.

Como se observa en la (Ilustración 3. 20) existe una relación directa acusada.

$$\text{Producción Energía} = 2444,19 + 0,0014335 * \text{POB}$$

$$\text{Coeficiente de Correlación} = 0,875934$$

Se realiza la (Ilustración 3. 21) X-Y donde el eje de la "X" se representa el número de empresas expresada en unidades por cada país de la Unión Europea como Alemania, Austria, Bélgica, Bulgaria, Croacia, Dinamarca, Eslovaquia, Eslovenia, España, Estonia, Francia, Grecia, Irlanda, Italia, Letonia, Lituania, Luxemburgo, Países Bajos, Polonia, Portugal, Reino Unido, República Checa, Rumania y Suecia y en el eje de la "Y" la producción primaria por parte de los recursos expresada en 1000 toneladas de petróleo equivalente. Para el año 2013.

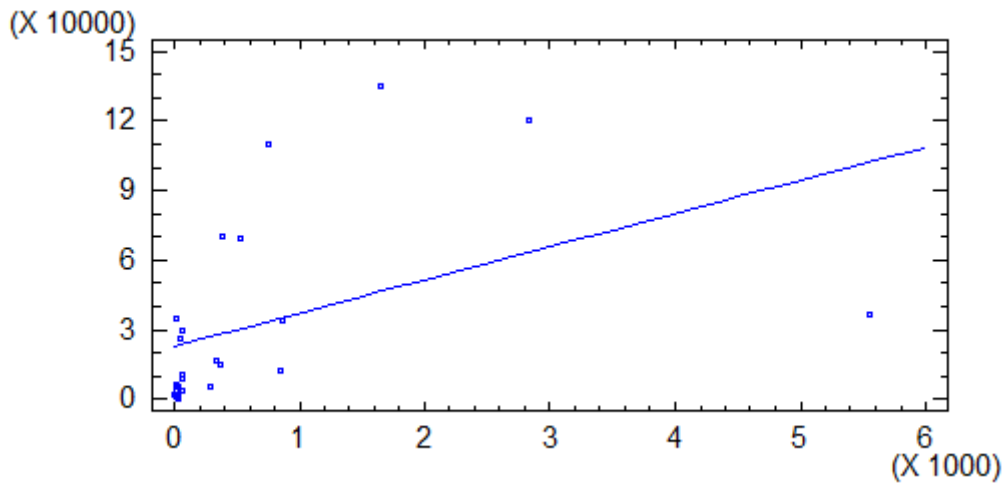


Ilustración 3. 21 Ilustración X-Y para el número de empresas (unidades) año 2013.

La (Ilustración 3. 21) tiene una relación directa.

$$\text{Producción Energía} = 22891,5 + 14,2909 * N^{\circ}EMP$$

$$\text{Coeficiente de Correlación} = 0,441965$$

El modelo teórico propuesto es el siguiente:

$$\begin{aligned}
 \text{PRODUCCIÓN ENERGÍA} = & \beta_0 + \beta_1 \text{COENG} + \beta_2 \text{COELECHOG} + \\
 & + \beta_3 \text{COENGHOGCOM} + \beta_4 \text{COENGTRANSP} + \beta_5 \text{PARTENGRE} + \\
 & + \beta_6 \text{TAIMP} + \beta_7 \text{PGAS} + \beta_8 \text{PELEC} + \beta_9 \text{POB} + \beta_{10} \text{N}^\circ \text{EMP} + U
 \end{aligned}$$

β_0 = El valor medio de producción de energía cuando todas las variables explicativas son 0.

β_1 = Coeficiente que acompaña a la variable COENG. El incremento medio de producción de energía cuando aumenta en una unidad el COENG y el resto de las variables explicativas permanecen constantes.

β_2 = Coeficiente que acompaña a la variable COELECHOG. El incremento de producción de energía cuando aumenta en una unidad el COELECHOG y el resto de las variables explicativas permanecen constantes.

β_3 = Coeficiente que acompaña a la variable COENGHOGCOM. El incremento de producción de energía cuando aumenta en una unidad el COENGHOGCOM y el resto de las variables explicativas permanecen constantes.

β_4 = Coeficiente que acompaña a la variable COENGTRANSP. El incremento de producción de energía cuando aumenta en una unidad el COENGTRANSP y el resto de las variables explicativas permanecen constantes.

β_5 = Coeficiente que acompaña a la variable PARTENGRENOV. El incremento de producción de energía cuando aumenta en una unidad la PARTENGRENOV y el resto de variables explicativas permanecen constantes.

β_6 = Coeficiente que acompaña a la variable TAIMP. El incremento de producción de energía cuando aumenta en una unidad la TAIMP y el resto de variables explicativas permanecen constantes.

β_7 = Coeficiente que acompaña a la variable PGAS. El incremento de producción de energía cuando aumenta en una unidad el PGAS y el resto de variables explicativas permanecen constantes.

β_8 = Coeficiente que acompaña a la variable PELEC. El incremento de producción de energía cuando aumenta en una unidad el PELEC y el resto de variables explicativas permanecen constantes.

β_9 = Coeficiente que acompaña a la variable POB. El incremento de producción de energía cuando aumenta en una unidad la POB y el resto de variables explicativas permanecen constantes.

β_{10} = Coeficiente que acompaña a la variable N° EMP. El incremento de producción de energía cuando aumenta en una unidad el N° EMP y el resto de variables explicativas permanecen constantes.

U= perturbación, la parte de los datos que no somos capaces de explicar con el modelo.

3.2.1 Multicolinealidad

Se realizará la (Tabla 3.12) que se calcula la matriz de correlación para saber si existe problema de multicolinealidad.

Tabla 3. 12 Matriz del tipo de una variable frente a otra.

	COENG	COELECHOG	COENGHOGCOM	COENGTRANSP	PARTENRENOV	TAIMP	PGAS	PELEC	POB	Nº EMP
COENG	1	0,94	0,10	-0,34	0,18	0,30	0,07	0,41	0,97	0,71
COELECHOG	0,94	1	0,10	-0,40	0,21	0,28	0,11	0,30	0,93	0,58
COENGHOGCOM	0,10	0,10	1	-0,04	-0,22	0,19	0,11	0,38	0,05	0,06
COENGTRANSP	-0,34	-0,40	-0,04	1	-0,12	-0,11	-0,21	-0,22	-0,32	-0,13
PARTENRENOV	0,18	0,21	-0,22	-0,12	1	0,18	0,49	0,20	0,10	0,08
TAIMP	0,30	0,28	0,19	-0,11	0,18	1	0,62	0,63	0,28	0,48
PGAS	0,07	0,11	0,11	-0,21	0,49	0,62	1	0,61	0,00	0,20
PELEC	0,41	0,30	0,38	-0,22	0,20	0,63	0,61	1	0,33	0,43
POB	0,97	0,93	0,05	-0,32	0,10	0,28	0,00	0,33	1	0,72
Nº EMP	0,71	0,58	0,06	-0,13	0,08	0,48	0,20	0,43	0,72	1

Como se observa en la (Tabla 3.12) la matriz de correlación tiene una elevada directa relación entre la variable consumo eléctrico hogar y la variable consumo de energía. También existe entre la variable población y consumo de energía, entre la variable población y consumo eléctrico hogar, entre la variable número de empresas y consumo de energía y por último entre la variable número de empresas y población ya que entre ellas también existe una elevada directa relación porque superan el límite de $|R_{ii}| \geq 0,7$.

Por lo tanto, concluimos que existe problema de multicolinealidad.

A continuación se realiza la matriz inversa en la (Tabla 3. 13) para saber si existe problema de multicolinealidad.

Tabla 3. 13 Matriz inversa del tipo de una variable frente al resto

COENG	42,61	-16,20	0,09	-0,26	-2,37	2,79	3,16	-7,23	-22,09	-3,64
COELECHOG	-16,20	18,78	-1,58	0,86	-0,18	-1,38	-3,36	4,96	-5,27	3,83
COENGHOGCOM	0,09	-1,58	1,52	-0,11	0,53	0,04	0,30	-0,96	1,62	-0,13
COENGTRANSP	-0,26	0,86	-0,11	1,30	-0,15	-0,24	0,33	0,24	0,02	-0,19
PARTENRENOV	-2,37	-0,18	0,53	-0,15	1,80	0,17	-1,10	0,27	1,92	0,22
TAIMP	2,79	-1,38	0,04	-0,24	0,17	2,51	-0,91	-1,07	-1,24	-0,90
PGAS	3,16	-3,36	0,30	0,33	-1,10	-0,91	3,69	-2,02	1,85	-0,94
PELEC	-7,23	4,96	-0,96	0,24	0,27	-1,07	-2,02	4,13	0,87	0,85
POB	-22,09	-5,27	1,62	0,02	1,92	-1,24	1,85	0,87	29,05	-2,58
Nº EMP	-3,64	3,83	-0,13	-0,19	0,22	-0,90	-0,94	0,85	-2,58	3,44

Como se observa en la (Tabla 3. 13) las variables consumo de energía, consumo eléctrico hogar y población tienen una elevada relación con el resto de variables ya que superan el valor límite $R_{ii}^{-1} \geq 10$.

Por lo tanto, concluimos que existe problema de multicolinealidad.

Por último se realiza el índice de acondicionamiento donde se observa la relación de todas las variables X a la vez, mediante la (Tabla 3. 14) llamada autovalores podremos sacar el índice de acondicionamiento para saber si existe problema de multicolinealidad.

Tabla 3. 14 Tabla autovalores.

	AUTOVALORES
Consumo Energía	4,21
Con Elec Hog	1,96
Con Ene Hog	1,32
Con Ene trans/PIB	0,96
Part Ene Renov	0,64
Tasa Impositiva	0,34
Precio Gas	0,31
Precio Electricidad	0,22
Población	0,03
Nº Empresas	0,02

$$I.C = \sqrt{\frac{\text{autovalormáximo}}{\text{autovalormínimo}}} = \sqrt{\frac{4,21}{0,02}} = 16,30$$

Como se observa existe problema de multicolinealidad moderada ya que el I.C está entre 10 y 20.

En conclusión, una vez realizado el análisis de la matriz de correlación, de la matriz inversa, y del índice de acondicionamiento, se ha de asumir que existe este problema. Se procede a solucionar el problema mediante la eliminación de variables explicativas.

Para eliminar una variable se debe observar la diagonal de la matriz inversa los valores mayores de 10 serían los candidatos a eliminar en este caso hay tres, pero existe un valor más grande y la diferencia entre esa variable y las otras dos es mayor a una unidad. Por lo tanto, se elimina la variable COENG.

MODELO TEORICO PROPUESTO AL ELIMINAR LA VARIABLE COENG

$$\begin{aligned} \text{PRODUCCIÓN ENERGÍA} &= \beta_0 + \beta_1 \text{COELECHOG} + \\ &+ \beta_2 \text{COENGHOGCOM} + \beta_3 \text{COENGTRANSP} + \beta_4 \text{PARTENGRE} + \\ &+ \beta_5 \text{TAIMP} + \beta_6 \text{PGAS} + \beta_7 \text{PELEC} + \beta_8 \text{POB} + \beta_9 \text{Nº EMP} + U \end{aligned}$$

Se realiza de nuevo una matriz de correlación que se muestra en la (Tabla 3. 15) para saber si existe multicolinealidad.

Tabla 3. 15 Matriz del tipo de una variable frente a otra.

	COELECHOG	COENHOGCOM	COENGRANSP	PARTENRENOV	TAIMP	PGAS	PELEC	POB	Nº EMP
COELECHOG	1	0,10	-0,40	0,21	0,28	0,11	0,30	0,93	0,58
COENHOGCOM	0,10	1	-0,04	-0,22	0,19	0,11	0,38	0,05	0,06
COENGRANSP	-0,40	-0,04	1	-0,12	-0,11	-0,21	-0,22	-0,32	-0,13
PARTENRENOV	0,21	-0,22	-0,12	1	0,18	0,49	0,20	0,10	0,08
TAIMP	0,28	0,19	-0,11	0,18	1	0,62	0,63	0,28	0,48
PGAS	0,11	0,11	-0,21	0,49	0,62	1	0,61	0,00	0,20
PELEC	0,30	0,38	-0,22	0,20	0,63	0,61	1	0,33	0,43
POB	0,93	0,05	-0,32	0,10	0,28	0,00	0,33	1	0,72
Nº EMP	0,58	0,06	-0,13	0,08	0,48	0,20	0,43	0,72	1

Como se observa en la (Tabla 3. 15) la matriz de correlación tiene una elevada directa relación entre la variable población y consumo eléctrico hogar. También existe entre la variable población y el número de empresas porque superan el límite de $|R_{ii}|=0,7/$.

Por lo tanto, se puede decir que existe problema de multicolinealidad.

A continuación volvemos a estudiar la matriz inversa de correlación en la (Tabla 3. 16) para saber si existe problema de multicolinealidad.

Tabla 3. 16 Matriz inversa del tipo de una variable frente a otra.

COELECHOG	12,62	-1,55	0,76	-1,08	-0,32	-2,16	2,20	-13,68	2,45
COENHOGCOM	-1,55	1,52	-0,11	0,53	0,04	0,29	-0,94	1,67	-0,13
COENGRANSP	0,76	-0,11	1,30	-0,17	-0,22	0,35	0,19	-0,12	-0,21
PARTENRENOV	-1,08	0,53	-0,17	1,67	0,33	-0,92	-0,13	0,69	0,02
TAIMP	-0,32	0,04	-0,22	0,33	2,33	-1,12	-0,60	0,21	-0,66
PGAS	-2,16	0,29	0,35	-0,92	-1,12	3,46	-1,48	3,49	-0,67
PELEC	2,20	-0,94	0,19	-0,13	-0,60	-1,48	2,90	-2,88	0,23
POB	-13,68	1,67	-0,12	0,69	0,21	3,49	-2,88	17,60	-4,47
Nº EMP	2,45	-0,13	-0,21	0,02	-0,66	-0,67	0,23	-4,47	3,12

Como se observa en la (Tabla 3. 16) la variable consumo de energía tiene una elevada relación con el resto de variables ya que la variable población superan el valor límite $R_{ii}^{-1}=10$.

Por lo tanto, existe problema de multicolinealidad.

Por último se realiza el índice de acondicionamiento donde se observa la relación de todas las variables X a la misma vez, mediante la (Tabla 3. 17) llamada autovalores podremos sacar el índice de acondicionamiento para saber si existe problema de multicolinealidad.

Tabla 3. 17 Tabla autovalores.

	AUTOVALORES
Cons Elec Hog	3,47
Cons Ene Hog	1,75
Cons Ene transp/PIB	1,32
Part Ene Renov	0,95
Tasa Impositiva	0,63
Precio Gas	0,34
Precio Electricidad	0,29
Población	0,22
Nº Empresas	0,03

$$I.C = \sqrt{\frac{\text{autovalormáximo}}{\text{autovalormínimo}}} = \sqrt{\frac{3,47}{0,03}} = 10,43$$

Como se observa existe problema de multicolinealidad moderada ya que el valor está entre 10 y 20.

En conclusión una vez realizado el análisis de la matriz de correlación, de la matriz inversa, y del índice de acondicionamiento, se ha de asumir existe este problema. Se procede a solucionar el problema mediante la eliminación de una variable explicativa.

Para eliminar una variable se debe observar la diagonal de la matriz inversa los valores mayores de 10 serían los candidatos a eliminar en este caso hay dos, pero existe un valor más grande que es la variable POB y la diferencia entre la variable POB y COELECHOG es mayor a una unidad. Por tanto, se elimina la variable POB.

MODELO TEORICO PROPUESTO AL ELIMINAR LA VARIABLE POB

$$\begin{aligned} \text{PRODUCCIÓN ENERGÍA} = & \beta_0 + \beta_1 \text{COELECHOG} + \\ & + \beta_2 \text{COENGHOOGCOM} + \beta_3 \text{COENGTRANSP} + \beta_4 \text{PARTENGRE} + \\ & + \beta_5 \text{TAIMP} + \beta_6 \text{PGAS} + \beta_7 \text{PELEC} + \beta_8 \text{Nº EMP} + U \end{aligned}$$

Se realiza de nuevo una matriz de correlación que se muestra en la (Tabla 3. 18) para saber si existe multicolinealidad.

Tabla 3. 18 Matriz del tipo de una variable frente a otra.

	COELECHOG	COENGHOGCOM	COENGTRANSP	PARTENGRENOV	TAIMP	PGAS	PELEC	Nº EMP
COELECHOG	1	0,10	-0,40	0,21	0,28	0,11	0,30	0,58
COENGHOGCOM	0,10	1	-0,04	-0,22	0,19	0,11	0,38	0,06
COENGTRANSP	-0,40	-0,04	1	-0,12	-0,11	-0,21	-0,22	-0,13
PARTENGRENOV	0,21	-0,22	-0,12	1	0,18	0,49	0,20	0,08
TAIMP	0,28	0,19	-0,11	0,18	1	0,62	0,63	0,48
PGAS	0,11	0,11	-0,21	0,49	0,62	1	0,61	0,20
PELEC	0,30	0,38	-0,22	0,20	0,63	0,61	1	0,43
Nº EMP	0,58	0,06	-0,13	0,08	0,48	0,20	0,43	1

Siguiendo el mismo criterio que en el análisis anterior, no existe multicolinealidad del tipo de relación de variables, ya que ninguna variable supera $|R_{ii}|=0,7$.

Por lo tanto, se concluye que no existe problema de multicolinealidad.

A continuación volvemos a estudiar la matriz inversa de correlación en la (Tabla 3. 19) para saber si existe problema de multicolinealidad.

Tabla 3. 19 Matriz inversa del tipo de una variable frente al resto.

COELECHOG	1,99	-0,25	0,67	-0,54	-0,16	0,55	-0,04	-1,02
COENGHOGCOM	-0,25	1,36	-0,10	0,47	0,02	-0,04	-0,67	0,30
COENGTRANSP	0,67	-0,10	1,30	-0,16	-0,22	0,37	0,17	-0,24
PARTENGRENOV	-0,54	0,47	-0,16	1,64	0,32	-1,06	-0,01	0,20
TAIMP	-0,16	0,02	-0,22	0,32	2,33	-1,16	-0,56	-0,61
PGAS	0,55	-0,04	0,37	-1,06	-1,16	2,77	-0,91	0,21
PELEC	-0,04	-0,67	0,17	-0,01	-0,56	-0,91	2,43	-0,50
Nº EMP	-1,02	0,30	-0,24	0,20	-0,61	0,21	-0,50	1,99

Siguiendo el mismo criterio que en el análisis anterior, no existe multicolinealidad del tipo relación de una variable con el resto.

Se concluye que no existe multicolinealidad.

Por último se realiza el índice de acondicionamiento donde se observa la relación de todas las variables X a la misma vez, mediante la (Tabla 3. 20) llamada autovalores podremos sacar el índice de acondicionamiento para saber si existe problema de multicolinealidad.

Tabla 3. 20 Tabla autovalores.

	AUTOVALORES
Cons Elec Hog	3,03
Cons Ene Hog	1,32
Cons Ene transp/PIB	1,27
Part Ene Renov	0,94
Tasa Impositiva	0,61
Precio Gas	0,34
Precio Electricidad	0,27
NºEmpresas	0,22

$$I.C = \sqrt{\frac{\text{autovalormáximo}}{\text{autovalormínimo}}} = \sqrt{\frac{3,03}{0,22}} = 3,72$$

Siguiendo el mismo criterio que en el análisis anterior, no existe multicolinealidad del tipo todas las variables entre todas.

Por lo tanto, se concluye que este modelo está libre de multicolinealidad.

3.3 Multivariante

$$\begin{aligned}
 \text{PRODUCCIÓN ENERGÍA} = & \beta_0 + \beta_1 \text{COELECHOG} + \\
 & + \beta_2 \text{COENGHOGCOM} + \beta_3 \text{COENGTRANSP} + \beta_4 \text{PARTENGRE} + \\
 & + \beta_5 \text{TAIMP} + \beta_6 \text{PGAS} + \beta_7 \text{PELEC} + \beta_8 \text{N}^\circ \text{EMP} + U
 \end{aligned}$$

3.3.1 Ajuste del Modelo

A continuación, se muestra los resultados del ajuste a un modelo de regresión lineal múltiple para describir la relación entre la variable producción y las 8 variables independientes.

$$\begin{aligned}
 \text{PRODUCCIÓN ENERGÍA} = & -186147 + 10,2181 \text{COELECHOG} - 539,329 \text{COENGHOGCOMB} + \\
 & + 296,186 \text{COENGTRANSP} + 1342,66 \text{PARTENGRE} + 46,7564 \text{TAIMP} - 1421,77 \text{PGAS} + \\
 & + 90006,5 \text{PELEC} - 5,887 \text{N}^\circ \text{EMP}
 \end{aligned}$$

R-cuadrado= 89,59%

3.3.2 Comprobación de Hipótesis

3.3.2.1 Significación

Por lo que se va a comprobar la hipótesis sobre el modelo en este caso existen dos métodos, pero se utilizará el segundo método del P-valor.

Tabla 3. 21 Contraste del modelo, coeficiente que acompaña a la variable explicativa.

Fuente	Coficiente-F	P-valor
Modelo	16,14	0,000

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$$

$$H_1 : \text{Algun } \beta_i \neq 0$$

P-valor (0,0000) < $\alpha=0,05$

Puesto que el P-valor es menor a α , se rechaza la hipótesis nula. Por lo que se concluye que el modelo es significativo.

A continuación, se comprobará la hipótesis sobre todos los parámetros del modelo también utilizando el mediante contraste de hipótesis.

Tabla 3. 22 Parámetros del modelo.

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	-18614,7	40230,9	-0,462696	0,6502
COELECHOG	10,2181	1,19806	8,52883	0,000
COENGHOGCOMB	-593,329	324,698	-1,82732	0,0876
COENGTRANSP	296,186	369,068	0,802526	0,4348
PARTENRENOV	1342,66	1306,39	1,02776	0,3204
TAIMP	46,7564	61,8162	0,756378	0,4611
PGAS	-1421,77	1024,19	-1,38819	0,1854
PELEC	90006,5	102630	0,876996	0,3943
Nº EMP	-5,887	3,80013	-1,54916	0,1422

Contraste β_0

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

P-valor (0'6502) > $\alpha=0,05$

Por tanto, se acepta la hipótesis nula. Se concluye que la constante no resulta significativa para el modelo.

Contraste $\beta_1=COELECHOG$

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

P-valor (0,0000) < $\alpha=0,05$

Por tanto, se rechaza la hipótesis nula. Se concluye que la variable COELECHOG sí que es significativa para el modelo.

Contraste $\beta_2=COENGHOGCOMB$

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

P-valor (0,0876) $>\alpha=0,05$

Por tanto, se acepta la hipótesis nula. Se concluye que la variable COENGHOGCOMB no resulta significativa para el modelo.

Contraste $\beta_3=COENGTRANS$

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

P-valor (0,4348) $>\alpha=0,05$

Por tanto, se acepta la hipótesis nula. Se concluye que la variable COENGTRANS no resulta significativa para el modelo.

Contraste $\beta_4=PARTENGRENOV$

$$H_0 : \beta_4 = 0$$

$$H_1 : \beta_4 \neq 0$$

P-valor (0,3204) $>\alpha=0,05$

Por tanto, se acepta la hipótesis nula. Se concluye que la variable PARTENGRENOV no resulta significativa para el modelo.

Contraste $\beta_5=TAIMP$

$$H_0 : \beta_5 = 0$$

$$H_1 : \beta_5 \neq 0$$

P-valor (0,4611) $>\alpha=0,05$

Por tanto, se acepta la hipótesis nula. Se concluye que la variable TAIMP no resulta significativa para el modelo.

Contraste $\beta_6=PGAS$

$$H_0 : \beta_6 = 0$$

$$H_1 : \beta_6 \neq 0$$

P-valor (0,1854) $>\alpha=0,05$

Por tanto, se acepta la hipótesis nula. Se concluye que la variable PGAS no resulta significativa para el modelo.

Contraste $\beta_7=PELEC$

$$H_0 : \beta_7 = 0$$

$$H_1 : \beta_7 \neq 0$$

P-valor (0,3943) $>\alpha=0,05$

Por tanto, se acepta la hipótesis nula. Se concluye que la variable PELEC no resulta significativa para el modelo.

Contraste $\beta_8=N^{\circ} EMP$

$$H_0 : \beta_8 = 0$$

$$H_1 : \beta_8 \neq 0$$

P-valor (0,1422) $>\alpha=0,05$

Por tanto, se acepta la hipótesis nula. Se concluye que la variable N^oEMP no resulta significativa para el modelo.

3.3.2.2 Normalidad de Residuos

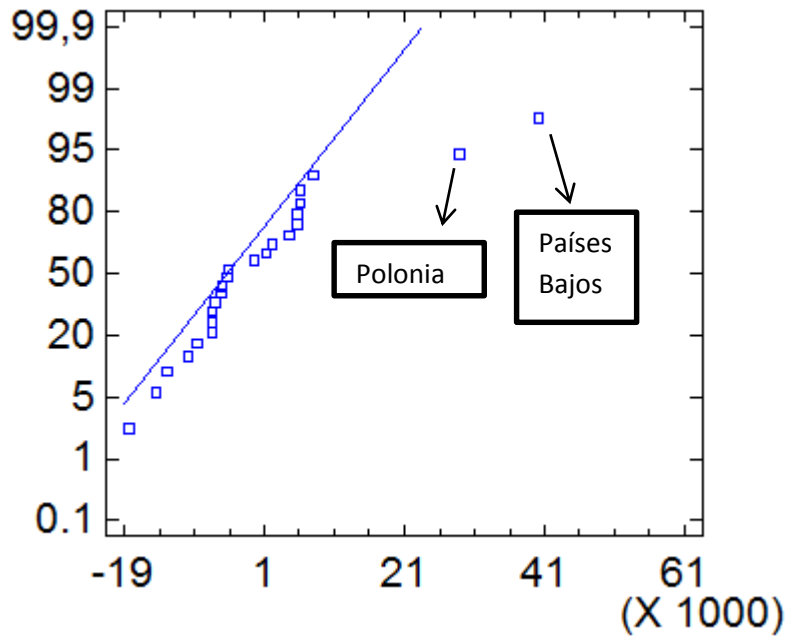


Ilustración 3. 22 Ilustración probabilístico de residuos.

Como se observa en la (Ilustración 3. 22) la mayoría de los puntos están sobre la línea de normalidad o bastante cerca se considera normalidad de residuos, aunque existen dos puntos que se alejan bastante que podrían ser puntos influyentes o atípicos que pertenecen a Polonia y los Países Bajos.

HISTOGRAMA

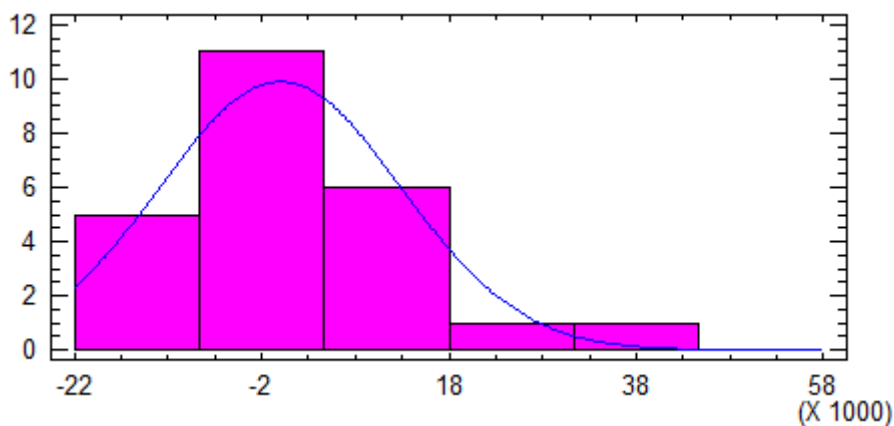


Ilustración 3. 23 Histograma para residuos.

Como se muestra en la (Ilustración 3. 23) el histograma no toma una forma semejante a una campana de Gauss, por lo tanto se podría asumir que no tienen un comportamiento de normalidad de residuos.

También se comprobará mediante el test de hipótesis.

H_0 : Los residuos son normales. } $\alpha=0,05$

H_1 : Los residuos no son normales. }

Tests para la Normalidad para RESIDUALS

Estadístico chi-cuadrado de bondad de ajuste = 20,3333

P-valor = 0,0409735

Estadístico W de Shapiro-Wilks = 0,844163

P-valor = 0,00127627

Puntuación Z para asimetría = 2,22307

P-valor = 0,0262112

Puntuación Z para curtosis = 2,58954

P-valor = 0,00961043



Se elige el test
más restrictivo.

P-valor (0,00127627) $< \alpha=0,05$

El test más restrictivo ha dado 0,00127627, por lo tanto, rechazamos la hipótesis nula, por lo tanto, los residuos no se distribuyen normalmente. Casi seguro que los residuos no son normales debido a las observaciones de Polonia y Países bajos, se comprobará más adelante.

3.3.2.3 Heteroscedasticidad

Se realiza un contraste de hipótesis para saber si existe heteroscedasticidad.

$H_0 : \sigma^2 = 0 \rightarrow$ No existe heteroscedasticidad

$H_1 : \sigma^2 \neq 0 \rightarrow$ Existe heteroscedasticidad

Se plantea un modelo nuevo.

$$\begin{aligned}
 \text{RESIDUOS}^2 = & \beta_0 + \beta_1 \text{COELECHOG} + \beta_2 \text{COENGHOGCOM} + \\
 & + \beta_3 \text{COENGTRANSP} + \beta_4 \text{PARTENGRE} + \beta_5 \text{TAIMP} + \\
 & + \beta_6 \text{PGAS} + \beta_7 \text{PELEC} + \beta_8 \text{N}^\circ \text{EMP} + U
 \end{aligned}$$

Para saber si existe heteroscedasticidad o no, nos fijamos en los parámetros de las variables explicativas. A continuación se muestra la (Tabla 3. 23) con los p-valor de cada una de las variables.

Tabla 3. 23 Tabla donde se muestra el P-valor de cada variable.

PARÁMETRO	P-VALOR
COELECHOG	0,6748
COENGHOGCOMB	0,2009
COENGTRANSP	0,5111
PARTENGRENOV	0,59
TAIMP	0,9755
PGAS	0,748
PELEC	0,6376
Nº EMP	0,6359

Como se observa en la (Tabla 3. 23) todos los P-valor que acompañan a las variables explicativas son mayores a $\alpha=0,05$, por lo se concluye que el modelo está libre de heteroscedasticidad.

3.3.2.4 Autocorrelación

Contraste de Autocorrelación de 1^{er} Orden

$H_0 : \rho_1 = 0 \rightarrow$ No hay Autocorrelación de 1^{er} orden

$H_1 : \rho_1 \neq 0 \rightarrow$ Si hay Autocorrelación de 1^{er} orden.

Estadístico de Durbin-Watson

0,915366 (P=0,0011)

P-valor (0,0011) < $\alpha=0,05$

En conclusión, puesto que el P-valor es menor a α , se rechaza la hipótesis nula. Por lo que existe Autocorrelación de 1^{er} orden.

A continuación se realiza el FAS (Ilustración 3.24) y el FAP (Ilustración 3.25) para saber si existe Autocorrelación de 1^{er} orden donde se releja en los retardos que se muestran en el eje de la "X".

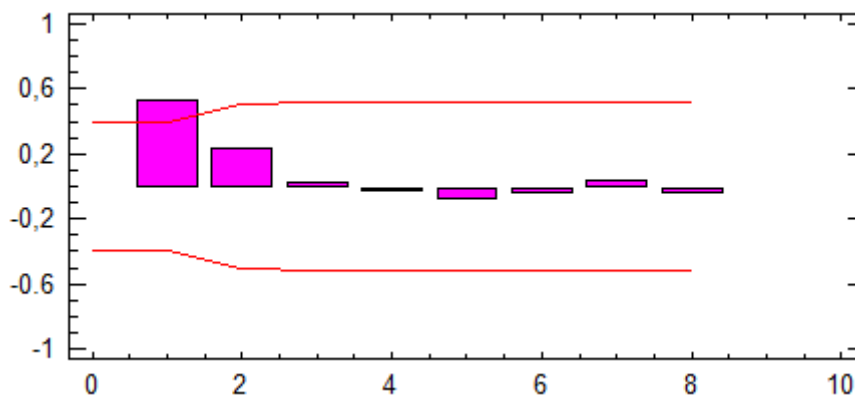


Ilustración 3. 24 FAS

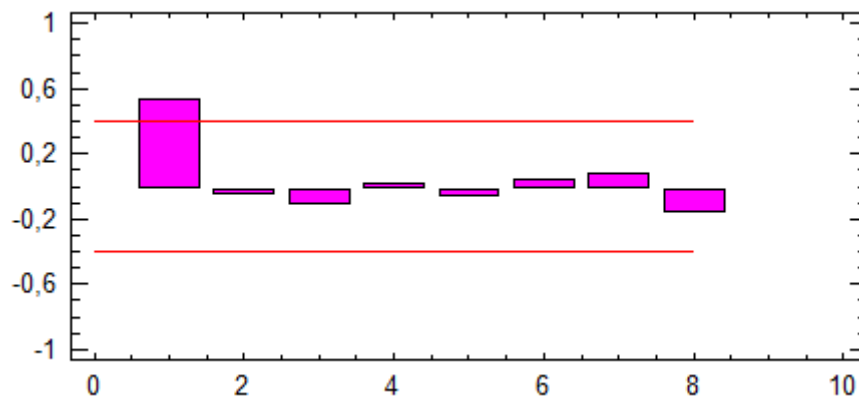


Ilustración 3. 25 FAP

Tanto en la (Ilustración 3.24) y la (Ilustración 3.25) se observa que existe Autocorrelación de 1^{er} orden positiva dado que el factor de Autocorrelación por el retardo 1 supera los límites de Autocorrelación.

Por lo que se concluye que hay problemas de Autocorrelación de 1^o orden positiva.

3.3.2.5 Puntos Influyentes

PUNTOS INFLUYENTES A PRIORI

Tabla 3. 24 Tabla donde se muestran los puntos influyentes.

País	FILA	Influencia	Distancia de Mahalanobis	DFITS
Dinamarca	6	0,749241	64,7773	-2,19349
Italia	14	0,950858	424,73	-5,57549
Países Bajos	18	0,142495	2,6993	1,50722
Suecia	24	0,814394	95,5741	-1,9894

Influencia media de un punto (\bar{n})=0,375

$$n_{ii} \geq 2\bar{n}$$

$$\bar{n} = 0,375$$

$$2\bar{n} = 2 \cdot 0,375 = 0,75$$

A priori

País	nii
Dinamarca	$0,749 < 2\bar{n}$
Italia	$0,951 > 2\bar{n}$
Países Bajos	$0,142 < 2\bar{n}$
Suecia	$0,814 > 2\bar{n}$

Por lo tanto, se concluye que Italia y Suecia son puntos influyentes a priori.

A posteriori

$$|DFITS| \geq 2\sqrt{\bar{n}}$$

$$2\sqrt{0,375} = 1,22$$

País	$ DFITS $
Dinamarca	$2,19 \geq 2\sqrt{\bar{n}}$
Italia	$5,57 \geq 2\sqrt{\bar{n}}$
Países Bajos	$1,50 \geq 2\sqrt{\bar{n}}$
Suecia	$1,98 \geq 2\sqrt{\bar{n}}$

Se concluye que Dinamarca, Italia, Países Bajos y Suecia son puntos influyentes a posteriori.

3.3.2.6 Residuos Atípicos

Tabla 3. 25 Tabla de residuos atípicos.

País	Fila	Y	Y Predicha	Residuo	Residuo Estudentizado
Países Bajos	18	69652,2	29397,3	40254,9	3,70
Polonia	19	70578,0	41688,6	28889,4	2,27

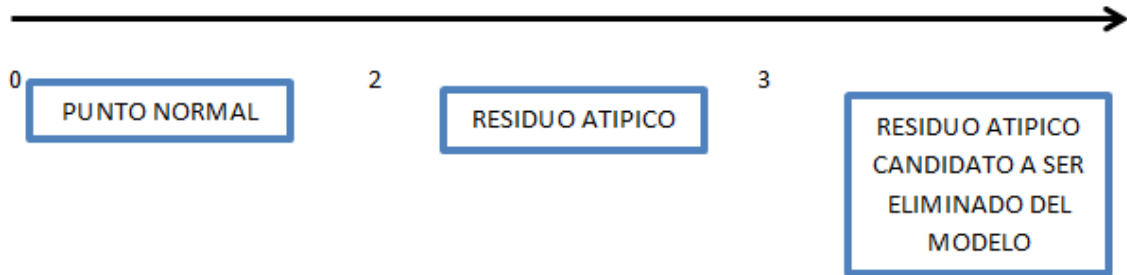


Ilustración 3. 26 Ilustración en la que indica si un punto es candidato a eliminar o no.

País	$ \text{Residuo Estudentizado} $
Países Bajos	3,70
Polonia	2,27

Por tanto, se concluye que Países Bajos y Polonia generan residuos atípicos y los Países Bajos sería un candidato a eliminar del modelo.

Los Países Bajos se eliminan de la (Tabla 2. 1) al ser un país que estaba dando problema a la hora de analizar los datos.

3.3.2.7 Puntos Influyentes

PUNTOS INFLUYENTES A PRIORI

Tabla 3. 26 Tabla donde se muestran los puntos influyentes.

País	FILA	Influencia	Distancia de Mahalanobis	DFITS
Italia	14	0,955862	453,821	-2,42866
Polonia	18	0,188639	3,92789	1,95847

Influencia media de un punto (\bar{n})=0,39

$$n_{ii} \geq 2\bar{n}$$

$$\bar{n} = 0,39$$

$$2\bar{n} = 2 \cdot 0,39 = 0,78$$

A priori

País **nii**

Italia $0,96 \geq 2\bar{n}$

Polonia $0,19 < 2\bar{n}$

Por lo tanto, se concluye que Italia son puntos influyentes a priori.

A posteriori

$$|DFITS| \geq 2\sqrt{\bar{n}}$$

$$2\sqrt{0,39} = 1,25$$

País **|DFITS|**

Italia $2,43 \geq 2\sqrt{\bar{n}}$

Polonia $1,96 \geq 2\sqrt{\bar{n}}$

Por lo tanto, se concluye que Italia y Polonia son puntos influyentes a posteriori.

3.3.2.8 Residuos Atípicos

Tabla 3. 27 Tabla de residuos atípicos.

País	Fila	Y	Y Predicha	Residuo	Residuo Estudentizado
Polonia	18	70578,0	40943,8	29634,2	4,06

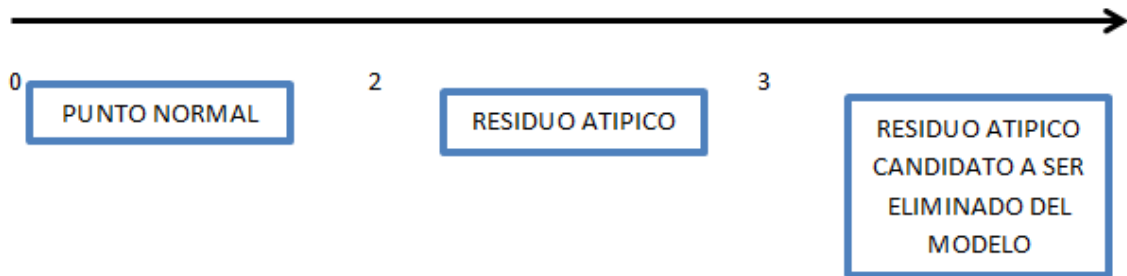


Ilustración 3. 27 Ilustración en la que indica si un punto es candidato a eliminar o no.

País $|Residuo\ Estudentizado|$

Polonia 4,06

Por tanto, se concluye que Polonia será un candidato a eliminar del modelo.

A continuación se elimina Polonia de la (Tabla 2. 1) porque también es otro país que estaba dando problemas a la hora de analizar los datos y volveremos a realizar análisis de normalidad de residuos y la Autocorrelación.

3.3.2.9 Normalidad de Residuos

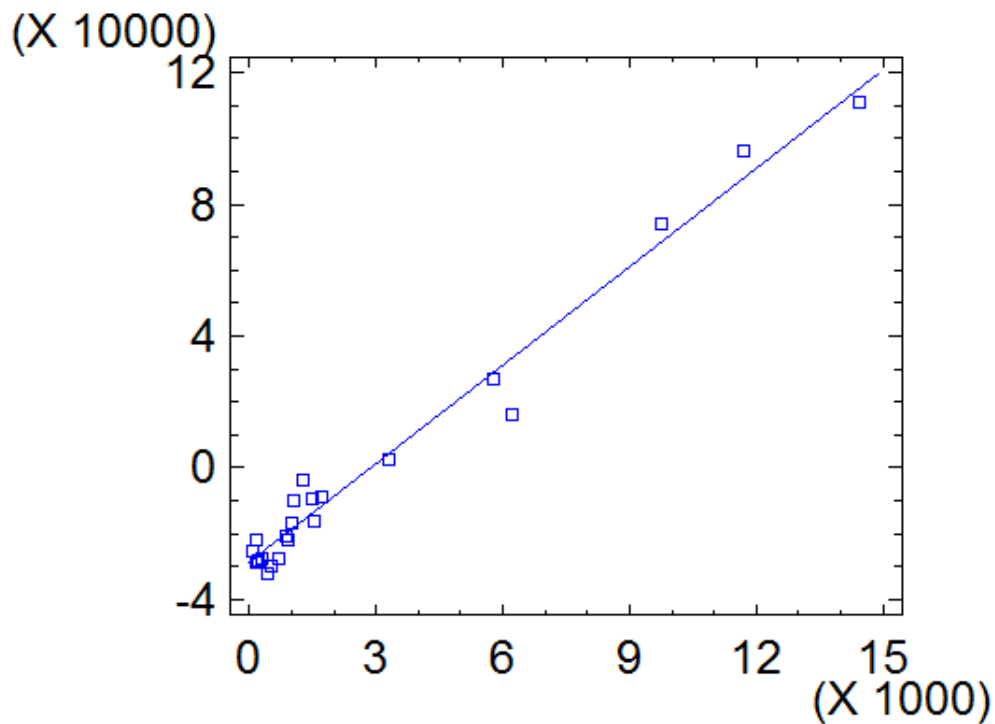


Ilustración 3. 28 Ilustración de normalidad de residuos.

Como se observa en la (Ilustración 3. 28) la mayoría de los puntos están sobre la línea de normalidad o bastante cerca, por lo tanto se considera normalidad de residuos.

HISTOGRAMA

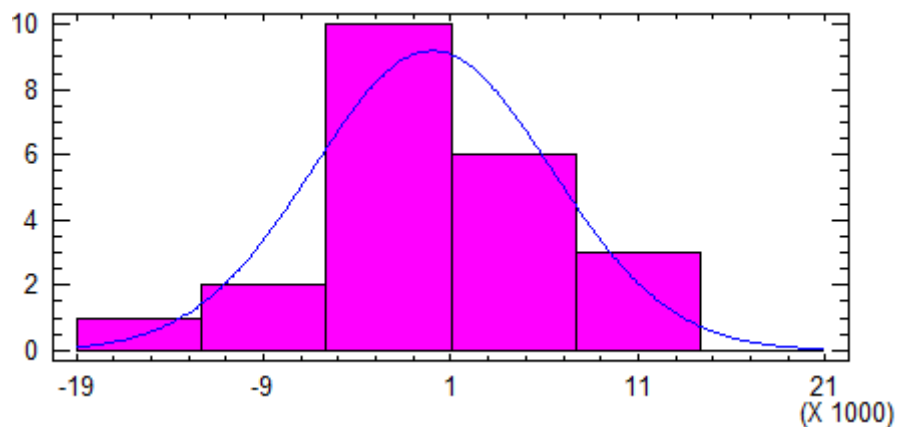


Ilustración 3. 29 Histograma para residuos.

Como se observa en la (Ilustración 3. 29) forma una campana casi perfecta, ya que la mayoría de los residuos están en la media, por lo que se puede suponer la normalidad de residuos.

A continuación, se realiza el test de hipótesis, del cual se sabrá con más certeza si existe normalidad de residuos o no.

H_0 : Los residuos son normales.

H_1 : Los residuos no son normales.


Tests para la Normalidad para RESIDUALS

Estadístico chi-cuadrado de bondad de ajuste = 11,0909
P-valor = 0,350479

Estadístico W de Shapiro-Wilks = 0,959309
P-valor = 0,470487

Puntuación Z para asimetría = 0,669785
P-valor = 0,502992

Puntuación Z para curtosis = 1,43303
P-valor = 0,151849



Se elige el P-valor más pequeño

En este caso elegimos el P-valor (0,151849) $>$ $\alpha=0,05$ por lo tanto se acepta la hipótesis nula, se concluye que los residuos se distribuyen normalmente.

3.3.2.10 Autocorrelación

Contraste de Autocorrelación de 1^{er} Orden

$H_0 : \rho_1 = 0 \rightarrow$ No hay Autocorrelación de 1^{er} orden

$H_1 : \rho_1 \neq 0 \rightarrow$ Si hay Autocorrelación de 1^{er} orden.

Estadístico de Durbin-Watson

1,8423 (P=0,3116)

P-valor (0,3116) > $\alpha=0,05$

En conclusión, puesto que el P-valor es mayor a α , se acepta la hipótesis nula. Por lo que no existe Autocorrelación de 1^{er} orden.

A continuación se realiza el FAS (Ilustración 3.30) y el FAP (Ilustración 3.31) para asegurarse si existe Autocorrelación de 1^{er} orden donde se refleja en los retardos que se muestran en el eje de la "X".

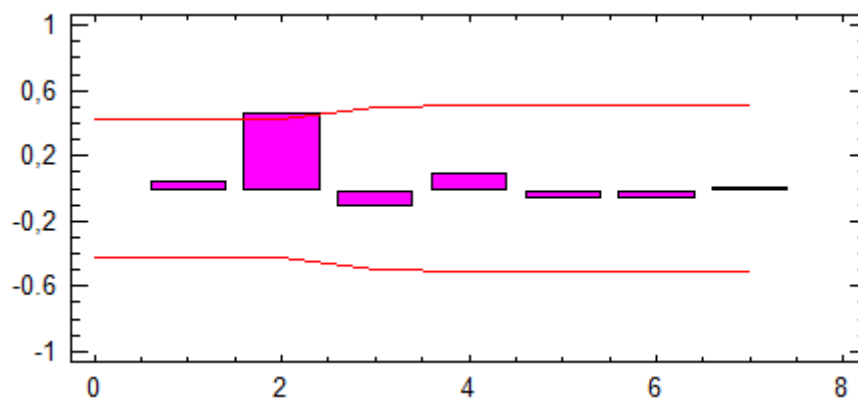


Ilustración 3. 30 Ilustración del FAS.

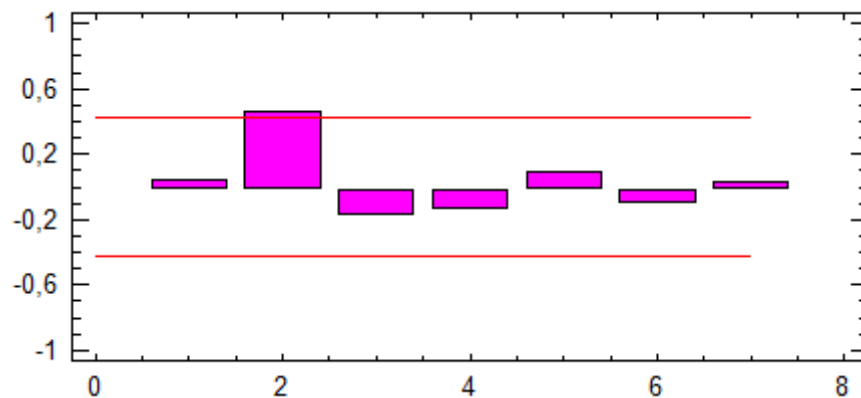


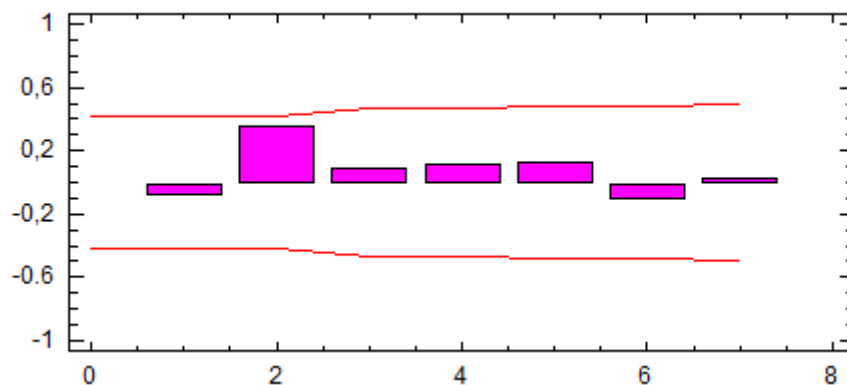
Ilustración 3. 31 Ilustración del FAP.

Tanto en la (Ilustración 3. 30) y la (Ilustración 3. 31) se observa que no existe Autocorrelación de 1^{er} orden positiva dado que el factor de Autocorrelación para el retardo 1 no supera los límites de Autocorrelación, pero se tiene que destacar que el retardo 2 sí que sobrepasa el límite de Autocorrelación, por lo que, se debe afirmar que si existe problema de Autocorrelación, pero este problema no tiene una explicación objetiva, porque aparece por cómo están ordenados los datos en la (Tabla 2. 1) , en este caso los datos de este proyecto han sido ordenados alfabéticamente y en el idioma Español.

A continuación se muestra un ejemplo, dónde los datos están ordenados aleatoriamente (Tabla 3.28) y se volverá a realizar la Autocorrelación (Ilustración 3. 32) y (Ilustración 3. 33) para que se compruebe que este problema no existe, como un problema que se debe solucionar.

Tabla 3. 28 Tabla de la base de datos del Eurostat.

Pais	Producción Energía	Con. Electrico Hogar	Con. Energetico Hogar Combustible	Con. energia transprte/PIB	Participacion Energías Renov.	Tasa Impositiva	Precio Gas	Precio electricidad	Nº empresas por pais
Estonia	5653,3	160,4	0,8	78,9	0,2	128,98	14,57	0,1351	29
Austria	12104	1520,8	18,5	105,2	7,5	168,98	21,32	0,2082	851
Belgica	14633,5	1703,1	31,5	85,6	4,3	127,74	18,32	0,2173	363
Bulgaria	10538	906	1,1	89,6	5,6	111,06	14,2397	0,0924	60
Suecia	34683,2	3284,9	0,6	77,9	16,7	230,55	34,088	0,2101	13
Dinamarca	16622,5	886,2	7,4	92,8	5,7	417,68	27,5854	0,3	328
Eslovaquia	6408,2	423,9	0,4	95,6	5,3	92,36	13,85	0,1698	8
Eslovenia	3550,5	277,6	15,8	117,1	3,4	216,58	18,57	0,161	8
España	34338,8	6235	18,4	80,8	0,4	170,22	20,34	0,2228	862
Reino Unido	109520,4	9755,2	6,5	78,6	4,4	247,21	14,7385	0,1741	757
Francia	135087,1	14437,1	16,5	85,2	7,2	214,26	18,83	0,1524	1654
Letonia	2143,5	153,3	4,2	86,2	3,1	101,55	14,0521	0,1378	7
Irlanda	2268,8	683,4	35,7	79,4	5	228,18	18,15	0,2295	32
Italia	36868,3	5759,5	7,9	91,2	5	363,12	23,17	0,2292	5550
Luxemburgo	139,7	75,4	38	99,1	3,9	214,63	17,34	0,1665	39
Lituania	1414,2	222,8	3,3	85,4	4,6	104,76	16,7458	0,137	16
Rumania	26110,9	1022,9	2,6	95,4	4,6	108,61	7,9114	0,1323	42
Portugal	5764,9	1058,8	19,5	95,4	0,7	133,13	23,23	0,2081	288
Croacia	3624,5	535,6	9	107	2,1	148,44	12,9191	0,1372	64
Republica Checa	29947,7	1265,3	0,1	99,2	5,7	127,41	17,8029	0,1525	65
Alemania	120566,3	11693,9	23,8	81,6	6,3	205,93	18,35	0,2919	2824
Grecia	9312,4	1500,1	26,3	85,5	1,1	257,19	21,44	0,1563	68


Ilustración 3. 32 Ilustración del FAP.

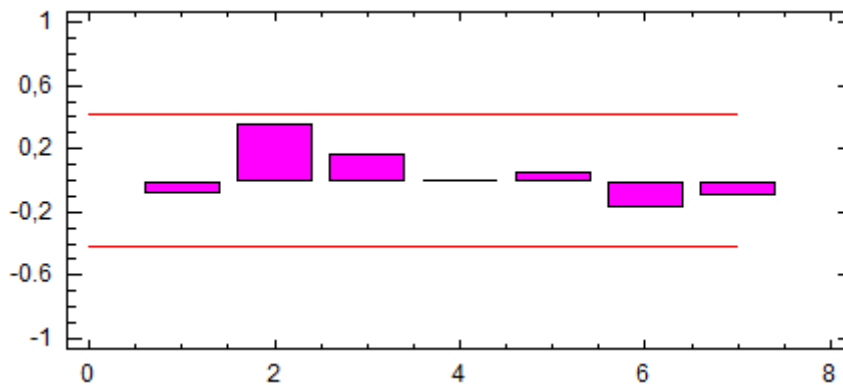


Ilustración 3. 33 Ilustración del FAS.

Para obtener el modelo definitivo se aplica una selección hacia adelante que introduce todas las variables explicativas de una en una, por mayor importancia, mientras sean significativas. Además, se elimina la constante del modelo, ya que no es significativa.

Tabla 3. 29 Contraste del modelo coeficiente que acompaña a la variable explicativa.

Fuente	Coefficiente-F	P-valor
Modelo	304,7	0,000

Tabla 3. 30 Parámetros del modelo.

Parámetro	Estimación	Error Estándar	T-Estadístico	P-valor
CONELECHOG	9,43419	0,541025	17,4376	0,000

El modelo definitivo es el siguiente y el R-cuadrado también se muestra a continuación, afirmar que es R-cuadrado muy bueno.

$$\text{Producción Energía} - \overline{\text{Producción Energía}} = 9,43419 * \text{COELECHOG} - \overline{\text{COELECHOG}}$$

R-Cuadrado= 93,83%

3.4 Predicciones

A partir del modelo económico definitivo, se van a realizar tres predicciones, para averiguar cuál es la producción de energía en ese país en el año 2013, se elegirá un país cualquiera para dichas predicciones, en este caso el país en el que se trabajará será Austria.

1^{er} Predicción en la cual la variable CONELECHOG aumenta de 1520,8 a 2250.

$$Y - 2824094 = 9,43419 \cdot (2250 - 2829,65)$$

$$Y = 9,43419 \cdot (2250 - 2829,65) + 2824094$$

$$Y = 9,43419 \cdot (-579,65) + 2824094$$

$$Y = -5468,53 + 2824094$$

$$Y = 2277241 \text{ toneladas}$$

Si aumenta el COELECHOG, la producción de energía siempre aumentará.

2ⁿ Predicción donde utilizamos el mismo dato de la (Tabla 2. 1) para la variable CONELECHOG.

En la siguiente (Tabla 3. 31) se muestra la predicción puntual y el intervalo mínimo y superior que ambos se utilizar para saber si nuestro modelo económico es bueno o no, aparte de información adicional de la variable COELECHOG.

Tabla 3. 31 Predicción puntual y pronóstico inferior y superior.

Parámetro	Estimación	Error Estándar	Estadísticos T	P-valor
COELECHOG	9,43419	0,527986	17,8683	0,000
	Predicción Puntual	Pronóstico Inferior	Pronóstico Superior	
	-12345,2	-33074	8383,49	

$$Y - 2824094 = 9,43419 \cdot (1520,8 - 2829,15)$$

$$Y = 9,43419 \cdot (1520,8 - 2829,15) + 2824094$$

$$Y = 9,43419 \cdot (-1308,35) + 2824094$$

$$Y = -12345,22 + 2824094$$

$$Y = 15895,77 \text{ toneladas}$$

Al mantener el mismo dato que en la (Tabla 2. 1) se observa que la producción de energía aumenta de 12104 a 15331,67, que es la producción real de energía.

Ahora se comprobará mediante el intervalo de confianza superior e inferior si la producción de energía se encuentra entre esos límites, en el caso de que sí, resultará ser un modelo económico bastante bueno.

Pronóstico Superior

$$\text{Producción Energía} - \overline{\text{Producción Energía}} = 8383,49$$

$$\text{Producción Energía} = 8383,49 + 28240,9 = 36624,39 \text{ toneladas}$$

Pronóstico Inferior

$$\text{Producción Energía} - \overline{\text{Producción Energía}} = -33074$$

$$\text{Producción Energía} = -33074 + 28240,9 = -4833,1 \text{ toneladas} \cong 0$$

Como un valor medido en toneladas no puede ser negativo en este caso se considera 0.

Intervalo de Confianza

$$[0 \leq \text{Producción Energía}_{\text{Austria}} \leq 36624,39]_{95\%}$$

La producción de Energía en Austria para el año 2013, con la misma cantidad de COELECHOG, esta entre el intervalo de confianza.

3^{er} Predicción donde disminuye el dato de la variable CONELECHOG de 1520,8 a 980,5.

$$Y - 28240,94 = 9,43419 \cdot (980,5 - 2829,65)$$

$$Y = 9,43419 \cdot (980,5 - 2829,65) + 28240,94$$

$$Y = 9,43419 \cdot (-1849,15) + 28240,94$$

$$Y = -17445,23 + 28240,94$$

$$Y = 10795,71 \text{ toneladas}$$

Claramente se observa que si disminuye el COELECHOG la producción de energía siempre disminuirá.



4. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

4.1 CONCLUSIONES

- Se ha conseguido un modelo econométrico donde la variable Producción de Energía se explica en función de la variable Consumo Eléctrico Hogar (COELECHOG).
- El modelo que se ha conseguido explica un 93,83% de los datos.
- La variable Consumo Eléctrico Hogar (COELECHOG) influye de manera directa en la Producción de Energía.
- El resto de variables explicativas no influyen de manera significativa en la Producción de Energía.
- Se ha comprobado que al realizar predicciones empleando los mismos valores que proporciona el Eurostat, los intervalos de confianza proporcionados por el modelo contienen el dato puntual del país.
- Se ha detectado una elevada correlación entre las variables Consumo de Energía y Población frente a Consumo eléctrico hogar, que podría haber afectado a la estimación del modelo. Se decidió eliminar estas variables.
- Se eliminó los datos de Polonia y Países Bajos ya que se comportaban de forma muy distinta al resto de países.

4.2 FUTURAS LÍNEAS DE INVESTIGACIÓN

Este proyecto tiene la ventaja de generar nuevas maneras de enfocar el estudio, además de abrir nuevos procedimientos de trabajo con las que poder complementarlo.

Estas nuevas maneras de enfocar el estudio o procedimientos de trabajo son los siguientes:

- Realización de este mismo proyecto mediante todas las estructuras (semilogarítmica, logarítmica, inversa y logarítmica inversa), ya que este solo ha sido realizado mediante la estructura lineal.
- Realizar el mismo proyecto, pero con datos entre los años 2013 y años actualizados en un futuro, para comprobar futuras predicciones de lo que ocurre en el apartado 3.4 continúa sucediendo.
- Resolver el problema de multicolinealidad por el método de componentes principales, en vez de por el método de eliminación de variables, ya que así obtendríamos una información más completa porque mediante la eliminación de variables se pierden datos que pueden llegar a ser importantes en el estudio.
- Realizar un análisis interno de Los Países Bajos y Polonia para conocer más datos sobre los factores internos que les puede llevar a producir energía.



5. BIBLIOGRAFÍA

- [1]<http://www.minetur.gob.es/energia/electricidad/Paginas/sectorElectrico.aspx>
- [2]<http://www.idae.es/index.php/idpag.16/relmenu.301/mod.pags/mem.detalle>
- [3]<http://www.idae.es/index.php/relcategoria.3932/id.702/relmenu.406/mod.pags/mem.detalle>
- [4]<http://www.minetur.gob.es/es-ES/GabinetePrensa/NotasPrensa/2016/Paginas/20160212-auditorias-eficiencias.aspx>
- [5]<http://www.statgraphics.net/caracteristicas/>
- [6]Martínez M, Ruiz R, Vallada E. <<Introducción a la Estadística>>. En: *Universidad Politécnica de Valencia, Valencia, España*.
- [7]<http://www.estadisticaparatodos.es/taller/graficas/cajas.html>
- [8]Novalés A, (2010). <<Econometría>>. En: *McGraw-Hill, D.L1993,2010., Madrid, España*
- [9]Chirivella V (2005). <<Econometría>>. En: *P.I.D 12059-C, Valencia, España*.
- [10]<http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=ten00076&plugin=1>
- [11]<http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=ten00095&plugin=1>
- [12]<http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdpc310&plugin=1>
- [13]http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=t2020_rk210&plugin=1
- [14]<http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdtr100&plugin=1>
- [15]<http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdcc340&plugin=1>
- [16]<http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tsdcc360&plugin=1>
- [17]<http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=ten00118&plugin=1>
- [18]<http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=ten00117&plugin=1>

Listado de Tablas.

Tabla 2. 1 Tabla de la base de datos del Eurostat de las variables con las que se va a trabajar a continuación.....	31
Tabla 3. 1 Resumen de estadísticos para la producción de energía primaria por parte de los recursos (1000 toneladas de petróleo equivalente) año 2013.....	32
Tabla 3. 2 Resumen de estadísticos para el consumo de energía final (1000 toneladas de petróleo equivalente) año 2013.....	34
Tabla 3. 3 Resumen de estadísticos para el consume de electricidad de los hogares (1000 toneladas de petróleo equivalente) año 2013.....	36
Tabla 3. 4 Resumen de estadísticos para el consumo final de energía en los hogares por combustible (%) año 2013.....	37
Tabla 3. 5 Resumen de estadísticos para el consume de energía del transporte en relación con el precio interior bruto (volúmenes encadenados a tipos de cambio 2000) año 2013.....	38
Tabla 3. 6 Resumen de estadísticos para la participación de energías renovables en el consumo de combustible de transporte (%) año 2013.....	39
Tabla 3. 7 Resumen estadísticos de la tasa impositiva (€/tonelada de petróleo) año 2013.....	40
Tabla 3. 8 Resumen estadísticos para los precios del gas por tipo de usuario (€ per KWh) año 2013.....	41
Tabla 3. 9 Resumen de Estadísticos de los precios de la electricidad por tipo de usuario (€ por KWh) año 2013.....	42
Tabla 3. 10 Resumen de estadísticos para el número de población (millones) año 2013.....	43
Tabla 3. 11 Resumen de Estadísticos para el número de empresas año 2013.....	44
Tabla 3. 12 Matriz del tipo de una variable frente a otra.....	57
Tabla 3. 13 Matriz inversa del tipo de una variable frente al resto.....	57
Tabla 3. 14 Tabla autovalores.....	58
Tabla 3. 15 Matriz del tipo de una variable frente a otra.....	59
Tabla 3. 16 Matriz inversa del tipo de una variable frente a otra.....	59
Tabla 3. 17 Tabla autovalores.....	60
Tabla 3. 18 Matriz del tipo de una variable frente a otra.....	61
Tabla 3. 19 Matriz inversa del tipo de una variable frente al resto.....	61
Tabla 3. 20 Tabla autovalores.....	62
Tabla 3. 21 Contraste del modelo, coeficiente que acompaña a la variable explicativa.....	63
Tabla 3. 22 Parámetros del modelo.....	64
Tabla 3. 23 Tabla donde se muestra el P-valor de cada variable.....	69
Tabla 3. 24 Tabla donde se muestran los puntos influyentes.....	72
Tabla 3. 25 Tabla de residuos atípicos.....	73
Tabla 3. 26 Tabla donde se muestran los puntos influyentes.....	74
Tabla 3. 27 Tabla de residuos atípicos.....	75
Tabla 3. 28 Tabla de la base de datos del Eurostat.....	80
Tabla 3. 29 Contraste del modelo coeficiente que acompaña a la variable explicativa.....	81
Tabla 3. 30 Parámetros del modelo.....	81
Tabla 3. 31 Predicción puntual y pronóstico inferior y superior.....	82

Listado de Ilustraciones:

Ilustración 1. 1 Proceso para suinistrar la energía.	3
Ilustración 2. 1 Algoritmo de Procesos.	8
Ilustración 2. 2 Los cuartiles dividen la muestra ordenada en cuatro segmentos con igual número de datos.	10
Ilustración 2. 3 Tipos de curtosis.	11
Ilustración 2. 4 Tipos de asimetría.	12
Ilustración 2. 5 Gráfico de Cajas y Bigotes.	13
Ilustración 2. 6 Relación entre dos variables.	14
Ilustración 2. 7 Comportamiento de las variables explicativas al explicar Y.	15
Ilustración 2. 8 Papel probabilístico normal.	20
Ilustración 2. 9 Histograma.	21
Ilustración 2. 10 Zona de aceptación/rechazo de autocorrelación en la prueba de Durbin Watson.	24
Ilustración 2. 11 Función de autocorrelación simple.	25
Ilustración 2. 12 Función autocorrelación parciales.	25
Ilustración 2. 13 Representación de un punto influyente (outlier).	27
Ilustración 2. 14 Ilustración para localizar los puntos atípicos o anómalos.	28
Ilustración 2. 15 Predicción puntual de Y , dado XP , YP , y por intervalos de confianza ($I. C$). 29	29
Ilustración 3. 1 Gráfico de Caja y Bigotes de la producción de energía (1000 toneladas por petróleo equivalente) en el año 2013.	33
Ilustración 3. 2 Gráfico de cajas y bigotes para el consumo de energía final en el año 2013. ...	35
Ilustración 3. 3 Gráfico de cajas y bigotes para el consumo eléctrico de los hogares (1000 toneladas de petróleo equivalente) año 2013.	36
Ilustración 3. 4 Gráfico de cajas y bigotes para el consumo de energía en los hogares por combustible (%) año 2013.	37
Ilustración 3. 5 Gráfico de caja y bigotes para el consumo de energía del transporte en relación con el precio interior bruto (volúmenes encadenados a tipos de cambio de 2000) para el año 2013.	38
Ilustración 3. 6 Gráfico de cajas y bigotes para la participación de energías renovables en consumo de combustible (%) para el año 2013.	39
Ilustración 3. 7 Gráfico de caja y bigotes para la tasa impositiva (€/tonelada de petróleo) año 2013.	40
Ilustración 3. 8 Gráfico de caja y bigotes de los precios del gas por tipo de usuario (€ per kWh) año 2013.	41
Ilustración 3. 9 Gráfico de cajas y bigotes de los precios de la electricidad por tipo de usuario (€ por kWh) año 2013.	42
Ilustración 3. 10 Gráfico de cajas y bigotes de la población (millones) año 2013.	43
Ilustración 3. 11 Gráfico de cajas y bigotes para el número de empresas por país (unidades) año 2013.	44
Ilustración 3. 12 Ilustración X-Y para el consumo de energía final (1000 toneladas por petróleo equivalente) año 2013.	45

Ilustración 3. 13 Ilustración X-Y para el consumo de electricidad de los hogares (1000 toneladas de petróleo equivalente) año 2013.....	46
Ilustración 3. 14 Ilustración X-Y para el consumo final de energía en los hogares por combustible (%) año 2013.....	47
Ilustración 3. 15 Ilustración X-Y para el consumo de energía del transporte en relación con el precio interior bruto (volúmenes encadenados a tipos de cambio de 2000) año 2013.....	48
Ilustración 3. 16 Ilustración X-Y para la participación de las energías renovables en el consumo de combustible de transporte (%) año 2013.....	49
Ilustración 3. 17 Ilustración X-Y para la tasa impositiva (€/tonelada de petróleo) año 2013.....	50
Ilustración 3. 18 Ilustración X-Y para los precios del gas por tipo de usuario (€) año 2013.	51
Ilustración 3. 19 Ilustración X-Y para los precios de la electricidad por tipo de usuario (€) l año 2013.....	52
Ilustración 3. 20 Ilustración X-Y número de población (millones) para cada país de la U.E citado anteriormente.....	53
Ilustración 3. 21 Ilustración X-Y para el número de empresas (unidades) año 2013.	54
Ilustración 3. 22 Ilustración probabilístico de residuos.	67
Ilustración 3. 23 Histograma para residuos.	67
Ilustración 3. 24 FAS.....	70
Ilustración 3. 25 FAP.....	71
Ilustración 3. 26 Ilustración en la que indica si un punto es candidato a eliminar o no.	73
Ilustración 3. 27 Ilustración en la que indica si un punto es candidato a eliminar o no.	75
Ilustración 3. 28 Ilustración de normalidad de residuos.....	76
Ilustración 3. 29 Histograma para residuos.	76
Ilustración 3. 30 Ilustración del FAS.	78
Ilustración 3. 31 Ilustración del FAP.	79
Ilustración 3. 32 Ilustración del FAP.....	80
Ilustración 3. 33 Ilustración del FAS.	81