

# *Análisis y predicción de la variabilidad del número de vehículos matriculados a nivel Europeo*

---

**MEMORIA PRESENTADA POR:**

*Sergio Vicent Sellés*

GRADO DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Convocatoria de defensa: Junio 2016

# ÍNDICE

<b>I. INTRODUCCIÓN</b> .....	3
<b>1. La Producción de automóviles en Europa</b> .....	4
1.1. Concepto de la producción de vehículos en Europa.....	4
1.2. Historia de la producción de automóviles en Europa.....	4-7
1.3. Noticias de actualidad sobre la producción de vehículos.....	7-8
<b>2. Formación requerida</b> .....	9
<b>II. OBJETIVOS</b> .....	10-11
<b>III. METODOLOGÍA Y TÉCNICAS</b> .....	12
<b>1. Búsqueda</b> .....	13
<b>2. Análisis Univariante</b> .....	14
<b>3. Análisis Bivariante</b> .....	14
3.1. Concepto de Multicolinealidad.....	14-15
3.1.1. Identificación.....	15
3.1.1.1. Matriz de Correlación.....	15-16
3.1.1.2. Matriz de Correlación Inversa.....	16
3.1.1.3. Índice de Acondicionamiento.....	16
3.2. Solución Multicolinealidad.....	17
<b>4. Análisis Multivariante</b> .....	17
4.1. Concepto Regresión múltiple.....	17
4.1.1. Otras estructuras de Regresión Múltiple.....	18
4.2. Estimación de los parámetros y validación del modelo.....	18
4.2.1. Significatividad.....	18
4.2.2. Análisis Gráfico.....	19-22
4.2.3. Normalidad de los Residuos.....	23
4.2.4. Heterocedasticidad.....	23-25
4.2.5. Autocorrelación.....	25-27
4.2.6. Puntos Atípicos e Influyentes.....	27
4.3. Predicción.....	28
4.4. Herramientas de Cálculo.....	28
<b>IV. CALCULOS Y RESULTADOS</b> .....	29
<b>1. Análisis Univariante</b> .....	30
1.1. Matriz de datos.....	30-58

<b>2. Análisis Bivariante</b> .....	58
2.1. Identificación.....	59-60
<b>3. Regresión Múltiple, primera estructura</b> .....	61
3.1. Modelo Teórico Lineal.....	61
3.2. Estimación de la primera ecuación de la Regresión Múltiple y validación del modelo.....	61
3.2.1. Análisis de la Significación del Modelo y Variables.....	62-65
3.2.2. Análisis gráficos de residuos.....	65-67
3.2.3. Normalidad de Residuos.....	67-68
3.2.4. Heterocedasticidad.....	68-69
3.2.5. Autocorrelación.....	69-70
<b>4. Regresión Múltiple, segunda estructura</b> .....	70
4.1. Modelo Teórico Logarítmico.....	70
4.2. Estimación de la primera ecuación de la Regresión Múltiple y validación del modelo.....	71
4.2.1. Análisis de la Significación del Modelo y Variables.....	71-72
4.2.2. Análisis gráficos de residuos.....	72-77
4.2.3. Normalidad de Residuos.....	78
4.2.4. Heterocedasticidad.....	79
4.2.5. Autocorrelación.....	80-81
<b>5. Regresión Múltiple, tercera estructura</b> .....	81
5.1. Modelo Teórico Inverso.....	81
5.2. Estimación de la primera ecuación de la Regresión Múltiple y validación del modelo.....	81
5.2.1. Análisis de la Significación del Modelo y Variables.....	82-84
5.2.2. Análisis gráficos de residuos.....	85-87
5.2.3. Normalidad de Residuos.....	87-88
5.2.4. Heterocedasticidad.....	88-89
5.2.5. Autocorrelación.....	89-90
<b>6. Interpretación y Predicción de los modelos</b> .....	91-93
<b>V. CONCLUSIONES</b> .....	94-95
<b>VI. FUTURAS LINEAS DE INVESTIGACIÓN</b> .....	96-97
<b>VII. BIBLIOGRAFÍA</b> .....	94-104

# I. Introducción

## **1. La Producción de vehículos en Europa.**

### **1.1 Concepto de la producción de vehículos en Europa.**

La producción de automóviles representa libertad y el crecimiento económico. Es un sector industrial muy exigente, particularmente competitivo y en constante evolución. Este sector requiere un capital humano y financiero muy importante.

Los automóviles son una tecnología liberadora para las personas de todo el mundo. El automóvil permite que las personas vivan, trabajen y jueguen en formas que eran inimaginables hace un siglo. Los automóviles proporcionan acceso a los mercados, a los puestos de trabajo.

La industria automotriz es el principal motor del crecimiento económico en el mundo. Es un sector clave de la economía de cada país, muy importante en el mundo. La industria sigue creciendo, registrando un aumento del 30 por ciento en la última década (1995-2005). La producción de automóviles crea gran número de empleados. La construcción de más de 60 millones de vehículos requiere el empleo de cerca de 9 millones de personas directamente en la fabricación de los vehículos y las partes que intervienen en ellos. Esto es más del 5 por ciento del empleo manufacturero total del mundo. Se estima que cada empleo directo de automóviles compatible con al menos otros 5 empleos indirectos en la comunidad, lo que resulta en más de 50 millones de puestos de trabajo debidos a la industria automotriz. Los autos se construyen utilizando los bienes de muchas industrias, incluyendo acero, hierro, aluminio, vidrio, plástico, vidrio, alfombras, textiles, chips de computadora, caucho y más.

### **1.2 Historia de la producción de automóviles en Europa.**

La historia en Europa está marcada por la producción de automóviles ya que en su momento fue una gran e importante innovación para la industria. [1]

La historia del automóvil empieza con los vehículos autopropulsados por vapor del siglo XVIII. En 1885 se crea el primer vehículo automóvil por motor de combustión interna con gasolina. Se divide en una serie de etapas marcadas por los principales hitos tecnológicos.

Uno de los inventos más característicos del siglo XX ha sido sin duda el automóvil. Los primeros prototipos se crearon a finales del siglo XIX, pero no fue hasta alguna década después cuando estos vehículos empezaron a ser vistos como algo "útil".

El intento de obtener una fuerza motriz que sustituyera a los caballos se remonta al siglo XVII. El automóvil recorre las tres fases de los grandes medios de propulsión: vapor, electricidad y gasolina.

El primer vehículo a vapor (1769) es el "Fardier", creado por Nicolás Cugnot, demasiado pesado, ruidoso y temible.

El automóvil pasó por varias Etapas:

### **-Etapa de invención**

Nicolas-Joseph Cugnot (1725-1804), escritor e inventor francés, dio el gran paso, al construir un automóvil de vapor, diseñado inicialmente para arrastrar piezas de artillería. El Fardier, como lo llamó Cugnot, comenzó a circular por las calles de París en 1769. Se trataba de un triciclo que montaba sobre la rueda delantera una caldera y un motor de dos cilindros verticales y 50 litros de desplazamiento; la rueda delantera resultaba tractora y directriz a la vez, trabajando los dos cilindros directamente sobre ella. En 1770 construyó un segundo modelo, mayor que el primero, y que podía arrastrar 4,5 toneladas a una velocidad de 4 km/h. Con esta versión se produjo el que podría considerarse primer accidente automovilístico de la historia, al resultar imposible el correcto manejo del monumental vehículo, que acabó chocando contra una pared que se derrumbó fruto del percance. Cugnot todavía tuvo tiempo de construir una tercera versión en 1771, que se conserva expuesta en la actualidad en el Museo Nacional de la Técnica de París.

En 1784 William Murdoch construyó un modelo de carro a vapor y en 1801 Richard Trevithick condujo un vehículo en Camborne (Reino Unido). En estos primeros vehículos se desarrollaron innovaciones como el freno de mano, las velocidades y el volante.

El belga Etienne Lenoir hizo funcionar un coche con motor de combustión interna alrededor de 1860, propulsado por gas de carbón.

Alrededor de 1870, en Viena, el inventor Siegfried Marcus hizo funcionar el motor de combustión interna a base de gasolina, conocido como el "Primer coche de Marcus". En 1883, Marcus patentó un sistema de ignición de bajo voltaje que se implantó en modelos subsiguientes.

## *Análisis y predicción de la variabilidad del número de vehículos matriculados a nivel Europeo*

Es comúnmente aceptado que los primeros automóviles con gasolina fueron casi simultáneamente desarrollados por ingenieros alemanes trabajando independientemente: Karl Benz construyó su primer modelo (el Benz Patent-Motorwagen) en 1885 en Mannheim. Benz lo patentó el 29 de enero de 1886 y empezó a producirlo en 1888. Poco después, Gottlieb Daimler y Wilhelm Maybach, de Stuttgart, diseñaron su propio automóvil en 1889.

### **-Etapa veterana**

En 1900, la producción masiva de automóviles ya había empezado en Francia y Estados Unidos. Las primeras compañías creadas para fabricar automóviles fueron las francesas Panhard et Levassor (1889), y Peugeot (1891). En 1908, Henry Ford comenzó a producir automóviles en una cadena de montaje, sistema totalmente innovador que le permitió alcanzar cifras de fabricación hasta entonces impensables.

En 1888, Bertha Benz viajó 80 km desde Mannheim hasta Pforzheim (Alemania) para demostrar el potencial del invento de su marido.

### **-Etapa del latón**

Ford modelo T. Así nombrada por el uso frecuente de latón para las carrocerías. En esta etapa la estética de los automóviles aún recordaba a la de los antiguos coches de caballos.

### **-Etapa de época**

Austin 7 Box saloon (1926). Comprende desde el final de la Primera guerra mundial hasta la Gran depresión de 1929.

Principales vehículos:

- 1922-1939 Austin 7
- 1924-1929 Bugatti Type 35
- 1927-1931 Ford T
- 1930 Cadillac V-16

### **-Etapa de preguerra**

Citroën Traction Avant

1929-1949 Desarrollo de los coches completamente cerrados y de forma más redondeada.

Automóviles relevantes:

Sergio Vicent Sellés

- 1932-1948 Ford B
- 1934-1940 Bugatti 57
- 1934-1956 Citroën Traction Avant
- 1938-2003 Volkswagen Tipo 1

Rolls-Royce Phantom IV de 1953

### **-Etapa moderna**

Caracterizada por el desarrollo de motores más seguros y eficientes y menos contaminantes.

- 1948-1990 Citroën 2CV
- 1955-1975 Citroën DS, en primer automóvil con frenos de disco
- 1961-1992 Renault 4
- 1965-1980 Renault 16, en primer modelo de automóvil con carrocería liftback, simula las líneas de un sedán pero tiene puerta trasera al igual que todos los hatchback3
- 1966-presente Toyota Corolla, Dodge Challenger, Ford Mustang
- 1970-presente Range Rover
- 1974-presente VW Golf
- 1975-1976 Cadillac Fleetwood Seventy-Five - uno de los autos más grandes fabricados
- 1976-presente Honda Accord
- 1986-presente Ford Taurus
- 1983-1998 Peugeot 205

### **1.3 Noticias de actualidad sobre la producción de vehículos.**

En esta sección se procede a explicar algunos artículos en los cuales se habla de la producción de vehículos en Europa en la actualidad. El primero de los artículos: Profesional Hoy – 9 de mayo del 2016. Este artículo se titula “La producción de vehículos en España llega casi al medio millón hasta febrero”

## *Análisis y predicción de la variabilidad del número de vehículos matriculados a nivel Europeo*

En este artículo se explica que España ha fabricado en los dos primeros meses del 2016 casi medio millón de unidades (497.874), lo que supone un 7,59 % más que en el mismo periodo del año 2015, según los datos facilitados hoy por la Asociación Nacional de Fabricantes de Automóviles y Camiones. Este aumento de la producción se ha debido al impulso del consumo interno, a la mejora en los principales mercados europeos y a que se están ampliando las exportaciones a otros países fuera de la UE. A pesar de estos buenos datos, Anfac considera que sigue siendo necesario articular medidas que favorezcan la competitividad de los principales sectores industriales, ya que la asignación de nuevos modelos para las plantas españolas se basa en la competitividad de cada factoría. [2]

*\*Anfac: Asociación Española de Fabricantes de Automóviles Turismos y Camiones.*

El segundo artículo fue escrito el 4 de mayo de 2015 por Javier Romera, en la página web "eleconomista.es". Este artículo se llama "La producción de automóviles en España lidera el crecimiento europeo", en él se explica que el sector del motor mete la sexta marcha en España. Nuestro país está liderando el crecimiento en la producción de coches en Europa, consolidándose como el segundo mayor fabricante del Viejo Continente solo por detrás de Alemania y muy por delante ya de Francia o el Reino Unido y el octavo del mundo. También cabe destacar que habla de recuperación tras las crisis ya que las previsiones apuntan, en cualquier caso, a que este año se alcanzará el mayor nivel de producción desde que se inició la crisis económica hace siete años y que la tendencia, además, seguirá al alza. En 2016 está previsto alcanzar 2,8 millones de coches y en 2017 se batirá la barrera de los tres millones. Todo ello después de que entre 2007 y 2014 la producción bajara un 16,8 por ciento, lo que equivale a casi medio millón de unidades. [3]

## 2. Formación Requerida

Tanto para la búsqueda de información como para su tratamiento, el alumno ha necesitado los conocimientos adquiridos en las asignaturas:

- **Investigación Comercial:** para marcar los objetivos, generar información, investigar sobre el proyecto.
- **Introducción a la Estadística:** para poder interpretar la normalidad, y todos los tipos de test de hipótesis.
- **Modelos Matemáticos:** para todo tipo de cálculos realizados a lo largo del proyecto.
- **Métodos Estadísticos para la Economía:** para interpretar los test de hipótesis planteados a lo largo del modelo.
- **Economía Española:** para el análisis de las varias regresiones múltiples.
- **Econometría:** para el análisis y la realización de los modelos de regresión.

## **II. Objetivos**

El objetivo principal de este trabajo es el análisis de varias variables que afectan a cada país de la UE, para poder determinar de qué depende que un país tengas más producción o menos de automóviles a nivel europeo. Este estudio nos ayuda principalmente a saber en qué país tenemos una gran producción de automóviles y si es viable crear una empresa de este tipo.

A partir de dicho objetivo se plantean los siguientes objetivos generales y específicos:

- ✓ Realizar un análisis univariante de las variables para entender, resumir y describir los datos encontrados.
- ✓ Realizar un análisis bivariante para determinar si las variables elegidas tienen algún tipo de problema y en este caso solucionarlo.
- ✓ Desarrollar varios modelos de Regresión Múltiple para explicar la producción de vehículos en la Unión Europea en función de las variables aptas para el modelo bivariante.
- ✓ Conseguir la información necesaria para ser capaz de explicar cuáles son los factores que más afectan en la producción de automóviles en cada país.
- ✓ Comprobar y validar los modelos de Regresión Múltiple, según los contrastes de hipótesis, significatividad, normalidad, autocorrelación, heterocedasticidad de estos.
- ✓ Realizar predicciones en base al/los modelo/s válido/s obtenido/s.

### **III. Metodología y Técnicas**

## 1. Búsqueda

En este proyecto, se hicieron varias búsquedas para que se pudieran encontrar las variables de forma correcta y que a la vez estas estuvieran relacionadas entre sí para poder explicar la producción de automóviles en la Unión Europea. Todas las variables se extrajeron del EUROSTAT (Oficina Europea de Estadística). [4]

La variable a explicar es:

1. **Número de Matriculados.** Unidades: número de vehículos matriculados

Las variables explicativas que se encontraron fueron:

1. **Motorización.** Unidades: coches por cada 1000 habitantes.
2. **Exportaciones – Importaciones.** Unidades: índice ratio
3. **Tasa de Paro.** Unidades: % de personas en paro.
4. **Tasa de Pobreza.** Unidades: número de unidades de consumo en un hogar, como la suma que se le adjudica a cada miembro y todo esto tendrá un valor final en %.
5. **PIB per Cápita.** Unidades: índice puntos.
6. **Tasa de Empleo.** Unidades: % de personas que tienen empleo.
7. **Comercio Internacional.** Unidades: peso y medida del vehículo.
8. **Contaminación, suciedad y otros problemas ambientales.** Unidades: en miligramos de dióxido de carbono (CO<sub>2</sub>) (milésima parte de un gramo).

## **2. Análisis Univariante.**

Se hace un análisis del modelo para ver una serie de descripciones estadísticas. Esto ayuda principalmente a obtener información, sintetizar información y lo más importante, aporta representaciones gráficas que la mayoría de las veces, esto dice más que los números (Histograma, Cajas y Bigotes, Gráfico de Dispersión). Los estadísticos son los siguientes:

- ❖ Estadísticos de centralización: Media, Mediana y Moda.
- ❖ Estadísticos de dispersión: Varianza, Desviación típica, Curtosis, Rango, Mínimo y Máximo.
- ❖ Estadísticos de simetría: Coeficiente de asimetría y Curtosis.

## **3. Análisis Bivariante.**

Se hace un análisis del modelo para ver si hay problema de multicolinealidad. Para evitar este problema las variables explicativas del modelo no deben estar relacionadas entre sí. Si este ocurre, afectaría a los parámetros del modelo, es decir, influiría el orden en el que se introduzcan las X (variables explicativas) y se diría que el modelo muestra problema de multicolinealidad.

### **3.1 Concepto de Multicolinealidad.**

Es una situación que se presenta cuando dos o más variables explicativas están relacionadas entre sí y toman valores semejantes o proporcionales, es entonces cuando se tiene problemas de multicolinealidad en el modelo.

Estos problemas afectarán de forma negativa a las estimaciones del modelo:

- ⊗ Malas estimaciones de los parámetros.
- ⊗ Mala estimación de la desviación típica de la perturbación.

La multicolinealidad desencadena unas consecuencias; por un lado que el contraste de los parámetros significativos y contraste del modelo, no es significativo o al revés o por otro lado que la estimación de un parámetro deje de ser significativa si se introducen las variables en otro orden.

Por último cabe destacar que hay dos tipos de multicolinealidad:

- Multicolinealidad exacta. Viene dada cuando al relación entre las variables explicativas es perfecta, es decir,  $R\text{-cuadrado} = 1$ .
- Multicolinealidad aproximada. Viene dada por cualquier relación entre las variables, es decir,  $R\text{-cuadrado} \neq 1$ .

### 3.1.1 Identificación.

Se identificarán los diferentes tipos de multicolinealidad mediante tres métodos. Y una vez estén todos realizados, si aparece algún problema en alguno de ellos se deberá solucionar, como se mostrará más adelante. Los métodos son los siguientes:

- Análisis Matriz de Correlación.
- Análisis Matriz Inversa de Correlación.
- Cálculo Índice de Acondicionamiento.

#### 3.1.1.1 Matriz de Correlación.

Este método se encarga de detectar un problema de relación entre variable. La matriz de correlación R muestra los coeficientes que correlacionan las variables entre sí. La matriz R tiene unos en la diagonal principal de la matriz y el resto de valores son los coeficientes de correlación de las variables.

$$\begin{array}{c} \mathbf{X1} \quad \mathbf{X2} \quad \mathbf{X3} \\ \mathbf{X1} \left( \begin{array}{ccc} \mathbf{R11} & R12 & R31 \\ R21 & \mathbf{R22} & R23 \\ R31 & R32 & \mathbf{R33} \end{array} \right) \\ \mathbf{X2} \\ \mathbf{X3} \end{array}$$

-Figura III. 1 Ejemplo de Matriz de Correlación.

El criterio que se emplea en este método para saber si hay problema de multicolinealidad, es que si cualquiera de los valores supera  $|0,7|$  en valor absoluto, significara que hay multicolinealidad, es decir,  $|R_{ii}| > 0,7$

### 3.1.1.2 Matriz de Correlación Inversa.

La matriz de correlación inversa es un método que se encarga de relacionar una variable explicativa con todas las demás a la vez. - Por ejemplo: X1 con X2 y X3.

$$\begin{matrix} \text{X1} \\ \text{X2} \\ \text{X3} \end{matrix} \begin{pmatrix} 14,21 & -18,75 & 0,57 \\ -18,75 & 27,43 & -0,45 \\ 0,57 & -0,45 & 2,05 \end{pmatrix}$$

-Figura III. 2 Ejemplo de Matriz de Correlación Inversa.

El criterio para saber si hay o no multicolinealidad se hará a través de la observación de la diagonal principal y si algún  $R_{ii}^{-1}$  supera el valor de 10, significará que hay multicolinealidad, es decir,  $R_{ii}^{-1} > 10$

### 3.1.1.3 Índice de Acondicionamiento.

El índice de Acondicionamiento es un método que se basa en los autovalores de la matriz de correlación, que además se encarga de relacionar todas las variables mediante la siguiente ecuación:

$$I.C. = \sqrt{(\text{Autovalor max}/\text{Autovalor min})}$$

El criterio para saber si hay multicolinealidad o no en este tercer método, es cuando se observa si la solución de la ecuación es menor que 10, y esto significa que sí que se haya un problema de multicolinealidad.

### 3.2 Solución Multicolinealidad.

Para solucionar los problemas de multicolinealidad, no importa el método que se haya detectado, se procede a eliminar las variables las cuales están generando el problema. El criterio de eliminación consiste en la búsqueda del valor superior en la diagonal principal de la matriz inversa, siempre y cuando este valor supere al menos en una unidad al siguiente. En caso contrario se realizan pruebas de regresión múltiple eliminando cada vez una de las variables candidatas, y se selecciona el modelo con mayor ajuste.

## 4. Regresión Múltiple.

Con posterioridad se va a explicar la importancia del método multivariante que se plantea en el proyecto.

### 4.1 Concepto Regresión Lineal Múltiple.

El modelo de regresión lineal múltiple pretende explicar el comportamiento de múltiples variables frente a una variable dependiente, el cual sirve para determinar una ecuación de regresión o predicción para predecir un valor de la variable dependiente a partir de un conjunto de variables independientes. La variable dependiente será representada por la letra Y. Esta variable será la que se pretende explicar con un conjunto de variables explicativas, las cuales serán las variables independientes y serán representadas por la letra X ( $X_1, X_2 \dots X_i$ ).

En el modelo de regresión uniecuacional la estructura lineal será la siguiente:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \dots + \beta_i \cdot X_i + U$$

- $\beta_0$  = Valor medio de la variable dependiente (Y) cuando el valor de todas las variables independientes ( $X_i$ ) valen 0.
- $\beta_i$  = Incremento medio de la variable dependiente (Y) cuando aumenta en 1 unidad la variable independiente ( $X_i$ ). Cuando las demás variables independientes permaneces constantes.
- $U$  = Perturbación del modelo econométrico, representa el resto de variables que no se han tenido en cuenta.

#### 4.1.1 Otras estructuras de Regresión múltiple.

La estructura Logarítmica será la siguiente:

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \dots + \beta_i \ln X_i + U$$

Por último, la estructura Inversa será la siguiente:

$$Y = \beta_0 + \beta_1 / X_1 + \dots + \beta_i / X_i + U$$

### 4.2 Estimación de los parámetros y validación del modelo.

#### 4.2.1 Significatividad.

En esta parte se va a explicar el contraste de hipótesis que determinara si nuestras variables explicativas ( $X_i$ ), es decir, las variables independientes son significativas en el modelo. Este contraste se encargará de especificar si todos los parámetros son iguales a 0 o contrarios a este.

$\left. \begin{array}{l} H_0: \beta_1 = \beta_2 = \dots = \beta_9 = 0 \\ H_1: \beta_1 \neq \beta_2 \neq \dots \neq \beta_9 \neq 0 \end{array} \right\}$	- Si aceptamos $H_0 \implies$	No hay significatividad en ninguna variable, por lo tanto, el modelo NO es significativo.
	- Si aceptamos $H_1 \implies$	Hay significatividad en al menos en una variable, por lo tanto, el modelo SI es significativo.

En cualquier contraste de hipótesis se acepta la hipótesis nula ( $H_0$ ), siempre que el p-Valor sea mayor o igual que  $\alpha$ .  $\alpha = 0,05$ .

$$\left. \begin{array}{l} H_0: \beta_0 = 0 \\ H_1: \beta_0 \neq 0 \end{array} \right\} \quad \left. \begin{array}{l} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{array} \right\}$$

Para terminar con la significación del modelo, se procede a hablar sobre la solución al problema, es decir, si sucede este problema se buscarán otras variables o se planteara otro tipo de estructura, estructuras no lineales.

#### 4.2.2 Análisis Gráfico.

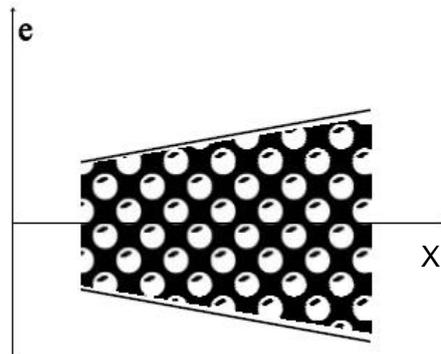
A partir de los análisis gráficos de los residuos del modelo, se verá si es posible tener problemas de Heterocedasticidad, Autocorrelación, Falta de linealidad y Puntos Atípicos e Influyentes.

En el gráfico de residuos frente a la estimación, se puede observar los siguientes problemas nombrados anteriormente.

- **Heterocedasticidad:** esto ocurre cuando la varianza de  $U$  no es constante, de forma que aumenta junto con el aumento de las variables  $X$  (Figura III. 3).

**e**  $\Rightarrow$  residuos estudentizados.

**X**  $\Rightarrow$  variable explicativa.



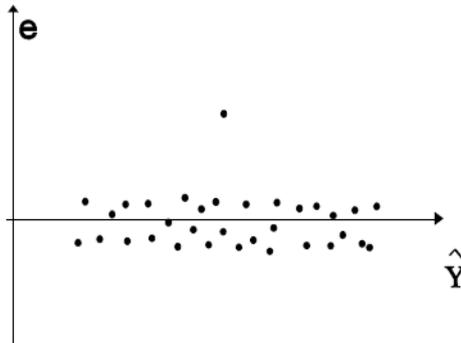
-Figura III. 3. Ejemplo de un gráfico de residuo con problemas de Heterocedasticidad.

*Análisis y predicción de la variabilidad del número de vehículos matriculados a nivel Europeo*

- **Punto Atípico o Anómalo:** se observan cuando alguna de las perturbaciones se aleja mucho de la pauta marcada por el resto. (Figura III.4).

**e**  $\Rightarrow$  residuos estudentizados.

**X**  $\Rightarrow$  variable explicativa.

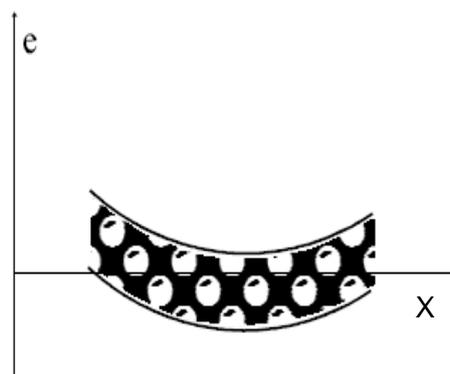


*-Figura III. 4. Ejemplo de residuos anómalos.*

- **Falta de Linealidad:** se produce por haber formulado mal el modelo. Para arreglarlo elevamos la variable explicativa al cuadrado (solo 1 variable explicativa). (Figura III. 5).

**e**  $\Rightarrow$  residuos estudentizados.

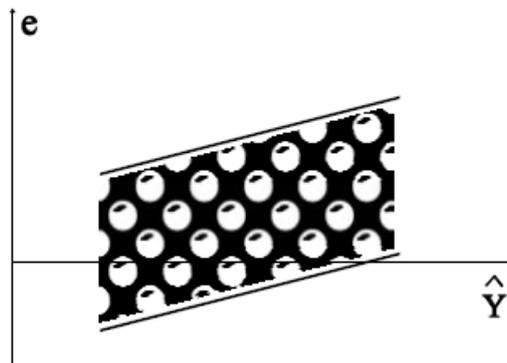
**X**  $\Rightarrow$  variable explicativa.



*-Figura III. 5. Gráfico ejemplo de falta de linealidad.*

- **Falta del término constante:** en el caso de que se haya planteado un modelo sin término independiente ( $\beta_0$ ), y este debería llevarlo, se observaría una curva ascendente de los residuos. Normalmente no sucede porque el programa utilizado, siempre plantea los modelos con términos independientes. (Figura III. 6).

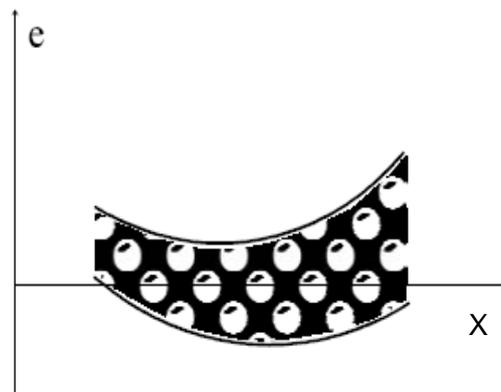
e  $\Rightarrow$  residuos estudentizados.  
Y  $\Rightarrow$  Predicción de variable dependiente.



-Figura III. 6. Gráfico del residuo de un ajuste sin contraste.

- **Combinación de problemas:** este caso se dará cuando se presente varios problemas de los anteriores a la vez. (Figura III. 7).

e  $\Rightarrow$  residuos estudentizados.  
X  $\Rightarrow$  variable explicativa.



-Figura III. 7. Gráfico con problemas de Heterocedasticidad y Falta de Linealidad.

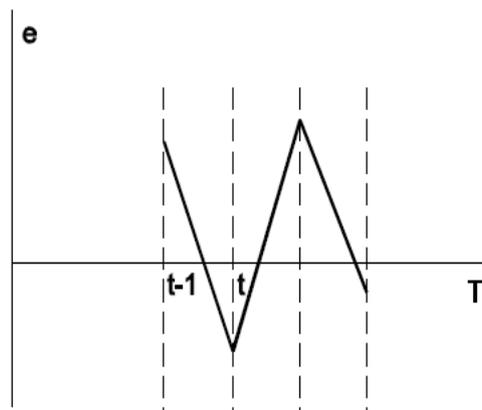
*Análisis y predicción de la variabilidad del número de vehículos matriculados a nivel Europeo*

Para terminar con los análisis gráficos, se presenta el gráfico de los residuos frente al número de filas, donde se puede observar los siguientes problemas:

- **Autocorrelación de 1º orden con signo negativo:** se observan cuando los residuos aparecen alternando su signo y tiene forma de zig-zag. (Figura III. 8).

**e**  $\Rightarrow$  residuo estudentizado.

**T**  $\Rightarrow$  número de fila.

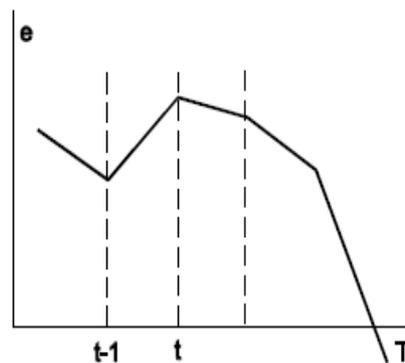


*-Figura III. 8 Gráfico de Autocorrelación de 1º orden con signo negativo.*

- **Autocorrelación de 1º orden con signo positivo:** se observan cuando los residuos toman forma sinusoidal prolongada. (Figura III. 9).

**e**  $\Rightarrow$  residuo estudentizado.

**T**  $\Rightarrow$  número de fila.



*-Figura III. 9. Gráfico de Autocorrelación de 1º orden con signo positivo.*

### 4.2.3 Normalidad de los Residuos.

En esta parte se va a explicar el contraste de hipótesis que se realiza para comprobar que los residuos siguen una distribución normal.

$H_0$  = Los residuos siguen una distribución normal.

$H_1$  = Los residuos no siguen una distribución normal.

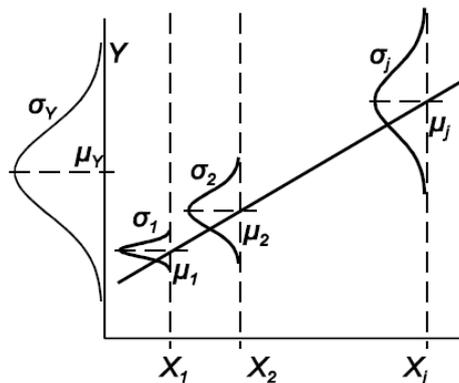
Se debe tener en cuenta que el programa estadístico solo saca una prueba de normalidad, entonces se debe cambiar la selección del test y seleccionar que este saque las cuatro pruebas:

- 1) Chi-Cuadrado.
- 2) Estadístico W de Shapiro-Wilk.
- 3) Valor Z para la asimetría.
- 4) Valor Z para curtosis.

### 4.2.4 Heterocedasticidad.

La heterocedasticidad, como ya se ha dicho anteriormente, es un problema que aparece cuando la varianza de las perturbaciones ( $U$ ) no es constante a lo largo de las observaciones (Figura III.10). 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 de algo que no lo es. La  $U$  tiene que tener una distribución normal de media cero y una desviación típica ( $\sigma^2$ ) constante. Si la desviación típica ( $\sigma^2$ ) no es constante, tendremos un error en el cálculo de:

1. Estimación de parámetros.
2. Predicciones.
3. Contraste de hipótesis.



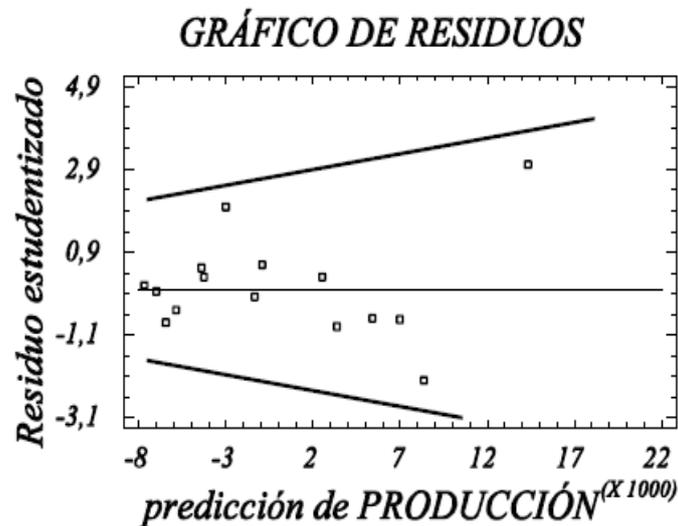
-Figura III. 10. Gráfico varianza no constante en Heterocedasticidad.

Seguidamente se exponen las causas de los problemas que hacen que se produzca la heterocedasticidad:

- ❖ Alguna de las variables explicativas tiene intrínseco en problema de heterocedasticidad, es decir, el problema viene de dentro de los datos.
- ❖ Que se introduzca en los datos debido a que los datos procedan de promediar un cierto número de datos.
- ❖ Error en el planteamiento inicial del modelo.
- ❖ Que aparezcan puntos atípicos e influyentes en los datos, da lugar a la heterocedasticidad.

La Heterocedasticidad se puede identificar a través del análisis gráfico y del contraste de hipótesis.

- ❖ **Análisis gráfico:** los residuos se relacionan con la variable dependiente o estos con las variables independientes para ver el motivo por el cual se produce el problema. (Figura III. 11).



-Figura III. 11. Gráfico problema de heterocedasticidad.

- ❖ **Contraste de hipótesis:** los residuos al cuadrado se relacionan con las variables dependientes y se formula este modelo.

$$❖ e^2 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$$

**H<sub>0</sub>**: desviación típica ( $\sigma^2$ ) = no existe heterocedasticidad.

**H<sub>1</sub>**: desviación típica ( $\sigma^2$ )  $\neq$  existe heterocedasticidad.

Finalmente, para terminar con la heterocedasticidad este problema se puede solucionar de la siguiente manera:

**1)** Se divide el modelo inicial entre la variable dependiente que genera los problemas de heterocedasticidad ( $X_p$ ) elevada a C.

$$C = (h / 2).$$

$$\frac{Y}{X_p^C} = \frac{\beta_0}{X_p^C} + \beta_1 \frac{X_1}{X_p^C} + \beta_2 \frac{X_2}{X_p^C} + \dots + \beta_K \frac{X_K}{X_p^C} + U$$

**2)** Para calcular h, se procede a ajustar los residuos al cuadrado frente a la variable dependiente que genera problemas la que se le va elevando a h.

h = este valor es de tipo Real, es decir, puede ser números enteros positivos, números negativos y fraccionarios.

*-Ejemplo:* 1, 2, 3, -1, -2, -3, (1/2), (1/3), (1/4)...

**3)** Con estos números se realiza una tabla donde como se ha dicho antes, h serán los números del ejemplo y al lado estará el ajuste del modelo ( $R^2$ ). Finalizaremos la tabla con el R-cuadrado mayor y por último repetiremos el análisis para comprobar que la solución es válida y que ya no tenemos heterocedasticidad.

#### **4.2.5 Autocorrelación.**

Cuando ocurre que los parámetros no tienen mínima varianza, los valores "t" y "F" son muy grandes y las predicciones no son adecuadas, tenemos problema de autocorrelación.

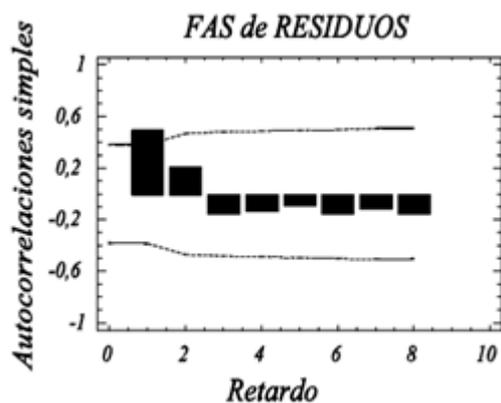
Las causas de la autocorrelación son las siguientes:

- ❖ Las variables tienen una estructura temporal.
- ❖ Los valores "t" y "F" son muy grandes y al ocurrir eso se pueden mal interpretar las variables dependientes (explicativas).
- ❖ La variable independiente se mide con error y no son válidas para el modelo.

Estas causas se pueden observar a través de:

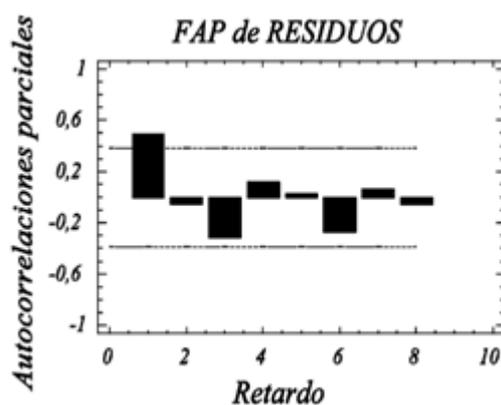
- ❖ **Métodos gráficos:** se representan los residuos frente a la estimación de la variable, frente a las variables dependientes y por esta causa se puede identificar la correlación entre ellos.
- ❖ **Pruebas de autocorrelación:** son varias pruebas con las que se comprueba la posible existencia de autocorrelación.
  - 1) Representación de la FAS y la FAP.
  - 2) Prueba de Durbin – Watson.
  - 3) Prueba de Wallis.
  - 4) Prueba de Durbin.
  - 5) Prueba de Ljung-Box.
  - 6) Prueba de Breusch-Godfrey.

Se destaca la primera prueba, la que es más general e importante en el proyecto. Esta prueba estudia la correlación tanto de primer orden como de segundo orden. Donde encontramos la función de autocorrelación simple (FAS) (Figura III. 12.A). Por otro lado, la función de autocorrelación parcial (FAP) (Figura III. 12.B).



-Figura III. 12A.

Gráfico autocorrelación simple.



-Figura III. 12B.

Gráfico autocorrelación parcial.

Para terminar con la autocorrelación, esta se podrá solucionar mediante cuatro métodos:

- 1) Transformación de Cochrane-Orcutt.

- 2) Método de las diferenciaciones.
- 3) Método de Cochrane-Orcutt dos etapas.
- 4) Método de Durbin dos etapas.

#### 4.2.6 Puntos Atípicos e Influyentes.

Se habla de **Puntos Atípicos** cuando en un modelo se ve que hay residuos muy grandes y, a causa de esto, es más difícil su identificación y eliminarlos.

- ❖ Si  $|rt_i| \geq 2$ : es un punto anómalo.
- ❖ Si  $|rt_i| \geq 3$ : es un punto anómalo candidato a eliminar del modelo.

En cuanto a los **Puntos influyentes** se encuentran cuando en una serie de modelos hay un par de puntos influyentes que determinan la estimación de los parámetros.

Existen dos métodos para detectar los puntos influyentes:

- ❖ **A priori**: si el leverage queda fuera del intervalo  $n_{ii} > 2\bar{n}$ , tal que  $\bar{n} \pm 3\hat{\sigma}_n$ , se aprecia que es un punto influyente. Estos puntos quedan dentro de la linealidad del modelo y no afectan a la tendencia o la pendiente del mismo.

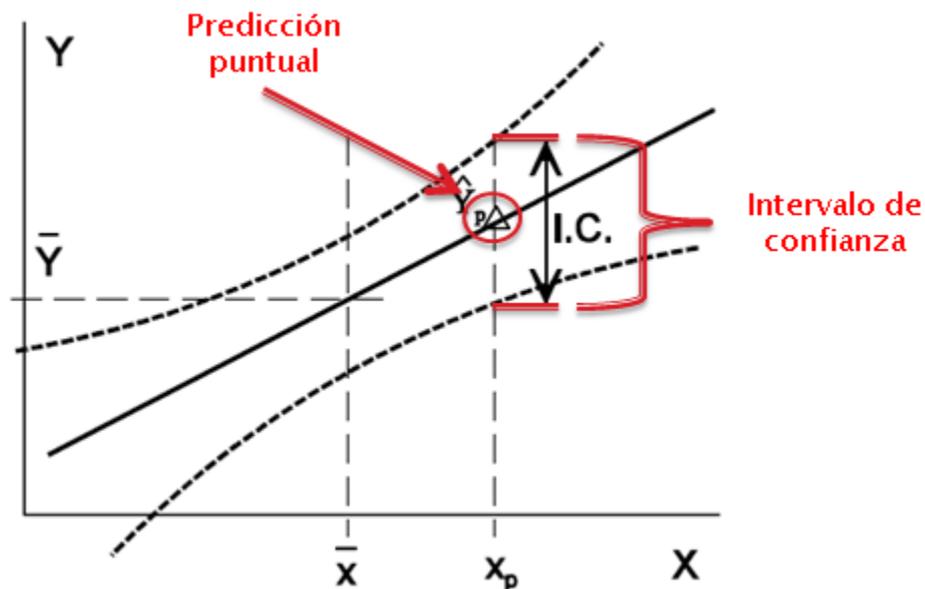
$n_{ii} \geq 2 \cdot$  influencia      Influencia  $\geq$  media de un solo punto.

- ❖ **A posteriori**: donde el método se basa en los parámetros para saber si es un punto influyente. Esta clase de puntos quedan fuera de la linealidad, afectando a la tendencia y la estimación de los parámetros del modelo.  $|DFITS| \geq 2\sqrt{\bar{n}}$

### 4.3 Predicción.

La predicción es un objetivo del proyecto porque permite hacer predicciones mediante una tabla final, determinar los valores de X (variables dependientes) y ajustar el modelo. En la Figura III. 13 se observa los dos tipos de predicciones que existen:

- 1) Predicción puntual.
- 2) Predicción por intervalos de confianza.



-Figura III. 13. Grafico predicciones puntual e intervalo de confianza.

### 4.4 Herramientas de cálculo.

En este punto se destacan los programas empleados que se han utilizado para realizar todo el proyecto.

- ❖ **Microsoft Excel:** programa que se ha utilizado para el análisis gráfico, realización de tablas y toda una serie de cálculos. La versión que se utiliza es Microsoft Excel 2013.
- ❖ **Statgraphics:** programa de gran importancia ya que ha sido el encargado de representar todas las regresiones múltiples que hay en el proyecto. La versión que se utiliza es Statgraphics XVI.

## **IV. Cálculos y Resultados**

## 1. Análisis Univariante.

A continuación se lista una tabla con los países de la unión europea seleccionados y las variables que pueden afectar a la producción de vehículos en la Unión Europea. En la tabla se observa la matriz de datos inicial con la que se inicia el trabajo y se ajusta el modelo, los datos informativos se encuentran en la página de [www.euroestat.com](http://www.euroestat.com).

### 1.1 Matriz de datos.

PAÍSES	MATRICULADOS (Número de vehículos)	MOTORIZACIÓN (Vehículos por cada 1000 habitantes)	EXPOR_IMPORTOR (Índice de un Ratio)	PARO (% de personas en Paro)
FINLAND	66.934	560	0.98	7.7
FRANCE	1.048.982	496	0.93	9.8
GERMANY	1.634.401	530	1.13	5.4
IRELAND	66.718	425	1.29	14.7
ITALY	815.213	621	1.04	10.7
PORTUGAL	53.375	406	0.98	15.8
SPAIN	406.071	476	1.02	24.8
BULGARIA	9.770	385	0.96	12.3
CROATIA	20.201	339	1.01	16
CYPRUS	6.426	549	0.93	11.9
CZECH REPUBLIC	94.233	448	1.08	7
ESTONIA	8.923	456	0.99	10
HUNGARY	27.785	301	1.08	11
LATVIA	5.462	305	0.94	15
MALTA	2.947	592	1.05	6.3
POLAND	148.710	486	1.01	10.1
ROMANIA	33.343	224	0.9	6.8
SLOVAKIA	34.316	337	1.06	14
SLOVENIA	28.327	518	1.07	8.9

-Tabla IV. 1. Matriz de datos inicial (Parte A)

<b>PAÍSES</b>	<b>POBREZA</b> (% de consumo en el hogar)	<b>PIB</b> (Índice de puntos)	<b>EMPLEO</b> (% de personas que tienen empleo)	<b>COMERCIO INTERNACIONAL</b> (Peso y medida)	<b>CONTAMINACIÓN</b> (En miligramos de CO <sub>2</sub> )
<b>FINLAND</b>	13.2	116	74	109.5	8.8
<b>FRANCE</b>	14.1	107	69.4	109.5	11.3
<b>GERMANY</b>	16.1	122	76.9	108.4	22.4
<b>IRELAND</b>	15.7	132	63.7	106.4	4.8
<b>ITALY</b>	19.5	101	60.9	110.7	17.3
<b>PORTUGAL</b>	17.9	77	66.3	110	14.9
<b>SPAIN</b>	20.8	92	59.6	109.6	8
<b>BULGARIA</b>	21.2	46	63	113.6	15
<b>CROATIA</b>	20.4	60	58.1	112.3	7.1
<b>CYPRUS</b>	14.7	91	70.2	109.9	15.5
<b>CZECH REPUBLIC</b>	9.6	82	71.5	109.5	15.5
<b>ESTONIA</b>	17.5	74	72.2	111.3	11.9
<b>HUNGARY</b>	14	65	61.6	105.9	11.8
<b>LATVIA</b>	19.2	60	68.1	111.4	22
<b>MALTA</b>	15.1	84	63.1	104.8	39.6
<b>POLAND</b>	17.1	67	64.7	107.4	11
<b>ROMANIA</b>	22.6	54	64.8	112.1	17.6
<b>SLOVAKIA</b>	13.2	75	65.1	105.4	15.1
<b>SLOVENIA</b>	13.5	81	68.3	107.9	16

-Tabla IV. 2. Matriz de datos inicial. (Parte B)

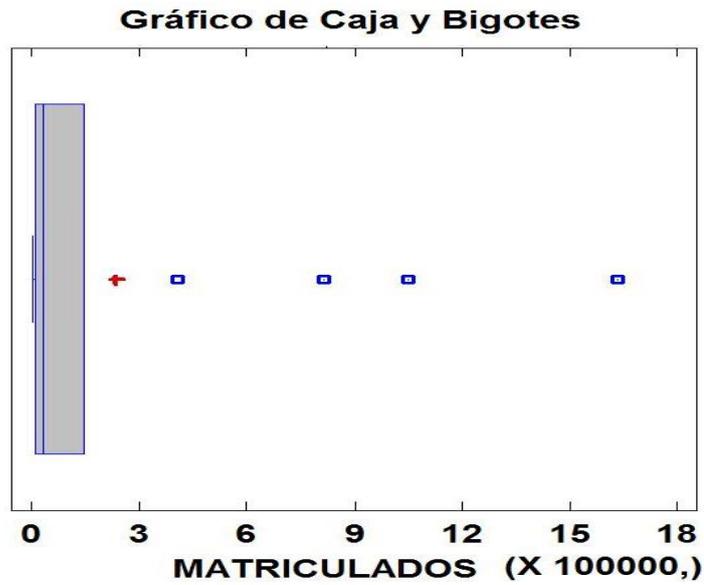
Ahora se van a describir las variables mediante una serie de estadísticos de centralización, dispersión y simetría que se han nombrado anteriormente.

→ Variable (Y) **MATRICULADOS**:

<b>MATRICULADOS</b>	
Recuento	19
Promedio	237481
Mediana	34316
Varianza	1,99E+11
Desviación Estándar	445773
Mínimo	2947
Máximo	1,63E+06
Rango	1,63E+06
Sesgo Estandarizado	4,15512
Curtosis Estandarizada	4,4819

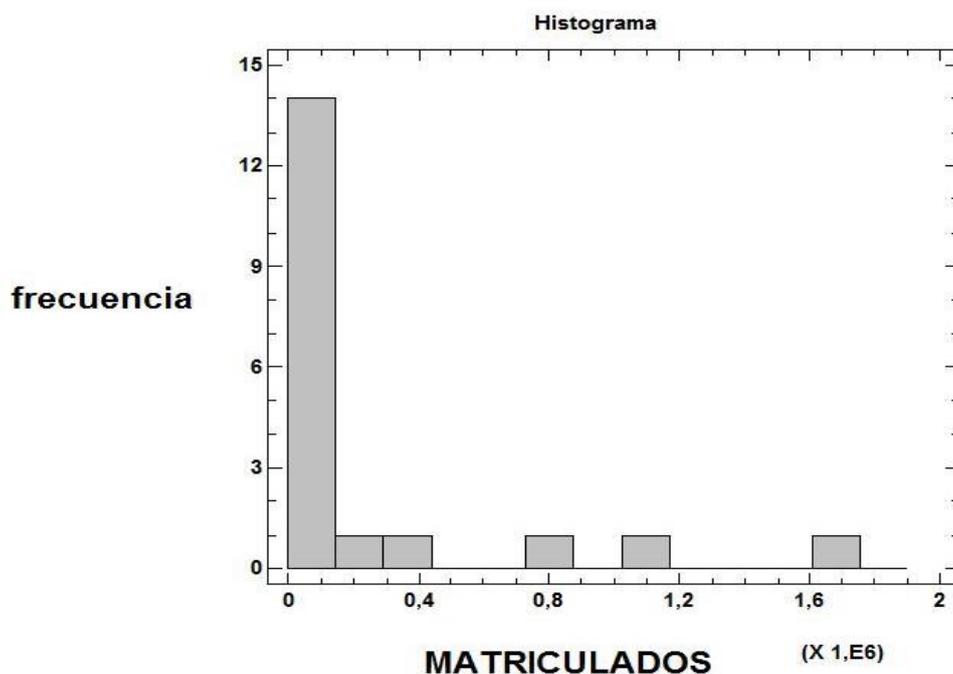
*-Tabla IV. 3. Variable Vehículos Matriculados.*

En la variable Y (nº vehículos Matriculados), se ha realizado un estudio de varios países de UE donde hay un recuento de 19 países. El promedio el cual consiste en el resultado que se obtiene al generar una división con la sumatoria de los vehículos matriculados en cada uno de los 19 países por el dígito que las represente, presenta un total de 237481. El sesgo es 4,15512. Esto quiere decir que presenta una asimetría hacia la derecha. El coeficiente curtosis también está por encima de 0, es decir, es positivo. Tiene forma de campana de Gauss (Leptocúrtica). Por último, la dispersión Rango = Max-Min. Es bastante grande.

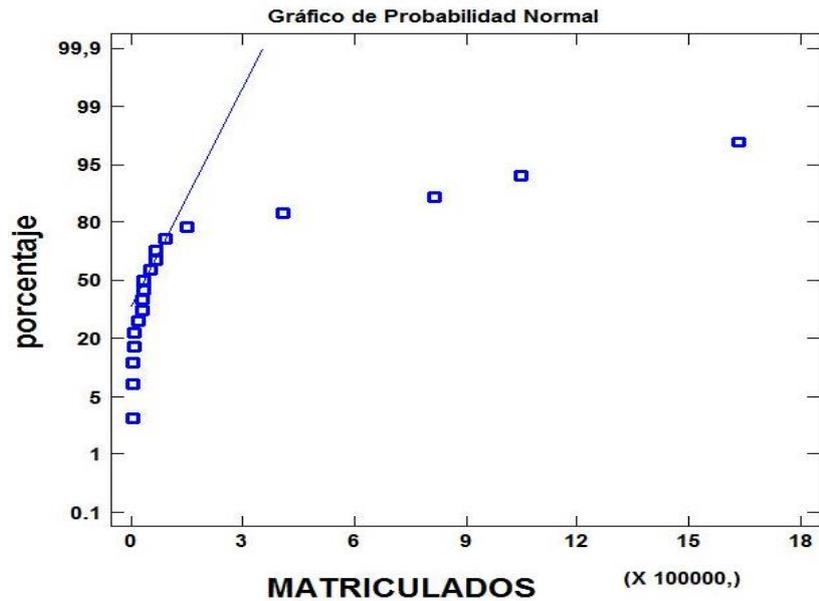


-Figura IV. 14. Gráfico de Caja y Bigotes Variable número de vehículos matriculados.

En este grafico (Gráfico IV. 14) de Caja y Bigotes, se puede observar que la Media y la Mediana están lejos, y además se observa la presencia de cuatro puntos anómalos destacados, que serán los causantes de la falta de normalidad de los datos de esta variable. Este fenómeno se observa también en la forma de distribución de las frecuencias del histograma (Gráfico IV. 45), que no se asemejan en nada a una campana de Gauss.

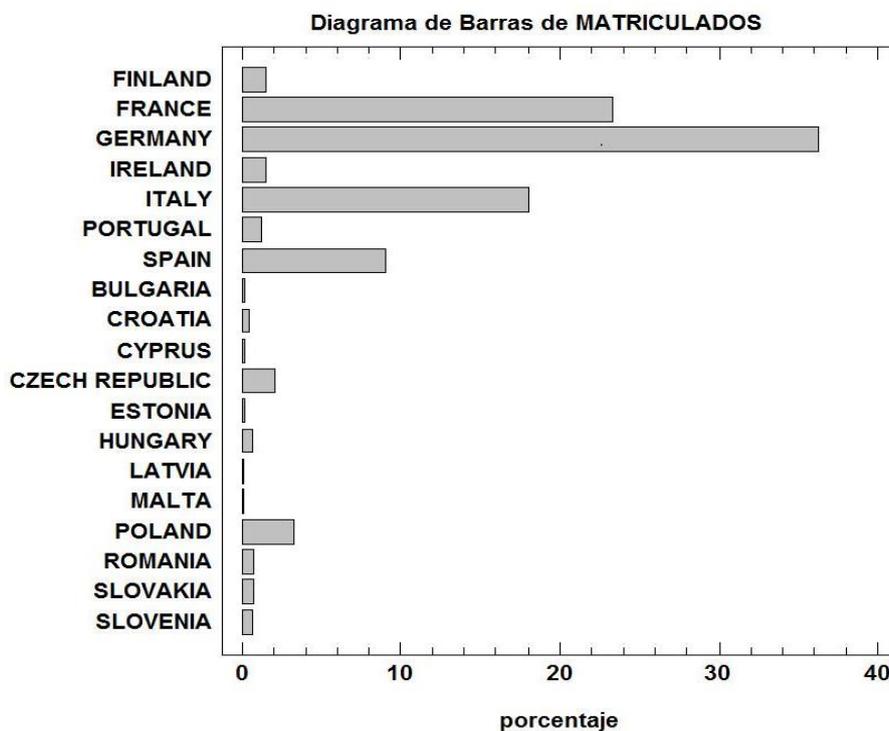


-Figura IV. 15. Histograma Variable número de vehículos Matriculados.



-Figura IV. 16. Gráfico Probabilidad Normal Variable número de vehículos Matriculados.

En el gráfico de probabilidad normal (Gráfico IV. 16) podemos ver que no sigue una distribución normal, ya que hay una serie de puntos muy lejos de la recta, estos puntos le llamaremos puntos anómalos.



-Figura IV. 17. Diagrama de barras, variable número de vehículos matriculados frente a los 19 países.

En el anterior diagrama de barras (Gráfico IV.17.) podemos se observa el nº de vehículos matriculados (%) frente a cada país de la UE.

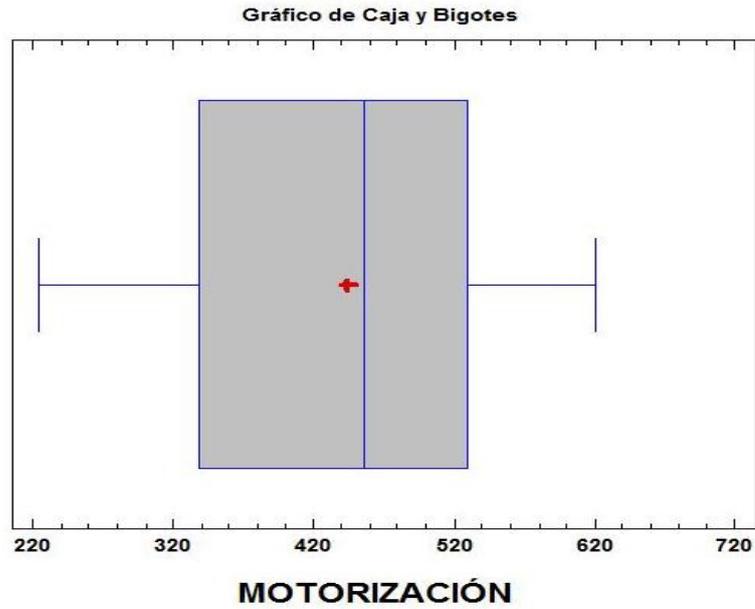
En primer lugar, se puede destacar, que el país que presenta el valor máximo sería Alemania seguido de Francia, Italia y España en cuarto lugar. Sobre el valor mínimo podemos destacar Malta y Latvia seguido de Bulgaria, Cyprus, y Estonia. Por último, decir que a simple vista se observa la gran diferencia entre la producción de vehículos matriculados entre unos países y otros.

→ Variable **MOTORIZACIÓN**:

<b>MOTORIZACIÓN</b>	
Recuento	19
Promedio	444,947
Mediana	456
Varianza	11786,2
Desviación Estándar	108,564
Mínimo	224
Máximo	621
Rango	397
Sesgo Estandarizado	-0,579813
Curtosis Estandarizada	-0,552063

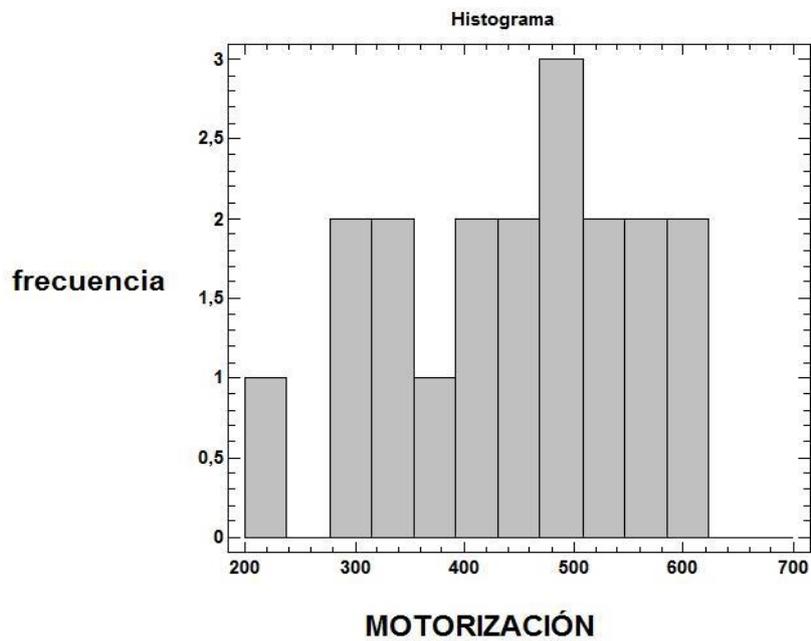
*-Tabla IV. 4. Variable Motorización.*

En esta variable la mediana está muy cerca de media. El sesgo es negativo, es decir, presenta una asimétrica hacia la izquierda. El coeficiente curtosis presenta una campana más plana (Platicúrtica).

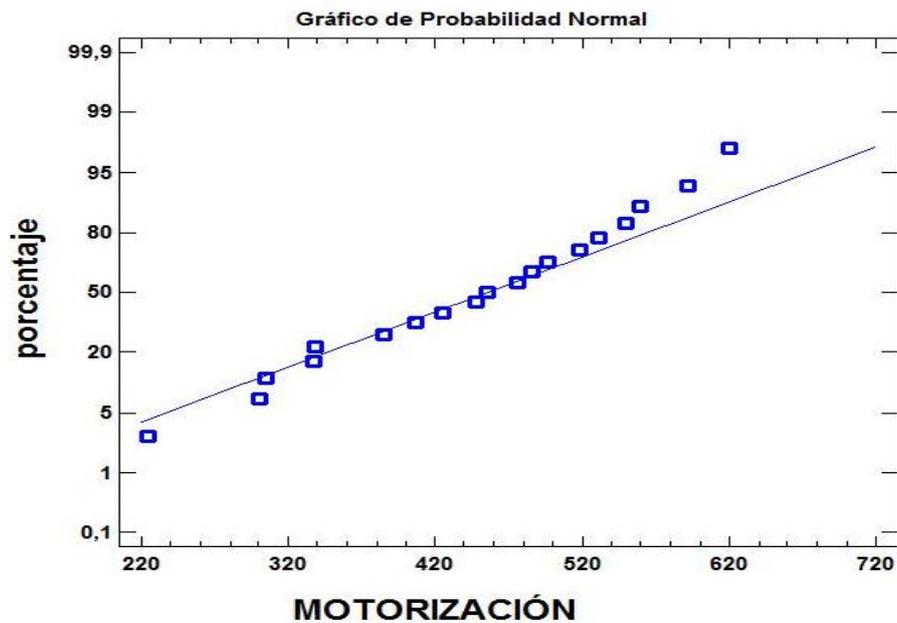


-Figura IV. 18. Gráfico de Caja y Bigotes Variable Motorización.

Se puede observar en el grafico que la media está muy cerca de la mediana.

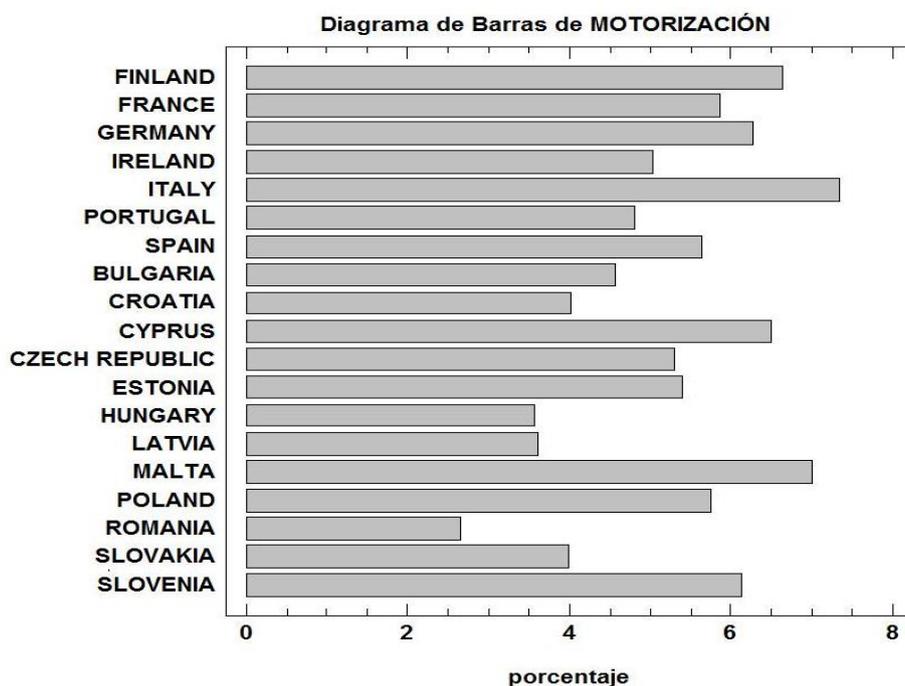


-Figura IV. 19. Histograma Variable Motorización.



-Figura IV. 20. Gráfico Probabilidad Normal Variable Motorización.

En este gráfico de Probabilidad Normal se observa que hay varios puntos anómalos, entonces tengo dudas de si sigue una distribución normal.



-Figura IV. 21. Diagrama de barras Variable Motorización frente a los 19 países.

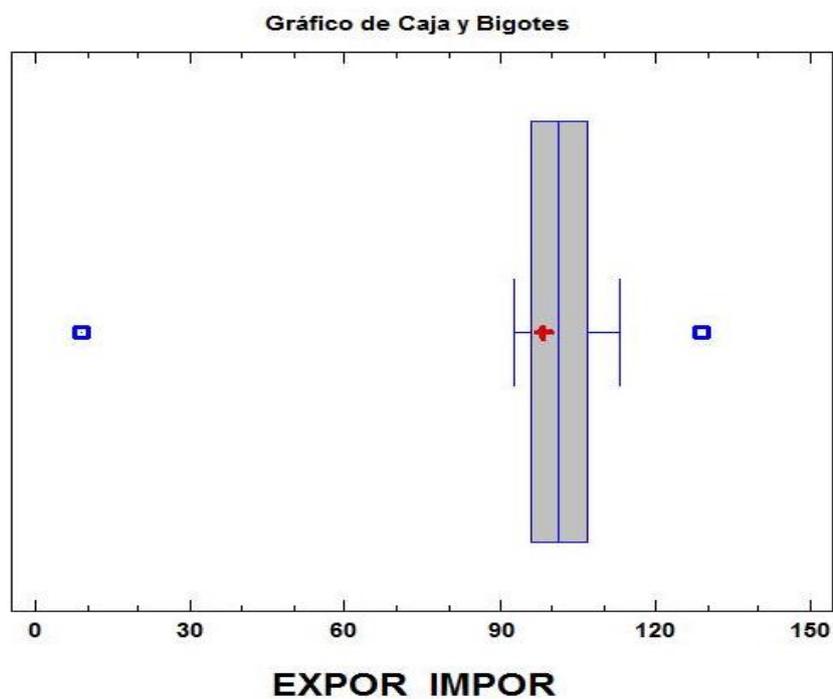
En este Diagrama de Barras se puede ver que el país con más motorización sería Italia, seguido de Malta. El país que menos porcentaje de motorización tiene sería Rumania.

→ Variable **EXPORTACIÓN-IMPORTACIÓN**:

<b>EXPOR-IMPOR</b>	
Recuento	19
Promedio	98,1053
Mediana	101
Varianza	535,655
Desviación Estándar	23,1442
Mínimo	9
Máximo	129
Rango	120
Sesgo Estandarizado	-5,98261
Curtosis Estandarizada	12,2324

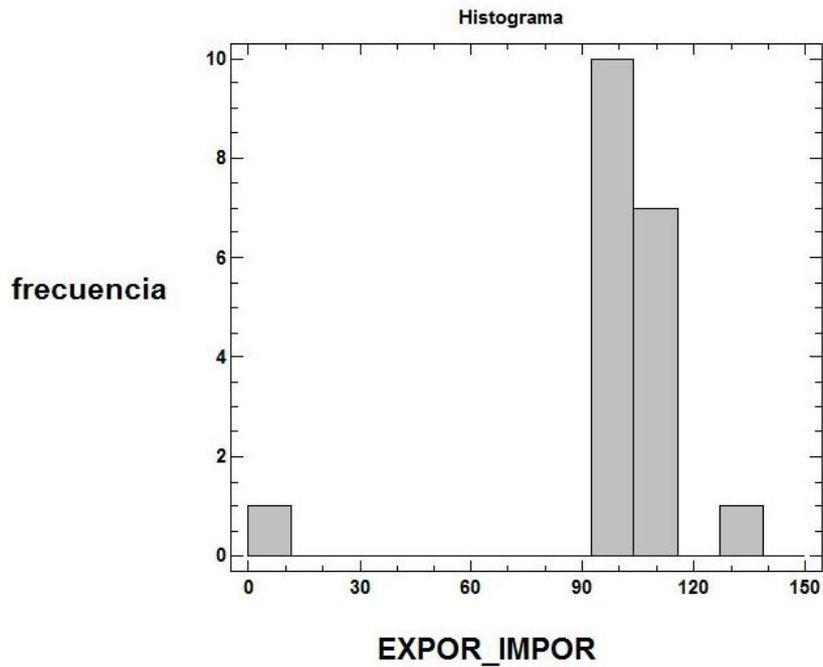
-Tabla IV. 5. Variable Exportación – Importación.

Se ve que en esta variable, la Mediana está muy cerca del promedio. El Sesgo Estandarizado es negativo, con esto se puede decir que presenta una asimétrica hacia la izquierda. Por otra lado la Curtosis Estandarizada al ser positivo presenta una campana Leptocurtica.

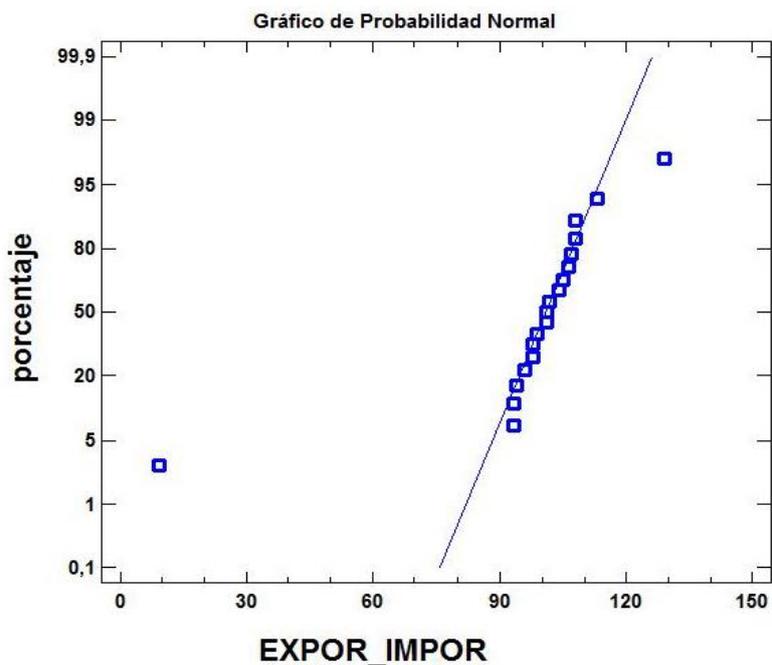


-Figura IV. 22. Gráfico de Caja y Bigotes Variable Exportación e Importación.

En el gráfico se aprecia claramente que la Media y la Mediana está muy cerca.

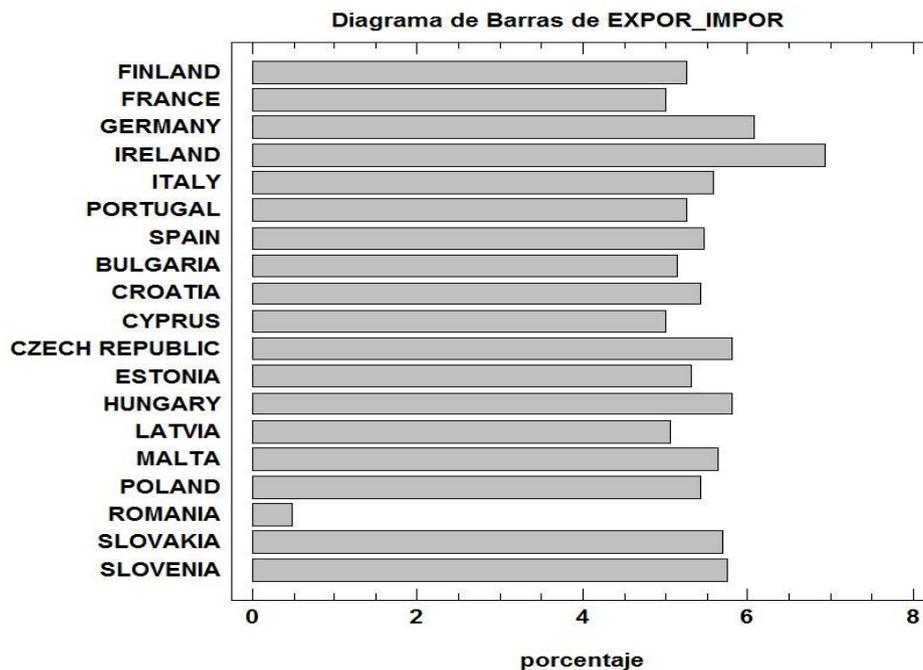


-Figura IV. 23. Histograma Variable Exportación e Importación.



-Figura IV. 24. Gráfico de Probabilidad Normal Variable Exportación e Importación.

Como se puede ver en el gráfico, todos los puntos siguen una distribución normal ya que están encima de la recta, quitados de dos puntos que serían valores anómalos. Estos valores anómalos sería el país Rumania como país que menos exportaciones e importaciones realiza con gran diferencia al resto y Irlanda como mayor país que más exportaciones e importaciones realiza diferenciándose del resto.



-Figura IV. 25. Diagrama de Barras Variable Exportación e Importación frente a los 19 países.

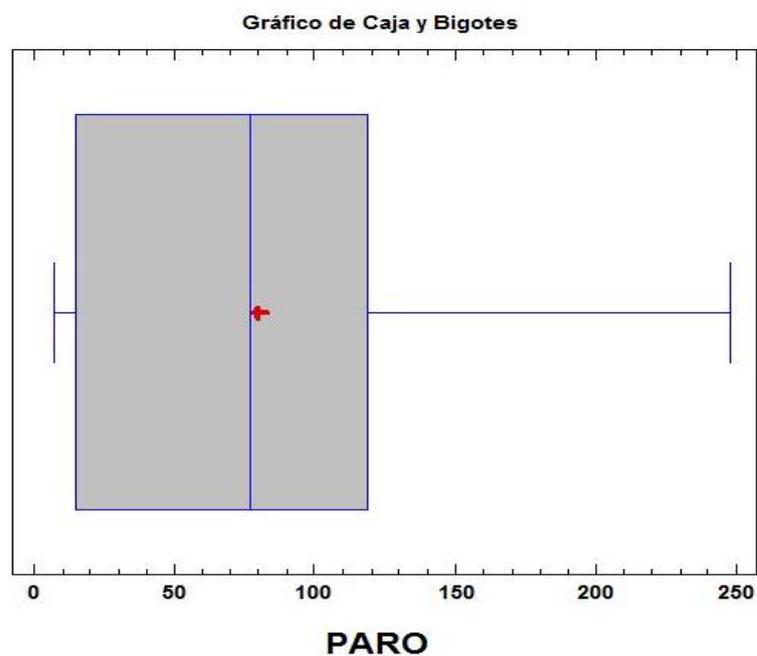
En este diagrama se ve claramente que Rumania es el país que menos exportaciones e importaciones realiza con gran diferencia frente al resto y que más o menos el resto de los 19 países tiene el mismo porcentaje, quitado de Irlanda que sería el más destacado, realizando más número de exportaciones e importaciones.

→ Variable **PARO**:

<b>PARO</b>	
Recuento	19
Promedio	80,2632
Mediana	77
Varianza	4029,2
Desviación Estándar	63,476
Mínimo	7
Máximo	248
Rango	241
Sesgo Estandarizado	1,60927
Curtosis Estandarizada	1,01303

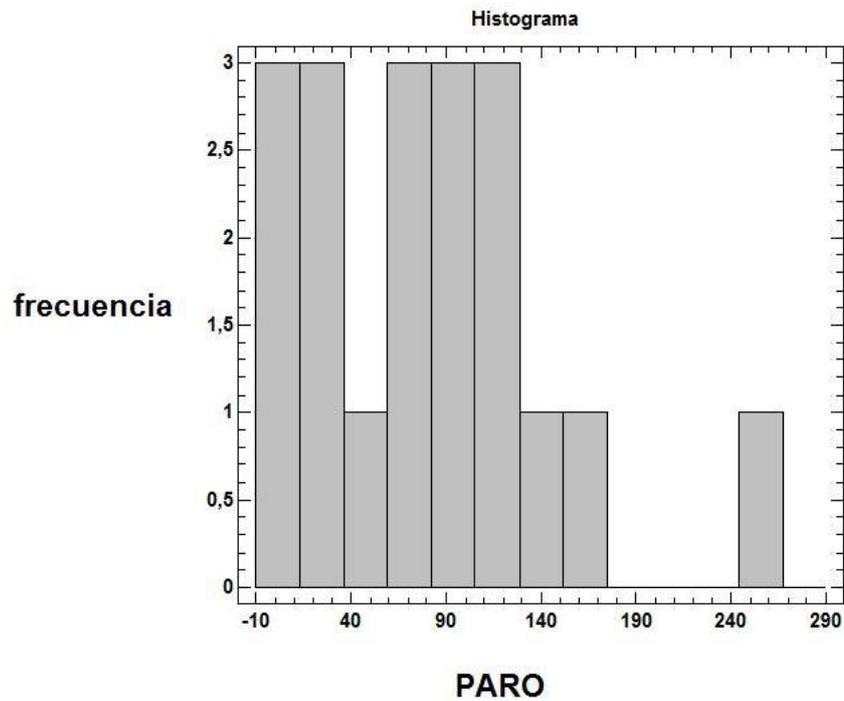
-Tabla IV. 6. Variable Paro.

En la variable Paro la Mediana está cerca del Promedio. El Sesgo Estandarizado es positivo, es decir, esta variable presenta una asimétrica hacia la derecha. La Curtosis presenta una campana Leptocurtica.

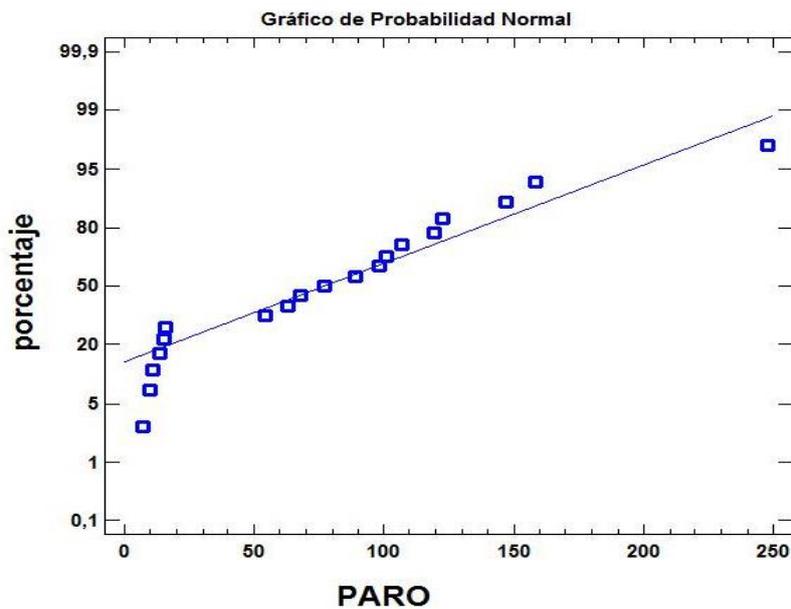


-Figura IV. 26. Gráfico de Caja y Bigotes Variable Paro.

En el grafico se puede ver que la Media está cerca de la Mediana.

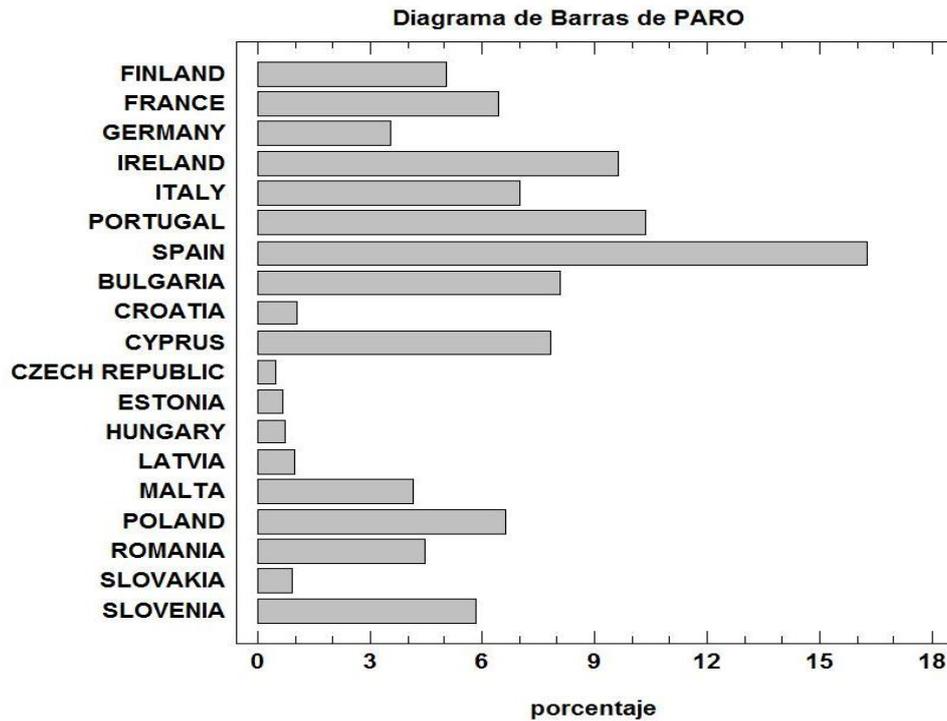


-Figura IV. 27. Histograma Variable Paro.



-Figura IV. 28. Gráfico de Probabilidad Normal Variable Paro.

Se ve que en el gráfico no todos los puntos están muy cerca o por encima de la recta, es decir, hay bastantes puntos anómalos por esto se puede decir que no sigue una distribución normal.



*-Figura IV. 29. Diagrama de Barras Variable Paro frente a los 19 países.*

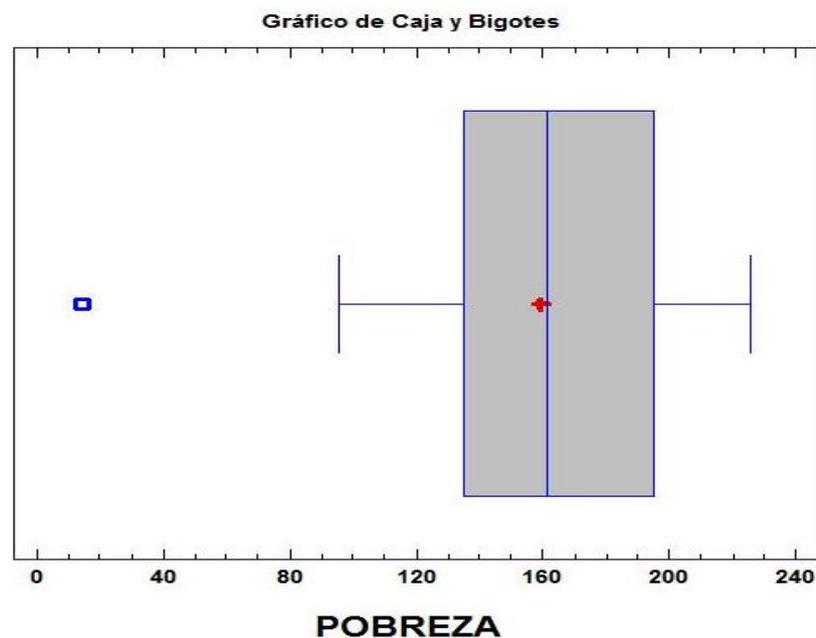
En este se puede observar qué país desafortunadamente tiene claramente más tasa de paro es España. Esto es alarmante ya que siendo el país que más tasa de paro tiene es un factor muy negativo para la producción de vehículos en el país, porque si hay mucha gente parada, es decir, sin trabajo esto propiciara que baje el porcentaje de gente que quiera un vehículo nuevo ya que no se lo podrán permitir y a la vez que las industrias de vehículos produzcan y matriculen menos vehículos. El país que va seguido de España es Portugal e Ireland. Todo lo contrario, pasa con Czech Republic, Estonia e Hungary y que serías los países con índices más bajos de paro. Algo muy positivo para estos países.

→ Variable **POBREZA**:

<b>POBREZA</b>	
Recuento	19
Promedio	159,368
Mediana	161
Varianza	2359,69
Desviación Estándar	48,5766
Mínimo	14
Máximo	226
Rango	212
Sesgo Estandarizado	-2,55642
Curtosis Estandarizada	3,07394

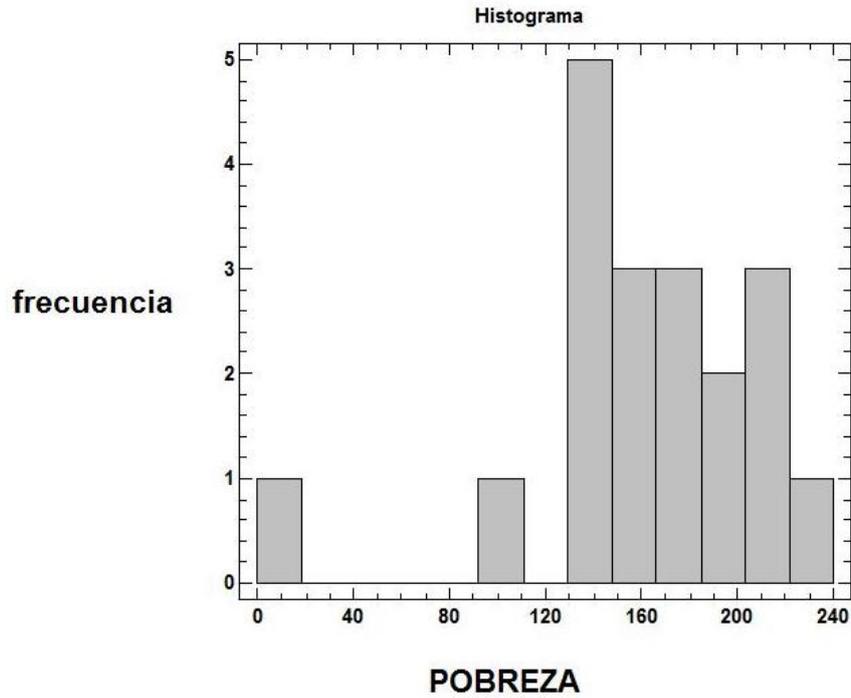
-Tabla IV. 7. Variable Pobreza.

En la tabla anterior de la variable Pobreza, la Mediana está muy cerca del Promedio. En este caso es 132. El Sesgo es negativo, entonces presenta una asimétrica hacia la izquierda. El coeficiente Curtosis es positivo, es decir, presenta una campana más en forma de cono (Leptocúrtica).

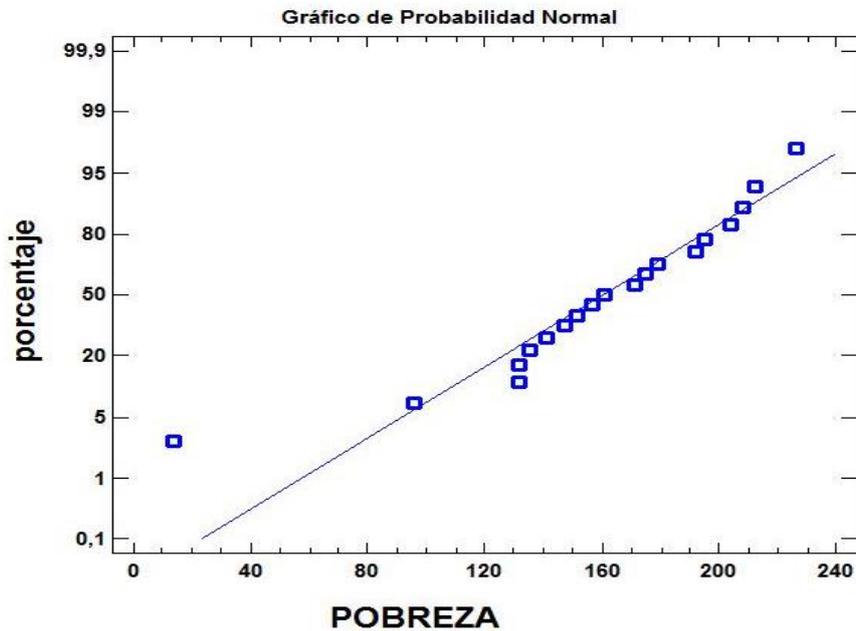


-Figura IV. 30. Gráfico de Caja y Bigotes Variable Pobreza.

En este gráfico la Media y Mediana están en el mismo punto.

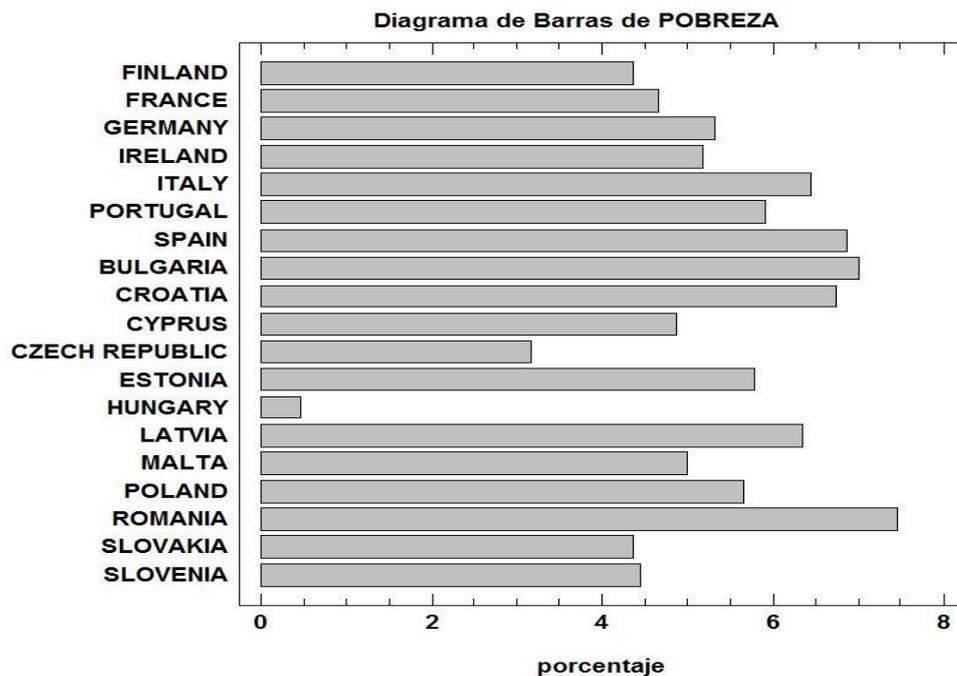


-Figura IV. 31. Histograma Variable Pobreza.



-Figura IV. 32. Gráfico de Probabilidad Normal Variable Pobreza.

En el gráfico es donde tenemos que fijarnos en la distribución que sigue. Parece que sigue una distribución normal, quitado de algún valor anómalo.



-Figura IV. 33. Diagrama de Barras Variable Pobreza frente a los 19 países.

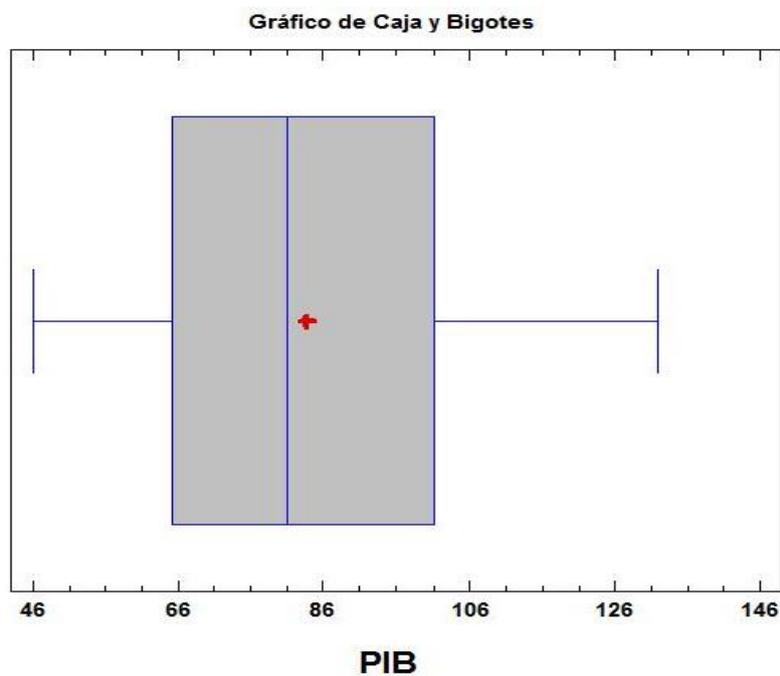
En el Diagrama de la variable se puede destacar claramente que el país con más porcentaje de pobreza es Hungary porque a menos porcentaje el país presenta más tasa de pobreza. Le sigue Czech Republic. Los países con menos tasa de pobreza son Slovakia, Bulgaria e Spain. Se considera que están en riesgo de pobreza aquellas personas que viven en hogares cuya renta es inferior al 60% de la renta mediana de su país o territorio, es decir, cuyos ingresos están por debajo del llamado "Umbral de pobreza".

→ Variable **PIB**:

<b>PIB</b>	
Recuento	19
Promedio	83,4737
Mediana	81
Varianza	560,374
Desviación Estándar	23,6722
Mínimo	46
Máximo	132
Rango	86
Sesgo Estandarizado	0,905387
Curtosis Estandarizada	-0,361204

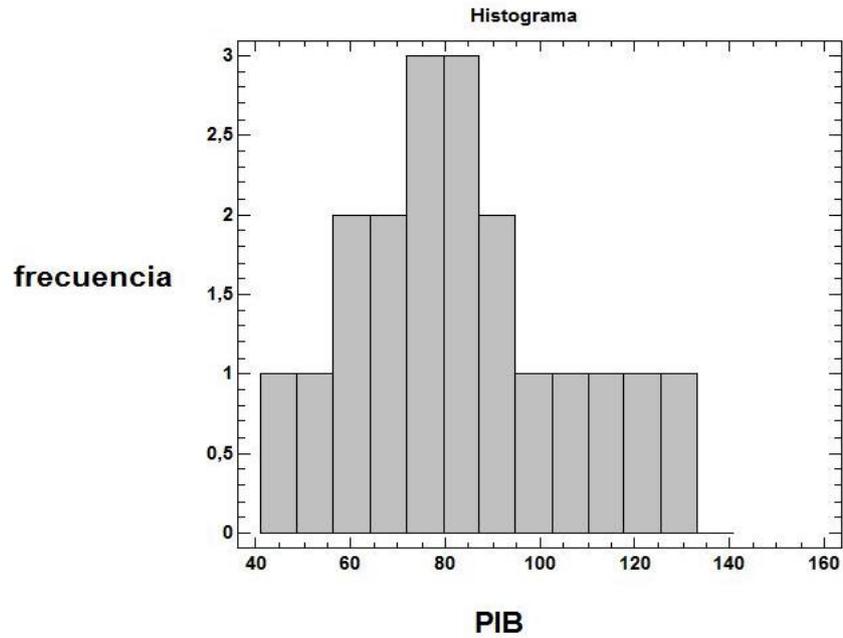
-Tabla IV. 8. Variable PIB.

El Promedio está muy cerca de la Mediana. El Sesgo es positivo, es decir, presenta una asimétrica hacia la derecha. La Curtosis da un valor negativo, es decir, es una campana más plana (Platicúrtica).

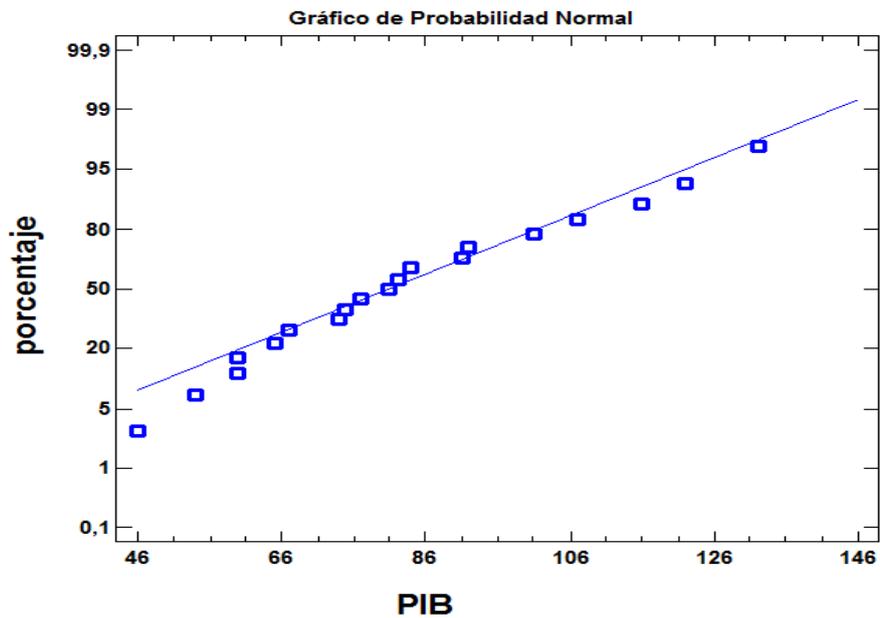


-Figura IV. 34. Gráfico de Caja y Bigotes Variable PIB.

En el grafico se ve que la Media está cerca de la Mediana.

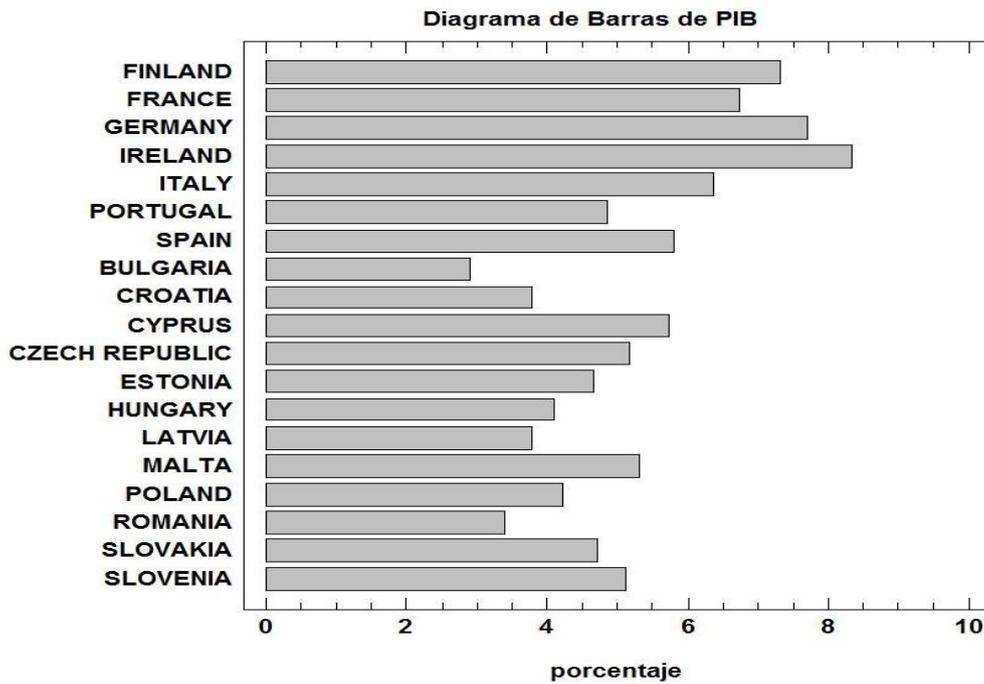


-Figura IV. 35. Histograma Variable PIB.



-Figura IV. 36. Gráfico de Probabilidad Normal Variable PIB.

En el gráfico la variable PIB no sigue una distribución normal porque hay bastantes puntos que no están por encima de la recta.



*-Figura IV. 37. Diagrama de barras Variable PIB frene a los 19 países.*

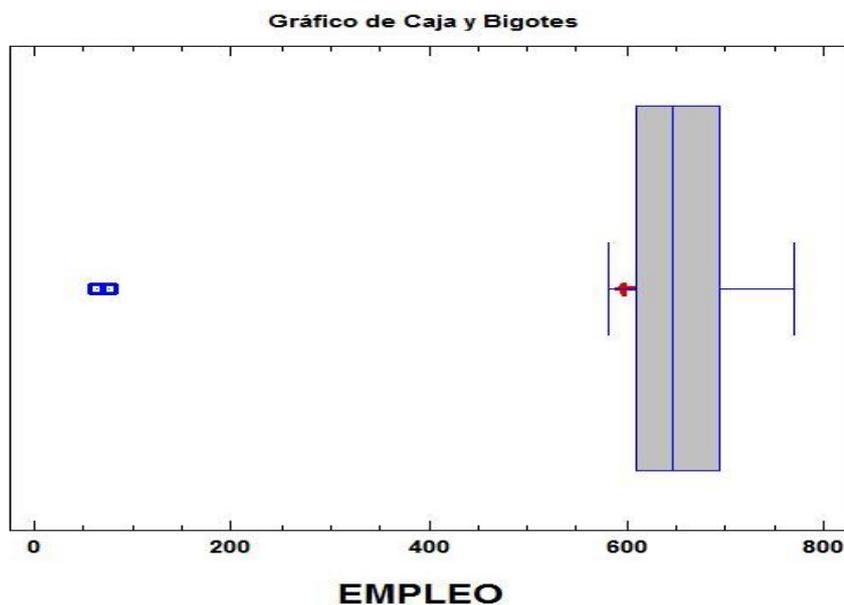
En el anterior grafico se observar que los países con gran porcentaje de producto interior bruto son Ireland, Germany y Finland seguidos de France e Italy. Por otro lado los que más bajo porcentaje tienen son Bulgaria, Romania y Croatia. España tiene un porcentaje normal de tasa de PIB.

→ Variable **EMPLEO**:

<b>EMPLEO</b>	
Recuento	19
Promedio	599,053
Mediana	648
Varianza	37139,7
Desviación Estándar	192,717
Mínimo	63
Máximo	769
Rango	706
Sesgo Estandarizado	-4,42517
Curtosis Estandarizada	4,83793

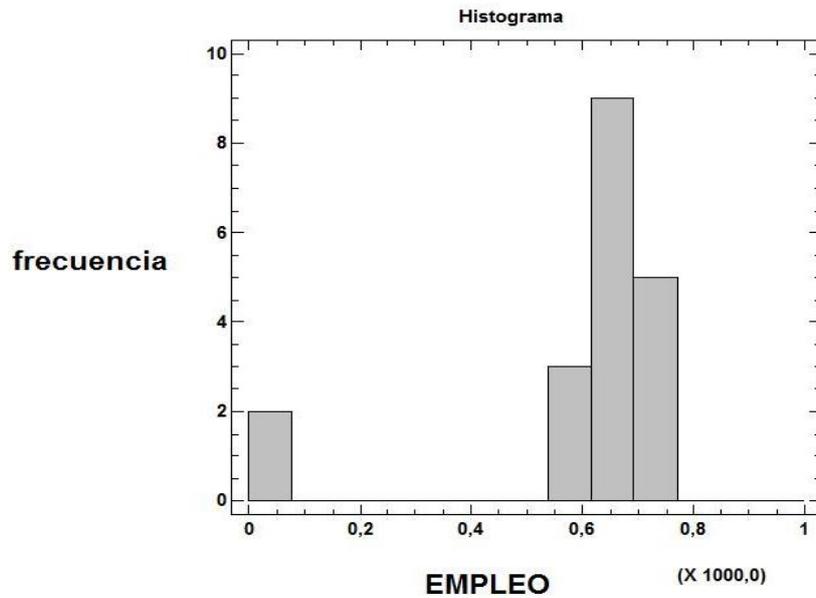
-Tabla IV. 9. Variable Empleo.

El variable Empleo, sigue siendo como la mayoría variables ya que el Promedio y la Mediana están bastante cerca los valores. El Sesgo Estandarizado presenta un valor negativo, entonces presenta una asimétrica hacia la izquierda. Por otro lado la Curtosis al ser positiva presenta una campana Leptocúrtica.

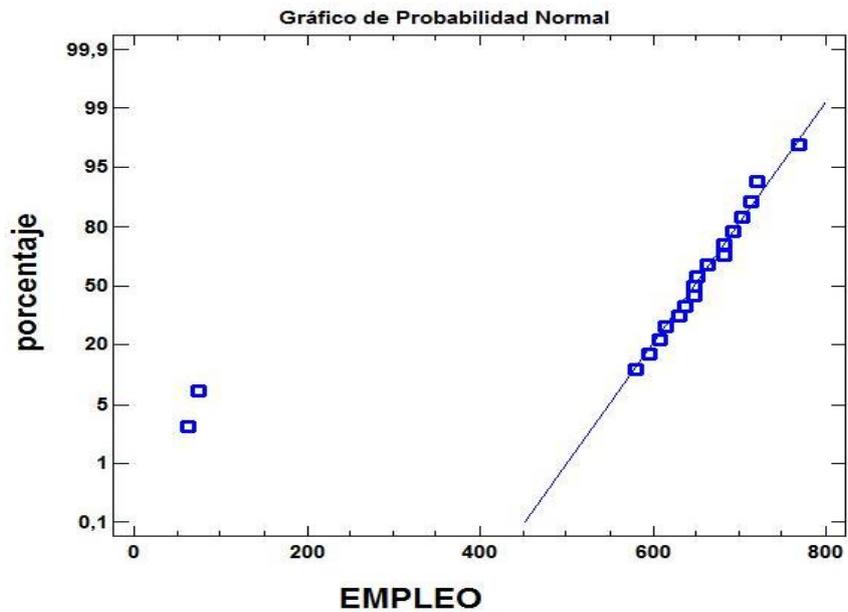


-Figura IV. 38. Gráfico de Caja y Bigotes Variable Empleo.

En este gráfico se ve que la Media y la Mediana están algo lejos. La tasa de empleo presenta en los gráficos dos puntos anómalos. Los cuales son Bulgaria y Finlandia que son los países con menos tasa de empleo presentan sobre el resto.

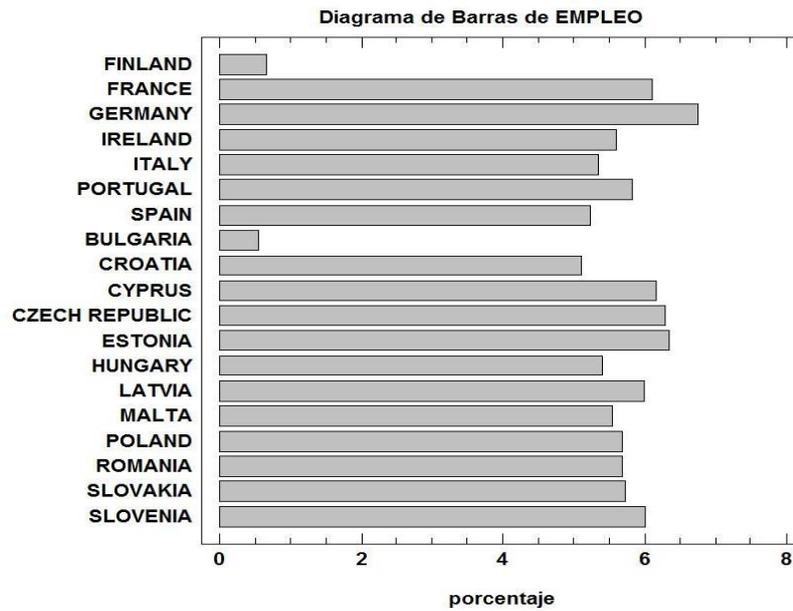


-Figura IV. 39. Histograma Variable Empleo.



-Figura IV. 40. Gráfico de Probabilidad Normal Variable Empleo.

En el grafico se puede observar claramente que, quitado de dos puntos anómalos, la variable sigue una distribución normal.



-Figura IV. 41. Histograma Variable Empleo frente a los 19 países.

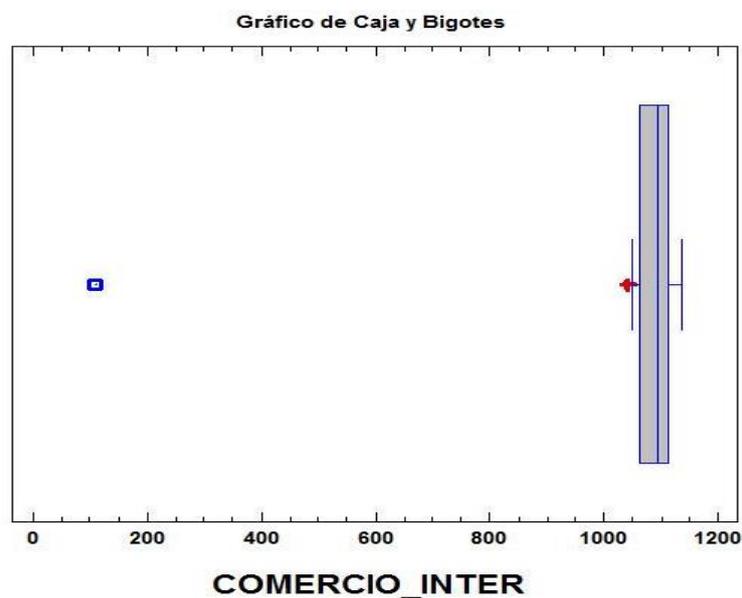
En este diagrama se puede observar la gran diferencia entre los países con más tasa de empleo los cuales son la gran parte y los que menos como Finland y Bulgaria. El país que más tasa de empleo tiene es Germany.

→ Variable **COMERCIO INTERNACIONAL**:

<b>COMERCIO INTERNACIONAL</b>	
Recuento	19
Promedio	1040,32
Mediana	1095
Varianza	51353,3
Desviación Estándar	226,613
Mínimo	110
Máximo	1136
Rango	1026
Sesgo Estandarizado	-7,60583
Curtosis Estandarizada	16,4439

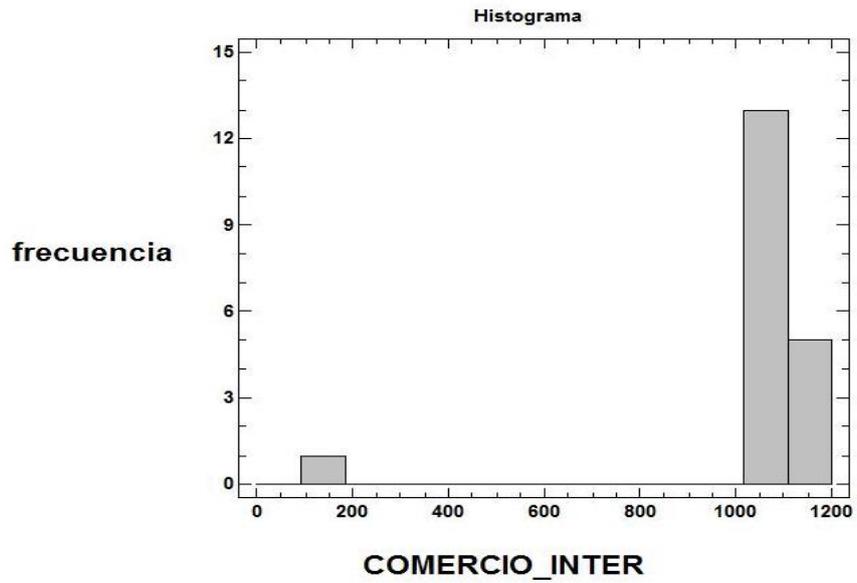
-Tabla IV. 10. Variable Comercio Internacional.

Como en la mayoría de las variables, en la variable Comercio Internacional también se repite que el valor Promedio está bastante cerca del valor Mediana. El Sesgo es negativo, es decir, una asimétrica hacia la izquierda. En cambio, la Curtosis es positiva, entonces podemos decir que la campana tiene mas forma de cono (Leptocúrtica).

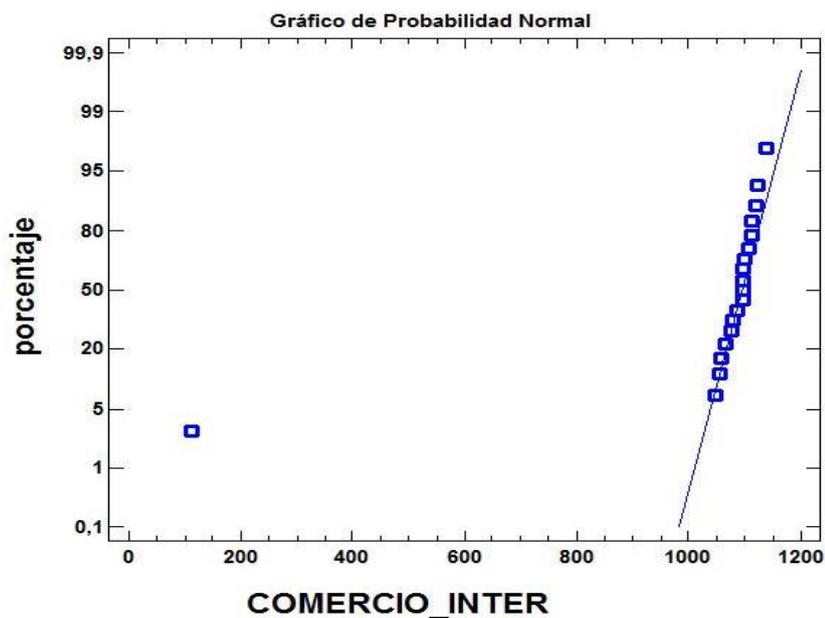


-Figura IV. 42. Gráfico de Caja y Bigotes Variable Comercio Internacional.

En el gráfico de Caja y Bigotes se ve que la Media y Mediana no están tan cerca.

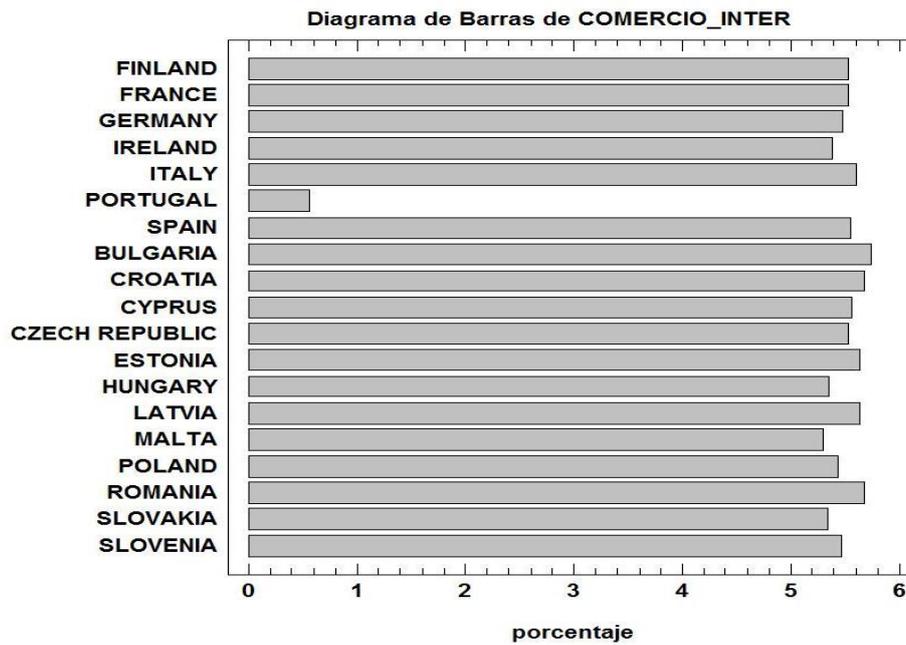


-Figura IV. 43. Histograma Variable Comercia Internacional.



-Figura IV. 44. Gráfico de Probabilidad Normal Variable Comercio Internacional.

Se ve en el gráfico que esta variable sigue una distribución normal porque quitado de un punto anómalo todos están muy cerca de la recta o por encima de ella.



*-Figura IV. 45. Diagramas de Barras Variable Comercio Internacional frente a los 19 países.*

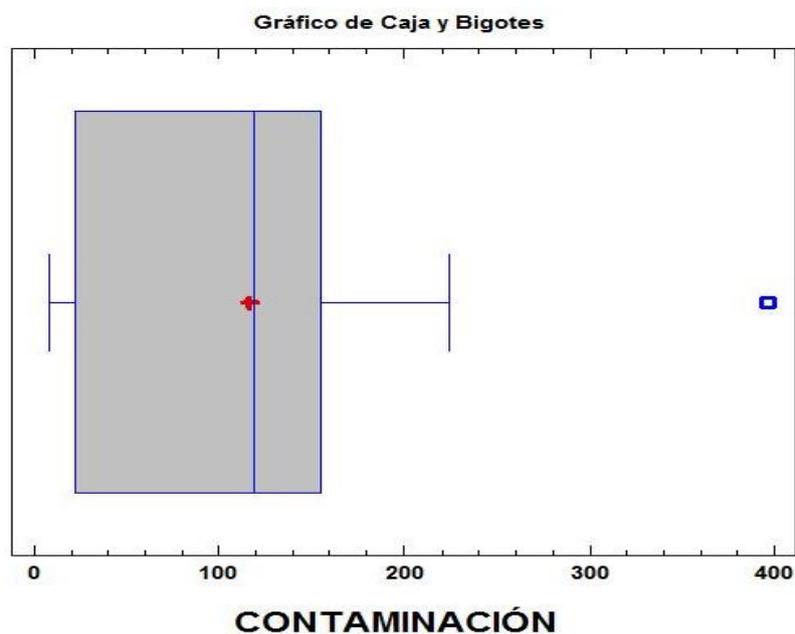
En el anterior diagrama se puede destacar que de los 19 países de la Unión Europea estudiados, más o menos presentan el mismo porcentaje, es decir, la diferencia es mínima. Quitado de Portugal que se puede ver como tiene muy bajo porcentaje frente al resto.

→ Variable **CONTAMINACIÓN**:

<b>CONTAMINACIÓN</b>	
Recuento	19
Promedio	116,211
Mediana	118
Varianza	8969,4
Desviación Estándar	94,7069
Mínimo	8
Máximo	396
Rango	388
Sesgo Estandarizado	2,34167
Curtosis Estandarizada	2,7125

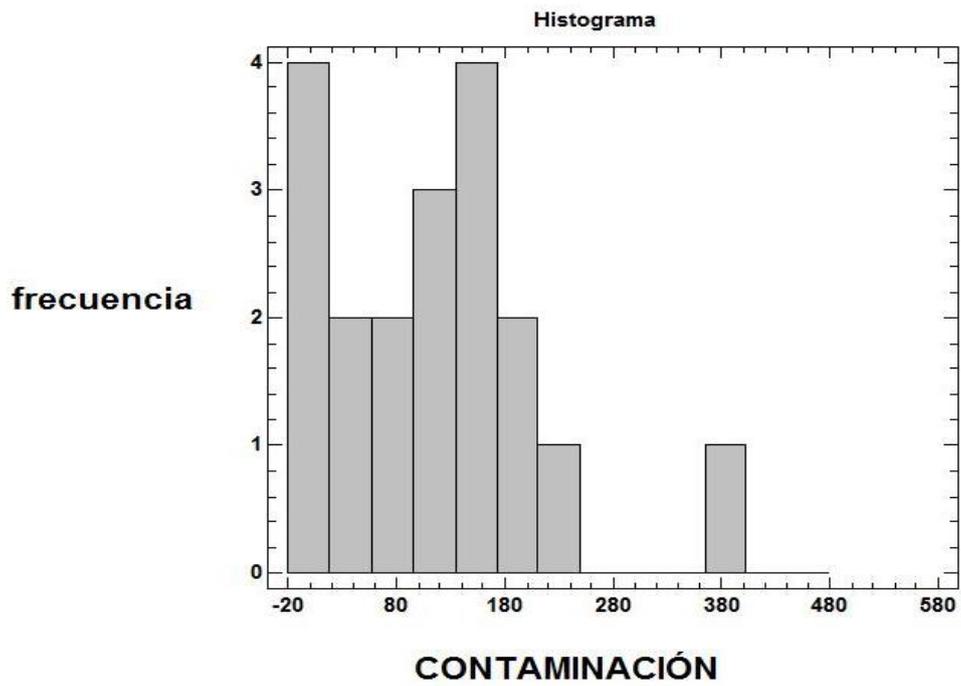
-Tabla IV. Variable Contaminación.

En la variable Contaminación podemos ver que el valor Promedio y el valor Mediana están muy cerca. Tanto el Sesgo como la Curtosis son positivo, es decir, presenta una asimétrica hacia la derecha y una campana tipo cono (Leptocúrtica).

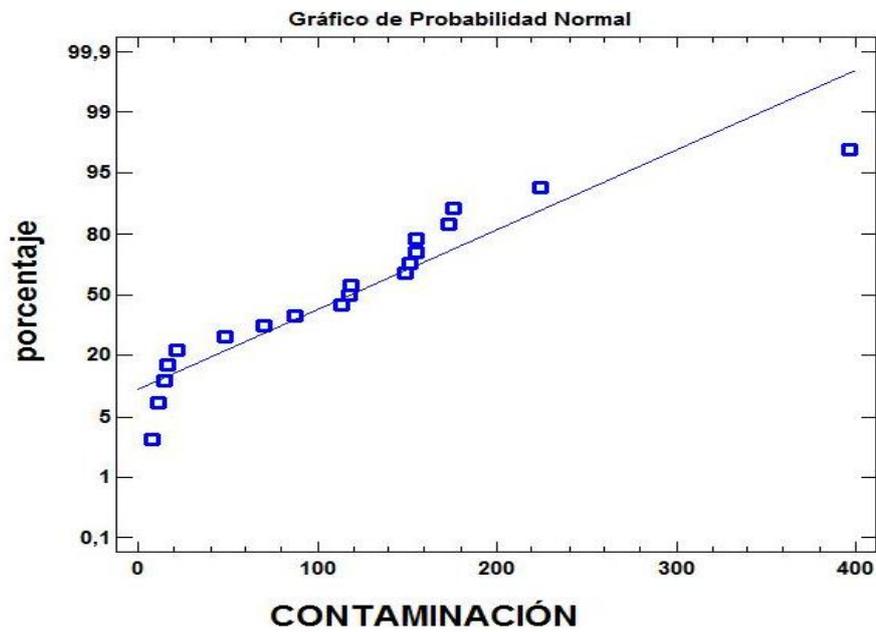


-Figura IV. 46. Gráfico de Caja y Bigotes Variable Contaminación.

Para terminar con la última variable, se puede destacar que la Mediana y Media se encuentran en el mismo punto.

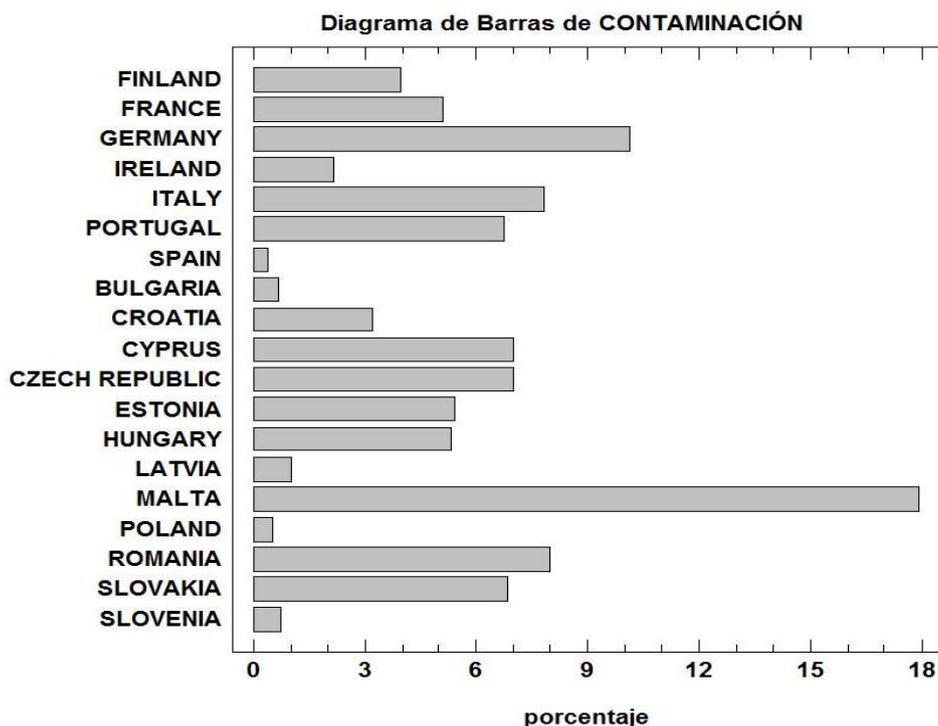


-Figura IV. 47. Histograma Variable Contaminación.



-Figura IV. 48. Gráfico de Probabilidad Normal Variable Contaminación.

En el gráfico de Probabilidad Normal se ve que no sigue una distribución normal, ya que muchos puntos no están cerca de la recta.



-Figura IV. 49. Diagrama de Barras Variable Contaminación.

En este último diagrama de barras se puede observar que el país que más alto índice de contaminación tiene es Malta con gran diferencia frente al resto, el país que le sigue en segundo lugar es Germany. También se puede destacar que el país que menos contamina es España, seguido de Poland, Bulgaria y Slovenia.

Sobre España destacar que esto es muy positivo porque cuando menos contaminación tiene país, mucho mejor para todos y para el mundo. Pero en la variable de vehículos matriculados, España es el cuarto país que más porcentaje tiene de vehículos matriculados y lo que quiere decir esto es que no es preocupante porque la tasa de contaminación está muy por debajo de mediana.

## 2. Análisis Bivariante.

En esta parte se va a realizar el análisis de multicolinealidad mediante las 3 pruebas para comprobar si existe o no el problema de multicolinealidad. Si existe, se solucionará el problema.

## 2.1 Identificación.

Inicialmente, se observa en la matriz de correlación inicial Tabla IV. 3, en la que todos los valores que sean mayores que  $|0.7|$  indican que hay problema de multicolinealidad. Por la tanto, el siguiente paso se procederá a la eliminación de las variables donde ocurra esto.

	MOTORIZACIÓN	EXPOR_IMPOR	PARO	POBREZA	PIB	EMPLEO	COMERCIO_INTER	CONTAMINACIÓN
MOTORIZACIÓN	1	0,456297	0,314791	-0,0423027	0,623735	-0,0365295	0,0624211	0,286306
EXPOR_IMPOR	0,456297	1	0,0488761	-0,397925	0,450534	0,0262552	-0,0482452	-0,106893
PARO	0,314791	0,0488761	1	0,391934	0,305675	-0,172991	-0,290172	-0,270662
POBREZA	-0,0423027	-0,397925	0,391934	1	-0,172878	-0,125504	-0,029873	-0,158091
PIB	0,623735	0,450534	0,305675	-0,172878	1	0,0989998	0,0258302	0,17401
EMPLEO	-0,0365295	0,0262552	-0,172991	-0,125504	0,0989998	1	-0,113902	0,27291
COMERCIO_INTER	0,0624211	-0,0482452	-0,290172	-0,029873	0,0258302	-0,113902	1	-0,121843
CONTAMINACIÓN	0,286306	-0,106893	-0,270662	-0,158091	0,17401	0,27291	-0,121843	1

-Tabla IV. 12. Matriz Correlación.

A continuación, para la eliminación de las variables se calcula la matriz de correlación inversa, donde dependiendo del valor de la diagonal, se sabrá que variables se deben de eliminar porque presenten problemas.

INVERSA	MOTORIZACIÓN	EXPOR_IMPOR	PARO	POBREZA	PIB	EMPLEO	COMERCIO_INTER	CONTAMINACIÓN
MOTORIZACIÓN	2,557348988	-1,064451115	-0,853128	-0,2681337	-0,7150265	0,25662303	-0,556867084	-1,132729263
EXPOR_IMPOR	-1,064451115	2,045374653	0,477109	0,65017986	-0,4524229	-0,0822286	0,436141907	0,909625782
PARO	-0,853128334	0,477108962	2,2611257	-0,6502742	-0,6968957	0,16462814	0,869946871	0,986788753
POBREZA	-0,268133666	0,65017986	-0,650274	1,57748563	0,32274662	-0,0268237	-0,085308327	0,160414811
PIB	-0,715026483	-0,452422894	-0,696896	0,32274662	2,00197685	-0,2437142	-0,285539156	-0,297887206
EMPLEO	0,256623033	-0,082228619	0,1646281	-0,0268237	-0,2437142	1,15720466	0,128572208	-0,299682772
COMERCIO_INTER	-0,556867084	0,436141907	0,8699469	-0,0853083	-0,2855392	0,12857221	1,402459625	0,613507901
CONTAMINACIÓN	-1,132729263	0,909625782	0,9867888	0,16041481	-0,2978872	-0,2996828	0,613507901	1,922359589

-Tabla IV. 13. Matriz de Correlación Inversa.

## *Análisis y predicción de la variabilidad del número de vehículos matriculados a nivel Europeo*

En primer lugar, como se puede observar claramente en la Tabla IV. 12. no hay ningún valor mayor que |0,7| entonces no habrá que eliminar ninguna variable ya que no existe problema de multicolinealidad por correlación entre pares de variables.

Seguidamente, como se puede comprobar en la Tabla IV. 13. no hay ningún valor de diagonal que sea >10 por lo tanto, no hay problema de multicolinealidad del tipo una variable con todas las demás.

Para terminar, se realizará la tercera prueba para ver si también existe o no multicolinealidad en este método, calculándose el índice de acondicionamiento (I.C)

$$I.C = \sqrt{\frac{AUTOVALOR\ MAX}{AUTOVALOR\ MIN}} = \sqrt{\frac{2,2329}{0,185467}} = 3,469774998$$

Después de calcular el valor del índice de acondicionamiento, se refleja claramente que en tercer método no existe problema de multicolinealidad ya que el valor no es >10.

### 3. Regresión Múltiple, primera estructura.

En este apartado se intenta explicar de qué depende el número de producción de vehículos en la Unión Europea a través de la regresión múltiple lineal. Se explica que influencia tienen varias variables independientes llamadas  $X_i$  frente a una variable dependiente llamada  $Y$ .

#### 3.1 Modelo Teórico Lineal.

En este trabajo la variable a explicar será los vehículos matriculados en la UE ( $Y$ ), la cual se medirá en el nº de vehículos matriculados. Por otro lado, para el modelo teórico se utilizará las siguientes variables explicativas ( $X_i$ ):

- ❖  $X_1$  = Motorización.
- ❖  $X_2$  = Exportaciones e Importaciones.
- ❖  $X_3$  = Tasa de Paro.
- ❖  $X_4$  = Tasa de Pobreza.
- ❖  $X_5$  = PIB per Cápita.
- ❖  $X_6$  = Tasa de Empleo.
- ❖  $X_7$  = Comercio Internacional.
- ❖  $X_8$  = Contaminación, suciedad y otros problemas ambientales.

$$\text{MATRICULADOS (Y)} = \beta_0 + \beta_1 \cdot \text{Motorización} + \beta_2 \cdot \text{Exportación e Importación} + \beta_3 \cdot \text{Paro} + \beta_4 \cdot \text{Pobreza} + \beta_5 \cdot \text{PIB} + \beta_6 \cdot \text{Empleo} + \beta_7 \cdot \text{Comercio Internacional} + \beta_8 \cdot \text{Contaminación} + U.$$

#### 3.2 Estimación de la primera ecuación de la Regresión Múltiple y validación del modelo.

En este punto, se estiman los parámetros del modelo teórico lineal utilizado en el programa Statgraphics. Seguidamente se observa la función que calculada del modelo que nos indica el programa.

$$\text{MATRICULADOS} = -1,43395\text{E}6 + 450,83 \cdot \text{MOTORIZACIÓN} - 1221,14 \cdot \text{EXPOR\_IMPOR} - 564,056 \cdot \text{PARO} + 1873,02 \cdot \text{POBREZA} + 10231,5 \cdot \text{PIB} + 476,167 \cdot \text{EMPLEO} + 169,685 \cdot \text{COMERCIO\_INTER} + 185,624 \cdot \text{CONTAMINACIÓN}.$$

### 3.2.1 Análisis de la Significación del Modelo y Variables.

En esta apartado se procederá a ver los P-valores de todas las variables dependientes (explicativas). Si estos P-valores son  $>0,05$  esto nos indicara que los parámetros del modelo no son significativos y, por lo contrario, si son  $<0,05$  sí que son significativos. También se comprobará el valor de la constante con la misma valoración para saber si también es significativa o no.

Parámetro	Estimación	Error	Estadístico	Valor-P
		Estándar	T	
CONSTANTE	-1,43395E6	1,10837E6	-1,29375	0,2248
MOTORIZACIÓN	450,83	1602,31	0,281362	0,7842
EXPOR_IMPOR	-1221,14	6721,74	-0,18167	0,8595
PARO	-564,056	2576,86	-0,218893	0,8311
POBREZA	1873,02	2812,5	0,665962	0,5205
PIB	10231,5	6501,72	1,57365	0,1466
EMPLEO	476,167	607,189	0,784215	0,4511
COMERCIO_INTER	169,685	568,458	0,298499	0,7714
CONTAMINACIÓN	185,624	1592,48	0,116563	0,9095

-Tabla IV. 14. Resumen Análisis Regresión Múltiple Lineal.

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,447E12	8	1,80876E11	0,85	0,5835
Residuo	2,12985E12	10	2,12985E11		
Total (Corr.)	3,57685E12	18			

-Tabla IV. 15. Resumen Análisis Varianza Regresión Múltiple Lineal.

- **Modelo Inicial: R-cuadrado = 40,4547%**

En la tabla IV. 15. se puede ver que el P-valor del modelo es  $>0,05$ , lo que indica que el modelo no es significativo, lo mismo ocurre con los P-valores de todas las variables y la constante, por lo que nada es significativo en esta primera estimación (tabla IV. 14).

Para que el modelo sea significativo se eliminarán los puntos anómalos (atípicos), en el caso de que haya y se verá que ocurre, es decir, si no sigue siendo significativo o si finalmente es significativo.

A continuación, se mira en la Tabla IV. 16A. para comprobar la existencia de Residuos Atípicos.

<b>Fila</b>	<b>Y</b>	<b>Y</b>		<b>Residuo</b>
		<i>Predicha</i>	<i>Residuo</i>	<i>Estudentizado</i>
3	1,6344E6	778031,	856370,	3,11
4	66718,0	654604,	-587886,	-2,60

-Tabla IV. 16A. Resumen Residuos Atípicos.

Como se puede observar, la fila 3 es un residuo atípico. Seguidamente se buscara que país europeo es la fila 3 y posteriormente será eliminado para comprobar si el modelo es significativo. La fila 3 corresponde a Alemania.

Una vez se elimine Alemania aparece otro residuo atípico como se puede ver en la Tabla IV. 16B.

<b>Fila</b>	<b>Y</b>	<b>Y</b>		<b>Residuo</b>
		<i>Predicha</i>	<i>Residuo</i>	<i>Estudentizado</i>
2	1,04898E6	371942,	677040,	3,21

-Tabla IV. 16B. Resumen Residuos Atípicos.

Por lo tanto, se realizará el mismo proceso que se ha realizado anteriormente, ahora eliminando ahora la fila 2, que corresponde a Francia.

Posteriormente, sale en la Tabla IV. 16C otro residuo atípico pero ahora en la fila 3, que corresponde a Italia. Se realizará el mismo criterio que antes y se eliminara Italia.

<b>Fila</b>	<b>Y</b>	<b>Y</b>		<b>Residuo</b>
		<i>Predicha</i>	<i>Residuo</i>	<i>Estudentizado</i>
3	815213,	294461,	520752,	6,49

-Tabla IV. 16C. Resumen Residuos Atípicos.

*Análisis y predicción de la variabilidad del número de vehículos matriculados a nivel Europeo*

Una vez eliminada la fila 3, es decir, Italia. Se ve que continúan saliendo filas con residuos atípicos, en esta última Tabla IV. 16D la fila 4, España. La pregunta que se formula ahora es, si se debe proceder con la eliminación de los datos de España.

		Y		Residuo
Fila	Y	Predicha	Residuo	Estudentizado
3	53375,0	27781,2	25593,8	2,13

*-Tabla IV. 16D. Resumen Residuos Atípicos.*

Si se elimina España, lo que ocurre es que sigue repitiéndose, es decir, siguen habiendo puntos anómalos.

Al eliminar tanto los países que generaban puntos anómalos, como las variables no significativas y la constante, quedan los resultados que se muestran en las tablas.

Parámetro	Estimación	Error	Estadístico	Valor-P
		Estándar	T	
MOTORIZACIÓN	-109,244	37,777	-2,8918	0,0147
PIB	936,807	199,162	4,70373	0,0006

*-Tabla IV. 17. Resumen Análisis Regresión Múltiple Final.*

Finalmente, se quedan dos variables explicativas (Tabla IV. 17), tanto la Motorización como el PIB. Los P-valores de las dos variables son <0,05, esto indica que son significativas.

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,4058E10	2	7,02898E9	34,38	0,0000
Residuo	2,24912E9	11	2,04466E8		
Total	1,63071E10	13			

*-Tabla IV. 18. Resumen Análisis Varianza Regresión Múltiple Final.*

Para terminar, en la Tabla IV. 18. se puede ver un resumen del análisis para el modelo regresión lineal, donde el P-valor que se ve es 0,000. Lo que quiere decir que el modelo, una vez eliminado los países que tenían puntos atípicos ahora sí que es

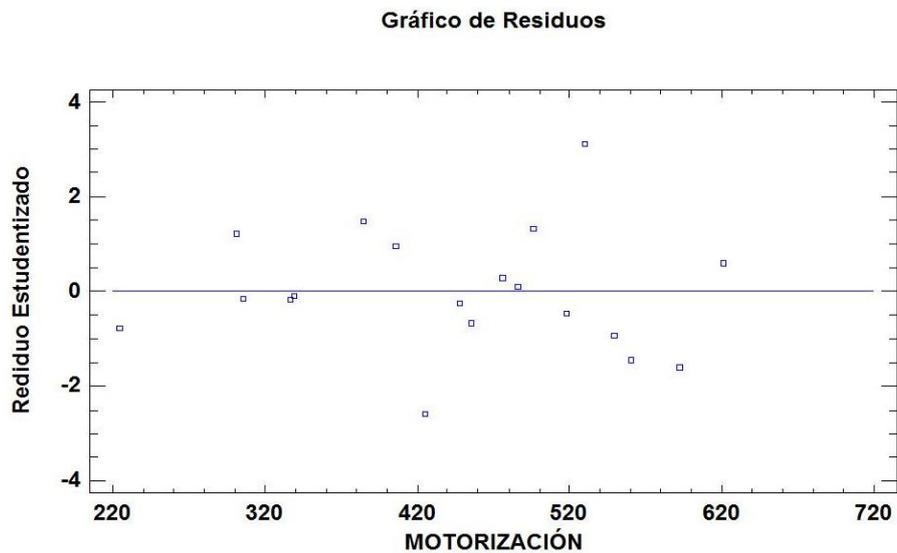
significativo, resultando además un modelo con ajuste muy elevado, por encima del 85%.

- **Modelo final:**  $R^2 = 86,2\%$   
 $R^2$  (ajustado para g.l.) = **84,9%**

$$\text{MATRICULADOS} = -109,244 * \text{MOTORIZACIÓN} + 936,807 * \text{PIB}$$

### 3.2.2 Análisis gráficos de residuos.

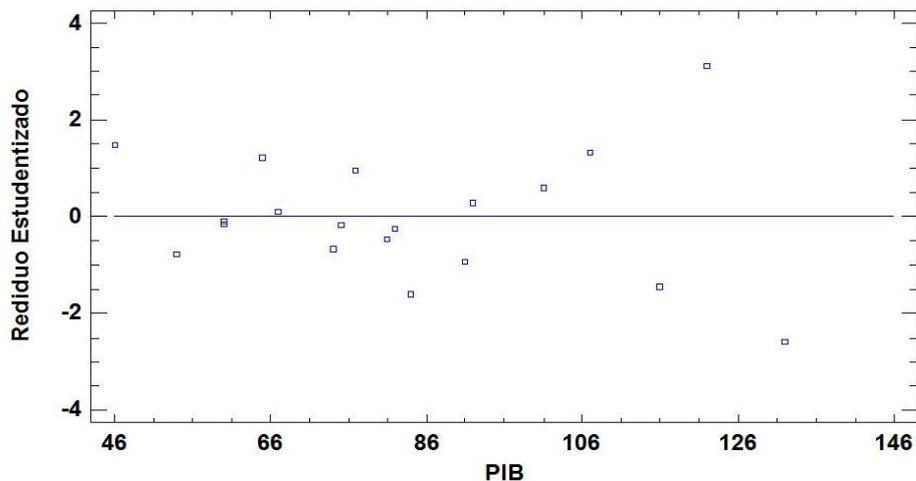
En este punto se exponen los gráficos de todas variables dependientes (explicativa) frente a los residuos. Con estos gráficos se pueden observar o ver la posibilidad de algún tipo de problema.



-Figura IV. 50. Gráfico Residuos vs Motorización Estructura Lineal.

En la Figura IV. 50. a simple vista se puede observar que puede presentar problemas de Heterocedasticidad.

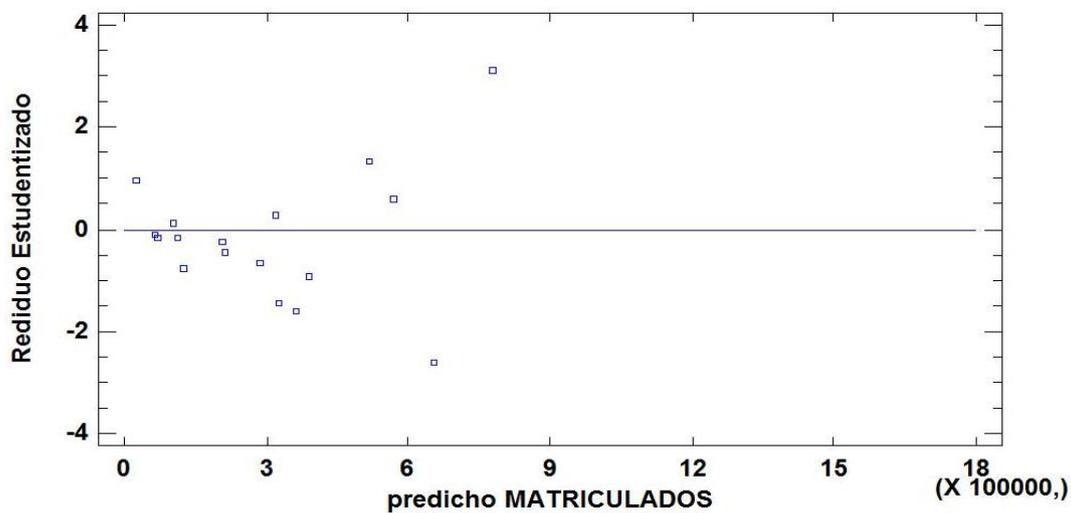
Gráfico de Residuos



-Figura IV. 51. Gráfico Residuos vs PIB Estructura Lineal.

En este Gráfico (Figura IV. 51.), como ocurre en el anterior se puede apreciar que en la variable PIB cabe la posibilidad de que presente Heterocedasticidad.

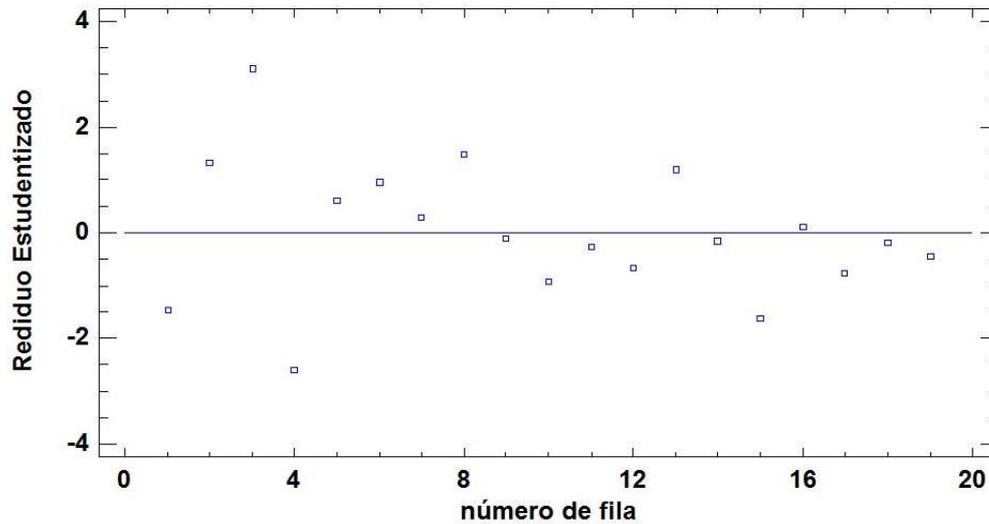
Gráfico de Residuos



-Figura IV. 52. Gráfico Residuos vs Predicho Matriculados Estructura Lineal.

En la Figura IV. 52. como ocurre en los dos gráficos anteriores se puede interpretar problemas de Heterocedasticidad.

Gráfico de Residuos



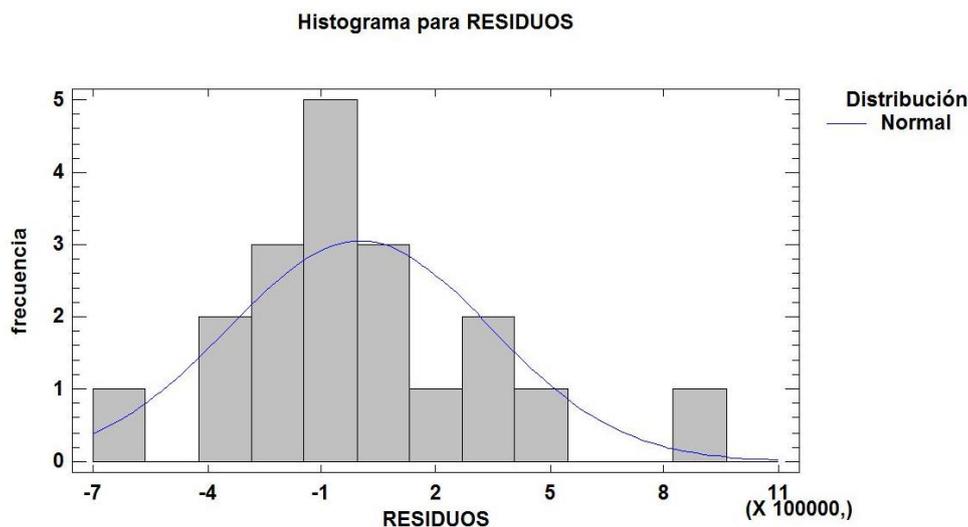
-Figura IV. 53. Gráfico Residuos vs Número de fila Estructura Lineal.

En este último gráfico (Figura IV. 53.) no se observa problema de Autocorrelación ni ningún otro problema.

### 3.2.3 Normalidad de Residuos.

Para estudiar si los residuos siguen una distribución normal, hay dos criterios que se deben de comprobar:

- Representación campana de GAUSS.
- Realización de cuatro test. Se escogerá el valor más restrictivo, es decir, el test con el P-valor más pequeño porque que más dificultad tendrá a la hora de verificar la hipótesis de normalidad ( $H_0$ ).



-Figura IV. 54. Histograma de Residuos Estructura Lineal.

Como se ve en el histograma, este no forma una campana de Gauss perfecta, pero como la mayoría de los residuos están en la media y a medida que nos alejamos de la medida la frecuencia disminuye, supondremos la normalidad de Residuos.

(Figura IV. 54.)

<b>Prueba</b>	<b>Estadístico</b>	<b>Valor-P</b>
Chi-Cuadrado	4,36842	0,885536
Estadístico W de Shapiro-Wilk	0,960337	0,577798
Valor-Z para asimetría	1,06838	0,285347
Valor-Z para curtosis	Datos Insuficientes	

-Tabla IV. 19. Test de Normalidad Lineal.

Una vez se tiene todos los P-valores, cogemos el P-valor más restrictivo que es el Valor-Z para asimetría que es 0.285347. Como este P-valor es  $>0,05$ , se saca la conclusión de que los Residuos siguen una distribución normal, es decir, se acepta  $H_0$ . Los residuos siguen una distribución normal (Tabla IV. 19.)

### 3.2.4 Heterocedasticidad.

Para la comprobación de si tiene el modelo, el problema de heterocedasticidad se calcula un modelo con los residuos al cuadrado frente a las variables dependientes

(explicativas) del modelo. Se obtienen los siguientes resultados que se ven en la Tabla IV. 20.

<b>Parámetro</b>	<b>Estimación</b>	<b>Error</b>	<b>Estadístico</b>	<b>Valor-P</b>
		<b>Estándar</b>	<b>T</b>	
CONSTANTE	2,23399E11	3,36799E11	0,663303	0,5221
MOTORIZACIÓN	1,21772E8	9,14987E8	0,133086	0,8968
PIB	-1,44669E9	4,32966E9	-0,334135	0,7452

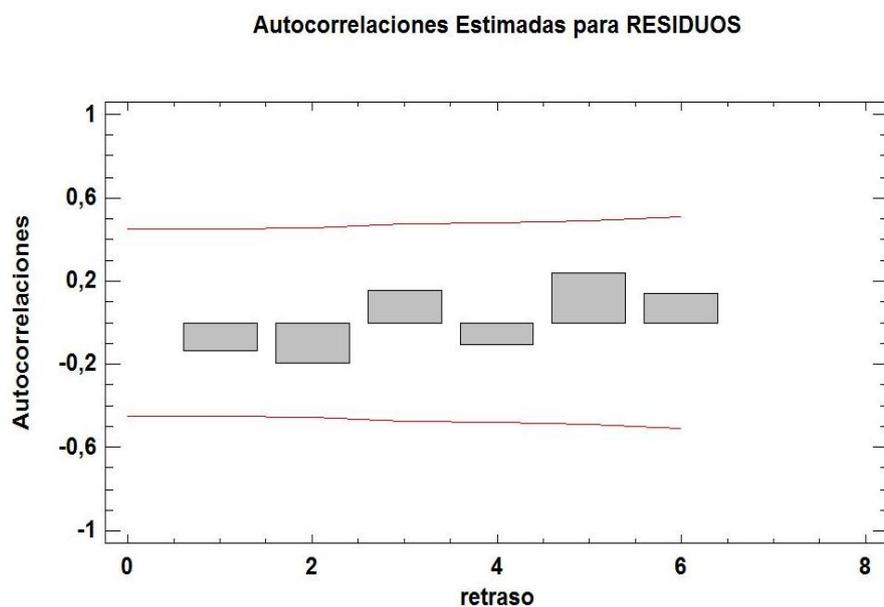
-Tabla IV. 20. Heterocedasticidad Lineal.

Como se ve en la Tabla IV. 20. ninguna de las dos variables explicativas, presenta problemas de heterocedasticidad ya que los P-valores son >0,05.

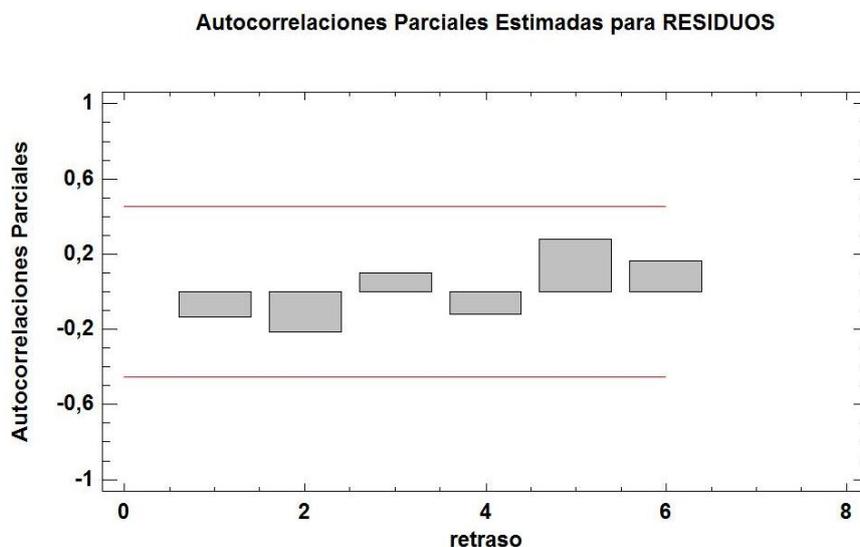
Entonces se acepta  $H_0$ , es decir, varianza constante.

### 3.2.5 Autocorrelación.

En este apartado se representa el gráfico de función de autocorrelación simple (FAS) (Figura IV. 55.) y el gráfico de función de autocorrelación parcial (FAP) (Figura IV. 56.) para encontrar en estos gráficos la presencia de autocorrelación.



-Figura IV. 55. Gráfico FAS Lineal.



-Figura IV. 56. Gráfico FAP Lineal.

Como se ve en los gráficos anteriores (Figura IV. 55.) y (Figura IV. 56.) en ninguno de los factores de autocorrelación sobrepasan los límites ni por arriba ni por debajo, por lo tanto, se puede decir que no hay problemas de autocorrelación en el modelo.

## 4. Regresión Múltiple, segunda estructura.

En este punto se intenta explicar lo mismo que se ha explicado anteriormente en la estructura lineal, es decir, el número de vehículos matriculados pero a través de la regresión múltiple logarítmica.

### 4.1 Modelo Teórico Logarítmico.

La estructura logarítmica tiene una forma similar a la lineal, sólo que en este caso tanto la variable dependiente como explicativas ( $X_i$ ) se encuentran en forma de logaritmo natural, aunque la función que se expresa en el software es **Ln** (Y o Xi).

$$\text{MATRICULADOS } \text{Ln } Y = \beta_0 + \beta_1 \text{LnMotorización} + \beta_2 \text{Ln Exportación e Importación} + \beta_3 \text{LnParo} + \beta_4 \text{LnPobreza} + \beta_5 \text{LnPIB} + \beta_6 \text{LnEmpleo} +$$

$$+ \beta_7 \text{LnComercio Internacional} + \beta_8 \text{LnContaminación} + U$$

#### 4.2 Estimación de la primera ecuación de la Regresión Múltiple y validación del modelo.

En esta parte de nuevo se realiza una estimación de los parámetros del modelo, mediante una función logarítmica que nos calcula el programa estadístico.

$$\begin{aligned} \text{Ln (Matriculados)} = & -3,00936 + 0,397491 * \text{Ln(MOTORIZACIÓN)} - \\ & - 0,964667 * \text{Ln(EXPOR\_IMPOR)} - 0,044381 * \text{Ln(PARO)} - 0,228908 * \text{Ln(POBREZA)} + \\ & + 4,5829 * \text{Ln(PIB)} + 0,213654 * \text{Ln(EMPLEO)} - 0,315261 * \text{Ln(COMERCIO\_INTER)} - \\ & - 0,504347 * \text{Ln(CONTAMINACIÓN)} \end{aligned}$$

##### 4.2.1 Análisis de la Significación del Modelo y Variables.

En este apartado se volverá a comprobar si los parámetros del modelo son significativos o no, es decir, si tienen un nivel de confianza del 95%.

Parámetro	Estimación	Error		Valor-P
		Estándar	Estadístico T	
CONSTANTE	-3,00936	13,1024	-0,229679	0,8230
LOG(MOTORIZACIÓN)	0,397491	3,3611	0,118262	<b>0,9082</b>
LOG(EXPOR_IMPOR)	-0,964667	1,52565	-0,6323	0,5414
LOG(PARO)	-0,044381	0,757363	-0,0585994	0,9544
LOG(POBREZA)	-0,228908	0,892839	-0,256382	0,8029
LOG(PIB)	4,5829	2,55694	1,79234	0,1033
LOG(EMPLEO)	0,213654	0,691721	0,308874	0,7638
LOG(COMERCIO_INTER)	-0,315261	1,10484	-0,285345	0,7812
LOG(CONTAMINACIÓN)	-0,504347	0,628818	-0,802055	0,4412

-Tabla IV. 21. Resumen Análisis Regresión Múltiple Logarítmica.

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	23,6892	8	2,96114	0,82	<b>0,6047</b>
Residuo	36,221	10	3,6221		
Total (Corr.)	59,9101	18			

-Tabla IV. 22. Resumen Análisis Varianza Regresión Múltiple Logarítmica.

- **Modelo Inicial:**  $R^2 = 39,5\%$

En la tabla IV. 21. se puede ver que el P-valor del modelo es  $> 0,05$ , lo que indica que el modelo no es significativo. Para que el modelo sea significativo eliminaremos los puntos atípicos, en el caso de que haya y se verá que ocurre, es decir, si no sigue siendo significativo o si finalmente es significativo.

		<b>Y</b>		<b>Residuo</b>
<b>Fila</b>	Y	<i>Predicha</i>	<i>Residuo</i>	<i>Estudentizado</i>

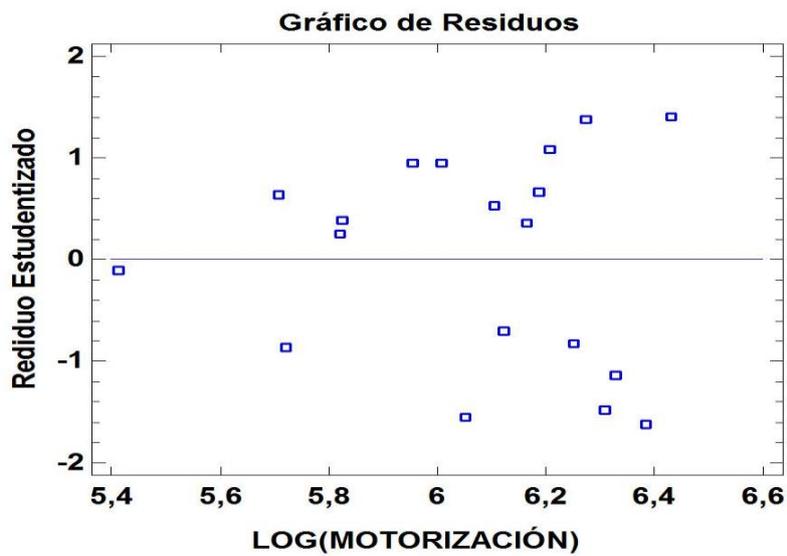
-Tabla IV. 23. Resumen Residuos Atípicos.

Como se puede ver (Tabla IV. 23.) no hay residuos superiores a dos, es decir, puntos anómalos, esto indica que no se puede eliminar ningún país para hacer que el modelo sea significativo y a causa de esto no se puede modificar ninguna variable explicativa.

Para terminar, esta estructura con un  $R^2$  insuficiente ya que está bastante por debajo de un 50%, no sería válida porque el ajuste del modelo sería muy bajito.

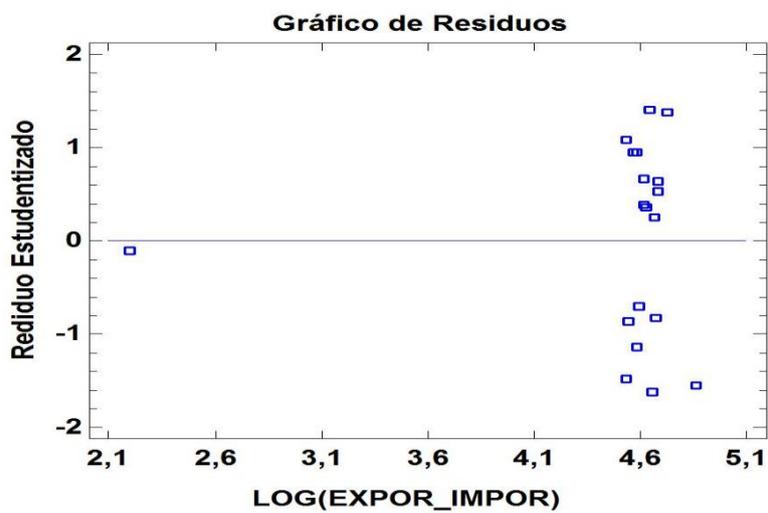
#### 4.2.2 Análisis gráficos de residuos.

Nuevamente en este punto se pueden analizar los gráficos de las variables explicativas utilizadas en este modelo frente a los residuos.



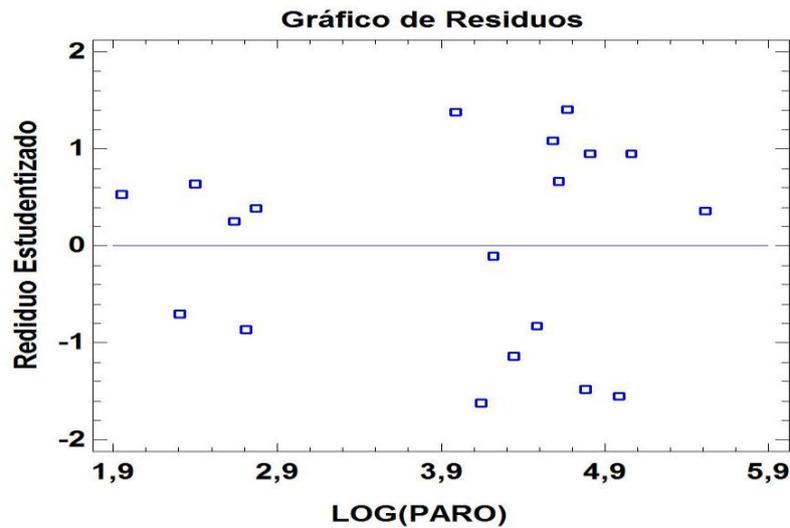
-Figura IV. 57. Gráfico Residuos vs Variable Motorización Estructura Logarítmica.

En esta Figura IV. 57. Se puede ver problema de Heterocedastidad en esta variable frente a los residuos.



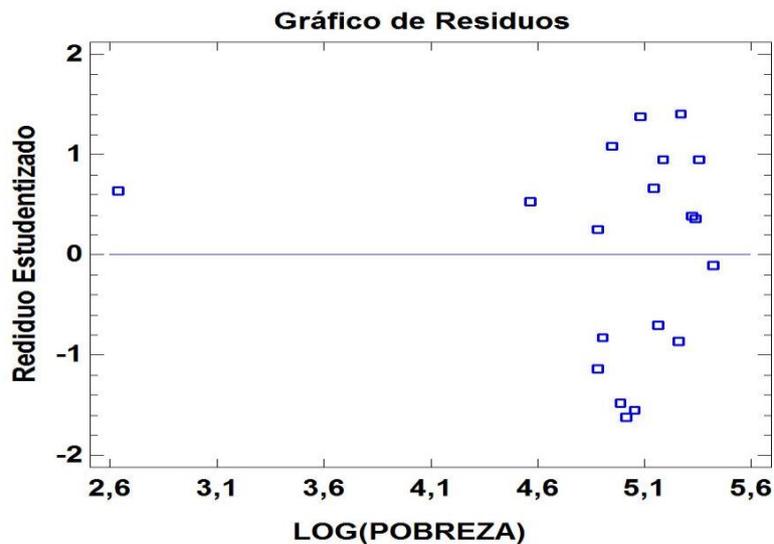
-Figura IV. 58. Gráfico Residuos vs Variable Exportaciones – Importaciones Estructura Logarítmica.

En este gráfico (Figura IV. 58.), se puede observar que los residuos no forman ninguna forma típica de presencia de ningún tipo de problema.



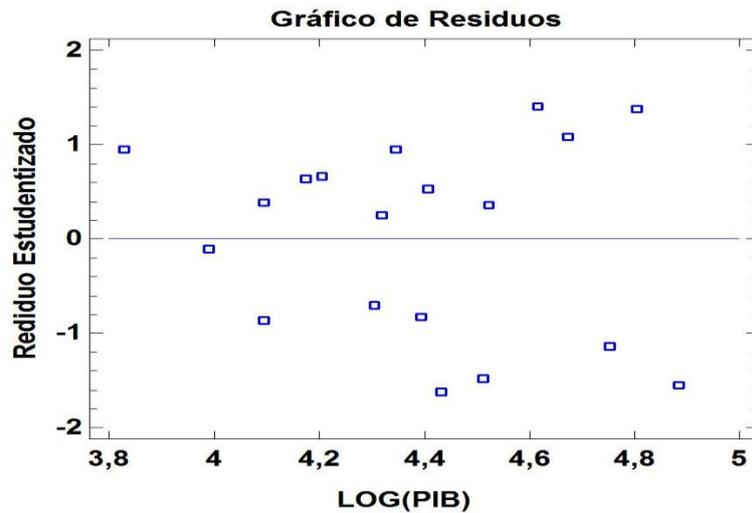
-Figura IV. 59. Gráfico Residuos vs Variable Paro Estructura Logarítmica.

Como ocurre en la Figura IV. 58. En este gráfico (Figura IV. 59.) los residuos no presentan forma de ningún problema.



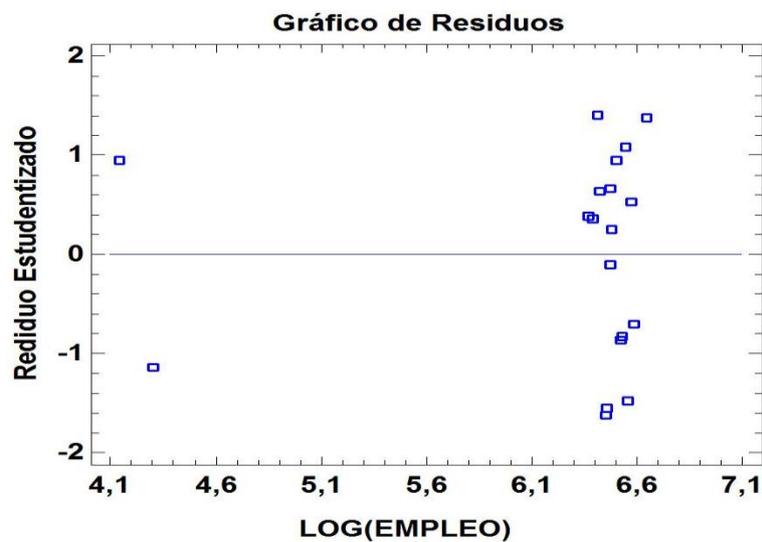
-Figura IV. 60. Gráfico Residuos vs Variable Pobreza Estructura Logarítmica.

En esta Figura IV. 60. Se repite el mismo caso del gráfico anterior, es decir, los residuos no tienen presencia de ningún problema.



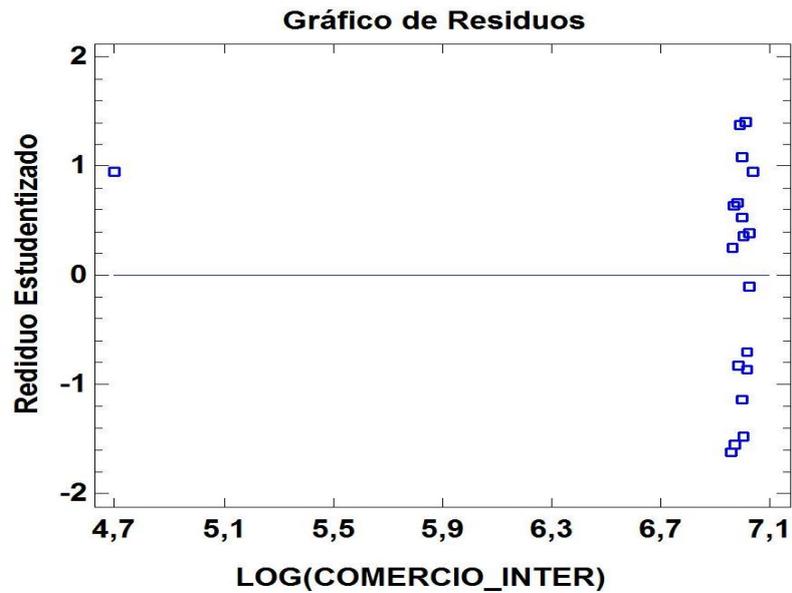
-Figura IV. 61. Gráfico Residuos vs Variable PIB Estructura Logarítmica.

En este gráfico (Figura IV. 61.) sí que se puede interpretar que la variable PIB tenga problemas de Heterocedasticidad.



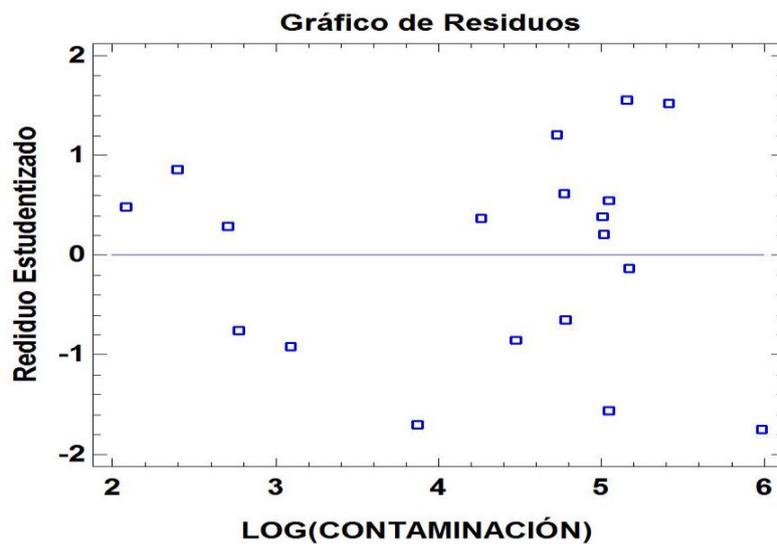
-Figura IV. 62. Gráfico Residuos vs Variable Empleo Estructura Logarítmica.

En este gráfico (Figura IV. 62.), no se observa a simple vista que la variable Empleo tenga ningún tipo de problema.



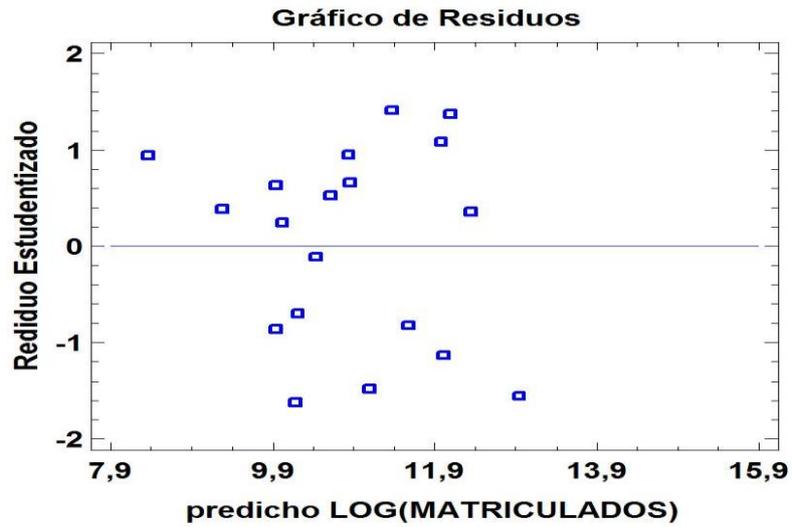
-Figura IV. 63. Gráfico Residuos vs Variable Comercio Internacional Estructura Logarítmica.

En esta Figura IV. 63. Ocurre lo mismo que el gráfico anterior (Figura IV. 62.) se puede interpretar que no se observa ningún problema.



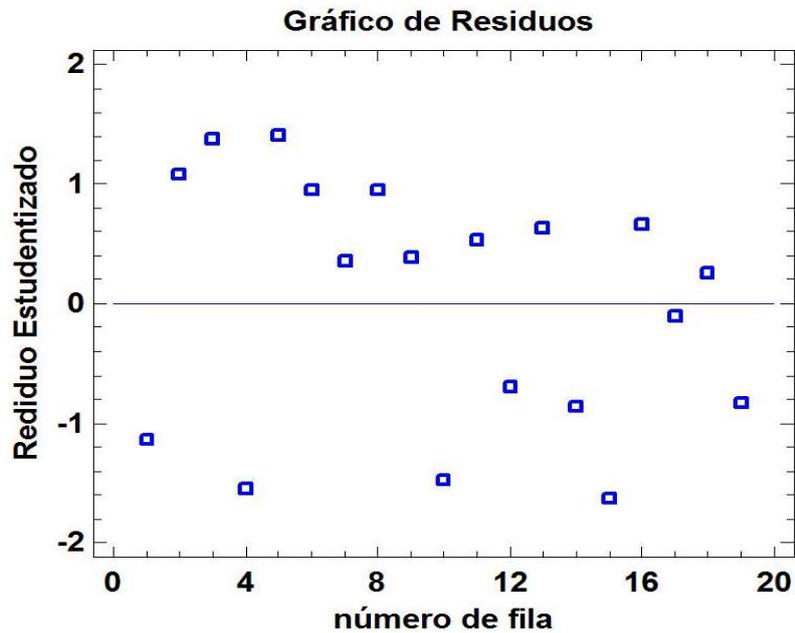
-Figura IV. 64. Gráfico Residuos vs Variable Contaminación Estructura Logarítmica.

En este Gráfico de Residuos (Figura IV. 64.), de nuevo se puede interpretar que no hay presencia de ningún problema.



-Figura IV. 65. Gráfico Residuos vs Predicho Matriculados Estructura Logarítmica.

En el gráfico de residuos vs Predicho Matriculados (Figura IV. 65.), no se observa ningún tipo de problema.

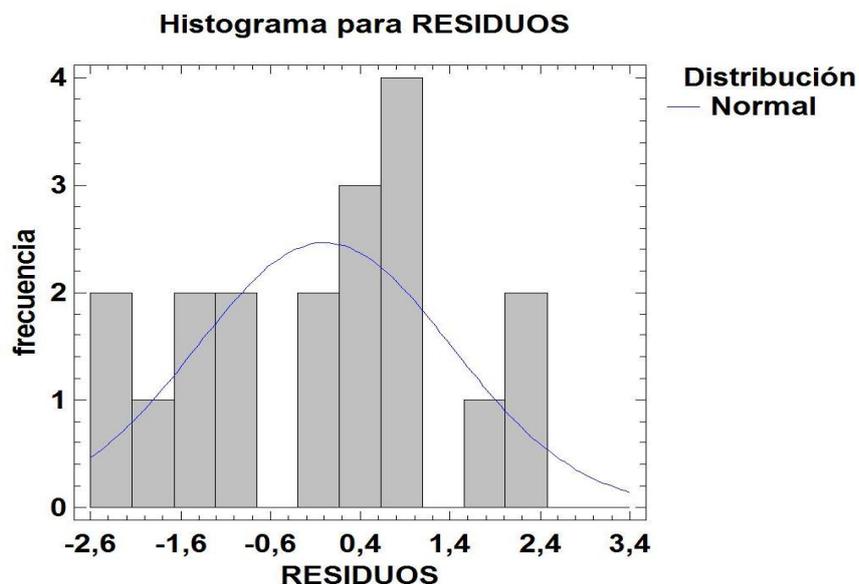


-Figura IV. 66. Gráfico Residuos vs Número de fila Estructura Logarítmica.

En este último gráfico de residuos (Figura IV. 66.), estos residuos no forman ninguna forma típica de presencia de problemas.

### 4.2.3 Normalidad de residuos.

Para ver si los residuos son normales, se volverán a realizar los dos criterios como se ha hecho en la estructura lineal y posteriormente se comprobará si son normales o no.



-Figura IV. 67. Histograma de Residuos Estructura Logarítmica.

El histograma (Figura IV. 67.) no forma una campana de Gauss perfecta, se observan huecos y el máximo no está del todo centrado, por lo que podría esperarse que los residuos no sigan una distribución normal.

<b>Prueba</b>	<b>Estadístico</b>	<b>Valor-P</b>
Chi-Cuadrado	6,89474	0,648078
Estadístico W de Shapiro-Wilk	0,947774	0,368958
Valor-Z para asimetría	0,130272	0,896346
Valor-Z para curtosis	Datos Insuficientes	

-Tabla IV. 24. Test de Normalidad Logarítmico.

Si volvemos a observar el test (Tabla IV. 23.) y se sigue el mismo método que en la estructura lineal, se cogerá el valor más restrictivo que en este caso es **Estadístico W de Shapiro-Wilk** con un P-valor de 0,368958.

Se saca como conclusión que los residuos son normales porque el P-valor es  $>0,05$ .

#### 4.2.4 Heterocedasticidad.

En este apartado se vuelve a hablar sobre el problema de heterocedasticidad en el modelo. Se calculará como se ha hecho anteriormente en la anterior estructura y se verá si tenemos problemas de heterocedasticidad o no en esta estructura a partir de los resultados de la Tabla IV. 25.

<b>Parámetro</b>	<b>Estimación</b>	<b>Error</b>		<b>Valor-P</b>
		<b>Estándar</b>	<b>Estadístico</b>	
			<b>T</b>	
CONSTANTE	-38,9259	7,50534	-5,18642	0,0004
LOG(MOTORIZACIÓN)	0,740135	1,9253	0,384426	0,7087
LOG(EXPOR_IMPOR)	1,78242	0,873921	2,03957	0,0687
LOG(PARO)	1,30124	0,433833	2,99941	0,0134
LOG(POBREZA)	0,510582	0,511436	0,99833	0,3417
LOG(PIB)	-1,32212	1,46466	-0,902677	0,3879
LOG(EMPLEO)	0,916887	0,396231	2,31402	0,0432
LOG(COMERCIO_INTER)	2,20257	0,632876	3,48026	0,0059
LOG(CONTAMINACIÓN)	1,29338	0,3602	3,59074	0,0049

-Tabla IV. 25. Heterocedasticidad Logarítmica.

Como se observa en la Tabla IV. 25. hay cuatro variables las cuales presentan problemas de Heterocedasticidad, ya que el P-valor es  $<0,05$  entonces la varianza no es constante.

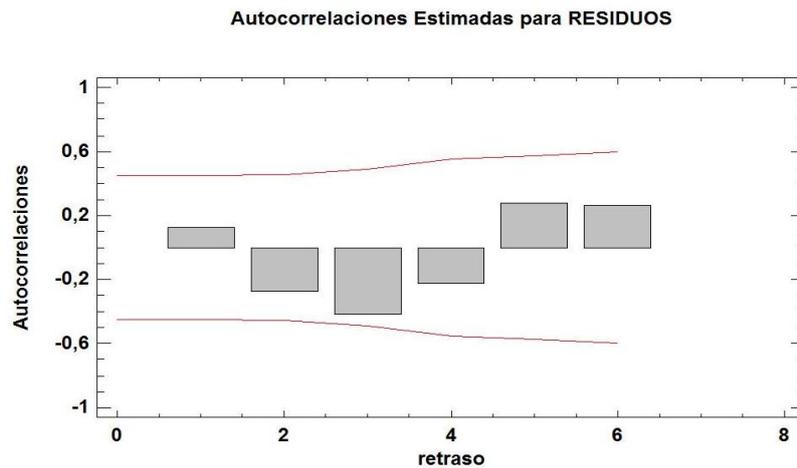
Cuando hay más de una variable que el P-valor sea inferior a 0,05 no se puede solucionar elevándola a H y luego mirando el R-cuadrado hasta saber con qué valor H nos quedamos. Se solucionará eliminando estas cuatro variables explicativas que tienen problema de heterocedasticidad y la estructura ya no presentara problema de heterocedasticidad como se ve en la siguiente Tabla IV. (X):

<b>Parámetro</b>	<b>Estimación</b>	<b>Error</b>		<b>Valor-P</b>
		<b>Estándar</b>	<b>Estadístico</b>	
			<b>T</b>	
CONSTANTE	-30,9214	7,72932	-4,00053	0,0015
LOG(MOTORIZACIÓN)	3,65191	2,02593	1,80259	0,0947
LOG(EXPOR_IMPOR)	-0,236736	0,833818	-0,283918	0,7809
LOG(POBREZA)	0,35373	0,635411	0,556696	0,5872
LOG(PIB)	1,799	1,62912	1,10427	0,2895
LOG(CONTAMINACIÓN)	0,496116	0,328692	1,50936	0,1551

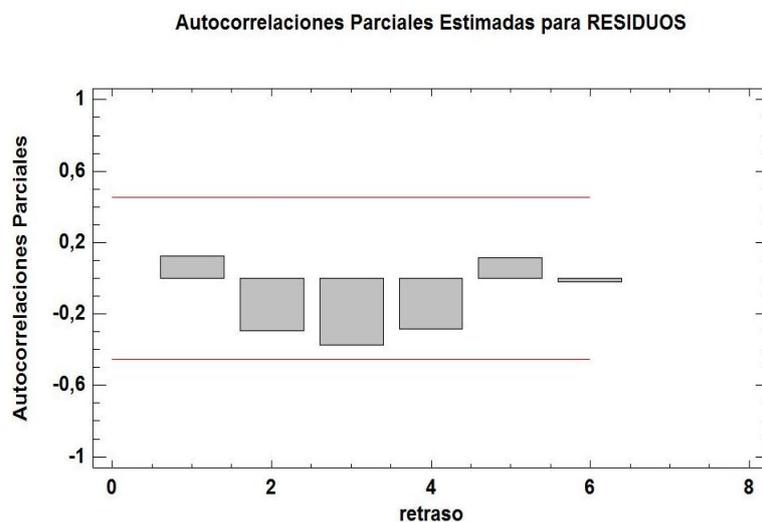
-Tabla IV. 26. Heterocedasticidad Logarítmica Final.

#### 4.2.5 Autocorrelación.

Para terminar con la estructura Logarítmica, se analizará el problema de autocorrelación para saber si este modelo presenta este problema o no. Como ya se ha realizado en la estructura anterior, se observarán los dos gráficos; el FAS (Figura IV. 68.) y el FAP (Figura IV. 69.) y mediante estas dos funciones gráficas, se mirará si el problema de autocorrelación está presente.



-Figura IV. 68. Gráfico FAS Logarítmico.



-Figura IV. 69. Gráfico FAP Logarítmico.

Como ocurre en la estructura lineal, la estructura logarítmica tampoco presenta problema de autocorrelación porque los retardos siguen sin sobrepasar los límites.

Con el intento de solución del único problema detectado de Heterocedasticidad, no se consigue un modelo con al menos una variable explicativa significativa, es decir, que con esta estructura no se puede obtener ningún modelo válido.

## 5. Regresión Múltiple tercera estructura.

En este apartado se intenta explicar la tercera estructura como se han explicado anteriormente las otras dos, diferenciándose en que esta vez se utiliza una estructura inversa.

### 5.1 Modelo Teórico Inverso.

Esta tercera estructura, la estructura inversa, sigue el principio de la estructura lineal lo que las variables explicativas se encuentran de forma inversa, es decir,  $(1/X_i)$ .

$$Y(\text{Matriculados}) = \beta_0 + \beta_1 1/\text{Motorización} + \beta_2 1/\text{Exportación e Importación} + \\ + \beta_3 1/\text{Paro} + \beta_4 1/\text{Pobreza} + \beta_5 1/\text{PIB} + \beta_6 1/\text{Empleo} + \\ + \beta_7 1/\text{Comercio Internacional} + \beta_8 1/\text{Contaminación} + U$$

### 5.2 Estimación de la primera ecuación de la Regresión Múltiple y validación del modelo.

En esta parte, se estiman los parámetros del modelo teórico inverso utilizado en el programa Statgraphics. Seguidamente se observa la función que calculada del modelo que nos indica el programa.

$$\text{MATRICULADOS} = 112088, + 7,97232\text{E}6/\text{MOTORIZACIÓN} + \\ + 1,56913\text{E}6/\text{EXPOR\_IMPOR} + 884063/\text{PARO} + 414331/\text{POBREZA} - \\ - 1,802\text{E}7/\text{PIB} + 2,20115\text{E}6/\text{EMPLEO} + 1,1987\text{E}7/\text{COMERCIO\_INTER} + \\ + 2,89739\text{E}6/\text{CONTAMINACIÓN}$$

### 5.2.1 Análisis de la Significación del Modelo y Variables.

En esta parte, como se ha hecho anteriormente se observa si los parámetros del modelo son significativos o no, analizando si los P-valores son  $<0,05$  los cuales nos indicaran que son significativos o por otro lado comprobando el intervalo de confianza de estos.

Parámetro	Estimación	Error	Estadístico	Valor-P
		Estándar	T	
CONSTANTE	112088,	78822,6	1,42202	0,1980
1/MOTORIZACIÓN	7,97232E6	4,54171E7	0,175536	<b>0,8656</b>
1/EXPOR_IMPOR	1,56913E6	1,12815E6	1,39089	0,2069
1/PARO	884063,	526649,	1,67866	0,1371
1/POBREZA	414331,	1,12856E6	0,367134	0,7244
1/PIB	-1,802E7	6,67697E6	-2,69883	0,0307
1/EMPLEO	2,20115E6	4,174E6	0,527348	0,6142
1/COMERCIO_INTER	1,1987E7	8,75559E6	1,36907	0,2133
1/CONTAMINACIÓN	2,89739E6	548965,	5,27792	0,0012

-Tabla IV. 27. Resumen Análisis Regresión Múltiple Inversa.

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,22262E11	8	1,52828E10	4,16	<b>0,0381</b>
Residuo	2,5729E10	7	3,67558E9		
Total (Corr.)	1,47991E11	15			

-Tabla IV. 28. Resumen Análisis Varianza Regresión Múltiple Inversa.

- **Modelo Inicial:  $R^2 = 82,6\%$**

En la tabla IV. 28. se puede ver que el P-valor del modelo es  $>0,05$ , lo que indica que el modelo no es significativo.

Como se ha hecho anteriormente en la estructura lineal, para que este modelo sea significativo, eliminaremos los puntos atípicos, en el caso de que haya.

Seguidamente, se observará la Tabla IV. 29A. para comprobar la existencia de los residuos anómalos (atípicos).

		<b>Y</b>		<b>Residuo</b>
<b>Fila</b>	<b>Y</b>	<b>Predicha</b>	<b>Residuo</b>	<b>Estudentizado</b>
3	1,6344E6	570068,	1,06433E6	3,31

-Tabla IV. 29A. Resumen Residuos Atípicos.

Como se puede ver en la Tabla IV. 29. el residuo anómalo está en la fila 3 que corresponde a Alemania. Seguiremos el mismo criterio que en los anteriores modelos y una vez eliminado este país del modelo, se volverá a comprobar si el modelo es significativo.

Una vez eliminamos Alemania aparece otro residuo anómalo como se puede ver en la Tabla VI. 29B

		<b>Y</b>		<b>Residuo</b>
<b>Fila</b>	<b>Y</b>	<b>Predicha</b>	<b>Residuo</b>	<b>Estudentizado</b>
2	1,04898E6	341724,	707258,	3,10

-Tabla IV. 29B. Resumen Residuos Atípicos.

Por lo tanto, se realizará el mismo proceso para eliminar la fila 2, es decir, Francia.

		<b>Y</b>		<b>Residuo</b>
<b>Fila</b>	<b>Y</b>	<b>Predicha</b>	<b>Residuo</b>	<b>Estudentizado</b>
3	815213,	223090,	592123,	11,44

-Tabla IV. 29C. Resumen Residuos Atípicos.

Seguidamente, se observa que sale otro punto pero aun peor (Tabla IV. 29C.) Se sigue eliminando esta fila, es decir, Italia y comprobaremos la prueba de normalidad de residuos.

Al comprobarlo se ve que con esto no solo se consigue que sean normales los residuos, sino que se mejora mucho el modelo, apareciendo más variables explicativas significativas, y mejorando notablemente el ajuste del modelo con los datos reales empleados.

*Análisis y predicción de la variabilidad del número de vehículos matriculados a nivel Europeo*

Fila	Y	Y		Residuo
		Predicha	Residuo	Estudentizado
4	406071,	287942,	118129,	2,96
16	28327,0	143061,	-114734,	-2,32

*-Tabla IV. 29D. Resumen Residuos Atípicos.*

Realizando de nuevo el análisis de la significatividad del modelo resultante se observa que quedan dos variables explicativas significativas, junto con la constante (Tabla IV. 30.).

Parámetro	Estimación	Error	Estadístico	Valor-P
		Estándar	T	
CONSTANTE	140793,	61770,6	2,27928	0,0402
1/PIB	-1,03179E7	4,39315E6	-2,34863	0,0353
1/CONTAMINACIÓN	2,0744E6	429693,	4,82762	0,0003

*-Tabla IV. 30. Resumen Análisis Regresión Múltiple.*

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	9,93672E10	2	4,96836E10	13,28	0,0007
Residuo	4,8624E10	13	3,74031E9		
Total (Corr.)	1,47991E11	15			

*-Tabla IV. 31. Resumen Análisis Varianza Regresión Múltiple.*

Para terminar, se puede ver un resumen del análisis para el modelo de regresión múltiple inversa, donde el P-valor que se ve es <0,05 lo que quiere decir que el modelo, una vez eliminado los países que tenían puntos atípicos ahora sí que es significativo (Tabla IV. 31.) En este caso el ajuste del modelo mejora de forma notable, comparándose con la primera estimación, pero no se consigue un ajuste tan elevado como en el caso de la estructura lineal final propuesta, aunque debe destacarse que en este modelo se han conseguido incluir mayor número de países en la modelización.

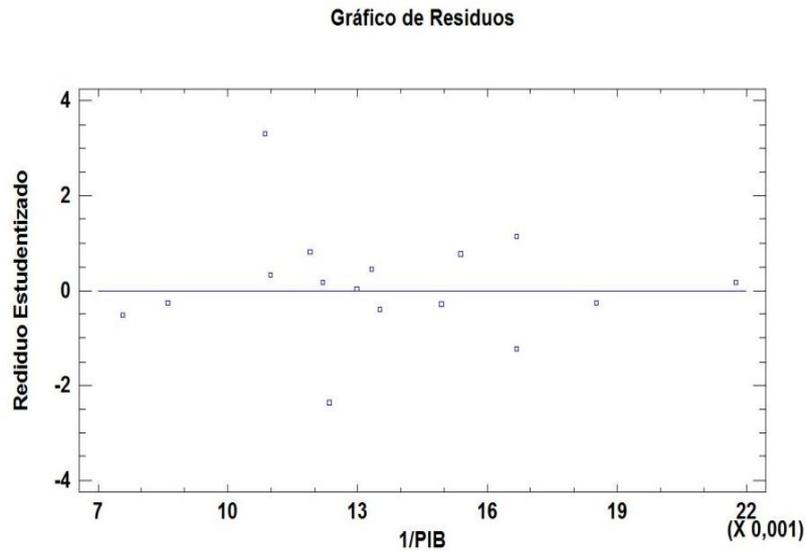
- **Modelo final:**  $R^2 = 67,1\%$

$$R^2 \text{ (ajustado para g.l.)} = 62,1\%$$

$$\text{MATRICULADOS} = 140793 - 1,03179E7/\text{PIB} + 2,0744E6/\text{CONTAMINACIÓN}$$

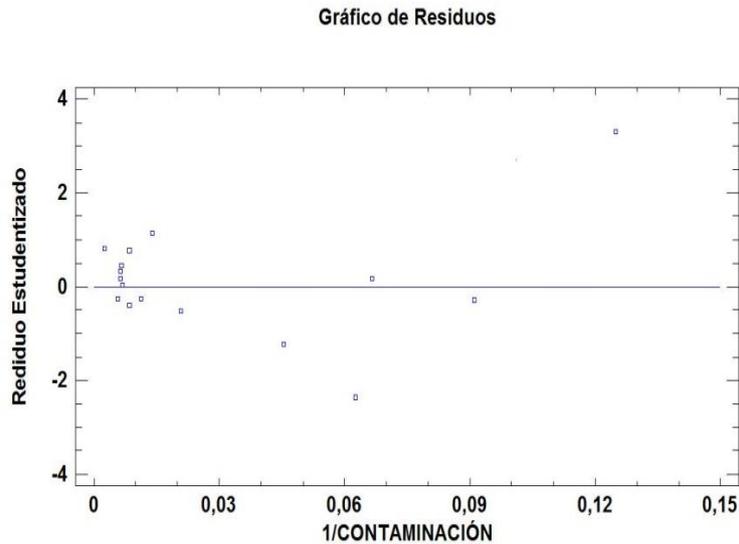
### 5.2.2 Análisis gráficos de residuos.

En este punto de nuevo como ya se ha realizado en las dos estructuras anteriores se muestran los gráficos de las variables dependientes (explicativas) utilizadas en el modelo frente a los residuos.



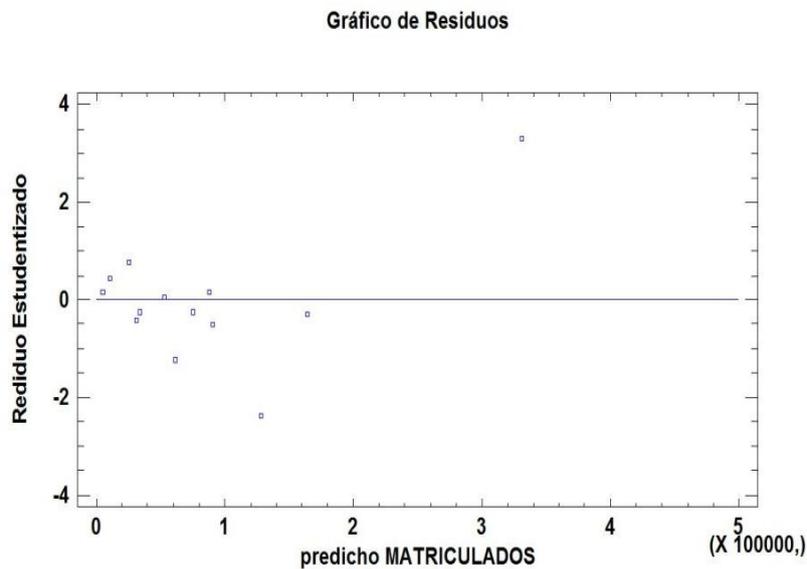
-Figura IV. 70. Gráfico Residuos vs Variable PIB Estructura Inversa.

En este gráfico (Figura IV. 70), se observa una distribución aleatoria, pero también se puede destacar que se ve un punto anómalo que viene causado porque se aleja del resto.



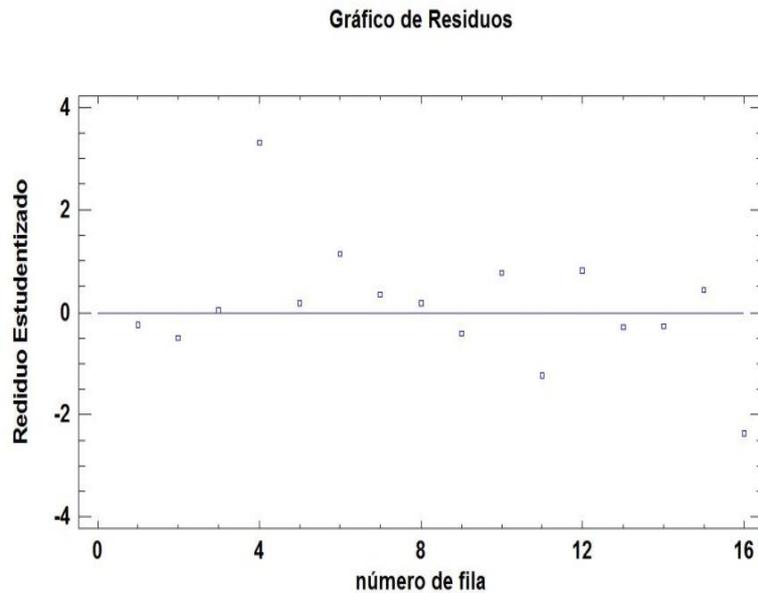
-Figura IV. 71. Gráfico Residuos vs Variable Contaminación Estructura Inversa.

En la Figura IV. 71. El gráfico de residuos sigue como el gráfico anterior (Figura IV. 70.), es decir, presenta una distribución aleatoria y tiene un punto anómalo.



-Figura IV. 72. Gráfico Residuos vs Predicho Matriculados Estructura Inversa.

En este gráfico de residuos frente al predicho de matriculados, se observa claramente un punto anómalo muy alejado del resto de puntos.

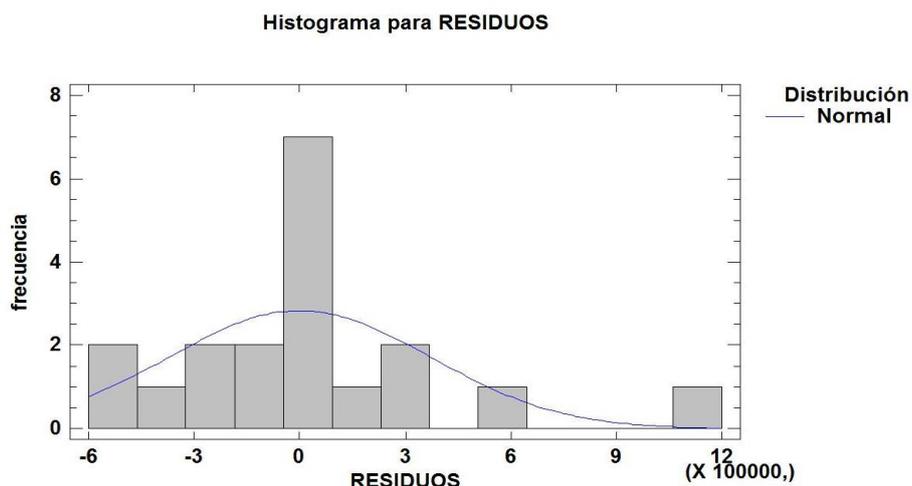


*-Figura IV. 73. Gráfico Residuos vs Número de fila Estructura Inversa.*

En esta última Figura IV. 73. Los residuos siguen una distribución aleatoria como el resto de gráficos de este modelo inverso y como en los demás gráficos sigue observándose un punto anómalo.

### **5.2.3 Normalidad de Residuos.**

Como se ha hecho anteriormente en las dos estructuras anteriores, se representará el Histograma para ver la forma que tiene la campana Gauss y seguidamente se volverá a observar el test para saber si los residuos siguen una distribución normal.



-Figura IV. 74. Histograma de Residuos Estructura Inversa.

Como se observa en la Figura IV. 74. este histograma no forma una campana de Gauss perfecta, pero como la mayoría de los residuos están en la media y a medida que nos alejamos de la medida la frecuencia disminuye, supondremos la normalidad de Residuos.

<b>Prueba</b>	<b>Estadístico</b>	<b>Valor-P</b>
Chi-Cuadrado	13,2105	0,153309
Estadístico W de Shapiro-Wilk	0,899483	0,0478661
Valor-Z para asimetría	1,57484	0,115292
Valor-Z para curtosis	Datos Insuficientes	

-Tabla IV. 32. Test de Normalidad Inverso

De seguida, una vez se tienen los P-valores de los cuatro test (Tabla IV. 32) se coge el más restrictivo que en este caso es Estadístico **W de Shapiro-Wilk** que es 0,0478661. Como este P-valor  $< 0,05$  se saca la conclusión de que los residuos no siguen una distribución normal, es decir, se acepta la  $H_1$ . Los residuos no siguen una distribución normal.

#### 5.2.4 Heterocedasticidad.

De nuevo para analizar si el modelo presenta problemas de heterocedasticidad o no, se realizará un análisis entre los residuos al cuadrado frente a las variables explicativas, como se ha hecho en las anteriores estructuras y se comprobarán los resultados de la (Tabla IV. 33.), es decir, los P-valores de cada variable.

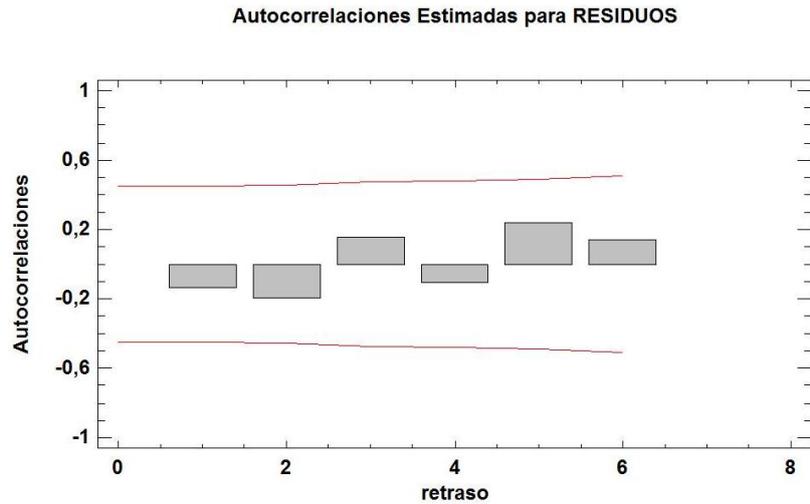
<b>Parámetro</b>	<b>Estimación</b>	<b>Error</b>	<b>Estadístico</b>	<b>Valor-P</b>
		<b>Estándar</b>	<b>T</b>	
CONSTANTE	2,3883E9	4,0882E9	0,584194	0,5774
1/MOTORIZACIÓN	8,66597E10	2,35559E12	0,0367889	<b>0,9717</b>
1/EXPOR_IMPOR	-1,81171E10	5,85124E10	-0,309628	0,7659
1/PARO	-9,53761E9	2,7315E10	-0,34917	0,7372
1/POBREZA	-1,55093E10	5,85335E10	-0,264965	0,7987
1/PIB	-1,14894E10	3,46306E11	-0,0331769	0,9745
1/EMPLEO	-2,09513E11	2,16488E11	-0,967782	0,3654
1/COMERCIO_INTER	-2,23094E11	4,54115E11	-0,491272	0,6383
1/CONTAMINACIÓN	3,18647E10	2,84725E10	1,11914	0,3000

-Tabla IV. 33. Heterocedasticidad Inversa.

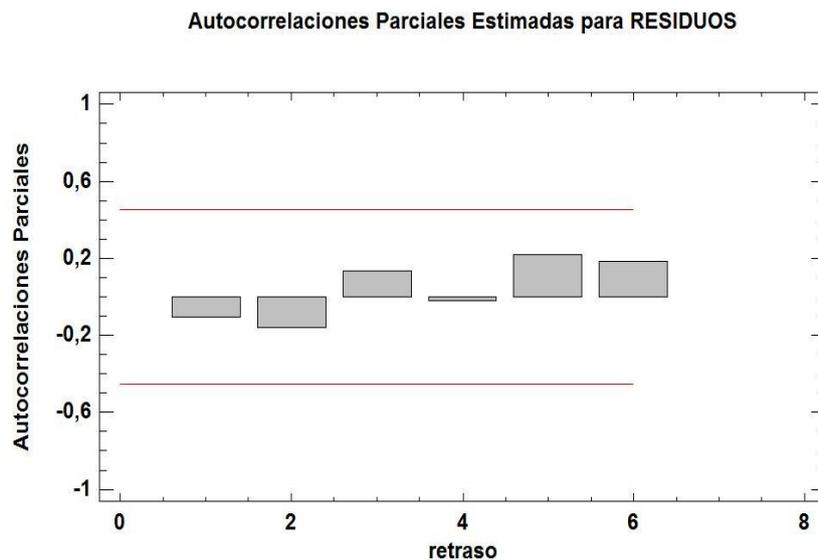
Como se observa en la Tabla IV. 33. ninguna de las variables explicativas, presenta problemas de heterocedasticidad ya que los P-valores son  $>0,05$  por esto se puede decir que el modelo está libre de heterocedasticidad.

### 5.2.5 Autocorrelación.

En esta parte, se representarán las dos funciones de autocorrelaciones estimadas (Figura IV. 75.) y (Figura IV. 76.) para observar si los retardos superan tanto los límites inferiores o superiores y saber así si hay problema de correlación como se ha hecho en las dos estructuras anteriores.



-Figura IV. 75. Gráfico FAS.



-Figura IV. 76. Gráfico FAP.

Como se ve en los gráficos anteriores (Figura IV. 75. y Figura IV. 76) ninguno de los factores de autocorrelación sobrepasan los límites ni por arriba ni por debajo, por lo tanto, se puede decir que ocurre como las anteriores estructuras, es decir, no está presenta el problema de autocorrelación.

## 6. Interpretación y Predicción de los modelos.

En este último punto de cálculos y resultados se pretende interpretar los parámetros de las estructuras, y también por otro lado tener una predicción concreta de los modelos. El objetivo principal del proyecto es poder realizar estas predicciones sobre el número de vehículos matriculados en Europa, y principalmente con interés en España. Se realizará un resumen de las estructuras con el fin de poder determinar aquella o aquellas más interesantes, con las que se modelizará y se obtendrán las predicciones pertinentes.

A continuación, se puede observar la Tabla IV. 34. de las tres estructuras con las que se destacara cual es la estructura más importante para el modelo del proyecto a partir de:

- 1) Significación.
- 2) Normalidad de Residuos.
- 3) Heterocedasticidad.
- 4) Autocorrelación.
- 5)  $R^2$ .

	SIGNIFICATIVO	NORMORMALIDAD	HETEROCEDASTICIDAD	AUTOCORRELACIÓN	$R^2$
LINEAL	V (SIN ESPAÑA)	V	V	V	86,2
LOGARÍTMICA	X	V	V	V	39,54
INVERSA	V	V	V	V	67,144

-Tabla IV. 34. Comparación de los 3 modelos.

Seguidamente una vez se ha observado la Tabla IV. 34. se elige la estructura lineal porque tiene el mayor R-cuadrado respecto al resto de estructuras, ya que lo que se busca es elegir un modelo con un índice alto de ajuste y aparte, el modelo es significativo, los residuos siguen una distribución normal, y no presenta ninguno problema, es decir, ni de heterocedasticidad ni de autocorrelación.

## *Análisis y predicción de la variabilidad del número de vehículos matriculados a nivel Europeo*

Como se ha explicado anteriormente, a causa de elegir esta estructura se ha eliminado España para obtener un modelo válido, donde los residuos cumplan el principio de normalidad. Por este motivo se eligen los datos de España para realizar una primera predicción:

	<b>AJUSTADO</b>	<b>INFERIOR 95.0%</b>	<b>SUPERIOR 95.0%</b>
<b>FILA/AÑO</b>	<b>Número de vehículos</b>	<b>LC para la MEDIA</b>	<b>LC para la MEDIA</b>
2012	34186,3	24316,7	44055,8
2014	33249,5	23457,2	43041,8
INVENCIÓN	28049,5	17156,4	38942,6

*-Tabla IV. 35. Predicción.*

En la Tabla IV. 35 se observan los valores de tres predicciones. En primer lugar, se obtienen los valores de los vehículos matriculados en España con los datos del 2012. El valor puntual de la predicción es 34186,3 número de vehículos que no coincide con el valor medio obtenido en principio para este año de 406071 número de vehículos. Sin embargo, tampoco queda el valor de la predicción dentro del intervalo de confianza calculado por el modelo seleccionado  $[24316,7 < Y < 44055,8]$  95%, debido a que el valor de los vehículos matriculados en España es tan superior al de los países que han quedado en el modelo, que quedaba precisamente como un punto anómalo. Para poder incluir estos datos se optaría por seleccionar la estructura inversa, pero es desaconsejable porque el ajuste no llega al 70% y las predicciones serían muy imprecisas.

Los siguientes valores de la predicción se siguen tomando con valores de España publicados, aunque se sabe que serían válidos sólo para los países que se han incluido en el modelo, en caso de que estos alcanzaran los valores de motorización o PIB de España. Por esto en segundo lugar, se realiza una prueba con el valor del PIB de España en 2014 que pasa de 92 a 91. Se realiza un cálculo sin variar el valor de 2012 de la motorización, para ver el efecto del cambio de esta variable en el modelo como ejemplo, no extrapolable a España. Se observa un ligero descenso de los vehículos matriculados con este cambio pasando a 33249,5 número de vehículos, con un intervalo de confianza de  $[23457,2 \leq Y \leq 43041,8]_{95\%}$ , de modo que un descenso de un punto en el PIB se traduce en un descenso de alrededor del 3% en el número de vehículos matriculados.

Por último, se realiza la prueba de predicción manteniendo fijo el valor de PIB en 91 puntos, y aumentando en un 10% la motorización en España (recordar que corresponde al número de vehículos por cada mil habitantes). Con esto se observa un cambio brusco del número de vehículos matriculados descendiendo aproximadamente en un 18%, con el aumento del número de vehículos por cada mil habitantes. Este fenómeno puede explicarse con el hecho de que si aumenta el número de vehículos por habitante, es lógico que descienda la necesidad de compra de nuevos vehículos, y con ello se de este descenso en el número de vehículos matriculados.

## **V. Conclusiones**

En este punto se destaca la lista de conclusiones obtenidas del trabajo:

- Se ha sido capaz de realizar la modelización de la producción número de vehículos de la Unión Europea en año 2012.
- Se ha conseguido determinar los factores más importantes para la producción de vehículos de los países europeos.
- Se han realizado tres modelos de regresión múltiple con estructuras; lineal, inversa y logarítmica para la explicación de la producción de número de vehículos en EU.
- De las tres estructuras planteadas, se confía el uso del modelo lineal con más alto índice de R-cuadrado entre los modelos.
- Para que el modelo lineal sea válido, se han eliminado países como España entre otros, ya que el valor de los vehículos es superior al resto de países del modelo y por esta causa es eliminado para que no diera problemas.
- Si se confía en el modelo inverso, se incluirían países como España pero esto es desaconsejable por que el ajuste del modelo no llega al 70%.
- Si desciende el PIB, esto implicará que el número de vehículos matriculados descenderá también como consecuencia de una relación directamente proporcional entre ambas variables.
- Si aumenta la Motorización en el país, afectara negativamente a la producción de vehículos matriculados, ya que con el aumento del número de vehículos por habitante, es lógico que descienda la necesidad de compra de nuevos vehículos, y con ello se de este descenso en el número de vehículos matriculados.

## **VI. Futuras Líneas de Investigación.**

En este apartado se listan por ultimo las posibles líneas de investigación derivadas de este trabajo:

- Búsqueda de modelos alternativos que expliquen el comportamiento de los vehículos matriculados en los países que han tenido que dejarse fuera del modelo como Alemania, España o Italia.
- Búsqueda de nuevos factores que puedan ser causantes de la matriculación de vehículos en un país que no se hayan tenido en cuenta en el trabajo como el nivel educativo de la población, o las ayudas que destinan los gobiernos al sector de automoción, por ejemplo.
- Modelización del número de vehículos matriculados en los países de principal interés mediante series temporales para realizar predicciones basadas únicamente en el registro de datos a lo largo del tiempo.

## **VII. Bibliografía.**

## Referencias:

[1] [https://es.wikipedia.org/wiki/Historia\\_del\\_autom%C3%B3vil](https://es.wikipedia.org/wiki/Historia_del_autom%C3%B3vil)

[2] <http://profesionaleshoy.es/blog/2016/03/30/la-produccion-de-vehiculos-en-espana-llega-casi-al-medio-millon-hasta-febrero/36732>

[3] <http://www.eleconomista.es/ecomotor/motor/noticias/7197214/12/15/La-produccion-de-automoviles-en-Espana-lidera-el-crecimiento-europeo.html>

[4] <http://ec.europa.eu/eurostat>

**→ Listado de Figuras:**

**Figura III. 1.** Ejemplo de Matriz Correlación.

**Figura III. 2.** Ejemplo de Matriz Correlación Inversa.

**Figura III. 3.** Ejemplo de un gráfico de Residuos con problemas de Heterocedasticidad.

**Figura III. 4.** Ejemplo de Residuos Anómalos.

**Figura III. 5.** Gráfico de Falta de Linealidad.

**Figura III. 6.** Gráfico del Residuo de un ajuste sin contraste.

**Figura III. 7.** Gráfico con problemas de Heterocedasticidad y Falta de Linealidad.

**Figura III. 8.** Gráfico de Autocorrelación de 1º orden con signo negativo.

**Figura III. 9.** Gráfico de Autocorrelación de 1º orden con signo positivo.

**Figura III. 10.** Gráfico de Varianza no Constante en Heterocedasticidad.

**Figura III. 11.** Gráfico Problema de Heterocedasticidad.

**Figura III. 12A.** Gráfico de Autocorrelación Simple.

**Figura III. 12B.** Gráfico de Autocorrelación Parcial.

**Figura III. 13.** Gráfico Predicciones Puntual e Intervalo Confianza.

**Figura IV. 14.** Gráfico de Caja y Bigotes Variable Números de Vehículos Matriculados.

**Figura IV. 15.** Histograma Variable Número de Vehículos Matriculados.

**Figura IV. 16.** Gráfico Probabilidad Normal Variable Número de Vehículos Matriculados.

**Figura IV. 17.** Diagrama de Barras Variable Número de Vehículos Matriculados frente a los 19 países.

**Figura IV. 18.** Gráfico de Caja y Bigotes Variable Motorización.

**Figura IV. 19.** Histograma Variable Motorización.

**Figura IV. 20.** Gráfico Probabilidad Normal Variable Motorización.

**Figura IV. 21.** Diagrama de Barras Variable Motorización frente a los 19 países.

**Figura IV. 22.** Gráfico de Caja y Bigotes Variable Exportaciones – Importaciones.

**Figura IV. 23.** Histograma Variable Exportaciones – Importaciones.

**Figura IV. 24.** Gráfico Probabilidad Normal Variable Exportaciones – Importaciones.

**Figura IV. 25.** Diagrama de Barras Variable Exportaciones – Importaciones frente a los 19 países.

**Figura IV. 26.** Gráfico de Caja y Bigotes Variable Paro.

**Figura IV. 27.** Histograma Variable Paro.

**Figura IV. 28.** Gráfico Probabilidad Normal Variable Paro.

**Figura IV. 29.** Diagrama de Barras Variable Paro frente a los 19 países.

**Figura IV. 30.** Gráfico de Caja y Bigotes Variable Pobreza.

**Figura IV. 31.** Histograma Variable Pobreza.

**Figura IV. 32.** Gráfico Probabilidad Normal Variable Pobreza.

**Figura IV. 33.** Diagrama de Barras Variable Pobreza frente a los 19 países.

**Figura IV. 34.** Gráfico de Caja Y Bigotes Variable PIB.

**Figura IV. 35.** Histograma Variable PIB.

**Figura IV. 36.** Gráfico Probabilidad Normal Variable PIB.

**Figura IV. 37.** Diagrama de Barras Variable PIB frente a los 19 países.

**Figura IV. 38.** Gráfico de Caja y Bigotes Variable Empleo.

**Figura IV. 39.** Histograma Variable Empleo.

**Figura IV. 40.** Gráfico Probabilidad Normal Variable Empleo.

**Figura IV. 41.** Diagrama de Barras Variable Empleo frente a los 19 países.

**Figura IV. 42.** Gráfico de Caja y Bigotes Variable Comercio Internacional.

**Figura IV. 43.** Histograma Variable Comercio Internacional.

**Figura IV. 44.** Gráfico Probabilidad Normal Variable Comercio Internacional.

**Figura IV. 45.** Diagrama de Barras Variable Comercio Internacional frente a los 19 países.

**Figura IV. 46.** Gráfico de Caja y Bigotes Variable Contaminación.

**Figura IV. 47.** Histograma Variable Contaminación.

**Figura IV. 48.** Gráfico Probabilidad Normal Variable Contaminación.

**Figura IV. 49.** Diagrama de Barras Variable Contaminación frente a los 19 países.

**Figura IV. 50.** Gráfico de Residuos vs Variable Motorización Estructura Lineal.

**Figura IV. 51.** Gráfico de Residuos vs Variable PIB Estructura Lineal.

**Figura IV. 52.** Gráfico de Residuos vs Predicho Matriculados Estructura Lineal.

**Figura IV. 53.** Gráfico de Residuos vs Número de Filas Estructura Lineal.

**Figura IV. 54.** Histograma de Residuos Estructura Lineal.

**Figura IV. 55.** Gráfico Fas Lineal.

**Figura IV. 56.** Gráfico Fap Lineal.

**Figura IV. 57.** Gráfico de Residuos vs Variable Motorización Estructura Logarítmica.

**Figura IV. 58.** Gráfico de Residuos vs Variable Exportaciones – Importaciones Estructura Logarítmica.

**Figura IV. 59.** Gráfico de Residuos vs Variable Paro Estructura Logarítmica.

**Figura IV. 60.** Gráfico de Residuos vs Variable Pobreza Estructura Logarítmica.

**Figura IV. 61.** Gráfico de Residuos vs Variable PIB Estructura Logarítmica.

**Figura IV. 62.** Gráfico de Residuos vs Variable Empleo Estructura Logarítmica.

**Figura IV. 63.** Gráfico de Residuos vs Variable Comercio Internacional Estructura Logarítmica.

**Figura IV. 64.** Gráfico de Residuos vs Variable Contaminación Estructura Logarítmica.

**Figura IV. 65.** Gráfico de Residuos vs Predicho de Matriculados Estructura Logarítmica.

**Figura IV. 66.** Gráfico de Residuos vs Número de Fila Estructura Logarítmica.

**Figura IV. 67.** Histograma de Residuos Estructura Logarítmica.

**Figura IV. 68.** Gráfico Fas Logarítmico.

**Figura IV. 69.** Gráfico Fap Logarítmico.

**Figura IV. 70.** Gráfico de Residuos vs Variable PIB Estructura Inversa.

**Figura IV. 71.** Gráfico de Residuos vs Variable Contaminación Estructura Inversa.

**Figura IV. 72.** Gráfico de Residuos vs Predicho Matriculados Estructura Inversa.

**Figura IV. 73.** Gráfico de Residuos vs Número de Filas Estructura Inversa.

**Figura IV. 74.** Histograma de Residuos Estructura Inversa.

**Figura IV. 75.** Gráfico Fas Inverso.

**Figura IV. 76.** Gráfico Fap Inverso.

**→ Listado de Tablas:**

**Tabla IV. 1.** Matriz de Datos Inicial (Parte A).

**Tabla IV. 2.** Matriz de Datos Inicial (Parte B).

**Tabla IV. 3.** Variable Vehículos Matriculados.

**Tabla IV. 4.** Variable Motorización.

**Tabla IV. 5.** Variable Exportación – Importación.

**Tabla IV. 6.** Variable Paro.

**Tabla IV. 7.** Variable Pobreza.

**Tabla IV. 8.** Variable PIB.

**Tabla IV. 9.** Variable Empleo.

**Tabla IV. 10.** Variable Comercio Internacional.

**Tabla IV. 11.** Variable Contaminación.

**Tabla IV. 12.** Matriz de Correlación.

**Tabla IV. 13.** Matriz de Correlación Inversa.

**Tabla IV. 14.** Resumen Análisis Regresión Múltiple Lineal.

**Tabla IV. 15.** Resumen Análisis Varianza Regresión Múltiple Lineal.

**Tabla IV. 16A.** Resumen Residuos Atípicos.

**Tabla IV. 16B.** Resumen Residuos Atípicos.

**Tabla IV. 16C.** Resumen Residuos Atípicos.

**Tabla IV. 16D.** Resumen Residuos Atípicos.

**Tabla IV. 17.** Resumen Análisis Regresión Múltiple Lineal Final.

**Tabla IV. 18.** Resumen Análisis Varianza Regresión Múltiple Lineal Final.

**Tabla IV. 19.** Test de Normalidad Lineal.

**Tabla IV. 20.** Heterocedasticidad Lineal.

**Tabla IV. 21.** Resumen Análisis Regresión Múltiple Logarítmica.

**Tabla IV. 22.** Resumen Análisis Varianza Regresión Múltiple Logarítmica.

**Tabla IV. 23.** Resumen Residuos Atípicos.

**Tabla IV. 24.** Test de la Normalidad Logarítmico.

**Tabla IV. 25.** Heterocedasticidad Logarítmica.

**Tabla IV. 26.** Heterocedasticidad Logarítmica Final.

**Tabla IV. 27.** Resumen Análisis Regresión Múltiple Inversa

**Tabla IV. 28.** Resumen Análisis Varianza Regresión Múltiple Inversa.

**Tabla IV. 29A.** Resumen Residuos Atípicos.

**Tabla IV. 29B.** Resumen Residuos Atípicos.

**Tabla IV. 29C.** Resumen Residuos Atípicos.

**Tabla IV. 29D.** Resumen Residuos Atípicos.

**Tabla IV. 30.** Resumen Análisis Regresión Múltiple Inversa Final.

**Tabla IV. 31.** Resumen Análisis Varianza Regresión Múltiple Inversa Final.

**Tabla IV. 32.** Test de Normalidad Inverso.

**Tabla IV. 33.** Heterocedasticidad Inversa.

**Tabla IV. 34.** Comparación 3 Modelos.

**Tabla IV. 35.** Predicción.

