



ÍNDICE

1	INTRODUCCIÓN Y OBJETIVOS	4
1.1	MOTIVACIÓN	5
1.2	OBJETIVOS	5
1.3	METODOLOGÍA	6
2	TÉCNICAS ESTADÍSTICAS	8
2.1	ANÁLISIS UNIVARIANTE	8
2.2	ANÁLISIS BIVARIANTE	12
2.3	ANÁLISIS MULTIVARIANTE	14
2.3.1	Regresión lineal múltiple	15
3	CÁLCULOS	18
3.1	ANÁLISIS UNIVARIANTE	18
3.2	ANÁLISIS BIVARIANTE	25
3.2.1	Comparación entre variables	25
3.2.2	Multicolinealidad	31
3.2.2	Resolución del problema de Multicolinealidad	33
3.3	ANÁLISIS MULTIVARIANTE	38
3.3.1	Modelo Teórico Propuesto e Interpretación de los Parámetros	38
3.3.2	Significatividad	39
3.3.3	Normalidad de los Residuos	42
3.3.4	Estudio de Gráficos	44
3.3.5	Heterocedasticidad	46
3.3.6	Autocorrelación	47
3.3.7	Puntos Influyentes y Residuos Atípicos	49
3.3.8	Resolución de los Problemas	51
3.3.9	R ² e Interpretación del Modelo	58
3.3.10	Predicción	58
4	DISCUSIÓN	62
5	ANÁLISIS DAFO DE LA CC.AA. ELEGIDA	66
6	CONCLUSIONES	73
7	FUTURAS LÍNEAS DE INVESTIGACIÓN	75
8	REFERENCIAS	77
9	ANEXO	79



Capítulo 1: Introducción y Objetivos

1 Introducción y Objetivos

Un modelo econométrico es aquel en el que se cuantifica la relación que liga una determinada variable (Y) con los valores de otras variables (X).

En este proyecto, se cuantifica la relación que existe entre la Población Ocupada con la Población Total, el Nº Empresas, el PIB per cápita, la Población Activa, la Población Parada y el IPC mediante una regresión múltiple, que es una parte fundamental del análisis multivariante. También se utilizan otras técnicas estadísticas como el análisis univariante y el análisis bivariante.

Los datos de estas variables se han extraído de la base de datos del Instituto Nacional de Estadística y corresponden al tercer trimestre del año 2014. (Ver **Tabla 23** en el Anexo)

Se considera que las variables elegidas explican muchas de las características económicas del país como el IPC (Índice de Precios al Consumo) que mide cómo evolucionan los precios de los bienes y servicios consumidos por la población residente en viviendas familiares en España; el PIB per cápita (Producto Interior Bruto per cápita) que indica la riqueza o el nivel de vida de la población; la Población Activa que mide cuántas personas componen el mercado de trabajo; la Población Parada que es el total de personas que tienen disposición de trabajar y estando en la edad, no realiza ninguna actividad laboral remunerada aunque activamente la buscan.

Una vez finalizado el análisis estadístico del modelo, se analiza también el macroentorno y el microentorno de aquella Comunidad Autónoma (CC.AA.) con menor Población Ocupada respecto a la variable Población Activa. La finalidad de este estudio concreto es saber cuáles son las oportunidades y amenazas y, las fortalezas y debilidades que existen en la Comunidad Autónoma elegida para así tener una visión más clara de la situación actual a la hora de emprender un negocio.

Por eso, definir el concepto de **emprendedurismo** es una cuestión clave. El emprendedurismo es convertir una idea en un proyecto, con o sin ánimo de lucro, aportando innovación y generando empleo.



1.1 Motivación

Uno de los motivos principales para efectuar este estudio es conocer la oportunidad de emprendedurismo que existe en las diferentes CC.AA., ya que si la tasa de ocupación de una CC.AA. es baja respecto a la tasa de actividad de la misma es porque no existe capacidad de empleo en dicha CC.AA., por lo que emprender en un negocio propio podría ser la clave para desatascar la situación actual de empleo de dicha CC.AA.

Por otro lado, podríamos decir que la experiencia del emprendedurismo nos puede llevar a cosechar grandes éxitos en la carrera profesional, ya que el grado de experiencia que se adquiere es muy elevado tanto en una actividad directiva, es decir, el dirigir tu propia empresa, como en una actividad más productiva como por ejemplo el ayudar en la cadena de producción de los bienes y servicios de la empresa.

1.2 Objetivos

El principal objetivo de este estudio es conocer la relación que existe entre una variable como es la Población Ocupada respecto a variables significativas para la sociedad en su conjunto, como por ejemplo el IPC, la Población Activa, el nº de Empresas totales, etc.

En segundo lugar, la idea de utilizar esta información es para que se pueda obtener una clara situación de empleabilidad que hay en nuestro país, es decir, en qué CC.AA. hay menos personas empleadas y, a su vez, en qué CC.AA. hay más personas empleadas.

Como último objetivo, saber en qué CC.AA. sería más factible emprender, teniendo en cuenta toda la serie de variables descritas en la **Tabla 23**, ya que a veces, no por ser la CC.AA. con mayor Población Total significa que ésta vaya a tener la mayor Población Activa y por consiguiente, vaya a ser la que cuente con mayor número de parados.



1.3 Metodología

La metodología usada en este proyecto se basará en una ciencia formal como es la estadística para analizar primeramente cada variable, lo que se llamará el análisis univariante. Posteriormente, se realizará un análisis bivariante donde se comparará la variable Población Ocupada (Y) con cada una de las variables independientes (X): Población Total, Nº de Empresas, PIB per cápita, Población Activa, Población Parada e IPC.

Antes de iniciar el análisis del modelo econométrico, será necesario comprobar si las variables independientes están relacionadas entre sí mediante el análisis de la multicolinealidad.

Una vez realizado este análisis y su posterior resolución, si existe el problema de multicolinealidad, se planteará el modelo econométrico con el que se realizará finalmente el estudio.

El estudio está compuesto por la formulación de una regresión lineal múltiple en la que analizaremos la significatividad de los parámetros y del modelo, la normalidad de los residuos, la heterocedasticidad, la autocorrelación, los puntos influyentes y los residuos atípicos. En caso de existir uno de estos problemas, se procederá a modificar y resolver el modelo econométrico para que se encuentre libre de cualquier problema anteriormente citado. Se acabará este estudio con la formulación de una predicción para el próximo trimestre, es decir, para el cuarto trimestre de 2014.

Para concluir este proyecto, se estudiará la situación actual de la CC.AA. con menor Población Ocupada mediante un análisis DAFO.



Capítulo 2: Técnicas Estadísticas

2 Técnicas Estadísticas

En este capítulo se van a describir brevemente las técnicas estadísticas utilizadas.

La estadística descriptiva es aquella que tiene por objetivo clasificar, presentar, describir, resumir y analizar datos relevantes a una o varias características sobre un conjunto de elementos.

2.1 Análisis Univariante

Es aquel análisis que se centra en una única característica del conjunto de elementos.

A este conjunto de elementos se le llamará **población**, y estará compuesta por individuos que son cada uno de los **componentes**.

La parte de la población que vaya a ser analizada se le llamará **muestra** y variará según la característica que se vaya a analizar.

Según el modelo, la variable estadística puede ser **cuantitativa o cualitativa** ya que depende de si toma o no, respectivamente, valor numérico y de si describe una cantidad o una cualidad.

Una variable estadística cuantitativa puede ser:

- **Discreta** si toma números finitos o numerables como por ejemplo el número total de estudiantes de la EPSA desde 1990 hasta la actualidad.
- **Continua** si toma número infinito o no numerable como por ejemplo el registro diario de las temperaturas en un año.

Una variable estadística cualitativa puede ser:

- **Ordinal** si los datos son ordenables como por ejemplo los días de la semana o los meses del año.
- **Nominal** si los datos no son ordenables como por ejemplo cuando los datos se agrupan según el género o la religión.

Los estadísticos son los valores en torno a los cuales se centran los datos.

Estos parámetros describen la posición y la dispersión de las observaciones.

Los principales **parámetros de POSICIÓN** son:

- **La media aritmética** es la suma total de los valores observados entre el total de datos.

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

- **La media geométrica** es la raíz enésima del producto de los valores observados.

$$MG = \bar{X} = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_n}$$

- **La mediana** es el valor central cuando se ordenan los valores de menor a mayor.

- ✓ Si n es impar $\rightarrow Me = x_{\left(\frac{n+1}{2}\right)}$

- ✓ Si n es par $\rightarrow Me = \frac{x_{\frac{n}{2}} + x_{\left(\frac{n}{2}+1\right)}}{2}$

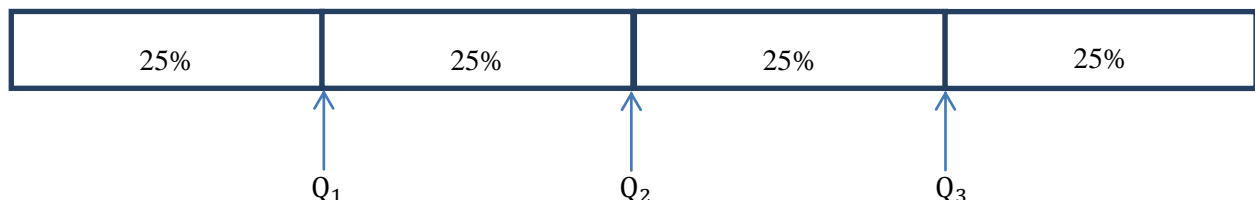
- **La moda** es el valor con mayor frecuencia.

- **Los cuartiles** dividen los valores observados en cuatro grupos. Éstos están ordenados de menor a mayor.

- ✓ Primer cuartil (Q_1), es el valor que separa el 25% de los valores de la muestra por la izquierda.

- ✓ Segundo cuartil (Q_2), es la mediana.

- ✓ Tercer cuartil (Q_3), es el valor que separa el 75% de los valores de la muestra por la izquierda.



Los principales **parámetros de DISPERSIÓN** son:

- **La varianza** proporciona el valor promedio de los cuadrados de las desviaciones alrededor de su media.

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{N - 1}$$

- **La desviación típica** es la raíz cuadrada de la varianza.

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N - 1}}$$

- **El rango intercuartílico** es la resta entre el tercer y el primer cuartil.

$$RI = Q_3 - Q_1$$

- **El error estándar de la media** es la desviación típica dividida entre la raíz cuadrada de la muestra.
- **El rango** es la diferencia entre el valor máximo y el valor mínimo.

$$Rango = (Max) - (Min)$$

- **El máximo** es el mayor valor que toma la variable.
- **El mínimo** es el valor más pequeño que toma la variable.

- **El coeficiente de asimetría** mide el grado de asimetría de la distribución en torno a su media. La distribución normal tomará valor 0. Una distribución que tome valor positivo significará que tiene una cola derecha larga. En cambio, una distribución que tome valor negativo significará que tiene una cola izquierda larga.

$$CA = \frac{\sum \frac{(X_i - \bar{X})^3}{(N - 1)}}{s^3}$$

- **El coeficiente de curtosis** mide el grado de agrupación de las observaciones en torno al punto central. La distribución normal tomará valor 0. Una curtosis positiva indica que se concentran más y tienen colas más largas que en una distribución normal. Una curtosis negativa indica que se agrupan menos y presentan colas más cortas.

$$CC = \frac{\sum \frac{(X_i - \bar{X})^4}{(N - 1)}}{s^4}$$

La **distribución de frecuencias de una variable** es el conjunto de sus categorías y las frecuencias de cada categoría. Para representarla, se utiliza el diagrama de caja y bigotes.

El diagrama de caja y bigotes (Box-Whisker) es la representación de los cuartiles, el mínimo valor y el máximo valor de un conjunto de datos en un rectángulo (caja) y dos segmentos (bigotes).

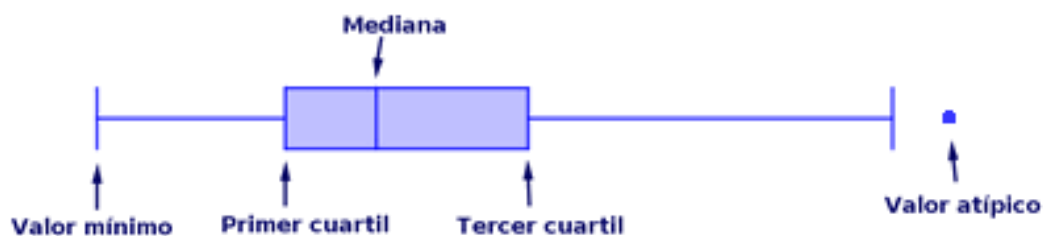


Ilustración 1. Diagrama Caja y Bigotes. Fuente: Google.

2.2 Análisis Bivariante

Es aquel análisis que se centra en dos características aleatorias del conjunto de elementos con el propósito de detectar posibles relaciones.

Describe el comportamiento de una variable, que será la dependiente (Y), en función de otra, que será la independiente (X).

- **La covarianza** es la medida de dependencia entre las variables. Si ésta es cero, afirma que las variables son independientes. Si es positiva quiere decir que la relación entre X e Y es directa; mientras que cuando es negativa, la relación es inversa.
 - ✓ **Relación directa** es cuando una variable crece, la otra también y viceversa.
 - ✓ **Relación inversa** es cuando una variable crece, la otra decrece y viceversa.

$$cov_{XY} = \frac{\sum(X_1 - \bar{X}) \cdot (Y_1 - \bar{Y})}{N - 1}$$

- **El coeficiente de correlación lineal** mide el grado de relación lineal que existe entre dos variables. Este valor siempre está comprendido entre -1 y 1, siendo un valor cercano a 0 el indicador de la inexistente o muy débil relación lineal.

$$R = \frac{\sum(X - \bar{X}) * (Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 * \sum(Y - \bar{Y})^2}}$$

Para representar este análisis, se utiliza el gráfico XY o diagrama de dispersión y la recta de regresión.

- **El diagrama de dispersión** es la representación en un plano; cuya abscisa será una variable y cuya ordenada será la otra, de cada observación con un punto en dicho plano.

- **La recta de regresión** es la que predice el valor que en promedio corresponde a una variable cuando la otra toma un valor determinado.

$$Y = a + b \cdot X$$

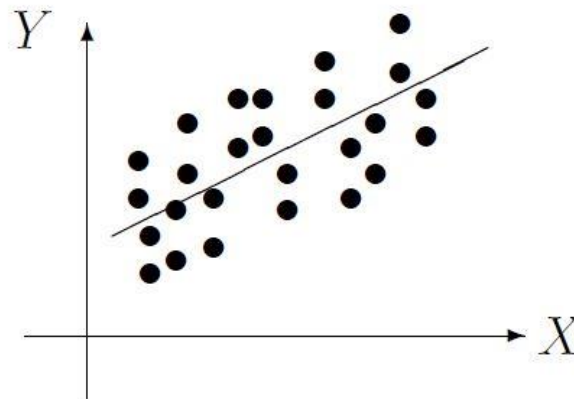


Ilustración 2. Diagrama de dispersión y Recta de regresión. Fuente: Google.

Para complementar este análisis, se puede utilizar el análisis de los residuos.

- **El residuo de un dato** (e_i) es la diferencia entre el valor que toma el mismo y el valor predicho para el valor medio de una variable con la recta de regresión.
- **El estudio de la Multicolinealidad** analiza si dos o más variables explicativas están relacionadas entre sí y toman valores semejantes o proporcionales.

Existen dos tipos de multicolinealidad:

- ✓ **La multicolinealidad exacta** que se da cuando la relación entre las variables explicativas es perfecta, es decir, cuando el $R^2 = 1$.
- ✓ **La multicolinealidad aproximada** que se da cuando en cualquier relación entre variables explicativas el R^2 sea menor a 1.

Con el fin de detectar un problema de relación entre variables, se utilizan tres métodos:

- ✓ **La Matriz de Correlación** que muestra los coeficientes que correlacionan las variables entre sí (coeficientes de correlación de Pearson). En este caso, existe el problema de multicolinealidad cuando un R_{X_i, X_i} en valor absoluto, por debajo de la diagonal supere un valor crítico de 0,7.

	X_1	X_2	X_3	X_4	...
X_1	1	R_{X_2, X_1}	R_{X_3, X_1}	R_{X_4, X_1}	R_{\dots, X_1}
X_2	R_{X_1, X_2}	1	R_{X_3, X_2}	R_{X_4, X_2}	R_{\dots, X_2}
X_3	R_{X_1, X_3}	R_{X_2, X_3}	1	R_{X_4, X_3}	R_{\dots, X_3}
X_4	R_{X_1, X_4}	R_{X_2, X_4}	R_{X_3, X_4}	1	R_{\dots, X_4}
...	$R_{X_1, \dots}$	$R_{X_2, \dots}$	$R_{X_3, \dots}$	$R_{X_4, \dots}$	1

- ✓ **La Matriz Inversa de Correlación** que muestra la relación de una variable explicativa con todas las demás. Se afirma la existencia de multicolinealidad cuando un R_{ii}^{-1} de la diagonal principal es mayor a 10.
- ✓ **El Método de Belsley, Kuh y Welsch** se basa en los autovalores de la matriz de correlación. Para ello, se define un Índice de Acondicionamiento (I.C.) en el que se analiza la relación de todas las variables a la vez. Si este I.C. está entre el intervalo $0 \leq I.C. \leq 10$, no existirá el problema. En cambio, si este I.C. se encuentra entre $10 \leq I.C. \leq 20$ el problema será moderado, mientras que si $I.C. > 20$, el problema será elevado.

Para resolver, en el caso de que exista, un problema de multicolinealidad se elimina la variable que, mayor R_{ii}^{-1} , de la diagonal principal, tenga en la Matriz Inversa de Correlación. Si estos valores R_{ii}^{-1} son muy cercanos, es decir, si la diferencia entre estos es menor a la unidad, se plantea un nuevo modelo eliminando, cada vez, una de las variables que genera el problema aceptando así, el modelo que mayor R^2 presente. Si ambos R^2 son iguales o no es posible calcularlos, se tendrá que obtener los Índices de Acondicionamiento con ambas variables y se eliminara la que de un I.C menor.

2.3 Análisis Multivariante

El análisis multivariante es el conjunto de métodos que tienen la función de **analizar a la vez** un conjunto de datos multivariantes en el que hay **varias variables medidas para cada individuo u objeto** a estudiar, es decir, relaciona la contribución de varios factores en un resultado.

Los factores pueden ser nombrados como **factores de riesgo, variables independientes o variables explicativas**. El resultado es la **variable dependiente** o la variable respuesta.

En Hair et al. (1999) decían:

“Las mujeres y hombres de negocios de hoy no pueden seguir aproximaciones ya pasadas en las que los consumidores eran considerados homogéneos y caracterizados por un número pequeño de variables demográficas. En su lugar, deben desarrollar estrategias que atraigan a numerosos segmentos de clientes con características demográficas y pictográficas diversas en un mercado con múltiples restricciones (legales, económicas, competitivas, tecnológicas, etc.). Sólo a través del análisis multivariante las relaciones múltiples de este tipo podrán ser examinadas adecuadamente para obtener un entendimiento más completo y real del entorno que permita tomar las decisiones más adecuadas.”

Existen diferentes **tipos de técnicas** multivariantes:

- **Métodos de Dependencia** que son lo que suponen que las variables son divididas en dependientes e independientes, además determina si el conjunto de las independientes afecta a las dependientes
- **Métodos de Interdependencia** que no suponen la división de las variables, sólo identifican la relación entre variables
- **Métodos Estructurales** que son una mezcla de los Métodos de dependencia y de los Métodos de interdependencia.

2.3.1 Regresión lineal múltiple

Se trata de cuantificar la influencia que ejercen las variables explicativas sobre una variable dependiente de carácter continuo, es decir, explica el comportamiento de una variable que será la que queramos explicar (variable dependiente = Y) en función de una serie de variables explicativas $X_1, X_2, X_3, \dots, X_k$.

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \dots + \beta_k \cdot X_k + U \quad \text{Siendo } U \text{ el término de perturbación.}$$

Para realizar un análisis de regresión lineal múltiple es necesario plantear una serie de consideraciones sobre los datos:



- **Linealidad:** si la variable dependiente sigue este modelo lineal.

$$Y = \beta \cdot X + U$$

- **Homocedasticidad:** si todas las perturbaciones tienen la misma varianza.

$$V(u_i) = \sigma^2$$

- **Independencia:** si son independientes entre sí las perturbaciones aleatorias.

$$E(u_i \cdot u_j) = 0, \forall i \neq j$$

- **Normalidad:** si la distribución de dicha perturbación es normal.

$$U \approx N(0, \sigma^2)$$

Una medida de la bondad del ajuste será la determinación del R^2 ya que indica la proporción de variabilidad que queda explicada por la regresión. Este valor estará entre el 0 y el 1, siendo mejor cuanto más próximo esté de 1.

Para medir la significatividad de los parámetros se usa el **estadístico t-Student** y el **estadístico F de Snedecor**.



Capítulo 3: Cálculos

3 Cálculos

3.1 Análisis Univariante

A. Población Total

Tabla 1. Análisis Univariante - Población Total

PARÁMETROS DE POSICIÓN		PARÁMETROS DE DISPERSIÓN	
Media Aritmética	2.480.510	Varianza	6.438.940.000.000
Media Geométrica	1.355.592	Desviación Típica	2.537.510
Mediana	1.472.050	Rango Intercuartílico	2.121.463
Moda	-	Error estándar de la Media	582.145
1er Cuartil	644.477	Rango	8.356.621
2do Cuartil	1.472.050	Máximo	8.440.300
3er Cuartil	2.765.940	Mínimo	83.679
		Coefficiente de Asimetría	2,46258
		Coefficiente de Curtosis	0,747243

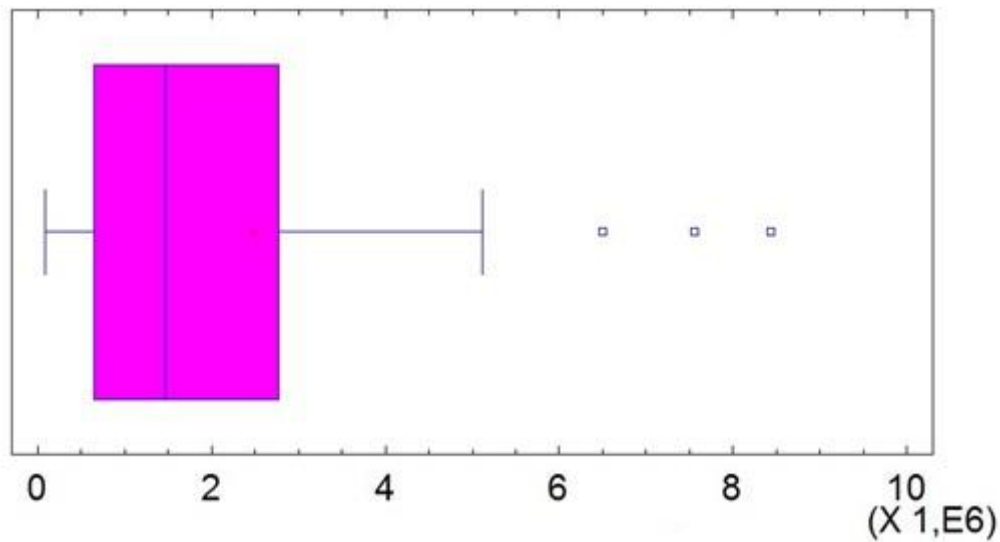


Ilustración 3. Gráfico Caja y Bigotes - Población Total

Destaca en la **Ilustración 3** los tres puntos anómalos que se observan que son primeramente la Comunidad de Madrid con aproximadamente 6,5 millones de habitantes, seguido por Cataluña con 7,5 millones de habitantes y finalmente, Andalucía con una población de más de 8 millones de habitantes.

B. Nº Empresas 2014

Tabla 2. Análisis Univariante - Nº Empresas 2014

PARÁMETROS DE POSICIÓN		PARÁMETROS DE DISPERSIÓN	
Media Aritmética	164.174	Varianza	30.262.000.000
Media Geométrica	84.352	Desviación Típica	173.960
Mediana	88.114	Rango Intercuartílico	150.163
Moda	-	Error estándar de la Media	39.909
1er Cuartil	41.582	Rango	572.975
2do Cuartil	88.114	Máximo	576.565
3er Cuartil	191.745	Mínimo	3.590
		Coefficiente de Asimetría	2,53108
		Coefficiente de Curtosis	0,799094

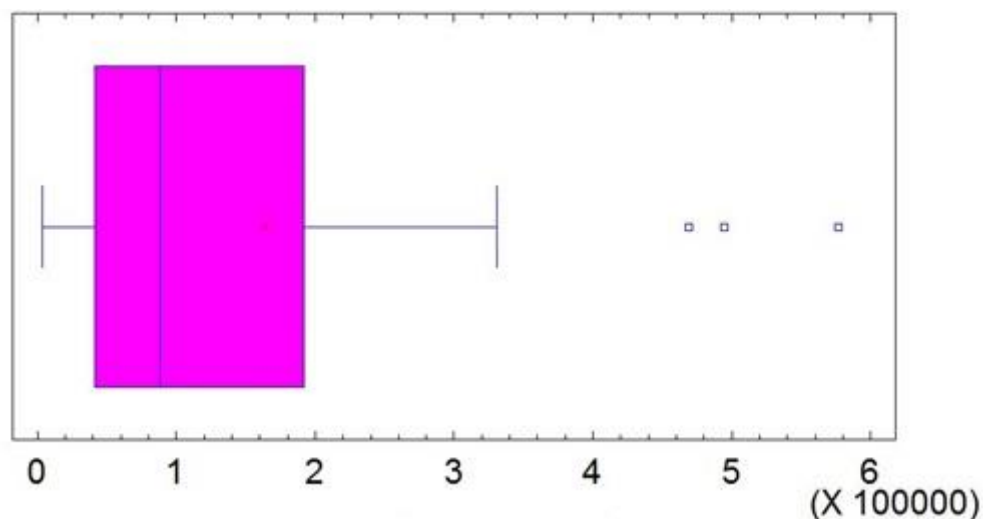


Ilustración 4. Gráfico Caja y Bigotes - Nº Empresas 2014

Destaca en la **Ilustración 4** los tres puntos anómalos que se observan que son de nuevo Andalucía con aproximadamente 470.000 empresas, seguido por la Comunidad de Madrid con 495.000 empresas y por último, Cataluña con 580.000 empresas.

C. PIB per cápita

Tabla 3. Análisis Univariante - PIB per cápita

PARÁMETROS DE POSICIÓN		PARÁMETROS DE DISPERSIÓN	
Media Aritmética	21.721,9	Varianza	20.342.300
Media Geométrica	21.290	Desviación Típica	4.510,24
Mediana	20.591	Rango Intercuartílico	7.376
Moda	-	Error estándar de la Media	1.035
1er Cuartil	17.901	Rango	14.933
2do Cuartil	20.591	Máximo	29.959
3er Cuartil	25.277	Mínimo	15.026
		Coficiente de Asimetría	0,798325
		Coficiente de Curtosis	-0,819241

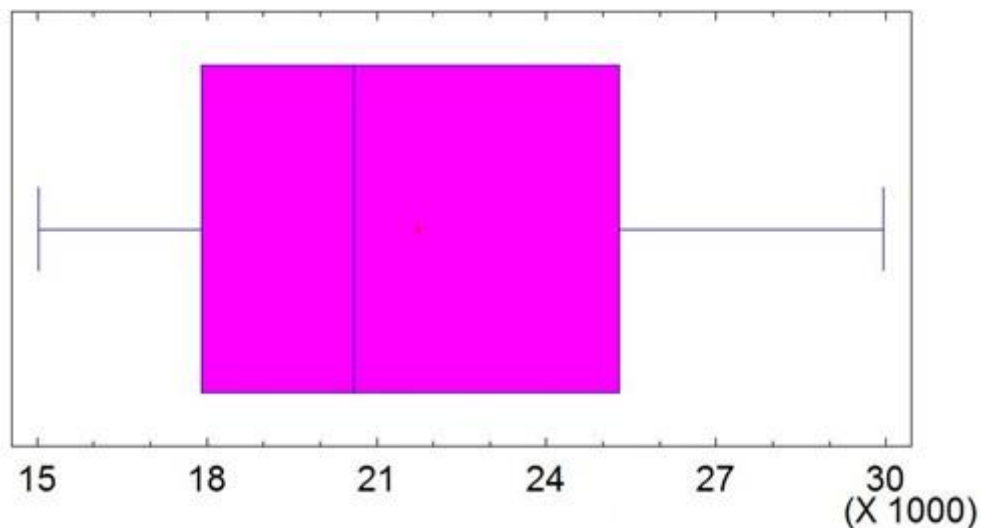


Ilustración 5. Gráfico Caja y Bigotes - PIB per cápita

Destaca en la **Ilustración 5** el tamaño de la caja, ya que cuanto más grande sea ésta más agrupados están los datos en de los cuartiles.

D. Población Activa

Tabla 4. Análisis Univariante - Población Activa

PARÁMETROS DE POSICIÓN		PARÁMETROS DE DISPERSIÓN	
Media Aritmética	1.477.600	Varianza	2.404.200.000.000
Media Geométrica	794.819	Desviación Típica	1.550.550
Mediana	895.742	Rango Intercuartílico	1.107.922
Moda	-	Error estándar de la Media	355.721
1er Cuartil	382.368	Rango	4.905.840
2do Cuartil	895.742	Máximo	4.951.080
3er Cuartil	1.490.290	Mínimo	45.237
		Coefficiente de Asimetría	2,53144
		Coefficiente de Curtosis	0,745147

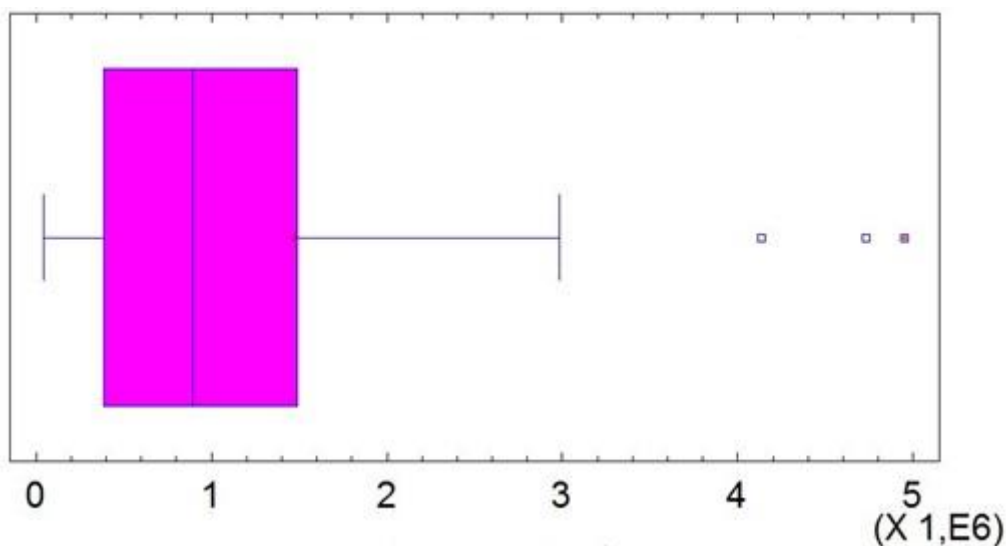


Ilustración 6. Gráfico Caja y Bigotes - Población Activa

Destaca en la **Ilustración 6** los tres puntos anómalos que se observan que son la Comunidad de Madrid con aproximadamente 4,1 millones de habitantes activos, seguido por Cataluña con 4,7 millones de habitantes activos y finalmente, Andalucía con una población activa de aproximadamente 4,9 millones de habitantes.

E. Población Parada

Tabla 5. Análisis Univariante - Población Parada

PARÁMETROS DE POSICIÓN		PARÁMETROS DE DISPERSIÓN	
Media Aritmética	589.277	Varianza	513.715.000.000
Media Geométrica	300.297	Desviación Típica	716.739
Mediana	333.355	Rango Intercuartílico	594.096
Moda	-	Error estándar de la Media	164.431
1er Cuartil	112.695	Rango	2.946.320
2do Cuartil	333.355	Máximo	2.971.830
3er Cuartil	706.791	Mínimo	25.514
		Coefficiente de Asimetría	4,12182
		Coefficiente de Curtosis	5,60638

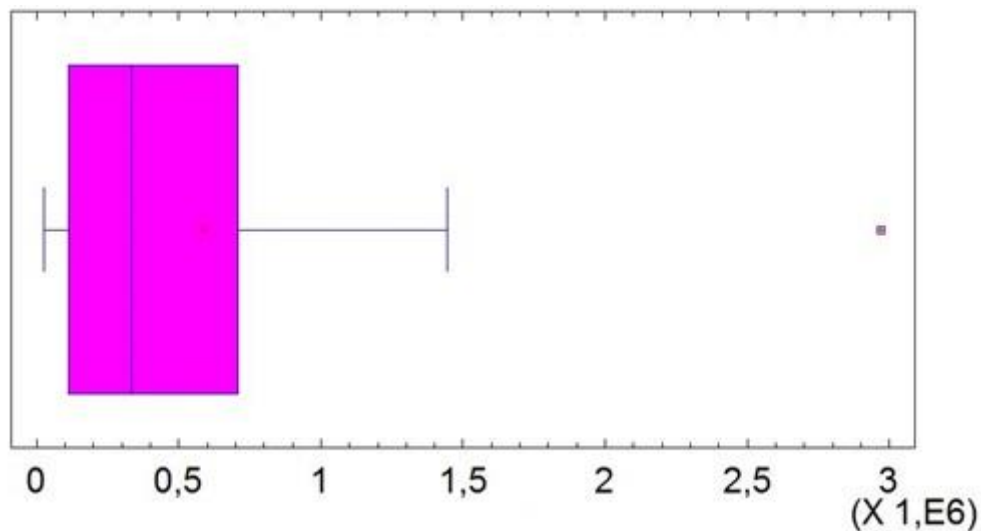


Ilustración 7. Gráfico Caja y Bigotes - Población Parada

Destaca en la **Ilustración 7** el único punto anómalo que se observa que es Andalucía con más de 2,9 millones de habitantes activos que están parados.

F. Población Ocupada

Tabla 6. Análisis Univariante - Población Ocupada

PARÁMETROS DE POSICIÓN		PARÁMETROS DE DISPERSIÓN	
Media Aritmética	921.263	Varianza	881.401.000.000
Media Geométrica	497.561	Desviación Típica	938.829
Mediana	532.700	Rango Intercuartílico	754.000
Moda	-	Error estándar de la Media	215.382
1er Cuartil	263.300	Rango	3.051.600
2do Cuartil	532.700	Máximo	3.074.800
3er Cuartil	1.017.300	Mínimo	23.200
		Coefficiente de Asimetría	2,4673
		Coefficiente de Curtosis	0,702388

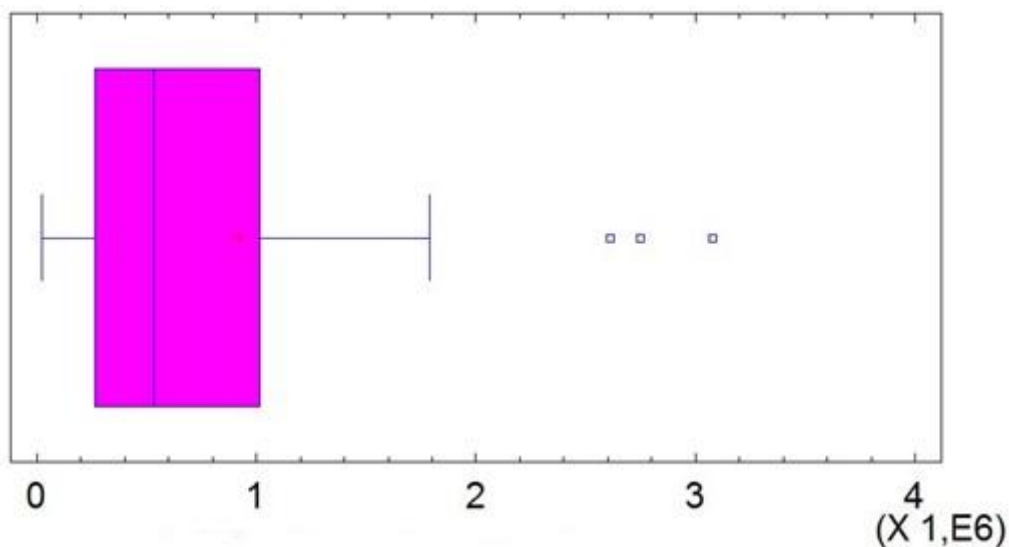


Ilustración 8. Gráfico Caja y Bigotes - Población Ocupada

Destaca en la **Ilustración 8** los tres puntos anómalos que se observan que son, en primer lugar Andalucía con un total aproximado de 2,6 millones de habitantes ocupados, seguido por la comunidad de Madrid con aproximadamente 2,7 millones de habitantes ocupados y, finalmente, Cataluña con una población ocupada de más de 3 millones de habitantes.

G. IPC

Tabla 7. Análisis Univariante - IPC

PARÁMETROS DE POSICIÓN		PARÁMETROS DE DISPERSIÓN	
Media Aritmética	103,924	Varianza	0,621616
Media Geométrica	103,921	Desviación Típica	0,788427
Mediana	103,852	Rango Intercuartílico	0,885
Moda	-	Error estándar de la Media	0,1809
1er Cuartil	103,68	Rango	3,026
2do Cuartil	103,852	Máximo	105,08
3er Cuartil	104,565	Mínimo	102,054
		Coefficiente de Asimetría	-1,51916
		Coefficiente de Curtosis	0,728187

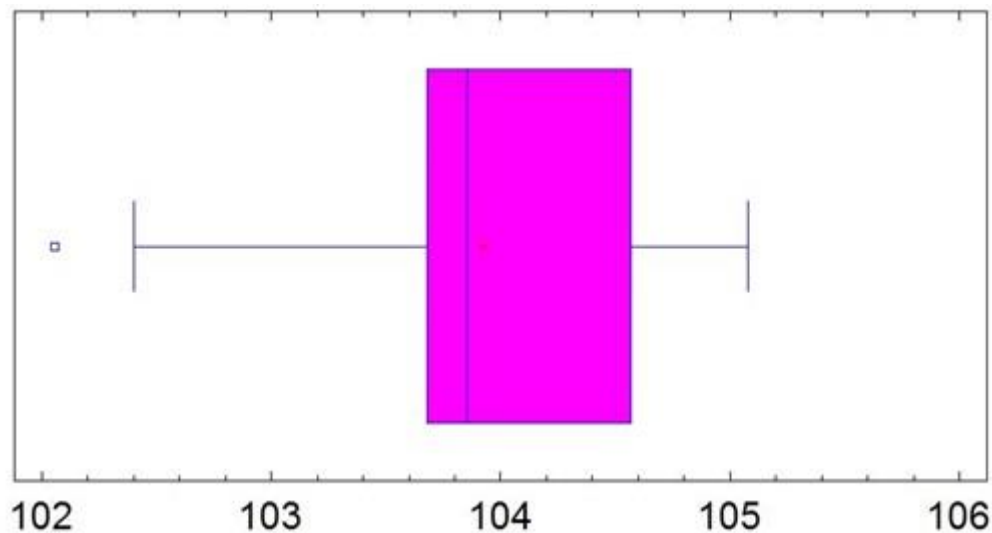


Ilustración 9. Gráfico Caja y Bigotes - IPC

Destaca en la **Ilustración 9** el punto anómalo que se observa que es la Ciudad Autónoma de Melilla donde el Índice de Precios al Consumo está por debajo del intervalo de todos los demás valores. El valor del IPC en Melilla es de 102,054 mientras que los otros valores se sitúan entre el 103 y el 105.

3.2 Análisis Bivariante

3.2.1 Comparación entre variables

- Población Ocupada

- ✓ **Población Ocupada – Población Total**

- Modelo Lineal Simple

$$\text{Población Ocupada} = \beta_1 * \text{Población Total} + \beta_0 + U$$

$$\text{Población Ocupada} = 0,364184 * \text{Población Total} + 17.900$$

- Análisis R^2

$$R^2 = 96,9\%$$

- Gráfico X, Y

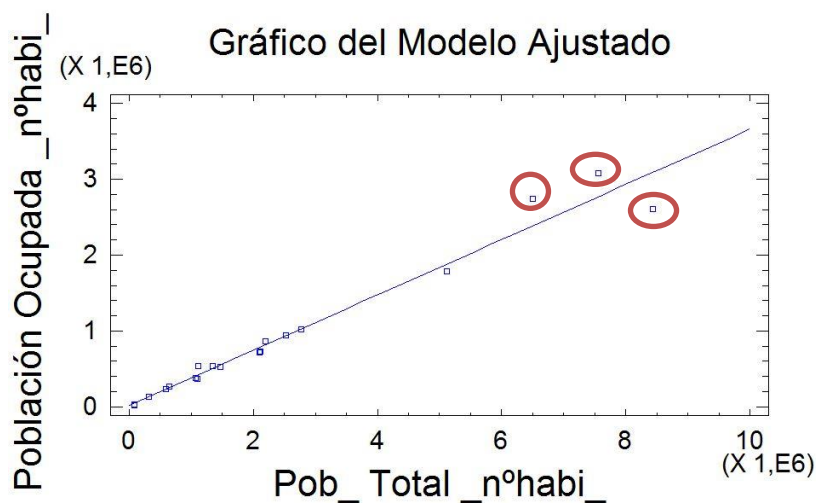


Ilustración 10. Gráfico Población Total (X), Población Ocupada (Y)

Destaca de la **Ilustración 10** la relación lineal que se observa, es decir, el modelo se ajusta muy bien. Además es posible destacar que esta relación es una dirección directa.

Los puntos que están redondeados (posibles puntos influyentes) pertenecen a las CCAA de Comunidad de Madrid, Cataluña y Andalucía.

✓ **Población Ocupada – Nº Empresas**

- Modelo Lineal Simple

$$\text{Población Ocupada} = \beta_1 * \text{N}^\circ \text{ Empresas} + \beta_0 + U$$

$$\text{Población Ocupada} = 5,39323 * \text{N}^\circ \text{ Empresas} + 35.834,7$$

- Análisis R^2

$$R^2 = 99,9\%$$

- Gráfico X, Y

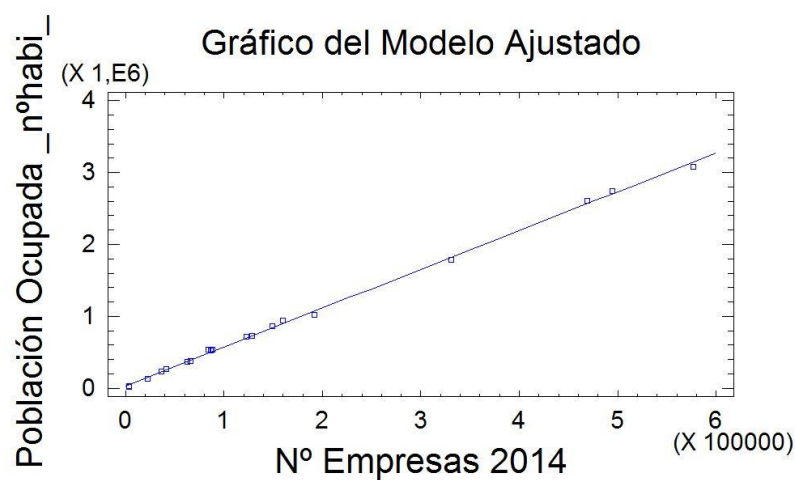


Ilustración 11. Gráfico Nº Empresas (X), Población Ocupada (Y)

En la **Ilustración 11** se observa que el modelo se ajusta casi a la perfección. Además, es posible destacar que esta relación es de proporcionalidad directa.

✓ **Población Ocupada – PIB**

- Modelo Lineal Simple

$$\text{Población Ocupada} = \beta_1 * \text{PIB per cápita} + \beta_0 + U$$

$$\text{Población Ocupada} = 50,2052 * \text{PIB per cápita} - 169.293$$

- Análisis R^2

$$R^2 = 5,8\%$$

- Gráfico X, Y

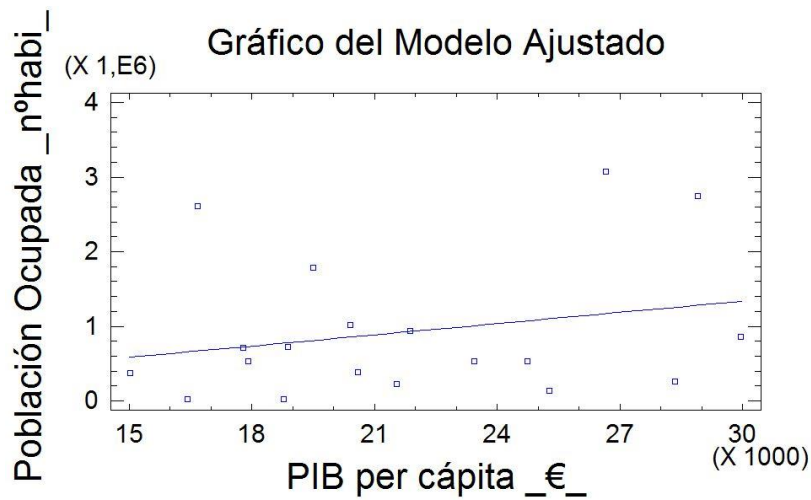


Ilustración 12. Gráfico PIB per cápita (X), Población Ocupada (Y)

Destacar de la **Ilustración 12** la no existencia de relación lineal.

✓ **Población Ocupada – Población Activa**

- Modelo Lineal Simple

$$Población\ Ocupada = \beta_1 * Población\ Activa + \beta_0 + U$$

$$Población\ Ocupada = 0,599883 * Población\ Activa + 34.878,9$$

- Análisis R^2

$$R^2 = 98,1\%$$

- Gráfico X, Y

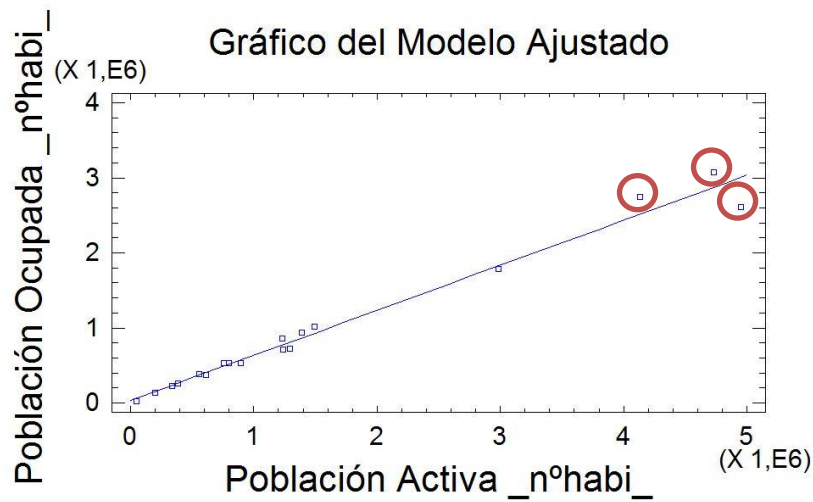


Ilustración 13. Gráfico Población Activa (X), Población Ocupada (Y)

Cabe destacar de la **Ilustración 13** que el modelo se ajusta muy bien. Además de que esta relación es de proporcionalidad directa.

Los puntos que están redondeados (posibles puntos influyentes) pertenecen a las CCAA de Comunidad de Madrid, Cataluña y Andalucía.

✓ **Población Ocupada – Población Parada**

- Modelo Lineal Simple

$$\text{Población Ocupada} = \beta_1 * \text{Población Parada} + \beta_0 + U$$

$$\text{Población Ocupada} = 1,11268 * \text{Población Parada} + 265.585$$

- Análisis R^2

$$R^2 = 72,1\%$$

- Gráfico X, Y

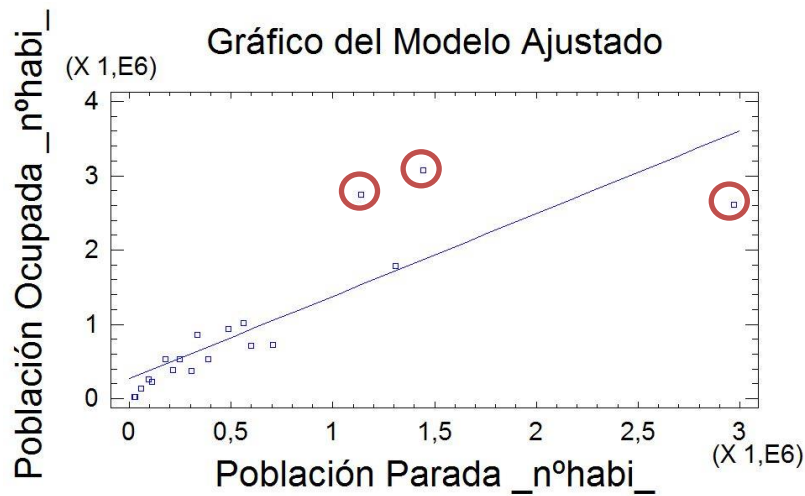


Ilustración 14. Gráfico Población Parada (X), Población Ocupada (Y)

De la **Ilustración 14** se puede destacar que hay una relación lineal muy elevada aunque hay una parte de los datos que no quedará explicada. Además se puede afirmar que está relación es de proporcionalidad directa.

Los puntos que están redondeados (posibles puntos influyentes) pertenecen a las CC.AA. de Comunidad de Madrid, Cataluña y Andalucía.

✓ **Población Ocupada – IPC**

- Modelo Lineal Simple

$$\text{Población Ocupada} = \beta_1 * \text{IPC} + \beta_0 + U$$

$$\text{Población Ocupada} = 356.591 * \text{IPC} - 36.137.000$$

- Análisis R^2

$$R^2 = 8,9\%$$

- Gráfico X, Y

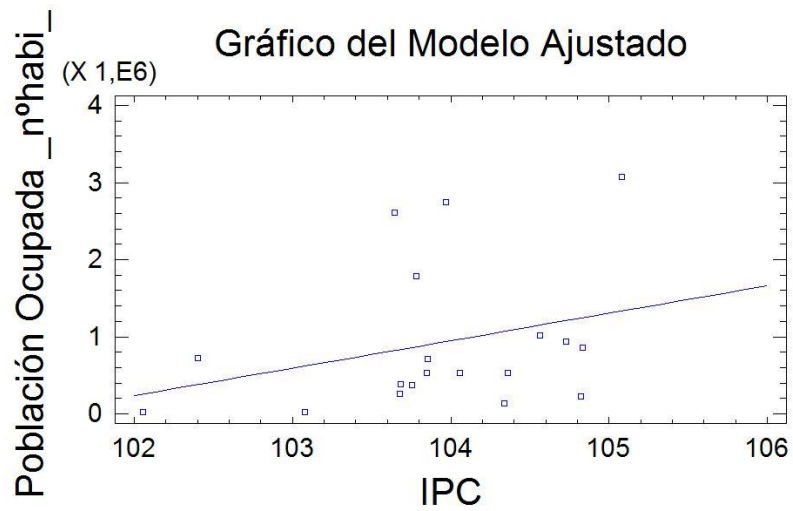


Ilustración 15. Gráfico IPC (X), Población Ocupada (Y)

Destacar de la **Ilustración 15** la no existencia de relación lineal.

3.2.2 Multicolinealidad

En este apartado se va a proceder a realizar el análisis de la multicolinealidad que constará de la realización de la matriz de correlación, la matriz inversa de correlación y el índice de acondicionamiento. Si alguno de estos tres métodos da positivo en problemas de correlación entre variables, se procederá a su solución.

✓ *Matriz de Correlación*

	<i>Población Total</i>	<i>Nº Empresas</i>	<i>PIB per cápita</i>	<i>Población Activa</i>	<i>Población Parada</i>	<i>IPC</i>
Población Total	1	0,9825	0,1305	0,998	0,9276	0,2449
Nº Empresas	0,9825	1	0,2374	0,989	0,8445	0,297
PIB per cápita	0,1305	0,2374	1	0,1558	-0,1201	0,5098
Población Activa	0,998	0,989	0,1558	1	0,9124	0,2424
Población Parada	0,9276	0,8445	-0,1201	0,9124	1	0,068
IPC	0,2449	0,297	0,5098	0,2424	0,068	1

Como se observa en los coeficientes que están por debajo de la diagonal principal, existe un alta correlación entre Población Total con Nº Empresas, Población Activa y Población Parada. Además, también entre Nº Empresas con Población Activa y Población Parada y, Población Activa con Población Parada.

Por tanto, se concluye que hay multicolinealidad porque estos R_{ii} (en letra roja), en valor absoluto, son superiores al valor crítico de 0,7.

✓ *Matriz Inversa de Correlación*

Población Total	1.018,725	-705,812	-4,495	25,225	-371,212	-18,439
Nº Empresas	-705,812	1.069,432	13,062	-783,676	467,713	6,731
PIB per cápita	-4,495	13,062	2,198	-15,798	7,858	-0,604
Población Activa	25,225	-783,676	-15,7983	1.022,736	-297,096	6,919
Población Parada	-371,212	467,713	7,858	-297,096	222,033	4,910
IPC	-18,439	6,731	-0,604	6,919	4,9107	1,813

Si observando los elementos de la diagonal principal, existe alguno con un valor superior a 10 (todos aquellos marcados en rojo) se afirmará que existe una elevada correlación entre la variable Población Total y el resto de variables, entre la variable N^o de empresas y el resto de variables, entre la variable Población Activa y el resto de variables y, entre la variable Población Parada y el resto de variables.

Por tanto, se puede concluir que existe una multicolinealidad de tipo una variable con todas las demás.

✓ **Índice de Acondicionamiento**

Tabla 8. Análisis Multicolinealidad – Autovalores

Autovalores	3,92104	1,4944	0,492034	0,0910838	0,00096902	0,00047368
--------------------	---------	--------	----------	-----------	------------	------------

$$\text{Índice de Acondicionamiento} = \sqrt{\frac{3,92104}{0,00047368}} = 90,982$$

Cómo el valor de este índice es mayor a 20, se concluye afirmando la elevada multicolinealidad de tipo todas las variables entre todas.

3.2.2 Resolución del problema de Multicolinealidad

Eliminación de parámetros

Para resolver el problema de multicolinealidad, se planteará primeramente un modelo sin la variable que mayor R_{ii} tiene en la diagonal principal de la Matriz Inversa de Correlación. En el caso de este estudio, se eliminará primeramente la variable **Nº Empresas**.

✓ **Matriz de Correlación**

	Población Total	PIB per cápita	Población Activa	Población Parada	IPC
Población Total	1	0,1305	0,998	0,9276	0,2449
PIB per cápita	0,1305	1	0,1558	-0,1201	0,5098
Población Activa	0,998	0,1558	1	0,9124	0,2424
Población Parada	0,9276	-0,1201	0,9124	1	0,068
IPC	0,2449	0,5098	0,2424	0,068	1

Como se observa en los coeficientes que están por debajo de la diagonal principal, existe un alta correlación entre Población Total con Población Activa y Población Parada. Además, también entre Población Activa con Población Parada.

Por tanto, se concluye que hay multicolinealidad porque estos R_{ii} (en letra roja), en valor absoluto, son superiores al valor crítico de 0,7.

✓ **Matriz Inversa de Correlación**

Población Total	552,897	4,125	-491,991	-62,527	-13,997
PIB per cápita	4,125	2,039	-6,225	2,145	-0,686
Población Activa	-491,991	-6,225	448,4610	45,642	11,852
Población Parada	-62,527	2,145	45,642	17,480	1,966
IPC	-13,997	-0,686	11,852	1,966	1,771

Si observando los elementos de la diagonal principal, existe alguno con un valor superior a 10 (todos aquellos marcados en rojo) se afirmará que existe una elevada correlación entre la variable Población Total y el resto de variables, entre la variable Población Activa y el resto de variables y, entre la variable Población Parada y el resto de variables.

Por tanto, se puede concluir que existe una multicolinealidad de tipo una variable con todas las demás.

✓ Índice de Acondicionamiento

Tabla 9. Análisis Multicolinealidad - Autovalores Modelo sin variable Nº Empresas

Autovalores	2,96441	1,49324	0,484951	0,0564266	0,00096699
--------------------	---------	---------	----------	-----------	------------

$$\text{Índice de Acondicionamiento} = \sqrt{\frac{2,96441}{0,00096699}} = 55,367$$

Cómo el valor de este índice es mayor a 20, se concluye afirmando la elevada multicolinealidad de tipo todas las variables entre todas.

Como aún eliminando la variable Nº Empresas se sigue apreciando el problema, se pasará a eliminar otra variable. En este segundo caso, la variable a eliminar es **Población Total** ya que es la que mayor R_{ii}^{-1} presenta en la Matriz Inversa de Correlación.

✓ Matriz de Correlación

	<i>PIB per cápita</i>	<i>Población Activa</i>	<i>Población Parada</i>	<i>IPC</i>
PIB per cápita	1	0,1558	-0,1201	0,5098
Población Activa	0,1558	1	0,9124	0,2424
Población Parada	-0,1201	0,9124	1	0,068
IPC	0,5098	0,2424	0,068	1

Como se observa en los coeficientes que están por debajo de la diagonal principal, existe un alta correlación entre Población Activa con Población Parada.

Por tanto, se concluye que hay multicolinealidad porque este R_{ii} (en letra roja), en valor absoluto, es superior al valor crítico de 0,7.

✓ **Matriz Inversa de Correlación**

PIB per cápita	2,008	-2,554	2,611	-0,582
Población Activa	-2,554	10,665	-9,997	-0,603
Población Parada	2,611	-9,997	10,409	0,383
IPC	-0,582	-0,603	0,383	1,416

Si observando los elementos de la diagonal principal, existe alguno con un valor superior a 10 (todos aquellos marcados en rojo) se afirmará que existe una elevada correlación entre la variable Población Activa y el resto de variables y, entre la variable Población Parada y el resto de variables.

Por tanto, se puede concluir que existe una multicolinealidad de tipo una variable con todas las demás.

✓ **Índice de Acondicionamiento**

Tabla 10. Análisis Multicolinealidad - Autovalores Modelo sin variable N° Empresas y Población Total

Autovalores	1,99051	1,48072	0,481697	0,0470769
--------------------	---------	---------	----------	-----------

$$\text{Índice de Acondicionamiento} = \sqrt{\frac{1,99051}{0,0470769}} = 6,502$$

Cómo el valor de este índice es menor a 10, se concluye afirmando que no existe el problema de multicolinealidad.

Como dos de los tres análisis afirman que existe el problema de multicolinealidad, se deberá eliminar otra variable. En este caso, como las dos variables tienen un número de R_{ii}^{-1} muy igualado, se plantearán dos modelos del que se elegirá el que mayor R^2 presente.

Modelo n°1 cuyo $R^2 = 84,63\%$

$$\text{Población Ocupada} = \beta_0 + \beta_1 * \text{PIB per Cápita} + \beta_2 * \text{Población Parada} + \beta_3 * \text{IPC} + U$$

Modelo nº2 cuyo $R^2 = 98,97\%$

$$\text{Población Ocupada} = \beta_0 + \beta_1 * \text{PIB per Cápita} + \beta_2 * \text{Población Activa} + \beta_3 * \text{IPC} + U$$

Como se puede comprobar, se elegirá el Modelo nº 2 para seguir con el estudio de la multicolinealidad.

✓ Matriz de Correlación

	<i>PIB per cápita</i>	<i>Población Activa</i>	<i>IPC</i>
<i>PIB per cápita</i>	1	0,1558	0,5098
<i>Población Activa</i>	0,1558	1	0,2424
<i>IPC</i>	0,5098	0,2424	1

Como se observa en los coeficientes que están por debajo de la diagonal principal, ninguno de estos R_{ii} es mayor, en valor absoluto, al valor crítico de 0,7, por lo que se concluye que no existe el problema de multicolinealidad.

✓ Matriz Inversa de Correlación

<i>PIB per cápita</i>	1,353	-0,046	-0,678
<i>Población Activa</i>	-0,046	1,064	-0,234
<i>IPC</i>	-0,678	-0,234	1,402

En la diagonal principal de esta matriz ninguno de los valores es superior a 10, por lo que no existe el problema de multicolinealidad.

✓ **Índice de Acondicionamiento****Tabla 11. Análisis Multicolinealidad - Autovalores Modelo nº 2**

Autovalores	1,63528	0,883045	0,48168
--------------------	---------	----------	---------

$$\text{Índice de Acondicionamiento} = \sqrt{\frac{1,63528}{0,48168}} = 1,842$$

Cómo el valor de este índice es menor a 10, se concluye afirmando que no existe el problema de multicolinealidad.

Con el Modelo nº 2, el problema de Multicolinealidad estaría resuelto, es decir, las variables PIB per Cápita, Población e IPC no presentan problemas de correlación entre ellas.

3.3 Análisis Multivariante

3.3.1 Modelo Teórico Propuesto e Interpretación de los Parámetros

El modelo teórico propuesto es el modelo nº 2 anteriormente citado, ya que está libre de multicolinealidad.

$$\text{Población Ocupada} = \beta_0 + \beta_1 * \text{PIB per Cápita} + \beta_2 * \text{Población Activa} + \beta_3 * \text{IPC} + U$$

En este caso, el significado de cada uno de los parámetros es:

- β_0 : constante del modelo. Valor medio de la Población Ocupada cuando todas las variables explicativas valen cero. (*nº habitantes*)
- β_1 : incremento medio de la Población Ocupada cuando aumenta en una unidad el PIB per Cápita mientras el resto de variables explicativas permanecen constantes.
 $\left(\frac{\text{nº habitantes}^2}{\text{€}}\right)$
- β_2 : incremento medio de la Población Ocupada cuando aumenta en una unidad la Población Activa mientras el resto de variables explicativas permanecen constantes.
(*adimensional*)
- β_3 : incremento medio de la Población Ocupada cuando aumenta en una unidad el IPC mientras el resto de las variables explicativas permanecen constantes.
 $\left(\frac{\text{nº habitantes}}{\text{índice}}\right)$
- U : perturbación aleatoria que recoge la influencia de las variables explicativas no contempladas sobre la Población Ocupada.

3.3.2 Significatividad

- **Todas las variables**

$$\text{Población Ocupada} = -3.305.460 + 16,0727 * \text{PIB per Cápita} + 0,589032 * \text{Población Activa} + 28.937 * \text{IPC}$$

✓ **Contraste de significatividad global (Modelo)**

$$\left. \begin{array}{l} H_0: \beta_1 = \beta_2 \dots = 0 \\ H_1: \text{alguno de los } \beta \neq 0 \end{array} \right\} \alpha (0,05)$$

El P-Valor de modelo es igual a 0,000. Al ser menor que $\alpha = 0,05$, se rechaza la hipótesis nula H_0 , por lo que el modelo en sí resulta significativo.

Fuente	Suma de cuadrados	GL	Cuadrado medio	Cociente-F	P-Valor
Modelo	1,57024E13	3	5,23412E12	482,07	0,0000
Residuo	1,62863E11	15	1,08575E10		
Total (Corr.)	1,58652E13	18			

Ilustración 16. Contraste de Significatividad Global

✓ **Contraste de significatividad individual (Parámetros)**

$$\left. \begin{array}{l} H_0: \beta_0 = 0 \\ H_1: \beta_0 \neq 0 \end{array} \right\} \alpha (0,05) \qquad \left. \begin{array}{l} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{array} \right\} \alpha (0,05)$$

$$\left. \begin{array}{l} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{array} \right\} \alpha (0,05) \qquad \left. \begin{array}{l} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{array} \right\} \alpha (0,05)$$

- Puesto que el P-Valor de β_0 es mayor a α ($0,3937 > 0,05$), se acepta la hipótesis nula. Por lo tanto, el parámetro β_0 no es significativo para el modelo. (**Ilustración 17**).
- Puesto que el P-Valor de β_1 es menor a α ($0,0228 < 0,05$), se rechaza la hipótesis nula. Por lo tanto, la variable PIB per Cápita sí resulta significativa para el modelo. (**Ilustración 17**).

- Puesto que el P-Valor de β_2 es menor a α ($0,0000 < 0,05$), se rechaza la hipótesis nula. Por lo tanto, la variable Población Activa sí resulta significativa para el modelo. (**Ilustración 17**).
- Puesto que el P-Valor de β_3 es mayor a α ($0,4451 > 0,05$), se acepta la hipótesis nula. Por lo tanto, la variable IPC no es significativa para el modelo. (**Ilustración 17**).

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	-3,30546E6	3,76376E6	-0,878233	0,3937
PIB per cápita _€	16,0727	6,33455	2,5373	0,0228
Población Activa	0,589032	0,0163386	36,0515	0,0000
IPC	28937,0	36894,7	0,784313	0,4451

Ilustración 17. Contraste de Significatividad Individual

Al presentarse una variable no significativa, se procede a la eliminación de la variable IPC del modelo, quedando como modelo final el siguiente, cuyo R^2 es 98,93%:

$$\text{Población Ocupada} = \beta_0 + \beta_1 * \text{PIB per Cápita} + \beta_2 * \text{Población Activa} + U$$

- Selección Hacia Delante

Con la selección hacia delante se observa el orden en el que se introducen las variables en el modelo. Si se mide la importancia con la que las variables llegan a formar parte del modelo, se obtiene que la variable Población Activa entra en el primer paso, siendo la variable más importante y, en el segundo paso, el PIB per Cápita, haciendo que la variable IPC quede excluida del modelo. El nuevo modelo ya estimado es:

$$\text{Población Ocupada} = -355.013 + 18,52 * \text{PIB per Cápita} + 0,591489 * \text{Población Activa}$$

✓ Contraste de significatividad global (Modelo)

$$\left. \begin{array}{l} H_0: \beta_1 = \beta_2 \dots = 0 \\ H_1: \text{alguno de los } \beta \neq 0 \end{array} \right\} \alpha (0,05)$$

El P-Valor de modelo es igual a 0,000. Al ser menor que $\alpha = 0,05$, se rechaza la hipótesis nula, por lo que el modelo en sí resulta significativo. (**Ilustración 18**).

Fuente	Suma de cuadrados	GL	Cuadrado medio	Cociente-F	P-Valor
Modelo	1,56957E13	2	7,84784E12	740,62	0,0000
Residuo	1,69542E11	16	1,05964E10		

Ilustración 18. Contraste de Significatividad Global - Selección Hacia Delante

✓ **Contraste de significatividad individual (Parámetros)**

$$\left. \begin{array}{l} H_0: \beta_0 = 0 \\ H_1: \beta_0 \neq 0 \end{array} \right\} \alpha (0,05) \qquad \left. \begin{array}{l} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{array} \right\} \alpha (0,05)$$

$$\left. \begin{array}{l} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{array} \right\} \alpha (0,05)$$

- Puesto que el P-Valor de β_0 es menor a α ($0,0089 < 0,05$), se rechaza la hipótesis nula. Por lo tanto, el parámetro β_0 sí es significativo para el modelo. (**Ilustración 19**).
- Puesto que el P-Valor de β_1 es menor a α ($0,0037 < 0,05$), se rechaza la hipótesis nula. Por lo tanto, la variable PIB per Cápita sí resulta significativa para el modelo. (**Ilustración 19**).
- Puesto que el P-Valor de β_2 es menor a α ($0,0000 < 0,05$), se rechaza la hipótesis nula. Por lo tanto, la variable Población Activa sí resulta significativa para el modelo. (**Ilustración 19**).

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	-355013,0	119319,0	-2,97531	0,0089
PIB per cápita_€	18,52	5,44602	3,40064	0,0037
Población Activa	0,591489	0,0158414	37,3381	0,0000

Ilustración 19. Contraste de Significatividad Individual - Selección Hacia Delante

3.3.3 Normalidad de los Residuos

- **Gráfico Probabilístico**

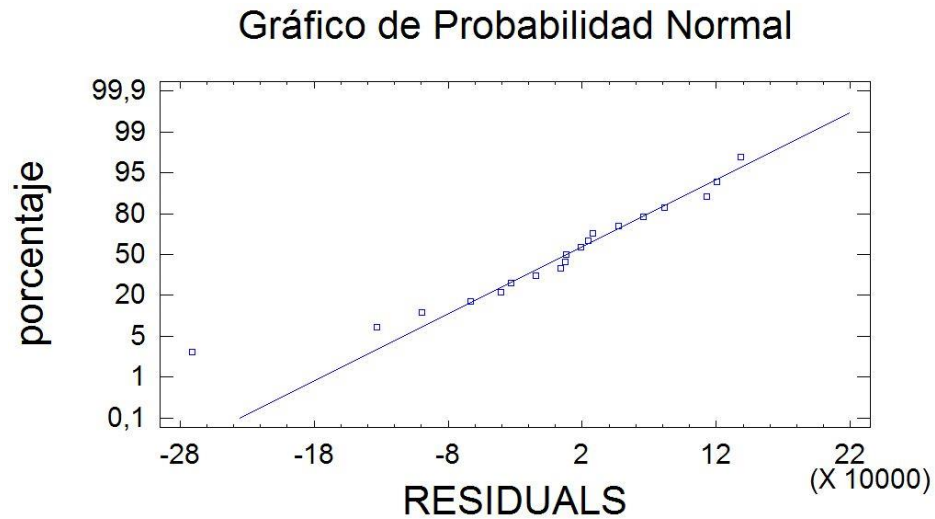


Ilustración 20. Análisis de la Normalidad de los Residuos - Gráfico Probabilístico

Como se observa en la **Ilustración 20**, los residuos muestran una linealidad, es decir, los residuos más o menos siguen una distribución normal aunque cabe destacar un punto que está muy alejado de la recta que posiblemente sea un punto atípico (Andalucía).

- **Histograma**

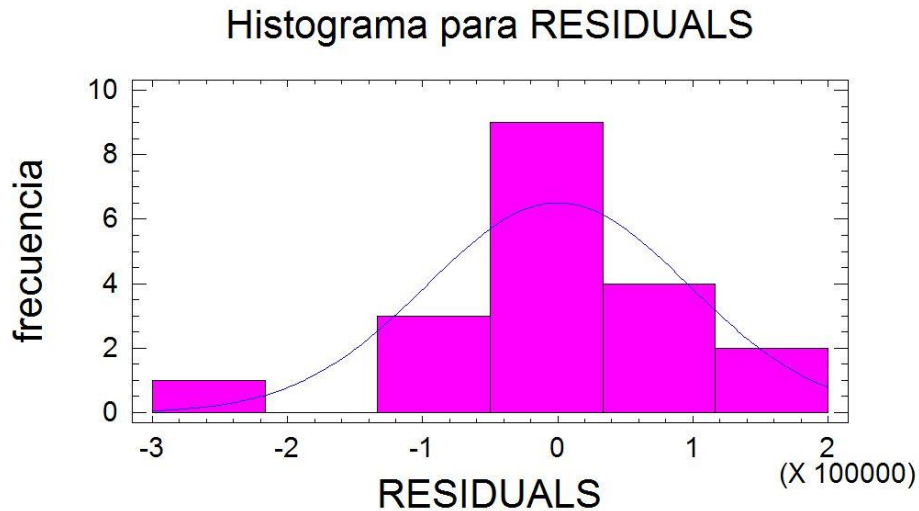


Ilustración 21. Análisis de la Normalidad de los Residuos – Histograma

Como se puede observar en la **Ilustración 21**, los residuos se distribuyen con normalidad ya que más o menos forman una campana de Gauss, además de que la mayoría de los residuos se centran sobre la media.

- **Contraste de Hipótesis**

$$\left. \begin{array}{l} H_0: \text{Los residuos son normales} \\ H_1: \text{Los residuos no son normales} \end{array} \right\} \alpha (0,05)$$

En el contraste de hipótesis se observa como el P-Valor de la puntuación Z para asimetría (que es el menor de todos los test para la normalidad) es mayor a α ($0,132804 > 0,05$), por lo que se acepta la hipótesis nula que afirma que los residuos se distribuyen de forma normal. (**Ilustración 22**).

Tests para la Normalidad para RESIDUALS

Estadístico chi-cuadrado de bondad de ajuste = 6,89474
P-valor = 0,648078

Estadístico W de Shapiro-Wilks = 0,926263
P-valor = 0,152981

Puntuación Z para asimetría = 1,50313
P-valor = 0,132804

Puntuación Z para curtosis no calculada.

Ilustración 22. Contraste para la Normalidad de los Residuos

3.3.4 Estudio de Gráficos

- **Residuos frente X**

✓ **PIB per Cápita**

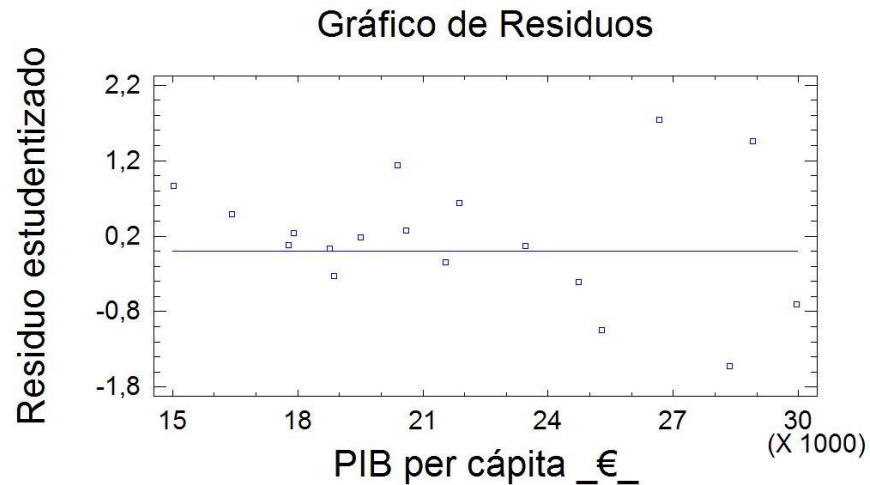


Ilustración 23. Estudio de Gráficos - Residuos frente PIB per Cápita

Como se puede observar en la **Ilustración 23**, no es destacable ningún problema de falta de linealidad o de heterocedasticidad. En esta ilustración, a simple vista, los residuos se distribuyen de forma aleatoria.

✓ **Población Activa**

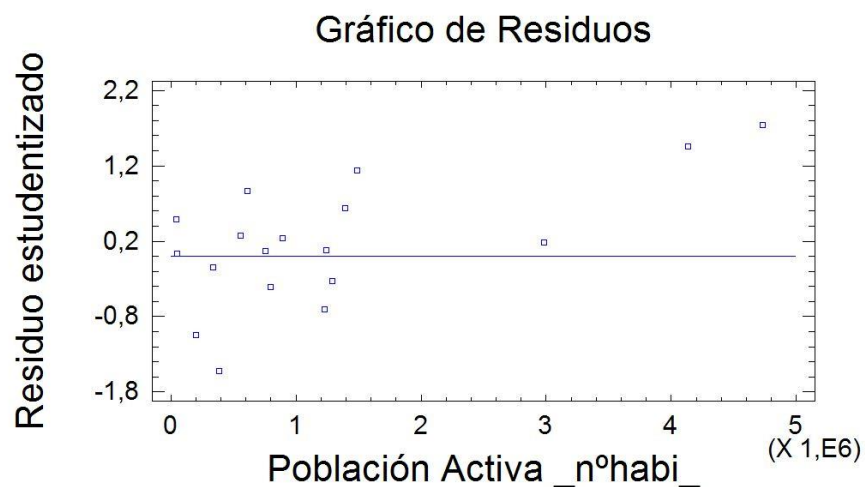


Ilustración 24. Estudio de Gráficos - Residuos frente Población Activa

Como se puede observar en la **Ilustración 24**, no es destacable ningún problema de falta de linealidad o de heterocedasticidad. También se observa la presencia de puntos atípicos: Comunidad de Madrid y Cataluña. En esta ilustración, a simple vista, los residuos se distribuyen de forma aleatoria.

- **Residuos frente Predicho**

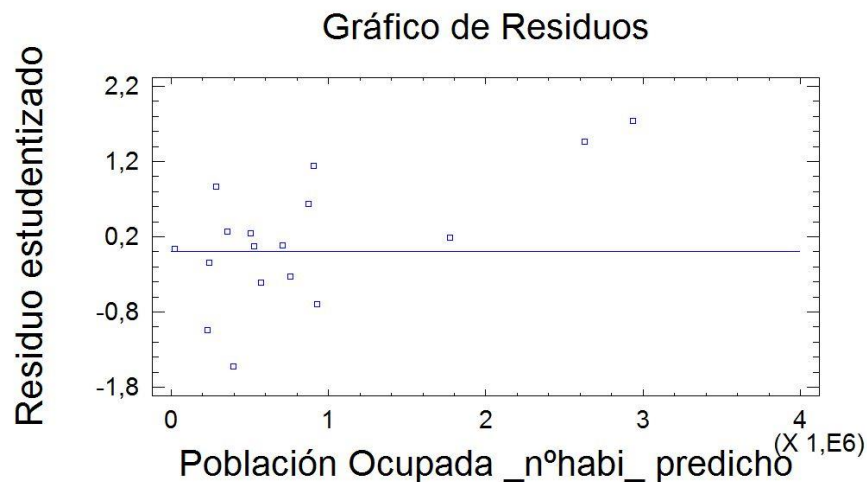


Ilustración 25. Estudio de Gráficos - Residuos frente Predicho

En la **Ilustración 25**, no se destaca ningún problema de falta de linealidad o de heterocedasticidad. En esta ilustración también es posible observar la presencia de puntos atípicos: Comunidad de Madrid y Cataluña. En este caso, no se observa que formas en los residuos que se puedan asociar a estos problemas.

- **Residuos frente Número de Fila**

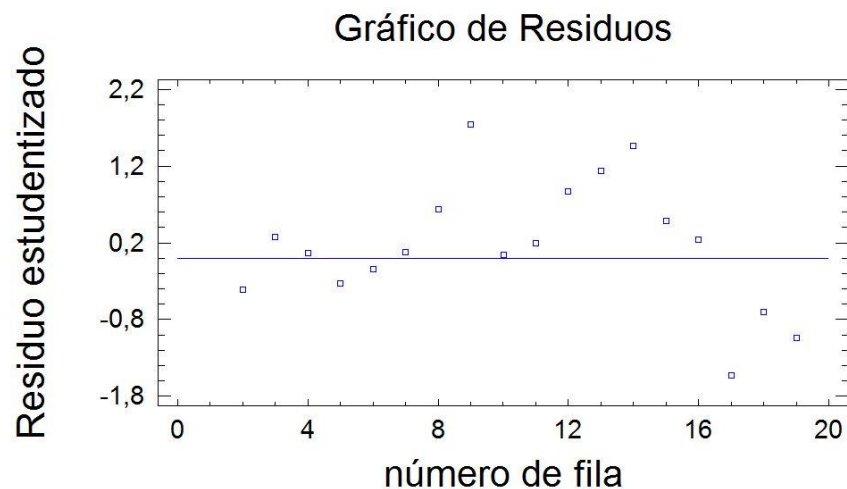


Ilustración 26. Estudio de Gráficos - Residuos frente Número de Fila

En la **Ilustración 26** se puede observar como los residuos se ordenan tomando forma sinusoidal, por lo que puede que exista el problema de autocorrelación de primer orden positiva.

3.3.5 Heterocedasticidad

Para hacer posible el estudio de la heterocedasticidad en el modelo, es necesario plantear otro modelo en el que los residuos se elevan al cuadrado. El nuevo modelo para hacer posible este análisis es:

$$\text{Residuos}^2 = \beta_0 + \beta_1 * \text{PIB per Cápita} + \beta_2 * \text{Población Activa} + U$$

- **Contraste de hipótesis**

$$\left. \begin{array}{l} H_0: \sigma^2 = 0 \rightarrow \text{no existe heterocedasticidad} \\ H_1: \sigma^2 \neq 0 \rightarrow \text{existe heterocedasticidad} \end{array} \right\} \alpha (0,05)$$

✓ **PIB per Cápita**

Como el P-Valor de la variable independiente PIB per Cápita es mayor a α ($0,4250 > 0,05$), se acepta la hipótesis nula por lo que la variable PIB per Cápita no produce problemas de heterocedasticidad. (**Ilustración 27**).

✓ **Población Activa**

En este caso el P-Valor de la variable independiente Población Activa es menor a α ($0,0020 < 0,05$), así que se rechaza la hipótesis nula por lo que la variable Población Activa sí produce problemas de heterocedasticidad. (**Ilustración 27**).

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	1,02919E10	1,52583E10	0,674515	0,5096
PIB per cápita _€	-570127,0	696424,0	-0,818649	0,4250
Población Activa	7455,03	2025,76	3,68011	0,0020

Ilustración 27. Contraste de Heterocedasticidad

3.3.6 Autocorrelación

- *Función de Autocorrelación Simple*

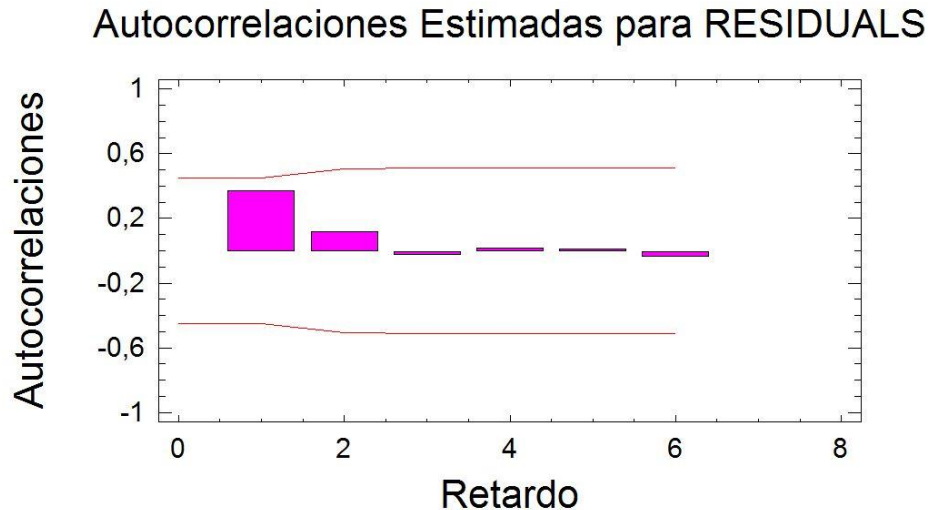


Ilustración 28. Análisis de la Autocorrelación - Gráfico FAS

Como se observa en la **Ilustración 28**, ningún retardo sobrepasa los límites de autocorrelación, aunque sí está muy próximo al límite el retardo 1. Aun así, el modelo no presenta problemas de autocorrelación de ningún orden.

- *Función de Autocorrelación Parcial*

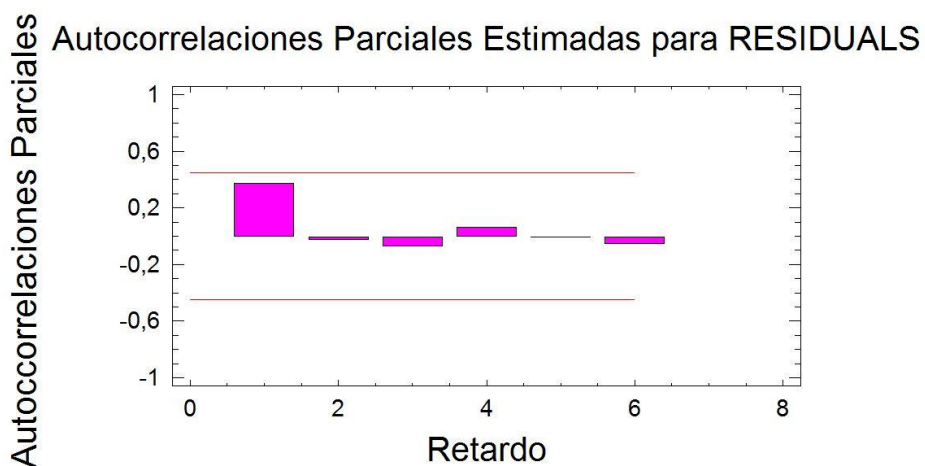


Ilustración 29. Análisis de la Autocorrelación - Gráfico FAP

En la **Ilustración 29**, se observa que ningún retardo sobrepasa los límites de autocorrelación, aunque sí está muy próximo al límite el retardo 1. Aun así, el modelo no presenta problemas de autocorrelación de ningún orden.

- **Contraste de Hipótesis**

$$\left. \begin{array}{l} H_0: \text{No existe autocorrelación de 1er orden} \\ H_1: \text{Existe autocorrelación de 1er orden} \end{array} \right\} \alpha (0,05)$$

En el contraste de la hipótesis, se analiza el P-Valor del Estadístico de Durbin-Watson. En este caso, al ser menor a α ($0,0009 < 0,05$), se rechaza la hipótesis nula por lo que se concluye que existe el problema de autocorrelación de 1er orden.

- **Estadístico Durbin- Watson**

Para sacar los valores de la tabla de Durbin-Watson, es necesario conocer el valor de k que son el número de variables (2), n que es el número de datos (19) y α (0,05).

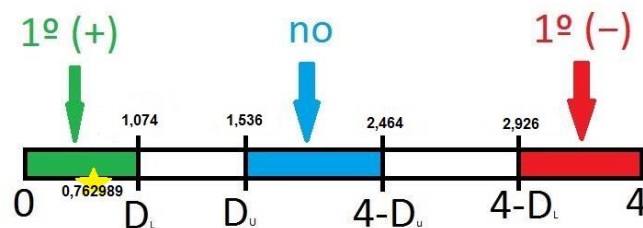


Ilustración 30. Análisis de la Autocorrelación - Estadístico Durbin-Watson

Como se observa en la **Ilustración 30**, el valor del Estadístico de Durbin-Watson (0,762989) está en el intervalo que asume el problema de autocorrelación de 1er orden positiva.

Como la prueba más fiable (el contraste de hipótesis) resulta ser positiva en el problema de autocorrelación, se asume que existe el problema de autocorrelación de 1er orden positiva.

3.3.7 Puntos Influyentes y Residuos Atípicos

- **Puntos Influyentes**

✓ **A Priori**

$$n_{ii} > 2 * \bar{n} \rightarrow \bar{n} = 0,157895$$

n_{ii} es la influencia y \bar{n} es la influencia media.

Tabla 12. Análisis de los Puntos Influyentes - A Priori

CCAA	n_{ii}	$2 * \bar{n}$	Punto Influyente A Priori
Andalucía	0,454474	0,31579	Sí
Cataluña	0,330781		Sí
Comunidad de Madrid	0,316131		Sí
Comunidad Foral de Navarra	0,222741		No

✓ **A posteriori**

$$|DFITS| > 2 * \sqrt{\bar{n}} \rightarrow \bar{n} = 0,157895$$

$DFITS$ es un estadístico que mide la variación de los coeficientes estimados si cada observación se sustrajera del conjunto de datos y, \bar{n} es la influencia media.

Tabla 13. Análisis de los Puntos Influyentes - A Posteriori

CCAA	$ DFITS $	$2 * \sqrt{\bar{n}}$	Punto Influyente A Posteriori
Andalucía	6,96002	0,794720	Sí
Cataluña	1,22458		Sí
Comunidad de Madrid	0,99393		Sí
Comunidad Foral de Navarra	0,816689		Sí

- **Residuos Atípicos**

$$|rt_j| \geq 2 \text{ o bien } |rt_j| \geq 3$$

rt_j es el valor del residuo estudentizado

Tabla 14. Análisis de los Residuos Atípicos

CCAA	$ rt_j $	Residuo Atípico
Andalucía	7,63	Sí y como es mayor a 3, candidato a eliminar del modelo

Resumen

Como se ha podido observar anteriormente, el modelo planteado tenía una variable no significativa (IPC) por lo que ha sido eliminada del modelo. Siguiendo con los análisis, se observa que los residuos se distribuyen de forma normal. Además, existen problemas de heterocedasticidad en la variable Población Activa y problemas de autocorrelación de 1er orden positiva. Finalizando el análisis, se obtiene un residuo atípico candidato a eliminar que pertenece a la CCAA de Andalucía.

Con esta información se pasa a resolver los problemas planteados en el modelo. Primeramente se resolverá la heterocedasticidad y se volverá a analizar la autocorrelación. Si ésta aún fuera existente, se pasará a resolverla también.

3.3.8 Resolución de los Problemas

Primeramente se va a solucionar el problema de Heterocedasticidad de la variable Población Activa.

Para ello, se plantea un nuevo modelo solución:

$$\frac{\text{Población Ocupada}}{\text{Población Activa}^c} = \frac{\beta_0}{\text{Población Activa}^c} + \frac{\beta_1 * \text{PIB per Cápita}}{\text{Población Activa}^c} + \frac{\beta_2 * \text{Población Activa}}{\text{Población Activa}^c} + U$$

$$\text{Dónde } c = \frac{h}{2}$$

Para saber que h es la más adecuada para el modelo, se deberá aproximar ésta al modelo siguiente:

$$\text{Residuos}^2 = \beta_0 + \beta_2 * \text{Población Activa}^h$$

Se aceptará la h que mayor R^2 proporcione.

Tabla 15. Cálculo de h

<i>h</i>	R^2	<i>h</i>	R^2	<i>h</i>	R^2
1	43,69%	13	83,28%	24	0%
2	55,22%	15	84,85%	1/2	32,89%
3	61,90%	18	86,66%	1/3	28,12%
4	66,65%	20	87,58%	-1	3,82%
10	80,16%	23	88,61%	-2	2,61%

Como 23 es el valor de la h que mayor R^2 proporciona, la c que se aplicará al nuevo modelo solución es de 11,5.

$$\frac{\text{Población Ocupada}}{\text{Población Activa}^{11,5}} = \frac{\beta_0}{\text{Población Activa}^{11,5}} + \frac{\beta_1 * \text{PIB per Cápita}}{\text{Población Activa}^{11,5}} + \frac{\beta_2 * \text{Población Activa}}{\text{Población Activa}^{11,5}}$$

Como es posible ver en la **Ilustración 31**, no es posible la solución del problema de heterocedasticidad con este valor de c ya que se presenta un problema de multicolinealidad muy elevado.

Resumen del Procedimiento

Variable dependiente: Población Ocupada $_n\text{ohabi_}/(\text{Población Activa } _n\text{ohabi_}^{\wedge}11,5)$

Variabes independientes:

$1/(\text{Población Activa } _n\text{ohabi_}^{\wedge}11,5)$

PIB per cápita $_€ /(\text{Población Activa } _n\text{ohabi_}^{\wedge}11,5)$

Población Activa $_n\text{ohabi_}/(\text{Población Activa } _n\text{ohabi_}^{\wedge}11,5)$

ERROR: Población Activa $_n\text{ohabi_}/(\text{Población Activa } _n\text{ohabi_}^{\wedge}11,5)$ es una combinación lineal de las otras variables.

Ilustración 31. Solución Heterocedasticidad

Al presentarse este inconveniente en la resolución del problema, se prueban los demás valores de h siendo el único que nos proporciona una solución al problema el valor de h=2, cuya c es igual a 1.

Planteando el nuevo modelo solución, se observa en la **Ilustración 32** como el R^2 pasa de ser casi del 99% a sólo un 42,68%, por lo que descartamos este modelo como el modelo solución.

$$\frac{\text{Población Ocupada}}{\text{Población Activa}} = \frac{\beta_0}{\text{Población Activa}} + \frac{\beta_1 * \text{PIB per Cápita}}{\text{Población Activa}} + \frac{\beta_2 * \text{Población Activa}}{\text{Población Activa}}$$

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	0,63702	0,0150315	42,3791	0,0000
$1/(\text{Población Acti}$	-35050,6	16065,5	-2,18173	0,0444
PIB per cápita $_€$	1,69041	0,918085	1,84124	0,0842

Análisis de Varianza

Fuente	Suma de cuadrados	GL	Cuadrado medio	Cociente-F	P-Valor
Modelo	0,0323473	2	0,0161737	5,96	0,0116
Residuo	0,0434361	16	0,00271476		
Total (Corr.)	0,0757834	18			

R-cuadrado = 42,6839 porcentaje
R-cuadrado (ajustado para g.l.) = 35,5194 porcentaje
Error estándar de est. = 0,0521033
Error absoluto medio = 0,0421207
Estadístico de Durbin-Watson = 2,08997 (P=0,4156)
Autocorrelación residual en Lag 1 = -0,182707

Ilustración 32. Solución Heterocedasticidad 2

Como al plantear esta solución del problema de Heterocedasticidad estamos asumiendo un residuo atípico candidato a eliminar, se procede a la eliminación de este dato.

Cuando se elimina el dato, es posible observar como el R^2 del modelo aumenta (99,73% frente al 98,93% del modelo con el residuo atípico y con problemas de heterocedasticidad y autocorrelación), como la significatividad de los parámetros y del modelo es positiva, como los residuos de este modelo se comportan de forma normal y, como a simple vista no existen problemas de heterocedasticidad ni autocorrelación.

$$\text{Población Ocupada} = -193.610 + 9,2957 * \text{PIB per Cápita} + 0,635565 * \text{Población Activa}$$

Variable dependiente: Población Ocupada _nºhabi_

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	-193610,0	59684,3	-3,2439	0,0055
PIB per cápita _€	9,2957	2,81973	3,29666	0,0049
Población Activa	0,635565	0,00939693	67,6354	0,0000

Análisis de Varianza

Fuente	Suma de cuadrados	GL	Cuadrado medio	Cociente-F	P-Valor
Modelo	1,28166E13	2	6,40831E12	2764,80	0,0000
Residuo	3,47673E10	15	2,31782E9		
Total (Corr.)	1,28514E13	17			

R-cuadrado = 99,7295 porcentaje
R-cuadrado (ajustado para g.l.) = 99,6934 porcentaje
Error estándar de est. = 48143,8
Error absoluto medio = 36119,1
Estadístico de Durbin-Watson = 1,98318 (P=0,4683)
Autocorrelación residual en Lag 1 = -0,0119434

Ilustración 33. Modelo sin el residuo atípico

```
Tests para la Normalidad para RESIDUALS
Estadístico chi-cuadrado de bondad de ajuste = 7,33333
P-valor = 0,602458

Estadístico W de Shapiro-Wilks = 0,952427
P-valor = 0,463707

Puntuación Z para asimetría = 0,8282
P-valor = 0,407555
```

Ilustración 34. Normalidad Modelo sin el residuo atípico

Variable dependiente: RESIDUALS^2

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	5,36612E9	2,91622E9	1,84009	0,0856
PIB per cápita _€	-199366,0	137774,0	-1,44705	0,1685
Población Activa	741,08	459,141	1,61405	0,1274

Ilustración 35. Heterocedasticidad Modelo sin residuo atípico

A continuación se analizará si el modelo propuesto a solución tiene dichos problemas:

- **Heterocedasticidad**

$$\text{Residuos}^2 = \beta_0 + \beta_1 * \text{PIB per Cápita} + \beta_2 * \text{Población Activa} + U$$

✓ **Contraste de hipótesis**

$$\left. \begin{array}{l} H_0: \sigma^2 = 0 \rightarrow \text{no existe heterocedasticidad} \\ H_1: \sigma^2 \neq 0 \rightarrow \text{existe heterocedasticidad} \end{array} \right\} \alpha (0,05)$$

▪ **PIB per Cápita**

Como el P-Valor de la variable independiente PIB per Cápita es mayor a α ($0,1685 > 0,05$), se acepta la hipótesis nula por lo que la variable PIB per Cápita no produce problemas de heterocedasticidad.

▪ **Población Activa**

Como el P-Valor de la variable independiente Población Activa es mayor a α ($0,1274 < 0,05$), se acepta la hipótesis nula por lo que la variable Población Activa no produce problemas de heterocedasticidad.

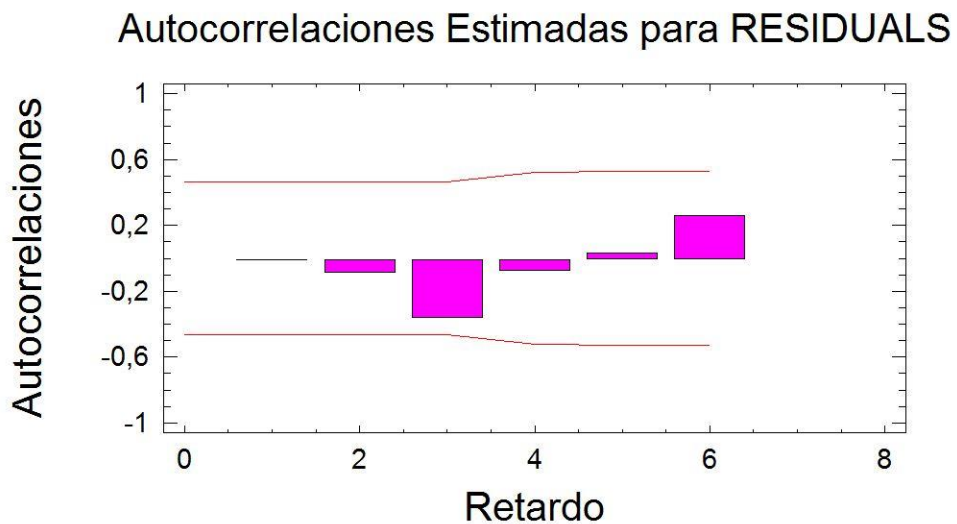
Variable dependiente: RESIDUALS^2

Parámetro	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	5,36612E9	2,91622E9	1,84009	0,0856
PIB per cápita _€	-199366,0	137774,0	-1,44705	0,1685
Población Activa	741,08	459,141	1,61405	0,1274

Ilustración 36. Análisis Heterocedasticidad - Modelo sin residuo atípico

- **Autocorrelación**

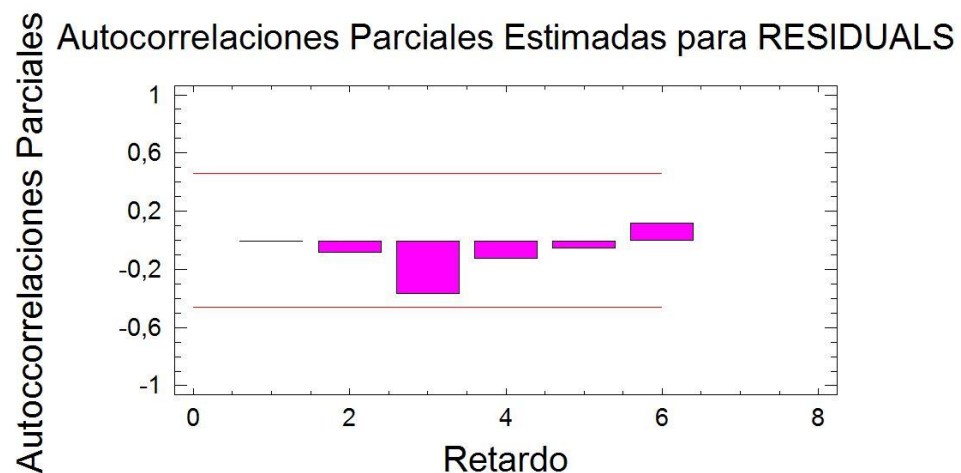
✓ **Función de Autocorrelación Simple**



Ilustraci6n 37. An6lisis de la Autocorrelaci6n Modelo sin residuo at6pico - Gr6fico FAS

Como se observa en la **ilustraci6n 37**, ning6n retardo sobrepasa los l6mites de autocorrelaci6n.

✓ **Funci6n de Autocorrelaci6n Parcial**



Ilustraci6n 38. An6lisis de la Autocorrelaci6n Modelo sin residuo at6pico - Gr6fico FAP

Como se observa en la **ilustraci6n 38**, ning6n retardo sobrepasa los l6mites de autocorrelaci6n.

✓ **Contraste de Hipótesis**

$$\left. \begin{array}{l} H_0: \text{No existe autocorrelación de 1er orden} \\ H_1: \text{Existe autocorrelación de 1er orden} \end{array} \right\} \alpha (0,05)$$

En el contraste de la hipótesis, se analiza el P-Valor del Estadístico de Durbin-Watson. En este caso, al ser menor a α ($0,4683 < 0,05$), se acepta la hipótesis nula por lo que se concluye que no existe el problema de autocorrelación de 1er orden.

✓ **Estadístico Durbin- Watson**

Para sacar los valores de la tabla de Durbin-Watson, es necesario conocer el valor de k que son el número de variables (2), n que es el número de datos (18) y α (0,05).

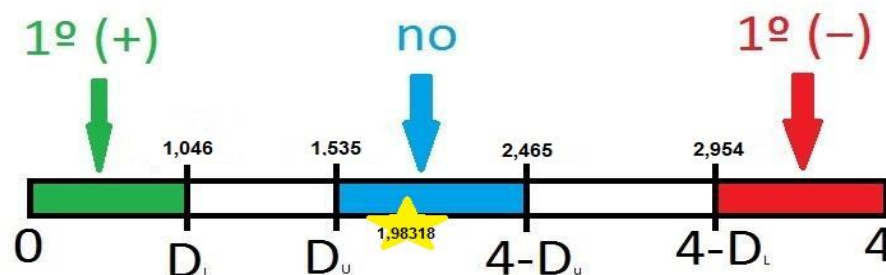


Ilustración 39. Análisis de la Autocorrelación Modelo sin residuo atípico- Estadístico Durbin-Watson

Como se observa en la **Ilustración 39**, el valor del Estadístico de Durbin-Watson (1,98318) está en el intervalo en el que no existe autocorrelación, por lo que no se asume el problema de autocorrelación de 1er orden positiva.

Como la prueba más fiable (el contraste de hipótesis) resulta ser negativa en ambos problemas (heterocedasticidad y autocorrelación), se concluye que el modelo propuesto está libre de cualquier problema.



Resumen

Como se ha podido observar en este apartado, el problema de heterocedasticidad del modelo nº2 es solucionable en el valor de $h=2$, es decir, en el valor de $c=1$, donde el R^2 del modelo es igual a 42,68%.

Por lo tanto, con este nuevo R^2 se asume un descenso del mismo en un 50% por lo que se cree conveniente desechar el modelo.

Para solucionar el nuevo problema que aparece en la resolución, se plantea un nuevo modelo sin el residuo atípico candidato a eliminar que se observa en el modelo nº2 (CC.AA. de Andalucía). Puesto que el R^2 de este nuevo modelo es de 99,73% se decide aceptarlo como el definitivo para la consecución de los objetivos de este proyecto. Además en este modelo final tanto el modelo en sí como los parámetros son significativos, los residuos se comportan de forma normal y no existen problemas ni de heterocedasticidad ni de autocorrelación.

El Modelo Final ya estimado es el siguiente:

$$Población\ Ocupada = -193.610 + 9,2957 * PIB\ per\ Cápita + 0,635565 * Población\ Activa$$

3.3.9 R^2 e Interpretación del Modelo

Como es posible observar en la **Ilustración 33**, el R^2 es igual a 99,73%. Este resultado tan elevado del R^2 es muy favorable ya que éste tiene que estar comprendido entre 0 y 1, siendo 1 el ajuste perfecto. En este caso, el ajuste es casi perfecto, es decir, con el modelo se explica casi la totalidad de los datos.

A la hora de explicar los parámetros del modelo vemos como tanto β_1 como β_2 son positivas, eso quiere decir que a medida que se aumente la Población Ocupada, el PIB per Cápita y la Población Activa también lo harán en la medida que determina el valor de su respectivo β .

3.3.10 Predicción

Para realizar una predicción en el modelo, es necesario conocer que CC.AA. están cercanas a la media ya que el programa Statgraphics se basa en la media del modelo para realizar dicha predicción.

Para ello, hay que conocer primeramente cuáles son los residuos de las CC.AA. y elegir en su caso, los que menor valor tengan, tanto por encima de la media como por debajo de la media.

Tabla 16. Residuos Modelo Final

CCAA	Residuo	CCAA	Residuo
Aragón	-9.250,54	Comunitat Valenciana	-94.369
Principado de Asturias	31.954,5	Extremadura	31.767,6
Illes Balears	28.833,5	Galicia	74.112
Canarias	-77.577,9	Comunidad de Madrid	43.333,4
Cantabria	7.546,24	Melilla	35.367,6
Castilla La Mancha	-43.774,4	Región de Murcia	-11.194,7
Castilla y León	44.576	Comunidad Foral de Navarra	-49.717,4
Cataluña	14.247,4	País Vasco	-2.724,74
Ceuta	13.333,3	La Rioja	-36.463

Una vez elegidas las dos CC.AA de las que se van a hacer las predicciones en la Población Ocupada, se deciden ahora los nuevos valores de las variables del siguiente modelo:

$$\text{Población Ocupada} = -193.610 + 9,2957 * \text{PIB per Cápita} + 0,635565 * \text{Población Activa}$$

La predicción en el caso en el que la variable PIB per Cápita aumentara su valor en 5.000 € sería:

- **País Vasco:** Población Activa = 1.230.629 habitantes; PIB per Cápita = 34.959€
 - ✓ **Predicción Puntual**
 - 913.503 habitantes.
 - ✓ **Predicción por Intervalos**
 - [782.191 habitantes ≤ Población Ocupada ≤ 1.044.820 habitantes]

- **Cantabria:** Población Activa = 339.448 habitantes; PIB per Cápita = 26.550€
 - ✓ **Predicción Puntual**
 - 268.932 habitantes
 - ✓ **Predicción por Intervalos**
 - [156.646 habitantes ≤ Población Ocupada ≤ 339.448 habitantes]

Como es posible observar, el aumento de la variable PIB per Cápita supone que los valores puntuales de la Población Ocupada aumenten. En el caso del País Vasco en 49.203 habitantes y, en el caso de Cantabria en 38.932 habitantes.

En el caso de que fuera la variable Población Activa la que aumentara en 5.000 habitantes, la predicción sería:

- **País Vasco:** Población Activa = 1.235.629 habitantes; PIB per Cápita = 29.959€
 - ✓ **Predicción Puntual**
 - 870.203 habitantes.
 - ✓ **Predicción por Intervalos**
 - [754.281 habitantes ≤ Población Ocupada ≤ 986.124 habitantes]

- **Cantabria:** Población Activa = 344.448 habitantes; PIB per Cápita = 21.550€
 - ✓ **Predicción Puntual**
 - 225.632 habitantes
 - ✓ **Predicción por Intervalos**
 - [118.680 habitantes \leq Población Ocupada \leq 332.583 habitantes]

Como es posible observar, el aumento de la variable Población Activa supone que el valor puntual de la Población Ocupada en el País Vasco aumente en 5.903 habitantes. Mientras, en el caso de Cantabria el valor puntual desciende en 4.368 habitantes ya que el valor predicho se encuentra más próximo a la media que el valor que inicialmente se obtiene en la **Tabla 23**.

En el caso en el que fueran ambas variables las que aumentarían en 5.000 unidades, las predicciones serían las siguientes:

- **País Vasco:** Población Activa = 1.235.629 habitantes; PIB per Cápita = 34.959€
 - ✓ **Predicción Puntual**
 - 916.681 habitantes.
 - ✓ **Predicción por Intervalos**
 - [785.392 habitantes \leq Población Ocupada \leq 1.047.970 habitantes]
- **Cantabria:** Población Activa = 344.448 habitantes; PIB per Cápita = 26.550€
 - ✓ **Predicción Puntual**
 - 272.110 habitantes
 - ✓ **Predicción por Intervalos**
 - [159.050 habitantes \leq Población Ocupada \leq 344.448 habitantes]

Como es posible observar, el aumento de ambas variables supone que los valores puntuales de la Población Ocupada aumenten. En el caso del País Vasco en 52.381 habitantes y, en el caso de Cantabria en 42.110 habitantes



Capítulo 4: Discusión

4 Discusión

La discusión que se plantea antes de realizar este proyecto es la siguiente: ¿se podría realizar un proyecto de características similares teniendo en cuenta la variación entre dos años de la Población Ocupada o con este proyecto se explicaría lo mismo que realizando el proyecto anteriormente nombrado?

Pues bien, con la realización de una predicción sobre la Población Ocupada y los parámetros de cada variable en unas Comunidades Autónomas elegidas tanto de un modelo como del otro, obtendremos la respuesta.

Primeramente, se realizará una primera predicción del modelo econométrico, al que llamaremos Modelo 2014, manteniendo el mismo valor de la Población Parada que en la **Tabla 23** mientras que en la segunda predicción se aumentará la Población Parada en mil unidades. Las Comunidades Autónomas elegidas son Galicia, Andalucía, Aragón, Cantabria y Castilla y León.

- **Modelo 2014:**

Tabla 17. Predicción Modelo 2014 - Discusión

PREDICCIÓN MODELO 2014				
		Predicción Puntual	Límite Inferior	Límite Superior
Galicia	Igual Pob. Parada	996.755	953.504	1.040.010
	+1000 Pob. Parada	996.317	953.025	1.039.610
Andalucía	Igual Pob. Parada	2.607.250	2.559.550	2.654.950
	+1000 Pob. Parada	2.606.810	2.559.100	2.654.530
Aragón	Igual Pob. Parada	533.836	497.115	570.556
	+1000 Pob. Parada	533.398	496.698	570.098
Cantabria	Igual Pob. Parada	233.581	194.486	272.676
	+1000 Pob. Parada	233.143	194.027	272.259
Castilla y León	Igual Pob. Parada	929.179	888.787	969.570
	+1000 Pob. Parada	928.741	888.408	969.073

- **Modelo Variación 2011-2014:**

Tabla 18. Predicción Modelo Variación 2011-2014 - Discusión

PREDICCIÓN MODELO VARIACIÓN 2011-2014				
		Predicción Puntual	Límite Inferior	Límite Superior
Galicia	Igual Pob. Parada	-76.861,5	-110.552	-43.170,7
	+1000 Pob. Parada	-77.335,8	-111.033	-43.638,6
Andalucía	Igual Pob. Parada	-164.203	-204.764	-123.641
	+1000 Pob. Parada	-164.677	-205.286	-124.068
Aragón	Igual Pob. Parada	-16.766,9	-47.365,4	13.831,16
	+1000 Pob. Parada	-17.241,2	-47.829,1	13.346,7
Cantabria	Igual Pob. Parada	-27.858,9	-62.312,6	6.594,73
	+1000 Pob. Parada	-28.333,2	-62.790	6.123,54
Castilla y León	Igual Pob. Parada	-71.705,8	-108.249	-35.162,5
	+1000 Pob. Parada	-72.180,1	-108.740	-35.620

Si se considera que la Población Ocupada en 2011 tiene un valor exacto, para estimar la Población Ocupada en 2014 es necesario calcularla mediante los valores estimados en el programa más los valores que tenemos en la Población Ocupada de 2011. Como se puede observar en la tabla, la estimación para el Modelo Variación 2011-2014 está dentro de los límites de confianza de la estimación del Modelo 2014.

- **Modelo Variación 2011-2014:**

Tabla 19. Predicción Modelo Variación 2011-2014 (2) - Discusión

PREDICCIÓN MODELO VARIACIÓN 2011-2014				
	Población Ocupada 2011		Estimación	Diferencia Modelos
Galicia	1.093.900	Igual Pob. Parada	1.017.038,5	20.283,5
		+1000 Pob. Parada	1.016.564,2	20.247,2
Andalucía	2.773.700	Igual Pob. Parada	2.609.497	2.247
		+1000 Pob. Parada	2.609.023	2.213
Aragón	558.400	Igual Pob. Parada	541.633,1	7.797,1
		+1000 Pob. Parada	541.158,8	7.760,8
Cantabria	246.900	Igual Pob. Parada	219.041,1	-14.539,9
		+1000 Pob. Parada	218.566,8	-14.576,2
Castilla y León	1.012.800	Igual Pob. Parada	941.094,2	11.915,2
		+1000 Pob. Parada	940.619,9	11.878,9

Al mismo modo que se ha analizado la estimación de la Población Ocupada, se analiza la estimación de los parámetros de cada una de las variables.

- **Modelo 2014:**

Tabla 20. Predicción parámetros Modelo 2014 - Discusión

PREDICCIÓN MODELO 2014			
	Predicción Puntual	Límite Inferior	Límite Superior
Constante	-151.800	-1.574.390	1.270.790
IPC	1.467,8	-12.444,1	15.379,7
Nº Empresas	0,44	-1,01	1,90
PIB	0,43	-2,22	3,09
Población Total	0,15	0,05	0,25
Población Activa	0,48	0,32	0,64
Población Parada	-0,44	-0,59	-0,27

- **Modelo Variación 2011-2014:**

Tabla 21. Predicción parámetros Modelo Variación 2011-2014 - Discusión

PREDICCIÓN MODELO VARIACIÓN 2011-2014			
	Predicción Puntual	Límite Inferior	Límite Superior
Constante	56.333,3	-12.325,4	124.992
IPC	-13.969,4	-29.628,8	1.690
Nº Empresas	-0,005	-0,02	0,01
PIB	27,06	-14,51	68,65
Población Total	-0,23	-0,78	0,31
Población Activa	1,04	0,82	1,27
Población Parada	-0,47	-0,56	-0,38



Capítulo 5: Análisis DAFO

5 Análisis DAFO de la CC.AA. elegida

El análisis externo analiza la parte del entorno que influye de alguna forma en la empresa, y que puede llegar a ser relevante a la hora de definir la estrategia empresarial.

Éste está compuesto por un análisis del macroentorno o análisis del entorno general que analiza todo lo que rodea a la empresa que afecta indirectamente a la misma y que se deriva del sistema socioeconómico en el que se desarrolla la actividad. Y el análisis del microentorno o análisis del entorno específico que analiza todo aquellos factores más próximos a la actividad desarrollada de la empresa.

En este caso, al analizar una CC.AA. no puede realizarse un análisis del microentorno, por lo que sólo se analizará el macroentorno de dicha CC.AA.

Para analizar el macroentorno, se utilizará la técnica del análisis PEST en el que se estudiarán los factores político-legales, económicos, socioculturales y tecnológicos que afectan a la CC.AA. En este caso, la CC.AA. elegida es Melilla ya que representa el menor porcentaje de Población Ocupada respecto a su Población Activa.

- **Análisis del Macroentorno**

✓ **Factores político-legales**

- El sistema político en España actualmente es una monarquía parlamentaria. El cambio del jefe de Estado es mediante sucesión directa por la Constitución. El Parlamento es escogido mediante elecciones legislativas y, sólo el jefe de Estado tiene la potestad de disolver las Cortes si éstas no tienen un criterio uniforme de Gobierno.
- La repartición del poder viene determinada por la convocatoria de elecciones cada cuatro años en las que se vota los tres niveles de poder: local, autonómico y estatal, ya que España tiene instaurada la soberanía popular como un pilar de su Constitución.
- El partido político que actualmente está en el poder es el Partido Popular, partido de centro-derecha tanto en España como en Melilla. Aunque haya varios partidos políticos, se podría considerar España como un sistema bipartidista ya que hay dos grandes partidos políticos que predominan.

- España al formar parte de la Unión Europea, de las muchas decisiones políticas que se toman algunas tienen que ser aprobadas y/o en su caso corregidas y/o sancionadas, por lo que puede llegar a ser esto una amenaza a nuestro país o una oportunidad, ya que por ejemplo la concesión de un préstamo es beneficiosa pero a la vez llega a ser una amenaza si los términos de devolución son abusivos.
 - Aparición de la Reforma Laboral que tratará de abaratar el despido improcedente y por causas económicas y, ayudar en las contrataciones de menores de 30 años y parados de larga duración, además de otras medidas.
 - Congelación del Salario Mínimo Interprofesional (SMI) en 654,3€ mensuales, lo que diariamente serían 21,51€ ó 9.034,2€ anuales considerando 14 pagas. En este caso, también existe un doble sentido ya que puede ser una oportunidad para negocios emergentes o en crisis que contratan a gente con dicho salario, pero al mismo tiempo es una amenaza para la sociedad ya que el nivel de poder adquisitivo disminuye.
 - Reforma en las edades de jubilación que afecta a la mayoría de las personas paradas con más de 50 años ya que se aumenta de 65 años la edad mínima a 67 años o, teniendo 65 años se debe acreditar un mínimo de 38 años y 6 meses cotizados. Además, con el paso de los años, se aplicará esta reforma de forma gradual.
- ✓ **Factores económicos**
- Posible crecimiento de la economía en una tasa intertrimestral del 0,6%. Este crecimiento situaría la tasa del PIB en un 1,9% en el último trimestre del año 2014. Cerrando así el año, el PIB español aumentaría en su conjunto un 1,4% el 2014.
 - Déficit público compuesto por el déficit del Estado, el déficit de las CC.AA y el déficit de los ayuntamientos. El déficit de la Administración Central es de 28.470 millones de euros, el 2,68% del PIB español; mientras que el déficit autonómico es de 14.167 millones de euros, el 1,33% del PIB español. Esto produce que haya una subida de impuestos a los que los ciudadanos tienen que responder bajando así su nivel de riqueza.
 - Incremento del Impuesto sobre el Valor Añadido que hace que la población española tenga que gastar más en sus compras.



- Con la congelación del SMI en tan sólo 654,3€ mensuales, muchas empresas cogen este mínimo a la hora de contratar a las personas, por lo que muchas de éstas han visto reducido su poder de adquisición.
- Aumento de la demanda interna privada española.
- La Tasa de desempleo en España se sitúa en un 23,67% en el tercer trimestre de 2014. El desempleo juvenil ha aumentado. La situación es dramática ya que el 22,5% de los jóvenes entre 15 y 29 años ni estudia ni trabaja.
- Debilitamiento de los mercados de exportación en el año 2015 que serán compensados por la posible depreciación del tipo de cambio del euro, por las mejoras en las condiciones de financiación y la reducción de los precios del petróleo. Esta última cuestión ha ayudado al descenso de la tasa de inflación española.
- Las empresas en Melilla representan el 0,1% del total españolas. Muchas de estas (95%) son microempresas (muchas son personas físicas). Hay más de 3.400 empresas, de las cuales más del 50% son del sector inmobiliario y hostelero lo cual plantea una fortaleza al poseer muchas opciones de turismo en la ciudad, además de una debilidad ya que el sector de la construcción está muy castigado en España.
- El grado de ocupación hotelera se situó en 2013 en el 41,2%.
- El paso fronterizo de Beni Enzar tiene como principal tránsito el de vehículos de contrabando, por lo que las compras en Melilla aumentan ya que muchos marroquíes compran los productos a menor costo para venderlos en su país, y al revés.
- Incremento del PMI (Purchasing Managers' Index) de la Industria Manufacturera mientras que el IPI (Índice de Producción Industrial) desciende.
- Línea ICO – Plan Avanza 2 (financiación por parte del Estado español para autónomos, pymes, asociaciones y agrupaciones que promuevan actividades económicas que ayuden al crecimiento y desarrollo del país además de mejorar la distribución de la riqueza nacional).

✓ **Factores socioculturales**

- Entre enero y febrero de 2014, 600 inmigrantes ilegales consiguieron saltar la valla de Melilla, ya que esta ciudad es la puesta de todas las miradas de la población africana a la hora de llegar a territorio europeo. En la misma frontera de Beni Enzar, se ofrece una oficina para aquellos inmigrantes que quieran solicitar asilo en España.
- La tasa de abandono escolar temprano es del 23,5% la cual duplica casi la media europea que se establece en un 12%.
- La sociedad española tiene un alto porcentaje de niños que están en situación de riesgo o exclusión social, ya que este se sitúa en un 32,7% en menores de 6 a 11 años y en un 37,6% en menores de 12 a 17 años. La tasa total de población en situación de riesgo o exclusión social es de 27,3%. El umbral de pobreza de los hogares españoles con 1 persona se sitúa en 8.114€ en 2013. La tasa de riesgo de pobreza en Melilla asciende al 21,7%
- En el año 2013, un 45,8% de la población española no puede permitirse ir de vacaciones fuera de casa al menos una semana al año; un 41% no tiene capacidad para afrontar gastos imprevistos; un 9,3% tiene retrasos en los pagos relacionados con la vivienda principal; y, un 16,9% tiene mucha dificultad para llegar a fin de mes. En la CC.AA. de Melilla, en el año 2013, un 35% de la población no puede permitirse ir de vacaciones fuera de casa al menos una semana al año; un 47,3% no tiene capacidad para afrontar gastos imprevistos; un 13,2% tiene retrasos en los pagos relacionados con la vivienda principal; y, un 15,3% tiene mucha dificultad para llegar a fin de mes.
- Inconformidad y malestar en la sociedad española con el Gobierno por la rebaja salarial, el aumento en los años de la jubilación o la subida de los impuestos cómo el IVA.
- Aumento de movimientos ciudadanos en contra de las políticas establecidas por el Gobierno central, que provocan un desequilibrio social importante, como por ejemplo: STOP Desahucios (Plataforma de Afectados por la Hipoteca), 15M, Podemos, etc.



- Implantación del III Plan Integral de Juventud (Melilla) que tratará de fomentar la formación, ayudar a la formación de nuevos jóvenes emprendedores y a impulsar la contratación juvenil.
- El número de jóvenes de entre 14 a 35 años alcanza la cifra de 27.743 habitantes, que representa un 32,89% de la población total melillense, un 5% más que la totalidad de los jóvenes españoles en su conjunto.

✓ **Factores tecnológicos**

- Eficiencia de las infraestructuras como puertos, aeropuertos y carreteras.
- Creación del corredor del Mediterráneo.
- Leyes en Melilla sobre la Propiedad Industrial en diseños industriales, marcas y nombres comerciales, patentes y modelos de utilidad y, topografías de semiconductores.
- Plan Especial de Ordenación del Puerto de Melilla (B.O.ME. nº 5.151 de 29/07/2014) por el que se hará una ampliación y mejora del puerto de Melilla.
- Proyecto PASI (Proyecto Integrado de desarrollo cooperativo en el ámbito de la Sociedad de la Información) que quiere dar soluciones en las TIC en la zona fronteriza de la Ciudad de Melilla y el norte de Marruecos.
- En marzo de 2014 se hizo una remodelación para la agilización del tránsito en el paso fronterizo de Beni Enzar.

Tabla 22. Análisis DAFO - Ciudad Autónoma de Melilla

AMENAZAS	OPORTUNIDADES
<p>País miembro de la Unión Europea.</p> <p>Nivel muy alto de déficit público.</p> <p>Subidas de los impuestos.</p> <p>Menor nivel de riqueza poblacional.</p> <p>Reforma en las edades de jubilación.</p> <p>Congelación del Salario Mínimo Interprofesional.</p> <p>Alta tasa de desempleo.</p> <p>Debilitamiento de los mercados de exportación.</p> <p>Debilitamiento de la tasa de inflación española.</p> <p>Descenso de IPI.</p> <p>Tasa de abandono escolar muy alta.</p> <p>Tasa muy elevada de población en situación de riesgo o exclusión social.</p> <p>Malestar e inconformidad social.</p> <p>Aumento de movimientos ciudadanos en contra del Gobierno Central.</p> <p>Mucho tránsito de vehículos de contrabando entre Melilla y Marruecos.</p>	<p>Soberanía popular y libertad de expresión.</p> <p>País miembro de la Unión Europea.</p> <p>Aparición de la nueva reforma laboral.</p> <p>Congelación del Salario Mínimo Interprofesional.</p> <p>Posible crecimiento de la economía española.</p> <p>Aumento de la demanda interna privada española.</p> <p>Aumento del PMI de la Industria Manufacturera.</p> <p>Aprobación de las Líneas ICO – Plan Avanza 2.</p> <p>Eficiencia en infraestructuras.</p> <p>Creación del Corredor del Mediterráneo.</p>
FORTALEZAS	DEBILIDADES
<p>Alto número de empresas dedicadas al sector hostelero e inmobiliario.</p> <p>Grado de ocupación hotelera medio.</p> <p>Mucho tránsito de vehículos de contrabando entre Melilla y Marruecos.</p> <p>Eficiencia en infraestructuras.</p> <p>Paso fronterizo de Beni Enzar.</p> <p>Implantación del III Plan Integral de Juventud.</p> <p>Plan Especial de Ordenación del Puerto de Melilla.</p> <p>Creación de Proyecto PASI.</p> <p>Alto número de personas en edad de trabajar menores de 35 años.</p> <p>Leyes sobre la Propiedad Industrial.</p>	<p>Tasa elevada de población en situación de riesgo o exclusión social.</p> <p>Alto número de empresas dedicadas al sector hostelero e inmobiliario.</p> <p>Mucho tránsito de vehículos de contrabando entre Melilla y Marruecos.</p> <p>Entrada de inmigrantes ilegales a Melilla a los que hay que dar asilo en el CETI.</p> <p>Tasa de pobreza muy alta en Melilla.</p>



Capítulo 6: Conclusiones

6 Conclusiones

Una vez finalizado el estudio, se obtiene del primer modelo propuesto que:

- Las variables que se plantean al principio del estudio están correlacionadas entre ellas, por lo que una vez solucionado el problema sólo 3 de las 6 variables propuestas son aptas (PIB per Cápita, Población Activa, IPC).
- En la significatividad de las variables resultan ser significativas por orden de entrada al modelo Población Activa y PIB per Cápita, siendo la variable IPC excluida del modelo.
- Después de realizar todos los análisis de las hipótesis, se procede a eliminar un residuo atípico que corresponde a la CC.AA. de Andalucía.

En el nuevo modelo sin Andalucía, cuyo R^2 es igual al 99,73%, se observa que:

- La significatividad de los parámetros es positiva.
- Se cumplen todas las hipótesis.

Con este modelo final, se predice que:

- Si se aumenta el PIB per Cápita en 5.000€, el valor de la Población Ocupada en el País Vasco y en Cantabria aumentará.
- Si se aumenta la Población Activa en 5.000 unidades, el valor de la Población Ocupada en el País Vasco aumentará, mientras que en Cantabria descenderá.
- Si se aumenta ambas variables en 5.000€ y 5.000 unidades, tanto el valor de la Población Ocupada en el País Vasco como el valor en Cantabria aumentará.

Cuando se finaliza la totalidad de los cálculos, se plantea una discusión en los datos, ya que podría ser posible explicar las mismas conclusiones mediante un estudio de la variación entre dos años de la Población Ocupada. Después de una serie de predicciones se observa que sí, que con ambos modelos se obtendrían los mismos resultados, por lo que posiblemente se plantee este problema como una futura línea sobre la investigación.

Lo más destacable del análisis DAFO es la tasa tan elevada de población en situación de riesgo o exclusión social que existe en Melilla, las masivas entradas de inmigrantes ilegales a los que hay que dar asilo, el elevado tránsito de vehículos de contrabando entre Melilla y Marruecos, además de un alto número de empresas que se dedican al sector hostelero e inmobiliario.



Capítulo 7: Futuras líneas de investigación



7 Futuras líneas de investigación

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

Estas nuevas formas de enfocar el estudio o vías de trabajo son las siguientes:

- Realización de este mismo proyecto mediante todas las estructuras (semilogarítmica, logarítmica, inversa y logarítmica inversa), ya que éste sólo ha sido realizado mediante la estructura lineal.
- Actualizar la base de datos de la que se han extraídos los datos, es decir, utilizar los datos más recientes ya que el proyecto se basa en datos del tercer trimestre del 2014.
- Realizar el mismo proyecto pero sobre los datos de la variación entre los años 2011 y 2014, así se podría hacer una comparación en las predicciones, además de asegurar que lo afirmado en el Capítulo 4 de este proyecto es cierto.
- Resolver el problema de multicolinealidad por el método de componentes principales en vez de por el método de eliminación de variables, ya que así obtendríamos una información más completa porque mediante la eliminación de variables se pierden datos que pueden llegar a ser importantes en el estudio.
- Realizar un análisis interno de Melilla para conocer más datos sobre los problemas internos que puede llegar a tener esta CC.AA. y por consiguiente, la poca Población Ocupada respecto a su Población Activa.



Capítulo 8: Referencias



8 Referencias

- www.todomktblog.com
- http://halweb.uc3m.es/esp/Personal/personas/aaribas/esp/docs/estl_grado/estlG_tema1.pdf
- <http://www5.uva.es/estadmed/datos/univariante/univar.htm>
- www.wikipedia.es
- http://web.educastur.princast.es/proyectos/formadultos/unidades/matematicas_2/ud3/2_8.html
- http://www.ditutor.com/estadistica/poligonos_frecuencias.html
- <http://www.uv.es/innomide/spss/SPSS/0201a/0201a.wiki>
- <http://ciberconta.unizar.es/leccion/anamul/inicio.html>
- http://www.uoc.edu/in3/emath/docs/T01_Reg_Lineal_Multiple.pdf
- www.ico.es
- www.melilla.es
- www.bde.es
- www.savethechildren.es
- www.salariminimo.es
- www.seg-social.es
- www.ine.es
- Libro: Estadística: Diseño de experimentos. Modelos de regresión – Rafael Romero Villafranca / Luisa Zúnica Ramajo. UPV
- Libro: Estrategia y diseño de la organización – Josep Capó Vicedo. UPV



Capítulo 9: Anexo

9 Anexo

Tabla 23. Datos utilizados para el modelo propuesto. Fuente: INE

CCAA	Pob. Total	Nº Empresas 2014	PIB per cápita	Población Activa	Población Parada	Población Ocupada	IPC
Andalucía	8.440.300	468.930	16.666	4.951.080	2.971.830	2.611.000	103,646
Aragón	1.347.150	88.114	24.732	794.819	247.337	532.200	103,849
Asturias, Principado de	1.068.165	66.342	20.591	557.689	213.206	384.200	103,685
Baleares, Illes	1.111.674	84.270	23.446	754.493	177.201	532.700	104,363
Canarias	2.118.679	128.518	18.873	1.289.640	706.791	723.900	102,403
Cantabria	591.888	36.698	21.550	339.448	112.695	230.000	104,824
Castilla - La Mancha	2.100.998	123.095	17.780	1.240.639	598.574	716.400	103,852
Castilla y León	2.519.875	159.473	21.879	1.392.231	488.856	939.200	104,729
Cataluña	7.553.650	576.565	26.666	4.730.096	1.442.747	3.074.800	105,080
Ceuta	84.180	3.590	18.771	50.643	26.896	26.400	103,080
Comunitat Valenciana	5.113.815	330.855	19.502	2.984.422	1.306.068	1.790.100	103,785
Extremadura	1.104.004	62.929	15.026	612.943	305.147	367.400	103,756
Galicia	2.765.940	191.745	20.399	1.490.288	560.103	1.017.300	104,565
Madrid, Comunidad de	6.495.551	494.509	28.915	4.134.418	1.138.670	2.746.200	103,968
Melilla	83.679	3.754	16.426	45.237	25.514	23.200	102,054
Murcia, Región de	1.472.049	86.782	17.901	895.742	386.413	530.900	104,057
Navarra, Comunidad Foral	644.477	41.582	28.358	382.368	96.156	263.300	103,680
País Vasco	2.191.682	149.245	29.959	1.230.629	333.355	864.300	104,831
Rioja, La	322.027	22.314	25.277	197.628	58.706	130.500	104,341