



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CAMPUS D'ALCOI

Búsqueda de los factores influyentes en las tasas de natalidad, fertilidad, mortalidad general e infantil en Europa.

MEMORIA PRESENTADA POR:

Mayra Jeanneth Chillagana Pilapanta

GRADO DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

TUTORA: BÁRBARA MICO

Convocatoria de defensa: Julio 2019

INDICE

1.	RESUMEN	2
2.	INTRODUCCIÓN	3
3.	METODOLOGÍA.....	5
3.1	VARIABLES EXPLICATIVAS.....	5
	MÉTODO DE INVESTIGACIÓN	7
4.	RESULTADOS	8
4.1	ANÁLISIS UNIVARIANTE Y BIVARIANTE	8
4.1.1	COMPARACIÓN DE VARIABLES A NIVEL EUROPEO	8
5.	MODELO TEÓRICO PROPUESTO (MORTALIDAD)	25
5.1	ESTUDIO DE MULTICOLINEALIDAD	26
5.2	DIAGNÓSTICO Y VALIDACIÓN DEL MODELO LINEAL.	28
5.2.1	ANÁLISIS DE LA SIGNIFICATIVIDAD:	28
5.3	ANÁLISIS DE RESIDUOS.....	30
5.4	CONTRASTE DE NORMALIDAD	32
5.5	ESTUDIO DE LA HETEROCEDASTICIDAD:	34
5.6	ESTUDIO DE AUTOCORRELACIÓN	37
5.7	PREDICCIONES MODELO MORTALIDAD	38
6.	MODELO TEÓRICO PROPUESTO (NATALIDAD)	40
6.1	ESTUDIO DE MULTICOLINEALIDAD	41
6.2	DIAGNÓSTICO Y VALIDACIÓN DEL MODELO LINEAL.	43
6.2.1	ANÁLISIS DE LA SIGNIFICATIVIDAD:	43
6.3	ANÁLISIS DE RESIDUOS.....	45
6.4	CONTRASTE DE NORMALIDAD	47
6.5	ESTUDIO DE LA HETEROCEDASTICIDAD:	49
6.6	ESTUDIO DE AUTOCORRELACIÓN	53
6.7	PREDICCIONES MODELO NATALIDAD	53
7.	SERIES TEMPORALES	55
7.1	SERIES TEMPORALES DE LA TASA DE MORTALIDAD INFANTIL	55
7.2	SERIES TEMPORALES DE LA TASA DE FERTILIDAD	60
8.	CONCLUSIONES	66
9.	BIBLIOGRAFÍA.....	67

1. Resumen

El presente estudio tiene como objetivo determinar las variables que afectan de algún modo a las tasas de natalidad, fertilidad, mortalidad general e infantil. En la actualidad, se conoce por varios medios de comunicación, redes sociales, prensa y otros, que en España existen serios problemas de longevidad poblacional, también se señala que cada vez los individuos retrasan más el hecho de tener descendencia.

Por ello, se realiza una comparativa a Nivel Europeo analizando los motivos existentes para el problema planteado y el porqué de las diferencias entre los países que más destacan y España. Se plantean 4 Modelos con la selección de variables elegidas como; empleo, índice de pobreza, actividad física, consumo de frutas y verduras, incluso el elevado consumo de cigarrillos, todas y cada una de ellas serán útiles para conocer si el problema es social o se debe a los hábitos personales, y de esta forma, obtener la explicación de cómo varían las tasas. Además, se plantean series temporales y se hacen predicciones futuras a corto plazo.

A través de las variables seleccionadas se puede conocer cómo influye para el futuro de los países Europeos, la variación de sus tasas anteriormente mencionadas, y llegar así a unas conclusiones y recomendaciones que serán de gran ayuda para la resolución del problema.

- **Palabras claves**

Tasa de Fertilidad, Tasa de Mortalidad, Tasa de Natalidad, Tasa de Empleo, Tasa de Mortalidad Infantil.

2. Introducción

La finalidad de este trabajo es crear un modelo econométrico viable y poder realizar estudios posteriores, el tema que se elige es la Búsqueda de los factores influyentes en las tasas de natalidad, fertilidad, mortalidad general e infantil en Europa, ya que en la actualidad existen menos niños y cada vez aumenta la longevidad y para esto queremos explicar cuáles son los factores significativos para la variación en las tasas mencionadas.

En la Unión europea España es uno de los países principales que tiene la tasa de natalidad y fertilidad por los suelos, las consecuencias son variadas y además públicas, por lo que se intentará explicar cuáles son las causas para este problema. Mientras que Francia y Alemania muestran todo lo contrario y con datos relevantes. Titulares como “España sufre en 2018 el menor número de nacimientos de la serie histórica” del confidencial, acreditan el hecho de que el fenómeno estudiado en este trabajo es un problema crítico en España. En la noticia hay datos alarmantes como España tiene un grave problema de natalidad y los datos del INE lo corroboran. En la primera mitad de 2018 se registraron 179.794 nacimientos de residentes en España, el peor dato desde que existen registros semestrales en 1941. Solo en el último periodo de 2018 se registró un descenso de la natalidad del 6%, lo que refleja hasta qué punto las políticas de estímulo de la maternidad y la paternidad no han dado ningún resultado (confidencial 2019)). (Figura 1)

Evolución de los nacimientos y las defunciones en la primera mitad de cada año desde 1975

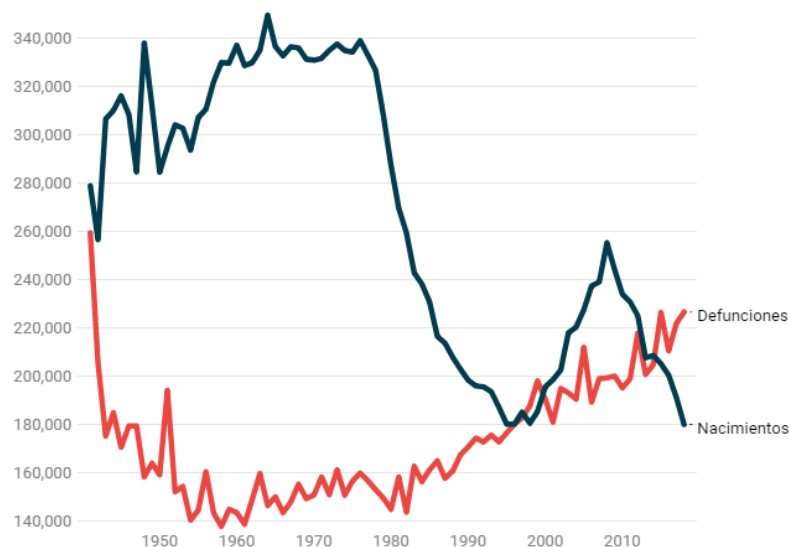


Figura 1. Evolución de la natalidad y la mortalidad en España desde 1975 hasta el 2018. Fuente el INE

Actualmente, nacen en España la mitad de los niños que lo hacían a mediados de los setenta, pero la comparativa no es mucho mejor si se miran los últimos años. Antes del estallido de la crisis, España había conseguido elevar los niveles de natalidad por encima de los 255.000 nacimientos (siempre contando la primera mitad del año), esto es, un 42% más que en la actualidad. En esos momentos, la inmigración y la bonanza económica ayudaron a elevar los índices de maternidad. Sin embargo, la crisis económica de 2008 expulsó a los inmigrantes y hundió más los niveles de natalidad de

los nacionales, lo que evidencia hasta qué punto los problemas económicos y laborales están en la base del drama de la natalidad.

La 'Encuesta de fecundidad' publicada en 2018 por el INE mostró claramente que las familias españolas quieren tener hijos, pero tienen muchos obstáculos para lograrlo. De media, las mujeres desean tener dos o más hijos. Sin embargo, un tercio de las mujeres de entre 25 y 34 años no han sido madres por motivos económicos (precariedad) y laborales (conciliación) (Figura 2) (CONFIDENCIAL 2018).

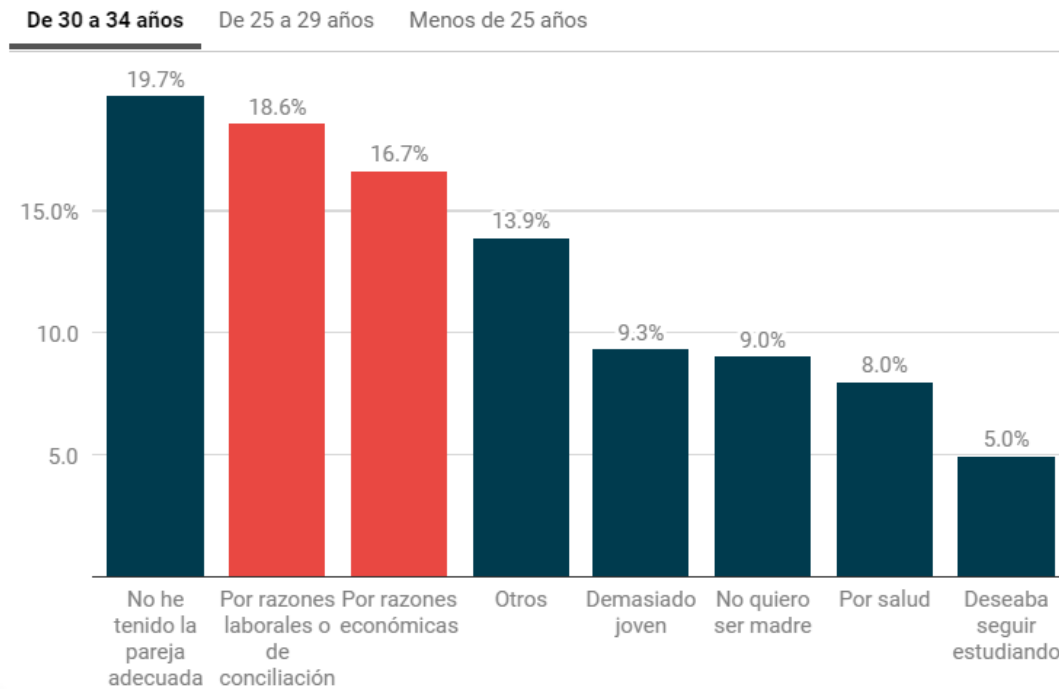


Figura 2 - Motivo principal por el cual las mujeres no han tenido hijos. Fuente el INE 2018.

El estudio realizado en este proyecto no se basa en análisis anteriores, pero sí es interesante ya que en la actualidad los medios de comunicación mencionan las situaciones conocidas que pueden influir en las tasas, son la precariedad laboral, falta de servicios sociales adecuados entre otros.

Se resuelve cualquier tipo de problema que presente los modelos planteados, como multicolinealidad, falta de Linealidad, Heterocedasticidad, Normalidad incluso Autocorrelación para tener un modelo correcto y fiable. Además, se validan series temporales de las tasas que mediante la regresión múltiple no se obtiene un R^2 alto, ya que no se pueden explicar sus datos.

Finalmente se obtiene un modelo válido y se realiza una serie de predicciones con exactitud al utilizar una base de datos reales, llegando a una serie de conclusiones sobre el estudio.

3. Metodología

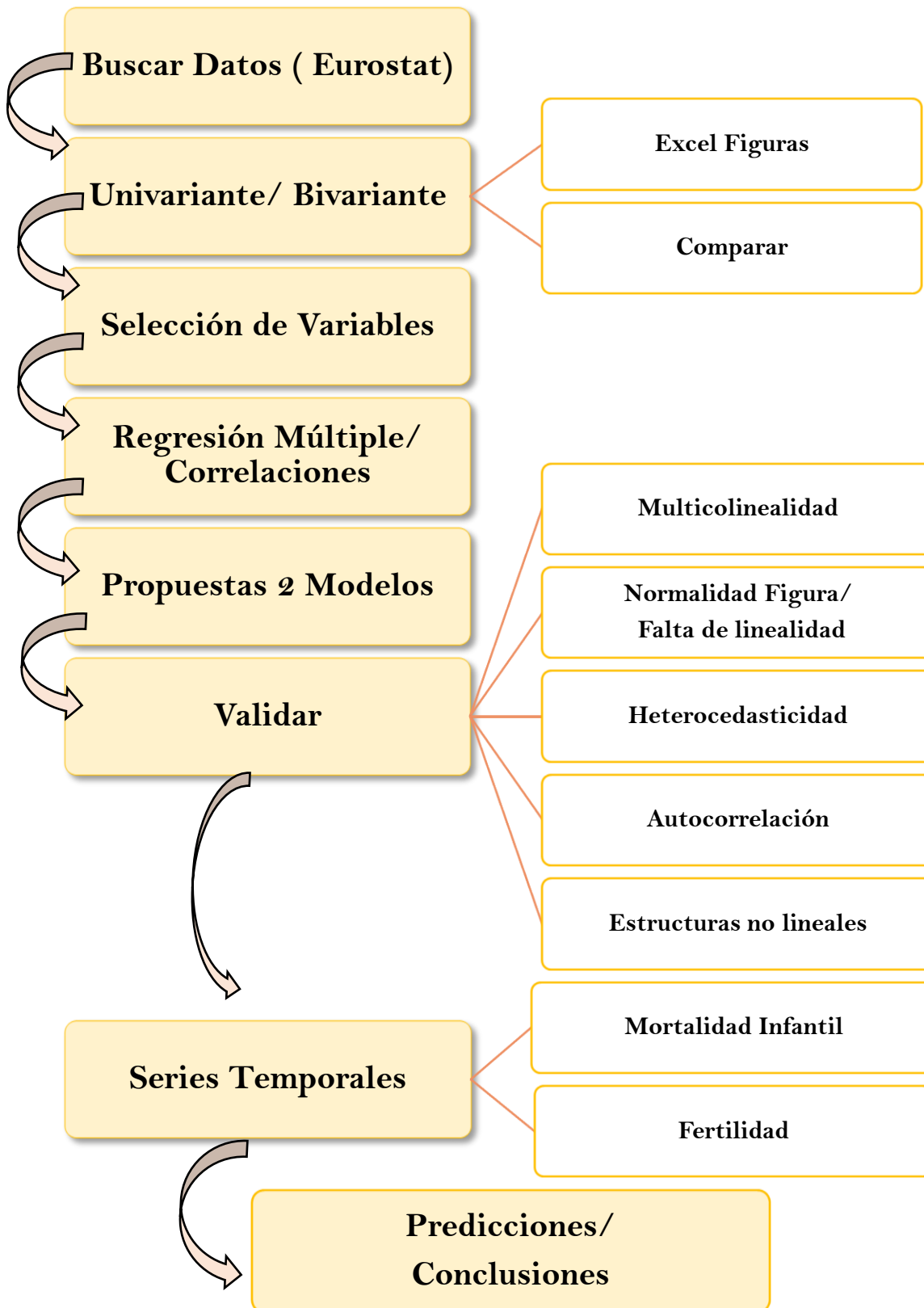
3.1 Variables explicativas

Las variables explicativas que se toma de referencia para dar a conocer el siguiente estudio se indica a continuación:

- a. Fertilidad.- Está variable proporciona una cifra para el número promedio de hijos que puede tener una mujer en su ciclo de poder dar a luz a niños, en cualquiera de las edades fértiles.
- b. Mortalidad.- Esta variable hace referencia al número de personas que fallecen con respecto al total de la población en un período de tiempo. Expresado en miles de muertes
- c. Natalidad.- Esta variables proporciona el promedio anual de nacimientos en cada uno de los países, esto depende tanto del nivel de fertilidad como de la edad de la población. Indicado en miles de nacimientos
- d. Mortalidad infantil.- La mortalidad infantil da referencia al número de muertes de bebés menores de un año en un año dado por cada 1000 nacidos vivos en cada año, esta tasa se usa con regularidad como indicador sobre el nivel de salud en cada uno de los países. Mencionado en porcentajes la muertes de niños.
- e. Consumo frutas y verduras.- Es importante el consumo de frutas y verduras para la salud de los individuos y llevar una dieta saludable para ayudar a prevenir diversas enfermedades como el cáncer, diabetes y la obesidad. Expresado en términos porcentuales
- f. Ejercicio Realizado.- En la actualidad que las personas ya sean adultos y niños tienen que llevar un ritmo de vida activa, realizando actividades físicas y así poder reducir el riesgo de muchos problemas de salud. Mencionado en términos porcentuales
- g. Fumadores.- Esta variable hace referencia a las personas que consumen tabaco y afecta directamente en la salud del individuo y de los que se encuentran a su alrededor. Ajustado en términos porcentuales
- h. Tasa de empleo.- se define a un coeficiente que permite indicar el porcentaje de personas están efectivamente empleadas. Se define la relación entre la población ocupada y población económicamente activa, y se puede formar parte del mercado laboral. Medido en miles de personas empleadas
- i. Índice de pobreza.- Se define para medir el nivel de vida de los países , incluso refleja mejor el nivel de privación de tres dimensiones básicas del Índice de Desarrollo Humano. Indicado en términos porcentuales

Se procede a tomar una base de datos de la web Eurostat que es una de las oficinas estadísticas de la Unión Europea que proporciona datos estadísticos fiables y objetivos. Estos datos permiten comparar países o regiones a nivel Europeo desde el año 2007 hasta 2016 en un campo específico de estudio y además ayuda para la toma de decisiones (EUROSTAT 2019).

• **Esquema del método**



Método de Investigación

Para la llegar al objetivo de esta investigación se clasifican las distintas etapas:

Etapa 1.- Mediante Fuentes de Información viables(Eurostat), se selecciona una serie de datos, fijando el espacio GeoFigura al cual se va a referir el estudio, en este caso se plantea a Nivel Europeo comprendido desde 2007 a 2016.

Etapa 2. Se realiza análisis univariante y bivariante comparando los datos seleccionados en la etapa anterior, para elaborar una serie de Figuras y poder plantear los modelos econométricos.

Etapa 3. Selección de Variables.- Se selecciona las variables relevantes ya sean cuantitativas o cualitativas, para el fenómeno que se pretende explicar la teoría planteada.

Etapa 4. Regresión Múltiple.- Se plantean 4 modelos para ver si las variables seleccionadas pueden llegar a explicar el 100% de los datos, o un aproximado. De los cuales 2 son válidos y significativos.

Etapa 5. Propuestas de 2 Modelos.- se obtiene 2 modelos con un R^2 muy alto que se aproxima a 1, entre ellos la tasa de natalidad y mortalidad general

Etapa 6. Se resuelve cualquier tipo de problema que presente los modelos planteados, como multicolinealidad, falta de Linealidad, Heterocedasticidad, Normalidad incluso Autocorrelación para tener un modelo correcto y fiable.

Etapa 7. Propuesta 2 Series Temporales.- la mortalidad infantil y Fertilidad da un R^2 muy lejano a 1 por tanto, estos modelos no son viables para determinar predicciones a futuro, ya que no se podrían explicar todos sus datos. Pero para estas tasas se realiza una serie temporal con datos seleccionados de 1975 hasta la actualidad

Etapa 8. Final. Tras elaborar modelos econométricos válidos ya sea por regresión múltiple o series temporales, se realiza una serie de predicciones y recomendaciones.

4. Resultados

4.1 Análisis Univariante y Bivariante

El análisis univariante y bivariante se realiza empleando los programas de cálculo y análisis estadístico de Excel y Statgraphics. La finalidad es realizar la descripción de las variables recopiladas en búsqueda de anomalías que puedan tenerse en cuenta o se corregidas. Además, se comprobará que si las variables recopiladas siguen una distribución normal para seleccionar la herramienta estadística más adecuada en cada tipo de comparación realizada.

4.1.1 Comparación de variables a nivel europeo

La unión Europea consta de 28 países los cuales de ellos se tendrá en cuenta aquellos que matizan del resto y se procede a realizar un contraste de hipótesis.

Se procede a realizar una comparativa entre países que más destacan en cada una de las variables cuantitativas que se efectúa el estudio:

- a) **Fertilidad.-** En este apartado se procede a realizar la comparación entre medias de la tasa de Fertilidad entre los años 2007 y 2016 de los países destacados en la Unión Europea (Figura 3).

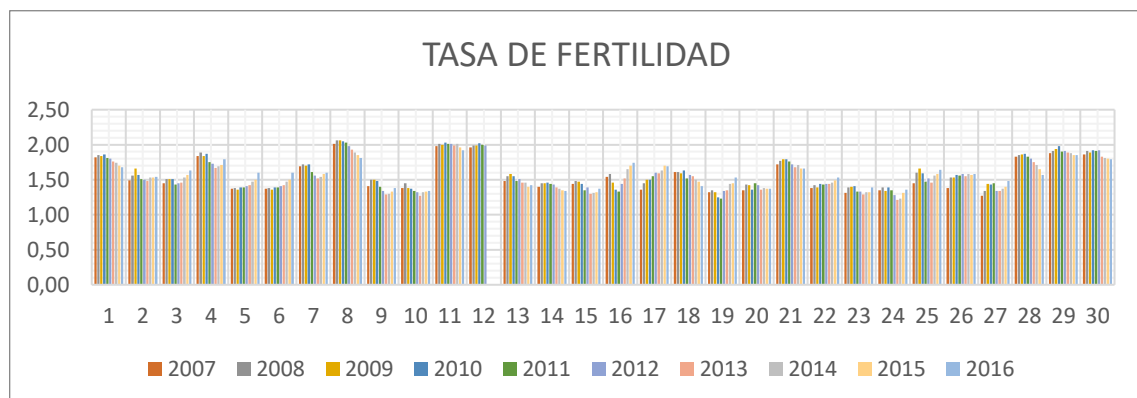


Figura 3. Tasa de Fertilidad- Fuente: Excel

Por lo tanto, se selecciona Alemania, Francia Italia , Rumania, Reino Unido y España, como se puede indicar en la Figura 3

Comparación de Varias Muestras (Variable Fertilidad)

- Muestra 1: Alemania
- Muestra 2: Bélgica
- Muestra 3: España
- Muestra 4: Francia
- Muestra 5: Irlanda
- Muestra 6: Países Bajos
- Muestra 7: Suecia

Contraste de hipótesis

$$\left. \begin{array}{l}
 H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 \\
 H_1: \text{Al menos } \mu_i \neq \mu_j
 \end{array} \right\} \alpha = 0,05$$

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	3,9131	6	0,652183	179,87	0,0000
Intra grupos	0,22843	63	0,00362587		
Total (Corr.)	4,14153	69			

Tabla 1. Análisis de Varianza -Fuente: Statgraphics

P. Valor = 0,000 < α ; **Se Rechaza H_0** , se concluye que en este modelo existe una diferencia significativa, por tanto ninguno de las muestras son iguales, como se puede observar en la Figura 4.

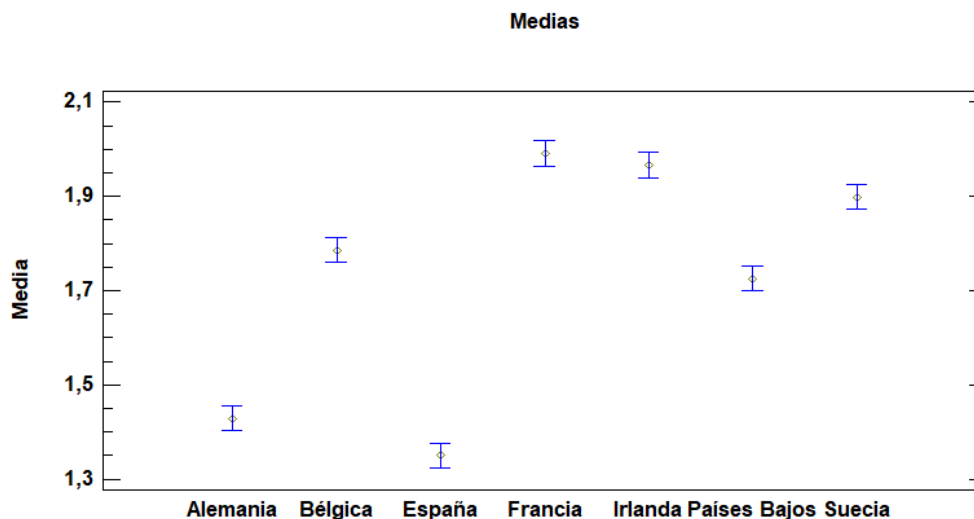


Figura 4. Comparación Medias -Fuente: Elaboración propia

El país que tiene una mayor fertilidad en la Unión Europea es Francia con una media de (1.99) por mujer, equivalente a dos hijos por cada francesa, con lo que se puede mantener la población sin necesidad a la inmigración con diferencia del resto.

En segundo lugar, de los países que encabezan una mayor fertilidad es Irlanda con una media de (1.96). Estas tasas son significativas ya que son países con una tasa de empleo femenino más alto, en estos países el compromiso de trabajo tiene mayor facilidad, las subvenciones de costes de los hijos son mucho más generosas, todo esto ayuda a tener una mejor calidad de vida tanto a los niños como a la familia en general.

A continuación, Suecia (1.90), Bélgica (1.79), Países bajos (1.73), Alemania (1.43), finalmente España se encuentra entre los países que más retrasa tener hijos en la zona europea con una media por mujer de (1.35).

Los países con una fecundidad por debajo del nivel de producción, lo que provoca es que aumente la población envejecida, induciendo desequilibrios en los gastos sanitarios y en las pensiones, esto influye en el futuro de las generaciones jóvenes, que cada vez son reducidas.

Además, en estos países con la tasa más baja se observa una tasa de empleo inferior, esto viene respaldado a que el coste por hijo es escaso , y los servicios monetarios son insuficientes incluso llegando a provocar la pobreza infantil (EL ECONOMISTA 2019).

b) Mortalidad.- En este apartado se procede a realizar la comparación entre medias de Mortalidad desde 2007 a 2016 de los países que destacan en la Unión Europea.

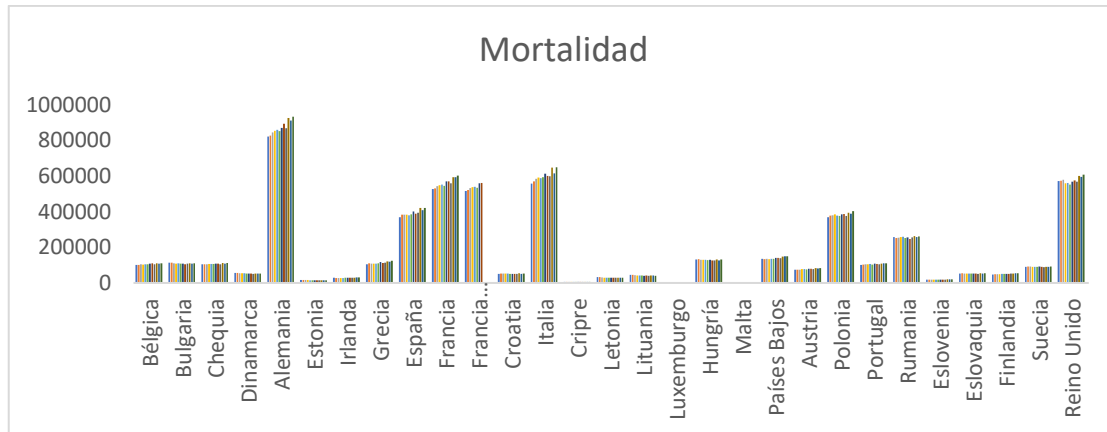


Figura 5. Tasa de Mortalidad- Fuente: Excel

Como se observa en la Figura 5, se selecciona Alemania, Francia, Italia, Reino Unido, Polonia, Rumania y España países que destacan con respecto al Resto en Europa.

Además, se comprueba que los valores del sesgo y curtosis están entre -2 y +2 y por ello se puede proceder con el ANOVA y comparar las media

Contraste de hipótesis

H0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

H1: Al menos $\mu_i \neq \mu_j$

$\alpha = 0,05$

Fuente	Suma de Cuadrados	GI	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	2,87533E12	6	4,79222E11	995,41	0,0000
Intra grupos	3,70701E10	77	4,8143E8		
Total (Corr.)	2,9124E12	83			

Tabla 2. Análisis de Varianza -Fuente: Elaboración propia

P. Valor = 0, 000 < α ; **Se rechaza Ho**, se concluye que en este modelo existe una diferencia significativa, por tanto ninguno de las muestras son iguales, como se puede observar en la Figura 6.

Gráfico de Medianas con Intervalos del 95,0% de Confianza

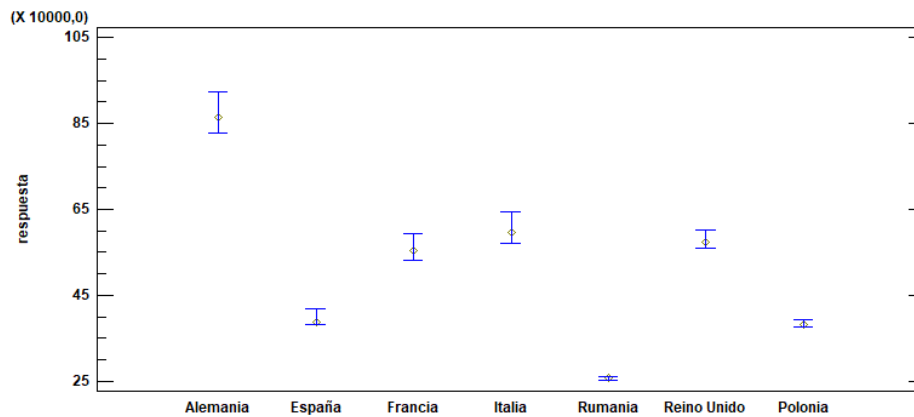


Figura 6. Comparación Medias -Fuente: Elaboración propia

			Error Est.		
	Casos	Media	(s agrupada)	Límite Inferior	Límite Superior
Alemania	12	871644	6333,97	862725,	880562,
España	12	393305	6333,97	384387,	402224,
Francia	12	561338	6333,97	552420,	570257,
Italia	12	600857	6333,97	591938,	609775,
Rumania	12	255823	6333,97	246904,	264741,
Reino Unido	12	576419	6333,97	567501,	585338,
Polonia	12	383298	6333,97	374380,	392217,
Total	84	520383			

Tabla 3. Medias -Fuente: Elaboración propia.

Como se observa el que mayor mortalidad tiene entre los países destacados en la unión Europea es Alemania con una media significativa de 871644 (miles) muertes entre el año 2006 y 2016, con diferencia del resto.

La organización mundial de la Salud hizo un informe detallando las 10 principales causas de muerte que son: cardiopatía isquémica, enfermedad cerebrovascular, infecciones del tracto respiratorio bajo (principalmente neumonía, absceso pulmonar y bronquitis aguda), VIH/sida, enfermedad pulmonar obstructiva crónica, enfermedades diarreicas, tuberculosis, cáncer de tráquea, de bronquio y de pulmón, paludismo y accidente de tránsito.

En Alemania la causa de muerte número uno son las enfermedades cardiovasculares según datos difundidos por la Oficina Central de Estadísticas de Wiesbaden. Esta media comprende tanto entre hombres y mujeres, las enfermedades más frecuentes están dolencia cardiovascular e infartos cardíacos, que se produce por tener insuficiencia circulatoria, esto para las personas en edad promedio de 65 años.

Otra de las causas de muerte que tiene un índice muy alto para los fallecimientos en Alemania, es el cáncer con una estimación del 25%, con respecto a las dolencias cardiovasculares, las causas más habituales son las afecciones coronarias que llevan al infarto cardíaco y la arterioesclerosis, y además los derrames cerebrales que llevan a producirse una apoplejía (MINDS 2019).

Con respecto al segundo país con mayor número de muertes, esta Italia con una media de 600857(miles) personas fallecidas entre hombres y mujeres. A continuación está

Reino Unido con una media de 576419 (miles) y Francia con 561338 (miles) fallecimientos. De las principales causas antes mencionadas, se concluye que en estos países las causas del fallecimiento son el cáncer y las enfermedades crónicas, como insuficiencia respiratoria o/y cardiaca (AGENCIA EFE 2019).

Por otro lado, los países con una menor tasa de mortalidad está España 393305, Polonia 383298 (miles) y Rumania 255823 (miles) muertes.

c) Natalidad.- En este apartado se procede a realizar la comparación entre medias de la tasa de natalidad desde 2007 a 2016 de los países que destacan en la Unión Europea.

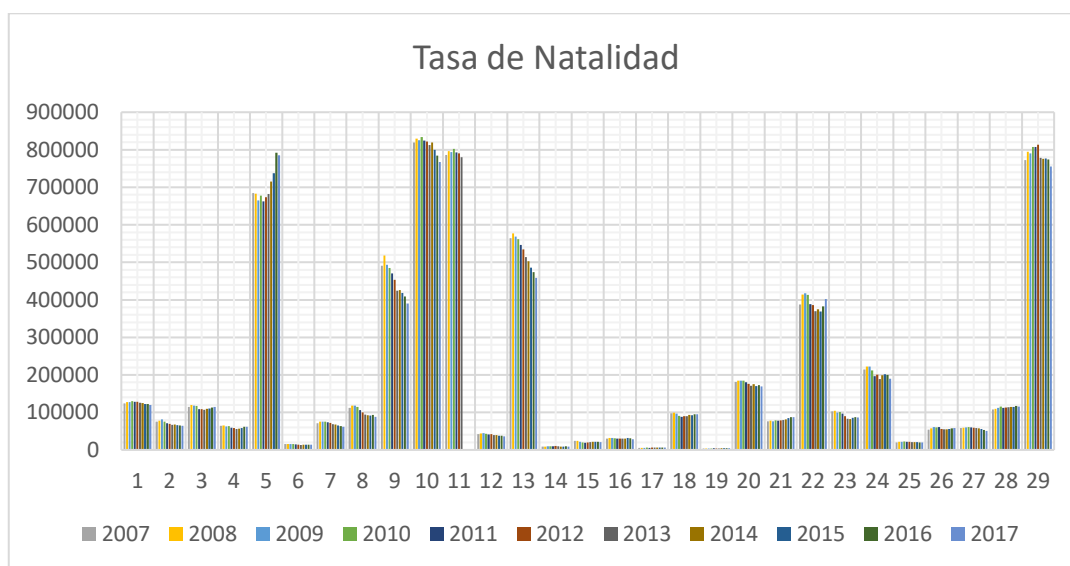


Figura 7. Tasa de Natalidad - Fuente: Excel.

En la Figura 7, se observa que entre los países que más destacan está Alemania, Francia, Italia, Polonia, Reino Unido y España

Comparación de Varias Muestras (Variable Natalidad)

- Muestra 1: Alemania
- Muestra 2: España
- Muestra 3: Francia
- Muestra 4: Italia
- Muestra 5: Polonia
- Muestra 6: Reino Unido

Contraste de hipótesis

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

$H_1: \text{Al menos } \mu_i \neq \mu_j$

$\alpha = 0,05$

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	1,90768E12	5	3,81537E11	351,47	0,0000
Intra grupos	7,16466E10	66	1,08555E9		
Total (Corr.)	1,97933E12	71			

Tabla 4. Análisis de Varianza -Fuente: Elaboración propia.

P. Valor = 0, 000 < α ; **Se Rechaza Ho**, se concluye que en este modelo existe una diferencia significativa, por tanto ninguno de las muestras son iguales, como se puede observar en la Figura 8.

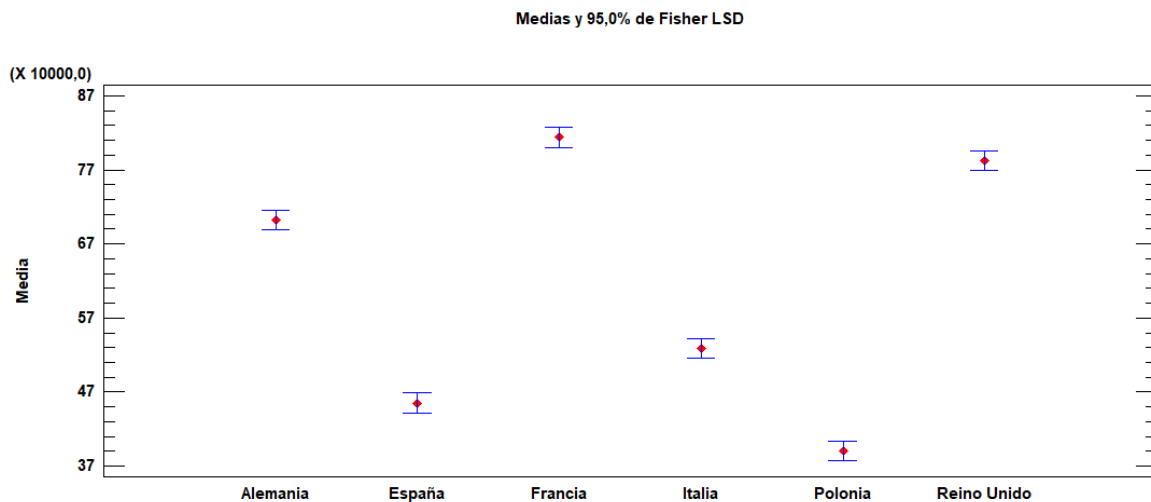


Figura 8. Comparación medias - Fuente: Elaboración propia

			Error Est.		
	Casos	Media	(s agrupada)	Límite Inferior	Límite Superior
Alemania	12	702593	9511,2	689165,	716021,
España	12	455126	9511,2	441698,	468554,
Francia	12	813991	9511,2	800563,	827418,
Italia	12	528871	9511,2	515443,	542298,
Polonia	12	390038	9511,2	376611,	403466,
Reino Unido	12	782821	9511,2	769393,	796249,
Total	72	612240,			

Tabla 5. Análisis de Medias -Fuente: Elaboración propia

Uno de los países con mayor tasa de natalidad al igual que la variable anterior la tasa de fertilidad se observa que es Francia es quien está en primer lugar, ya que nacieron entre una media de 813991 (miles) niños (varones y mujeres) entre 2006 y 2016, aunque en el último año va descendiendo en un 1.82%.

El segundo y tercer país con un mayor número de nacimientos es Reino Unido con 782821 (miles) y Alemania 702893 (miles) niños y niñas, estos países con mayor natalidad pueden llegar a tener una población estable.

Esto es debido a que existe una mayor tasa de empleo y los padres tienen trabajos estables por tanto, tienen más recursos económicos y pueden cubrir gastos ocasionados por los niños con mayor facilidad, en estos países existen un mayor número de ayudas o subvenciones por hijo por lo que ayuda a mejorar la forma de vida familiar.

Los que menos nacimientos tienen son Italia con 528871 (miles), España con 455126 (miles) y Polonia 390038 (miles) de niños y niñas, entre los países que más destacan en Europa. Este índice mientras más bajo sea puede provocar una población muy envejecida y que se agote el fondo de las pensiones, las mujeres se plantean mucho en

tomar la decisión de tener hijos ya que en la actualidad es complicado la incorporación al mundo laboral.

A diferencia de los países que tienen la tasa más elevada, esto puede deberse al volumen y cuantías económicas que se ofrece (pobre en comparación con Francia, Reino Unido y Alemania). En estos países las familias se acogen a menos beneficios estatales, por tanto la disposición de tener hijos es más difícil de tomarla.

Otro problema en la actualidad es que existen puestos de empleo temporales, por lo que puede provocar un aumento a la pobreza y paro, influye también en que no hay igualdad, sin duda menos partidas económicas y por último se menciona la brecha salarial que existe principalmente en España que una mujer cobra por debajo en comparación de un hombre (EXPANSION 2019).

d) Mortalidad infantil.- En este apartado se procede a realizar la comparación entre medias de la tasa de mortalidad Infantil desde 2007 a 2016 de los países que destacan en la Unión Europea.

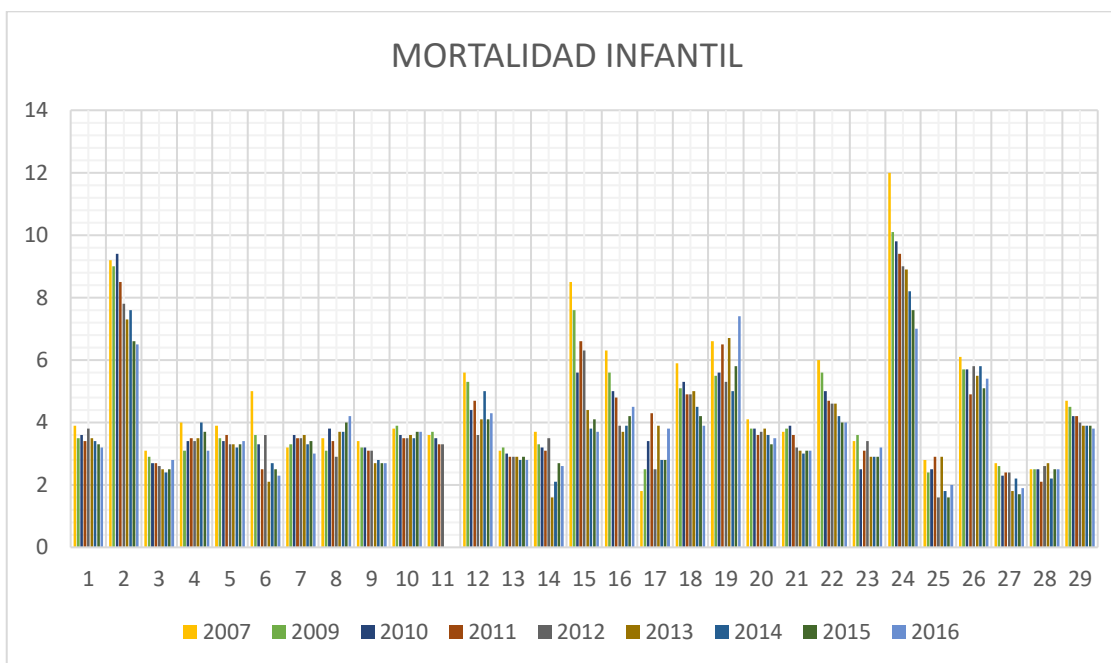


Figura 9. Mortalidad Infantil - Fuente: Elaboración propia

Como se observa en la Figura 9. Mortalidad Infantil los países que destacan son: Bulgaria Eslovenia, España Croacia, Letonia, Malta y Rumania.

Comparación de Varias Muestras

- Muestra 1: Bulgaria
- Muestra 2: Eslovenia
- Muestra 3: España
- Muestra 4: Croacia
- Muestra 5: Letonia
- Muestra 6: Malta
- Muestra 7: Rumania

Contraste de hipótesis

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$

$H_1: \text{Al menos } \mu_i \neq \mu_j$

$\alpha = 0,05$

Tabla ANOVA

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	537,688	6	89,6147	48,50	0,0000
Intra grupos	142,261	77	1,84754		
Total (Corr.)	679,949	83			

Tabla 6. Análisis de Varianza de mortalidad infantil -Fuente: Elaboración propia

P. Valor = 0, 000 < α ; **Se rechaza H_0** , se concluye que en este modelo existe una diferencia significativa, por tanto, ninguno de las muestras son iguales, como se puede observar en la Figura 10.

Gráfico de Medianas con Intervalos del 95,0% de Confianza

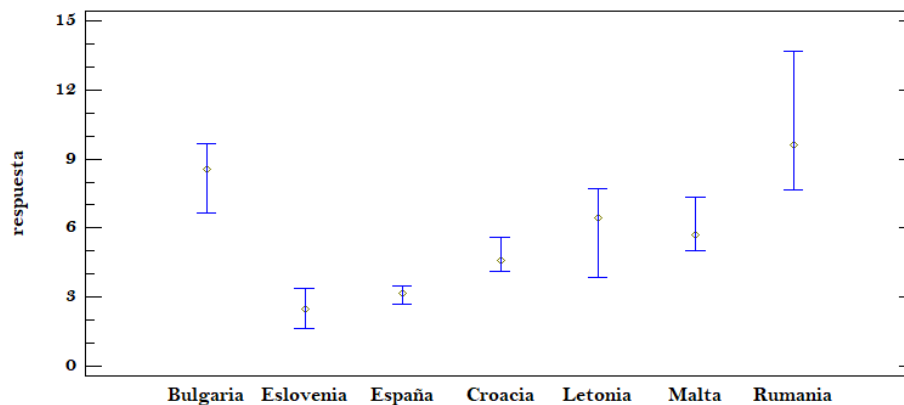


Figura 10. Mortalidad Infantil - Fuente: Statgraphics

	Tamaño de Muestra	Rango Promedio
Bulgaria	12	67,1667
Eslovenia	12	9,70833
España	12	15,9167
Croacia	12	35,625
Letonia	12	47,4583
Malta	12	47,375
Rumania	12	74,25

Tabla 7. Análisis de Medias de mortalidad infantil -Fuente: Elaboración propia

Con la mayor tasa de mortalidad infantil a nivel Europeo es Rumania con una media del 74,25 % y Bulgaria con un 67,18% entre los años 2005 y 2016 esto con diferencia a los adultos los niños son más vulnerables a las amenazas ambientales, uso sistemático de los sistemas de retención infantil para el coche, enfermedades y otros, por lo que esto depende a la legislación que adopte a las necesidades de cada uno de los países.

A continuación, se muestra Letonia con una media del 47,45% y Malta con un 47,37% de muertes de niños, principalmente se debe a afecciones no transmisibles o

nutricionales, además de enfermedades respiratorias crónicas que se producen la mayoría de las muertes en el mundo desarrollado.

Como se puede observar sigue Croacia con 35,62% que tiene una media inferior a los países anteriores, esto se debe a que en este país se han comprometido a reconocer y respetar principalmente los derechos humanos fundamentales. Aunque el cumplimiento de los derechos infantiles no han estado a la altura de los objetivos marcados (BANCO MUNDIAL 2019).

Finalmente, por debajo está España con una media de 15,92% y Eslovenia con 9,70%, las causas son maformidades congénitas, deformidades y anomalías cromosómicas, sepsis bacteriana y hemorragia del recién nacido. Destacando entre los países con menor tasa de mortalidad. El 10% de los niños sufren accidentes en su primer año de vida principalmente golpes, caídas o quemaduras (SALUD 2019).

e) Consumo frutas y verduras.- En este apartado se procede a realizar la comparación entre medias sobre el consumo de frutas y verduras por porciones inferiores a 1, entre 1 a 4 y superiores a 5 porciones

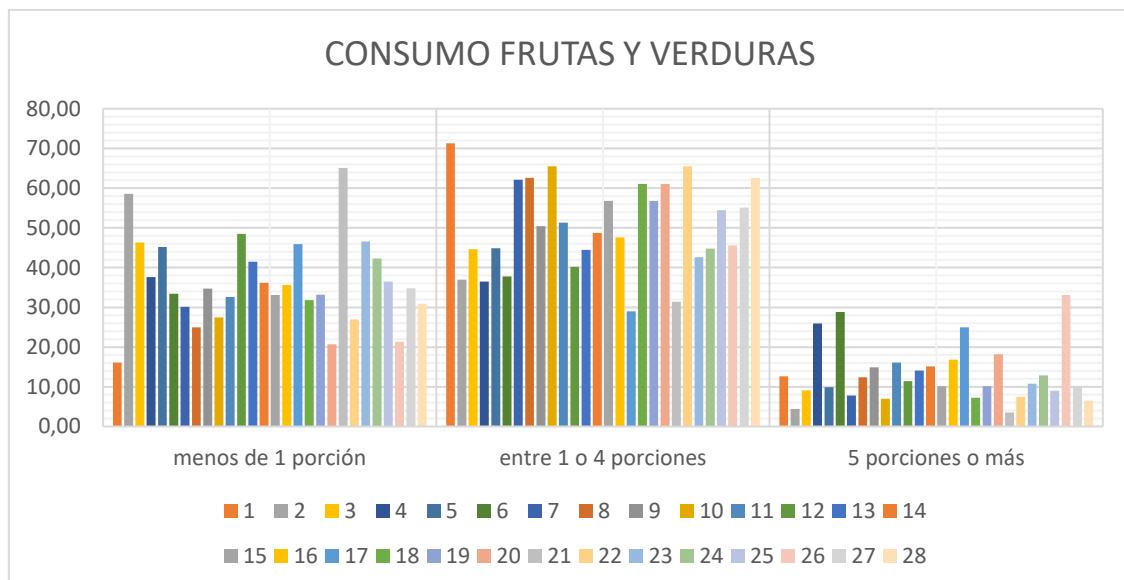


Figura 11. Consumo de Frutas y Verduras - Fuente: Elaboración Propia

Como se observa en la Figura 11. Consumo de Frutas y Verduras destacan que los ciudadanos de la Unión Europea consumen entre 1 a 4 porciones con mayor regularidad.

Comparación de Varias Muestras

- Muestra 1: 5 porciones o más
- Muestra 2: entre 1 o 4 porciones
- Muestra 3: menos de 1 porción

Contraste de hipótesis

H0: $\mu_1 = \mu_2 = \mu_3$
H1: Al menos $\mu_i \neq \mu_j$



$\alpha = 0,05$

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	19757,8	2	9878,89	99,92	0,0000
Intra grupos	8008,35	81	98,8685		
Total (Corr.)	27766,1	83			

Tabla 8. Análisis de Consumo de frutas y ver. -Fuente: Elaboración propia

P. Valor = **0,000** < α ; *Rechazamos H_0* , concluimos que en este modelo existe una diferencia significativa, por tanto, ninguno de las muestras son iguales, como se puede observar en la Figura 12.

Medias y 95,0% de Fisher LSD

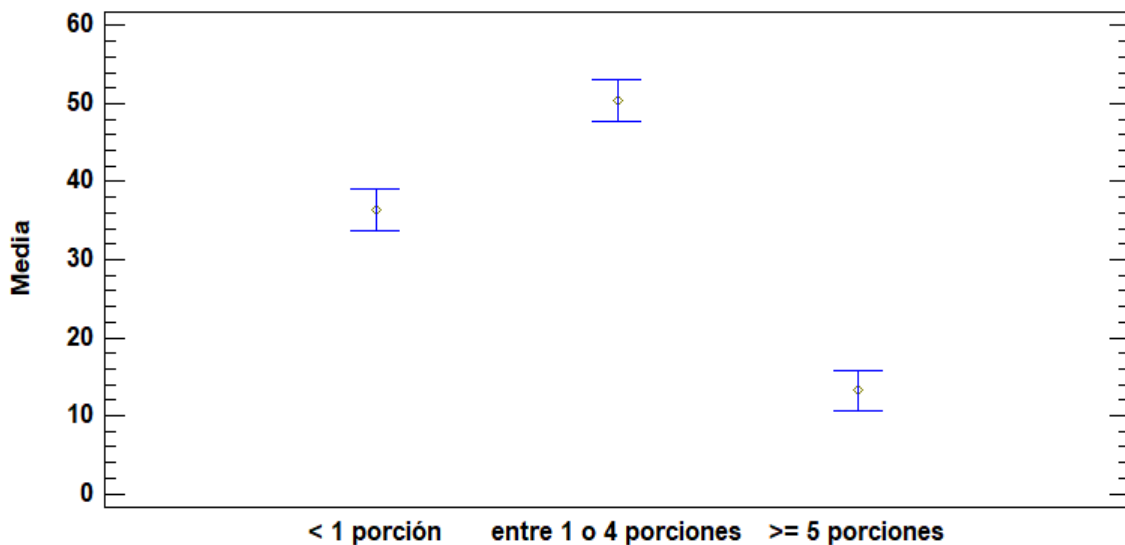


Figura 12. Comparación Medias- Fuente: Elaboración propia

Muchos Organismos internacionales recomienda que lo mejor es llevar una dieta que incluya más de 5 porciones de frutas y verduras al día llevando una media del 13%, otros componentes de las frutas y verduras son esenciales para la salud humana, como se puede observar en la Figura 12 el consumo las de más de 5 porciones es la media más baja a comparación con los otros.

La media que se destaca en el consumo de frutas de las personas en la Unión Europea es entre una porción a 4 porciones con un 50,42%, por lo que se concluye que es muy bajo de lo recomendado para llevar una vida saludable, también lo que ha llamado la atención sobre un aspecto, es el desperdicio de frutas y verduras que tienen los hogares llegando a terminar en la basura. En general son las mujeres y los que comprenden más de 40 años quienes superan más de una ración,

Mientras que hay un 36,36% de personas europeas que consumen menos de una porción de frutas y verduras al día, puede ser alarmante porque lo que consumen es la bollería incluyendo dulces de manera regular a su dieta, llegando a producir con mayor facilidad enfermedades cardiovasculares, colesterol afectando al mal funcionamiento del corazón (EL ECONOMISTA 2019).

f) **Ejercicio Realizado.-** En este apartado se procede a realizar la comparación entre medias sobre qué actividades físicas se realiza con mayor frecuencia en la Unión Europea. Entre ellas están caminar, ciclismo, deporte aeróbicos y fortalecimiento muscular.

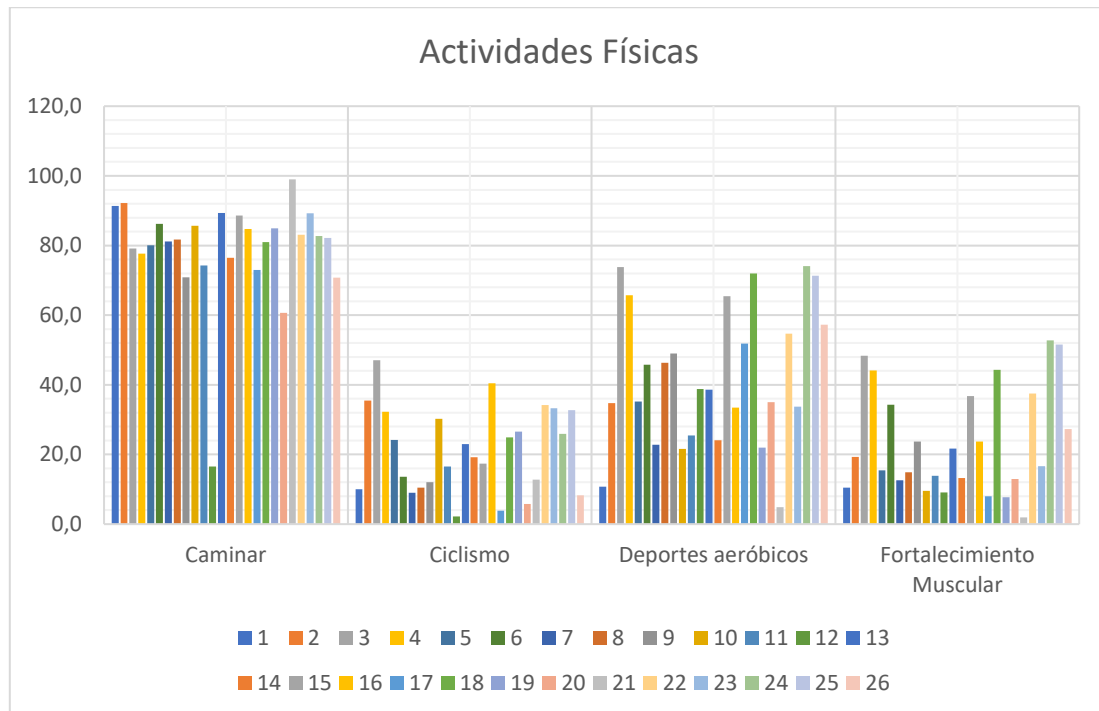


Figura 13. Actividades Físicas - Fuente: Elaboración propia

Como se observa en la Figura 13 los Europeos la actividad física que realizan más a menudo es caminar a los sitios, seguido de deportes aeróbicos y fortalecimiento muscular, en menos proporción ciclismo

Comparación de Varias Muestras

Muestra 1: Aeróbicos
Muestra 2: Caminar
Muestra 3: Ciclismo
Muestra 4: Musculación

Contraste de hipótesis

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

$H_1: \text{Al menos } \mu_i \neq \mu_j$

$\alpha = 0,05$

Tabla ANOVA

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	56345,3	3	18781,8	74,93	0,0000
Intra grupos	25065,4	100	250,654		
Total (Corr.)	81410,6	103			

Tabla 9 Análisis de Ejercicio Realizado - Fuente: Elaboración propia

P. Valor = $0,000 < \alpha$; **Se rechaza H_0** , se concluye que en este modelo existe una diferencia significativa, por tanto ninguno de las muestras son iguales, como se puede observar en la Figura 14.

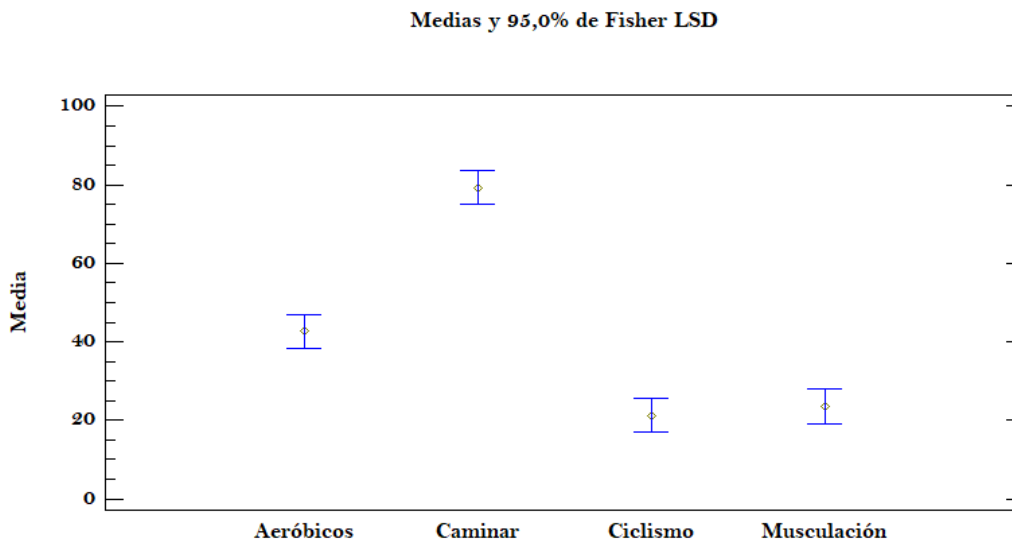


Figura 14. Medias ejercicio realizado- Fuente: Elaboración propia

Resumen Estadístic

	Recuento	Promedio	Desviación Estándar	Rango
Aeróbicos	26	42,6308	19,911	69,2
Caminar	26	79,3346	15,0858	82,5
Ciclismo	26	21,2115	12,1311	44,9
Musculación	26	23,5308	15,2126	50,9
Total	104	41,6769	28,114	97,1

Tabla 10. Promedio Ejercicio Real- Fuente: Elaboración propia

Según el informe global de la Organización Mundial de la Salud, es importante realizar el suficiente ejercicio físico para mantener el bienestar físico y reducir enfermedades del corazón diabetes de tipo dos y algunos tipos de cáncer. La falta de actividad física es el cuarto elemento de riesgo que aumenta la tasa de mortalidad.

Se observa que un 80% de personas que realizan actividades caminar para ir y volver de los sitios, esto ayudarán a envejecer mejor a partir de los 30 años, ya que el metabolismo va cada vez más lento. Destacan al hacer ejercicio las mujeres con diferencia de los hombres.

A continuación, la actividad que más gente lo realiza con un promedio de 42,63% son deportes aeróbicos. Son ejercicios de media intensidad y de una duración larga que ayuda a quemar grasas e hidratos, esto hace obtener energía y para ello necesita oxígeno.

Aunque con una media inferior está el ciclismo y el fortalecimiento muscular con un 21%y 23% respectivamente. Este deporte lo realizan para disminuir grasa ayudando a preservar y mejorar la masa muscular a cualquier edad. También se menciona que el deporte va relacionado con el nivel económico y educativo (MAGAZINE 2019).

- g) Fumadores.-** En este apartado se procede a realizar la comparación entre medias sobre los fumadores en la Unión Europea.

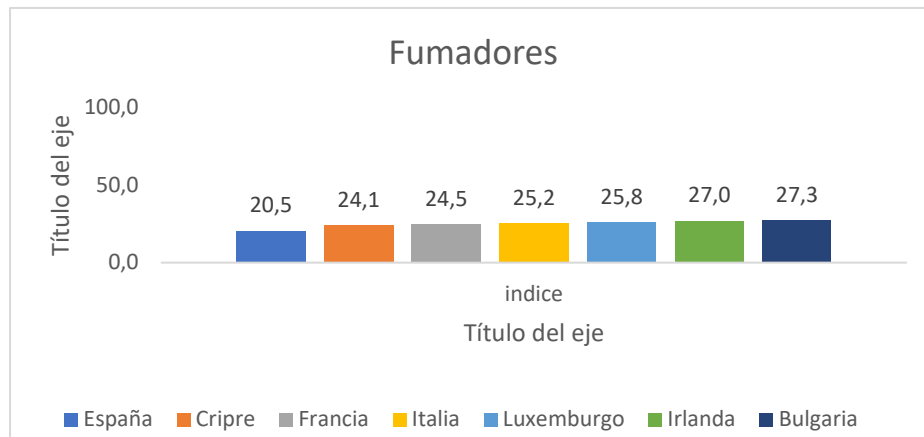


Figura 15. Fumadores - Fuente Elaboración propia

Como se puede observar en la Figura 15. El país con mayor número de fumadores es Bulgaria con un 27,3%, seguido de Irlanda con un 27% , Italia y Luxemburgo tienen un 25,2% y 25,8% aproximadamente.

En Italia y Francia representa un 24,5 y 24,1% de fumadores, entre los países destacados de la Unión Europea España tiene un 20,5% de fumadores. Que a diferencia de otros países muestra un nivel preocupante para la sociedad.

Aunque en la actualidad hay varias campañas para concienciar a la gente que deje de fumar no es suficiente, ya que la media que presenta es muy elevada, Los individuos no tienen en cuenta que es perjudicial para la salud no solo para el que lo hace, sino para quien se encuentre a su lado.

- g) Tasa de empleo.-** En este apartado se procede a realizar la comparación entre medias de la tasa de empleo desde 2007 a 2016 de los países que destacan en la Unión Europea.

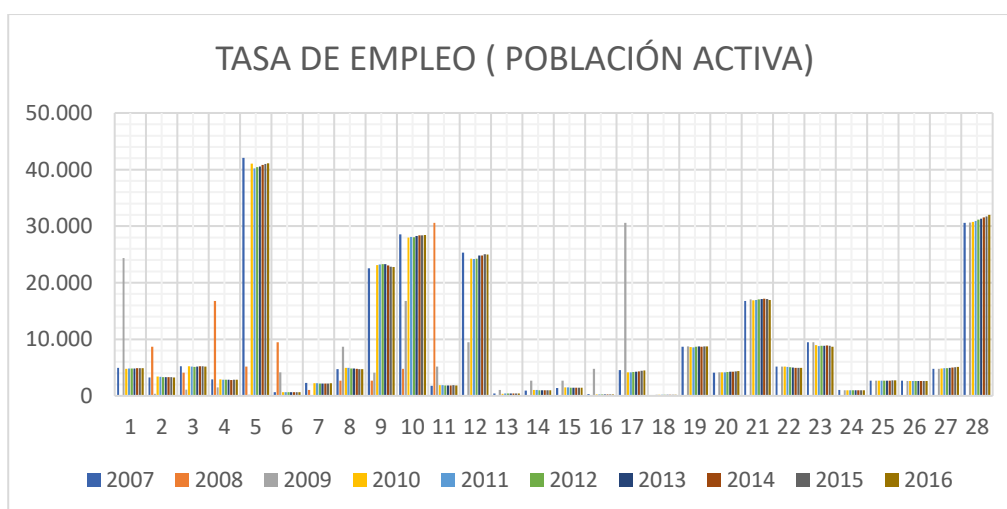


Figura 16. Tasa de empleo - Fuente – Elaboración propia

Como se observa en la Figura 16. Tasa de empleo los países que destacan son Alemania, Francia, Reino Unido, Rumania, Italia, Polonia y España.

Comparación de Varias Muestras

Muestra 1: Alemania
 Muestra 2: España
 Muestra 3: Francia
 Muestra 4: Reino Unido
 Muestra 5: Rumania
 Muestra 6: Italia
 Muestra 7: Polonia

Contraste de hipótesis

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

$H_1: \text{Al menos } \mu_i \neq \mu_j$

$\alpha = 0,05$

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	5,63093E9	6	9,38488E8	7470,75	0,0000
Intra grupos	7,03481E6	56	125622,		
Total (Corr.)	5,63796E9	62			

Tabla 11. Análisis de Tasa de empleo - Fuente: Elaboración propia

P. Valor = **0, 000** < α ; **Se rechaza H_0** , se concluye que en este modelo existe una diferencia significativa, por tanto, ninguno de las muestras son iguales, como se puede observar en el Figura 15. Medias Tasa de empleo.

Gráfico de Medianas con Intervalos del 95,0% de Confianza

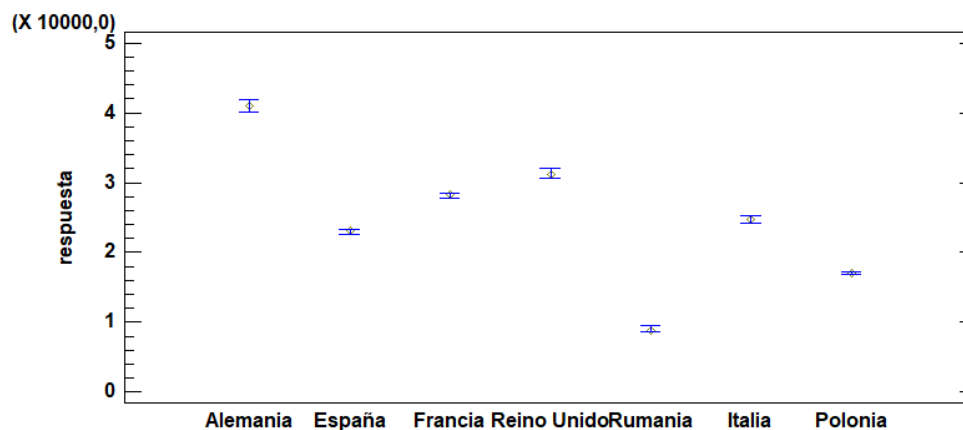


Figura 17. Medias Tasa de empleo - Fuente: Elaboración propia

	Recuento	Promedio
Alemania	9	40896,4
España	9	23007,4
Francia	9	28197,6
Reino Unido	9	31186,8
Rumania	9	8979,67
Italia	9	24665,1
Polonia	9	17007,0
Total	63	24848,6

Tabla 12. Comparación Medias Tasa empleo - Fuente: Elaboración propia

Uno de los países con mayor tasa empleo es Alemania quien está en primer lugar, ya que tiene un promedio de 40896.4 (miles) de personas esto se debe a que la Economía en este país tiene una estabilidad muy buena. El paro en este país muestra un notable descenso y actualmente tiene un PIB elevado.

El segundo y tercer país con un mayor empleo es Reino Unido con 31186.8(miles) y Francia (28197.6)miles) de personas, Una notable diferencia entre los países , En estos países siguen entre los mejores países con una tasa de empleo alta. El empleo ha incrementado para los hombres, aunque de las mujeres ha permanecido estable.

Italia y España tiene un 24665,1 (miles) , 23007,4 (miles) personas respectivamente, por lo contrario el empleo de hombres en estos países se ha reducido, pasando el aumento de contratación de mujeres. Aunque es más considerable cuando empieza la temporada de verano, ya que son países que tienen una temperatura muy agradable para los turistas. Por otra parte, La ayuda para las empresas que contraten a personas menores de 30 años a beneficiado a reducir la tasa de paro (EL PAIS 2019).

Finalmente entre los destacados está Polonia con 17007 (miles) personas e Rumania 8979,67(miles) personas, el bajo desarrollo económico afecta directamente en el empleo ya que se ha reducido puestos de empleo, recortes de salarios públicos también afectan en el bajo índice (LA VANGUARDIA 2019).

- h) Índice de pobreza.-** En este apartado se procede a realizar la comparación entre medias de la tasa de Índice de pobreza desde 2007 a 2016 de los países que destacan en la Unión Europea.

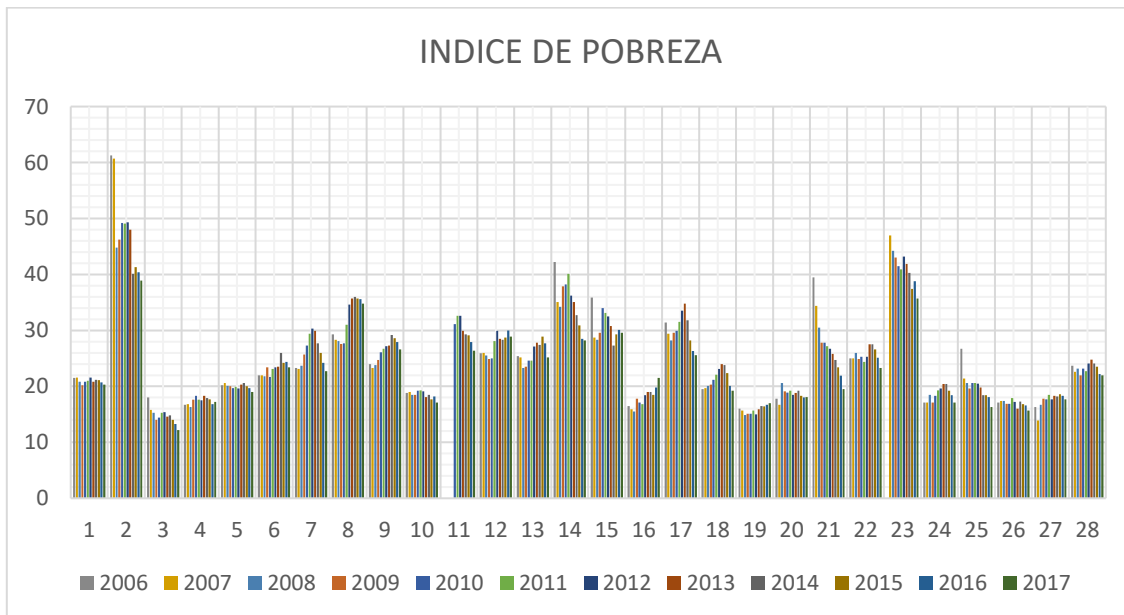


Figura 18. Índice de Pobreza - Fuente: Elaboración propia

En la Figura 18, se indica entre los países que destacan con un mayor índice de pobreza están Bulgaria Hungría, Grecia, Lituania, Portugal, Rumania, Letonia y España.

Comparación de Varias Muestras

Muestra 1: Bulgaria

Muestra 2: España

Muestra 3: Hungría

Muestra 4: Grecia
Muestra 5: Lituania
Muestra 6: Portugal
Muestra 7: Rumania
Muestra 8: Letonia

Contraste de hipótesis

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

$H_1: \text{Al menos } \mu_i \neq \mu_j$

$\alpha = 0,05$

Tabla ANOVA

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	4577,89	7	653,984	48,23	0,0000
Intra grupos	1071,13	79	13,5586		
Total (Corr.)	5649,02	86			

Tabla 13. Análisis Índice de pobreza - Fuente: Elaboración propia

P. Valor = **0,000 < α** ; **Se rechaza H_0** , se concluye que en este modelo existe una diferencia significativa, por tanto, ninguno de las muestras es iguales, como se puede observar en la Figura 19.

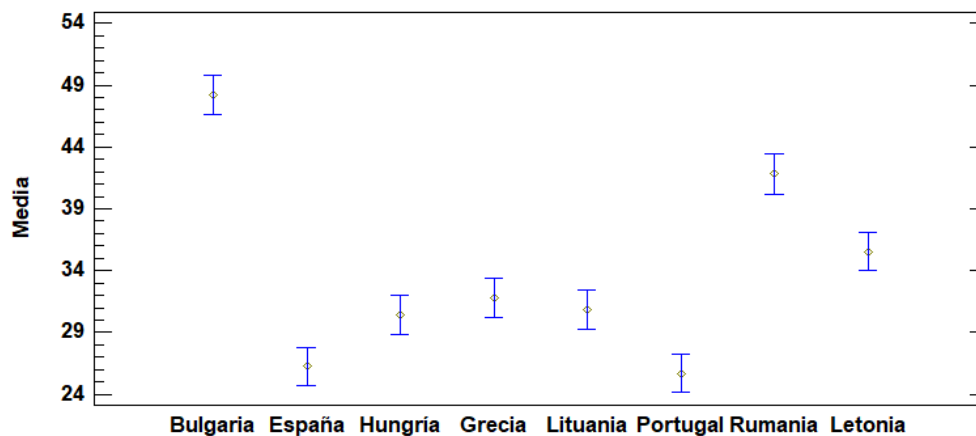


Figura 19. Medias de Índice de Pobreza - Fuente: Elaboración propia

Resumen Estadístico

	Recuento	Promedio
Bulgaria	11	48,2182
España	11	26,2545
Grecia	11	31,7818
Rumania	10	41,82
Letonia	11	35,5545
Total	87	33,7345

Tabla 14. Resumen Elaboración propia

Bulgaria con un 48,21% es uno de los países con mayor pobreza, de la Unión Europea, su situación económica es considerada pésima, por lo que los ciudadanos la mejor opción para solucionar este inconveniente es emigrar a un país vecino, ya que en su país no les da una mejor calidad de vida. Los problemas que presentan es que aumenta la población longeva, cada vez existe menos creación de nuevos. No obstante lo que más perjudica a la sociedad es la corrupción y el crimen organizado que lastran a la comunidad Búlgara (EL PERIODICO 2019).

Rumania con un 41,82 % al igual que Bulgaria los problemas que presentan para tener una pobreza tan elevada son similares, que además están obligados a emigrar para mejorar su economía familiar.

Letonia con un 35,55% y Grecia con un 31,78% de índice de pobreza, los hogares de estos países no obtienen suficiente Renta para mediar con los gastos. Los ingresos que tienen estos ciudadanos están por debajo del Umbral de pobreza. Principalmente la pobreza se presenta con mayor notoriedad en las mujeres a diferencia de los hombres aunque los mayores de 65 son quienes tienen la mayor tasa (EXPANSION 2019).

España son países con mayor tasa de pobreza con un 26,2% que se encuentra por encima de la media europea, aunque con la comparación de otros países tiene una media inferior al 30%, Los menores de 18 años son quienes presentan la mayor tasa de riesgo de pobreza, por lo que hay familias que tienen muchos más gastos que ingresos y no llegan a fin de mes. Se busca mejorar la eficiencia haciendo compatible prestaciones con empleo, simplificar el sistema de rentas mínimas, esto para evitar solapamientos entre las prestaciones de las diferentes administraciones (MUNDO 2019).

5. Modelo Teórico Propuesto (Mortalidad)

En este apartado se va a proponer el primer modelo de regresión múltiple para explicar la variabilidad de la mortalidad en Europa. El modelo inicial propuesto es el siguiente teniendo en cuenta las variables recopiladas:

$$\text{Mortalidad} = \beta_0 + \beta_1 * \text{ACTIV. FISICAS} + \beta_2 * \text{CIGARRILLO} + \beta_3 * \text{CONSUMO} + \beta_4 * \text{EMPLEO} + \beta_5 * \text{POBREZA} + U$$

La primera estimación del modelo lineal propuesto resulta como sigue:

$$\text{Mortalidad} = -10382,3 + 234,007 * \text{actividades} + 695,181 * \text{cigarillo} - 1236,13 * \text{consumo} + 21,046 * \text{empleo} + 1824,05 * \text{pobreza}$$

- **Interpretación de los parámetros y unidades físicas**

Seguidamente se procede con la validación del modelo planteado, empezando por la interpretación y validación de los parámetros estimados.

β_0 = Es el valor medio de la tasa de Mortalidad cuando las variables X valen cero.

β_1 = Es el incremento medio de Mortalidad cuando aumenta en una unidad de **act. Físicas** y el resto de las variables se mantiene constante.

β_2 = Es el incremento medio de Mortalidad cuando aumenta en un 1% el **Consumo de cigarillo** y el resto de las variables se mantiene constante.

β_3 = Es el incremento medio de Mortalidad cuando aumenta en un 1% el índice del **Consumo de frutas y verduras** y el resto de las variables se mantiene constante.

β_4 = Es el incremento medio de Mortalidad cuando aumenta en una unidad el **empleo** y el resto de las variables se mantiene constante.

β_5 = Es el incremento medio de Mortalidad cuando aumenta en 1% la **pobreza** y el resto de las variables se mantiene constante.

U = Perturbación del modelo econométrico, parte del modelo que no hemos tenido en cuenta en el modelo.

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,4649E12	5	2,92979E11	337,74	0,0000
Residuo	1,73494E10	20	8,6747E8		
Total (Corr.)	1,48224E12	25			

Tabla 15. Análisis de varianza Fuente: Statgraphics

R-cuadrada = 98,8295 por ciento

R-cuadrado (ajustado para g.l.) = 98,5369 por ciento

Error estándar del est. = 29452,8

Error absoluto medio = 17481,5

Estadístico Durbin-Watson = 2,07281 (P=0,5552)

Autocorrelación de residuos en retraso 1 = -0,13823

TABLA DE DATOS

PAISES	Mortalidad (miles)	Actividades (%)	Cigarrillo (%)	Consumo (%)	Empleo (miles)	Pobreza (%)
Bulgaria	107580	91,40	27,30	37,00	3200	40,40
Chequia	107750	92,20	21,20	44,60	5226	13,30
Dinamarca	52824	79,10	12,30	36,50	2934	16,80
Alemania	910902	77,70	15,00	44,90	41932	19,70
Estonia	15392	80,10	22,70	37,80	658	24,40
Irlanda	30540	86,20	27,00	62,10	2260	24,20
Grecia	118785	81,20	22,20	62,60	4732	35,60
España	408231	81,70	20,50	50,40	22657	27,90
Francia	594005	70,90	24,50	65,50	28471	18,20
Croacia	51542	85,70	17,40	51,30	1806	27,90
Italia	615261	74,30	25,20	40,20	25243	30,00
Chipre	5471	16,50	24,10	44,50	408	27,70
Letonia	28580	89,30	20,20	48,70	957	28,50
Lituania	41106	76,50	13,80	56,80	1433	30,10
Luxemburgo	3967	88,60	25,80	47,60	277	19,80
Hungría	127098	84,70	18,90	29,00	4543	26,30
Malta	3342	73,00	17,20	61,10	215	20,10
Austria	80669	81,00	23,90	61,10	4412	18,00
Polonia	388009	84,90	21,90	31,40	16961	21,90
Portugal	110573	60,70	16,30	65,50	4940	25,10
Rumania	257215	99,00	19,80	42,60	8696	38,80
Eslovenia	19689	83,10	18,00	44,80	982	18,40
Eslovaquia	52351	89,20	22,60	54,50	2738	18,10
Finlandia	53923	82,70	11,60	45,60	2615	16,60
Suecia	90982	82,20	8,70	55,10	5100	18,30
Reino Unido	595655	70,80	13,70	62,60	32005	22,20

Tabla 16. Base datos modelo Mortalidad - Fuente: Creación propia (2019)

5.1 ESTUDIO DE MULTICOLINEALIDAD

Mediante el planteamiento de la hipótesis se identifica si las variables están relacionadas entre sí. Si esto pasa presenta un problema de multicolinealidad, la cual hay que solucionar. Esto puede provocar malas estimaciones de los parámetros o mala estimación de la desviación típica de la perturbación (Núñez 2017).

- **Matriz de correlación**

Mediante esta matriz se analiza si existe una relación entre parejas de variables.

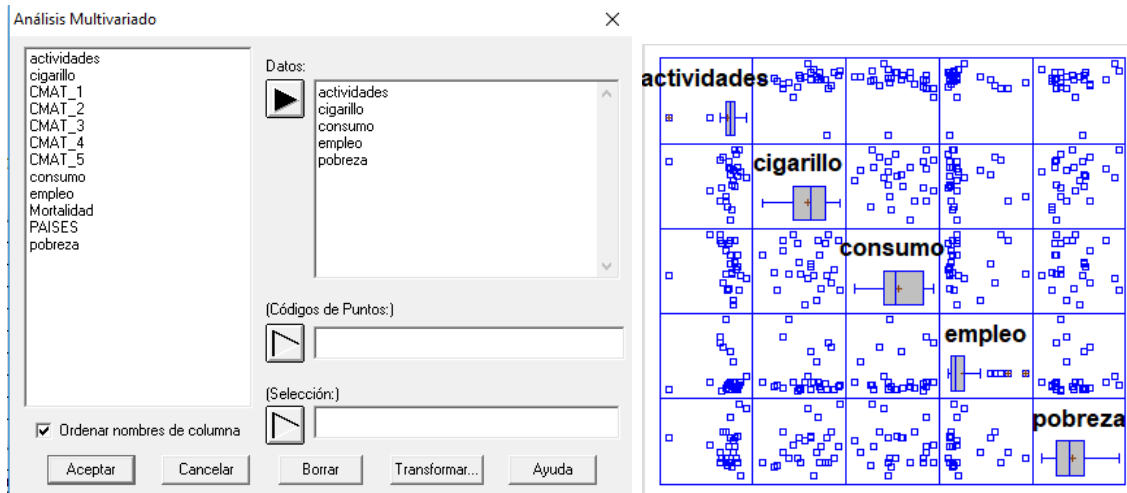


Figura 20. Análisis Multivariado - Fuente: Elaboración propia

Se realiza un Análisis Multivariado, en el cual se seleccionan las variables explicativas del modelo y se obtiene la matriz de correlación.

MATRIZ DE CORRELACIÓN					
	Actividades	Cigarrillo	Consumo	Empleo	Pobreza
Actividades	1	0,0047833	-0,169778	-0,0627676	0,0311607
Cigarrillo	0,0047833	1	-0,0705767	-0,0655636	0,310363
Consumo	-0,169778	-0,0705767	1	0,0598484	-0,126178
Empleo	-0,0627676	-0,0655636	0,0598484	1	-0,0682485
Pobreza	0,0311607	0,310363	-0,126178	-0,0682485	1

Después de realizar la matriz de correlación se observa que no existe relación entre ninguna de las variables ya que $[R_{ii}] \geq 0.7$, por lo que ninguna de las variables presenta problema de correlación con otra de la base de datos.

- **Matriz de correlación inversa**

Estudia la relación que existe entre las variables explicativas "X" y el resto de las variables explicativas.

MATRIZ INVERSA DE CORRELACIÓN					
Actividades	1,032862387	0,014242365	0,171673	0,054724737	-0,011208781
Cigarrillo	0,014242365	1,110282301	0,035316738	0,048556635	-0,337264235
Consumo	0,171673	0,035316738	1,048421255	-0,041935397	0,11311521
Empleo	0,054724737	0,048556635	-0,041935397	1,012337555	0,047023751
Pobreza	-0,011208781	-0,337264235	0,11311521	0,047023751	1,122505565

La matriz inversa de correlación indica que no existe ningún número en su diagonal principal mayor a 10, [$R_{ii}^{-1} \geq 10$]. No existe problema de **Multicolinealidad**, del tipo una variable con el resto.

- **Índice de acondicionamiento**

Mide la relación entre todas las variables a la vez.

Análisis de Componentes Principales			
Número	Eigenvalor	Varianza	Acumulado
1	1,42195	28,439	28,439
2	1,11821	22,364	50,803
3	0,958901	19,178	69,981
4	0,816819	16,336	86,318
5	0,684114	13,682	100

IC ≤ 10

$$IC = \sqrt{\frac{1,42195}{0,684114}}$$

$$IC = 1,441710043$$

Tabla 17. Análisis de componente principales - Fuente: Elaboración propia

El Índice de acondicionamiento es de **1,441710043 < 10** por lo que el modelo no tiene problema de multicolinealidad, no existe relación entre todas las variables a la vez.

Después de comprobar la prueba de multicolinealidad a través de los tres pasos se concluye que en el modelo planteado no existe **problema de multicolinealidad**.

5.2 Diagnóstico y Validación del Modelo LINEAL.

5.2.1 Análisis de la significatividad:

Al no eliminar o reagrupar las variables explicativas por multicolinealidad, se puede realizar la validación del modelo a partir de la primera estimación planteada. Se realiza en primer lugar la comprobación de la significatividad del modelo mediante el contraste de hipótesis.

- **Contraste significación global (modelo).**

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_1: \text{Algún de } \beta_i \neq 0$$

$$\alpha = 0,05$$

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,4649E12	5	2,92979E11	337,74	0,0000
Residuo	1,73494E10	20	8,6747E8		
Total (Corr.)	1,48224E12	25			

Tabla 18. Análisis de significatividad - Fuente: Elaboración propia

P. Valor = 0,000 < α ; Se rechaza H_0 , se concluye que este modelo es significativa.

- **Contraste significación individual (parámetros)**

En este apartado se procede con la valoración individual de la significación de los parámetros del modelo por contraste de hipótesis.

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	-10382,3	55250,7	-0,187912	0,8528
actividades	234,007	396,836	0,589682	0,5620
cigarrillo	695,181	1235,37	0,562731	0,5799
consumo	-1236,13	564,681	-2,18907	0,0406
Empleo	21,046	0,512807	41,0407	0,0000
Pobreza	1824,05	895,412	2,03711	0,0551

Tabla 19. Contraste significatividad - Fuente: Elaboración propia

➤ **Contraste de Parámetro β_0 (CONSTANTE)**

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

P. Valor = 0,8528 > α ; se **Acepta H_0** , se concluye que la constante NO es significativa.

➤ **Contraste de Parámetro β_1 (ACTIVIDADES FÍSICAS)**

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

P. Valor = 0,5620 > α ; se **Acepta H_0** , se concluye que la variable actividades físicas NO es significativa.

➤ **Contraste de Parámetro β_2 (Consumo de Cigarrillo)**

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

P. Valor = 0,5799 > α ; se **Acepta H_0** , se concluye que la variable Consumo de cigarrillo NO es significativa.

➤ **Contraste de Parámetro β_3 (Consumo frutas y verduras)**

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

P. Valor = 0,0406 < α ; **Se Rechaza H_0** , se concluye la variable consumo de frutas y verduras es significativo.

➤ **Contraste de Parámetro β_4 (empleo)**

$$\left. \begin{array}{l} H_0: \beta_4 = 0 \\ H_1: \beta_4 \neq 0 \end{array} \right\} \alpha = 0,05$$

P. Valor = 0,000 < α ; **Rechazamos H_0** , se concluye que la variable empleo es significativo.

➤ **Contraste de Parámetro β_5 (pobreza)**

$$\left. \begin{array}{l} H_0: \beta_5 = 0 \\ H_1: \beta_5 \neq 0 \end{array} \right\} \alpha = 0,05$$

P. Valor = 0,0551 > α ; **Se acepta H_0** , se concluye que la variable pobreza NO es significativa.

5.3 Análisis de Residuos

En este apartado se procede a la validación del modelo con el comportamiento de los residuos. En primer lugar, se buscarán problemas de falta de linealidad mediante el análisis Figura de los residuos frente a cada variable.

- **Residuos vs X**

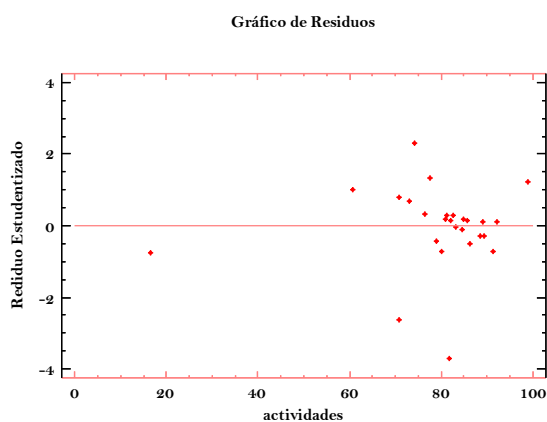


Figura 21. Residuos Actividades

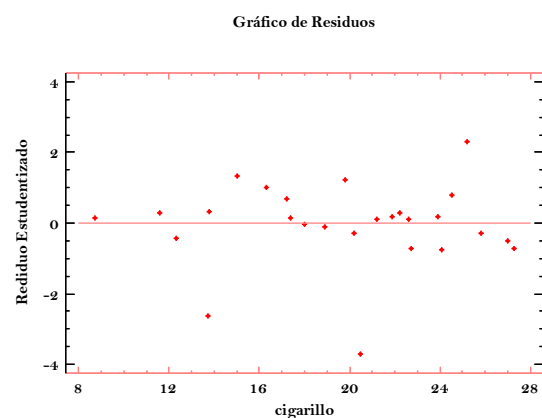


Figura 22. Residuos Cigarillo

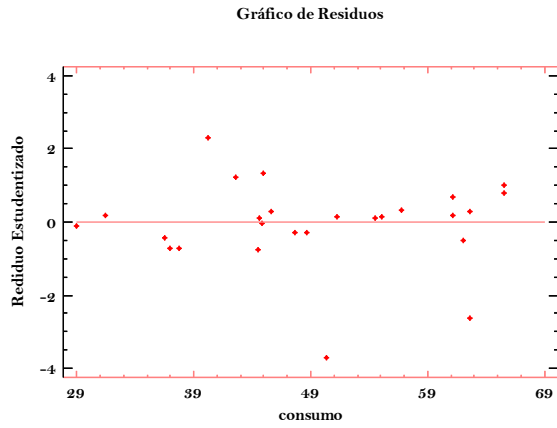


Figura 23. Residuos Consumo



Figura 24. Residuos Empleo

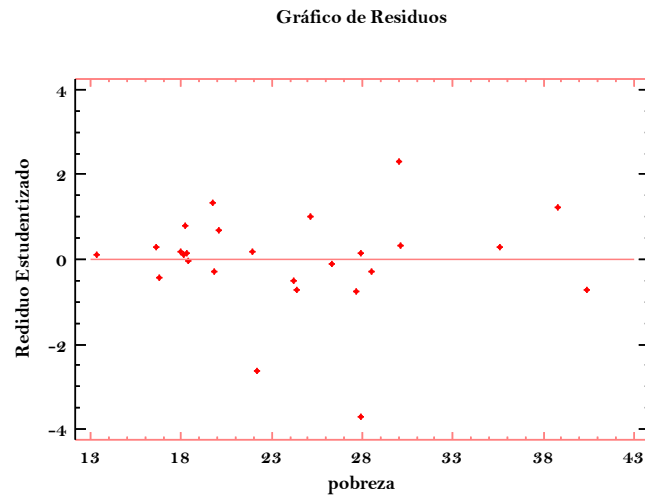


Figura 25. Residuos Pobreza

En el modelo se observa que ninguna de las variables explicativas presenta problema de falta de linealidad. Tampoco se observan forma cónica, sinusoidal o parábola, ya que la dispersión de los residuos es aleatoria.

- **Residuos Vs Y predicho**

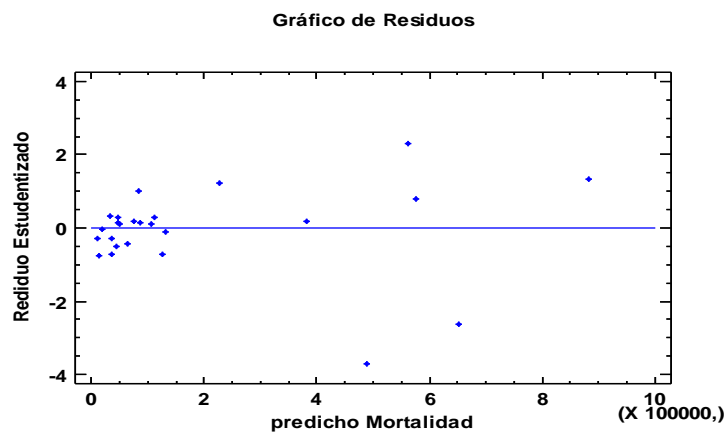


Figura 26. Residuos vs Y predicho -Fuente: Elaboración propia

En la Figura 26 se observa una forma cónica al parecer puede presentar un problema de heterocedasticidad que se corroborará más adelante con el contraste de hipótesis correspondiente.

- **Residuos Vs nº fila**

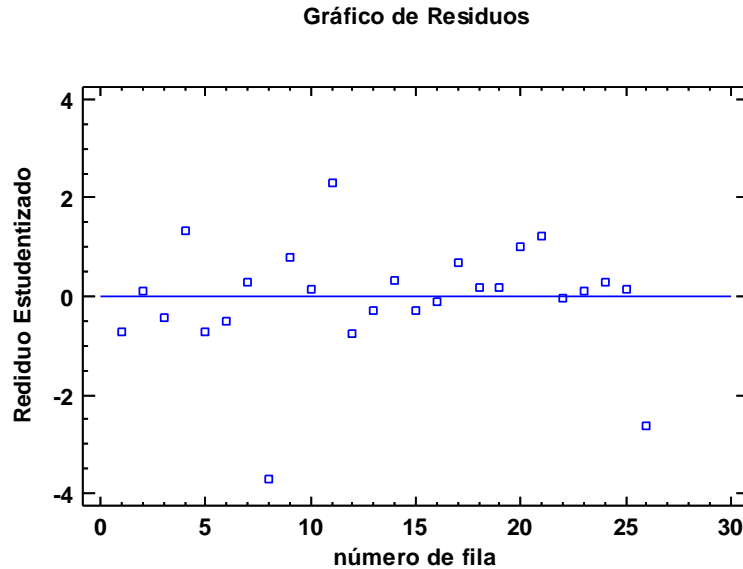


Figura 27. Residuos vs nºfila - Fuente: Elaboración propia

En la Figura 27 presenta una distribución uniforme y aleatoria y no parece que haya autocorrelación solo se observa algunos puntos Anómalos.

5.4 Contraste de Normalidad

Seguidamente se procede a analizar la distribución de los residuos, para comprobar si siguen una distribución normal.

- **Gráfico probabilístico**

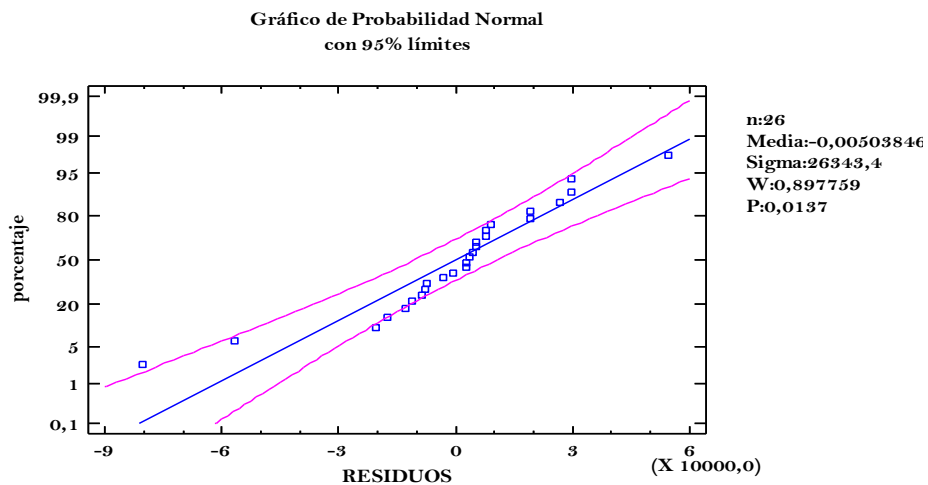


Figura 28. Probabilidad Normal - Fuente: Statgraphics

Parece que los residuos siguen una distribución normal, ya que la mayoría de los residuos se encuentran sobre la línea de normalidad o muy cerca de ella.

- Histograma de residuos

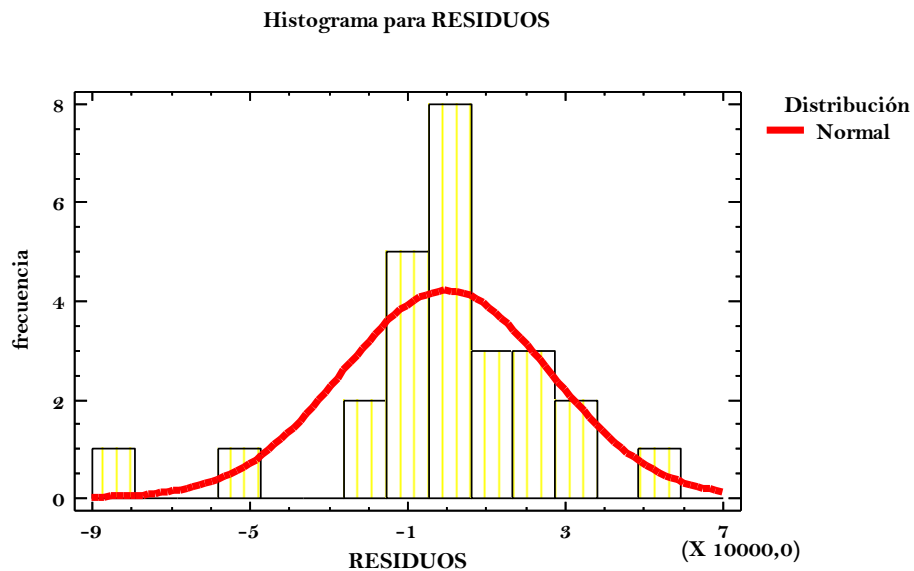


Figura 29. Histograma de Residuos . Fuente: Elaboración propia

Aparentemente la Figura 29 muestra que la mayoría de los residuos siguen una distribución normal, ya que las frecuencias toman la forma de la campana de Gaus.

- **Test de Normalidad. Contraste de hipótesis**

H0: Los residuos si Distribución Normal

H1: Los residuos no una Distribución Normal

$\alpha = 0,05$

Pruebas de Normalidad para RESIDUOS

Prueba	Estadístico	Valor-P
Chi-Cuadrado	14,9231	0,186046
Estadístico W de Shapiro-Wilk	0,897759	0,0137432
Valor-Z para asimetría	1,64899	0,0991489
Valor-Z para curtosis	2,44283	0,0145726

Tabla 20. Prueba de Normalidad -Fuente: Elaboración propia

Mediante el análisis de normalidad se selecciona la prueba de Estadístico W de Shapiro-Wilk ya que tiene el P.valor más restrictivo (el que tiene el P.Valor más pequeño)

P. Valor = 0,0137432 < α ; Se rechaza Ho, se concluye que los residuos no siguen una Distribución Normal.

5.5 Estudio de la heterocedasticidad:

En este apartado se analiza si la varianza de perturbación es constante, el motivo es que una de las hipótesis del modelo es que esta no dependa de los residuos. Para comprobar esta hipótesis se estima el modelo resultante de los residuos al cuadrado resultantes del modelo analizado en función de las variables explicativas del modelo.

$$\text{Residuos } 2 = \beta_0 + \beta_1 * \text{ACTIV. FISICAS} + \beta_2 * \text{CIGARRILLO} + \beta_3 * \text{CONSUMO} + \beta_4 * \text{EMPLEO} + \beta_5 * \text{POBREZA} + U$$

- **Hipótesis**

$$\left. \begin{array}{l} \text{Ho: } \sigma^2 = \text{CONSTANTE} \\ \text{H1: } \sigma^2 \neq \text{CONSTANTE} \end{array} \right\} \alpha = 0,05$$

$$\text{Residuos } 2 = -1,27891\text{E}9 - 1,2201\text{E}6 * \text{actividades} - 8,14843\text{E}6 * \text{cigarrillo} + 7,95445\text{E}6 * \text{consumo} + 71779,8 * \text{empleo} + 4,91755\text{E}7 * \text{pobreza}$$

Parámetro	Estimación	Error	Estadístico	Valor-P
		Estándar	T	
actividades	-1,2201E6	1,7094E7	-0,0713756	0,9438
cigarrillo	-8,14843E6	5,32146E7	-0,153124	0,8798
consumo	7,95445E6	2,43241E7	0,327019	0,7471
empleo	71779,8	22089,6	3,24948	0,0040
pobreza	4,91755E7	3,85706E7	1,27495	0,2169

Tabla 21. Contraste Heterocedasticidad - Fuente: Elaboración propia

➤ ACTIVIDADES FISICAS

P. Valor = 0,9438 > α ; **Se acepta Ho**, se concluye que esta variable no genera problema de Heterocedasticidad

➤ CONSUMO DE CIGARRILLO

P. Valor = 0,8798 > α ; **Se acepta Ho**, se concluye que esta variable no genera problema de Heterocedasticidad

➤ CONSUMO DE FRUTAS Y VERDURAS

P. Valor = 0,7471 > α ; **Se acepta Ho**, se concluye que esta variable no genera problema de Heterocedasticidad

➤ EMPLEO

P. Valor = 0,0040 > α ; **Se rechaza Ho**, se concluye que esta variable genera problema de Heterocedasticidad

➤ **POBREZA**

P. Valor = 0,2169 > α; **Se acepta Ho**, se concluye que esta variable no genera problema de Heterocedasticidad

• **Solución de Heterocedasticidad**

Tras comprobar la hipótesis e indicar que presenta heterocedasticidad se plantea la solución. Para solucionar se divide el modelo inicial entre la variable que genera problema de heterocedasticidad (empleo) elevada a C.

$$\frac{\text{Mortalidad}}{\text{EMPLEO}^c} = \frac{\beta_0}{\text{EMPLEO}^c} + \frac{\beta_1 \cdot \text{ACTIV.FISICAS}}{\text{EMPLEO}^c} + \frac{\beta_2 \cdot \text{CIGARRILLO}}{\text{EMPLEO}^c} + \frac{\beta_3 \cdot \text{CONSUMO}}{\text{EMPLEO}^c} + \frac{\beta_4 \cdot \text{EMPLEO}}{\text{EMPLEO}^c} + \frac{\beta_5 \cdot \text{POBREZA}}{\text{EMPLEO}^c} + U$$

Puesto que c no puede calcularse directamente se calcula a partir de h. $c = \frac{h}{2}$

Para calcular h.- ajustar los **residuos**² frente a la variable explicativa que genera el problema de heterocedasticidad (empleo) elevado a h. Este puede tomar valores de positivos, negativos y números fraccionarios.

Para elegir el valor correcto se prueba valores de “h” y se elige el que tiene un **R²** más alto.

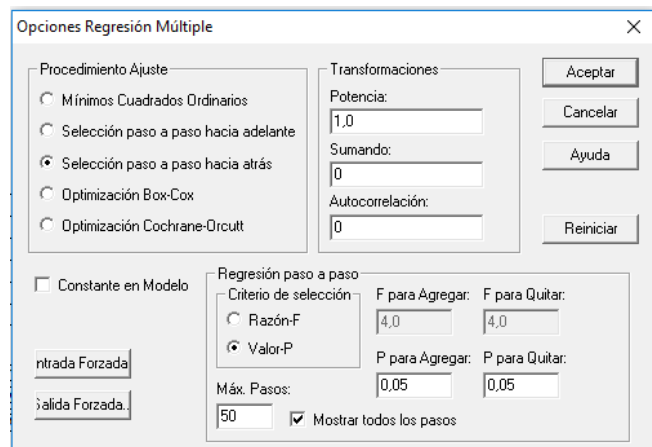
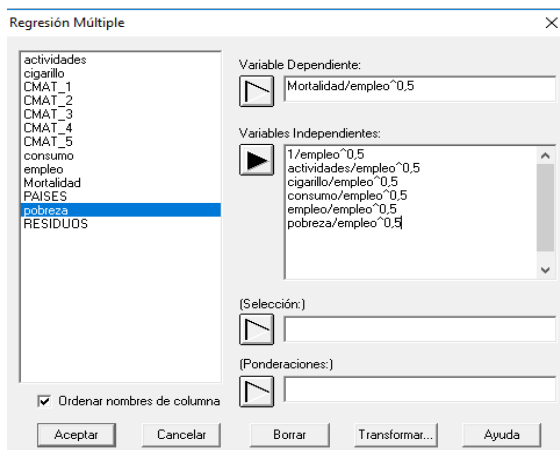
$$\text{Residuos}^2 = \beta_0 + \beta_1 * \text{Empleo}^c + U$$

$$C = \frac{h}{2} = \frac{1}{2} = \mathbf{0,5}$$

H	R ²
-2	1,43
-1	4,52
1	32,20
2	20,67

• **Modelo Solución Heterocedasticidad**

$$\frac{\text{Mortalidad}}{\text{EMPLEO}^{0,5}} = \frac{\beta_0}{\text{EMPLEO}^{0,5}} + \frac{\beta_1 \cdot \text{ACTIV.FISICAS}}{\text{EMPLEO}^{0,5}} + \frac{\beta_2 \cdot \text{CIGARRILLO}}{\text{EMPLEO}^{0,5}} + \frac{\beta_3 \cdot \text{CONSUMO}}{\text{EMPLEO}^{0,5}} + \frac{\beta_4 \cdot \text{EMPLEO}}{\text{EMPLEO}^{0,5}} + \frac{\beta_5 \cdot \text{POBREZA}}{\text{EMPLEO}^{0,5}} + U$$



Al plantear el nuevo modelo para resolver la heterocedasticidad, se observa que no lo soluciona, por lo que se aplica un paso hacia atrás y se procede a quitar la constante.

Comprobar heterocedasticidad en la solución planteada.

Se guarda los nuevos residuos, del modelo válido y se comprueba si el modelo presenta y se ha solucionado el problema de heterocedasticidad.

- **Contraste de Hipótesis**

$$\left. \begin{array}{l} H_0: \sigma^2 = \text{CONSTANTE} \\ H_1: \sigma^2 \neq \text{CONSTANTE} \end{array} \right\} \alpha = 0,05$$

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	-12357,7	71321,3	-0,173268	0,8641
1/empleo ^{0,5}	-3,35413E6	4,1435E6	-0,809492	0,4273
actividades/empleo ^{0,5}	16690,1	29818,2	0,559729	0,5816
empleo/empleo ^{0,5}	773,615	480,661	1,60948	0,1224
pobreza/empleo ^{0,5}	106603,	139184,	0,765913	0,4523

Tabla 22. Contraste hipótesis Heterocedasticidad - Fuente: Elaboración propia

P.valor > 0,05 por lo que **se acepta H₀**, se concluye que las variables no presentan problema de Heterocedasticidad.

Se valida el modelo final, y se comprueba si las variables son significativas y se utiliza para hacer predicciones.

- **Contraste de hipótesis.**

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

Fuente	Suma de Cuadrados	GI	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,06503E8	4	2,66257E7	399,54	0,0000
Residuo	1,46609E6	22	66640,4		
Total	1,07969E8	26			

Tabla 23. Contraste de Hipótesis - Fuente Elaboración propia

Se observa P. Valor del modelo 0,000 < α; **se rechaza H₀**, Se concluye que el modelo es significativo

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
1/empleo ^{0,5}	-52269,1	12821,0	-4,07683	0,0005
actividades/empleo ^{0,5}	205,318	87,5171	2,34603	0,0284
empleo/empleo ^{0,5}	21,121	0,603767	34,982	0,0000
pobreza/empleo ^{0,5}	1665,19	402,228	4,13991	0,0004

Tabla 24. Contraste de Hipótesis de parámetros - Fuente Elaboración propia

R-cuadrada = 98,6421 por ciento
R-cuadrado (ajustado para g.l.) = 98,457 por ciento

Mediante la Tabla 23. Contraste de Hipótesis de parámetros indica que las variables presentan P. Valor $0.000 < \alpha$; **se rechaza H_0** , Se concluye que las variables son significativas.

Finalmente el modelo que se plantea para poder explicar la variable mortalidad es válido y fiable ya que presenta un **R^2 de un 98.46%**.

- **Modelo válido tras la solución de la heterocedasticidad:**

$$\frac{\text{Mortalidad}}{\text{EMPLEO}^{0.5}} = \frac{\beta_0}{\text{EMPLEO}^{0.5}} + \frac{\beta_1 * \text{ACTIV.FISICAS}}{\text{EMPLEO}^{0.5}} + \frac{\beta_2 * \text{EMPLEO}}{\text{EMPLEO}^{0.5}} + \frac{\beta_3 * \text{POBREZA}}{\text{EMPLEO}^{0.5}} + U$$

$$\frac{\text{Mortalidad}}{\text{EMPLEO}^{0.5}} = -52269,1 * \frac{1}{\text{EMPLEO}^{0.5}} + \frac{205,318 * \text{ACTIV.FISICAS}}{\text{EMPLEO}^{0.5}} + \frac{21,121 * \text{EMPLEO}}{\text{EMPLEO}^{0.5}} + \frac{1665,19 * \text{POBREZA}}{\text{EMPLEO}^{0.5}}$$

5.6 Estudio de Autocorrelación

Para que el modelo sea válido no puede tener autocorrelación, esto quiere decir que no dependerá de datos anteriores. Puede haber correlación de 1er y 2do orden.

- **Presentación gráfica del FAS Y FAP**

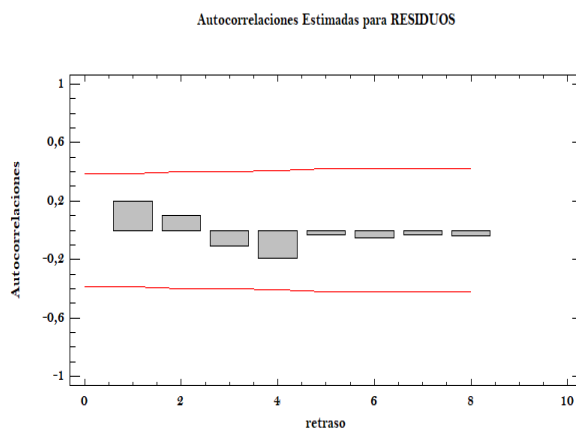


Figura 31. FAS - Fuente: Elaboración propia

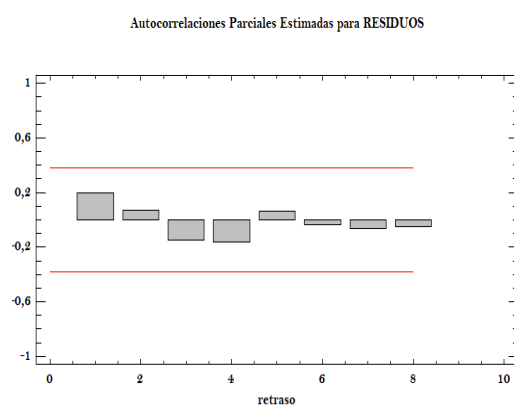


Figura 30 FAP - Fuente: Elaboración propia

Las Figura 31 y Figura 30 indican que el Factor de Autocorrelación de primer orden NO supera los límites de autocorrelación por la parte superior, y el resto de los factores de autocorrelación no superan los límites de autocorrelación.

- **Puntos anómalos.**

		Y		Residuo
Fila	Y	Predicha	Residuo	Estudentizado
6	232450,	38394,0	194056,	2,78
8	291523,	110624,	180900,	2,62

Tabla 25. Residuos atípicos - Fuente Elaboración propia

R-cuadrada = 98,6421 por ciento

La *Tabla 25*. residuos atípicos nos indica que hay residuos mayores a 2 en valor absoluto nos indica un punto anómalo en la fila 6 y otra en la fila 8 al no ser mayor de 3 no son candidatos para ser eliminados.

Pruebas de Normalidad para RESIDUOS

Prueba	Estadístico	Valor-P
Chi-Cuadrado	9,53846	0,572317
Estadístico W de Shapiro-Wilk	0,945775	0,197877
Valor-Z para asimetría	0,71384	0,475324
Valor-Z para curtosis	0,845076	0,398066

Tabla 26. Pruebas de Normalidad

Se selecciona la prueba del Estadístico W de Shapiro-Wilk ya que tiene el P. Valor más restrictivo (el que tiene el P - valor más pequeño)

P. Valor = 0,197877 > α; Se acepta Ho y se concluye que los residuos siguen una Distribución Normal

El modelo ya no presenta ningún problema y se descarta eliminar algún punto anómalo, por lo que se puede hacer predicciones ya que se obtiene un $R^2 = 98,6421\%$.

5.7 Predicciones Modelo Mortalidad

Después de la validación del modelo se propone una predicción actual sobre la mortalidad ya que puede ser viable por tener un R^2 muy bueno de un 98,64%.

$$\frac{\text{Mortalidad}}{\text{EMPLEO}^{0,5}} = \frac{\beta_0}{\text{EMPLEO}^{0,5}} + \frac{\beta_1 * \text{ACTIV.FISICAS}}{\text{EMPLEO}^{0,5}} + \frac{\beta_2 * \text{EMPLEO}}{\text{EMPLEO}^{0,5}} + \frac{\beta_3 * \text{POBREZA}}{\text{EMPLEO}^{0,5}} + U$$

$$\frac{\text{Mortalidad}}{\text{EMPLEO}^{0,5}} = -52269,1 * \frac{1}{\text{EMPLEO}^{0,5}} + \frac{205,318 * \text{ACTIV.FISICAS}}{\text{EMPLEO}^{0,5}} + \frac{21,121 * \text{EMPLEO}}{\text{EMPLEO}^{0,5}} + \frac{1665,19 * \text{POBREZA}}{\text{EMPLEO}^{0,5}}$$

¿ **Qué porcentaje de Mortalidad tiene España?** Con una aumento de un 11% de la tasa de empleo, Índice de pobreza se reduce a un 25% y las actividades físicas se mantienen con un 82%

¿**Si se reduce el Índice de pobreza a un 25%?**

Fila	Ajustado	Error Est. LC para Pronóstico	Inferior 95,0% LC para Pronóstico	Superior 95,0% LC para Pronóstico	Inferior 95,0% LC para la Media
27	3418,97	281,587	2833,38	4004,57	3216,47
28	3387,57	281,342	2802,49	3972,66	3186,55

$$\frac{\text{Mortalidad}}{\text{EMPLEO}^{0,5}} (\text{fila 27}) = 3418,58$$

$$\text{Mortalidad} = 3418,58 * \text{EMPLEO}^{0,5}$$

$$\text{Mortalidad} = 3418,58 * 25149^{0,5}$$

$$\text{Mortalidad} = 542133 \text{ (MILES) DE MUERTES}$$

Tras plantear un modelo para poder explicar la Mortalidad a nivel europeo, las variables seleccionadas Actividad Física, Empleo y Pobreza son finalmente las significativas para las predicciones.

Ya que si se aumenta a 25149 (miles) de personas empleadas, que corresponde un 11% más del empleo con diferencia del último año (2016). Si el porcentaje de las actividades físicas se mantienen en un 82% y se aumenta el índice de pobreza a un 28% en España nos indica que la Mortalidad aumenta a 542133(miles) de Muertes en 2017 con un 95% de fiabilidad.

$$\frac{\mathbf{Mortalidad}}{\mathbf{EMPLEO}^{0,5}} (\mathbf{fila\ 28}) = \mathbf{3387,57}$$

$$\mathbf{Mortalidad} = 3418,58 * \mathbf{EMPLEO}^{0,5}$$

$$\mathbf{Mortalidad} = 3418,58 * 25149^{0,5}$$

$$\mathbf{Mortalidad} = 537215 (\mathbf{MILES}) \mathbf{DE\ MUERTES}$$

Por otro lado si el índice de pobreza disminuye a un 25%, manteniendo el empleo 25149 (miles) de personas empleadas y manteniendo las actividades físicas en un 82% se prevé un descenso de la Mortalidad en 537215 (MILES) con un 95% de fiabilidad.

6. Modelo Teórico Propuesto (Natalidad)

En este apartado se va a proponer el primer modelo de regresión múltiple para explicar la variabilidad de la Natalidad en Europa. El modelo inicial propuesto es el siguiente teniendo en cuenta las variables recopiladas:

$$Natalidad = \beta_0 + \beta_1 * ACTIV.FISICAS + \beta_2 * CIGARRILLO + \beta_3 * CONSUMO + \beta_4 * EMPLEO + \beta_5 * POBREZA + U$$

La primera estimación del modelo lineal propuesto resulta como sigue:

$$Natalidad = -70998,1 - 15,5141 * actividades + 1472,23 * cigarillo + 1499,3 * consumo + 21,3376 * empleo + 1203,39 * pobreza$$

- **Interpretación de los parámetros y unidades físicas**

β_0 = Es el valor medio de la Natalidad cuando las variables X valen cero.

β_1 = Es el incremento medio de Natalidad cuando aumenta en un 1% las **act. Físicas** y el resto de las variables se mantiene constante.

β_2 = Es el incremento medio de Natalidad cuando aumenta en un 1% el **Consumo de cigarillo** y el resto de las variables se mantiene constante.

β_3 = Es el incremento medio de Natalidad cuando aumenta en un 1% el índice del **Consumo de frutas y verduras** y el resto de las variables se mantiene constante.

β_4 = Es el incremento medio de Natalidad cuando aumenta en una unidad el **empleo** y el resto de las variables se mantiene constante.

β_5 = Es el incremento medio de Mortalidad cuando aumenta en 1% de **pobreza** y el resto de las variables se mantiene constante.

U = Perturbación del modelo econométrico, parte del modelo que no hemos tenido en cuenta en el modelo

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,54253E12	5	3,08506E11	118,01	0,0000
Residuo	5,22853E10	20	2,61426E9		
Total (Corr.)	1,59481E12	25			

Tabla 27. Análisis Varianza - Fuente: Elaboración propia

R-cuadrada = 96,7215 por ciento

R-cuadrado (ajustado para g.l.) = 95,9019 por ciento

Error estándar del est. = 51129,9

Error absoluto medio = 30914,7

Estadístico Durbin-Watson = 2,49859 (P=0,8929)

Autocorrelación de residuos en retraso 1 = -0,309572

TABLA DE DATOS

PAISES	Natalidad (miles)	Empleo (%)	Pobreza (%)	Consumo (%)	Actividades (%)	Cigarrillo (%)
Bulgaria	64984	3200	40,4	37,00	91,4	27,3
Chequia	112663	5226	13,3	44,60	92,2	21,2
Dinamarca	61614	2934	16,8	36,50	79,1	12,3
Alemania	792141	41932	19,7	44,90	77,7	15,0
Estonia	14053	658	24,4	37,80	80,1	22,7
Irlanda	63836	2260	24,2	62,10	86,2	27,0
Grecia	92898	4732	35,6	62,60	81,2	22,2
España	408734	22657	27,9	50,40	81,7	20,5
Francia	784325	28471	18,2	65,50	70,9	24,5
Croacia	37537	1806	27,9	51,30	85,7	17,4
Italia	473438	25243	30	40,20	74,3	25,2
Cripre	9455	408	27,7	44,50	16,5	24,1
Letonia	21968	957	28,5	48,70	89,3	20,2
Lituania	30623	1433	30,1	56,80	76,5	13,8
Luxemburgo	6050	277	19,8	47,60	88,6	25,8
Hungría	95361	4543	26,3	29,00	84,7	18,9
Malta	4476	215	20,1	61,10	73,0	17,2
Austria	87675	4412	18	61,10	81,0	23,9
Polonia	382257	16961	21,9	31,40	84,9	21,9
Portugal	87126	4940	25,1	65,50	60,7	16,3
Rumania	200009	8696	38,8	42,60	99,0	19,8
Eslovenia	20345	982	18,4	44,80	83,1	18,0
Eslovaquia	57557	2738	18,1	54,50	89,2	22,6
Finlandia	52814	2615	16,6	45,60	82,7	11,6
Suecia	117425	5100	18,3	55,10	82,2	8,7
Reino Unido	774386	32005	22,2	62,60	70,8	13,7

Tabla 28. datos modelo Natalidad - Fuente: Elaboración Propia

6.1 ESTUDIO DE MULTICOLINEALIDAD

Mediante el planteamiento de la hipótesis se identifica si las variables están relacionadas entre sí. Si esto pasa presenta un problema de multicolinealidad la cual hay que solucionar. Esto puede provocar malas estimaciones de los parámetros o mala estimación de la desviación típica de la perturbación (Núñez 2017).

- **Matriz de correlación**

Mediante esta matriz se analiza si existe una relación entre parejas de variables.

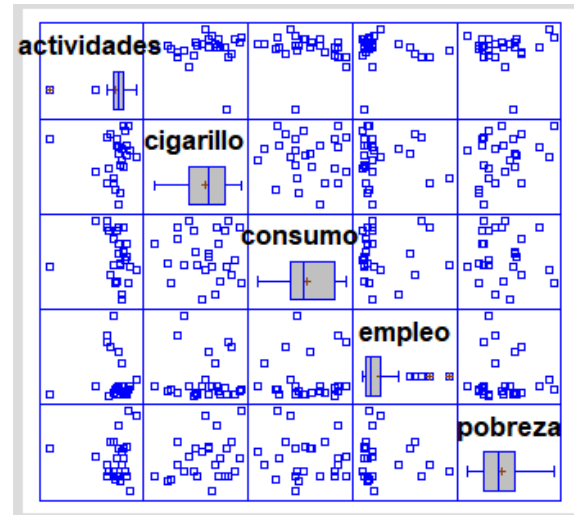
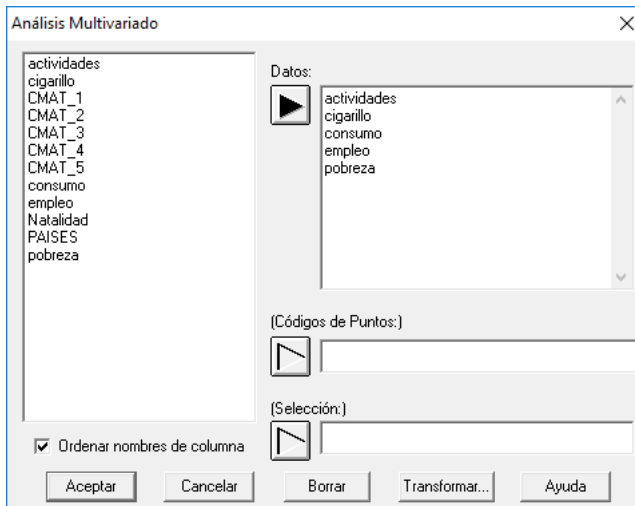


Figura 32. Análisis Multivariado - Fuente: Elaboración propia

MATRIZ DE CORRELACIÓN					
	Actividades	Cigarillo	Consumo	Empleo	Pobreza
Actividades	1	0,0047833	-0,169778	-0,0627676	0,0311607
Cigarillo	0,0047833	1	-0,0705767	-0,0655636	0,310363
Consumo	-0,169778	-0,0705767	1	0,0598484	-0,126178
Empleo	-0,0627676	-0,0655636	0,0598484	1	-0,0682485
Pobreza	0,0311607	0,310363	-0,126178	-0,0682485	1

Después de realizar la matriz de correlación se observa que no existe relación entre ninguna de las variables ya que $[R_{ii}] \geq 0.7$, por lo que ninguna de las variables presenta problema de correlación con otra de la base de datos.

- **Matriz de correlación inversa**

La matriz estudia la relación que existe entre las variables explicativas "X" y el resto de las variables explicativas

MATRIZ INVERSA DE CORRELACIÓN					
Actividades	1,03286239	0,014242365	0,171673	0,054724737	-0,01120878
Cigarillo	0,01424236	1,110282301	0,035316738	0,048556635	-0,33726423
Consumo	0,171673	0,035316738	1,048421255	-0,0419354	0,11311521
Empleo	0,05472474	0,048556635	-0,0419354	1,012337555	0,047023751
Pobreza	0,01120878	-0,33726423	0,11311521	0,047023751	1,122505565

La matriz inversa de correlación indica que no existe ningún número en su diagonal principal mayor a 10, $[R_{ii}^{-1} \geq 10]$. No existe problema de Multicolinealidad, del tipo una variable con el resto.

Índice de acondicionamiento

Mide la relación entre todas las variables a la vez.

Análisis de Componentes Principales			
Número	Eigenvalor	Varianza	Acumulado
1	1,42195	28,439	28,439
2	1,11821	22,364	50,803
3	0,958901	19,178	69,981
4	0,816819	16,336	86,318
5	0,684114	13,682	100

$$IC \leq 10$$

$$IC = \sqrt{\frac{1,42195}{0,684114}}$$

$$IC = 1,441710043$$

Tabla 29. Componentes principales - Fuente: Elaboración propia

El Índice de acondicionamiento es de **1,441710043** < 10 por lo que no tiene problema de multicolinealidad, no existe relación entre todas las variables a la vez.

Después de comprobar la prueba de multicolinealidad a través de los tres pasos se indica que en el modelo planteado no existe **problema de multicolinealidad**.

6.2 Diagnóstico y Validación del Modelo LINEAL.

6.2.1 Análisis de la significatividad:

Al no eliminar o reagrupar las variables explicativas por multicolinealidad, se puede realizar la validación del modelo a partir de la primera estimación planteada. Se realiza en primer lugar la comprobación de la significatividad del modelo mediante el contraste de hipótesis.

- **Contraste significación global (modelo).**

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_1: \text{Algún de } \beta_i \neq 0$$

$$\alpha = 0,05$$

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,54253E12	5	3,08506E11	118,01	0,0000
Residuo	5,22853E10	20	2,61426E9		
Total (Corr.)	1,59481E12	25			

Tabla 30. Significatividad Modelo - Fuente: Elaboración propia

P. Valor = 0,000 < α ; Se rechaza H_0 , se concluye que este modelo es significativa

• **Contraste significación individual (parámetros)**

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	-70998,1	95914,6	-0,740222	0,4678
actividades	-15,5141	688,903	-0,02252	0,9823
cigarrillo	1472,23	2144,59	0,686487	0,5003
consumo	1499,3	980,281	1,52946	0,1418
empleo	21,3376	0,890229	23,9686	0,0000
pobreza	-1203,39	1554,43	-0,774172	0,4479

Tabla 31. Significatividad Modelo - Fuente: Elaboración propia

Contraste de Parámetro β_0 (CONSTANTE)

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

P. Valor = 0,4678 > α ; **Se acepta H_0** , se concluye que la constante NO es significativa.

➤ **Contraste de Parámetro β_1 (ACTIVIDADES FÍSICAS)**

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

P. Valor = 0,9823 > α ; **se Acepta H_0** , se concluye que la variable actividades físicas NO es significativo.

➤ **Contraste de Parámetro β_2 (Consumo de Cigarrillo)**

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

P. Valor = 0,5003 > α ; **Se acepta H_0** , se concluye que la variable Consumo de cigarrillo NO es significativo.

➤ **Contraste de Parámetro β_3 (Consumo frutas y verduras)**

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

P. Valor = 0,1418 < α ; **Se rechaza H_0** , se concluye que la variable consumo de frutas y verduras es significativo.

➤ **Contraste de Parámetro β_4 (empleo)**

$$\left. \begin{array}{l} H_0: \beta_4 = 0 \\ H_1: \beta_4 \neq 0 \end{array} \right\} \alpha = 0,05$$

P. Valor = 0,000 < α ; **Se rechaza H_0** , se concluye que la variable empleo es significativo.

➤ **Contraste de Parámetro β_5 (pobreza)**

$$\left. \begin{array}{l} H_0: \beta_5 = 0 \\ H_1: \beta_5 \neq 0 \end{array} \right\} \alpha = 0,05$$

P. Valor = 0,4479 > α ; **Se acepta H_0** , se concluye que la variable pobreza NO es significativo.

6.3 Análisis de Residuos

En este apartado se procede a la validación del modelo con el comportamiento de los residuos. En primer lugar, se buscarán problemas de falta de linealidad mediante el análisis Figura de los residuos frente a cada variable

- **Residuos vs x**

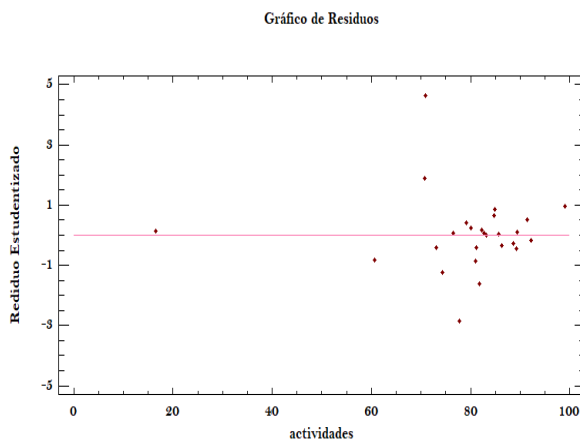


Figura 34. Residuos Actividades

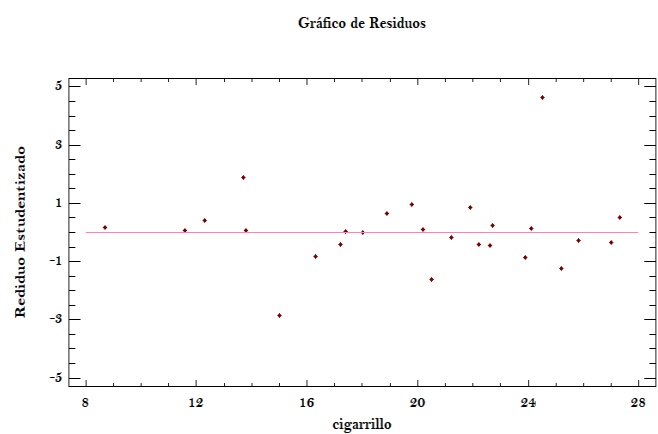


Figura 33. Residuos Cigarrillo - Fuente Elaboración propia

Gráfico de Residuos

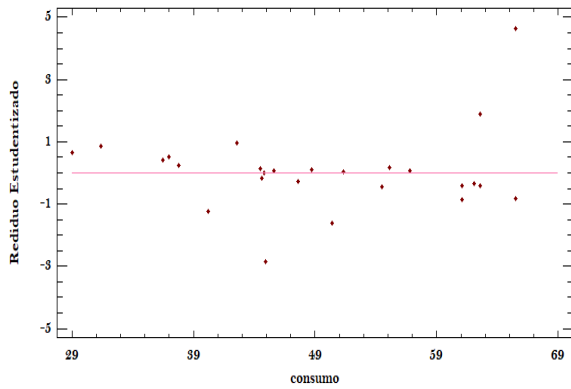


Figura 36. Residuos Consumo

Gráfico de Residuos

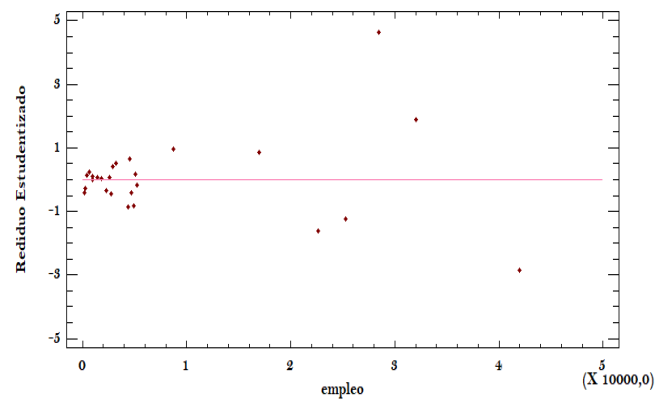


Figura 35. Residuos Empleo - Fuente Elaboración propia

Gráfico de Residuos

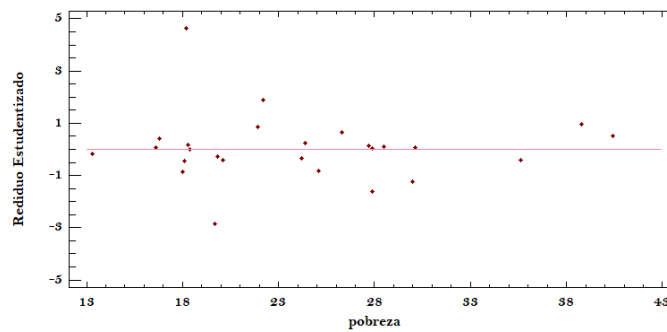


Figura 37. Residuos - Fuente Elaboración propia

En el modelo se observa que ninguna de las variables explicativas presenta problema de falta de linealidad. Tampoco se observan la forma cónica, sinusoidal o parábola, ya que la dispersión de los residuos es aleatoria.

- **Residuos Vs Y predicho**

Gráfico de Residuos

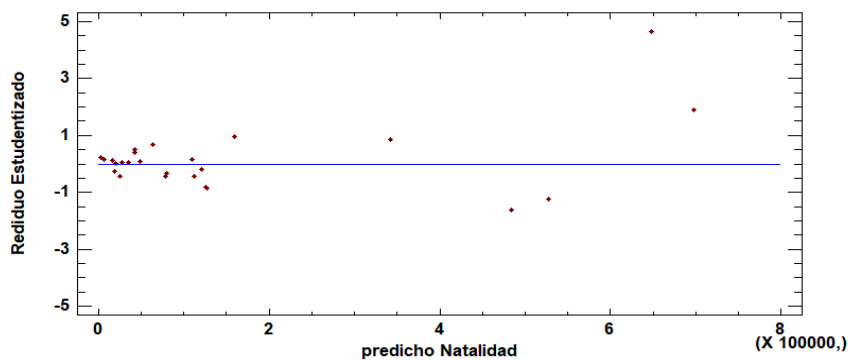


Figura 38. . Residuos vs Y predicho -Fuente: Elaboración propia

En la Figura 38 se observa una forma cónica al parecer puede presentar un problema de heterocedasticidad que se corroborará más adelante con el contraste de hipótesis correspondiente.

- **Residuos Vs nº fila**

Gráfico de Residuos

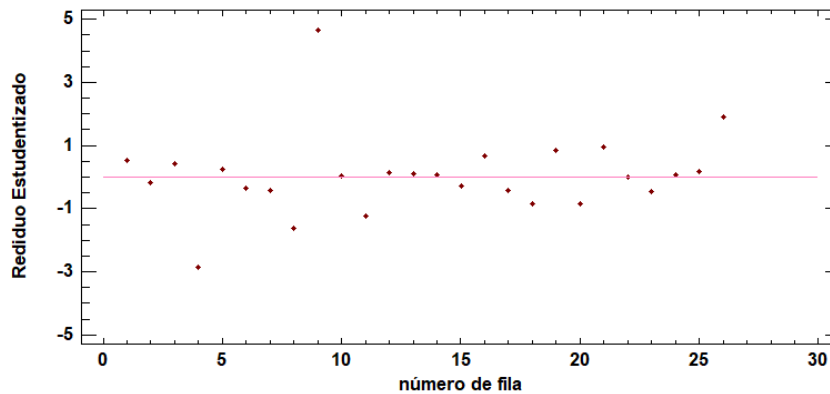


Figura 39. Residuos vs n fila - Fuente: Elaboración propia

La Figura 39 presenta una distribución uniforme y aleatoria y no parece que haya autocorrelación solo se observa algunos puntos Anómalos.

6.4 Contraste de Normalidad

Mediante el contraste de normalidad se procede analizar la distribución de los residuos, para comprobar si siguen una distribución normal.

- **Figura probabilístico**

Gráfico de Probabilidad Normal
con 95% límites

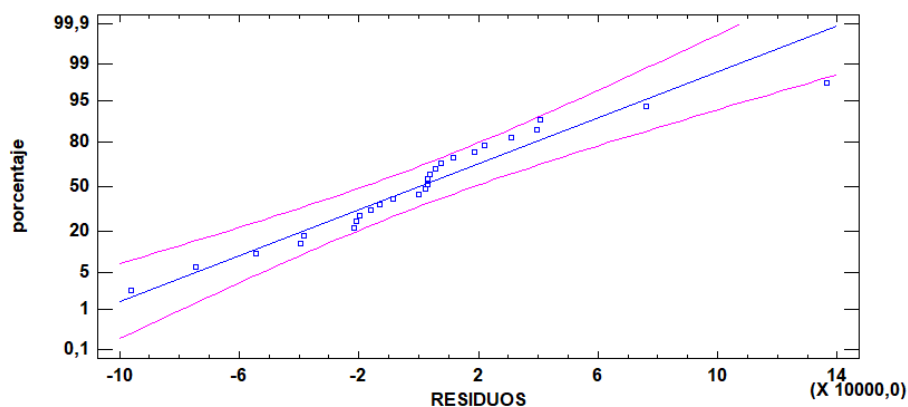


Figura 40. Probabilístico - Fuente: Elaboración propia

La Figura 40 indica que los residuos siguen una distribución normal, ya que la mayoría de los residuos se encuentran sobre la línea de normalidad o muy cerca de ella.

- **Histograma de residuos**

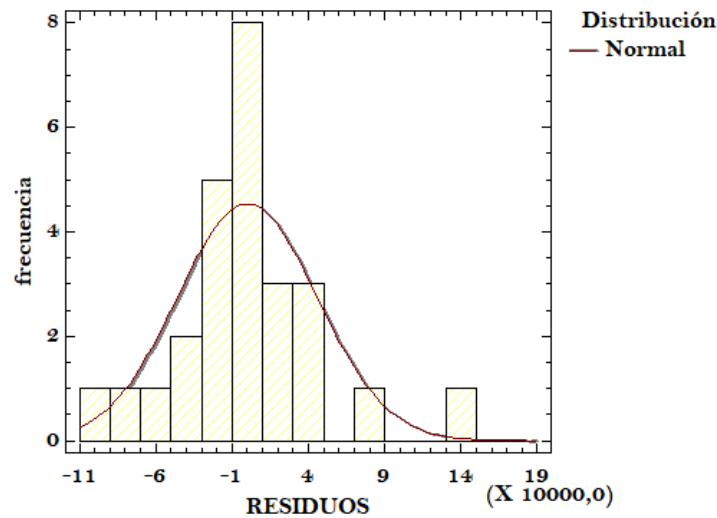


Figura 41. Histograma de residuos - Fuente: Elaboración propia

Aparentemente la Figura 41 muestra que la mayoría de los residuos siguen una distribución normal, ya que las frecuencias toman la forma de la campana de Gaus.

- **Test de Normalidad. Contraste de hipótesis**

H0: Los residuos si Distribución Normal
H1: Los residuos no una Distribución Normal

$\alpha = 0,05$

Prueba	Estadístico	Valor-P
Chi-Cuadrado	22,4615	0,0210318
Estadístico W de Shapiro-Wilk	0,939279	0,138812
Valor-Z para asimetría	1,10797	0,267874
Valor-Z para curtosis	2,16612	0,0303019

Tabla 32. Normalidad de Residuos - Fuente: Elaboración propia

Mediante el análisis de normalidad se selecciona la prueba de Chi-cuadrado ya que tiene el P.valor más restrictivo (el que tiene el P.Valor más pequeño)

P. Valor = 0,0210318 < α ; Se rechaza Ho y se concluye que los residuos no siguen una Distribución Normal.

6.5 Estudio de la heterocedasticidad:

$$Residuos^2 = \beta_0 + \beta_1 * ACTIV.FISICAS + \beta_2 * CIGARRILLO + \beta_3 * CONSUMO + \beta_4 * EMPLEO + \beta_5 * POBREZA + U$$

• Hipótesis

$$\left. \begin{array}{l} H_0: \sigma^2 = \text{CONSTANTE} \\ H_1: \sigma^2 \neq \text{CONSTANTE} \end{array} \right\} \alpha = 0,05$$

$$Residuos^2 = -4,96013E9 - 9,26079E6*actividades + 1,52775E8*cigarrillo + 8,45353E7*consumo + 270770,*empleo - 7,54379E7*pobreza$$

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	-4,96013E9	4,64504E9	-1,06783	0,2983
actividades	-9,26079E6	3,33628E7	-0,277578	0,7842
cigarrillo	1,52775E8	1,0386E8	1,47097	0,1569
consumo	8,45353E7	4,74739E7	1,78067	0,0902
empleo	270770,	43112,8	6,2805	0,0000
pobreza	-7,54379E7	7,52791E7	-1,00211	0,3283

Tabla 33. Hipótesis Elaboración propia

➤ ACTIVIDADES FISICAS

P. Valor = 0,7842 > α ; se **Acepta Ho**, se concluye que esta variable no genera problema de Heterocedasticidad

➤ CONSUMO DE CIGARRILLO

P. Valor = 0,1569 > α ; se **Acepta Ho**, se concluye que esta variable no genera problema de Heterocedasticidad

➤ CONSUMO DE FRUTAS Y VERDURAS

P. Valor = 0,0902 > α ; se **Acepta Ho**, se concluye que esta variable no genera problema de Heterocedasticidad

➤ EMPLEO

P. Valor = 0,000 > α ; se **Rechaza Ho**, se concluye que esta variable genera problema de Heterocedasticidad

➤ POBREZA

P. Valor = 0,3283 > α ; se **Acepta Ho**, se concluye que esta variable no genera problema de Heterocedasticidad

• **Solución de Heterocedasticidad de la variable empleo**

Tras el análisis de cada P.valor se plantea la solución para el problema de heterocedasticidad.

Para solucionar se divide el modelo inicial entre la variable que genera problema de heterocedasticidad (empleo) elevada a C.

$$\frac{Natalidad}{EMPLEO^c} = \frac{\beta_0}{EMPLEO^c} + \frac{\beta_1 * ACTIV.FISICAS}{EMPLEO^c} + \frac{\beta_2 * CIGARRILLO}{EMPLEO^c} + \frac{\beta_3 * CONSUMO}{EMPLEO^c} + \frac{\beta_4 * EMPLEO}{EMPLEO^c} + \frac{\beta_5 * POBREZA}{EMPLEO^c} + U$$

Puesto que c no puede calcularse directamente se calcula a partir de h. $C = \frac{h}{2}$

Para calcular h.- ajustar los **residuos**² frente a la variable explicativa que genera el problema de heterocedasticidad (empleo) elevado a h. Este puede tomar valores de positivos, negativos y números fraccionarios.

Para elegir el valor correcto se prueba valores de “h” y se elige el que tiene un **R²** más alto.

$$Residuos^2 = \beta_0 + \beta_1 * Empleo^c + U$$

$$C = \frac{h}{2} = \frac{1}{2} = 0,5$$

H	R ²
-2	2,53188
-1	6,5084
1	61,1303
2	55,6132

• **Modelo Solución Heterocedasticidad**

$$\frac{Natalidad}{EMPLEO^{0,5}} = \frac{\beta_0}{EMPLEO^{0,5}} + \frac{\beta_1 * ACTIV.FISICAS}{EMPLEO^{0,5}} + \frac{\beta_2 * CIGARRILLO}{EMPLEO^{0,5}} + \frac{\beta_3 * CONSUMO}{EMPLEO^{0,5}} + \frac{\beta_4 * EMPLEO}{EMPLEO^{0,5}} + \frac{\beta_5 * POBREZA}{EMPLEO^{0,5}} + U$$

Parámetro	Estimación	Error		Valor-P
		Estándar	Estadístico T	
1/empleo ^{0,5}	-10658,6	29453,4	-0,361881	0,7212
actividades/empleo ^{0,5}	-11,6966	113,566	-0,102994	0,9190
cigarrillo/empleo ^{0,5}	322,931	615,199	0,524922	0,6054
consumo/empleo ^{0,5}	152,539	310,895	0,490646	0,6290
empleo/empleo ^{0,5}	21,6368	0,791174	27,3478	0,0000
pobreza/empleo ^{0,5}	-132,215	537,581	-0,245945	0,8082

Tabla 34. Heterocedasticidad Variables - Fuente: Elaboración Propia

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,0457E8	6	1,74284E7	155,52	0,0000
Residuo	2,24126E6	20	112063,		
Total	1,06811E8	26			

Tabla 35. Heterocedasticidad Modelo - Fuente: Elaboración Propia

Después de plantear la solución sobre el problema de heterocedasticidad, se concluye que no tiene solución ya que sigue presentando heterocedasticidad por lo que se procede a eliminar puntos anómalos.

- **Puntos anómalos.**

		Y		Residuo
Fila	Y	Predicha	Residuo	Estudentizado
4	792141,	888218,	-96076,8	-2,86
9	784325,	647775,	136550,	4,65

Tabla 36. Residuos Atípicos - Fuente: Elaboración propia

La tabla de residuos atípicos indican que hay residuos mayores a -2 en valor absoluto por lo que corresponde un punto anómalo en la fila 4. Pero la fila 9 correspondiente al País **Francia**, al tener un 4,65 es mayor a 3 por lo que es candidato para eliminar.

		Y		Residuo
Fila	Y	Predicha	Residuo	Estudentizado
4	792141,	852114,	-59973,2	-2,59
25	774386,	660665,	113721,	8,76

Tabla 37. Residuos Atípicos - Fuente: Elaboración propia

Al eliminar a Francia la tabla de residuos atípicos indican que hay residuos mayores a -2 en valor absoluto por lo que corresponde un punto anómalo en la fila 4. Pero la fila 25 correspondiente al País **Reino Unido**, al tener un 8,76 es mayor a 3 por lo que es candidato para eliminar.

		Y		Residuo
Fila	Y	Predicha	Residuo	Estudentizado
18	382257,	335588,	46668,7	4,92

Tabla 38. Residuos Atípicos - Fuente: Elaboración propia

Al eliminar a Reino Unido la tabla de residuos atípicos indican que hay residuos mayores a 3 en valor absoluto, esto muestra un punto anómalo en la fila 18 candidato para ser eliminado que corresponde al país **Polonia**.

		Y		Residuo
Fila	Y	Predicha	Residuo	Estudentizado
8	408734,	428755,	-20021,4	-2,34
11	9455,0	149,061	9305,94	2,62
19	200009,	176248,	23761,4	3,45

Tabla 39. Residuos Atípicos - Fuente: Elaboración propia

Al eliminar Polonia en la tabla de residuos atípicos indican que hay residuos mayores a 2 en valor absoluto muestra un punto anómalo en la fila 8 y la fila 11. Pero la fila 19 correspondiente al País Rumania, al tener un 3,45 es mayor a 3 por lo que es candidato para eliminar.

		Y		Residuo
Fila	Y	Predicha	Residuo	Estudentizado
6	63836,0	47794,1	16041,9	2,62
8	408734,	425777,	-17042,7	-2,69
22	117425,	102640,	14785,0	2,37

Tabla 40. Residuos Atípicos - Fuente: Elaboración propia

Finalmente la tabla de residuos Atípicos no presenta puntos anómalos candidatos a ser eliminados ya que ninguno tiene un valor mayor a 3. Por lo que se determina que no hay puntos anómalos para eliminar.

- **Pruebas de Normalidad para RESIDUOS**

Prueba	Estadístico	Valor-P
Chi-Cuadrado	13,4545	0,199356
Estadístico W de Shapiro-Wilk	0,915113	0,0579063
Valor-Z para asimetría	0,563658	0,572984
Valor-Z para curtosis	1,43568	0,151093

Tabla 41. Prueba de Normalidad - Fuente: Elaboración propia

Mediante la prueba del Estadístico W de Shapiro-Wilk ya que tiene el P. Valor más restrictivo (el que tiene el P - valor más pequeño)

P. Valor = 0,0579063 > α ; Se acepta H_0 y se concluye que los residuos siguen una Distribución Normal

Cuando se elimina lo puntos anómalos se comprueba de nuevo la heterocedasticidad, finalmente:

- **Contraste de Hipótesis**

$$\left. \begin{array}{l} H_0: \sigma^2 = \text{CONSTANTE} \\ H_1: \sigma^2 \neq \text{CONSTANTE} \end{array} \right\} \alpha = 0,05$$

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	2,29012E16	5	4,58024E15	0,56	0,7315
Residuo	1,31641E17	16	8,22758E15		
Total (Corr.)	1,54543E17	21			

Tabla 42. Contraste Heterocedasticidad - Fuente: Elaboración propia

El modelo presenta un **P. Valor = 0,7315 > α ; se acepta H_0** , se concluye que este modelo no tiene problema de heterocedasticidad. Se aplica un paso hacia adelante para validar las variables significativas.

- **Contraste significación global (modelo).**

$$\left. \begin{array}{l} H_0: \beta_1 = \beta_2 = 0 \\ H_1: \text{Algún de } \beta_i \neq 0 \end{array} \right\} \alpha = 0,05$$

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	7,6419E11	1	7,6419E11	12086,50	0,0000
Residuo	1,26454E9	20	6,32268E7		
Total (Corr.)	7,65455E11	21			

Tabla 43. Contraste significatividad Modelo - Fuente: Elaboración Propia

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	5689,68	2004,63	2,83827	0,0102
empleo	18,58	0,169003	109,939	0,0000

Tabla 44. Contraste significatividad - Fuente: Elaboración Propia

La constante **P. Valor** = 0,000 < α ; **se rechaza H_0** , se concluye que este modelo es significativa

La variable Empleo **P. Valor** = 0,000 < α ; **se rechaza H_0** , se concluye que este modelo es significativa

El Modelo **P. Valor** = 0,000 < α ; **se Rechaza H_0** , se concluye que este modelo es significativa

No se podía solucionar la heterocedasticidad, Para intentar conseguir un modelo válido en este caso lo que se hace es eliminar 4 puntos anómalos candidatos a eliminar que surgían con forme eliminaba el anterior. Así quedando fuera del análisis los países de Francia, Polonia, Reino Unido y Rumanía, es posible conseguir un modelo directamente sin heterocedasticidad y con una distribución normal de los residuos. Quedará la variable empleo sólo como significativa al final y se podrá hacer predicciones directas.

6.6 Estudio de Autocorrelación

Para que el modelo sea válido no puede tener autocorrelación, esto quiere decir que no dependerá de datos anteriores. Puede haber correlación de 1er y 2do orden.

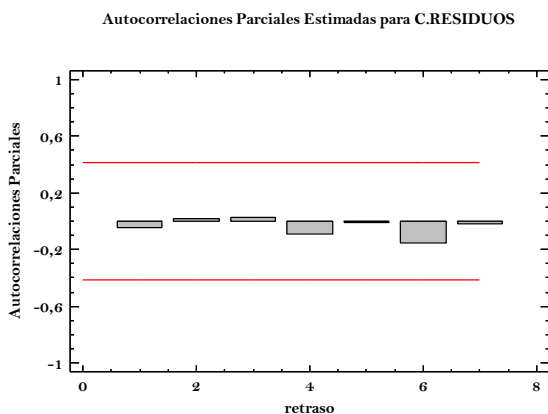


Figura 43. FAS - Fuente: Elaboración propia

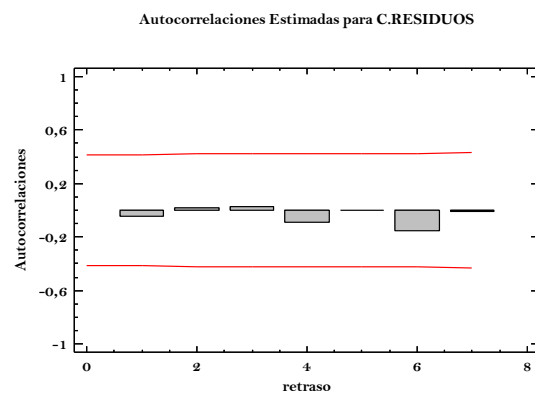


Figura 42. FAP - Fuente: Elaboración propia

La Figura 43 y Figura 42 indican que el Factor de Autocorrelación de primer orden NO supera los límites de autocorrelación por la parte superior, y el resto de los factores de autocorrelación no superan los límites de autocorrelación.

6.7 Predicciones modelo Natalidad

Después de la validación del modelo se propone una predicción actual sobre la Natalidad ya que puede ser viable por tener un R^2 muy bueno de un 98,83%.

$$\mathbf{Natalidad} = \beta_0 + \beta_1 * \text{Empleo} + U$$

$$\mathbf{Natalidad} = 5689,68 + 18,58 * \text{Empleo}$$

¿Qué porcentaje de Natalidad tiene España? Si el empleo en 2017 llega a 22.558 (miles) personas.

	Ajustado	Error Est.	Inferior 95,0%
<i>Fila</i>		<i>LC para Pronóstico</i>	<i>LC para Pronóstico</i>
23	424817,	8580,34	406919,

Tabla 45. Pronóstico Natalidad - Fuente: Elaboración propia

Natalidad = 424817(miles) nacimientos

$$\left\{ 406919 \text{ (miles)} \leq \mathbf{Natalidad} \leq 442716 \text{ (miles)} \right\} \text{ 95\% confianza}$$

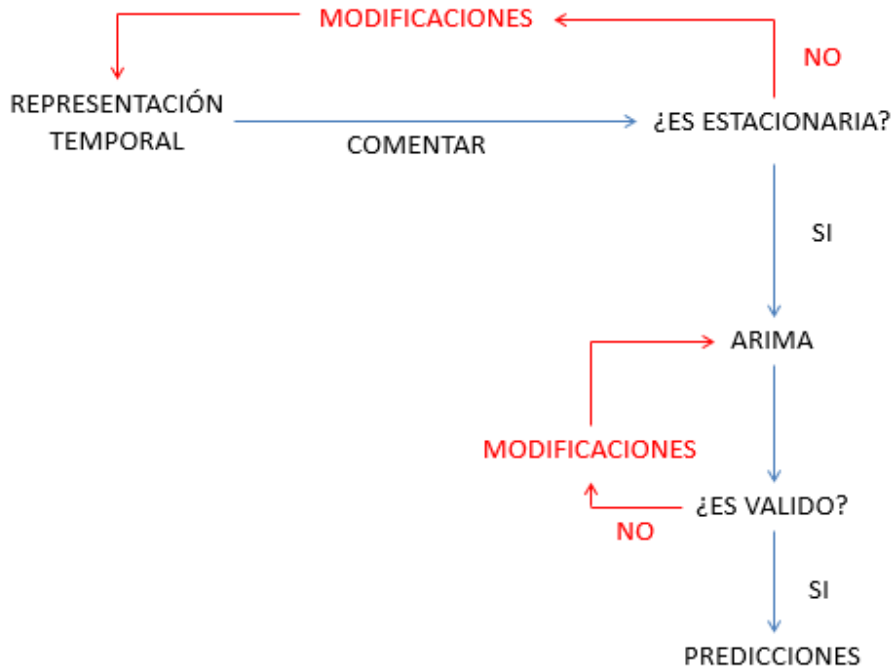
Tras plantear un modelo para poder explicar la Natalidad a nivel europeo, la variable Empleo es la única significativa para poder utilizar en las predicciones. Por tanto se interpreta que en la variación de la tasa de Natalidad afecta directamente el hecho de que las personas trabajen o no.

Ya que se incrementa para una predicción en 2017 de 22558 (miles) personas empleadas se tendría 424817(miles) de nacimientos de niños.

7. Series Temporales

Este estudio se realiza a una serie de datos de una variable ordenada cronológicamente, para poder construir un modelo que describa el comportamiento histórico de la variable, posteriormente sirvan para hacer predicciones de valores futuros.

Esquema de series temporales



7.1 Series Temporales de la tasa de Mortalidad Infantil

En este apartado se plantea una serie temporal de la Mortalidad Infantil, se muestra datos de 1975 a 2018.

Representación Gráfica.- Se representa los datos (variable Mortalidad Infantil), se indica cómo se estructuran en el tiempo.

Gráfica de Serie de Tiempo para Ambos sexos

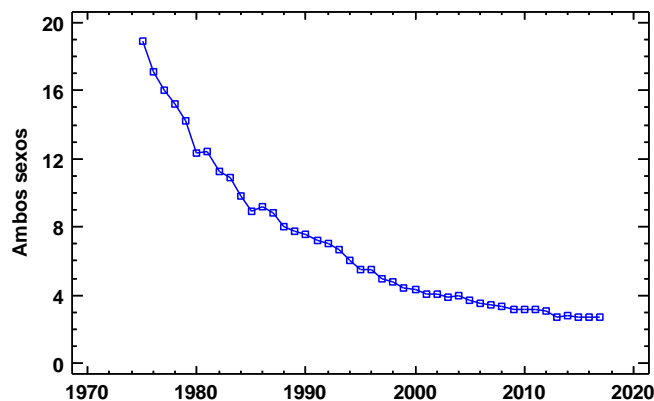


Figura 44. Serie de Tiempo Mortalidad Infantil - Fuente: Elaboración Propia

Al representar la serie temporal se observa que la mortalidad Infantil va descendiendo ya que presenta claramente una tendencia negativa, que desde 1975 a 2000 se observa más acusada el descenso, por otro lado, aunque la tendencia sigue siendo negativa es menos acusada entre 2010 y 2020 indica que es constante.

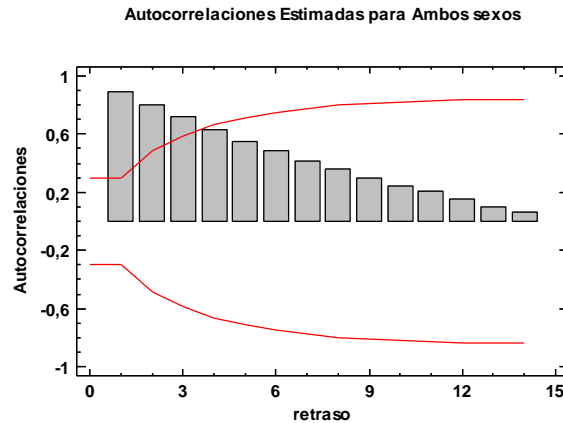


Figura 45. FAS Mortalidad infantil - Fuente: Elaboración propia

Se observa la Figura 45 en la que el Factor de Autocorrelación de primer orden supera los límites de autocorrelación por la parte superior, y el resto de los factores de autocorrelación superan los límites de autocorrelación. En este caso el Figura confirma que la mortalidad infantil tiene un descenso acentuado de los factores de autocorrelación que tiende a cero.

¿El modelo es Estacionaria?

1. Varianza Constante, no presenta forma cónica
2. **Tendencia constante o nula, no cumple tiene tendencia negativa**
3. No debe tener variaciones estacionales, al ser una datos anuales no presenta variaciones estacionales
4. Tener autocorrelación, el Figura de FAS presenta autocorrelación

Se observa de los 4 requisitos que debe cumplir el modelo hay una de ellas que no lo cumple, por lo que se procede a realizar una modificación y solucionar la tendencia negativa que muestra.

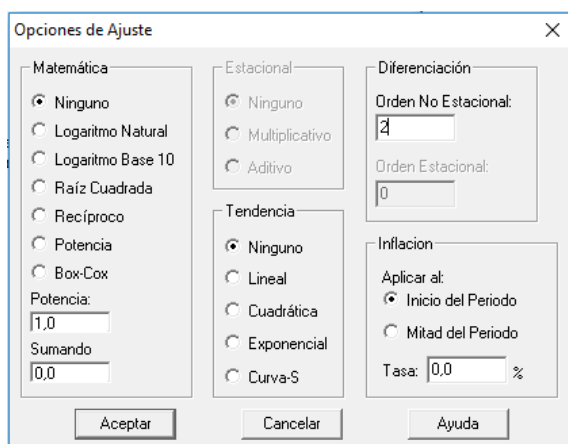


Figura 47. Opciones de ajustes

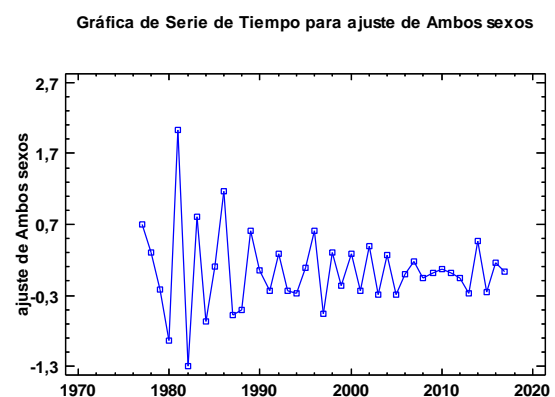


Figura 46. Serie de tiempo Mortalidad Inf.

Para solucionar la tendencia negativa, se aplican 2 diferenciaciones de orden no estacional y se soluciona, ahora el modelo cumple con los requisitos para ser estacionaria.

- **Propuesta del MODELO ARIMA**

Para el planteamiento del modelo es imprescindible que sea una serie estacionaria, Para proponer el modelo se observa las siguiente

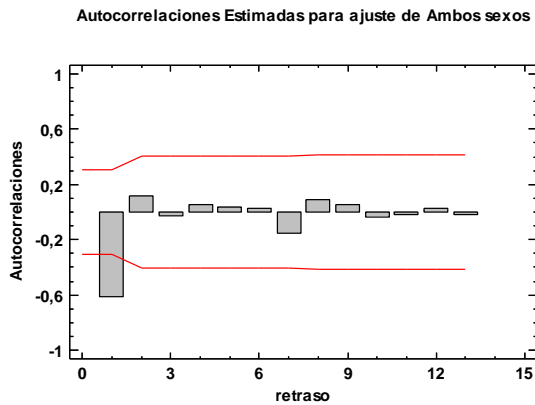


Figura 48 FAS Mortalidad Infantil

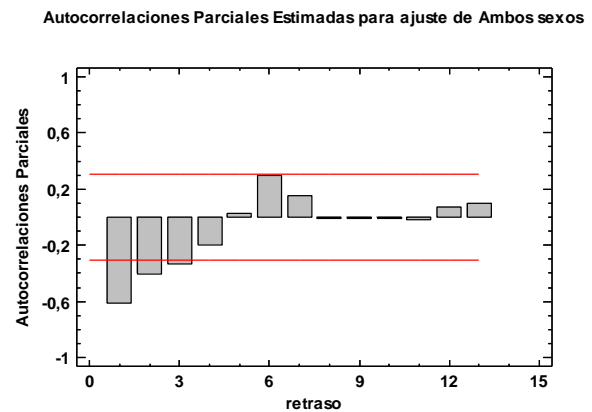
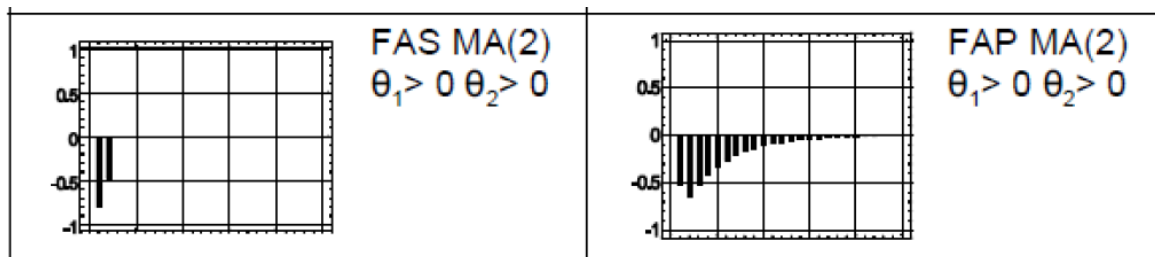


Figura 49. FAP Mortalidad Infantil



Introducir el modelo ARIMA teórico en el programa Statgraphics:

ARIMA (p , d , q) x (P , D , Q) con constante

ARIMA (0 , 2 , 2) con constante

- p = Orden del AR
- d = Orden no estacional
- q = Orden del MA
- P = Orden del SAR
- D = Orden estacional
- Q = Orden del SMA

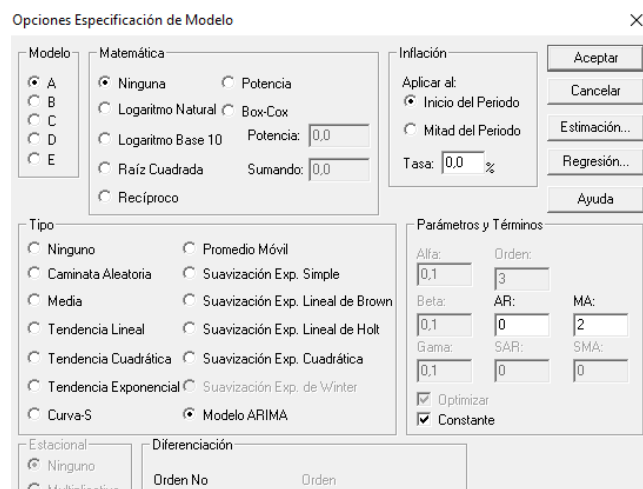


Figura 50. Opciones Especificos del modelo- Fuente: Statgraphics

- **Validación del modelo ARIMA**

a) Significatividad del modelo

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

Resumen de Modelo ARIMA

Parámetro	Estimado	Error Estd.	t	Valor-P
MA(1)	1,43374	0,0570221	25,1436	0,000000
MA(2)	-0,933929	0,0476911	-19,5829	0,000000

Tabla 46. Modelo ARIMA

Para que el modelo sea válido y viable se elimina la constante y el modelo es significativa.

ARIMA (0 , 2 , 2)

MA(1) P.Valor = 0,0000 < 0,05 → Rechazamos **H₀** se concluye que el modelo MA(1) es significativa

MA(2) P.Valor = 0,0000 < 0,05 → Rechazamos **H₀** se concluye que el modelo MA(2) es significativa

b) No tener Autocorrelación

Autocorrelaciones Residuos para ajuste de Mortalidad Infantil
ARIMA(0,2,2)

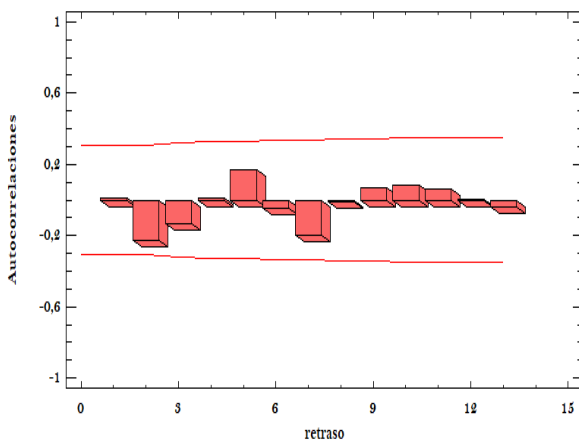


Figura 51. FAS ARIMA- Fuente: Elaboración propia

Autocorrelaciones Residuos Parciales para ajuste de Mortalidad Infantil
ARIMA(0,2,2)

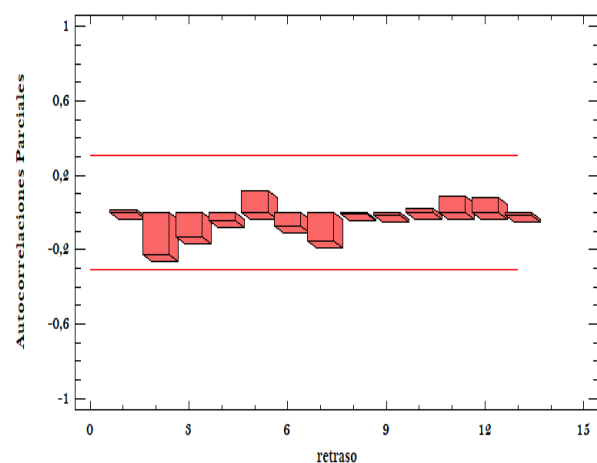


Figura 52. FAP ARIMA- Fuente: Elaboración propia

Se observan la Figura y Figura, que el Factor de Autocorrelación de primer orden NO supera los límites de autocorrelación por la parte superior, y el resto de los factores de autocorrelación no superan los límites de autocorrelación.

c) Principio de Homocedasticidad y media nula

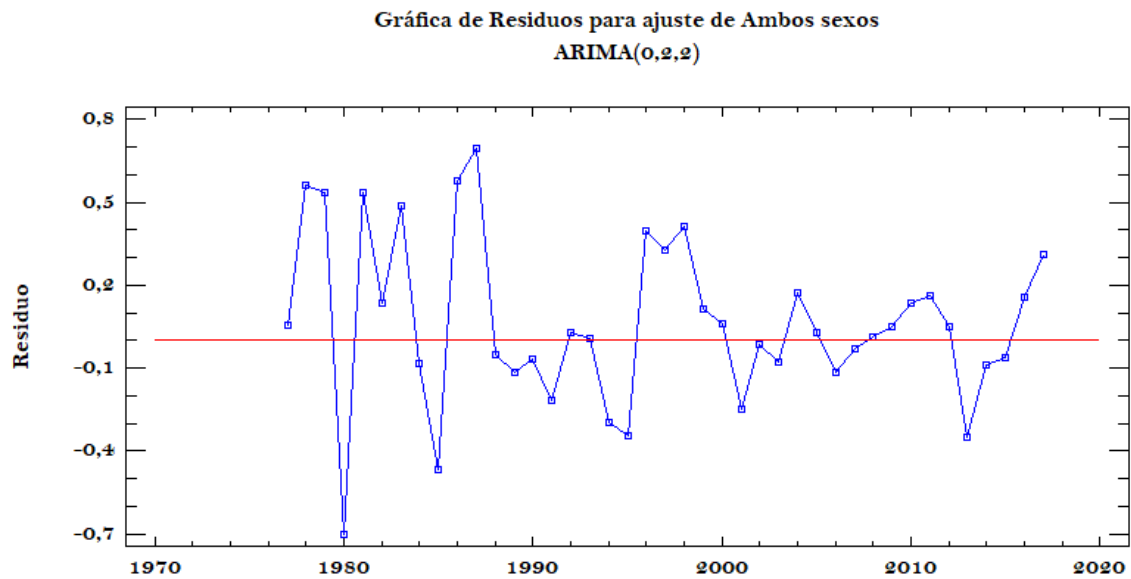


Figura 51. Residuos ARIMA - Fuente: Elaboración propia

Se observa en la Figura 51, que el modelo tiene homocedasticidad y los residuos tiene media nula, ya que la línea se encuentra cerca de cero.

d) Normalidad de Residuos

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

Prueba	Estadístico	Valor-P
Chi-Cuadrado	19,122	0,160342
Estadístico W de Shapiro-Wilk	0,970039	0,460436
Valor-Z para asimetría	0,074602	0,940526
Valor-Z para curtosis	0,423903	0,671633

Tabla 47. Prueba de Normalidad -Fuente: Elaboración propia

Mediante el análisis de normalidad se selecciona la prueba de Estadístico W de Shapiro-Wilk ya que tiene el P.valor más restrictivo (el que tiene el P.Valor más pequeño)

P. Valor = 0,160342 < α ; **se acepta Ho** y se concluye que los residuos son normales.

Después de analizar los requisitos para que sea un modelo válido, se concluye que el modelo de serie temporal de mortalidad Infantil es significativo, con sus residuos normales, no presenta autocorrelación y si representa homocedasticidad.

- **Predicciones modelo Mortalidad Infantil**

Finalmente se valida el modelo y se concluye que para hacer las predicciones es viable para los años posteriores, aunque esta predicción es a corto plazo con un máximo de 3 años siguientes.

		<i>Límite en 95,0%</i>	<i>Límite en 95,0%</i>
<i>Periodo</i>	<i>Pronóstico</i>	<i>Inferior</i>	<i>Superior</i>
2018	2,46004	1,80979	3,11029
2019	2,49061	1,74335	3,23788
2020	2,52119	1,50173	3,54064

Tabla 48. Pronóstico Mortalidad Infantil - Fuente Elaboración propia

Mortalidad Infantil 2018= 2,46% muertes de niños

Intervalo $\left(1,80\% \leq \text{muertes} \leq 3,11\% \right)$ 95% confianza

Mortalidad Infantil 2019= 2,49% muertes de niños

Intervalo $\left(1,74\% \leq \text{muertes} \leq 3,23\% \right)$ 95% confianza

Mortalidad Infantil 2020= 2,52% muertes de niños

Intervalo $\left(1,50\% \leq \text{muertes} \leq 3,54\% \right)$ 95% confianza

Mediante la serie temporal válida, se predice datos posteriores pero a corto plazo por lo que se concluye que para 2018 se tendrá un 2,46% de Mortalidad Infantil, para 2019 un 2,45% y 2020 2,52% , esto indica que permanece constante con mínimas variaciones.

7.2 Series temporales de la tasa de Fertilidad

Representación Gráfica.- Se representa los datos (variable Fertilidad), se indica cómo se estructuran en el tiempo.

Gráfica de Serie de Tiempo para Ambas nacionalidades

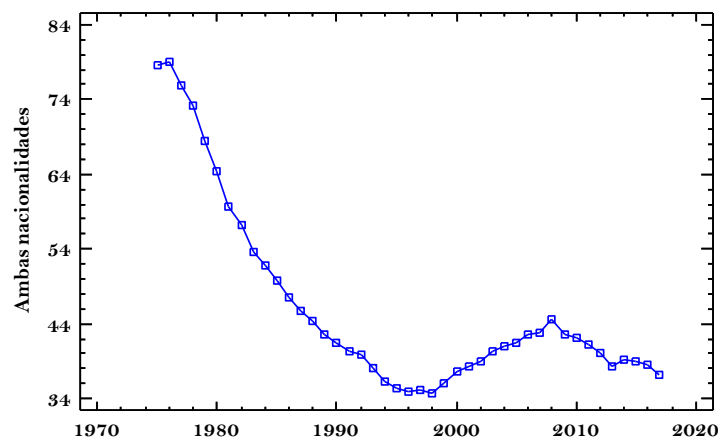


Figura 52. Serie de Tiempo Fertilidad - Fuente: Elaboración Propia

Al representar la serie temporal se observa que la fertilidad va descendiendo ya que presenta claramente una tendencia negativa muy pronunciada desde 1975 hasta 1998, por otro lado, a partir de 1999 se observa un aumento hasta aproximadamente 2007, esto

según datos en el INE es por el aumento de la migración en dichos años. Desde 2008 se ve un mínimo descenso aunque se muestra constante en los últimos años.

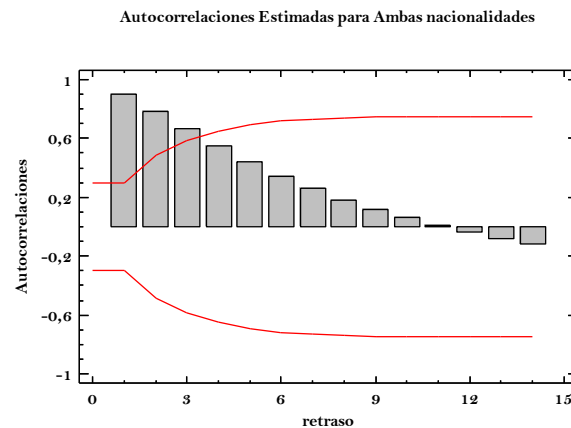


Figura 53. FAS Fertilidad - Fuente: Elaboración propia

En la Figura 53 se observa que el Factor de Autocorrelación de primer orden supera los límites de autocorrelación por la parte superior, y el resto de los factores de autocorrelación superan los límites de autocorrelación. En este caso la Figura 53 confirma que la tasa de fertilidad tiene un descenso acentuado de los factores de autocorrelación.

¿El modelo es Estacionaria?

1. Varianza Constante, no presenta forma cónica
2. **Tendencia constante o nula, no cumple tiene tendencia negativa**
3. No debe tener variaciones estacionales, al ser una datos anuales no presenta variaciones estacionales
4. Tener autocorrelación, el Figura de FAS presenta autocorrelación

Se observa de los 4 requisitos que debe cumplir el modelo hay una de ellas que no lo cumple, por lo que se procede a realizar una modificación y solucionar la tendencia negativa que muestra.

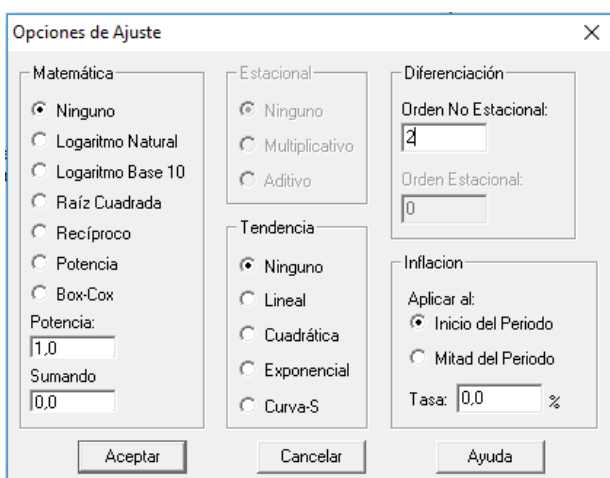


Figura 54. - Fuente: Elaboración propia

Gráfica de Serie de Tiempo para ajuste de Ambas nacionalidades

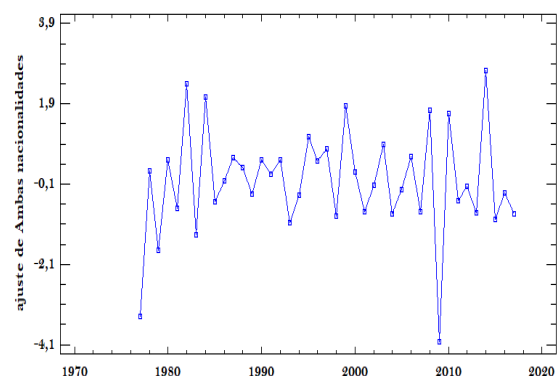


Figura 55. Serie Tiempo Fertilidad- Fuente: Elaboración propia

Para solucionar la tendencia negativa, aplicamos 2 diferenciaciones de orden no estacional y se observa que lo soluciona, ahora el modelo cumple con los requisitos para ser estacionaria.

- **Propuesta del MODELO ARIMA**

Para el planteamiento del modelo es imprescindible que sea una serie estacionaria, para proponer el modelo nos fijamos en los siguientes Figuras FAS Y FAP.

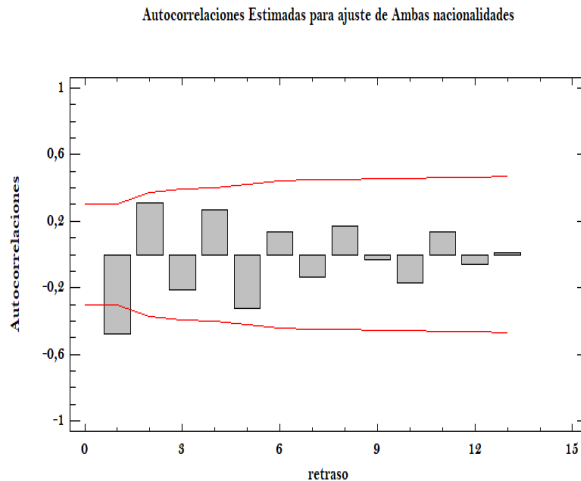


Figura 57. FAS Fertilidad - Fuente: Elaboración propia

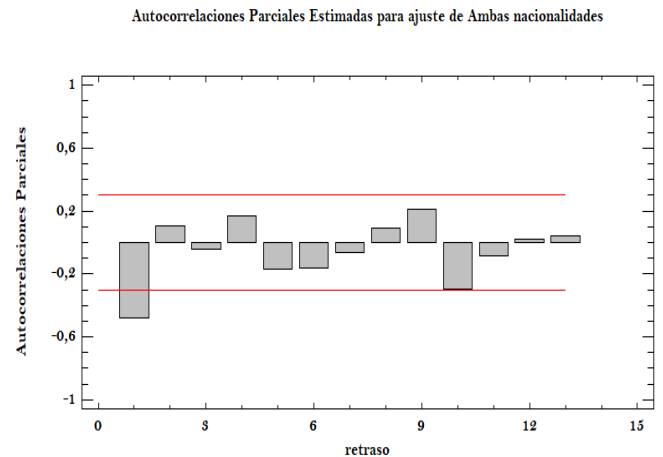
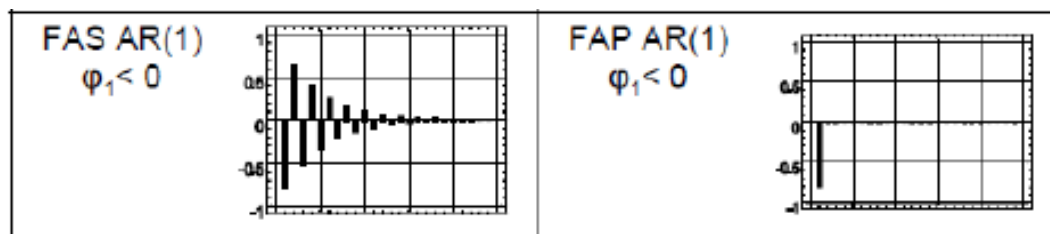


Figura 56. - FAP fertilidad - Fuente: Elaboración propia



Introducir el modelo ARIMA teórico en el programa Statgraphics:

ARIMA (p , d , q) x (P , D , Q) con constante

ARIMA (1 , 2 , 0) con constante

p = Orden del AR

d = Orden no estacional

q = Orden del MA

P = Orden del SAR

D = Orden estacional

Q = Orden del SMA

Opciones Especificación de Modelo

Modelo: Matemática

A Ninguna Potencia

B Logaritmo Natural Box-Cox

C Logaritmo Base 10 Potencia: 0,0

D Raíz Cuadrada Sumando: 0,0

E Recíproco

Inflación

Aplicar al: Inicio del Periodo Mitad del Periodo

Tasa: 0,0 %

Tipo

Ninguno Promedio Móvil

Caminata Aleatoria Suavización Exp. Simple

Media Suavización Exp. Lineal de Brown

Tendencia Lineal Suavización Exp. Lineal de Holt

Tendencia Cuadrática Suavización Exp. Cuadrática

Tendencia Exponencial Suavización Exp. de Winter

Curva-S Modelo ARIMA

Parámetros y Términos

Alla: 0,1 Orden: 3

Beta: AR: 1 MA: 0

Gama: SAR: 0 SMA: 0

Optimizar Constante

Estacional: Ninguno Multiplicativo Aditivo

Diferenciación: Orden No 2 Orden 0

Aceptar Cancelar Estimación... Regresión... Ayuda

Figura 58. Opciones Específicos del modelo- Fuente: Statgraphics

• **Validación del modelo ARIMA**

a. Significatividad del modelo

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

Resumen de Modelo ARIMA

Parámetro	Estimado	Error Estd.	t	Valor-P
AR(1)	-0,606688	0,120953	-5,01589	0,000011

Tabla 49. Contraste Arima – Fuente Elaboración propia

Para que el modelo sea válido y viable se elimina la constante y el modelo es significativa.

ARIMA (1 , 2 , 0)

AR(1) P.Valor = **0,0000 < 0,05** → Se rechaza *H₀* se concluye que el modelo AR(1) es significativa

b. No tener Autocorrelación

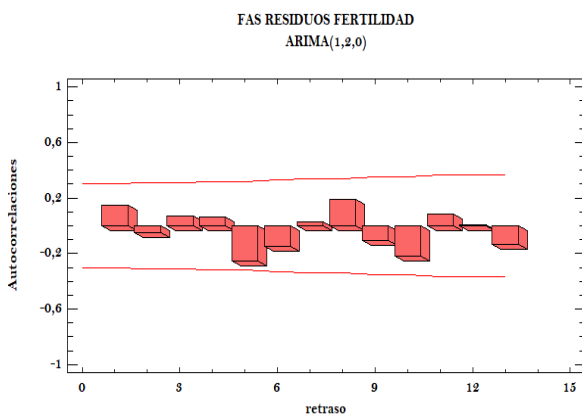


Figura 59.FAS ARIMA - Fuente Elaboración propia

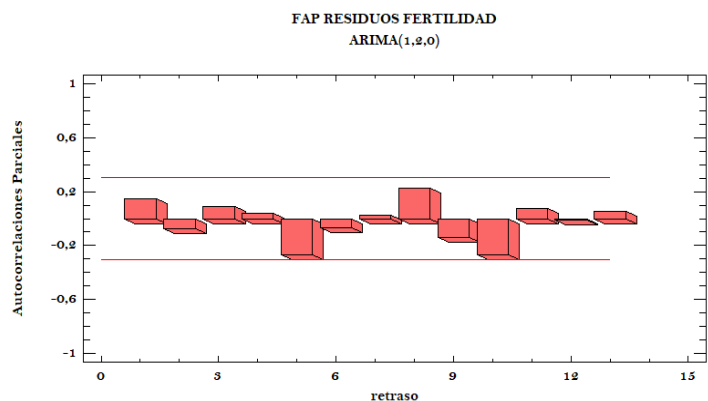


Figura 60.FAP ARIMA - Fuente Elaboración propia

Se observa la Figura 59 y Figura 60, que el Factor de Autocorrelación de primer orden NO supera los límites de autocorrelación por la parte superior, y el resto de los factores de autocorrelación no superan los límites de autocorrelación.

c. Principio de Homocedasticidad y media nula

Gráfica de Residuos para ajuste de Ambas nacionalidades
ARIMA(1,2,0)

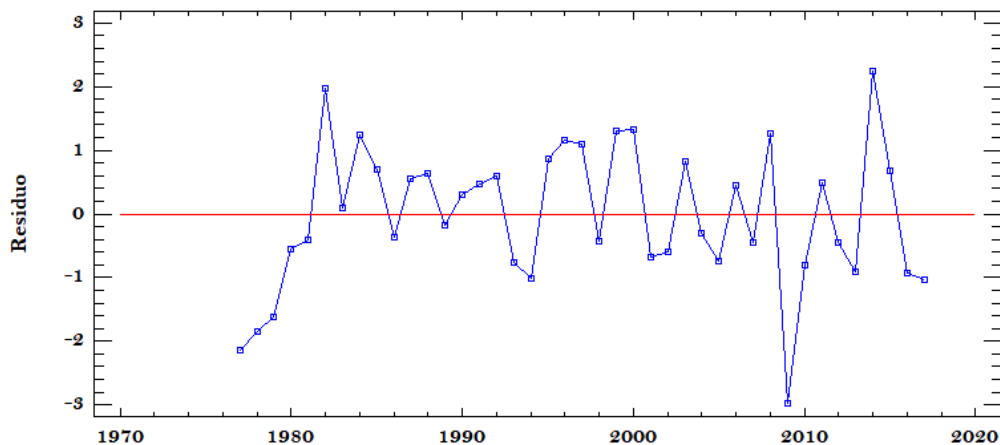


Figura 61. Residuos Arima Fuente Elaboración propia

Se observa en la Figura 61 de Residuos muestra que el modelo tiene homocedasticidad y los residuos tiene media nula.

d. Normalidad de Residuos

Pruebas de Normalidad para RESIDUOS

H_0 : Los residuos normales
 H_1 : Los residuos no normales
 }
 $\alpha = 0,05$

Prueba	Estadístico	Valor-P
Chi-Cuadrado	15,8049	0,325435
Estadístico W de Shapiro-Wilk	0,977018	0,683403
Valor-Z para asimetría	0,628056	0,529965
Valor-Z para curtosis	0,452197	0,651124

Tabla 50. Pruebas de Normalidad

Mediante el análisis de normalidad se selecciona la prueba Chi-cuadrado ya que tiene el P.valor más restrictivo (el que tiene el P.Valor más pequeño) **P. Valor = 0,3254 > α** ; **Aceptamos H_0** y se concluye que los residuos son normales.

Después de analizar los requisitos para que sea un modelo válido, se concluye que el modelo de serie temporal de la tasa de fertilidad es significativo, con sus residuos normales, no presenta autocorrelación y si representa homocedasticidad.

• Predicciones del Modelo Fertilidad

Periodo	Pronóstico	Límite en 95,0%	
		Inferior	Superior
2018	36,4196	34,1341	38,7051
2019	35,3101	31,3904	39,2297
2020	34,3881	28,094	40,6821

Tabla 51. Predicciones Fertilidad

Fertilidad 2018 = 36,42% de hijos nacidos por mujer fértil
Intervalo $\left(34,13\% \leq \text{nacimientos} \leq 38,70\% \right)_{95\%}$

Fertilidad 2019 = 35,31% de hijos nacidos por mujer fértil
Intervalo $\left(31,39\% \leq \text{nacimientos} \leq 39,22\% \right)_{95\%}$

Fertilidad 2020 = 34,38% de hijos nacidos por mujer fértil
Intervalo $\left(28,09\% \leq \text{nacimientos} \leq 40,68\% \right)_{95\%}$

Mediante la serie temporal válida, se predice datos posteriores pero a corto plazo por lo que se concluye que para 2018 se tendrá un 36,42% de tasa de fertilidad, para 2019 un 35,31% y 2020 34,38% , esto indica que irá en descenso.

8. CONCLUSIONES

En la Búsqueda de factores influyentes sobre las tasas de natalidad, fertilidad, mortalidad general e infantil, se plantean modelos econométricos que ayudan a explicar sus variaciones, entre las variables que se elige está: el empleo, las actividades físicas, consumo de frutas y verduras, consumo de cigarrillos y el índice de pobreza, para determinar si el problema puede ser social o de hábitos personales.

Se plantea 4 modelos mediante la regresión múltiple, entre ellos la Mortalidad y Natalidad da un R^2 cercano a 1 con un 98,64%.y un 98,83%. Por lo que para predicciones posteriores puede explicar prácticamente la mayoría de sus datos.

Sobre la estructura lineal para explicar la Mortalidad se comenta que no presenta problemas de multicolinealidad, ni alinealidad, los residuos son normales y no tiene autocorrelación, pero si muestra problemas de Heterocedasticidad por lo que se divide el modelo inicial entre la variable que genera el problema elevada a C. Con esto se determina que no se soluciona la heterocedasticidad, se aplica un paso hacia atrás y se elimina la constante, con esto se obtienen variables significativas entre ellas están: el empleo, las actividades físicas y el índice de pobreza llevando a obtener un modelo válido y factible.

Respecto al modelo para explicar la Natalidad se expone que no presenta problemas de multicolinealidad, ni alinealidad, ni autocorrelación, pero si muestra que los residuos no siguen una distribución normal y problemas de Heterocedasticidad por lo que se divide el modelo inicial entre la variable que genera el problema elevada a C.

Con esto se determina que no se soluciona la heterocedasticidad, Para intentar conseguir un modelo válido en este caso lo que se hace es eliminar 4 puntos anómalos candidatos a eliminar que surgían con forme eliminaba el anterior. Así quedando fuera del análisis los países de Francia, Polonia, Reino Unido y Rumanía, es posible conseguir un modelo directamente sin heterocedasticidad y con una distribución normal de los residuos. Quedará la variable empleo sólo como significativa al final y se podrá hacer predicciones directas.

Por otro lado, la mortalidad infantil da un R^2 muy lejano a 1 con 31,25% y la fertilidad un 20,47 % por tanto estos modelos no son viables para determinar predicciones a futuro, ya que no se podrían explicar todos sus datos. Pero para estas tasas se realiza una serie temporal con datos seleccionados de 1975 hasta la actualidad.

Se Plantea un Modelo Arima MA(2) para la tasa de mortalidad Infantil y para la Tasa de Fertilidad un Arima AR(1) y se procede a validar los modelos. Los modelos de serie temporal son significativos, con sus residuos normales, no presentan autocorrelación y tienen homocedasticidad.

Finalmente, se concluye que en la tasa de mortalidad a medida que aumente la pobreza y disminuya el empleo incrementará la mortalidad. La actividad física también influye en la variable por lo que serie interesante concienciar a las personas a realizar deporte.

Para la tasa de natalidad se prevé que si hay menos empleo las familias deciden tener menos niños ya que la economía en los hogares puede ser insuficientes. Esto lleva a explicar por qué en España y otros países de Europa retrasan el hecho de tener hijos y provocan el problema de una población muy envejecida.

9. Bibliografía

- «AGENCIA EFE.» 14 de MARZO de 2019.
<https://www.efe.com/efe/espana/sociedad/espana-es-el-segundo-pais-con-menor-tasa-de-mortalidad-la-union-europea/10004-3552540>.
- «BANCO MUNDIAL.» 21 de Marzo de 2019.
<https://datos.bancomundial.org/indicador/SP.DYN.IMRT.IN>.
- confidencial, EL. «https://www.elconfidencial.com/economia/2018-12-11/espana-menor-numero-nacimientos-historia-defunciones_1698198/.» 22 de MAYO de 2019.
- «EL ECONOMISTA.» 22 de MARZO de 2019.
<https://www.eleconomista.es/economia/noticias/9756090/03/19/Espana-tiene-la-segunda-tasa-de-fertilidad-mas-baja-de-Europa-y-las-madres-primerizas-mas-mayores-.html>.
- «EL ECONOMISTA.» 25 de MARZO de 2019.
<https://www.eleconomista.es/economia/noticias/8679980/10/17/Economia-Consumo-La-exportacion-de-frutas-y-verduras-posiciona-a-Espana-como-principal-proveedor-de-Europa.html>.
- «EL PAIS.» 10 de MAYO de 2019.
https://elpais.com/economia/2019/01/03/actualidad/1546541410_156906.html.
- «EL PERIODICO.» 26 de mayo de 2019.
<https://www.elperiodico.com/es/sociedad/20161119/25-millones-de-menores-sufren-riesgo-de-exclusion-en-europa-5638694>.
- «EXPANSION.» 25 de marzo de 2019.
<https://datosmacro.expansion.com/demografia/natalidad/zona-euro>.
- «LA VANGUARDIA.» 10 de MAYO de 2019.
<https://www.lavanguardia.com/economia/20190109/454041157703/creacion-empleo-2019-menos-2018.html>.
- «MAGAZINE.» 25 de MARZO de 2019. <https://iymagazine.es/noticia/2633/id-universitario/el-nivel-de-actividad-fisica-y-deporte-de-los-europeos-por-paises-a-estudio.html>.
- MINDS, DW MADE FOR. «DW MADE FOR MINDS .» 12 de MARZO de 2019.
<https://www.dw.com/es/oms-esperanza-de-vida-de-las-mujeres-es-mayor-que-la-de-los-hombres/a-48206298>.
- MUNDO, EL. «<https://www.elmundo.es>.» 26 de MAYO de 2019.
- Núñez, Jorge Jordán. «Apuntes Econometricos.» 2017.
- «SALUD.» 22 de MARZO de 2019. <https://www.bebesymas.com/salud-infantil/mortalidad-infantil-en-europa-los-accidentes-son-la-principal-causa>.
- WIKIPEDIA. «<https://es.wikipedia.org>.» 12 de MAYO de 2019.