



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CAMPUS D'ALCOI

Modelización del flujo de matriculados universitarios a nivel europeo, mediante variables socioeconómicas del entorno

MEMORIA PRESENTADA POR:

Raquel Tena García

GRADO DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Convocatoria de defensa: Julio del 2016

Índice

1. INTRODUCCIÓN Y OBJETIVO DEL PROYECTO	6
1.1. NOTICIAS.....	6
1.1.1. “¿QUÉ ESTUDIAR PARA SER MILLONARIO?”	6
1.1.2. “EL NÚMERO DE ESTUDIANTES QUE BUSCA EMPLEO CRECE UN 28% Y ALCANZA CIFRAS RÉCORD”	7
1.1.3. “¿POR QUÉ ESTUDIAR UNA CARRERA UNIVERSITARIA?”	7
1.2.1. LAS ACTITUDES PARA EL APRENDIZAJE Y LA EDUCACIÓN PROFESIONAL	8
2. OBJETIVOS	13
2.1. OBJETIVOS SECUNDARIOS	13
3. METODOLOGÍA UTILIZADA EN EL PROYECTO	15
3.1. BÚSQUEDA DE DATOS.....	15
3.2 TÉCNICAS DE ANÁLISIS.....	17
3.2.1. ANÁLISIS BIVARIANTE	17
3.2.2. ANÁLISIS MULTIVARIANTE	18
4. CÁLCULOS Y RESULTADOS	25
4.1 BÚSQUEDA DE VARIABLES.....	25
4.2. MULTICOLINEALIDAD	26
4.2.1 IDENTIFICACIÓN	26
4.2.2 SOLUCIÓN DE MULTICOLINEALIDAD.....	29
4.2.3 COMPROBACIÓN.....	30
4.3. MODELO LINEAL	31
4.3.1. SIGNIFICACIÓN	32
4.3.2. FALTA DE LINEALIDAD	33
4.4.3. NORMALIDAD	40
4.4.4. HETEROCEDASTICIDAD	41
4.4.5. AUTOCORRELACIÓN	41
4.5. MODELO LINEAL SIN CONSTANTE.....	43

4.5.1. SIGNIFICATIVIDAD	43
4.5.2. FALTA DE LINEALIDAD	44
4.5.3. NORMALIDAD	51
4.5.4. HETEROCEDASTICIDAD.....	52
4.5.5. AUTOCORRELACIÓN	53
4.6. MODELO LINEAL SIN CONSTANTE NI VARIABLES PROBLEMATICAS.....	54
4.6.1. SIGNIFICATIVIDAD	54
4.6.2. FALTA DE LINEALIDAD	55
4.6.3. NORMALIDAD	58
4.6.4. HETEROCEDASTICIDAD.....	59
4.6.5. AUTOCORRELACIÓN	59
4.7 MODELO LINEAL SIN MALTA.....	60
4.7.1. SIGNIFICATIVIDAD	60
4.7.2. FALTA DE LINEALIDAD	61
4.7.3. NORMALIDAD	65
4.7.4. HETEROCEDASTICIDAD.....	65
4.7.5. AUTOCORRELACIÓN	66
4.7.6. MODELO FINAL LINEAL.....	67
4.8. MODELO LOGARITMICO.....	67
4.8.1. SIGNIFICATIVIDAD	67
4.8.2. FALTA DE LINEALIDAD	68
4.8.3. NORMALIDAD	71
4.8.4. HETEROCEDASTICIDAD.....	71
4.8.5. AUTOCORRELACIÓN	72
4.9.6. MODELO LOGARÍTMICO	73
4.9. MODELO INVERSO	73
4.9.1. SIGNIFICATIVIDAD.....	73
4.9.2. FALTA DE LINEALIDAD	74
4.9.3. NORMALIDAD	75
4.9.4. HETEROCEDASTICIDAD.....	76

4.9.5. AUTOCORRELACIÓN	76
4.9.6. MODELO ESTRUCTURA INVERSA	77
4.10. RESUMEN DE LAS ESTRUCTURAS Y ELECCIÓN DEL MODELO	78
4.10.1. LINEAL	78
4.10.2. LOGARÍTMICO	78
4.10.3. INVERSA	78
5. PREDICCIÓN	81
5.1. CÁLCULO DE LA PREDICCIÓN	81
6. CONCLUSIONES	84

I. Introducción

1. INTRODUCCIÓN Y OBJETIVO DEL PROYECTO

Este proyecto está basado en una investigación macroeconómica, en la que se pretende comprender el por qué la gente estudia en las universidades en el entorno europeo. La base y motivación ha sido el resultado de encontrar noticias, estudios y estadísticas sobre las carreras universitarias, y la necesidad de querer estudiar y entender el comportamiento de las personas europeas ya sea, por gustos, costumbre familiar, etc., y de esta manera, poder realizar estudios completos y complejos, que ayuden a comprender fenómenos como es el interés por la formación superior en los países con elevado nivel de desarrollo.

1.1. NOTICIAS

En este apartado se podrán ver algunas de las noticias encontradas en internet, las cuales motivaron la temática principal del proyecto.

1.1.1. “¿QUÉ ESTUDIAR PARA SER MILLONARIO?”

Esta noticia encontrada en la página web de educación www.universia.com, resalta que la mayoría de millonarios son europeos y con carreras universitarias. Esto podría ser un punto de motivación para la gente, a la hora de querer empezar estudios universitarios. La empresa Approved Index analiza la lista de las 100 personas más ricas del mundo que elabora anualmente la revista Forbes, en la que se observa que el 68% de los listados tienen una titulación universitaria. Según estima el Censo Wealth-X y UBS Billionaire las 2.325 personas más ricas del mundo acumulan un valor neto total de 7,29 billones de dólares, casi la décima parte del Producto Interno Bruto (PIB) mundial.

Además en esta noticia se afirma que más de la quinta parte de los millonarios (22%) ha estudiado Ingenierías, prácticamente duplicando el porcentaje de la segunda carrera en el listado, Negocios (12%). En el 3º puesto se ubican las Artes (9%), Ciencias económicas (8%), Finanzas (3%), y un 2% de los hombres más ricos ha optado por Ciencia, Matemáticas o Abogacía.

El 29% de los hombres más ricos del mundo son europeos. La directora de Approved Index, Amy Catlow, sostiene que “estos resultados indudablemente agregan una nueva dimensión al debate acerca de la relevancia y el valor actual de las titulaciones, y sugiere que para lograr una economía próspera y diversa, necesitamos incentivar una amplia variedad de especializaciones” [1]. Con esto se plantea una de las primeras hipótesis de este trabajo, de forma que una de las causas por las que en los países

I.Introducción

desarrollados como son los que conforman la Unión Europea, asciende de forma progresiva el interés por los habitantes en la obtención de titulaciones universitarias.

1.1.2. “EL NÚMERO DE ESTUDIANTES QUE BUSCA EMPLEO CRECE UN 28% Y ALCANZA CIFRAS RÉCORD”

Esta noticia encontrada en la página web www.20minutos.es da muchas esperanzas de obtener empleo cuando los estudiantes universitarios acaben sus estudios universitarios. Sin embargo cabe resaltar que, el empleo obtenido no sea de la misma rama del grado estudiado por los estudiantes, y puede que estén empleados en otros sectores, como por ejemplo, hostelería.

En 2015, el número de estudiantes demandantes de empleo ha registrado la cifra más alta de los últimos ocho años, tras tres periodos de aumentos consecutivos. En total, el mercado laboral nacional registra 81.600 estudiantes buscando empleo, lo que supone un 28% más que en 2014, cuando la cifra se situaba en 63.800.

Un análisis de Randstad basado en los datos del Servicio de Empleo Público Estatal, pone de manifiesto que este colectivo ha duplicado sus cifras respecto a 2008, ya que el aumento se sitúa en el 107%. Hace ocho años, el número de profesionales alcanzaba los 39.400. Por otro lado, dos de cada cinco estudiantes desempleados lleva menos de seis meses buscando un puesto de trabajo, el 40,8% de ellos.

En concreto, 33.272 estudiantes se encuentran en esta situación, un 13,4% más que en mayo de 2014. Por su parte, el 23% de los profesionales de este colectivo demanda un empleo desde hace más de 6 meses y menos de un año, mientras que el 37% restante lleva más de un año en esta situación [2]. Así que la segunda hipótesis lógica que se extrae, es que otra de las causas principales por las que los europeos buscan obtener una titulación universitaria, es por la esperanza de que ésta les proporcione garantías de encontrar un empleo remunerado.

1.1.3. “¿POR QUÉ ESTUDIAR UNA CARRERA UNIVERSITARIA?”

Ésta noticia encontrada en la página web www.mentesliberadas.com.ar expone el por qué y las expectativas futuras de la gente, a la hora de adentrarse en el mundo estudiantil y elegir una carrera universitaria.

Hay decisiones en la vida que son trascendentes. Una de ellas puede llegar a ser elegir una carrera. Quienes sigan el camino de la educación académica, deberán definir en qué área especializarse. Las posibilidades son amplias y muy variadas. Pero lo más importante, es el motivo que lleva a una persona a empezar una carrera.

Como puede verse, la opción “**Porque me gustan los temas de la carrera**”, está asociada directamente con la vocación y la pasión que uno tiene por determinadas áreas del conocimiento

I.Introducción

humano. Se sabe que lo ideal sería que todos estudien por este motivo, pero en la realidad esto no ocurre así.

“Para ampliar las oportunidades laborales y económicas” es la segunda opción y suele ser bastante común en estos días. Los jóvenes se inclinan por carreras que pueden llegar a abrir las puertas del mercado laboral, para así tener un futuro económicamente próspero. Claro que esto no está garantizado ni mucho menos.

La tercera opción es **“Por reconocimiento social”**. Hay personas que se sienten demasiado presionadas por su entorno. Se encuentran rodeadas de amigos y conocidos que estudian carreras universitarias. Esta presión los lleva a seguir estudiando para no sentirse en inferioridad de condiciones.

La cuarta es **“Porque mi familia me lo exige”**. No es raro que aún en estos días, los padres presionen a los hijos para que se inclinen por la educación universitaria, y hasta los obliguen a especializarse en un área determinada.

Finalmente, aparece la opción **“Otros”** que es la más importante para esta nota. Justamente porque agrupa demasiadas posibilidades. Por razones de espacio, no se pueden agregar todas a la encuesta, pero sí crear este espacio para que puedan comentar aquí aquellos que hayan votado, y quieran especificar alguna otra razón que los motive a estudiar una carrera universitaria [3]. Es decir, que hay una parte relevante de los encuestados, que señala tener otras razones, a parte de las que podían parecer más evidentes, para decidirse por ampliar sus estudios a nivel universitario. Encontrar estos motivos, y comprender las relaciones entre los factores que afectan al incremento del número de universitarios en Europa, es una de las principales motivaciones de este proyecto.

1.2. Estudio europeo

En este sub-apartado se va a comentar uno de los estudios encontrados más afín al tema objeto de este proyecto. Éste estudio encontrado en la base de datos www.ec.europa.eu, también está basado en las influencias y motivaciones de las personas a la hora de elegir qué estudiar. La Investigación está realizada en los 27 Estados Miembros de Europa, encuestando a 26.840 personas a partir de 15 años en el 2011. Este documento no se muestra completo aquí. El estudio fue solicitado por la Dirección General de Educación y Cultura, y coordinado por la Dirección General de Comunicación, y llevada a cabo por TNS Opinion & Social. A continuación se pasa a comentar con detalle los principales resultados del estudio analizado:

1.2.1. LAS ACTITUDES PARA EL APRENDIZAJE Y LA EDUCACIÓN PROFESIONAL

Calidad de los servicios de orientación en general

En primer lugar se cuestionan si los encuestados consideran suficiente la información, o el asesoramiento que reciben los jóvenes, a la hora de escoger una carrera universitaria. Una mayoría de

I.Introducción

los encuestados de la UE (52%) cree que los jóvenes reciben suficiente asesoramiento sobre su aprendizaje, y oportunidades profesionales de las escuelas y los servicios de empleo. Por otro lado el 39% de encuestados (encuestados con un mayor nivel de educación y gente con problemas económicos), piensan que no reciben suficiente asesoramiento, y el restante 10% no sabe si la gente joven recibe suficiente asesoramiento.

Factores que influyen en la elección de un itinerario de aprendizaje

La fuente más común de los encuestados de la UE son sus padres u otro miembro de la familia: el 41% de las personas dicen que recibieron asesoramiento de esta fuente. Sin embargo, 56% de los ciudadanos de la UE dicen que no recibieron asesoramiento sobre itinerarios de aprendizaje de su familia.

De las personas que no han sido aconsejadas por su familia, el 20% de los encuestados de la UE dicen que tomaron consejo educativo de alguien desde el mundo laboral, el 28% desde un profesor, el 19% desde un amigo, el 14% desde un orientador, el 10% por su director y el resto se han visto aconsejadas de otra manera.

También algunos de los factores son tales como las oportunidades de empleo, que son un factor importante que determina la elección de un camino vocacional, el tipo de enseñanza (práctica o académica), la imagen de la escuela, institución o empleador influenciada por la misma escuela o institución, la duración de los estudios, los costes o lo lejos que esté la universidad de la ciudad donde vive el estudiante (Figura I. 1).

Las fuentes de información utilizadas para la elección de un itinerario de aprendizaje

La familia es la fuente de información que los encuestados de la UE más utilizan a la hora de elegir. Las escuelas se citan como la fuente de información más importante en un 14%, mientras que el 10% de los encuestados dicen que utilizan Internet y las redes sociales. Sólo el 6% de los encuestados identifican a los amigos como su principal fuente de información. Eventos y jornadas de puertas abiertas en las escuelas, se citan únicamente en un 4% de los encuestados. Otro 4% indican que la fuente procede desde la experiencia laboral. La prensa es nombrada por el 3% de los encuestados, repartiéndose entre la televisión con un 2%, y exposiciones o ferias con el 1%. Además la radio es mencionado por sólo el 2% de los encuestados. El 4% de los encuestados dicen que usan otras fuentes para obtener su información, y finalmente el 19% de las personas dicen que no utilizan ninguna fuente de información educativa (Figura I. 2) [4].

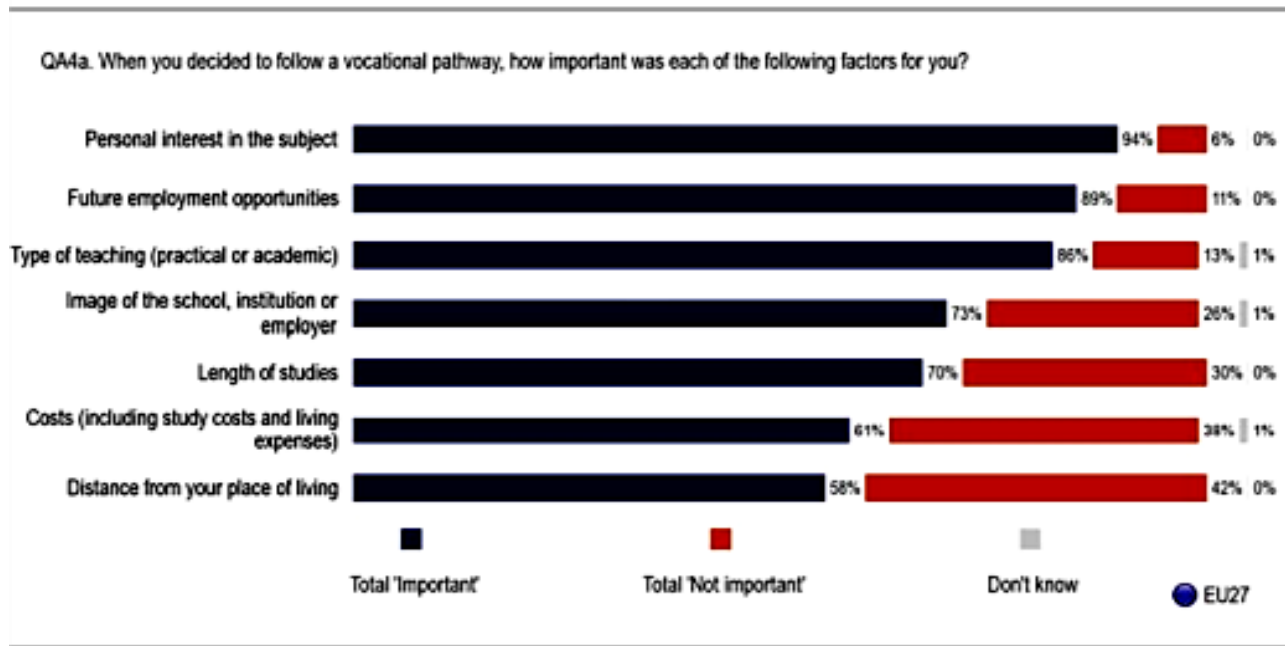


Figura I. 1 Gráfico de barras sobre la encuesta del estudio, a la respuesta de cuándo decides seguir una vía profesional

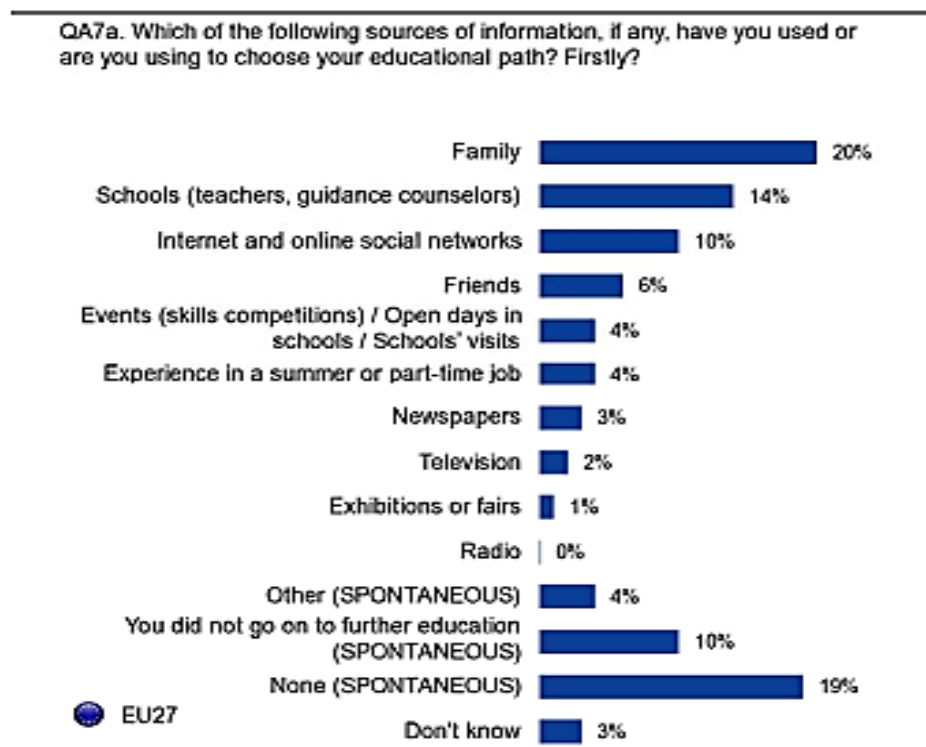


Figura I. 2 A quién se le pide consejo para elegir la vía profesional

I.Introducción

En definitiva se observa que tanto las noticias destacadas, como el estudio encontrado, trabajan con estadísticos univariantes, aportando información comparativa con gráficos de frecuencias o sectores, y comparando los porcentajes obtenidos para las encuestas y los datos analizados. En ningún caso se cuestiona la relevancia entre las diferencias encontradas en las frecuencias en qué responden los encuestados, o se cuestiona la significatividad o relevancia real de los factores analizados en principio. En todos los casos se centran en el análisis puramente descriptivo de las cuestiones planteadas en cada tipo de encuesta. En este proyecto se propone hacer un estudio del entorno de la población, que motiva a la iniciación de estudios universitarios. Además de realizar un estudio con variables generales del entorno socioeconómico de los países Europeos, se buscará encontrar relaciones entre más de dos variables, en el comportamiento de los estudiantes universitarios. De este modo se establecen los objetivos de este trabajo, que se desglosarán a continuación, en los que se intentará aportar innovación en la temática citada.

II. Objetivos

2. OBJETIVOS

Después de ver en el apartado introductorio las noticias y estudios previos en los que se basa la temática del proyecto, se plantea el objetivo principal del trabajo que es modelizar mediante técnicas de regresión múltiple, el comportamiento de los estudiantes universitarios, medidos como el número de matriculados, en función de causas socioeconómicas globales, no analizadas directamente desde encuestas de opinión, como ocurre con la mayoría de los estudios encontrados.

2.1. OBJETIVOS SECUNDARIOS

En este apartado se listan los objetivos secundarios que surgen a partir del objetivo principal, y son los siguientes:

- El primer objetivo secundario de este proyecto será obtener más de un modelo válido en el que se determinen las causas del número de matriculados en carreras universitarias en Europa.
- Realizar las pruebas de validación del modelo o modelos planteados.
- Determinar cuáles de las variables seleccionadas para modelizar el número de matriculados universitarios europeos es significativa realmente, y además se determinará en tipo de influencia sobre la variabilidad del número de matriculados.
- Por último se pretende realizar predicciones con los modelos validados realizando a la finalización del proyecto, comparaciones con los datos reales obtenidos durante la realización del proyecto. Se pretende realizar predicciones cuyos intervalos de confianza contengan el valor medio real encontrado durante la realización del trabajo.

III. Metodología

3. METODOLOGÍA UTILIZADA EN EL PROYECTO

En este apartado se incluye toda la metodología utilizada para la recopilación de información, el planteamiento de las hipótesis principales en forma de formulación teórica de los modelos matemáticos, y la realización y validación de los cálculos en el proyecto.

3.1. BÚSQUEDA DE DATOS

En este estudio se van a proponer los factores que en hipótesis previa, influirán sobre la variable a analizar. Las variables dependientes inicialmente planteadas serán: la renta, ayudas percibidas, horas dedicadas a la cultura de los futuros estudiantes (niños), empleo de egresados, el PIB del país, indicadores de trabajo, número de universidades en el territorio, el nivel de desempleo y por último las tarifas, o gasto universitario por alumno exigido. La variable independiente (Y) a analizar será el número de matrículas universitarias. Los datos correspondientes se han obtenido de las bases de datos <http://ec.europa.eu/eurostat> y <http://datos.bancomundial.org/> [6]. La descripción detallada de cada una de las variables es la siguiente:

- **Número de matrículas:** Tasa bruta de matrícula, educación superior total (Nivel 5 y 6). Correspondiente al número total de estudiantes matriculados en educación superior, independientemente de su edad, expresado como porcentaje de la población total.
- **Renta:** Tasa neta de rentas expresadas en miles de euros. Correspondiente a la media entre el total de la población (miles de personas).
- **Ayudas:** Ayuda financiera a los estudiantes expresado en porcentaje del gasto público total en educación, en el nivel terciario de la educación.
- **Horas dedicadas a la cultura:** Expresado como el número medio de horas semanales desde los 3 años hasta la edad mínima de escolaridad obligatoria.
- **Empleo egresados:** Presenta las tasas de empleo de las personas de 20 a 34 años, después de haber alcanzado la educación superior como el más alto nivel de educación, sin haber recibido ningún tipo de educación o formación en las cuatro semanas anteriores a la encuesta. El indicador se calcula con base en datos de la Encuesta de Población Activa de la UE. Expresado en porcentaje.

III. Metodología

- **PIB:** Tasa de crecimiento porcentual del PIB a precios de mercado en moneda local, a precios constantes. Expresado en dólares americanos a precios constantes del año 2005. El PIB es la suma del valor agregado bruto de todos los productores residentes en la economía, más todo impuesto a los productos, menos todo subsidio no incluido en el valor de los productos. Se calcula sin hacer deducciones por depreciación de bienes manufacturados o por agotamiento y degradación de recursos naturales.
- **Poco trabajo:** Número de personas que viven en hogares con intensidad de trabajo muy bajas (población de 18 a 59 años). Expresado en porcentaje.
- **Número de universidades:** El número total incluye también instituciones que no son exactamente universidades (Escuelas de negocio, de arte, radio y televisión, música, politécnicos), por lo que las cifras de cada país suelen ser mayores que las cifras oficiales proporcionadas por los respectivos Ministerios de Educación. Expresado en unidades.
- **Gasto universitario por alumno:** El gasto público por estudiante es el gasto público corriente en educación, dividido por la cantidad total de estudiantes por nivel, como porcentaje del PIB per cápita. El gasto público (corriente y de capital) incluye el gasto del Gobierno en instituciones educativas (tanto públicas como privadas), administración educativa y subsidios para entidades privadas (estudiantes/hogares y otras entidades privadas). Expresado en porcentaje.
- **Desempleo:** Proporción de la población activa que no tiene trabajo pero que busca trabajo y está disponible para realizarlo. Expresado en porcentaje, desempleo total (porcentaje de la población activa total).

Con estas variables y la ayuda de un software de cálculo estadístico (*Statgraphics Centurión XVI.II*) se realizarán los cálculos de los coeficientes de los modelos de regresión múltiple planteados, así como su validación. Seguidamente se realizarán las predicciones de los modelos validados y se comprobará si coinciden con los resultados encontrados posteriormente (datos publicados del 2013) y, de esta manera, se verá si verdaderamente son modelos fiables.

3.2 TÉCNICAS DE ANÁLISIS

En este apartado se explicará todas las técnicas de análisis estadísticas empleadas a lo largo de la realización del en el proyecto, para la obtención y validación de los modelos matemáticos planteados.

3.2.1. ANÁLISIS BIVARIANTE

Los análisis de multicolinealidad se realizan para asegurar que no se da el fenómeno que ocurre cuando dos o más variables explicativas están relacionadas entre sí, y toman valores semejantes o proporcionales. Es entonces cuando se tendrán problemas de multicolinealidad.

Hay tres formas de encontrar problemas de multicolinealidad. La primera mediante la obtención de la matriz de correlaciones, dónde se obtienen los coeficientes de correlación de cada pareja de variables R_{ii} . El criterio establecido para identificar el problema será cuando se encuentren coeficientes de correlación con valores absolutos superiores a 0,7.

El segundo método de identificación de multicolinealidad, es mediante el cálculo de la matriz inversa de correlaciones (1). En este caso el criterio se tendrá en cuenta con los valores de la diagonal principal de la matriz, de forma que se dirá que existe multicolinealidad de una variable con el resto si $R_{ii}^{-1} > 10$.

$$R_{ii}^{-1} = \frac{1}{1 - R_{i\text{resto}}^2} \quad (1)$$

El último método para identificar multicolinealidad es mediante el cálculo del Índice de acondicionamiento (2) o número de condición de la matriz R . Este estadístico se basa en los autovalores de la matriz de correlación. El criterio para la detección de multicolinealidad en este caso será $IC > 10$.

$$I.C. = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \quad (2)$$

El problema se subsana eliminando la variable problemática, eligiendo entre ellas, la que menor ajuste (R^2) proporcione con su presencia en el modelo.

III. Metodología

3.2.2. ANÁLISIS MULTIVARIANTE

Para hablar del análisis multivariante se toma un conjunto de datos que se pueden clasificar en un subconjunto de variables Y, junto con otro subconjunto de variables X.

3.2.2.1. REGRESIÓN MÚLTIPLE LINEAL

La técnica de la regresión múltiple lineal se emplea para aproximar la relación de dependencia entre una variable dependiente Y, las variables independientes X_i . En la mayoría de las relaciones económicas intervienen más de dos variables. Donde:

- Y_t : Variable dependiente, explicada o regresando.
- $X_1, X_2 \dots X_i$: Variables explicativas, independientes o regresores.
- $\beta_1, \beta_2 \dots \beta_i$: Parámetros, miden la influencia que las variables explicativas tienen sobre el regrediendo.

La ecuación general para una regresión múltiple lineal queda como: $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + U$, donde "i" son números enteros positivos, y toma valores de 1 a ∞ .

3.2.2.2. VALIDACIÓN

En este apartado se explicarán los métodos que se utilizarán para verificar él o los modelos planteados, asegurando la robustez de los mismos.

3.2.2.2.1. SIGNIFICACIÓN

A continuación se va a describir el test con el que se determinará si el modelo y los parámetros escogidos para el modelo son significativos, basándose en el P-valor obtenido en cada caso. El criterio de aceptación para la H_0 será p-Valor superior a 0,05.

Si se rechaza H_0 y alguna β_i del modelo es distinta de 0, el modelo si es significativo (3). Lo mismo pasa con las variables del modelo y con la constante (4).

$$H_0: \beta_1 = \beta_2 = \dots \beta_i = 0$$

$$H_1: \beta_1 \neq \beta_2 \neq \dots \beta_i \neq 0 \quad (3)$$

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad (4)$$

III. Metodología

El problema está solucionado cuando se eliminan esas variables que no son significativas, y por tanto no deben estar presentes en el modelo. Aunque esta solución debe aplicarse sólo después de comprobar que no hay otros problemas presentes en el modelo, que pueden afectar a la estimación de los parámetros del modelo, pudiendo provocar falsos positivos o negativos en los test de significación.

3.2.2.2. ANÁLISIS GRÁFICO DE LOS RESIDUOS

Si los residuos (U) no están distribuidos de forma aleatoria y tienen de media el valor cero, puede significar que los parámetros se están calculando de forma errónea, y dar por significativas cosas que no lo son y viceversa, como se ha comentado en el apartado anterior. Por esto mismo, primero se analizan los gráficos para observar si existe algún problema. Se pueden observar los posibles problemas presentes un modelo de estas características de tres formas diferentes gráficamente (Figura III.1, Figura III.2, Figura III.3).

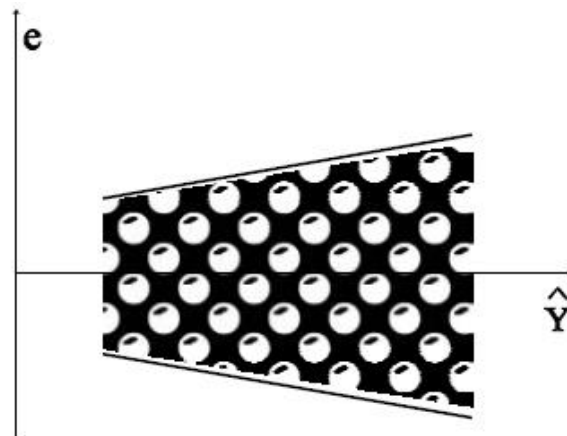


Figura III.1 Forma cónica, que puede indicar heterocedasticidad

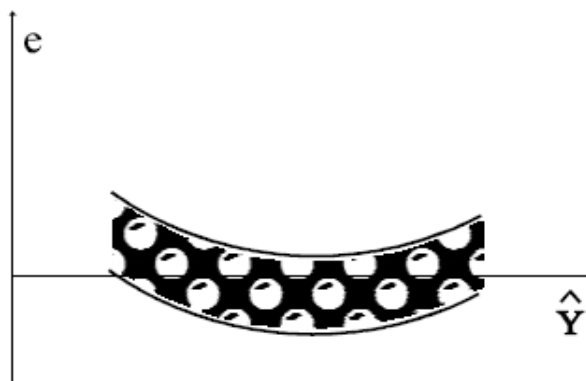


Figura III.2 Forma parabólica, que implica falta de linealidad

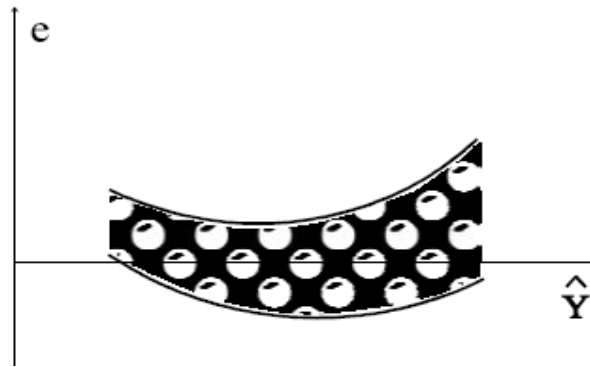


Figura III.3 Combinación de problemas de heterocedasticidad y falta de linealidad

El análisis gráfico se realiza principalmente en búsqueda de problemas de falta de linealidad, ya que éste fenómeno sólo puede detectarse de forma gráfica. Éste problema se puede solucionar elevando a un exponente la variable que crea este tipo de problema y, así, ver como los gráficos van perdiendo la forma problemática para el modelo. Los gráficos estarán divididos en tres maneras diferentes para observar si existe el problema:

- Residuos Vs X: Se analizarán todas las variables detectando desviaciones.
- Residuos Vs Y predicho: Mide la variabilidad de Y detectando desviaciones.
- Residuos Vs n^o de filas: Mide la variabilidad de las filas detectando desviaciones.

3.2.2.2.3. TEST DE NORMALIDAD

En este apartado se procede con el test (5) que determinará si los residuos siguen una distribución normal.

H_0 : Los residuos siguen una distribución normal

H_1 : Los residuos no siguen una distribución normal (5)

La hipótesis de normalidad de los residuos debe cumplirse porque de lo contrario, se pueden tener problemas en la estimación de los parámetros. Se realizarán todos los test estadísticos disponibles para asegurarse del cumplimiento de ésta hipótesis.

III. Metodología

3.2.2.2.4. CONTRASTE DE HETEROCEDASTICIDAD

Es un problema que aparece cuando la varianza de la perturbación no es constante. El ajuste de los residuos al cuadrado frente a las variables explicativas. El test es el siguiente (6):

H_0 : varianza= constante (no existe heterocedasticidad)

H_1 : varianza \neq constante (existe heterocedasticidad) (6)

Cuando los P-valores de las variables del modelo planteado sean superiores al nivel de significación (0,05) se acepta la hipótesis de homocedasticidad de los residuos. La ecuación general para obtener los P-valores requeridos para la realización de este análisis es: $RES^2 = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + U$

3.2.2.2.5. PUNTOS INFLUYENTES Y ANÓMALOS

Se observan cuando alguna de las perturbaciones se aleja mucho de la pauta marcada por el resto (Figura III.4). Los hay de distintas clases, y deben identificarse para poder explicar comportamientos inusuales en el modelo, y poder decidir si deben eliminarse en determinados casos. En este apartado se muestran las ecuaciones para la identificación de los distintos puntos influyentes y anómalos.

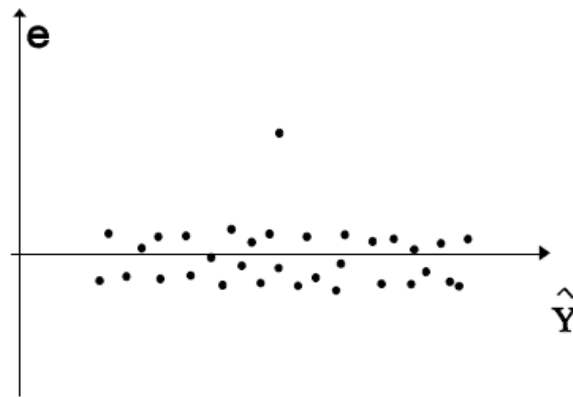


Figura III.4 Puntos anómalos

- Puntos influyentes a priori:

$$n_{ii} \geq 2 \frac{k+1}{n} \quad (7)$$

- Puntos influyentes a posteriori:

III. Metodología

$$|D_{(i)}| \geq 2 \sqrt{\frac{k+1}{n}} \quad (8)$$

- Valores de Y_i anómalos:

$$|rt_i| \geq 2 \text{ o bien } |rt_i| \geq 3 \quad (9)$$

Los puntos influyentes se dejarán identificados, al igual que aquellos anómalos con valores de los residuos estudentizados entre 2 y 3 en valor absoluto (9). Tan sólo se podrá plantear la eliminación de algún dato en caso de que se encuentren residuos con valores absolutos superiores a 3 en valor absoluto. Esto no implica la necesidad inherente de eliminar el dato correspondiente a esta clase de residuos, sino que en estos casos se puede plantear la eliminación, siempre y cuando se considere conveniente. La decisión queda supeditada a la calidad de la información perdida, que en este trabajo sería perder la información correspondiente a alguno de los países incluidos en el análisis, y de la mejora del ajuste del modelo planteado a los datos reales que se decidan mantener.

3.2.2.2.6 AUTOCORRELACIÓN

Hace referencia a que la perturbación en un momento dado, no debe depender de los valores tomados anteriormente, si ocurre esto, existen problemas de autocorrelación. Es un fenómeno es muy común cuando los datos recopilados se encuentran ordenados en el tiempo. En estos casos es mejor cambiar de estrategia y optar por otro tipo de modelización, como son por ejemplo, las series temporales.

El análisis del problema de autocorrelación se realiza de forma gráfica. La función de autocorrelación simple (FAS) es un conjunto de coeficientes de autocorrelación r_s , que mide la relación lineal entre el residuo en diferentes instantes e_t y e_{t-s} . La función de autocorrelación parcial (FAP) mide lo mismo que FAS, pero hace una pequeña corrección eliminando relaciones entre residuos intermedios. En ambos casos el análisis se realiza observando los factores de autocorrelación. Si en algún caso se superan los límites de autocorrelación representados, se podrá confirmar el problema de autocorrelación del orden en el que se encuentre el factor, y positiva o negativa, según se supere el límite superior o inferior, respectivamente (Figura III.5).

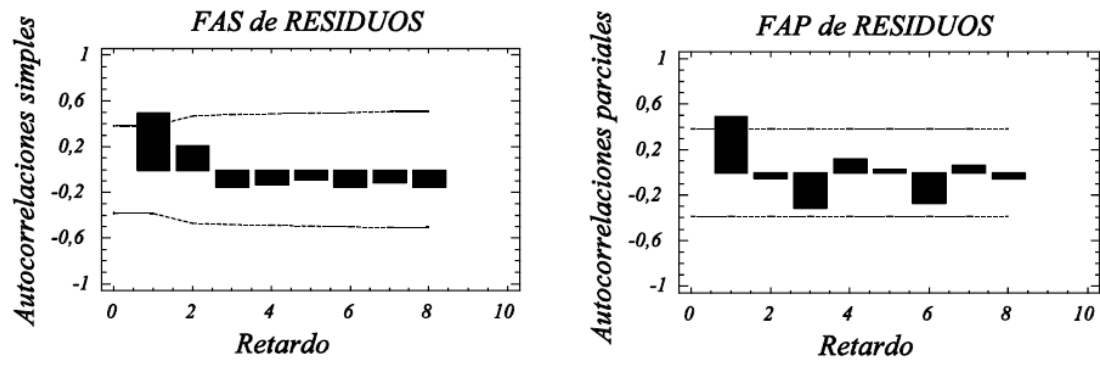


Figura III.5 Patrones de ejemplo de autocorrelación positiva de primer orden el FAS y el FAP.

3.2.2.3 ESTRUCTURAS

En este apartado se muestran las tres estructuras (10) que se han utilizado para los modelos de regresión múltiple planteados:

- Modelo Lineal: $Y = \beta_0 + \beta_1 X_1 \dots + \beta_n X_n + U$
- Logarítmica: $\text{Ln}Y = \beta_0 + \beta_0 + \beta_1 \text{Ln}X_1 \dots + \beta_n \text{Ln}X_n + U$
- Inversa: $Y = \beta_0 + \beta_1 1/X_1 \dots + \beta_2 1/X_n + U$ (10)

IV. Cálculos y resultados

4. CÁLCULOS Y RESULTADOS

En este apartado van a mostrar los resultado obtenidos para el planteamiento y validación de las diferentes estructuras planteadas, en el intento de modelización del comportamiento de los estudiantes universitarios en Europa.

4.1 BÚSQUEDA DE VARIABLES

En primer lugar se realizó la búsqueda de las variables que van a formar parte de los modelos matemáticos planteados. Para facilitar la interpretación y modificación de los parámetros de los modelos, codifican las diferentes variables según las leyendas expresadas entre paréntesis a continuación:

- X1: Renta (RENTA) en euros
- X2: PIB (PIB) en dólares
- X3: Horas cultura de niños (HORAS_CULTURA) en horas
- X4: Poco trabajo en hogares (POCO_TRABAJO) en horas
- X5: Ayuda para alumnos de nivel terciario (AYUDAS) en porcentaje
- X6: Desempleo (PARO) en porcentaje
- X7: Número de universidades en Europa (Nº_UNIVERSIDADES) en unidades
- X8: Gasto por alumno (GASTO_ALUMNO) en porcentaje
- X9: Empleo egresados (EMPLEO_EGRESADOS) en porcentaje
- Y: Número de matrículas (Nº_MATRICULAS) en porcentaje

Los datos que se consiguieron recopilar tras la búsqueda descrita en la metodología, se listan en la Tabla IV. 1. Comentar que no se encontraron datos más actuales a nivel Europeo en todas las variables que se pretendían incluir. Por este motivo los modelos planteados, se basan en los datos representados en el año 2011.

Durante la realización del proyecto estas bases de datos actualizaron los datos, de forma que se realizarán las predicciones pertinentes con los datos actualizados de las variables explicativas del modelo, y se compararán los datos reales obtenidos, con los predichos finalmente.

IV.Cálculos y resultados

Tabla IV. 1 Tabla datos países UE

Países/EUROPA 2011	Nº Matriculas (%)	Renta (mill €)	PIB (\$)	Poco trabajo (%)	Horas cultura (h)	Ayudas (%)	Paro (%)	Nº universidades (U)	Gasto alumno (%)	Empleo egresados (%)
Belgium	69	20.008	1,6	13,8	31,2	14,4	7,2	91,0	33,4	80,8
Bulgaria	60	2.914	2	10,2	23,9	18,3	11,3	58,0	16	59,2
Czech Republic	65	7.451	2	6,6	22,2	1,5	6,7	83,0	26	80,7
Denmark	77	26.394	1,2	13,0	33,6	28,4	7,6	90,0	51,3	83
Germany		19.043	3,6	11,4	25,0	21,9	5,8	407,0		88,3
Estonia	75	5.598	8,3	10,4	35,3	9,3	12,3	35,0	24,6	75,1
Ireland	73	19.726	2,8	23,6	17,8	13,3	14,7	49,0	29,6	70,9
Greece	114	10.985	-8,9	13,6	21,2		17,9	64,0		50,4
Spain	83	13.929	-0,6	14,1	24,6	9,4	21,4	236,0	27,2	67,1
France	57	19.995	2,1	9,8	28,7	8,0	9,2	619,0	36,4	77,3
Croatia	59	5.579	-0,3	16,7	17,6	5,5	13,7	25,0	25,6	62,7
Italy	64	15.972	0,6	11,3	31,5	22,2	8,4	212,0	24,2	57,7
Cyprus	47	16.990	0,3	5,5	22,9	52,6	7,9	17,0	39,2	72,5
Latvia	67	4.195	5	12,9	27,9	14,0	16,2	58,0	20	71,6
Lithuania	77	3.857	6,1	13,2	26,7	10,1	15,4	49,0	23,8	69,5
Hungary	60	4.535	1,8	11,7	26,0	12,4	11	76,0	28,4	73,4
Malta	40	10.862	1,4	8,6	20,1	18,4	6,4	4,0	40,1	91,4
Netherlands	76	20.310	1,7	9,8	18,5	28,8	5	148,0	34,4	91,5
Austria	71	21.463	3,1	9,1	21,5	9,8	4,6	77,0	35	90,5
Poland	74	5.025	4,8	8,2	15,1	12,7	9,7	448,0	20,6	75,3
Portugal	69	8.410	-1,8	8,6	31,3	15,4	12,9	110,0	26,9	75,8
Romania	52	2.116	2,3	7,4	14,0	9,3	7,2	111,0	19,7	70,8
Slovenia	85	11.999	0,6	8,6	32,7	23,4	8,2	39,0	25,8	76
Slovakia	55	6.306	2,7	7,8	25,5	16,7	13,7	33,0	22,2	70,1
Finland	96	21.826	2,6	11,1	32,6	13,7	7,8	49,0	36,3	78,4
Sweden	74	22.506	2,7	7,5	31,8	24,7	7,8	50,0	38,5	84,6
United Kingdom	61	17.136	1,6	10,6	18,5	39,2	8,1	313,0	32	81
Iceland	81	18.973	2,4	6,0	36,5	26,2	7,1	9,0	23,2	87,5
Norway	73	36.453	1	8,3	30,3	42,8	3,3	66,0	42,2	
Switzerland	54	33.936	1,8	4,8	14,7	2,2	4	107,0	39,9	85,3

4.2. MULTICOLINEALIDAD

Se va a realizar este análisis para asegurar que dos o más variables explicativas están relacionadas entre sí. La primera de las hipótesis que debe cumplir un modelo de regresión múltiple, es la de independencia de las variables explicativas, así que, tal y como se ha descrito en el apartado de metodología, antes de proponer las estructuras para la modelización, se procede con el análisis, y solución si procede, de los distintos tipos de multicolinealidad posibles.

4.2.1 IDENTIFICACIÓN

En este punto se verán los cálculos necesarios realizados en una hoja de cálculo de Excel para, de este modo, poder detectar algún problema de multicolinealidad. Como se ha comentado se realizarán tres pruebas.

IV.Cálculos y resultados

4.2.1.1 MATRIZ DE CORRELACIÓN (1º PRUEBA)

La primera prueba consiste en la evaluación de la correlación, entre las posibles combinaciones por parejas de las variables seleccionadas. Para esto se representa la que se denomina matriz de correlaciones (Tabla IV. 2).

Tabla IV. 2 Primera prueba

	RENTA	PIB	HORAS CULTURA	POCO TRABAJO	AYUDAS	DESEMPLEO	Nº UNIVERSIDADES	GASTO ALUMNO	EMPLEO EGRESADOS
RENTA	1	-0,243296	0,0995568	-0,0591717	0,229279	-0,499811	0,0965357	0,779402	0,542866
PIB	-0,243296	1	0,0855444	0,0321122	-0,244308	0,0837013	-0,0256876	-0,253131	0,0632218
HORAS CULTURA	0,0995568	0,0855444	1	-0,0361713	0,156756	0,0928993	-0,17745	0,0753392	0,00185083
POCO TRABAJO	-0,0591717	0,0321122	-0,0361713	1	-0,194208	0,558946	-0,0389717	-0,0542781	-0,374991
AYUDAS	0,229279	-0,244308	0,156756	-0,194208	1	-0,233216	-0,0881996	0,304674	0,128169
DESEMPLEO	-0,499811	0,0837013	0,0928993	0,558946	-0,233216	1	0,00415429	-0,467111	-0,638485
Nº UNIVERSIDADES	0,0965357	-0,0256876	-0,17745	-0,0389717	-0,0881996	0,00415429	1	0,0151382	-0,0441578
GASTO ALUMNO	0,779402	-0,253131	0,0753392	-0,0542781	0,304674	-0,467111	0,0151382	1	0,603987
EMPLEO EGRESADOS	0,542866	0,0632218	0,00185083	-0,374991	0,128169	-0,638485	-0,0441578	0,603987	1

Se observa que el valor entre las variables RENTA (X1) y GASTO ALUMNO(X8) es superior a 0,7 y que se relacionan en un 77,94%, por tanto, se puede concluir que entre ellas existe un problema de correlación (Tabla IV. 2). Se debe determinar cuál de ellas eliminar para subsanar dicho problema.

4.2.1.2 MATRIZ DE CORRELACIÓN INVERSA (2º PRUEBA)

En segundo lugar, se realiza el cálculo de la matriz inversa de correlaciones, para identificar la correlación del tipo, una variable con todas las demás, y además poder adquirir un criterio de eliminación, con las variables que han presentado problemas en el apartado anterior.

Tabla IV. 3 Segunda prueba

	RENTA	PIB	HORAS CULTURA	POCO TRABAJO	AYUDAS	DESEMPLEO	Nº UNIVERSIDADES	GASTO ALUMNO	EMPLEO EGRESADOS
RENTA	3,01658666	0,30447571	-0,36337567	-0,513193771	0,0627345	0,816411056	-0,353577413	-1,717497979	-0,313663492
PIB	0,30447571	1,27602293	-0,223284775	-0,218138085	0,1948581	0,066610825	-0,058205883	0,396670612	-0,551947426
HORAS CULTURA	-0,3633757	-0,2232848	1,181080668	0,284899262	-0,1743363	-0,432610357	0,242528328	-0,068429953	0,114199048
POCO TRABAJO	-0,5131938	-0,2181381	0,284899262	1,851919521	0,2186315	-1,136392933	0,222672259	-0,50925977	0,550139756
AYUDAS	0,06273451	0,19485809	-0,174336318	0,218631532	1,247228	0,149218998	0,102747796	-0,427986883	0,234385565
DESEMPLEO	0,81641106	0,06661083	-0,432610357	-1,136392933	0,149219	2,612490985	-0,166141728	0,101503459	0,707518339
Nº UNIVERSIDADES	-0,3535774	-0,0582059	0,242528328	0,222672259	0,1027478	-0,166141728	1,103075988	0,005254065	0,204964189
GASTO ALUMNO	-1,717498	0,39667061	-0,068429953	-0,50925977	-0,4279869	0,101503459	0,005254065	3,288937983	-1,150128841
EMPLEO EGRESADOS	-0,3136635	-0,5519474	0,114199048	0,550139756	0,2343856	0,707518339	0,204964189	-1,150128841	2,536670969

La matriz inversa en principio indica que no existe relación entre ninguna de las variables con el resto, puesto que ningún valor de la diagonal principal es mayor a 10. No puede tomar una decisión directa sobre qué variable de las anteriores eliminar, ya que las dos que generaban el problema entre ellas no

IV.Cálculos y resultados

difieren en más de una unidad, en el valor de la diagonal de matriz inversa (Tabla IV. 3). Por tanto, el criterio de eliminación deberá ser mediante comprobación de ajuste del modelo, en función de la presencia de una u otra variable (RENTA o GASTO ALUMNO).

4.2.1.2 ÍNDICE DE ACONDICIONAMIENTO (3ª PRUEBA)

La última comprobación de multicolinealidad, se realiza mediante el cálculo del índice de acondicionamiento. Como puede verse, el índice de correlación se sitúa entre 1 y 10 (4,018). Decir que el valor señalado en amarillo en la tabla, es el resultante del cálculo del índice como se ha indicado en el apartado de metodología (2). Según este resultado, no existe correlación entre las combinaciones de más de dos de las variables seleccionadas (Tabla IV. 4).

Tabla IV. 4 Índice de correlación

IC	
3,07368	4,01848185
1,3252	
1,10128	
1,08556	
0,64449	
0,325366	
0,25409	
0,190342	

4.2.2 SOLUCIÓN DE MULTICOLINEALIDAD

En este apartado se va a solucionar el problema de multicolinealidad entre dos variables, es decir, la relación entre sí de las variables RENTA (X1) y GASTO ALUMNO (X8).

4.2.2.1 REGRESIÓN MÚLTIPLE

A fin de determinar cuál es la variable a eliminar se observa el ajuste, mediante el valor del R-cuadrado resultante tras eliminar la variable X1. Posteriormente se prueba el modelo eliminando la variable X8 y se observa nuevamente el R-cuadrado. Se elimina la variable que proporcione un modelo ajustado propuesto con menor R-cuadrado pues, ya que sería el modelo propuesto menos preciso a la hora de realizar predicciones. Los resultados obtenidos de las pruebas fueron los siguientes:

- **Sin eliminar ninguna variable:** R-cuadrada = 52,0963 % y R-cuadrado (ajustado para g.l.) = 26,7356 %.
- **Eliminando la variable Gasto por Alumno (X8):** R-cuadrada = 41,3759 % y R-cuadrado (ajustado para g.l.) = 15,3208 %.
- **Eliminando la variable Renta (X1):** R-cuadrada = 50,9223 % y R-cuadrado (ajustado para g.l.) = 27,8269 %.

A la vista de los resultados se decide eliminar X1 (RENTA) puesto que al quitar dicha variable el modelo explica mejor la variabilidad del Número de Matrículas (Y) (43,3865% quitando X1 frente al 50,9223% quitando X8). Comentar además que aunque se aparenta un empeoramiento del ajuste del modelo inicial, que al incluir todas las variables tiene un ajuste del 52,0963%, no es real, puesto que para comparar estructuras con diferente número de variables explicativas, se debe realizar la comparación del R-cuadrado ajustado. En este caso, si se observa que se pasa del 26,7357% al 27,8269%. Mejora el ajuste con la solución de la multicolinealidad, pero aún queda mucho trabajo para conseguir un modelo válido, ya que el ajuste está muy por debajo de lo que se empezaría a considerar aceptable.

IV.Cálculos y resultados

4.2.3 COMPROBACIÓN

Ante el problema de multicolinealidad detectado anteriormente se ha eliminado la variable RENTA y, para ver que el problema ha sido solucionado, se vuelve a repetir las tres pruebas realizadas en el apartado anterior.

4.2.2.1 MATRIZ DE CORRELACIÓN (1ª PRUEBA)

En la nueva matriz de correlación calculada, se observa que ahora no existe ningún valor que supere el 0,7 en valor absoluto, y por tanto esto indica que los problemas de correlación han sido subsanados.

Tabla IV. 5 Primera prueba comprobación problemas de multicolinealidad

	PIB	POCO TRABAJO	HORAS CULTURA	AYUDAS	DESEMPLEO	Nº UNIVERSIDADES	GASTO ALUMNO	EMPLEO EGRESADOS
PIB	1	0,0321122	0,0855444	-0,244308	0,0837013	-0,0256876	-0,253131	0,0632218
POCO TRABAJO	0,0321122	1	-0,0361713	-0,194208	0,558946	-0,0389717	-0,0542781	-0,374991
HORAS CULTURA	0,0855444	-0,0361713	1	0,156756	0,0928993	-0,17745	0,0753392	0,00185083
AYUDAS	-0,244308	-0,194208	0,156756	1	-0,233216	-0,0881996	0,304674	0,128169
DESEMPLEO	0,0837013	0,558946	0,0928993	-0,233216	1	0,00415429	-0,467111	-0,638485
Nº UNIVERSIDADES	-0,0256876	-0,0389717	-0,17745	-0,0881996	0,00415429	1	0,0151382	-0,0441578
GASTO ALUMNO	-0,253131	-0,0542781	0,0753392	0,304674	-0,467111	0,0151382	1	0,603987
EMPLEO EGRESADO	0,0632218	-0,374991	0,00185083	0,128169	-0,638485	-0,0441578	0,603987	1

4.2.2.2 MATRIZ DE CORRELACIÓN INVERSA (2ª PRUEBA)

Aun así se debe corroborar mediante los otros dos análisis, de forma que se representa seguidamente la matriz inversa (Tabla IV. 6). Se observa que ningún valor de la diagonal principal es mayor a 10, por tanto, siguen sin existir los problemas de multicolinealidad del tipo una variable frente a todas las demás.

Tabla IV. 6 Segunda prueba comprobación problemas de multicolinealidad

	PIB	POCO TRABAJO	HORAS CULTURA	AYUDAS	DESEMPLEO	Nº UNIVERSIDADES	GASTO ALUMNO	EMPLEO EGRESADOS
PIB	1,245291019	-0,16633946	-0,186607869	0,18852606	-0,01579269	-0,022517953	0,5700243	-0,520288162
POCO TRABAJO	-0,16633946	1,764612947	0,223080342	0,22930418	-0,99750182	0,162520258	-0,801447386	0,49677807
HORAS CULTURA	-0,186607869	0,223080342	1,137308719	-0,1667794	-0,33426612	0,19993667	-0,275318416	0,076415389
AYUDAS	0,188526055	0,229304177	-0,166779368	1,24592337	0,13224049	0,110100976	-0,392268899	0,240908675
DESEMPLEO	-0,015792688	-0,997501824	-0,33426612	0,13224049	1,239153694	-0,070449298	0,566328274	0,792408439
Nº UNIVERSIDADES	-0,022517953	0,162520258	0,19993667	0,11010098	-0,0704493	1,061632794	-0,196055747	0,168199349
GASTO ALUMNO	0,5700243	-0,801447386	-0,275318416	-0,3922689	0,56632827	-0,196055747	1,2311078025	-1,328713604
EMPLEO EGRESADO	-0,520288162	0,49677807	0,076415389	0,24090868	0,79240844	0,168199349	-1,328713604	1,2504056363

IV.Cálculos y resultados

4.2.2.3 ÍNDICE DE ACONDICIONAMIENTO (3º PRUEBA)

La última prueba para corroborar la ausencia de multicolinealidad es la del cálculo del índice de acondicionamiento, que como se observa en la Tabla IV.7, con los nuevos autovalores obtenidos, el valor del IC sigue siendo inferior a 10, quedando confirmada la correcta solución del primer problema detectado.

Tabla IV. 7 Tercera prueba IC

IC	
2,53916	1,59347419
1,28653	
1,20666	
0,956583	
0,856359	
0,638162	
0,292429	
0,224123	

4.3. MODELO LINEAL

Con todo, ya se puede proceder a la estimación (11) y validación del primer modelo propuesto para explicar el número de matriculados universitarios en Europa, con las variables explicativas independientes.

Modelo resultante tras solucionar el problema de multicolinealidad detectado:

$$\begin{aligned} \text{Nº_MATRICULAS} = & 14.8927 + 1.0706 * \text{PIB} + 1.68063 * \text{POCO_TRABAJO} + 1.08093 * \\ & * \text{HORAS_CULTURA} + 0.0283943 * \text{AYUDAS} - 1.31358 * \text{PARO} + 0.00277473 * \text{Nº_UNIVERSIDADES} - \\ & - 0.595986 * \text{GASTO_ALUMNO} + 0.445569 * \text{EMPLEO_EGRESADOS} \end{aligned} \quad (11)$$

Así que tal y como se ha comentado en el apartado de

IV.Cálculos y resultados

3. METODOLOGÍA utilizada en el proyecto se procede con el análisis y validación del modelo con estructura lineal.

4.3.1. SIGNIFICACIÓN

En primer lugar, se procede a analizar, si tanto las variables como el modelo, son significativos para el modelo.

Tabla IV. 8 Prueba de significatividad

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	14,8927	31,5902	0,471435	0,6433
PIB	1,0706	1,22317	0,875272	0,3936
POCO_TRABAJO	1,68063	0,732655	2,29388	0,0348
HORAS_CULTURA	1,08093	0,327399	3,30157	0,0042
AYUDAS	0,0283943	0,207425	0,13689	0,8927
PARO	-1,31358	1,0733	-1,22388	0,2377
Nº_UNIVERSIDADES	0,00277473	0,0153544	0,180712	0,8587
GASTO_ALUMNO	-0,595986	0,391653	-1,52172	0,1465
EMPLEO_EGRESADOS	0,445569	0,369817	1,20484	0,2448

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>de Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	1956,12	8	244,515	2,20	0,0812
Residuo	1885,26	17	110,898		
Total (Corr.)	3841,38	25			

R-cuadrada = 50,9223 por ciento

R-cuadrado (ajustado para g.l.) = 27,8269 por ciento

Las únicas variables significativas del modelo son POCO_TRABAJO y HORAS_CULTURA. La falta de ajuste es muy grande, ni siquiera el modelo es significativo. Como la constante no es significativa se

IV.Cálculos y resultados

eliminará, puesto que por experiencia, este fenómeno suele producir la falta de ajuste observada, y se esperará una gran mejora en el modelo. Aún así, se continúa con los siguientes análisis, para detectar otros posibles problemas, que estén causando este mal ajuste inicial, y que a lo sumo pudieran estar afectando a la estimación de los parámetros del modelo, y por tanto a su significatividad

4.3.2. FALTA DE LINEALIDAD

En esta parte del análisis se analizarán los gráficos de residuos posibles, y, con ellos principalmente se determinará si existe algún problema de falta de linealidad en el modelo.

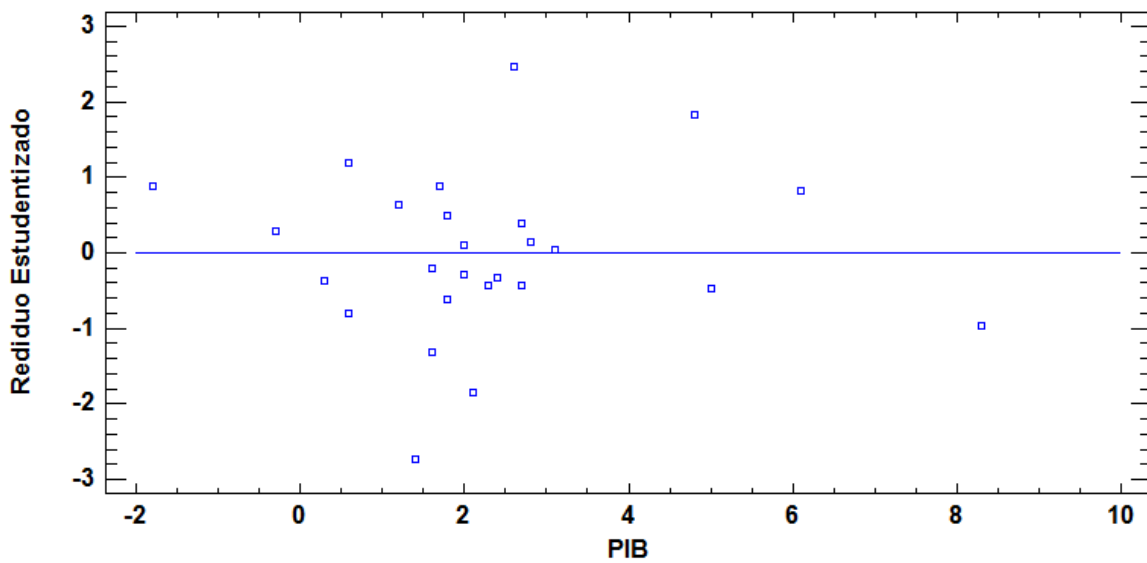


Figura IV. 1 PIB

IV.Cálculos y resultados

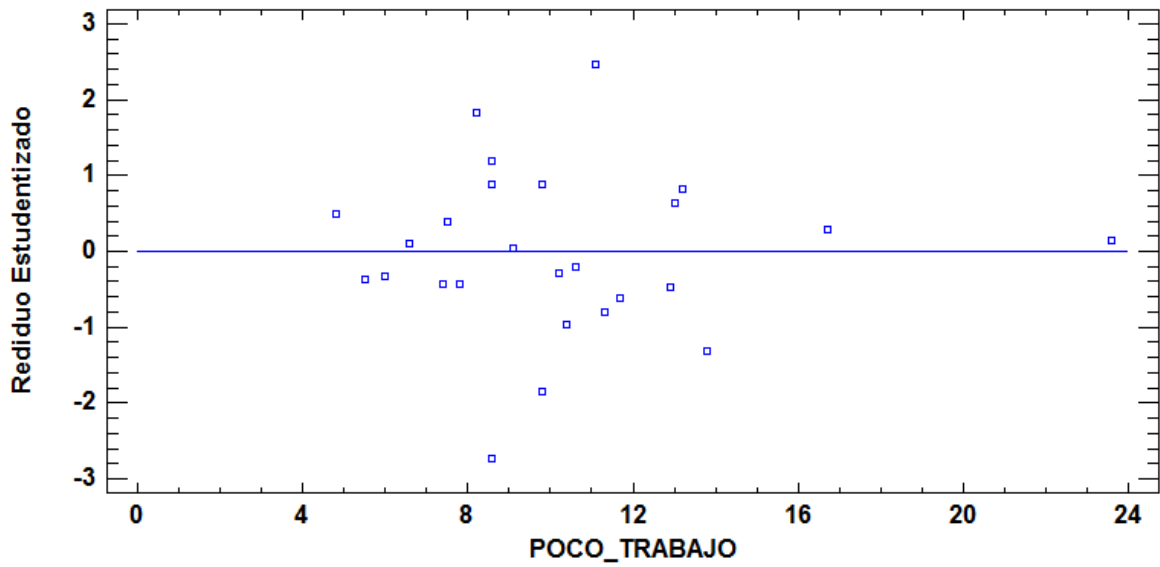


Figura IV. 2 Poco trabajo

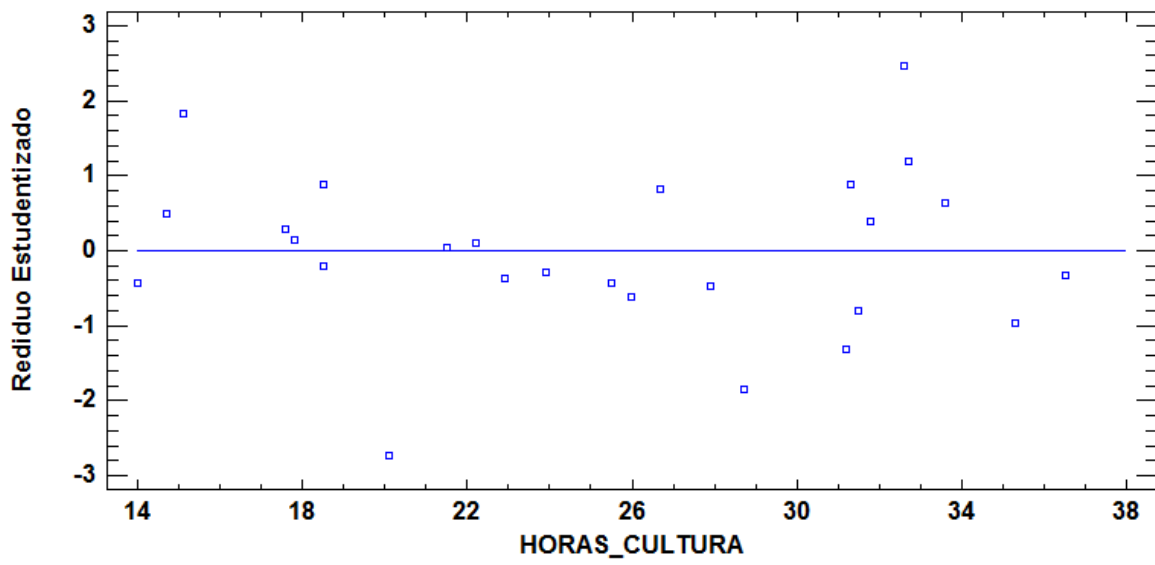


Figura IV. 3 Horas cultura niños

IV.Cálculos y resultados

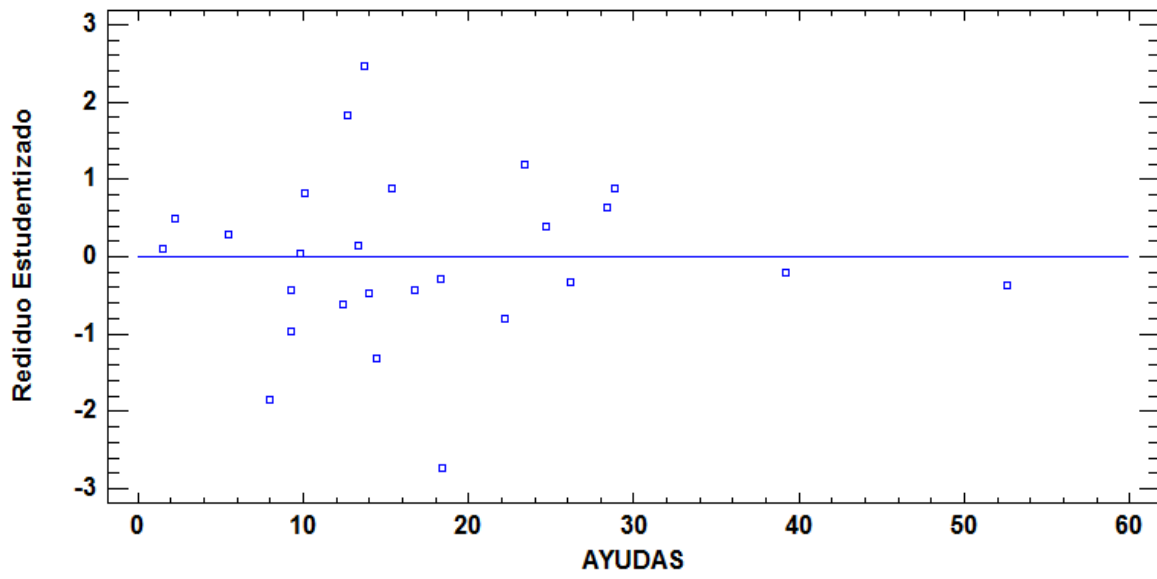


Figura IV. 4 Ayudas

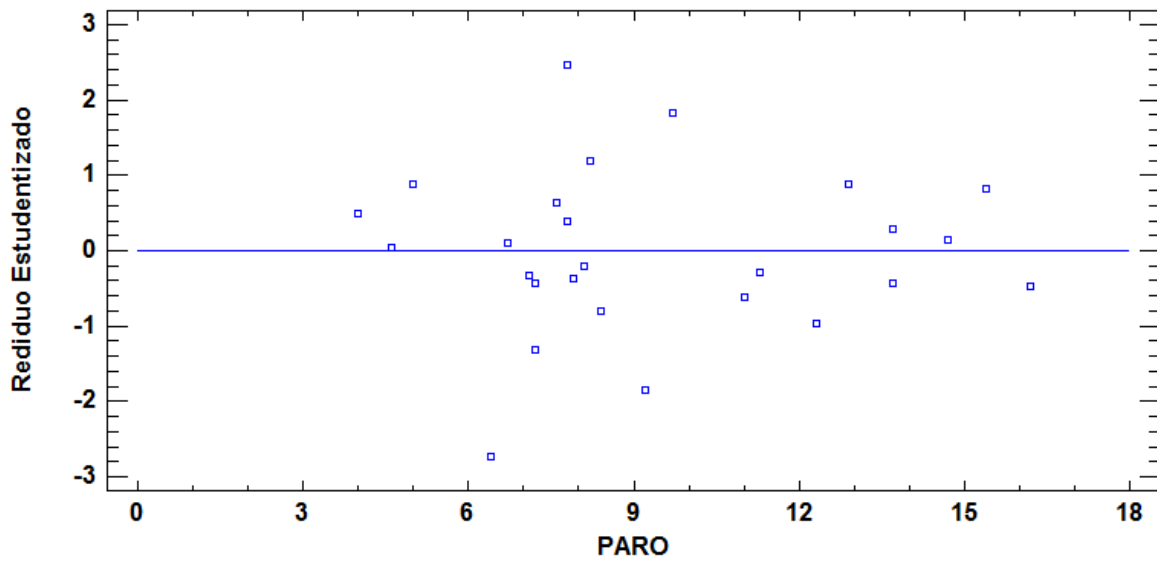


Figura IV. 5 Desempleo

IV.Cálculos y resultados

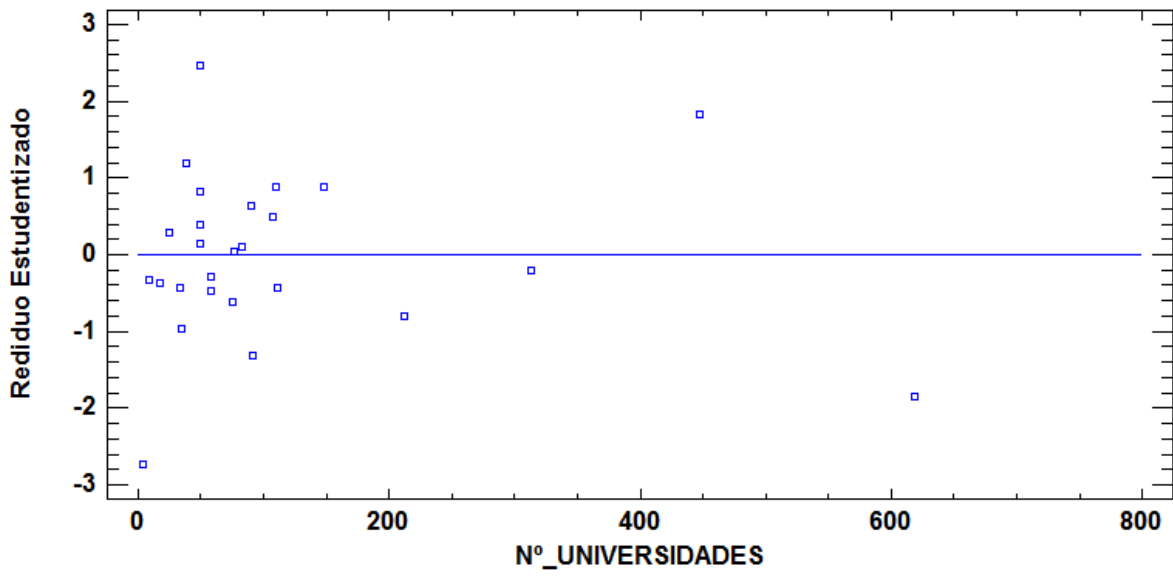


Figura IV. 6 Número de universidades

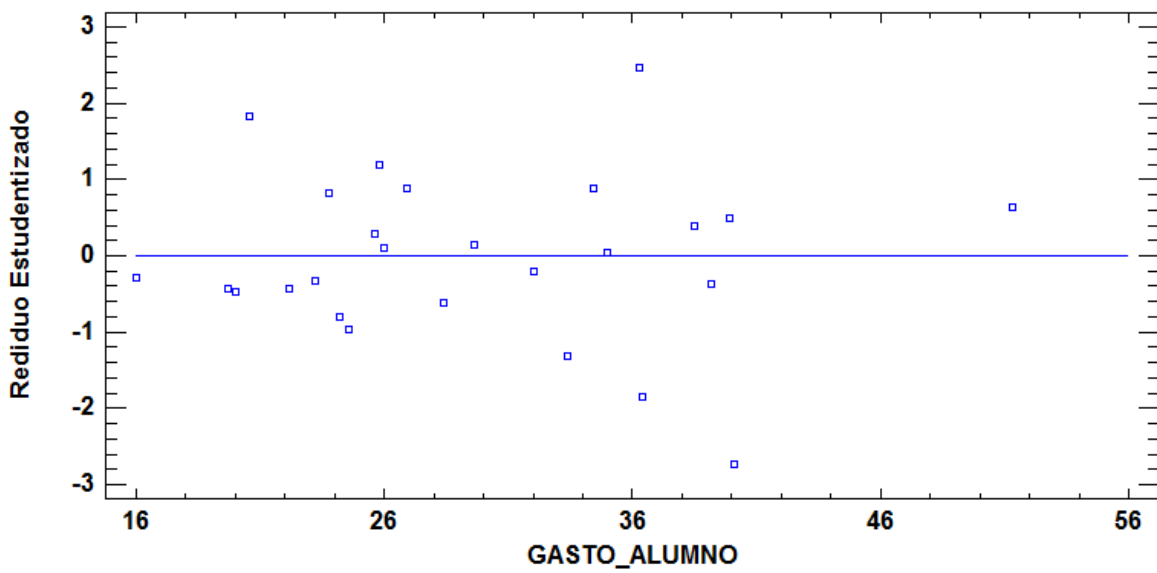


Figura IV. 7 Gasto por alumno

IV.Cálculos y resultados

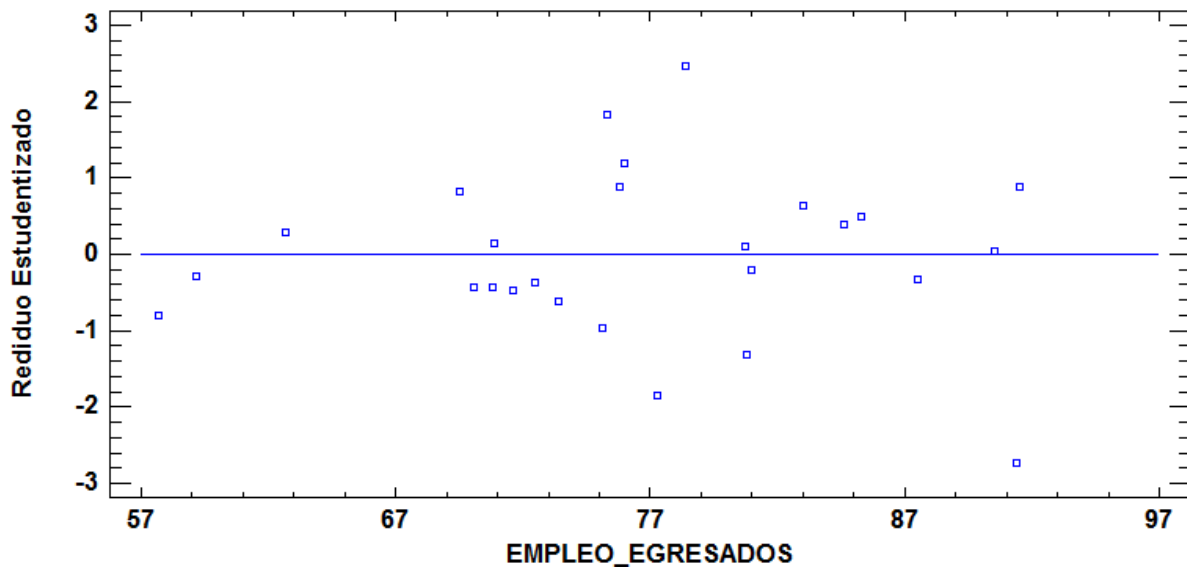


Figura IV. 8 Empleo egresados

En el gráfico de HORAS_CULTURA (Figura IV. 3) de residuos frente a predicho se puede apreciar cierta forma parabólica que podría estar indicando que posiblemente haya falta de linealidad. Por las demás variables (Figura IV. 1, 2, 4, 5, 6, 7,8) no se ve ningún tipo de problema. Con ello, para corroborar el problema, y solucionarlo si es el caso, se elevará la variable de HORAS_CULTURA a un exponente (potencias de 2) hasta que se vea una solución o el R-cuadrado empiece a disminuir. Los pasos seguidos y los resultados han sido:

- R-cuadrado sin modificar la variable HORAS_CULTURA: R-cuadrada = 50.9223 %
- R-cuadrado modificando la variable (HORAS_CULTURA²): R-cuadrada = 52,6602 %
- R-cuadrado modificando la variable (HORAS_CULTURA⁴): R-cuadrada = 52,9762 %
- R-cuadrado modificando la variable (HORAS_CULTURA⁸): R-cuadrada = 46,8356 %

Se observa que a partir de la potencia 8 el ajuste o R-cuadrado del modelo baja y, de este modo, se deja la solución como: HORAS_CULTURA⁴. Mediante este método se comprueba que había una evidente falta de linealidad debida a esta variable, y se consigue mejorar ligeramente el ajuste del modelo a los datos reales.

El nuevo gráfico de residuos frente a la variable en cuestión, quedará modificado como se observa en la Figura IV. 9. Los puntos quedan más dispersos y el gráfico no tiene tanta forma parabólica, por lo que se ve que el problema está subsanado.

IV. Cálculos y resultados

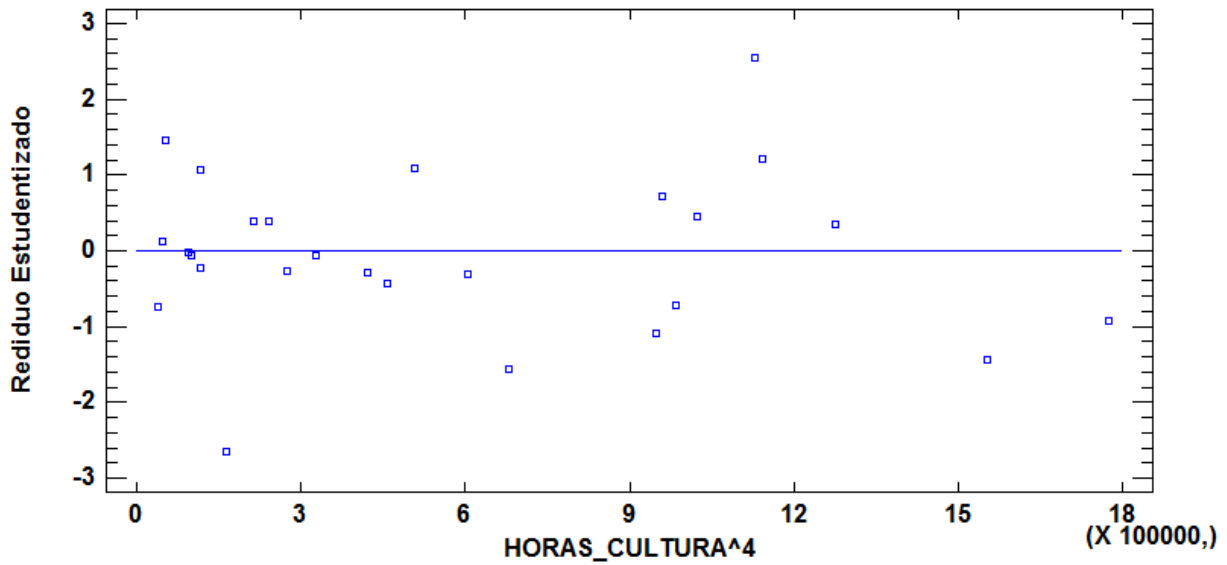


Figura IV. 9 Horas cultura elevado a 4

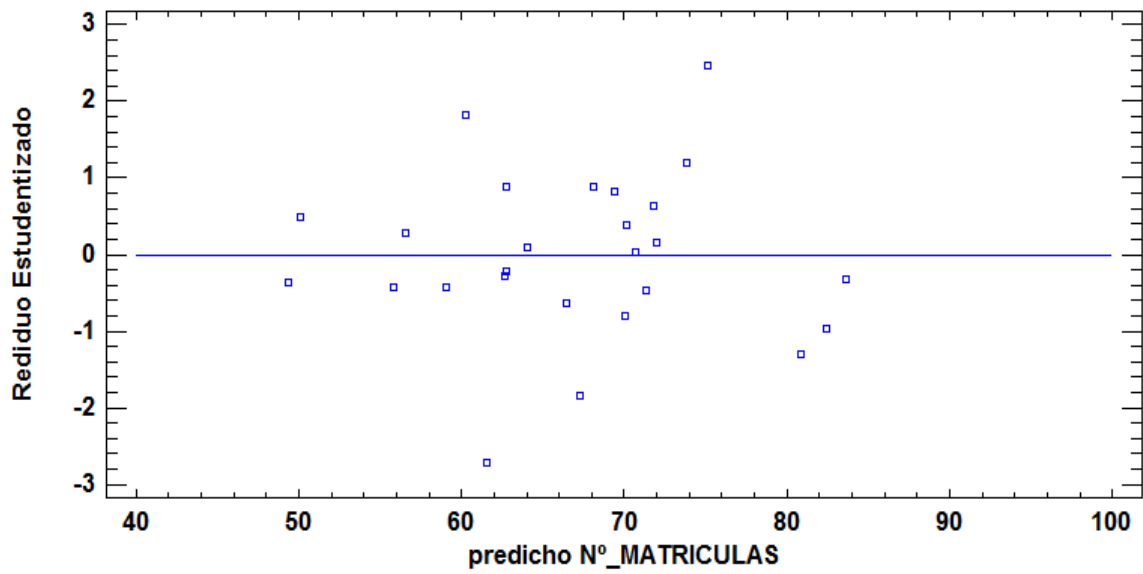


Figura IV. 10 Vs Y predicho

IV.Cálculos y resultados

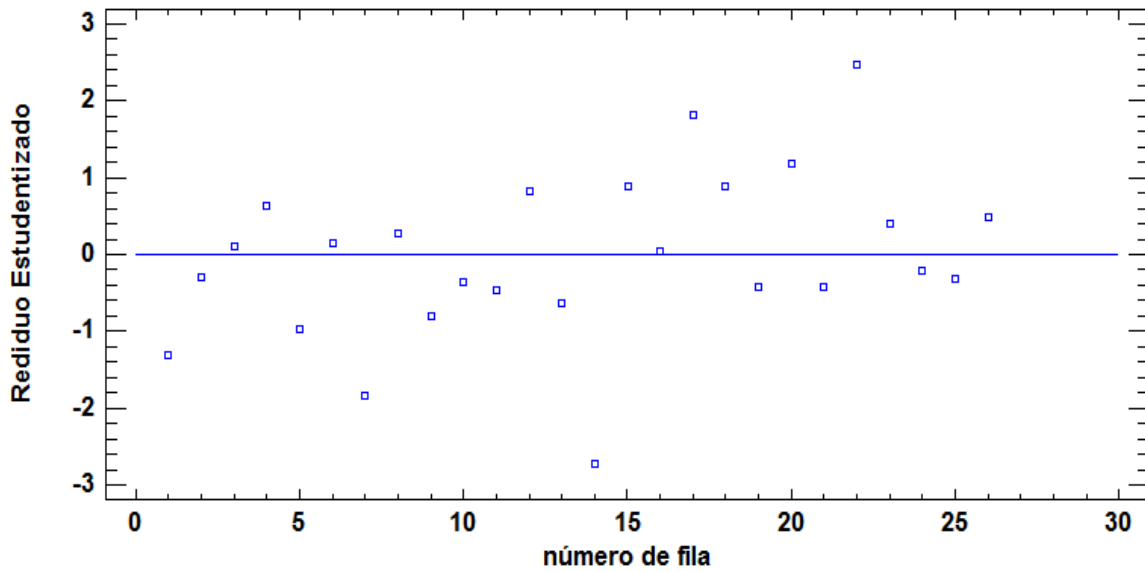


Figura IV. 11 Vs número de filas

En el resto de los gráficos obtenidos, no se observa ninguna forma de parabólica ni cónica, por lo que a priori, no existe ningún otro tipo de problema en el modelo (Figura IV. 10 y Figura IV. 11).

Tabla IV. 9 Resolución significatividad con HORAS_CULTURA⁴

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	31,7658	30,4406	1,04353	0,3113
PIB	0,839972	1,20551	0,696779	0,4954
POCO_TRABAJO	1,56529	0,711609	2,19965	0,0420
HORAS_CULTURA ⁴	0,0000148157	0,00000425588	3,48123	0,0029
AYUDAS	0,0270599	0,202987	0,133309	0,8955
PARO	-0,932597	1,03264	-0,903118	0,3791
Nº_UNIVERSIDADES	0,00349015	0,0150512	0,231885	0,8194
GASTO_ALUMNO	-0,49814	0,379038	-1,31422	0,2062
EMPLEO_EGRESADOS	0,405933	0,361916	1,12162	0,2776

IV.Cálculos y resultados

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>de Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	2035,02	8	254,377	2,39	0,0620
Residuo	1806,37	17	106,257		
Total (Corr.)	3841,38	25			

R-cuadrada = 52,9762 por ciento

R-cuadrado (ajustado para g.l.) = 30,8473 por ciento

Con la variable HORAS_CULTURA⁴ se observa que el R-cuadrado aumenta, y que el problema de falta de linealidad en esta variable se ha solucionado. Aunque el modelo sigue sin ser significativo junto con algunas variables.

4.4.3. NORMALIDAD

Seguidamente se procederá a determinar si los residuos se distribuyen de forma normal en este apartado.

Se opta por realizar el análisis en función de los test de Normalidad que pueden realizarse, con el software empleado. Como puede observarse en la Tabla IV. 10 Los P-valor de las cuatro pruebas son mayores a 0,05 y por lo tanto, no hay nada que nos indique el problema de falta de normalidad.

Tabla IV.10 Pruebas de normalidad para el primer modelo lineal

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	11,6923	0,38721
Estadístico W de Shapiro-Wilk	0,98221	0,913234
Valor-Z para asimetría	0,207192	0,835856
Valor-Z para curtosis	1,17919	0,238321

IV.Cálculos y resultados

4.4.4. HETEROCEDASTICIDAD

En este apartado se verá si la varianza de la perturbación es constante realizando un modelo con los residuos al cuadrado y las variables explicativas. Confirmar que los residuos analizados en cuestión, han sido los generados por el modelo con la solución de falta de linealidad detectada, y que estos al cuadrado se representan en función de la variable HORAS_CULTURA, con la corrección aplicada HORAS_CULTURA ⁴.

Los P-valor de la Tabla IV. 11 son superiores a 0,05, por lo tanto, no hay problema de heterocedasticidad generados por ninguna variable. Este problema no es el causante de la falta de ajuste en el modelo lineal que se está analizando. Se procederá con las siguientes pruebas de validación.

Tabla IV. 11 Prueba de heterocedasticidad

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	-129,753	383,812	-0,338063	0,7395
PIB	6,9499	14,8611	0,467657	0,6460
POCO_TRABAJO	1,87952	8,90154	0,211145	0,8353
HORAS_CULTURA	2,36932	3,9778	0,595637	0,5593
AYUDAS	-1,33658	2,52015	-0,530359	0,6027
PARO	-5,89968	13,0402	-0,452422	0,6567
Nº_UNIVERSIDADES	-0,0382324	0,186552	-0,204942	0,8401
GASTO_ALUMNO	1,31952	4,75846	0,2773	0,7849
EMPLEO_EGRESADOS	1,92798	4,49317	0,429092	0,6732

4.4.5. AUTOCORRELACIÓN

En este apartado se van a representar los factores de autocorrelación calculados al relacionar los residuos con sus mismos datos anteriores en diferentes grados, en busca de autocorrelación.

Autocorrelaciones Estimadas para RESIDUOS

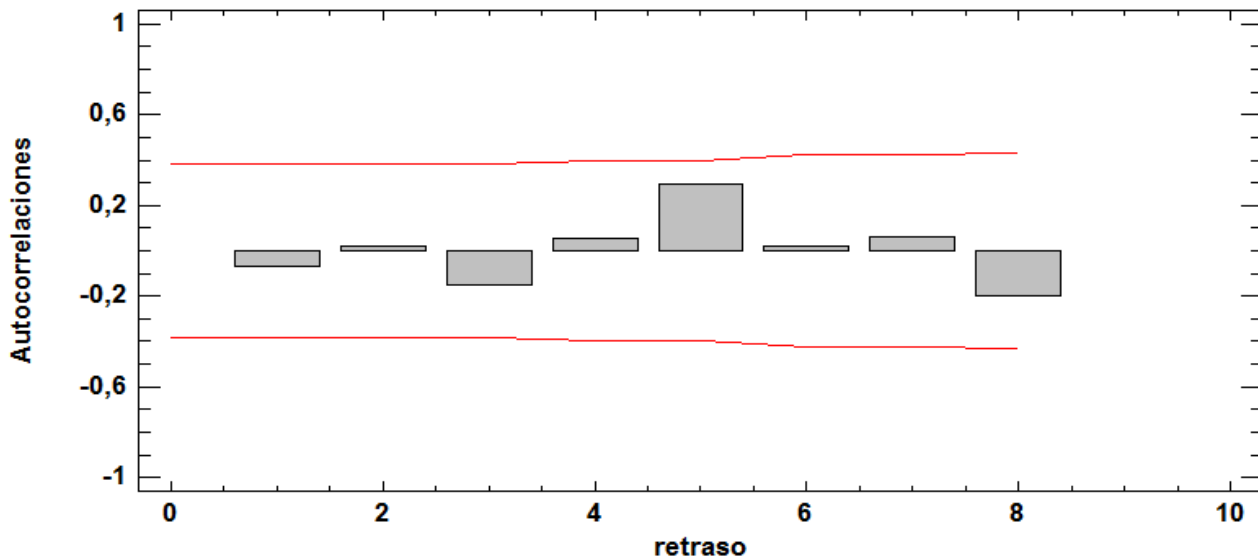


Figura IV. 12 FAS

Autocorrelaciones Parciales Estimadas para RESIDUOS

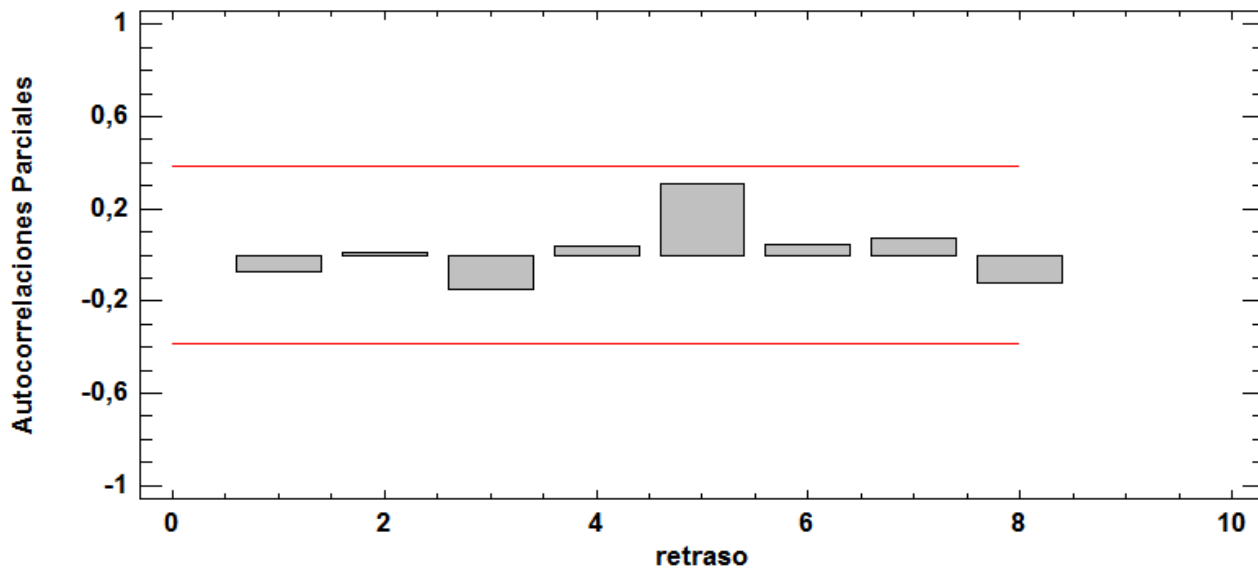


Figura IV. 13 FAP

No se detectan problemas de autocorrelación ni positiva ni negativa en ninguno de los dos gráficos (Figura IV. 12 y Figura IV. 13) ya que los factores de autocorrelación no sobrepasan los límites establecidos.

4.5. MODELO LINEAL SIN CONSTANTE

Después de realizar todas las pruebas requeridas y ver los problemas del modelo, se eliminará la constante y algunas variables que no son significativas para el modelo, puesto que tras realizar las únicas modificaciones posibles, siguen resultando no significativas. Es en este momento cuando es correcto proceder con la eliminación del parámetro que corresponde a la constante, así como de las variables que no han resultado ser significativas. El orden de acción será eliminar en primer lugar la constante, y tras comprobar los efectos en el ajuste del modelo y la significatividad del resto de las variables, ya se procederá con la eliminación si procede de algunas de ellas.

4.5.1. SIGNIFICATIVIDAD

El cambio y mejora pronosticada después del primer análisis de significatividad se cumple, ya que como puede verse en la Tabla IV.12, por fin se consigue que con esto el modelo también sea significativo. Las variables que finalmente quedan en el modelo al ser significativas serán POCO_TRABAJO, HORAS_CULTURA y EMPLEO_EGRESADOS. A este efecto se le suma el considerable aumento del R-cuadrado alcanza un valor hasta el 98.4087% de la variabilidad de la Emigración explicada, por lo tanto, el ajuste del modelo a los datos reales es muy superior al anterior con todos los parámetros iniciales.

Tabla IV. 12 Prueba de significatividad sin constante

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
POCO_TRABAJO	1,72467	0,710801	2,42638	0,0260
HORAS_CULTURA	1,10815	0,315228	3,51539	0,0025
EMPLEO_EGRESADOS	0,597597	0,177075	3,37482	0,0034

IV.Cálculos y resultados

PIB	0,929585	1,16012	0,801283	0,4334
AYUDAS	0,0410479	0,201188	0,204027	0,8406
PARO	-1,02485	0,862143	-1,18872	0,2500
Nº_UNIVERSIDADES	0,00449686	0,0145878	0,308261	0,7614
GASTO_ALUMNO	-0,621781	0,379341	-1,63911	0,1186

Análisis de Varianza

Fuente	Suma de Cuadrados	de Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	118110	8	14763,8	139,14	0,0000
Residuo	1909,91	18	106,106		
Total	120020	26			

R-cuadrada = 98,4087 por ciento

R-cuadrado (ajustado para g.l.) = 97,7898 por ciento

4.5.2. FALTA DE LINEALIDAD

Seguidamente, deben realizarse todas las comprobaciones anteriores con el nuevo modelo planteado, para determinar si la solución planteada produce algún efecto negativo, por lo que deberían tomarse otras medidas para conseguir una modelización correcta. Igual que en el caso anterior se empieza mediante el análisis gráfico, en búsqueda de la posible falta de linealidad.

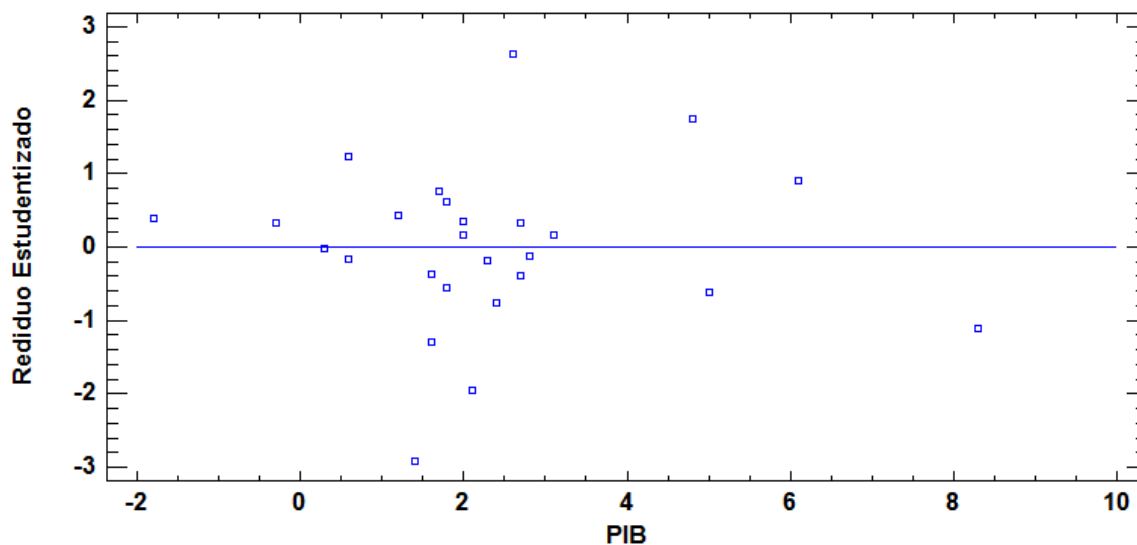


Figura IV. 14 PIB sin constante

IV.Cálculos y resultados

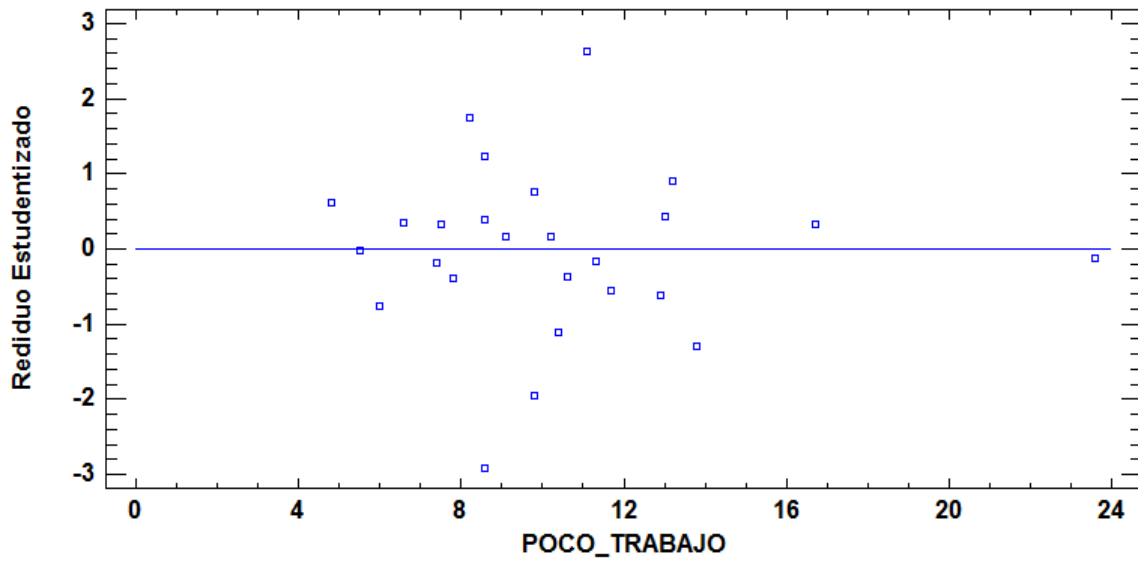


Figura IV. 15 Poco trabajo sin constante

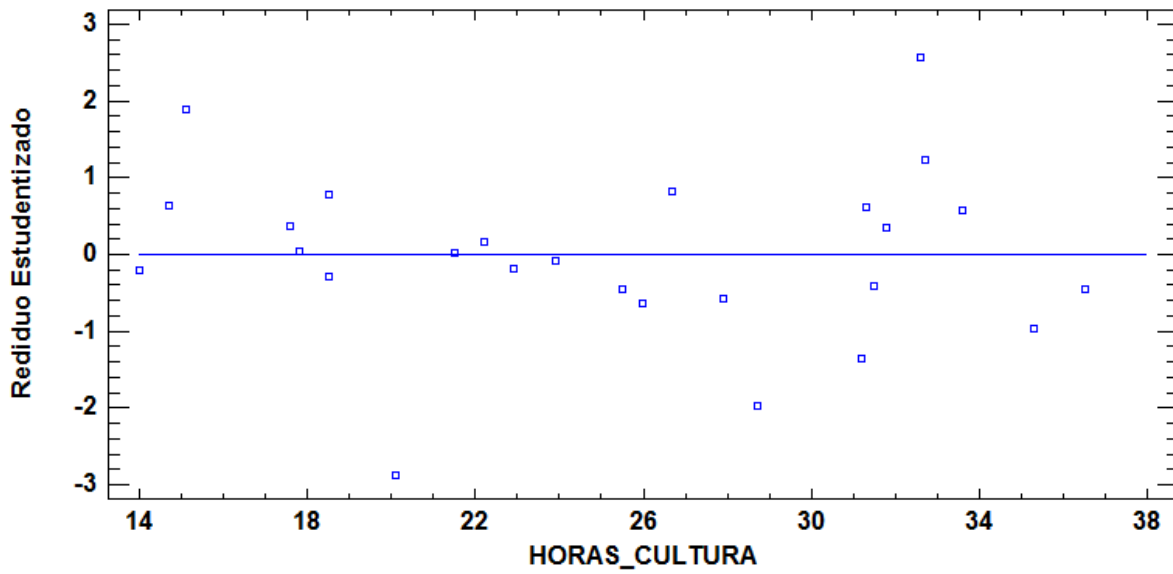


Figura IV. 16 Horas cultura niños sin constante

IV.Cálculos y resultados

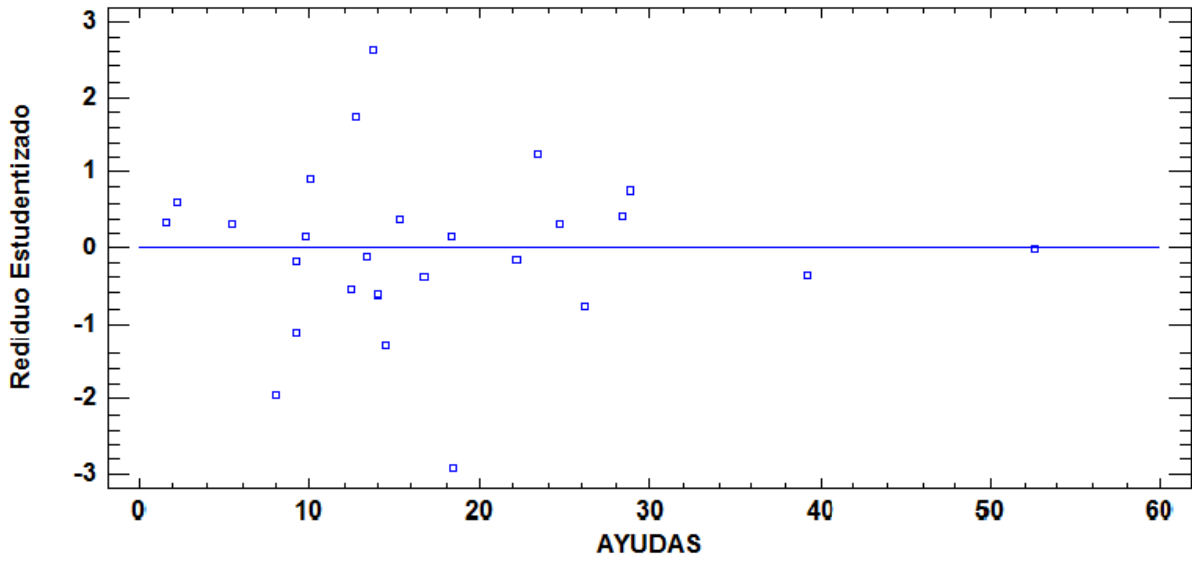


Figura IV. 17 Ayudas sin constante

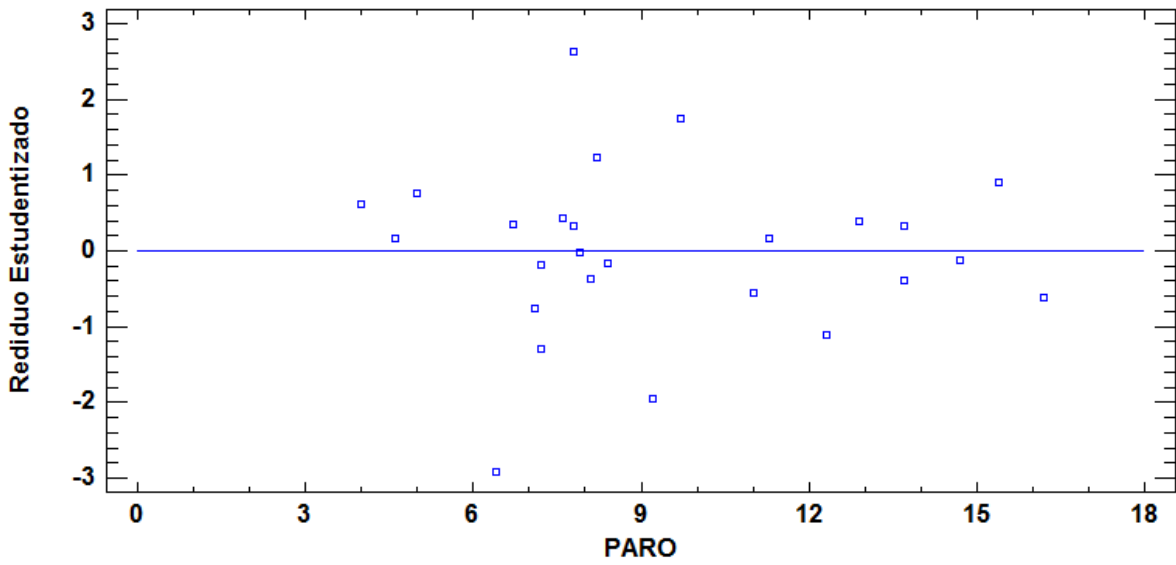


Figura IV. 18 Desempleo sin constante

IV.Cálculos y resultados

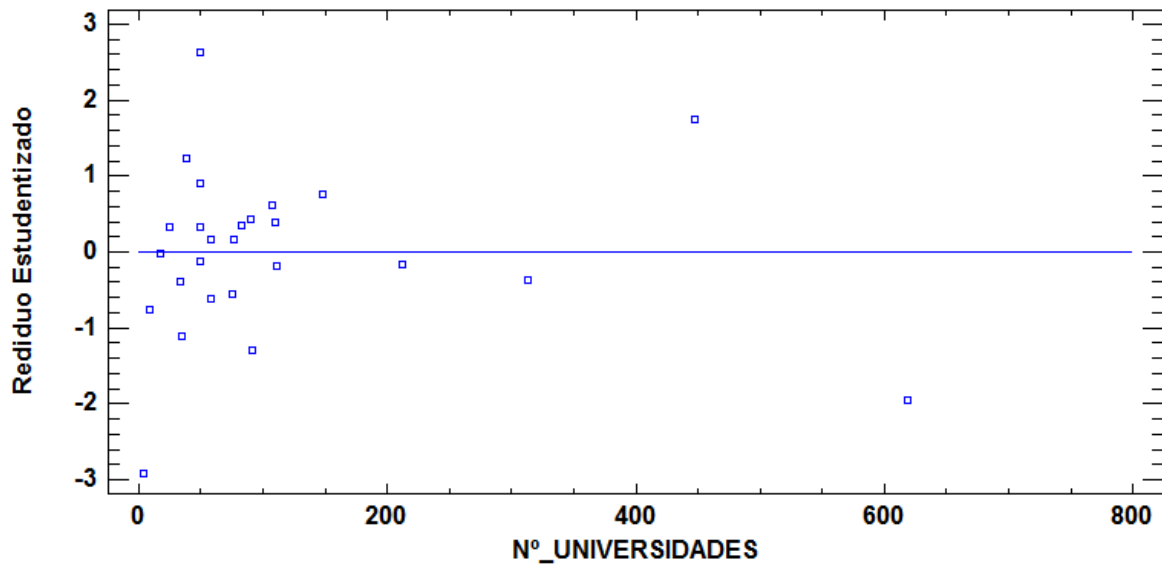


Figura IV. 19 Número de universidades sin constante

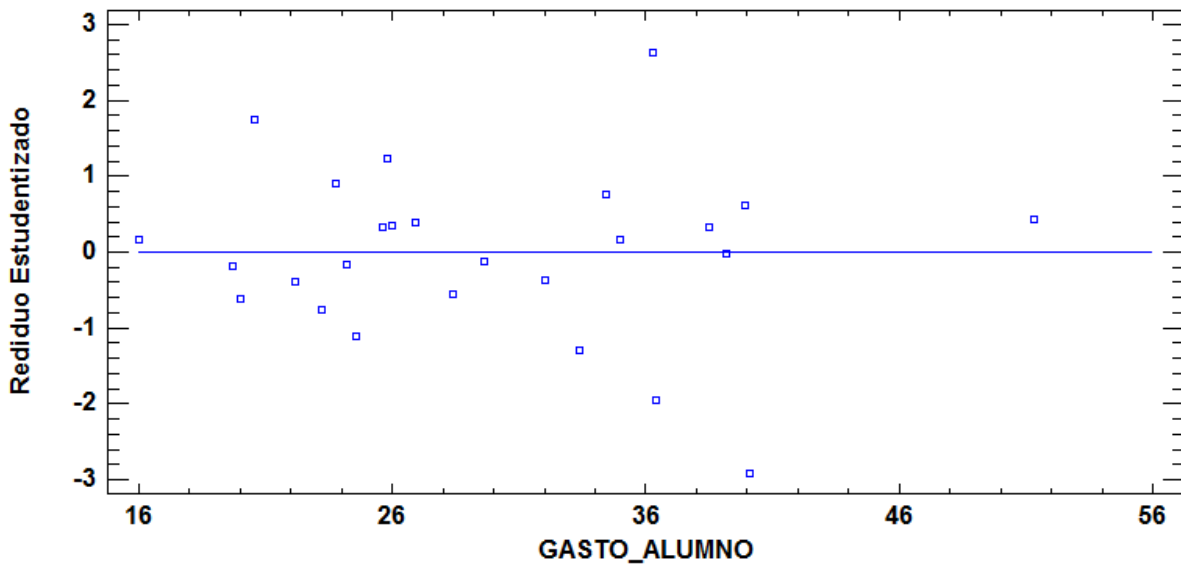


Figura IV. 20 Gasto por alumno sin constante

IV.Cálculos y resultados

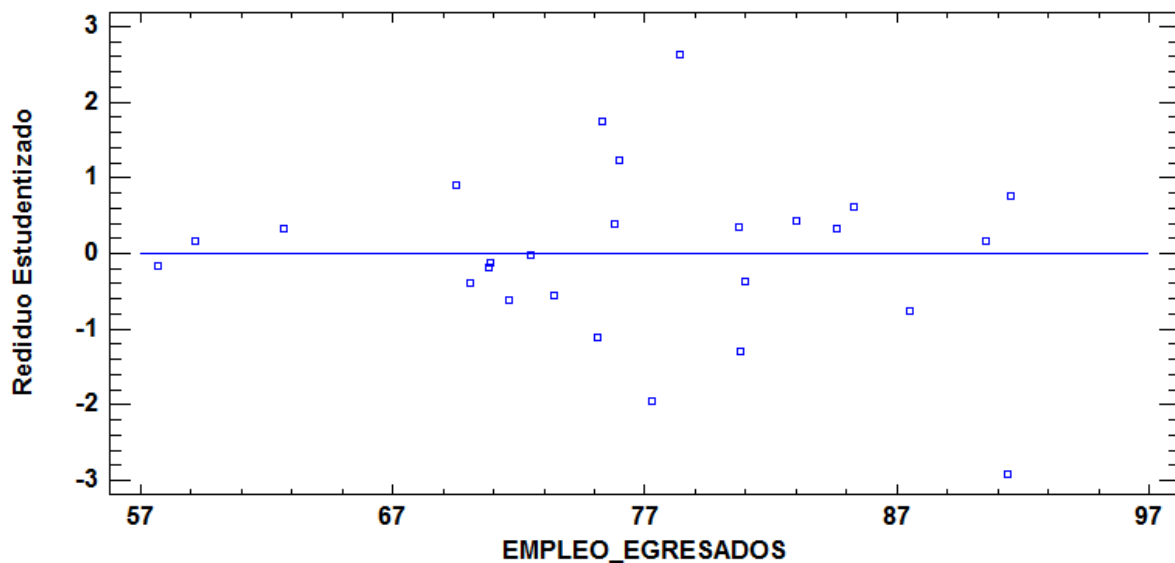


Figura IV. 21 Empleo egresados sin constante

Se observa que posiblemente la variable HORAS_CULTURA (Figura IV. 16) tiene forma parabólica, igual que ocurría con el primer modelo lineal, y que las demás variables no parecen tener ninguna forma parabólica ni cónica, por lo tanto, no tienen problemas de linealidad. Sólo es posible detectar algunos puntos anómalos.

Como se puede ver, la variable HORAS_CULTURA tiene posibles problemas de falta de linealidad por lo que se va a intentar subsanar el problema, con el mismo procedimiento seguido anteriormente. En este caso el resultado es el siguiente:

- R-cuadrado sin modificar la variable HORAS_CULTURA: R-cuadrada = 98,4087 %
- R-cuadrado modificando la variable (HORAS_CULTURA²):R-cuadrada = 98,4196%
- R-cuadrado modificando la variable (HORAS_CULTURA⁴): R-cuadrada = 98,3985%

Se observa que en este caso a partir de la cuarta potencia el R-cuadrado baja y, de este modo, se quedará con HORAS_CULTURA². Al elevar HORAS_CULTURA al cuadrado se observa que el R-cuadrado es mayor que anteriormente con un 98,4196% por lo que se sigue con la mejora del ajuste al ir aplicando las soluciones a los problemas detectados (Tabla IV.13).

IV.Cálculos y resultados

Tabla IV.13 Resolución significatividad con HORAS_CULTURA^2

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
PIB	0,734841	1,16101	0,632935	0,5347
POCO_TRABAJO	1,72506	0,708283	2,43555	0,0255
HORAS_CULTURA^2	0,022215	0,00626626	3,54517	0,0023
AYUDAS	0,0561623	0,199871	0,280993	0,7819
PARO	-0,645447	0,823765	-0,783533	0,4435
Nº_UNIVERSIDADES	0,00587998	0,0145899	0,403018	0,6917
GASTO_ALUMNO	-0,611807	0,377544	-1,62049	0,1225
EMPLEO_EGRESADOS	0,713925	0,170488	4,18754	0,0006

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	118123,	8	14765,4	140,12	0,0000
Residuo	1896,77	18	105,376		
Total	120020,	26			

R-cuadrada = 98,4196 por ciento

R-cuadrado (ajustado para g.l.) = 97,805 por ciento

IV.Cálculos y resultados

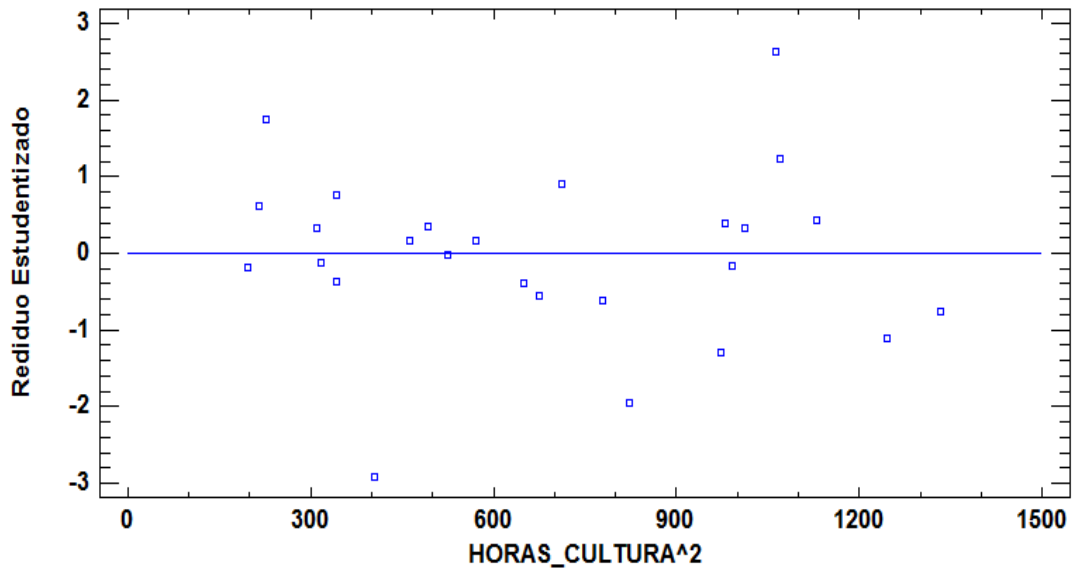


Figura IV. 22 Horas cultura elevado a 2

Al realizar esta modificación se ve reflejado un pequeño cambio en el gráfico de residuos correspondiente, al expandir un poco más los puntos y deshacer la nombrada forma parabólica (Figura IV. 22).

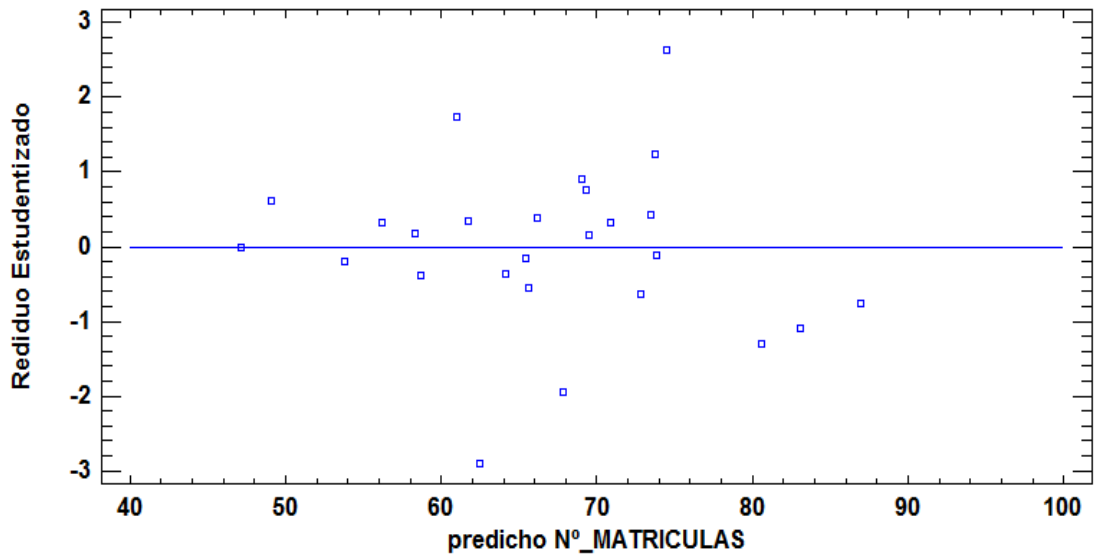


Figura IV. 23 Vs Y predicho sin constante

IV.Cálculos y resultados

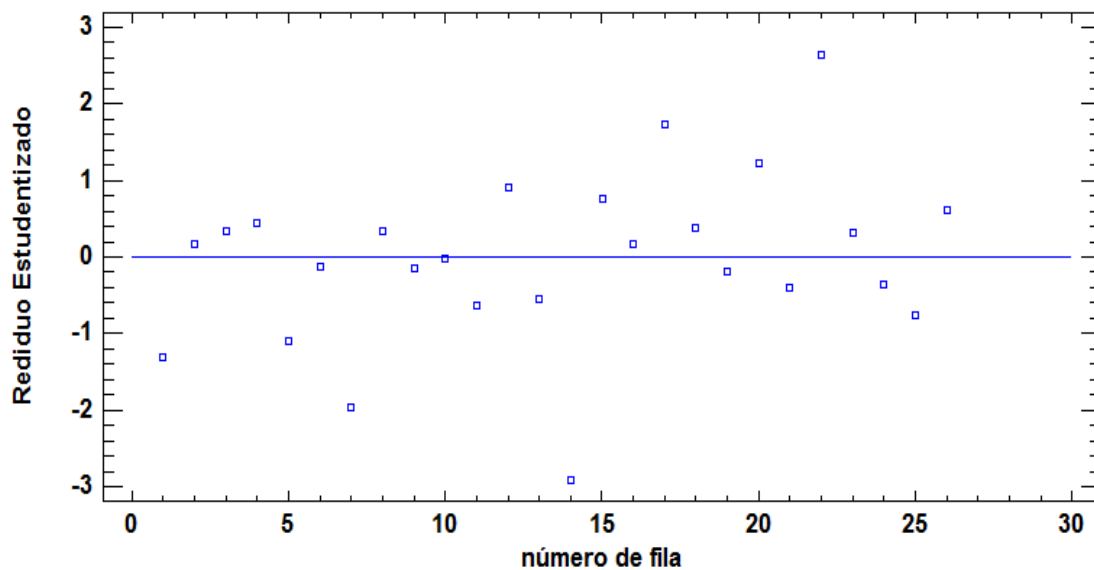


Figura IV. 24 Vs nº de filas sin constante

No se observa ninguna otra forma correspondiente a los posibles problemas generados por la perturbación del modelo en el análisis gráfico, por lo tanto, no hay falta de linealidad en el gráfico Vs Y predicho (Figura IV. 23). Aunque en la Figura IV. 24 hay una posible forma sinusoidal que advierte de un posible problema de autocorrelación de primer orden positiva.

4.5.3. NORMALIDAD

En el caso del análisis de la normalidad de los residuos, se observa que los P-valor de las cuatro pruebas realizadas, de nuevo son mayores a 0,05 (Tabla IV. 14), y por lo tanto, sigue sin detectarse problema alguno que indique falta de normalidad.

Tabla IV. 14 Prueba normalidad sin constante

Pruebas de Normalidad para RESIDUOS

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	9,53846	0,572317
Estadístico W de Shapiro-Wilk	0,975639	0,776021
Valor-Z para asimetría	0,140899	0,887945
Valor-Z para curtosis	1,59767	0,110115

IV.Cálculos y resultados

4.5.4. HETEROCEDASTICIDAD

Los P-valor de las variables explicativas en el modelo de los residuos elevados al cuadrado, son superiores a 0,05 (Tabla IV.15), y por tanto, de nuevo se observa que no hay problema de heterocedasticidad pese a los cambios realizados en el modelo.

Tabla IV.15 Prueba heterocedasticidad sin constante

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	-222,492	439,984	-0,505681	0,6196
PIB	4,29575	17,0361	0,252156	0,8039
POCO_TRABAJO	1,28919	10,2043	0,126338	0,9009
HORAS_CULTURA	1,90234	4,55996	0,417182	0,6818
AYUDAS	-1,84191	2,88898	-0,637566	0,5323
PARO	-2,86478	14,9487	-0,191641	0,8503
Nº_UNIVERSIDADES	0,00184282	0,213854	0,00861719	0,9932
GASTO_ALUMNO	2,47622	5,45488	0,453946	0,6556
EMPLEO_EGRESADOS	2,72998	5,15076	0,530015	0,6030

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>de Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	58187,9	8	7273,48	0,34	0,9387
Residuo	365713,	17	21512,5		
Total (Corr.)	423901,	25			

R-cuadrada = 13,7268 porciento

R-cuadrado (ajustado para g.l.) = 0 porciento

4.5.5. AUTOCORRELACIÓN

Igual que ocurría con el modelo completo, se vuelve a comprobar que no hay problemas de autocorrelación ni positiva ni negativa en ninguno de los dos gráficos (Figura IV. 25 y Figura IV. 26).

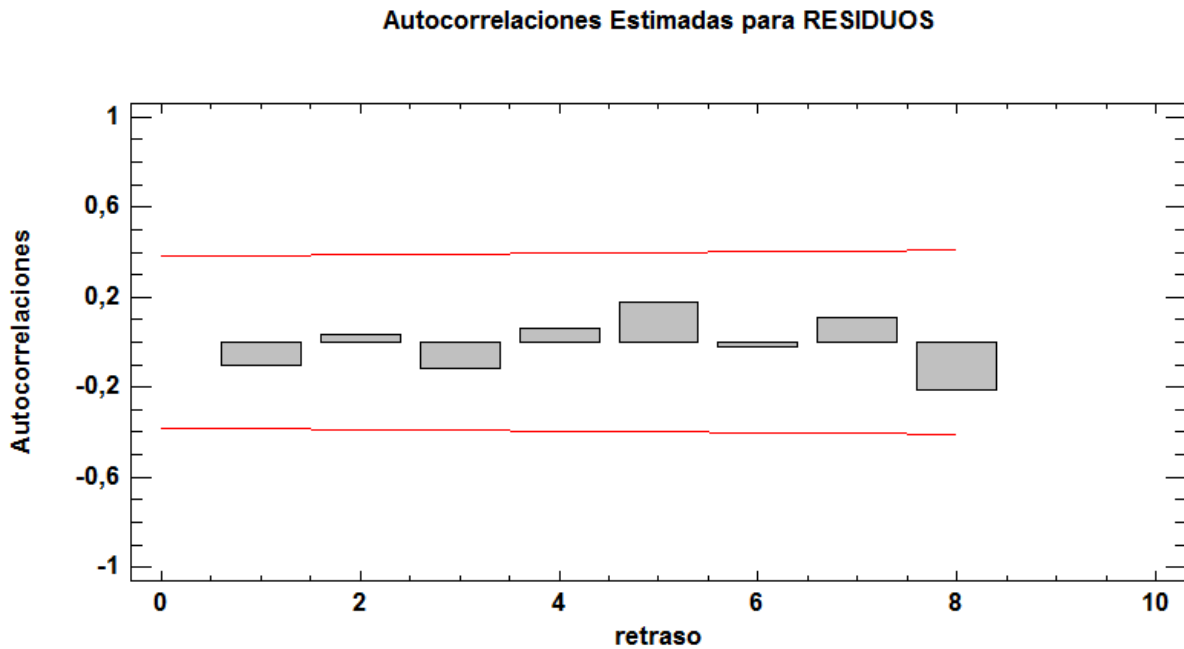


Figura IV. 25 FAS sin constante

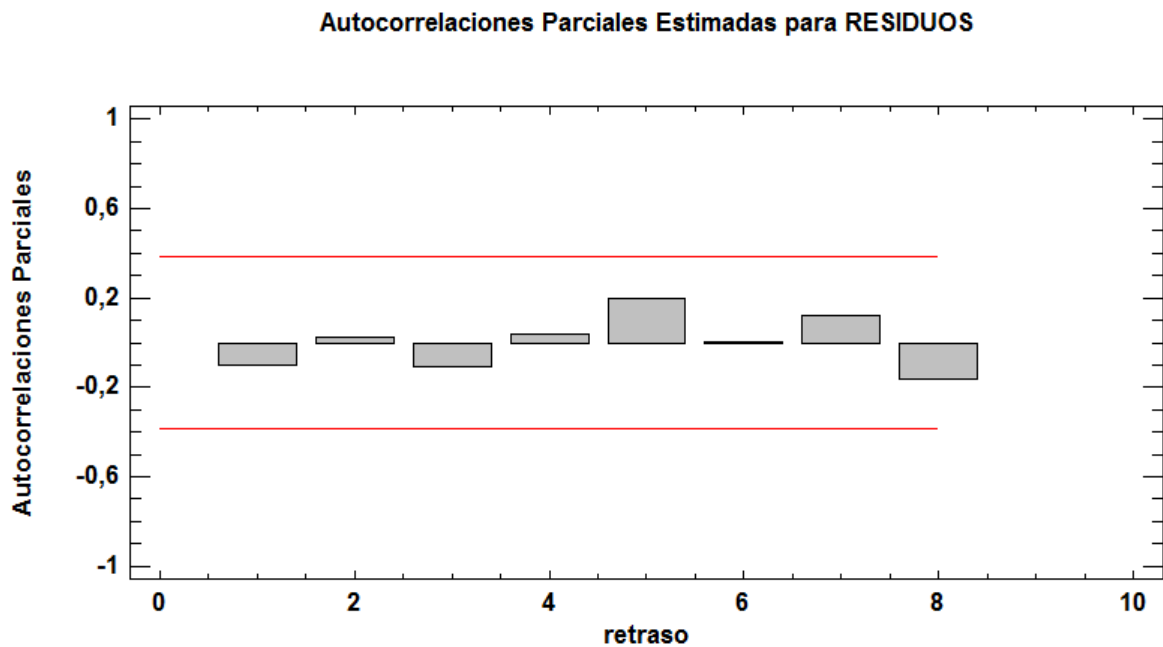


Figura IV. 26 FAP sin constante

4.6. MODELO LINEAL SIN CONSTANTE NI VARIABLES PROBLEMATICAS

En resumen finalmente se decide eliminar la constante junto con las variables PIB, AYUDAS, PARO, N° UNIVERSIDADES y GASTO POR ALUMNO, debido a que no son significativas, y no se debe a la presencia de ningún problema en el modelo lineal sin la constante. Se procede con la comprobación de la validación del modelo con todo significativo.

4.6.1. SIGNIFICATIVIDAD

Se observa que tanto el modelo como las variables son significativos. El R-cuadrado ajustado es de 97,8433%, cuando en el anterior, sin contante era de 97,7898%, por lo que con la eliminación de las variables no significativas, está mejorándose el ajuste del modelo lineal planteado (Tabla IV.16).

Tabla IV. 16 Prueba de significatividad sin constante ni variables

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
POCO_TRABAJO	1,10722	0,466137	2,37531	0,0263
HORAS_CULTURA	0,981823	0,276725	3,54801	0,0017
EMPLEO_EGRESADOS	0,399137	0,101331	3,93893	0,0007

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Gl Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	117639,	3	39212,9	378,72	0,0000
Residuo	2381,42	23	103,54		
Total	120020,	26			

R-cuadrada = 98,0158 por ciento

R-cuadrado (ajustado para g.l.) = 97,8433 por ciento

4.6.2. FALTA DE LINEALIDAD

En el gráfico HORAS_CULTURA (Figura IV. 28) se puede visualizar una leve forma parabólica que indica un posible problema de falta de linealidad de esta variable, por lo demás, no se ve ninguna forma cónica ni parabólica en las otras dos variables pero, si un punto anómalo, que será candidato a eliminar, en la fila 14 (Malta) con un -3.39 (Tabla IV. 17).

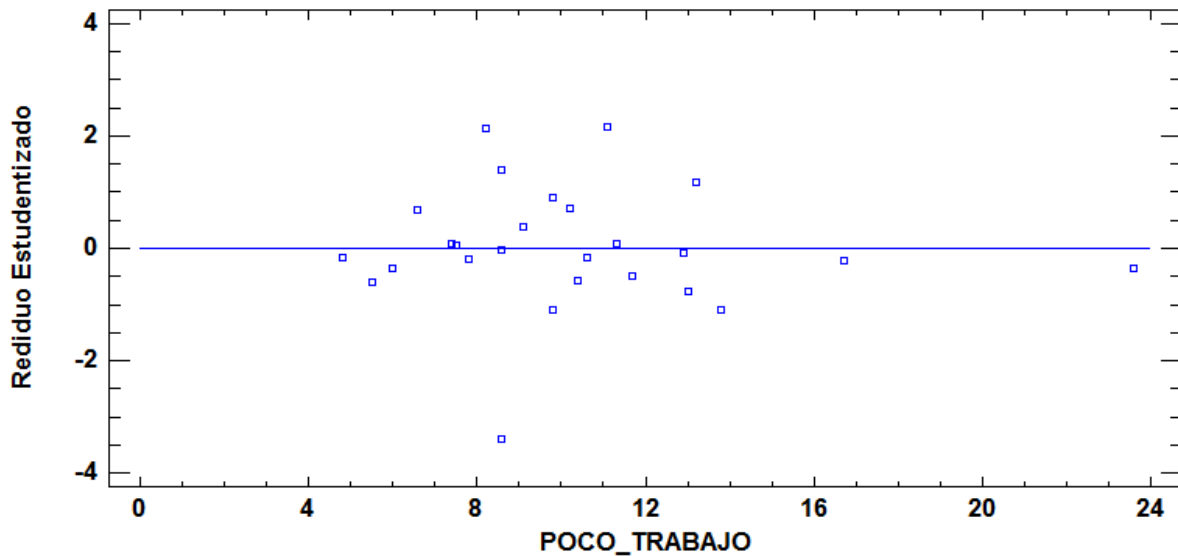


Figura IV. 27 Poco trabajo sin variables

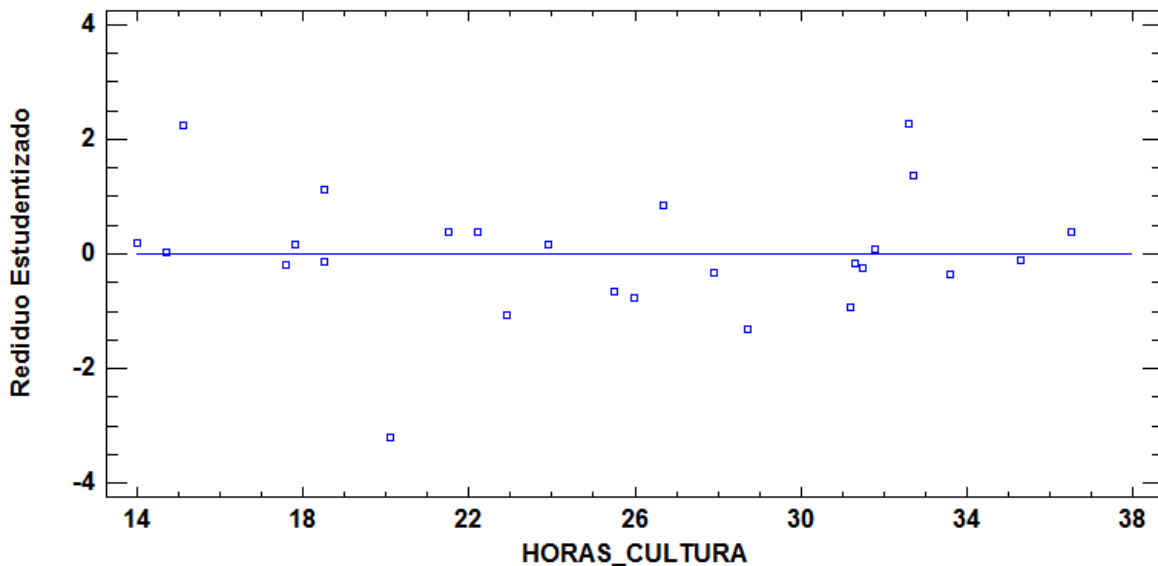


Figura IV. 28 Horas cultura sin variables

IV.Cálculos y resultados

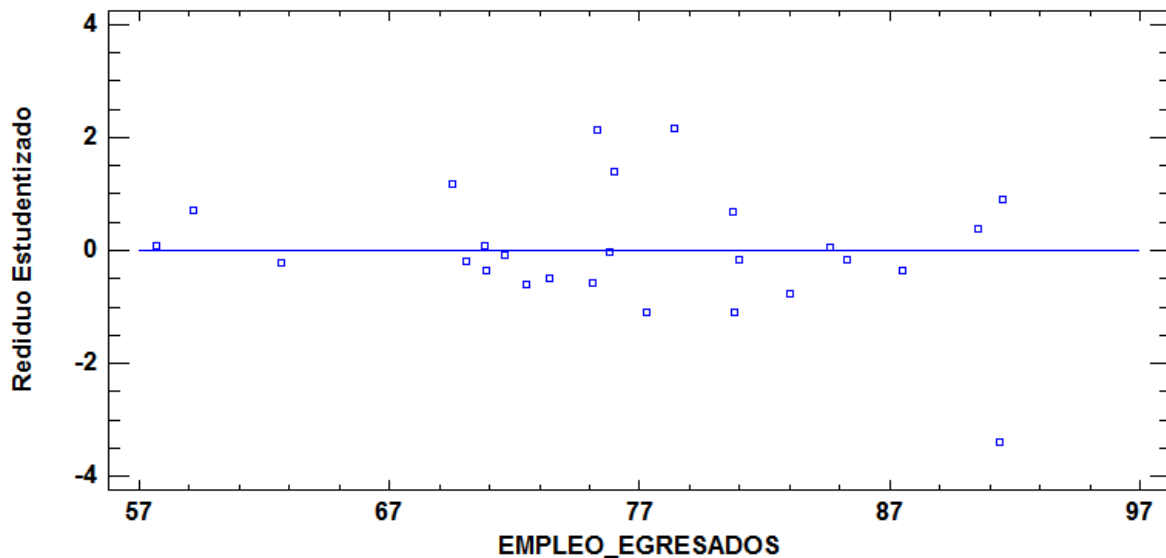


Figura IV. 29 Empleo egresados sin variables

Tabla IV. 17 Puntos anómalos

		<i>Y</i>		<i>Residuo</i>
<i>Fila</i>	<i>Y</i>	<i>Predicha</i>	<i>Residuo</i>	<i>Estudentizado</i>
14	40,0	66,2967	-26,2967	-3,39
17	74,0	55,028	18,972	2,14
22	96,0	76,8022	19,1978	2,17

Para solucionar el problema se planteará eliminar Malta de la base de datos ya que, según dicen las pruebas, no sirve para nuestro modelo.

En el caso de la variable HORAS_CULTURA, se modificará elevándola a un exponente para solucionar el problema como ha venido haciéndose. En este caso se obtiene de nuevo que la mejor solución resulte elevando la variable a la cuarta. No se vuelven a mostrar todos los resultados obtenidos del ajuste con los pasos de la solución para que no resulte tan repetitivo el trabajo. Decir entonces únicamente, que con HORAS_CULTURA⁴, en el modelo con todo significativo, el R-cuadrado alcanza un valor del 98,0948%

Igual que en los casos anteriores, a nivel gráfico también se corrobora la solución de la falta de linealidad al incorporar la variable HORAS_CULTURA⁴ (Figura IV. 30).

IV.Cálculos y resultados

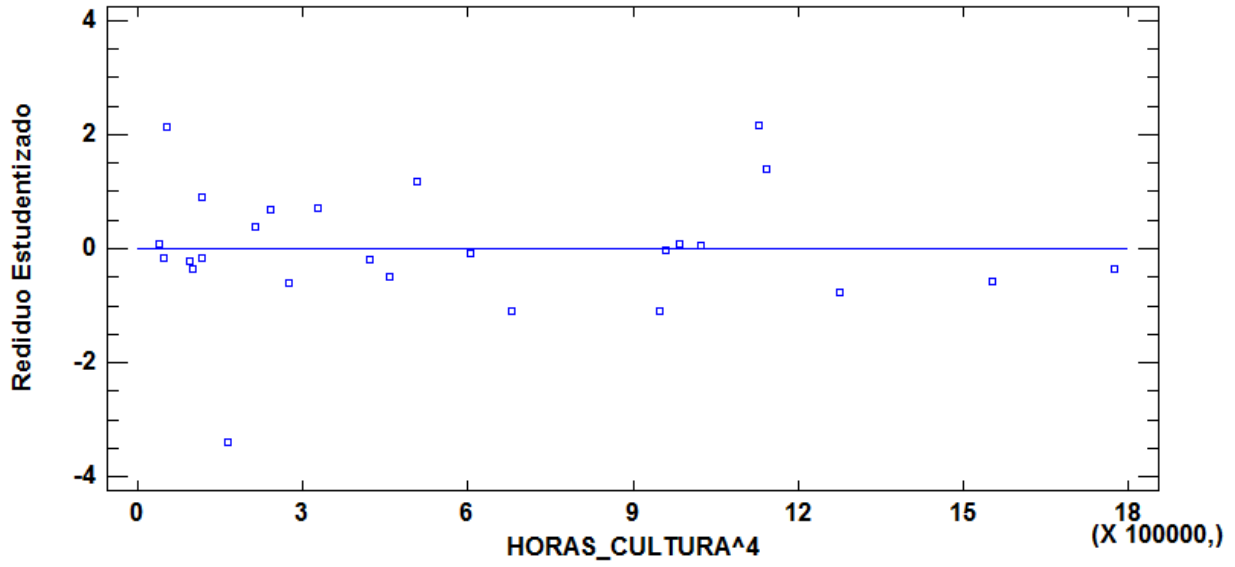


Figura IV. 30 Horas cultura elevado a 4

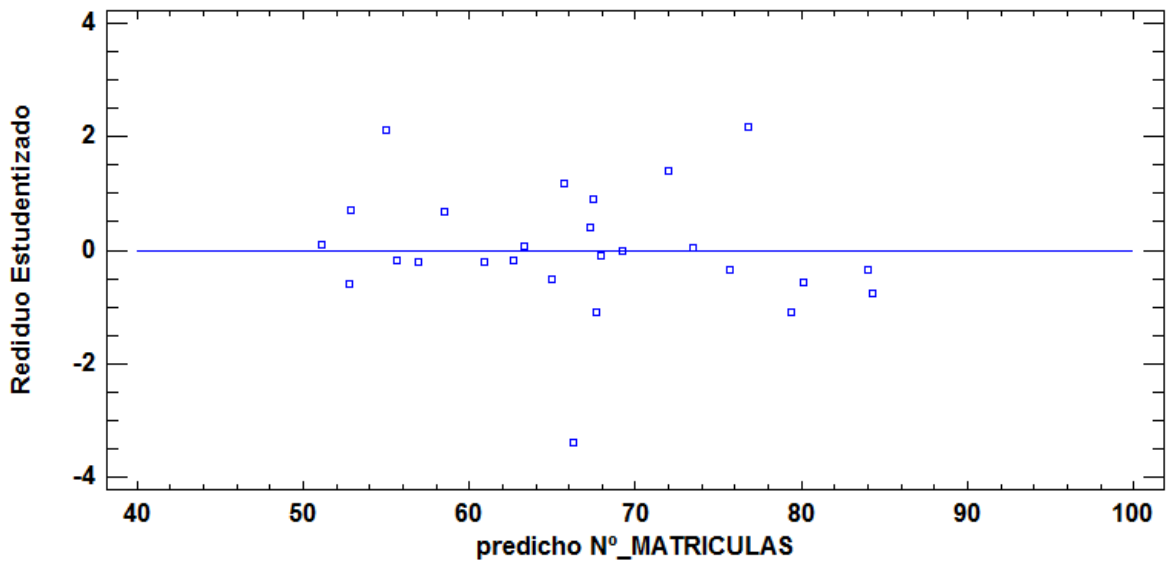


Figura IV. 31 Vs Y predicho sin variables

IV.Cálculos y resultados

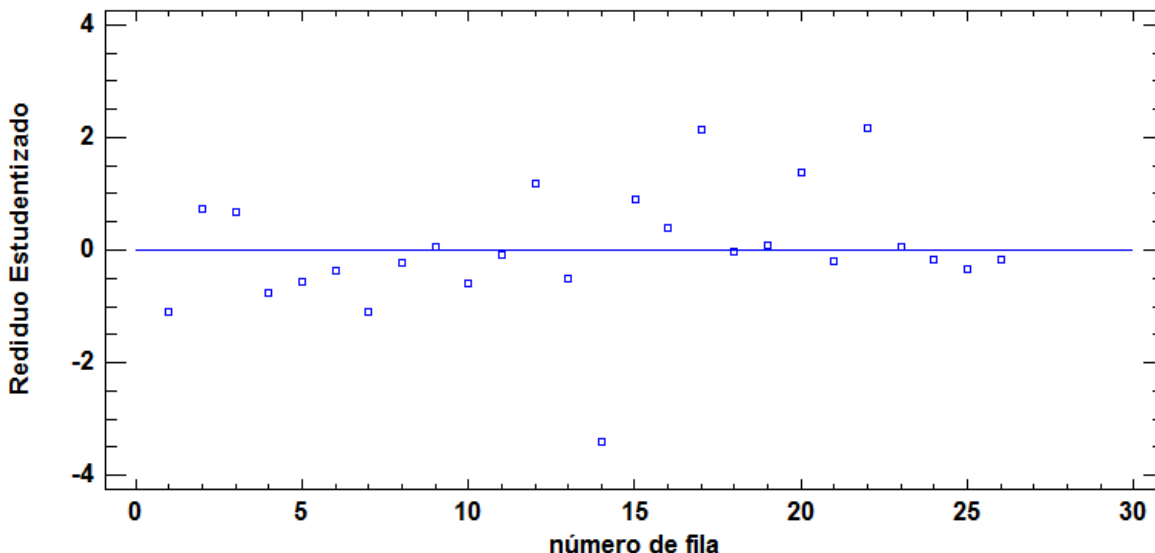


Figura IV. 32 Vs nº de filas sin variables

No se detecta falta de linealidad en la Figura IV. 31, pero si un punto anómalo en la fila 14 (Malta) como se ha visto anteriormente. Y posiblemente exista una forma de zigzag que pueda estar indicando presencia de autocorrelación negativa de primer orden (Figura IV. 32). El fenómeno se corroborará en los gráficos del FAS y FAP correspondientes.

4.6.3. NORMALIDAD

Los P-valor de las cuatro pruebas (Tabla IV. 18) son mayores a 0,05, por lo tanto, no existe ningún problema que nos indique falta de normalidad en los residuos de este modelo.

Tabla IV. 18 Prueba normalidad sin variables

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	14,9231	0,186046
Estadístico W de Shapiro-Wilk	0,943287	0,17292
Valor-Z para asimetría	0,336352	0,736602
Valor-Z para curtosis	1,65037	0,098867

IV.Cálculos y resultados

4.6.4. HETEROCEDASTICIDAD

De nuevo se representan los residuos del modelo al cuadrado frente a las variables explicativas, que se han dejado finalmente en el modelo, y como puede observarse en la (Tabla IV. 19) los P-valores de todas las variables explicativas son mayores a 0,05, confirmando que no hay problema de heterocedasticidad.

Tabla IV. 19 Prueba de heterocedasticidad sin variables

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	-272,541	356,784	-0,763881	0,4531
POCO_TRABAJO	0,771832	8,81803	0,0875288	0,9310
HORAS_CULTURA	-1,90593	4,67151	-0,407991	0,6872
EMPLEO_EGRESADOS	5,22989	3,78121	1,38312	0,1805

4.6.5. AUTOCORRELACIÓN

Por último, siguen sin detectarse problemas de autocorrelación ni positiva ni negativa, pese a lo que se había mencionado en el gráfico de residuos frente al número de fila (Figura IV. 33 y Figura IV. 34).

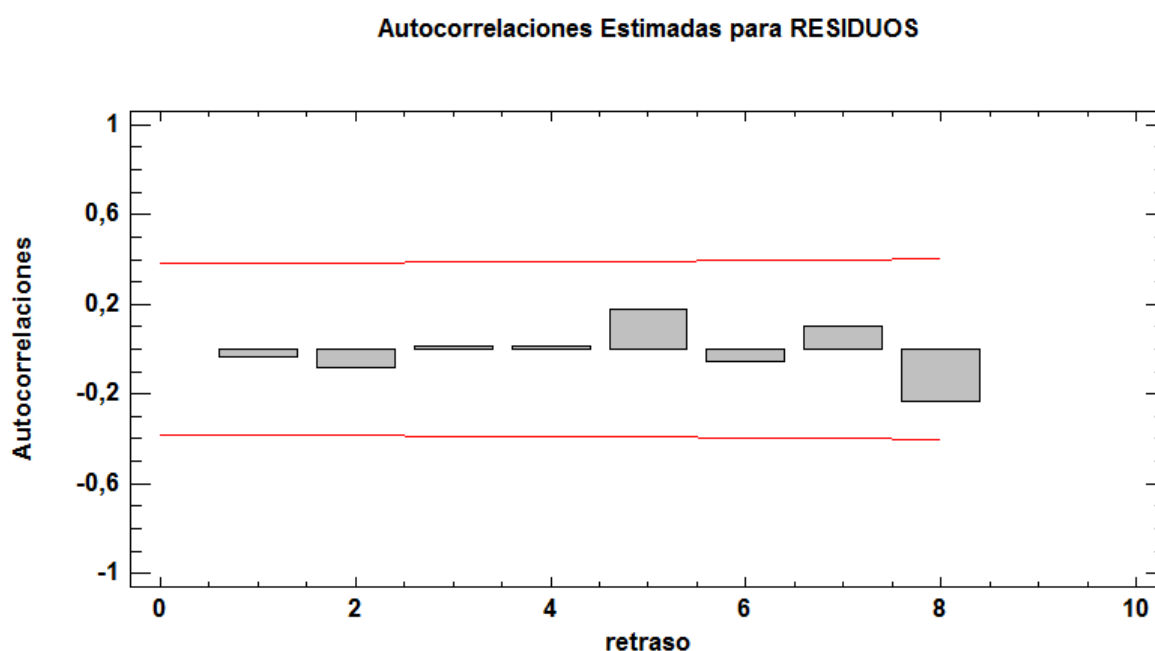


Figura IV. 33 FAS sin variables problemáticas

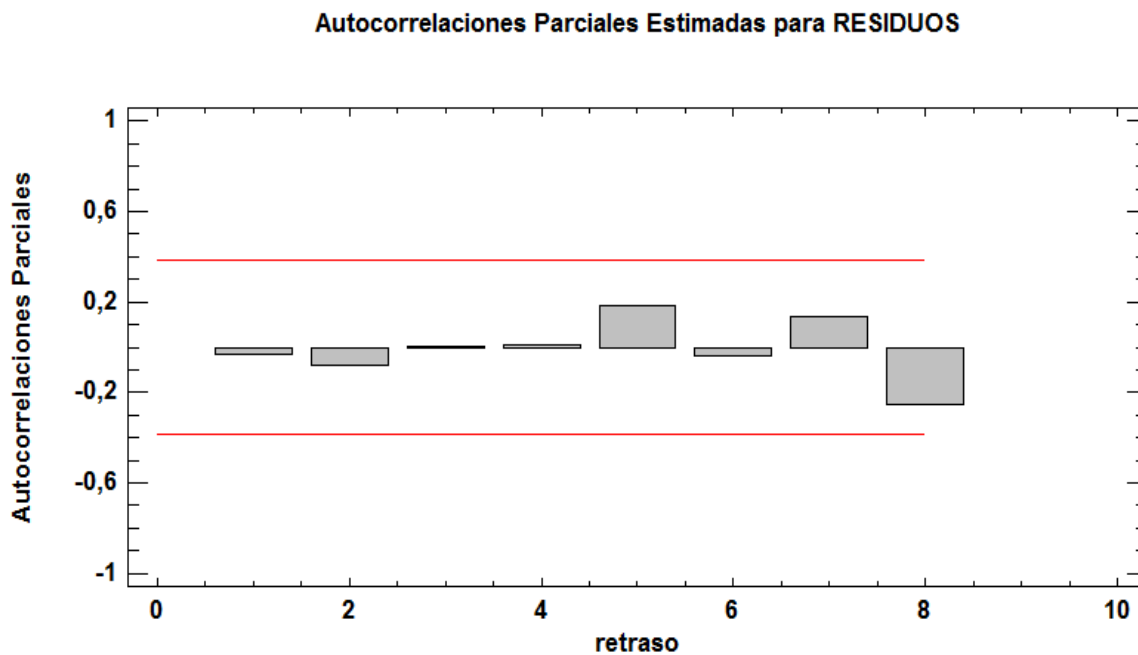


Figura IV. 34 FAP sin variables problemáticas

4.7 MODELO LINEAL SIN MALTA

En el proceso anterior, se ha observado que existía un punto anómalo candidato a eliminar, pero que el modelo era válido. En esta última parte del modelo lineal se han eliminado, aparte de las ya comentadas variables no significativas y la constante, el punto anómalo candidato a eliminar, que corresponde con ellos datos del país Malta, para ver si así mejora aún más el ajuste del modelo lineal.

4.7.1. SIGNIFICATIVIDAD

Con la eliminación del punto anómalo el R-cuadrado ajustado pasa de 97,9291% (en el modelo con la falta de linealidad solucionada incluida), a 98,5076%, con lo que se puede confirmar una mejora en el ajuste del modelo con esta decisión. Dónde todo continúa siendo significativo, como en el caso anterior (Tabla IV.20). Decir que esta decisión deja fuera uno de los países analizados en el modelo, pero si se corrobora su validez, mejorando el ajuste del modelo, las predicciones obtenidas serán más fiables y ajustadas. Es por este motivo que se repetirán las pruebas de validación realizadas hasta ahora, con tan de poder tener al final otra alternativa, aún más ajustada que la anterior.

Tabla IV. 20 Prueba de significatividad sin Malta

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
POCO_TRABAJO	0,974937	0,395247	2,46665	0,0219
HORAS_CULTURA	0,794482	0,240529	3,30307	0,0032
EMPLEO_EGRESADOS	0,494866	0,090491	5,46867	0,0000

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>de Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	116800,	3	38933,3	528,73	0,0000
Residuo	1619,98	22	73,6354		
Total	118420,	25			

R-cuadrada = 98,632 por ciento

R-cuadrado (ajustado para g.l.) = 98,5076 por ciento

4.7.2. FALTA DE LINEALIDAD

Se detecta como viene dándose de nuevo la falta de linealidad en la variable HORAS_CULTURA (Figura IV. 35) al mostrar una forma parabólica en el gráfico. Igual que en el caso anterior la mejor solución encontrada para este fenómeno, fue elevar la variable a la cuarta potencia, de forma que se pasa de tener un R-cuadrado de 98,632% al 98,7325 %. El problema se soluciona como puede verse de nuevo en la gráfica de los residuos frente a HORAS_CULTURA⁴ (Figura IV. 38) y no es posible mejorar el ajuste del modelo con las siguientes potencias de dos analizadas.

IV.Cálculos y resultados

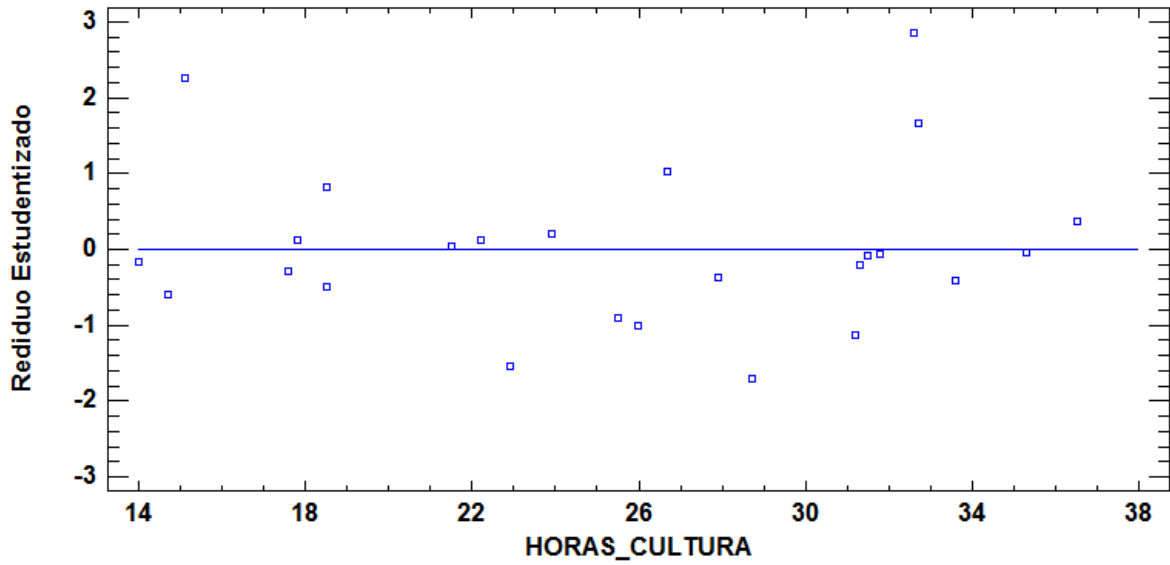


Figura IV. 35 Horas cultura sin Malta

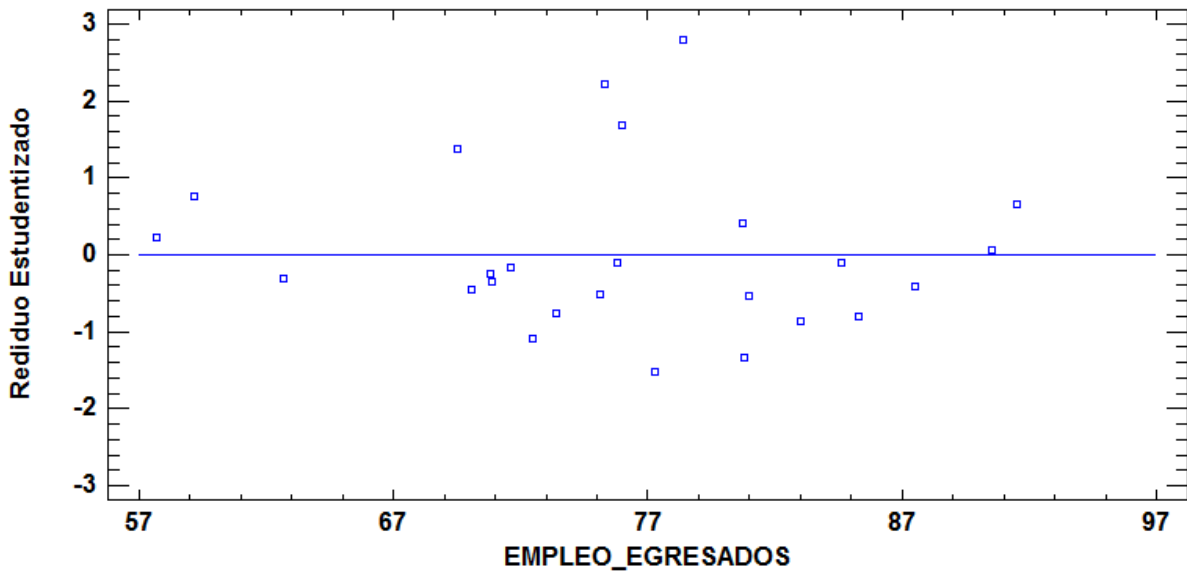


Figura IV. 36 Empleo egresados sin Malta

IV.Cálculos y resultados

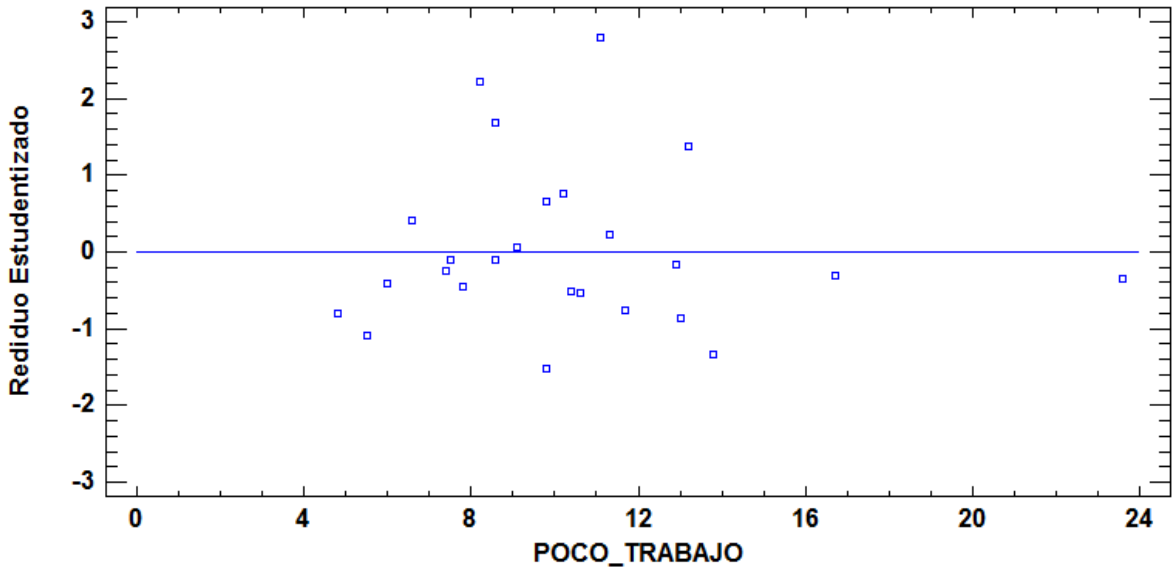


Figura IV. 37 Poco trabajo sin Malta

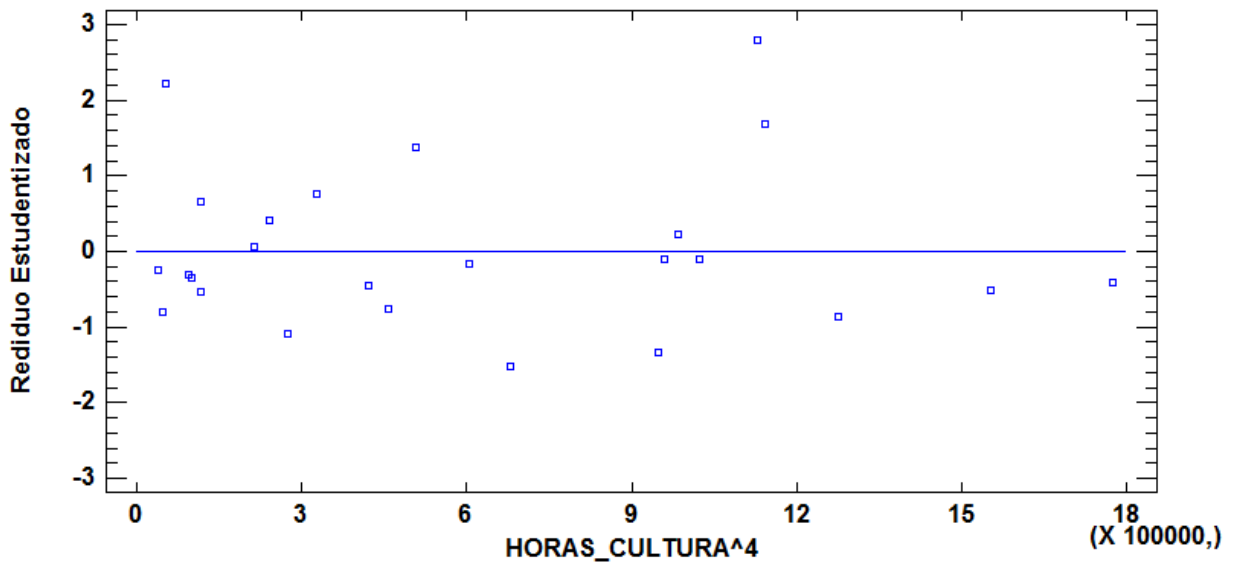


Figura IV. 38 Horas cultura elevado a 4

IV.Cálculos y resultados

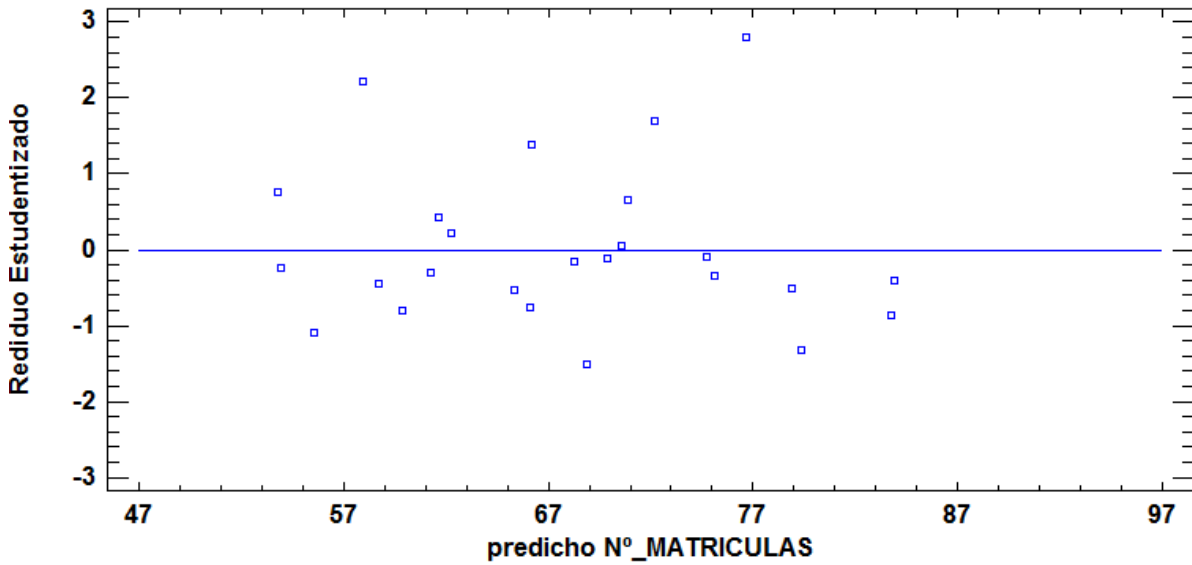


Figura IV. 39 Vs Y predicho sin Malta

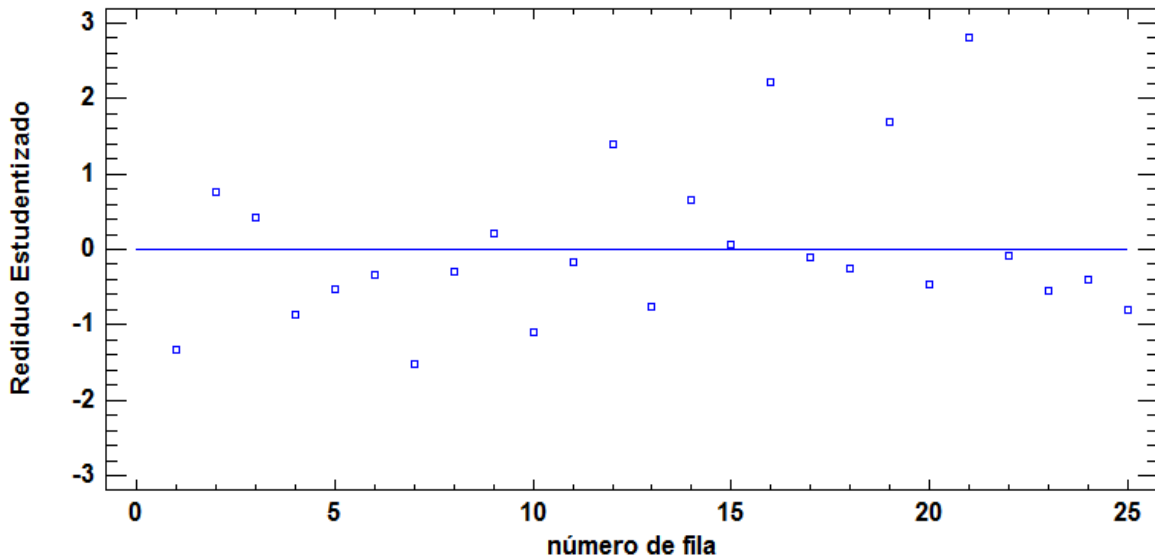


Figura IV. 40 Vs nº de filas sin Malta

No se detecta falta de linealidad al no observar forma parabólica, ni heterocedasticidad con una forma cónica en la Figura IV. 39, tampoco puntos anómalos candidatos a eliminar. En la Figura IV. 40 no se detectan posibles problemas de autocorrelación de primer orden positiva o negativa.

IV.Cálculos y resultados

4.7.3. NORMALIDAD

En este caso el test para la normalidad más restrictivo resulta ser el del Estadístico W de Shapiro-Wilk, que supera el valor de significación (0,05), y por lo tanto puede aceptarse esta hipótesis de partida de normalidad de los residuos, también en esta propuesta (Tabla IV. 21).

Tabla IV. 21 Prueba de normalidad sin Malta

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	9,16	0,607126
Estadístico W de Shapiro-Wilk	0,923975	0,065442
Valor-Z para asimetría	1,41399	0,157365
Valor-Z para curtosis	0,758988	0,447857

4.7.4. HETEROCEDASTICIDAD

Los P-valores de las variables explicativas frente a los nuevos residuos que se obtienen, son mayores a 0,05 también, con lo que se sigue asumiendo la ausencia de heterocedasticidad en los residuos del modelo (Tabla IV. 22).

Tabla IV. 22 Prueba de heterocedasticidad sin Malta

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	13,0792	219,567	0,0595683	0,9531
POCO_TRABAJO	-0,595957	5,31034	-0,112226	0,9117
HORAS_CULTURA	1,59391	2,84974	0,559318	0,5819
EMPLEO_EGRESADOS	0,163556	2,40651	0,0679641	0,9465

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	3277,3	3	1092,43	0,12	0,9493

4.7.5. AUTOCORRELACIÓN

De nuevo se observa que, pese a las sospechas anteriores, no se detectan problemas de autocorrelación de ningún orden, ni positiva ni negativa, en ninguno de los dos gráficos (Figura IV. 9 y Figura IV. 42).

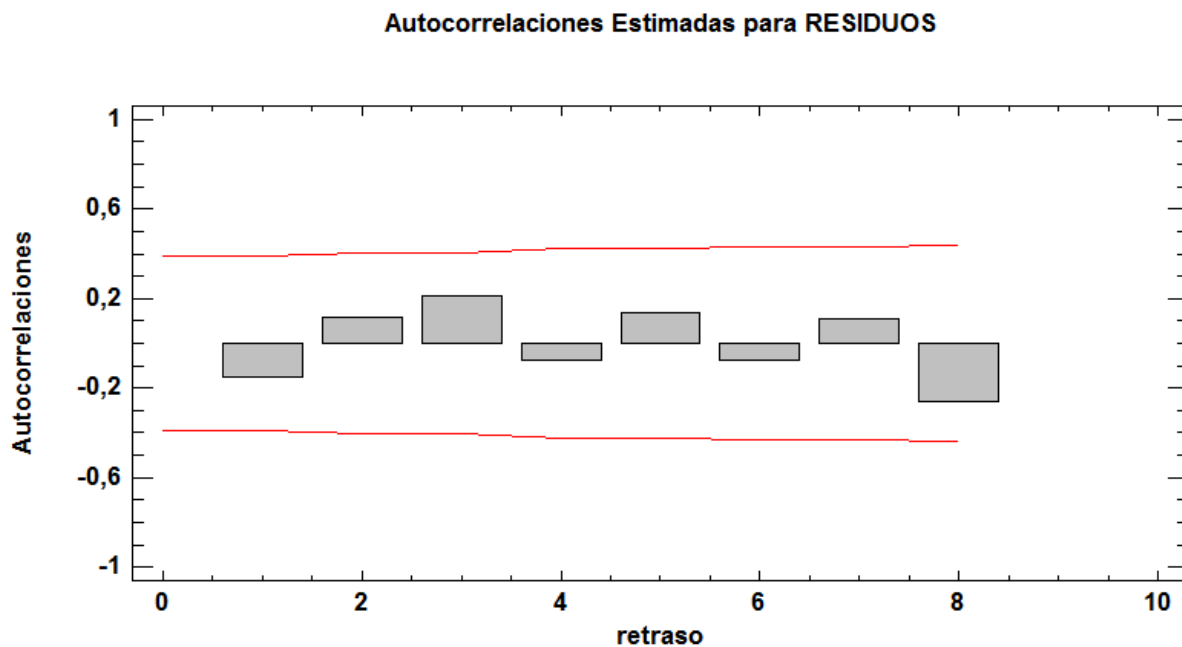


Figura IV. 41 FAS sin Malta

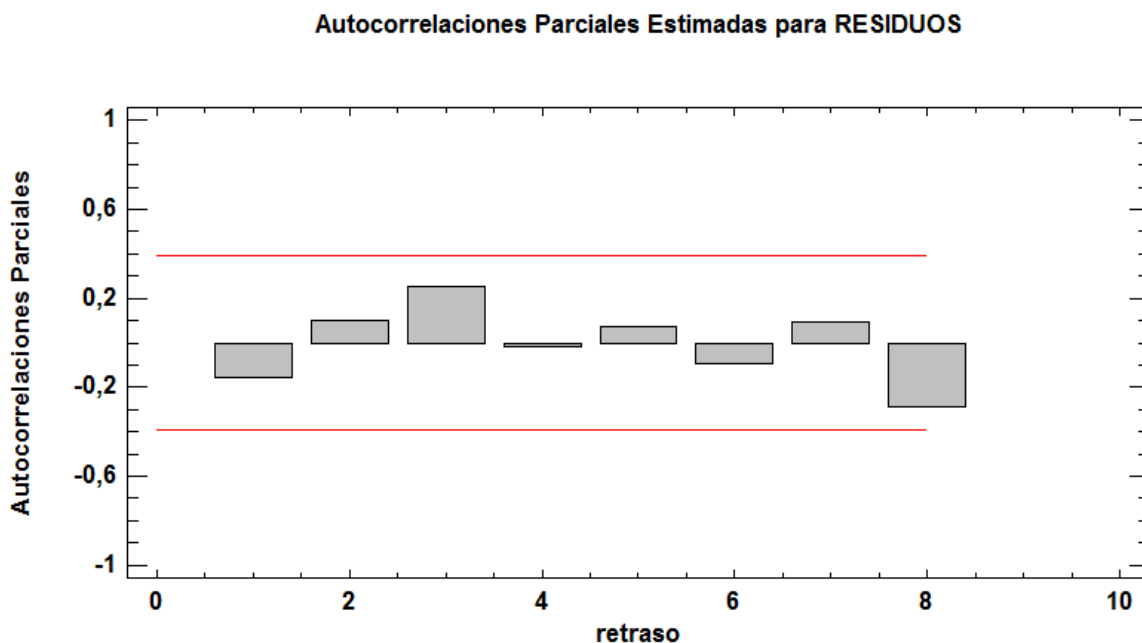


Figura IV. 42 FAP sin Malta

IV.Cálculos y resultados

4.7.6. MODELO FINAL LINEAL

La estimación del modelo lineal final propuesto, libre de todos los problemas detectados es el siguiente (12):

$$N^{\circ}\text{MATRICULAS}=1,25326*\text{POCO_TRABAJO}+0,0000122175*\text{HORAS_CULTURA}^4+0,624782*\text{EMPLEO_EGRESADOS}. \quad (12)$$

4.8. MODELO LOGARITMICO

Se realizará ahora el modelo logarítmico con las mismas pruebas realizadas anteriormente con el modelo lineal, pero en este caso solo se ha reflejado la parte final de los cálculos, es decir, se muestra el modelo con las soluciones aplicadas después de la detección de los problemas, con la misma metodología que en la estructura anterior, eliminando la constante y las variables no significativas, pero, sin eliminar el país Malta de la base de datos, puesto que en este caso no fue candidato a eliminar.

4.8.1. SIGNIFICATIVIDAD

En la propuesta final del modelo logarítmico ya no hay problemas de significatividad, ya que tanto las variables como el modelo son significativos al obtener P-Valores menores a 0,05, con lo que se rechazan todas las H_0 planteadas. Además se puede ver que el ajuste del modelo es muy elevado, alcanzando un R-cuadrado de 99,92% (Tabla IV.23). Curiosamente, pese al cambio de estructura, las variables que quedan significativas en la propuesta final del modelo son las mismas, luego con el cambio de forma, no se han conseguido incluir variables distintas en la modelización.

Tabla IV. 23 Prueba de significatividad logarítmico

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
LOG(POCO_TRABAJO)	0,205458	0,0686958	2,99083	0,0067
LOG(HORAS_CULTURA)	0,244653	0,0860299	2,84381	0,0094
LOG(EMPLEO_EGRESADOS)	0,683258	0,0663509	10,2976	0,0000

IV.Cálculos y resultados

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	442,361	3	147,454	9412,63	0,0000
Residuo	0,344641	22	0,0156655		
Total	442,706	25			

R-cuadrada = 99,9222 por ciento

R-cuadrado (ajustado para g.l.) = 99,9151 por ciento

4.8.2. FALTA DE LINEALIDAD

Igualmente, se procede seguidamente con el análisis gráfico, en búsqueda de problemas de falta de linealidad. Además según la experiencia previa, cabría esperar encontrar este problema en la variable HORAS_CULTURA, aunque habrá que ver, como ha afectado a este fenómeno el cambio de forma en la estructura del modelo de regresión.

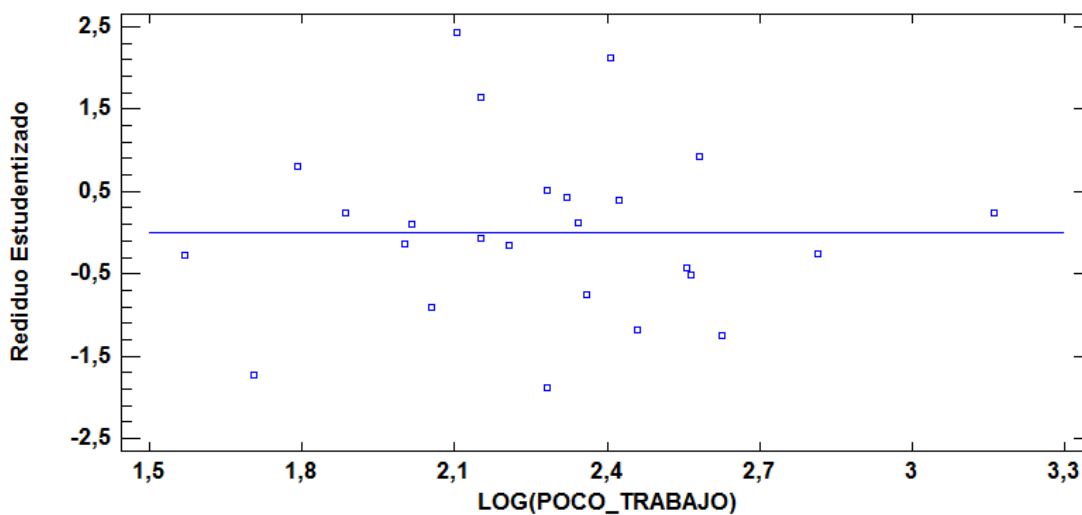


Figura IV. 43 Poco trabajo Log

IV.Cálculos y resultados

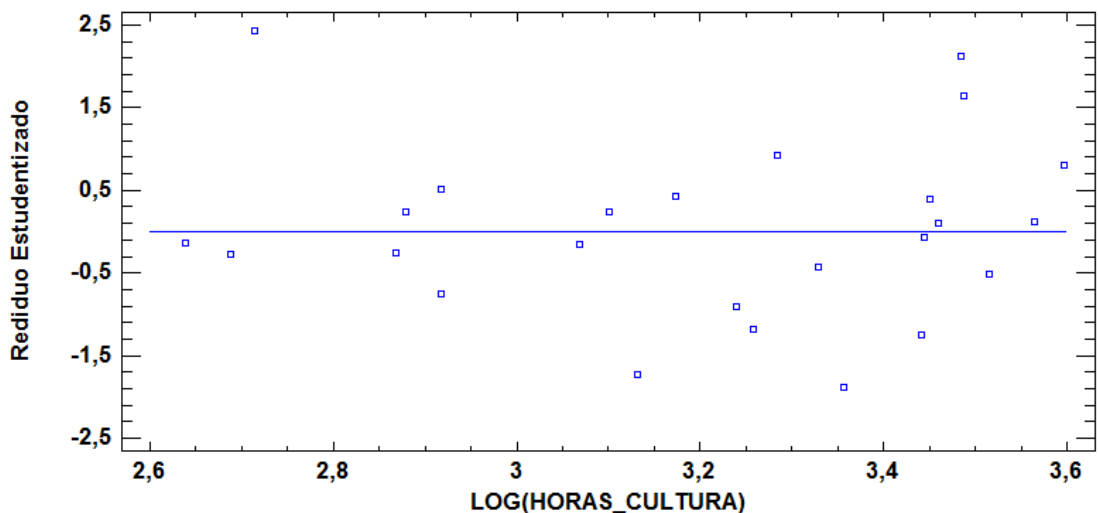


Figura IV. 44 Horas cultura Log

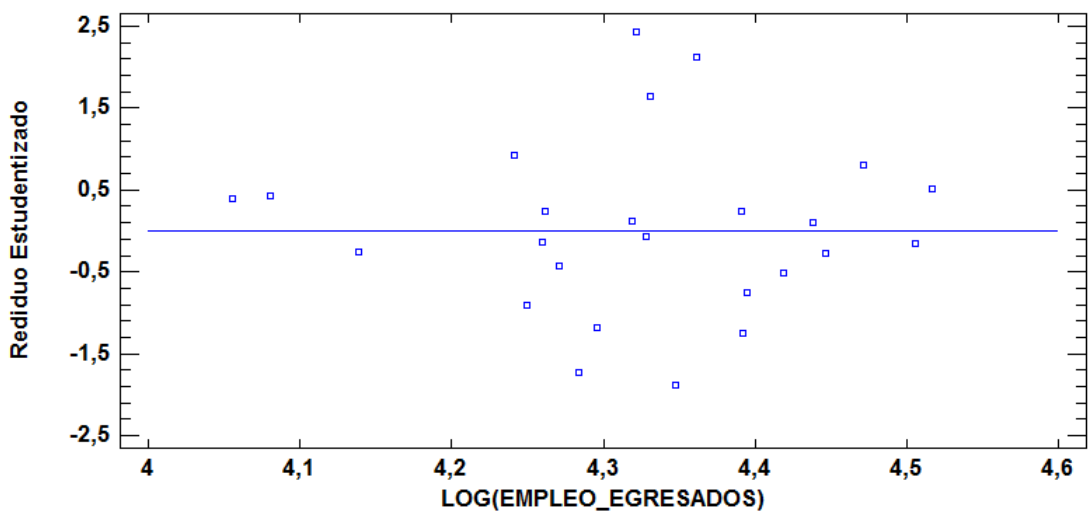


Figura IV. 45 Empleo egresados Log

Por el hecho comentado anteriormente, aunque no se ve de forma tan clara como en el modelo lineal, la posible falta de linealidad en la variable LOG(HORAS_CULTURA), con la ligera sospecha (Figura IV. 44), se realizó la prueba de elevar la variable al cuadrado. Se pudo comprobar que el ajuste del modelo descendía, indicando que estas sospechas no estaban acertadas en este caso, y por tanto se dejó la variable como estaba en un principio.

IV.Cálculos y resultados

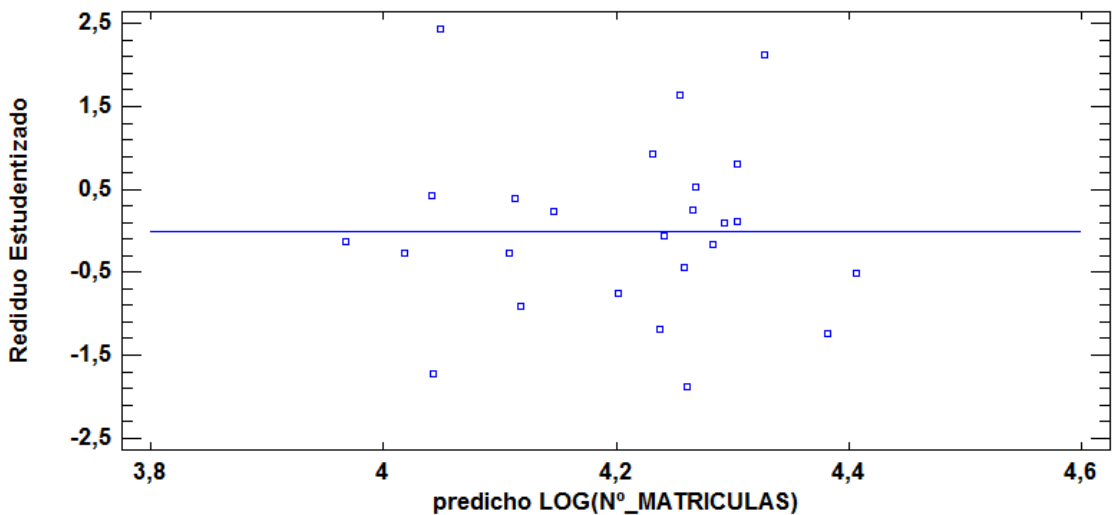


Figura IV. 46 Vs Y predicho Log

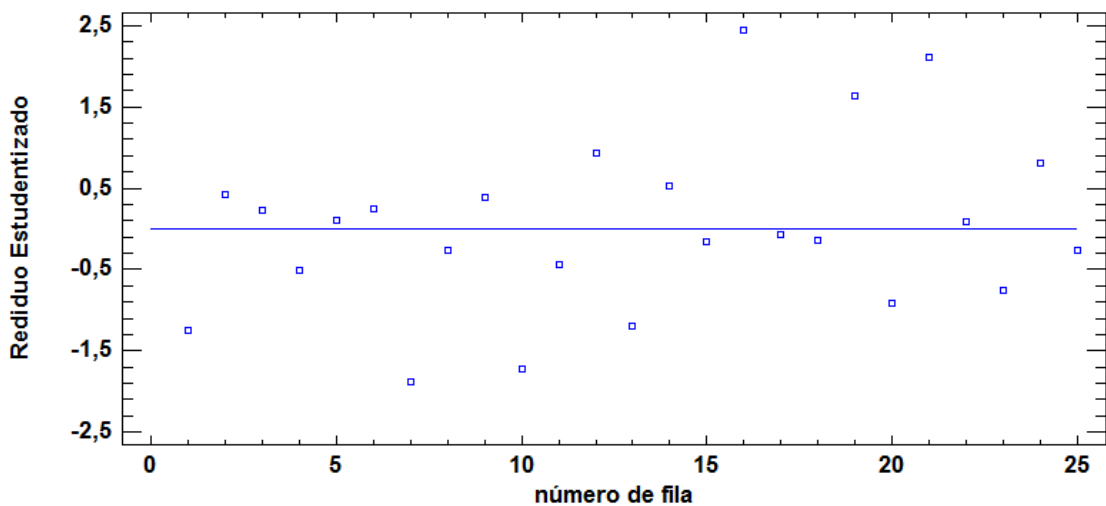


Figura IV. 47 Vs nº de filas Log

En el resto de los gráficos de residuos, no se detectan formas típicas de la presencia de los problemas analizados, observándose en todos los casos una distribución aleatoria y uniforme de los residuos. (Figura IV. 43, 45, 46). Tampoco se observan las formas típicas que señalan el fenómeno de la autocorrelación de primer orden, positiva o negativa en los residuos del modelo analizado en la Figura IV. 47.

IV.Cálculos y resultados

4.8.3. NORMALIDAD

Los residuos de la estructura logarítmica pasan todos los test de normalidad planteados, por lo que se acepta la hipótesis de la distribución normal de los mismos (Tabla IV. 24).

Tabla IV. 24 Prueba de normalidad Log

Pruebas de Normalidad para RESIDUOS

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	5,8	0,88637
Estadístico W de Shapiro-Wilk	0,96973	0,645007
Valor-Z para asimetría	0,548082	0,583633
Valor-Z para curtosis	0,379459	0,704344

4.8.4. HETEROCEDASTICIDAD

En esta estructura tampoco se detectan problemas por la presencia de heterocedasticidad por la varianza de los residuos (Tabla IV. 25).

Tabla IV. 25 Prueba de heterocedasticidad Log

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	0,0147337	0,178636	0,082479	0,9350
LOG(POCO_TRABAJO)	-0,00588196	0,0126899	-0,463516	0,6478
LOG(HORAS_CULTURA)	0,00458785	0,0142954	0,320931	0,7514
LOG(EMPLEO_EGRESADOS)	-0,000522744	0,0378504	-0,0138108	0,9891

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	0,000134154	3	0,000044718	0,11	0,9551
Residuo	0,00878712	21	0,000418434		
Total (Corr.)	0,00892127	24			

4.8.5. AUTOCORRELACIÓN

Por último en el caso de la estructura logarítmica, tampoco se observa ningún tipo de problema de autocorrelación (Figura IV. 48 y Figura IV. 49).

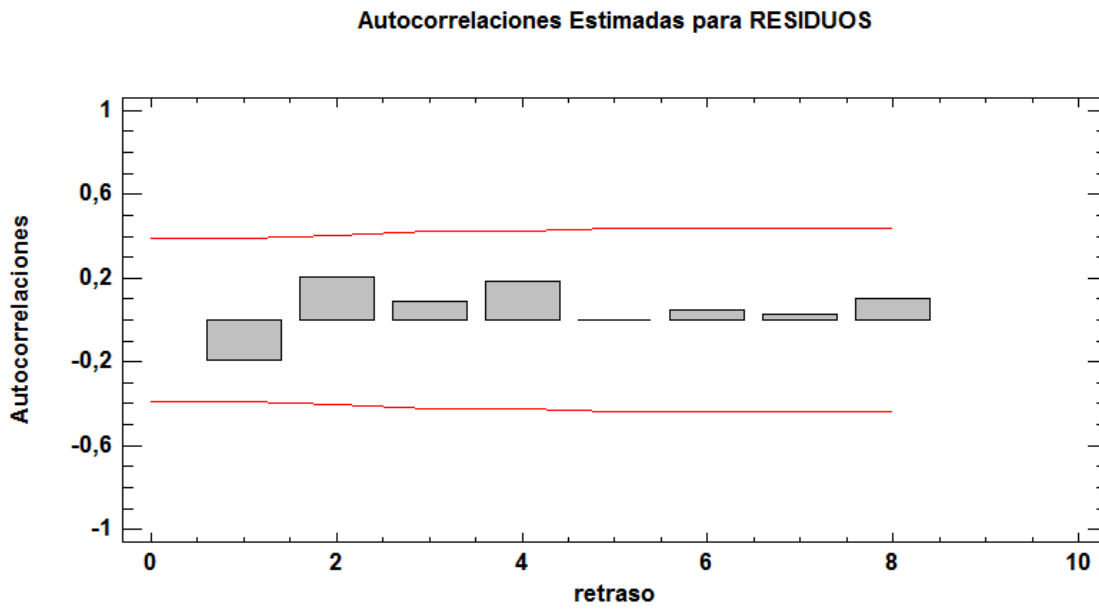


Figura IV. 48 FAS Log

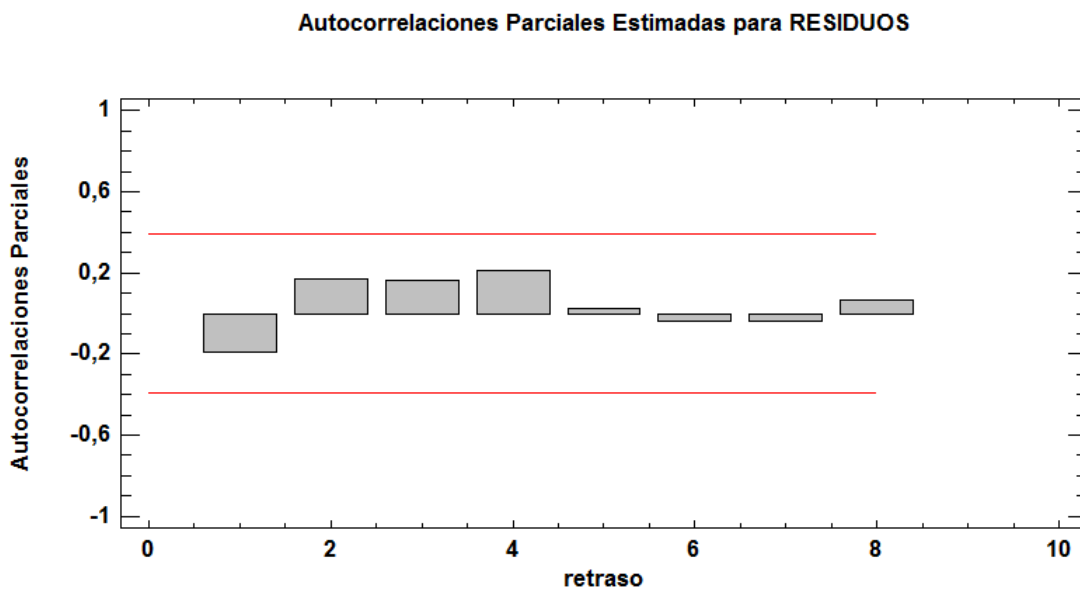


Figura IV. 49 FAP Log

IV.Cálculos y resultados

4.9.6. MODELO LOGARÍTMICO

Después de la validación del modelo con todo significativo queda la siguiente estimación (13):

$$\text{LOG(Nº_MATRICULAS)} = 0,205458 * \text{LOG(POCO_TRABAJO)} + 0,244653 * \text{LOG(HORAS_CULTURA)} + 0,683258 * \text{LOG(EMPLEO_EGRESADOS)}. \quad (13)$$

4.9. MODELO INVERSO

Se realizará ahora el modelo con la estructura inversa, validado a partir de las mismas pruebas realizadas anteriormente con el modelo lineal y el modelo logarítmico, añadiendo solo la solución final de este modelo (como se ha hecho con el modelo logarítmico). Por ello se muestran las pruebas con el modelo sin constante, y sin las variables que no eran significativas.

4.9.1. SIGNIFICATIVIDAD

En la estructura inversa eliminando la constante, solo ha quedado una variable significativa, 1/EMPLEO_EGRESADOS. Con ello, el modelo es significativo y el R-cuadrado aumenta de forma considerable con respecto a la del modelo inverso inicial estimado.

Tabla IV. 26 Prueba de significatividad inversa

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
1/EMPLEO_EGRESADOS	4956,63	235,532	21,0444	0,0000

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	113607,	1	113607	442,86	0,0000
Residuo	6413,18	25	256,527		
Total	120020,	26			

R-cuadrada = 94,6566 por ciento

R-cuadrado (ajustado para g.l.) = 94,6566 por ciento

4.9.2. FALTA DE LINEALIDAD

En principio no se observa de forma clara la forma de parábola que indique la presencia del problema de falta de linealidad (Figura IV. 50 y Figura IV. 51). Pero, se detecta un principio de parábola en los gráficos, y se decide comprobar de nuevo si hay problema mediante la elevación al cuadrado de la variable explicativa $1/EMPLEO_EGRESADOS$, pero como ha ocurrido anteriormente, se descarta la acción en cuanto se observó que el ajuste del modelo disminuía de forma considerable.

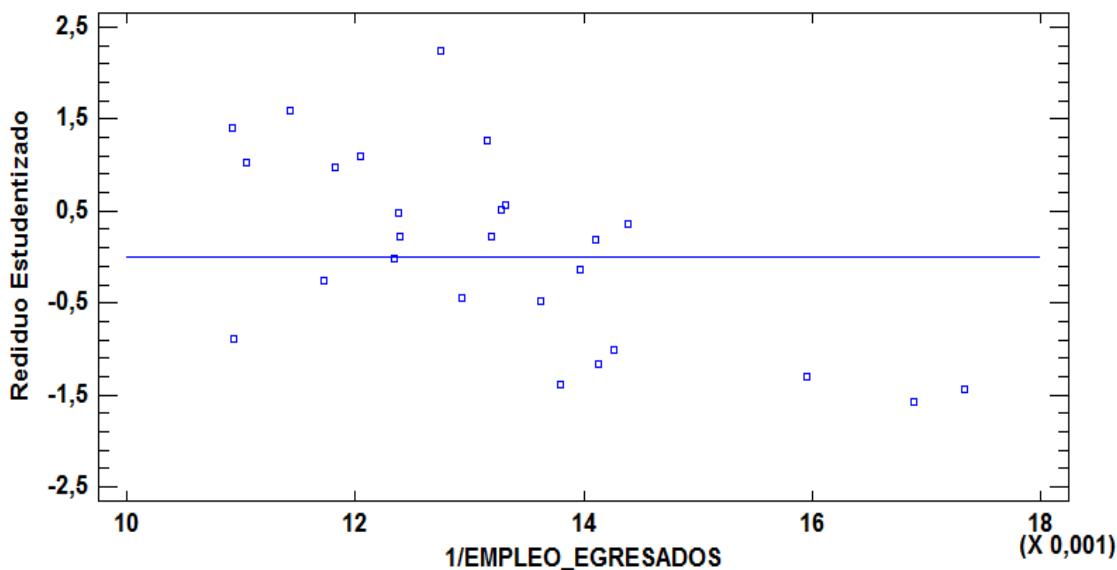


Figura IV. 50 Empleo egresados inversa

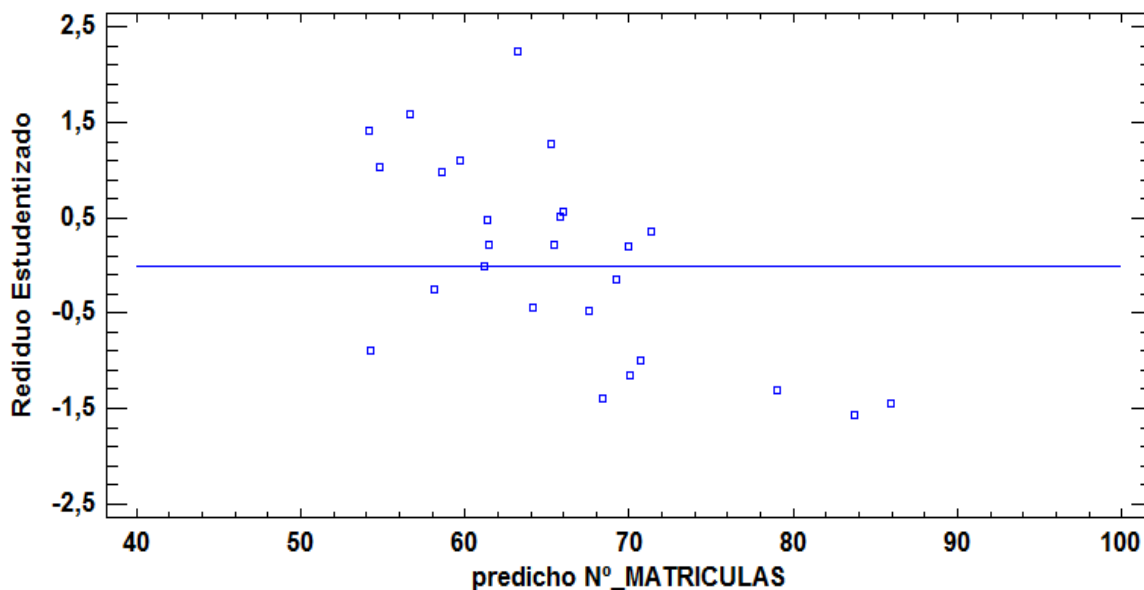


Figura IV. 51 Vs Y predicho inversa

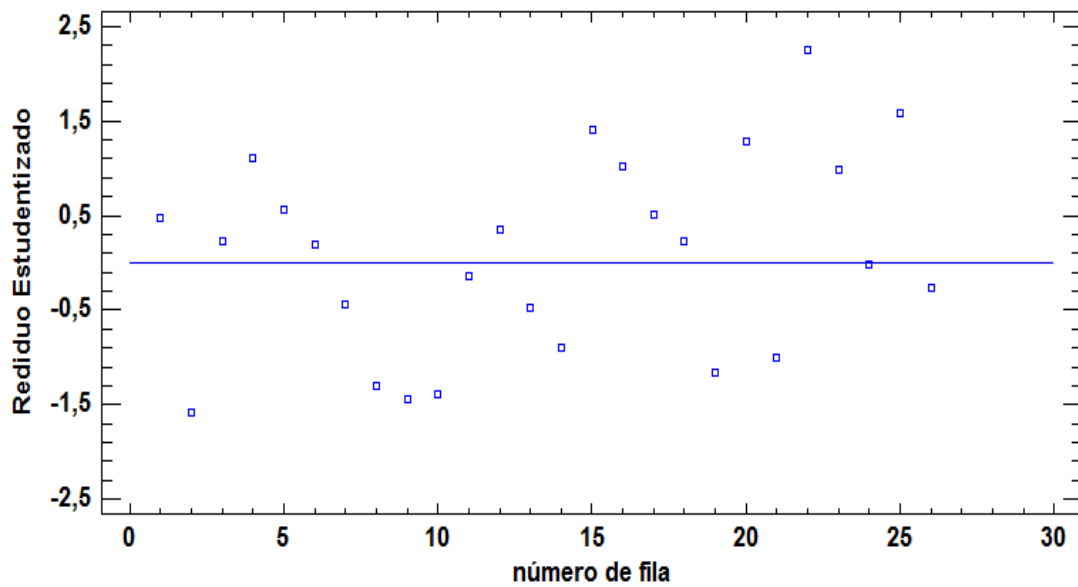


Figura IV. 52 Vs nº de filas inversa

Por último, en el gráfico de la Figura IV. 52, no se detecta la presencia de autocorrelación positiva, aunque esto deberá comprobarse en los gráficos del FAS y FAP correspondientes.

4.9.3. NORMALIDAD

En esta estructura tampoco se detectan problemas de falta de normalidad de los residuos, ya que todas las pruebas obtienen P-Valores mayores al nivel de significación tomado y por lo tanto, se afirma que los residuos se distribuyen normalmente.

Tabla IV. 27 Prueba de normalidad inversa

Pruebas de Normalidad para RESIDUOS

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	10,6154	0,476024
Estadístico W de Shapiro-Wilk	0,961657	0,444651
Valor-Z para asimetría	0,0700601	0,94414
Valor-Z para curtosis	-1,28532	0,198682

IV.Cálculos y resultados

4.9.4. HETEROCEDASTICIDAD

La varianza de los residuos del modelo inverso es constante, como puede verse en el P-Valor de la variable, o el modelo de los residuos frente a 1/EMPLEO_EGRESADOS (Tabla IV. 28).

Tabla IV. 28 Prueba de heterocedasticidad

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	-28,1212	418,105	-0,0672587	0,9469
1/EMPLEO_EGRESADOS	20760,6	31351,3	0,662191	0,5142

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	29886,4	1	29886,4	0,44	0,5142
Residuo	1,63576E6	24	68156,5		
Total (Corr.)	1,66564E6	25			

4.9.5. AUTOCORRELACIÓN

Por último, se realizan las pruebas de autocorrelación.

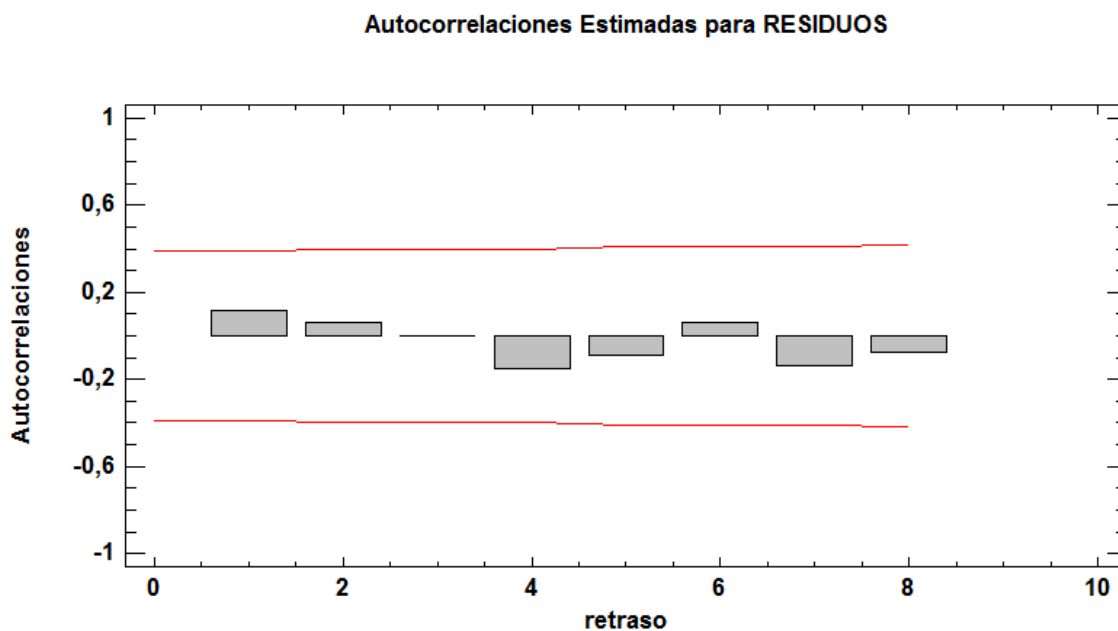


Figura IV. 53 FAS inversa

Autocorrelaciones Parciales Estimadas para RESIDUOS

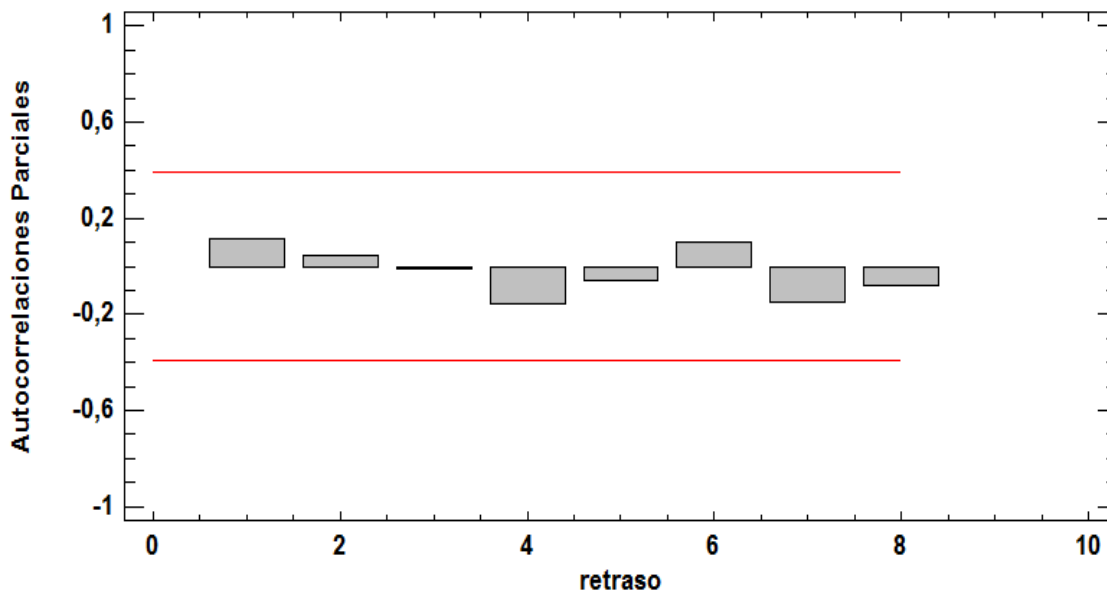


Figura IV. 54 FAP inversa

Se comprueba que no hay problemas de autocorrelación mediante los gráficos, que representan los factores de autocorrelación de los distintos grados posibles en el modelo de la estructura inversa (Figura IV. 53 y Figura IV. 54).

4.9.6. MODELO ESTRUCTURA INVERSA

Así pues, después de eliminar la constante y las variables no significativas en la estructura inversa se obtiene la siguiente estimación validada (14). Comentar que este modelo es válido, pero sólo se relaciona el número de matrículas con una de las variables explicativas inicialmente propuestas. Esto puede ser una ventaja por la menor exigencia de información del modelo en algunos casos, y una desventaja porque no se tienen en cuenta otros factores, que en las otras estructuras, han resultado ser determinantes en el número de estudiantes matriculados en las universidades Europeas.

$$N^{\circ}_MATRICULAS = 4956,63 * 1 / EMPLEO_EGRESADOS \tag{14}$$

4.10. RESUMEN DE LAS ESTRUCTURAS Y ELECCIÓN DEL MODELO

En este apartado se verá como han quedado los tres modelos finales, y se comparará y elegirá la mejor opción entre las tres ecuaciones planteadas, para hacer una predicción mediante la cual se terminará de comprobar la fiabilidad del procedimiento seguido en la modelización. En resumen se han obtenido las siguientes estimaciones de las estructuras válidas:

4.10.1. LINEAL

$$\begin{aligned} \text{N}^\circ_MATRICULAS = & 1,25326 * \text{POCO_TRABAJO} + 0,0000122175 * \text{HORAS_CULTURA}^4 + \\ & + 0,624782 * \text{EMPLEO_EGRESADOS} \end{aligned} \quad (15)$$

4.10.2. LOGARÍTMICO

$$\begin{aligned} \text{LOG}(\text{N}^\circ_MATRICULAS) = & 0,205458 * \text{LOG}(\text{POCO_TRABAJO}) + 0,244653 * \text{LOG}(\text{HORAS_CULTURA}) + \\ & + 0,683258 * \text{LOG}(\text{EMPLEO_EGRESADOS}) \end{aligned} \quad (16)$$

4.10.3. INVERSA

$$\text{N}^\circ_MATRICULAS = 4956,63 * 1 / \text{EMPLEO_EGRESADOS} \quad (17)$$

Para tomar la decisión en la Tabla IV. 29 se indica si existe algún problema, tras la finalización y validación de los modelos planteados. Además se recoge el ajuste de cada uno de ellos. En verde se señala aquel que resulta más interesante según los resultados obtenidos.

Tabla IV. 29 Estructuras de los modelos

	NORMALIDAD	SIN HETEROCEDASTICIDAD	SIN AUTOCORRELACIÓN	R^2
LINEAL	NO	NO	NO	98,7325%
LOGARITMICO	NO	NO	NO	99,9222%
INVERSA	NO	NO	NO	94,6566%

IV.Cálculos y resultados

Después de realizar las tres pruebas necesarias para obtener los tres modelos se observa que todos tienen un modelo muy significativo para realizar predicciones, con lo que la decisión final se basará en el ajuste del modelo a los datos reales empleados en la modelización. Así que, se selecciona la estructura del modelo logarítmico para la realización de predicciones del flujo de Número de matrículas universitarias. Recordar además, en la estructura lineal, pese a obtenerse un ajuste del 98,7325%, este se consigue acosta de eliminar información del modelo, en concreto la del país Malta. Por lo que con el Logarítmico, no solo se está obteniendo un modelo más ajustado a los datos reales observados, sino que se tienen en cuenta todos los países de los que se ha dispuesto de la información introducida inicialmente.

V. Predicciones

5. PREDICCIÓN

En este apartado se realizará una predicción con el modelo escogido (18), el logarítmico, ya que ha sido el escogido puesto que su R-cuadrado (99,9222%), siendo el más alto de los tres modelos calculados anteriormente. Y posteriormente, se escogerán unos datos reales de las variables del modelo encontradas en la base de datos del Eurostat del año 2013 los cuales se utilizarán en esta predicción [5].

El modelo ha quedado de esta manera:

$$\text{LOG}(\text{N}^{\circ}\text{MATRICULAS})=0,205458*\text{LOG}(\text{POCO_TRABAJO}) + 0,244653*\text{LOG}(\text{HORAS_CULTURA}) + 0,683258*\text{LOG}(\text{EMPLEO_EGRESADOS}) \quad (18)$$

5.1. CÁLCULO DE LA PREDICCIÓN

Se escogerán los datos de Alemania y España de la base de datos del 2013, ya que son los datos completos más próximos al día de hoy. Los cuáles serán:

- Horas de cultura niños: España 25,9 y Alemania 27,3 (media de horas semanales)
- Poco trabajo: España 16,7 y Alemania 10,3 (media de horas de trabajo en porcentaje)
- Empleo egresados: España 59,9 y Alemania 89,7 (tasa en porcentaje de graduados que han sido empleados)

Posteriormente, se pondrán al final de la tabla de la base de datos (Tabla IV. 1) creada para este proyecto, con las variables y países escogidos, en el programa Statgraphics y así, poder calcular una regresión múltiple con todos los datos juntos (año 2011 y año 2013) y, de esta manera, obtener el pronóstico (Tabla V. 1) calculado para la variable independiente “número de matrículas”.

Tabla V. 1 Pronóstico ajustados por el software del LOG(EMIGRACIÓN)

		Ajustado	Inferior 95,0%	Superior 95,0%
País	Fila		LC para Pronóstico	LC para Pronóstico
ESPAÑA	26	4,17096	3,89011	4,45181
ALEMANIA	27	4,36044	4,09529	4,6256

El cálculo de la variable Y (Número de matrículas) se realizó en la plantilla Excel con la herramienta EXP. Con lo que los valores reales de la predicción, tanto del valor puntual estimado, como el de los intervalos de confianza, son los que se muestran en la Tabla V. 2.

Tabla V. 2 Pronóstico calculado

		Ajustado	Inferior 95,0%	Superior 95,0%
País	Fila		LC para Pronóstico	LC para Pronóstico
ESPAÑA	26	65	50	86
ALEMANIA	27	78	60	102

Después de estos cálculos realizados se contrastará con los datos reales del número de matrículas universitarias encontradas en la base de datos de los países España y Alemania del año 2013.

El dato real encontrado de España es 87,06577% que, por muy poco, se sale del intervalo calculado anteriormente en la Tabla V. 2, por lo que se puede dudar si este modelo podría ser fiable.

El dato real encontrado de Alemania es 61,05642% que, aunque muy cerca de la parte inferior, queda dentro del intervalo calculado en la Tabla V. 2, por lo que se puede decir que es un modelo fiable.

VI. Conclusiones

6. CONCLUSIONES

Este proyecto se ha basado en una investigación macroeconómica sobre el por qué la gente estudia en las universidades en el entorno europeo; su base y motivación ha sido encontrar noticias, estudios y estadísticas referidas al entorno universitario en los países de Europa.

Básicamente, el objetivo de este proyecto ha sido modelizar mediante técnicas de regresión múltiple, el comportamiento de los estudiantes universitarios a la hora de matricularse en una carrera universitaria. Con la ayuda de cálculos y pruebas de validación del modelo se han podido determinar cuáles de las variables dependientes elegidas al principio han sido significativas para los tres modelos planteados (lineal, logarítmico e inverso), los cuales han resultado ser válidos, y que han tenido una relación con la variable independiente.

En este estudio se han empleado variables dependientes que han ayudado a las predicciones tales como la renta, ayudas, horas dedicadas a la cultura de los futuros estudiantes (niños), empleo de egresados, PIB, poco trabajo, número de universidades, desempleo y gasto universitario por alumno mientras que, la variable independiente es el número de matrículas universitarias.

Después de que se hayan realizado las técnicas de análisis necesarias en los tres modelos, cada uno de ellos ha obtenido un resultado diferente, pero todos ellos con un R-cuadrado bastante elevado, lo que quiere decir que las predicciones que se han realizado son fiables.

En el primero, modelo lineal, se han eliminado la constante, el país Malta de la base de datos y unas variables dependientes que no han sido significativas para el modelo, las cuales han sido PIB, Ayudas, Paro, Nº Universidades y Gasto por alumno. Con la eliminación de estos problemas el R-cuadrado ajustado ha resultado ser 98,7325% y, finalmente, el modelo lineal final ha sido:

$$N^{\circ}\text{MATRICULAS}=1,25326*\text{POCO_TRABAJO}+0,0000122175*\text{HORAS_CULTURA}^4+0,624782*\text{EMPLEO_EGRESADOS}.$$

En el segundo, el modelo logarítmico, se han eliminado las mismas variables y la constante, pero, no ha sido eliminado el país Malta. El R-cuadrado ha resultado ser de 99,9222% después de haber subsanado los problemas. Este modelo ha sido el más exacto de los tres y ha sido el que se ha utilizado para hacer predicciones futuras. El modelo logarítmico final ha sido:

$$\text{LOG}(N^{\circ}\text{MATRICULAS})=0,205458*\text{LOG}(\text{POCO_TRABAJO})+0,244653*\text{LOG}(\text{HORAS_CULTURA})+0,683258*\text{LOG}(\text{EMPLEO_EGRESADOS}).$$

La manera en la que se ha visto afectado el modelo y la variable independiente con sus variables dependientes significativas ha sido de forma positiva, es decir, para poder aumentar el número de matriculados universitarios la gente debe de tener un horario laboral bajo (pocas horas de trabajo), un buen nivel de cultura, ya que, a más cultura que tenga la gente, más posibilidad hay de que el número

VI. Conclusiones

de matriculados suba y, para la mayoría de personas se podría decir que es la variable más importante, es que al finalizar los estudios universitarios los egresados obtengan un empleo de la rama que han estudiado.

En el tercer y último modelo que se ha realizado, el inverso, se ha eliminado la constante y todas las demás variables menos una, la única que ha resultado significativa para este modelo. El R-cuadrado ha resultado ser de 94,6566% y el modelo inverso final ha sido:

$$N^{\circ}_MATRICULAS = 4956,63 * 1/EMPLEO_EGRESADOS.$$

Y por último, se ha hecho una predicción con el mejor modelo. Se han escogido unos datos reales de las variables del modelo encontradas en la base de datos del Eurostat del año 2013 y, con ellos, se han sustituido las variables del modelo final. Se han elegido los datos de Alemania y España. Aunque uno de los dos no han estado dentro del intervalo, hay que recalcar que los valores reales encontrados en 2012-2013 han estado muy cerca en el caso de España, y dentro de los intervalos de confianza calculados en el modelo seleccionado en el caso de los datos de Alemania, en la que el resultado real, ha estado muy cerca de la parte inferior del intervalo del pronóstico.

Y para finalizar, comentar el estudio y las noticias encontradas sobre las influencias y motivaciones de las personas a la hora de escoger una carrera universitaria con variables socioeconómicas en Europa, comparado con el modelo que se ha realizado con variables macroeconómicas europeas. En general, en la parte social, las personas eligen una carrera universitaria según la influencia de sus padres o amigos, por el reconocimiento social o, simplemente, porque les gustaría estudiar un tipo de carrera específico. Por otra parte, también hay otras variables socioeconómicas que influyen a la hora de elegir una carrera universitaria, ya sean los costes, la distancia o la imagen de la universidad a las que les gustaría ir.

Basándose en el modelo logarítmico que se ha realizado, el cual ha sido el que mejor resultado ha obtenido, las variables macroeconómicas que han influido a matricularse en la universidad son tres; tener un horario laboral mínimo, con el que se pueda, quizás, compaginar los estudios y, a su vez, pagarlos; las horas de cultura que los niños hayan dedicado durante su infancia y el futuro empleo de los egresados en un trabajo de la rama que se ha estudiado.

VII. Bibliografía

VII. Bibliografía

[1] <http://noticias.universia.es/educacion/noticia/2015/04/27/1123936/estudiar-millonario.html>.
Fecha de consulta 2015

[2] <http://www.20minutos.es/noticia/2495406/0/aumenta-numero-estudiantes/buscando-empleo/>
Fecha de consulta 2015

[3] <http://www.mentesliberadas.com.ar/2013/05/19/por-que-estudiar-una-carrera-universitaria/>
Fecha de consulta 2015

[4] http://ec.europa.eu/public_opinion/archives/ebs/ebs_369_en.pdf. Fecha de consulta 2015

[5] <http://ec.europa.eu/eurostat> . Fecha de consulta 2016

[6] <http://datos.bancomundial.org/>. Fecha de consulta 2015

<http://www.webometrics.info/es/node/35> Fecha de consulta 2015