



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

CAMPUS D'ALCOI

# *Análisis e identificación de los principales factores que afectan las visitas turísticas en España*

---

**MEMORIA PRESENTADA POR:**

*Paula Micó Penadés*

GRADO DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Convocatoria de defensa: Julio 2018



## RESUMEN

Este proyecto consta de dos partes, una de análisis de los factores que afectan al turismo y otra de la identificación de estos factores.

En primer lugar, se analizan los factores mediante búsqueda de información, proyectos, etc. Mientras que, en segundo lugar, se realiza una identificación de cuáles son los factores más importantes.

A partir de ahí, se realizan las diferentes pruebas econométricas para modelizar y conseguir un modelo óptimo.

Una vez obtenido el modelo óptimo se realizan unas predicciones y, por último, se sacan unas conclusiones.

**PALABRAS CLAVE:** análisis estadístico, modelos de regresión, turismo, multicolinealidad, identificación

## **SUMMARY**

This project consists of two parts, one of analysis of the factors that affect tourism and another of the identification of these factors.

First, the factors are analyzed through information search, projects, etc. While, in the second place, an identification is made of which are the most important factors.

From there, the different econometric tests are performed to model and achieve an optimal model.

Predictions are made and, finally, conclusions are drawn.

**KEY WORDS:** statistical analysis, regression models, tourism, multicollinearity, identification

## INDICE

Capítulo 1: Introducción .....	1
1.1 INTRODUCCIÓN .....	2
1.2 JUSTIFICACIÓN.....	3
1.3 OBJETIVOS .....	4
1.4 ESTRUCTURA.....	5
Capítulo 2: Metodología y Técnicas .....	7
2.1 ESTADO DEL ARTE .....	8
2.2 TÉCNICAS DE ANÁLISIS.....	10
Capítulo 3: Elaboración del modelo econométrico.....	15
3.1 DESCRIPCIÓN DE LOS DATOS .....	16
3.2 ANÁLISIS DESCRIPTIVO DE LAS VARIABLES CONSIDERADAS .....	18
3.4 BIVARIANTE.....	21
3.4 MULTIVARIANTE – MODELO DE REGRESIÓN.....	25
3.5 RESULTADOS DE MULTICOLINEALIDAD .....	38
Capítulo 4: Identificación de los principales factores.....	41
4.1 IDENTIFICACIÓN DE LOS FACTORES.....	42
4.2 ELECCIÓN MODELO OPTIMO .....	44
4.3 COMPROBACIÓN DE LAS HIPÓTESIS INICIALES PARA EL MODELO FACTORIAL.....	45
4.4 PREDICCIONES.....	55
Capítulo 5: Conclusiones.....	59
BIBLIOGRAFÍA .....	61
ANEXO.....	63

## TABLAS

Tabla 1. Resumen estadístico Viajeros .....	18
Tabla 2. Resumen Estadístico Gasto total viaje.....	19
Tabla 3. Resumen Estadístico Habitantes .....	20
Tabla 4. Matriz de Correlaciones .....	24
Tabla 5. Matriz de correlación .....	28
Tabla 6. Matriz inversa de correlación .....	28
Tabla 7. Matriz inversa de correlación .....	29
Tabla 8. Matriz inversa de correlación .....	29
Tabla 9. Matriz inversa de correlación .....	30
Tabla 10. Matriz inversa de correlación .....	30
Tabla 11. Matriz inversa de correlación .....	31
Tabla 12. Matriz inversa de correlación .....	31
Tabla 13. Matriz inversa de correlación .....	32
Tabla 14. Matriz inversa de correlación .....	32
Tabla 15. Porcentajes componentes.....	34
Tabla 16. Composición de los componentes .....	34
Tabla 17. Porcentajes factores.....	37
Tabla 18. Composición factores .....	37
Tabla 19. Método eliminación variables .....	38
Tabla 20. Método componentes principales .....	39
Tabla 21. Método análisis factorial .....	40
Tabla 22. Normalidad de residuos .....	46
Tabla 23. Contraste de hipótesis.....	50
Tabla 24. Significatividad del modelo .....	52
Tabla 25. Significatividad de los factores .....	53
Tabla 26. Modelo óptimo .....	54
Tabla 27. Regresión Viajeros Grupo 1 .....	55
Tabla 28. Regresión Viajeros (Grupo 2).....	56
Tabla 29. Regresión Viajeros (Grupo 3).....	57

## GRAFICOS

Gráfico 1. Histograma Viajeros.....	18
Gráfico 2. Histograma Gasto total viaje .....	19
Gráfico 3. Histograma Habitantes.....	20
Gráfico 4. Viajeros vs Gasto total viaje.....	21
Gráfico 5. Viajeros vs Gasto medio viajes .....	22
Gráfico 6. Viajeros vs Gasto medio diario.....	23
Gráfico 7. Sedimentación .....	33
Gráfico 8. Sedimentación .....	36
Gráfico 9. Gráfico de Probabilidad Normal .....	45
Gráfico 10. Histograma de Residuos.....	46
Gráfico 11. Residuos vs Factor 1 .....	47
Gráfico 12. Residuos vs Factor 2 .....	47
Gráfico 13. Residuos vs Factor 3 .....	48
Gráfico 14. Residuos vs Factor 4 .....	48
Gráfico 15. Residuos vs Viajeros .....	49
Gráfico 16. Residuos vs nº fila.....	50
Gráfico 17. FAS.....	51
Gráfico 18. FAP.....	52





# Capítulo 1

## Introducción

En este capítulo se hace una introducción sobre el trabajo y un resumen de todas las partes que van a constar en él, así como la importancia del trabajo y por qué se ha elegido este tema.

También están marcados los objetivos que se pretenden alcanzar con la realización del mismo.

## 1.1 INTRODUCCIÓN

El turismo, según la Organización Mundial del Turismo (OMT), “es un fenómeno social, cultural y económico relacionado con el movimiento de las personas a lugares que se encuentran fuera de su lugar de residencia habitual por motivos personales o profesionales”.

Según una noticia de la página web “deloitte” afirma que el sector turístico en España ha batido un nuevo récord. Durante 2016 ha incrementado la visita de turistas un 10,6% con respecto a 2015, siendo un total de 75 millones de turistas. Se espera que esta tendencia positiva continúe. [1]

Los turistas residentes y los extranjeros se alojan en los hoteles españoles prácticamente en la misma proporción. Los españoles suponen un 49,71% mientras que los extranjeros un 50,29%. Por otra parte, los hoteles de 4 estrellas es el preferido mayoritariamente por ambos, y el de 5 estrellas es más frecuente en los turistas extranjeros.

Madrid es el principal destino turístico urbano español, y uno de los principales de Europa. Por otra parte, Málaga también tiene muchos turistas, es denominada la capital de la Costa del Sol y busca quitarse la etiqueta de destino de sol y playa, para posicionarse como destino de turismo cultural, una ciudad de vanguardia.

He elegido el tema del turismo en España, ya que es un tema de actualidad, es una actividad que desde los últimos años ha experimentado un gran crecimiento. Son muchas las ciudades que sobreviven gracias al turismo, así como el sector de la hostelería.

## 1.2 JUSTIFICACIÓN

La elección de este tema ha sido principalmente conocer las estadísticas sobre el turismo en España, ya que dentro de este sector es una de las principales potencias del mundo.

Me interesa conocer tal y como está el turismo en España actualmente, cuales son los lugares más visitados, etc.

Una de las elecciones de este proyecto ha sido que está relacionado con las asignaturas de Introducción a la Estadística, Métodos Estadísticos, Economía y Econometría, la cual me gustó mucho y me parece interesante ya que, mediante un estudio de los datos, se puede sacar mucha información y realizar predicciones que es la finalidad de este trabajo, sacar un buen modelo en el cual se puedan realizar buenas predicciones.

En concreto, el turismo es muy importante en la economía española, representa el 11% del PIB, genera 2,3 millones de puestos de trabajo y contribuye a equilibrar la balanza de pagos. [14]

### 1.3 OBJETIVOS

El objetivo que se pretende conseguir con este trabajo es **analizar estadísticamente variables más importantes que afectan al turismo en España.**

Se persigue explicar el comportamiento de la demanda turística en España, medida como número de viajeros, a través de factores relevantes, como por ejemplo el índice de Precios Hoteleros (IPH), el gasto total del viaje, la tasa riesgo de pobreza, la duración media de los viajes, la tasa de actividad, la tasa de paro, etc.

Los objetivos específicos que se persiguen son:

- Buscar información sobre el turismo en España.
- Analizar las variables más importantes.
- Buscar otros trabajos relacionados.
- Identificar las variables más importantes.
- Generar el modelo óptimo aplicando las diferentes técnicas
- Realizar predicciones

## 1.4 ESTRUCTURA

Se estructura en diferentes partes. El capítulo 1 consta de una introducción al trabajo. En el capítulo 2, se ha realizado un largo análisis de búsqueda de información, como ha sido seleccionar diferentes trabajos sobre estudios de turismo en España, con esos trabajos se han evaluado las variables más importantes y las cuales se repetían en los distintos trabajos. Entonces, una vez evaluadas las variables se han buscado a través del INE.

También se explican las técnicas y la metodología a seguir para resolver el modelo creado a través de las diferentes variables.

En el capítulo 3, se describen los distintos tipos de datos que se han utilizado, así como de la variable dependiente y las variables independientes.

En el capítulo 4, se identifican los principales factores y cuál de los métodos de resolución de multicolinealidad ha sido el que consigue un modelo mejor.

En el capítulo 5, una vez finalizados los trabajos de búsqueda de información y análisis de ello, se han realizado los cálculos prácticos sobre los datos a analizar.

Se calcula un análisis de correlación, con el fin de comprobar que variables están más relacionadas.

Se aplican las diferentes técnicas de normalidad, autocorrelación, heterocedasticidad, etc.

Y, por último, en los capítulos 6 y 7 se realizan unas predicciones y con ello, unas conclusiones respectivamente.



# Capítulo 2

## Metodología y Técnicas

Es uno de los principales capítulos, ya que se analizan los estudios en los cuales se basa este trabajo.

En este capítulo se van a explicar las diferentes técnicas estadísticas estudiadas para realizar este trabajo (análisis univariante, bivariante, multivariante), que serán aplicadas en el capítulo 3.

Se exponen las técnicas que se utilizan para eliminar el problema de la multicolinealidad como son: análisis de componentes principales, análisis factorial y método de eliminación de parámetros.

## 2.1 ESTADO DEL ARTE

En este trabajo primeramente está centrado en obtener un análisis de búsqueda de datos, con ello se pretende que los datos no estén censurados y elegir las variables más importantes y que aporten mayor información.

Para ello, se han buscado trabajos con características similares, de los cuales, se han encontrado 3 TFG, y todos están centrados en hacer un modelo de regresión sobre el turismo en España.

En el primero de ellos, el documento de **Modelo Econométrico del Turismo en España** realizado por Silvia Vezanzones González en la Universidad de Valladolid, año 2015, la variable dependiente son los viajeros, mientras que las independientes son nº establecimientos abiertos en España, personal contratado, IPH, PIB y el empleo. [11]

El objetivo que tiene es explicar el comportamiento de la demanda turística en España, medida a través de los viajeros.

El trabajo está estructurado en diferentes partes tal y como lo está este mismo.

Como conclusiones que se obtienen de este trabajo son:

- La situación económica del país influye mucho en la demanda turística.
- Los precios de los servicios turísticos también afectan de forma importante.
- La crisis económica afecta al turismo en España.

En el segundo, el documento de **Modelo econométrico de demanda turística para España** realizado por Alfredo de Paz Sastre en la Universidad de Valladolid, año 2015 la variable dependiente son los viajeros, y las independientes son IPC, IPH, PIB, población, tasa paro, tasa actividad, establecimientos hoteleros, plazas hoteleras y personal ocupado. [12]

El objetivo principal de este trabajo es analizar los factores que influyen en la demanda turística en España, y para ello se elabora un modelo econométrico que reúna las variables de mayor relevancia para el análisis.

El trabajo es estructurado en 3 partes diferentes.

Y como conclusiones de este trabajo se obtienen las siguientes:

- En España la actividad turística es una de las más importantes.
- Las CCAA que reciben más viajeros son Cataluña, Andalucía, las Islas (Baleares y Canarias), y Madrid.
- El modelo presentaba problemas de heterocedasticidad pero se han resuelto con las variables en logaritmos.



Y en el tercero, el documento **Modelo Econométrico sobre el Turismo** realizado por Ruth Rubio Rodríguez y Miriam Gómez Sánchez la variable dependiente es el precio hotelero y las independientes son nº de pernoctaciones, estancia media de los turistas, personal empleado, ingresos turismo y gastos en turismo. [13]

Este trabajo está enfocado al turismo en Granada, es más concreto ya que se centra solamente en una única ciudad.

El objetivo principal es observar la evolución del sector turístico en Granada en los últimos años.

Los tres utilizan un tipo de estructura similar, en el cual todos tienen una introducción y explicación sobre de que va a constar el trabajo, descripción de las variables y planteamiento del modelo. Seguidamente todos han realizado las pruebas necesarias para validar el modelo y por último y más importante las previsiones realizadas sobre el modelo econométrico válido.

A partir de estos tres trabajos, se ha podido comprobar qué variables eran las más importantes y con ello, se buscan los datos en el INE. Por lo tanto, las conclusiones que se obtienen de estos trabajos son las siguientes:

- La variable dependiente a analizar más importante son los viajeros.
- Otras variables de interés han sido PIB, IPH, tasa de paro, tasa de actividad, estancia media de los turistas...
- La estructura a seguir.

## 2.2 TÉCNICAS DE ANÁLISIS

Ahora se van a explicar las diferentes técnicas estadísticas utilizadas para el proyecto:

El **análisis univariante** está constituido por un conjunto de técnicas cuyo objetivo es clasificar, presentar, describir, resumir y analizar los datos relativos a una o más características de individuos de una población, a partir de la información sobre todos y cada uno de ellos. [8]

Con tal de explicar bien el análisis univariado, para ello vamos a definir varias características sobre un conjunto de elementos:

- Variable estadística: la característica de la población que se quiere estudiar. Pueden ser cualitativas (si los valores que toma no son numéricos) o cuantitativa (si los valores que toma son numéricos). A su vez, las variables cuantitativas pueden ser discretas (si toma un número finito o numerable de valores) o continua (si toma un número infinito no numerable de valores).
- Diagrama de barras: representa la distribución de frecuencias de una variable cualitativa mediante barras de longitud proporcional a la frecuencia de cada categoría.
- Histograma: representa la distribución de frecuencias de una variable cuantitativa continua. Se construye representando sobre cada intervalo un rectángulo de área proporcional a la frecuencia de la categoría.
- Media: es la suma total de los valores observados dividida entre el número total de datos.
- Mediana: de una muestra es el valor de “en medio” (el que deja 50% por arriba y 50% por debajo) cuando los valores están ordenados.
- Moda: es el valor con mayor frecuencia de una variable estadística.
- Rango: es la diferencia entre el mayor y el menor valor de la muestra.
- Rango intercuartílico: es la diferencia entre el tercer y el primer cuartil de la muestra.
- Primer cuartil: es el valor que deja por debajo de sí en el 25% de los valores de la muestra.
- Tercer cuartil: es el valor que deja por debajo de sí el 75% de los valores de la muestra.
- Segundo cuartil: es la mediana.

- Varianza: la media de las distancias al cuadrado de los valores a la media.
- Desviación típica: es la raíz cuadrada de la varianza.

Seguidamente se explican las técnicas de **análisis bivariante**, las cuales permiten el análisis conjunto de dos características de los individuos de una población con el propósito de detectar posibles relaciones entre ellas. [9]

Los datos bivariantes provienen de la observación simultánea de dos variables (X, Y) en una muestra de  $n$  individuos.

Se usan para describir las dos variables conjuntamente o una variable en función de la otra.

Se intenta describir el comportamiento de una de las variables (variable dependiente Y), en función de la otra variable (variables independientes o explicativas X). [11]

- Diagrama de dispersión: la representación gráfica más común para dos variables cuantitativas.
- Varianza residual: es una medida del error cometido en la predicción de los datos.
- Coefficiente de determinación ( $R^2$ ): mide el porcentaje de la varianza total recogido en el modelo.
- Gráfico de residuos: para ver si el modelo de regresión es adecuado, se hace un diagrama de dispersión de residuos frente a valores predichos. Los puntos parecen aleatorios indica que la recta de regresión se ajusta correctamente.

Por último, se explicará el **análisis multivariante** (modelos de regresión):

Son todas aquellas técnicas estadísticas que simultáneamente analizan múltiples resultados en los individuos u objetos bajo investigación.

Para que un análisis se considere multivariado todas las variables deben ser aleatorias y relacionadas de tal manera que el efecto que producen no pueda ser interpretado de manera individual.

¿Para qué sirve?

- Representar datos de forma inteligible.
- Esclarecer la distribución real de “varias variables”.
- Desarrollar un modelo de predicción basado en múltiples variables (modelos de regresión).
- Hallar las relaciones de causa-efecto entre variables.

El propósito del análisis multivariante es medir, explicar y predecir el grado de relación que existe entre la variación.

Etapas análisis multivariante:

- a) **Objetivos del análisis:** definir el problema, especificar los objetivos, elección de técnicas multivariantes...
- b) **Diseño del análisis:** determinar el tamaño muestral, las ecuaciones a estimar, las distancias a calcular...
- c) **Hipótesis del análisis:** escoger la formulación de hipótesis subyacente a la técnica.
  - **Multicolinealidad:** se hace referencia a la existencia de relaciones aproximadamente lineales entre dos o más variables del modelo, cuando la estimación y la precisión de estas se ven afectados. Con la finalidad de resolver este problema entre las variables se ha utilizado el método de componentes principales el cual ya se explica más tarde.
  - **Normalidad de residuos:** esta hipótesis se contrasta a partir de los residuos estandarizados. Se observan los gráficos de normalidad y el histograma.
  - **Autocorrelación:** conecta con la idea de que los errores contienen cierta persistencia y, por tanto, no se deben a factores puramente aleatorios, desconectados los unos de los otros. Así pues, cuando existe autocorrelación, el error cometido en un momento del tiempo está influido por el error de períodos previos.
  - **Heterocedasticidad:** cuando la varianza de las perturbaciones no es constante a lo largo de las observaciones. Esto implica el incumplimiento de una de las hipótesis básicas sobre las que se asienta el modelo de regresión lineal.
- d) **Realización del análisis:** se estima el modelo y se evalúa el ajuste a los datos.
- e) **Interpretar los resultados**
- f) **Validación del análisis:** establecer la validez de los resultados obtenidos analizando si los resultados obtenidos con la muestra se generalizan a la población de la que procede.

En cuanto al modelo de regresión, con tal de eliminar la correlación existente entre las variables del modelo utilizaremos tres técnicas con el fin de encontrar el mejor modelo y con mayor  $R^2$ , el cual nos explique mejor todas las variables del modelo. Las técnicas son las siguientes:

### a) Método de eliminación de variables:

Para realizar este método primeramente hay que realizar 3 pruebas para identificar la multicolinealidad, si en una de ellas indica que existe problema de multicolinealidad entonces ya existe ese problema, aunque con las otras pruebas no exista el problema de multicolinealidad. Estas pruebas son las siguientes:

1. Matriz de correlación: muestra si existe relación entre las variables.

$|R_{ii}| > 0.7$  entonces existe multicolinealidad entre dos variables.

2. Matriz inversa de correlación: muestra la relación que existe entre una variable explicativa con todas las demás.  
Se observan los elementos de la diagonal principal.

$R_{ii}^{-1} > 10$  entonces existe multicolinealidad entre una variable y el resto.

3. Índice de acondicionamiento: mide la relación de todas las variables explicativas a la vez.

$$I.C. = \sqrt{\frac{\text{Autovalor max}}{\text{Autovalor min}}}$$

Una vez realizadas las tres pruebas y que los resultados hayan dado problemas de multicolinealidad, se elimina una de las variables correlacionadas y se realiza el ajuste. Para saber cuál de las variables se ha de eliminar:

- 1º criterio: se elimina la variable más grande de la diagonal principal de la matriz inversa de correlación.  
Pero si los dos valores entre las variables más grandes son muy cercanos (la diferencia es menor que 1) entonces este criterio no sirve.
- 2º criterio: se hacen dos regresiones eliminando cada vez una de las variables que generan problemas y se elige el modelo que tiene mayor  $R^2$

### b) Método de componentes principales:

Este método está basado en obtener nuevas variables que sean combinación lineal de las variables que estén correlacionadas.

Estas nuevas variables obtenidas se utilizarán para modelizar el modelo.

Una vez obtenidas las nuevas variables de componentes principales se vuelve a hacer la regresión múltiple pero esta vez utilizando como variables explicativas los nuevos componentes y así quedarán eliminados los problemas de multicolinealidad.

**c) Método análisis factorial:**

Es una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables.

Los grupos homogéneos se forman con las variables que correlacionan mucho entre sí y procurando, inicialmente, que unos grupos sean independientes de otros.

Su propósito consiste en buscar el número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos. [10]

# Capítulo 3

## Elaboración del modelo econométrico

En este capítulo se describen los datos y las variables del modelo.

También se expone como se ha eliminado la multicolinealidad mediante los tres métodos (eliminación de variables, componentes principales y análisis factorial), elección del modelo óptimo y realización de las diferentes pruebas como son: normalidad de residuos, heterocedasticidad, autocorrelación...

### 3.1 DESCRIPCIÓN DE LOS DATOS

Se ha buscado información previa al trabajo y las variables han sido elegidas a través de los diferentes TFG de otros alumnos de otras universidades, muchas de las cuales se repetían en todos los trabajos.

La fuente de datos que se ha utilizado ha sido básicamente el INE (Instituto Nacional de Estadística), el cual proporciona la información suficiente para obtener los datos de las diferentes variables por Comunidades Autónomas. [15]

Para este trabajo los datos que se han utilizado han sido anuales (2016), ya que era el año más actual que aparecían más datos, porque los años 2017 y 2018 aun no contienen toda la información y referidos a las 17 Comunidades Autónomas, no se han considerado convenientes estudiar Ceuta y Melilla, porque en algunas de las variables no existía información para estas dos Comunidades Autónomas.

En relación con la multitud de variables turísticas que existen se han seleccionado las más importantes con acuerdo a la demanda turística, y son las siguientes:

- Variable dependiente:

**Los viajeros:** toda persona que se desplaza dos o más países distintos (viajero internacional) o entre dos o más lugares dentro de su país de residencia habitual (viajero interno). También podría definirse como persona que realiza una o más pernoctaciones (cada noche que un viajero se aloja en un establecimiento) seguidas en el mismo alojamiento. Los tipos de viajeros relacionados con el turismo se denominan visitantes, ya sean turistas (visitantes que pernoctan) o excursionistas (visitantes de un día).

El calificativo de viajero suele utilizarse para definir a una persona aventurera que con frecuencia cambia de lugar de destino en busca de nuevas experiencias. También puede ocurrir que una persona viajera tenga que realizar desplazamientos habituales por motivos de trabajo al hacer viajes de negocios.

- Variables independientes:

**Gasto total del viaje:** esta variable representa el gasto total que han tenido los viajeros durante su viaje. Esta medida en millones de €.

**Gasto medio por viaje:** es el gasto medio que se gastan los viajeros durante sus viajes en un año. Esta medido en €.

**Gasto medio diario por persona:** es el gasto medio que realizan los viajeros por día. Esta medido en €.

**Duración media de los viajes:** los días que están de viaje según la Comunidad Autónoma. Esta medido por días.

**Índice Precios Hoteleros (IPH):** permite conocer la evolución temporal de los precios recibidos por los empresarios hoteleros considerando todos sus clientes.



**Viajeros camping españoles:** esta medido en personas y representa a los españoles que cuando viajan se alojan en un camping.

**Renta media por persona:** es la renta media que reciben las personas de cada Comunidad Autónoma durante un año. Esta medida en €.

**Tasa riesgo de pobreza:** se caracteriza porque tiene una renta por debajo de lo que se considera nivel de vida medio.

**Habitantes:** está formado por todos los habitantes residentes en cada Comunidad Autónoma. Esta medido por personas.

**Índice Precios Trabajo (IPT):** mide el cambio en el precio de la mano de obra en el tiempo como consecuencia exclusivamente de las presiones del mercado laboral.

**Índice Precios Consumo (IPC):** es un índice económico en el que se valoran los precios de un predeterminado conjunto de bienes y servicios determinado sobre la base de los hogares, que una cantidad de consumidores adquiere de manera regular, y la variación con respecto del precio de cada uno, respecto de una muestra anterior.

**Tasa de actividad:** es un índice que medie el nivel de actividad en el empleo de un país. Se calcula como el cociente entre la población activa (PA) y la población en edad de trabajar o mayor de 16 años.

**Tasa de paro:** expresa el nivel de desocupación entre la población económicamente activa. Su cálculo es el cociente entre el nº de desempleados y la población activa (PA).

**PIB:** Producto Interior Bruto. Es una magnitud macroeconómica que expresa el valor monetario de la producción de bienes y servicios de demanda final de un país o región durante un período determinado, normalmente un año.

**Viajeros hotel:** esta medido en personas, y representa a los españoles que cuando viajan se alojan en un hotel.

**Viajeros turismo rural:** esta medido en personas, y representa a los españoles que cuando viajan se alojan en una casa rural.

### 3.2 ANÁLISIS DESCRIPTIVO DE LAS VARIABLES CONSIDERADAS

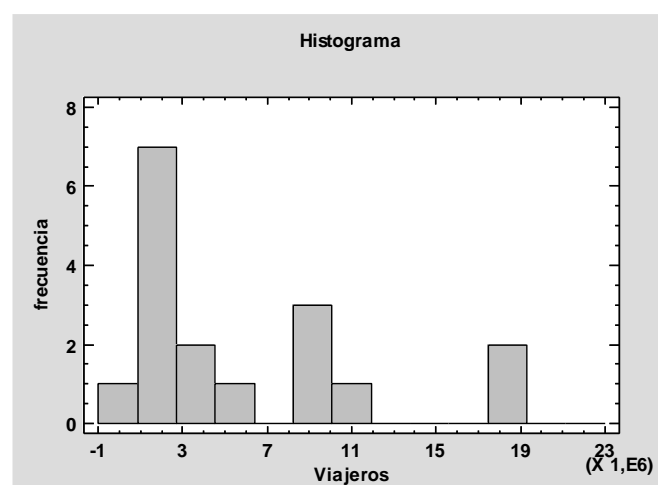
En el análisis univariante, trata de explicar las diferentes variables de una forma más descriptiva, calculando de cada una de ellas la media, la mediana, el mínimo, el máximo, etc. Y también un histograma.

- **Viajeros:**

*Tabla 1. Resumen estadístico Viajeros*

Resumen Estadístico para Viajeros	
Recuento	17
Promedio	5,86507E6
Mediana	2,83919E6
Desviación Estándar	5,88866E6
Coefficiente de Variación	100,402%
Mínimo	581746,
Máximo	1,90944E7
Rango	1,85126E7
Cuartil Inferior	1,3219E6
Cuartil Superior	9,35169E6
Rango Intercuartílico	8,0298E6
Sesgo Estandarizado	2,09858
Curtosis Estandarizada	0,507008

En esta tabla se puede observar la media de viajeros con unos 5.860.000 viajeros, la mediana con unos 2.830.000 viajeros y las demás técnicas estadísticas explicadas en el capítulo anterior.



*Gráfico 1. Histograma Viajeros*

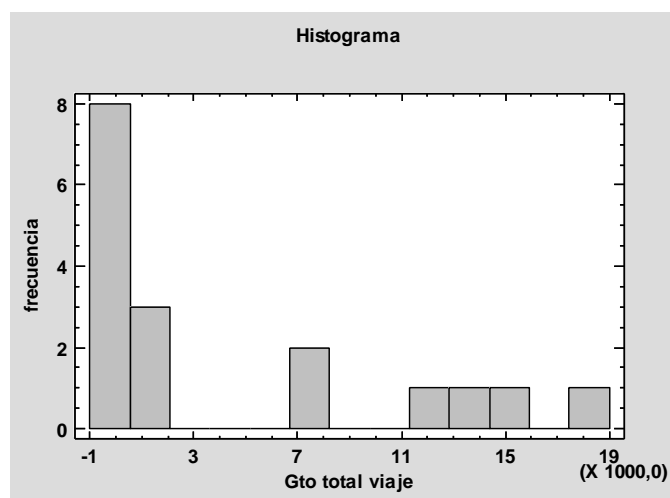
En el histograma se observa cómo se distribuyen los viajeros. No tienen una distribución normal ya que no siguen el modelo de la campana de Gauss, y se podría concluir que existen 3 grupos entre los viajeros ya que hay un grupo (Andalucía y Cataluña) que reciben muchos viajeros, otro grupo (Comunidad Valenciana, Madrid, Islas Baleares e Islas Canarias) que reciben menos viajeros y otro grupo (Aragón, Asturias, Cantabria, Castilla y León, Castilla la Mancha, Extremadura, Galicia, Murcia, Navarra, País Vasco y la Rioja) que son los que menos viajeros reciben.

- **Gasto total del viaje:**

*Tabla 2. Resumen Estadístico Gasto total viaje*

Resumen Estadístico para Gasto total viaje	
Recuento	17
Promedio	4553,53
Mediana	835,71
Desviación Estándar	6096,91
Coefficiente de Variación	133,894%
Mínimo	93,09
Máximo	17466,8
Rango	17373,7
Cuartil Inferior	315,26
Cuartil Superior	7597,41
Rango Intercuartílico	7282,15
Sesgo Estandarizado	1,85559
Curtosis Estandarizada	-0,274578

En esta tabla se observan las técnicas univariantes básicas, y se podrían destacar que la media que se gastan los viajeros en un viaje es de 4553.53€, el mínimo y el máximo son 93.09€ y 17466.8€ respectivamente.



*Gráfico 2. Histograma Gasto total viaje*

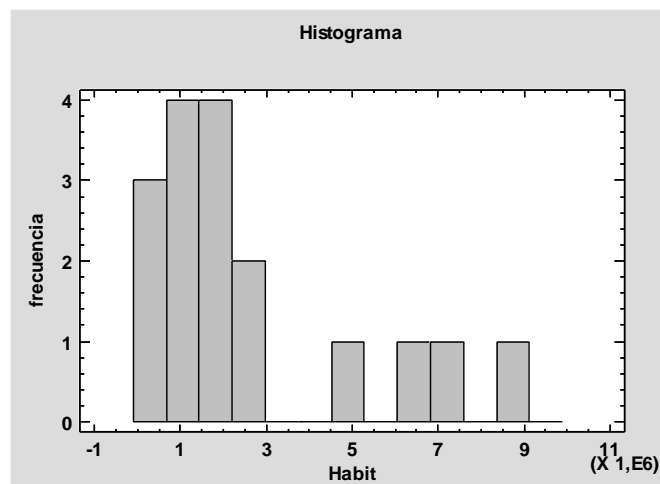
En el histograma también se puede ver que no sigue una distribución normal y se concluiría que se puede observar los tres grupos distintos de Comunidades Autónomas nombrados anteriormente.

- **Habitantes:**

*Tabla 3. Resumen Estadístico Habitantes*

Resumen Estadístico para Habitantes	
Recuento	17
Promedio	2,72947E6
Mediana	2,03148E6
Desviación Estándar	2,52779E6
Coefficiente de Variación	92,611%
Mínimo	315381,
Máximo	8,37982E6
Rango	8,06444E6
Cuartil Inferior	1,07992E6
Cuartil Superior	2,70834E6
Rango Intercuartílico	1,62842E6
Sesgo Estandarizado	2,25365
Curtosis Estandarizada	0,476233

En esta tabla tenemos las técnicas univariantes aplicadas a los habitantes, la media es de 2.720.000 habitantes y la mediana 2.030.000 habitantes.



*Gráfico 3. Histograma Habitantes*

En el histograma de habitantes se observan también los tres grupos de turistas, con lo cual nos puede indicar que haya diferentes niveles de turistas.

Se han descrito tres de las principales variables mediante técnicas univariantes, las demás tablas e histogramas del resto de variables están en el anexo.

### 3.4 BIVARIANTE

#### 3.4.1 Relación de viajeros frente a cada una de las variables

En el análisis bivalente primeramente se analiza la relación existente entre los viajeros y cada una de las variables, con la finalidad de comprobar si existe o no correlación entre ellas.

##### - **Viajeros vs Gasto total viaje:**

La correlación entre viajeros y gasto total viaje es de 0.8958, lo que significa que estas dos variables están correlacionadas.

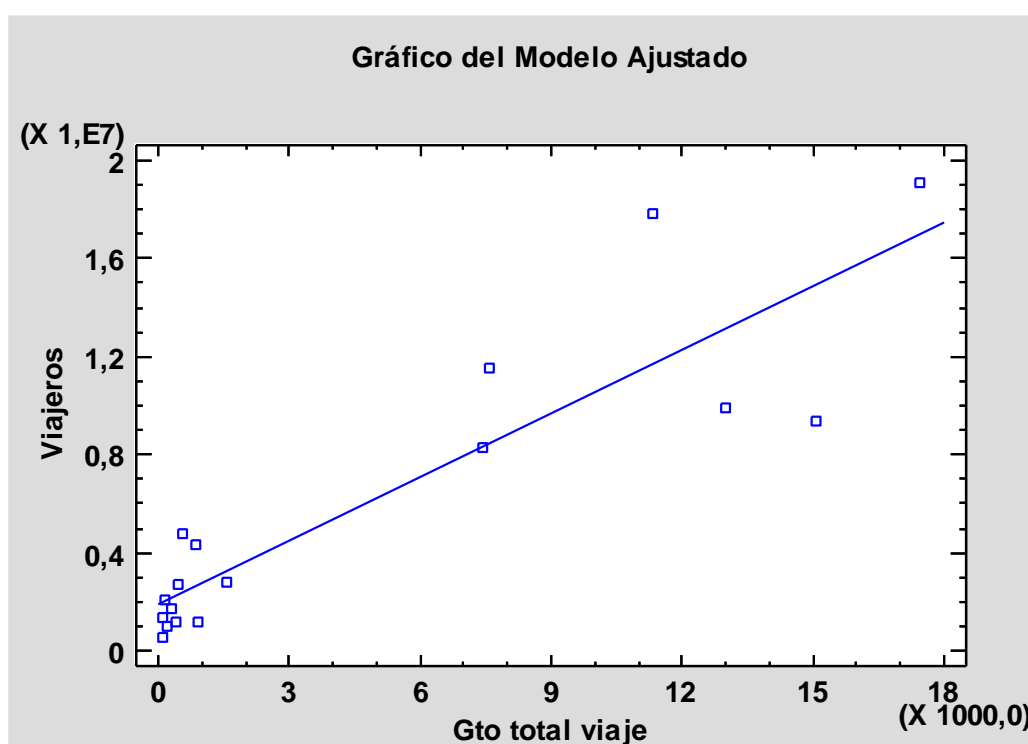
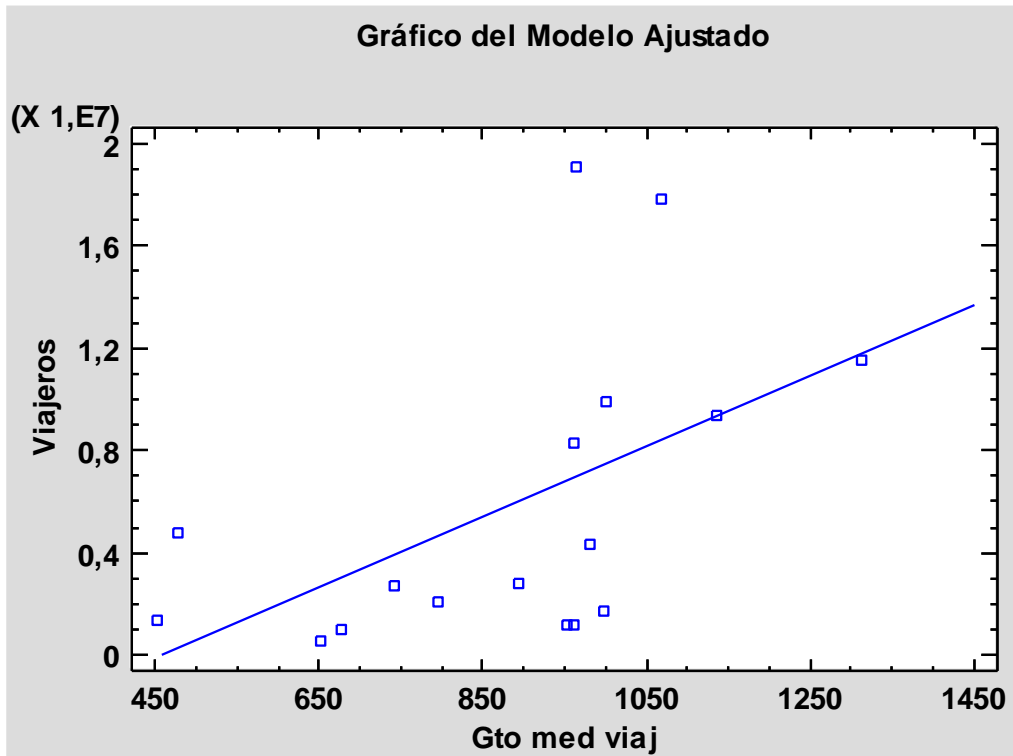


Gráfico 4. Viajeros vs Gasto total viaje

Como se observa en el gráfico los puntos están muy cerca de la línea hay puntos muy juntos, lo cual indican que tienen relación estas variables, pero también se observan puntos extremos (Andalucía y Cataluña) que son los que más turistas reciben.

- **Viajeros vs Gasto medio viaje:**

La correlación entre Viajeros y gasto medio viaje es de 0.5605, entonces se concluye que estas variables no tienen mucha relación.

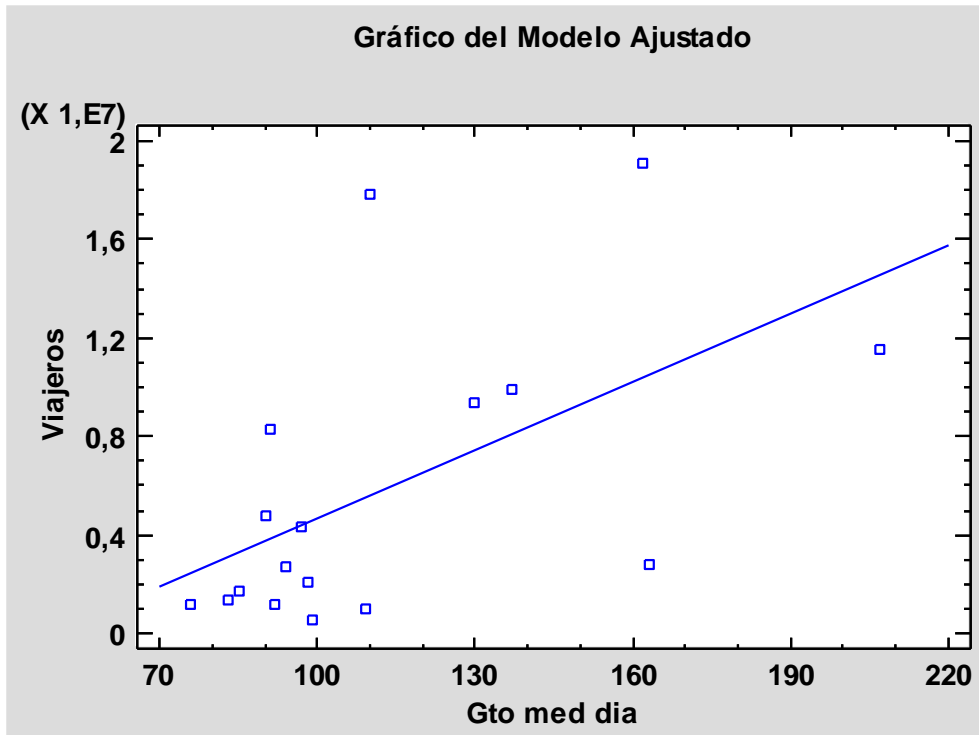


*Gráfico 5. Viajeros vs Gasto medio viajes*

En el gráfico se observan que los puntos están dispersos y no siguen una distribución cerca de la línea, lo que también indica que no tienen relación lineal, pero se puede observar cierta relación logarítmica.

- **Viajeros vs Gasto medio diario:**

La correlación existente entre viajeros y gasto medio diario es de 0.0193 entonces como es un valor cercano a 0 no existe relación entre estas dos variables.



*Gráfico 6. Viajeros vs Gasto medio diario*

En este gráfico también se distribuyen los puntos aleatoriamente y no hay ningún tipo de relación lineal.

Como se puede observar en las tres variables analizadas existe relación en una y hay dos en las cuales no existe relación. Las demás variables están en el anexo con sus respectivas correlaciones y sus gráficos de cada una.

### 3.4.2 Relación de las variables independientes

Para analizar la relación existente entre las variables independientes se analiza la matriz de correlación ya que nos muestra para cada una de las variables como está relacionada con el resto.

Tabla 4. Matriz de Correlaciones

MATRIZ DE CORRELACIONES																
	Gto total viaj	Gto med viaj	Gto med día	Dur med	IPH	Viaj camp	Rent med	Triesgpob	Habit	IPT	IPC	Tasaact	Tasaparo	PIB	Viaj hotel	Viaj turismru
Gto total viaj	1	0,565756	0,553661	-0,0824115	<b>0,734591</b>	0,577639	-0,0310971	0,239491	0,640842	0,117449	0,328257	0,587942	0,207291	0,564548	<b>0,895779</b>	0,0679184
Gto med viaj	0,565756	1	0,552877	0,436735	0,543203	0,156815	0,029158	0,147047	0,494865	0,0669086	0,0537517	0,371619	0,0461654	0,486237	0,530736	-0,316465
Gto med día	0,553661	0,552877	1	-0,487968	0,24315	0,282641	0,554493	-0,251462	0,50521	0,255273	0,22579	0,586236	-0,261373	0,676184	0,560492	-0,0249865
Dur med	-0,0824115	0,436735	-0,487968	1	0,254478	-0,117339	-0,491108	0,334205	-0,045776	-0,201889	-0,240311	-0,261316	0,238976	-0,195872	-0,111637	-0,323482
IPH	<b>0,734591</b>	0,543203	0,24315	0,254478	1	0,239109	-0,11332	0,232208	0,358595	0,263828	0,175851	0,142132	0,281398	0,230534	0,572964	-0,0888352
Viaj camp	0,577639	0,156815	0,282641	-0,117339	0,239109	1	0,112702	-0,0833065	<b>0,707235</b>	-0,0231934	0,401278	0,232934	-0,0355323	<b>0,713949</b>	<b>0,715393</b>	0,327418
Rent med	-0,0310971	0,029158	0,554493	-0,491108	-0,11332	0,112702	1	<b>-0,91852</b>	-0,0794806	0,339647	0,16552	0,0671381	<b>-0,849331</b>	0,177916	-0,0242607	0,0757427
Triesgpob	0,239491	0,147047	-0,251462	0,334205	0,232208	-0,0833065	<b>-0,91852</b>	1	0,263488	-0,300492	-0,176319	0,103984	<b>0,934702</b>	0,00843917	0,231004	-0,0992043
Habit	0,640842	0,494865	0,50521	-0,045776	0,358595	<b>0,707235</b>	-0,0794806	0,263488	1	-0,101541	0,298866	0,325899	0,256932	<b>0,936086</b>	<b>0,890637</b>	0,246405
IPT	0,117449	0,0669086	0,255273	-0,201889	0,263828	-0,0231934	0,339647	-0,300492	-0,101541	1	0,494909	0,0087539	-0,320262	-0,0809154	-0,025536	0,0163425
IPC	0,328257	0,0537517	0,22579	-0,240311	0,175851	0,401278	0,16552	-0,176319	0,298866	0,494909	1	-0,0113002	-0,176608	0,29751	0,413972	0,602161
Tasaact	0,587942	0,371619	0,586236	-0,261316	0,142132	0,232934	0,0671381	0,103984	0,325899	0,0087539	-0,0113002	1	-0,0890758	0,404292	0,453047	-0,271918
Tasaparo	0,207291	0,0461654	-0,261373	0,238976	0,281398	-0,0355323	<b>-0,849331</b>	<b>0,934702</b>	0,256932	-0,320262	-0,176608	-0,0890758	1	0,00523882	0,224472	-0,0587009
PIB	0,564548	0,486237	0,676184	-0,195872	0,230534	<b>0,713949</b>	0,177916	0,00843917	<b>0,936086</b>	-0,0809154	0,29751	0,404292	0,00523882	1	<b>0,81061</b>	0,242525
Viaj hotel	<b>0,895779</b>	0,530736	0,560492	-0,111637	0,572964	<b>0,715393</b>	-0,0242607	0,231004	<b>0,890637</b>	-0,025536	0,413972	0,453047	0,224472	<b>0,81061</b>	1	0,264065
Viaj turismru	0,0679184	-0,316465	-0,0249865	-0,323482	-0,0888352	0,327418	0,0757427	-0,0992043	0,246405	0,0163425	0,602161	-0,271918	-0,0587009	0,242525	0,264065	1

Como se puede observar en la matriz de correlación las variables marcadas con rojo son las que tienen correlación, por lo tanto, en este caso son muchas las que tienen relación entre ellas.

Las variables que a simple vista no tienen relación son:

- Gasto medio viaje
- Gasto medio diario
- Duración media
- IPT
- IPC
- Tasa actividad
- Viajeros turismo rural

Por lo demás, el resto de variables todas tienen algún tipo de relación con alguna otra.

Las variables que más pueden dar problemas serían el PIB, viajeros hotel, viajeros camping, ya que tienen relación con 3 o más variables.

Estos problemas ya se estudiarán más adelante cuando se resuelva el problema de la multicolinealidad.



### 3.4 MULTIVARIANTE – MODELO DE REGRESIÓN

El modelo de regresión planteado ha sido el siguiente:

$$\text{Viajeros} = \beta_0 + \text{Gasto total viaje} * \beta_1 + \text{Gasto medio viaje} * \beta_2 + \text{Gasto medio diario} * \beta_3 + \text{Duración media} * \beta_4 + \text{IPH} * \beta_5 + \text{Viajeros camping} * \beta_6 + \text{Renta media} * \beta_7 + \text{Tasa riesgo pobreza} * \beta_8 + \text{Habitantes} * \beta_9 + \text{IPT} * \beta_{10} + \text{IPC} * \beta_{11} + \text{Tasa de actividad} * \beta_{12} + \text{Tasa de paro} * \beta_{13} + \text{PIB} * \beta_{14} + \text{Viajeros hotel} * \beta_{15} + \text{Viajeros turismo rural} * \beta_{16} + U$$

$\beta_0$  → valor medio de viajeros cuando las variables explicativas valen 0.

$\beta_1$  → incremento medio de los viajeros cuando aumenta 1 millón € el gasto total del viaje y el resto de variables explicativas permanecen constantes.

$\beta_2$  → incremento medio de los viajeros cuando aumenta 1€ el gasto medio por persona y el resto de variables explicativas permanecen constantes.

$\beta_3$  → incremento medio de los viajeros cuando aumenta 1€ el gasto medio diario por persona y el resto de variables explicativas permanecen constantes.

$\beta_4$  → incremento medio de los viajeros cuando aumenta 1 día la duración media de los viajes y el resto de variables explicativas permanecen constantes.

$\beta_5$  → incremento medio de los viajeros cuando aumenta 1 unidad el IPH y el resto de variables explicativas permanecen constantes.

$\beta_6$  → incremento medio de los viajeros cuando aumenta 1 persona el nº de viajeros camping españoles y el resto de variables explicativas permanecen constantes.

$\beta_7$  → incremento medio de los viajeros cuando aumenta 1€ la renta media por persona y el resto de variables explicativas permanecen constantes.

$\beta_8$  → incremento medio de los viajeros cuando aumenta 1% la tasa riesgo de pobreza y el resto de variables explicativas permanecen constantes.

$\beta_9$  → incremento medio de los viajeros cuando aumenta 1 habitante y el resto de variables explicativas permanecen constantes.

$\beta_{10}$  → incremento medio de los viajeros cuando aumenta 1% el IPT y el resto de variables explicativas permanecen constantes.

$\beta_{11}$  → incremento medio de los viajeros cuando aumenta 1% el IPC y el resto de variables explicativas permanecen constantes.

$\beta_{12}$  → incremento medio de los viajeros cuando aumenta 1 unidad la tasa de actividad y el resto de variables explicativas permanecen constantes.

$\beta_{13}$  → incremento medio de los viajeros cuando aumenta 1 unidad la tasa de paro y el resto de variables explicativas permanecen constantes.

$\beta_{14}$  → incremento medio de los viajeros cuando aumenta 1€ el PIB y el resto de variables explicativas permanecen constantes.

$\beta_{15}$  → incremento medio de los viajeros cuando aumenta 1 unidad los viajeros hotel y el resto de variables explicativas permanecen constantes.

$\beta_{16}$  → incremento medio de los viajeros cuando aumenta 1 unidad los viajeros turismo rural y el resto de variables explicativas permanecen constantes.

U → perturbación del modelo econométrico.

- Unidades físicas de los parámetros:

$\beta_0$  → se mide en personas

$\beta_1$  → se mide en millones €

$\beta_2$  → se mide en €

$\beta_3$  → se mide en €

$\beta_4$  → se mide en días.

$\beta_5$  → se mide en %

$\beta_6$  → se mide en personas

$\beta_7$  → se mide en €

$\beta_8$  → se mide en %

$\beta_9$  → se mide en personas

$\beta_{10}$  → se mide en %

$\beta_{11}$  → se mide en %

$\beta_{12}$  → se mide en %

$\beta_{13}$  → se mide en %

$\beta_{14}$  → se mide en €

$\beta_{15}$  → se mide en personas

$\beta_{16}$  → se mide en personas

Una vez descrito el modelo, se analiza la multicolinealidad del modelo, para ello se utilizan 3 métodos distintos:

- Método de eliminación de parámetros.
- Método de componentes principales.
- Método de análisis factorial.

### 3.4.3 Multicolinealidad – Método eliminación parámetros

Tabla 5. Matriz de correlación

MATRIZ DE CORRELACIÓN																
	Gto total viaj	Gto med viaj	Gto med dia	Dur med	IPH	Viaj camp	Rent med	Triesgpob	Habit	IPT	IPC	Tasaact	Tasaparo	PIB	Viaj hotel	Viaj turismru
Gto total viaj	1	0,565756	0,553661	-0,0824113	<b>0,734591</b>	0,577641	-0,0310974	0,239491	0,640844	0,117448	0,328257	0,587941	0,207291	0,564549	<b>0,89578</b>	0,0679187
Gto med viaj	0,565756	1	0,552877	0,436735	0,543203	0,156815	0,029158	0,147047	0,494865	0,0669086	0,0537517	0,371619	0,0461654	0,486237	0,530736	-0,316465
Gto med dia	0,553661	0,552877	1	-0,487968	0,24315	0,282641	0,554493	-0,251462	0,50521	0,255273	0,22579	0,586236	-0,261373	0,676184	0,560492	-0,0249865
Dur med	-0,0824113	0,436735	-0,487968	1	0,254478	-0,117339	-0,491108	0,334205	-0,045776	-0,201889	-0,240311	-0,261316	0,238976	-0,195872	-0,111637	-0,323482
IPH	<b>0,734591</b>	0,543203	0,24315	0,254478	1	0,239109	-0,11332	0,232208	0,358595	0,263828	0,175851	0,142132	0,281398	0,230534	0,572964	-0,0888352
Viaj camp	0,577641	0,156815	0,282641	-0,117339	0,239109	1	0,112702	-0,0833065	<b>0,707235</b>	-0,0231934	0,401278	0,232934	-0,0355323	<b>0,713949</b>	<b>0,715393</b>	0,327418
Rent med	-0,0310974	0,029158	0,554493	-0,491108	-0,11332	0,112702	1	-0,91852	-0,0794806	0,339647	0,16552	0,0671381	<b>-0,849331</b>	0,177916	-0,0242607	0,0757427
Triesgpob	0,239491	0,147047	-0,251462	0,334205	0,232208	-0,0833065	-0,91852	1	0,263488	-0,300492	-0,176319	0,103984	0,934702	0,00843917	0,231004	-0,0992043
Habit	0,640844	0,494865	0,50521	-0,045776	0,358595	<b>0,707235</b>	-0,0794806	0,263488	1	-0,101541	0,298866	0,325899	0,256932	<b>0,936086</b>	<b>0,890637</b>	0,246405
IPT	0,117448	0,0669086	0,255273	-0,201889	0,263828	-0,0231934	0,339647	-0,300492	-0,101541	1	0,494909	0,0087539	-0,320262	-0,0809154	-0,025536	0,0163425
IPC	0,328257	0,0537517	0,22579	-0,240311	0,175851	0,401278	0,16552	-0,176319	0,298866	0,494909	1	-0,0113002	-0,176608	0,29751	0,413972	0,602161
Tasaact	0,587941	0,371619	0,586236	-0,261316	0,142132	0,232934	0,0671381	0,103984	0,325899	0,0087539	-0,0113002	1	-0,0809758	0,404292	0,453047	-0,271918
Tasaparo	0,207291	0,0461654	-0,261373	0,238976	0,281398	-0,0355323	-0,849331	<b>0,934702</b>	0,256932	-0,320262	-0,176608	-0,0890758	1	0,00523882	0,224472	-0,0587009
PIB	0,564549	0,486237	0,676184	-0,195872	0,230534	<b>0,713949</b>	0,177916	0,00843917	<b>0,936086</b>	-0,0809154	0,29751	0,404292	0,00523882	1	<b>0,81061</b>	0,242525
Viaj hotel	<b>0,89578</b>	0,530736	0,560492	-0,111637	0,572964	<b>0,715393</b>	0,0757427	0,231004	<b>0,890637</b>	-0,025536	0,413972	0,453047	0,224472	<b>0,81061</b>	1	0,264065
Viaj turismru	0,0679187	-0,316465	-0,0249865	-0,323482	-0,0888352	0,327418	0,0757427	-0,0992043	0,246405	0,0163425	0,602161	-0,271918	-0,0587009	0,242525	0,264065	1

Tabla 6. Matriz inversa de correlación

MATRIZ INVERSA DE CORRELACIÓN																
Gto total viaj	<b>281,156517</b>	-141,312208	80,3029751	117,6941	-57,2424915	-74,5358918	30,3628208	-42,9505947	208,818992	-31,1627626	43,6173631	-31,2110407	25,1694108	-23,5336903	-285,489497	-14,102667
Gto med viaj	-141,312208	<b>171,929915</b>	-152,473786	-149,510997	33,2715466	47,2321413	1,34693393	40,2873001	-133,638778	20,3464121	-28,8357542	16,3784477	-16,859222	54,9394293	122,382597	11,0026397
Gto med dia	80,3029751	-152,473786	<b>181,139707</b>	133,111583	-16,3190739	-21,567782	-37,0565908	-51,2195738	141,296394	-22,7787423	20,7605689	-11,318462	4,51322852	-107,629419	-84,5379051	-4,1154235
Dur med	117,6941	-149,510997	133,111583	<b>137,289254</b>	-31,9439823	-41,703956	1,87707264	-46,9993177	112,844857	-14,4878355	23,6363781	-7,32597761	29,5847034	-49,0086612	-99,1035504	5,56135281
IPH	-57,2424915	33,2715466	-16,3190739	-31,9439823	<b>19,9383861</b>	19,9428453	-5,39056369	18,2849314	-26,3330588	0,44356018	-5,71432164	5,62638045	-15,2863653	-6,03314851	40,0761706	2,57523562
Viaj camp	-74,5358918	47,2321413	-21,567782	-41,703956	19,9428453	<b>29,0825025</b>	-8,63890662	19,7377854	-39,3800954	3,70271316	-10,2275904	6,82143637	-15,3977666	-9,13104212	58,1232986	4,46099975
Rent med	30,3628208	1,34693393	-37,0565908	1,87707264	-5,39056369	-8,63890662	<b>55,7340566</b>	35,101077	-2,79256293	-1,6493627	9,91807012	3,06979652	8,81972656	31,532561	-34,2608234	-2,84937966
Triesgpob	-42,9505947	40,2873001	-51,2195738	-46,9993177	18,2849314	19,7377854	35,101077	<b>96,229636</b>	-71,2469536	0,13895003	-1,40505914	-12,8398958	-52,8194129	45,8612122	41,504983	-9,04202261
Habit	208,818992	-133,638778	141,296394	112,844857	-26,3330588	-39,3800954	-2,79256293	-71,2469536	<b>314,911643</b>	-47,7562379	45,3564663	-7,83599767	20,5489345	-163,085945	-292,115605	-1,10479061
IPT	-31,1627626	20,3464121	-22,7787423	-14,4878355	0,44356018	3,70271316	-1,6493627	0,13895003	-47,7562379	<b>11,6522015</b>	-9,82739024	3,93648886	5,2802334	25,0350576	48,3181998	2,30973169
IPC	43,6173631	-28,8357542	20,7605689	23,6363781	-5,71432164	-10,2275904	9,91807012	-1,40505914	45,3564663	-9,82739024	<b>13,1405276</b>	-3,43549883	2,28791394	-13,960574	-54,2852034	-3,98023986
Tasaact	-31,2110407	16,3784477	-11,318462	-7,32597761	5,62638045	6,82143637	3,06979652	-12,8398958	-7,83599767	3,93648886	-3,43549883	<b>16,0574649</b>	17,5109784	-7,68811105	21,4221496	6,94762365
Tasaparo	25,1694108	-16,859222	4,51322852	29,5847034	-15,2863653	-15,3977666	8,81972656	-52,8194129	20,5489345	5,2802334	2,28791394	17,5109784	<b>54,6508126</b>	-2,26750028	-20,1532213	7,72760033
PIB	-23,5336903	54,9394293	-107,629419	-49,0086612	-6,03314851	-9,13104212	31,532561	45,8612122	-163,085945	25,0350576	-13,960574	-7,68811105	-2,26750028	<b>147,564044</b>	84,8784103	-7,18539557
Viaj hotel	-285,489497	122,382597	-84,5379051	-99,1035504	40,0761706	58,1232986	-34,2608234	41,504983	-292,115605	48,3181998	-54,2852034	21,4221496	-20,1532213	84,8784103	<b>360,96916</b>	7,80653926
Viaj turismru	-14,102667	11,0026397	-4,1154235	-5,56135281	2,57523562	4,46099975	-2,84937966	-9,04202261	-1,10479061	2,30973169	-3,98023986	6,94762365	7,72760033	-7,18539557	7,80653926	6,2801594

El índice de acondicionamiento es de 79.71.

Se observa que existe problema de multicolinealidad en las 3 pruebas realizadas, ya que en la primera muchas variables sobrepasan el valor de 0,7. En cuando a la matriz inversa de correlación se observa la diagonal principal y hay variables que son mayores que 10. Y por último el Índice de Acondicionamiento es mayor que 10.

Entonces con la finalidad de solucionar la multicolinealidad se elimina la variable IPT por la matriz inversa de correlación y se vuelven a realizar las 3 pruebas.

Se elimina la variable viajeros hotel. [4]

Tabla 7. Matriz inversa de correlación

MATRIZ INVERSA DE CORRELACIÓN															
Gto total viaj	55,3636742	-44,5201557	13,442183	39,3133799	-25,5463606	-28,5663379	3,26603164	-10,1244199	-22,2144207	7,05196617	0,68334793	-14,2683226	9,23028457	43,5964064	-7,92849699
Gto med viaj	-44,5201557	130,43745	-123,812146	-115,911036	19,6841647	27,5260805	12,9626867	26,2154952	-34,6002172	3,96466142	-10,4309571	9,11550513	-10,0264955	26,1623442	8,35591896
Gto med dia	13,442183	-123,812146	161,341181	109,901826	-6,93335377	-7,95547813	-45,0803739	-41,4992301	72,8837874	-11,4627635	8,04713547	-6,30145815	-0,20659603	-87,7511478	-2,2871552
Dur med	39,3133799	-115,911036	109,901826	110,08052	-20,9411288	-25,7462901	-7,52918587	-35,6041859	32,6449488	-1,22214501	8,73245558	-1,44455804	24,0516662	-25,7054189	-3,41807897
IPH	-25,5463606	19,6841647	-6,93335377	-20,9411288	15,488977	13,4897754	-1,58679663	13,6768901	6,09872782	-4,92091034	0,31262833	3,24801184	-13,0488779	-15,4566728	1,70852391
Viaj camp	-28,5663379	27,5260805	-7,95547813	-25,7462901	13,4897754	19,7234818	-3,12222463	13,0546479	7,65639521	-4,07748933	-1,48658024	3,37203921	-12,1526925	-22,7981742	3,20398984
Rent med	3,26603164	12,9626867	-45,0803739	-7,52918587	-1,58679663	-3,12222463	52,4822441	39,0404576	-30,5182587	2,93668368	4,76567493	5,10304635	6,90691508	39,5886626	-2,1084342
Triesgprob	-10,1244199	26,2154952	-41,4992301	-35,6041859	13,6768901	13,0546479	39,0404576	91,4573069	-37,6588953	-5,41677685	4,83676619	-15,303059	-50,5021537	36,1017165	-9,93963469
Habit	-22,2144207	-34,6002172	72,8837874	32,6449488	6,09872782	7,65639521	-30,5182587	-37,6588953	78,516029	-8,65455907	1,42596819	9,49995474	4,23986696	-94,3977834	5,21267965
IPT	7,05196617	3,96466142	-11,4627635	-1,22214501	-4,92091034	-4,07748933	2,93668368	-5,41677685	-8,65455907	5,1844788	-2,56094313	1,06898709	7,9778804	13,6734997	1,26477283
IPC	0,68334793	-10,4309571	8,04713547	8,73245558	0,31262833	-1,48658024	4,76567493	4,83676619	1,42596819	-2,56094313	4,97671847	-0,21387805	-0,74287605	-1,19593294	-2,80623493
Tasaact	-14,2683226	9,11550513	-6,30145815	-1,44455804	3,24801184	3,37203921	5,10304635	-15,303059	9,49995474	1,06898709	-0,21387805	14,7861417	18,7069956	-12,7253225	6,48433518
Tasaparo	9,23028457	-10,0264955	-0,20659603	24,0516662	-13,0488779	-12,1526925	6,90691508	-50,5021537	4,23986696	7,9778804	-0,74287605	18,7069956	53,5256408	2,47133498	8,16344618
PIB	43,5964064	26,1623442	-87,7511478	-25,7054189	-15,4566728	-22,7981742	39,5886626	36,1017165	-94,3977834	13,6734997	-1,19593294	-12,7253225	2,47133498	127,605706	-9,02102784
Viaj turismru	-7,92849699	8,35591896	-2,2871552	-3,41807897	1,70852391	3,20398984	-2,1084342	-9,93963469	5,21267965	1,26477283	-2,80623493	6,48433518	8,16344618	-9,02102784	6,11133042

El índice de acondicionamiento es de 46.54.

En este caso es igual, sigue habiendo problemas de multicolinealidad y se elimina la variable gasto medio diario.

Tabla 8. Matriz inversa de correlación

MATRIZ INVERSA DE CORRELACIÓN														
Gto total viaj	54,2437352	-34,2047141	30,1568803	-24,9687064	-27,9035251	7,02191487	-6,66690065	-28,2867526	8,00698932	0,01289872	-13,743315	9,24749717	50,9074164	-7,73794206
Gto med viaj	-34,2047141	35,4248358	-31,5731083	14,3635553	21,4210997	-21,6316914	-5,63073705	21,3303149	-4,83178677	-4,25563899	4,27980819	-10,1850359	-41,1773011	6,60077133
Dur med	30,1568803	-31,5731083	35,2179759	-16,2182909	-20,3272051	23,1785067	-7,33588435	-17,0017765	6,58602042	3,25093577	2,84784735	24,1923946	34,0686033	-1,86012254
IPH	-24,9687064	14,3635553	-16,2182909	15,1910283	13,1479027	-3,52404649	11,8935336	9,23078052	-5,41350246	0,65843984	2,97721775	-13,057756	-19,2276242	1,61023743
Viaj camp	-27,9035251	21,4210997	-20,3272051	13,1479027	19,3312097	-5,34506647	11,0083866	11,2501793	-4,64270004	-1,08978874	3,06132429	-12,1628794	-27,1250443	3,09121384
Rent med	7,02191487	-21,6316914	23,1785067	-3,52404649	-5,34506647	39,8863272	27,4451488	-10,1537841	-0,26612951	7,01412677	3,34235458	6,84919003	15,0700957	-2,74748872
Triesgprob	-6,66690065	-5,63073705	-7,33588435	11,8935336	11,0083866	27,4451488	80,7831189	-18,912156	-8,36516148	6,90660306	-16,923883	-50,5552932	13,5308821	-10,5279233
Habit	-28,2867526	21,3303149	-17,0017765	9,23078052	11,2501793	-10,1537841	-18,912156	45,5917228	-3,47640425	-2,20922095	12,3465568	4,33319404	-54,7573398	6,24587237
IPT	8,00698932	-4,83178677	6,58602042	-5,41350246	-4,64270004	-0,26612951	-8,36516148	-3,47640425	4,37008693	-1,98922047	0,62128909	7,96320243	7,43906754	1,10227793
IPC	0,01289872	-4,25563899	3,25093577	0,65843984	-1,08978874	7,01412677	6,90660306	-2,20922095	-1,98922047	4,57535542	0,10041671	-0,73257176	3,18078829	-2,69215961
Tasaact	-13,743315	4,27980819	2,84784735	2,97721775	3,06132429	3,34235458	-16,923883	12,3465568	0,62128909	0,10041671	14,5400273	18,6989266	-16,152595	6,39500638
Tasaparo	9,24749717	-10,1850359	24,1923946	-13,057756	-12,1628794	6,84919003	-50,5552932	4,33319404	7,96320243	-0,73257176	18,6989266	53,5253763	2,35897037	8,1605175
PIB	50,9074164	-41,1773011	34,0686033	-19,2276242	-27,1250443	15,0700957	13,5308821	-54,7573398	7,43906754	3,18078829	-16,152595	2,35897037	79,8791192	-10,2649787
Viaj turismru	-7,73794206	6,60077133	-1,86012254	1,61023743	3,09121384	-2,74748872	-10,5279233	6,24587237	1,10227793	-2,69215961	6,39500638	8,1605175	-10,2649787	6,07890795

El índice de acondicionamiento es de 31.01.

Existe problema de multicolinealidad, pero esta vez no se puede eliminar la mayor variable de la diagonal principal ya que la diferencia entre las 2 mayores es menor a 1. Entonces hay que utilizar otro criterio, se trata de generar dos modelos, uno eliminando la variable tasa riesgo pobreza, y otro eliminando la variable PIB. Una vez generados los dos modelos, se elige el que tenga un R<sup>2</sup> mayor.

Modelo A → Se elimina la variable tasa riesgo pobreza → R<sup>2</sup>= 99,68%

Modelo B → Se elimina la variable PIB → R<sup>2</sup>= 99,67%

Entonces se elige el Modelo A y se elimina la variable Tasa riesgo pobreza.

Tabla 9. Matriz inversa de correlación

MATRIZ INVERSA DE CORRELACIÓN													
Gto total viaj	53,6935266	-34,6694098	29,5514616	-23,9871522	-26,9950207	9,28691878	-29,8475424	7,31662603	0,58288954	-15,1400158	5,07525027	52,0240983	-8,60679461
Gto med viaj	-34,6694098	35,0323627	-32,0844334	15,1925572	22,1884052	-19,7187123	20,0121017	-5,41485444	-3,77423562	3,10018135	-13,708836	-40,2341728	5,86695505
Dur med	29,5514616	-32,0844334	34,551807	-15,1382437	-19,3275377	25,6707903	-18,7191822	5,82638328	3,87812177	1,31099596	19,6014876	35,2973376	-2,81615925
IPH	-23,9871522	15,1925572	-15,1382437	13,4399676	11,5271604	-7,56473971	12,0151786	-4,18191682	-0,35840522	5,46888648	-5,61460344	-21,2197483	3,16024208
Viaj camp	-26,9950207	22,1884052	-19,3275377	11,5271604	17,8310872	-9,08504103	13,8273555	-3,50277213	-2,03095761	5,3675567	-5,27366515	-28,9689095	4,5258632
Rent med	9,28691878	-19,7187123	25,6707903	-7,56473971	-9,08504103	30,562149	-3,72859355	2,57583927	4,66768668	9,09205197	24,0247777	10,4731319	0,82925387
Habit	-29,8475424	20,0121017	-18,7191822	12,0151786	13,8273555	-3,72859355	41,1641933	-5,43477428	-0,59231439	8,38450234	-7,50231817	-51,5896216	3,78117763
IPT	7,31662603	-5,41485444	5,82638328	-4,18191682	-3,50277213	2,57583927	-5,43477428	3,50386726	-1,27403578	-1,13119356	2,72815832	8,84020203	0,01210241
IPC	0,58288954	-3,77423562	3,87812177	-0,35840522	-2,03095761	4,66768668	-0,59231439	-1,27403578	3,98487109	1,54733463	3,58968452	2,0239571	-1,79206825
Tasaact	-15,1400158	3,10018135	1,31099596	5,46888648	5,3675567	9,09205197	8,38450234	-1,13119356	1,54733463	10,9945116	8,10770559	-13,3179054	4,18942996
Tasaparo	5,07525027	-13,708836	19,6014876	-5,61460344	-5,27366515	24,0247777	-7,50231817	2,72815832	3,58968452	8,10770559	21,887112	10,8268003	1,57198445
PIB	52,0240983	-40,2341728	35,2973376	-21,2197483	-28,9689095	10,4731319	-51,5896216	8,84020203	2,0239571	-13,3179054	10,8268003	77,6127451	-8,50158933
Viaj turismru	-8,60679461	5,86695505	-2,81615925	3,16024208	4,5258632	0,82925387	3,78117763	0,01210241	-1,79206825	4,18942996	1,57198445	-8,50158933	4,70687416

El índice de acondicionamiento es 30.83.

Existe problema de multicolinealidad y se elimina la variable PIB ya que es la mayor de la diagonal principal.

Tabla 10. Matriz inversa de correlación

MATRIZ INVERSA DE CORRELACIÓN													
Gto total viaj	21,8001496	-7,96218626	8,44476586	-12,7148323	-10,6165803	-2,5823427	-15,2902087	6,61041101	3,26602932	-2,01423573	-3,44917446	7,74411195	-1,19601285
Gto med viaj	-7,96218626	14,1981355	-14,0109074	4,45180766	7,43827005	-13,8631297	1,34436749	-6,89682989	-0,9969833	-2,61595795	-4,0467767	-8,96899805	1,30922429
Dur med	8,44476586	-14,0109074	20,6876488	-8,01767096	-8,75831973	16,7510809	-13,1068325	6,35233786	3,41324331	1,89432323	9,73696126	23,186289	2,51791129
IPH	-12,7148323	4,45180766	-8,01767096	10,5627658	6,61865998	0,10346141	15,150539	-3,94980485	-3,62285183	1,42408426	-0,9108575	-12,4899306	-0,86063548
Viaj camp	-10,6165803	7,43827005	-8,75831973	6,61865998	10,1201914	-0,2276213	15,6031516	-7,34410766	-2,11657007	-0,00966887	-2,42371185	-11,3618293	-0,39452817
Rent med	-2,5823427	-13,8631297	16,7510809	0,10346141	-0,2276213	37,0431839	24,8923955	0,17680485	-1,6695934	6,41403522	6,38972359	6,4041437	-0,81088486
Triesgpob	-15,2902087	1,34436749	-13,1068325	15,150539	15,6031516	24,8923955	78,491096	-9,6367018	-9,62527986	6,36780304	-14,187763	-50,9548838	-8,78911825
Habit	6,61041101	-6,89682989	6,35233786	-3,94980485	-7,34410766	0,17680485	-9,6367018	8,05542686	1,6230955	-0,02878247	1,27391175	5,95027423	-0,79079666
IPT	3,26602932	-0,9969833	3,41324331	-3,62285183	-2,11657007	-1,6695934	-9,62527986	1,6230955	3,67729354	-2,2854443	2,12556513	7,74351373	2,05824577
IPC	-2,01423573	-2,61595795	1,89432323	1,42408426	-0,00966887	6,41403522	6,36780304	-0,02878247	-2,2854443	4,44869636	0,7436134	-0,82650601	-2,28340793
Tasaact	-3,44917446	-4,0467767	9,73696126	-0,9108575	-2,42371185	6,38972359	-14,187763	1,27391175	2,12556513	0,7436134	11,2737629	19,175941	4,31929443
Tasaparo	7,74411195	-8,96899805	23,186289	-12,4899306	-11,3618293	6,4041437	-50,9548838	5,95027423	7,74351373	-0,82650601	19,175941	53,4557117	8,46366031
Viaj turismru	-1,19601285	1,30922429	2,51791129	-0,86063548	-0,39452817	-0,81088486	-8,78911825	-0,79079666	2,05824577	-2,28340793	4,31929443	8,46366031	4,75979242

El índice de acondicionamiento es 23.21.

Existe problema de multicolinealidad y se elimina la variable Tasa riesgo pobreza.

Tabla 11. Matriz inversa de correlación

MATRIZ INVERSA DE CORRELACIÓN												
Gto total viaj	18,821589	-7,70030102	5,89153097	-9,76347964	-7,57705791	2,26674135	4,73316392	1,39100725	-0,77377608	-6,21297657	-2,18199238	-2,90814912
Gto med viaj	-7,70030102	14,1751096	-13,7864183	4,19231463	7,17102484	-14,2894778	-6,7317759	-0,83212494	-2,72502341	-3,80377376	-8,09626098	1,45976117
Dur med	5,89153097	-13,7864183	18,4990049	-5,48775882	-6,15282809	20,9077366	4,7431536	1,80596633	2,95765057	7,36781823	14,6775898	1,05026073
IPH	-9,76347964	4,19231463	-5,48775882	7,63837257	3,60690236	-4,70132826	-2,08970564	-1,76495731	0,1949557	1,82769842	-2,65449711	0,83586113
Viaj camp	-7,57705791	7,17102484	-6,15282809	3,60690236	7,0184595	-5,17595098	-5,42843917	-0,20317214	-1,27551686	0,39665667	-1,23256857	1,35265019
Rent med	2,26674135	-14,2894778	20,9077366	-4,70132826	-5,17595098	29,1488954	3,23295522	1,38293465	4,39457213	10,8891818	22,5637998	1,97646579
Habit	4,73316392	-6,7317759	4,7431536	-2,08970564	-5,42843917	3,23295522	6,87228602	0,44135698	0,7530211	-0,46798316	-0,30568403	-1,86987589
IPT	1,39100725	-0,83212494	1,80596633	-1,76495731	-0,20317214	1,38293465	0,44135698	2,49695568	-1,50456737	0,38573481	1,49497037	0,98044552
IPC	-0,77377608	-2,72502341	2,95765057	0,1949557	-1,27551686	4,39457213	0,7530211	-1,50456737	3,93209107	1,89463415	3,30734713	-1,57036688
Tasaact	-6,21297657	-3,80377376	7,36781823	1,82769842	0,39665667	10,8891818	-0,46798316	0,38573481	1,89463415	8,70923486	9,96552291	2,73060561
Tasaparo	-2,18199238	-8,09626098	14,6775898	-2,65449711	-1,23256857	22,5637998	-0,30568403	1,49497037	3,30734713	9,96552291	20,3767981	2,7579367
Viaj turismru	-2,90814912	1,45976117	1,05026073	0,83586113	1,35265019	1,97646579	-1,86987589	0,98044552	-1,57036688	2,73060561	2,7579367	3,77562219

El índice de acondicionamiento es 16.59.

Existe problema de multicolinealidad y se elimina la variable Renta media.

Tabla 12. Matriz inversa de correlación

MATRIZ INVERSA DE CORRELACIÓN											
Gto total viaj	18,6453176	-6,58909082	4,26565698	-9,39788448	-7,17455407	4,48175565	1,2834644	-1,11551659	-7,05976538	-3,93664885	-3,06184746
Gto med viaj	-6,58909082	7,17006975	-3,53695145	1,88761236	4,63365128	-5,14690481	-0,15417767	-0,57070024	1,53436063	2,96504721	2,42867143
Dur med	4,26565698	-3,53695145	3,50243489	-2,11561962	-2,44025449	2,42423979	0,81402365	-0,19446052	-0,44270568	-1,50683079	-0,36740623
IPH	-9,39788448	1,88761236	-2,11561962	6,88011099	2,77209047	-1,56827306	-1,54190838	0,9037416	3,58397829	0,98474301	1,15463872
Viaj camp	-7,17455407	4,63365128	-2,44025449	2,77209047	6,09936915	-4,85436535	0,04239469	-0,49517546	2,33024182	2,77407113	1,70360998
Habit	4,48175565	-5,14690481	2,42423979	-1,56827306	-4,85436535	6,51371327	0,28797327	0,26561138	-1,67572145	-2,80827472	-2,08908918
IPT	1,2834644	-0,15417767	0,81402365	-1,54190838	0,04239469	0,28797327	2,43134399	-1,71306261	-0,13088946	0,42445774	0,88667446
IPC	-1,11551659	-0,57070024	-0,19446052	0,9037416	-0,49517546	0,26561138	-1,71306261	3,26955261	0,25294947	-0,09443686	-1,8683446
Tasaact	-7,05976538	1,53436063	-0,44270568	3,58397829	2,33024182	-1,67572145	-0,13088946	0,25294947	4,64135241	1,53634182	1,99225532
Tasaparo	-3,93664885	2,96504721	-1,50683079	0,98474301	2,77407113	-2,80827472	0,42445774	-0,09443686	1,53634182	2,91043946	1,22797895
Viaj turismru	-3,06184746	2,42867143	-0,36740623	1,15463872	1,70360998	-2,08908918	0,88667446	-1,8683446	1,99225532	1,22797895	3,64160624

El índice de acondicionamiento es 11.43

Existe problema de multicolinealidad y se elimina la variable Gasto total viaje.

Tabla 13. Matriz inversa de correlación

MATRIZ INVERSA DE CORRELACIÓN										
Gto med viaj	4,84154317	-2,02950589	-1,43351714	2,09822715	-3,56309188	0,29938733	-0,96491397	-0,9604982	1,57387022	1,34664155
Dur med	-2,02950589	2,52654218	0,03441892	-0,79886724	1,39890825	0,52039396	0,06074624	1,17242033	-0,60620823	0,33308014
IPH	-1,43351714	0,03441892	2,14325242	-0,84413274	0,69068668	-0,89499797	0,34148269	0,02561254	-0,99946406	-0,38863822
Viaj camp	2,09822715	-0,79886724	-0,84413274	3,3386639	-3,12982523	0,53626049	-0,92441642	-0,38629375	1,25928331	0,52543803
Habit	-3,56309188	1,39890825	0,69068668	-3,12982523	5,43643832	-0,02053174	0,53374695	0,02122703	-1,86202652	-1,35311609
IPT	0,29938733	0,52039396	-0,89499797	0,53626049	-0,02053174	2,34299575	-1,63627519	0,35507477	0,69543991	1,09743902
IPC	-0,96491397	0,06074624	0,34148269	-0,92441642	0,53374695	-1,63627519	3,20281321	-0,16942388	-0,32995964	-2,05152955
Tasaact	-0,9604982	1,17242033	0,02561254	-0,38629375	0,02122703	0,35507477	-0,16942388	1,96827983	0,04578972	0,83293344
Tasaparo	1,57387022	-0,60620823	-0,99946406	1,25928331	-1,86202652	0,69543991	-0,32995964	0,04578972	2,0792815	0,58152077
Viaj turismru	1,34664155	0,33308014	-0,38863822	0,52543803	-1,35311609	1,09743902	-2,05152955	0,83293344	0,58152077	3,13880387

El índice de acondicionamiento es 5.96

Ya casi está solucionado el problema de multicolinealidad pero aun no, como se puede observar en la matriz de correlaciones entre viajeros camping y habitantes. Entonces se generan dos modelos y se elige el que mayor  $R^2$  obtenga.

Modelo A → Se elimina la variable Viajeros camping →  $R^2= 95,09\%$

Modelo B → Se elimina la variable Habitantes →  $R^2= 92,90\%$

El modelo A es el que tiene mayor  $R^2$  y se elimina la variable Viajeros camping.

Tabla 14. Matriz inversa de correlación

MATRIZ INVERSA DE CORRELACIÓN										
Gto med viaj	3,52288478	-1,52744729	-0,90301084	-1,59611214	-0,03763262	-0,38395234	-0,71772682	0,7824571	1,01642312	
Dur med	-1,52744729	2,3353912	-0,16756307	0,65001141	0,64870904	-0,16044584	1,07998892	-0,30489003	0,45880566	
IPH	-0,90301084	-0,16756307	1,9298257	-0,10064423	-0,75941228	0,1077574	-0,07205623	-0,68107254	-0,25578884	
Habit	-1,59611214	0,65001141	-0,10064423	2,50238856	0,48218482	-0,33284578	-0,34090344	-0,68151335	-0,86054503	
IPT	-0,03763262	0,64870904	-0,75941228	0,48218482	2,2568609	-1,48779424	0,41712177	0,4931722	1,01304249	
IPC	-0,38395234	-0,16044584	0,1077574	-0,33284578	-1,48779424	2,94685881	-0,27638172	0,01871343	-1,90604514	
Tasaact	-0,71772682	1,07998892	-0,07205623	-0,34090344	0,41712177	-0,27638172	1,92358445	0,1914927	0,89372825	
Tasaparo	0,7824571	-0,30489003	-0,68107254	-0,68151335	0,4931722	0,01871343	0,1914927	1,60430273	0,38333509	
Viaj turismru	1,01642312	0,45880566	-0,25578884	-0,86054503	1,01304249	-1,90604514	0,89372825	0,38333509	3,05611057	

El índice de acondicionamiento es 3.98

Ahora ya se ha resuelto el problema de multicolinealidad por el método de eliminación de variables.



### 3.4.4 Multicolinealidad – Método componentes principales

Con el método de componentes principales se trata de resolver la multicolinealidad creando 5 componentes que contengan toda la información de las 16 variables, de esta manera no se elimina ninguna variable y se crea un nuevo modelo que no tenga problemas de multicolinealidad (que no tengan correlación entre ellos) y que explique toda la información recopilada anteriormente.

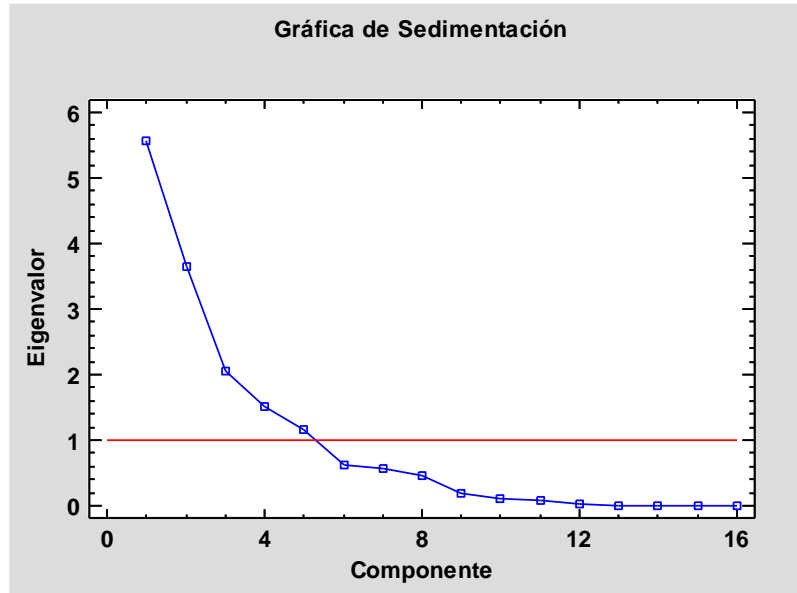


Gráfico 7. Sedimentación

En el gráfico se observa que serían 5 las componentes principales ya que son las que están por encima de la línea roja.

Tabla 15. Porcentajes componentes

Número	Eigenvalor	% Varianza	% Acumulado
1	5,5575	34,734	34,734
2	3,63708	22,732	57,466
3	2,04892	12,806	70,272
4	1,52473	9,53	79,802
5	1,15278	7,205	87,006
6	0,632793	3,955	90,961
7	0,572554	3,578	94,54
8	0,459591	2,872	97,412
9	0,188825	1,18	98,592
10	0,119668	0,748	99,34
11	0,0725162	0,453	99,794
12	0,01664	0,104	99,898
13	0,00777008	0,049	99,946
14	0,00452456	0,028	99,974
15	0,00322628	0,02	99,995
16	0,00087476	0,005	100

Como se observa en la tabla 15, hay 5 componentes con un valor propio mayor que 1, que indican el número de componentes a elegir. Estas explican el 87% de la variabilidad.

Tabla 16. Composición de los componentes

	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5
Gto total viaje	0,375284	0,0854554	-0,0766276	0,12433	-0,177374
Gto med viaj	0,261443	0,112357	-0,398998	0,154578	0,274788
Gto med dia	0,305098	-0,220067	-0,243605	-0,137761	-0,178279
Dur med	-0,0694536	0,318182	-0,177154	0,277401	0,561539
IPH	0,238131	0,158582	-0,162376	0,460442	-0,0303455
Viaj camp	0,303261	-0,0498058	0,247935	-0,081254	0,260406
Rent med	0,0498704	-0,483665	-0,159767	-0,0283021	0,135227
Triesgpob	0,0464277	0,482331	0,086574	-0,0327468	-0,275975
Habit	0,372887	0,0940666	0,139084	-0,146256	0,175355
IPT	0,0476007	-0,234233	-0,0808089	0,562507	-0,307977
IPC	0,184314	-0,181773	0,347218	0,400938	-0,0889465
Tasaact	0,231706	-0,0190234	-0,30169	-0,284641	-0,375734
Tasaparo	0,0366067	0,463249	0,155456	-0,0185936	-0,253352
PIB	0,369425	-0,0428289	0,0780393	-0,242487	0,206664
Viaj hotel	0,40766	0,0744475	0,0931777	-0,00923081	0,0136473
Viaj turismrural	0,0813254	-0,130793	0,589027	0,0736737	0,031139

Las variables que tienen más peso en cada componente son las siguientes:

- Componente 1: está formado por Gasto total viaje (0.37), Gasto medio viaje (0.26), Gasto medio diario (0.30), Viajeros camping (0.30), Habitantes (0.37), PIB (0.37) y Viajeros hotel (0.41).
- Componente 2: está formado por la Duración media (0.32), Renta media (-0.48), Tasa riesgo pobreza (0.48), Tasa de paro (0.46).
- Componente 3: está formado por Gasto medio viaje (-0.39), IPC (0.34), Tasa de actividad (-0.30), Viajeros turismo rural (0.59).
- Componente 4: está formado por la Duración media (0.28), IPH (0.46), IPT (0.56), IPC (0.40).
- Componente 5: está formado por el Gasto medio viaje (0.27), Duración media (0.56), Tasa riesgo pobreza (-0.27), IPT (-0.30), Tasa de actividad (-0.37).

### 3.4.5 Multicolinealidad – Método análisis factorial

El método de análisis factorial consiste en crear un nuevo modelo explicando todas las variables creando unas nuevas llamadas factores. De todas las variables se modelan como combinaciones lineales de factores.

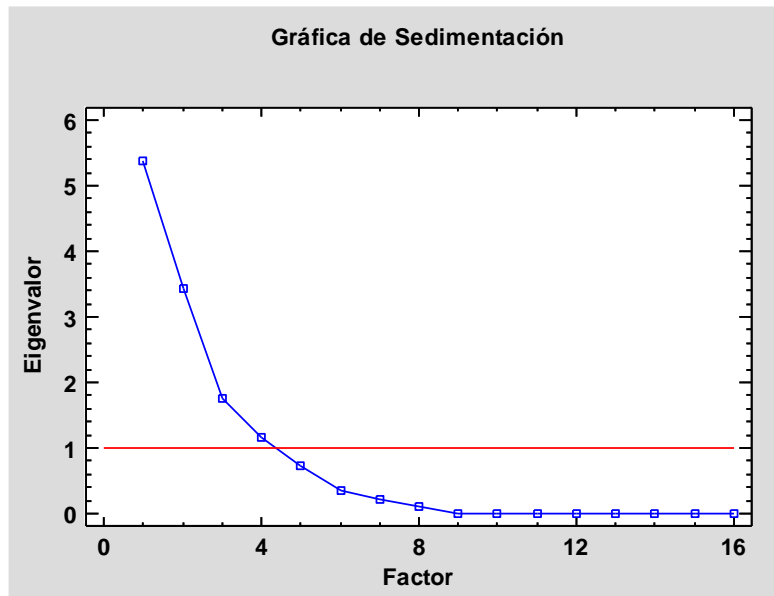


Gráfico 8. Sedimentación

En este caso, se han creado 4 factores que explican las 16 variables.

Se eligen 4 factores ya que son los 4 que están por encima de la línea roja como se observa en el gráfico.

En el siguiente apartado se pueden observar los resultados y se elige el mejor método con tal de realizar las siguientes pruebas.

Tabla 17. Porcentajes factores

Número	Eigenvalor	% Varianza	% Acumulado
1	5,36568	40,797	40,797
2	3,4295	26,076	66,872
3	1,74524	13,27	80,142
4	1,15845	8,808	88,95
5	0,731267	5,56	94,51
6	0,366279	2,785	97,295
7	0,231498	1,76	99,055
8	0,124275	0,945	100

Como se observa en la tabla 17, hay 4 factores con un valor propio mayor que 1, que indican el número de factores a elegir. Estos explican casi el 89% de la variabilidad.

En cuanto al peso que tiene cada variable se refleja en la siguiente tabla:

Tabla 18. Composición factores

COEFICIENTES DEL FACTOR				
	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
Gto total viaje	0,874044	0,150245	-0,125488	0,206002
Gto med viaj	0,593447	0,175222	-0,534672	0,136081
Gto med dia	0,700419	-0,419177	-0,330969	-0,144217
Dur med	-0,143937	0,510192	-0,214746	0,195156
IPH	0,543407	0,274586	-0,237147	0,533097
Viaj camp	0,671936	-0,0930355	0,299225	-0,0976336
Rent med	0,111043	-0,946999	-0,209724	-0,0421181
Triesgpob	0,116002	0,919568	0,0894633	-0,0244096
Habit	0,879957	0,168349	0,18394	-0,226804
IPT	0,0981737	-0,390632	-0,074937	0,545971
IPC	0,41892	-0,322109	0,484212	0,465562
Tasaact	0,495188	-0,0378901	-0,343217	-0,20764
Tasaparo	0,0926239	0,866091	0,174344	-0,0104289
PIB	0,872225	-0,0949427	0,106655	-0,354883
Viaj hotel	0,977958	0,135081	0,124449	0,00019719
Viaj turismrural	0,187736	-0,224172	0,800768	0,0841108

### 3.5 RESULTADOS DE MULTICOLINEALIDAD

#### 3.5.1 Método eliminación variables

Los resultados que se obtienen mediante este método son los siguientes:

Viajeros =  $-6,31137E8 + 1617,52 * \text{Gto med viaj} - 323577 * \text{Dur med} + 1,17853E6 * \text{IPH} + 1,31995 * \text{Habit} - 472197 * \text{IPT} + 6,46525E6 * \text{IPC} + 312797 * \text{Tasaact} - 5782,25 * \text{Tasaparo} + 2,19622 * \text{Viaj turismrural}$

Y este modelo tiene un  $R^2 = 95,09\%$

*Tabla 19. Método eliminación variables*

		Error	Estadístico		
Parámetro	Estimación	Estándar	T	Valor-P	
CONSTANTE	-631137000	426133000	-1,48108	0,1821	
Gto med viaj	1617,52	4088,81	0,395596	0,7042	
Dur med	-323577	319892	-1,01152	0,3454	
IPH	1178530	374138	3,14998	0,0162	
Habit	1,31995	0,308611	4,27708	0,0037	
IPT	-472197	290631	-1,62473	0,1482	
IPC	6465250	4363430	1,48169	0,182	
Tasaact	312797	202967	1,54112	0,1672	
Tasaparo	-5782,25	132135	-0,0437602	0,9663	
Viaj turismrural	2,19622	5,51672	0,398102	0,7024	
<b>Análisis de Varianza</b>					
Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	5,27584E+14	9	5,86204E+13	15,07	0,0009
Residuo	2,72376E+13	7	3,89109E+12		
Total (Corr.)	5,54822E+14	16			

Si se observa la significatividad de las variables hay muchas no significativas y esto puede ser un problema más adelante, pero el modelo sí que es significativo.

### 3.5.2 Método componentes principales

Los resultados de la resolución de la multicolinealidad mediante el método de componentes principales son:

El modelo sería el siguiente:

$$\text{Viajeros} = 5,86507\text{E}6 + 2,40057\text{E}6*\text{PCOMP}_1 + 438395*\text{PCOMP}_2 + 548691*\text{PCOMP}_3 - 54358,7*\text{PCOMP}_4 + 80363,1*\text{PCOMP}_5$$

Con un  $R^2 = 96,19\%$

*Tabla 20. Método componentes principales*

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P	
CONSTANTE	5,87E+06	336330	17,4385	0	
PCOMP_1	2,40E+06	147058	16,3239	0	
PCOMP_2	438395	181783	2,41164	0,0345	
PCOMP_3	548691	242196	2,26548	0,0447	
PCOMP_4	-54358,7	280759	-0,193614	0,85	
PCOMP_5	80363,1	322892	0,248886	0,808	
Análisis de Varianza					
Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	5,34E+14	5	1,07E+14	55,5	0
Residuo	2,12E+13	11	1,92E+12		
Total (Corr.)	5,55E+14	16			

También se observa que los componentes son significativos y que el modelo lo es también ya que es importante para la elección del mejor modelo de los tres.

### 3.5.3 Método análisis factorial

En cuanto a la resolución de la multicolinealidad según el método análisis factorial se obtiene el siguiente resultado:

El modelo sería el siguiente:

$$\text{Viajeros} = 5,86507E6 + 1,04058E6 * \text{FSCORE}_1 + 230272 * \text{FSCORE}_2 + 388788 * \text{FSCORE}_3 - 20699,9 * \text{FSCORE}_4$$

Con un  $R^2 = 96.15\%$

Tabla 21. Método análisis factorial

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P	
CONSTANTE	5,87E+06	323333	18,1394	0	
FSCORE_1	1,04E+06	61061,3	17,0416	0	
FSCORE_2	230272	94512,6	2,43642	0,0314	
FSCORE_3	388788	176518	2,20255	0,0479	
FSCORE_4	-20699,9	253270	-0,0817306	0,9362	
Análisis de Varianza					
Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	5,33E+14	4	1,33E+14	75,05	0
Residuo	2,13E+13	12	1,78E+12		
Total (Corr.)	5,55E+14	16			

Si se observa la imagen hay tres factores significativos y uno que no es significativo. El modelo también es significativo, ya que el p-valor es inferior a 0.05.

Entonces este modelo también sería válido ya que tiene un  $R^2$  del 96.15% y los factores y el modelo son significativos.



# Capítulo 4

## Identificación de los principales factores

Se estudian los diferentes factores de cada método con la finalidad de obtener el mejor modelo.

Una vez elegido el mejor método, se aplican sobre él (modelo análisis factorial) las diferentes técnicas como son: autocorrelación, heterocedasticidad, normalidad, etc.

Y, por último, se realizan las predicciones con la finalidad de validar el modelo.

## 4.1 IDENTIFICACIÓN DE LOS FACTORES

Ahora se debe plantear lo siguiente:

¿Qué modelo interesa elegir?

Para ello, se ha analizado mediante los tres métodos, para que se escoja el modelo más óptimo.

### 4.1.1 Método eliminación de variables

Con respecto a este método las variables significativas hacen una buena explicación del modelo, pero como se observa son dos las variables, el IPH (Índice Precios Hoteleros) y el número de habitantes por comunidad Autónoma que explican casi el 95% del modelo.

También cabe destacar el error que en este caso sería del  $1.97258 \cdot 10^6$  que si se compara con los otros métodos este error sería el más grande. Entonces, esto ya sería un punto en contra a la hora de elegir este método.

### 4.1.2 Método componentes principales

En este método se utiliza la información de todas las variables para juntarla y crear 5 nuevos componentes que expliquen la misma información.

Los componentes están creados principalmente de las siguientes variables:

- Componente 1: Gasto total viaje (0.37), Gasto medio diario (0.31), Habitantes (0.37), PIB (0.37).
- Componente 2: Duración media (0.32), Renta media (-0.48), Tasa riesgo pobreza (0.48), Tasa paro (0.46).
- Componente 3: Gasto medio viaje (-0.39), IPC (0.35), Viajeros turismo rural (0.58).
- Componente 4: IPH (0.46), IPT (0.56), IPC (0.40).
- Componente 5: Duración media (0.56), IPT (-0.30), Tasa actividad (-0.37).

Se pueden observar que todos los componentes están creados con la información de las diferentes variables, por lo tanto, mediante este método no se pierde información de ninguna de ellas.

La resolución de la multicolinealidad mediante este método es una de las mejores elecciones.

Con este método se obtiene un modelo muy bueno ya que explica el 96.18%, y con un error menor que en el método anterior (método de eliminación de variables), exactamente el error es de 914.411.

Con respecto a la significatividad, este modelo tiene 3 componentes significativos y 2 que no lo son.

Por lo tanto, se puede afirmar que este método sería mejor elección que el anterior.

#### 4.1.3 Método análisis factorial

Con respecto a este método también se utiliza la información de todas las variables y esta vez, se crean factores, en concreto 4 factores que explican la misma información que las 16 variables.

Los factores están formados por las siguientes variables:

- Factor 1: Gasto total viaje (0.87), Gasto medio viaje (0.59), Gasto medio diario (0.70), IPH (0.54), Habitantes (0.87), Viajeros hotel (0.97).

Se puede decir que el factor 1 está compuesto principalmente por todos los gastos.

- Factor 2: Renta media (-0.94), Tasa riesgo pobreza (0.91).

En este caso, el factor 2 estaría creado principalmente por la renta media que actúa en negativo frente a la tasa riesgo de pobreza que tiene un peso de 0.91 en positivo.

- Factor 3: Viajeros turismo rural (0.80), IPC (0.48), Gasto medio viaje (-0.53).

Este factor estaría compuesto por una mezcla entre los viajeros de turismo rural y el IPC.

- Factor 4: IPH (0.53), IPT (0.55), IPC (0.46).

En este factor claramente se observa que está formado por los diferentes índices.

En este método, los factores son significativos menos el último, por lo tanto, esto es muy bueno para el modelo ya que tiene que ser significativos.

El error 908.525 es menor que en el método de componentes principales, por lo que esto también es mejor.

Y, por último, estos factores explican el 96.16% que es casi igual que en el método de componentes principales.

## 4.2 ELECCIÓN MODELO OPTIMO

Después de realizar un buen análisis de cada uno de los métodos de resolución de la multicolinealidad, se tiene que elegir uno de ellos.

En este caso, el mejor método ha sido el de análisis factorial, el cual tiene los factores significativos, el error es el más pequeño que se comete de los 3, y el conjunto de los 4 factores explican el 96.16% del modelo, por lo tanto, es muy buen modelo.

También una de las causas de elegir el método de análisis factorial ha sido que hay 16 variables, y son muchas frente a 17 Comunidades Autónomas, por lo tanto, solo habría un error de 1. Entonces con este método se pasa de 16 variables a 3 factores que son los significativos y que mantienen la información del conjunto de variables explicativas.

### 4.3 COMPROBACIÓN DE LAS HIPÓTESIS INICIALES PARA EL MODELO FACTORIAL

#### 4.3.1 Estudio de la normalidad de los residuos:

##### Gráfico probabilístico

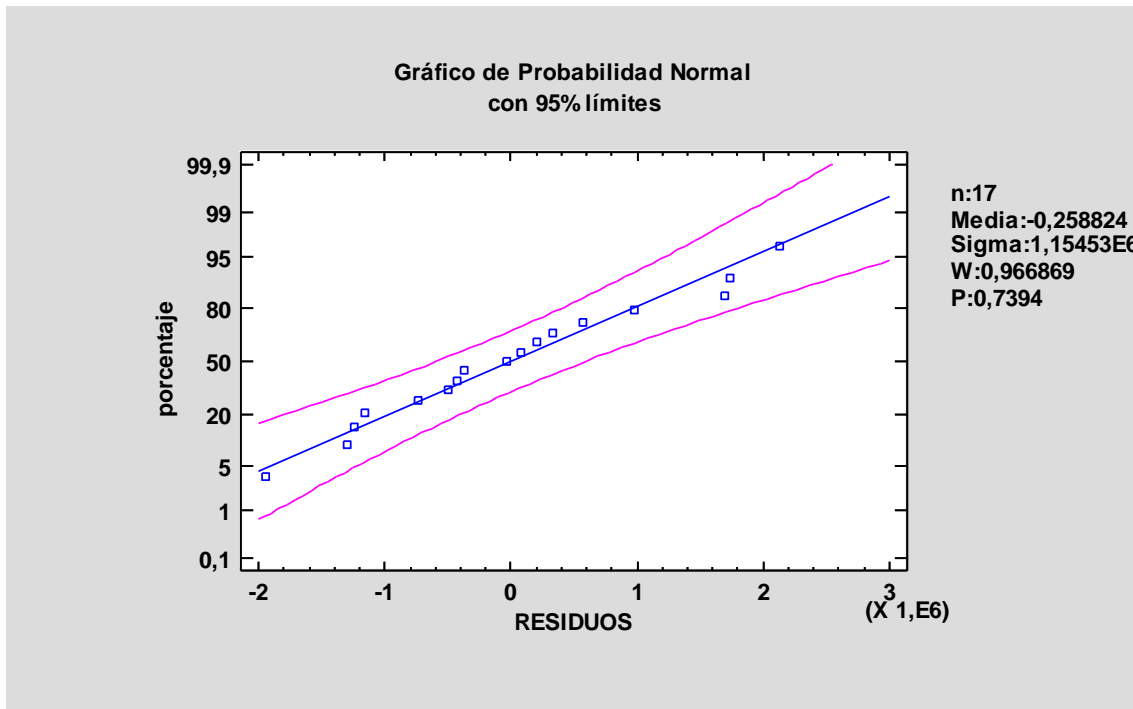


Gráfico 9. Gráfico de Probabilidad Normal

En este gráfico se puede ver como todos los puntos están dentro de los límites, y pegados a la línea, por lo tanto, se observa perfectamente que los residuos son normales. [5]

## Histograma de residuos

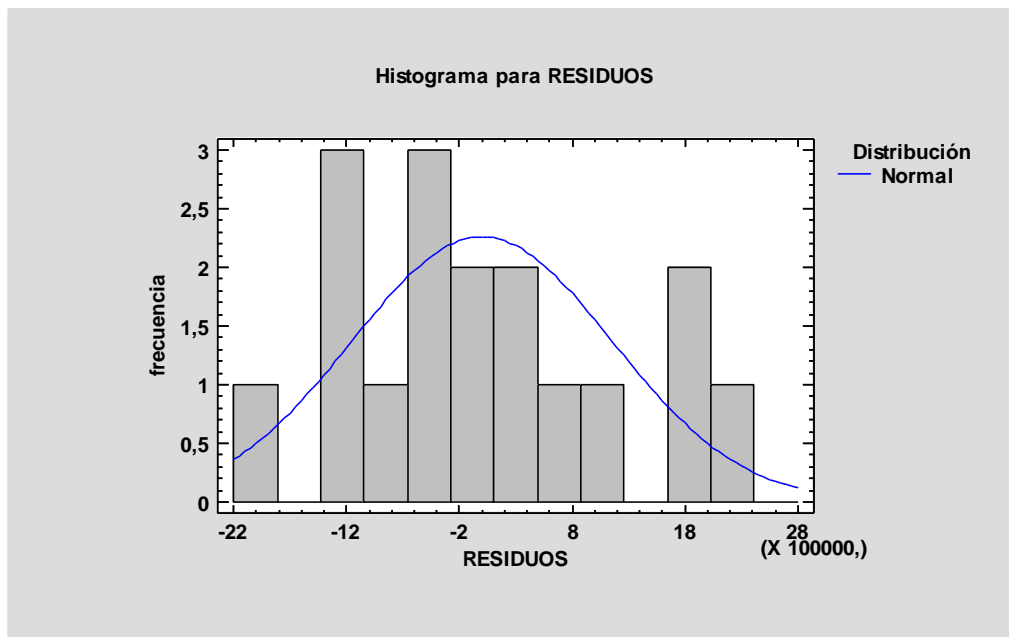


Gráfico 10. Histograma de Residuos

En el histograma también se observa que siguen una distribución normal ya que más o menos se asemeja a la campana de Gauss.

## Contraste de hipótesis

Tabla 22. Normalidad de residuos

Pruebas de Normalidad para RESIDUOS		
Prueba	Estadístico	Valor-P
Chi-Cuadrado	7,70588	0,564036
Estadístico W de Shapiro-Wilk	0,966869	0,739362
Valor-Z para asimetría	0,44882	0,653558
Valor-Z para curtosis	Datos Insuficientes	

Por último, se realiza esta prueba para comprobar que los residuos son normales, en la cual se coge el p-valor más restrictivo, en este caso el de Chi-Cuadrado (0.564036) y como es mayor que  $\alpha$  (0.05), se concluye que los residuos son normales.

### 4.3.2 Detectar gráficamente posibles problemas:

#### Residuos vs X

En este apartado se van a representar los gráficos de los residuos frente a los factores:

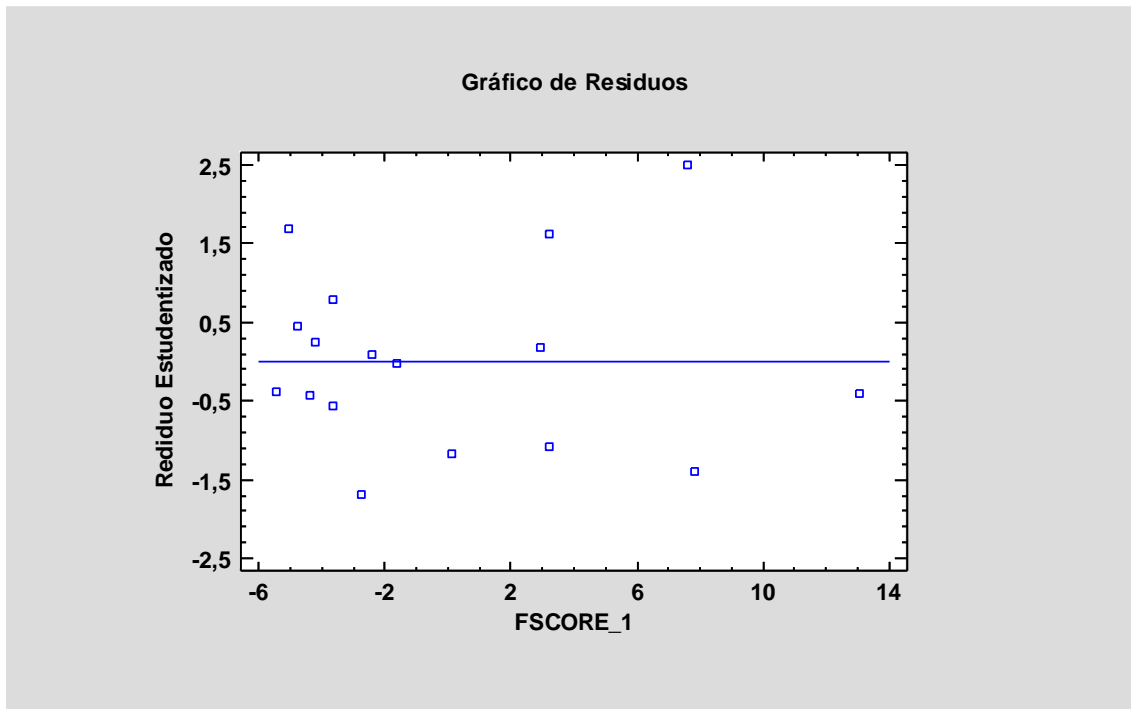


Gráfico 11. Residuos vs Factor 1

Como se puede observar en este gráfico todos los puntos están dispersos y no forman ninguna figura en concreto que nos permita saber si existen problemas.

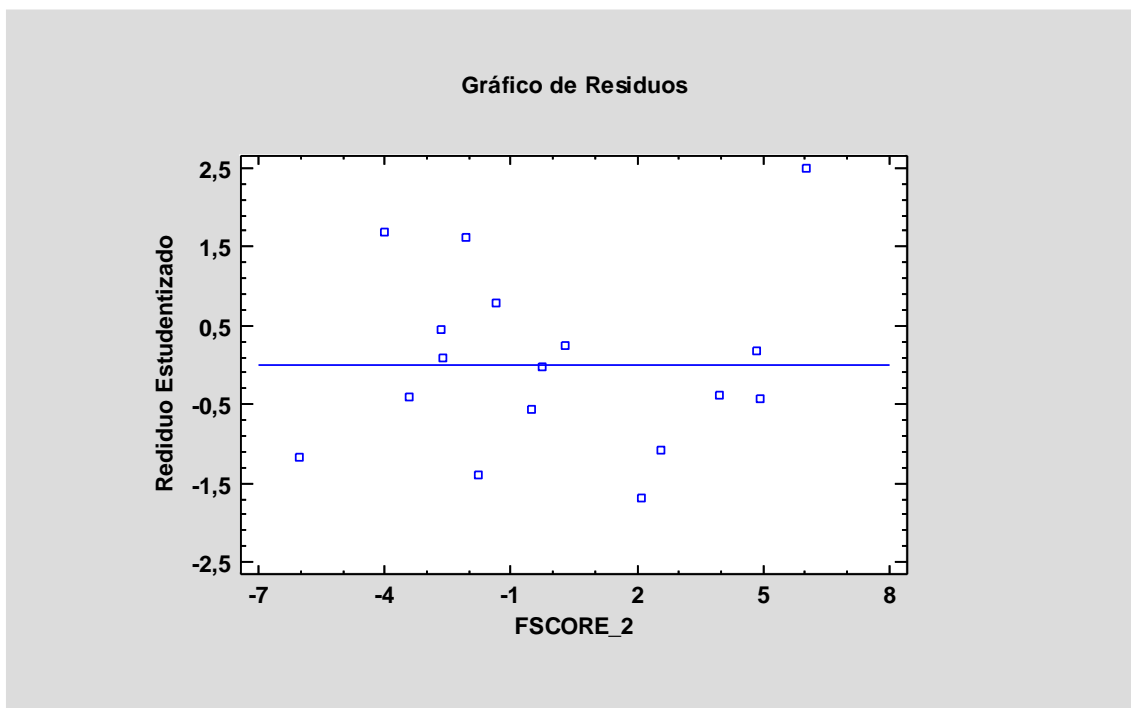


Gráfico 12. Residuos vs Factor 2

En este gráfico también hay una dispersión aleatoria de los puntos por lo cual tampoco encontramos ningún problema, no existen puntos anómalos, pero sí que podríamos observar algún punto extremo.

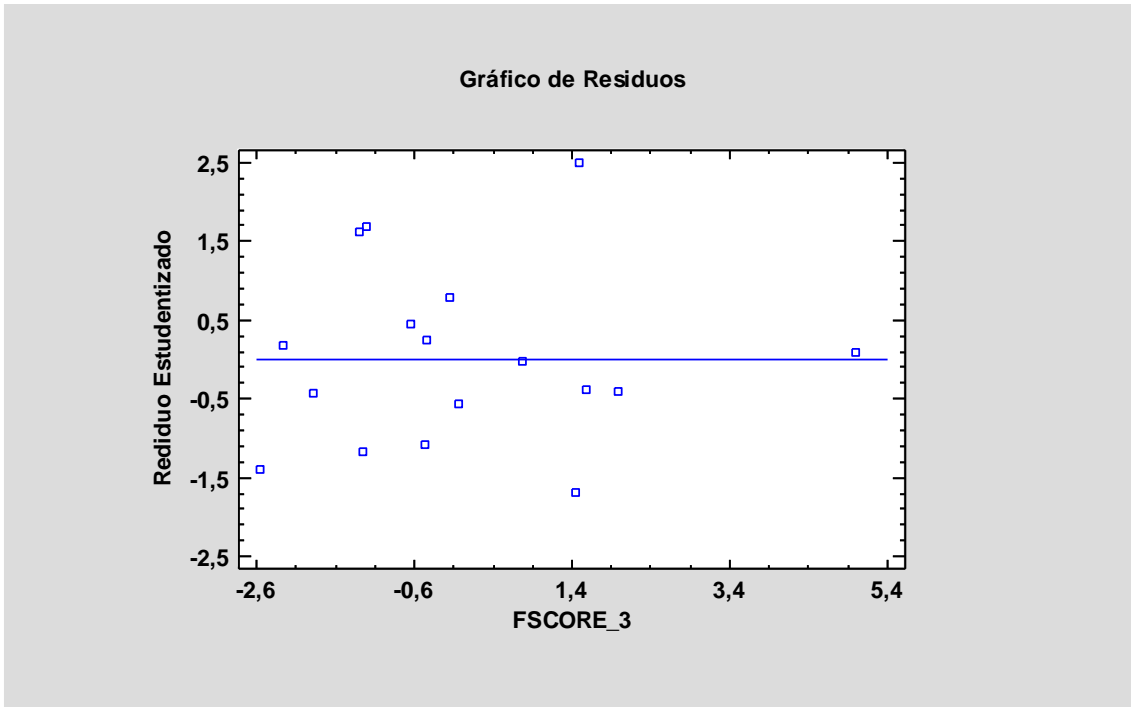


Gráfico 13. Residuos vs Factor 3

En este gráfico se observa una dispersion de puntos que no forman ninguna figura.

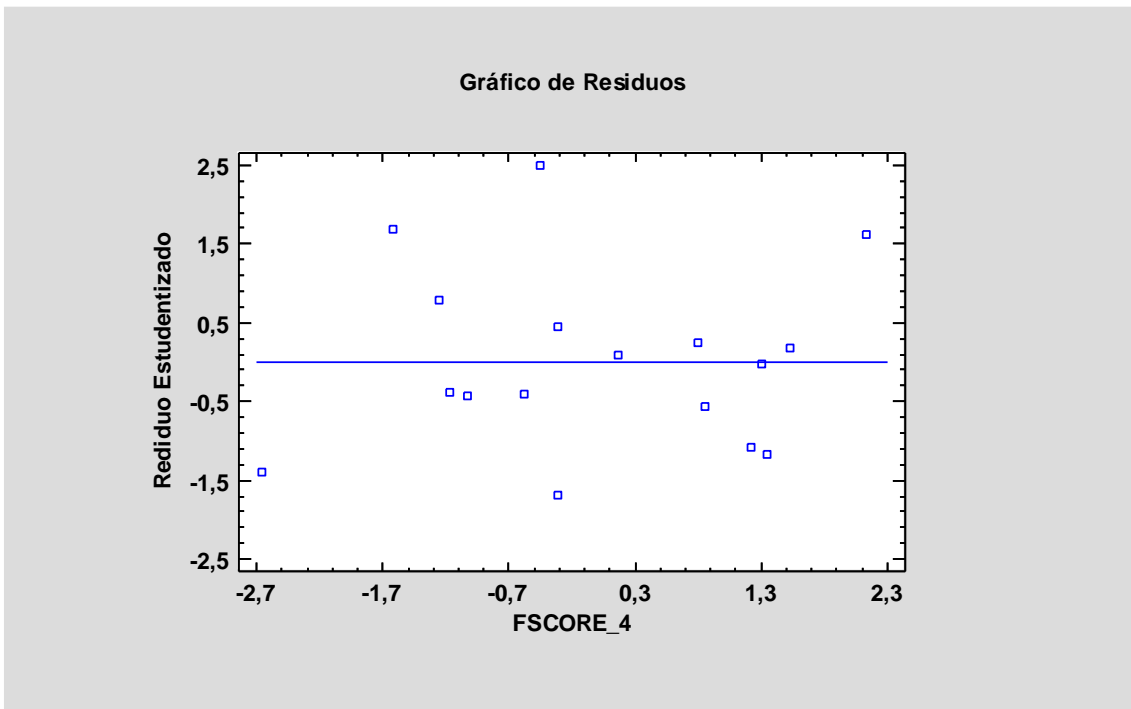


Gráfico 14. Residuos vs Factor 4



Y, por último, en este gráfico tampoco existe ninguna forma en la cual se pueda detectar algún problema, por lo tanto, se concluye que los gráficos son normales y tienen una distribución de puntos dispersos.

### Residuos vs Y predicho

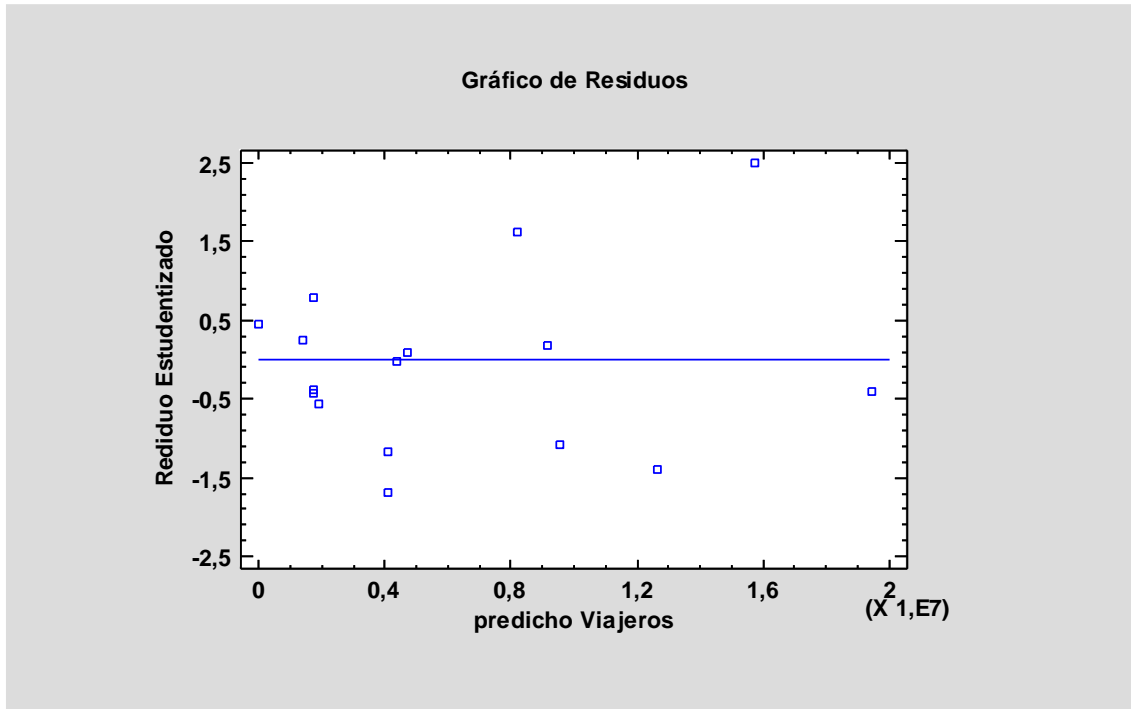


Gráfico 15. Residuos vs Viajeros

En este gráfico se muestran los residuos frente a los viajeros, en el cual también se observa una dispersión de los puntos.

## Residuos vs nº fila

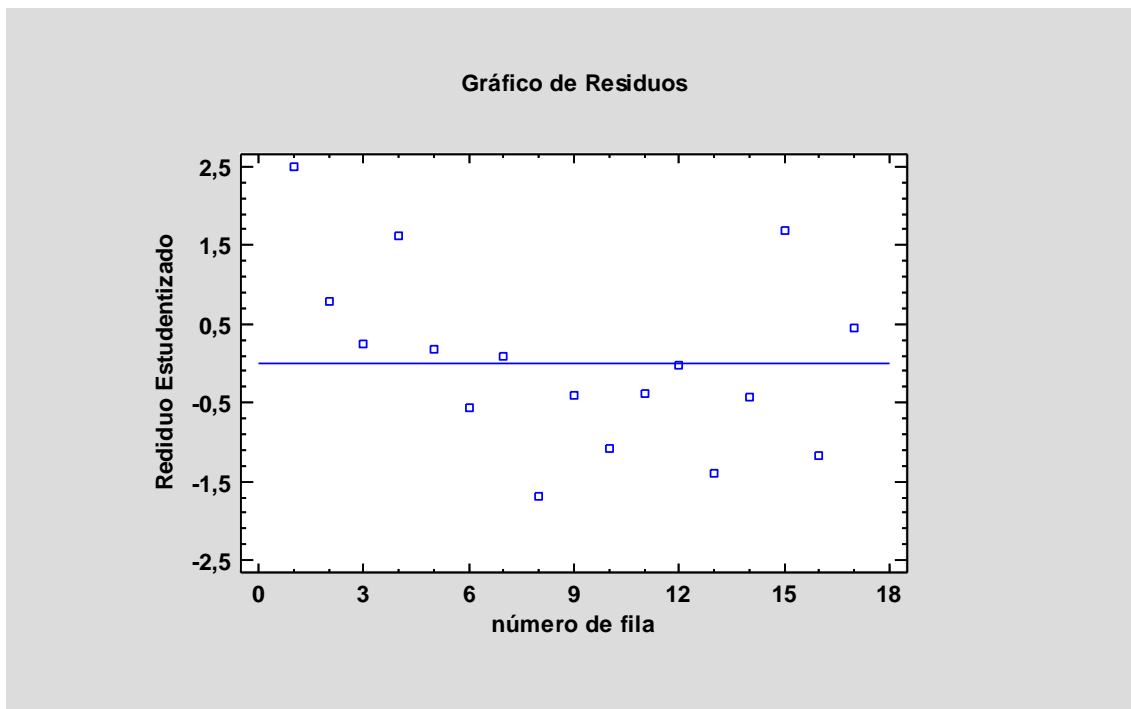


Gráfico 16. Residuos vs nº fila

En estos gráficos no se detecta ningún problema, ya que todos los puntos están dispersos y no forman ninguna figura en concreto. Tampoco se observa ningún punto anómalo.

### 4.3.3 Heterocedasticidad:

#### Contraste de hipótesis modelo análisis factorial

El contraste de hipótesis que se realiza para comprobar posibles problemas de heterocedasticidad, es hacer una regresión múltiple elevando los Residuos<sup>2</sup> frente a los factores. [7]

El modelo es el siguiente:

$$\text{Residuos}^2 = \beta_0 + \text{Factor 1} \cdot \beta_1 + \text{Factor 2} \cdot \beta_2 + \text{Factor 3} \cdot \beta_3 + \text{Factor 4} \cdot \beta_4 + U$$

Tabla 23. Contraste de hipótesis

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	1,25E+12	3,93E+11	3,18853	0,0078
FSCORE_1	5,58E+10	7,43E+10	0,750489	0,4674
FSCORE_2	4,46E+10	1,15E+11	0,388217	0,7047
FSCORE_3	-3,00E+10	2,15E+11	-0,139831	0,8911
FSCORE_4	-9,57E+10	3,08E+11	-0,310363	0,7616

Se deben observar los p-valores de los diferentes componentes: [7]

p-valor Factor 1 = 0,4674 >  $\alpha$  → Se acepta  $H_0$  y se concluye que la variable Factor 1 no genera problemas de heterocedasticidad.

p-valor Factor 2 = 0,7047 >  $\alpha$  → Se acepta  $H_0$  y se concluye que la variable Factor 2 no genera problemas de heterocedasticidad.

p-valor Factor 3 = 0,8911 >  $\alpha$  → Se acepta  $H_0$  y se concluye que la variable Factor 3 no genera problemas de heterocedasticidad.

p-valor Factor 4 = 0,7616 >  $\alpha$  → Se acepta  $H_0$  y se concluye que la variable Factor no genera problemas de heterocedasticidad.

Como ninguna de las variables es significativa, no hay problemas de heterocedasticidad.

#### 4.3.4 Autocorrelación:

En la autocorrelación, se generan los gráficos FAS y FAP con la finalidad de comprobar si alguna de las variables sobrepasa los límites de autocorrelación y solucionarlo, para generar un modelo óptimo. [6]

#### **FAS**

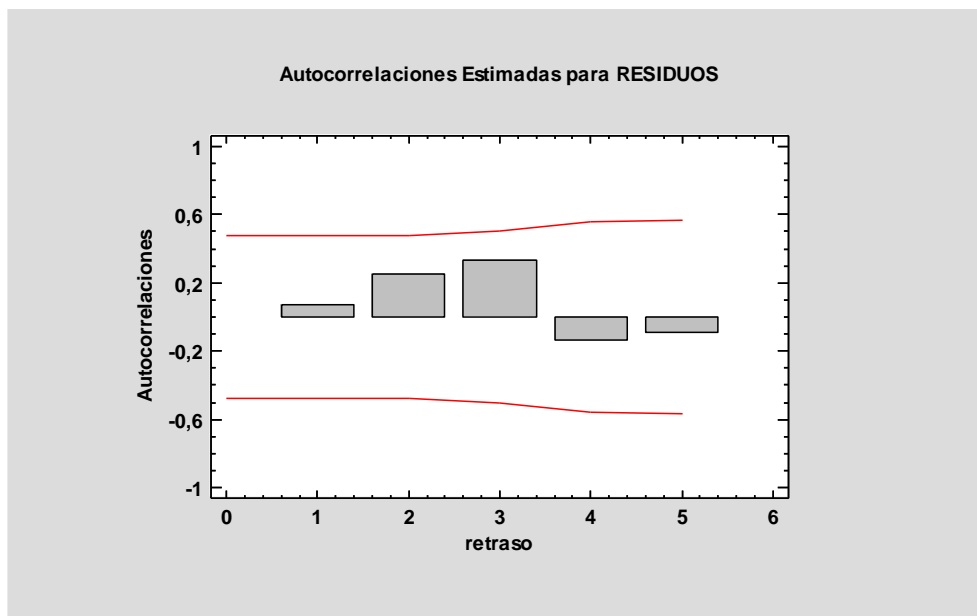


Gráfico 17. FAS

En el gráfico FAS se observan que los factores no sobrepasan los límites de autocorrelación, por lo tanto, a simple vista se puede observar que no existen problemas de autocorrelación.

## FAP

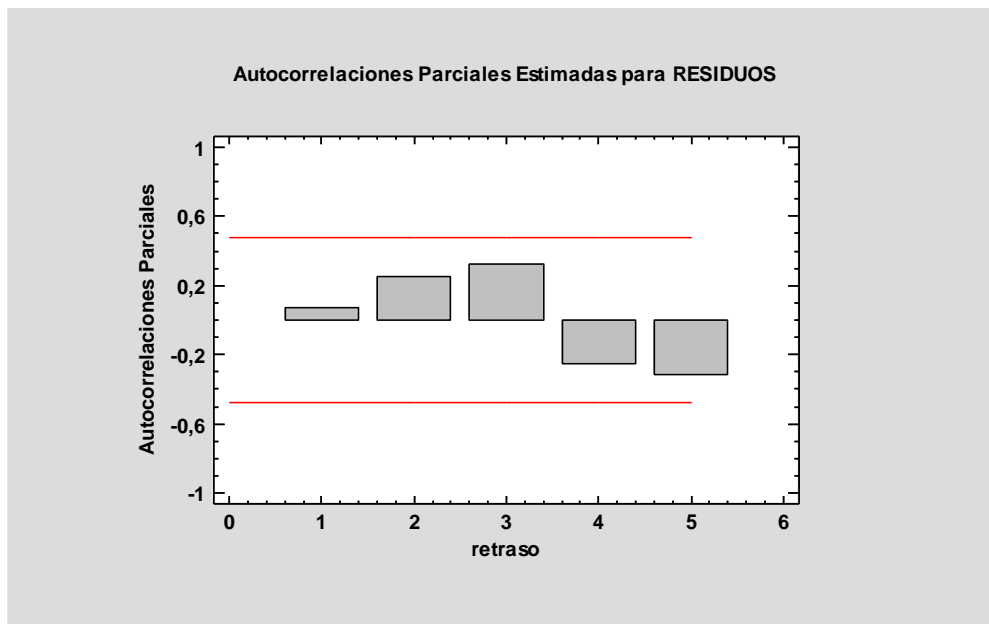


Gráfico 18. FAP

En cuanto al gráfico FAP, se observa lo mismo que en el gráfico FAS, por lo tanto, se concluye que no existen problemas de autocorrelación.

### 4.3.5 Análisis de la significatividad:

#### **Contraste significación global (modelo)**

Tabla 24. Significatividad del modelo

Análisis de Varianza					
Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	5,33E+14	4	1,33E+14	75,05	0
Residuo	2,13E+13	12	1,78E+12		
Total (Corr.)	5,55E+14	16			

Ho:  $\beta_1 = \beta_2 = \dots = \beta_k = 0$

H1: Al menos uno es  $\neq 0$

El p-valor del modelo es 0, por lo tanto, se rechaza Ho y se concluye que el modelo es significativo, ya que es menor que 0,05.

## Contraste significación individual (parámetros)

Tabla 25. Significatividad de los factores

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	5,87E+06	323333	18,1394	0
FSCORE_1	1,04E+06	61061,3	17,0416	0
FSCORE_2	230272	94512,6	2,43642	0,0314
FSCORE_3	388788	176518	2,20255	0,0479
FSCORE_4	-20699,9	253270	-0,0817306	0,9362

Constante  $\rightarrow$  p-valor = 0 <  $\alpha(0,05)$ , Se rechaza  $H_0$  y se concluye que la Constante es significativa.

Factor 1  $\rightarrow$  p-valor = 0 <  $\alpha(0,05)$ , Se rechaza  $H_0$  y se concluye que el Factor 1 es significativo.

Factor 2  $\rightarrow$  p-valor = 0.0314 <  $\alpha(0,05)$ , Se rechaza  $H_0$  y se concluye que el Factor 2 es significativo.

Factor 3  $\rightarrow$  p-valor = 0.0479 <  $\alpha(0,05)$ , Se rechaza  $H_0$  y se concluye que el Factor 3 es significativo.

Factor 4  $\rightarrow$  p-valor = 0.9362 >  $\alpha(0,05)$ , Se acepta  $H_0$  y se concluye que el Factor 4 no es significativo.

El modelo final sería el siguiente:

Después de aplicar la selección paso a paso hacia adelante, en la cual se elimina el factor 4 que no es significativo.

*Tabla 26. Modelo óptimo*

		Error	Estadístico		
Parámetro	Estimación	Estándar	T	Valor-P	
CONSTANTE	5,87E+06	310734	18,8749	0	
FSCORE_1	1,04E+06	58681,2	17,7324	0	
FSCORE_2	230313	90828,7	2,53569	0,0249	
FSCORE_3	388803	169640	2,29194	0,0392	
Análisis de Varianza					
Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	5,33E+14	3	1,78E+14	108,34	0
Residuo	2,13E+13	13	1,64E+12		
Total (Corr.)	5,55E+14	16			
R-cuadrada = 96,1539 por ciento					
R-cuadrado (ajustado para g.l.) = 95,2664 por ciento					
Error estándar del est. = 1,28119E6					
Error absoluto medio = 906095,					
Estadístico Durbin-Watson = 1,6388 (P=0,1943)					
Autocorrelación de residuos en retraso 1 = 0,0659287					

Las variables que identificamos como más importantes son: Gasto total viaje, Gasto medio viaje, Gasto medio diario, IPH, Habitantes, Viajeros hotel, Renta media, Tasa riesgo pobreza, Viajeros turismo rural y IPC, ya que son las que más peso tienen en los tres factores.

#### 4.4 PREDICCIONES

En este apartado se van a realizar 3 predicciones, en la cual una estará enfocada al grupo de Comunidades Autónomas que reciben más viajeros, otra a los que reciben menos viajeros y la última predicción a las CCAA que reciben el menor número de viajeros.

Con tal de realizar las predicciones se utilizarán los datos de 2015 ya se tienen de todas las variables, y así se puede comprobar que el modelo es válido si el número total de viajeros se acerca a los datos que se han sacado del INE. [15]

Según los datos de 2015, se han calculado los factores. Una vez expuestos y realizar la regresión múltiple con la finalidad de averiguar el número de viajeros, los resultados han sido los siguientes:

##### **Grupo 1 (Andalucía y Cataluña):**

Los valores de las variables son: Gasto total viaje (14256,25), Gasto medio viaje (985), Gasto medio diario (128), Duración media (7,5), IPH (5,1), Viajeros camping (1.259.500), Renta media (10.530), Tasa riesgo pobreza (24,1), Habitantes (7.967.452), IPT (98,53), IPC (102,14), Tasa actividad (59,21), Tasa de paro (18,52), PIB (181.638.895), Viajeros hotel (18.480.329), Viajeros turismo rural (331.058).

Obteniendo a partir de estos los siguientes valores para los factores:

F1: 9.0562

F2: 3.889

F3: 0.658

Se obtiene un número de viajeros total para 2015 de 16.440.100 viajeros.

*Tabla 27. Regresión Viajeros Grupo 1*

Resultados de la Regresión de Viajeros (Grupo 1)					
	Ajustado	Error Est.	Inferior 95,0%	Superior 95,0%	Inferior 95,0%
		LC para Pronóstico	LC para Pronóstico	LC para Pronóstico	LC para la Media
Fila 18	16440100	1470200	13263900	19616200	14882200

##### **Grupo 2 (Comunidad Valenciana, Madrid, Islas Canarias e Islas Baleares):**

Los valores de las variables son: Gasto total viaje (10.562), Gasto medio viaje (1.053), Gasto medio diario (130), Duración media (7,9), IPH (5,8), Viajeros camping (193.421), Renta media (10.650), Tasa riesgo pobreza (23,2), Habitantes (3.670.023), IPT (101),

IPC (101,46), Tasa actividad (59,99), Tasa de paro (16,85), PIB (89.340.123), Viajeros hotel (9.750.423), Viajeros turismo rural (151.563).

Obteniendo a partir de estos los siguientes valores para los factores:

F1: 1.33

F2: 1.25

F3: 0.12

Se obtiene un número de viajeros total para 2015 de 7.583.560 viajeros.

*Tabla 28. Regresión Viajeros (Grupo 2)*

Resultados de la Regresión de Viajeros (Grupo 2)					
	Ajustado	Error Est.	Inferior 95,0%	Superior 95,0%	Inferior 95,0%
		LC para Pronóstico	LC para Pronóstico	LC para Pronóstico	LC para la Media
Fila 18	7583560	1325740	4719470	10447600	6847350

**Grupo 3 (Aragón, Asturias, Cantabria, Castilla y León, Castilla la Mancha, Extremadura, Galicia, Murcia, Navarra, País Vasco y La Rioja):**

Los valores de las variables son: Gasto total viaje (4.358), Gasto medio viaje (785), Gasto medio diario (97), Duración media (7,9), IPH (3,1), Viajeros camping (141.652), Renta media (9.985), Tasa riesgo pobreza (18,56), Habitantes (1.436.852), IPT (99,93), IPC (101,24), Tasa actividad (56,21), Tasa de paro (14,52), PIB (33.090.256), Viajeros hotel (2.586.452), Viajeros turismo rural (215.348).

Obteniendo a partir de estos los siguientes valores para los factores:

F1: -3.35

F2: -2.589

F3: -0.56

Se obtiene un número de viajeros total para 2015 de 1.565.200 viajeros.



Tabla 29. Regresión Viajeros (Grupo 3)

Resultados de la Regresión de Viajeros (Grupo 3)					
	Ajustado	Error Est.	Inferior 95,0%	Superior 95,0%	Inferior 95,0%
		LC para Pronóstico	LC para Pronóstico	LC para Pronóstico	LC para la Media
Fila 18	1565200	1357360	-1367190	4497600	596727

Como conclusión final de estas tres predicciones se puede afirmar que el modelo es bueno, ya que las predicciones y los datos obtenidos se acercan a los viajeros de 2015, por lo tanto, el modelo es bueno para realizar predicciones.



# Capítulo 5

## Conclusiones

En este capítulo se resumen las conclusiones obtenidas durante todo el trabajo.

Los objetivos específicos que se han perseguido en este trabajo han sido los siguientes:

- Búsqueda de información sobre el turismo en España.
- Análisis de las variables más importantes.
- Búsqueda de otros trabajos relacionados.
- Identificación de las variables más importantes.
- Generación del modelo óptimo aplicando las diferentes técnicas
- Realización de predicciones

Como conclusiones finales de este trabajo se pueden sacar las siguientes:

- El modelo presenta problemas de multicolinealidad que se han resuelto mediante el método de análisis factorial, una vez realizadas las demás técnicas, se obtiene un modelo válido con 3 factores significativos y del cual no se obtiene ningún problema.
- Existen tres grupos de recepción de turistas según Comunidad Autónoma. Un grupo que recibe más turistas (Grupo 1), otro grupo que recibe menos (Grupo 2) y, por último, un grupo formado por más Comunidades que son los que menos turistas reciben (Grupo 3).
- El modelo creado es válido, ya que las predicciones realizadas se ajustan a los datos de 2015.
- Las variables más importantes serían la de los Gastos, los índices de precios y los habitantes. Estas son las variables que mejor explican el comportamiento de los viajeros por Comunidades Autónomas.
- Y, por último, como valoración personal de este proyecto creo que ha sido muy interesante, he aplicado los conceptos aprendidos en diferentes asignaturas, así como las prácticas de Econometría. Es un buen proyecto para concluir definitivamente mi etapa universitaria en el Grado de Administración y Dirección de Empresas en la Escuela Politécnica Superior de Alcoy.

# BIBLIOGRAFÍA

[1] **Noticia sobre el turismo:**

<https://www2.deloitte.com/es/es/pages/consumer-business/articles/tendencias-y-evolucion-del-turismo.html> (Junio 2018)

[2] **Informe de la actividad turística y el empleo 2016:**

[https://static.hosteltur.com/web/uploads/2017/02/2016\\_Informe\\_propuestas\\_Industria\\_turísticas\\_al\\_nivel\\_nacional\\_final.pdf](https://static.hosteltur.com/web/uploads/2017/02/2016_Informe_propuestas_Industria_turísticas_al_nivel_nacional_final.pdf)

[3] **Datos PIB:**

<https://www.datosmacro.com/pib/espana-comunidades-autonomas>

[4] **Multicolinealidad:**

<https://www.uv.es/uriel/material/multicolinealidad3.pdf>

[5] **Normalidad:**

[http://dm.udc.es/asignaturas/estadistica2/sec9\\_4.html](http://dm.udc.es/asignaturas/estadistica2/sec9_4.html)

[6] **Conceptos básicos sobre la autocorrelación en el modelo básico de regresión lineal:**

[https://www.uam.es/personal\\_pdi/economicas/rarce/pdf/autocorrel.pdf](https://www.uam.es/personal_pdi/economicas/rarce/pdf/autocorrel.pdf)

[7] **Concepto Heterocedasticidad:**

<https://es.wikipedia.org/wiki/Heterocedasticidad>

[8] **Estadística univariante:**

[http://halweb.uc3m.es/esp/Personal/personas/aarribas/esp/docs/estI\\_grado/estIG\\_tema1.pdf](http://halweb.uc3m.es/esp/Personal/personas/aarribas/esp/docs/estI_grado/estIG_tema1.pdf)

[9] **Análisis de datos bivariantes:**

[http://www.est.uc3m.es/esp/nueva\\_docencia/getafe/economia/estadistica\\_I/doc\\_generica/tema3.pdf](http://www.est.uc3m.es/esp/nueva_docencia/getafe/economia/estadistica_I/doc_generica/tema3.pdf)

[10] **Análisis factorial:**

<http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/FACTOREAL/analisis-factorial.pdf>

[11] **Modelo Econométrico del Turismo en España presentado por Silvia Véganzones González, Universidad de Valladolid, 2015**

[12] **Modelo econométrico de demanda turística para España presentado por Alfredo de Paz Sastre, Universidad de Valladolid, 2015**

[13] **Modelo econométrico sobre el Turismo, realizado por Ruth Rubio y Miriam Gómez**

[14] **Noticia sobre la importancia del turismo en España:**

<https://www.probuen.es/blog/la-importancia-del-turismo-para-la-economia-espanola/>

[15] **INE:**

<http://www.ine.es/dyngs/INEbase/listaoperaciones.htm>

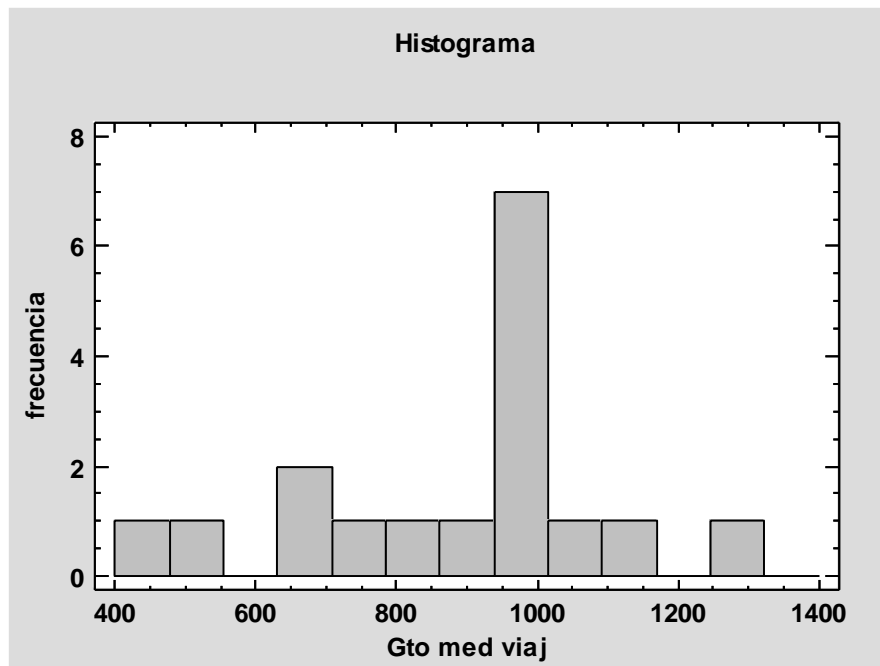
# ANEXO

UNIVARIANTE:

- Gasto medio del viaje:

## Resumen Estadístico para Gto med viaj

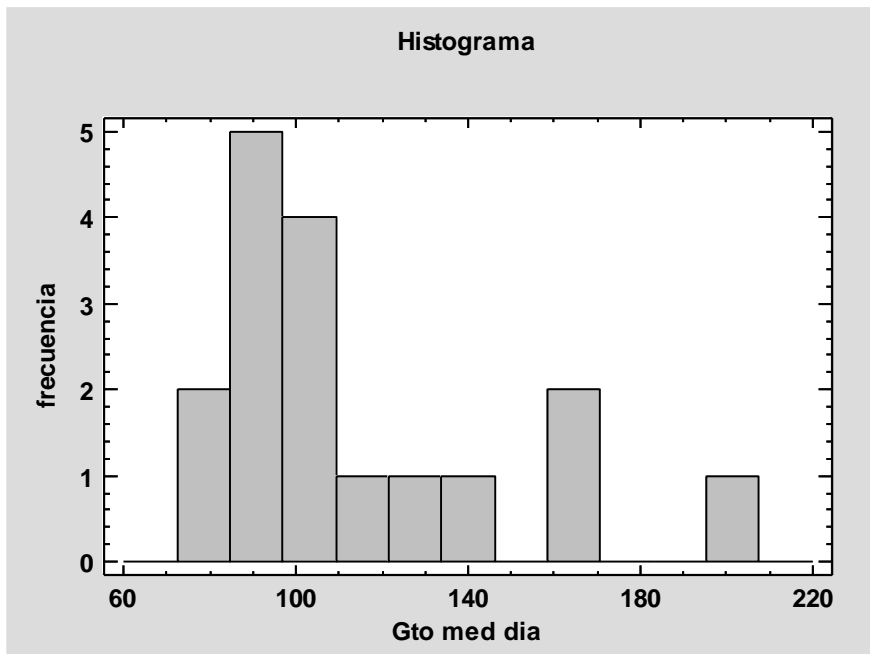
Recuento	17
Promedio	883,941
Mediana	961,0
Desviación Estándar	226,375
Coefficiente de Variación	25,6097%
Mínimo	452,0
Máximo	1314,0
Rango	862,0
Cuartil Inferior	742,0
Cuartil Superior	997,0
Rango Intercuartilico	255,0
Sesgo Estandarizado	-0,671457
Curtosis Estandarizada	0,0616958



- Gasto medio diario:

**Resumen Estadístico para Gto med dia**

Recuento	17
Promedio	113,118
Mediana	98,0
Desviación Estándar	35,5068
Coefficiente de Variación	31,3893%
Mínimo	76,0
Máximo	207,0
Rango	131,0
Cuartil Inferior	91,0
Cuartil Superior	130,0
Rango Intercuartilico	39,0
Sesgo Estandarizado	2,49304
Curtosis Estandarizada	1,47215

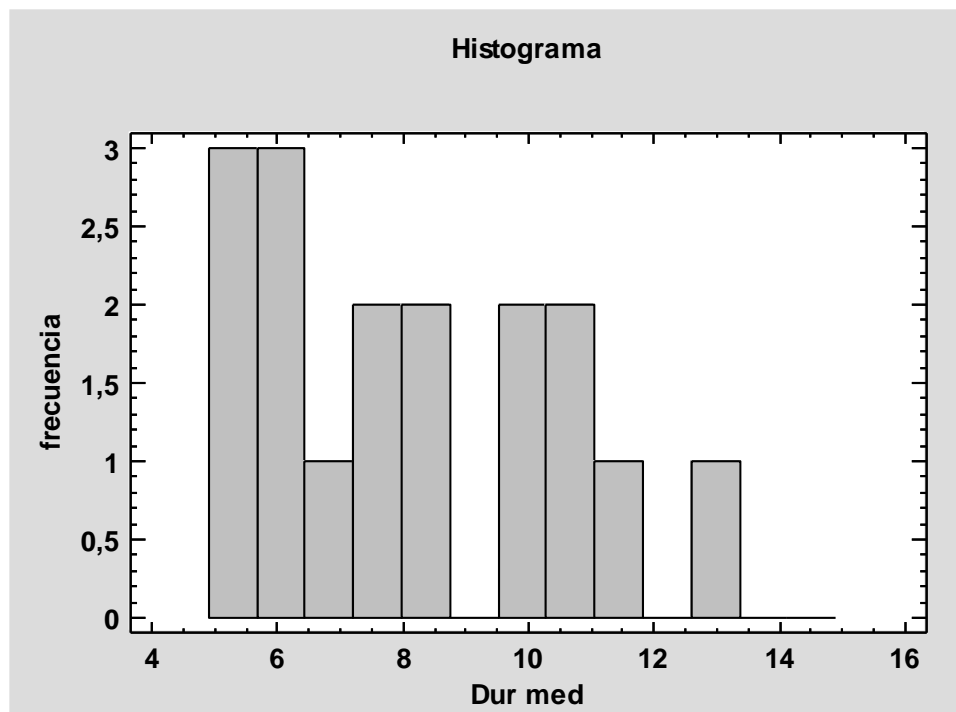




- Duración media:

### Resumen Estadístico para Dur med

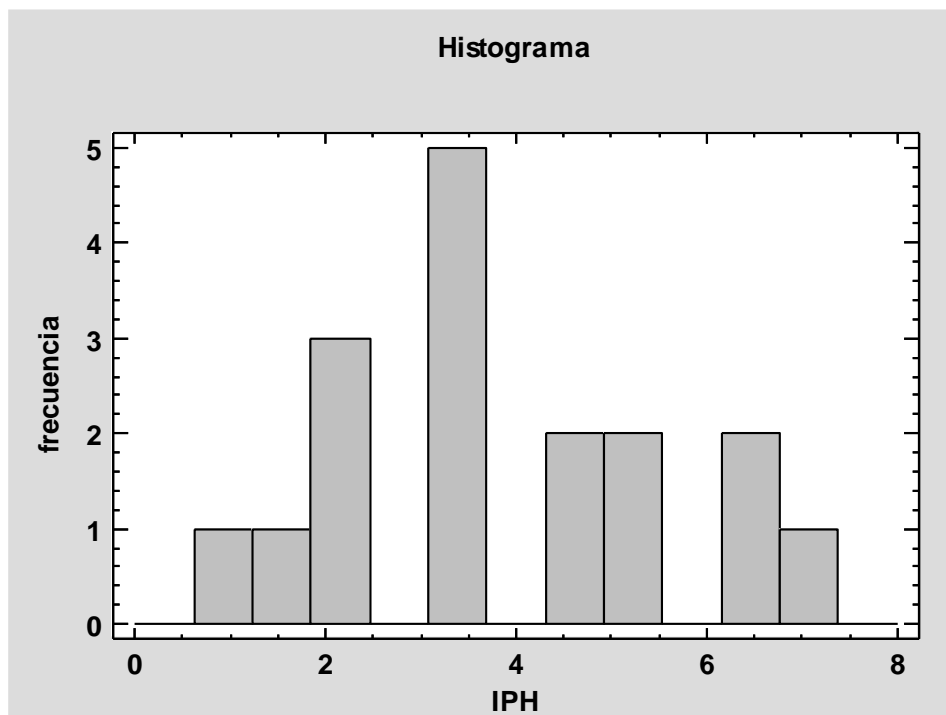
Recuento	17
Promedio	8,14706
Mediana	7,9
Desviación Estándar	2,35588
Coefficiente de Variación	28,9169%
Mínimo	5,3
Máximo	12,7
Rango	7,4
Cuartil Inferior	6,2
Cuartil Superior	10,1
Rango Intercuartílico	3,9
Sesgo Estandarizado	0,817918
Curtosis Estandarizada	-0,819722



- IPH:

### Resumen Estadístico para IPH

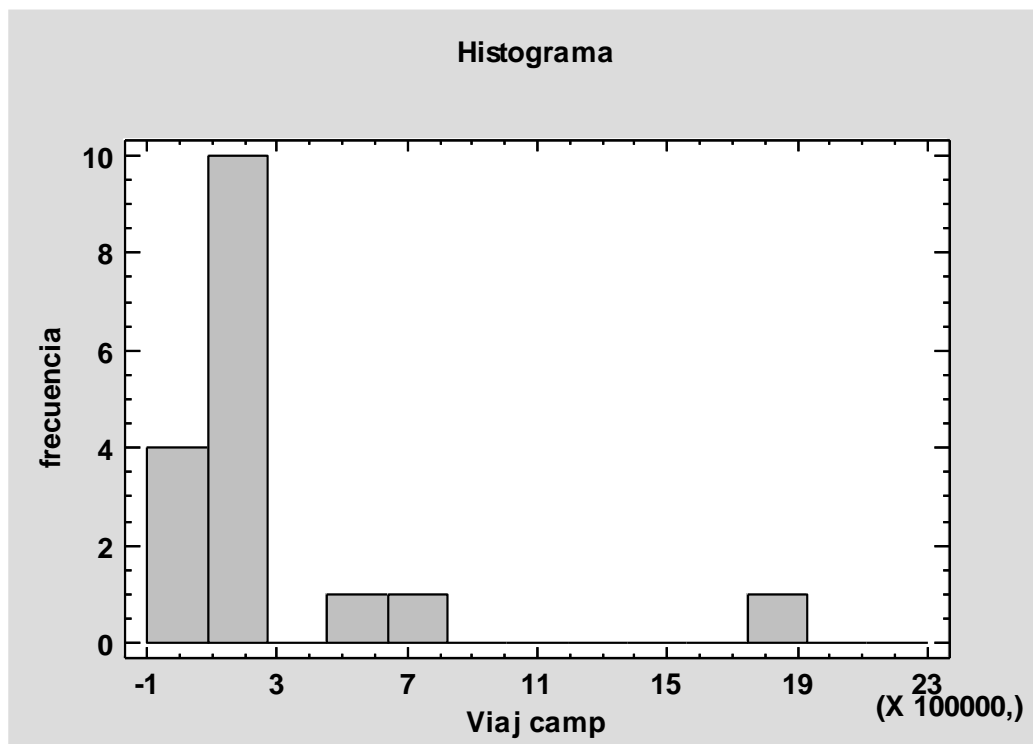
Recuento	17
Promedio	3,88235
Mediana	3,4
Desviación Estándar	1,83106
Coefficiente de Variación	47,1638%
Mínimo	1,1
Máximo	6,8
Rango	5,7
Cuartil Inferior	2,4
Cuartil Superior	5,1
Rango Intercuartílico	2,7
Sesgo Estandarizado	0,485552
Curtosis Estandarizada	-0,806851



- Viajeros camping:

### Resumen Estadístico para Viaj camp

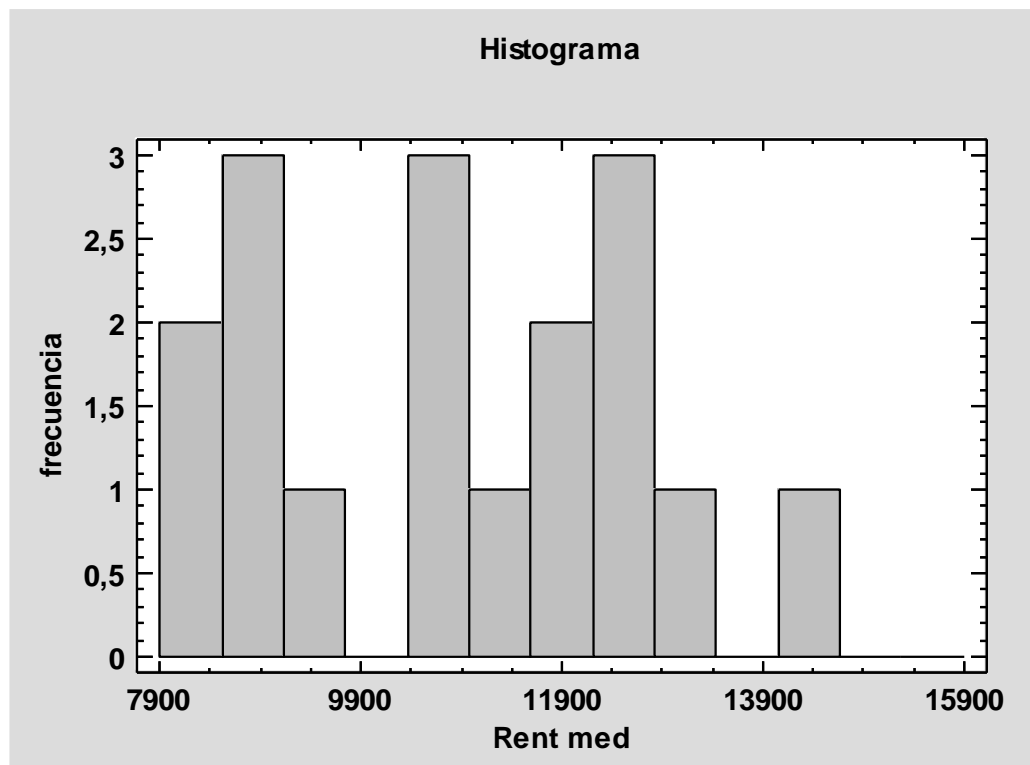
Recuento	17
Promedio	287237,
Mediana	140455,
Desviación Estándar	447468,
Coefficiente de Variación	155,783%
Mínimo	21228,0
Máximo	1,89081E6
Rango	1,86958E6
Cuartil Inferior	94604,0
Cuartil Superior	219048,
Rango Intercuartílico	124444,
Sesgo Estandarizado	5,47051
Curtosis Estandarizada	9,64904



- Renta media:

### Resumen Estadístico para Rent med

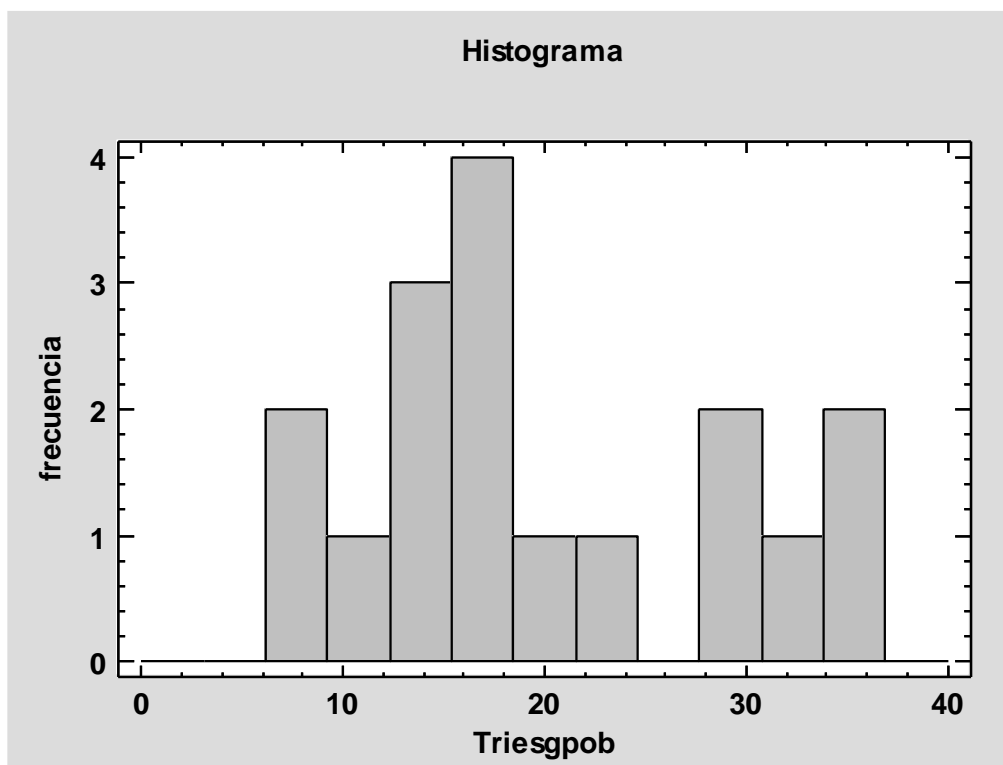
Recuento	17
Promedio	10855,7
Mediana	10815,0
Desviación Estándar	1920,31
Coefficiente de Variación	17,6894%
Mínimo	8273,0
Máximo	14345,0
Rango	6072,0
Cuartil Inferior	8731,0
Cuartil Superior	12222,0
Rango Intercuartilico	3491,0
Sesgo Estandarizado	0,164641
Curtosis Estandarizada	-0,993512



- Tasa riesgo de pobreza:

### Resumen Estadístico para Triesgpob

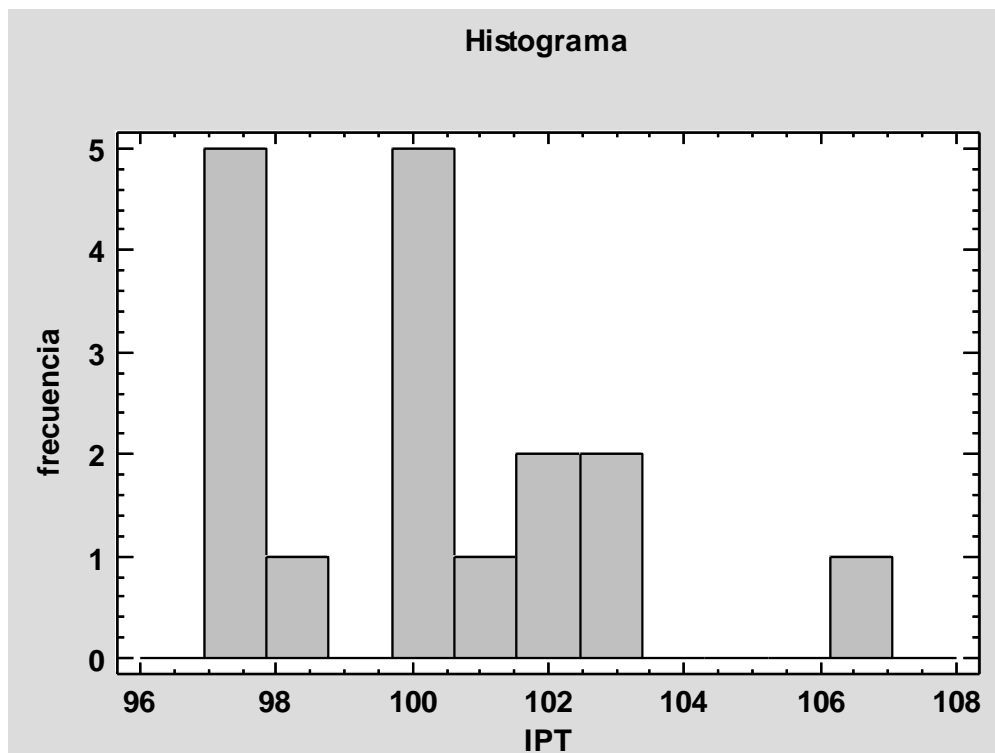
Recuento	17
Promedio	20,2176
Mediana	17,7
Desviación Estándar	8,91889
Coefficiente de Variación	44,1144%
Mínimo	9,0
Máximo	35,4
Rango	26,4
Cuartil Inferior	13,4
Cuartil Superior	28,9
Rango Intercuartílico	15,5
Sesgo Estandarizado	0,945126
Curtosis Estandarizada	-0,926641



- IPT:

### Resumen Estadístico para IPT

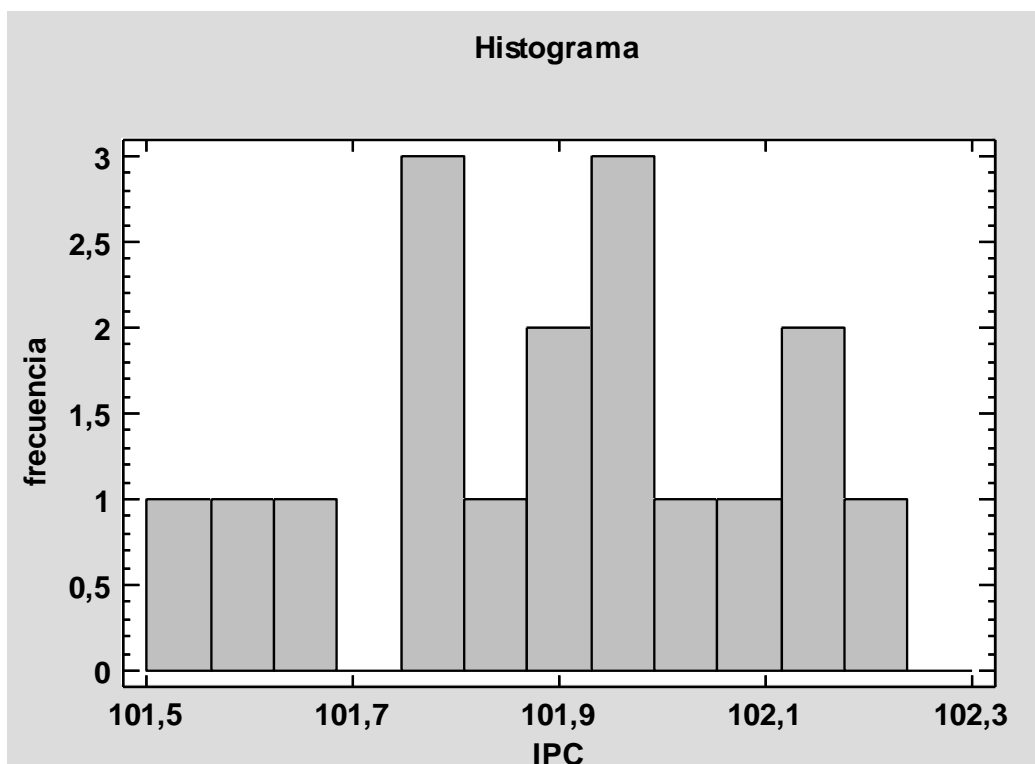
Recuento	17
Promedio	100,113
Mediana	100,313
Desviación Estándar	2,54909
Coefficiente de Variación	2,54622%
Mínimo	97,119
Máximo	106,685
Rango	9,566
Cuartil Inferior	97,543
Cuartil Superior	101,591
Rango Intercuartílico	4,048
Sesgo Estandarizado	1,48804
Curtosis Estandarizada	1,03805



- IPC:

### Resumen Estadístico para IPC

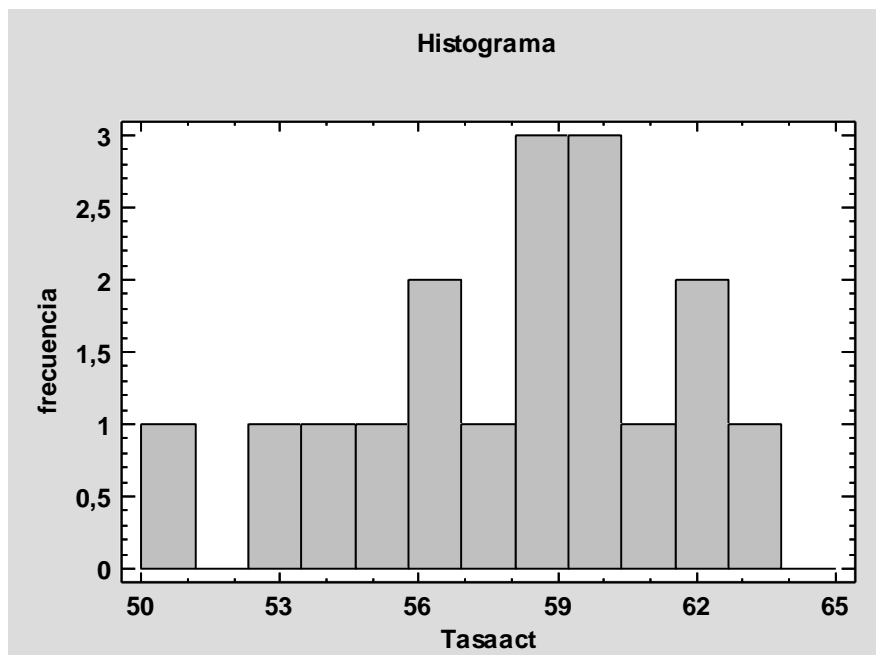
Recuento	17
Promedio	101,902
Mediana	101,93
Desviación Estándar	0,194011
Coefficiente de Variación	0,190391%
Mínimo	101,54
Máximo	102,19
Rango	0,65
Cuartil Inferior	101,79
Cuartil Superior	102,05
Rango Intercuartílico	0,26
Sesgo Estandarizado	-0,60461
Curtosis Estandarizada	-0,586156



- Tasa de actividad:

### Resumen Estadístico para Tasaact

Recuento	17
Promedio	57,8735
Mediana	58,85
Desviación Estándar	3,36981
Coefficiente de Variación	5,82272%
Mínimo	50,67
Máximo	63,32
Rango	12,65
Cuartil Inferior	55,85
Cuartil Superior	59,5
Rango Intercuartílico	3,65
Sesgo Estandarizado	-0,825416
Curtosis Estandarizada	-0,152576

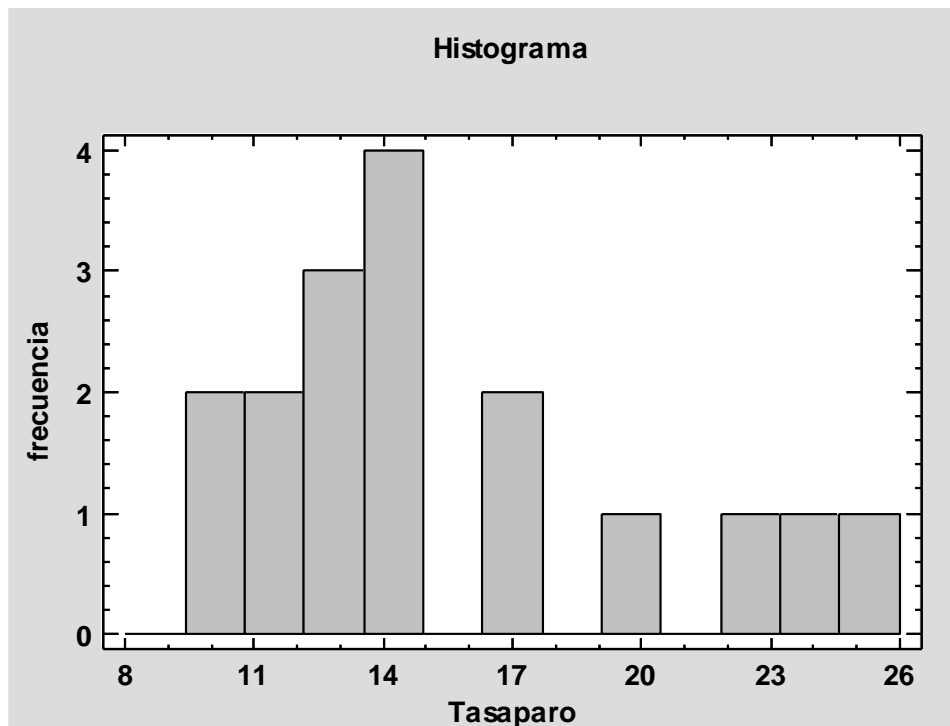




- Tasa de paro:

### Resumen Estadístico para Tasaparo

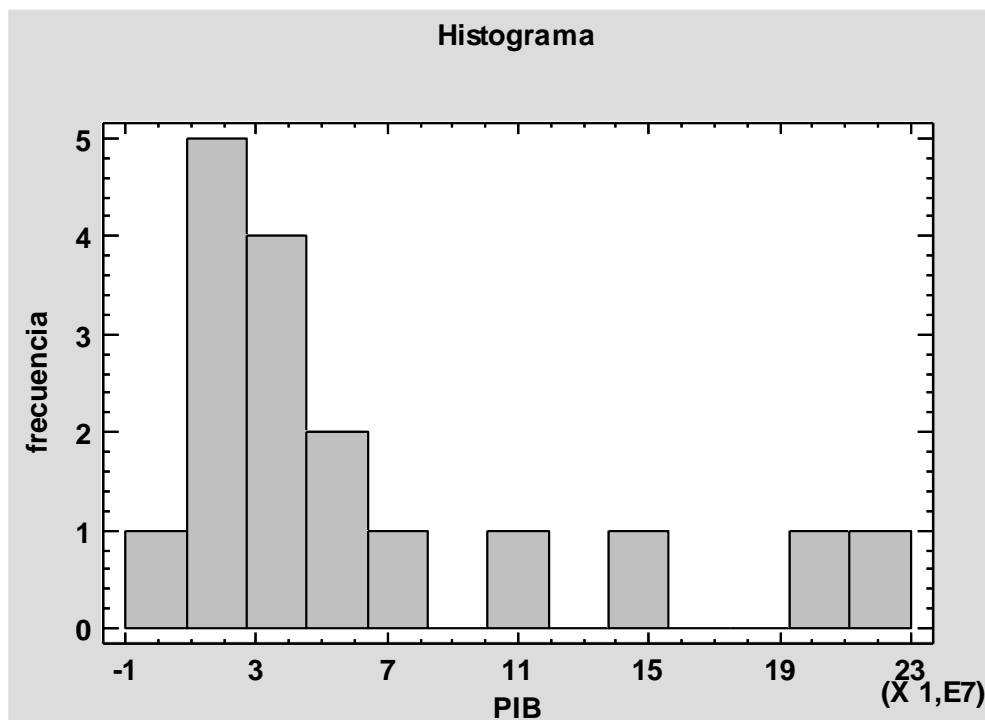
Recuento	17
Promedio	15,5247
Mediana	13,75
Desviación Estándar	4,72717
Coefficiente de Variación	30,4493%
Mínimo	9,63
Máximo	25,12
Rango	15,49
Cuartil Inferior	12,61
Cuartil Superior	17,21
Rango Intercuartílico	4,6
Sesgo Estandarizado	1,59676
Curtosis Estandarizada	-0,0934982



- PIB:

### Resumen Estadístico para PIB

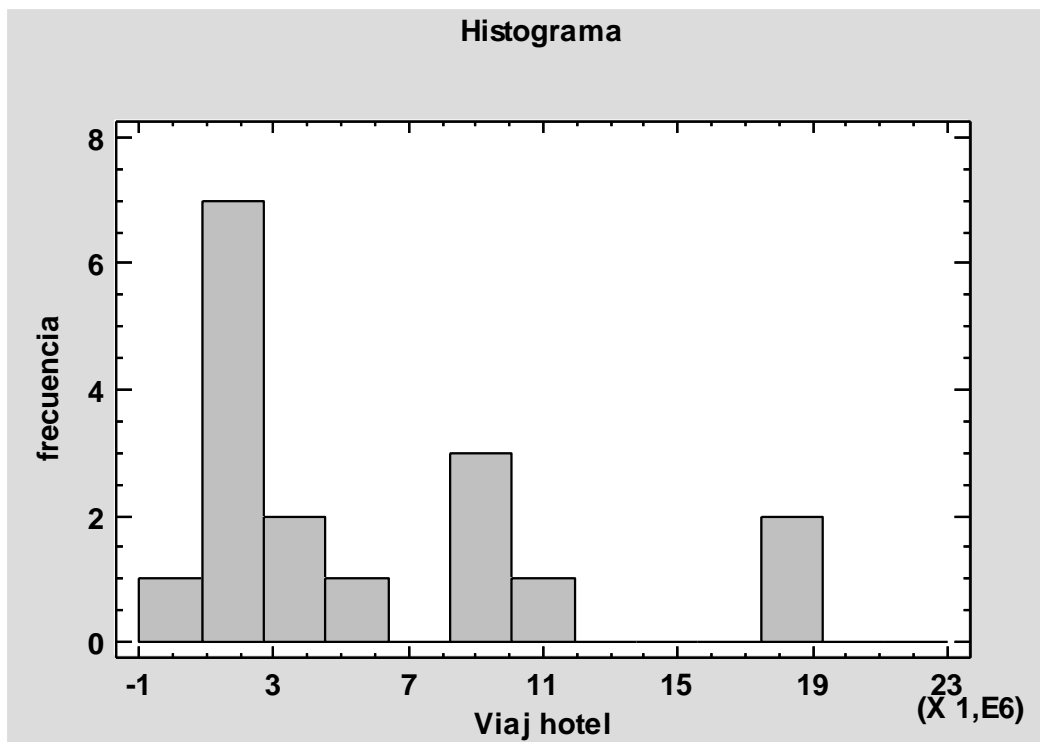
Recuento	17
Promedio	6,38011E7
Mediana	3,43681E7
Desviación Estándar	6,69172E7
Coefficiente de Variación	104,884%
Mínimo	7,91511E6
Máximo	2,13766E8
Rango	2,0585E8
Cuartil Inferior	1,91524E7
Cuartil Superior	6,88172E7
Rango Intercuartílico	4,96648E7
Sesgo Estandarizado	2,57141
Curtosis Estandarizada	1,04293



- Viajeros hotel:

### Resumen Estadístico para Viaj hotel

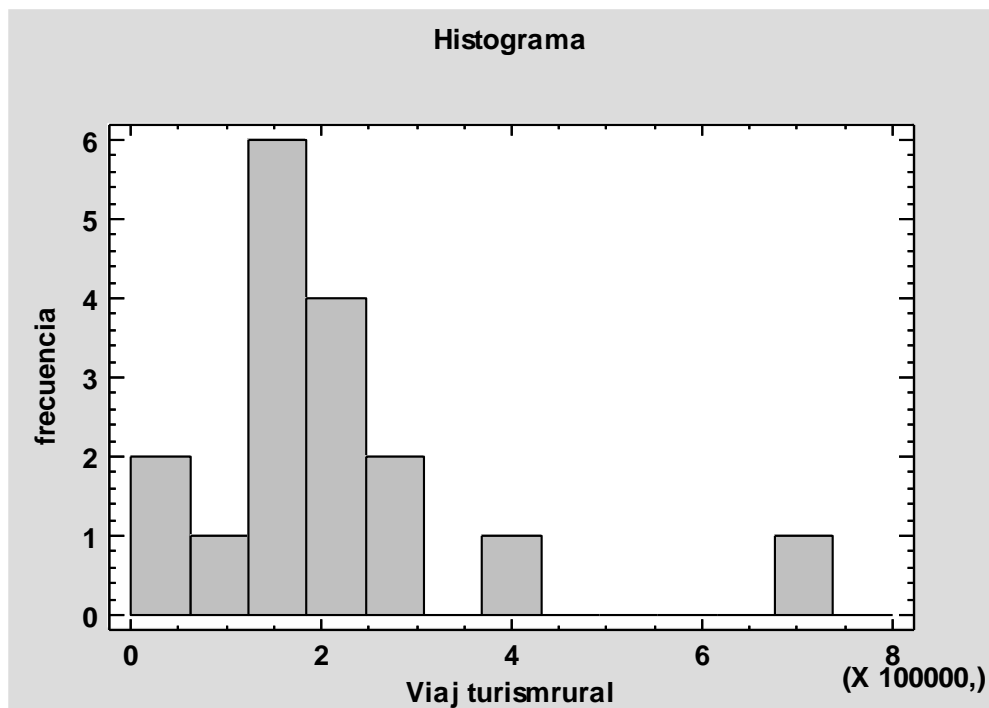
Recuento	17
Promedio	5,86507E6
Mediana	2,83919E6
Desviación Estándar	5,88866E6
Coefficiente de Variación	100,402%
Mínimo	581746,
Máximo	1,90944E7
Rango	1,85126E7
Cuartil Inferior	1,3219E6
Cuartil Superior	9,35169E6
Rango Intercuartílico	8,0298E6
Sesgo Estandarizado	2,09858
Curtosis Estandarizada	0,507008



- Viajeros turismo rural:

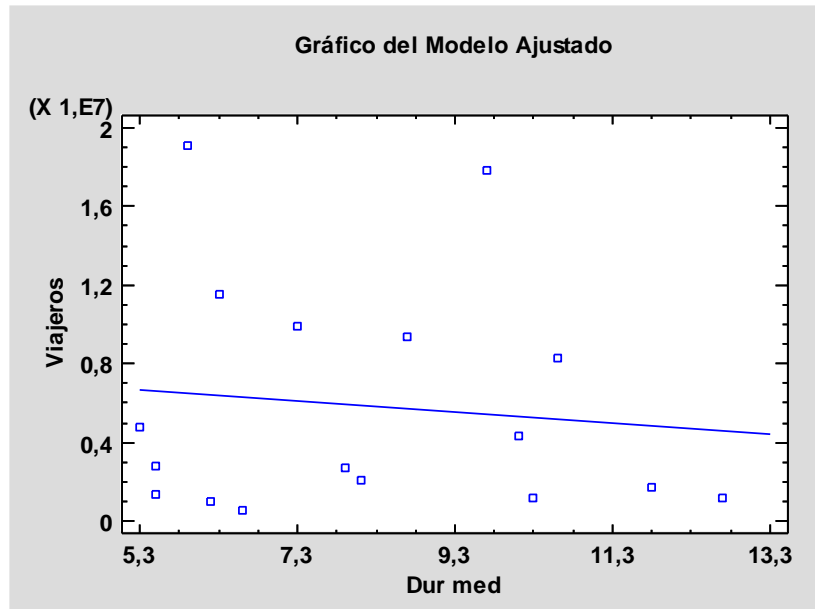
### Resumen Estadístico para Viaj turismrural

Recuento	17
Promedio	213742,
Mediana	172563,
Desviación Estándar	156271,
Coefficiente de Variación	73,1119%
Mínimo	35728,0
Máximo	715702,
Rango	679974,
Cuartil Inferior	144565,
Cuartil Superior	231517,
Rango Intercuartílico	86952,0
Sesgo Estandarizado	3,69662
Curtosis Estandarizada	5,51089

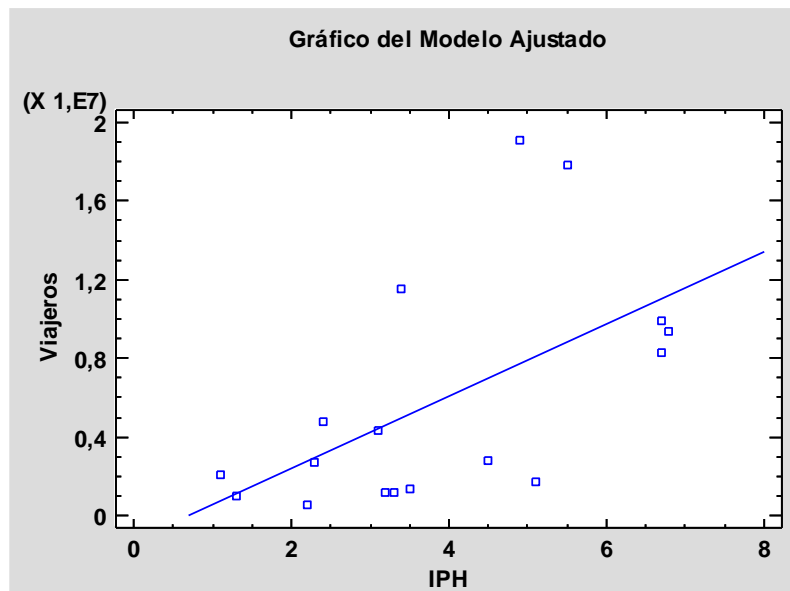


BIVARIANTE:

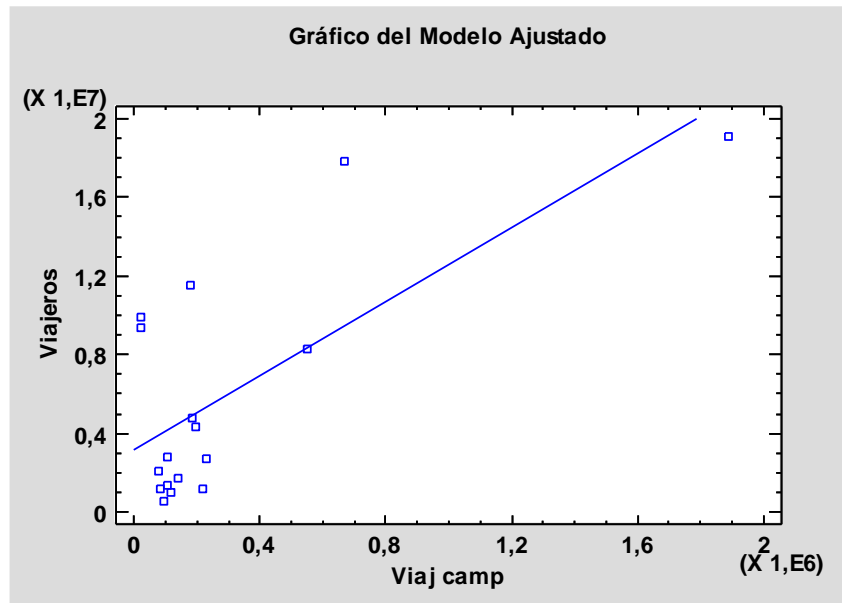
- Viajeros vs Duración media:



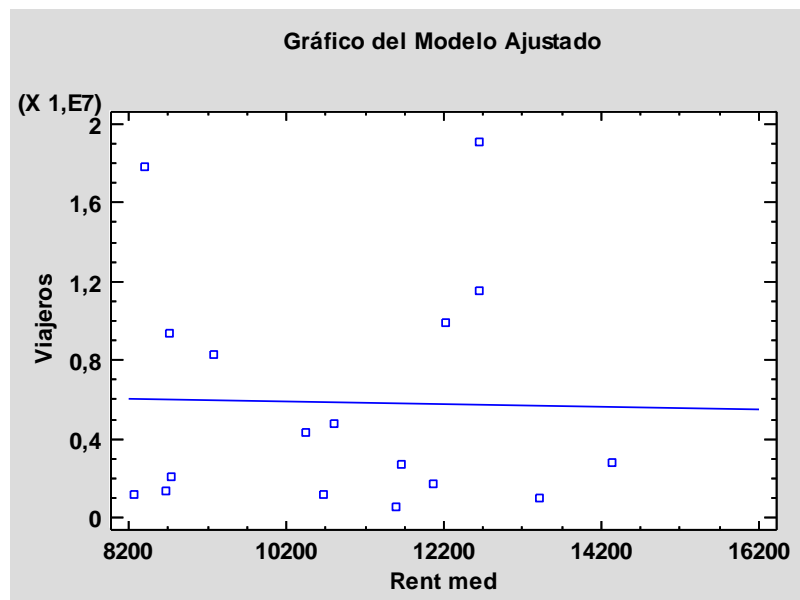
- Viajeros vs IPH:



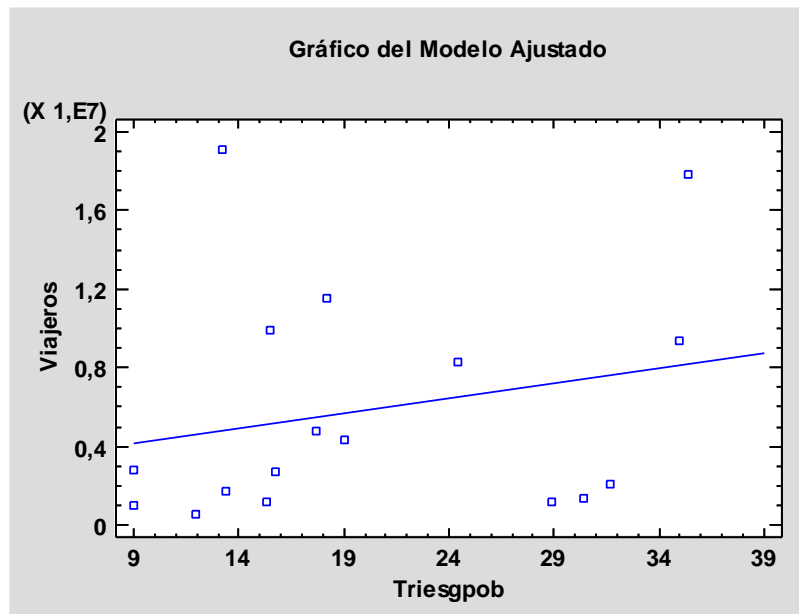
- Viajeros vs Viajeros camping:



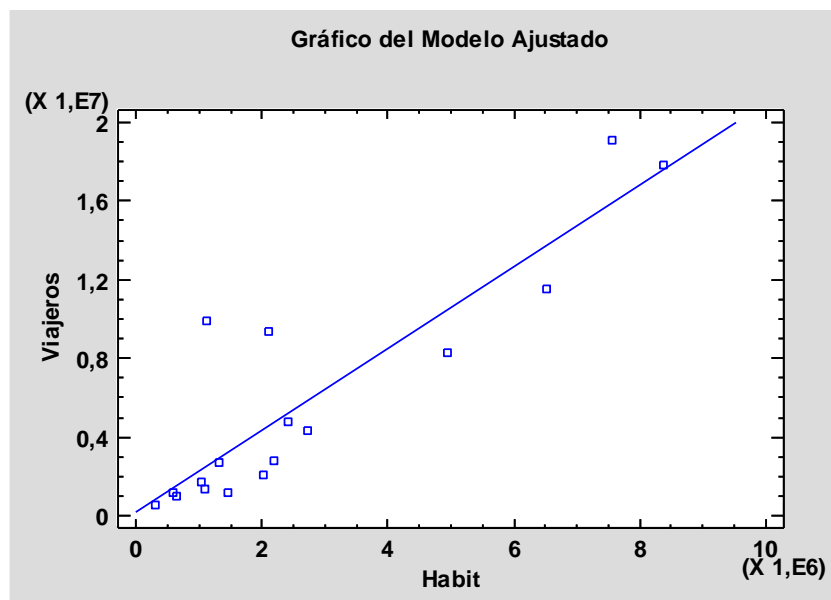
- Viajeros vs Renta media:



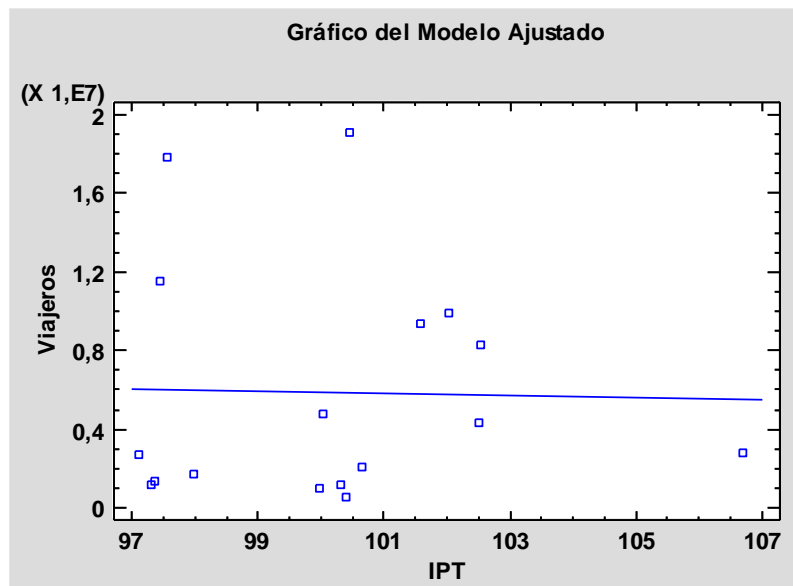
- Viajeros vs Tasa riesgo pobreza:



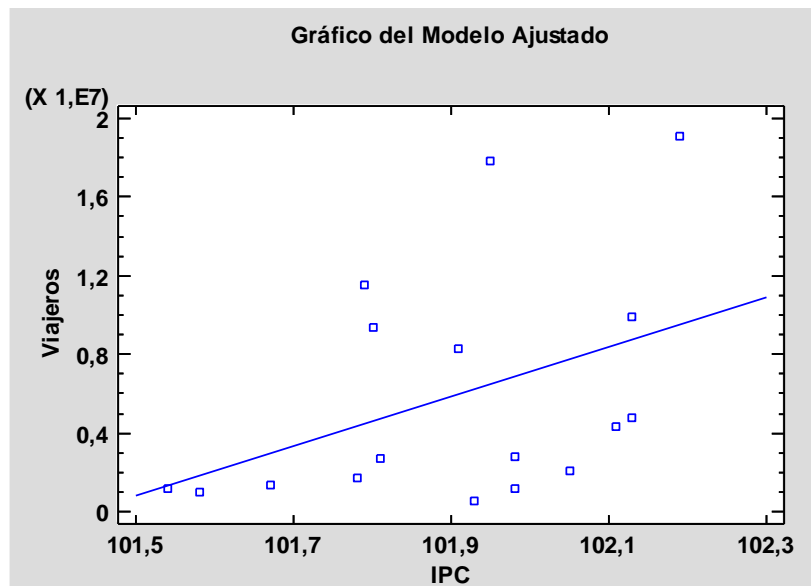
- Viajeros vs Habitantes:



- Viajeros vs IPT:

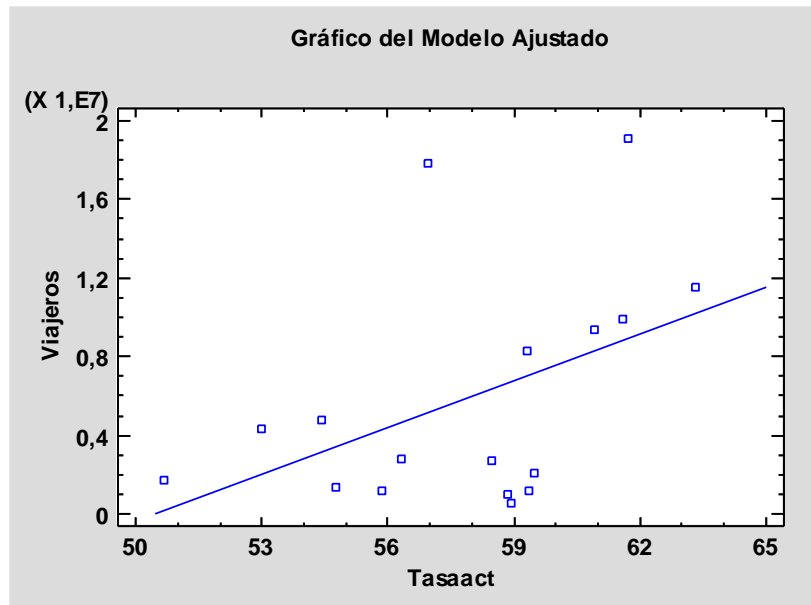


- Viajeros vs IPC:

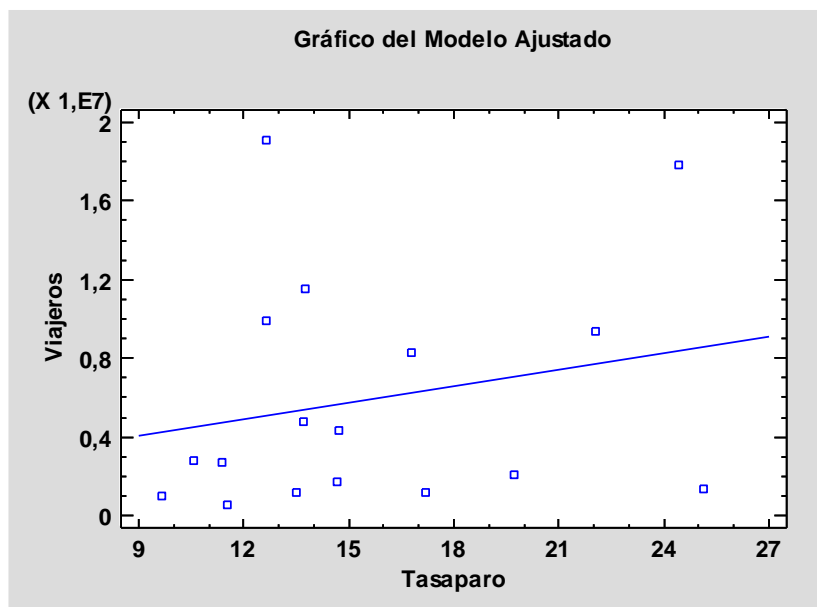




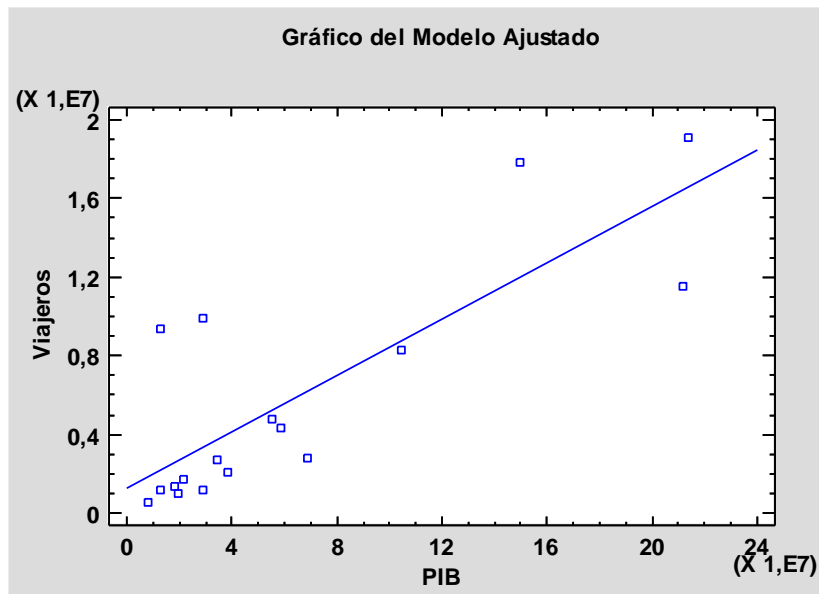
- Viajeros vs Tasa actividad:



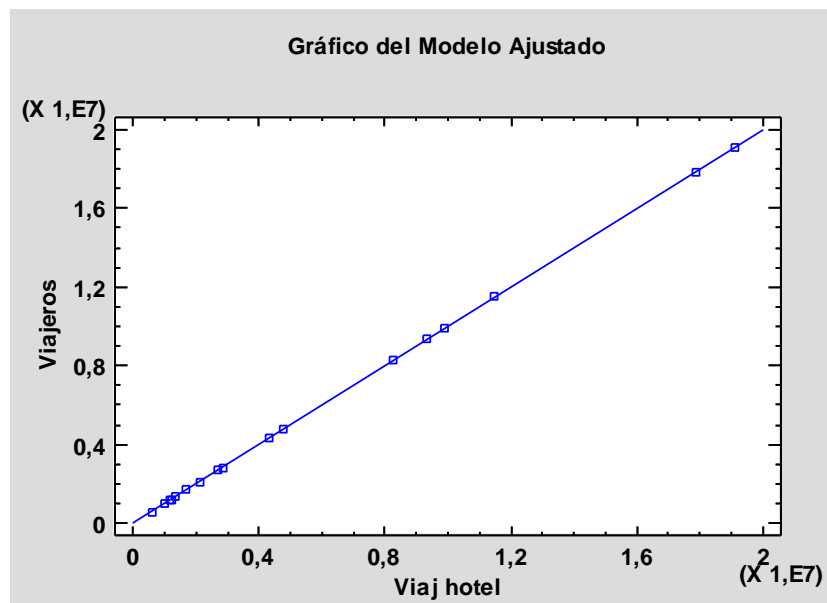
- Viajeros vs Tasa paro:



- Viajeros vs PIB:



- Viajeros vs Viajeros hotel:



- Viajeros vs Viajeros turismo rural:

