



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CAMPUS D'ALCOI

Estudio econométrico sobre los visitantes turísticos en España en el año 2014

MEMORIA PRESENTADA POR:

Asunción Biosca Barberá

GRADO DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Convocatoria de defensa: Julio 2017

ÍNDICE

1. Introducción y Objetivos	1
1.1 Motivación	2
1.2 Hipótesis.....	7
1.3 Objetivos	8
2. Metodología y Técnicas.....	9
2.1 Metodología	10
2.2 Técnicas	11
3. Cálculos, Resultados y Explotación	26
3.1 Análisis Univariante.....	28
3.2 Análisis Bivariante	44
3.3 Análisis Multivariante.....	79
3.4 Predicciones	90
4. Conclusiones y Futuras líneas de investigación	92
4.1 Conclusiones.....	93
4.2 Futuras Líneas de Investigación	94
5. Bibliografía	95

1. Introducción y Objetivos

1.1 Motivación

La realización de este Trabajo de Fin de Grado trata de recoger parte de lo aprendido durante los cursos realizados en el Grado de Administración y Dirección de empresas.

Concretamente en este caso, la rama que se ha elegido es la del Dpto. de Estadística e Investigación Operativa Aplicadas y Calidad, ya que parece interesante poder hacer este tipo de análisis para conocer y mostrar que variables son las más influyentes en el tema a analizar, que en este caso, será el Turismo.

El Turismo según la OMT (La Organización Mundial del Turismo) 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 de negocios/profesionales. Estas personas se denominan visitantes (que pueden ser **turistas o excursionistas; residentes o no residentes**) y el turismo tiene que ver con sus actividades, de las cuales algunas implican un gasto turístico”*.**[1]**

Este tema, el Turismo, se ha elegido porque es un ítem muy importante en España. Se considera una de las mayores fuentes de ingresos en el territorio español.

Gracias al territorio que posee y la ubicación en la que se encuentra, España, recibe millones de visitas de Turistas por Comunidad Autónoma. Destacar que, es uno de los principales destinos turísticos del mundo, gracias a los diversos atractivos que posee, como por ejemplo sus playas, paisajes, montañas, patrimonio histórico o clima entre otros.

Por este motivo, para hacer el análisis se elegirán distintas variables con valores numéricos que ‘a priori’ influyen a la hora de valorar la recepción de los visitantes (turistas o excursionistas/residentes o no residentes) que visitan el país Español. Se hará una división por comunidades autónomas.

A continuación, se empezará con una breve explicación de los principales términos/valores importantes en la economía. Para poder estimar la situación económica actual tanto a nivel mundial, como a nivel nacional. Así, poder mostrar que el Turismo se encuentra dentro de uno de los sectores productivos más importantes a nivel nacional.

Los Sectores Productivos

En España existen diferentes sectores productivos, principalmente hay tres. Explicaremos brevemente cada uno de ellos y destacaremos aquel donde se encuentra el ítem a analizar.

➤ ***El Sector Agrario***

Una de las actividades más importantes del sector son la agricultura y la pesca.

España tiene una gran superficie de terreno útil para poder cultivar. Alrededor de 23 millones de hectáreas, casi la mitad del territorio español.

Además de que la rodean múltiples mares y dentro del territorio español hay multitud de ríos. Por esta ubicación la pesca es uno de los empleos vigentes en nuestro país. Además, como se ha indicado anteriormente, la ubicación marítima del país hace que sea un destino más apetecible para las visitas turísticas **[2]**.

➤ **El Sector Industrial**

El sector industrial (secundario) representa actualmente alrededor de un 40% del PIB en España y genera sobre el 26,5% de empleo para la población activa.

En las últimas décadas se ha producido un fuerte crecimiento de las industrias vinculadas al I+D+i (Investigación, desarrollo e innovación), aunque respecto al resto de países en España existe poca capacidad tecnológica. Se pueden destacar diferentes sectores como el automóvil, farmacéuticos, petroquímicos e industrias ligadas a la tecnología como la informática, las telecomunicaciones, la electrónica o la biotecnología.

La recuperación industrial española se ha caracterizado por el dinamismo de las PYMES y por la tendencia a la deslocalización (traslado de la producción fuera del territorio español) y la terciarización industrial. [3]

➤ **El Sector Servicios o sector terciario**

Se define como el sector que incluye todas las actividades económicas necesarias para cubrir los servicios que demanda la población.

Dentro de los servicios se encuentran los sectores financiero, el turismo, la hostelería, el transporte, sanidad, educación y administración.

Destacar que España cuenta con productos de denominación de origen. Productos alimentarios con sello español que se venden en muchos de los casos en el extranjero. Productos como el aceite de oliva, el vino, el queso, el jamón, las naranjas, los plátanos....etc. Gran parte de los Turistas, acuden a España gracias a la oferta culinaria y la gran cantidad de productos españoles que existen y están reconocidos mundialmente. Estas visitas estarían incluidas en el sector del turismo y de la Hostelería. [4]

Como se puede observar el Turismo pertenece al sector servicios por lo que influirá en la situación económica del país. En primer lugar influye en los servicios y a continuación en la economía.

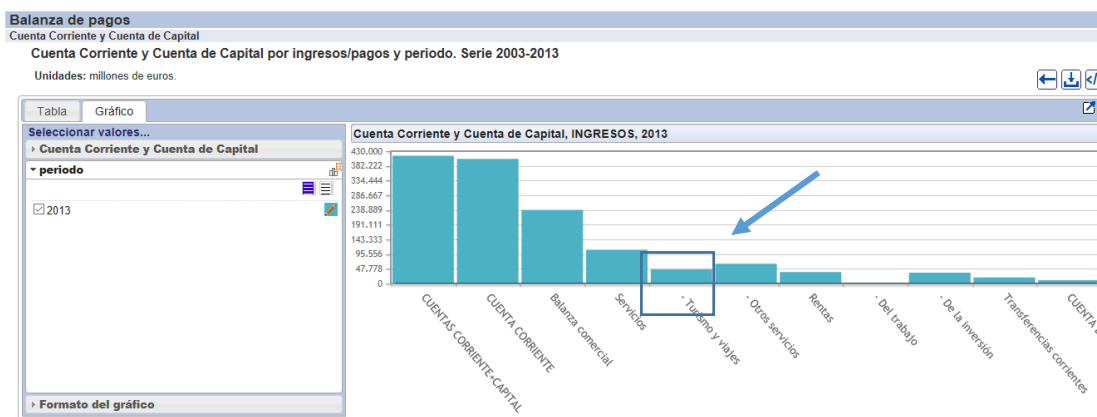


Ilustración 1.1 Balanza de Pagos 2013. Fuente INE (Instituto Nacional de Estadística)

Como muestran el gráfico de la **Ilustración 1.1**. El Turismo representa gran parte de los servicios y es la tercera variable más influyente en la cuenta corriente, después de la balanza comercial y el total de Servicios. Por lo que, como se ha indicado anteriormente los ingresos por turismo son relevantes en la economía. [5]

El Sistema Financiero

Paralelamente, se debe tener en cuenta que El sistema financiero español está formado principalmente por instituciones, instrumentos y mercados. Un resumen de la acción del sistema financiero sería la siguiente:



Ilustración 1.2 Resumen Funciones Sistema Financiero

La labor del sistema financiero es relacionar a las unidades económicas ahorradoras (familias) con las unidades económicas deficitarias. Con esto se canalizarán los ahorros de unos a otros y posibilitará la transferencia de fondos y de riesgos para que se dirijan hacia la inversión.

El objetivo principal es que se asignen de forma más eficiente los recursos para poder obtener mayores rendimientos económicos. El sistema debe de contribuir en la estabilidad monetaria y financiera. En resumen familias, empresas, particulares....prefieren:

+ LIQUIDEZ, + RENTABILIDAD Y – RIESGO

Esta labor que pretende conseguir el sistema financiero se logrará gracias a las partes que intervienen dentro de los diferentes Sectores indicados anteriormente. [6]

Como se puede observar en los gráficos que aparecen a continuación, el Turismo influye en los ingresos y El Producto Interior bruto del país (partes del Sistema Financiero). Al incluir esta variable (PIB) en el modelo, tras todos los análisis que se van a realizar, se comprobará si influyen en las entradas de Turistas y a su vez en la economía del país.

En el gráfico siguiente se muestra la diferencia que existe en el PIB de las distintas Comunidades Autónomas.

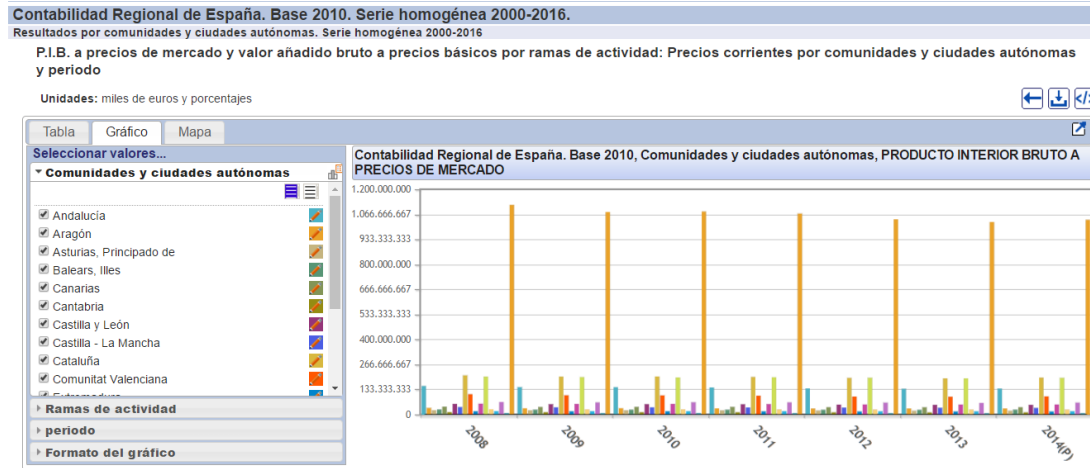


Ilustración 1.3 Producto Interior Bruto (Base 2010) Fuente INE (Instituto Nacional de Estadística)

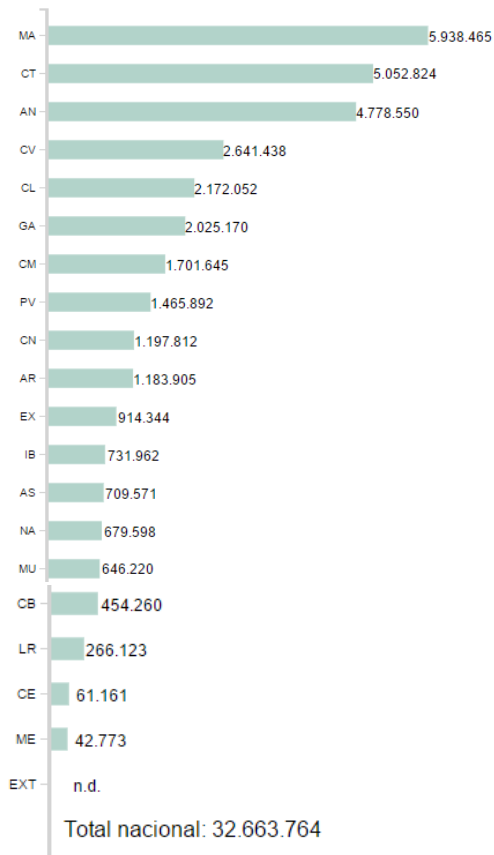
Como se puede observar en el gráfico de la **Ilustración 1.3** las Comunidades Autónomas que destacan respecto al resto teniendo en cuenta el PIB son Cataluña, Comunidad de Madrid, Andalucía y Comunidad Valenciana (ordenadas de mayor a menor influencia) [7]

1^{er} trimestre de 2015



Número de viajes por Comunidad Autónoma

Turistas residentes, T1/2015



- ▲ El número de viajes creció respecto al mismo trimestre del año anterior
- ▼ El número de viajes decreció respecto al mismo trimestre del año anterior

Ilustración 1.4 Número de Viajes por CCAA (1er Trimestre 2015), Fuente INE (Instituto Nacional de Estadística)

Como se puede observar en el gráfico de la **Ilustración 1.4** las Comunidades Autónomas que destacan respecto al número de viajes por CCAA siguen siendo las indicadas en el gráfico anterior. La única diferencia en este caso es que Madrid está por delante de Cataluña. **[8]**

1.2 Hipótesis

Explicar la variable Entradas de Turistas por Comunidad Autónoma ($TUR = Y$). La explicación se hará a través de otras variables explicativas influyentes en la variable a estudiar.

Se comprobará si la Población Total (PTOTAL), los parados (PAR), los ocupados (OCU), el Producto Interior Bruto (PIB) o el Índice de Precios al Consumo (IPC) de las distintas Comunidades Autónomas influye en las entradas de los Turistas. Si a mayor valor de estas variables las entradas de Turistas aumentan, se entiende que a mayor número de personas existentes en cada Comunidad mayor número de visitantes/turistas.

Por otra parte se comprobará también si el número de empresas, concretamente las relacionadas con el Sector Servicios (Turismo) influye de forma directa en la entrada de Turistas. Si a mayor número de empresas del sector, existen mayor número de entradas.

Además, se comprobará que tipo de gastos de turistas influyen más en la variable dependiente (entrada de Turistas).

En los siguientes puntos se mostrarán las variables dependientes (X) que se han tenido en cuenta para explicar la variable entradas de Turistas por Comunidad Autónoma (Y).

1.3 Objetivos

El trabajo consiste en conocer la influencia del Turismo en España, pero más concretamente las diferencias que existen entre las distintas Comunidades Autónomas y cuáles son las variables principales que determinan estas diferencias.

Este objetivo se conseguirá analizando diferentes variables que se consideran importantes a la hora de explicar el turismo.

Durante el desarrollo del trabajo se pretende explicar este índice a través de la recopilación de información, proveniente de fuentes fiables como puedan ser el INE o la Base de Datos del Ministerio.

Una vez se hayan recopilado todos los datos a través de Internet, y más concretamente a partir del INE, se elaborará una matriz de datos. Para nuestra matriz de datos, todos los datos son reales y lo más actuales posible (desfasados en dos años aproximadamente).

Se realizará un estudio de regresión múltiple que permita obtener un modelo más simple y apropiado para lo que se quiere analizar. Tras conseguirse el modelo se realizarán diferentes análisis, donde se predecirá la evolución e influencia del turismo.

Para conseguir los objetivos planteados se deben seguir estos pasos:

- En primer lugar, se definirán las variables explicativas que pueden influir en el nivel de turismo de cada una de las Comunidades Autónomas. A partir de las variables independientes (X) y la variables dependiente (Y), se realizará un análisis descriptivo de las variables. (Donde se especificará la α , la cual es igual al nivel de significación).
- En segundo lugar, se comprueba la relación lineal entre las variables independientes (X) y las Entradas de Turistas (TUR=Y)
- En tercer lugar, se realizará el estudio de multicolinealidad, donde se analizará la relación que existe entre las variables. En este punto se irán elaborando y analizando matrices inversas de correlaciones.
- Tras eliminar las correspondientes variables explicativas a partir del estudio de multicolinealidad, se realiza el ajuste y la validación del modelo econométrico con las variables X resultantes frente la variables dependiente (Y=TUR).
- Identificar si alguna Comunidad Autónoma se comporta distinto al resto a través del análisis de puntos anómalos.
- A continuación, se efectuaran predicciones a partir del modelo definitivo para ver cuál sería el resultado si una Comunidad Autónoma cambiara su comportamiento.
- Finalmente se elaboraran las conclusiones a través de los resultados obtenidos en el modelo y se efectuará su interpretación económica.

2. Metodología y Técnicas

2.1 Metodología

En este punto se van a mostrar los diferentes pasos que se van a seguir durante el trabajo. Se detallará de forma teórica cada uno de los pasos que se han seguido en los diferentes análisis.

A continuación, en el esquema siguiente se resumen todos los pasos efectuados:

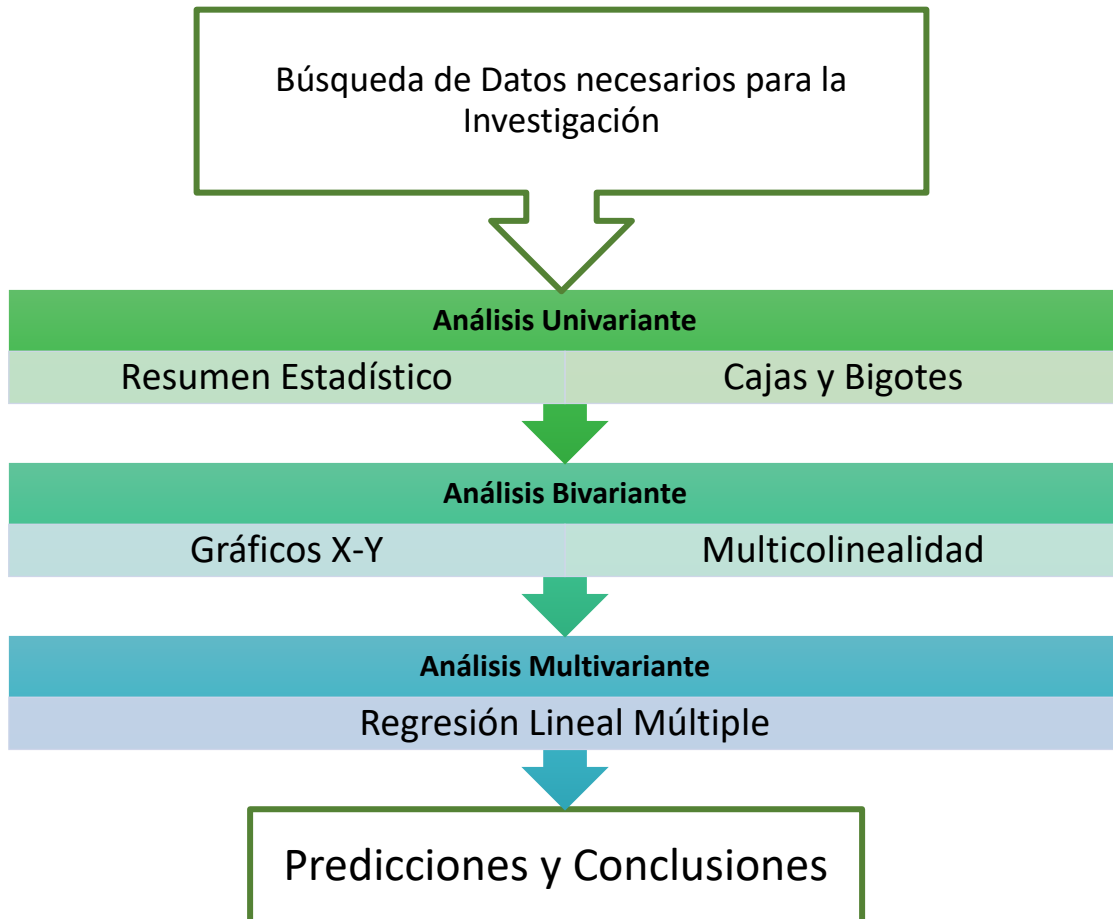


Ilustración 2.1 Resumen de Procesos

2.2 Técnicas

A continuación, se van a explicar las técnicas utilizadas para realizar el trabajo de fin de grado.

Indicar que los análisis se han desarrollado a través del programa **Statgraphics Centurion XVI.11** y del **Microsoft Excel**.

2.2.1 Análisis Univariante

Este estudio consiste en analizar el comportamiento de las variables de forma individual, tanto de la variable dependiente (Y) como de las variables independientes (X). En este análisis se incluirá el resumen de estadísticos que está compuesto por las medidas de centralización, dispersión y asimetría. Por otra parte, se analizará el correspondiente gráfico de Caja y Bigotes de cada una de las variables con los datos del año 2014 para cada Comunidad Autónoma española. (Gráfico que resume las medidas anteriormente nombradas).

2.2.1.1 Resumen de Estadísticos

El resumen de estadísticos resumirá la información más importante de la variable, midiendo valores que acumulan una determinada probabilidad. Los datos del análisis aparecerán en una tabla indicando cada uno de los resultados.

Existen dos tipos de medidas dentro del resumen. Se muestran ambas medidas y en cada una de ellas las partes que la contienen.

- Medidas de centralización o Tendencia Central.

Muestran la posición que representa la variable en el total de la muestra, es decir, da información sobre donde se encuentra la variable. Estas medidas pretenden resumir en un solo valor un conjunto de valores. Las medidas que se pueden calcular dentro de la Tendencia Central son: **[9] [10]**

- **Media:** Es una de las medidas que más se utiliza a la hora de hacer estudios o análisis estadísticos. Se trata del valor promedio de las variables, es decir, el resultado obtenido de dividir la suma total de los resultados entre el número de variables que forman la muestra analizada. Cuando existan puntos anómalos o atípicos en la muestra, el valor promedio tenderá a ser más alto o más bajo de lo que debería, a causa de, las grandes diferencias entre los datos. En este caso el valor promedio podría ser poco representativo. El valor de la media siempre se encontrará entre el valor mínimo y el máximo. Por lo que la función que expresa la media es:

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

- **Mediana:** Se trata del valor que se encuentra justo en la mitad de los valores observados, una vez se han ordenado de menor a mayor. Por lo que una el valor de la mediana, divide la distribución en dos partes iguales dejando el mismo número de valores por arriba y por abajo (50%). La mediana también se puede definir como el segundo cuartil. Como no tiene en cuenta el valor de las variables, aunque haya puntos anómalos no influye en su resultado.
 - Si n es impar: $Me = X_{\left(\frac{n+1}{2}\right)}$
 - Si n es par: $Me = \frac{x_{\frac{n}{2}} + x_{\left(\frac{n}{2}+1\right)}}{2}$

➤ Medidas de dispersión o variación.

Sirven como indicadores de la variabilidad de los datos, indican como se alejan los datos respecto a la media aritmética. A continuación, se explicarán las medidas de dispersión y los valores más utilizados que son: **[10] [11]**

- **El Rango:** Se calcula como el valor resultante de la diferencia entre el valor máximo y el mínimo de la variable X de la muestra. Es una medida muy sencilla pero no proporciona información sobre la posible desigualdad que puede existir entre los valores de la estimación realizada (puntos anómalos/valores atípicos). La denotación/función correspondiente a esta variable es la siguientes:

$$R = x_{max} - x_{min}$$

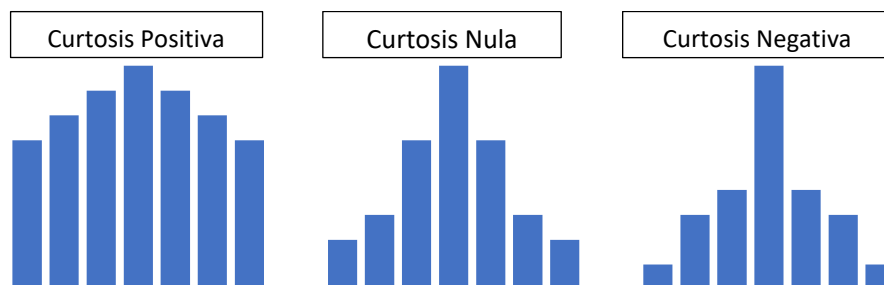
- **La Varianza:** Corresponde a la media aritmética de los cuadrados de las desviaciones con respecto a la media, es decir, es la distancia que existe entre los valores y la media. Se toman los valores al cuadrado para que las cantidades negativas, no anulen las distancias positivas. Por ello, su expresión matemática es:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

- **La desviación típica:** Es el valor resultante de calcularla raíz cuadrada de la varianza. Mide el grado de dispersión de los datos con respecto a la media. Se define como se expresa a continuación y se denota como σ para la población y como s para la muestra.

$$s = \sqrt{s^2}$$

- **Primer cuartil (1er Cuartil = C1):** se trata del punto donde quedan recogidos el 25% de los datos observados.
- **Tercer cuartil (3er Cuartil = C3):** se trata del punto donde quedan recogidos el 75% de los datos observados.
- **Mínimo:** Es el valor mínimo de la variable a estudiar.
- **Máximo:** Es el valor máximo de la variable a estudiar.
- **Asimetría:** Medida que indica si los valores de las variables se distribuyen de forma simétrica con respecto a la media aritmética (valor central). Dependiendo de la forma en la que se distribuyan los valores del análisis pueden surgir 3 tipos de curtosis según la asimetría (**Ilustración 2.2**):



A continuación, se muestra la función correspondiente a esta medida:

$$CA = \frac{\sum_{i=1}^N (X_i - \bar{X})^3}{(N - 1) * S^3}$$

- **Curtosis Tipificada:** Sirve para medir si existe la misma concentración de elementos a la izquierda y a la derecha de la media. Dependiendo de la forma en la que se distribuyan los valores del análisis pueden surgir 3 tipos de distribuciones según la asimetría. Se muestra la forma gráfica (**Ilustración 2.2.2**) y la función correspondiente: **[11][12]**

$$CC = \frac{\sum_{i=1}^N (X_i - \bar{X})^4}{(N - 1) * S^4} - 3$$

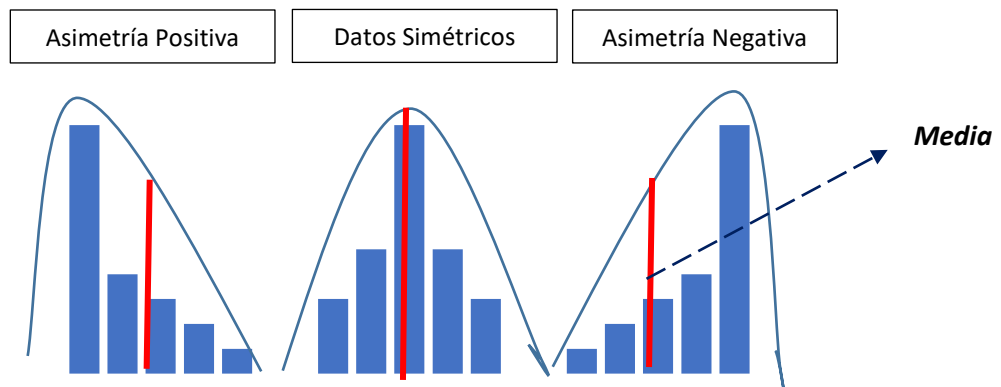


Ilustración 2.3 Tipos de Asimetría

2.2.1.2 Gráfico de Cajas y Bigotes

Tras haber explicado los estadísticos anteriores, a través del gráfico de cajas y bigotes se muestra un resumen de varios de ellos. Concretamente, mediante este gráfico, se representan los valores del 1^{er} Cuartil (C1), 2^{ndo} cuartil (C2/Mediana), 3^{er} cuartil (C3), mediana, media, máximo y mínimo. Por lo que, se puede indicar que el gráfico analizado muestra tanto la dispersión como la posición de los valores analizados. Además, muestra gráficamente los valores atípicos (en el caso de que existan).

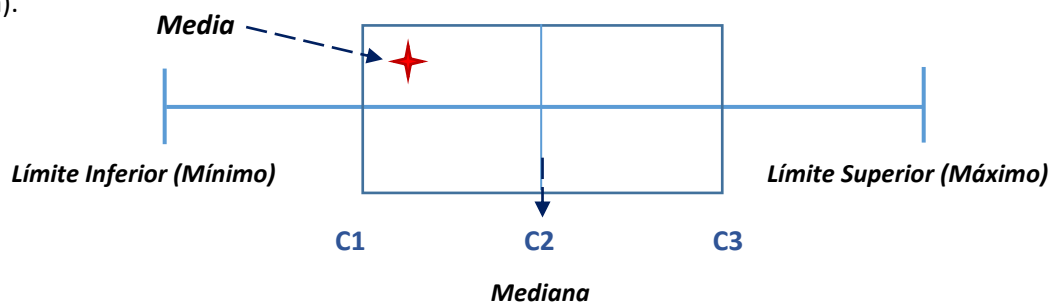


Ilustración 2.4 Gráfico de Caja y Bigotes

Tener en cuenta que en este punto, también se analizarán los datos del *Resumen Estadístico*, donde aparecen los datos concretos de cada uno de los estadísticos analizados (media, mediana, máximo, mínimo...). **[13]**

2.2.2 Análisis Bivariante

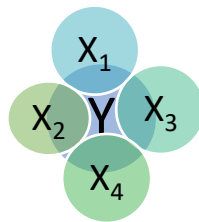
Hasta el momento, en el punto anterior se había explicado como analizar la importancia o normalidad de cada una de las variables del modelo de forma individual. Este estudio consiste en analizar las variables X e la variable Y de manera conjunta para ver la relación que existe entre dos o más variables. A continuación se muestran las pruebas a realizar en este punto:

2.2.2.1 Gráfico X-Y

Muestra la relación existente entre cada una de las variables X (de forma individual) contra la variable Y. Se trata de un gráfico bidimensional donde las variables que irán cambiando serán las del eje horizontal (variables independientes), ya que, la variable Y permanecerá constante en todos los gráficos a realizar.

2.2.2.2 Multicolinealidad

Existe multicolinealidad cuando las variables independientes (X) tienen relación entre ellas, en este análisis no se tiene en cuenta la variable dependiente (Y). Pueden existir dos tipos de relación, exacta o casi multicolinealidad. Este problema, puede provocar problemas en el ajuste del modelo, ya que, existe la misma información en más de una variable. Lo ideal sería que se cumpliera la relación que aparece en el siguiente gráfico:



Como muestra la **Ilustración 2.5** la variable Y está explicada con parte de todas las variables X. Destacar que entre las variables X no existe ningún tipo de relación, por lo que no existe ningún tipo de Multicolinealidad, ya que, no coinciden ocupando parte de la Y. La parte de la Y no explicada por las X es la perturbación (U) del modelo. Para poderlo ver de forma más clara, se muestran a continuación las relaciones que pueden existir entre 2 variables:

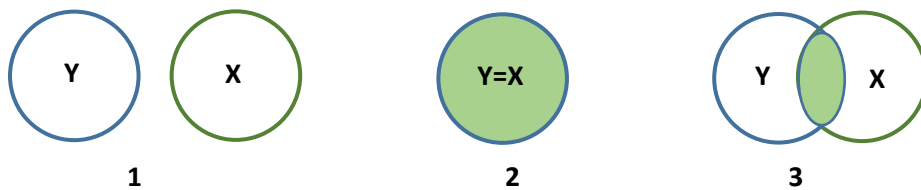


Ilustración 2.6 Relación entre variable X-Y

En la **Ilustración 2.6** se puede observar que en el caso 1 la variable X no tiene ningún tipo de relación con la Y, por lo que, está variable no servirá para explicar el modelo. En el caso 2 ocurre todo lo contrario, la X explica el total de la Y. En el caso 3, se muestra la situación más habitual donde la variable X explica parte de la variable Y. No explica el total de la variable dependiente por lo que, se necesitará analizar el resto de variables X y observar si explican otra parte del modelo o bien coinciden con la parte explicada por la variable X inicial. En el caso de que ocurra esto, existirá problema de multicolinealidad en el modelo, esta deberá de identificarse siguiendo los pasos que se indicarán a continuación. En el caso de que exista deberá eliminarse. [14] [15]

Pasos para identificar los problemas de Multicolinealidad:

➤ **Matriz de correlación (R) (Relacionar variables dependientes (X) 2 a 2)**

Se calculan las Correlaciones a través de **Statgraphics**, para poder realizar la matriz de correlación (R), la cual muestra la relación entre una variable X con otra. Por lo que en este paso, se mostrará a través del coeficiente de correlación si existe relación lineal directa entre las variables explicativas. La función que define este análisis es la siguiente:

$$R_{ij} = \frac{cov(X_i X_j)}{\sqrt{var(X_i)var(X_j)}}$$

En el resultado de la matriz de correlación (R), se puede observar una matriz donde la diagonal principal está compuesta por valores=1. El resto de los elementos son los coeficientes de correlación resultantes de la relación entre cada pareja de variables explicativas (X). Aparecerán los mismos resultados tanto en la parte inferior como a la superior, ya que, se trata de una matriz simétrica. Se tendrá en cuenta una de las dos partes, como por ejemplo la parte inferior, para poder identificar si existe algún $|R_{ij}| \geq 0,7$. En este caso, existirán problemas de Multicolinealidad y en el caso contrario (valores $< 0,7$) existirá baja correlación.

➤ **Matriz Inversa de Correlación (R⁻¹)**

A diferencia de la Matriz de Correlación (R), que solo tiene en cuenta la relación de una X contra otra, en este caso se tiene en cuenta que una variable X sea explicada por el resto de las variables. La función que define R⁻¹ es la siguiente:

$$R_{ii}^{-1} = \frac{1}{1 - R_{i-resto}^2}$$

Donde, los valores que hay que analizar son los $R_{ii}^{-1} \geq 10$ que se encuentran en la diagonal principal. Este valor indica que la X analizada tiene mucha relación con el resto de las variables, por lo que, hay problemas de multicolinealidad.

➤ **Método de Belsley, Kug y Welsch**

Este método consiste en calcular el Índice de Acondicionamiento (I.C), a partir de los autovalores de la Matriz de Correlación. Los autores Belsley, Kug y Welsch definieron el I.C como la raíz del autovalor máximo entre el autovalor mínimo:

$$I.C = \sqrt{\frac{Autovalor_{máximo}}{Autovalor_{mínimo}}}$$

Según el resultado que se obtenga, se podrá indicar si existe o no multicolinealidad y de qué tipo es. [14][15]

- Si $1 \leq I.C \leq 10$ → No existe Multicolinealidad
- Si $10 \leq I.C \leq 20$ → Existe Multicolinealidad Moderada
- Si $20 \leq I.C$ → Existe Multicolinealidad Elevada

Tras realizar los 3 pasos que se acaban de indicar, se podrá detectar si las variables están relacionadas (multicolinealidad) o no. En el caso de que existan problemas de Multicolinealidad en el modelo, para poder solucionarlo se deben eliminar variables del modelo teórico inicial planteado.

Para poder corregir el modelo inicial se deben tener en cuenta los valores de la diagonal de la Matriz Inversa de Correlación (R^{-1}). Los valores candidatos a ser eliminados del modelo son los superiores a 10. En el caso de que exista más de una variable con valor superior al indicado, se eliminará la variable que tenga el valor más elevado. Si las 2 variables con el valor más elevado son iguales o se diferencian en menos de una unidad, deberán de plantearse dos modelos de regresión múltiple. En cada una de las regresiones se eliminará una de las 2 variables (la variable con valor superior). Tras eliminar cada una de las variables en las regresiones, se tendrá en cuenta el resultado del R^2 , el modelo que obtenga un R^2 superior será el que se debe de continuar analizando. Por lo que la variable a eliminar del modelo será la que se haya incluido en la regresión múltiple con un R^2 inferior.

Este proceso se volvería a realizar hasta el punto que dejarán de aparecer los problemas de multicolinealidad. Para que el modelo quede libre de Multicolinealidad, en ninguna de las 3 pruebas debe de darse el problema. Una vez no exista el problema, se tendrá el modelo definitivo para poder realizar la Regresión Lineal Múltiple y otros pasos para que el modelo quede libre de todos los problemas posibles.

2.2.3 Análisis Multivariante

Este análisis se realiza principalmente para predecir las respuestas de las variables independientes (X) frente la variable dependiente (Y). Es decir, para saber cómo influyen sobre ella. Se pretende conseguir un modelo matemático a través de la Regresión Lineal Múltiple y otras pruebas para conocer la validez del modelo.

Para que el modelo sea válido se deben cumplir características como el nivel de significación, normalidad de los residuos, que no exista Heteroscedasticidad ni autocorrelación o que se puedan localizar puntos influyentes o anómalos. Todos estos conceptos se van a explicar a continuación.

Para localizar todos estos problemas se van a realizar contrastes de Hipótesis, donde a través del p-valor, se aceptará o rechazará la H_0 . (Teniendo en cuenta un nivel de significación $\alpha = 0,05$).

Además, para conocer la bondad del modelo se tendrá que analizar el R^2 (Coeficiente de determinación/bondad de ajuste). Siendo:

$$R^2 = \frac{\text{Suma de Cuadrados Medios (SCM)}}{\text{Suma de Cuadrados Total (SCT)}} * 100$$

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

El valor de R^2 siempre está entre 0 y 1. Cuando el valor se encuentre más cerca de 1, más bueno será el modelo.

2.2.3.1 Significación del Modelo

La primera prueba a realizar dentro del análisis Multivariante será el contraste de Significación. Tras haber eliminado las variables que generaban problemas en el apartado de Multicolinealidad, queda definido el modelo teórico propuesto. Se realiza la Regresión Lineal Múltiple para conocer la significación de las β que acompañan a las variables y del modelo. Por lo que se realizarán contrastes de Hipótesis de forma global (Test de Significación Global) y de forma individual (Test de Significación Individual).

Existen diferentes formas de identificar la significación:

- Se compara el estadístico F_{calc} frente $T_{teórico}$, análisis utilizado para el modelo.
- Se compara el estadístico T_{calc} frente $T_{teórico}$, análisis utilizado para las variables (de forma individual).
- P-valor para el modelo y para las variables. (Cuando menos sea el valor \rightarrow Mejor)

Como se ha indicado anteriormente, en este proyecto, se utiliza el P-valor. Siendo:

$$P - Value = P(F_{k,n-k-1}^{\alpha} \geq F_{calc})$$

En primer lugar, se realiza el contraste de Hipótesis del modelo y a continuación de cada una de las β . Ejemplo:

	\swarrow Modelo		\swarrow Parámetros
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$		$H_0 : \beta_i = 0$	
$H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq \beta_5 \neq 0$		$H_1 : \beta_i \neq 0$	

P-valor $> \alpha$ (0,05) \rightarrow Se acepta H_0 , por lo que el modelo/parámetros no son significativos.

2.2.3.2 Normalidad de los Residuos

En este punto, se analizarán los gráficos del papel probabilístico normal y el Histograma para poder indicar visualmente si los datos siguen una distribución normal. Para poder confirmar la normalidad de los residuos, se realiza el contraste de Hipótesis.

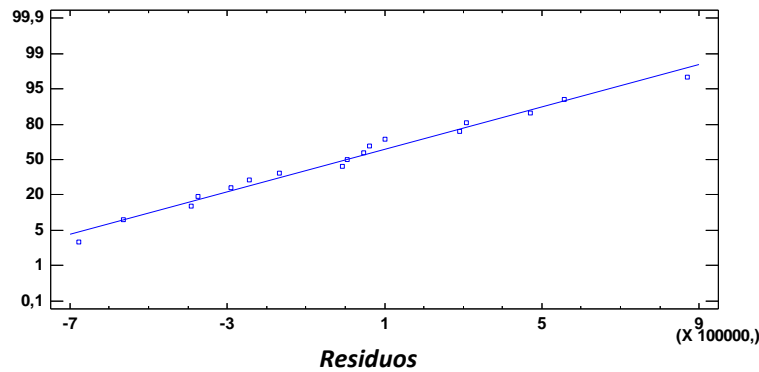


Ilustración 2.7 Papel Probabilístico Normal (Plot de Daniel)

La **Ilustración 2.7** muestra la normalidad de los residuos a través de los puntos azules que están cercanos a la diagonal. Cuando la campana de Gauss se convierte en una línea, tiene como resultado este gráfico.

En este caso los datos siguen una distribución normal, ya que, están distribuidos a través de la línea de tendencia y no existe ningún punto alejado de la línea, en la situación contraria los datos no seguirán una distribución normal. En el caso de que existan puntos alejados de la línea se definirán como puntos atípicos o anómalos que pueden influir en la distribución del resto de variables (desvían la normalidad), a causa de ser valores extremos. Se explicarán en puntos posteriores.

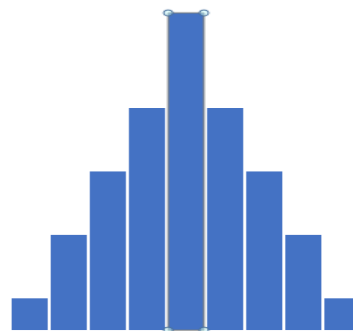


Ilustración 2.8 Histograma de residuos

En el caso de que el Histograma siga la forma de una Campana de Gauss perfecta o casi perfecta, como muestra la **Ilustración 2.8**, existirá normalidad de los residuos.

Como muchas veces, analizando un modelo no puede verse claro de forma gráfica si los residuos siguen una distribución normal, se efectuará este análisis a través del contraste de Hipótesis, teniendo en cuenta el P-valor del test de normalidad más restrictivo. **[10]**

Existen 3 pruebas de normalidad:

1. Chi-Cuadrado
2. Estadístico de Shapiro-Wilk
3. Valor Z para asimetría



Se elegirá el Test más Restrictivo

Una vez se conozca el p-valor del test más restrictivo, se realizará el siguiente contraste de Hipótesis:

H_0 : Los residuos se distribuyen normalmente.

H_1 : Los residuos no se distribuyen normalmente.

Si $p\text{-valor} \geq \alpha (0,05) \rightarrow$ Acepto H_0

En el caso de que el p-valor sea $\geq \alpha (0,05) \rightarrow$ Se acepta H_0 , por lo que se concluye que los residuos se distribuyen normalmente.

2.2.3.3 Heterocedasticidad

Este problema aparece cuando la varianza de la perturbación no es constante. Cuando ocurre esto, tendremos un error en el cálculo de: estimación de parámetros, predicciones y contraste de hipótesis.

Siempre es mejor para estimar el modelo que la varianza de U sea constante, es decir, lo que le conviene al modelo es que no existan problemas de heterocedasticidad.

¿Cuáles pueden ser las causas de la Heterocedasticidad?

1. Alguna de las variables tiene intrínseco el problema.
2. Que el problema se introduzca en los datos debido a que estos provienen de promediar diferentes muestras de datos.
3. Se ha planteado mal el modelo.
4. Existen puntos anómalos entre los datos que provocan el problema.

Se puede identificar el problema a través de dos formas:

- **Mediante análisis gráfico** → La identificación suele ser fácil debido a que el gráfico tiene una forma cónica.
- **Mediante contraste de hipótesis** → Es el análisis que se utilizará en el proyecto y el que se detalla a continuación.

Se realiza un modelo lineal (a través del Statgraphics) que relaciona los residuos² del modelo con las variables explicativas:

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

Considerando el siguiente contraste de hipótesis:

$H_0: \sigma^2$ es constante → No existe heterocedasticidad

$H_1: \sigma^2$ no es constante → Existe heterocedasticidad

En el caso de que el p-valor sea $\geq \alpha$ (0,05) → Se acepta H_0 , por lo que se concluye que no hay heterocedasticidad. Cuando alguno de los parámetros β (que no sea β_0) sea significativo, se aceptaría que existe heterocedasticidad, siendo el p-valor de alguna variable menor a $\alpha = (0,05)$.

[14] [15]

2.2.3.4 Autocorrelación

Existe el problema cuando una de las hipótesis del modelo hace referencia a que la perturbación en un momento dado no debe depender de los valores tomados anteriormente. Algunos de los problemas que puede causar la autocorrelación son las siguientes:

- Las estimaciones de los parámetros no tendrán varianza mínima.
- Los valores “t” y “F” ya no sirvan, ya que dejan de seguir la distribución correspondiente y se hacen muy grandes, por lo que, pueden aceptarse como variables explicativas aquellas que no lo son.
- Las predicciones que ofrece el modelo no son adecuadas.

¿Cuáles pueden ser las causas de la Autocorrelación?

1. Las variables estudiadas son series temporales (no es el caso de este proyecto).
2. El modelo no está bien planteado.
3. La variable a explicar se ha medido con error.

En este proyecto, para poder localizar este problema se han utilizado

➤ **Contraste de hipótesis mediante el test de Durbin Watson.**

La prueba de Durbin Watson se utiliza para detectar la presencia de autocorrelación de primer orden entre los residuales del modelo lineal de regresión múltiple. El estadístico se define de la siguiente manera:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

Siendo:

$$d = 2(1 - \rho_1)$$

Se detecta autocorrelación de 1^{er} orden a través del siguiente contraste de Hipótesis:

$H_0: \rho_1=0$ no hay autocorrelación de 1^{er} orden

$H_1: \rho_1 \neq 0$ hay autocorrelación de 1^{er} orden.

Si p-valor > α (0,05) → Acepto H_0

En el caso de aceptar la H_0 se confirma que no existe autocorrelación de 1^{er} orden.

Seguidamente, para poder confirmar que no existen problemas de autocorrelación de otro orden en el modelo, se van a realizar los gráficos de autocorrelación (FAS y FAP). [14] [15]

➤ **Los gráficos de FAS y FAP.**

Siendo FAS: Función de Autocorrelación Simple, conjunto de coeficientes de autocorrelación ρ_s que miden la relación lineal entre el residuo e_t y e_{t-s} .

Siendo FAP: Función de Autocorrelación Parcial, el cual mide lo mismo que la Fas pero haciendo una pequeña corrección eliminando relaciones entre residuos intermedios.

Se muestran ejemplos gráficos a continuación que muestran si el modelo semilogarítmico se ve afectado por la correlación:

FAS:

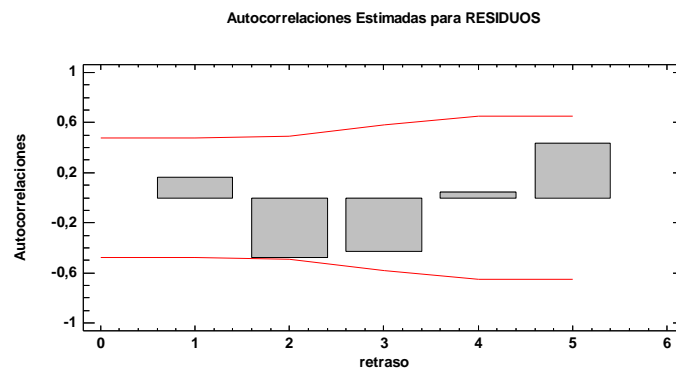


Ilustración 2.9 Función de Autocorrelación Simple

FAP:

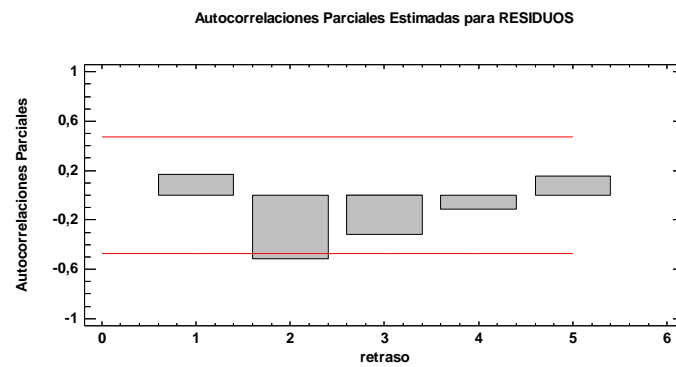


Ilustración 2.10 Función de Autocorrelación Parcial

Los gráficos sirven para poder analizar visualmente si existe autocorrelación de algún orden. Se detecta el problema si alguno de los factores de autocorrelación supera los límites (líneas en rojo).

Como se puede observar en los ejemplos, los gráficos de la **Ilustración 2.9 y 2.10** muestran autocorrelación de segundo orden ya que los factores superan el límite. Aunque en este caso, se debe indicar que no se trata de una autocorrelación fuerte. Además sería de orden negativo ya que el factor de autocorrelación supera el límite por la parte inferior. En el caso de que fuera por la parte superior se trataría de una autocorrelación positiva.

El gráfico siguiente (Ilustración 2.11) muestra el ejemplo de un modelo libre de autocorrelación:

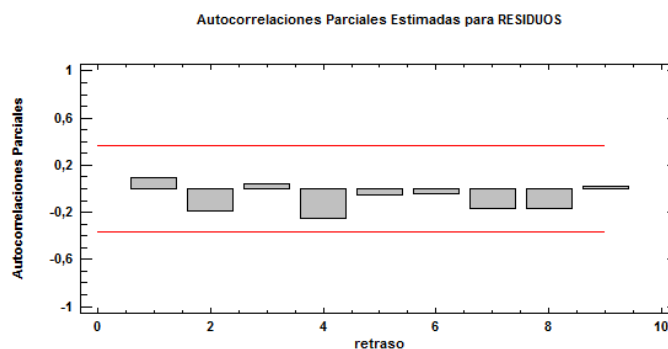


Ilustración 2.11 Función de Autocorrelación Simple

2.2.3.5 Puntos influyentes y Residuos atípicos o anómalos:

Los puntos influyentes se pueden detectar a través del gráfico de normalidad de los residuos, que es resultado de realizar una regresión simple. Cuando hacemos este análisis de regresión de los residuos, se pueden observar datos que no forman parte de la relación lineal (puntos alejados de la línea de normalidad).

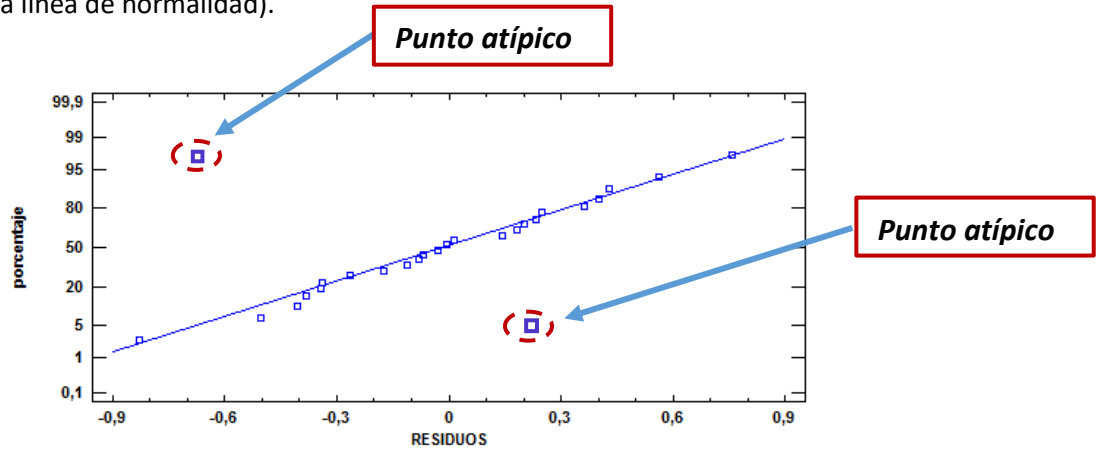


Ilustración 2.12 Representación de puntos atípicos en el gráfico de normalidad de los residuos.

Como muestra la **Ilustración 2.12** existen puntos atípicos entre los datos, los cuales pueden ser influyentes a priori o a posteriori. Cada tipo influye de forma diferente en la tendencia del modelo. Los puntos se calcularán mediante el **Statgraphics** y dependiendo de la información que proporcionen pertenecerán a un tipo o a otro.

PUNTOS INFLUYENTES A PRIORI

Serán aquellos que nos indiquen si los puntos se alejan mucho de la media de “x”. Influyen modificando el valor de los parámetros que acompañan a las variables explicativas (B_i). Se identifican cuando cumplen:

$$n_{ii} > 2\bar{n}$$

Siendo \bar{n} un valor proporcionado a través de Statgraphics siguiendo la ruta correspondiente.

Cuando exista este tipo de puntos lo que puede ocurrir es que cambie la pendiente de la recta de normalidad de los residuos.

PUNTOS INFLUYENTES A POSTERIORI

Los puntos influyentes a posteriori nos indicarán como varía la pendiente del modelo. Influyen modificando el valor de la constante (B₀). Se identifican cuando cumplen:

$$|DFITS| \geq 2\sqrt{\bar{n}}$$

Siendo $|DFITS|$ un valor proporcionado a través de Statgraphics siguiendo la ruta correspondiente.

Cuando exista este tipo de puntos lo que puede ocurrir es que se desplace hacia arriba o hacia abajo la recta de normalidad de los residuos.

Además, como los puntos atípicos o anómalos son residuos mal estimados (residuo muy grande) que influyen en la validez/bondad del modelo. Se puede calcular el valor estudentizado mediante los residuos para identificarlos según muestra la **Ilustración 2.13**:



Ilustración 2.13 Muestra los rangos de los puntos candidatos a eliminar o no del modelo

Por lo que al analizar el residuo estudentizado, se concluye si hay puntos anómalos y si son candidatos a eliminar o no del modelo.

En el caso de que existan residuos atípicos candidatos a eliminar del modelo, se tendrán que volver a realizar los análisis teniendo en cuenta el nuevo modelo propuesto (sin la variable eliminada). [14] [15]

2.2.4 Predicciones

Uno de los objetivos que se quiere lograr con el modelo definitivo es poder hacer predicciones de los valores de la variable modificando los datos que la componen, es decir, saber cómo varía el modelo si los datos cambian.

Se van a utilizar diferentes formas para hacer predicciones:

1. **Predicción Puntual:** Se modifican, aumentando o disminuyendo, los datos de una variable que esté incluida en el modelo definitivo. En el caso de este proyecto se elegirá un Comunidad Autónoma y se variarán los datos de una de las variables incluidas en el modelo final para poder observar como varia el resultado de la variable dependiente (Y).
Si la variable aparece en el modelo con signo positivo, si aumenta → aumentará el valor de Y. En el caso de que la variable aparezca en negativo cuando aumente → más pequeño será el valor de Y.
2. **Predicción por intervalos:** Se modifican datos de una de las variables definitivas, a través de Statgraphics se realiza la predicción puntual y el valor resultante se compara con el valor real de la variable para ver si se encuentra entre el límite inferior y superior. En el caso de que se encuentre dentro del límite se podrá indicar que el modelo es lo bastante “bueno” como para poder hacer predicciones en el caso de cambios en las variables.

A partir de este punto donde se indicará si el modelo sirve o no para poder hacer predicciones, se indicarán las conclusiones del modelo y se propondrán futuras líneas de investigación.

3. Cálculos, Resultados y Explotación

Los análisis que se van a realizar van a hacerse con datos de variables cuantitativas, ya que todas las variables toman valores numéricos.

A estas variables se les llama población. La población es el conjunto de elementos sobre el cual se va a realizar el análisis o el estudio. Dentro de la población están presentes los individuos que serán cada uno de los componentes (variables X) que forman la población.

Para obtener los datos se hace una muestra de la población, es decir, que el estudio se efectúa sobre un grupo/subconjunto de la población total.

Estos componentes, variables X, se han obtenido a partir de fuentes de datos fiables como:

- El INE (Instituto Nacional de Estadística) y más concretamente del apartado DIRCE (Directorio Central de Empresas: explotación estadística).
- Instituto de Estudios Turísticos

Seguidamente, se muestra un pequeño resumen explicativo sobre los dos tipos de variables que se podrían encontrar a la hora de hacer un estudio. En el caso de este estudio, como se ha indicado anteriormente las variables son cuantitativas.

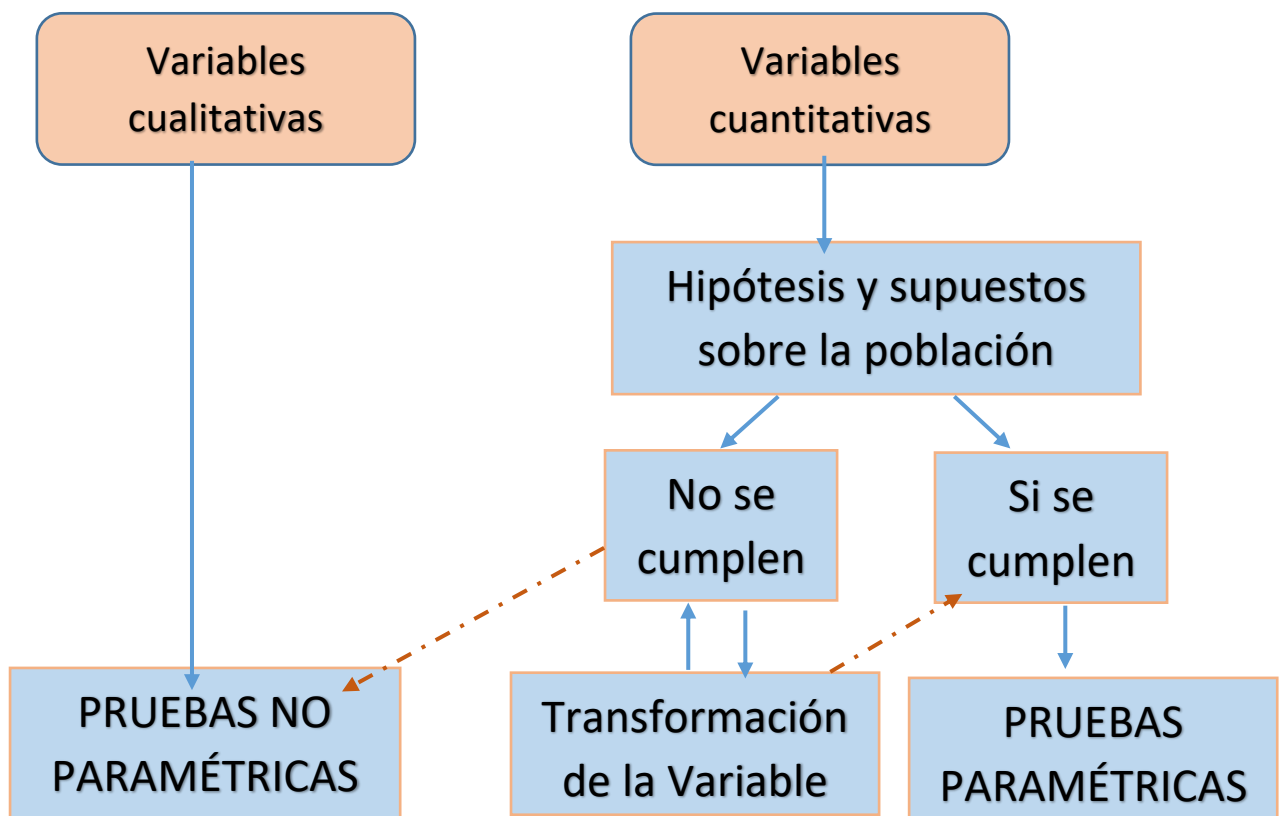


Ilustración 3.1. Elaboración propia a partir de los apuntes de la asignatura métodos Estadísticos en Economía

Según el esquema anterior, los análisis de las variables se harán mediante pruebas paramétricas. Los requisitos para este tipo de pruebas son:

- Normalidad de los datos, es decir, una distribución normal de las variables.
- Independencia de las observaciones
- Igualdad de varianzas en las distintas poblaciones estudiadas.
- Los datos estén dentro de una escala de medida métrica.

Suponemos que en el caso de los datos existentes, se cumplen los requisitos anteriores. Por lo que se realizarán los análisis correspondientes.

En el primer análisis a realizar (Univariante) se hará una definición de la variable Independiente y de las variables dependientes.

Para ello, se han elegido 13 variables explicativas que, según mi entender, están directamente relacionadas con el Turismo y pueden influir en su comportamiento. Las variables se muestran en la siguiente tabla. [16][17]

Tabla 3.1 *Tabla de Base de Datos del Ine y Dirce de las variables a analizar en el proyecto.*

CCAA	TUR	PTOTAL	IPC	NEC	NETR	NES	NEAI	PAR	OCU	PIB	GTOT	GTOM	GTOMD	EM
Andalucía	8.498.724	8.399.618	102,892	110.187	26.650	4.015	23.007	1395,7	2681,8	16.577	9.336	1.098	102	11
Aragón	326.690	1.326.403	103,009	15.741	5.561	1.030	5.108	120,7	526,5	24.713	293	898	101	9
Asturias, Principado de	234.489	1.049.875	103,284	12.865	4.366	1.371	3.023	98,4	375,2	19.727	314	1.340	98	14
Baleares, Illes	11.348.259	1.124.972	103,958	14.020	3.772	2.474	5.937	111,4	478,5	23.498	10.355	912	110	8
Canarias	11.474.965	2.126.144	102	27.334	8.314	2.242	7.930	342,2	758,7	19.238	12.443	1.084	110	10
Cantabria	369.964	585.359	104,381	6.974	2.263	738	1.716	51,2	226,8	20.237	337	912	98	9
Castilla - La Mancha	182.606	2.062.767	102,88	26.137	8.070	1.283	4.182	283,1	710,3	17.636	230	1.261	97	13
Castilla y León	995.596	2.478.079	103,953	32.353	9.406	2.293	5.359	234,3	921	21.063	702	705	90	8
Cataluña	16.793.818	7.396.991	104,616	94.982	34.463	4.820	41.562	756,5	3048,1	26.624	15.114	900	120	7
Comunitat Valenciana	6.226.081	4.939.674	103,083	64.022	14.436	1.970	18.986	569,4	1855,4	19.693	5.383	865	78	11
Extremadura	196.265	1.091.623	102,868	14.929	3.269	631	1.894	151,8	354,8	15.457	116	589	89	7
Galicia	1.048.878	2.734.656	103,774	38.031	10.570	1.624	6.237	263,8	1000,4	19.661	898	856	97	9
Madrid, Comunidad de	4.545.820	6.385.298	103,134	68.396	26.320	1.813	31.176	612,3	2789	30.755	5.477	1.205	168	7
Murcia, Región de	805.365	1.463.773	103,401	17.483	4.514	329	3.633	196,9	525,3	18.325	875	1.087	86	13
Navarra, Comunidad For	261.460	636.402	103,228	7.792	3.066	817	1.409	46,4	264,3	27.709	177	678	114	6
País Vasco	1.568.945	2.165.100	104,095	25.952	9.704	1.031	3.734	174,1	874,5	29.277	927	591	124	5

Los datos obtenidos son cifras de referencia de operaciones estadísticas del INE (Instituto Nacional de Estadística), del DIRCE (Directorio Central de Empresas) y del Instituto de Turismo de España. Datos que se transmiten a nivel nacional como datos oficiales de la población. [16][17]

3.1 Análisis Univariante

El primer análisis a realizar en el punto 3, será el análisis univariante.

$Y = \text{Turismo (TUR)}$ → Representa el total de las entradas de Turistas en cada una de las Comunidades Autónomas españolas en el año 2014. Se mide en número de personas y representan cifras oficiales registradas.

Se calcula el Gráfico de Caja y Bigotes y el resumen de estadísticos para la variable Independiente a analizar, que son, las entradas de Turistas por Comunidades Autónomas en el año 2014.

Gráfico de Caja y Bigotes

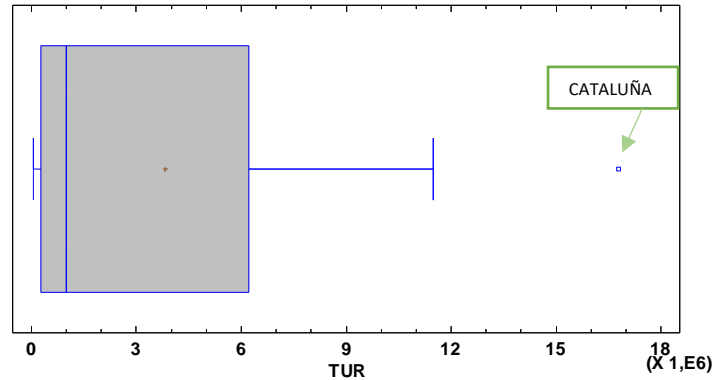


Ilustración 3.2. Gráfico de Caja y Bigotes de las entradas de turistas según Comunidad Autónoma en España en el año 2014

Resumen Estadístico para TUR:

Tabla 3.2 Resumen de estadísticos para la entrada de turistas según Comunidad Autónoma en España en el año 2014

ESTADISTICO	VALOR
Recuento	17
Promedio	$3,82 \cdot 10^6$
Mediana	995596
Varianza	$2,698 \cdot 10^{13}$
Desviación Estándar	$5,194 \cdot 10^6$
Mínimo	61020
Máximo	$1,679 \cdot 10^7$
Rango	$1,673 \cdot 10^7$
Cuartil Inferior	261460
Cuartil Superior	$6,226 \cdot 10^6$
Sesgo	1,3996
Curtosis	0,956612

Como se puede observar en la **Ilustración 3.2** hay un punto fuera del gráfico de Caja y Bigotes, este punto se llama punto anómalo o atípico. El punto indica que hay una Comunidad Autónoma que no se comporta como el resto. En este caso se trata de Cataluña y se puede indicar que destaca respecto al resto de Comunidades por el número de entradas de Turistas (elevado número de personas).

Hay una gran diferencia entre el límite inferior (0 en el gráfico) y el límite superior (11,5 en el gráfico). Esta diferencia puede ser causada porque hay grandes diferencias entre los valores de los datos analizados.

Se puede observar que la mediana (valor que se encuentra a la mitad de las observaciones analizadas) es aproximadamente 1 (valor del gráfico), aparece señalado en la línea vertical del interior del rectángulo. Si se compara con la media (valor resultante de $\frac{\Sigma}{n}$ de Todos los valores/ número de observaciones), que se encuentra desplazada a la derecha en el gráfico (Punto rojo de la **Ilustración 3.2**) y se tiene en cuenta el valor de la tabla del Resumen de Estadísticos se puede indicar que existe una asimetría derecha.

En este caso comparando la media con la desviación estándar se puede indicar que los datos están dispersos, ya que, existe poca diferencia entre ambos estadísticos.

Indicar que el valor mínimo de la observación estaría ubicado en La Rioja y el máximo en Cataluña (Punto Anómalo).

$X_1 = \text{Población Total (PTOTAL)}$ → Representa el total de la población residente en cada una de las Comunidades Autónomas españolas en el año 2014 (Enero de 2015). Se mide en número de personas y representan cifras oficiales de población resultantes de los datos indicados en el Padrón Municipal.

Se calcula el Gráfico de Caja y Bigotes y el resumen de estadísticos para la variable Dependiente a analizar, que es, la población total existente en cada Comunidad Autónoma en España en el año 2014.

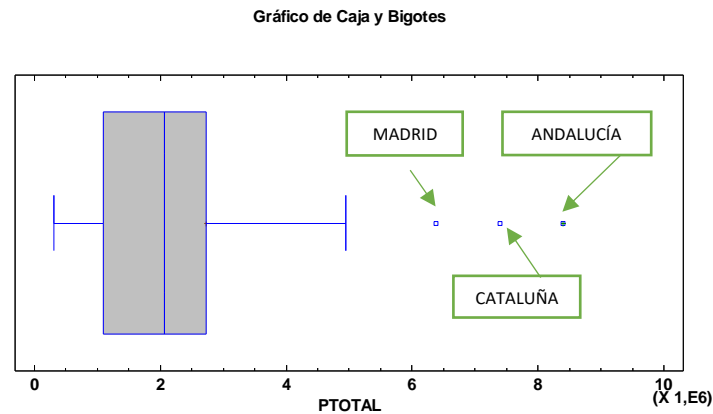


Ilustración 3.3. Gráfico de Caja y Bigotes de la población total existente en cada Comunidad Autónoma en España en el año 2014

Resumen Estadístico para PTOTAL:

Tabla 3.3. Resumen de estadísticos para la población total según Comunidad Autónoma en España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	$2,72 \cdot 10^6$
Mediana	$2,063 \cdot 10^6$
Varianza	$6,24 \cdot 10^{12}$
Desviación Estándar	$2,498 \cdot 10^6$
Mínimo	313569
Máximo	$8,4 \cdot 10^6$
Rango	$8,086 \cdot 10^6$
Cuartil Inferior	$1,091 \cdot 10^6$
Cuartil Superior	$2,735 \cdot 10^6$
Sesgo	1,33546
Curtosis	0,604737

Como se puede observar en la **Ilustración 3.3** hay tres puntos fuera del gráfico de Caja y Bigotes, estos puntos anómalos indican que hay tres Comunidades Autónomas que no se comportan como el resto. En este caso se trata de Madrid, Andalucía y Cataluña. Se puede indicar que destacan respecto al resto porque son las Comunidades que poseen mayor número de personas con respecto a la Población Total.

Se puede observar que la diferencia entre el límite inferior (0,25 en el gráfico) y el límite superior (5 en el gráfico), no es tan elevada como en el caso anterior.

Si se compara la mediana y la media se puede indicar que apenas existe diferencia. Indicar que la media se encuentra desplazada a la derecha respecto a la mediana y coincide con el tercer cuartil. Significa que existe una asimetría a la derecha.

En este caso comparando la media con la desviación estándar se puede indicar que los datos están muy dispersos, ya que, apenas existe diferencia entre ambos estadísticos.

Indicar que el valor mínimo de la observación estaría ubicado en La Rioja y el máximo en Andalucía (Punto Anómalo, junto con Madrid y Cataluña, que serían los siguientes máximos).

X₂ = Índice de Precios al Consumo (IPC) → Se expresa en valor numérico (%) y representa el índice general de las variaciones interanuales que han experimentado los precios en las distintas Comunidades Autónomas españolas en el año 2014 (último cuatrimestre). Tienen como objetivo medir la evolución del nivel de los precios de los bienes y servicios de consumo adquiridos por los hogares residentes en cada una de las regiones indicadas.

Gráfico de Caja y Bigotes

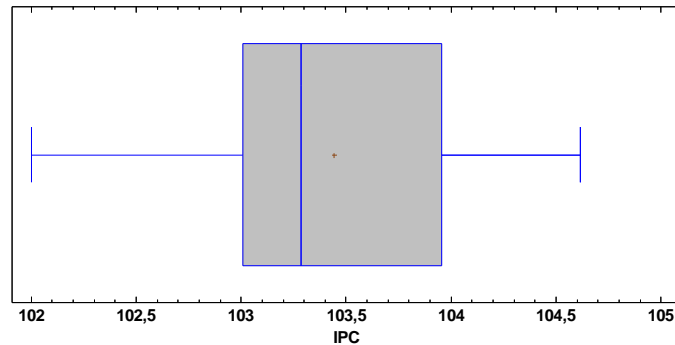


Ilustración 3.4. Gráfico de Caja y Bigotes del IPC (Índice General) de cada Comunidad Autónoma en España en el año 2014

Resumen Estadístico para IPC:

Tabla 3.4 Resumen de estadísticos para el IPC según Comunidad Autónoma en España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	103,44
Mediana	103,284
Varianza	0,441938
Desviación Estándar	0,664784
Mínimo	102
Máximo	104,616
Rango	2,616
Cuartil Inferior	103,009
Cuartil Superior	103,953
Sesgo	-0,124842
Curtosis	-0,0437481

Como se puede observar en la **Ilustración 3.4** no hay puntos fuera del gráfico de Caja y Bigotes, por lo que no hay puntos anómalos. No existen Comunidades Autónomas que destaquen del resto.

Se puede observar que el primer cuartil y el tercer cuartil se encuentran centrados entre el límite inferior y superior.

Si se compara la mediana y la media se puede indicar que apenas existe diferencia. Indicar que la media se encuentra desplazada a la derecha respecto a la mediana y coincide aproximadamente con el centro de la caja. Significa que existe una pequeña asimetría a la derecha.

En este caso comparando la media con la desviación estándar se puede indicar que los datos están poco dispersos, ya que, la desviación estándar es muy inferior a la media. Esto es debido a que la diferencia entre valores es baja.

Indicar que el valor mínimo de la observación estaría ubicado en Canarias y el máximo en Cataluña.

X_3 = Número de empresas de Comercio (NEC) → Se expresa en número de empresas y representa las empresas de comercio al por menor (excepto de vehículos de motor y motocicletas) en las distintas Comunidades Autónomas españolas en el año 2014.

Gráfico de Caja y Bigotes

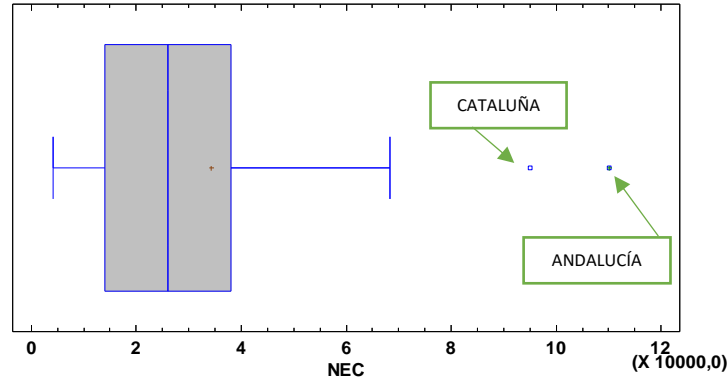


Ilustración 3.5. Gráfico de Caja y Bigotes de Comercio al por menor, excepto de vehículos de motor y motocicletas en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para NEC:

Tabla 3.5. Resumen de estadísticos para Comercio al por menor, excepto de vehículos de motor y motocicletas según Comunidad Autónoma en España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	34197,9
Mediana	25952
Varianza	9,9501*10 ⁸
Desviación Estándar	31543,7
Mínimo	4167
Máximo	110187
Rango	106020
Cuartil Inferior	14020
Cuartil Superior	38031
Sesgo	1,41464
Curtosis	1,10028

Como se puede observar en la **Ilustración 3.5** hay dos puntos anómalos fuera del gráfico de Caja y Bigotes. Las Comunidades Autónomas que destacan del resto son Cataluña y Andalucía ya que poseen un número de empresas del sector muy superior.

Se puede observar que el primer cuartil y el tercer cuartil se encuentran más cercanos al límite inferior. Esto se debe a que en la mayoría de Comunidades el número de empresas de este sector están más cerca del valor mínimo, que son 4167 empresas.

Si se compara la mediana y la media se puede indicar la media se encuentra desplazada a la derecha. Significa que existe una pequeña asimetría a la derecha. Esto es debido a que hay una gran diferencia entre el valor mínimo (4167 empresas en La Rioja) y el máximo (110.187 empresas en Andalucía).

En este caso comparando la media con la desviación estándar se puede indicar que los datos están muy dispersos, ya que, apenas existe diferencia entre ambos estadísticos.

X₄ = Número de empresas de Transporte (NETR) → Se expresa en número de empresas y representa las empresas de Transporte terrestre y por tubería, Transporte marítimo y por vías navegables interiores y Transporte aéreo en las distintas Comunidades Autónomas españolas en el año 2014.

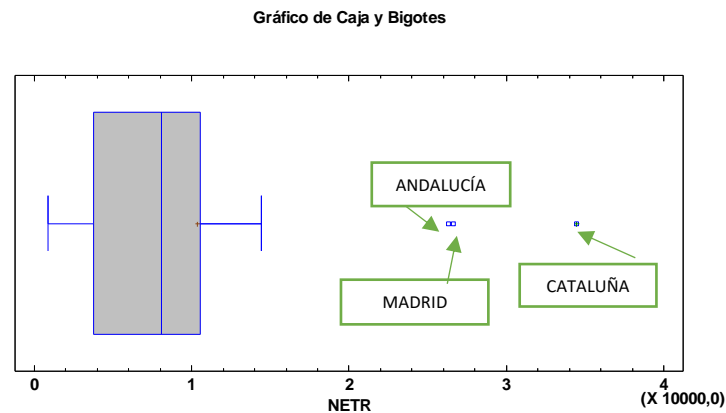


Ilustración 3.6 Gráfico de Caja y Bigotes de Transporte terrestre y por tubería; Transporte marítimo y por vías navegables interiores y Transporte aéreo en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para NETR:

Tabla 3.6 Resumen de estadísticos de Transporte terrestre y por tubería; Transporte marítimo y por vías navegables interiores; Transporte aéreos en cada Comunidad Autónoma en España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	10331,2
Mediana	8070,0
Varianza	9,524*10 ⁷
Desviación Estándar	9758,95
Mínimo	886,0
Máximo	34463,0
Rango	33577,0
Cuartil Inferior	3772,0
Cuartil Superior	10570,0
Sesgo	1,49615
Curtosis	1,34293

Como se puede observar en la **Ilustración 3.6** hay tres puntos anómalos fuera del gráfico de Caja y Bigotes. Las Comunidades Autónomas que destacan del resto son Cataluña, Andalucía y Madrid ya que poseen un número de empresas del sector superior al resto de Comunidades.

Se puede observar que el primer cuartil y el tercer cuartil se encuentran centrados entre el límite inferior y superior. Indicar que la mediana está más cerca del valor máximo y la media coincide con el tercer cuartil.

Respecto al gráfico indicar que la media se encuentra desplazada a la derecha respecto a la mediana, esto significa que existe una asimetría a la derecha. El mínimo de empresas del Sector se encuentra en La Rioja (886 empresas) y el máximo en Cataluña (34.463 empresas).

En este caso comparando la media con la desviación estándar se puede indicar que los datos están muy dispersos, ya que, apenas existe diferencia entre ambos estadísticos.

X_5 = Número de empresas de Servicio (NES) → Se expresa en número de empresas y representa las empresas de Servicios de alojamiento en las distintas Comunidades Autónomas españolas en el año 2014.

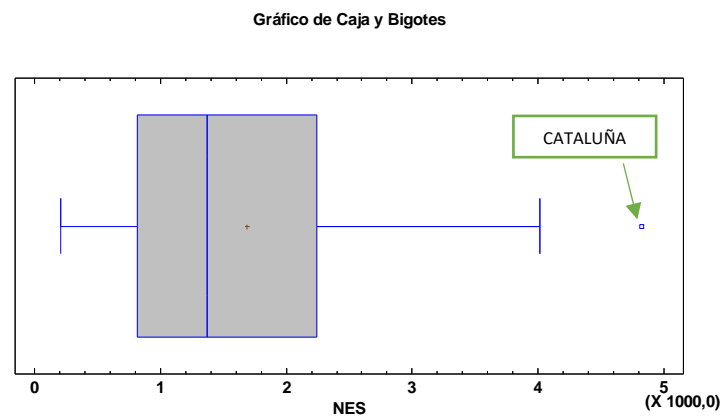


Ilustración 3.7. Gráfico de Caja y Bigotes de Servicios de alojamiento en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para NES:

Tabla 3.7 Resumen de estadísticos de Servicios de alojamiento en cada Comunidad Autónoma en España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	1687,65
Mediana	1371,0
Varianza	1,535*10 ⁶
Desviación Estándar	1238,85
Mínimo	209,0
Máximo	4820,0
Rango	4611,0
Cuartil Inferior	817,0
Cuartil Superior	2242,0
Sesgo	1,30911
Curtosis	1,65743

Como se puede observar en la **Ilustración 3.7** hay un punto anómalo fuera del gráfico de Caja y Bigotes. La Comunidad Autónoma que destaca del resto es Cataluña ya que posee mayor número de empresas de Servicios de Alojamiento.

Se puede observar que el primer cuartil y el tercer cuartil se encuentran más cercanos al límite inferior. Esto se debe a que en la mayoría de Comunidades el número de empresas de este sector están más cerca del valor mínimo, que son 209 empresas.

Respecto al gráfico y la tabla de Resumen de Estadísticos indicar que la media se encuentra desplazada a la derecha respecto a la mediana, esto significa que existe una asimetría a la derecha.

En este caso comparando la media con la desviación estándar se puede indicar que los datos están muy dispersos, ya que, apenas existe diferencia entre ambos estadísticos.

El mínimo de empresas del Sector se encuentra en La Rioja (209 empresas) y el máximo en Cataluña (4.820 empresas).

X₆ = Número de empresas de Actividades Inmobiliarias (NEAI) → Se expresa en número de empresas y representa las empresas de Actividades inmobiliarias, Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con los mismos en las distintas Comunidades Autónomas españolas en el año 2014.

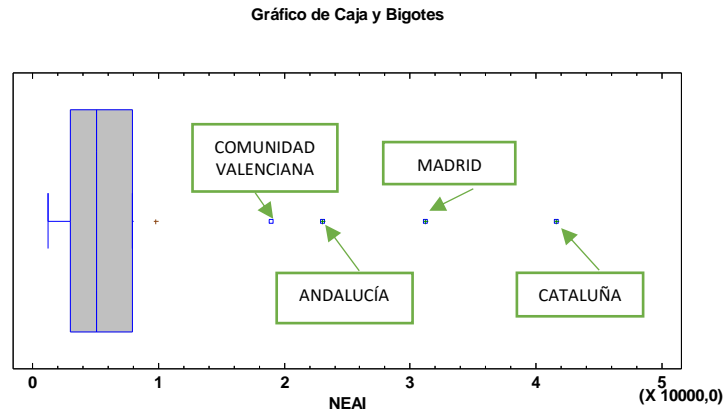


Ilustración 3.8. Gráfico de Caja y Bigotes de Actividades inmobiliarias y Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con los mismos en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para NEAI:

Tabla 3.8 Resumen de estadísticos de Actividades inmobiliarias y Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con los mismos en cada Comunidad Autónoma de España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	9774,18
Mediana	5108
Varianza	1,388*10 ⁸
Desviación Estándar	11781,3
Mínimo	1268
Máximo	41562
Rango	40294
Cuartil Inferior	3023
Cuartil Superior	7930
Sesgo	1,8011
Curtosis	2,45972

Como se puede observar en la **Ilustración 3.8** hay varios puntos anómalos fuera del gráfico de Caja y Bigotes. Las Comunidades Autónomas que destacan del resto son Comunidad Valenciana, Andalucía, Madrid y Cataluña ya que posee un número de empresas de Servicios de Actividades Inmobiliarias mucho mayor que el resto de las Comunidades.

Respecto al gráfico y la tabla de Resumen de Estadísticos, indicar que la media se encuentra muy desplazada a la derecha respecto a la mediana, esto significa que existe una asimetría a la derecha. En el gráfico no aparece un límite superior, ya que hay valores muy dispersos.

El mínimo de empresas del Sector se encuentra en La Rioja (1.268 empresas) y el máximo en Cataluña (41.562 empresas).

En este caso comparando la media con la desviación estándar se puede indicar que los datos están dispersos, ya que, apenas existe diferencia entre ambos estadísticos.

$X_7 = \text{Parados (PAR)}$ → Se expresa en miles de personas y representa a los parados que hay en cada Comunidad Autónoma. Se consideran “parados” aquellas personas que están desempleadas, desocupadas o en paro dentro del mercado laboral. Se determinan a través de los datos recogidos en la EPA (Encuesta de Población Activa).

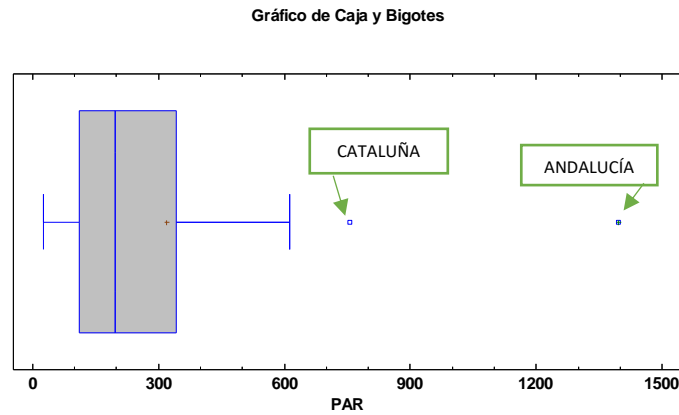


Ilustración 3.9 Gráfico de Caja y Bigotes de Parados (miles de personas) en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para PAR:

Tabla 3.9 Resumen de estadísticos de Parados (miles de personas) en cada Comunidad Autónoma de España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	319,682
Mediana	196,9
Varianza	121128
Desviación Estándar	348,034
Mínimo	26,4
Máximo	1395,7
Rango	1369,3
Cuartil Inferior	111,4
Cuartil Superior	342,2
Sesgo	2,12814
Curtosis	5,11808

Como se puede observar en la **Ilustración 3.9** hay dos puntos anómalos fuera del gráfico de Caja y Bigotes. Las Comunidades Autónomas que destacan del resto son Cataluña y Andalucía ya que poseen mayor número de personas paradas.

Respecto al gráfico y la tabla de Resumen de Estadísticos, indicar que la media se encuentra desplazada a la derecha respecto a la mediana, esto significa que existe una asimetría a la derecha.

El mínimo de personas paradas se encuentra en La Rioja y el máximo en Andalucía.

En este caso comparando la media con la desviación estándar se puede indicar que los datos están muy dispersos, ya que, apenas existe diferencia entre ambos estadísticos.

$X_8 = \text{Ocupados (OCU)}$ → Se expresa en miles de personas y representa a los ocupados que hay en cada Comunidad Autónoma. Se consideran “ocupados” aquellas personas que están empleadas, ocupadas o desempeñando algún tipo de actividad remunerada, dentro del mercado laboral. Se determinan a través de los datos recogidos en la EPA (Encuesta de Población Activa).

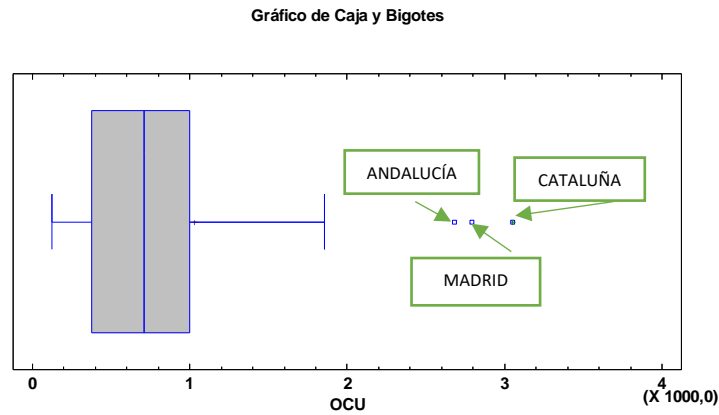


Ilustración 3.10 Gráfico de Caja y Bigotes de Ocupados (miles de personas) en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para OCU:

Tabla 3.10 Resumen de estadísticos de Ocupados (miles de personas) en cada Comunidad Autónoma de España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	1030,48
Mediana	710,3
Varianza	908266,
Desviación Estándar	953,03
Mínimo	127,5
Máximo	3048,1
Rango	2920,6
Cuartil Inferior	375,2
Cuartil Superior	1000,4
Sesgo	1,30419
Curtosis	0,32728

Como se puede observar en la **Ilustración 3.10** hay varios puntos anómalos fuera del gráfico de Caja y Bigotes. Las Comunidades Autónomas que destacan del resto son Cataluña, Andalucía y Madrid ya que poseen mayor número de personas ocupadas.

Respecto al gráfico y la tabla de Resumen de Estadísticos, indicar que la media se encuentra desplazada a la derecha respecto a la mediana, esto significa que existe una asimetría a la derecha.

El mínimo de personas ocupadas se encuentra en La Rioja y el máximo en Cataluña.

En este caso comparando la media con la desviación estándar se puede indicar que los datos están dispersos, ya que, existe poca diferencia entre ambos estadísticos.

X₉ = Producto Interior Bruto (PIB) → Se expresa en euros (€) y representa el PIB per cápita en las distintas Comunidades Autónomas. El PIB per cápita puede definirse como el resultado de dividir el valor de los bienes y servicios generados en un periodo de tiempo, en este caso un año, entre la población total de cada una de las Comunidades.

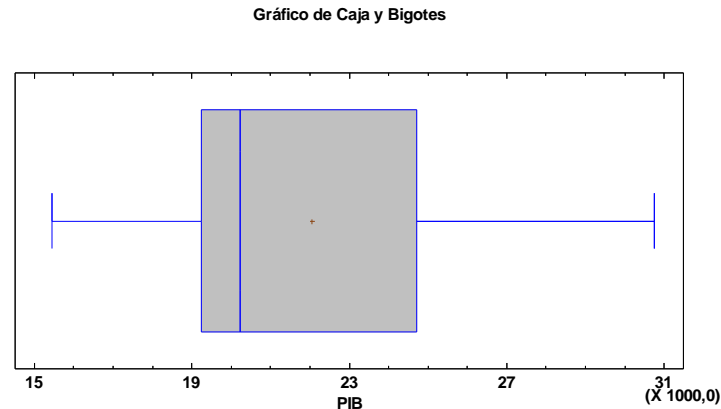


Ilustración 3.11 Gráfico de Caja y Bigotes del Crecimiento del PIB per cápita (euros) en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para PIB:

Tabla 3.11 Resumen de estadísticos del Crecimiento del PIB per cápita (euros) en cada Comunidad Autónoma de España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	22046,5
Mediana	20237,0
Varianza	2,07*10 ⁷
Desviación Estándar	4556,97
Mínimo	15457,0
Máximo	30755,0
Rango	15298,0
Cuartil Inferior	19238,0
Cuartil Superior	24713,0
Sesgo	0,523504
Curtosis	-0,792182

Como se puede observar en la **Ilustración 3.11** no hay puntos anómalos fuera del gráfico de Caja y Bigotes. Esto significa que no hay ninguna Comunidad Autónoma que destaque respecto al resto, teniendo en cuenta los valores del Producto Interior Bruto.

Se puede observar que la mediana se encuentra más cerca del primer cuartil. Destacar que la caja (cuartil 1 y 3) se encuentra centrada entre el límite inferior y superior.

Respecto al gráfico y la tabla de Resumen de Estadísticos, indicar que la media se encuentra desplazada a la derecha respecto a la mediana, esto significa que existe una asimetría a la derecha.

El mínimo de Crecimiento del PIB per cápita (euros) se encuentra en Extremadura y el máximo en la Comunidad de Madrid.

En este caso comparando la media con la desviación estándar se puede indicar que los datos están poco dispersos, ya que, existe diferencia entre ambos estadísticos.

X_{10} = **Gasto Total de los Turistas (GTOT)** → Se expresa en millones de euros (€) y representa el Gasto Total de los Turistas según destino principal en las distintas Comunidades Autónomas. Puede definirse como el gasto total realizado por los turistas internacionales que visitaron España en el año 2014. Los datos están distribuidos según Comunidades Autónomas.

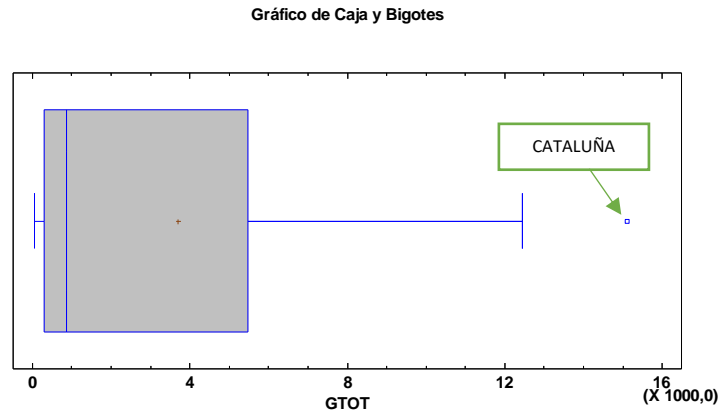


Ilustración 3.12. Gráfico de Caja y Bigotes del Gasto Total (millones de euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para GTOT:

Tabla 3.12 Resumen de estadísticos del Gasto Total (millones de euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	3707,94
Mediana	875
Varianza	$2,5388 \cdot 10^7$
Desviación Estándar	5038,7
Mínimo	58
Máximo	15114
Rango	15056,0
Cuartil Inferior	293
Cuartil Superior	5477
Sesgo	1,24071
Curtosis	0,162226

Como se puede observar en la **Ilustración 3.12** hay un punto anómalo fuera del gráfico de Caja y Bigotes. En este caso es Cataluña y significa que esta Comunidad Autónoma destaca respecto al resto, teniendo en cuenta los valores del Gasto Total de los Turistas según destino principal.

Respecto al gráfico y la tabla de Resumen de Estadísticos, indicar que la media se encuentra desplazada a la derecha respecto a la mediana, esto significa que existe una asimetría a la derecha.

El mínimo de Crecimiento del Gasto Total de los Turistas (millones de euros) se encuentra en La Rioja y el máximo en Cataluña (el cual coincide con el punto anómalo).

En este caso comparando la media con la desviación estándar se puede indicar que los datos están dispersos, ya que, existe poca diferencia entre ambos estadísticos.

X_{11} = Gasto Total Medio de los Turistas (GTOM) → Se expresa en euros (€) y representa el Gasto Total Medio de los Turistas según destino principal en las distintas Comunidades Autónomas. Puede definirse como el gasto total medio realizado por los turistas internacionales que visitaron España en el año 2014. Los datos están distribuidos según Comunidades Autónomas.

Gráfico de Caja y Bigotes

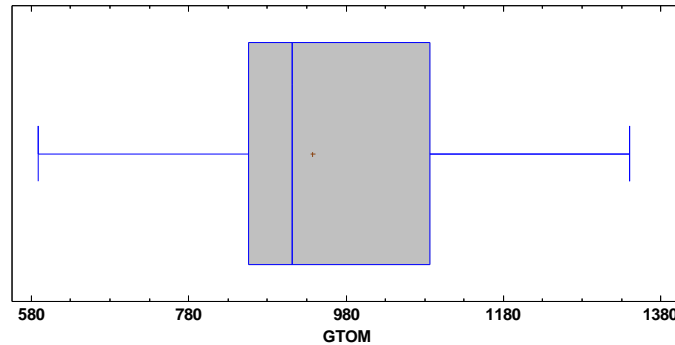


Ilustración 3.13. Gráfico de Caja y Bigotes del Gasto Total Medio (euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para GTOM:

Tabla 3.13 Resumen de estadísticos del Gasto Total Medio (euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	937,353
Mediana	912,0
Varianza	49196,0
Desviación Estándar	221,802
Mínimo	589,0
Máximo	1340,0
Rango	751,0
Cuartil Inferior	856,0
Cuartil Superior	1087,0
Sesgo	0,0917581
Curtosis	-0,62237

Como se puede observar en la **Ilustración 3.13** no hay puntos anómalos fuera del gráfico de Caja y Bigotes. Esto significa que no hay ninguna Comunidad Autónoma que destaque respecto al resto, los valores son similares teniendo en cuenta los valores del Gasto Total Medio de los Turistas según destino principal.

Respecto al gráfico y la tabla de Resumen de Estadísticos, indicar que la media se encuentra casi a la par con respecto a la mediana, se podría indicar que existe una simetría.

El mínimo de Crecimiento del Gasto Total Medio de los Turistas (euros) se encuentra en Extremadura y el máximo en el Principado de Asturias.

En este caso comparando la media con la desviación estándar se puede indicar que los datos están poco dispersos, ya que, existe bastante diferencia entre ambos estadísticos.

X_{12} = **Gasto Total Medio Diario de los Turistas (GTOMD)** → Se expresa en euros (€) y representa el Gasto Total Medio Diario de los Turistas según destino principal en las distintas Comunidades Autónomas. Puede definirse como el gasto total medio diario realizado por los turistas internacionales que visitaron España en el año 2014. Los datos están distribuidos según Comunidades Autónomas.

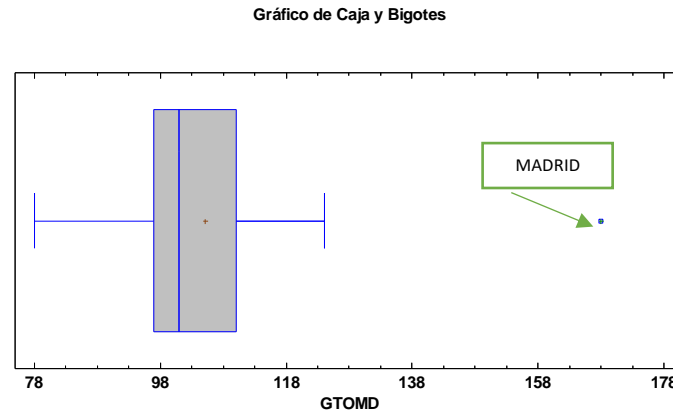


Ilustración 3.14. Gráfico de Caja y Bigotes del Gasto Total Medio Diario (euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para GTOMD:

Tabla 3.14 Resumen de estadísticos del Gasto Total Medio Diario (euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	105,059
Mediana	101,0
Varianza	406,809
Desviación Estándar	20,1695
Mínimo	78,0
Máximo	168,0
Rango	90,0
Cuartil Inferior	97,0
Cuartil Superior	110,0
Sesgo	1,93631
Curtosis	5,52194

Como se puede observar en la **Ilustración 3.14** hay un punto anómalo fuera del gráfico de Caja y Bigotes. En este caso es la Comunidad de Madrid y esto significa que destaca respecto al resto, teniendo en cuenta los valores del Gasto Total Medio Diario de los Turistas según destino principal.

Respecto al gráfico y la tabla de Resumen de Estadísticos, indicar que la media se encuentra cerca de la mediana, aunque se puede ubicar a la derecha y esto indica que existe una asimetría derecha.

El mínimo de Crecimiento del Gasto Total Medio Diario de los Turistas (euros) se encuentra en La Comunidad Valenciana y el máximo en la Comunidad de Madrid (el cual coincide con el punto anómalo).

En este caso comparando la media con la desviación estándar se puede indicar que los datos están poco dispersos, ya que, existe bastante diferencia entre ambos estadísticos.

X_{13} = **Estancia Media (EM)** → Se expresa en noches y representa la estancia media de los Turistas según destino principal en las distintas Comunidades Autónomas. Puede definirse como el número medio de noches realizado por los turistas internacionales que visitaron España en el año 2014. Los datos están distribuidos según comunidades autónomas.

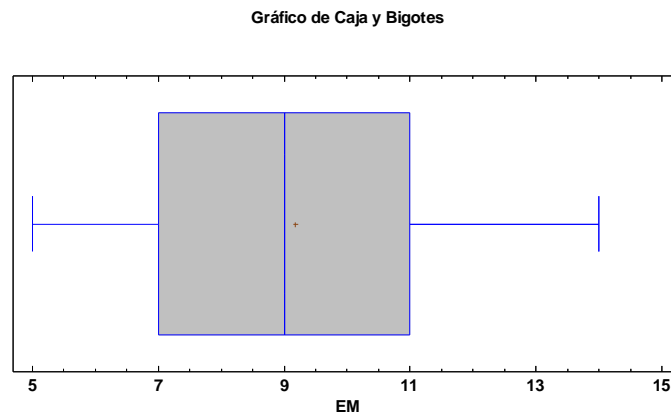


Ilustración 3.15 Gráfico de Caja y Bigotes de la Estancia Media (noches) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014

Resumen Estadístico para EM:

Tabla 3.15 Resumen de estadísticos de la Estancia Media (noches) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014

ESTADÍSTICO	VALOR
Recuento	17
Promedio	9,17647
Mediana	9
Varianza	6,52941
Desviación Estándar	2,55527
Mínimo	5
Máximo	14
Rango	9
Cuartil Inferior	7
Cuartil Superior	11
Sesgo	0,414353
Curtosis	-0,520041

Como se puede observar en la **Ilustración 3.15** no hay puntos anómalos fuera del gráfico de Caja y Bigotes. Esto significa que no hay ninguna Comunidad Autónoma que destaque respecto al resto, los valores son similares teniendo en cuenta los valores de la Estancia Media (noches) de los turistas según destino principal por Comunidades de España en el año 2014.

Respecto al gráfico y la tabla de Resumen de Estadísticos, indicar que la media se encuentra a la par de la mediana, por lo que existe una simetría.

El mínimo de Estancia Media en noches de los Turistas se encuentra en El País Vasco y el máximo en el Principado de Asturias.

En este caso comparando la media con la desviación estándar se puede indicar que los datos están poco dispersos, ya que, existe bastante diferencia entre ambos estadísticos.

3.2 Análisis Bivariante

A continuación, el análisis a realizar en el punto 3, será el análisis bivariante. Se harán los siguientes análisis para extraer datos de:

- Gráfico X-Y (La variable Y comparada con cada variable X)
- Modelo, se trata de la ecuación matemática.
- Coeficiente de correlación (R)

Todos estos pasos realizados en el año 2014 para cada una de las Comunidades Autónomas Españolas. Que son Andalucía, Aragón, Principado de Asturias, Illes Balears, Canarias, Cantabria, Castilla La Mancha, Castilla y León, Cataluña, Comunidad Valenciana, Extremadura, Galicia, Comunidad de Madrid, Región de Murcia, Navarra, País Vasco y La Rioja.

Entradas de turistas según Comunidad autónoma VS Población Total (TUR VS PTOTAL)

$$TUR = 347390 + 1,27556 * PTOTAL$$

Coeficiente de Correlación = **0,613**

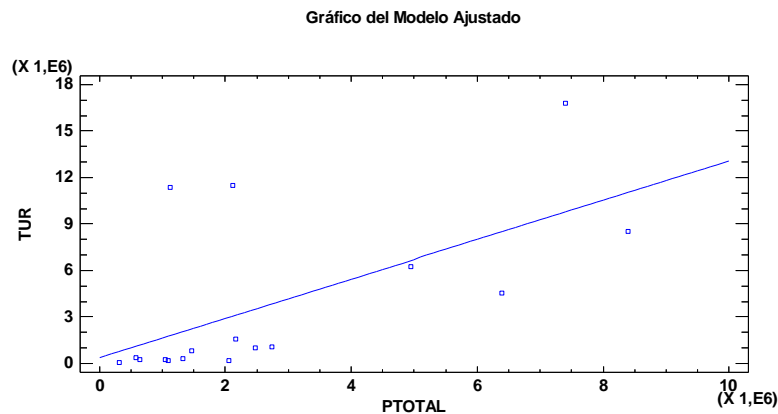


Ilustración 3.16 Gráfico X-Y para la PTOTAL

Según muestra el gráfico (**Ilustración 3.16**) y el coeficiente de correlación, se puede indicar que existe una relación directa moderada entre las entradas de Turistas y la población Total.

Además, indicar que la pendiente es positiva.

Entradas de turistas según Comunidad autónoma VS Índice de Precios al Consumo General (TUR VS IPC)

$$TUR = -3,42962 \cdot 10^7 + 368485 \cdot IPC$$

Coefficiente de Correlación = **0,047**

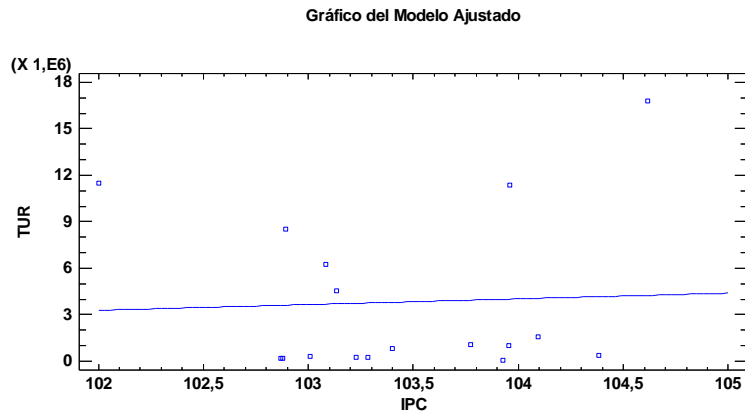


Ilustración 3.17 Gráfico X-Y para la IPC

Según muestra el gráfico (**Ilustración 3.17**) y el coeficiente de correlación, se puede indicar que existe una relación directa pequeña o prácticamente nula entre las entradas de Turistas y el IPC.

Además, indicar que la pendiente es horizontal.

Entradas de turistas según Comunidad autónoma VS Número de empresas de Comercio al por menor, excepto de vehículos de motor y motocicletas (TUR VS NEC)

$$TUR = 295558 + 103,058 * NEC$$

Coefficiente de Correlación = **0,63**

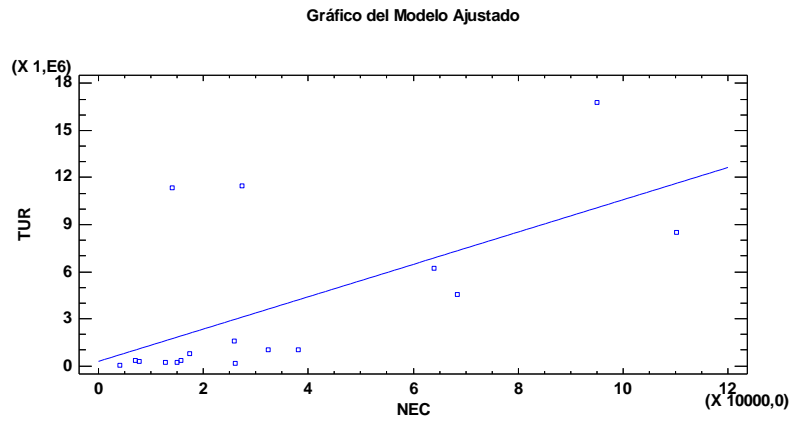


Ilustración 3.18 Gráfico X-Y para la NEC

Según muestra el gráfico (**Ilustración 3.18**) y el coeficiente de correlación, se puede indicar que existe una relación directa moderada entre las entradas de Turistas y los Comercio al por menor, excepto de vehículos de motor y motocicletas.

Además, indicar que la pendiente es positiva.

Entradas de turistas según Comunidad autónoma VS Número de empresas de Transporte terrestre y por tubería; Transporte marítimo y por vías navegables interiores; Transporte aéreo (TUR VS NETR)

$$TUR = 239774 + 346,54 * NETR$$

Coefficiente de Correlación = 0,65

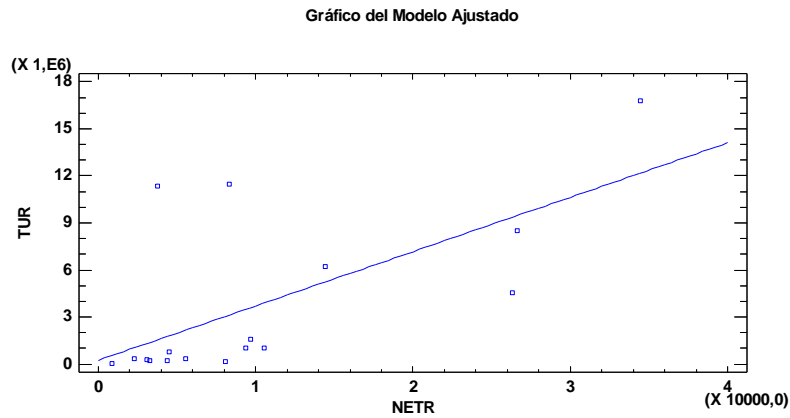


Ilustración 3.19 Gráfico X-Y para la NETR

Según muestra el gráfico (**Ilustración 3.19**) y el coeficiente de correlación, se puede indicar que existe una relación directa moderada entre las entradas de Turistas y las empresas de Transporte terrestre y por tubería, Transporte marítimo y por vías navegables interiores y Transporte aéreo.

Además, indicar que la pendiente es positiva.

Entradas de turistas según Comunidad autónoma VS Número de empresas de Servicios de alojamiento (TUR VS NES)

$$TUR = -2,17079 \cdot 10^6 + 3549,75 \cdot NES$$

Coefficiente de Correlación = **0,85**

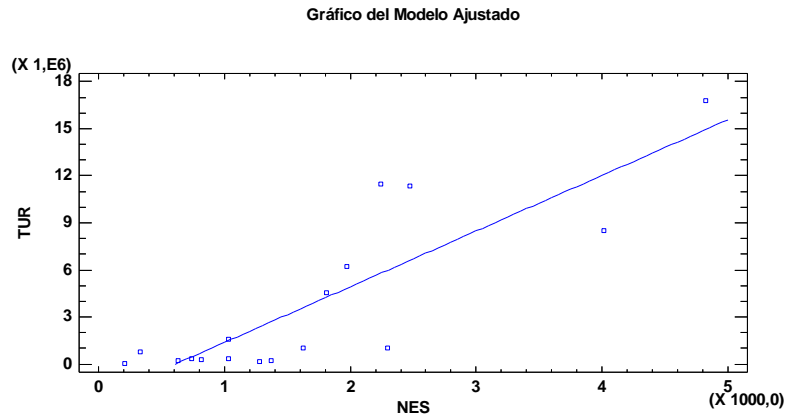


Ilustración 3.20 Gráfico X-Y para la NES

Según muestra el gráfico (**Ilustración 3.20**) y el coeficiente de correlación, se puede indicar que existe una relación directa muy elevada entre las entradas de Turistas y las empresas de Servicios de alojamiento.

Además, indicar que la pendiente es positiva.

Entradas de turistas según Comunidad autónoma VS Número de empresas de Actividades inmobiliarias y agencias de viajes (TUR VS NEAI)

$$TUR = 728756 + 316,26 * NEAI$$

Coefficiente de Correlación = **0,72**

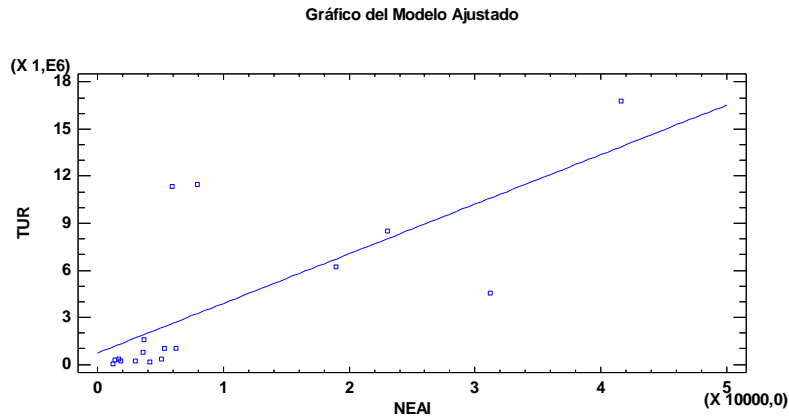


Ilustración 3.21 Gráfico X-Y para la NEAI

Según muestra el gráfico (**Ilustración 3.21**) y el coeficiente de correlación, se puede indicar que existe una relación directa elevada entre las entradas de Turistas y las empresas de Actividades inmobiliarias, Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas.

Además, indicar que la pendiente es positiva.

**Entradas de turistas según Comunidad autónoma VS Número de Parados (miles de personas)
(TUR VS PAR)**

$$TUR = 1,10745 \cdot 10^6 + 8484,95 \cdot PAR$$

Coefficiente de Correlación = **0,57**

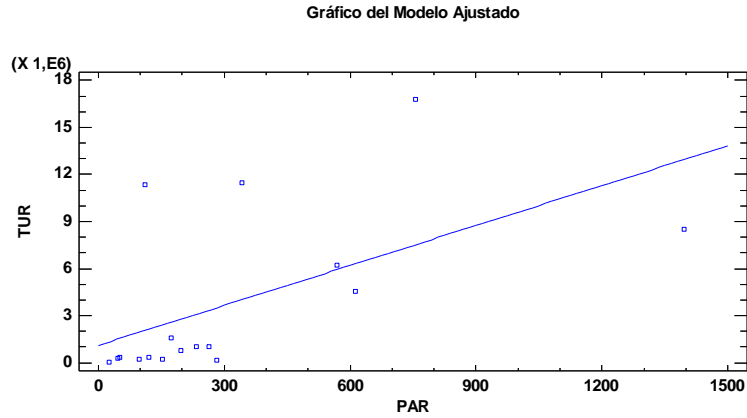


Ilustración 3.22 Gráfico X-Y para la PAR

Según muestra el gráfico (**Ilustración 3.22**) y el coeficiente de correlación, se puede indicar que existe una relación directa moderada entre las entradas de Turistas y los parados.

Además, indicar que la pendiente es positiva.

**Entradas de turistas según Comunidad autónoma VS Número de Ocupados (miles de personas)
(TUR VS OCU)**

$$TUR = 287870 + 3427,61 * OCU$$

Coefficiente de Correlación = **0,63**

Gráfico del Modelo Ajustado

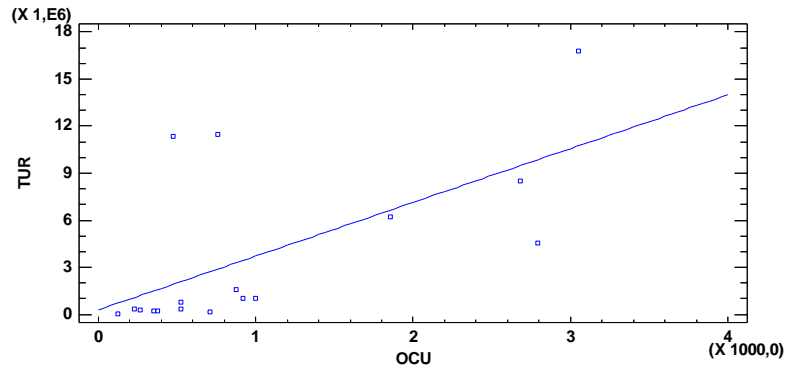


Ilustración 3.23 Gráfico X-Y para la OCU

Según muestra el gráfico (**Ilustración 3.23**) y el coeficiente de correlación, se puede indicar que existe una relación directa moderada entre las entradas de Turistas y los ocupados.

Además, indicar que la pendiente es positiva.

Entradas de turistas según Comunidad autónoma VS Crecimiento Producto Interior Bruto per cápita (euros) (TUR VS PIB)

$$TUR = 925037 + 131,309 * PIB$$

Coefficiente de Correlación = **0,11**

Gráfico del Modelo Ajustado

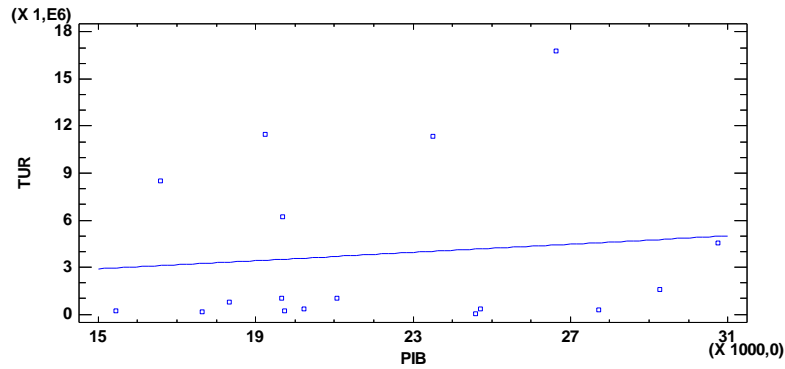


Ilustración 3.24 Gráfico X-Y para la PIB

Según muestra el gráfico (**Ilustración 3.24**) y el coeficiente de correlación, se puede indicar que existe una relación directa pequeña entre las entradas de Turistas y el PIB.

Además, indicar que la pendiente es horizontal/positiva.

Entradas de turistas según Comunidad autónoma VS Gasto Total de los turistas según destino principal (en millones de euros) (TUR VS GTOT)

$$TUR = 29153,6 + 1022,34 * GTOT$$

Coefficiente de Correlación = **0,99**

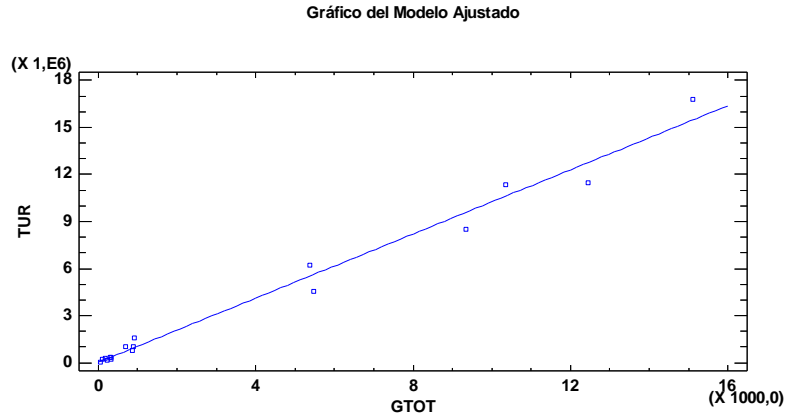


Ilustración 3.25 Gráfico X-Y para la GTOT

Según muestra el gráfico (**Ilustración 3.25**) y el coeficiente de correlación, se puede indicar que existe una relación directa muy elevada entre las entradas de Turistas y el Gasto Total de los Turistas.

Además, indicar que la pendiente es positiva.

Entradas de turistas según Comunidad autónoma VS Gasto Medio de los turistas según destino principal (en euros) (TUR VS GTOM)

$$TUR = 1,2527 \cdot 10^6 + 2738,82 \cdot GTOM$$

Coefficiente de Correlación = **0,12**

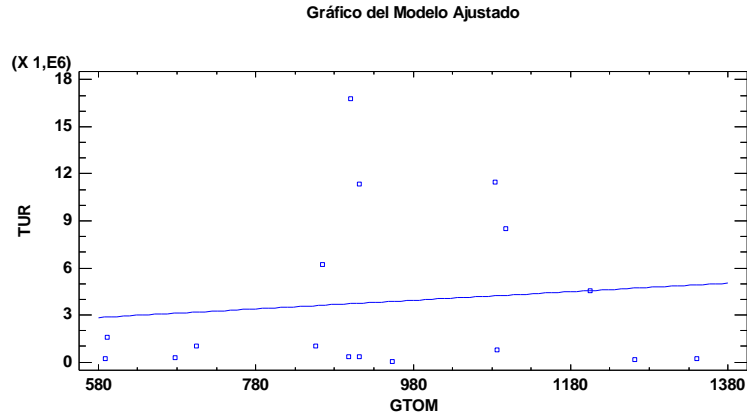


Ilustración 3.26 Gráfico X-Y para la GTOM

Según muestra el gráfico (**Ilustración 3.26**) y el coeficiente de correlación, se puede indicar que existe una relación directa pequeña entre las entradas de Turistas y el Gasto Medio por persona (Turistas).

Además, indicar que la pendiente es horizontal/positiva.

Entradas de turistas según Comunidad autónoma VS Gasto Medio Diario de los turistas según destino principal (en euros) (TUR VS GTOMD)

$$TUR = -3,25495 \cdot 10^6 + 67342,1 \cdot GTOMD$$

Coefficiente de Correlación = **0,26**

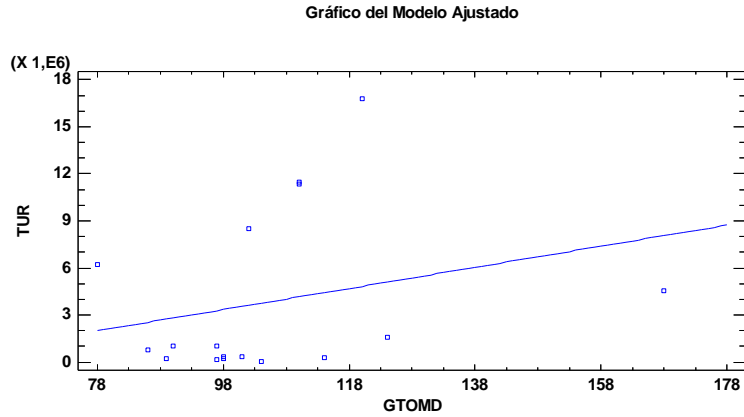


Ilustración 3.27 Gráfico X-Y para la GTOMD

Según muestra el gráfico (**Ilustración 3.27**) y el coeficiente de correlación, se puede indicar que existe una relación directa pequeña entre las entradas de Turistas y el Gasto Medio Diario por persona (Turistas).

Además, indicar que la pendiente es horizontal/positiva.

Entradas de turistas según Comunidad autónoma VS Estancia Media de los turistas según destino principal (noches) (TUR VS EM)

$$TUR = 6,26945 \cdot 10^6 - 266934 \cdot EM$$

Coefficiente de Correlación = - 0,13

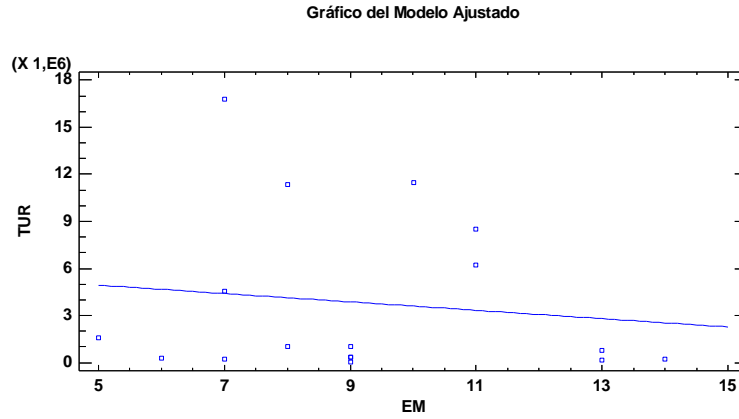


Ilustración 3.28 Gráfico X-Y para la EM

Según muestra el gráfico (**Ilustración 3.28**) y el coeficiente de correlación, se puede indicar que existe una relación indirecta pequeña entre las entradas de Turistas y la estancia media por turista.

Además, indicar que la pendiente es negativa.

3.2.1 Análisis de Multicolinealidad

Dentro del análisis Bivariante, se encuentra el análisis de la Multicolinealidad.

Se trata de relacionar cada una de las variables X con el resto de las X. Se debe indicar que la variable Y no se incluye en ningún momento en el análisis.

Para realizar esta parte del análisis se va a utilizar el Statgraphics y el Microsoft Excel. Las pruebas a realizar serán las siguientes:

- Matriz de Correlación
- Matriz Inversa de Correlación
- Índice de Acondicionamiento

Se seguirán los pasos anteriormente señalados a partir del modelo propuesto inicial. A partir del primer análisis se irán eliminando variables hasta conseguir un modelo libre de problemas de multicolinealidad.

El modelo teórico propuesto inicial es el siguiente:

$$\text{TUR} = \beta_0 + \beta_1 \text{PTOTAL} + \beta_2 \text{IPC} + \beta_3 \text{NEC} + \beta_4 \text{NETR} + \beta_5 \text{NES} + \beta_6 \text{NEAI} + \beta_7 \text{PAR} + \beta_8 \text{OCU} + \beta_9 \text{PIB} + \beta_{10} \text{GTOT} + \beta_{11} \text{GTOM} + \beta_{12} \text{GTOMD} + \beta_{13} \text{EM} + U$$

Interpretación de los parámetros y unidades físicas:

TUR → (Número de personas). Será nuestra variable dependiente, por lo tanto, la que tratamos de explicar a través de las variables independientes. Representa el total de las entradas de Turistas en cada una de las Comunidades Autónomas españolas en el año 2014. Se mide en número de personas y representan cifras oficiales registradas

β_0 → Valor medio de las entradas de Turistas cuando el resto de variables explicativas son 0.

β_1 → (Número de personas). Coeficiente que acompaña a la variable **PTOTAL**. Es el valor promedio de la Población Total cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable **PTOTAL** y el resto de variables explicativas se mantienen constantes.

β_2 → (%). Coeficiente que acompaña a la variable **IPC**. Es el valor promedio del Índice de Precios al Consumo cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable **IPC** y el resto de variables explicativas se mantienen constantes.

β_3 → (Número de empresas). Coeficiente que acompaña a la variable **NEC**. Es el valor promedio del número de empresas de comercio al por menor cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable **NEC** y el resto de variables explicativas se mantienen constantes.

β_4 → (Número de empresas). Coeficiente que acompaña a la variable **NETR**. Es el valor promedio del número de empresas de Transporte terrestre y por tubería, Transporte marítimo y por vías navegables interiores y Transporte aéreo cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable **NETR** y el resto de variables explicativas se mantienen constantes.

$\beta_5 \rightarrow$ **(Número de empresas)**. Coeficiente que acompaña a la variable **NES**. Es el valor promedio del número de empresas de Servicios de alojamiento cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable NES y el resto de variables explicativas se mantienen constantes.

$\beta_6 \rightarrow$ **(Número de empresas)**. Coeficiente que acompaña a la variable **NEAI**. Es el valor promedio del número de empresas de Actividades inmobiliarias y Actividades de agencias de viajes cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable NEAI y el resto de variables explicativas se mantienen constantes.

$\beta_7 \rightarrow$ **(Miles de Personas)**. Coeficiente que acompaña a la variable **PAR**. Es el valor promedio de miles de personas paradas cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable PAR y el resto de variables explicativas se mantienen constantes.

$\beta_8 \rightarrow$ **(Miles de Personas)**. Coeficiente que acompaña a la variable **OCU**. Es el valor promedio de miles de personas ocupadas cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable OCU y el resto de variables explicativas se mantienen constantes.

$\beta_9 \rightarrow$ **(Euros)**. Coeficiente que acompaña a la variable **PIB**. Es el valor promedio de euros de PIB per cápita cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable PIB y el resto de variables explicativas se mantienen constantes.

$\beta_{10} \rightarrow$ **(Millones de Euros)**. Coeficiente que acompaña a la variable **GTOT**. Es el valor promedio de millones de euros de Gasto Total de los Turistas cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable GTOT y el resto de variables explicativas se mantienen constantes.

$\beta_{11} \rightarrow$ **(Euros)**. Coeficiente que acompaña a la variable **GTOM**. Es el valor promedio de euros de Gasto Total Medio de los Turistas cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable GTOM y el resto de variables explicativas se mantienen constantes.

$\beta_{12} \rightarrow$ **(Euros)**. Coeficiente que acompaña a la variable **GTOMD**. Es el valor promedio de euros de Gasto Total Medio Diario de los Turistas cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable GTOMD y el resto de variables explicativas se mantienen constantes.

$\beta_{13} \rightarrow$ **(Noches)**. Coeficiente que acompaña a la variable **EM**. Es el valor promedio de noches de Estancia Media de los Turistas cuando las demás variables sean 0, es decir, representa el incremento de las entradas de Turistas cuando aumenta en una unidad la variable EM y el resto de variables explicativas se mantienen constantes.

$U \rightarrow$ **Perturbación del modelo**. Son todas las variables explicativas que no se tienen en cuenta en el modelo pero sí que sirven para explicar el TUR. Es decir, la parte de los datos que se pueden explicar con el modelo.

3.2.1.1 Matriz de Correlación

A través de este estudio se va a tratar de analizar si las variables explicativas/dependientes (X), tienen una fuerte relación entre ellas, es decir, identificar si hay problemas de Multicolinealidad. En caso de identificar el problema, se expondrán todos los pasos para poder solucionarlo.

El análisis se hace a partir de las trece variables X existentes en el estudio.

Para determinar el grado de relación lineal que existe relacionando las variables 2 a 2, y así estudiar las que suponen un problema en el modelo, se empieza calculando la matriz de correlaciones (R1). Mediante la matriz se ha obtenido que las variables están muy relacionadas entre ellas. Se indica marcado en rojo los valores que superan el límite de $|R_{ii}| \geq 0,7$.

Tabla 3.16 Matriz de Correlación 1ª de una variable X frente otra X.

	EM	GTOM	GTOMD	GTOT	IPC	NEAI	NEC	NES	NETR	OCU	PAR	PIB	PTOTAL
EM	1	0,76	-0,50	-0,10	-0,37	-0,15	-0,01	-0,07	-0,14	-0,12	0,11	-0,66	-0,04
GTOM	0,76	1	0,16	0,18	-0,33	0,21	0,18	0,13	0,19	0,20	0,26	-0,20	0,21
GTOMD	-0,50	0,16	1	0,30	0,07	0,48	0,26	0,19	0,47	0,45	0,18	0,78	0,34
GTOT	-0,10	0,18	0,30	1	-0,04	0,72	0,64	0,84	0,66	0,64	0,61	0,09	0,63
IPC	-0,37	-0,33	0,07	-0,04	1	0,12	-0,01	0,12	0,09	0,04	-0,18	0,39	-0,02
NEAI	-0,15	0,21	0,48	0,72	0,12	1	0,89	0,79	0,96	0,96	0,77	0,30	0,92
NEC	-0,01	0,18	0,26	0,64	-0,01	0,89	1	0,84	0,95	0,96	0,96	0,03	0,99
NES	-0,07	0,13	0,19	0,84	0,12	0,79	0,84	1	0,82	0,79	0,78	0,02	0,81
NETR	-0,14	0,19	0,47	0,66	0,09	0,96	0,95	0,82	1	0,99	0,85	0,24	0,97
OCU	-0,12	0,20	0,45	0,64	0,04	0,96	0,96	0,79	0,99	1	0,88	0,22	0,98
PAR	0,11	0,26	0,18	0,61	-0,18	0,77	0,96	0,78	0,85	0,88	1	-0,13	0,95
PIB	-0,66	-0,20	0,78	0,09	0,39	0,30	0,03	0,02	0,24	0,22	-0,13	1	0,09
PTOTAL	-0,04	0,21	0,34	0,63	-0,02	0,92	0,99	0,81	0,97	0,98	0,95	0,09	1

Como se observa en la **(Tabla 3.16)** la matriz de correlación muestra la relación que existe entre las distintas variables. Existe una elevada relación entre:

- Estancia Media y Gasto Total Medio (**EM y GTOM**)
- Gasto Medio Diario y Producto Interior Bruto (**GTOMD y PIB**)
- Gasto Total con empresas de Actividades inmobiliarias y Actividades de agencias de viajes con empresas de Servicios de alojamiento (**GTOT con NEAI y NES**)
- Empresas de Actividades inmobiliarias y Actividades de agencias de viajes con empresas de Servicios de alojamiento, empresas de Transporte (terrestre y por tubería; marítimo y por vías navegables interiores y aéreo), con Ocupados y Parados. (**NEAI con NEC, NES, NETR, OCU y PAR**)
- Empresas de Comercio al por menor (excepto vehículos de motor y motocicletas) con empresas de Servicios de alojamiento, empresas de Transporte (terrestre y por tubería; marítimo y por vías navegables interiores y aéreo), con Ocupados, Parados y Población Total (**NEC con NES, NETR, OCU, PAR y PTOTAL**)
- Empresas de Servicios de alojamiento con empresas de Transporte (terrestre y por tubería; marítimo y por vías navegables interiores y aéreo), con Ocupados, Parados y Población Total (**NES con NETR, OCU, PAR y PTOTAL**)

- Empresas de Transporte (terrestre y por tubería; marítimo y por vías navegables interiores y aéreo), con Ocupados, Parados y Población Total (**NETR con OCU, PAR y PTOTAL**)
- Ocupados con Parados y Población Total (**OCU con PAR y PTOTAL**)
- Parados y Población Total (**PAR y PTOTAL**)

Con las variables indicadas se puede indicar que existe fuerte relación entre ellas. Por lo que existe problema de Multicolinealidad.

3.2.1.2 Matriz Inversa de Correlación

Para poder asegurar que variables se deberían de eliminar del modelo se ha calculado la matriz inversa de correlaciones (R⁻¹)-1. Esta matriz representa la relación de una variable X frente al resto de las X. Donde, para estudiar los problemas de multicolinealidad se han marcado en rojo los números de la diagonal ≥ 10 . Este valor indica que hay problemas de multicolinealidad.

Tabla 3.17 Matriz Inversa de Correlación 1ª de una variable X frente el resto de X

EM	186,61	-159,08	154,64	-60,91	5,13	5,76	601,76	68,47	-138,03	1011,83	364,77	-16,90	-1838
GTOM	-159,08	137,97	-132,30	49,41	-4,20	-16,37	-460,66	-56,68	112,09	-765,83	-282,76	15,69	1409
GTOMD	154,64	-132,30	144,80	-50,61	4,00	32,79	531,55	54,07	-136,02	687,18	240,38	-20,90	-1356
GTOT	-60,91	49,41	-50,61	50,36	-0,74	47,31	-469,11	-48,48	88,51	-978,94	-336,79	4,25	1630
IPC	5,13	-4,20	4,00	-0,74	2,14	-2,70	12,90	0,92	-3,74	39,95	15,59	-1,11	-61
NEAI	5,76	-16,37	32,79	47,31	-2,70	248,23	-684,33	-44,98	42,72	-1979,25	-636,71	-12,94	2960
NEC	601,76	-460,66	531,55	-469,11	12,90	-684,33	5862,50	415,57	-956,73	11066,88	3634,04	-40,34	-18696
NES	68,47	-56,68	54,07	-48,48	0,92	-44,98	415,57	54,25	-89,16	931,46	321,12	-4,48	-1522
NETR	-138,03	112,09	-136,02	88,51	-3,74	42,72	-956,73	-89,16	252,52	-1640,27	-530,54	16,59	2815
OCU	1011,83	-765,83	687,18	-978,94	39,95	-1979,25	11066,88	931,46	-1640,27	25632,24	8607,13	-9,52	-41122
PAR	364,77	-282,76	240,38	-336,79	15,59	-636,71	3634,04	321,12	-530,54	8607,13	2952,51	-0,67	-13830
PIB	-16,90	15,69	-20,90	4,25	-1,11	-12,94	-40,34	-4,48	16,59	-9,52	-0,67	7,69	49
PTOTAL	-1837,89	1408,61	-1355,85	1630,39	-61,21	2960,49	-18695,64	-1522,37	2815,16	-41122,18	-13829,77	49,35	66971

Como se puede observar en la **Tabla 3.17** las variables: PTOTAL, PAR, OCU, NETR, NES, NEC, NEAI, FTOT, GTOMD, GTOM, EM se han obtenido valores muy elevados que superan el valor de 10. Esto significa las variables están muy relacionadas, hay problemas de multicolinealidad.

Para continuar con el análisis se calcula el índice de acondicionamiento (I.C). Esta prueba indica la relación de todas las variables X a la vez.

3.2.1.3 Índice de Acondicionamiento (IC)

Lo calcularemos mediante la raíz del autovalor máximo entre el autovalor mínimo y según su resultado obtendremos que tipo de multicolinealidad existe.

$$IC = \sqrt{\frac{\text{Autovalor M\u00e1ximo}}{\text{Autovalor M\u00ednimo}}} = \sqrt{\frac{7,16557}{0,000829345}} = 92,9518$$

Tabla 3.18 Autovalores 1ª para realizar el Índice de Acondicionamiento.

	AUTOVALOR	IC
		92,9518
1	7,16557 MÁXIMO	
2	2,71885	
3	1,28581	
4	0,802962	
5	0,635072	
6	0,162611	
7	0,121528	
8	0,0790976	
9	0,0152904	
10	0,0093181	
11	0,00305637	
12	0,000829345 MÍNIMO	

Si el Índice de Acondicionamiento es mayor a 10, hay problemas de multicolinealidad. Por lo que en el caso del valor del IC (92,95), se puede indicar que las variables están muy relacionadas.

Tras haber realizado las 3 pruebas planteadas al inicio del análisis de Multicolinealidad se puede afirmar que las variables están muy relacionadas, por lo que, para poder solucionar el problema de multicolinealidad se deben eliminar variables del modelo teórico inicial planteado.

Para poder corregir el modelo inicial se deben tener en cuenta los valores de la diagonal de la Matriz Inversa de Correlación (MIC). Los valores candidatos a ser eliminados del modelo son los superiores a 10. La variable inicial a eliminar es la variable con valor superior, por lo que en este caso es la variable de Población Total (PTOTAL).

El modelo teórico propuesto al eliminar la variable PTOTAL es el siguiente:

$$\text{TUR} = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{NEC} + \beta_3 \text{NETR} + \beta_4 \text{NES} + \beta_5 \text{NEAI} + \beta_6 \text{PAR} + \beta_7 \text{OCU} + \beta_8 \text{PIB} + \beta_9 \text{GTOT} + \beta_{10} \text{GTOM} + \beta_{11} \text{GTOMD} + \beta_{12} \text{EM} + U$$

Se realiza de nuevo la Matriz de Correlación. Al igual que en la matriz inicial se marcan los valores que superan el límite de $|R_{ii}| \geq 0,7$, valores que indican que continúan existiendo problemas de Multicolinealidad.

Tabla 3.19 Matriz de Correlación 2ª de una variable X frente otra X. (Eliminando del modelo la variable PTOTAL)

	EM	GTOM	GTOMD	GTOT	IPC	NEAI	NEC	NES	NETR	OCU	PAR	PIB
EM	1	0,76	-0,50	-0,10	-0,37	-0,15	-0,01	-0,07	-0,14	-0,12	0,11	-0,66
GTOM	0,76	1	0,16	0,18	-0,33	0,21	0,18	0,13	0,19	0,20	0,26	-0,20
GTOMD	-0,50	0,16	1	0,30	0,07	0,48	0,26	0,19	0,47	0,45	0,18	0,78
GTOT	-0,10	0,18	0,30	1	-0,04	0,72	0,64	0,84	0,66	0,64	0,61	0,09
IPC	-0,37	-0,33	0,07	-0,04	1	0,12	-0,01	0,12	0,09	0,04	-0,18	0,39
NEAI	-0,15	0,21	0,48	0,72	0,12	1	0,89	0,79	0,96	0,96	0,77	0,30
NEC	-0,01	0,18	0,26	0,64	-0,01	0,89	1	0,84	0,95	0,96	0,96	0,03
NES	-0,07	0,13	0,19	0,84	0,12	0,79	0,84	1	0,82	0,79	0,78	0,02
NETR	-0,14	0,19	0,47	0,66	0,09	0,96	0,95	0,82	1	0,99	0,85	0,24
OCU	-0,12	0,20	0,45	0,64	0,04	0,96	0,96	0,79	0,99	1	0,88	0,22
PAR	0,11	0,26	0,18	0,61	-0,18	0,77	0,96	0,78	0,85	0,88	1	-0,13
PIB	-0,66	-0,20	0,78	0,09	0,39	0,30	0,03	0,02	0,24	0,22	-0,13	1

Para poder asegurar que variables se deberían de eliminar del modelo se ha calculado la matriz inversa de correlaciones (R1)-1. Donde, los problemas de multicolinealidad se han marcado en amarillo y rojo los números de la diagonal ≥ 10 . Este valor indica que hay problemas de multicolinealidad.

Tabla 3.20 Matriz Inversa de Correlación 2ª de una variable X frente el resto de X. (Eliminando del Modelo la variable PTOTAL)

EM	136,17	-120,43	117,43	-16,17	3,45	87,00	88,69	26,69	-60,78	-116,69	-14,77	-15,55
GTOM	-120,43	108,34	-103,78	15,11	-2,91	-78,64	-67,43	-24,66	52,88	99,09	8,13	14,65
GTOMD	117,43	-103,78	117,35	-17,60	2,77	92,73	153,05	23,25	-79,02	-145,35	-39,61	-19,90
GTOT	-16,17	15,11	-17,60	10,67	0,75	-24,76	-13,97	-11,42	19,98	22,16	-0,11	3,05
IPC	3,45	-2,91	2,77	0,75	2,08	0,01	-4,18	-0,47	-1,16	2,36	2,95	-1,07
NEAI	87,00	-78,64	92,73	-24,76	0,01	117,36	142,12	22,32	-81,73	-161,42	-25,36	-15,12
NEC	88,69	-67,43	153,05	-13,97	-4,18	142,12	643,43	-9,42	-170,85	-412,78	-226,67	-26,57
NES	26,69	-24,66	23,25	-11,42	-0,47	22,32	-9,42	19,64	-25,17	-3,32	6,74	-3,36
NETR	-60,78	52,88	-79,02	19,98	-1,16	-81,73	-170,85	-25,17	134,18	88,32	50,80	14,52
OCU	-116,69	99,09	-145,35	22,16	2,36	-161,42	-412,78	-3,32	88,32	382,04	115,25	20,78
PAR	-14,77	8,13	-39,61	-0,11	2,95	-25,36	-226,67	6,74	50,80	115,25	96,62	9,52
PIB	-15,55	14,65	-19,90	3,05	-1,07	-15,12	-26,57	-3,36	14,52	20,78	9,52	7,66

Se vuelve a calcular el Índice de Acondicionamiento mediante la raíz del autovalor máximo entre el autovalor mínimo.

$$IC = \sqrt{\frac{\text{Autovalor M\u00e1ximo}}{\text{Autovalor M\u00ednimo}}} = \sqrt{\frac{6,22531}{0,000815581}} = 87,366903$$

Tabla 3.21 Autovalores 2ª para realizar el Índice de Acondicionamiento.

	AUTOVALOR		IC	87,366903
1	6,22531	MÁXIMO		
2	2,70847			
3	1,28184			
4	0,801894			
5	0,593553			
6	0,16258			
7	0,121271			
8	0,0781529			
9	0,0149073			
10	0,00815532			
11	0,00304964			
12	0,000815581	MÍNIMO		

Según el resultado del IC se puede confirmar que siguen existiendo los problemas de multicolinealidad. Ya que el valor del IC es superior a 10.

Tras volverse a realizar las tres pruebas (Matriz de Correlación, Matriz Inversa de Correlación e Índice de Acondicionamiento), que muestran los problemas de Multicolinealidad, se puede afirmar que el problema sigue existiendo a pesar de que se ha eliminado del modelo la variable explicativa PTOTAL.

Por lo que, se debe de eliminar otra variable explicativa (X) para poder corregir los problemas del modelo. Para poder corregir el modelo inicial se deben tener en cuenta los valores de la diagonal de la Matriz Inversa de Correlación (MIC). Los valores candidatos a ser eliminados del modelo son los superiores a 10. La variable a eliminar es la variable con valor superior, por lo que en este caso es la variable número de empresas de Comercio al por menor (excepto de vehículos de motor y motocicletas) (NEC).

El modelo teórico propuesto al eliminar la variable NEC es el siguiente:

$$\text{TUR} = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{NETR} + \beta_3 \text{NES} + \beta_4 \text{NEAI} + \beta_5 \text{PAR} + \beta_6 \text{OCU} + \beta_7 \text{PIB} + \beta_8 \text{GTOT} + \beta_9 \text{GTOM} + \beta_{10} \text{GTOMD} + \beta_{11} \text{EM} + U$$

Se realiza de nuevo la Matriz de Correlación. Al igual que en la matriz inicial se marcan los valores que superan el límite $|R_{ii}| \geq 0,7$. Valores que indican que continúan existiendo problemas de Multicolinealidad.

Tabla 3.22 Matriz de Correlación 3ª de una variable X frente otra X. (Eliminando del modelo la variable NEC)

	EM	GTOM	GTOMD	GTOT	IPC	NEAI	NES	NETR	OCU	PAR	PIB
EM	1	0,76	-0,50	-0,10	-0,37	-0,15	-0,07	-0,14	-0,12	0,11	-0,66
GTOM	0,76	1	0,16	0,18	-0,33	0,21	0,13	0,19	0,20	0,26	-0,20
GTOMD	-0,50	0,16	1	0,30	0,07	0,48	0,19	0,47	0,45	0,18	0,78
GTOT	-0,10	0,18	0,30	1	-0,04	0,72	0,84	0,66	0,64	0,61	0,09
IPC	-0,37	-0,33	0,07	-0,04	1	0,12	0,12	0,09	0,04	-0,18	0,39
NEAI	-0,15	0,21	0,48	0,72	0,12	1	0,79	0,96	0,96	0,77	0,30
NES	-0,07	0,13	0,19	0,84	0,12	0,79	1	0,82	0,79	0,78	0,02
NETR	-0,14	0,19	0,47	0,66	0,09	0,96	0,82	1	0,99	0,85	0,24
OCU	-0,12	0,20	0,45	0,64	0,04	0,96	0,79	0,99	1	0,88	0,22
PAR	0,11	0,26	0,18	0,61	-0,18	0,77	0,78	0,85	0,88	1	-0,13
PIB	-0,66	-0,20	0,78	0,09	0,39	0,30	0,02	0,24	0,22	-0,13	1

Para poder asegurar que variables se deberían de eliminar del modelo se ha calculado la matriz inversa de correlaciones (R1)-1. Donde, los problemas de multicolinealidad se han marcado en amarillo y rojo los números de la diagonal ≥ 10 . Este valor indica que hay problemas de multicolinealidad.

Tabla 3.23 Matriz Inversa de Correlación 3ª de una variable X frente el resto de X. (Eliminando del Modelo la variable NEC)

EM	123,95	-111,13	96,33	-14,24	4,02	67,41	27,99	-37,23	-59,79	16,48	-11,89
GTOM	-111,13	101,27	-87,74	13,65	-3,35	-63,74	-25,65	34,98	55,83	-15,63	11,87
GTOMD	96,33	-87,74	80,94	-14,27	3,76	58,92	25,49	-38,38	-47,16	14,31	-13,58
GTOT	-14,24	13,65	-14,27	10,36	0,66	-21,67	-11,62	16,27	13,20	-5,03	2,47
IPC	4,02	-3,35	3,76	0,66	2,06	0,93	-0,53	-2,27	-0,32	1,48	-1,24
NEAI	67,41	-63,74	58,92	-21,67	0,93	85,97	24,40	-43,99	-70,24	24,71	-9,26
NES	27,99	-25,65	25,49	-11,62	-0,53	24,40	19,50	-27,67	-9,36	3,43	-3,75
NETR	-37,23	34,98	-38,38	16,27	-2,27	-43,99	-27,67	88,82	-21,28	-9,39	7,46
OCU	-59,79	55,83	-47,16	13,20	-0,32	-70,24	-9,36	-21,28	117,23	-30,17	3,74
PAR	16,48	-15,63	14,31	-5,03	1,48	24,71	3,43	-9,39	-30,17	16,76	0,16
PIB	-11,89	11,87	-13,58	2,47	-1,24	-9,26	-3,75	7,46	3,74	0,16	6,56

Se vuelve a calcular el Índice de Acondicionamiento mediante la raíz del autovalor máximo entre el autovalor mínimo.

$$IC = \sqrt{\frac{\text{Autovalor Máximo}}{\text{Autovalor Mínimo}}} = \sqrt{\frac{5,3409}{0,00235193}} = 47,6535$$

Tabla 3.24 Autovalores 3ª para realizar el Índice de Acondicionamiento.

	AUTOVALOR	IC
		47,6535
1	5,3409 MÁXIMO	
2	2,67764	
3	1,25535	
4	0,801125	
5	0,542678	
6	0,162402	
7	0,118557	
8	0,0766689	
9	0,0144452	
10	0,00787546	
11	0,00235193 MÍNIMO	

Según el resultado del IC se puede confirmar que siguen existiendo los problemas de multicolinealidad. Ya que el valor del IC es superior a 10.

Tras volverse a realizar las tres pruebas (Matriz de Correlación, Matriz Inversa de Correlación e Índice de Acondicionamiento), que muestran los problemas de Multicolinealidad, se puede afirmar que el problema sigue existiendo a pesar de que se ha eliminado del modelo la variable explicativa NEC.

Por lo que, se debe de eliminar otra variable explicativa (X) para poder corregir los problemas del modelo. Para poder corregir el modelo inicial se deben tener en cuenta los valores de la diagonal de la Matriz Inversa de Correlación (MIC). Los valores candidatos a ser eliminados del modelo son los superiores a 10. La variable a eliminar es la variable con valor superior, por lo que en este caso es la variable Gasto de los Turistas según Estancia Media (noches) (EM).

El modelo teórico propuesto al eliminar la variable EM es el siguiente:

$$\text{TUR} = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{NETR} + \beta_3 \text{NES} + \beta_4 \text{NEAI} + \beta_5 \text{PAR} + \beta_6 \text{OCU} + \beta_7 \text{PIB} + \beta_8 \text{GTOT} + \beta_9 \text{GTOM} + \beta_{10} \text{GTOMD} + U$$

Se realiza de nuevo la Matriz de Correlación. Al igual que en la matriz inicial se marcan los valores que superan el límite de $|R_{ii}| \geq 0,7$. Valores que indican que continúan existiendo problemas de Multicolinealidad.

Tabla 3.25 Matriz de Correlación 4ª de una variable X frente otra X. (Eliminando del modelo la variable EM)

	GTOM	GTOMD	GTOT	IPC	NEAI	NES	NETR	OCU	PAR	PIB
GTOM	1	0,16	0,18	-0,33	0,21	0,13	0,19	0,20	0,26	-0,20
GTOMD	0,16	1	0,30	0,07	0,48	0,19	0,47	0,45	0,18	0,78
GTOT	0,18	0,30	1	-0,04	0,72	0,84	0,66	0,64	0,61	0,09
IPC	-0,33	0,07	-0,04	1	0,12	0,12	0,09	0,04	-0,18	0,39
NEAI	0,21	0,48	0,72	0,12	1	0,79	0,96	0,96	0,77	0,30
NES	0,13	0,19	0,84	0,12	0,79	1	0,82	0,79	0,78	0,02
NETR	0,19	0,47	0,66	0,09	0,96	0,82	1	0,99	0,85	0,24
OCU	0,20	0,45	0,64	0,04	0,96	0,79	0,99	1	0,88	0,22
PAR	0,26	0,18	0,61	-0,18	0,77	0,78	0,85	0,88	1	-0,13
PIB	-0,20	0,78	0,09	0,39	0,30	0,02	0,24	0,22	-0,13	1

Se vuelve a calcular la matriz inversa de correlaciones (R1)-1, para asegurar que variables se deberían de eliminar del modelo. Los problemas de multicolinealidad se han vuelto a marcar en amarillo y rojo, teniendo en cuenta números de la diagonal ≥ 10 . Este valor indica que hay problemas de multicolinealidad.

Tabla 3.26 Matriz Inversa de Correlación 4ª de una variable X frente el resto de X. (Eliminando del Modelo la variable EM)

GTOM	1,63	-1,37	0,88	0,26	-3,30	-0,55	1,60	2,23	-0,85	1,21
GTOMD	-1,37	6,08	-3,21	0,63	6,53	3,74	-9,45	-0,69	1,50	-4,35
GTOT	0,88	-3,21	8,73	1,12	-13,93	-8,40	11,99	6,33	-3,14	1,10
IPC	0,26	0,63	1,12	1,93	-1,26	-1,44	-1,07	1,62	0,94	-0,86
NEAI	-3,30	6,53	-13,93	-1,26	49,30	9,18	-23,74	-37,73	15,75	-2,79
NES	-0,55	3,74	-8,40	-1,44	9,18	13,18	-19,26	4,14	-0,29	-1,06
NETR	1,60	-9,45	11,99	-1,07	-23,74	-19,26	77,64	-39,24	-4,44	3,89
OCU	2,23	-0,69	6,33	1,62	-37,73	4,14	-39,24	88,39	-22,22	-1,99
PAR	-0,85	1,50	-3,14	0,94	15,75	-0,29	-4,44	-22,22	14,57	1,74
PIB	1,21	-4,35	1,10	-0,86	-2,79	-1,06	3,89	-1,99	1,74	5,42

De nuevo se calcula el Índice de Acondicionamiento mediante la raíz del autovalor máximo entre el autovalor mínimo.

$$IC = \sqrt{\frac{\text{Autovalor Máximo}}{\text{Autovalor Mínimo}}} = \sqrt{\frac{5,31879}{0,00767478}} = 26,3253$$

Tabla 3.27 Autovalores 4ª para realizar el Índice de Acondicionamiento.

	AUTOVALOR	IC
		26,3253
1	5,31879 MÁXIMO	
2	1,95492	
3	1,20508	
4	0,632454	
5	0,541835	
6	0,146608	
7	0,108082	
8	0,074763	
9	0,00978841	
10	0,00767478 MÍNIMO	

Según el resultado del IC se puede confirmar que siguen existiendo los problemas de multicolinealidad. Ya que el valor del IC es superior a 10.

Tras volverse a realizar las tres pruebas por cuarta vez (Matriz de Correlación, Matriz Inversa de Correlación e Índice de Acondicionamiento), que muestran los problemas de Multicolinealidad, se puede afirmar que el problema sigue existiendo a pesar de que se ha eliminado del modelo la variable explicativa EM.

Por lo que, se debe de eliminar otra variable explicativa (X) para poder corregir los problemas del modelo. Para hacerlo, se volverán a tener en cuenta los valores de la diagonal de la Matriz Inversa de Correlación (MIC). Los valores candidatos a ser eliminados del modelo son los superiores a 10. La variable a eliminar es la variable con valor superior, por lo que en este caso es la variable Ocupados (OCU).

El modelo teórico propuesto al eliminar la variable OCU es el siguiente:

$$\text{TUR} = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{NETR} + \beta_3 \text{NES} + \beta_4 \text{NEAI} + \beta_5 \text{PAR} + \beta_6 \text{PIB} + \beta_7 \text{GTOT} + \beta_8 \text{GTOM} + \beta_9 \text{GTOMD} + U$$

Se realiza de nuevo la Matriz de Correlación. Al igual que en la matriz inicial se marcan los valores que superan el límite de $|R_{ii}| \geq 0,7$. Valores que indican que continúan existiendo problemas de Multicolinealidad.

Tabla 3.28 Matriz de Correlación 5ª de una variable X frente otra X. (Eliminando del modelo la variable OCU)

	GTOM	GTOMD	GTOT	IPC	NEAI	NES	NETR	PAR	PIB
GTOM	1	0,16	0,18	-0,33	0,21	0,13	0,19	0,26	-0,20
GTOMD	0,16	1	0,30	0,07	0,48	0,19	0,47	0,18	0,78
GTOT	0,18	0,30	1	-0,04	0,72	0,84	0,66	0,61	0,09
IPC	-0,33	0,07	-0,04	1	0,12	0,12	0,09	-0,18	0,39
NEAI	0,21	0,48	0,72	0,12	1	0,79	0,96	0,77	0,30
NES	0,13	0,19	0,84	0,12	0,79	1	0,82	0,78	0,02
NETR	0,19	0,47	0,66	0,09	0,96	0,82	1	0,85	0,24
PAR	0,26	0,18	0,61	-0,18	0,77	0,78	0,85	1	-0,13
PIB	-0,20	0,78	0,09	0,39	0,30	0,02	0,24	-0,13	1

Se vuelve a calcular la matriz inversa de correlaciones (R1)-1, para asegurar que variables se deberían de eliminar del modelo. Los problemas de multicolinealidad se han vuelto a marcar en amarillo y rojo, teniendo en cuenta números de la diagonal ≥ 10 . Este valor indica que hay problemas de multicolinealidad.

Tabla 3.29 Matriz Inversa de Correlación 5ª de una variable X frente el resto de X. (Eliminando del Modelo la variable OCU)

GTOM	1,58	-1,36	0,72	0,22	-2,35	-0,66	2,59	-0,30	1,26
GTOMD	-1,36	6,07	-3,16	0,65	6,23	3,77	-9,76	1,33	-4,36
GTOT	0,72	-3,16	8,27	1,01	-11,23	-8,70	14,80	-1,55	1,25
IPC	0,22	0,65	1,01	1,90	-0,56	-1,51	-0,35	1,35	-0,82
NEAI	-2,35	6,23	-11,23	-0,56	33,20	10,95	-40,49	6,26	-3,64
NES	-0,66	3,77	-8,70	-1,51	10,95	12,99	-17,43	0,75	-0,97
NETR	2,59	-9,76	14,80	-0,35	-40,49	-17,43	60,21	-14,30	3,01
PAR	-0,30	1,33	-1,55	1,35	6,26	0,75	-14,30	8,99	1,24
PIB	1,26	-4,36	1,25	-0,82	-3,64	-0,97	3,01	1,24	5,37

De nuevo se calcula el Índice de Acondicionamiento mediante la raíz del autovalor máximo entre el autovalor mínimo.

$$IC = \sqrt{\frac{\text{Autovalor M\u00e1ximo}}{\text{Autovalor M\u00ednimo}}} = \sqrt{\frac{4,40261}{0,00966213}} = 21,34610639$$

Tabla 3.30 Autovalores 5ª para realizar el Índice de Acondicionamiento.

	AUTOVALOR	IC
		21,34610639
1	4,40261 MÁXIMO	
2	1,95489	
3	1,20451	
4	0,631897	
5	0,473969	
6	0,140519	
7	0,107207	
8	0,0747397	
9	0,00966213 MÍNIMO	

Según el resultado del IC se puede confirmar que siguen existiendo los problemas de multicolinealidad. Ya que el valor del IC es superior a 10.

Tras volverse a realizar las tres pruebas por quinta vez (Matriz de Correlación, Matriz Inversa de Correlación e Índice de Acondicionamiento), que muestran los problemas de Multicolinealidad, se puede afirmar que el problema sigue existiendo a pesar de que se ha eliminado del modelo la variable explicativa OCU.

Por lo que, se debe de eliminar otra variable explicativa (X) para poder corregir los problemas del modelo. Para hacerlo, se volverán a tener en cuenta los valores de la diagonal de la Matriz Inversa de Correlación (MIC). Los valores candidatos a ser eliminados del modelo son los superiores a 10. La variable a eliminar es la variable con valor superior, por lo que en este caso es la variable número de empresas de Transporte terrestre y por tubería; Transporte marítimo y por vías navegables interiores y Transporte aéreo (NETR).

El modelo teórico propuesto al eliminar la variable NETR es el siguiente:

$$\text{TUR} = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{NES} + \beta_3 \text{NEAI} + \beta_4 \text{PAR} + \beta_5 \text{PIB} + \beta_6 \text{GTOT} + \beta_7 \text{GTOM} + \beta_8 \text{GTOMD} + U$$

Se realiza de nuevo la Matriz de Correlación. Al igual que en la matriz inicial se marcan los valores que superan el límite de $|R_{ii}| \geq 0,7$. Valores que indican que continúan existiendo problemas de Multicolinealidad.

Tabla 3.31 Matriz de Correlación 6ª de una variable X frente otra X. (Eliminando del modelo la variable NETR)

	GTOM	GTOMD	GTOT	IPC	NEAI	NES	PAR	PIB
GTOM	1	0,16	0,18	-0,33	0,21	0,13	0,26	-0,20
GTOMD	0,16	1	0,30	0,07	0,48	0,19	0,18	0,78
GTOT	0,18	0,30	1	-0,04	0,72	0,84	0,61	0,09
IPC	-0,33	0,07	-0,04	1	0,12	0,12	-0,18	0,39
NEAI	0,21	0,48	0,72	0,12	1	0,79	0,77	0,30
NES	0,13	0,19	0,84	0,12	0,79	1	0,78	0,02
PAR	0,26	0,18	0,61	-0,18	0,77	0,78	1	-0,13
PIB	-0,20	0,78	0,09	0,39	0,30	0,02	-0,13	1

Se vuelve a calcular la matriz inversa de correlaciones (R1)-1, para asegurar que variables se deberían de eliminar del modelo.

Tabla 3.32 Matriz Inversa de Correlación 6ª de una variable X frente el resto de X. (Eliminando del Modelo la variable NETR)

GTOM	1,47	-0,94	0,09	0,23	-0,61	0,09	0,32	1,14
GTOMD	-0,94	4,49	-0,76	0,59	-0,33	0,95	-0,99	-3,88
GTOT	0,09	-0,76	4,64	1,09	-1,27	-4,42	1,97	0,51
IPC	0,23	0,59	1,09	1,89	-0,80	-1,62	1,27	-0,80
NEAI	-0,61	-0,33	-1,27	-0,80	5,97	-0,77	-3,35	-1,62
NES	0,09	0,95	-4,42	-1,62	-0,77	7,95	-3,39	-0,10
PAR	0,32	-0,99	1,97	1,27	-3,35	-3,39	5,59	1,95
PIB	1,14	-3,88	0,51	-0,80	-1,62	-0,10	1,95	5,22

Los problemas de multicolinealidad no aparecen en esta Matriz Inversa de Correlación, ya que, si se tienen en cuenta los números de la diagonal ≥ 10 , no se encuentra ningún caso.

A continuación, se hace el Índice de Acondicionamiento para confirmar si existen o no problemas de Multicolinealidad.

De nuevo se calcula el Índice de Acondicionamiento mediante la raíz del autovalor máximo entre el autovalor mínimo.

$$IC = \sqrt{\frac{\text{Autovalor M\u00e1ximo}}{\text{Autovalor M\u00ednimo}}} = \sqrt{\frac{3,5038}{0,0679077}} = 7,18307074$$

Tabla 3.33 Autovalores 6ª para realizar el Índice de Acondicionamiento.

	AUTOVALOR	IC
		7,18307074
1	3,5038 MÁXIMO	
2	1,95384	
3	1,20123	
4	0,631833	
5	0,399195	
6	0,136126	
7	0,106073	
8	0,0679077 MÍNIMO	

Según el resultado del IC se puede indicar que para este modelo no existen problemas de multicolinealidad, ya que el valor del IC es inferior a 10.

Tras volverse a realizar las tres pruebas por sexta vez (Matriz de Correlación, Matriz Inversa de Correlación e Índice de Acondicionamiento), que muestran los problemas de Multicolinealidad, se puede observar que siguen existiendo problemas como muestra la Matriz de Correlaciones.

Por lo que este modelo sigue sin ser el definitivo, ya que, existe relación entre los datos. La variable a eliminar del modelo en este caso va a ser la que tenga mayor valor en la diagonal de la Matriz Inversa de Correlaciones, será la variable número de empresas de Servicios de alojamiento (NES).

El modelo teórico propuesto al eliminar la variable NES es el siguiente:

$$\text{TUR} = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{NEAI} + \beta_3 \text{PAR} + \beta_4 \text{PIB} + \beta_5 \text{GTOT} + \beta_6 \text{GTOM} + \beta_7 \text{GTOMD} + U$$

Se realiza de nuevo la Matriz de Correlación. Al igual que en la matriz inicial se marcan los valores que superan el límite de $|R_{ii}| \geq 0,7$. Valores que indican que continúan existiendo problemas de Multicolinealidad.

Tabla 3.34 Matriz de Correlación 7ª de una variable X frente otra X. (Eliminando del modelo la variable NES)

	GTOM	GTOMD	GTOT	IPC	NEAI	PAR	PIB
GTOM	1	0,16	0,18	-0,33	0,21	0,26	-0,20
GTOMD	0,16	1	0,30	0,07	0,48	0,18	0,78
GTOT	0,18	0,30	1	-0,04	0,72	0,61	0,09
IPC	-0,33	0,07	-0,04	1	0,12	-0,18	0,39
NEAI	0,21	0,48	0,72	0,12	1	0,77	0,30
PAR	0,26	0,18	0,61	-0,18	0,77	1	-0,13
PIB	-0,20	0,78	0,09	0,39	0,30	-0,13	1

Se vuelve a calcular la matriz inversa de correlaciones (R1)-1, para asegurar que variables se deberían de eliminar del modelo.

Tabla 3.35 Matriz Inversa de Correlación 7ª de una variable X frente el resto de X. (Eliminando del Modelo la variable NES)

GTOM	1,47	-0,95	0,14	0,25	-0,60	0,36	1,14
GTOMD	-0,95	4,38	-0,23	0,78	-0,24	-0,58	-3,86
GTOT	0,14	-0,23	2,18	0,19	-1,70	0,08	0,45
IPC	0,25	0,78	0,19	1,57	-0,95	0,58	-0,82
NEAI	-0,60	-0,24	-1,70	-0,95	5,89	-3,68	-1,63
PAR	0,36	-0,58	0,08	0,58	-3,68	4,15	1,91
PIB	1,14	-3,86	0,45	-0,82	-1,63	1,91	5,22

De nuevo se calcula el Índice de Acondicionamiento mediante la raíz del autovalor máximo entre el autovalor mínimo.

$$IC = \sqrt{\frac{\text{Autovalor M\u00e1ximo}}{\text{Autovalor M\u00ednimo}}} = \sqrt{\frac{2,78193}{0,0886874}} = 5,60069777$$

Tabla 3.36 Autovalores 7ª para realizar el Índice de Acondicionamiento.

	AUTOVALOR	IC
		5,60069777
1	2,78193 MÁXIMO	
2	1,92147	
3	1,05793	
4	0,63076	
5	0,387284	
6	0,13193	
7	0,0886874 MÍNIMO	

Según el resultado del IC se puede indicar que para este modelo no existen problemas de multicolinealidad, ya que el valor del IC es inferior a 10.

Tras volverse a realizar las tres pruebas por séptima vez (Matriz de Correlación, Matriz Inversa de Correlación e Índice de Acondicionamiento), que muestran los problemas de Multicolinealidad, se puede observar que siguen existiendo problemas como muestra la Matriz de Correlaciones.

Por lo que este modelo sigue sin ser el definitivo, ya que, existe relación entre los datos. La variable a eliminar del modelo en este caso va a ser la que tenga mayor valor en la diagonal de la Matriz Inversa de Correlaciones, será la variable número de empresas de Actividades inmobiliarias; Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con los mismos (NEAI).

El modelo teórico propuesto al eliminar la variable NEAI es el siguiente:

$$\text{TUR} = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{PAR} + \beta_3 \text{PIB} + \beta_4 \text{GTOT} + \beta_5 \text{GTOM} + \beta_6 \text{GTOMD} + U$$

Se realiza de nuevo la Matriz de Correlación. Al igual que en la matriz inicial se marcan los valores que superan el límite de $|R_{ii}| \geq 0,7$. Valores que indican que continúan existiendo problemas de Multicolinealidad.

Tabla 3.37 Matriz de Correlación 8ª de una variable X frente otra X. (Eliminando del modelo la variable NEAI)

	GTOM	GTOMD	GTOT	IPC	PAR	PIB
GTOM	1	0,16	0,18	-0,33	0,26	-0,20
GTOMD	0,16	1	0,30	0,07	0,18	0,78
GTOT	0,18	0,30	1	-0,04	0,61	0,09
IPC	-0,33	0,07	-0,04	1	-0,18	0,39
PAR	0,26	0,18	0,61	-0,18	1	-0,13
PIB	-0,20	0,78	0,09	0,39	-0,13	1

Se vuelve a calcular la matriz inversa de correlaciones (R1)⁻¹, para asegurar que variables se deberían de eliminar del modelo.

Tabla 3.38 Matriz Inversa de Correlación 8ª de una variable X frente el resto de X. (Eliminando del Modelo la variable NEAI)

GTOM	1,41	-0,97	-0,04	0,15	-0,01	0,97
GTOMD	-0,97	4,37	-0,30	0,74	-0,73	-3,93
GTOT	-0,04	-0,30	1,69	-0,08	-0,98	-0,02
IPC	0,15	0,74	-0,08	1,41	-0,02	-1,09
PAR	-0,01	-0,73	-0,98	-0,02	1,84	0,89
PIB	0,97	-3,93	-0,02	-1,09	0,89	4,77

De nuevo se calcula el Índice de Acondicionamiento mediante la raíz del autovalor máximo entre el autovalor mínimo.

$$IC = \sqrt{\frac{\text{Autovalor Máximo}}{\text{Autovalor Mínimo}}} = \sqrt{\frac{1,99438}{0,106592}} = 4,32555312$$

Tabla 3.39 Autovalores 8ª para realizar el Índice de Acondicionamiento.

	AUTOVALOR	IC
		4,32555312
1	1,99438 MÁXIMO	
2	1,91357	
3	0,99226	
4	0,626159	
5	0,367039	
6	0,106592 MÍNIMO	

Según el resultado del IC se puede indicar que para este modelo no existen problemas de multicolinealidad, ya que el valor del IC es inferior a 10.

Tras volverse a realizar las tres pruebas por octava vez (Matriz de Correlación, Matriz Inversa de Correlación e Índice de Acondicionamiento), que muestran los problemas de Multicolinealidad, se puede observar que siguen existiendo problemas como muestra la Matriz de Correlaciones.

Por lo que debe de eliminarse otra variable dependiente del modelo, ya que, existe relación entre los datos. La variable a eliminar del modelo en este caso va a ser la que tenga mayor valor en la diagonal de la Matriz Inversa de Correlaciones, será la variable Gasto Medio Diario de los turistas según destino principal (GTOMD).

El modelo teórico propuesto al eliminar la variable GTOMD es el siguiente:

$$\text{TUR} = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{PAR} + \beta_3 \text{PIB} + \beta_4 \text{GTOT} + \beta_5 \text{GTOM} + U$$

Se realiza de nuevo la Matriz de Correlación. Al igual que en la matriz inicial se marcan los valores que superan el límite de $|R_{ij}| \geq 0,7$. Valores que indican que continúan existiendo problemas de Multicolinealidad.

Tabla 3.40 Matriz de Correlación 9ª de una variable X frente otra X. (Eliminando del modelo la variable GTOMD)

	GTOM	GTOT	IPC	PAR	PIB
GTOM	1	0,18	-0,33	0,26	-0,20
GTOT	0,18	1	-0,04	0,61	0,09
IPC	-0,33	-0,04	1	-0,18	0,39
PAR	0,26	0,61	-0,18	1	-0,13
PIB	-0,20	0,09	0,39	-0,13	1

Se vuelve a calcular la matriz inversa de correlaciones (R1)-1, para asegurar que variables se deberían de eliminar del modelo.

Tabla 3.41 Matriz Inversa de Correlación 9ª de una variable X frente el resto de X. (Eliminando del Modelo la variable GTOMD)

GTOM	1,19	-0,10	0,32	-0,18	0,10
GTOT	-0,10	1,67	-0,03	-1,03	-0,29
IPC	0,32	-0,03	1,28	0,11	-0,42
PAR	-0,18	-1,03	0,11	1,72	0,23
PIB	0,10	-0,29	-0,42	0,23	1,24

De nuevo se calcula el Índice de Acondicionamiento mediante la raíz del autovalor máximo entre el autovalor mínimo.

$$IC = \sqrt{\frac{\text{Autovalor Máximo}}{\text{Autovalor Mínimo}}} = \sqrt{\frac{1,91831}{0,355723}} = 2,32222033$$

Tabla 3.42 Autovalores 9ª para realizar el Índice de Acondicionamiento.

	AUTOVALOR	IC
1	1,91831 MÁXIMO	2,32222033
2	1,39209	
3	0,75215	
4	0,581728	
5	0,355723 MÍNIMO	

Según el resultado del IC se puede indicar que para este modelo no existen problemas de multicolinealidad, ya que el valor del IC es inferior a 10.

Tras volverse a realizar las tres pruebas por novena vez (Matriz de Correlación, Matriz Inversa de Correlación e Índice de Acondicionamiento), que muestran los problemas de Multicolinealidad, se puede observar que ya no existe el problema.

Por tanto, se concluye que tras eliminar la variable Gasto Medio Diario de los turistas según destino principal (GTOMD) del modelo, este queda libre de los problemas de la Multicolinealidad.

Las variables que quedan para hacer el modelo definitivo son GTOM (Gasto Medio de los turistas según destino principal), GTOT (Gasto Total de los turistas según destino principal), IPC (Índice de Precios al Consumo), PAR (número de parados), PIB (Producto Interior Bruto).

3.3 Análisis Multivariante

3.3.1 Modelo teórico propuesto

El modelo teórico ha quedado finalmente de la siguiente manera:

$$\text{TUR} = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{PAR} + \beta_3 \text{PIB} + \beta_4 \text{GTOT} + \beta_5 \text{GTOM} + U$$

A continuación, a través del Statgraphics se va a ajustar el modelo mediante un análisis de regresión múltiple (más de una variable independiente), para saber cómo influyen las variables X en la Y.

3.3.2 Ajuste del Modelo

Seguidamente, se muestra el modelo de Regresión lineal múltiple para describir la relación entre la variable Entradas de turistas por Comunidad autónoma (TUR=Y) y las 5 variables independientes. (X)

$$\text{TUR} = -6,41343\text{E}7 - 872,808 \cdot \text{GTOM} + 1054,55 \cdot \text{GTOT} + 632989 \cdot \text{IPC} - 473,096 \cdot \text{PAR} - 21,0081 \cdot \text{PIB}$$

R-cuadrada = 99,37 %

El coeficiente de la R-cuadrada indica la proporción de ajuste que existe en el modelo propuesto. A mayor valor, mejor explican las variables independientes (X) la variable dependiente (Y). Por lo que en este caso el ajuste del modelo es muy elevado.

3.3.3 Validación del Modelo

A continuación, se van a realizar 10 pruebas para analizar la validación del modelo.

3.3.3.1 Significación

Es importante tener en cuenta la bondad de ajuste del modelo. Este análisis se hará a través del contraste de Hipótesis, donde se plantea la hipótesis nula (H_0) y la hipótesis alternativa (H_1), para aceptar una de las dos debe tenerse en cuenta el p-valor del modelo y de cada una de las variables.

- *Modelo:*

Análisis de Varianza

Tabla 3.43 Contraste del Modelo según el coeficiente de la Variable Explicativa

Fuente	Suma de Cuadrados	Razón-F	Valor-P
Modelo	4,28999E14	345,84	0,0000

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

$$H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq \beta_5 \neq 0$$

P-valor < α (0,05) \rightarrow Se rechaza H_0

Se tiene en cuenta un $\alpha=0,05$

Como el p-valor del modelo es menor que α entonces se rechaza H_0 por lo que al menos alguna β_i es distinta a cero. Esto indica que el modelo es significativo.

- *Contraste significación individual (parámetros):*

Tabla 3.44 Parámetros del Modelo

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	-6,41343*10 ⁷	2,19142*10 ⁷	-2,92662	0,0138
GTOM	-872,808	612,477	-1,42504	0,1819
GTOT	1054,55	31,9257	33,0314	0,0000
IPC	632989,	212290,	2,98171	0,0125
PAR	-473,096	469,268	-1,00816	0,3350
PIB	-21,0081	30,3746	-0,691635	0,5035

Contraste $\beta_0 = \text{CONSTANTE}$

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

P-valor = 0,0138 como es $< \alpha \rightarrow$ Rechazamos H_0 . Se concluye que la constante es significativa para el modelo.

Contraste β_1 : GTOM

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

P-valor = 0,1819 como es $> \alpha \rightarrow$ Aceptamos H_0 . Se concluye que la variable GTOM no es significativa para el modelo.

Contraste β_2 : GTOT

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

P-valor = 0,0000 como es $< \alpha \rightarrow$ Rechazamos H_0 . Se concluye que la variable GTOT es significativa para el modelo.

Contraste β_3 : IPC

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

P-valor = 0,0125 como es $< \alpha \rightarrow$ Rechazamos H_0 . Se concluye que la variable IPC es significativa para el modelo.

Contraste β_4 : PAR

$$H_0 : \beta_4 = 0$$

$$H_1 : \beta_4 \neq 0$$

P-valor = 0,3350 como es $>\alpha \rightarrow$ Aceptamos H_0 . Se concluye que la variable PAR no es significativa para el modelo.

Contraste β_5 : PIB

$$H_0 : \beta_5 = 0$$

$$H_1 : \beta_5 \neq 0$$

P-valor = 0,5035 como es $>\alpha \rightarrow$ Aceptamos H_0 . Se concluye que la variable PIB no es significativa para el modelo.

3.3.3.2 Normalidad de Residuos

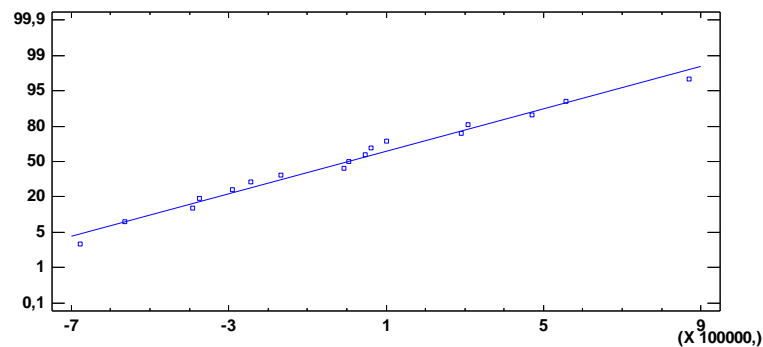


Ilustración 3.29. Gráfico Probabilístico (Probabilidad Normal)

Como se puede observar en la **(Ilustración 3.29)** los residuos se distribuyen muy cerca de la línea de normalidad, por lo que, aparentemente los datos siguen una distribución normal.

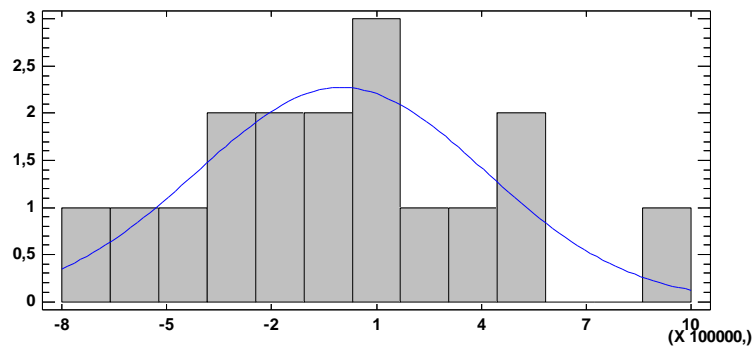


Ilustración 3.30 Histograma de Residuos

Como se puede observar en la **Ilustración 3.30** el Histograma no tiene una forma de campana de Gauss perfecta, ya que se puede identificar que a la parte derecha existen residuos que no se encuentran dentro de la distribución normal.

Aunque debe de indicarse que la mayoría de los datos se concentran en la parte media del gráfico. Según este análisis se podría indicar que la los residuos siguen una distribución normal.

Para poder asegurar la normalidad de los residuos se efectuará el contraste de hipótesis siguiente:

H_0 : Los residuos se distribuyen normalmente.

H_1 : Los residuos no se distribuyen normalmente.

Teniendo en cuenta que si $p\text{-valor} > \alpha (0,05) \rightarrow$ Acepto H_0

Para ver que hipótesis se acepta se van a realizar todos los Test para la normalidad de los residuos. Se muestran a continuación:

Pruebas de Normalidad para RESIDUOS

Tabla 3.45 Resumen de Test de Normalidad de Residuos

Prueba	Estadístico	Valor-P
Chi-Cuadrado	6,29412	0,710159
Estadístico W de Shapiro-Wilk	0,98146	0,951878
Valor-Z para asimetría	0,471084	0,637578
Valor-Z para curtosis	Datos Insuficientes	

De todos los Test que aparecen en la **Tabla 3.45** se elige el Test más restrictivo teniendo en cuenta el p-valor. En este caso el test más restrictivo es el Test de Valor-Z para asimetría.

Como el p-valor del test más restrictivo (0,637578).es mayor que $\alpha=0,05$, se acepta la Hipótesis nula (H_0), por lo que, se concluye tras todos los análisis que los residuos se distribuyen normalmente.

3.3.3.3 Heteroscedasticidad

A continuación se realiza un contraste de hipótesis para comprobar si existe heteroscedasticidad.

$H_0: \sigma^2$ es constante = 0 → No existe heteroscedasticidad

$H_1: \sigma^2$ no es constante $\neq 0$ → Existe heteroscedasticidad

Para comprobar si existe heteroscedasticidad se plantea un nuevo modelo que queda de la siguiente manera:

$$\text{RESIDUOS}^2 = \beta_0 + \beta_1 \text{IPC} + \beta_2 \text{PAR} + \beta_3 \text{PIB} + \beta_4 \text{GTOT} + \beta_5 \text{GTOM} + U$$

A continuación, a través del Statgraphics, se van a estudiar las variables del modelo para comprobar si existe o no heteroscedasticidad. Los resultados se muestran en la **Tabla 3.46**:

Tabla 3.46 P-valor de las variables para el análisis de heteroscedasticidad

Parámetro	P-Valor
CONSTANTE	0,9697
GTOM	0,7241
GTOT	0,6252
IPC	0,9858
PAR	0,8263
PIB	0,5571

Tras realizar la regresión múltiple del nuevo modelo propuesto, se puede observar que todos los p-valor de las variables explicativas y de la constante son mayores que $\alpha(0,05)$, y que por tanto la σ^2 es constante. Es decir que se acepta H_0 y se puede indicar que el modelo está libre de heteroscedasticidad, ya que, ninguna de las variables está produciendo este problema.

3.3.3.4 Autocorrelación

El análisis de autocorrelación se va a realizar con las 2 pruebas siguientes:

1. Test de Durbin Watson
2. FAS y FAP

Contraste de hipótesis para el Test de Durbin-Watson:

En el caso de este análisis, solamente se detecta autocorrelación de 1^{er} orden:

$H_0: \rho_1=0$ no hay autocorrelación de 1^{er} orden

$H_1: \rho_1 \neq 0$ hay autocorrelación de 1^{er} orden.

Si p-valor > α (0,05) → Acepto H_0

El p-valor del estadístico de Durbin Watson es 0,2848

En este caso, como el p-valor es superior a α , se puede aceptar la H_0 por lo que, se confirma que no existe autocorrelación de 1^{er} orden.

Seguidamente, para poder confirmar que no existen problemas de autocorrelación de otro orden en el modelo, se van a realizar los gráficos de autocorrelación (FAS y FAP).

FAS Y FAP

- FAS

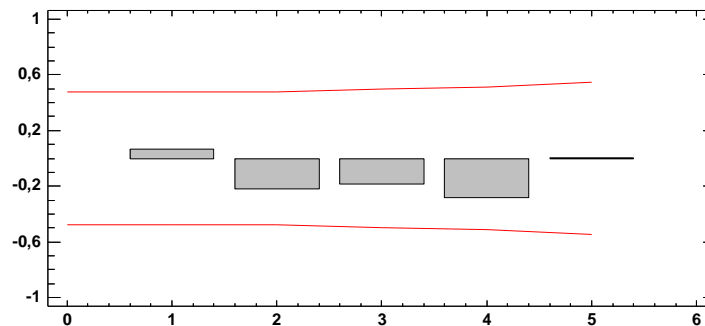


Ilustración 3.31. Función de Autocorrelación (FAS)

- FAP

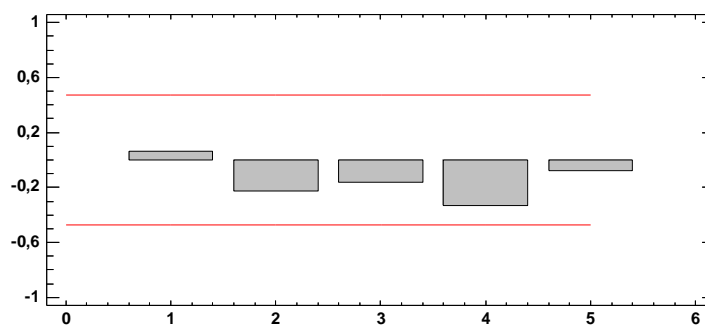


Ilustración 3.32 Función de Autocorrelación Parcial (FAP)

Como se puede observar en la **Ilustración 3.31 y 3.32** el FAS y el FAP muestran que no existe autocorrelación de ningún orden. Ningún factor de autocorrelación supera los límites por lo que se puede confirmar que no existe el problema.

3.3.3.5 Puntos Influyentes

Indicar que existen Puntos influyentes a Priori y a Posteriori, los analizaremos en el orden que se acaba de indicar.

Tabla 3.47 Tabla que muestra los puntos influyentes

Comunidad Autónoma	Fila	Influencia	Distancia de Mahalanobis	DFITS
Andalucía	1	0,739942	41,7419	-2,64498
Canarias	5	0,699082	33,9099	-3,83928
Cataluña	9	0,600854	21,6428	1,96659
Comunidad de Madrid	13	0,559009	18,0768	-1,35708

Influencia media de un solo punto (\bar{n}) = **0,352941**

PUNTOS INFLUYENTES A PRIORI

Los puntos influyentes a priori serán aquellos que nos indiquen si los puntos se alejan mucho de la media de "x". Influyen modificando el valor de los parámetros que acompañan a las variables explicativas (B_i).

$$n_{ii} > 2\bar{n}$$

Siendo $\bar{n} = 0,352941$, por lo que $2\bar{n} = 2 * 0,352941 = 0,705882$

Tabla 3.48 Resumen Puntos Influyentes a Priori

Comunidad Autónoma	Fila	$n_{ii} > 2\bar{n}$
Andalucía	1	$0,739942 > 2\bar{n}$
Canarias	5	$0,699082 < 2\bar{n}$
Cataluña	9	$0,600854 < 2\bar{n}$
Comunidad de Madrid	13	$0,559009 < 2\bar{n}$

Por lo tanto, se concluye que Andalucía es un punto influyente a Priori.

PUNTOS INFLUYENTES A POSTERIORI

Los puntos influyentes a posteriori nos indicarán como varía la pendiente del modelo. Influyen modificando el valor de la constante (B_0).

$$|DFITS| \geq 2\sqrt{\bar{n}} \rightarrow 2\sqrt{0,352941} = 1,18817675$$

Tabla 3.49 Resumen Puntos Influyentes a Posteriori

Comunidad Autónoma	Fila	$ DFITS $
Andalucía	1	$2,64498 > 2\sqrt{\bar{n}}$
Canarias	5	$3,83928 > 2\sqrt{\bar{n}}$
Cataluña	9	$1,96659 > 2\sqrt{\bar{n}}$
Comunidad de Madrid	13	$1,35708 > 2\sqrt{\bar{n}}$

Como se puede observar, Andalucía, Canarias, Cataluña y La Comunidad de Madrid son puntos influyentes a posteriori.

3.3.3.6 Residuos atípicos o puntos anómalos

Los puntos anómalos serán los encargados de medir si el punto genera un residuo muy grande, y por tanto, si influye en el resultado del modelo.

Los puntos, se identificarán a través del residuo estudentizado en valor absoluto y guiándose a partir de esta tabla:

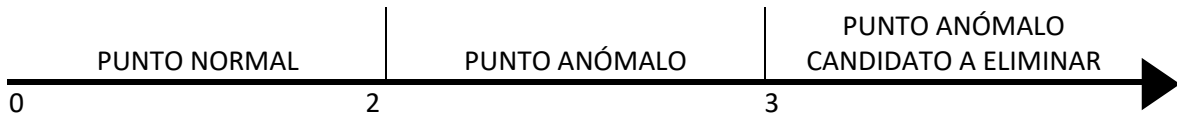


Ilustración 3.33 Muestra los rangos de los puntos candidatos a eliminar o no del modelo

En este caso los resultados de los residuos atípicos son los siguientes:

Tabla 3.50 Residuos Estudentizados

Comunidad Autónoma	Fila	Y	Y Predicha	Residuo	Residuo Estudentizado
Canarias	5	1,1475*10 ⁷	1,20396*10 ⁷	-564604	-2,52
Comunidad Valenciana	10	6,22608*10 ⁶	5,35512*10 ⁶	870964	2,18

Tabla 3.51 Residuos Estudentizados en valor absoluto

Comunidad Autónoma	Fila	Residuo Estudentizado
Canarias	5	2,52
Comunidad Valenciana	10	2,18

Por lo que al analizar el residuo estudentizado, se concluye que de la fila 5 correspondiente a Canarias y la fila 10 correspondiente a la Comunidad Valenciana son puntos anómalos del modelo. En este caso, como no tienen un valor absoluto superior a 3, no son puntos anómalos candidatos a eliminar del modelo.

Por lo que no se elimina ninguna Comunidad Autónoma (fila) del modelo. A continuación, se mostrarán los problemas que han aparecido a lo largo de los análisis de validación del modelo (3.3.3).

En el único punto en el que han aparecido problemas ha sido en el apartado **3.3.3.1 Significación**. En este punto, ha habido problemas en la significatividad de algunas de las variables del modelo. Las variables que han dado problemas por tener un p-valor > α (0,05) son:

β_1 : GTOM

GTOM no es significativa para el modelo.

β_4 : PAR

PAR no es significativa para el modelo.

β_5 : PIB

PIB no es significativa para el modelo.

Para eliminar las variables no significativas del modelo y así, poder obtener el modelo definitivo, se hará un análisis a través de Statgraphics que se define como el método selección paso a paso hacia adelante.

Se trata de un método de selección por pasos que selecciona la entrada basándose en la significación de la variable, es decir, se introducen las variables significativas de forma individual y por orden de importancia (de mayor a menor), y no se introducen las no significativas.

El Statgraphics tiene en cuenta el valor del CME (Cuadrado Medio Explicado) para definir el modelo. La primera variable que entra en el modelo es la que tiene el CME más grande y así sucesivamente.

En este caso se han eliminado las variables GTOM, PAR y PIB. El modelo definitivo y el R-cuadrado han quedado de la forma siguiente:

$$\text{TUR} = -7,32038 \cdot 10^7 + 1026,44 \cdot \text{GTOT} + 707827 \cdot \text{IPC}$$

R-cuadrada = 99,16 %

Tabla 3.52 P-valor de los parámetros del modelo

Parámetro	Estimación	Error Estándar	T-Estadístico	P-Valor
CONSTANTE	$-7,32038 \cdot 10^7$	$1,9809 \cdot 10^7$	-3,69548	0,0024
GTOT	1026,44	25,2605	40,6341	0,0000
IPC	707827	191457	3,69706	0,0024

Tabla 3.53 Análisis de la varianza del modelo.

Fuente	Razón-F	P-Valor
Modelo	827,42	0,0000

Como se puede observar en las tablas anteriores (**Tabla 3.52 y 3.53**), los p-valor tanto del modelo como de las variables, son inferiores a α , por lo que rechazamos H_0 y se concluye que la constante, las variables y el modelo son significativos. Siguiendo este proceso se observa que la variable más importante es GTOT (primera en introducirse) y a continuación el IPC.

3.4 Predicciones

Para poder comprobar si las variables independientes (X) explican de forma correcta la variable dependiente (Y) y saber si el modelo definitivo es “bueno”, se van a realizar 3 predicciones modificando los datos iniciales de una de las Comunidades Autónomas.

En este caso, aprovechando que vivimos en la Comunidad Valenciana, los datos a modificar para hacer dichas predicciones, van a ser los de la comunidad indicada.

1. Predicción donde utilizamos el mismo dato de la tabla inicial para la variable GTOT.

Siguiendo la ruta correspondiente en el programa Statgraphics, se muestra a través de la siguiente tabla la predicción puntual. La cual se comparará con el límite inferior y superior del intervalo para saber si el modelo económico es bueno o no.

Tabla 3.54 Predicción donde utilizamos el mismo dato de la tabla inicial para la variable GTOT.

Parámetro	Predicción Puntual (Ajustado)	Error Est.(LC para Pronóstico)	Pronóstico Inferior	Pronóstico Superior
GTOT	$5,28651 \cdot 10^6$	529276	$4,15132 \cdot 10^6$	$6,42169 \cdot 10^6$

Según muestra la **Tabla 3.54** la predicción puntual ($5,28651 \cdot 10^6$), se encuentra dentro del Intervalo [$4,15132 \cdot 10^6 - 6,42169 \cdot 10^6$] del pronóstico. Por lo que según esta predicción, se puede indicar que el modelo se ajusta bien y es bastante bueno.

2. Predicción donde la variable IPC aumenta y GTOT disminuye.

Siendo el modelo definitivo de la forma siguiente, se podría indicar si aumentan una de las 2 variables independientes definitivas (IPC o GTOT), el valor de TUR (variable dependiente) aumentará, ya que, las variables van acompañadas por un valor β_i positivo. En el caso que disminuyan de valor, el valor de Y también debería disminuir.

$$TUR = -7,32038 \cdot 10^7 + 1026,44 \cdot GTOT + 707827 \cdot IPC$$

Para comprobar estas predicciones se va a utilizar el **Statgraphics** modificando los siguientes datos:

- Aumenta el IPC de 103,083 a 105:

Tabla 3.55 Predicción donde la variable IPC aumenta.

Parámetro	Predicción Puntual (Ajustado)	Error Est.(LC para Pronóstico)	Pronóstico Inferior	Pronóstico Superior
IPC ↑	$6,64341 \cdot 10^6$	604970	$5,34588 \cdot 10^6$	$7,94094 \cdot 10^6$

Según la **Tabla 3.55** se puede observar que la predicción puntual ha aumentado a causa de un aumento en el IPC. El valor se encuentra dentro del intervalo por lo que se considera que tras el cambio, el modelo sigue siendo bueno.

- Disminuye el GTOT de 5383,09 a 4800:

Tabla 3.56 Predicción donde la variable GTOT disminuye.

Parámetro	Predicción Puntual (Ajustado)	Error Est.(LC para Pronóstico)	Pronóstico Inferior	Pronóstico Superior
GTOT ↓	$4,688 \cdot 10^6$	528386	$3,55472 \cdot 10^6$	$5,82128 \cdot 10^6$

Según la **Tabla 3.56** se puede observar que la predicción puntual ha disminuido a causa de una disminución en el GTOT. El valor se encuentra dentro del intervalo por lo que se considera que tras el cambio, el modelo sigue siendo bueno.

3. Predicción donde se tienen en cuenta todas las Comunidades Autónomas.

A continuación, se va a realizar la predicción puntual de todas las Comunidades Autónomas para comprobar si se encuentran dentro del mínimo y máximo del intervalo.

Tabla 3.57 Predicción donde se tienen en cuenta todas las Comunidades Autónomas.

CCAA	TUR (Y)	Límite Inferior para las Predicciones	Límite Superior para las Predicciones	¿Está dentro del Intervalo?
Andalucía	8498724	8026300	10390800	OK
Aragón	326690	-1142890	1162320	OK
Asturias, Principado de	234489	-913825	1365650	OK
Balears, Illes	11348259	9808730	12210100	OK
Canarias	11474965	10421100	13111300	OK
Cantabria	369964	-172445	2224620	OK
Castilla - La Mancha	182606	-1309190	1016540	OK
Castilla y León	995596	-54967	2249740	OK
Cataluña	16793818	14980700	17738300	OK
Comunidad Valenciana	6226081	4151320	6421690	OK
Extremadura	196265	-1437340	892590	OK
Galicia	1048878	31438	2311960	OK
Madrid, Comunidad de	4545820	4285580	6551850	OK
Murcia, Región de	805365	-248556	2017790	OK
Navarra, Comunidad Foral	261460	-1096950	1188390	OK
País Vasco	1568945	266081	2591170	OK
Rioja, La	61020	-736808	1574740	OK

Según la **Tabla 3.57** se puede observar que la predicción puntual está dentro del intervalo en todas las Comunidades Autónomas (OK).

Por lo que, se considera un modelo bueno y con buen ajuste tras realizar distintas predicciones en el modelo definitivo.

4. Conclusiones y Futuras líneas de investigación

4.1 Conclusiones

- Se ha conseguido un modelo econométrico donde la variable *Entradas de turistas por Comunidad autónoma (TUR=Y)* se explica en función de las variables independientes Gasto Total de los turistas según destino principal (GTOT, en millones de Euros) y según el Índice de Precios al Consumo (IPC).
- El modelo que se ha conseguido tras todos los pasos explica un $99,16\%$ (R^2) de los datos. Se puede considerar un ajuste muy bueno.
- Las variables independientes Gasto Total de los turistas según destino principal (GTOT) y el Índice de Precios al Consumo (IPC) influyen de manera significativa en la entrada de turistas según Comunidad Autónoma (TUR).
- El resto de variables independientes (X) no influyen de manera significativa en la entrada de turistas según Comunidad Autónoma (TUR).
- Según el estudio de multicolinealidad, existe una fuerte relación entre las variables independientes (X) planteadas inicialmente. Por este motivo se han eliminado múltiples variables en este paso dejando como modelo definitivo para realizar la *Regresión Lineal Múltiple* las variables GTOM (Gasto Medio de los turistas según destino principal), GTOT (Gasto Total de los turistas según destino principal), IPC (Índice de Precios al Consumo), PAR (número de parados), PIB (Producto Interior Bruto).
- Tras realizar las distintas pruebas del análisis multivariante y tener problemas en la significación de las variables GTOM (Gasto Medio de los turistas según destino principal), PAR (número de parados), PIB (Producto Interior Bruto) que tenían un p-valor $> \alpha$ (0,05), se utiliza un método de selección por pasos para eliminar las variables no significativas del modelo. En este caso se eliminan las 3 variables y quedan como variables definitivas GTOT e IPC, sobre las cuales se realizan las predicciones.
- Aunque existen puntos atípicos e anómalos en el modelo, no se han eliminado variables de los datos ya que el comportamiento no influía de forma directa.
- Tras realizar las predicciones con los mismos valores que proporciona el INE y el DIRCE, las predicciones puntuales obtenidas a través del *Statgraphics* se encuentran dentro del mínimo y el máximo del Intervalo de Confianza.
- Además, modificando los valores de las variables independientes del modelo se puede indicar que la predicción puntual sigue encontrándose dentro del Intervalo de Confianza. Por lo que las predicciones indican que el modelo definitivo es bueno y se ajusta bien.

4.2 Futuras Líneas de Investigación

En este proyecto se han seguido todos los pasos indicados en el punto 2. Metodología y Técnicas.

Sería interesante generar un nuevo enfoque en el estudio. Por lo que a continuación, se van a indicar nuevos procedimientos de trabajo con los que se podría completar el análisis realizado.

Los procedimientos a realizar podrías ser:

- Realizar este mismo proyecto mediante todas las estructuras posibles que son la semilogarítmica, logarítmica, inversa y logarítmica inversa, ya que, en este análisis solamente se ha realizado la estructura lineal.
- Realizar el mismo proyecto con datos de años posteriores que estén más actualizados. Como se ha indicado inicialmente en este caso los datos más recientes eran los que se han utilizado. Por lo que, se podría hacer un análisis recopilando datos recientes para comprobar que las predicciones realizadas son correctas.
- Resolver la multicolinealidad mediante el método de componentes principales, ya que, se puede obtener información más completa que la que aparece en el método de eliminación de variables (método utilizado en este proyecto).
La información resultante del método de componentes principales podría llegar a ser importante para el modelo y podrías mostrar más datos sobre el porqué se eliminan las variables
- Se podría añadir más variables que se consideren influyentes a la hora de explicar la variable *Entradas de turistas por Comunidad autónoma (TUR=Y)*. Variables como por ejemplo la *Superficie de terreno* que posee cada una de las Comunidades Autónomas.

5. Bibliografía

- [1] (unwto.org) <http://media.unwto.org/es/content/entender-el-turismo-glosario-basico> → 09/05/2017
- [2] <http://origin.magrama.gob.es/es/cambio-climatico/temas/mitigacion-politicas-y-medidas/agricola.aspx> → 09/05/2017
- [3] http://recurso+stic.educacion.es/secundaria/edad/3esohistoria/quincena10/quincena10_contenidos_3d.htm → 18/06/2017
- [4] Producto e información a partir de <http://www.origenespana.es/> → 18/06/17
- [5] <http://www.ine.es/jaxi/Datos.htm?path=/t35/a122/a1998/l0/&file=01001.px> → 06/05/17
- [6] Apuntes economía española (Tema 5) Antonio Benito Benito, 2015
- [7] <http://www.ine.es/jaxi/Datos.htm?path=/t35/p010/base2010/homoge/l0/&file=01001.px> → 06/05/11
- [8] <http://www.ine.es/infografias/turismo/desktop/panel.html?t=4&lang=es> → Datos Turistas residentes. → 04/05/17
- [9] <https://www.medwave.cl/link.cgi/Medwave/Series/MBE04/4934> → 25/05/2017
- [10] Martínez M, Ruiz R, Vallada E. <<Introducción a la Estadística>>. En: Universidad Politécnica de Valencia, Valencia, España.
- [11] https://www.ecured.cu/Medidas_de_dispersi%C3%B3n → 27/06/2017
- [12] <http://www.universoformulas.com/estadistica/descriptiva/asimetria-curtosis/> → 27/06/2017
- [13] <http://www.estadisticaparatodos.es/taller/graficas/cajas.html> → 30/05/2017
- [14] Novales A, (2010). <<Econometría>>. En: McGraw-Hill, D.L 1993, 2010. Madrid, España.
- [15] Chirivella V, (2005). <<Econometría>>. En: P.I.D 12059-C, Valencia, España.
- [16] <http://estadisticas.tourspain.es/es-ES/Paginas/default.aspx> → 01/03/2017
- [17] <http://www.ine.es/> → 01/03/2017

LISTADO DE ILUSTRACIONES:

Ilustración 1.1 Balanza de Pagos 2013. Fuente INE (Instituto Nacional de Estadística)	3
Ilustración 1.2 Resumen Funciones Sistema Financiero.....	4
Ilustración 1.3 Producto Interior Bruto (Base 2010) Fuente INE (Instituto Nacional de Estadística)	5
Ilustración 1.4 Número de Viajes por CCAA (1er Trimestre 2015), Fuente INE (Instituto Nacional de Estadística)	5
Ilustración 2.1 Resumen de Procesos.....	10
Ilustración 2.2 Según la Asimetría, Tipos de Curtosis.....	12
Ilustración 2.3 Tipos de Asimetría.....	13
Ilustración 2.4 Gráfico de Caja y Bigotes.....	13
Ilustración 2.5 Relación perfecta. No Multicolinealidad	14
Ilustración 2.6 Relación entre variable X-Y	14
Ilustración 2.7 Papel Probabilístico Normal (Plot de Daniel)	18
Ilustración 2.8 Histograma	18
Ilustración 2.9 Función de Autocorrelación Simple	22
Ilustración 2.10 Función de Autocorrelación Parcial.....	22
Ilustración 2.11 Función de Autocorrelación Simple	22
Ilustración 2.12 Representación de puntos atípicos en el gráfico de normalidad de los residuos.	23
Ilustración 2.13 Muestra los rangos de los puntos candidatos a eliminar o no del modelo	24
Ilustración 3.1. Elaboración propia a partir de los apuntes de la asignatura métodos Estadísticos en Economía.....	26
Ilustración 3.2. Gráfico de Caja y Bigotes de las entradas de turistas según Comunidad Autónoma en España en el año 2014.....	28
Ilustración 3.3. Gráfico de Caja y Bigotes de la población total existente en cada Comunidad Autónoma en España en el año 2014.....	30
Ilustración 3.4. Gráfico de Caja y Bigotes del IPC (Índice General) de cada Comunidad Autónoma en España en el año 2014.....	32
Ilustración 3.5. Gráfico de Caja y Bigotes de Comercio al por menor, excepto de vehículos de motor y motocicletas en cada Comunidad Autónoma de España en el año 2014.....	33
Ilustración 3.6 Gráfico de Caja y Bigotes de Transporte terrestre y por tubería; Transporte marítimo y por vías navegables interiores y Transporte aéreo en cada Comunidad Autónoma de España en el año 2014	34
Ilustración 3.7. Gráfico de Caja y Bigotes de Servicios de alojamiento en cada Comunidad Autónoma de España en el año 2014.....	35
Ilustración 3.8. Gráfico de Caja y Bigotes de Actividades inmobiliarias y Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con los mismos en cada Comunidad Autónoma de España en el año 2014.....	36
Ilustración 3.9 Gráfico de Caja y Bigotes de Parados (miles de personas) en cada Comunidad Autónoma de España en el año 2014.....	37
Ilustración 3.10 Gráfico de Caja y Bigotes de Ocupados (miles de personas) en cada Comunidad Autónoma de España en el año 2014.....	38
Ilustración 3.11 Gráfico de Caja y Bigotes del Crecimiento del PIB per cápita (euros) en cada Comunidad Autónoma de España en el año 2014.....	39
Ilustración 3.12. Gráfico de Caja y Bigotes del Gasto Total (millones de euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014.....	40
Ilustración 3.14. Gráfico de Caja y Bigotes del Gasto Total Medio Diario (euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014.....	42
Ilustración 3.15 Gráfico de Caja y Bigotes de la Estancia Media (noches) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014	43
Ilustración 3.16 Gráfico X-Y para la PTOTAL.	44
Ilustración 3.17 Gráfico X-Y para la IPC.	45
Ilustración 3.18 Gráfico X-Y para la NEC.	46
Ilustración 3.19 Gráfico X-Y para la NETR.	47
Ilustración 3.20 Gráfico X-Y para la NES.	48
Ilustración 3.21 Gráfico X-Y para la NEAI.....	49
Ilustración 3.22 Gráfico X-Y para la PAR.	50

Ilustración 3.23 Gráfico X-Y para la OCU	51
Ilustración 3.24 Gráfico X-Y para la PIB	52
Ilustración 3.25 Gráfico X-Y para la GTOT.....	53
Ilustración 3.26 Gráfico X-Y para la GTOM.....	54
Ilustración 3.27 Gráfico X-Y para la GTOMD.....	55
Ilustración 3.28 Gráfico X-Y para la EM.....	56
Ilustración 3.29. Gráfico Probabilístico (Probabilidad Normal)	82
Ilustración 3.30 Histograma de Residuos.....	82
Ilustración 3.31. Función de Autocorrelación (FAS)	85
Ilustración 3.32 Función de Autocorrelación Parcial (FAP)	85
Ilustración 3.33 Muestra los rangos de los puntos candidatos a eliminar o no del modelo	88

LISTADO DE TABLA:

Tabla 3.1 Tabla de Base de Datos del Ine y Dirce de las variables a analizar en el proyecto.....	27
Tabla 3.2 Resumen de estadísticos para la entrada de turistas según Comunidad Autónoma en España en el año 2014.....	28
Tabla 3.3. Resumen de estadísticos para la población total según Comunidad Autónoma en España en el año 2014.....	30
Tabla 3.4 Resumen de estadísticos para el IPC según Comunidad Autónoma en España en el año 2014.....	32
Tabla 3.5. Resumen de estadísticos para Comercio al por menor, excepto de vehículos de motor y motocicletas según Comunidad Autónoma en España en el año 2014.....	33
Tabla 3.6 Resumen de estadísticos de Transporte terrestre y por tubería; Transporte marítimo y por vías navegables interiores; Transporte aéreos en cada Comunidad Autónoma en España en el año 2014.....	34
Tabla 3.7 Resumen de estadísticos de Servicios de alojamiento en cada Comunidad Autónoma en España en el año 2014.....	35
Tabla 3.8 Resumen de estadísticos de Actividades inmobiliarias y Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con los mismos en cada Comunidad Autónoma de España en el año 2014.....	36
Tabla 3.9 Resumen de estadísticos de Parados (miles de personas) en cada Comunidad Autónoma de España en el año 2014.....	37
Tabla 3.10 Resumen de estadísticos de Ocupados (miles de personas) en cada Comunidad Autónoma de España en el año 2014.....	38
Tabla 3.11 Resumen de estadísticos del Crecimiento del PIB per cápita (euros) en cada Comunidad Autónoma de España en el año 2014.....	39
Tabla 3.12 Resumen de estadísticos del Gasto Total (millones de euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014.....	40
Tabla 3.13 Resumen de estadísticos del Gasto Total Medio (euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014.....	41
Tabla 3.14 Resumen de estadísticos del Gasto Total Medio Diario (euros) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014.....	42
Tabla 3.15 Resumen de estadísticos de la Estancia Media (noches) de los turistas según destino principal en cada Comunidad Autónoma de España en el año 2014.....	43
Tabla 3.16 Matriz de Correlación 1ª de una variable X frente otra X.....	59
Tabla 3.17 Matriz Inversa de Correlación 1ª de una variable X frente el resto de X.....	61
Tabla 3.18 Autovalores 1ª para realizar el Índice de Acondicionamiento.....	62
Tabla 3.19 Matriz de Correlación 2ª de una variable X frente otra X. (Eliminando del modelo la variable PTOTAL).....	63
Tabla 3.20 Matriz Inversa de Correlación 2ª de una variable X frente el resto de X. (Eliminando del Modelo la variable PTOTAL).....	63
Tabla 3.21 Autovalores 2ª para realizar el Índice de Acondicionamiento.....	64
Tabla 3.22 Matriz de Correlación 3ª de una variable X frente otra X. (Eliminando del modelo la variable NEC).....	65
Tabla 3.23 Matriz Inversa de Correlación 3ª de una variable X frente el resto de X. (Eliminando del Modelo la variable NEC).....	65
Tabla 3.24 Autovalores 3ª para realizar el Índice de Acondicionamiento.....	66
Tabla 3.25 Matriz de Correlación 4ª de una variable X frente otra X. (Eliminando del modelo la variable EM).....	67
Tabla 3.26 Matriz Inversa de Correlación 4ª de una variable X frente el resto de X. (Eliminando del Modelo la variable EM).....	67
Tabla 3.27 Autovalores 4ª para realizar el Índice de Acondicionamiento.....	68
Tabla 3.28 Matriz de Correlación 5ª de una variable X frente otra X. (Eliminando del modelo la variable OCU).....	69
Tabla 3.29 Matriz Inversa de Correlación 5ª de una variable X frente el resto de X. (Eliminando del Modelo la variable OCU).....	69
Tabla 3.30 Autovalores 5ª para realizar el Índice de Acondicionamiento.....	70
Tabla 3.31 Matriz de Correlación 6ª de una variable X frente otra X. (Eliminando del modelo la variable NETR).....	71

Tabla 3.32 Matriz Inversa de Correlación 6ª de una variable X frente el resto de X. (Eliminando del Modelo la variable NETR).....	71
Tabla 3.33 Autovalores 6ª para realizar el Índice de Acondicionamiento.....	72
Tabla 3.34 Matriz de Correlación 7ª de una variable X frente otra X. (Eliminando del modelo la variable NES).....	73
Tabla 3.35 Matriz Inversa de Correlación 7ª de una variable X frente el resto de X. (Eliminando del Modelo la variable NES).....	73
Tabla 3.36 Autovalores 7ª para realizar el Índice de Acondicionamiento.....	74
Tabla 3.37 Matriz de Correlación 8ª de una variable X frente otra X. (Eliminando del modelo la variable NEAI).....	75
Tabla 3.38 Matriz Inversa de Correlación 8ª de una variable X frente el resto de X. (Eliminando del Modelo la variable NEAI).....	75
Tabla 3.39 Autovalores 8ª para realizar el Índice de Acondicionamiento.....	76
Tabla 3.40 Matriz de Correlación 9ª de una variable X frente otra X. (Eliminando del modelo la variable GTOMD).....	77
Tabla 3.41 Matriz Inversa de Correlación 9ª de una variable X frente el resto de X. (Eliminando del Modelo la variable GTOMD).....	77
Tabla 3.42 Autovalores 9ª para realizar el Índice de Acondicionamiento.....	77
Tabla 3.43 Contraste del Modelo según el coeficiente de la Variable Explicativa	79
Tabla 3.44 Parámetros del Modelo.....	80
Tabla 3.45 Resumen de Test de Normalidad de Residuos.....	83
Tabla 3.46 P-valor de las variables para el análisis de heteroscedasticidad.....	84
Tabla 3.47 Tabla que muestra los puntos influyentes.....	87
Tabla 3.48 Resumen Puntos Influyentes a Priori.....	87
Tabla 3.49 Resumen Puntos Influyentes a Posteriori	87
Tabla 3.50 Residuos Estudentizados	88
Tabla 3.51 Residuos Estudentizados en valor absoluto.....	88
Tabla 3.52 P-valor de los parámetros del modelo.....	89
Tabla 3.53 Análisis de la varianza del modelo.	89
Tabla 3.54 Predicción donde utilizamos el mismo dato de la tabla inicial para la variable GTOT.	90
Tabla 3.55 Predicción donde la variable IPC aumenta.	90
Tabla 3.56 Predicción donde la variable GTOT disminuye.	90
Tabla 3.57 Predicción donde se tienen en cuenta todas las Comunidades Autónomas.....	91