



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CAMPUS D'ALCOI

Estudio de las pernoctaciones hoteleras según el turismo de ocio, recreo y vacaciones

MEMORIA PRESENTADA POR:

David Burgada Izquierdo

TUTOR/A:

Jorge Jordán Núñez

GRADO DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Convocatoria de defensa: Julio 2021



RESUMEN EN CASTELLANO

En este trabajo se va a buscar información acerca de los factores que afectan al turismo y pernoctaciones hoteleras en España y se realizará una descripción en función de los datos encontrados. Posteriormente se tratará de modelizar las pernoctaciones mediante la técnica de regresión múltiple. Finalmente se realizarán una serie de supuestos para ver distintos escenarios en las comunidades autónomas españolas.

PALABRAS CLAVE EN CASTELLANO

Turismo; Pernoctaciones; Estadística; Regresión múltiple; Econometría

RESUMEN EN INGLÉS

In this work, information about the factors that affect tourism and hotel stays in Spain will be sought and a description will be made based on the data found. Subsequently, an attempt will be made to model the overnight stays using the multiple regression technique. Finally, a series of assumptions will be made to see different scenarios in the Spanish autonomous communities.

PALABRAS CLAVE EN INGLÉS

Tourism; Overnight stays; Statistics; Multiple regression; Econometrics



ÍNDICE

1. INTRODUCCIÓN Y OBJETIVOS	4
1.1 Motivación	5
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	22
3.1 Univariante	24
3.2 Bivariante	42
3.3 Multivariante	62
3.4 Predicciones	71
4. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	73
4.1 Conclusiones	74
4.2 Futuras Líneas de Investigación	75
5. BIBLIOGRAFÍA	76



1. INTRODUCCIÓN Y OBJETIVOS



1.1 Motivación

El sector turístico en España es uno de los pilares sobre los cuales se basa la economía española, aportando al PIB español más de un 10%.

Además, de ser un tema candente y actual en la sociedad española, España cuenta con el privilegio de ser uno de los países más visitados en todo el mundo, la decisión de tomar el turismo como referente para la realización de este proyecto, viene condicionada también por mis estudios universitarios. Ser alumno del doble grado en Turismo y Administración de empresas, impulsaba mi motivación por lograr la combinación perfecta para poder entender este sector tan importante en España e intentar lograr como la forma de incrementar las pernoctaciones hoteleras en nuestro país, basado en una serie de variables que a priori pueden condicionar a los millones de turistas que acoge nuestro país.

El sector turístico en España ha sufrido una serie de cambios desde los años 60, donde una España marcada por la inestabilidad política del momento quiso darse a conocer para recuperar el terreno perdido sobre los demás países vecinos y europeos.[1]

La apertura de España al exterior supuso un gran crecimiento económico, gracias al Plan de Estabilización (1959), donde esta serie de medidas puso fin a la autarquía y se apostó principalmente por la industria y el sector turístico, favorecido por el buen clima, el turismo era principalmente de sol y playa. Gracias a la llegada masiva de turistas en esos años se pudo invertir en rehabilitar los establecimientos hoteleros, así como favorecer la construcción de muchos más.[1]

En los años 80 se produce el boom turístico en nuestro país, gracias a la transición política donde España daba un gran paso hacia la democracia, hizo más llamativo el territorio español como destino turístico, acompañado también por las medidas adoptadas por los diferentes gobiernos que aportaron una etapa de bonanza económica.[1]

A nivel internacional se produjo una globalización mundial, variable clave para el turismo en España, ya que introdujo el marketing para nuestros destinos principales, favoreciendo así el turismo nacional, despertando el interés por no solo turistas internacionales sino también nacionales.[1]

Desde los años 90, en Europa y en el mundo surgen nuevos destinos que compiten directamente con el turismo que hasta la fecha había reinado en España, el turismo de sol y playa, por eso se decide apostar por otro tipo de turismo como el rural o cultural, ya que nuestro país cuenta con los recursos necesarios como para impulsar ese tipo de turismo.[1]

Con esta arriesgada apuesta se pretende combatir la imagen que España tiene al exterior de venderse como un turismo de poco poder adquisitivo de sol y playa, acto que hace despertar el interés cada día más notable en cuanto al turismo nacional o turismo de interior. Sobre el turismo de sol y playa cada vez tiende a encarecerse ya que no solo se busca la llegada masiva como antaño, la tendencia es competir con otros destinos atractivos en exclusividad o calidad.[1]



Una vez puesto en situación sobre la evolución del turismo en España, se centrará la atención en conocer los alojamientos turísticos más utilizados en el país para albergar a la gran cantidad de turistas que recibe cada año.

El principal y sobre el que está basada nuestra variable explicativa del modelo que se creará son los hoteles. En todo el territorio español se pueden encontrar diversos alojamientos hoteleros desde 1 hasta 5 estrellas, teniendo el turista la posibilidad de escoger cualquiera de ellos según sean sus necesidades y/o motivo del viaje, ocio, recreo, vacaciones, negocios, universitarios, turistas de tercera edad, entre otros.[2]

Los alojamientos turísticos también son una opción muy latente en la sociedad española, perfectos para viajes generalmente de quincenas o meses, perfectamente equipados con todo tipo de accesorios para sentirte como en casa y mucho más baratos que los hoteles. Suelen escogerse en destinos donde predomina el turismo de sol y playa.[2]

En menor medida y dependiendo del tipo de turismo que se vaya a realizar se pueden encontrar: paradores, alojamientos que recuerdan a lugares únicos de alguna época pasada con la peculiaridad de que están equipados con las comodidades del siglo XXI. Albergues y campings turísticos, donde predominan en turismo rural para que exista una conexión más cercana entre la naturaleza y el turista. Por último, alojamientos singulares, donde dependiendo de la región que se visite el turista se dormirá o descansará en alojamientos típicos de dicha zona.[2]



1.2 Hipótesis

A continuación, se detallarán una serie de hipótesis previas a la realización del proyecto, una vez realizado el estudio, se comprobará si se cumplen o no las hipótesis planteadas.

- Explicar la variable pernотaciones hoteleras según el tipo de turismo (ocio, recreo y vacaciones) a partir de otras variables.
- Ver si los kilómetros de playa accesibles hacen aumentar las pernотaciones hoteleras, ya que el turismo predominante en España es de sol y playa
- Ver hasta qué punto afecta de manera negativa en Cataluña el elevado número de delitos en las pernотaciones hoteleras.
- La relación entre el turismo culturas y el número de bienes declarados Patrimonio de la Humanidad afecta de manera positiva a las pernотaciones hoteleras.
- Los territorios con las temperaturas y las horas de sol más elevadas son un condicionante clave a la hora de elegir establecimiento hotelero en España.
- El gasto medio por turista suele ser menor en los territorios que no colindan con la playa y poseen un elevado número de rutas de senderismo afectando de manera negativa a las pernотaciones hoteleras, ya que el turismo rural suele ser económico.



1.3 Objetivos

El objetivo principal del presente proyecto es estudiar y analizar las pernoctaciones hoteleras en los alojamientos del territorio español según el turismo de ocio, recreo y vacaciones para la obtención de un modelo econométrico que explique dichas pernoctaciones a partir de unos factores que pueden influir, llamados variables explicativas. A partir de ahí se plantean los siguientes objetivos secundarios:

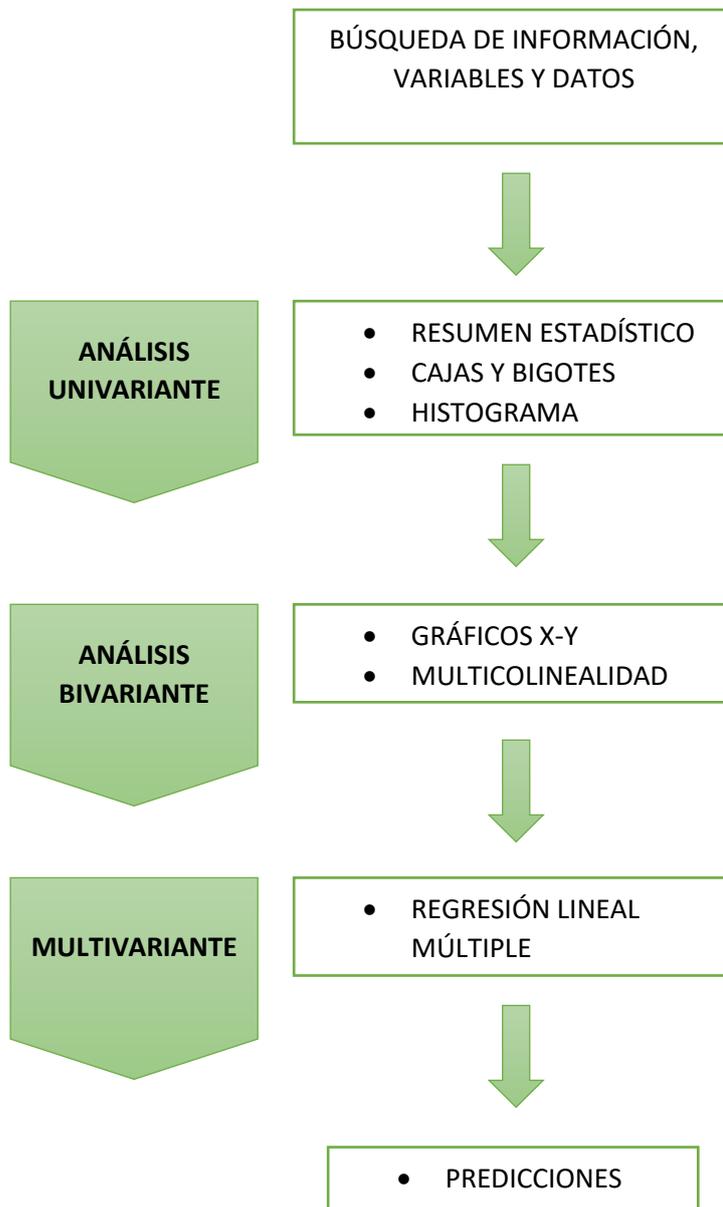
- Explicar las pernoctaciones hoteleras a través de un análisis descriptivo de las demás variables.
- Demostrar si existe relación lineal entre las pernoctaciones hoteleras (variable dependiente) y las variables independientes.
- Realizar un estudio de multicolinealidad para determinar el grado de relación que hay entre las variables independientes.
- Ajustar el modelo econométrico y comprobar la validez, dependiendo de las variables restantes al estudio de multicolinealidad frente a la variable dependiente, las pernoctaciones hoteleras.
- Determinar si existen anomalías en alguna comunidad autónoma del territorio español.
- Predecir con el modelo final resultante para ver que se podría conseguir con un territorio que modificara o viera alterada su planificación turística.



2. METODOLOGÍA Y TÉCNICAS

2.1 Metodología

A continuación, se procede a analizar las técnicas y herramientas que se ha seguido para realizar el trabajo, mostrando los procesos que han servido de guía para obtener el modelo.



1- Desglose de Procesos

2.2 Técnicas

2.2.1 Análisis Univariante

El análisis de una variante o análisis univariante mide el comportamiento de todas de las variables ya sean independientes o dependientes del modelo una a una, es decir, por separado. Está compuesto por el resumen estadístico donde los datos relevantes para el modelo son, el promedio, mediana, desviación estándar, mínimo, máximo, rango, cuartil inferior, cuartil superior, sesgo y curtosis. El análisis se complementa además con el gráfico de cajas y bigotes e histograma.

Resumen Estadístico

El resumen estadístico tal y como se ha citado anteriormente aporta la información importante para el modelo. Hay dos tipos de medidas: medidas de centralización y medidas de dispersión.

- Las medidas de centralización, aportan la información necesaria para saber situar una variable, es decir, un valor aproximado sobre dónde estaría el valor de la muestra.
 - Media: Se obtiene mediante la suma de todos y cada uno de los valores que se quieren analizar, dividido por el número total, en este caso las comunidades autónomas de España. El valor obtenido debe estar comprendido entre el máximo y el mínimo de los valores analizados.

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

- Mediana: La mediana es el valor situado en el centro de los valores de la muestra cuando estos están ordenados de menor a mayor. Se le conoce también como segundo cuartil, ya que representa el 50% de los datos. La mediana no se ve alterada o afectada por puntos anómalos.

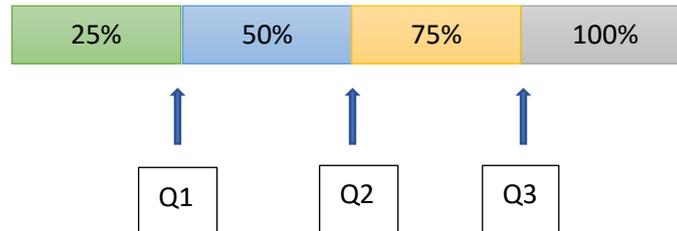
$$\text{Si } n \text{ es impar: } Me = X\left(\frac{n+1}{2}\right)$$

$$\text{Si } n \text{ es par: } Me = \frac{x_{\frac{n}{2}} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

- Medidas de dispersión.
 - Rango: Es la diferencia que existe ente el valor más elevado y el valor más pequeño, es una medida de variación muy simple y a la que le afecta la presencia de datos anómalos.

$$R = x_{\text{Max}} - x_{\text{Min}}$$

- Cuartiles 1 (Q1) y 3 (Q3): Son los valores que junto al cuartil 2 (Q2 o mediana), dividen los datos, siempre que estén ordenados, en cuatro partes. Q1 indica que el 25% de los datos son inferiores o iguales a dicho valor. Q3 indica que el 75% de los datos obtenidos están por debajo a ese valor.



2- Representación de los Cuartiles

- Varianza: La varianza mide la distancia existente en los valores de la muestra ($x_1, x_2, x_3, \dots, x_n$) a la media, es decir, mide la dispersión de estos datos alrededor de la media. Por lo tanto, cuanto mayor sea la varianza, mayor dispersión habrá en los datos. La varianza es igual a la desviación estándar elevado a 2.[3]

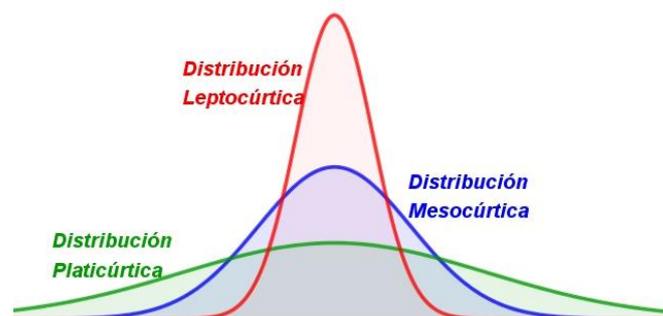
$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

- Desviación típica: En una muestra de valores ($x_1, x_2, x_3, \dots, x_n$), la desviación típica es la raíz cuadrada de la varianza.[3]

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

- Mínimo y máximo: En el caso del mínimo es el valor más pequeño y en el máximo el valor más alto de la variable.

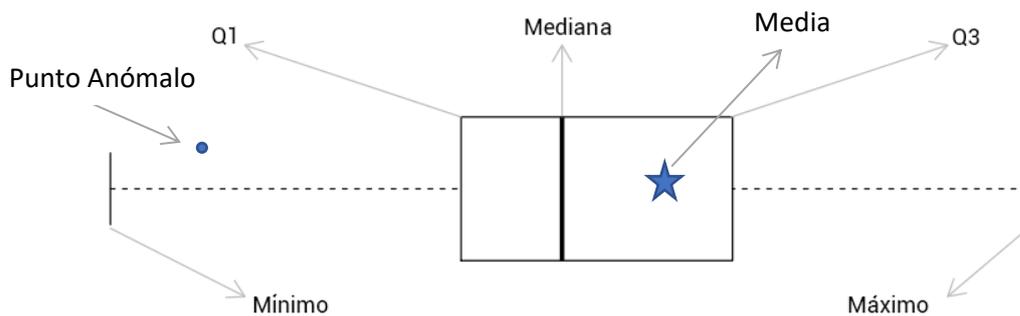
- Curtosis: La curtosis de una variable, trata de medir la frecuencia en la que se distribuyen los valores en torno a un valor central.[4]



3- Tipos de curtosis

Gráfico de cajas y bigotes

El gráfico de cajas y bigotes [5] es una herramienta visual que es capaz de mostrar varias características importantes a la vez, tanto de dispersión como de posición. Como se puede observar en la imagen, el gráfico muestra el valor mínimo y el valor máximo representado por una línea horizontal. El rectángulo o también llamado caja representa los cuartiles Q1, Q2 o también llamado mediana, y Q3. Además, en el gráfico se pueden observar los puntos anómalos, así como la media y la simetría del gráfico.



4- Gráfico de Cajas y Bigotes

2.2.2 Anàlisis Bivariante

En el análisis univariante, se explicaba cada variable por separado. En este apartado se pretende analizar la relación que existe entre variables.

Gráfico X-Y

Representa gráficamente dos variables, una variable X con la variable Y del modelo econométrico, de este modo se puede ver que relación existe entre ellas.

Multicolinealidad

El problema de multicolinealidad aparece cuando se tiene una alta relación entre las variables explicativas del modelo. Esto provoca problemas de ajuste del modelo, ya que es posible que la información que se quiere obtener esté duplicada por 2 o más variables de manera simultánea.

Para detectar los problemas de multicolinealidad existen varios métodos que pueden ser utilizados, pero en este caso se va a proceder a la utilización de; la matriz de correlación R y su inversa R^{-1} y el método de Belsley, Kug, Welsch.

- La matriz de correlación R: La matriz de correlación, mide el grado de relación existente entre ambas variables, es decir, mínimo hacen falta 2 variables ya que se analiza por parejas. De manera que en la diagonal de la matriz aparecerán 1 y en los elementos no diagonales el coeficiente de correlación R_{ij} . Para comprobar que dicha matriz esté bien elaborada, debe ser simétrica tanto por encima de la diagonal de 1 como por debajo.

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

En definitiva, aparecerán problemas de multicolinealidad cuando por debajo de la diagonal haya algún valor $R_{ij} > 0.7$.

- La matriz inversa de correlación R^{-1} , compara una variable explicativa con el resto de variables del modelo.

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

Si los valores de $R_{ii}^{-1} \geq 10$, aparecen problemas de multicolinealidad ya que significa que está muy relacionada con las demás variables.

- Método de Belsley, Kug, Welsch

Para aplicar este método, son necesarios los autovalores de las variables. Estos autores propusieron el llamado Índice de Acondicionamiento (I.C.)

$$I.C. = \sqrt{\frac{\text{Autovalor Max}}{\text{Autovalor Min}}}$$

Para identificar problemas de multicolinealidad se deben tener en cuenta:

Si $1 \leq I.C. \leq 10 \rightarrow$ No hay multicolinealidad

Si $10 \leq I.C. \leq 20 \rightarrow$ Aparece multicolinealidad moderada

Si $20 \leq I.C. \rightarrow$ Aparece multicolinealidad elevada

Para solucionar estos problemas que van apareciendo en el modelo ya que como se ha citado anteriormente están altamente relacionadas y la información se duplica, la mejor manera es la eliminación de una variable. La matriz inversa R^{-1} será la encargada de facilitar la información necesaria para saber que variable eliminar, ya que de todos los valores de cuya diagonal superen 10 serán candidatos, eliminando la variable que tenga el valor más elevado. Este proceso se deberá repetir hasta que no haya ningún error en los dos métodos anteriormente nombrados.

2.2.3 Análisis Multivariante

Se pretende obtener un modelo matemático a través de un modelo de Regresión Lineal Múltiple, en el que se pretende explicar una variable dependiente Y a través de unas variables explicativas X.

El modelo es válido cuando no se obtiene problemas de significatividad, de normalidad de residuos, de heterocedasticidad, autocorrelación, de puntos influyentes y de puntos anómalos, a continuación, se explicará más detalladamente cada apartado.

Otro análisis que se efectuará es el de coeficiente de determinación R^2 , este es muy útil para saber si el modelo a estudiar es bastante bueno o no.

El coeficiente de determinación siempre debe estar entre 0 y 1, cuanto más próximo esté de 1, mejor será el modelo a estudiar. Dependiendo si el signo del valor es positivo o negativo, indicará la pendiente del modelo.

Significatividad del modelo

Una vez eliminadas las variables adecuadas según el análisis de multicolinealidad, se realiza el modelo matemático, es decir, la regresión múltiple para determinar que variables son significativas y cuáles no.

Hay varios métodos para obtener las variables significativas, como comparar el F_{calc} con el $F_{teórico}$, utilizado para el modelo o comparar el T_{calc} con el $T_{teórico}$ utilizado para los parámetros. En esta memoria se analizará mediante el tercer método, utilizando el p-valor que sirve tanto para el modelo como para los parámetros.

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

Contraste del Modelo

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

$$H_1: \text{Algún } \beta_i \neq 0$$

Contraste de uno de los parámetros

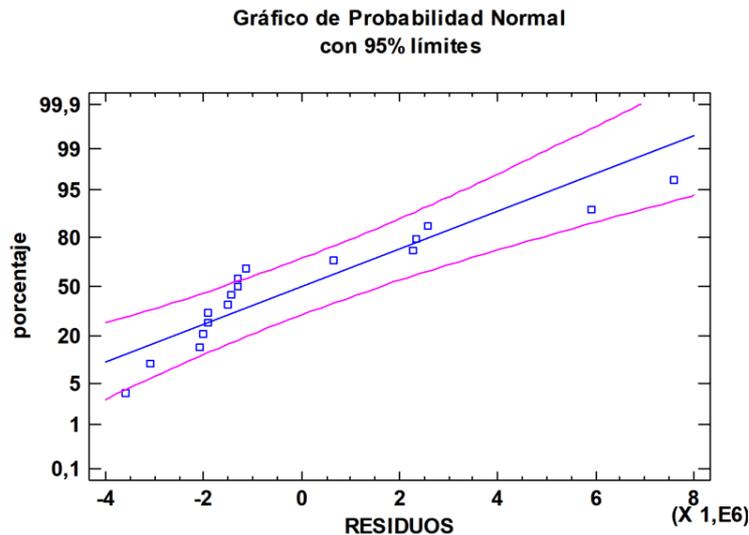
$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

El valor de α es 0'05, si el p-valor $\geq \alpha$ H_0 se acepta, concluyendo que el modelo y/o los parámetros no son significativos.

Normalidad de residuos

En el siguiente apartado, se va a explicar el gráfico de normalidad de residuos o papel probabilístico normal acompañado del contraste de hipótesis, esto ayudará al modelo a conocer si los residuos siguen una distribución normal o no.



5- Gráfico de Normalidad de Residuos

En la ilustración vienen representados los residuos mediante los puntos azules, que indican la normalidad de los mismos.

Este gráfico, aporta una gran cantidad de información, ya que se puede ver si los residuos siguen una distribución normal cuanto más próximos a la línea de tendencia estén, también se pueden ver los posibles problemas de residuos que tenga el modelo. Además, permite conocer los puntos anómalos, estos serán los que más alejados de la línea se encuentren.

El contraste de hipótesis, es otra manera de detectar los posibles problemas que aparezcan en la normalidad de residuos, que a través de un test realizado en el programa Statgraphics [6], este calcula 3 pruebas llamadas "Chi-cuadrado, Shapiro Wilks y Z para asimetría". Donde se escogerá el que tenga el p-valor más bajo.

H_0 : Los residuos son normales

H_1 : Los residuos no son normales

p-valor $\geq \alpha$, donde Alpha es 0'05, se aceptará la hipótesis nula, concluyendo de esta manera que los residuos siguen una distribución normal.

Heterocedasticidad

En los modelos de regresión lineal como este, se dice que hay problemas de heterocedasticidad cuando después de realizar varias observaciones, la varianza de los errores no es constante, con lo que se está incumpliendo una de las hipótesis del modelo de regresión.[7]

Para conocer si existe heterocedasticidad o no hay varias formas de analizarlo: Mediante análisis gráfico o bien mediante contraste de hipótesis, que es la que se utilizará en este modelo econométrico.

- Contraste de Hipótesis: en primer lugar, se guardarán los residuos del modelo inicial en el programa Statgraphics [6] y se tendrán que relacionar con las demás variables explicativas del modelo en cuestión, planteando un modelo lineal que será el siguiente.

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + U$$

Como se ha citado anteriormente, una vez se hayan guardado los residuos, se aplicará el siguiente modelo para comprobar el contraste de hipótesis.

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

$H_0: \sigma^2 = \text{cte} \rightarrow$ No hay heterocedasticidad

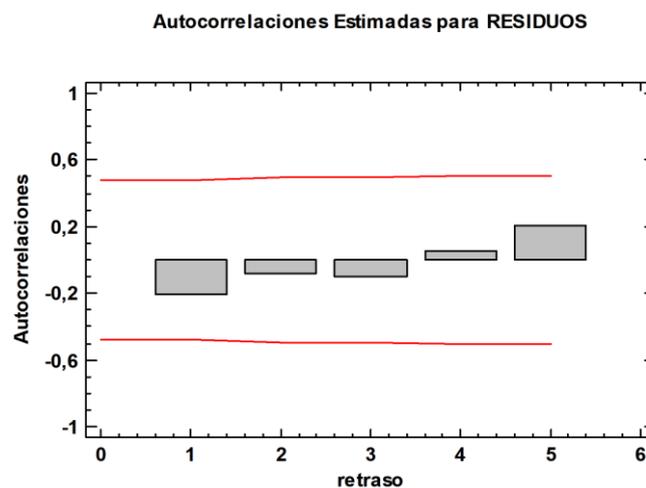
$H_0: \sigma^2 \neq \text{cte} \rightarrow$ Si hay heterocedasticidad

En definitiva, se podría concluir que si alguno de los parámetros beta sin tener en cuenta la constante, el p-valor de alguna variable es menor a 0'05, siendo de esta manera significativo, se podría aceptar que existe heterocedasticidad.

Autocorrelación

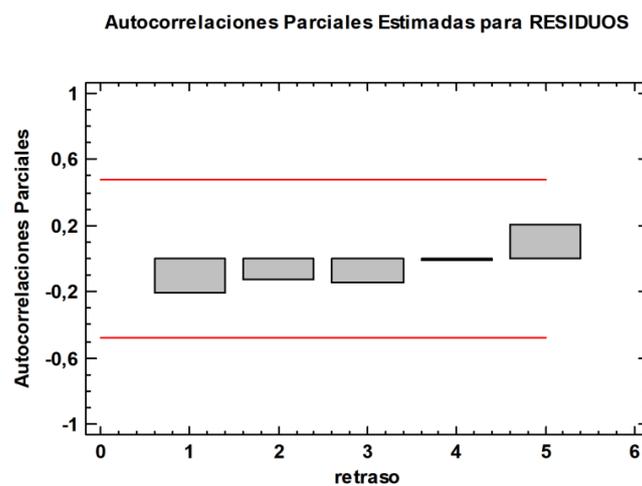
Para determinar la autocorrelación se pueden utilizar 2 formas: por un lado, el contraste de autocorrelación de 1er orden, por otro lado, los gráficos FAS (Funciones de Autocorrelaciones simples) y FAP (Funciones de Autocorrelaciones Parciales) [14].

El gráfico FAS representa todos los coeficientes de autocorrelación simple ρ_s , la relación lineal entre el residuo en el instante $t-s$, e_{t-s} y el residuo en el instante de tiempo t , e_t .



6- Gráfico FAS

Por el contrario, el gráfico FAP representa el conjunto de autocorrelaciones parciales, pero mide la relación directa entre los residuos citados anteriormente, con la particularidad de que elimina el efecto de los residuos intermedios en caso de que estos tuvieran algún efecto.



7- Gráfico FAP



Puntos influyentes y puntos anómalos

Los puntos influyentes son los puntos que al realizar el gráfico de residuos son relevantes para el modelo de regresión ya que afectan a la relación lineal existente en el modelo. Los puntos influyentes a priori afectan a la pendiente del modelo, es decir, a los parámetros que afectan a variables explicativas. Por otro lado, los puntos influyentes a posteriori, afectan a la constante del modelo.

Existen puntos influyentes a priori y a posteriori, que se pueden calcular de manera sencilla con el programa que se ha utilizado para realizar este proyecto de manera sencilla, donde de toda la información existente se deberá tener en cuenta solo la siguiente: La influencia (nii), DFITS y la influencia media de un punto (\bar{n}).

Para calcular los puntos influyentes a priori se tiene en cuenta la influencia $\rightarrow 2 \times (\bar{n})$.

Si la influencia es mayor que $2 \times (\bar{n}) \rightarrow$ Será un punto influyente a priori.

Si la influencia es menor que $2 \times (\bar{n}) \rightarrow$ No será un punto influyente a priori.

Para calcular los puntos influyentes a posteriori se tiene en cuenta DFITS $\rightarrow 2\sqrt{(\bar{n})}$.

Si DFITS es mayor que $2\sqrt{(\bar{n})} \rightarrow$ Será un punto influyente a posteriori.

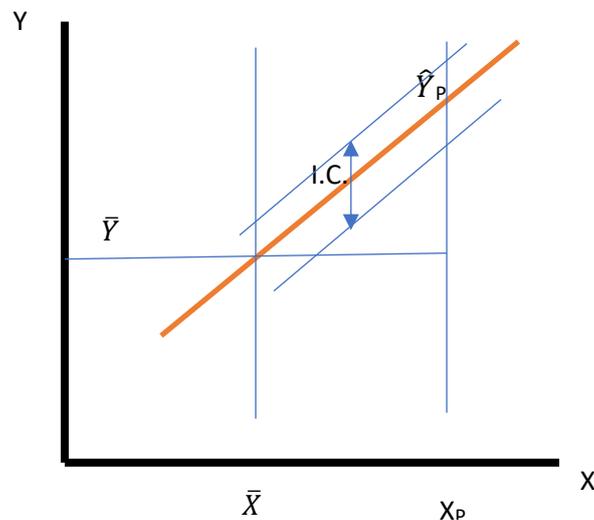
Si DFITS es menor que $2\sqrt{(\bar{n})} \rightarrow$ No será un punto influyente a posteriori.

Los residuos anómalos son los puntos que para el modelo están mal estimados, donde el programa calcula a través de los residuos un valor estudentizado, donde si este es mayor a 3 (en valor absoluto), es un punto anómalo candidato a ser eliminado del modelo.

Predicciones

El objetivo de este proyecto, así como de otros modelos económicos es poder hacer predicciones de las variables y que están al llevarlas a cabo se asemejen lo máximo posible a la predicción realizada.

Por lo tanto, partiendo de la base del modelo ajustado, se predice un valor medio, es decir, si la variable en cuestión aumenta o disminuye, en el caso de esta investigación el valor predicho está en porcentaje. Existen dos tipos de predicciones, puntuales y por intervalos.



8- Gráfico Predicción, puntual y por intervalos de confianza

La predicción de la variable explicativa es el punto representado en la ilustración de arriba por el punto X_p . Obteniéndose con el corte de la vertical de la recta \hat{Y}_p . Con todo esto, la predicción será, el valor esperado de la variable que se haya explicado, para los valores asignados a las demás variables explicativas.

El intervalo de confianza I.C., contiene una probabilidad determinada, el valor verdadero desconocido, de la variable determinada para los valores asignados a las demás variables explicativas del modelo, el valor esperado debe estar incluido en el intervalo de confianza y en la ilustración se encuentra entre las dos curvas de la recta de regresión.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CAMPUS D'ALCOI

3. CÁLCULOS, RESULTADOS Y EXPLOTACIÓN



En la tabla que hay a continuación, se muestran ordenadas todas las variables que se utilizarán para la realización de este proyecto y realizar el modelo econométrico. Los datos han sido aportados por el INE en base al año 2019 [8], donde la información aportada va en base a las comunidades autónomas del territorio español. Las comunidades autónomas aparecen a la derecha de la tabla ordenadas alfabéticamente, mientras que en la parte superior se encuentran las variables obtenidas, la primera de la izquierda, PH (Pernoctaciones Hoteleras), será la variable Y del modelo, es decir, la variable dependiente. Las demás variables son las variables independientes o variables X, es decir, las variables explicativas.

CCAA	PH	VIS. RES.	PLAYA	TEM. MED.	VIS. EXT.	Nº AERO	EMP. CULT.	Nº PAT. HUM.	GM PAX	Nº DELITOS	Nº SEND.	H. SOL
01 Andalucía	21003259	29728070	910	18,65	12023,2	6	16933	9	1032	15010	455816	9,14
02 Aragón	3433945	6690201	0	15	544,8	2	3107	6	672	1813	221396	8,34
03 Asturias, Principado de	2252686	4441749	401	13,9	317,9	1	2657	3	952	1681	131879	5,3
04 Balears, Illes	3936363	3622222	1283	17,7	13679,8	4	3420	2	1085	2165	106714	7,6
05 Canarias	7715656	6931553	1501	21,6	13146,9	8	5544	3	1282	3604	152518	8,95
06 Cantabria	1917330	2133001	284	15,1	392,5	1	1199	2	887	798	83964	6,1
07 Castilla y León	4235780	11297345	0	12,83	1286,2	4	5307	8	509	3267	296945	7,72
08 Castilla - La Mancha	1760285	8422734	0	15,78	262	1	3420	7	854	1979	107673	8,32
09 Cataluña	10731093	33448057	699	16,57	19375,2	4	25585	8	1100	24893	1136217	8,23
10 Comunitat Valenciana	13985612	17838521	518	18,54	9535,5	2	12467	5	1002	10649	397852	9
11 Extremadura	1171326	3842266	0	16,95	469,4	1	1920	3	583	1203	72827	8,9
12 Galicia	4796217	9854554	1629	14,53	1624,7	3	7033	5	943	3818	180951	6,58
13 Madrid, Comunidad de	4278329	35343853	0	15,8	7641	2	27715	5	1368	15974	270257	8,4
14 Murcia, Región de	1166684	4691885	274	19,7	1037,8	1	2667	1	1073	2177	115490	9
15 Navarra, Comunidad Foral de	1096042	2973847	0	13,7	415,1	1	1644	1	630	710	133367	6,6
16 País Vasco	2216342	10742903	246	14,1	1611,3	3	6010	5	800	4796	288751	5,44
17 Rioja, La	591761	1310083	0	15	139,5	1	701	2	704	493	33933	7,4

Tabla 1: Tabla de la base de datos del INE de las variables escogidas para el proyecto

3.1 Univariante

A continuación, se procede a la realización del análisis de una variable, es decir, el análisis univariante. A través del cual, se puede obtener el resumen estadístico para cada una de estas variables, ya sea dependiente o independiente. El estudio se complementa con el gráfico de cajas y bigotes y el histograma para el año 2019 en las 17 comunidades autónomas.

La variable “Y” o variable dependiente de este modelo son las pernoctaciones hoteleras, según el turismo de ocio, recreo y vacaciones. Englobando todo tipo de actividades que se realicen fuera del ámbito laboral, pudiendo ser realizadas en familia, que impliquen pernoctación, por ejemplo, actividades que se pueden realizar en parajes naturales, en grandes urbes o incluso en ciudades costeras.

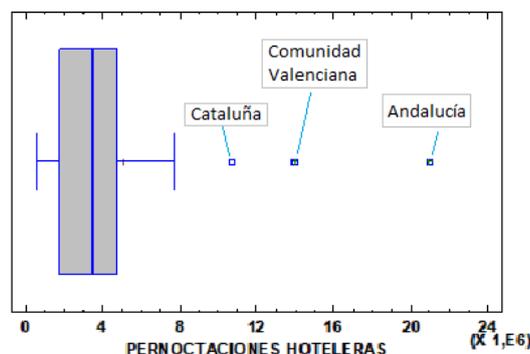
En la tabla de abajo (Tabla 2), se calcula el resumen estadístico para las pernoctaciones hoteleras durante el año 2019 de los territorios previamente nombrados.

Recuento	17
Promedio	5,07581E6
Mediana	3,43395E6
Desviación Estándar	5,48304E6
Mínimo	591761,
Máximo	2,10033E7
Rango	2,04115E7
Cuartil Inferior	1,76029E6
Cuartil Superior	4,79622E6
Sesgo	1,9568
Curtosis	3,71968

Tabla 2: Resumen Estadístico de PH

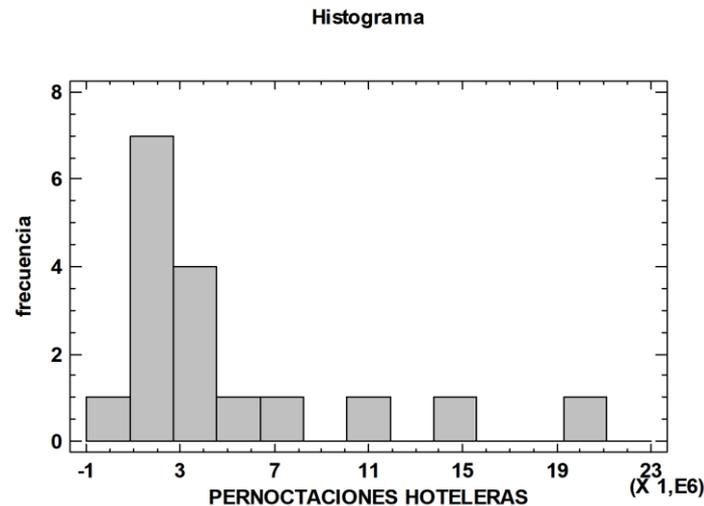
En la ilustración inferior (Ilustración 9), está representado el gráfico de cajas y bigotes, en se puede observar el comportamiento de las diferentes comunidades en referencia a la variable de pernoctaciones hoteleras, destacando el comportamiento de Cataluña, Comunidad Valenciana y Andalucía. Estos puntos son llamados anómalos, ya que poseen un mayor índice de pernoctación que las demás, por lo tanto, sobresalen del bigote de la derecha.

Gráfico de Caja y Bigotes



9- Gráfico de cajas y bigotes para PH

El histograma de las pernoctaciones hoteleras muestra que en la parte izquierda sí que se puede ver reflejada una especie de campana de gauss, en cambio, en la parte derecha, vemos las comunidades que forman los puntos anómalos en el gráfico de cajas y bigotes.



10- Histograma variable PH

Una vez realizados los cálculos para la variable dependiente del modelo, se procederá a analizar las variables independientes, es decir, todas las que pueden influir en el resultado final de la variable "Y".

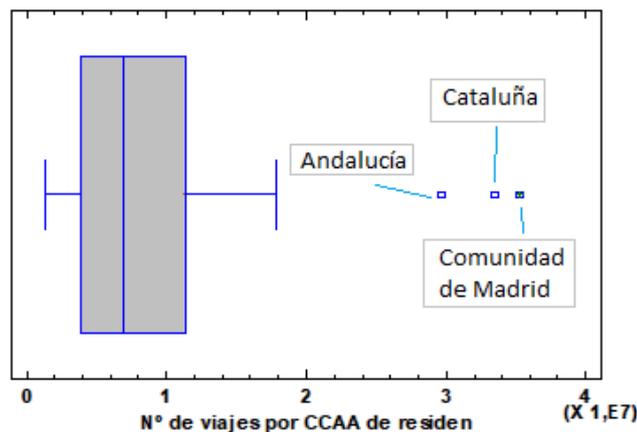
En la Tabla 3, se puede ver el resumen estadístico sobre el número de viajes por comunidad autónoma de residencia. Se puede apreciar que hay poca diferencia entre el máximo y el mínimo.

Recuento	17
Promedio	1,13713E7
Mediana	6,93155E6
Desviación Estándar	1,10689E7
Mínimo	1,31008E6
Máximo	3,53439E7
Rango	3,40338E7
Cuartil Inferior	3,84227E6
Cuartil Superior	1,12973E7
Sesgo	1,40243
Curtosis	0,696788

Tabla 3: Resumen Estadístico variable Vis. Res.

En la Ilustración 11, cabe destacar la aparición de 3 puntos anómalos, Andalucía, Cataluña y la Comunidad de Madrid, esto puede deberse a la gran densidad de población de dichas comunidades, haciendo de ellas lugares más atractivos para visitar y con más flujo de movimiento. Por lo demás, el gráfico representa simetría, destacando que el bigote de la derecha es más largo que el de la izquierda.

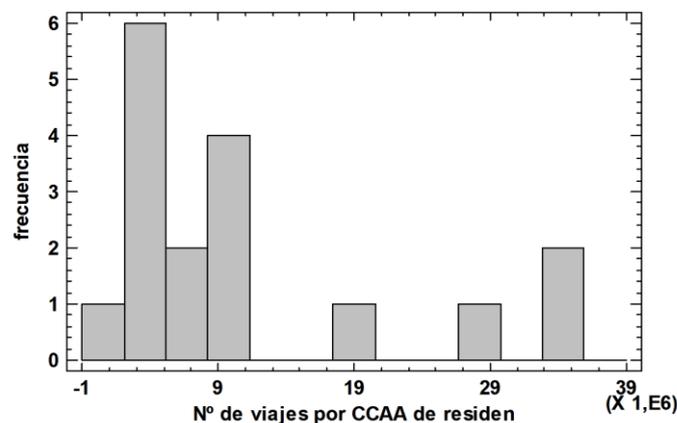
Gráfico de Caja y Bigotes



11- Gráfico de cajas y bigotes variable Vis. Res.

El histograma del N.º de viajes por CCAA de residencia, es similar al anterior, ya que en la parte izquierda se puede ver una campana de gauss con una pendiente creciente inclinada. También se pueden apreciar los puntos anómalos, representados a la derecha del histograma.

Histograma



12- Histograma variable Vis. Res.

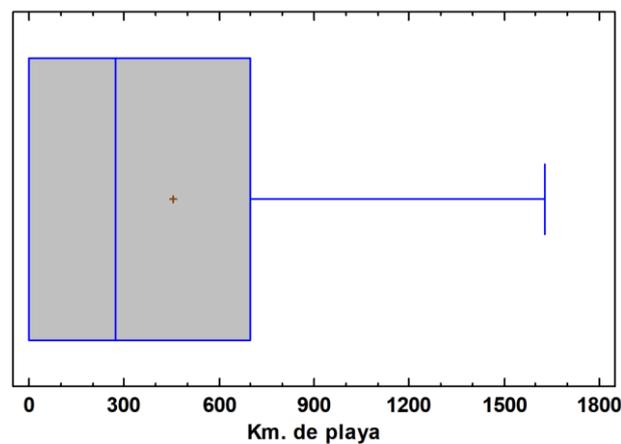
En la variable de Kilómetros de playa por comunidad autónoma, se puede ver en la Tabla 4 la gran diferencia que existe entre el mínimo y el máximo, esto se debe a que algunas de las comunidades no limitan con la costa, aun así, el promedio es de 455'6 km. de costa, debido principalmente a los 2 archipiélagos que provocan un aumento del promedio.

Recuento	17
Promedio	455,588
Mediana	274,0
Desviación Estándar	558,249
Mínimo	0
Máximo	1629,0
Rango	1629,0
Cuartil Inferior	0
Cuartil Superior	699,0
Sesgo	1,11998
Curtosis	0,00983636

Tabla 4: Resumen Estadístico variable KM. PLAYA

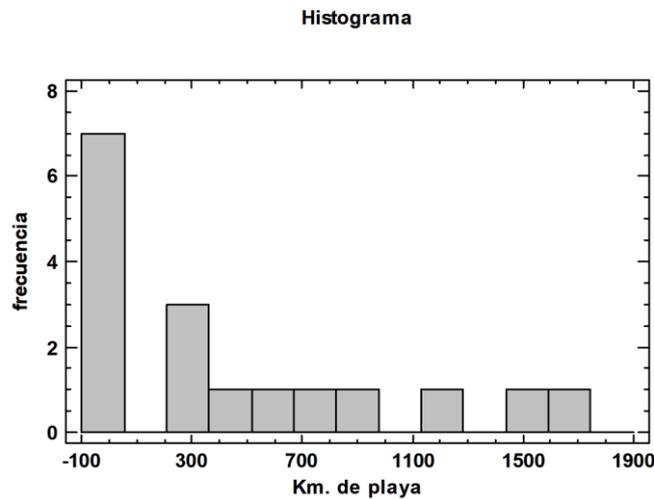
El gráfico de cajas y bigotes de la variable Km. de playa, es asimétrico a la derecha, ya que el bigote de la izquierda coincide con el valor del mínimo, siendo en este caso 0.

Gráfico de Caja y Bigotes



13- Gráfico de cajas y bigotes para KM. PLAYA

En el histograma, se puede observar el final de una campana de gauss con pendiente descendente, hasta llegar a ser constante.



14 Histograma variable KM. PLAYA

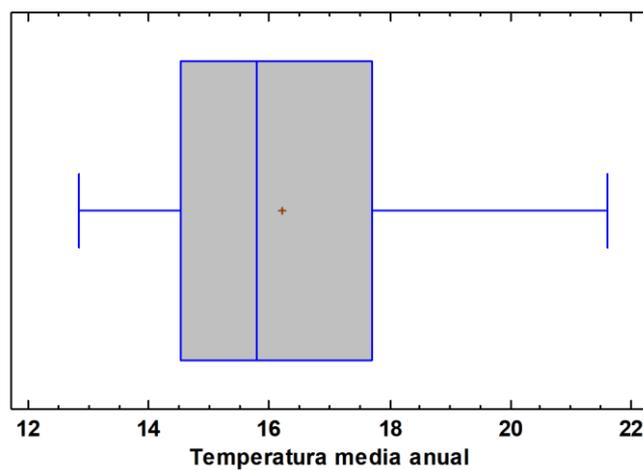
La siguiente variable es la temperatura media anual, esta puede afectar al modelo de manera muy significativa, en España se puede observar en el resumen estadístico de la Tabla 5 que no hay una gran diferencia entre el mínimo y el máximo de temperatura media, apenas 10 grados.

Recuento	17
Promedio	16,2029
Mediana	15,78
Desviación Estándar	2,37644
Mínimo	12,83
Máximo	21,6
Rango	8,77
Cuartil Inferior	14,53
Cuartil Superior	17,7
Sesgo	0,770565
Curtosis	0,0520116

Tabla 5: Resumen Estadístico variable TEMP. MED.

En el siguiente gráfico (Ilustración 15), se puede comprobar que no existe una gran diferencia entre las comunidades en cuanto a temperatura se refiere, ya que no existe ninguna anomalía. A pesar de que el bigote de la derecha es más largo que el de la izquierda, se podría decir que existe un grado de simetría, esto es debido al buen clima que existe en España, que provoca que no existan diferencias elevadas entre comunidades.

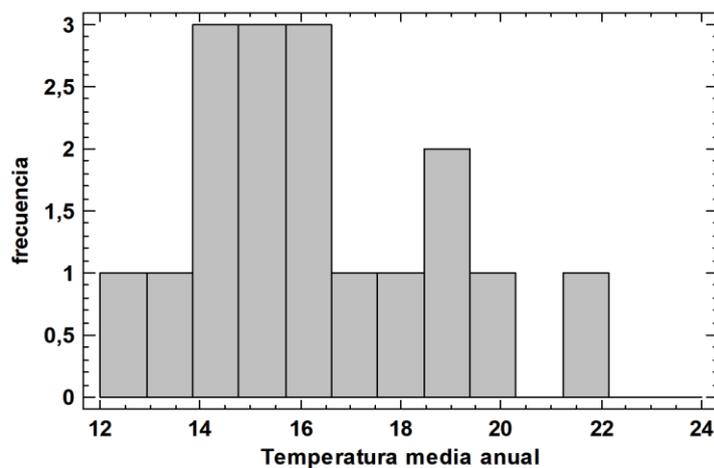
Gráfico de Caja y Bigotes



15- Gráfico cajas y bigotes TEMP. MED.

El histograma de la temperatura representa una campana de gauss, obteniendo el pico máximo de la campana de 3 comunidades, con valores entre 14 y 17 grados aproximadamente.

Histograma



16- Histograma variable TEMP. MED.

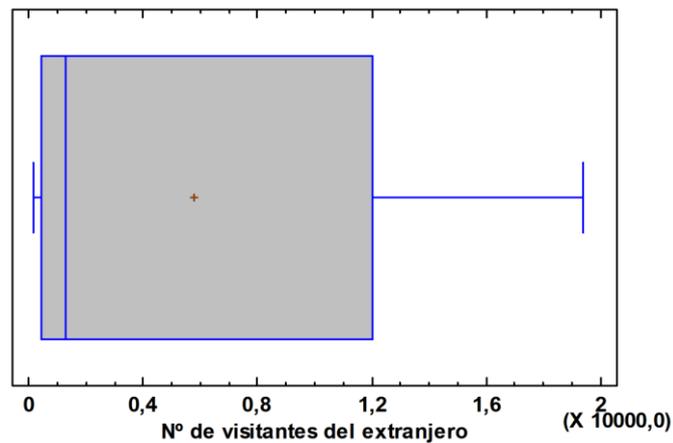
En cuanto a la siguiente variable, N.º de visitantes del extranjero, está expresada en miles, a pesar de ser un país muy turístico, observando los datos que nos ofrece el resumen estadístico, aparecen diferencias importantes entre el mínimo y el máximo.

Recuento	17
Promedio	5764,97
Mediana	1286,2
Desviación Estándar	6755,94
Mínimo	139,5
Máximo	19375,2
Rango	19235,7
Cuartil Inferior	415,1
Cuartil Superior	12023,2
Sesgo	0,795014
Curtosis	-0,964349

Tabla 6: Resumen Estadístico VIST. EXT.

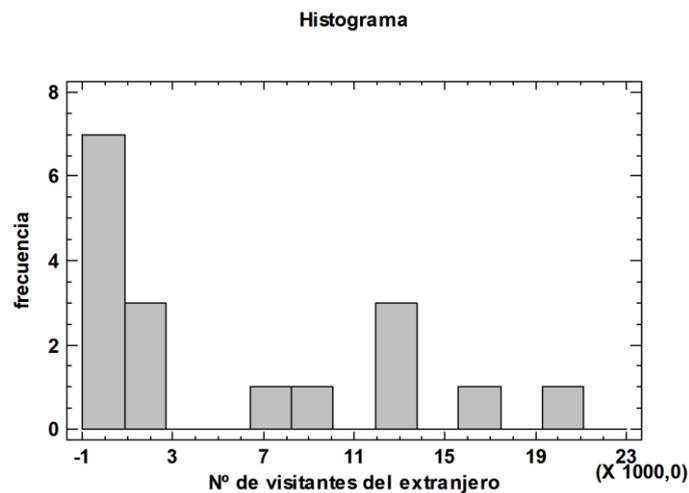
La imagen 17, muestra el gráfico de cajas y bigotes de la variable visitantes del extranjero, se puede ver que es asimétrico a la izquierda, ya que el bigote y la caja de la derecha es mucho mayor que el de la izquierda. En cuanto puntos anómalos no obtenemos ninguno.

Gráfico de Caja y Bigotes



17- Gráfico cajas y bigotes VIST. EXT.

En cuanto al histograma de la variable, se puede observar una gran dispersión en los datos, ya que no se repite un patrón en el comportamiento del mismo, aunque en la parte derecha se puede ver un cierto grado de simetría.



18- Histograma variable VIST. EXT.

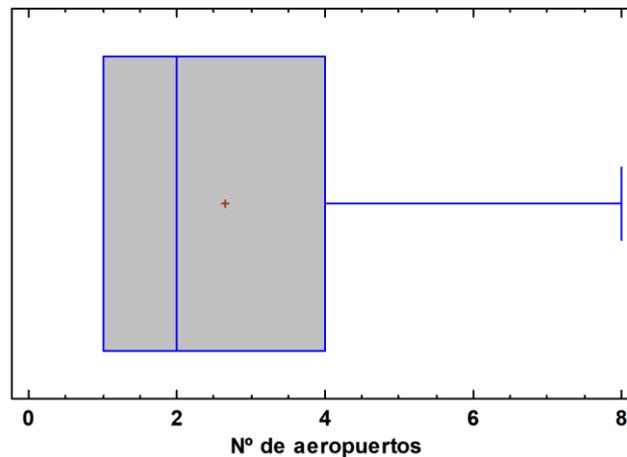
La siguiente variable que se analizará en el modelo, son el n.º aeropuertos existentes en cada una de las comunidades autónomas y que están gestionados por AENA [9], es una variable muy importante ya que la mayoría de los turistas que recibe el país provienen a través del transporte aéreo.

Recuento	17
Promedio	2,64706
Mediana	2,0
Desviación Estándar	2,0292
Mínimo	1,0
Máximo	8,0
Rango	7,0
Cuartil Inferior	1,0
Cuartil Superior	4,0
Sesgo	1,41154
Curtosis	1,76557

Tabla 7: Resumen Estadístico para N.º de Aeropuertos

En la siguiente ilustración, se puede apreciar un gráfico asimétrico, ya que el cuartil inferior coincide con el mínimo. Es importante destacar que no aparece ningún punto anómalo, a pesar de que el 50% de las comunidades poseen menos de 2 aeropuertos.

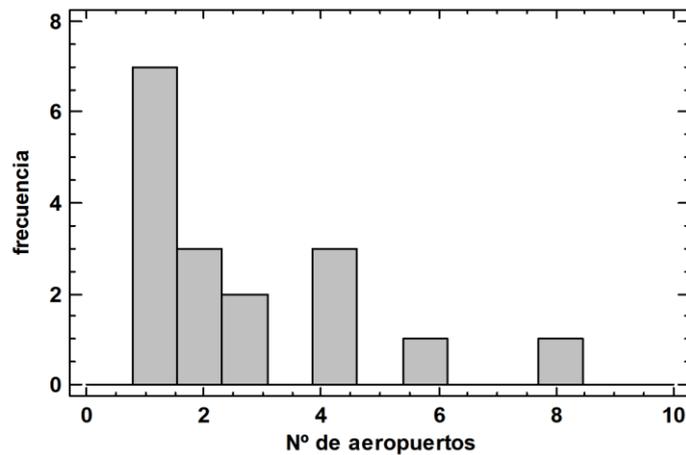
Gráfico de Caja y Bigotes



19- Gráfico de cajas y bigotes variable Nº AERO.

El histograma de dicha variable representa una mitad de campana de gauss con pendiente descendente.

Histograma



20- Histograma variable Nº AERO.

El N.º de empresas culturales es la siguiente variable del modelo, en este grupo se encuentran todo tipo de empresas que aportan valor añadido a un determinado lugar, pudiendo ser museos, cines o teatros, además ofrecen enseñanzas sobre todo tipo de temas, como arte, historia o música [10].

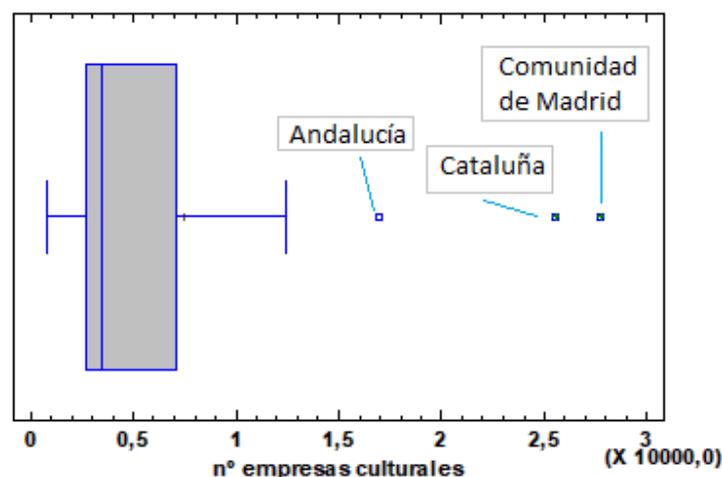
En el resumen estadístico se puede ver que hay una notable diferencia entre el mínimo y el máximo, esto puede deberse a que en dichas comunidades el principal tipo de turismo receptor no sea el cultural, provocando que se apueste por otro tipo de empresas captadoras de turistas.

Recuento	17
Promedio	7489,94
Mediana	3420,0
Desviación Estándar	8333,67
Mínimo	701,0
Máximo	27715,0
Rango	27014,0
Cuartil Inferior	2657,0
Cuartil Superior	7033,0
Sesgo	1,68447
Curtosis	1,83679

Tabla 8: Resumen Estadístico Nº EMP. CUL.

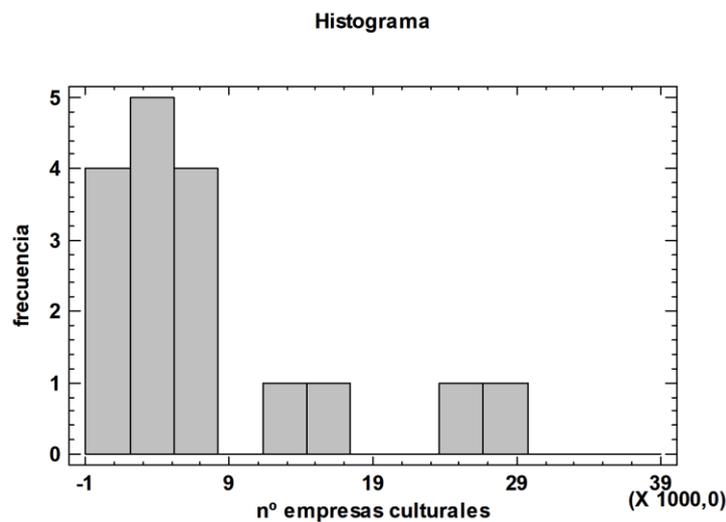
En el gráfico, se puede ver una asimetría a la derecha, ya que el bigote de la derecha es mayor que el de la izquierda. Además, es destacable la aparición de 3 puntos anómalos, Andalucía, Cataluña y Madrid, esto puede deberse a que son unas de las comunidades más visitadas de nuestro país, destacando el caso de Madrid, al no disponer de turismo de sol y playa, encuentra su atractivo en los monumentos y eventos culturales.

Gráfico de Caja y Bigotes



21- Gráfico de cajas y bigotes Nº EMP. CUL.

En el histograma, en la parte izquierda podemos ver una especie de campana de gauss perfecta, obteniendo el pico en 5 comunidades, en la parte derecha tenemos 4 comunidades, siendo 3 de ellas los puntos anómalos citados anteriormente.



22- Histograma variable EMP. CUL.

La variable, N.º de bienes patrimonio de la humanidad declarados por la UNESCO [11] está muy relacionada con la de empresas culturales, ya que engloba el mismo tipo de turismo, el cultural, la diferencia se encuentra en que esta variable se centra únicamente en monumentos y principalmente su atractivo radica en la historia de los mismos, bien por antigüedad o bien por costumbres autóctonas del lugar.

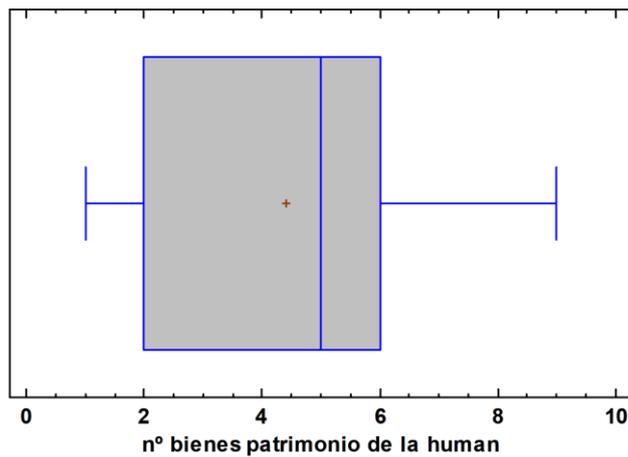
Prestando atención al resumen estadístico, vemos que en general, en España hay una media de 4'41 bienes declarados patrimonio de la Humanidad en cada comunidad, ya que compensa la diferencia existente entre comunidades.

Recuento	17
Promedio	4,41176
Mediana	5,0
Desviación Estándar	2,55095
Mínimo	1,0
Máximo	9,0
Rango	8,0
Cuartil Inferior	2,0
Cuartil Superior	6,0
Sesgo	0,346475
Curtosis	-1,05299

Tabla 9: Resumen Estadístico Nº PAT. HUM.

El gráfico de cajas y bigotes, se puede ver una asimetría a la izquierda, ya que el bigote de la izquierda es más corto que el de la derecha y en este caso hay más comunidades por encima de la media.

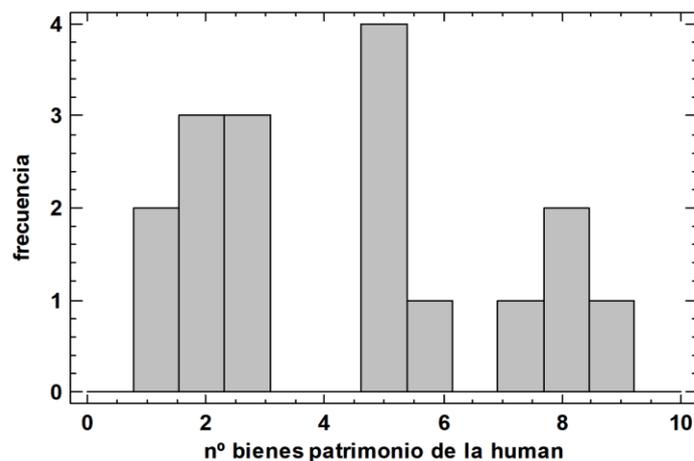
Gráfico de Caja y Bigotes



23- Gráfico de cajas y bigotes N° PAT. HUM.

En el histograma se puede apreciar una campana de gauss, teniendo en el centro del gráfico el máximo de la campana, a pesar de existir una campana de gauss el histograma no es simétrico, ya que en los datos de la derecha se observa otra pequeña campana de gauss mientras que en el lado izquierdo no.

Histograma



24- Histograma variable N° PAT. HUM.

La variable del gasto medio por persona de los turistas internacionales, facilita tener una visión de los que los turistas que visitan los diferentes territorios de España gastan en nuestro país de media. Observando el resumen estadístico, el promedio en las comunidades autónomas es de 910'353 euros.

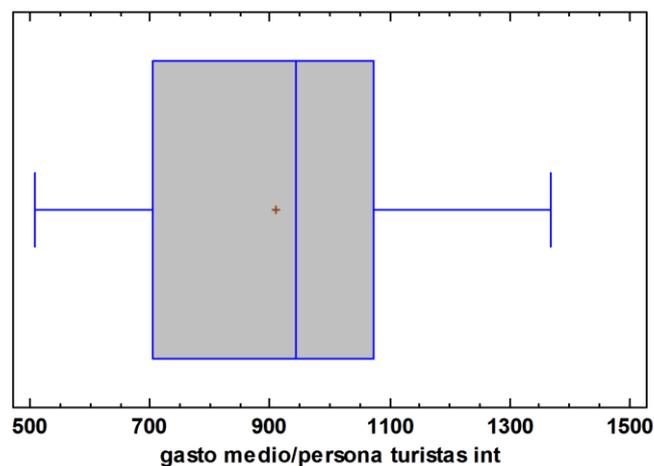
La diferencia entre el mínimo y el máximo es significativa ya que son aproximadamente unos 860 euros, esto puede avanzar que hay diferencias entre los lugares más turísticos de España, que, por ende, los turistas gastan más dinero.

Recuento	17
Promedio	910,353
Mediana	943,0
Desviación Estándar	240,912
Mínimo	509,0
Máximo	1368,0
Rango	859,0
Cuartil Inferior	704,0
Cuartil Superior	1073,0
Sesgo	0,104766
Curtosis	-0,548791

Tabla 10: Resumen Estadístico GM. PAX.

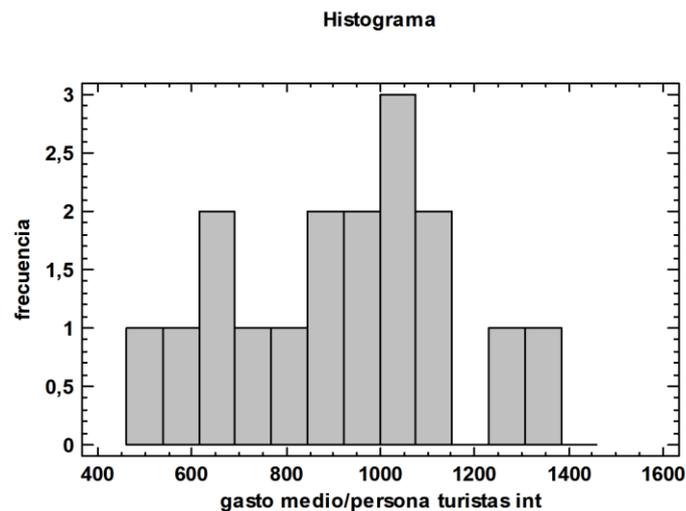
El gráfico de cajas y bigotes de la variable de gasto medio por turista internacional no presenta ningún punto anómalo. Además, se puede apreciar un grado importante de simetría, ya que entre el bigote de la izquierda y el de la derecha no existen grandes diferencias.

Gráfico de Caja y Bigotes



25- Gráfico cajas y bigotes para GM PAX.

En el histograma, se pueden observar dos campanas de gauss, la primera entre 500 y 850 euros de gasto, obteniendo el pico en 650 euros con una frecuencia de 2 comunidades. La segunda, englobaría desde los 850 a los 1400 euros, obteniendo el pico en 3 comunidades con un gasto de 1050 aproximadamente.



26- Histograma variable GM. PAX.

La siguiente variable que afecta a nuestro estudio es el n.º de delitos (robo y hurto), teniendo como referencia los delitos de robo y hurto solamente. Esta variable pretende medir cuanto afecta la seguridad de un territorio para que este sea visitado o no. Por norma, general los territorios con más afluencia de gente y con más turistas suelen ser objeto de robos y hurtos, con lo que si se cometen demasiados actos delictivos puede influir en el número final de visitantes.

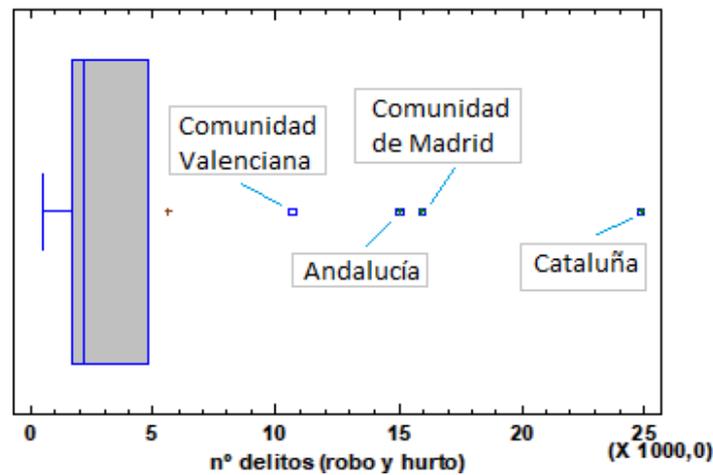
El resumen estadístico, muestra un promedio de 5590 actos delictivos en las comunidades autónomas de España, pero teniendo en cuenta la gran diferencia que existe entre el máximo y el mínimo, estos se cometen con mayor frecuencia en algunas comunidades concretas como pueden ser Cataluña y Madrid, incrementando la media nacional.

Recuento	17
Promedio	5590,0
Mediana	2177,0
Desviación Estándar	6918,46
Mínimo	493,0
Máximo	24893,0
Rango	24400,0
Cuartil Inferior	1681,0
Cuartil Superior	4796,0
Sesgo	1,83254
Curtosis	2,78509

Tabla 11: Resumen Estadístico para Nº DEL.

El gráfico de cajas y bigotes muestra 4 puntos anómalos, es decir, 4 comunidades (Comunidad Valenciana, Andalucía, Comunidad de Madrid y Cataluña) sobrepasan este máximo significativamente, dando lugar a un gráfico asimétrico a la izquierda.

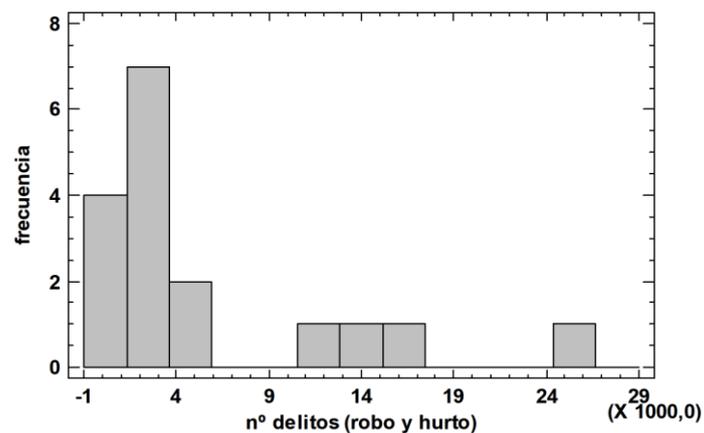
Gráfico de Caja y Bigotes



27- Gráfico cajas y bigotes Nº. DEL.

En el histograma, en la parte izquierda, se puede ver una especie de campana de gauss con el pico en 7 comunidades, las cajas de la parte central y derecha muestran los puntos anómalos indicados en el gráfico anterior.

Histograma



28- Histograma variable Nº DEL.

Respecto a la variable, n.º de rutas de senderismo, se aprecia el gran número de rutas existentes en nuestro país, esto es debido a todo el sistema montañoso y con gran desnivel que encontramos en las comunidades autónomas. El senderismo y otros deportes de montaña, son una pieza clave en el turismo de ocio y recreo, ya que combinas la actividad física y saludable, con actividades (en muchos casos) culturales [12].

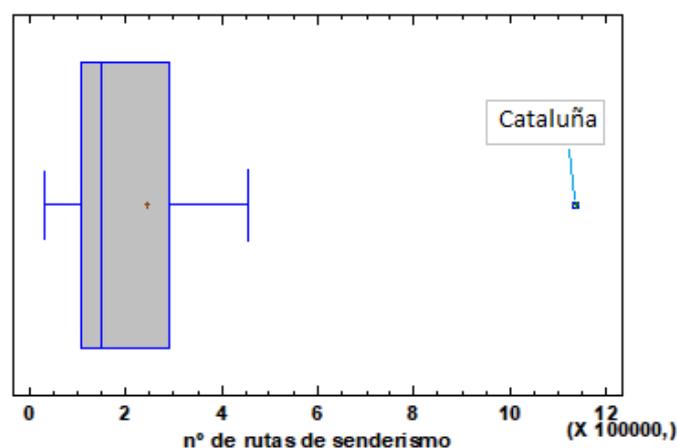
El promedio de rutas en las 17 comunidades autónomas según el resumen estadístico, es de 246268 rutas. La diferencia entre el máximo y el mínimo es significativa ya que existen diferencias entre la superficie del territorio de las comunidades autónomas, provocando que no aparezcan rutas.

Recuento	17
Promedio	246268,
Mediana	152518,
Desviación Estándar	257469,
Mínimo	33933,0
Máximo	1,13622E6
Rango	1,10228E6
Cuartil Inferior	107673,
Cuartil Superior	288751,
Sesgo	2,8512
Curtosis	9,49961

Tabla 12: Resumen Estadístico para Nº SEN.

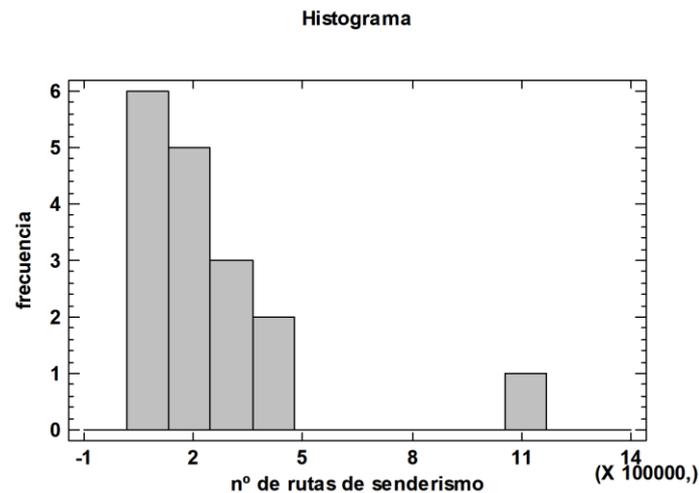
El gráfico de cajas y bigotes enseña un grado de simetría, aunque el bigote de la derecha es ligeramente mayor. Destacar la aparición de un punto anómalo, este es la comunidad de Cataluña, ya que posee un índice elevado de rutas de senderismo.

Gráfico de Caja y Bigotes



29- Gráfico cajas y bigotes de Nº SEN

Tal y como se aprecia en el histograma, no existe una campana de gauss, sino que se puede ver una pendiente descendente, desde el pico situado en 6 comunidades hasta situarlo en 2. La caja apartada en el lado derecho del gráfico, es el punto anómalo mostrado anteriormente, la comunidad de Cataluña.



30- Histograma variable Nº SEN

Para terminar con el análisis univariante, tenemos la variable horas de sol media al día, gracias al clima existente en España, las horas de sol al día son importantes, a pesar de ello, existen diferencias entre las CCAA [13].

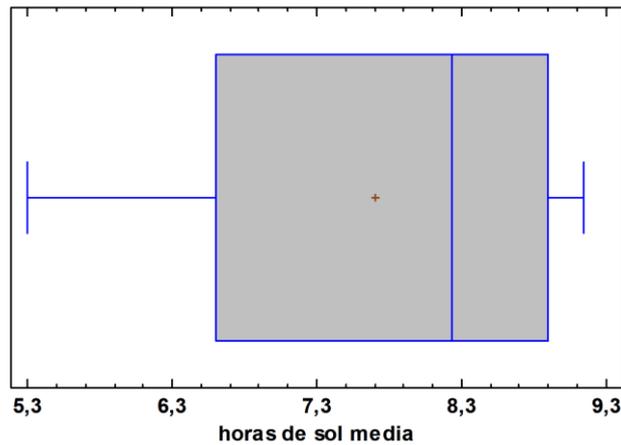
Como se puede ver en el resumen estadístico, la media de horas de sol al día durante el año 2019 en las comunidades autónomas fue de 7,7 horas al día, es decir, aproximadamente un tercio del día. La diferencia entre el mínimo y el máximo es de 4 horas, que en términos de la variable que estamos tratando es significativa.

Recuento	17
Promedio	7,70706
Mediana	8,23
Desviación Estándar	1,27506
Mínimo	5,3
Máximo	9,14
Rango	3,84
Cuartil Inferior	6,6
Cuartil Superior	8,9
Sesgo	-0,708153
Curtosis	-0,742101

Tabla 13: Resumen Estadístico para H. SOL.

En el gràfic de caixes y bigotes para la variable horas de sol al día, no existe ningún punto anómalo, existiendo asimetría a la izquierda, ya que el bigote de la izquierda es mayor al de la derecha.

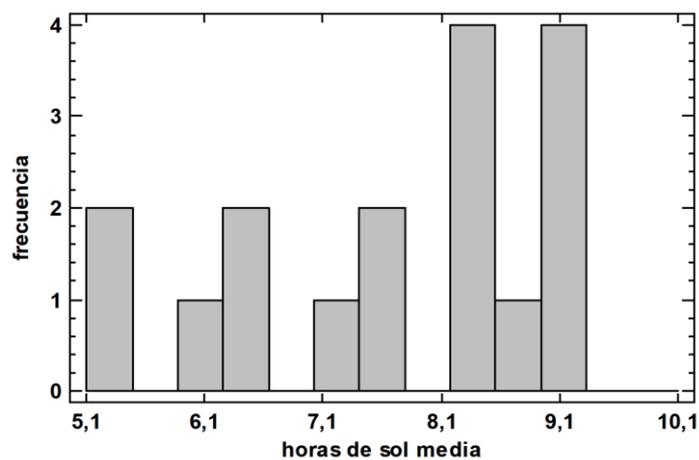
Gráfico de Caja y Bigotes



31- Gráfico cajas y bigotes de H. SOL

En el histograma, no se aprecia la existencia de una campana de gauss, en la parte izquierda del gráfico, se puede apreciar un patrón que se repite cuanto menor son las horas de sol, a medida que las horas de sol aumentan (8'1 y 9'1), la frecuencia se dispara englobando a 4 las comunidades autónomas en cada caso.

Histograma



32- Histograma variable H. SOL

3.2 Bivariante

El siguiente análisis que se va a realizar es el bivariante, es decir, se procederá a relacionar cada una de las variables X (independientes), con la variable Y (dependiente), para ver cómo afecta y que relación tienen entre ellas.

Las opciones de relación que existen en el modelo son diversas, pudiendo estar entre el siguiente rango de valores: entre 1 y 0'7 (correlación alta y directa), entre 0'4 y 0'7 (correlación moderada y directa), entre 0'4 y -0'4 (no hay correlación), entre -0'4 y -0'7 (correlación moderada e indirecta) y entre -0'7 y -1 (correlación alta e indirecta).

3.2.1 Gráficos X-Y

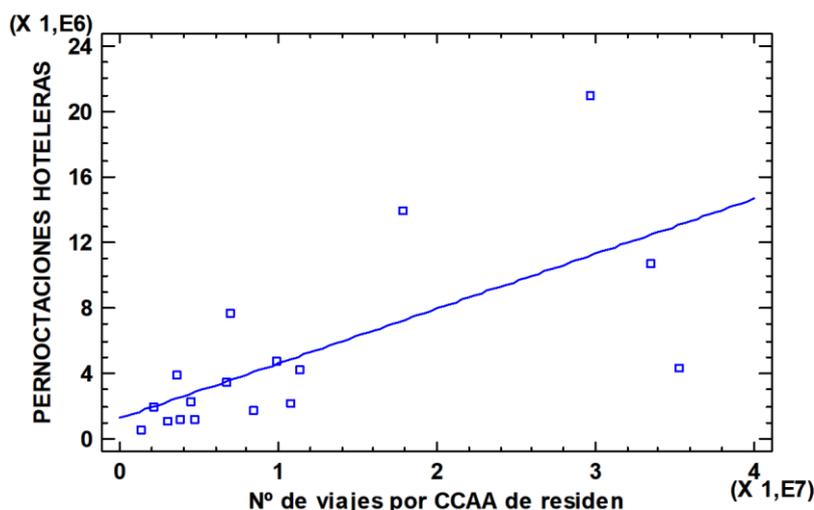
La primera variable que se va a analizar son las pernoctaciones hoteleras, representada en el eje Y, en el eje X se expresa el N.º de viajes por CCAA de residencia. Además, se podrá ver el modelo correspondiente y el coeficiente de correlación (R), para cada una de las comunidades autónomas en el año 2019.

La relación existente entre ambas variables es directa, ya que como se puede observar en el gráfico de dispersión, la pendiente es ascendente, esto quiere decir que a medida que aumenta la variable X, aumenta la variable Y.

El coeficiente de correlación, se encuentra muy próximo al 0'7, con lo que la relación existente es moderada-alta.

Coefficiente de Correlación = 0,674899.

PERNOCTACIONES HOTELERAS = 1,27421E6 + 0,334314*Nº de viajes por CCAA de residen



33- Gráfico X-Y de la variable VIS. RES.

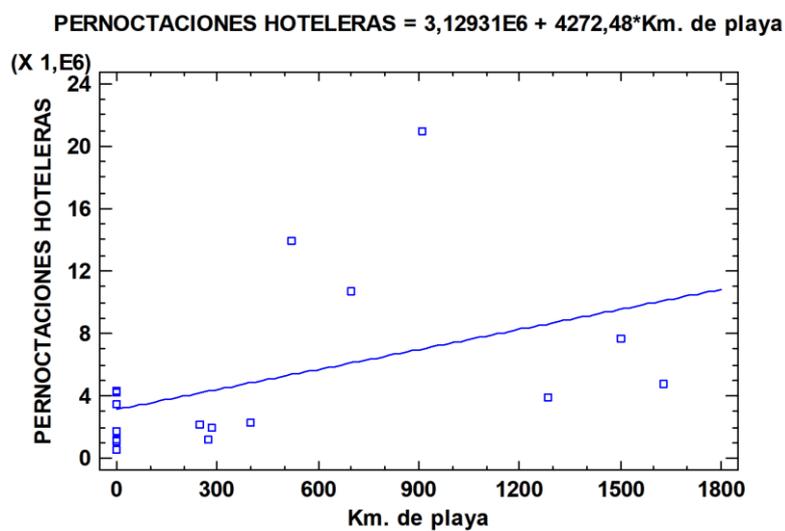


En el siguiente gráfico del análisis, se puede apreciar representado en el eje X, los km. De playa en los diferentes territorios citados anteriormente del estado español. En el eje Y nuestra variable dependiente, las pernoctaciones hoteleras.

La pendiente mostrada en el gráfico de dispersión es ascendente, esto quiere decir que la relación entre ambas variables es directa. Cuantos más km. De playa disponibles, más pernoctaciones hoteleras existen.

El coeficiente de correlación se encuentra en un valor cercano al 0'4, esto quiere decir que hay correlación moderada-baja o inexistente.

Coeficiente de Correlación = 0,434997



34- Gráfico X-Y de la variable KM. PLAYA

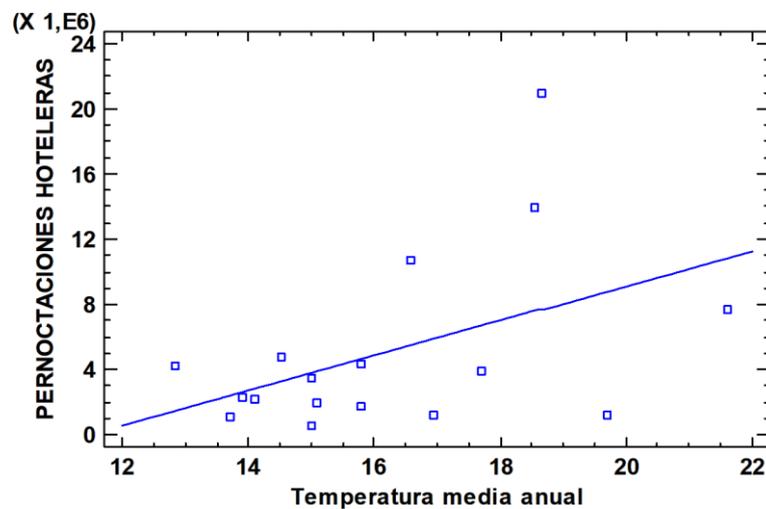
La siguiente variable planteada por el modelo es la temperatura media anual, representada en el eje X, por el contrario, en el eje Y, pernoctaciones hoteleras, nuestra variable dependiente, para las 17 CCAA en el año 2019.

La relación existente entre ambas variables es directa, ya que la pendiente del gráfico asciende, a medida que existe mayor temperatura en un territorio, mayores pernoctaciones hoteleras se encuentran en el territorio.

El coeficiente de correlación que aparece es un valor, aunque un poco mayor que en la anterior variable, sigue siendo próximo a 0'4. Concluyendo con una correlación moderada-baja o inexistente.

Coefficiente de Correlación = 0,461342

$$\text{PERNOCTACIONES HOTELERAS} = -1,21711\text{E}7 + 1,06443\text{E}6 \cdot \text{Temperatura media anual}$$



35- Gráfico X-Y de la variable TEMP. MED.

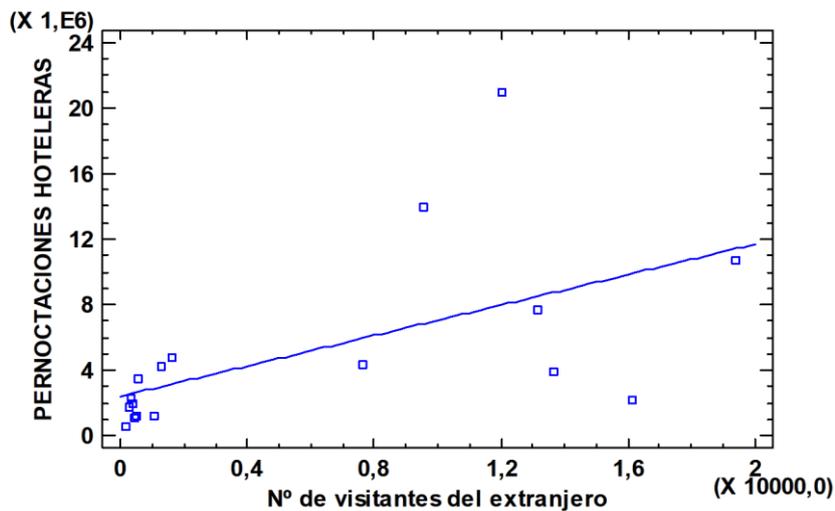
A continuació, se mostra en el gràfic de dispersió la variable n.º de visitants del estranger, en el eje X, y la variable dependiente pernoctacions hoteleras en el eje Y, para cada una de las comunidades autónomas en el año 2019.

La relació existent entre las dos variables del modelo es directa, ya que, como en anteriores ocasiones, la pendiente es ascendente, esto significa que, a medida que la variable representada en el eje X aumenta, la variable Y también aumenta.

El coeficiente de correlación es de 0'57, lo que significa un índice moderado, al situarse dentro del rango 0'4 y 0'7.

Coeficiente de Correlación = 0,572608

$$\text{PERNOCTACIONES HOTELERAS} = 2,3967E6 + 464,722 \cdot \text{N}^\circ \text{ de visitantes del extranjero}$$



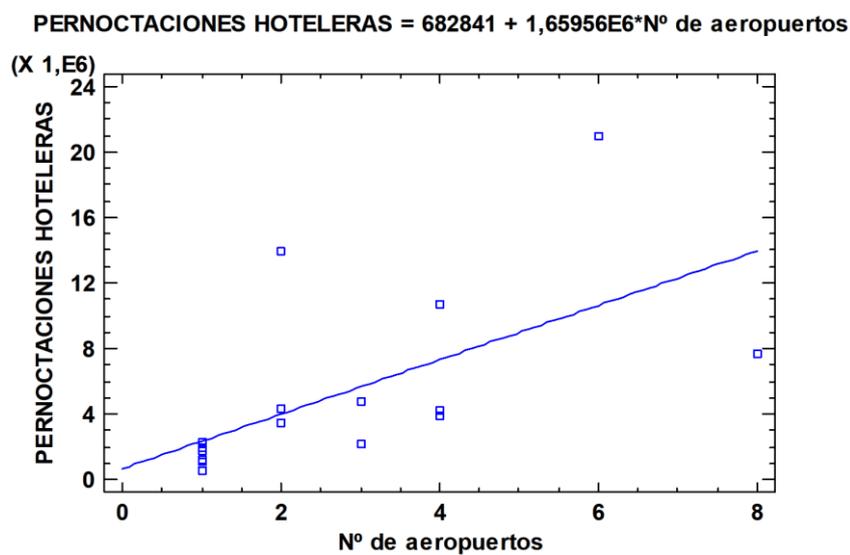
36- Gráfico X-Y de la variable VIST. EXT.

El siguiente gráfico de dispersión, muestra la relación existente entre el n.º de aeropuertos, representado en el eje X y la variable dependiente en el eje Y. Para cada una de las comunidades autónomas en el año 2019.

La pendiente se puede apreciar que es ascendente, dando lugar a una relación directa, ya que cuantos más aeropuertos, más índice de pernoctación hotelera encontramos.

El coeficiente de correlación entre ambas es de 0'61, lo que indica una correlación moderada, al situarse dentro del rango 0'4 y 0'7 explicado anteriormente.

Coefficiente de Correlación = 0,614182



37- Gráfico X-Y de la variable Nº AERO.

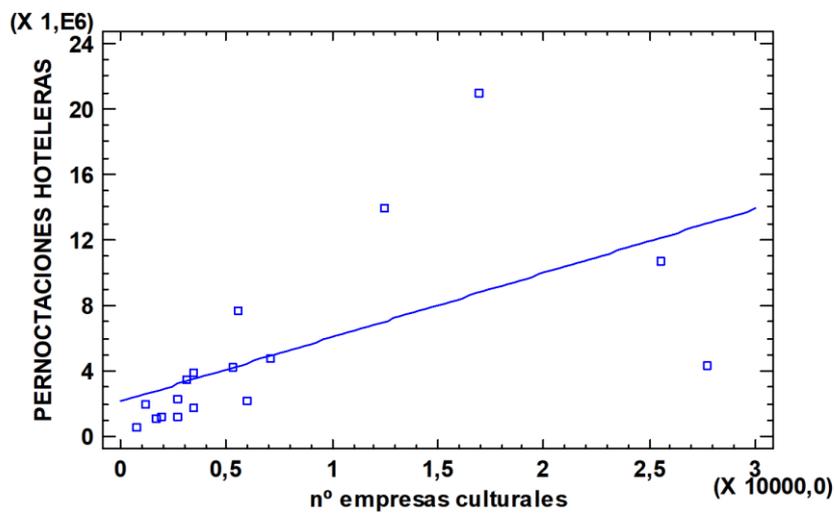
El anàlisis bivariante contínua con midiendo la relación de las siguientes variables. Por un lado, en el eje X, el n.º de empresas culturales. Por otro lado, en el eje Y la variable dependiente del modelo, las pernoctaciones hoteleras.

Como se puede aprecia en el gráfico, la pendiente es ascendente, esto quiere decir que la relación de las variables es directa.

El coeficiente de correlación se sitúa entre 0'4 y 0'7, con lo que consideramos que la correlación es moderada.

Coefficiente de Correlación = 0,59521

$$\text{PERNOCTACIONES HOTELERAS} = 2,14266\text{E}6 + 391,611 \cdot n^{\circ} \text{empresas culturales}$$



38- Gráfico X-Y de la variable Nº EMP. CULT.

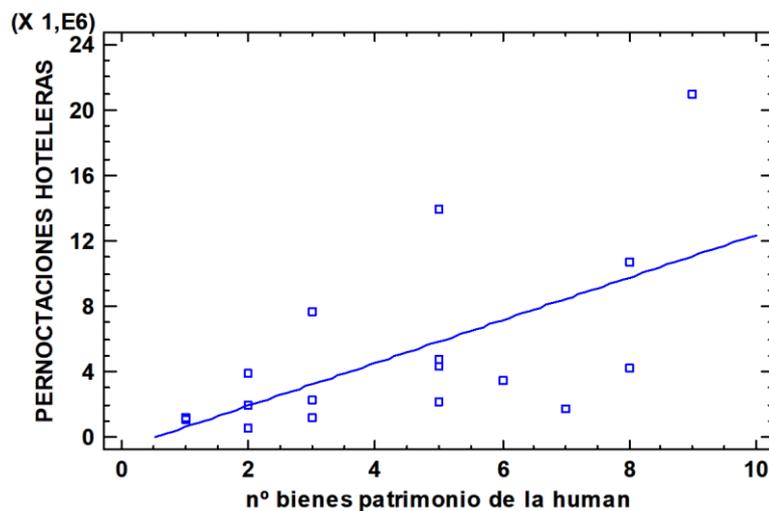
El siguiente gráfico muestra la relación entre la variable dependiente del modelo, pernoctaciones hoteleras, representada en el eje Y, y la variable n.º de bienes declarados Patrimonio de la Humanidad por la UNESCO, para cada una de las comunidades autónomas en el año 2019.

La relación es directa, ya que al tener mayor número de bienes patrimonio de la humanidad, aumentan las pernoctaciones hoteleras.

El coeficiente de correlación es de 0'6, con lo que la correlación es moderada, al situarse entre los valores 0'4 y 0'7.

Coefficiente de Correlación = 0,607528

PERNOCTACIONES HOTELERAS = -685195 + 1,30583E6*nº bienes patrimonio de la human



39- Gráfico X-Y de la variable Nº PAT. HUM.



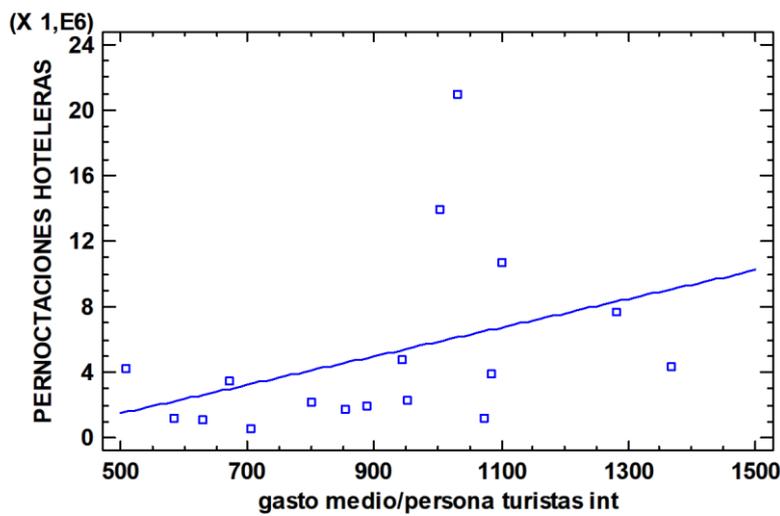
La siguiente variable que se va a analizar es el gasto medio por persona de los turistas internacionales, en el eje X, junto con la variable de las pernoctaciones hoteleras, variable dependiente del modelo, en el eje Y.

La relación entre ambas variables es directa, a medida que asciende la variable X, asciende la variable Y.

El coeficiente de correlación es de 0'38, un valor muy próximo al 0'4 que marca el rango, con lo que podríamos definir como una correlación moderada-baja o inexistente.

Coefficiente de Correlación = 0,383491

$$\text{PERNOCTACIONES HOTELERAS} = -2,86982\text{E}6 + 8728,08 \cdot \text{gasto medio/persona turistas int}$$



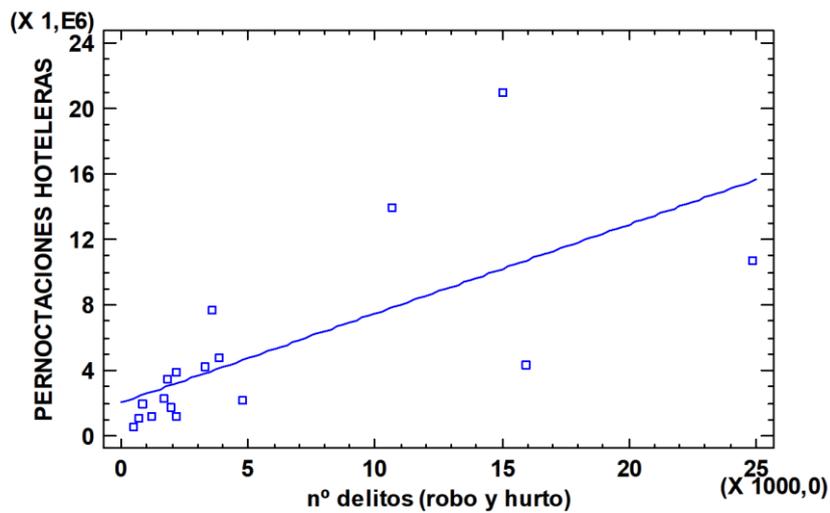
40- Gráfico X-Y de la variable GM. PAX.

En cuanto al análisis bivalente entre las variables n.º de delitos (robo y hurto), representado en el eje X, y la variable dependiente pernoctaciones hoteleras en el eje Y, para las comunidades autónomas en el año 2019. Se puede concluir según los datos ofrecidos por el gráfico de dispersión, que la relación es directa, ya que al aumentar la variable X, aumenta la Y.

El coeficiente de correlación es de un valor muy próximo a 0,7, con lo que la correlación es moderada-alta.

Coefficiente de Correlación = 0,686161

$$\text{PERNOCTACIONES HOTELERAS} = 2,03597\text{E}6 + 543,798 \cdot \text{n}^\circ \text{delitos (robo y hurto)}$$



41- Gráfico X-Y de la variable Nº DEL.

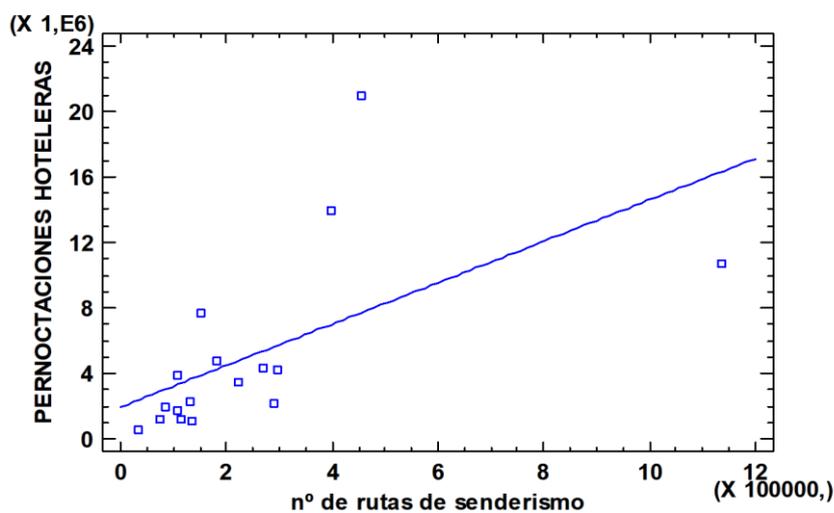


Respecto a la variable X, n.º de rutas de senderismo, y la variable Y dependiente del estudio, pernoctaciones hoteleras, para cada una de las comunidades autónomas en el año 2019. Se puede apreciar en el gráfico de dispersión una relación directa, ya que, al ser la pendiente ascendente, significa que a medida que aumentan las rutas de senderismo, aumentan las pernoctaciones en establecimientos hoteleros.

El coeficiente de correlación se encuentra en el rango de correlación moderada, ya que está entre 0'4 y 0'7.

Coefficiente de Correlación = 0,59296

$$\text{PERNOCTACIONES HOTELERAS} = 1,96603\text{E}6 + 12,6276 \cdot \text{n}^\circ \text{ de rutas de senderismo}$$



42- Gráfico X-Y de la variable Nº SEND.

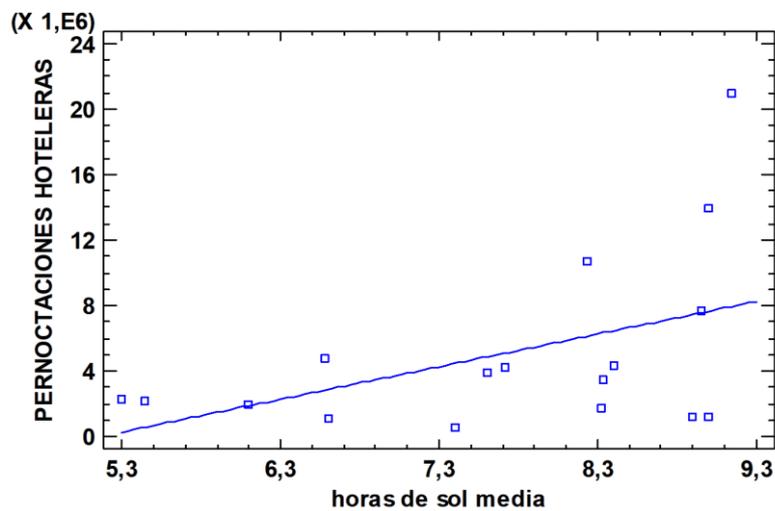
Para terminar con el análisis bivalente, tenemos la variable horas de sol media al día, representada en el eje X y la variable dependiente en el eje Y, pernoctaciones hoteleras. Para cada una de las comunidades autónomas en el año 2019.

Observando el gráfico, la pendiente es ascendente, con lo que se podría concluir que la relación existente entre ambas variables es directa.

El coeficiente de correlación se sitúa entre el rango marcado (0'4 y 0'7) con o que la correlación es moderada.

Coefficiente de Correlación = 0,463154

$$\text{PERNOCTACIONES HOTELERAS} = -1,0274E7 + 1,99166E6 \cdot \text{horas de sol media}$$



43- Gráfico X-Y de la variable H. SOL

3.2.2 Multicolinealidad

A continuació, se va a proceder a realitzar el anàlisi de multicolinealidad. En primer lugar, se realiza la matriz de correlación, donde todos los valores por debajo de la diagonal de 1 superiores a 0'7 indicarán problemas de multicolinealidad.

Tal y como se puede observar en la matriz inferior, la matriz de correlación muestra una relación directa existente entre: el número de empresas culturales y visitantes por lugar de residencia. Entre el número de delitos de robo y hurto con visitantes por lugar de residencia, también entre número de delitos y número de empresas culturales.

Las siguientes relaciones son entre, número de rutas de senderismo y visitantes por lugar de residencia, entre rutas de senderismo y empresas culturales y entre rutas de senderismo y número de delitos. Por último, hay relación entre horas de sol y temperatura media. Todas ellas están señaladas en rojo ya que su relación es superior al 0'7.

Por lo tanto, se concluye que existe multicolinealidad.

	VIS. RES.	PLAYA	TEMP. MED.	VIS. EXT.	Nº AERO.	EMP. CULT.	Nº PAT. HUM.	GM PAX.	Nº DEL.	Nº. SEN.	H. SOL.
VIS. RES.	1	0,09	0,15	0,58	0,35	0,98	0,68	0,52	0,95	0,76	0,37
PLAYA	0,09	1	0,48	0,48	0,69	0,13	0,03	0,54	0,17	0,15	0,05
TEMP. MED.	0,15	0,48	1	0,42	0,50	0,17	-0,11	0,62	0,22	0,08	0,71
VIS. EXT.	0,58	0,48	0,42	1	0,67	0,60	0,34	0,55	0,68	0,66	0,16
Nº AERO.	0,35	0,69	0,50	0,67	1	0,32	0,39	0,42	0,36	0,35	0,31
EMP. CULT.	0,98	0,13	0,17	0,60	0,32	1	0,56	0,60	0,96	0,75	0,34
Nº PAT. HUM.	0,68	0,03	-0,11	0,34	0,39	0,56	1	0,00	0,61	0,63	0,28
GM PAX.	0,52	0,54	0,62	0,55	0,42	0,60	0,00	1	0,53	0,26	0,27
Nº DEL.	0,95	0,17	0,22	0,68	0,36	0,96	0,61	0,53	1	0,89	0,35
Nº. SEN.	0,76	0,15	0,08	0,66	0,35	0,75	0,63	0,26	0,89	1	0,21
H. SOL.	0,37	0,05	0,71	0,16	0,31	0,34	0,28	0,27	0,35	0,21	1

44- Matriz inicial de correlación del modelo, todas las variables. (una variable frente a otra).

El siguiente paso es la matriz inversa, para saber cuál es el grado de relación existente entre una variable, con el resto.

Como se puede observar en la ilustración inferior, las variables visitantes por lugar de residencia, temperatura media, número de empresas culturales, número de bienes patrimonio de la humanidad, gasto medio de los turistas, número de delitos, número de rutas de senderismo y horas medias de sol, poseen una elevada relación con el resto de variables, ya que superan el límite marcado en 10.

Por lo tanto, se concluye, que existen problemas de multicolinealidad.

VIS. RES.	199,46	11,94	-29,80	10,05	-8,34	-137,92	-42,60	20,83	-98,80	53,72	24,21
PLAYA	11,94	3,94	-1,83	1,26	-2,73	-5,37	-2,46	-0,65	-8,03	3,06	2,17
TEMP. MED.	-29,80	-1,83	39,06	-7,06	-3,35	74,01	12,21	-21,67	-38,99	7,88	-25,36
VIS. EXT.	10,05	1,26	-7,06	5,89	-2,42	-12,39	-2,15	3,16	-1,82	-0,01	5,66
Nº AERO.	-8,34	-2,73	-3,35	-2,42	5,54	-4,10	-0,56	1,80	16,62	-5,28	0,66
EMP. CULT.	-137,92	-5,37	74,01	-12,39	-4,10	215,43	36,74	-45,72	-66,71	5,12	-47,33
Nº PAT. HUM.	-42,60	-2,46	12,21	-2,15	-0,56	36,74	13,06	-6,03	11,54	-9,42	-8,86
GM PAX.	20,83	-0,65	-21,67	3,16	1,80	-45,72	-6,03	16,78	16,09	-0,05	13,84
Nº DEL.	-98,80	-8,03	-38,99	-1,82	16,62	-66,71	11,54	16,09	219,59	-85,77	16,57
Nº. SEN.	53,72	3,06	7,88	-0,01	-5,28	5,12	-9,42	-0,05	-85,77	39,91	-1,62
H. SOL.	24,21	2,17	-25,36	5,66	0,66	-47,33	-8,86	13,84	16,57	-1,62	18,30

45- Matriz inversa de correlación inicial, todas las variables. (una variable frente al resto)

Para terminar, se realiza el índice de acondicionamiento (I.C.), en él, se puede ver la relación de todas las variables independientes a la vez (variables X). Se calcula la tabla de autovalores y se procede a aplicar la fórmula.

	AUTOVALORES
VIS. RES.	5,5800
PLAYA	2,2379
TEMP. MED.	1,2223
VIS. EXT.	0,9519
Nº AERO.	0,4575
EMP. CULT.	0,2752
Nº PAT. HUM.	0,1388
GM PAX.	0,0934
Nº DEL.	0,0375
Nº. SEN.	0,0030
H. SOL.	0,0025

Tabla 14: Autovalores de todas las variables

$$I. C. = \sqrt{\frac{\text{Autovalor Max}}{\text{Autovalor Min}}} = \sqrt{\frac{5,58}{0,0025}} = 46'87$$



Una vez aplicada la fórmula se puede ver que existe multicolinealidad elevada ya que el índice de acondicionamiento supera 10 con un valor mucho mayor, 46'87. Una vez realizados los 3 pasos, se ve claramente que existen problemas de multicolinealidad., donde la solución pasa por proceder a la eliminación de variables explicativas del modelo.

Para determinar que variable se ha de eliminar, hay que situarse sobre la matriz inversa de correlación, donde cada valor superior a 10, además de significar problemas de multicolinealidad, son candidatos a eliminar. Eliminando finalmente la que posea ese número más elevado. En este caso N.º de delitos por robo y hurto. Se procederá a repetir estos 3 pasos (matriz de correlación, matriz inversa e índice de acondicionamiento), hasta que no existan problemas de multicolinealidad.

Se vuelve a realizar la matriz de correlación para determinar si hay problemas con el nuevo modelo propuesto sin la variable n.º de delitos.

En este caso, tal y como se observa en la matriz inferior, la relación existente con valores superiores a 0'7, indicando problemas de multicolinealidad, son: entre la variable empresas culturales y visitantes por lugar de residencia. Entre rutas de senderismo y visitantes por lugar de residencia, entre rutas de senderismo y empresas culturales. Por último, entre horas de sol y temperatura media.

Se concluye que existe problema de multicolinealidad.

	VIS. RES.	PLAYA	TEMP. MED.	VIS. EXT.	Nº AERO.	EMP. CULT.	Nº PAT. HUM.	GM PAX.	Nº SEN.	H. SOL
VIS. RES.	1	0,09	0,15	0,58	0,35	0,98	0,68	0,52	0,76	0,37
PLAYA	0,09	1	0,48	0,48	0,69	0,13	0,03	0,54	0,15	0,05
TEMP. MED.	0,15	0,48	1	0,42	0,50	0,17	-0,11	0,62	0,08	0,71
VIS. EXT.	0,58	0,48	0,42	1	0,67	0,60	0,34	0,55	0,66	0,16
Nº AERO.	0,35	0,69	0,50	0,67	1	0,32	0,39	0,42	0,35	0,31
EMP. CULT.	0,98	0,13	0,17	0,60	0,32	1	0,56	0,60	0,75	0,34
Nº PAT. HUM.	0,68	0,03	-0,11	0,34	0,39	0,56	1	0,00	0,63	0,28
GM PAX.	0,52	0,54	0,62	0,55	0,42	0,60	0,00	1	0,26	0,27
Nº SEN.	0,76	0,15	0,08	0,66	0,35	0,75	0,63	0,26	1	0,21
H. SOL	0,37	0,05	0,71	0,16	0,31	0,34	0,28	0,27	0,21	1

46- Matriz de correlación 2 sin la variable Nº DEL.

A continuación, se procede a realizar la matriz inversa de correlación. Se puede apreciar que las variables: visitantes por lugar de residencia, temperatura media, empresas culturales, número de bienes patrimonio de la humanidad, gasto medio por turistas y horas de sol, están muy relacionadas con el resto de variables.

Como hay variables que superan el límite de 10, e concluye que existe problemas de multicolinealidad, siendo también candidatas a eliminar.

VIS. RES.	155,01	8,33	-47,35	9,23	-0,86	-167,93	-37,41	28,07	15,13	31,67
PLAYA	8,33	3,65	-3,26	1,19	-2,12	-7,81	-2,03	-0,06	-0,08	2,78
TEMP. MED.	-47,35	-3,26	32,13	-7,39	-0,40	62,16	14,26	-18,82	-7,35	-22,42
VIS. EXT.	9,23	1,19	-7,39	5,88	-2,28	-12,95	-2,05	3,30	-0,72	5,80
Nº AERO.	-0,86	-2,12	-0,40	-2,28	4,28	0,95	-1,43	0,58	1,21	-0,59
EMP. CULT.	-167,93	-7,81	62,16	-12,95	0,95	195,17	40,24	-40,84	-20,94	-42,30
Nº PAT. HUM.	-37,41	-2,03	14,26	-2,05	-1,43	40,24	12,46	-6,88	-4,91	-9,73
GM PAX.	28,07	-0,06	-18,82	3,30	0,58	-40,84	-6,88	15,60	6,24	12,63
Nº SEN.	15,13	-0,08	-7,35	-0,72	1,21	-20,94	-4,91	6,24	6,41	4,85
H. SOL	31,67	2,78	-22,42	5,80	-0,59	-42,30	-9,73	12,63	4,85	17,05

47- Matriz inversa de correlación 2, sin la variable Nº DEL

Por último, antes de eliminar otra variable, se procede a realizar el índice de acondicionamiento (I.C.), para ver la relación de todas las variables X a la vez.

	AUTOVALORES
VIS. RES.	4,7572
PLAYA	2,1173
TEMP. MED.	1,2205
VIS. EXT.	0,9221
Nº AERO.	0,4523
EMP. CULT.	0,2647
Nº PAT. HUM.	0,1359
GM PAX.	0,0933
Nº SEN.	0,0342
H. SOL	0,0026

Tabla 15: Autovalores sin variable Nº DEL.

$$I. C. = \sqrt{\frac{\text{Autovalor Max}}{\text{Autovalor Min}}} = \sqrt{\frac{4,7572}{0,0026}} = 43'17$$

Como el valor del I.C. es superior a 10, con un valor de 43'17, se concluye que hay problemas de multicolinealidad.

Por lo tanto, observando la matriz inversa, la variable a eliminar del proceso es número de empresas culturales.

Se realiza de nuevo todo el proceso para detectar los posibles problemas de multicolinealidad en el modelo, esta vez sin la variable número de empresas culturales en el modelo.

En la matriz inferior, se puede observar la relación existente entre la variable rutas de senderismo y visitantes por lugar de residencia. También la relación entre horas de sol y temperatura media. Se concluye que hay problemas de multicolinealidad, ya que hay valores por encima de 0'7.

	VIS. RES.	PLAYA	TEMP. MED.	VIS. EXT.	Nº AERO	Nº PAT. HUM.	GM PAX	Nº SEN.	H. SOL
VIS. RES.	1	0,09	0,15	0,58	0,35	0,68	0,52	0,76	0,37
PLAYA	0,09	1	0,48	0,48	0,69	0,03	0,54	0,15	0,05
TEMP. MED.	0,15	0,48	1	0,42	0,50	-0,11	0,62	0,08	0,71
VIS. EXT.	0,58	0,48	0,42	1	0,67	0,34	0,55	0,66	0,16
Nº AERO	0,35	0,69	0,50	0,67	1	0,39	0,42	0,35	0,31
Nº PAT. HUM.	0,68	0,03	-0,11	0,34	0,39	1	0,00	0,63	0,28
GM PAX	0,52	0,54	0,62	0,55	0,42	0,00	1	0,26	0,27
Nº SEN.	0,76	0,15	0,08	0,66	0,35	0,63	0,26	1	0,21
H. SOL	0,37	0,05	0,71	0,16	0,31	0,28	0,27	0,21	1

48- Matriz de correlación 3, si variable Nº DEL y Nº EMP CUL

A continuación, se procede a realizar la matriz inversa de correlación del modelo propuesto para ver el grado de relación tiene una variable con el resto del modelo.

En este caso, se observa en la matriz inferior, que los números no son tan elevados como en los modelos anteriores, aun así, sigue existiendo problemas de multicolinealidad, ya que la variable, visitantes por lugar de residencia y temperatura media, superan el límite marcado en 10.

VIS. RES.	10,52	1,61	6,14	-1,91	-0,05	-2,79	-7,07	-2,88	-4,73
PLAYA	1,61	3,34	-0,77	0,67	-2,08	-0,42	-1,70	-0,91	1,08
TEMP. MED.	6,14	-0,77	12,33	-3,26	-0,70	1,44	-5,81	-0,68	-8,94
VIS. EXT.	-1,91	0,67	-3,26	5,02	-2,22	0,62	0,59	-2,11	2,99
Nº AERO	-0,05	-2,08	-0,70	-2,22	4,28	-1,62	0,78	1,31	-0,39
Nº PAT. HUM.	-2,79	-0,42	1,44	0,62	-1,62	4,16	1,54	-0,60	-1,01
GM PAX	-7,07	-1,70	-5,81	0,59	0,78	1,54	7,06	1,86	3,78
Nº SEN.	-2,88	-0,91	-0,68	-2,11	1,31	-0,60	1,86	4,16	0,32
H. SOL	-4,73	1,08	-8,94	2,99	-0,39	-1,01	3,78	0,32	7,88

49- Matriz inversa de correlación 3, sin variable Nº DEL y Nº EMP CUL.

Por último, se procede a realizar el I.C., para ver la relación de las variables X del modelo todas a la vez, aplicando la fórmula se obtiene:

	AUTOVALORES
VIS. RES.	4,116
PLAYA	1,945
TEMP. MED.	1,212
VIS. EXT.	0,795
Nº AERO	0,434
Nº PAT. HUM.	0,264
GM PAX	0,114
Nº SEN.	0,085
H. SOL	0,034

Tabla 16: Autovalores sin variable Nº DEL y Nº EMP CUL.

$$I.C. = \sqrt{\frac{\text{Autovalor Max}}{\text{Autovalor Min}}} = \sqrt{\frac{4,116}{0,034}} = 10,98$$

Se concluye que hay problema de multicolinealidad en grado moderado ya que se supera el límite fijado en 10 por apenas 1 unidad (10'98).

En este caso, se ha podido comprobar que en los 3 pasos no hay tantos números marcados con problemas de multicolinealidad. Aun así, el modelo sigue dando problemas de multicolinealidad en los 3 pasos realizados. Se procederá a eliminar la variable temperatura media del modelo para ver si los problemas de multicolinealidad desaparecen.

Se realiza la matriz de correlación nuevamente, esta vez sin la variable temperatura media, para saber qué relación existe entre variables. En este caso están relacionadas con un valor superior a 0'7 las variables número de rutas de senderismo con visitantes por lugar de residencia.

Por lo tanto, se concluye, que existen problemas de multicolinealidad.

	VIS. RES.	PLAYA	VIST. EXT.	Nº AERO	Nº PAT. HUM.	GM. PAX.	Nº SEN	H. SOL
VIS. RES.	1	0,09	0,58	0,35	0,68	0,52	0,76	0,37
PLAYA	0,09	1	0,48	0,69	0,03	0,54	0,15	0,05
VIST. EXT.	0,58	0,48	1	0,67	0,34	0,55	0,66	0,16
Nº AERO	0,35	0,69	0,67	1	0,39	0,42	0,35	0,31
Nº PAT. HUM.	0,68	0,03	0,34	0,39	1	0,00	0,63	0,28
GM. PAX.	0,52	0,54	0,55	0,42	0,00	1	0,26	0,27
Nº SEN	0,76	0,15	0,66	0,35	0,63	0,26	1	0,21
H. SOL	0,37	0,05	0,16	0,31	0,28	0,27	0,21	1

50- Matriz de correlación 4, sin variable Nº DEL, Nº EMP CUL y TEMP MED.

A continuación, se procede a la matriz inversa de correlación, para saber cuánto está relacionada una variable con el resto de variables del modelo. En este caso, se observa en la matriz inferior que las variables tienen todos los valores de la diagonal dentro del rango, es decir, inferior a 10. Por lo tanto, se concluye que no hay problemas de multicolinealidad.

VIS. RES.	7,46	1,99	-0,29	0,30	-3,50	-4,17	-2,54	-0,27
PLAYA	1,99	3,29	0,47	-2,12	-0,33	-2,06	-0,96	0,53
VIST. EXT.	-0,29	0,47	4,15	-2,41	1,00	-0,95	-2,29	0,63
Nº AERO	0,30	-2,12	-2,41	4,24	-1,54	0,45	1,27	-0,90
Nº PAT. HUM.	-3,50	-0,33	1,00	-1,54	3,99	2,22	-0,52	0,03
GM. PAX.	-4,17	-2,06	-0,95	0,45	2,22	4,32	1,53	-0,44
Nº SEN	-2,54	-0,96	-2,29	1,27	-0,52	1,53	4,13	-0,18
H. SOL	-0,27	0,53	0,63	-0,90	0,03	-0,44	-0,18	1,40

51- Matriz inversa sin las variables Nº DEL, Nº EMP CUL y TEMP MED

Por último, se procede a realizar el I.C. para determinar la relación existente entre todas las variables X a la vez.

	AUTOVALORES
VIS. RES.	3,816
PLAYA	1,619
VIST. EXT.	0,933
Nº AERO	0,795
Nº PAT. HUM.	0,398
GM. PAX.	0,262
Nº SEN	0,105
H. SOL	0,072

Tabla 17: Autovalores sin variable Nº DEL, Nº EMP CUL y TEMP MED

$$I.C. = \sqrt{\frac{\text{Autovalor Max}}{\text{Autovalor Min}}} = \sqrt{\frac{3,816}{0,072}} = 7,29$$

Una vez realizado el I.C. se observa que no hay problemas de multicolinealidad, ya que el valor obtenido es inferior a 10, siendo este de 7,29.

A pesar de que se cumplen 2 de los requisitos para detectar problemas de multicolinealidad, en la matriz de correlación se puede ver como hay un valor por encima de 0,7. Por este motivo, a pesar de que en la matriz inversa todos los valores cumplen con el valor inferior a 10, se procede a eliminar del modelo la variable, visitantes por lugar de residencia.

Nuevamente, se procede a realizar la matriz de correlación, esta vez sin la variable visitantes por lugar de residencia. Para determinar la relación existente entre las variables por parejas y se concluye que no hay problemas de multicolinealidad, ya que ningún valor por debajo de la diagonal supera 0'7.

	PLAYA	VIS. EXT.	Nº AERO	Nº PAT. HUM.	GM PAX.	Nº SEN.	H. SOL
PLAYA	1	0,48	0,69	0,03	0,54	0,15	0,05
VIS. EXT.	0,48	1	0,67	0,34	0,55	0,66	0,16
Nº AERO	0,69	0,67	1	0,39	0,42	0,35	0,31
Nº PAT. HUM.	0,03	0,34	0,39	1	0,00	0,63	0,28
GM PAX.	0,54	0,55	0,42	0,00	1	0,26	0,27
Nº SEN.	0,15	0,66	0,35	0,63	0,26	1	0,21
H. SOL	0,05	0,16	0,31	0,28	0,27	0,21	1

52- Matriz de correlación final con las variables restantes del modelo

A continuación, se procede a realizar la matriz inversa de correlación para ver en qué grado están relacionadas las variables con las demás, y se observa en la matriz inferior que ninguna supera el límite fijado en 10. Por lo tanto, se concluye que no hay problemas de multicolinealidad.

PLAYA	2,75	0,55	-2,20	0,60	-0,95	-0,28	0,60
VIS. EXT.	0,55	4,14	-2,40	0,87	-1,11	-2,39	0,62
Nº AERO	-2,20	-2,40	4,22	-1,40	0,61	1,37	-0,89
Nº PAT. HUM.	0,60	0,87	-1,40	2,35	0,26	-1,71	-0,09
GM PAX.	-0,95	-1,11	0,61	0,26	1,99	0,11	-0,59
Nº SEN.	-0,28	-2,39	1,37	-1,71	0,11	3,26	-0,27
H. SOL	0,60	0,62	-0,89	-0,09	-0,59	-0,27	1,39

53- Matriz inversa con las variables restantes del modelo

Por último, se procede a realizar el I.C. para determinar qué relación tienen las variables X todas a la vez. Se obtiene el siguiente resultado:

	AUTOVALORES
PLAYA	3,243
VIS. EXT.	1,432
Nº AERO	0,931
Nº PAT. HUM.	0,689
GM PAX.	0,348
Nº SEN.	0,253
H. SOL	0,103

Tabla 18: Autovalores de las variables restantes del modelo

$$I.C. = \sqrt{\frac{\text{Autovalor Max}}{\text{Autovalor Min}}} = \sqrt{\frac{3,243}{0,103}} = 5,62$$

Se concluye que no hay problemas de multicolinealidad.

Una vez completado el análisis de multicolinealidad, se obtiene el modelo final, ya que se han realizado los 3 análisis para detectar problemas de multicolinealidad y en ninguno de los 3 pasos existen problemas, por lo tanto, se pasará a realizar el análisis multivariante con las variables restantes: kilómetros de playa, visitantes del extranjero, n.º de aeropuertos, n.º de bienes declarados patrimonio de la humanidad, gasto medio por persona, n.º de rutas de senderismo y horas medias de sol.

3.3 Multivariante

3.3.1 Ajuste del Modelo

A continuación, una vez realizado el modelo de regresión lineal múltiple se muestran los resultados, para describir la variable dependiente o Y pernoctaciones hoteleras (PH) y las 8 variables explicativas o independientes:

$$PH = -11136300 + 2699,22 * PLAYA + 88,3637 * VIS. EXT. + 132334 * N^{\circ} AERO + 786669 * N^{\circ} PAT. HUM. + 1002,59 * GM PAX. + 3,47757 * N^{\circ} SEN + 1152580 * H. SOL$$

R-cuadrada = 67,218 %

3.3.2 Validación

3.3.2.1 Significatividad

SIGNIFICATIVIDAD DE PARÁMETROS

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	-1,11363E7	7,22953E6	-1,54039	0,1579
PLAYA	2699,22	3111,18	0,867588	0,4082
VIS. EXT.	88,3637	315,256	0,280292	0,7856
N ^o AERO	132334,	1,05982E6	0,124864	0,9034
N ^o PAT. HUM.	786669,	628623,	1,25142	0,2423
GM PAX.	1002,59	6125,91	0,163663	0,8736
N ^o SEN	3,47757	7,33662	0,474002	0,6468
H. SOL	1,15258E6	966514,	1,19251	0,2636

Tabla 19: Parámetros del Modelo

Análisis de la constante (β_0)

H₀: $\beta_0 = 0$

H₁: $\beta_0 \neq 0$

Puesto que el p-valor es 0'1579 que es mayor que alfa 0'05, se acepta la hipótesis nula (H₀) y se concluye que la constante no es significativa.

Análisis de la variable Playa

H₀: $\beta_1 = 0$

H₁: $\beta_1 \neq 0$

Dado que el p-valor es de 0'4082 que es mayor que alfa 0'05, se acepta la hipótesis nula (H₀) y se puede concluir que la variable playa no es significativa.

Análisis de la variable Vis. Ext.

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

Debido a que el p-valor de la variable visitantes del extranjero es 0'7856, siendo mayor que alfa 0'05, se acepta la hipótesis nula (H_0) y se puede decir que la variable, visitantes del extranjero no es significativa.

Análisis Nº de Aeropuertos

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0$$

El p-valor de la variable Nº de aeropuertos es 0'9034 siendo así mayor que alfa 0'05, se puede afirmar que la variable no es significativa, ya que se acepta la hipótesis nula (H_0).

Análisis Nº de Bienes declarados Patrimonio de la Humanidad

$$H_0: \beta_4 = 0$$

$$H_1: \beta_4 \neq 0$$

Como en la variable Nº de bienes declarados patrimonio de la humanidad, el p-valor es 0'2423 siendo de esta manera mayor a alfa 0'05, se acepta la hipótesis nula (H_0) y se concluye que la variable no es significativa.

Análisis Gasto Medio por personas

$$H_0: \beta_5 = 0$$

$$H_1: \beta_5 \neq 0$$

En la variable gasto medio por persona el p-valor es 0'8736, sobrepasando el valor de alfa 0'05, se acepta la hipótesis nula (H_0) y se afirma que la variable gasto medio por persona no es significativa.

Análisis Nº de Rutas de Senderismo

$$H_0: \beta_6 = 0$$

$$H_1: \beta_6 \neq 0$$

Puesto que la variable nº de rutas de senderismo posee un p-valor de 0'6468, superando excesivamente el valor de alfa 0'05, se puede concluir que la hipótesis nula (H_0) se acepta y se concluye que la variable nº de rutas de senderismo no es significativa.

Análisis Horas de Sol al día

$$H_0: \beta_7 = 0$$

$$H_1: \beta_7 \neq 0$$

Por último, la variable horas de sol al día tiene un p-valor de 0'2636, superando así el valor de alfa situado en 0'05, por lo tanto, se acepta la hipótesis nula (H_0) y se concluye que la variable horas de sol al día no es significativa.

SIGNIFICATIVIDAD DEL MODELO

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	3,23332E14	7	4,61903E13	2,64	0,0885
Residuo	1,57688E14	9	1,75209E13		
Total (Corr.)	4,8102E14	16			

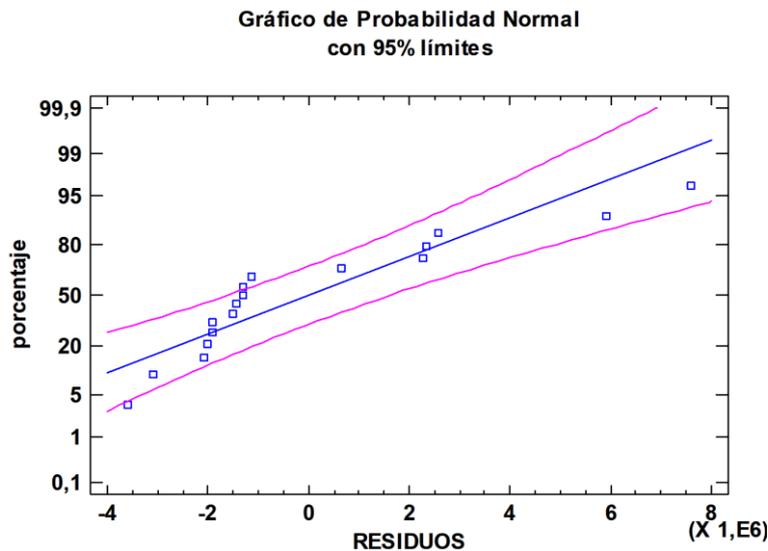
Tabla 20: Significatividad del Modelo

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

$H_1: \text{Al menos 1 es } \neq 0$

Puesto que el p-valor es 0'0885 que es mayor que alfa 0'05, se acepta la hipótesis nula (H_0) y se concluye que el modelo no es significativo.

3.3.2.2 Normalidad de residuos



54- Gráfico de Normalidad de Residuos

Como se puede observar en la tabla, la mayoría de los puntos están situados sobre la línea de normalidad. También se puede apreciar que algunos de los puntos están alejados de la línea citada anteriormente, pero se encuentran dentro o próximos a las líneas de confianza al 95%.

3.3.2.3 Contraste de hipótesis

Prueba	Estadístico	Valor-P
Chi-Cuadrado	27,4706	0,00116874
Estadístico W de Shapiro-Wilk	0,839519	0,00666305
Valor-Z para asimetría	1,60726	0,107998
Valor-Z para curtosis	Datos Insuficientes	

Tabla 21: Contraste de Hipótesis para Normalidad de Residuos

H_0 : Los residuos son normales

H_1 : Los residuos no son normales

Como el p-valor más pequeño obtenido en las pruebas de normalidad para residuos es “Chi-cuadrado”, que es menor que Alpha, se rechaza la hipótesis nula, concluyendo que los residuos no siguen una distribución normal. De esta manera se pueden apreciar problemas de normalidad de residuos que posteriormente se analizarán y se resolverán con un análisis más exhaustivo.

3.3.2.4 Heterocedasticidad

$$\text{RESIDUOS}^2 = -2,94663\text{E}13 + 7,62597\text{E}9 * \text{PLAYA} + 4,88637\text{E}8 * \text{VIS. EXT.} - 1,92926\text{E}12 * \text{N}^\circ \text{ AERO} + 3,23531\text{E}12 * \text{N}^\circ \text{ PAT. HUM.} - 1,55772\text{E}9 * \text{GM PAX.} - 1,00301\text{E}7 * \text{N}^\circ \text{ SEN} + 3,52565\text{E}12 * \text{H. SOL}$$

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	-2,94663E13	2,70426E13	-1,08962	0,3042
PLAYA	7,62597E9	1,16376E10	0,655286	0,5287
VIS. EXT.	4,88637E8	1,17924E9	0,414366	0,6883
Nº AERO	-1,92926E12	3,96435E12	-0,486653	0,6381
Nº PAT. HUM.	3,23531E12	2,35141E12	1,3759	0,2021
GM PAX.	-1,55772E9	2,29144E10	-0,0679796	0,9473
Nº SEN	-1,00301E7	2,74432E7	-0,365485	0,7232
H. SOL	3,52565E12	3,61532E12	0,975197	0,3549

Tabla 22: Heterocedasticidad, donde se muestra el p-valor de cada parámetro

$H_0: \sigma^2 = \text{constante} \rightarrow$ La variable no genera Heterocedasticidad

$H_1: \sigma^2 \neq \text{constante} \rightarrow$ La variable genera Heterocedasticidad

Análisis Variable Playa

Puesto que p-valor es mayor que Alpha ($0'5287 > 0'05$) se acepta la hipótesis nula y se puede concluir que la variable playa no genera problemas de heterocedasticidad.

Análisis Variable Visitantes del Extranjero

Al visualizar que el p-valor de la variable es $0'6883$ es mayor a alfa $0'05$, se acepta la hipótesis nula y se concluye que no existen problemas de heterocedasticidad en el modelo.

Análisis Nº de aeropuertos

Como se puede observar en la tabla el p-valor de la variable es $0'6381$, superando el valor de alfa situado en $0'05$, de este modo se acepta la hipótesis nula y se concluye que el modelo no genera heterocedasticidad.

Análisis Nº de Bienes Patrimonio de la Humanidad

El p-valor de la variable nº de bienes patrimonio de la humanidad es $0'2021$ siendo mayor que el valor referente de alfa $0'05$, concluyendo de esta manera que la variable no genera problemas de heterocedasticidad ya que se acepta la hipótesis nula.

Análisis Gasto Medio por personas

En la variable gasto medio se puede ver que el p-valor es $0'9473$, siendo de esta manera excesivamente superior al valor de alfa $0'05$, de este modo se acepta la hipótesis nula y se concluye que no hay problemas de heterocedasticidad.



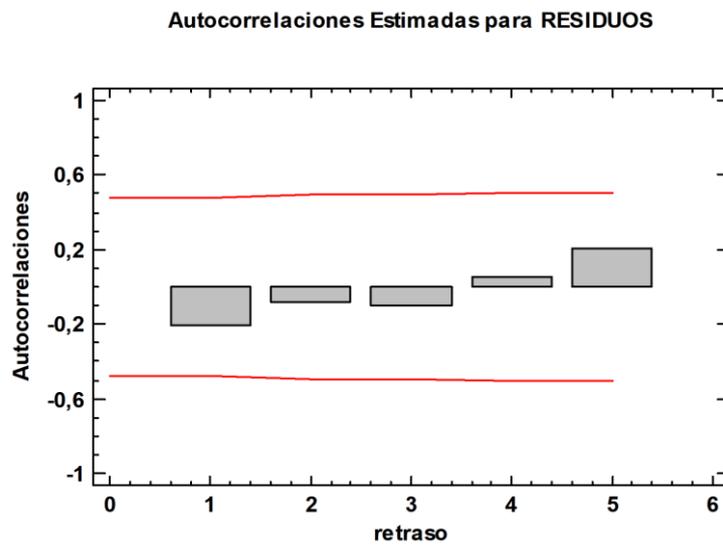
Análisis Nº de Rutas de Senderismo

Puesto que el p-valor de la variable es 0'7232, superando de esta manera el valor de alfa 0'05, se acepta la hipótesis nula y se puede afirmar que no existen problemas de heterocedasticidad.

Análisis Horas de sol al día

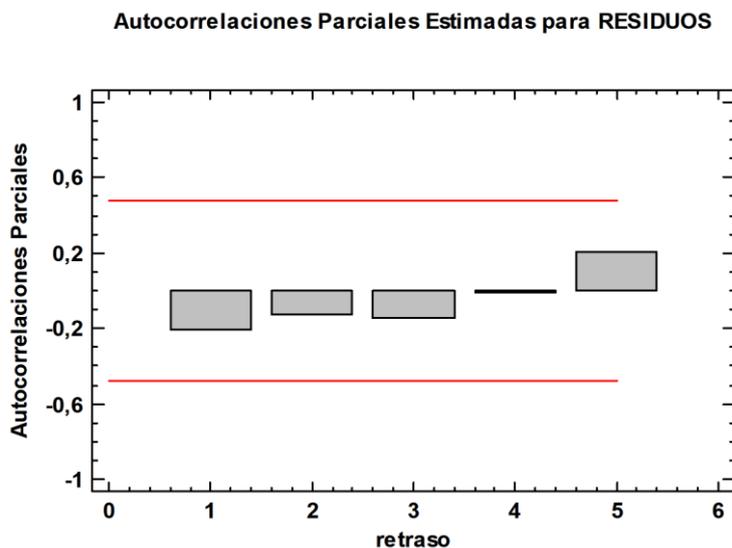
Para terminar el análisis de heterocedasticidad de las variables, se observa que el p-valor de la variable es 0'3549, siendo de esta manera superior al valor de alfa 0'05, se puede concluir que la variable no genera problemas de heterocedasticidad ya que se acepta la hipótesis nula.

3.3.2.5 Autocorrelación



55- Gráfico FAS del modelo

Tal y como se puede observar en ambos gráficos (FAS y FAP), ningún factor de autocorrelación supera los límites ni superior ni inferior, se puede concluir que no existe problema de autocorrelación de ningún orden.



56- Gráfico FAP del modelo

3.3.2.6 Puntos Influyentes

A priori

Fila	Influencia (n_{ii})		
Andalucía	0,360854	$<2\bar{n}$	No influyente a priori
Canarias	0,764987	$<2\bar{n}$	No influyente a priori
Cataluña	0,887475	$<2\bar{n}$	No influyente a priori
Galicia	0,753282	$<2\bar{n}$	No influyente a priori

Tabla 23: Puntos Influyentes A Priori

Influencia media de un solo punto = 0,470588

Como $2 \cdot 0,470588$ es mayor a la influencia de cualquiera de las comunidades, se puede concluir que no hay ningún punto influyente a priori.

A posteriori

Fila	DFITS		
Andalucía	2,45674	$>2\sqrt{\bar{n}}$	Influyente a posteriori
Canarias	-1,68741	$>2\sqrt{\bar{n}}$	Influyente a posteriori
Cataluña	-8,59714	$>2\sqrt{\bar{n}}$	Influyente a posteriori
Galicia	-1,76434	$>2\sqrt{\bar{n}}$	Influyente a posteriori

Tabla 24: Puntos Influyentes A Posteriori

Influencia media de un solo punto = 0,470588

$2 \cdot \text{Raíz de influencia media} = 1,37198833814$

Como todos los DFITS son mayores a la raíz de influencia media por 2, se puede decir que los puntos son influyentes a posteriori.

3.3.2.7 Puntos Anómalos

Fila	Residuo Estudentizado		
Andalucía	3,27	> 3	Candidato a eliminar
Cataluña	-3,06	> 3	Candidato a eliminar

Tabla 25: Puntos Anómalos candidatos a eliminar

Se observa que ambas comunidades tienen residuos estudentizados en valor absoluto mayor a 3, por lo que son candidatas a ser eliminadas del modelo.

3.3.3 Solución de Problemas

Como se ha podido observar en el estudio del modelo, se han obtenido problemas de significatividad, parámetros y modelo, normalidad de residuos y puntos anómalos. Por lo que a continuación, se va a proceder a solucionar los problemas eliminando puntos anómalos. Se han eliminado las CCAA de Andalucía, Comunidad Valenciana y Murcia.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	-3,23467E6	530691,	-6,09519	0,0009
PLAYA	821,875	208,532	3,94123	0,0076
VIS. EXT.	-122,994	21,4765	-5,72691	0,0012
Nº AERO	652755,	75983,6	8,59074	0,0001
Nº PAT. HUM.	-4666,64	46844,5	-0,0996197	0,9239
GM PAX.	2544,87	405,297	6,27903	0,0008
Nº SEN	7,786	0,489478	15,9067	0,0000
H. SOL	192330,	74222,0	2,59129	0,0411

Tabla 26: Significatividad de Parámetros solucionando los problemas

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,00948E14	7	1,44211E13	197,75	0,0000
Residuo	4,37548E11	6	7,29246E10		
Total (Corr.)	1,01385E14	13			

Tabla 27: Significatividad del modelo solucionando los problemas

R-cuadrada = 99,5684%

$$PH = -3,23467E6 + 821,875*PLAYA - 122,994*VIS. EXT. + 652755,*Nº AERO - 4666,64*Nº PAT. HUM. + 2544,87*GM PAX. + 7,786*Nº SEN + 192330,*H. SOL$$

Se observa que sigue apareciendo una variable que no es significativa, nº de bienes patrimonio de la humanidad, ya que su p-valor es 0'9239 y supera el valor de alfa 0'05 y se procede a eliminarla del modelo.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	-3,2431E6	485429,	-6,6809	0,0003
PLAYA	826,74	187,849	4,40108	0,0032
VIS. EXT.	-122,325	18,9037	-6,47099	0,0003
Nº AERO	650449,	67057,9	9,69981	0,0000
GM PAX.	2546,08	375,375	6,78275	0,0003
Nº SEN	7,75462	0,347159	22,3374	0,0000
H. SOL	191632,	68465,1	2,79897	0,0266

Tabla 28: Significatividad de parámetros una vez resueltos los problemas



Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1,00947E14	6	1,68245E13	268,72	0,0000
Residuo	4,38271E11	7	6,26102E10		
Total (Corr.)	1,01385E14	13			

Tabla 29: Significatividad del modelo una vez resuelto los problemas

R-cuadrada = 99,5677%

$PH = -3,2431E6 + 826,74*PLAYA - 122,325*VIS. EXT. + 650449,*N^{\circ} AERO + 2546,08*GM PAX. + 7,75462*N^{\circ} SEN + 191632,*H. SOL$

Se comprueba que al eliminar la variable patrimonio de la humanidad, todas las restantes son significativas. Por último, se va a comprobar la normalidad de residuos.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	15,8571	0,0654711
Estadístico W de Shapiro-Wilk	0,949127	0,519354
Valor-Z para asimetría	0,945327	0,34449
Valor-Z para curtosis	Datos Insuficientes	

Tabla 30: Pruebas de Normalidad de residuos

Por último, para solucionar los problemas en la normalidad de residuos, es decir, si siguen una distribución normal o no. Se escoge el p-valor más bajo, en este caso es el valor de "Chi-cuadrado", al ser 0'0654711 mayor que alfa 0'05 ligeramente, se acepta la hipótesis nula y se concluye que los residuos siguen una distribución normal.

3.4 Predicciones

SUPUESTO 1

¿Qué pasaría si Asturias disminuye un 2% sus km. de playa accesibles o practicables?

La variable Y (Pernoctaciones Hoteleras) sería 2155550.

Con un intervalo de confianza $[1761100 \leq PH \leq 2550000]$ 95%

Se puede observar que al disminuir un 2% los km. de playa accesibles, disminuyen las pernoctaciones hoteleras en la comunidad de Asturias en un 4'31%.

SUPUESTO 2

¿Qué ocurriría si en la Comunidad de Madrid aumenta el número de rutas de senderismo en un 3%?

La variable Y (Pernoctaciones Hoteleras) sería 4374460.

Con un intervalo de confianza $[3877340 \leq PH \leq 4871580]$ 95%

Al aumentar las rutas de senderismo en un 3% por el territorio de la Comunidad de Madrid, se observa que las pernoctaciones hoteleras aumentan en un 2'19%.

SUPUESTO 3

¿Qué sucedería si en Canarias aumentara el gasto medio de los turistas que visitan el archipiélago en un 5%?

La variable Y (Pernoctaciones Hoteleras) sería 7918070.

Con un intervalo de confianza $[7372400 \leq PH \leq 8463730]$ 95%

Al incrementar el gasto medio de los turistas que visitan Canarias en un 5%, se vería reflejado un aumento de un 2'55% en las pernoctaciones hoteleras.

SUPUESTO 4

¿Qué pasaría si Extremadura a fin de mejorar sus comunicaciones aumentara en 1 el número de aeropuertos? ¿Y si Cataluña quisiera reducir las emisiones de carbono cerrando 2 de sus aeropuertos?

La variable Y (Pernoctaciones Hoteleras) sería 1755000.

Con un intervalo de confianza $[1397260 \leq PH \leq 2112750]$ 95%



En caso de que Extremadura apueste en un futuro por el sector turístico, si quisiera mejorar sus comunicaciones tanto con el resto de España y Portugal, como con el resto de países, si aumentara en 1 el número de aeropuertos, se vería reflejado un aumento de un 33'25% el número de pernoctaciones hoteleras.

En cuanto a Cataluña si apostara por una campaña sostenible y cerrara 2 aeropuertos, la variable Y (Pernoctaciones Hoteleras) sería 9454350.

Con un intervalo de confianza de $[8754810 \leq PH \leq 10153900]$ 95%

Las pernoctaciones hoteleras en Cataluña se verían reducidas en un 11'89% si se tomase la decisión de cerrar 2 aeropuertos.

SUPUESTO 5

¿Qué ocurriría con las pernoctaciones hoteleras si en el País Vasco se redujera un 3% el gasto medio de los turistas, pero en ese año se vieran incrementadas un 20% las horas de sol?

La variable Y (Pernoctaciones Hoteleras) sería 2406860.

Con un intervalo de confianza $[1926510 \leq PH \leq 2887200]$ 95%

A pesar de que se reduzca el gasto medio por turista en el País Vasco, se puede observar que las pernoctaciones hoteleras aumentarían un 7'91%, ya que primaría tener más horas de sol en el País Vasco.

SUPUESTO 6

¿Qué sucedería si en la Comunidad Foral de Navarra se declararan 2 bienes (sumados al que ya poseen) Patrimonio de la Humanidad?

La variable Y (Pernoctaciones Hoteleras) sería 1259580.

Con un intervalo de confianza $[1014190 \leq PH \leq 1504960]$ 95%

Las pernoctaciones hoteleras en Navarra se verían incrementadas en un 12'98% si el número de bienes declarados patrimonio de la humanidad pasase de 1 a 3 en dicha comunidad autónoma.



4. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN



4.1 Conclusiones

Para finalizar el proyecto, se procederá a citar una serie de conclusiones a las que se ha llegado a través de la investigación realizada y los datos e información aportada. Siendo las siguientes:

- En variable dependiente del modelo Pernoctaciones Hoteleras (PH) se encuentran 3 puntos anómalos las comunidades autónomas de Cataluña, Comunidad Valenciana y Andalucía.
- En el modelo final obtenido con las variables que han quedado, se explican un 99'5677% de los datos.
- La correlación existente entre las horas medias de sol en las comunidades autónomas de España es directa y en grado moderada.
- Se puede observar que cuyas comunidades autónomas que si disminuyeran los kilómetros de playa accesibles de cuyas comunidades limitan con la costa, las pernoctaciones hoteleras descenderían.
- Se ha comprobado que las comunidades verían incrementadas las pernoctaciones hoteleras, si se habilitan más zonas rurales con posibilidad de efectuar rutas de senderismo.
- Se ha detectado que, si las comunidades mejoraran sus comunicaciones aéreas, bien con el resto de España o con otros países, las pernoctaciones hoteleras aumentarían considerablemente.
- Las variables eliminadas en los análisis de multicolinealidad estaban altamente correlacionadas con las demás, si hubieran seguido perteneciendo al modelo, habrían afectado a la estimación del mismo.
- Se han eliminado los datos de Andalucía, Comunidad Valenciana y Murcia para poder solucionar los problemas en el modelo.



4.2 Futuras Líneas de Investigación

En definitiva, el proyecto realizado tiene la suerte y la ventaja de ser objeto cuyo tema principal es el Turismo. Un sector muy importante en España, además de actual y en constante cambio, de ahí que se habrá la posibilidad de hacer un estudio con diferente enfoque a partir de los datos aquí presentados.

Para poder realizar otros enfoques o estudios relacionados con el tema en cuestión, los procedimientos a seguir son:

En primer lugar, la actualización de la base de datos, ya que el sector turístico es un sector muy sensible y cambiante, por lo que si los años escogidos se remontan a datos antiguos no tendría sentido la investigación.

La búsqueda de nuevas variables explicativas es otro procedimiento clave, para así poder contrastar la información con la presentada en esta investigación. Así como la utilización de otros métodos más precisos para obtener la información más completa, como por el ejemplo, en multicolinealidad hacer uso del método de componentes principales.

Por último, probar distintos modelos (logarítmico, semilogarítmico, logarítmico inversa y el de la inversa), ya que en este proyecto solo se ha hecho uso de la estructura lineal.



5. BIBLIOGRAFÍA

- [1] Ponferrada, M. L. V. (2015). Evolución del turismo en España: el turismo cultural. *International journal of scientific management and tourism*, 1(4), 75-95.
- [2] <https://www.fp-santagema.es/tipos-de-alojamientos-turisticos-en-espana/> (14/02/21)
- [3] <https://economipedia.com/definiciones/varianza.html> (14/02/21)
- [4] Martínez Gómez, M., García, R. R., & Regalado, E. V. (2010). *Introducción a la Estadística*. Editorial UPV.
- [5] <https://www.estadisticaparatodos.es/taller/graficas/cajas.html> (28/02/21)
- [6] <https://www.statgraphics.com/> (18/03/21)
- [7] <https://www.WoltersKluwer.es> (31/03/21)
- [8] <https://www.ine.es/> (16/04/21)
- [9] <https://portal.aena.es/es/aerolineas/red-aeropuertos.html> (25/04/21)
- [10] <https://mediterrani.com/blog/turismo-cultural/> (25/04/21)
- [11] <https://es.unesco.org/> (02/05/21)
- [12] <https://rutacultural.com/senderismo-salud-y-cultura/> (02/05/21)
- [13] <https://www.epdata.es/horas-luz-provincias-2019/6841c77e-e946-420e-a0c4-cff4de1154f3> (23/05/21)
- [14] Chirivella, V. (2005). *Econometría>>>. PID 12059-C, Valencia, España.*



LISTADO DE TABLAS

Tabla 1: Tabla de la base de datos del INE de las variables escogidas para el proyecto	23
Tabla 2: Resumen Estadístico de PH	24
Tabla 3: Resumen Estadístico variable Vis. Res.....	25
Tabla 4: Resumen Estadístico variable KM. PLAYA	27
Tabla 5: Resumen Estadístico variable TEMP. MED.....	28
Tabla 6: Resumen Estadístico VIST. EXT.....	30
Tabla 7: Resumen Estadístico para N.º de Aeropuertos	31
Tabla 8: Resumen Estadístico Nº EMP. CUL.....	33
Tabla 9: Resumen Estadístico Nº PAT. HUM.....	34
Tabla 10: Resumen Estadístico GM. PAX.....	36
Tabla 11: Resumen Estadístico para Nº DEL.....	37
Tabla 12: Resumen Estadístico para Nº SEN.....	39
Tabla 13: Resumen Estadístico para H. SOL.....	40
Tabla 14: Autovalores de todas las variables	54
Tabla 15: Autovalores sin variable Nº DEL.....	56
Tabla 16: Autovalores sin variable Nº DEL y Nº EMP CUL.....	58
Tabla 17: Autovalores sin variable Nº DEL, Nº EMP CUL y TEMP MED.....	59
Tabla 18: Autovalores de las variables restantes del modelo.....	61
Tabla 19: Parámetros del Modelo.....	62
Tabla 20: Significatividad del Modelo	63
Tabla 21: Contraste de Hipótesis para Normalidad de Residuos	64
Tabla 22: Heterocedasticidad, donde se muestra el p-valor de cada parámetro.....	65
Tabla 23: Puntos Influyentes A Priori.....	68
Tabla 24: Puntos Influyentes A Posteriori.....	68
Tabla 25: Puntos Anómalos candidatos a eliminar	68
Tabla 26: Significatividad de Parámetros solucionando los problemas.....	69
Tabla 27: Significatividad del modelo solucionando los problemas	69
Tabla 28: Significatividad de parámetros una vez resueltos los problemas	69
Tabla 29: Significatividad del modelo una vez resuelto los problemas	70
Tabla 30: Pruebas de Normalidad de residuos	70

LISTADO DE ILUSTRACIONES

1- Desglose de Procesos.....	10
2- Representación de los Cuartiles.....	12
3- Tipos de curtosis	12
4- Gráfico de Cajas y Bigotes	13
5- Gráfico de Normalidad de Residuos.....	17
6- Gráfico FAS.....	19
7- Gráfico FAP.....	19
8- Gráfico Predicción, puntual y por intervalos de confianza	21
9- Gráfico de cajas y bigotes para PH.....	24
10- Histograma variable PH.....	25
11- Gráfico de cajas y bigotes variable Vis. Res.	26
12- Histograma variable Vis. Res.....	26



13- Gráfico de cajas y bigotes para KM. PLAYA.....	27
14 Histograma variable KM. PLAYA.....	28
15- Gráfico cajas y bigotes TEMP. MED.	29
16- Histograma variable TEMP. MED.	29
17- Gráfico cajas y bigotes VIST. EXT.....	30
18- Histograma variable VIST. EXT.	31
19- Gráfico de cajas y bigotes variable Nº AERO.	32
20- Histograma variable Nº AERO.	32
21- Gráfico de cajas y bigotes Nº EMP. CUL.....	33
22- Histograma variable EMP. CUL.	34
23- Gráfico de cajas y bigotes Nº PAT. HUM.....	35
24- Histograma variable Nº PAT. HUM.	35
25- Gráfico cajas y bigotes para GM PAX.....	36
26- Histograma variable GM. PAX.....	37
27- Gráfico cajas y bigotes Nº. DEL.	38
28- Histograma variable Nº DEL.....	38
29- Gráfico cajas y bigotes de Nº SEN.....	39
30- Histograma variable Nº SEN.....	40
31- Gráfico cajas y bigotes de H. SOL.....	41
32- Histograma variable H. SOL.....	41
33- Gráfico X-Y de la variable VIS. RES.	42
34- Gráfico X-Y de la variable KM. PLAYA.....	43
35- Gráfico X-Y de la variable TEMP. MED.	44
36- Gráfico X-Y de la variable VIST. EXT.	45
37- Gráfico X-Y de la variable Nº AERO.	46
38- Gráfico X-Y de la variable Nº EMP. CULT.....	47
39- Gráfico X-Y de la variable Nº PAT. HUM.	48
40- Gráfico X-Y de la variable GM. PAX.....	49
41- Gráfico X-Y de la variable Nº DEL.	50
42- Gráfico X-Y de la variable Nº SEND.	51
43- Gráfico X-Y de la variable H. SOL.....	52
44- Matriz inicial de correlación del modelo, todas las variables. (una variable frente a otra). 53	
45- Matriz inversa de correlación inicial, todas las variables. (una variable frente al resto) 54	
46- Matriz de correlación 2 sin la variable Nº DEL.....	55
47- Matriz inversa de correlación 2, sin la variable Nº DEL 56	
48- Matriz de correlación 3, si variable Nº DEL y Nº EMP CUL 57	
49- Matriz inversa de correlación 3, sin variable Nº DEL y Nº EMP CUL..... 57	
50- Matriz de correlación 4, sim variable Nº DEL, Nº EMP CUL y TEMP MED. 58	
51- Matriz inversa sin las variables Nº DEL, Nº EMP CUL y TEMP MED..... 59	
52- Matriz de correlación final con las variables restantes del modelo 60	
53- Matriz inversa con las variables restantes del modelo..... 60	
54- Gráfico de Normalidad de Residuos..... 64	
55- Gráfico FAS del modelo..... 67	
56- Gráfico FAP del modelo..... 67	