



Universidad de Valladolid

Facultad de Ciencias

Trabajo Fin de Grado

Grado en
ESTADÍSTICA

**UN MODELO PARA ESTIMAR
EL PRECIO DE VEHÍCULOS**

Autor:
D. Fernando Garro García

Tutora:
D.^a Cristina Rueda Sabater

RESUMEN

En este trabajo se analiza la relación existente entre las características técnicas de los automóviles y el precio final recomendado por los fabricantes Volkswagen, Peugeot y Renault.

Para conseguir un modelo de estimación del precio de los coches se ha utilizado como método estadístico la regresión lineal.

Las características técnicas que son importantes en la estimación del precio y que aparecen en todos los modelos de regresión son la potencia del motor, la longitud y la velocidad máxima de los vehículos.

El análisis de los datos refleja que eliminando el efecto de las características técnicas, Peugeot es la marca que fija precios más económicos para sus modelos y que, por el contrario, Volkswagen establece los precios más altos. Esas diferencias son aún más acentuadas en los vehículos de gama intermedia y gama alta.

ABSTRACT

This paper analyzes the relationship between the technical features of vehicles and the recommended final prices of the following manufacturers: Volkswagen, Peugeot and Renault.

The multiple regression method has been applied in order to achieve an estimation model for the price of cars.

The technical characteristics relevant for the estimation of prices (present in all regression models) are the motor's power, the length and maximum speed of vehicles.

The analysis of data shows that, ignoring the effect of the technical features, Peugeot is the brand that fixes the lowest rates for its models and, on the contrary, Volkswagen sets the highest ones. These differences are even more notorious for medium and high-end cars.

INDICE

RESUMEN.....	2
ABSTRACT.....	2
1 INTRODUCCIÓN.....	5
2 OBJETIVOS DEL TRABAJO.....	7
3 DESCRIPCIÓN DE LOS DATOS.....	7
3.1 Modelos de Automóvil.....	7
3.2 Descripción de Variables.....	7
4 METODOLOGÍA.....	10
5 MODELO DE REGRESION DE RENAULT.....	12
5.1 Análisis de Variables.....	13
5.2 Resultados de la Regresión de Renault.....	19
5.2.1 Estadísticos Descriptivos.....	19
5.2.2 Matriz de Correlaciones.....	19
5.2.3 Resumen del Modelo.....	20
5.2.4 Análisis de la Varianza.....	21
5.2.5 Coeficientes de la Regresión.....	21
5.2.6 Ecuación Resultante de la Regresión.....	21
5.2.7 Diagnóstico del Modelo de Regresión.....	22
5.2.8 Intervalo de Confianza.....	24
5.3 Interpretación del Modelo de Renault.....	25
6 MODELO DE REGRESION DE PEUGEOT.....	25
6.1 Análisis de Variables.....	26
6.2 Resultados de la Regresión de Peugeot.....	32
6.2.1 Estadísticos Descriptivos.....	32
6.2.2 Matriz de Correlaciones.....	32
6.2.3 Resumen del Modelo de Peugeot.....	33
6.2.4 Análisis de la Varianza.....	34
6.2.5 Coeficientes de la Regresión.....	34
6.2.6 Ecuación de la Regresión.....	34
6.2.7 Diagnóstico del Modelo de Regresión.....	35
6.2.8 Intervalo de Confianza.....	37
6.3 Interpretación del Modelo de Peugeot.....	38
7 MODELO DE REGRESION DE VOLKSWAGEN.....	39
7.1 Análisis de Variables.....	39
7.2 Resultados de la Regresión de Volkswagen.....	45
7.2.1 Estadísticos Descriptivos.....	45
7.2.2 Matriz de Correlaciones.....	46
7.2.3 Resumen del Modelo de Volkswagen.....	46
7.2.4 Análisis de la Varianza.....	47
7.2.5 Coeficientes de Regresión.....	47
7.2.6 Ecuación de Regresión.....	47
7.2.7 Diagnóstico del Modelo de Regresión.....	48
7.2.8 Intervalo de Confianza.....	50
7.3 Interpretación del Modelo de Volkswagen.....	50

8	MODELO DE REGRESION DE VOLKSWAGEN, PEUGEOT Y RENAULT.....	51
8.1	Análisis de Variables.....	51
8.2	Resultados de la Regresión de Marcas Conjuntas.....	55
8.2.1	Estadísticos Descriptivos.....	56
8.2.2	Matriz de Correlaciones.....	56
8.2.3	Resumen del Modelo Conjunto de Marcas.....	57
8.2.4	Análisis de la Varianza.....	57
8.2.5	Coefficientes de la regresión.....	58
8.2.6	Diagnóstico del Modelo de Regresión.....	58
8.2.7	Intervalos de Confianza.....	60
8.3	Análisis ACP.....	61
8.4	Análisis Cluster.....	62
8.4.1	Estudio del Precio de los Coches según la Gama.....	65
8.4.2	Interpretación de Resultados de Regresiones por Gama.....	67
9	CONCLUSIONES.....	67
10	BIBLIOGRAFÍA.....	69

1 INTRODUCCIÓN

Durante el año 2012 se vendieron un total de 81,7 millones de vehículos en todo el mundo, según las estimaciones de OICA (Organización Internacional de Constructores de Automóviles). De ellos, 699.589 automóviles se vendieron en España.

El siguiente cuadro refleja la evolución en las ventas de turismos en España de los últimos 4 años, tendencia decreciente motivada por la crisis actual.

Matriculación de vehículos por tipo y procedencia						
<i>New registration of vehicles by type and origin</i>						
Tipo de Vehículo/ Type	Origen /Origin	2008	2009	2010	2011	2012
Turismos + Todo terreno/ Passenger cars + 4wd (Jeep type)	Nacional / National	266.077	249.622	242.942	180.109	135.382
	Importado UE / Imported EU	762.868	604.798	638.430	547.240	494.713
	Importado no UE / Imported not EU	132.231	98.352	100.643	80.702	69.494
Total		1.161.176	952.772	982.015	808.051	699.589

Según los datos del cuadro anterior ofrecidos por ANFAC, Asociación Nacional de Fabricantes de Automóviles y Camiones, los vehículos adquiridos en España en 2012 provenían en un 90% de origen europeo y de ellos, el 21,5 % eran de fabricación nacional.

En 2012, Volkswagen fue la marca líder en ventas en España con 62.490 unidades, desplazando a Seat, que mantenía históricamente el liderazgo y que cayó al tercer puesto. En segundo lugar por número de ventas le sigue Peugeot con 55.644 automóviles. Ambos fabricantes, Volkswagen y Peugeot, también ocupan las mismas posiciones como líderes de ventas en Europa.

Por modelos, en España el Renault “Megane” fue el más vendido en 2012 por cuarto año consecutivo (30.146 unidades) y es Renault la cuarta marca en importancia por número de ventas.

El mercado del automóvil español evoluciona hacia un mercado más eficiente desde el punto de vista energético. El grupo de turismos de mayor demanda es el situado en el tramo del compacto de matriculaciones que va desde los 120 gr/CO2 hasta los 160 gr/CO2, que representa una cuota del 48%.

En cuanto al combustible utilizado, el mercado de turismos en España sigue siendo esencialmente diésel, con un 68,9% del total de turismos. Sin embargo, el porcentaje de vehículos con motorizaciones de gasolina se va incrementando paulatinamente cada año.

La estrategia comercial de los fabricantes de automóviles en los últimos tiempos ha consistido en diversificar su producción en diversos modelos para los que se amplía su oferta comercial con numerosas versiones para adaptarse a la demanda del mercado.

Los fabricantes de automóviles intentan disminuir parte de los costes de producción utilizando las mismas plataformas para distintos modelos y comercializando versiones con algunas modificaciones sobre las características técnicas (potencia del motor, cilindrada, dimensiones exteriores, etc.) o distintos acabados.

A pesar de la numerosa información que el sector del automóvil revela en cuanto al catálogo disponible de los modelos y versiones de las distintas marcas, comunicación comercial de cada modelo, datos de pruebas técnicas o de conducción realizadas, etc. no existe un estudio estadístico conocido en el que se cuantifique de alguna forma la importancia o peso que sobre el precio final de los vehículos tienen sus características técnicas principales.

En este trabajo se analiza la relación entre las características técnicas de los automóviles y el precio final recomendado por los fabricantes.

El método de estudio estadístico utilizado para conseguir un modelo de estimación del precio de los coches ha sido el de regresión lineal multivariante.

Mediante un Análisis Cluster se van a generar unos grupos de vehículos en relación a sus características técnicas, una nueva variable, para hacer luego análisis por grupos.

Los análisis se han realizado utilizando el software IBM SPSS.

Las marcas seleccionadas para este estudio han sido las correspondientes a los tres fabricantes más representativos de Europa conforme a los datos de ventas de 2012. Por ello, el estudio se ha realizado con las marcas Volkswagen, Peugeot y Renault.

2 OBJETIVOS DEL TRABAJO

El objetivo principal de este trabajo es desarrollar un modelo de predicción del precio de los automóviles de las marcas Volkswagen, Peugeot y Renault en función de las características técnicas de sus coches.

Además se pretende identificar si las marcas objeto de este trabajo establecen precios significativamente diferentes, más caros o baratos que las otras, considerando únicamente esas características técnicas. En caso afirmativo, averiguar si las diferencias son más acusadas segmentando los vehículos por gama.

3 DESCRIPCIÓN DE LOS DATOS

3.1 Modelos de Automóvil

Los vehículos incluidos en el trabajo pertenecen a la categoría de turismos y, por tanto, han quedado fuera del análisis tanto los vehículos comerciales o combi-comerciales como los todoterreno.

Debido a que las versiones de turismos con motorizaciones eléctricas o híbridas (con motor de combustible y motor eléctrico) tienen unas características técnicas muy específicas, además de su escasa representación en porcentaje de ventas, se han excluido igualmente del estudio.

Las fuentes para la confección de la base de datos para este trabajo han sido las revistas especializadas del motor “Altagama Motor” y “Motor Mundial” además de la información general que ofrecen los fabricantes en sus respectivas páginas Web.

La base de datos engloba a un total de 540 versiones diferentes de los modelos que Peugeot, Renault y Volkswagen comercializaban en España en enero de 2013.

3.2 Descripción de Variables

Las variables que han sido consideradas para este estudio, además del precio final de los automóviles, son aquellas referidas a las principales características técnicas de los mismos, conocidas por casi todos los futuros clientes, y que podrían influir en sus decisiones de compra.

A continuación se detallan las variables con el nombre por el que se conocen en el estudio así como su descripción.

a) Variable objetivo o predictora del modelo:

- PVP_MILES: es el precio de venta al público recomendado por los fabricantes, expresado en miles de euros.

Cada una de las marcas de automóviles que se comercializan en nuestro país tiene un listado de precios oficial. Dicho listado se compone de un precio franco fábrica al que hay que sumársele el IVA para obtener como resultado el PVP o Precio de Venta al Público. El PVP no es obligatorio para los concesionarios, pues la marca no puede obligar a que éstos vendan sus coches a dicho precio. Por ello suele también denominarse PVP recomendado. A este precio, además, se les suelen sumar o restar determinadas cantidades. La tramitación de la matriculación y las opciones que se deseen incorporar hacen aumentar el PVP recomendado, mientras que los descuentos que el concesionario aplique hacen que el PVP se reduzca.

b) Variables candidatas a ser consideradas en el modelo de regresión o explicativas:

- CCUBIC: indica la cilindrada del vehículo, es decir, los centímetros cúbicos del motor.

Es la denominación que se da a la suma del volumen útil de todos los cilindros de un motor alternativo.

La cilindrada ha sido habitualmente usada en muchos países como punto de partida para el cálculo de impuestos. En España en particular se aplica para calcular el llamado Impuesto Municipal de Vehículos de Tracción Mecánica.

- CABALL: representa la potencia en caballos de vapor (CV) del motor de cada turismo.

La potencia máxima está directamente relacionada con el par motor y con el número de revoluciones a las que gira la mecánica. Así, un motor puede tener mucha potencia mediante dos vías, el par motor y las revoluciones de giro. Por ello, un motor con mucho par y que no alcanza demasiadas revoluciones puede llegar a tener igual potencia máxima que otro con poco par y capaz de girar a muy altas vueltas. Esa es la gran diferencia entre los motores diésel (que son los primeros) y los motores de gasolina (que son los segundos).

- CONSUM: El consumo es la cantidad de combustible, medido en litros, que utiliza el vehículo para recorrer 100 kilómetros.

Una de las obligaciones por parte de los fabricantes de automóviles que recoge el Real Decreto 837/2002 de 2 de agosto (BOE núm. 185 de 3 de agosto), que incorpora al ordenamiento jurídico español la Directiva 1999/94/CE, es la relativa a la información de consumo de combustible y emisiones de CO₂ a facilitar al consumidor al comercializar turismos nuevos.

Para obtener un consumo homologado equitativo entre los diferentes vehículos se realizan pruebas bajo las mismas condiciones y supuestos. Éstas se realizan en un lugar cerrado, situando el automóvil sobre un banco dinamométrico. El automóvil en cuestión tiene que haber rodado entre 3000 y 15000 kilómetros y haber permanecido a una temperatura de entre 20 y 30 grados centígrados entre 6 y 30 horas antes de la realización de la prueba. Antes de proceder se debe comprobar que la temperatura de refrigerante y lubricante sea, como máximo, dos grados superior o inferior a la ambiental. El banco dinamométrico está destinado a absorber parte de la potencia y tiene 22 posibilidades de calibración dependiendo del peso del vehículo. Con él se simulan las condiciones de rodadura.

Se procede a la realización de una prueba de consumo urbano y otra de consumo extraurbano. Posteriormente se calcula el consumo mixto que es el resultado de multiplicar el consumo urbano por 0,368 y el consumo

extraurbano por 0,632. Se considera así que un recorrido mixto habitual es 36,8% de recorrido urbano y 63,2% de recorrido extraurbano.

Para este estudio se ha utilizado el dato de consumo de combustible extraurbano (en carretera a 90 km/h.)

- CO₂: señala la cantidad de gases contaminantes que desprenden los vehículos. Se mide en gramos de CO₂ expulsados por kilómetro recorrido.

La Directiva 1999/94 CE, publicada en el Diario Oficial de las Comunidades Europeas el 18 de enero de 2000, tiene por objetivo informar sobre el consumo de carburante y las emisiones de CO₂ de los turismos nuevos, para que los futuros compradores consideren la adquisición de los coches más eficientes energéticamente.

- VELMAX: es la velocidad máxima que pueden alcanzar los coches, medido en km/hora.

La velocidad máxima homologada de un automóvil es la capacidad real del vehículo para alcanzar dicha velocidad en línea recta, por sus propios medios y sin la existencia de pendientes que favorezcan o se opongan a dicha capacidad. Un automóvil, dependiendo de las circunstancias propias y de la calzada, podría alcanzar velocidades superiores a la homologada. Así, con pendiente favorable y contando con un cambio de relaciones largas, un vehículo podría superar claramente su velocidad máxima, aunque ésta no habría sido alcanzada por sus propios medios, sino gracias a la existencia de la comentada pendiente. De hecho, se suele considerar que un automóvil dispone de un cambio de relación larga cuando la velocidad máxima se consigue por debajo del régimen de potencia máxima, mientras que dispondría de relaciones de cambio cortas si la velocidad máxima se consigue a un régimen de giro superior al de la potencia máxima.

- LONGIT: indica la longitud total del coche medida en centímetros.

Esta característica suele tener una relación directa con parámetros como la habitabilidad del vehículo aunque, debido a la existencia de diferentes tipos de carrocería, un coche más pequeño puede en muchas ocasiones ofrecer mayor espacio interior que otro más largo.

- ANCHO: es el ancho total del vehículo (medido en centímetros).

Existen dos modos de medirlo, o bien con los espejos retrovisores desplegados o bien sin los espejos retrovisores. La medición más habitual y la que utilizado para este trabajo es la anchura sin espejos retrovisores.

4 METODOLOGÍA

El método estadístico que se ha utilizado para generar el modelo de predicción del precio de los automóviles frente a las características técnicas ha sido el de Regresión Lineal Múltiple.

Este modelo ha servido para estudiar el grado de dependencia simultánea entre una serie de variables (características técnicas de los coches) y, en base al grado en que dependan, establecer una ecuación, la de regresión, que relaciona las variables entre sí, de forma que los valores que presenta la variable objetivo (precio de los coches) puedan ser predichos, con exactitud relativa, en función de las demás.

La ecuación que suministra el modelo de regresión lineal multivariante es de la forma:

$$Y = B_1X_1 + B_2X_2 + \dots + B_pX_p + B_0$$

Siendo Y la variable explicada; X_1, X_2, \dots, X_p todas las variables explicativas; y B_j cada coeficiente que mide el efecto marginal sobre la respuesta de un aumento en X_j cuando el resto de variables permanece constante.

Con objeto de evaluar la influencia de las variable explicativas en el precio en cada marca, se ha realizado un análisis de regresión individualizado por cada una de las marcas objeto del estudio.

De este modo, se pretende obtener un modelo de predicción por marca del precio de los turismos respecto a las características técnicas.

Se han analizado las variables de forma univariante desde el punto de vista descriptivo y la relación de cada una con la variable respuesta mediante una regresión simple.

A continuación, se ha llevado a cabo un análisis de correlaciones de las variables. Estos análisis permiten determinar las variables que son susceptibles de entrar como explicativas en el modelo multivariante.

Una vez seleccionadas las variables susceptibles de entrar en el modelo de regresión, se ha procedido a realizar la regresión correspondiente.

El método de regresión utilizado en SPSS para desarrollar el modelo de regresión ha sido el llamado “paso a paso” o Stepwise.

El criterio de entrada y salida de las variables con este método es el siguiente:

1º Las variables de la ecuación se introducen una a una; se escoge en primer lugar la variable con una mayor correlación con el criterio.

2º Se calculan las correlaciones parciales entre el resto de las variables independientes y el criterio; la variable que se neutraliza por medio de la correlación parcial es la escogida en primer lugar; la segunda variable seleccionada es la que tiene la mayor correlación parcial (es decir, suponiendo que los sujetos son iguales en la variable ya introducida) con el criterio.

3º Por cada nueva variable introducida se verifica la significación estadística de las variables previamente incluidas y se eliminan si dejan de ser estadísticamente significativas al añadir una nueva variable.

Así se procede sucesivamente: se va escogiendo la variable con una mayor correlación con el criterio una vez neutralizadas (igualando a los sujetos en) las ya escogidas mediante correlaciones parciales y a la vez se van eliminando las introducidas si ya no son estadísticamente significativas al introducir nuevas variables. (Morales, 2012)

Para validar cada modelo de regresión obtenido se han analizado las tablas de la varianza - Anova.

El objetivo principal del Anova es contrastar si existen diferencias entre las diferentes medias de los niveles de las variables (factores).

La variación observada en la respuesta se asume que es debida al efecto de las variables independientes, aunque también se acepta que existe cierto error aleatorio que explica la variación residual. Se asume también que dicho error aleatorio sigue una distribución normal con media 0 y varianza constante.

Para estudiar la validez del modelo se ha realizado un análisis de los residuos (valores predichos - valores observados): normalidad, tendencias, etc. y contraste de homocedasticidad (homogeneidad de varianzas entre los grupos). (Tejedor, 1999)

Una vez realizadas las regresiones para obtener el modelo de predicción de precios de cada marca en particular, se ha llevado a cabo otro análisis de regresión con el fin de conseguir un modelo único para predecir el precio de los coches de las tres marcas agrupadas. En este caso las variables candidatas han sido las mismas que las analizadas en las anteriores regresiones. El método utilizado mediante SPSS para la introducción de las variables candidatas en el modelo ha sido también Stepwise.

Para reducir la información de las variables explicativas resultantes de este modelo se ha realizado un análisis factorial. El análisis factorial es una técnica de reducción de la dimensionalidad de los datos. Su propósito último consiste en buscar el número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos.

Para ello se ha utilizado la herramienta descriptiva Análisis de Componentes Principales (ACP).

El principal atractivo del ACP recae en la simplicidad y capacidad de reducción de dimensión, minimizando el error cuadrático de reconstrucción producido por una combinación lineal de variables latentes, conocidas como componentes principales, las cuales se obtienen a partir de una combinación lineal de los datos originales. Los parámetros del modelo pueden calcularse directamente de la matriz de datos centralizada X , bien sea por descomposición en valores singulares o por la diagonalización de la matriz de covarianza (positiva semidefinida) (Jolliffe, 2002).

Posteriormente, se ha realizado un Análisis Cluster o Análisis de conglomerados. La finalidad de su uso consiste en ordenar objetos en grupos (conglomerados o clusters) de forma que el grado de semejanza entre miembros del mismo cluster sea más fuerte que el grado de asociación entre miembros de diferentes clusters.

El método elegido para realizar el Análisis de Conglomerados ha sido el de K-medias. En el caso del algoritmo de k-medias, cada una de las k agrupaciones de observaciones que se desea revelar está representada por un punto μ_l para $l = 1, \dots, k$, denominado centroide. Cada una de las observaciones x_i de la muestra se asigna al grupo de puntos C_l para el que se cumple que $\min_i(dist(x_i, \mu_l))$. (Peña, 2002)

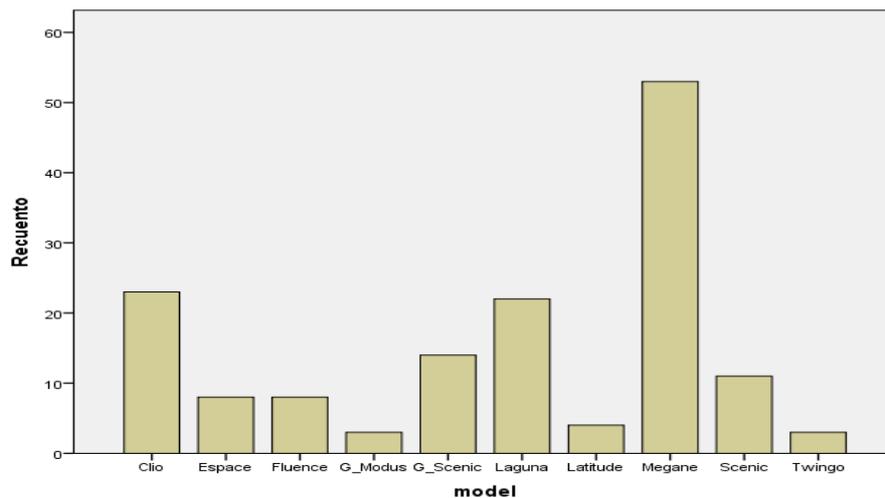
Por último, con objeto de evaluar la influencia de las variables explicativas en el precio de todos los coches en función del agrupamiento obtenido del Análisis Cluster, se han realizado regresiones multivariantes con las variables resultantes del modelo general de las tres marcas por cada tipo de conglomerado.

5 MODELO DE REGRESION DE RENAULT

La primera marca analizada ha sido Renault. La base de datos de Renault se compone de 149 vehículos, distribuidos en 10 modelos conocidos comercialmente.

En el siguiente gráfico se muestra su distribución:

Figura 1 Histograma de modelos de Renault



Como puede verse en el gráfico destacan por el número de versiones puestas a la venta de Renault, los modelos “Megane”, “Clio” y “Laguna” que coinciden con los de mayores ventas de este fabricante.

5.1 Análisis de Variables

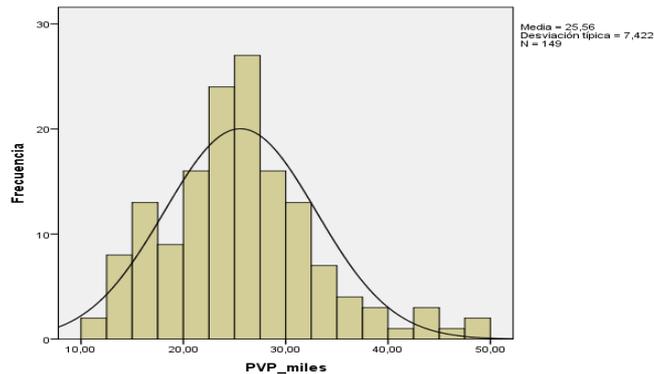
En primer lugar se han analizado las variables a considerar en la regresión, de forma univariante desde el punto de vista descriptivo, interpretando los histogramas de las variables y tomando transformaciones para ver si mejora sustancialmente la normalidad de las mismas.

A continuación, para analizar la relación de cada una de con la variable respuesta, se ha realizado una regresión lineal de cada variable frente al precio de los turismos.

5.1.1 Variable Objetivo: Precio

La distribución que sigue la variable objetivo, PVP_miles, se puede observar en el siguiente histograma.

Figura 2. Histograma de variable objetivo PVP_miles



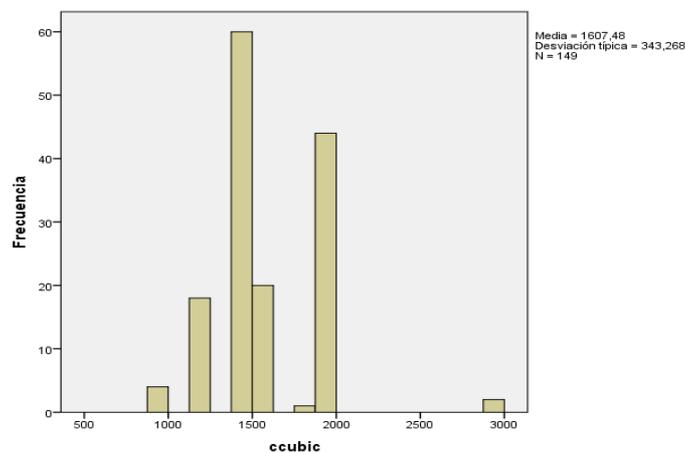
Comparamos el histograma con la distribución normal y vemos que se asemeja bastante.

Se han realizado transformaciones de esta variable, pero los resultados obtenidos no mejoran sustancialmente los datos de normalidad y tampoco los de la regresión posterior. Por tanto, se ha utilizado la variable origen sin transformar.

5.1.2 Centímetros Cúbicos del Motor

El histograma de los centímetros cúbicos (CCUBIC) es el siguiente:

Figura 3. Histograma variable ccubic



Atendiendo a la frecuencia de esta variable, se aprecia que los coches de Renault se agrupan principalmente en 5 tipos diferenciados por su cilindrada: 900 CC, 1200 CC, 1500 CC, 1600 CC y 2000 CC.

El resumen del modelo de la regresión resultante entre esta variable y el precio, se refleja en la siguiente tabla.

Tabla 1. Resumen del modelo con variable ccubic

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,779 ^a	,607	,605	4,66719

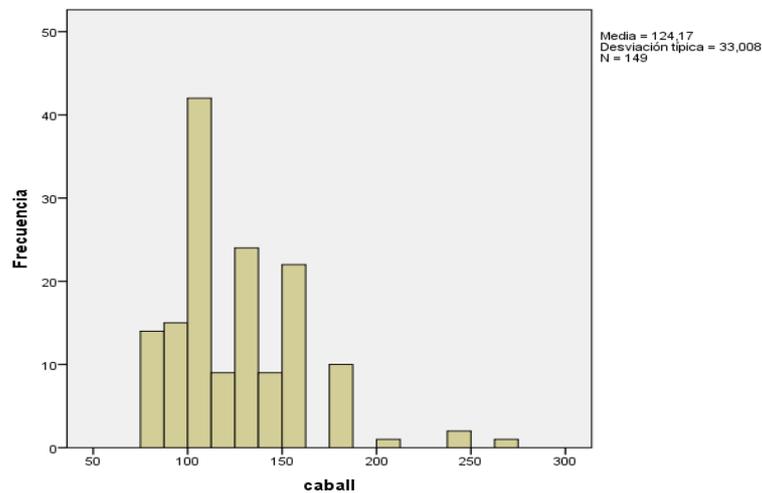
a. Variables predictoras: (Constante), ccubic

Esta variable recoge un 78% de variabilidad del precio con un coeficiente de determinación R^2 del 60,7% por lo que parece bastante significativa para el estudio.

5.1.3 Potencia del Motor

El histograma de la variable CABALL nos indica la distribución que siguen los vehículos de Renault respecto a la potencia del motor.

Figura 4. Histograma de variable caball



Como se observa, la potencia de los coches se agrupa en 8 valores principales distribuidos entre 75 y 180 CV, siendo el grupo con mayor número de turismos los que tienen un motor de 110 CV que representan casi un tercio del total.

Tan sólo hay 4 coches con una potencia superior a 200 CV que no son muy representativos y que se han dejado fuera del análisis, tras verificar que no interfieren en la regresión en aras de mejorar la normalidad de la variable.

Tabla 2. Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,831 ^a	,691	,689	3,95282

a. Variables predictoras: (Constante), caball

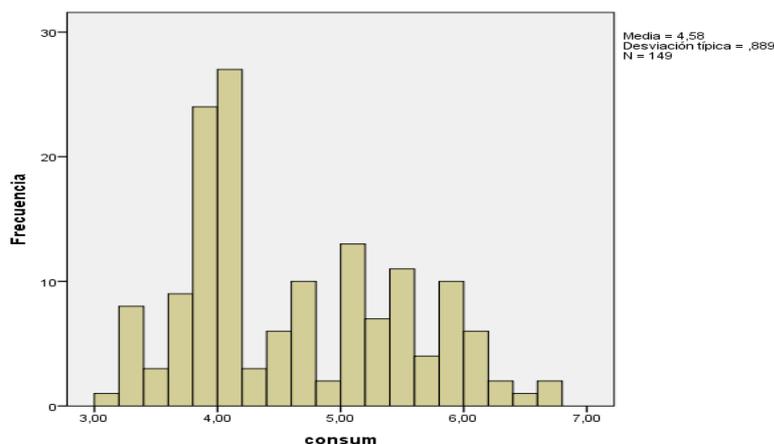
b. Variable dependiente: PVP_miles

El resumen del modelo de regresión entre los caballos del motor y el precio revela que según el coeficiente de determinación, R^2 , esta variable recoge un 69% de la variación del precio; por este motivo se considera muy importante su inclusión como variable candidata al modelo de regresión.

5.1.4 Consumo de Combustible

El histograma del consumo (CONSUM), por cada 100 Km. en carretera es:

Figura 5. Histograma de variable consum



El consumo es una variable con un amplio abanico de valores que van desde 3 litros hasta 6,7 litros consumidos por cada 100 Km. recorridos.

Tabla 3. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,362 ^a	0,131	0,125	6,94298

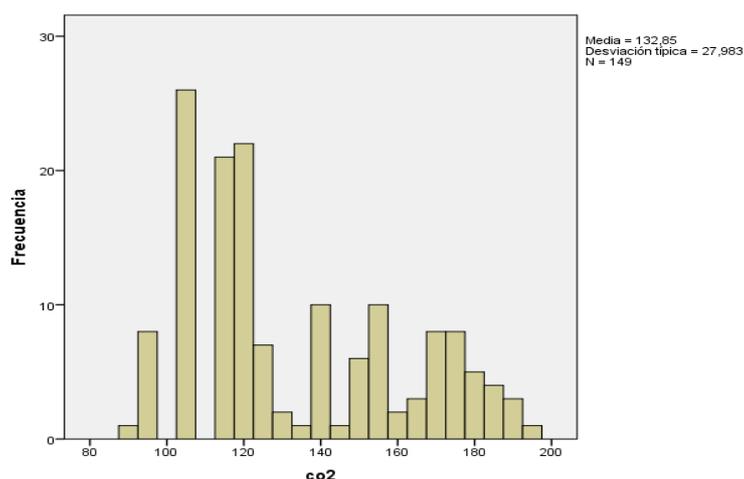
a. Variables predictoras: (Constante), consum

La variable “consum” explicaría en un 13%, según el coeficiente de determinación R^2 , la variabilidad de PVP_miles.

5.1.5 Emisión de Gases Contaminantes

Los valores de emisión de gases CO2 se muestran en el histograma:

Figura 6. Histograma de variable CO2



A tenor del histograma se observa que la mayoría de las versiones que Renault oferta emiten entre 110 y 160 gr/CO2 por kilómetro recorrido.

Este dato es bastante lógico en la aplicación de su política comercial ya que casi el 50 % de las ventas de turismos en España durante el 2012 fueron de vehículos que se encontraban entre esos valores de emisiones de CO2.

El resumen del modelo de regresión entre el CO2 y el precio es:

Tabla 4. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,497 ^a	0,247	0,241	6,46412

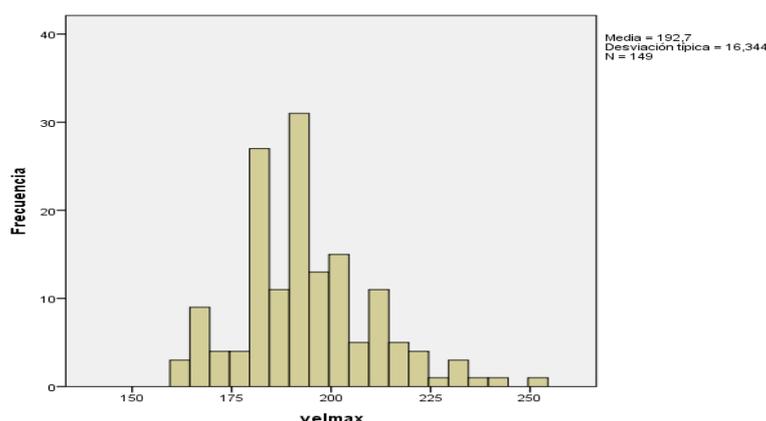
a. Variables predictoras: (Constante), co2

La variable CO2 explica en un 25% la variabilidad del precio conforme al coeficiente de determinación R², mejorando el dato de la variable “Consum” pero con un error de estimación mayor.

5.1.6 Velocidad Máxima

El histograma de la Velocidad Máxima que alcanzan los coches (VELMAX) se presenta en la Figura 7.

Figura 7. Histograma de variable velmax



La velocidad máxima a la que pueden llegar los coches de Renault va desde los 162 km/h para los coches menos veloces hasta los 230 km/h para los más deportivos.

Las velocidades máximas más frecuentes son 180 km/h y 190 km/h.

Tabla 5. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,662 ^a	0,438	0,434	5,58234

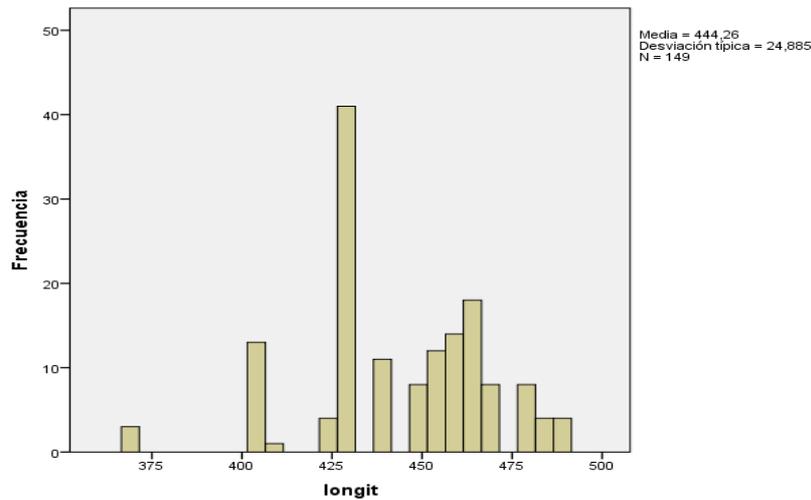
a. Variables predictoras: (Constante), velmax

La velocidad máxima recoge en gran medida la variabilidad del precio, un 66% según el coeficiente de determinación R².

5.1.7 Longitud de los Vehículos

El histograma con la longitud de los coches (LONGIT) queda representado en la Figura 8.

Figura 8. Histograma de variable longit



La longitud de los coches va desde los 369 cm del modelo “Twingo” hasta los 490 cm del Renault “Latitude”.

Tabla 6. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,732 ^a	0,536	0,533	5,07013

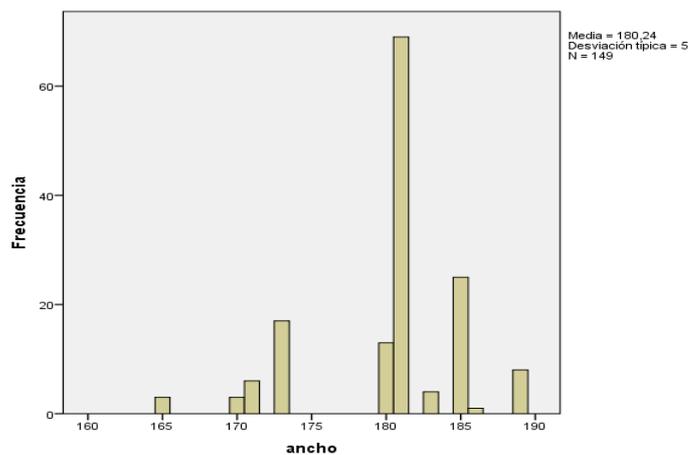
a. Variables predictoras: (Constante), longit

Se puede observar en los datos resumen de la tabla 6. del modelo de regresión que el coeficiente de determinación R^2 tiene un valor de explicación del 53,6%, lo que indica bastante relación con la variable objetivo PVP_miles.

5.1.8 Ancho de los Coches

A continuación se reflejan las frecuencias del ancho de los vehículos (ANCHO).

Figura 9. Histograma de variable ancho



El ancho que tienen los coches Renault se sitúan entre los 165 cm del modelo “Twingo” hasta los 189 cm del modelo “Gran Espace”.

Esta variable está totalmente relacionada con el modelo del vehículo por lo que se ha decidido no incluir una variable que recoja el “nombre del modelo” en la regresión.

Tabla 7. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,712 ^a	0,506	0,503	5,23182

a. Variables predictoras: (Constante), ancho

El coeficiente de determinación R^2 revela que la variación del precio está explicada por el ancho de los coches en un 50,6%. Parece ser el ancho una variable bastante influyente en el precio de los coches.

Conforme a los resultados del análisis de las variables, se puede señalar que todas ellas pueden ser candidatas a entrar en el modelo de regresión de Renault.

5.2 Resultados de la Regresión de Renault

Utilizando las posibilidades que ofrece el programa SPSS en regresión, he optado por utilizar el método “paso a paso” por su criterio de entrada y salida de variables en el modelo en función del grado de correlación parcial de las variables y su grado de significación.

5.2.1 Estadísticos Descriptivos

En la siguiente tabla recoge las variables que se van a correlacionar así como los tres índices que sintetizan las distribuciones: media, desviación típica y tamaño de la muestra.

Tabla 8. Estadísticos Descriptivos de Renault

	Media	Desviación típica	N
PVP_miles	25,2494	7,09013	145
Ccubic	1585,57	304,638	145
Caball	121,08	27,316	145
Consum	4,5448	,87193	145
Co2	131,56	27,018	145
Velmax	191,42	14,554	145
Longit	444,14	24,681	145
Ancho	180,23	5,011	145

El número de versiones de automóviles de Renault analizadas son 145 unidades.

El precio medio de los coches de Renault se sitúa sobre los 25.000 €.

5.2.2 Matriz de Correlaciones

La siguiente tabla recoge las correlaciones entre las variables candidatas de la regresión de Renault.

Tabla 9. Matriz de Correlaciones Regresión Renault

	PVP_miles	ccubic	caball	consum	co2	velmax	longit	ancho
PVP_miles	1	0,75	0,831	0,331	0,454	0,652	0,744	0,744
Ccubic	0,75	1	0,812	0,404	0,55	0,702	0,551	0,461
Caball	0,831	0,812	1	0,549	0,639	0,901	0,58	0,589
Consum	0,331	0,404	0,549	1	0,968	0,417	0,199	0,311
Co2	0,454	0,55	0,639	0,968	1	0,508	0,283	0,34
Velmax	0,652	0,702	0,901	0,417	0,508	1	0,471	0,386
Longit	0,744	0,551	0,58	0,199	0,283	0,471	1	0,717
Ancho	0,744	0,461	0,589	0,311	0,34	0,386	0,717	1

Se puede ver en la Tabla 9. que existe alta correlación (marcada en negrita) entre las variables CO2 y CONSUM.

Es lógico pensar que existe una relación entre la emisión de partículas contaminantes y el consumo de combustible. Aún así, se han introducido las variables en el estudio por ser consideradas muy significativas para el mismo aunque se deberá tener en cuenta para su interpretación en caso de que entren las variables en el modelo de regresión.

5.2.3 Resumen del Modelo

La siguiente tabla muestra el modelo final resultante.

Tabla 10. Resumen del Modelo de RENAULT

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
6	,939 ^f	,882	,876	2,49191

f. Variables predictoras: (Constante), caball, longit, velmax, consum, co2, ancho

g. Variable dependiente: PVP_miles

R^2 , el coeficiente de determinación es 0,882. Es decir, la proporción de la variación de la variable dependiente (PVP_miles) explicada por el modelo de regresión es del 88,2%. Según este resultado, podemos decir que el modelo se ajusta bastante bien a los datos.

Con los datos de la tabla 11., Anova, se puede calcular directamente el valor de R^2 .

$$R^2 = \frac{SC_{\text{exp}}}{SC_t} = \frac{\sum_{i=1}^N (Y_i^3 - \bar{Y})^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} = \frac{6381,941}{7238,866} = 0,882$$

R^2 corregida (coef. de determinación ajustado) indica que las variables predictoras explican un 87,6 % de la varianza que presenta el precio.

Es una corrección a la baja de R^2 basada en el número de casos n y de variables independientes p según la fórmula siguiente:

$$R^2_{\text{corregida}} = R^2 - \frac{p(1 - R^2)}{n - p - 1} = 0,882 - \frac{6*(1 - 0,882)}{145 - 6 - 1} = 0,876$$

Error típico de la estimación: Nos da la desviación típica residual que se mide con la raíz cuadrada de la varianza residual. En nuestro caso el resultado es 2,491.

5.2.4 Análisis de la Varianza

Esta es la tabla de Análisis de la Varianza (Anova):

Tabla 11. ANOVA Modelo Renault

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
6 Regresión	6381,941	6	1063,657	171,29	,000 ^a
Residual	856,925	138	6,21		
Total	7238,866	144			

a. Variable dependiente: PVP_miles

g. Variables predictoras: (Constante), caball, longit, velmax, consum, co2, ancho

El modelo se muestra válido para representar los datos. El valor de significación obtenido (Sig=0,000) indica que el conjunto de las variables predictoras introducidas explican parte de la variabilidad del precio de los coches, PVP_miles.

5.2.5 Coeficientes de la Regresión

Los coeficientes resultantes en el modelo de regresión son:

Tabla 12. Coeficientes de la Regresión de Renault

Modelo	Coeficientes estandarizados		Coeficientes tipificados	T	Sig.
	B	Error típ.	Beta		
(Constante)	-44,991	12,898		-3,488	,001
Caball	,235	,025	,907	9,235	,000
Longit	,052	,013	,182	3,957	,000
Velmax	-,180	,038	-,370	-4,804	,000
Consum	-7,032	1,065	-,865	-6,604	,000
co2	,202	,037	,770	5,450	,000
Ancho	,324	,071	,229	4,575	,000

a. Variable dependiente: PVP_miles

Han entrado en el modelo todas las variables candidatas salvo CO2.

El conjunto de variables regresoras resultantes en el modelo tienen un nivel de significación de 0,000.

Los coeficientes de las variables Consum y Velmax tienen valores negativos. Esto significa que si no tuvieran variación las restantes características técnicas del modelo,

un aumento del consumo o de la velocidad podría suponer una disminución del precio de venta de algunas versiones de los coches.

5.2.6 Ecuación Resultante de la Regresión

La ecuación resultante, conforme a los coeficientes del modelo, es:

$$\text{PVP_miles} = 0,235 \text{ caball} + 0,052 \text{ longit} - 0,180 \text{ velmax} - 7,032 \text{ consum} + 0,202 \text{ CO2} + 0,324 \text{ ancho} - 44,991$$

5.2.7 Diagnóstico del Modelo de Regresión

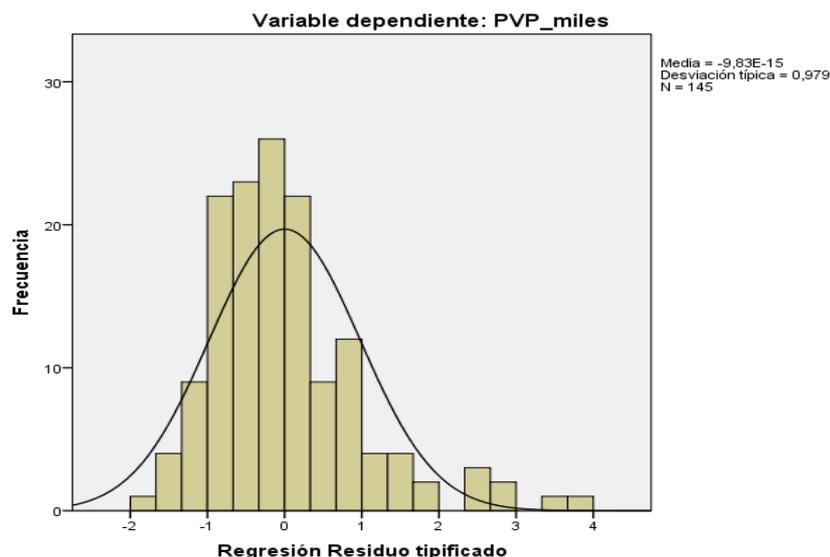
Para estudiar la validez del modelo de regresión se ha realizado un análisis de los residuos (valores predichos - valores observados): normalidad, tendencias, y contraste de homocedasticidad (homogeneidad de varianzas entre los grupos).

A continuación se exponen los distintos gráficos de residuos que ayudarán para diagnosticar el modelo.

5.2.7.1 Histograma de Residuos estandarizados

El Histograma presenta un gráfico comparativo entre las frecuencias observadas frente a los valores estandarizados que corresponden a tales frecuencias.

Figura 10. Histograma de Residuos tipificados

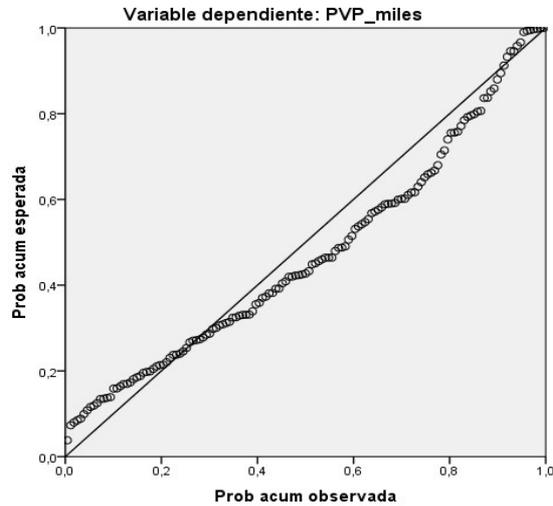


Contrastando los residuales observados y los esperados, se observa que se asemejan bastante a la ley normal; por lo que se puede decir en este caso que la distribución sigue una normalidad.

5.2.7.2 Gráfico de Normalidad

En la Figura 11. se expresan los valores reales en ordenadas, y los esperados en abscisas (esperados si la distribución fuera normal).

Figura 11. Gráfico P-P normal de regresión Residuo tipificado

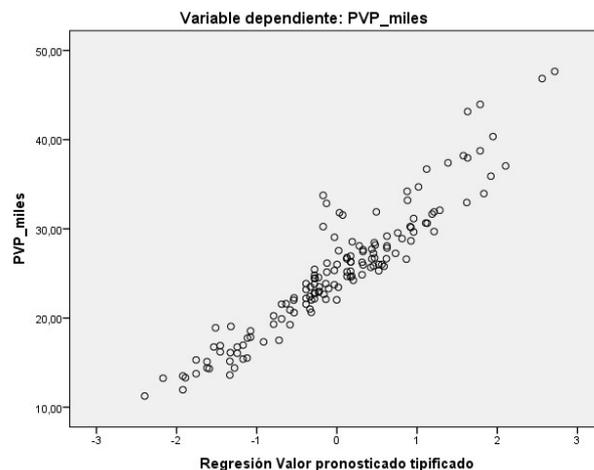


Comparamos luego con la recta diagonal que parte del origen (0,0) los valores observados y vemos que se aproxima bastante. Esto significa que nuestra distribución sigue una ley normal.

5.2.7.3 Gráficos de Dispersión

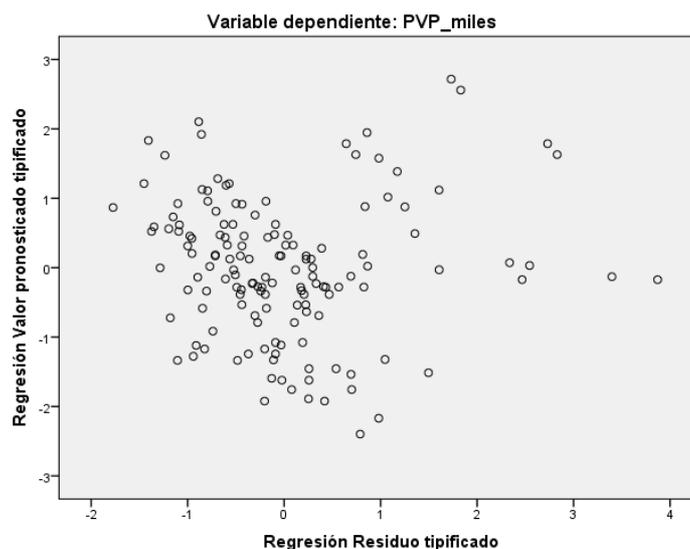
Los siguientes Gráficos de dispersión muestran la relación entre los valores observados y los predichos.

Figura 12. Gráfico de Dispersión valores pronosticados por valores observados



En la Figura 12. se muestran los valores pronosticados tipificados frente a los valores reales de la variable objetivo, PVP_miles. Vemos que se aproxima mucho a la recta diagonal y, por tanto, no existen desviaciones significativas.

Figura 13. Gráfico dispersión de residuos tipificados por pronósticos tipificados



En el gráfico anterior (Figura 13.) se observa que los residuos se distribuyen de forma simétrica a ambos lados del eje 0 y a lo largo de todo el rango de valores de la estimación y la variabilidad parece constante.

Observamos que existe homocedasticidad, no parece haber tendencia.

5.2.8 Intervalo de Confianza

El programa SPSS permite obtener intervalos de confianza para las puntuaciones, que pasan a formar parte de la base de datos, siendo incluidos sus límites como nuevas variables en el archivo.

En la siguiente tabla se muestran los 5 vehículos con mayores residuos:

Tabla 13. Vehículos con mayores residuos

Versión vehículo	RES_1	LMCI_1	UMCI_1	LICI_1	UICI_1
Megane 1.5 dCi+ Flord CC	9,65	23,39	24,82	19,12	29,08
Megane 1.5 dCi+ Pr CC EDC	8,47	23,73	25,04	19,41	29,36
Espace 2.0 dCi+ Initiale	7,05	35,05	37,14	31,06	41,14
G Espace 2.0 dCi+ In	6,80	36,09	38,20	32,11	42,19
Laguna 1.5 dCi Emo Coupe	6,34	24,43	26,49	20,43	30,49

Las columnas etiquetadas como LMCI_1 y UMCI_1, representan los límites inferior y superior, respectivamente, para la media de la recta de regresión. Las etiquetadas como LICI_1 y UICI_1, límites inferior y superior del intervalo de confianza para la predicción.

Los IC para el precio de coches con unas determinadas características de ccubic, caball, consum, co2, velmax, longit y ancho tienen una amplitud de aproximadamente 10.000 euros.

5.3 Interpretación del Modelo de Renault

El valor de R^2 o coeficiente de determinación es de 0,882, lo que indica que con el conjunto de los seis predictores podemos explicar el 88,2% de la variación de los resultados, siendo la proporción de variación no explicada o error del 11,20%, que se deberá a errores aleatorios y a la influencia de variables no incluidas en el modelo.

El consumo tiene un coeficiente con signo negativo en la ecuación del modelo de regresión de Renault. Esto significa que aquellos coches cuyo consumo sea mayor al de otros vehículos con el resto de características técnicas iguales, normalmente asociado al tipo de motorización/combustible usado, tendrán proporcionalmente un precio bastante inferior.

El resto de variables (salvo la velocidad máxima), presentan una relación positiva proporcional respecto a la variación del precio. Las que tienen los valores más destacables sobre el precio son el ancho de los vehículos y su potencia. A mayor ancho o potencia de los vehículos mayor será su precio.

Por ejemplo, el modelo de regresión indica que si el resto de valores considerados en la ecuación de regresión no variara, un incremento de 10 CV de potencia en un vehículo podría suponer un aumento en el precio de 2.350 €

Atendiendo a los resultados de confianza para las predicciones individuales, se estima que la variabilidad debida a otros aspectos no considerados, como el nivel de equipamiento, tipo de cambio o la tracción de los coches, puede hacer que un vehículo cambie hasta en 10.000 € su precio de venta.

Por tanto, el modelo de regresión de Renault ofrece unos resultados bastante aceptables considerando que la variabilidad del precio de los coches es explicada por las variables predictoras en un 88% aunque con un intervalo de confianza ancho.

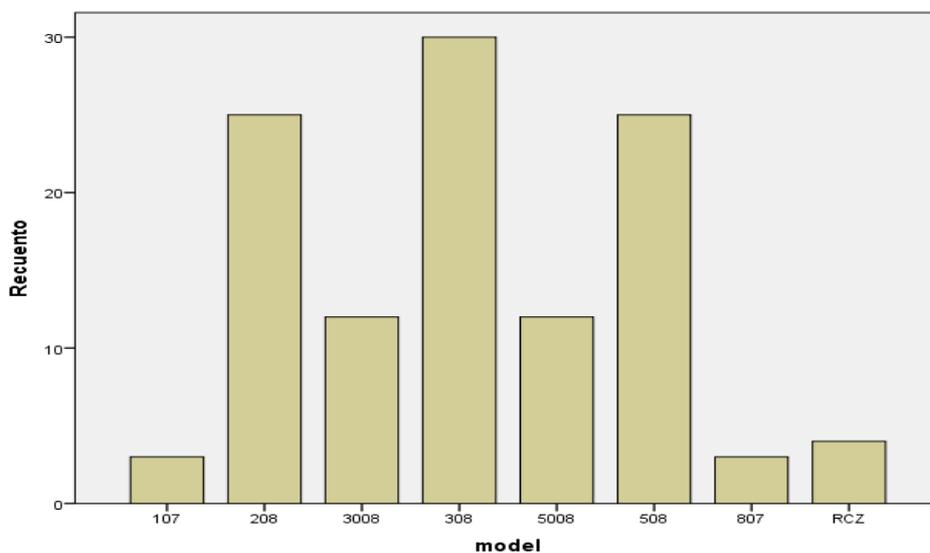
6 MODELO DE REGRESION DE PEUGEOT

La siguiente marca analizada para conseguir un modelo de predicción del precio de sus coches en función de las características técnicas ha sido Peugeot.

La base de datos de Peugeot está compuesta por 114 vehículos que se distribuyen en 8 modelos diferenciados comercialmente.

En el siguiente gráfico se muestra su distribución:

Figura 14. Histograma de modelos de Peugeot



Destacan especialmente por el número de versiones ofertadas, los modelos Peugeot “308”, “208” y “508”.

6.1 Análisis de Variables

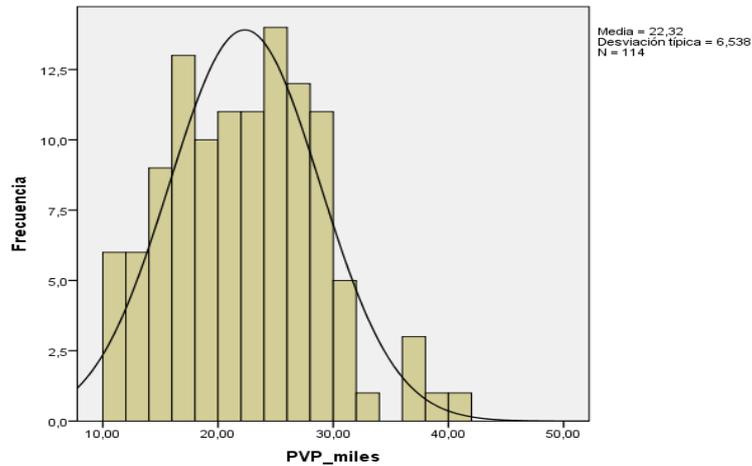
Al igual que se realizó con la marca Renault, se ha realizado un estudio de las variables analizando los histogramas de las mismas y tomando transformaciones para ver si mejora sustancialmente su distribución.

A continuación, se ha realizado una regresión con cada variable para ver si tienen relación significativa con la variable objetivo.

6.1.1 Variable Objetivo: Precio

El histograma de la variable PVP_miles, es el siguiente:

Figura 15. Histograma de PVP_miles

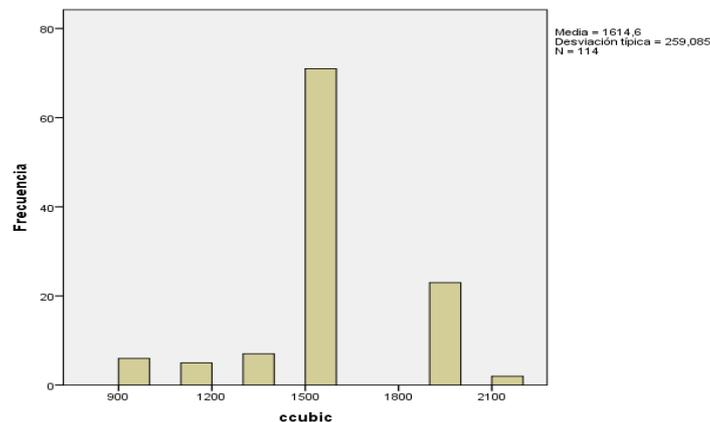


He realizado transformaciones de esta variable, tanto con la raíz cuadrada como con el logaritmo neperiano de la misma. Los datos obtenidos no mejoran sustancialmente los datos de normalidad y tampoco los de la regresión posterior. Por ello, he utilizado la variable origen sin transformar.

6.1.2 Centímetros Cúbicos del Motor

El histograma de los centímetros cúbicos (CCUBIC) es el siguiente:

Figura 16. Histograma de variable ccubic



Atendiendo a la frecuencia de esta variable, se aprecia que los coches de la marca Peugeot se agrupan principalmente en 6 tipos diferenciados por su cilindrada: 1000 CC, 1200 CC, 1400 CC, 1600 CC, 2000 CC y 2200 CC. Se aprecia que el modelo de motor más frecuente es el de 1600 CC.

Este es el modelo de la regresión resultante:

Tabla 14. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,781 ^a	,609	,606	4,10530

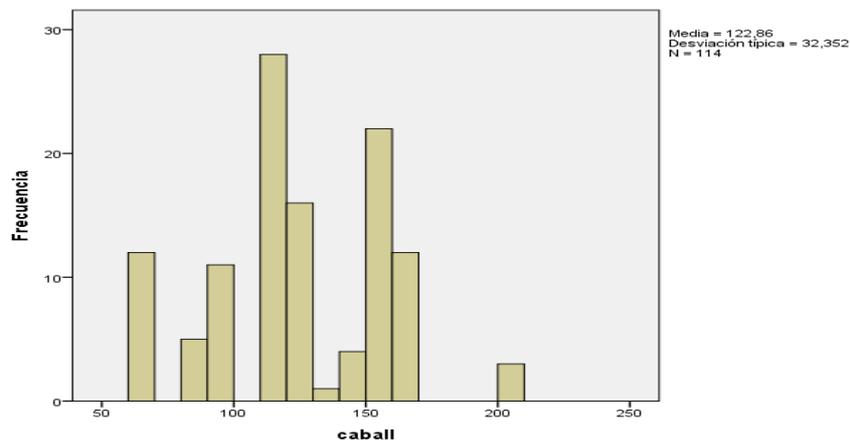
a. Variables predictoras: (Constante), ccubic

Esta variable recoge un 78% de variabilidad del precio con un coeficiente de determinación R^2 del 60,9%. Esto indica que es una variable importante para ser incluida como candidata al modelo de regresión de Peugeot.

6.1.3 Potencia del Motor

El histograma de la variable que indica los caballos del motor (CABALL) es:

Figura 17. Histograma de variable caball



La potencia de los coches se distribuye entre los 68 CV del modelo “107” hasta los 204 CV de las dos versiones de Peugeot “508 más potentes.

En el siguiente cuadro se refleja la regresión resultante entre “Caball” y “PVP_miles”.

Tabla 15. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,843 ^a	,710	,708	3,53540

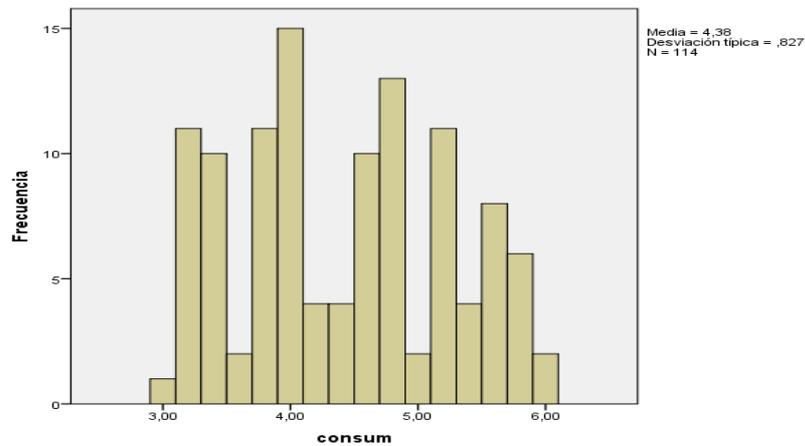
a. Variables predictoras: (Constante), caball

La variable que indica la potencia del motor recoge un 69% de la variación del precio conforme al coeficiente de determinación, R^2 .

6.1.4 Consumo de Combustible

El histograma del consumo de carburante por cada 100 Km. en carretera es:

Figura 18. Histograma de variable consum



El consumo es una variable con un amplio espectro de valores que van desde 3 litros del Peugeot “308 1.6 e-HDI” hasta los 6 litros consumidos cada 100 Km. recorridos por el modelo “3008 1.6 THP Allure Aut.”.

Tabla 16. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,445 ^a	,198	,191	5,88013

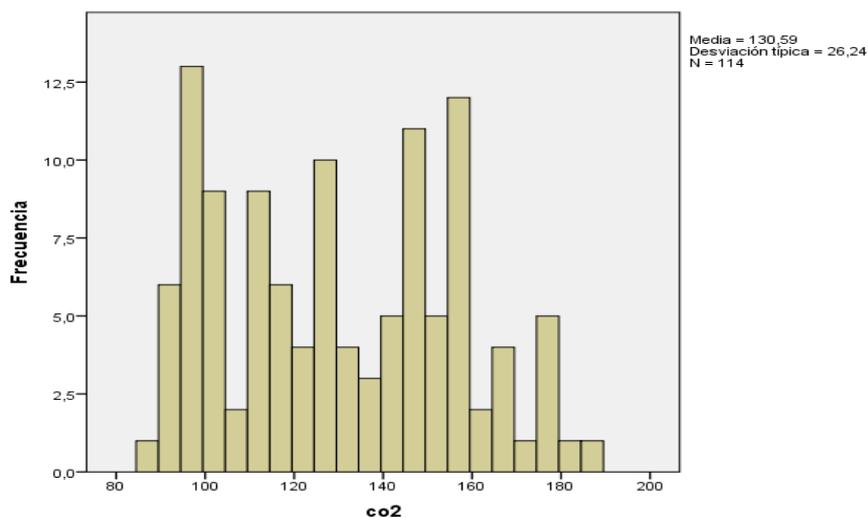
a. Variables predictoras: (Constante), consum

Aunque la variación en el precio de los coches influenciada por el consumo aparentemente tiene un valor menor que las variables hasta ahora analizadas, merece la pena su inclusión en el modelo de regresión a realizar posteriormente.

6.1.5 Emisión de Gases Contaminantes

Los valores de emisión de gases CO2 se muestran en este histograma.

Figura 19. Histograma de variable CO2



La emisión de partículas de CO2 va desde 87 gr/Km. del Peugeot “208 1.4 e-HDI Act BL 5p C” hasta los 189 gr/km del modelo “807 2.0 HDI+ Active Aut”.

Tabla 17. resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,601 ^a	,361	,355	5,25061

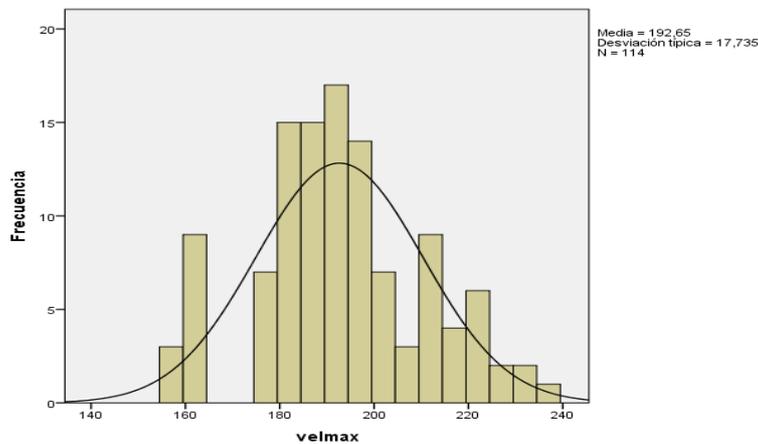
a. Variables predictoras: (Constante), co2

La variable CO2 explica en un 36% la variabilidad del precio conforme al coeficiente de determinación R².

6.1.6 Velocidad Máxima

En el siguiente gráfico se representa el histograma de la Velocidad Máxima:

Figura 20. Histograma de variable velmax



La velocidad máxima que alcanzan los coches de Peugeot oscila entre los 157 km/h del Peugeot “107” hasta los 237 km/h que alcanza el más deportivo Peugeot “RCZ 1.6 THP+”.

La gran mayoría de los coches pueden alcanzar una velocidad máxima de 180 a 200 km/h.

La regresión de esta variable con el precio es:

Tabla 18. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,764 ^a	,584	,580	4,23699

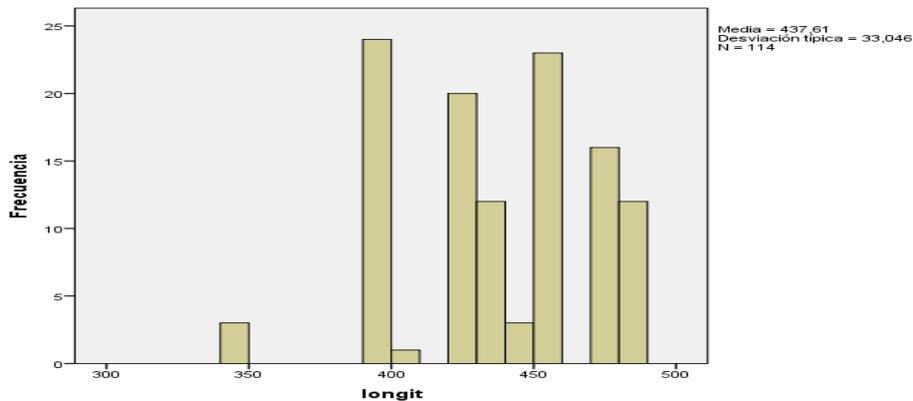
a. Variables predictoras: (Constante), velmax

Conforme al cuadro resumen del modelo, la velocidad máxima recoge en gran medida la variabilidad del precio, un 58,4% según el coeficiente de determinación R².

6.1.7 Longitud de los Vehículos

El histograma con la longitud de los coches (LONGIT) es:

Figura 21. Histograma de variable longit



La longitud de los coches va desde los 345 cm del Peugeot “107” hasta los 481 cm del Peugeot “508”.

Tabla 19. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,775 ^a	,601	,597	4,15041

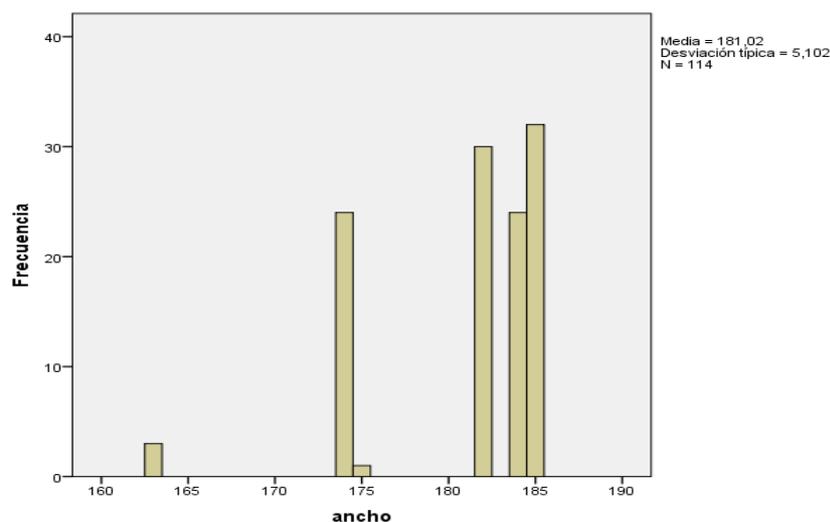
a. Variables predictoras: (Constante), longit

Según los datos del resumen del modelo de regresión, R^2 tiene un valor de explicación del 60,1%; La variable “Longit” influye relativamente en el precio de los automóviles, “PVP_miles”.

6.1.8 Ancho de los Vehículos

A continuación se reflejan las frecuencias del ancho de los vehículos (ANCHO):

Figura 22. Histograma de variable ancho



El ancho de los coches Peugeot difiere entre los 163 cm. del modelo “107” y los 185 cm. de los modelos más anchos (807, RCZ y 508).

Tabla 20. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,770 ^a	,593	,589	4,18911

a. Variables predictoras: (Constante), ancho

El coeficiente de determinación R^2 revela que la variación del precio estaría explicada por el ancho de los coches en un 77%.

Por tanto, es una variable a considerar en el modelo de regresión que se realizará posteriormente.

6.2 Resultados de la Regresión de Peugeot

A continuación se muestran los resultados obtenidos de la Regresión de Peugeot tras la introducción de las variables consideradas en el punto anterior frente al precio.

6.2.1 Estadísticos Descriptivos

Los datos de los estadísticos descriptivos se muestran en la siguiente tabla:

Tabla 21. Estadísticos Descriptivos

	Media	Desviación típica	N
PVP_miles	22,3184	6,53824	114
Ccubic	1614,60	259,085	114
Caball	122,86	32,352	114
Consum	4,3789	,82671	114
co2	130,59	26,240	114
Velmax	192,65	17,735	114
Longit	437,61	33,046	114
Ancho	181,02	5,102	114

En este cuadro se ven las variables que se van a correlacionar, aportando los tres índices que sintetizan las distribuciones: media, desviación típica y tamaño de la muestra.

El número de versiones analizadas de Peugeot es de 114 unidades.

El precio medio de los vehículos de Peugeot está entorno a 22.300 €, casi 3000 € menos que el precio medio de los vehículos de Renault.

6.2.2 Matriz de Correlaciones

En el siguiente cuadro (tabla 22.) se recogen los datos de correlaciones entre las variables.

Tabla 22. Matriz de Correlaciones Peugeot

	PVP_miles	ccubic	caball	consum	co2	velmax	longit	ancho
PVP_miles	1	0,781	0,843	0,445	0,601	0,764	0,775	0,77
Ccubic	0,781	1	0,795	0,315	0,527	0,712	0,641	0,666
Caball	0,843	0,795	1	0,624	0,78	0,913	0,658	0,71
Consum	0,445	0,315	0,624	1	0,928	0,44	0,337	0,46
co2	0,601	0,527	0,78	0,928	1	0,599	0,483	0,587
Velmax	0,764	0,712	0,913	0,44	0,599	1	0,684	0,678
Longit	0,775	0,641	0,658	0,337	0,483	0,684	1	0,905
Ancho	0,77	0,666	0,71	0,46	0,587	0,678	0,905	1

En la Tabla 22. se ve la alta correlación existente entre las variables CO2 y “consum”. También vemos que la velocidad máxima (velmax) tiene bastante correlación con los caballos del motor (caball). A pesar de ello, he introducido las variables en la regresión por considerarlas significativas para su estudio y poder analizar sus resultados.

6.2.3 Resumen del Modelo de Peugeot

El siguiente cuadro, nos indica el modelo final resultante de la regresión de Peugeot.

Tabla 23. Resumen del Modelo de PEUGEOT

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
6	,906 ^f	0,822	0,815	2,81239

f. Variables predictoras: (Constante), caball, longit, velmax, co2

R (coeficiente de correlación múltiple): 0,906

R², coeficiente de determinación es 0,822. Significa que la proporción de la variación de la variable dependiente (PVP_miles) es explicada por el modelo de regresión, conjunto de variables predictoras, en un 82,2%.

Según este resultado, podemos decir que el modelo se ajusta bastante bien a los datos.

R² corregida (coef. de determinación ajustado): 0,815; indica que las variables predictoras explican el 81,5 % de la varianza que presenta el precio, teniendo en cuenta el ajuste por el número de variables explicativas del modelo.

Error típico de la estimación: Nos da la desviación típica residual medida como la raíz cuadrada de la varianza residual. En el modelo resultante es de 2,812.

6.2.4 Análisis de la Varianza

Esta es la tabla de Análisis de la Varianza (ANOVA):

Tabla 24. ANOVA del Modelo de PEUGEOT

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
6 Regresión	3968,447	4	992,112	125,432	,000 ^g
Residual	862,142	109	7,91		
Total	4830,59	113			

a. Variable dependiente: PVP_miles

g. Variables predictoras: (Constante), caball, longit, velmax, co2

El modelo se muestra válido para representar los datos. El valor de significación obtenido ($p=0,000$) indica que las variaciones en la variable PVP_miles es explicada parcialmente por el conjunto de variables predictivas identificadas (caball, longit, velmax y CO2).

6.2.5 Coeficientes de la Regresión

En el siguiente cuadro se encuentran los coeficientes resultantes del modelo de regresión de Peugeot.

Tabla 25. Coeficientes de la Regresión de Peugeot

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
6 (Constante)	-6,597	6,377		-1,034	0,303
Caball	0,223	0,029	1,104	7,799	0
Longit	0,086	0,011	0,437	7,801	0
Velmax	-0,15	0,042	-0,407	-3,538	0,001
co2	-0,057	0,018	-0,227	-3,15	0,002

Las variables dependientes explican con un nivel de significación suficiente, menor que el requerido para el modelo (0,05), la variabilidad del precio.

Las variables que no han entrado en el modelo son las correspondientes a la cilindrada (ccubic), consumo de combustible (consum) y ancho de los vehículos (ancho).

Debido al signo de los coeficientes de las variables CO2 y Velmax, estas variables influyen de manera inversamente proporcional a la variable dependiente PVP_miles.

6.2.6 Ecuación de la Regresión

La ecuación resultante de la regresión es:

$$\text{PVP_miles} = 0,223 \text{ caball} + 0,086 \text{ longit} - 0,150 \text{ velmax} + \\ - 0,057 * \text{CO2} - 6,597$$

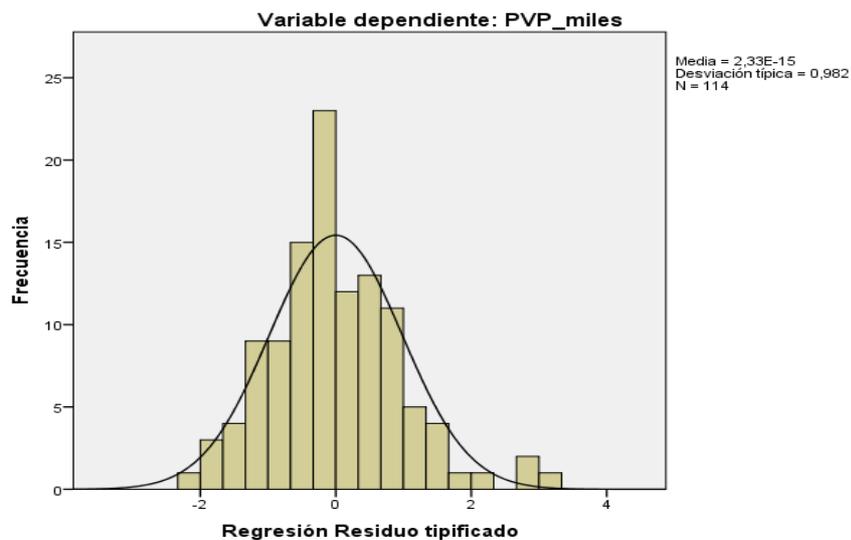
6.2.7 Diagnóstico del Modelo de Regresión

A continuación se visualizan los gráficos que sobre la regresión resultante de Peugeot van a permitir diagnosticar el modelo obtenido.

6.2.7.1 Histograma

El gráfico representado en la tabla 26. corresponde al Histograma de residuos estandarizados:

Tabla 26. Histograma de Residuos



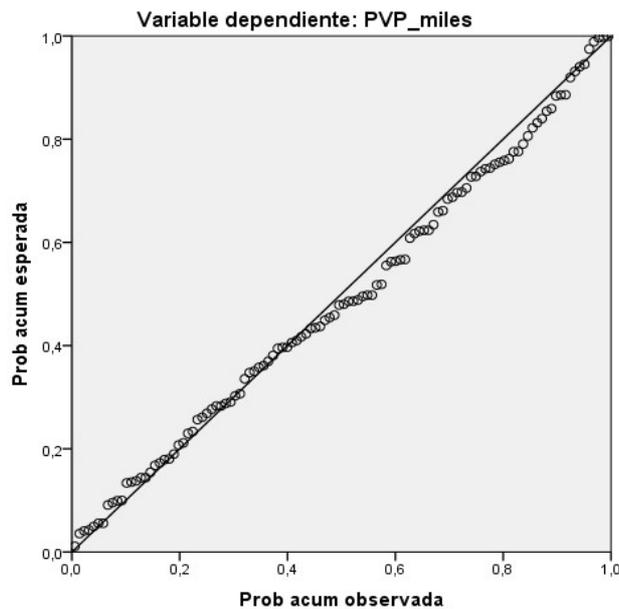
El Histograma presenta el gráfico comparativo entre las frecuencias observadas frente a los valores estandarizados que corresponden a tales frecuencias.

Contrastando los residuales observados y los esperados, vemos que se asemejan bastante a la ley normal.

6.2.7.2 Gráfico de Normalidad

En la Figura 23. se observan los valores reales en ordenadas, y los esperados en abscisas (esperados si la distribución fuera normal).

Figura 23. Gráfico P-P normal de regresión Residuo tipificado

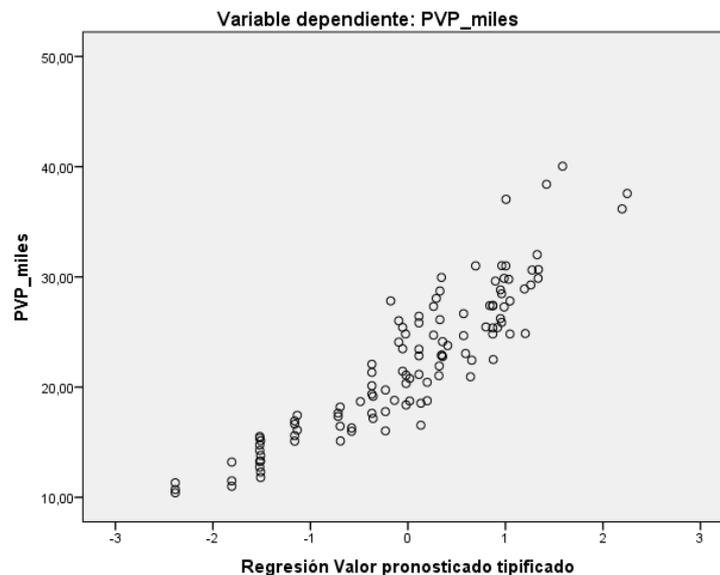


Comparando la recta diagonal que parte del origen (0,0) y los valores observados, se aprecia que se aproximan bastante. Esto significa que la distribución sigue una ley normal.

6.2.7.3 Gráficos de Dispersión

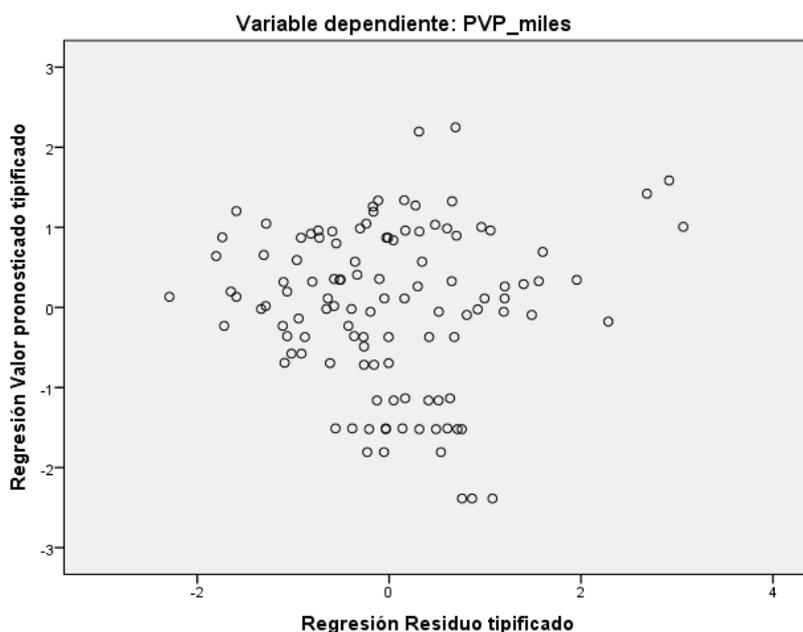
En la Figura 24. se muestra el gráfico de dispersión entre valores observados y los predichos.

Figura 24. Grafico de Dispersión



En el gráfico siguiente, Figura 25., se observa que los residuos se distribuyen de forma simétrica a ambos lados del eje 0 y a lo largo de todo el rango de valores de la estimación y la variabilidad parece constante.

Figura 25. Gráfico de Dispersión-2



Se observa que no parece haber tendencia u homocedasticidad.

6.2.8 Intervalo de Confianza

En la siguiente tabla se representan los 5 vehículos de Peugeot con mayores residuos:

Tabla 27. Vehículos con mayores residuos

Versión vehículo	RES_1	LMCI_1	UMCI_1	LICI_1	UICI_1
807 2.0 HDI Active	8,78	26,97	29,55	22,43	34,08
807 2.0 HDI+ Active Aut	8,36	30,01	33,35	25,76	37,60
807 2.0 HDI+ Active	7,70	29,55	31,85	24,91	36,50
308 1.6 e-HDI Act BL CC	6,54	20,61	21,94	15,56	26,99
RCZ 1.6 THP Aut	5,60	23,10	25,60	18,53	30,16

Las columnas etiquetadas como LMCI_1 y UMCI_1, representan los límites inferior y superior, respectivamente, para la media de la recta de regresión. Las etiquetadas como LICI_1 y UICI_1, límites inferior y superior del intervalo de confianza para la predicción.

Los IC para el precio de coches con unas determinadas características de caball, longit, velmax y co2 tienen una amplitud de aproximadamente 11.600 euros.

6.3 Interpretación del Modelo de Peugeot

El valor de R^2 o coeficiente de determinación es de 0,822, lo que indica que con el conjunto de los cuatro predictores podemos explicar el 82,2% de la variación de los resultados, siendo la proporción de variación no explicada o error del 17,80%, que se deberá tanto a errores aleatorios como a la influencia de variables no incluidas en el modelo.

La ecuación resultante de la regresión era:

$$\text{PVP_miles} = 0,223 \text{ caball} + 0,086 \text{ longit} - 0,150 \text{ velmax} + \\ - 0,057 \text{ CO2} - 6,597$$

Tanto la emisión de partículas de CO2 como la velocidad máxima tienen coeficientes con signo negativo. Esto implica que en aquellos coches cuya velocidad máxima o emisión de CO2 se incrementara, si el resto de variables no sufrieran variaciones, tendrían un precio bastante inferior.

El resto de variables presentan una relación positiva proporcional respecto al precio. Destaca por su coeficiente la potencia del motor (caball). Parece lógico que a mayor potencia de los vehículos estos tengan un precio superior.

El modelo indica, por ejemplo, que si el resto de valores considerados en la ecuación de regresión no variara, una subida en 10 CV de potencia sobre un vehículo podría suponer un aumento en el precio de 2.230 €

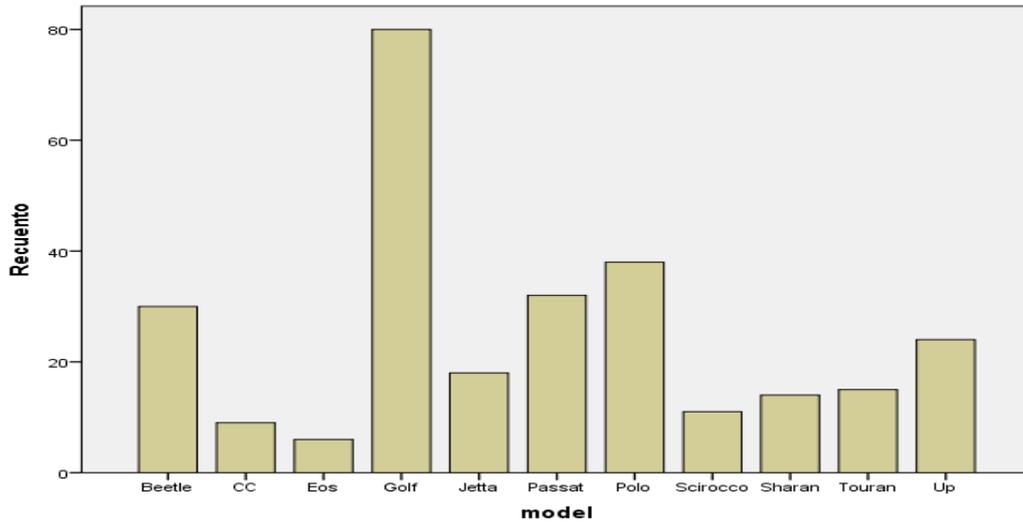
Atendiendo a los resultados de confianza para las predicciones individuales, se estima que la variabilidad debida a otros aspectos no considerados, como el nivel de equipamiento, tipo de cambio o la tracción de los coches, puede hacer que un vehículo cambie hasta en 11.600 € su precio de venta.

El modelo obtenido ofrece unos resultados aceptables teniendo en cuenta que la variabilidad del precio de los coches es explicado en un 82% por las cuatro variables predictoras.

7 MODELO DE REGRESION DE VOLKSWAGEN

El total de turismos que componen la base de datos de Volkswagen es el mayor de las tres marcas analizadas con 277 turismos repartidos en 11 modelos diferentes.

Figura 26. Histograma de modelos de vehículos de Volkswagen



El modelo Golf, líder en ventas en Europa, es el que más versiones comerciales posee con gran diferencia sobre el resto.

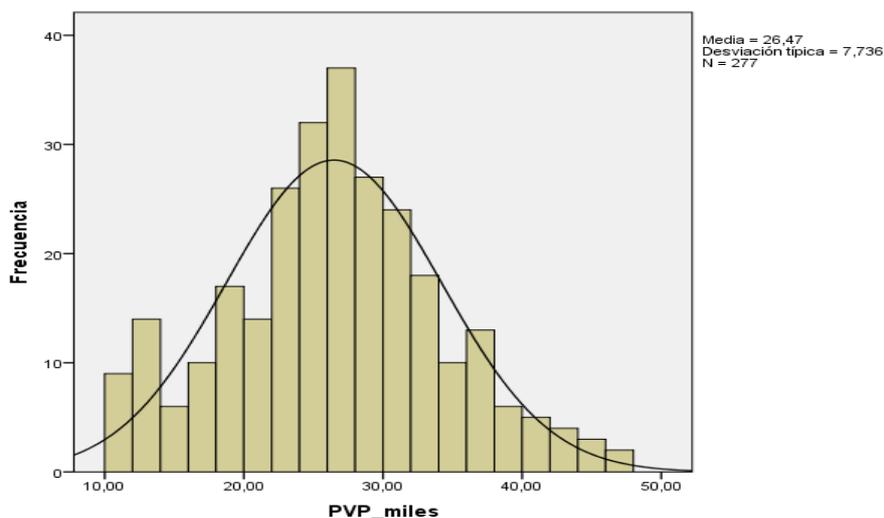
7.1 Análisis de Variables

A continuación se analizan las variables para incluir en la regresión de Volkswagen, como en los casos anteriores.

7.1.1 Variable Objetivo: Precio

El histograma de la variable objetivo, PVP_miles, es el siguiente:

Figura 27. Histograma de variable PVP_miles



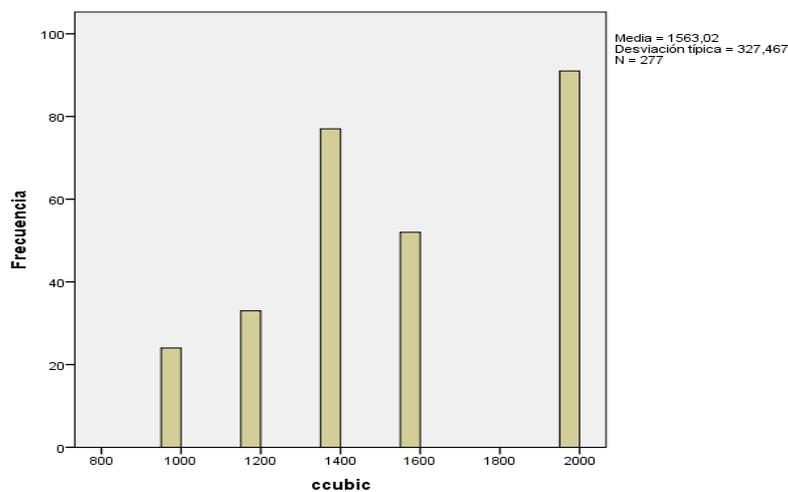
En este caso, se ve que la distribución en el histograma se asemeja bastante con la curva normal. Por ello, no hace falta realizar transformaciones en esta variable.

A continuación se muestran los Histogramas y Regresiones de cada una de las variables susceptibles de entrar en el modelo de regresión frente al precio de los coches.

7.1.2 Centímetros Cúbicos del Motor

El histograma de los centímetros cúbicos (CCUBIC) es el siguiente:

Figura 28. Histograma de variable ccubic



Como se muestra en el gráfico, los coches de la marca Volkswagen se agrupan principalmente en 5 tipos diferenciados por su cilindrada: 1000 CC, 1200 CC, 1400 CC, 1600 CC y 2000 CC.

La regresión resultante entre “ccubic” y el precio se muestran en la tabla 28.

Tabla 28. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,750 ^a	,562	,561	5,12830

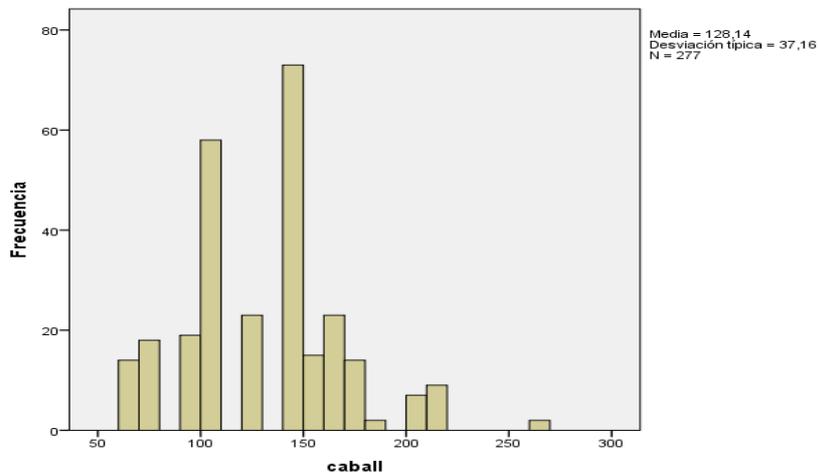
a. Variables predictoras: (Constante), ccubic

Esta variable recoge un 75% de variabilidad del precio con un coeficiente de determinación R^2 del 56,2% por lo que parece bastante significativa para el estudio.

7.1.3 Potencia del Motor

En la siguiente tabla se muestra la frecuencia de la potencia del motor (CABALL).

Figura 29. Histograma de variable caball



La potencia de los coches se distribuye entre los 60 CV de los modelos Volkswagen “Up” y “Polo” hasta los 265 CV de las versiones más deportivas del modelo “Scirocco”.

Tabla 29. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,783 ^a	0,613	0,611	4,82389

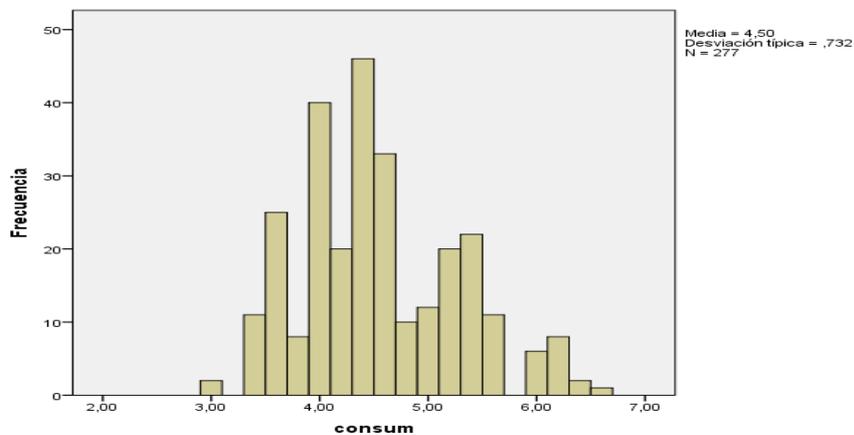
a. Variables predictoras: (Constante), caball

La variable “caball” explica en un 69% la variabilidad del precio, PVP_miles, conforme al coeficiente de determinación, R².

7.1.4 Consumo de Combustible

El siguiente gráfico corresponde al histograma del consumo de carburante extraurbano, variable “consum”.

Figura 30. Histograma de variable consum



El consumo es una variable con una frecuencia de sus datos muy dispersa. Toma valores que van desde 3 litros de consumo del Volkswagen “Polo 1.2 TDI Blue” hasta los 6,60 litros que consume el modelo “Sharan 2.0 TSI Sport DSG”.

Tabla 30. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,545 ^a	0,297	0,294	6,49863

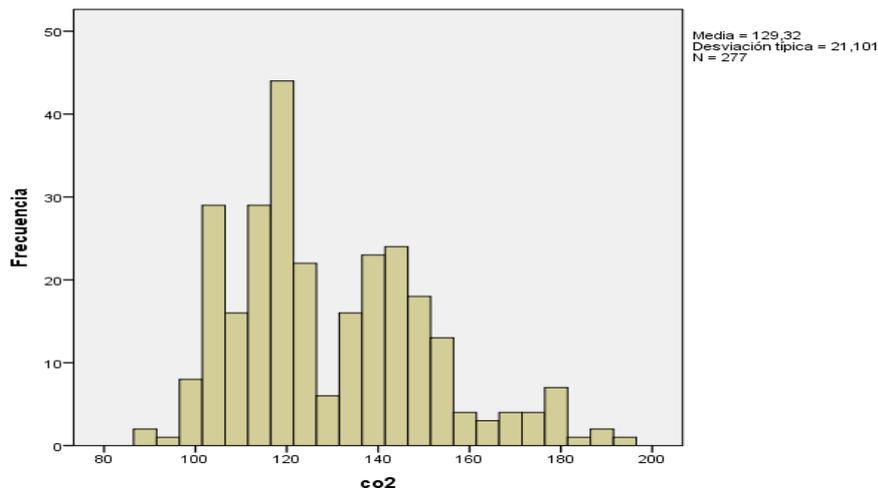
a. Variables predictoras: (Constante), consum

La variable “consum” explica en un 30% la variabilidad del precio de los coches.

7.1.5 Emisión de Gases Contaminantes

La distribución de los coches Volkswagen en función de la emisión de gases CO₂ se muestra en este gráfico:

Tabla 31. Histograma de variable CO₂



La emisión de partículas de CO₂ se extiende desde 89 gr/km hasta 196 gr/km para el vehículo que más contamina.

La regresión de CO₂ frente a PVP_miles se muestra en la tabla 32.

Tabla 32. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,677 ^a	0,458	0,456	5,70425

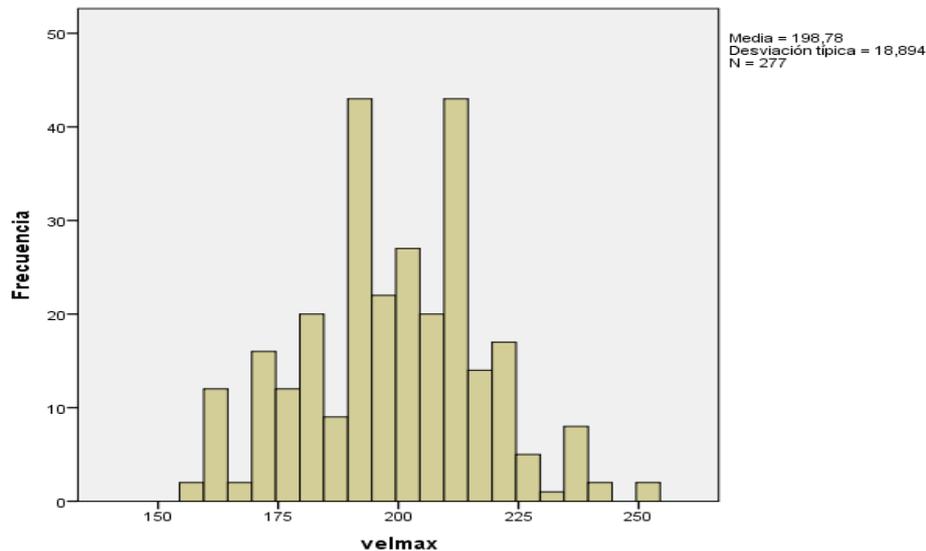
a. Variables predictoras: (Constante), co2

La variable CO₂ explica en un 46% la variabilidad del precio conforme al coeficiente de determinación R².

7.1.6 Velocidad Máxima

Este es el gráfico con el histograma de la Velocidad Máxima:

Figura 31. Histograma de variable velmax



La velocidad que puede alcanzar un turismo de Volkswagen tiene un amplio abanico de posibilidades que va desde los 157 km/h que alcanza el Polo 1.2 Advance hasta los 265 km/h que alcanza el vehículo deportivo “Scirocco 2.0 TSI R”.

La regresión de esta variable con el precio es:

Tabla 33. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,692 ^a	,479	,477	5,59307

a. Variables predictoras: (Constante), velmax

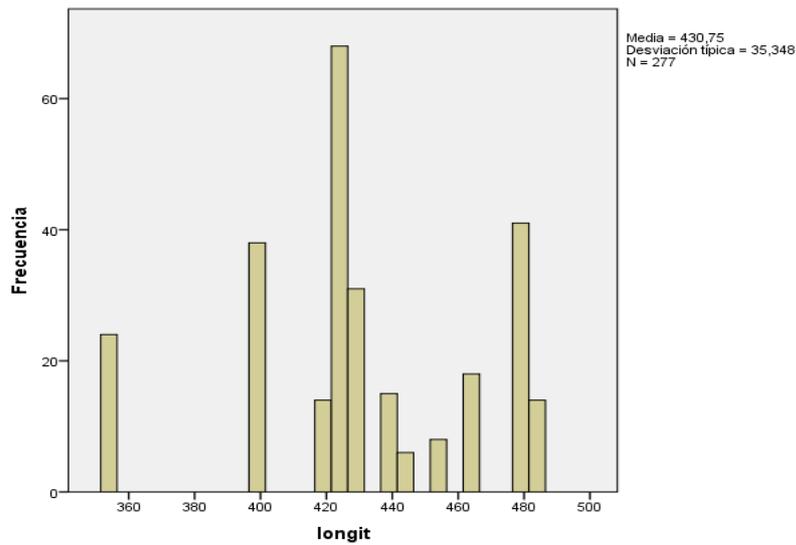
Se puede observar en los datos del modelo de regresión que el coeficiente de determinación R^2 tiene un valor de explicación de casi el 48%: Esto significa que la variable velmax explica en ese porcentaje la variabilidad de PVP_miles.

7.1.7 Longitud de los Vehículos

Como puede verse en el histograma de la Figura 32., la longitud de los coches Volkswagen va desde los 354 cm del modelo UP hasta los 485 cm de las versiones del modelo “Sharan”.

La longitud parece ser un parámetro estrechamente relacionado en esta marca con el modelo de coche.

Figura 32. Histograma de variable longit



El resultado del modelo de Regresión de la longitud de los coches con el precio se refleja en la siguiente tabla:

Tabla 34. resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,835 ^a	,697	,696	4,26666

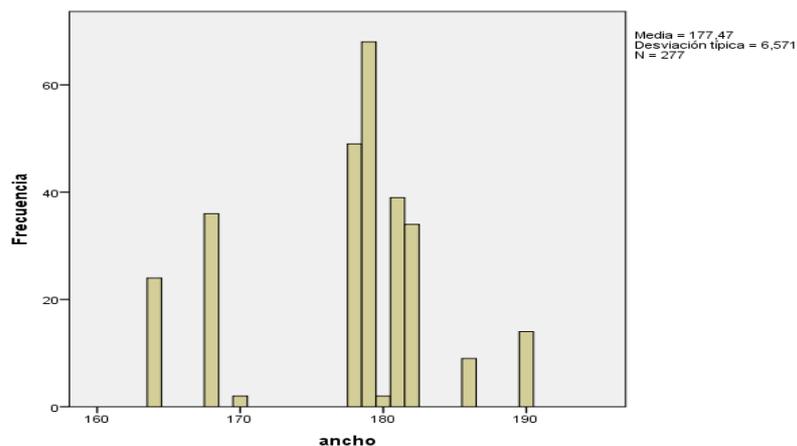
a. Variables predictoras: (Constante), longit

Según el resumen de este modelo, la variabilidad del precio es recogida en un 70% por la longitud del vehículo por lo que es importante su inclusión en la regresión posterior.

7.1.8 Ancho de los Vehículos

A continuación se reflejan las frecuencias del ancho de los vehículos (ANCHO).

Figura 33. Histograma de variable ancho



El ancho de los coches Volkswagen, al igual que la longitud, va en consonancia con el modelo de coche que sea. En esta marca los coches oscilan entre 164 cm del modelo UP y 190 cm del modelo Sharan.

Este es el modelo de regresión lineal simple entre la variable ancho y el precio:

Tabla 35. Resumen del Modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,866 ^a	,750	,749	3,87360

a. Variables predictoras: (Constante), ancho

La variación del precio está explicada por el ancho de los coches en un 75% conforme al coeficiente de determinación.

7.2 Resultados de la Regresión de Volkswagen

A continuación se muestran los distintos indicadores del modelo de Volkswagen.

7.2.1 Estadísticos Descriptivos

En la siguiente tabla se detallan las variables que se van a correlacionar, aportando los tres índices que sintetizan las distribuciones: media, desviación típica y tamaño de la muestra.

Tabla 36. Estadísticos descriptivos de Volkswagen

	Media	Desviación típica	N
PVP_miles	26,4684	7,73616	277
ccubic	1563,02	327,467	277
caball	128,14	37,160	277
consum	4,4971	,73217	277
co2	129,32	21,101	277
velmax	198,78	18,894	277
longit	430,75	35,348	277
ancho	177,47	6,571	277

El número de automóviles que forman la muestra es de 277 unidades.

El precio medio de los automóviles de Volkswagen es de casi 26.500 €, el precio medio más alto de las tres marcas objeto de este estudio.

7.2.2 Matriz de Correlaciones

La Tabla 37. contiene la matriz de correlaciones de las variables.

Tabla 37. Matriz de Correlaciones de Volkswagen

	PVP_miles	ccubic	caball	Consum	co2	velmax	longit	ancho
PVP_miles	1	0,75	0,78	0,54	0,68	0,69	0,83	0,87
Ccubic	0,75	1	0,65	0,21	0,42	0,59	0,62	0,65
Caball	0,78	0,65	1	0,67	0,74	0,94	0,59	0,66
Consum	0,54	0,21	0,67	1	0,94	0,53	0,35	0,45
co2	0,68	0,42	0,74	0,94	1	0,57	0,47	0,54
Velmax	0,69	0,59	0,94	0,53	0,57	1	0,62	0,61
Longit	0,83	0,62	0,59	0,35	0,47	0,62	1	0,88
Ancho	0,87	0,65	0,66	0,45	0,54	0,61	0,88	1

Como en las anteriores regresiones, se ve una correlación alta entre las variables “co2” y “consum” y también entre la velocidad máxima (velmax) y los caballos del motor (caball).

Cuanto mayor sea la potencia de un vehículo más probabilidad tendrá de alcanzar una velocidad tope superior.

Al considerar que estas variables son muy significativas para el estudio se ha decidido incluirlas como candidatas al modelo de regresión.

7.2.3 Resumen del Modelo de Volkswagen

El siguiente cuadro, nos indica el modelo final resultante (Modelo 7):

Tabla 38. Resumen del Modelo de Volkswagen

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
7	0,948	0,899	0,897	2,48604

Variables predictoras: (Constante), ancho, caball, ccubic, longit, velmax

El coeficiente de correlación múltiple, **R**, es 0,948

R², coeficiente de determinación arroja un resultado de 0,899. Según este dato, la variación de la variable dependiente (PVP_miles) es explicada en un 90% por las variables predictoras del modelo de regresión. Es decir, el modelo se ajusta bastante bien a los datos,

R² corregida, coeficiente de determinación ajustado, da un valor de 0,897; indica que las variables predictoras explican un 89,7 % de la varianza que presenta el precio considerando el número de variables predictoras.

Error típico de la estimación: da un resultado de 2,486, y corresponde con la raíz cuadrada de la varianza residual.

7.2.4 Análisis de la Varianza

Esta es la tabla de Análisis de la Varianza (ANOVA):

Tabla 39. ANOVA de Volkswagen

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
7 Regresión	14843,23	5	2968,646	480,334	,000 ^h
Residual	1674,881	271	6,18		
Total	16518,111	276			

a. Variable dependiente: PVP_miles

h. Variables predictoras: (Constante), ancho, caball, ccubic, longit, velmax

El modelo se muestra válido para representar los datos. El valor de significación obtenido ($p=0,000$) indica que la probabilidad de que el conjunto de las variables predictoras introducidas no sea suficiente para aportar explicación por los valores predichos de PVP_miles es nula. Dicho de otro modo, las variaciones en la variable PVP_miles es explicada significativamente por el conjunto de variables predictivas identificadas (ancho, caball, ccubic y longit).

7.2.5 Coeficientes de Regresión

Los coeficientes resultantes en el modelo de regresión son:

Tabla 40. Coeficientes de la Regresión de VOLKSWAGEN

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
7 (Constante)	-29,770	8,183		-3,638	0,000
ancho	0,191	0,056	0,163	3,424	0,001
caball	0,183	0,014	0,878	12,820	0,000
ccubic	0,004	0,001	0,151	5,373	0,000
longit	0,098	0,010	0,447	9,748	0,000
velmax	-0,246	0,026	-0,600	-9,325	0,000

El conjunto de variables predictoras resultantes en el modelo tienen un nivel de significación menor o igual a 0,001.

La variación de la variable Velmax tiene una relación inversamente proporcional a la variación de la variable dependiente PVP_miles.

7.2.6 Ecuación de Regresión

La ecuación resultante de la regresión de Volkswagen es:

$$\text{PVP_miles} = 0,191 \text{ ancho} + 0,183 \text{ caball} + 0,004 \text{ ccubic} + \\ + 0,098 \text{ longit} - 0,246 \text{ velmax} - 29,770$$

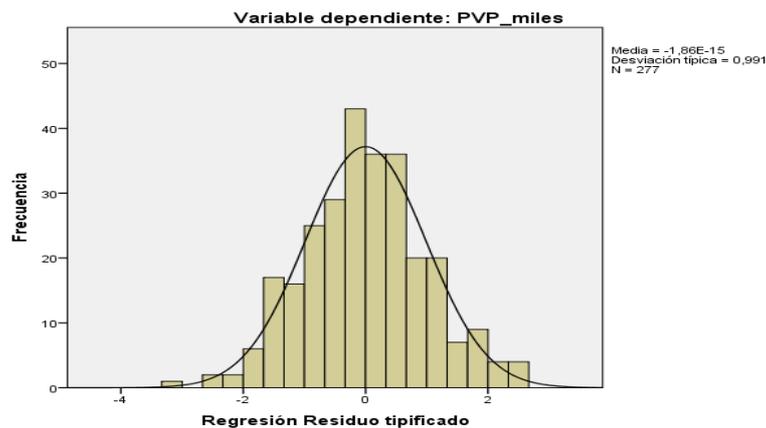
7.2.7 Diagnóstico del Modelo de Regresión

Se ha realizado un análisis de los residuos (valores predichos - valores observados): normalidad, tendencias, y contraste de homocedasticidad (homogeneidad de varianzas entre los grupos) para estudiar la validez del modelo de regresión.

7.2.7.1 Histograma

El siguiente gráfico es el Histograma de los residuos estandarizados.

Figura 34. Histograma de Residuos



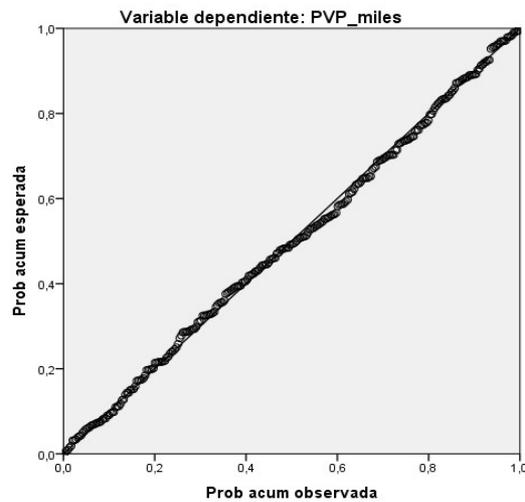
En este gráfico podemos comparar las frecuencias observadas frente a los valores estandarizados que corresponden a esas frecuencias.

El gráfico refleja que los residuales observados y los esperados se asemejan bastante a una distribución normal.

7.2.7.2 Gráfico de Normalidad

En este gráfico, Figura 35., se expresan los valores reales en ordenadas, y los esperados en abscisas (esperados si la distribución fuera normal).

Figura 35. Gráfico P-P normal de regresión Residuo tipificado

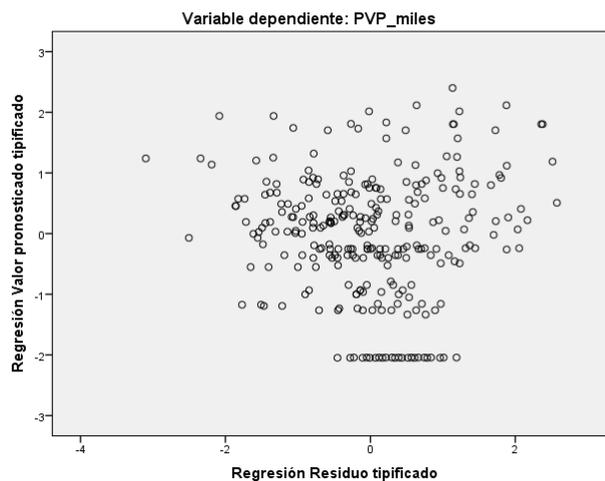


Los valores observados siguen la recta diagonal que parte del origen (0,0), por lo que se puede afirmar que la distribución sigue una ley normal.

7.2.7.3 Gráfico de Dispersión

El siguiente Gráfico de dispersión muestra la relación entre los residuos tipificados y los valores pronosticados tipificados.

Tabla 41. Gráfico de dispersión



Los residuos se distribuyen de forma simétrica a ambos lados del eje 0 y a lo largo de todo el rango de valores de la estimación y la variabilidad parece constante, por lo que no se aprecia que haya tendencia u homocedasticidad.

7.2.8 Intervalo de Confianza

En esta tabla 42. se muestran los 5 vehículos con mayores residuos de Volkswagen.

Tabla 42. Vehículos con mayores residuos

Versión vehículo	RES_1	LMCI_1	UMCI_1	LICI_1	UICI_1
Eos 2.0 TDI BM DSG	6,40	29,66	30,72	25,27	35,12
Eos 2.0 TFSI DSG	6,25	34,29	36,05	30,20	40,14
Sharan 2.0 TDI BM Sp DSG	5,91	38,73	40,67	34,71	44,69
Sharan 2.0 TDI Sport 4M	5,86	38,73	40,67	34,71	44,69
Golf 2.0 TDI+ BM Cab DSG	5,39	27,48	28,70	23,16	33,03

Las columnas etiquetadas como LMCI_1 y UMCI_1, representan los límites inferior y superior, respectivamente, para la media de la recta de regresión. Las etiquetadas como LICI_1 y UICI_1, límites inferior y superior del intervalo de confianza para la predicción.

Los IC para el precio de coches con unas determinadas características de ancho, caball, ccubic, longit y velmax tienen una amplitud de aproximadamente 10.000 euros.

7.3 Interpretación del Modelo de Volkswagen

El valor de R^2 o coeficiente de determinación es de 0,899, lo que indica que con el conjunto de los cinco predictores podemos explicar el 89,9% de la variación de los resultados, siendo la proporción de variación no explicada o error del 10,10%, que se deberá tanto a errores aleatorios como a la influencia de variables no incluidas en el modelo.

La ecuación resultante de la regresión era:

$$\text{PVP_miles} = 0,191 \text{ ancho} + 0,183 \text{ caball} + 0,004 \text{ ccubic} + 0,098 \text{ longit} + \\ - 0,246 \text{ velmax} - 29,770$$

La velocidad máxima tiene un coeficiente muy representativo y su signo es negativo. Esto indica que, por ejemplo, si un coche aumenta su velocidad máxima respecto a otros con el resto de características inalteradas puede tener un precio menor. Posiblemente este dato refleje la relación existente entre los diferentes tipos de motorización y las prestaciones del vehículo.

El resto de variables presentan una relación positiva proporcional respecto al precio. Destaca por su coeficiente la potencia del motor (caball) que refleja que ante una variación de la potencia de los coches el precio de los mismos se vería directamente afectado por esa variación.

El modelo indica, por ejemplo, que si el resto de valores considerados en la ecuación de regresión permanecieran estables, una subida en 10 CV de potencia sobre un vehículo podría suponer un aumento en su precio de unos 1.800 €

Atendiendo a los resultados de confianza para las predicciones individuales, se estima que la variabilidad debida a errores de estimación u otros aspectos no considerados en el modelo, puede suponer que un vehículo cambie hasta en 10.000 € su precio de venta.

El modelo ofrece unos resultados bastante aceptables considerando que la variabilidad del precio de los coches es explicado por las cinco variables predictoras en un 90%.

8 MODELO DE REGRESION DE VOLKSWAGEN, PEUGEOT Y RENAULT

Una vez realizadas las regresiones individualizadas de las marcas Volkswagen, Renault y Peugeot se ha realizado una regresión con todos los coches de las tres marcas. El fin es obtener un modelo de estimación del precio de los coches de las tres marcas y comparar resultados frente a las regresiones individuales de cada marca.

La base de datos está compuesta por un total de 540 versiones de vehículos de las tres marcas comerciales analizadas.

8.1 Análisis de Variables

Se ha incluido en el análisis las mismas variables de las regresiones individuales y añadido las variables categorizadas de las marcas de los fabricantes.

Las variables categorizadas introducidas en la regresión son dos (n-1 marcas):

Marca_R: coches de Renault

Marca_W: coches de Volkswagen

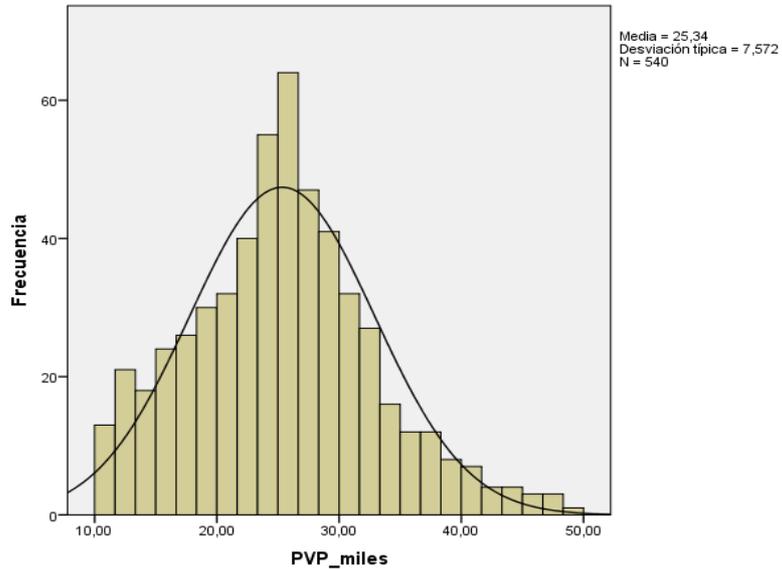
Obviamente, ambas variables son dicotómicas y sus posibles valores son “0” y “1”. El valor “1” representa la pertenencia de un vehículo a una marca y “0” la no pertenencia a la misma.

Como el resto de variables candidatas a incluirse en la regresión son las mismas que las de los modelos anteriores, tan sólo se detallará a continuación las frecuencias de cada variable (histograma) y su interpretación.

8.1.1 Variable Objetivo: Precio

El histograma de la variable objetivo, PVP_miles, es el siguiente:

Figura 36. Histograma de variable PVP_miles

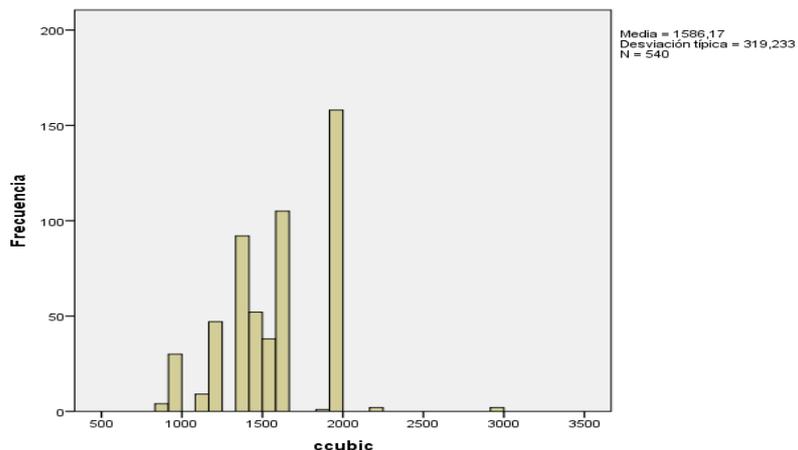


Comparamos el histograma con la distribución normal y vemos que se asemeja mucho por lo que no es necesario realizar transformaciones de esta variable.

8.1.2 Centímetros Cúbicos del Motor

El histograma de los centímetros cúbicos (CCUBIC) es el siguiente:

Figura 37. Histograma de variable ccubic

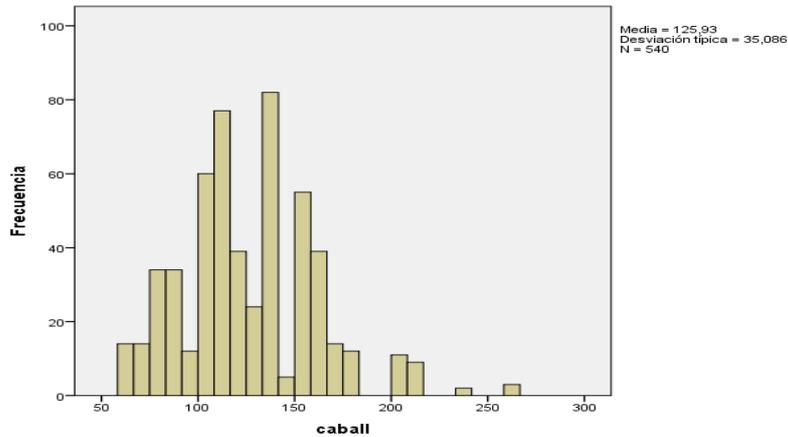


Como se muestra en el gráfico, los coches de las tres marcas se agrupan principalmente en 5 tipos diferenciados por su cilindrada: 1000 CC, 1200 CC, 1400 CC, 1600 CC y 2000 CC.

8.1.3 Potencia del Motor

El histograma de la variable que indica la potencia del motor (CABALL) es:

Figura 38. Histograma de variable caball

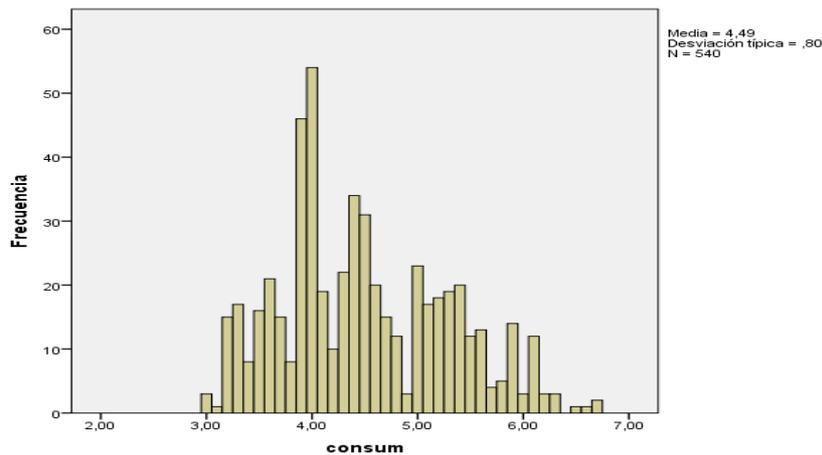


La mayor parte de los vehículos de las tres marcas, concentran su potencia entre potencia entre 100 CV y 170 CV.

8.1.4 Consumo de Combustible

El grafico que sigue corresponde al histograma del consumo de carburante.

Figura 39. Histograma de variable consum

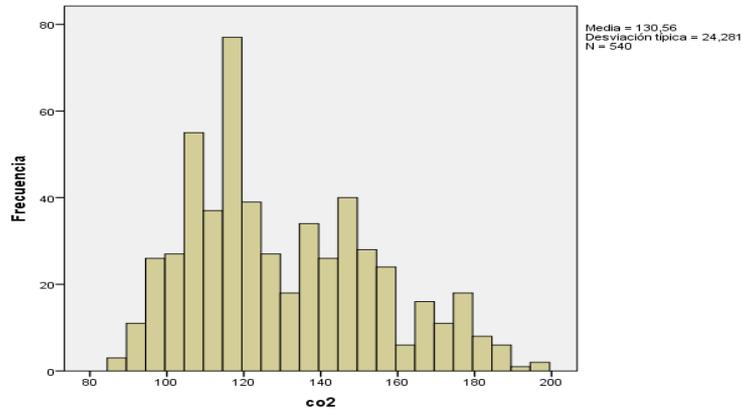


Destacan especialmente en cuanto a su número los vehículos con un consumo cercano a los 4 litros por cada 100 km. recorridos.

8.1.5 Emisión de Gases Contaminantes

La distribución de emisión de gases CO2 se muestra en este gráfico:

Figura 40. Histograma de variable CO2

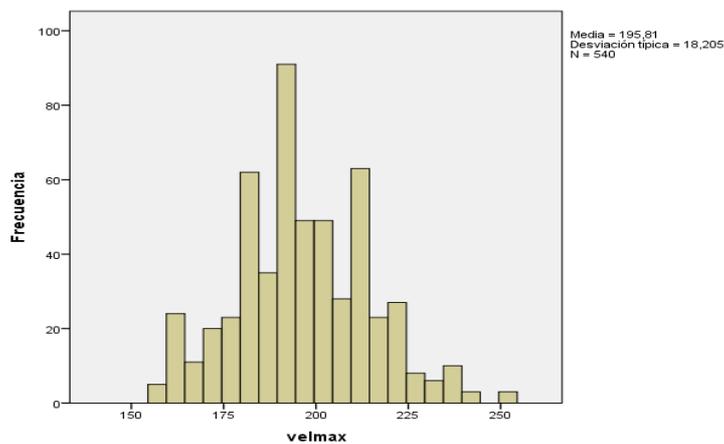


Los valores más frecuentes de emisión de CO2 son los 110 gr/km y 120 gr/km.

8.1.6 Velocidad Máxima

Según el histograma de la variable VELMAX, la mayoría de los vehículos alcanza una velocidad máxima comprendida entre 180 km/h y 215 km/h.

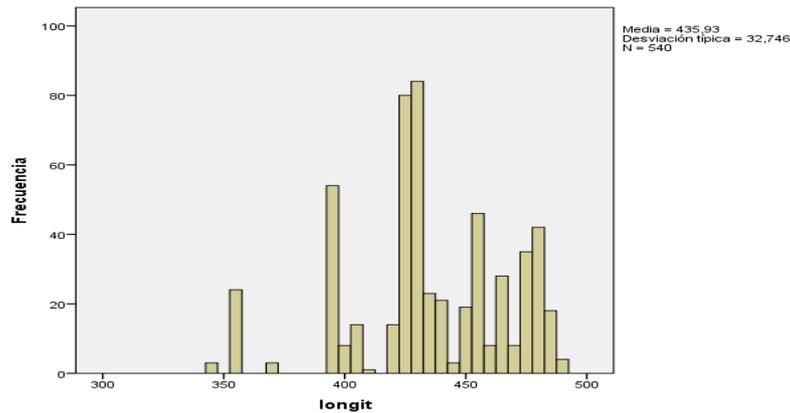
Figura 41. Histograma de variable velmax



8.1.7 Longitud de los Vehículos

El histograma con la longitud de los coches (LONGIT), que se reflejan en la Figura 42., indica que la longitud de los coches tiene una gran variación que va desde 345 cm del modelo Peugeot “107” hasta los 490 cm. del Renault Latitude.

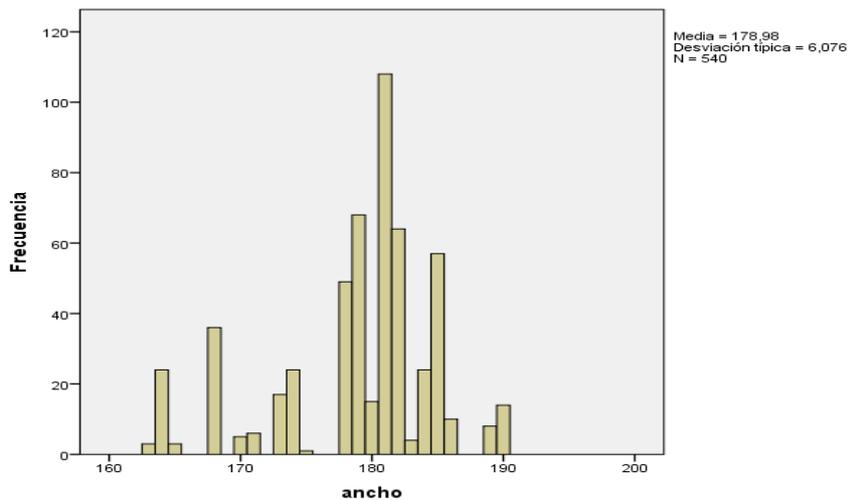
Figura 42. Histograma de variable longit



8.1.8 Ancho de los Vehículos

A continuación se reflejan las frecuencias del ancho de los vehículos (ANCHO):

Figura 43. Histograma de variable ancho



Como puede verse en el gráfico la distribución que siguen los automóviles es bastante irregular. El modelo más estrecho corresponde al Peugeot “107” con 163 cm. y el más ancho es el Volkswagen “Sharan” con 190 cm.

8.2 Resultados de la Regresión de Marcas Conjuntas

A continuación se detallan los resultados más significativos de la regresión de los vehículos de las marcas Volkswagen, Renault y Peugeot conjuntamente.

8.2.1 Estadísticos Descriptivos

En el siguiente cuadro se observan las variables que se van a correlacionar y los índices que sintetizan las distribuciones: media, desviación típica y tamaño de la muestra.

Tabla 43. Estadísticos Descriptivos

	Media	Desviación típica	N
PVP_miles	25,3425	7,57181	540
Ccubic	1586,17	319,233	540
Caball	125,93	35,086	540
Consum	4,4946	0,79969	540
co2	130,56	24,281	540
Velmax	195,81	18,205	540
Longit	435,93	32,746	540
Ancho	178,98	6,076	540
Marca_R	0,28	0,447	540
Marca_W	0,51	0,5	540

El número de automóviles que forman la muestra es de 540 unidades. Los vehículos de Renault suponen un 28% del total de ellos, el 51% son de Volkswagen y el 21% corresponden a Peugeot.

El precio medio de los automóviles de las tres marcas conjuntas es de 25.350 €.

8.2.2 Matriz de Correlaciones

Como puede observarse en la Tabla 44., existe una alta correlación, señalada en negrita, entre las variables consumo y emisión de gases contaminantes.

Tabla 44. Correlaciones de VOLKSWAGEN, PEUGEOT y RENAULT

	PVP_miles	ccubic	caball	consum	co2	velmax	longit	ancho	Marca_R	Marca_W
PVP_miles	1	0,73	0,787	0,466	0,576	0,696	0,748	0,719	0,018	0,153
Ccubic	0,73	1	0,708	0,296	0,493	0,621	0,594	0,588	0,041	-0,075
Caball	0,787	0,708	1	0,626	0,706	0,927	0,562	0,59	-0,031	0,065
Consum	0,466	0,296	0,626	1	0,942	0,482	0,301	0,384	0,065	0,003
co2	0,576	0,493	0,706	0,942	1	0,548	0,415	0,473	0,058	-0,052
Velmax	0,696	0,621	0,927	0,482	0,548	1	0,546	0,496	-0,106	0,168
Longit	0,748	0,594	0,562	0,301	0,415	0,546	1	0,849	0,157	-0,162
Ancho	0,719	0,588	0,59	0,384	0,473	0,496	0,849	1	0,128	-0,256
Marca_R	0,018	0,041	-	0,065	0,058	-0,106	0,157	0,128	1	-0,634
Marca_W	0,153	-0,075	0,031 0,065	0,003	-	0,168	-	-	-0,634	1
					0,052		0,162	0,256		

Al igual que en las anteriores regresiones, las he introducido en el análisis por ser muy significativas para el estudio.

8.2.3 Resumen del Modelo Conjunto de Marcas

La regresión para explicar el precio en función de las características técnicas que se obtiene con SPSS mediante el método “paso a paso” es la siguiente:

Tabla 45. Resumen del modelo Volkswagen, Peugeot y Renault

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
7	,929 ^a	0,864	0,862	2,81239

Variables predictoras: (Constante), caball, longit, Marca_W, velmax, ccubic, Marca_R, ancho

El modelo resultante se ajusta muy bien a los datos.

El coeficiente de determinación es 0,864. Es decir, la proporción de la variación de la variable dependiente (PVP_miles) es explicada por el modelo de regresión, con el conjunto de variables predictoras, en un 86,4%.

8.2.4 Análisis de la Varianza

La tabla 46. muestra los datos del Análisis de la varianza.

Tabla 46. ANOVA modelo de Volkswagen, Peugeot y Renault

Modelo	Suma de cuadrados	Gl	Media cuadrática	F	Sig.
7 Regresión	26694,242	7	3813,463	482,135	,000 ^h
Residual	4207,874	532	7,91		
Total	30902,116	539			

a. Variable dependiente: PVP_miles

h. Variables predictoras: (Constante), caball, longit, Marca_W, velmax, ccubic, Marca_R, ancho

El modelo se muestra válido para representar los datos. El valor de significación obtenido ($p=0,000$) indica que las variaciones en la variable PVP_miles es explicada parcialmente por el conjunto de variables predictivas identificadas.

8.2.5 Coeficientes de la regresión

Los coeficientes resultantes en el modelo de regresión se presentan en la siguiente tabla.

Tabla 47. Coeficientes Regresión de Volkswagen, Peugeot y Renault

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
7 (Constante)	-45,734	6,614		-6,915	,000
Caball	,151	,011	,701	13,278	,000
Longit	,081	,008	,348	10,170	,000
Marca_W	5,924	,341	,391	17,373	,000
Velmax	-,166	,020	-,400	-8,290	,000
Ccubic	,005	,001	,197	8,043	,000
Marca_R	2,718	,361	,161	7,526	,000
Ancho	,214	,044	,172	4,866	,000

a. Variable dependiente: PVP_miles

Las interacciones con las variables marcas no aportan demasiado y dan lugar a un modelo bastante menos razonable de interpretar.

En cambio, si es interesante la interpretación de los coeficientes de las variables Marca_R (Renault) y Marca_W (Peugeot) que ponen de manifiesto que a igual características técnicas de las variables estudiadas, estas marcas son más caras que Peugeot.

La marca Volkswagen es la marca más cara de las tres y, según este modelo, puede tener hasta unos 6000 € de sobreprecio frente a un vehículo de Peugeot con iguales características.

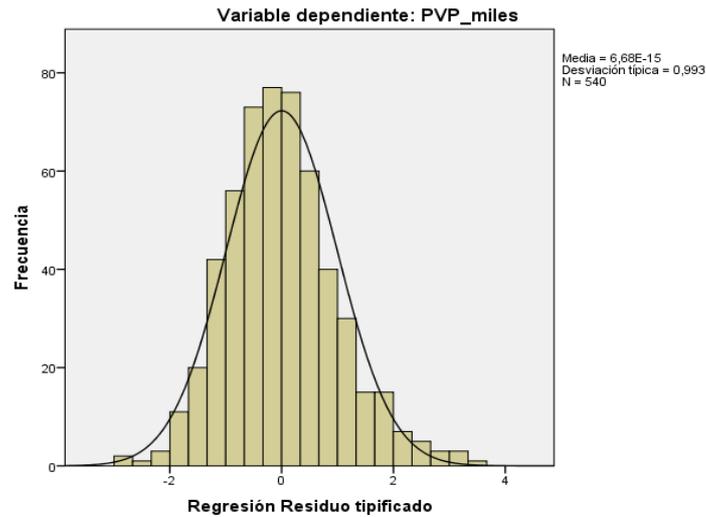
8.2.6 Diagnóstico del Modelo de Regresión

Para diagnosticar la validez del modelo de regresión, se ha realizado un análisis de los residuos (valores predichos - valores observados): normalidad, tendencias, y contraste de homocedasticidad (homogeneidad de varianzas entre los grupos).

8.2.6.1 Histograma

El histograma de la Figura 44., se muestra la comparativa entre las frecuencias observadas frente a los valores estandarizados que corresponden a tales frecuencias.

Figura 44. Histograma de Residuos

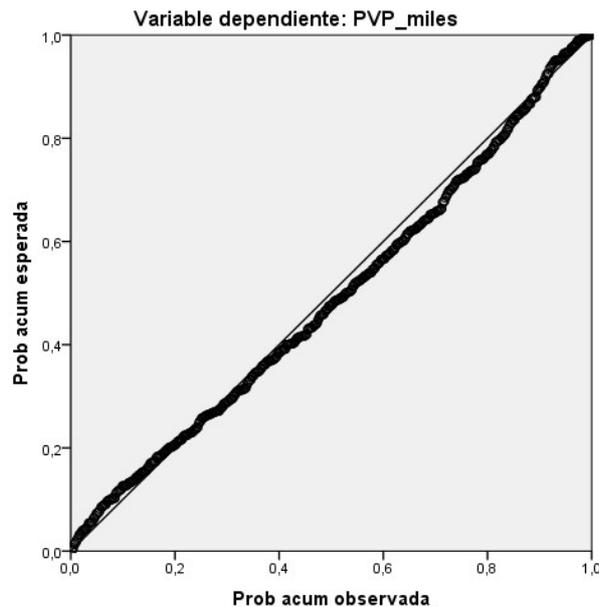


Como puede verse en el gráfico, los residuales observados y los esperados se asemejan bastante a la ley normal, y como resultado podríamos decir que en nuestro caso la distribución sigue una normalidad.

8.2.6.2 Gráfico de Normalidad

Los valores observados siguen la recta diagonal, por lo que, como era de esperar como en las regresiones individuales por marca realizadas anteriormente, se puede afirmar que la distribución sigue una ley normal.

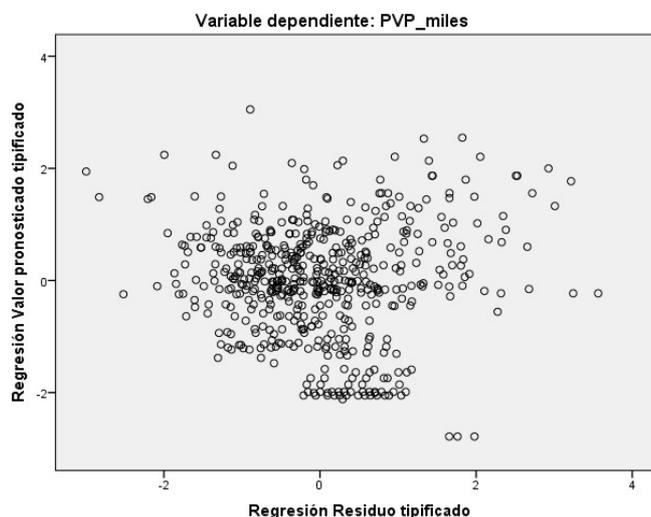
Figura 45. Gráfico P-P normal de regresión Residuo tipificado



8.2.6.3 Gráfico de Dispersión

En el siguientes gráfico, Figura, 46, muestra la dispersión residual entre los valores observados y los predichos.

Figura 46. Gráfico de dispersión



Como puede observarse los residuos se distribuyen de forma simétrica a ambos lados del eje 0 y a lo largo de todo el rango de valores de la estimación y la variabilidad parece constante, no parece existir tendencia.

8.2.7 Intervalos de Confianza

En la siguiente tabla se muestran los 10 vehículos con mayores residuos del modelo.

Tabla 48. Vehículos con Mayores Residuos del Modelo de Regresión

MARCA	VERSION	RES_2	LMCI_2	UMCI_2	LICI_2	UICI_2
Renault	Megane 1.5 dCi+ Flord CC	10,02	23,22	24,24	18,18	29,28
Renault	Megane 1.5 dCi+ Pr CC EDC	9,12	23,22	24,24	18,18	29,28
Renault	Espace 2.0 dCi* In Aut	9,04	36,9	38,71	32,21	43,4
Renault	Espace 2.0 dCi+ Initiale	8,46	33,96	35,42	29,12	40,26
Renault	G Espace 2.0 dCi* In Aut	8,23	38,48	40,36	33,81	45,02
Renault	G Espace 2.0 dCi+ In	7,65	35,58	37,02	30,73	41,87
Renault	Laguna 1.5 dCi Emo Coupe	7,53	23,68	24,87	18,71	29,83
Peugeot	807 2.0 HDI Active	7,47	28,88	30,26	24	35,14
Volkswagen	Sharan 2.0 TDI BM Sp DSG	7,11	37,66	39,34	32,91	44,09
Volkswagen	Sharan 2.0 TDI Sport 4M	7,06	37,66	39,34	32,91	44,09

Las columnas etiquetadas como LMCI_2 y UMCI_2, representan los límites inferior y superior, respectivamente, para la media de la recta de regresión. Las etiquetadas como LICI_2 y UICI_2, límites inferior y superior del intervalo de confianza para la predicción.

El Intervalo de Confianza sobre estimaciones individuales del precio de los automóviles de Volkswagen, Peugeot y Renault según este modelo de regresión tiene una amplitud de aproximadamente 11.000 euros.

8.3 Análisis ACP

El ACP es una herramienta descriptiva interesante que permite reducir la información de las variables utilizadas en el modelo por un número menor. Normalmente las dos primeras son las más interesantes porque se puede hacer una representación en el plano.

En este caso esas dos primeras componentes recogen el 86% de la variabilidad como puede verse en la siguiente tabla.

Tabla 49. Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	4,395	73,248	73,248	4,395	73,248	73,248
2	0,798	13,306	86,554	0,798	13,306	86,554
3	0,394	6,565	93,119			
4	0,205	3,422	96,541			
5	0,165	2,744	99,284			
6	0,043	0,716	100			

Método de extracción: Análisis de Componentes principales.

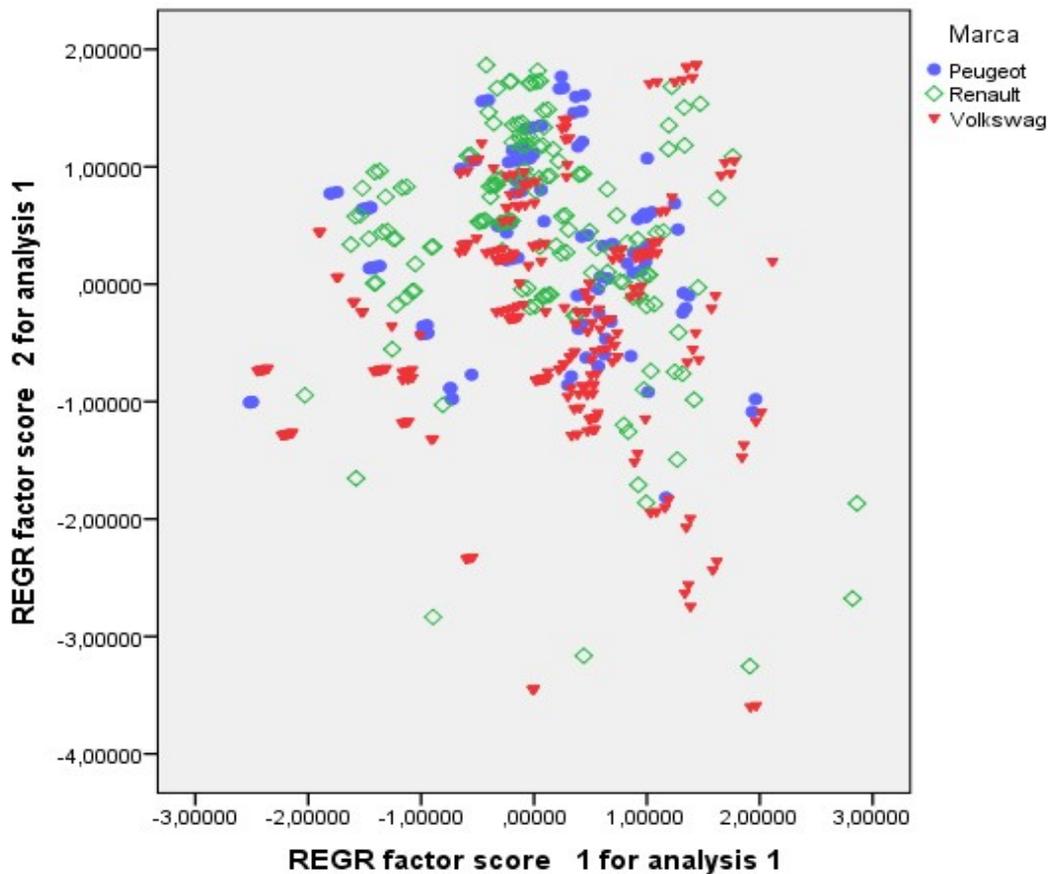
La matriz de componentes reducida a dos componentes principales se muestra en la tabla 50.

Tabla 50. Matriz de Componentes

	Componente	
	1	2
Ccubic	0,825	-0,092
PVP_miles	0,915	0,036
Caball	0,894	-0,388
Velmax	0,838	-0,454
Longit	0,835	0,456
Ancho	0,824	0,473

Con los datos guardados de la reducción a los dos componentes principales se puede elaborar el siguiente gráfico de dispersión por marca de coche.

Figura 47. Gráfico de dispersión por Marca



La primera componente principal está muy relacionada con el precio y la segunda inversamente con los caballos y la velocidad máxima.

Este gráfico puede ser muy útil para encontrar vehículos con determinadas características de precio, potencia o tamaño y comparar los vehículos disponibles de cualquiera de las tres marcas.

En este caso, por ejemplo, si se quiere adquirir un coche barato pero potente en caballos se debería buscar desde la esquina inferior a la izquierda.

8.4 **Análisis Cluster**

Se ha realizado un análisis de conglomerados con el objeto de agrupar los coches en relación a sus características técnicas y hacer luego análisis por grupos.

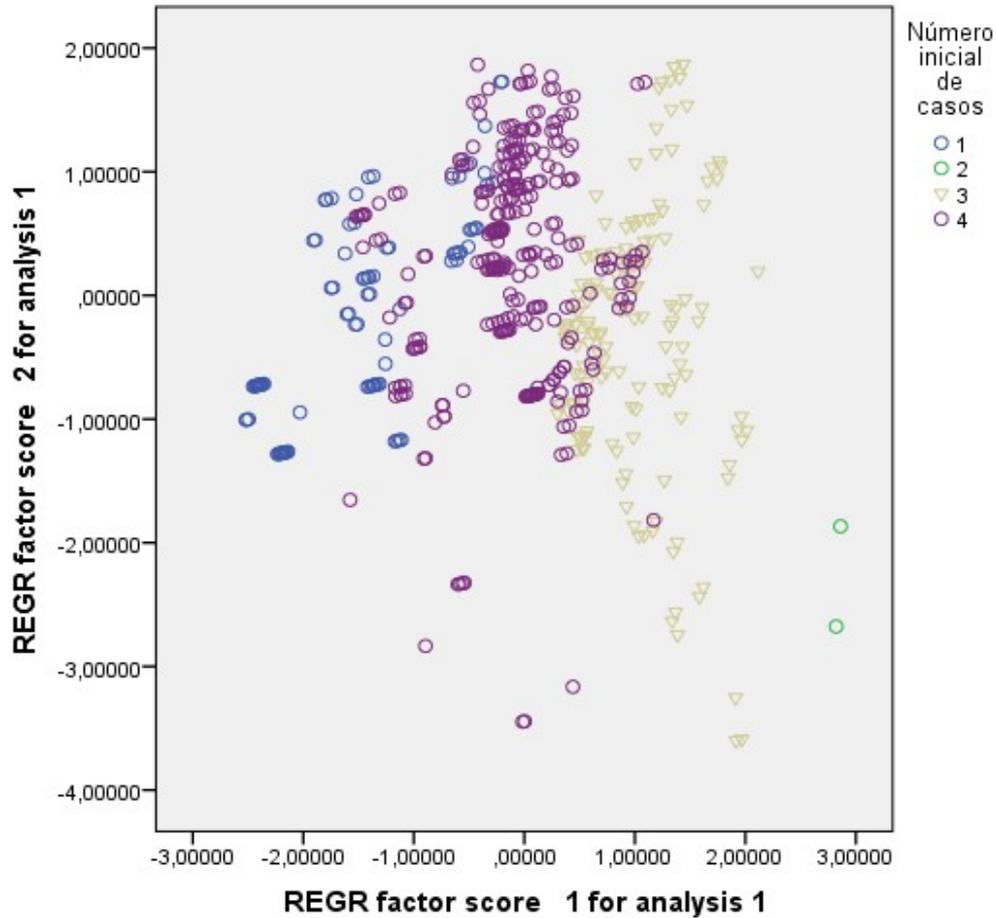
El método empleado para realizar el Análisis de Conglomerados ha sido el de K-medias.

Un resultado razonable es el que nos da la clasificación en 4 grupos que corresponde a lo que habitualmente se entiende por coches de gama baja, intermedia y alta además de

una clase formada por dos coches que pueden considerarse fuera de esta clasificación y que son Renault (Laguna 3.0 dci GT Coupé y Renault Laguna 3.0 dci Aut). Ambas versiones de Renault tienen unas características técnicas muy específicas en relación a la deportividad (motor de alta cilindrada, mucha potencia, alto consumo y velocidad máxima elevada).

Los grupos resultantes se representan claramente en el plano de las dos primeras componentes principales que se incluye a continuación (figura 48):

Figura 48. Gráfico de dispersión por Gama de vehículo



Los tres grupos correspondientes a la Gama de vehículo son:

- 1: Gama Baja
- 4: Gama Intermedia;
- 3: Gama Alta

Si agrupamos los datos de acuerdo a la nueva variable GAMA y a la variable MARCA obtenemos la siguiente tabla de contingencia.

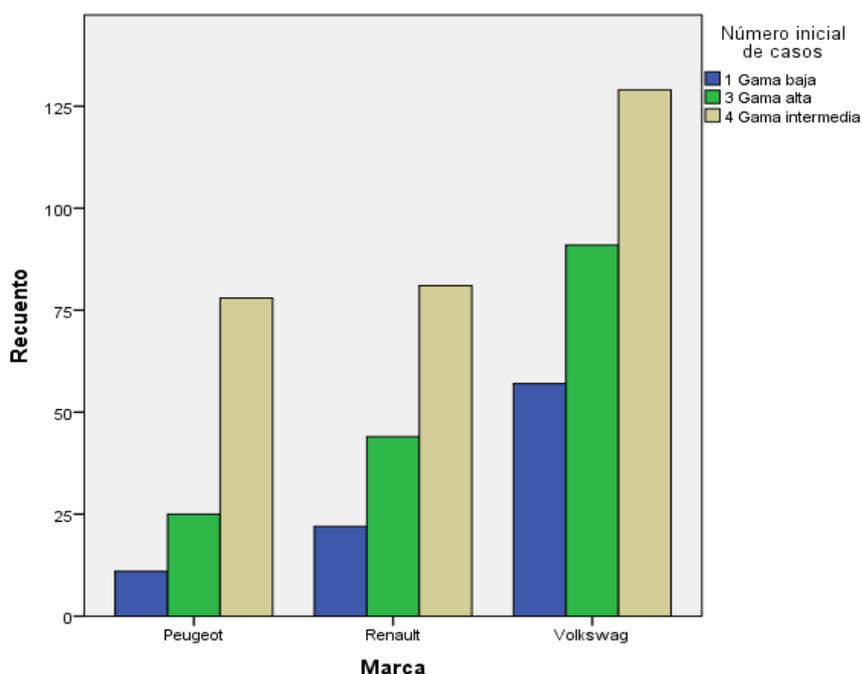
Tabla 51. Contingencia variable Marca * Numero Inicial de Casos

		Número inicial de casos			Total	
		1	3	4		
Marca	Peugeot	Recuento	11	25	78	114
		Frecuencia esperada	19,1	33,9	61	114
	Renault	Recuento	22	44	81	147
		Frecuencia esperada	24,6	43,7	78,7	147
	Volkswagen	Recuento	57	91	129	277
		Frecuencia esperada	46,3	82,4	148,3	277
Total		Recuento	90	160	288	538
		Frecuencia esperada	90	160	288	538

La diferencia entre las marcas es evidente. Peugeot es la marca que menos coches tiene de gama alta y de gama baja de los esperados en un reparto proporcional y Volkswagen la que más tiene.

En la Figura 49. se muestra el gráfico de distribución por Marca con la frecuencia de coches de cada gama , donde se visualizan claramente las diferencias.

Figura 49. Grafico distribución Marca y Gama



8.4.1 Estudio del Precio de los Coches según la Gama

Con el fin de analizar si el precio de los coches tiene una relación diferenciada por gama, se han realizado las regresiones de los coches por gama baja, alta e intermedia.

8.4.1.1 Regresión de Coches Gama Baja

Tabla 52. Resumen del modelo vehículos Gama Baja

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
4	,946 ^d	0,895	0,89	1,35208

a. Variables predictoras: (Constante), caball

d. Variables predictoras: (Constante), caball, Marca_W, longit, velmax

El coeficiente de determinación indica que las variables predictoras explican un 89,5% la variabilidad del precio de los coches de gama baja.

Tabla 53. Coeficientes

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
4 (Constante)	1,178	5,459		0,216	0,830
caball	0,228	0,027	1,000	8,496	0,000
Marca_W	3,443	0,346	0,409	9,963	0,000
longit	0,057	0,008	0,447	7,155	0,000
velmax	-0,166	0,038	-0,446	-4,355	0,000

a. Variable dependiente: PVP_miles

El conjunto de variables predictoras resultantes en el modelo tienen un nivel de significación de 0,000.

El coeficiente de la variable Velmax, como en los modelos de regresión individualizados por marca, tiene signo negativo. Es decir, supone una minoración en el precio ante un aumento de la velocidad máxima.

Ha entrado en la ecuación resultante de este modelo la variable Marca_W (variable dicotómica de los coches de Volkswagen).

8.4.1.2 Regresión de Coches de Gama Alta

Tabla 54. Resumen del modelo vehículos Gama Alta

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
6	,822 ^f	0,676	0,664	3,22487

f. Variables predictoras: (Constante), ancho, Marca_W, Marca_R, caball, longit, velmax

El coeficiente de determinación es menor que en el caso de los coches de gama baja. En este modelo las variables predictoras explican un 67,6% de la variabilidad del precio de los automóviles de gama alta.

En este caso el error de estimación es bastante superior al que tenía el modelo de regresión de gama baja.

Tabla 55. Coeficientes

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
6 (Constante)	-102,733	21,842		-4,703	0,000
Ancho	0,537	0,125	0,332	4,31	0,000
Marca_W	6,296	0,75	0,563	8,393	0,000
Marca_R	3,414	0,822	0,276	4,153	0,000
Caball	0,134	0,021	0,622	6,424	0,000
Longit	0,102	0,017	0,425	6,189	0,000
Velmax	-0,167	0,04	-0,421	-4,174	0,000

a. Variable dependiente: PVP_miles

El conjunto de variables predictoras resultantes en el modelo tienen un nivel de significación de 0,000. Es decir, el modelo recoge parte de la variabilidad del precio.

En este caso han entrado en la ecuación resultante de este modelo tanto la variable Marca_W (variable dicotómica de los coches de Volkswagen) como la variable Marca_R (coches de Renault).

8.4.1.3 Regresión de Coches de Gama Intermedia

La siguiente tabla contiene el resumen del modelo de regresión de los coches de gama intermedia:

Tabla 56. Resumen del modelo vehículos Gama Intermedia

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
6	,866 ^f	0,749	0,744	2,44834

f. Variables predictoras: (Constante), caball, longit, Marca_W, Marca_R, ancho, velmax

El coeficiente de determinación es 0,749. Es decir, la proporción de la variación de la variable dependiente (PVP_miles) es explicada por el modelo de regresión, con el conjunto de variables predictoras, en un 75%.

Tabla 57. Coeficientes

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
6 (Constante)	-51,753	8,591		-6,024	0,000
caball	0,122	0,016	0,61	7,473	0,000
longit	0,065	0,009	0,336	7,108	0,000
Marca_W	5,56	0,403	0,573	13,797	0,000
Marca_R	2,894	0,401	0,269	7,219	0,000
ancho	0,263	0,054	0,251	4,903	0,000
velmax	-0,092	0,028	-0,271	-3,245	0,001

a. Variable dependiente: PVP_miles

El conjunto de variables predictoras resultantes en el modelo tienen un nivel de significación menor o igual a 0,001.

Las variables que han entrado en este modelo de regresión son las mismas que lo hacían en el modelo anterior de coches de gama alta.

8.4.2 Interpretación de Resultados de Regresiones por Gama

El precio de los coches (PVP_miles) es mucho menos variable en la gama baja y más fácil de predecir. No hay diferencias en precio entre las marcas Peugeot y Renault en vehículos de gama baja.

En los vehículos de gama alta afecta el ancho de los vehículos y existen diferencias significativas de precio entre todas las marcas, siendo Volkswagen la más cara y Peugeot la más barata.

Para la gama intermedia, los resultados son similares a los de la gama alta. Y, de igual modo, podemos decir que Renault y especialmente Volkswagen son las marcas que tienen precios más altos a los que ofrece Peugeot.

9 CONCLUSIONES

En este trabajo se han diseñado diferentes modelos para explicar el precio de los coches en términos de diferentes características y marcas. El porcentaje de varianza explicada en cada caso es muy alto.

Las variables introducidas de cada modelo de regresión explican la variación del precio final recomendado de los fabricantes en un 82% para la marca Peugeot, 88% en Renault y casi un 90% en la marca Volkswagen.

Según el modelo de Renault, se deduce que el consumo de combustible es un factor importante en este fabricante para establecer el precio de sus turismos. Las versiones de sus vehículos con menor consumo, normalmente relacionadas con motorizaciones diesel, tienen precios significativamente más altos.

También es relevante en Renault la relación directa que tienen sobre el precio de venta tanto el ancho de los vehículos, relacionado con el modelo del automóvil, como la potencia del motor de las distintas versiones de sus turismos.

A tenor de los resultados del modelo de Peugeot, parece que en los coches de esta marca influyen más a la hora de establecer el precio de sus automóviles la longitud de los mismos, asociado al tipo o modelo de vehículo, y la potencia que consiguen las motorizaciones de sus distintas versiones. También influyen aunque en menor medida otras características técnicas, de manera inversamente proporcional a la variación de las mismas sobre el precio, como son las emisiones de gases contaminantes y la velocidad máxima que alcanzan sus vehículos.

En los turismos de Volkswagen, según su modelo de regresión, las características técnicas que condicionan más el precio de venta son el ancho, la potencia del motor y la velocidad máxima.

Esta última variable, la velocidad máxima, como en el resto de los modelos individuales de regresión de las tres marcas, influye ajustando relativamente el precio.

Los vehículos de todas las marcas tendrían una disminución en el precio de venta al aumentar la velocidad máxima que podrían alcanzar cuando el resto de características técnicas no tuvieran variación. Esto puede deberse al tipo de motorización empleado en los vehículos ya que, generalmente, los vehículos con motor de gasolina poseen mejores prestaciones (mayor velocidad punta) que los vehículos diesel de igual potencia.

El intervalo de confianza sobre previsiones individuales de los tres modelos de regresión obtenidos por cada marca es ancho; oscilan entre 10.000 y 12.000 € debido tanto a los posibles errores de estimación como a la variabilidad del precio de los coches por otras características que no han sido incluidas en el estudio, como el nivel de equipamiento, tipo de tracción, etc.

De los resultados obtenidos en el análisis de regresión conjunto de los datos de las tres marcas, Volkswagen, Renault y Peugeot, se puede extraer la conclusión que, eliminando el efecto de las características técnicas, es claramente el fabricante Peugeot el que ofrece los vehículos más baratos, a los que seguiría la marca Renault. Y son, como cabía esperar en principio, más caros los vehículos de Volkswagen.

Según los modelos resultantes de predicción del precio de los coches segmentados por marca y gama, se deduce que en los coches de gama baja hay un mayor equilibrio de precios entre las marcas Renault y Peugeot, en los que las diferencias de costes finales son poco significativas.

Tanto en los coches de gama intermedia como en los de gama alta el ancho de los vehículos influye especialmente en su precio final recomendado; en cambio, esta variable no es representativa para establecer el precio en los modelos de gama baja.

Tras el estudio de resultados de las regresiones tanto de las tres marcas individuales como de la conjunta, se puede concluir que los modelos obtenidos se ajustan bastante bien a los datos y explican la varianza del precio en un porcentaje muy alto aunque con un intervalo de confianza ancho.

Posiblemente para intentar mejorar los modelos y reducir su intervalo de confianza, se podría realizar una clasificación más detallada del tipo de vehículo y peculiaridades exclusivas para añadir como variables a considerar en el estudio. Por ejemplo, estableciendo diferencias entre los automóviles por tipo de carrocería (coupé, berlina, familiar, etc.) o acabados exteriores (con techo normal, panorámico, descapotables), tipo de cambio (manual, automático), etc.

10 BIBLIOGRAFÍA

Memoria Anual ANFAC 2012.

Recuperado de: <http://www.anfac.com/documents/tmp/memoria2012.zip>

Revista “Altagama Motor” (2013). Num. 122, enero 2013. Ed. Grupo V

Revista “Motor Mundial”.(2013). Num. 770, diciembre 2012 - enero 2013. Ed. La Primera del Motor.

Morales, P. (2012). Correlación y regresión, simple y múltiple. UPM

Recuperado de: <http://www.upcomillas.es/personal/peter/investigacion/Regresion.pdf>

Tejedor, F.J. (1999). Análisis de varianza. Schaum. Madrid: La Muralla S.A.

Peña, D. (2002), Análisis de datos multivariantes, McGraw-Hill.

Jolliffe, I.T. (2002), Principal Component Analysis, Springer Series in Statistics, second edn, Springer.

Del Barrio, M.; Clar M. y Suriñach J. (2002). Modelo de Regresión Lineal Múltiple: especificación, estimación y contraste. Universitat Oberta de Catalunya.

Montes, F. (2004). Estadística Básica

Recuperado de: <http://www.uv.es/montes/biomecanica2004/anova>

Chen X.; Ender P.; Mitchell M. y Wells C. Regression with SPSS. Idre UCLA

Recuperado de: <http://www.ats.ucla.edu/stat/spss/webbooks/reg/default.htm>

Peña D. (2010). Regresión y diseño de experimentos. Ed. Alianza Editorial

Anderson, T.W. (2003). An Introduction to Multivariate Statistical Analysis. Hoboken, NJ: J. Wiley & Sons.

Diccionario Motor Giga. Ed. Red Giga

Disponible en: <http://diccionario.motorgiga.com>

Cuadras C.M. Nuevos métodos de Análisis Multivariante. CMC Editions

Recuperado de: <http://www.ub.edu/stat/personal/cuadras/metodos.pdf>