



Universidad de Jaén

Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado

**ANÁLISIS ESTADÍSTICO
DEL PRECIO DEL
METRO CUADRADO
DE LA VIVIENDA EN
LAS PROVINCIAS ANDALUZAS**

Alumno: Julián Esono Sango

Julio, 2015

AGRADECIMIENTOS

Quiero expresar mi enorme gratitud hacia:

Mi tutor Antonio Conde Sánchez, sin su apoyo incondicional y plena confianza este trabajo fin de grado no podía haberse realizado. Su empeño constante en que me esforzara y le dedicara tiempo han dado como resultado este trabajo.

Agradecer a mi familia, especialmente a mi tío Laureano Sango y a mis padres quienes han sido los principales sustentadores de mis estudios, han apoyados mis decisiones y en momentos difíciles y de equivocación siempre han sabido aconsejarme.

Gracias a mi hermana en estos cuatro años de carrera también he vivido momentos realmente divertidos. Ella me ha enseñado a valorar lo que se tiene en el momento, poco o mucho, siempre podía contar con ella. Gracias Javiera Verónica Esono Sango.

Y gracias a esos compañeros que me han sabido valorar, Sergio García, Miguel Teixeira, Miguel Núñez y Juan Cristóbal, vosotros me habéis ayudado a crecer.

ÍNDICE

ÍNDICE.....	2
RESUMEN	3
1. INTRODUCCIÓN	4
1.1 MOTIVACIÓN.....	6
1.2 ESTRUCTURA DEL TRABAJO	7
2. MARCO TEÓRICO.....	8
2.1 DESCRIPCIÓN DEL SECTOR	8
2.2 BREVE DESCRIPCIÓN DEL ANÁLISIS REGRESIÓN.....	12
2.2.1 ORIGEN	12
2.2.2 CONCEPTO DE LA REGRESIÓN	13
2.2.3 METODOLOGÍA DEL ANÁLISIS DE REGRESIÓN	14
2.3 PREPARACIÓN DE LOS DATOS	17
3. ANÁLISIS ESTADÍSTICO.....	22
3.1 METODOLOGÍA.....	22
3.2 ANÁLISIS EXPLORATORIO.....	23
3.3 AJUSTE DEL MODELO DE REGRESIÓN	24
3.3.1 AJUSTE DEL MODELO DE REGRESIÓN MÚLTIPLE	24
3.3.2 CONTRASTES	40
4. CONCLUSIONES	43
5. BIBLIOGRAFÍA.....	46
6. ANEXO.....	48
6.1 RESULTADOS.....	52

UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado
Estadística y Empresa

RESUMEN

En este trabajo fin de grado se analiza la evolución del precio de los inmuebles de la comunidad autónoma de Andalucía. Para ello se emplean una serie de variables de carácter importante para la explicación del precio del metro cuadrado de la vivienda. Los datos necesarios para realizar este análisis se encuentran en el INE (Instituto Nacional de Estadística), en la página oficial del ministerio de Fomento y en la página del Instituto de Estadística y Cartografía de la Junta de Andalucía. Para el desarrollo del trabajo se ha descrito brevemente las variables empleadas y su principal relación con el tema en cuestión. Como el objetivo principal es saber si existe una variación de precios considerable entre las provincias, se ha realizado primero un análisis descriptivo de esta variable para saber cómo evoluciona a través del tiempo, posteriormente se emplean los demás factores que afectan más o menos a la variable original y se utiliza el método de regresión para determinar el comportamiento y las relaciones de estas variables y el precio del metro cuadrado de la vivienda.

RÉSUMÉ

Dans ce travail de fin de cours on analyse l'évolution du prix des immeubles de la communauté autonome d'Andalucía, pour cela on utilise une série de variables genre important pour l'explication du prix du mètre carré d'une maison. Les éléments nécessaires pour réaliser cet analyse sont publiés sur INE (Institut National de Statistique), sur le site officiel du Ministre de Fomento et sur le site de l'Institut de Statistique et Cartographie du conseil d'Andalucía. Pour le développement du travail on a décrit en résumé les variables employées et la relation principale avec le sujet en question. Vu que l'objectif principal est celui de savoir si il existe une variation des prix remarquable entre les communautés, premièrement on a fait une analyse descriptif de cette variable pour connaître son évolution à travers du temps, après on utilise les autres facteurs qui affectent plus ou moins à la variable originale et on utilise la méthode de régression pour déterminer le comportement et les relations de ces variables et le prix du mètre carré de la maison.

UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado
Estadística y Empresa

1. INTRODUCCIÓN

En este presente trabajo fin de grado se va a tratar de analizar la evolución de los precios del metro cuadrado de las viviendas en la comunidad autónoma de Andalucía. El precio es determinante en el mercado inmobiliario porque afecta de manera directa a las transacciones que se realizan en dicho mercado. Este mercado es de gran interés, pues su elevada participación en el PIB a través de la construcción y la estrecha relación entre la actividad financiera (concesión de préstamos) y la inmobiliaria, lo convierten en un sector importante en la economía española y en la economía de cualquier país.

Antes de enmarcarnos en lo que viene a ser el estudio de regresión con los datos, se va a realizar una descripción breve del sector inmobiliario a fin de ubicarnos en el marco temporal y justificar el motivo del estudio realizado, así como las variables empleadas y el período de tiempo seleccionado.

Si en períodos de burbuja inmobiliaria se registraron subidas de precios excesivas, en la actualidad, aunque han caído los precios de las viviendas, siguen manteniéndose algo elevados. El motivo de este trabajo se centra en analizar el patrón que siguen dichos precios y si guardan o están relacionados con alguna variable o indicador externo. Más concretamente, se trata de observar este cambio en las distintas provincias andaluzas.

Se parte de un objetivo principal, identificar las variables necesarias que afectan de manera directa o indirecta al precio de los inmuebles y a partir de ellas determinar el motivo de las diferencias o discrepancias en el precio de las viviendas entre las distintas provincias andaluzas. El objetivo no es identificar la manera de determinar el precio del metro cuadrado de la vivienda, sino identificar primero las fluctuaciones en el precio del metro cuadrado de la vivienda y explicar de manera detallada los factores que influyen en esas fluctuaciones del precio.

La economía española atraviesa situaciones de desempleo elevadas, PIB per cápita bajos sobre todo en Andalucía y dificultad en obtener créditos bancarios, estos factores y

algunos más pueden ser influyentes en la determinación del rumbo del mercado inmobiliario.

Gran parte de los datos empleados para realizar este análisis se obtienen fundamentalmente de bases de datos importantes como la del Instituto de Estadística y Cartografía de la Junta de Andalucía, en esta base se obtienen datos anuales sobre el precio de la vivienda y otros datos de interés como el PIB per cápita o el salario medio anual en cada una de las distintas provincias andaluzas. A parte de la junta de Andalucía, también se emplea como fuente de información y de obtención de datos el INE (Instituto Nacional de Estadística) donde aparecen publicados datos sobre este sector y datos demográficos de Andalucía como la población anual en cada provincia. Aparte de estas bases de datos también se han empleado páginas de registros económicos para recopilar información sobre tipos de intereses hipotecarios. Estos datos, junto con los obtenidos del ministerio de fomento, que a su vez obtiene los datos del registro de la propiedad donde se recogen las ventas firmadas ante notario entre dos y tres meses antes, se emplean para elaborar el presente análisis.

El período de recogida ha sido anual partiendo desde el año 2000 al año 2014. Se ha considerado este rango de años porque la escasez de datos relacionados con este sector sólo permitía disponer de datos comprendidos en este período de tiempo para todas las variables, por lo tanto, para encontrar la concordancia temporal entre todas las variables del estudio sólo se han tomado valores desde el año 2000 al año 2014.

Todos los datos se usan posteriormente para la formación de las variables relevantes para el estudio en cuestión, usando para la constitución y correcta ordenación de las mismas, programas informáticos como el Excel y el open Office y para el ajuste de regresión el programa estadístico R-commander.

Estas variables se detallan posteriormente en el estudio siendo las principales el precio del metro cuadrado de la vivienda, el año, y la localidad de la misma que se emplea como factor.

A este conjunto de variables cuantitativas y cualitativas, se le suman otras de carácter cuantitativo como el número de compraventas realizadas, el PIB per cápita provincial, la

tasa de paro provincial, el Euribor, la población, el índice de construcción de la vivienda y el salario medio anual en cada provincia. Todas estas son empleadas para localizar alguna relación de dependencia con el precio del metro cuadrado de la vivienda que pueda explicar de manera clara los distintos valores de este que se obtienen cada año en cada provincia. Para realizar un correcto análisis se emplea un modelo de regresión apropiado y se usan las herramientas de regresión que se estimen oportunas y válidas para explicar la adecuada relación entre estas variables y su influencia en la evolución del precio del metro cuadrado de la vivienda.

1.1 MOTIVACIÓN

Los aspectos que me han empujado o que justifican la elección de mi trabajo fin de grado se resumen a continuación:

Mi estancia en la empresa Gestiones y Punto, empresa en la que realicé las prácticas de curriculares, pude contemplar, en concreto, en el departamento de inmobiliaria de esta empresa que el stock de viviendas existente crecía a medida que pasaba el tiempo y es más había viviendas que en años anteriores tenían un determinado valor y que al siguiente se les ponía otro valor generalmente más bajo que el del año anterior. Otro aspecto que me llamó la atención fue que existían viviendas con iguales características en otras zonas de Jaén pero que tomaban precios distintos al que tomaban las de Jaén. Esta primera impresión del mercado inmobiliario en Jaén despertó en mi la curiosidad por saber el comportamiento de este mercado. Por lo que he decidido realizar un estudio retrospectivo basado en datos históricos de este sector en la Comunidad Autónoma de Andalucía poder identificar una relación entre el precio del metro cuadrado de la vivienda y otras variables, ya que esta es una de las metas de la carrera, poder recopilar y analizar de forma correcta series de datos a fin de obtener conclusiones claras y que sean de utilidad para estudios futuros. Basándome en la asignatura de marketing vista en la carrera, he empleado el precio del metro cuadrado de la vivienda como variable respuesta ya que este es en última instancia el que determina la importancia que le concede el consumidor al producto.

UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado
Estadística y Empresa

1.2 ESTRUCTURA DEL TRABAJO

En este trabajo fin de grado se ha recogido una serie de precios del metro cuadrado de la vivienda de las provincias andaluzas desde el año 2000 al 2014. Desde el supuesto de que el precio de la vivienda ha variado mucho estos años, se contempla la mucha o poca diferencia de precios que existe entre las provincias andaluzas, las variables que afectan a este precio del metro cuadrado de la vivienda y que relación guardan con él.

Antes de empezar el estudio se recoge información relacionado con el sector inmobiliario y se distinguen las variables relevantes para este estudio.

En concreto, en el apartado 2 que compone la parte teórica de este trabajo se divide en tres partes. En el punto 2.1 se realiza una breve descripción del sector inmobiliario que recoge, por un lado, información sobre las fechas más relevantes en las que se inició la publicación de información del sector inmobiliario (compraventas, precios de viviendas, índices de construcción, etc.) y por otro alguno de los motivos por los que ha variado tanto el precio del metro cuadrado de la vivienda. En el punto 2.2 se describe brevemente el concepto de regresión para hacerse una idea de esta técnica y en el punto 2.3 se identifican las variables empleadas para el estudio y la recopilación de los datos.

En la sección 3 se llega a la parte del estudio. En el punto 3.1 se describe la metodología empleada para realizar el estudio, en el 3.2 un análisis exploratorio previo con objeto de ver las relaciones más evidentes entre las variables y el punto 3.3 se corresponde con el análisis propiamente dicho.

Una vez concluido el análisis, en la sección 4 se muestran las conclusiones de este trabajo fin de grado.

En las secciones 5 y 6 aparecen la bibliografía empleada y los anexos.

2. MARCO TEÓRICO

2.1 DESCRIPCIÓN DEL SECTOR

El sector inmobiliario influye de manera considerable en la economía de cualquier país tanto en términos de producción con su elevada participación en el PIB a través de la construcción y la generación de empleo, como en términos de valor añadido.

Vergés R. (2012) afirma que “la difusión de información inmobiliaria en España es un hecho relativamente reciente tanto en lo cuantitativo como pueden ser número y características de las operaciones, como en lo cualitativo, es decir valor de las mismas”.

Haciendo un seguimiento de las primeras apariciones documentadas públicamente sobre transacciones de viviendas podemos observar que “no es hasta el año 1960 cuando algunas empresas deciden compartir las actividades realizadas internamente, convirtiéndolas en estadísticas válidas tanto para el público como para sectores relacionados con la actividad inmobiliaria” (Vergés, 2012).

A partir de 2004 el Ministerio de Fomento y el Colegio de Registradores publican trimestralmente y anualmente información sobre transacciones realizadas de viviendas, estos datos son obtenidos de los registradores y de la base Ancert (Agencia Notarial de Certificación) del Notario. En la base del Ministerio de Fomento se publican el total de transacciones inmobiliarias clasificadas por comunidades autónomas, por provincias y por municipios.

❖ **Desagregación territorial por provincias y comunidades autónomas.**

Aparecen trimestralmente el número de compraventas realizadas. En este apartado se encuentran publicadas la siguiente serie de datos:

- Número total de transacciones inmobiliarias.

- Número total de transacciones inmobiliarias de vivienda libre o de vivienda de protección oficial.
- Número total de transacciones inmobiliarias de viviendas nuevas y de viviendas de segunda mano.
- Número total de transacciones inmobiliarias según la residencia del comprador.
- Número total de transacciones inmobiliarias según la tipología del edificio.
- Número total de transacciones inmobiliarias.

Todos estos aspectos pueden cruzarse entre sí dando lugar a una extensa fuente de datos.

❖ **Desagregación territorial por municipios:** En este apartado sólo se alcanza publicar transacciones según el tipo de vivienda (nueva, segunda mano, de protección oficial o libre).

Finalmente, a partir del año 2007 el Instituto Nacional de Estadística difunde también datos sobre las transacciones mensuales de viviendas según comunidad autónoma y provincia y según si se trata de una vivienda de protección oficial, una vivienda libre, una vivienda usada y una vivienda nueva.

Por otro lado, en lo que a valor del metro cuadrado de la vivienda respecta, “no es hasta 1974 cuando la consultora Tecnigrama publica sus informes de coyuntura inmobiliaria, aunque sólo se limitaba a las grandes ciudades, Madrid, Barcelona, Valencia y Sevilla” (Vergés, 2012).

“Desde mediados de los 80, la empresa TINSA una importante sociedad de tasación, recoge datos sobre tasación de las viviendas que se emplean después por el ministerio de fomento desde 1987 para determinar los precios oficiales del metro cuadrado de la vivienda en España”. (García-Montalvo, 2001).

En este punto es importante comentar que la tasación no son los precios finales de las viviendas pero que esta es fundamental para el cálculo posterior de estos, según Pérez y

González (1996) “la tasación tiene por objetivo conocer el valor de realización de una vivienda por lo que aproxima bien el valor de mercado”.

Finalmente, en 2004 en el Ministerio de Fomento se inicia la publicación de las series inmobiliarias de los notarios y los registradores.

En esta página web del Ministerio de Fomento se pueden encontrar publicados datos trimestrales sobre los precios del metro cuadrado de la vivienda libre y de la vivienda de protección oficial por comunidades autónomas y provincias desde el año 1995. En cambio, cuando ya se le añade la componente antigüedad los datos sólo aparecen desde 2013.

El precio de la vivienda en sí, no se puede considerar homogéneo, este valor varía según localidad, estructura, calidad del edificio, tamaño... Al trabajar con los precios del metro cuadrado, las series estadísticas corrigen sólo el factor diferencial del tamaño ofreciendo el valor del precio medio de la vivienda por metro cuadrado.

En el presente trabajo, nos centramos en los precios brutos del mercado por metro cuadrado de vivienda, el hecho de que puedan ser directamente observables los hace más eficientes, aunque cabe destacar que no pueden competir con conceptos más elaborados como los precios hedónicos del INE que tiene como objetivo la medición de la evolución de los precios de compraventa de las viviendas de precio libre, tanto nuevas como de segunda mano a lo largo del tiempo y los precios de ventas repetidas registradas de la ERI (Estadística Registral Inmobiliaria).

En la actualidad el mercado inmobiliario español está atravesando situaciones de recesión que se deben fundamentalmente al efecto de la crisis económica que afectó a la economía española en general. Mientras que entre el año 2004 y 2005 se vendieron un total de 1656411 viviendas de régimen libre aproximadamente, entre 2012, 2013, 2014 y 2015 sólo se han vendido 1342337 viviendas aproximadamente, según los datos del Ministerio de Fomento.

Aunque entre los años 1986 y 1992 se produjo una burbuja inmobiliaria, esta no afectó al volumen de viviendas construidas como la burbuja inmobiliaria iniciada en el año 1998,

(la comprendida entre 1986 y 1992 sólo afectó a los precios de las viviendas). Por tanto en este trabajo se considera que el sector inmobiliario español se encuentra dividido en dos partes fundamentalmente, la época de la burbuja inmobiliaria entre 1998 y 2007 aproximadamente y la posterior a 2007. En la primera parte de este mercado, sus principales componentes, bancos, empresas de construcción e inmobiliarias se vieron cegados por las condiciones favorables del momento persiguieron beneficios rápidos y altos en el corto plazo que dieron como consecuencia la segunda parte de este mercado. En esta segunda parte, se atraviesan situaciones de difícil concesión de créditos, listas amplias de hipotecas impagadas, descensos del precio del metro cuadrado de la vivienda, etc.

Una de las principales causas de la burbuja inmobiliaria fue un impresionante incremento de la demanda de las viviendas que superaban el total de las ofertas de viviendas existentes en este momento. Las ofertas de viviendas no eran lo suficientemente elevadas como para satisfacer el total de demandas, eso trajo consigo un incremento del precio de las viviendas. “Pudiendo registrarse un incremento de entre 159,5 y un 104,1 por 100 en el precio de la vivienda libre entre 1998 y 2005”. (Bernardos, 2009).

Esta elevada demanda de viviendas supuso un crecimiento en la construcción de viviendas, aumentando la inversión en este mercado y creando en última instancia muchos puestos de trabajo. En esas fechas los indicadores más relevantes de la economía se encontraban en auge económico. Crecimiento de la ocupación que provocó un incentivo entre los inmigrantes aumentando la cifra de llegadas a España de estos, incremento de la renta disponible, etc. Esto, por un lado. Por otro, los elevados precios de las viviendas permitían justificar las elevadas financiaciones bancarias que se concedieron en este momento y a un tipo de interés muy reducido.

Una vez atravesado este periodo de increíble incremento de la demanda de viviendas, llega el estallido de la burbuja inmobiliaria caracterizado principalmente por un excesivo número de ofertas de viviendas a precios elevados a las que los ciudadanos ya no podían hacer frente por las increíbles subidas del tipo de interés de los préstamos hipotecarios, entonces llega el período conocido como crisis inmobiliaria.

López, G. M. (2013) afirma que “la crisis de sector inmobiliario no fue la que causó la crisis financiera sino al contrario. Las políticas monetarias, la regulación y supresión del sistema financiero y la actuación de las administraciones públicas en la política fiscal fueron entre otros los motivos del inicio de la crisis financiera”.

La crisis inmobiliaria española se inició en el año 2007 aproximadamente, cuando se restringió la capacidad para encontrar financiación extranjera a las entidades financieras españolas. Aunque en gran medida la crisis financiera extranjera (EEUU) afectó al mercado español, no fue el principal causante del estallido de la burbuja inmobiliaria en España.

Bernardos G. (2009) afirma que “en municipios sin ningún atractivo turístico y con una población inferior a 15000 habitantes, el exceso de oferta es tan elevado y la demanda natural tan reducida que la crisis inmobiliaria está asegurada al menos durante los próximos 15 años”.

2.2 BREVE DESCRIPCIÓN DEL ANÁLISIS REGRESIÓN

2.2.1 ORIGEN

Se puede considerar que el concepto regresión emana de los trabajos de Sir Francis Galton, quien en 1885 realizó un estudio sobre las relaciones de estaturas entre padres y sus hijos. Observó que padres demasiado altos, tenían descendientes bajos y que padres que era más bajos tenían hijos más altos que ellos. Eso dio como resultado su “teoría hacia la mediocridad”, posteriormente empleó ese término en otros trabajos. Aunque este autor fue el primero en acuñar el concepto de regresión, el origen de esta técnica puede situarse años más atrás. En el siglo XVIII donde matemáticos franceses como Laplace realizaron tareas que podían denominarse regresión.

2.2.2 CONCEPTO DE LA REGRESIÓN

El análisis de regresión es una técnica estadística para estudiar y modelar la relación entre variables. Esta técnica posee diversas aplicaciones desde la ingeniería, la física, la economía, la administración, hasta las ciencias biológicas y las sociales. Esta técnica puede ser considerada la más usada en la estadística (Montgomery, Peck y Vining, 2005).

La regresión aparece cuando deseamos estudiar la relación entre dos o más variables, donde se pretende explicar una de ellas en función de las demás (Sánchez Carrión, 2008).

La expresión empleada para representar la relación entre las variables esas variables es la siguiente:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

Ecuación (1)

Donde:

Y es la variable respuesta.

X_1 y X_2 las variables independientes.

β_0 es el término constante.

β_1 y β_2 los coeficientes de regresión de X_1 y X_2 .

e representa el error aleatorio.

i representa la observación i ésima.

2.2.3 TIPOS DE MODELOS DE REGRESIÓN

Existen múltiples criterios de clasificación de los modelos de regresión. Algunos de los más empleados son:

Según la forma de recogida muestral:

- Deterministas: Las variables regresoras toman valores predeterminados.
- Aleatorios: Las variables regresoras toman valores azarosos.

Según el número de variables independientes:

- Simple: Este modelo sólo incluye una variable regresora o independiente y tienen la forma:

$$Y = \beta_0 + \beta_1 X + U$$

Ecuación (2)

- Múltiple: Con más de una variable independiente. Similar al modelo recogido en la ecuación 1.

2.2.4 METODOLOGÍA DEL ANÁLISIS DE REGRESIÓN

La metodología del análisis de regresión comprende los siguientes pasos (Ruiz, 2014):

Debido a que esta metodología se realiza de forma exhaustiva en nuestro estudio, en esta sección sólo se ilustra de manera breve los aspectos importantes del estudio de regresión.

1. Enunciado de la teoría.
2. Especificación del modelo.
3. Obtención de los datos muestrales
4. Estimación y análisis de los parámetros del modelo.
5. Diagnóstico y validación del modelo.
6. Interpretación y empleo del modelo para los fines para los que ha sido creado.

1. ENUNCIADO DE LA TEORÍA

En este punto se define el problema en cuestión objeto de estudio y se determinan las variables importantes del modelo.

2. ESPECIFICACIÓN DEL MODELO

Consiste en determinar la relación que guardan las variables del modelo, elegir de forma adecuada la forma funcional apropiada para el modelo y sus variables.

3. OBTENCIÓN DE LOS DATOS MUESTRALES

Este es un aspecto fundamental en la regresión ya que todo análisis de regresión es tan bueno como lo son los datos empleados.

Se pueden emplear tres métodos fundamentales para recoger los datos del estudio:

- a) **Estudio retrospectivo basado en datos históricos** disponibles de las variables de interés. La ventaja de este es que al estar disponibles los datos el coste se minimiza. Entre sus ventajas más notables está la posible presencia de datos faltantes y la fiabilidad y calidad de los datos, etc.
- b) **Estudio observacional**, que como su nombre indica consiste en observar el área de estudio para conseguir los datos. Existen dos tipos de estudios observacionales:
 - **Confirmatorios.** En este tipo de estudios se pretende confirmar o rechazar hipótesis que estudios previos han realizado. Para ello se trabaja con variables que estudios anteriores han considerado que tienen un efecto significativo sobre la respuesta y se añaden variables nuevas relacionadas con la hipótesis que se desea confirmar. Por ejemplo, si se desea explicar los accidentes en carretera y se sabe por estudios previos que el alcohol y el sexo y se pretende confirmar la hipótesis de si el tipo de vehículo es significativo para los accidentes en carretera.

- **Exploratorios.** En estos estudios no se dispone de estudios previos, el investigador ha de recopilar y determinar las variables que son importantes para la respuesta. Por ejemplo, las ventas de un tipo de producto pueden estar afectados por la población, el país, la renta de los habitantes, etc.
- c) **Experimento diseñado.** Esta forma de recogida de datos sostiene que se pueden manipular favorablemente las variables independientes de acuerdo a una estrategia bien definida asegurando separar los efectos de las variables independientes o factores.

4. ESTIMACIÓN Y ANÁLISIS DE LOS PARÁMETROS DEL MODELO

Una vez recopilado los datos y seleccionadas las variables del estudio, se da paso a lo que viene a ser el ajuste del modelo.

Consiste en emplear las técnicas apropiadas para estimar los parámetros del modelo dependiendo del tipo de relación detectada. El modelo de mínimos cuadrados es el más empleado para problemas en los que la relación entre las variables es lineal. Otros métodos son los modelos de respuesta logarítmica donde la variable respuesta generalmente es categórica.

El programa estadístico R-commander empleado para realizar el estudio recoge todos los distintos modelos de regresión y las formas de estimar los parámetros de dichos modelos.

5. DIAGNOSIS Y VALIDACIÓN DEL MODELO

Una vez definida la función que recoge la relación entre las variables del modelo se procede a la comprobación de conceptos y de las hipótesis básicas que han de cumplirse para que el modelo sea considerado como bueno. Entre los aspectos más importantes a comprobar en la diagnosis del modelo están la normalidad, la homocedasticidad, la linealidad, la aleatoriedad y la autocorrelación. En este punto también es importante

mencionar el gran papel que juegan las observaciones atípicas e influyentes, así como la multicolinealidad.

6. INTERPRETACIÓN Y EMPLEO DEL MODELO PARA LOS FINES PARA LOS QUE HA SIDO CREADO

La última fase del estudio de regresión consiste en interpretar de forma adecuada el modelo final obtenido una vez superados todos los obstáculos encontrados.

2.3 PREPARACIÓN DE LOS DATOS

Los datos se recopilan de distintas fuentes con objeto de ser agrupados, analizados e interpretados según los objetivos del estudio en cuestión.

A. DESCRIPCIÓN DE LAS VARIABLES UTILIZADAS

Los datos extraídos de las distintas fuentes se emplean para formar las variables numéricas del estudio. Estos datos se han extraído considerando que de una manera u otra afectan o influyen en las fluctuaciones del precio del metro cuadrado de la vivienda. Se tiene en cuenta que el precio del metro cuadrado de la vivienda ha variado mucho estos últimos años. Ahora bien, con las variables identificadas, se pretende determinar una relación de dependencia y en caso de que la hubiera explicar el motivo de dicha relación empleando para ello un modelo apropiado de regresión.

Diversos autores han realizado un análisis del precio de la vivienda, ya sea de tipo descriptivo o un análisis más completo, sea cual fuere el tipo de análisis estos autores coinciden en que diversas variables afectan e influyen en la determinación del precio del metro cuadrado de la vivienda.

Pagés Martínez J. y Maza L. A. (2003) consideran que “tanto la renta, la ocupación, el coste de uso de la vivienda, el tipo de interés de los préstamos bancarios, el número de viviendas ofertadas, la rentabilidad de la bolsa, la población, la riqueza financiera, la deuda hipotecaria y el índice de costes reales de la construcción entre otros influyen en la determinación del precio de la vivienda”.

Montalvo García J. (2001) considera que “la demografía, la evolución de la renta per cápita, la capacidad de endeudamiento, el tipo de interés, los costes de construcción y el volumen de viviendas, son algunos factores que determinan el precio de la vivienda”.

Estos autores y algunos más coinciden en varios factores. De esta información y en base a los datos disponibles en las bases de datos de los Ministerios de Fomento, Junta de Andalucía y el INE, se han considerado las siguientes variables para el análisis:

Precio del metro cuadrado: El precio del metro cuadrado de la vivienda se obtiene dividiendo el valor desglosado de la vivienda entre la superficie total construida. Nuestra variable respuesta es el precio real del metro cuadrado de las viviendas tasadas en las distintas provincias andaluzas por las entidades tasadoras que colaboran con el Ministerio de Fomento. Como se indicó arriba estos valores se obtienen de la base de datos de la Junta de Andalucía.

Número de transacciones: Cantidad total de viviendas vendidas (compraventas) en cada provincia andaluza anualmente.

El número total de las transacciones queda dividido en transacciones totales de vivienda libre, transacciones totales de vivienda libre nueva y transacciones totales de vivienda libre de segunda mano. Hemos empleado el total de transacciones de viviendas libre, ya que incluyen las de primera y segunda mano.

Tasa de Paro: Número de personas con edad para trabajar que no encuentran un puesto de trabajo. Estos valores sacados del INE también se han extraído para las provincias andaluzas en nuestro período de estudio.

Euribor: Los préstamos bancarios a los ciudadanos para la adquisición de las viviendas, conocidos fundamentalmente como hipotecas, se conceden en base a un tipo de interés, entre estos, el Euribor, es el más utilizado. Desde la página web www.euribordiarario.com se publican todos los valores que ha alcanzado este tipo de interés hipotecario cada día del mes. A partir de esos valores se ha hallado un promedio del valor anual del Euribor en cada año del estudio desde el 2000 al 2014.

Salario medio: Se ha considerado también el salario medio anual de los andaluces, puesto que afecta de manera directamente proporcional a la capacidad del ciudadano para hacer frente a las deudas.

PIB per cápita: La riqueza personal en Andalucía ha sido una de las más bajas de todo el territorio español durante años, según estadísticas del INE, siendo incluso el PIB per cápita en 2014 de 16577 euros, un 26% por debajo del registro nacional situado en 22412 euros. Se han tomado los valores anuales del PIB per cápita para cada provincia andaluza para el estudio.

Población: Esta variable demográfica se considera fundamentalmente por ser una indicadora directa del crecimiento demográfico de una sociedad, se han obtenido el total de población mayor de 16 años, edad a partir de la cual se considera a un ciudadano como activo, o sea capacitado para ejercer un trabajo. La fuente de estos datos ha sido La Junta de Andalucía, donde aparecen publicados anualmente la población total de cada una de las provincias andaluzas.

Costes de construcción: Para concluir esta primera parte de descripción de variables utilizadas para este trabajo, mencionar que se ha tenido en cuenta un índice de coste de la construcción elaborado por el Ministerio de Fomento que tiene como objetivo medir la evolución económicamente del sector de la construcción.

Ámbito temporal: El período de tiempo que se emplea para realizar el estudio. En este caso son 15 años del 2000 al 2014.

Provincias: Las ocho provincias que integran todo el territorio andaluz. Almería, Cádiz, Córdoba, Granada, Huelva, Jaén, Málaga y Sevilla.

B. RECOPIACIÓN DE LOS DATOS

Los datos para llevar a cabo el análisis recogen información del sector inmobiliario en la comunidad autónoma de Andalucía. La forma principal de recogida de estos datos ha sido estratificada en función de la provincia.

El período de recogida ha sido anual, contando con datos desde el año 2000 al 2014, un total de 15 años (incluyendo el año 2014) para estudiar las relaciones existentes entre las variables empleadas en el análisis en las fluctuaciones del precio del metro cuadrado de la vivienda en las distintas provincias andaluzas.

Los datos necesarios para realizar este análisis se han recogido de fuentes distintas, siendo estas las principales fuentes:

Del **INE (Instituto Nacional de Estadística)** que a su vez emplea como fuente algunos ministerios y bases de datos sobre viviendas escrituradas que proporciona el Consejo General del Notariado, se han extraído los datos sobre la tasa de paro trimestral en cada una de las provincias andaluzas en particular y la tasa de paro total de la

comunidad autónoma en general, esa tasa de paro también se encuentra publicada en función del sexo del individuo y de la edad del mismo, pero en este estudio carece de relevancia incluir estas características. Posteriormente se han transformado los datos trimestrales obtenidos del año 2000 al 2014 en anuales para obtener una serie de datos en sinergia anual.

De la base de datos **de la Junta de Andalucía** “Instituto de Estadística y Cartografía de Andalucía” se han extraído los precios del metro cuadrado de la vivienda libre en las distintas provincias, estos precios aparecen publicados también en el Ministerio de Fomento y a diferencia del Ministerio de Fomento que publica los precios del metro cuadrado trimestrales, La Junta de Andalucía ofrece sólo datos anuales. De esta base de datos también se ha extraído el PIB por habitante en cada provincia y el salario medio anual.

Del **Ministerio de Fomento** se han extraído el número de transacciones (compraventas) inmobiliarias de viviendas realizadas trimestralmente en todo el período de estudio en las provincias andaluzas. Sólo se han considerado el total de transacciones inmobiliarias de vivienda libre y no las de protección oficial, ya que el precio de estas últimas lo determina generalmente el estado. Puesto que necesitamos datos anuales, los datos obtenidos trimestralmente del Ministerio de Fomento se han empleado para calcular el total de compraventas anuales realizadas en las provincias andaluzas. De esta misma base de datos se han extraído los costes de la construcción en este caso de Andalucía en general. El Ministerio de Fomento obtiene toda la información sobre tasaciones y compraventas de organismos de tasación importantes como TINSA.

Por último, del **Euribor Diario** se han obtenido los valores medios anuales en los que ha variado el Euribor durante los quince años del estudio.

Una vez recogidos todos los datos, se ha empleado el Microsoft Excel para dar el formato apropiado a todos los datos en este caso transformar una serie de datos trimestrales en anuales o datos diarios en anuales y generar las variables necesarias para realizar el estudio.

3. ANÁLISIS ESTADÍSTICO

En este apartado que comprende el análisis propiamente dicho, se va a emplear el modelo de regresión antes comentado para explicar las relaciones entre el precio del metro cuadrado y las variables consideradas como independientes.

3.1 METODOLOGÍA

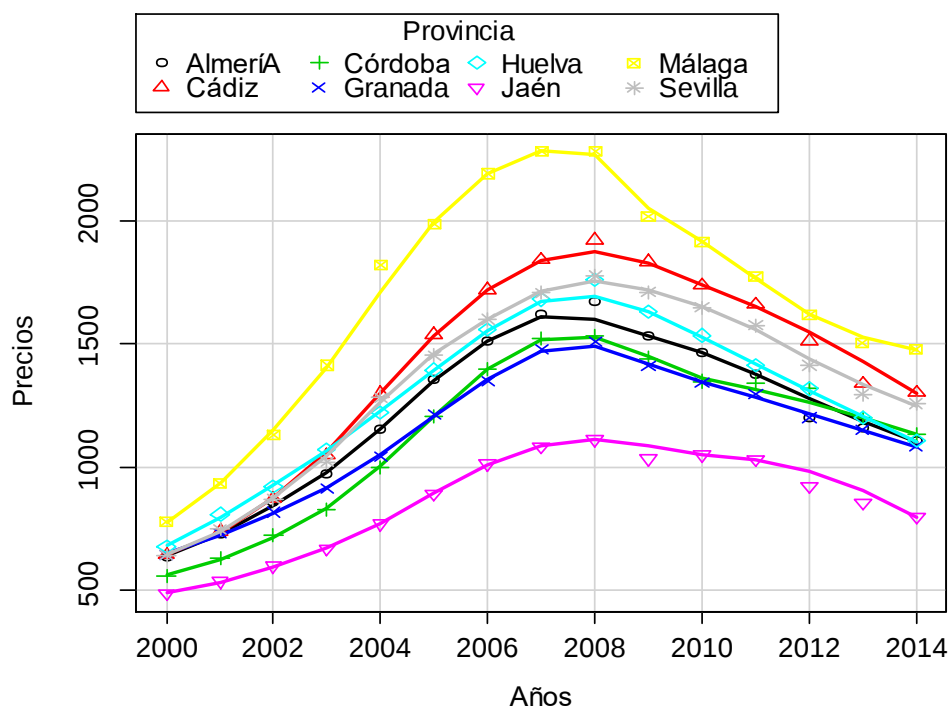
Para realizar el estudio partimos de nuestros datos recogidos en un fichero Excel. Este está compuesto por 120 observaciones, debemos tener en cuenta que se trata de una observación por año y que esas 120 observaciones se dividen 15 observaciones por provincia ya que son ocho el total de provincias andaluzas.

Se realiza un análisis previo con los gráficos apropiados para hacerse una idea de los datos.

Tras ese previo análisis se realiza un ajuste de regresión múltiple de nuestras variables que incluirá unos análisis de multicolinealidad e influencia para conocer posibles problemas que guardan los datos tras el primer modelo planteado y que puedan afectar a la validez del modelo, así como una diagnosis del mismo esta parte concluirá con la presentación del modelo definitivo que cumpla todos los requisitos.

Para finalizar el análisis realizaremos contrastes de regresión a modo de observar si se comportan de igual manera los resultados por provincias.

3.2 ANÁLISIS EXPLORATORIO



Gráfica 1. Diagrama de dispersión

Fuente. Elaboración propia

El diagrama de dispersión que aparece arriba muestra la correlación existente entre las variables precio del metro cuadrado y el Año entre las distintas provincias andaluzas. Se aprecia como la evolución del precio del metro cuadrado es similar en las distintas provincias. En el año 2000 se empieza con incrementos de precios del metro cuadrado de la vivienda hasta alcanzar el pico en el año 2007 donde se empieza a notar un claro decrecimiento de estos valores. Este decrecimiento probablemente se deba al estallido de la burbuja inmobiliaria (inicio de la crisis en España). Eso sí, Málaga es con diferencia la provincia que mayores precios del metro cuadrado de la vivienda ha registrado y Jaén la provincia con precios más bajos registrados.

3.3 AJUSTE DEL MODELO DE REGRESIÓN

Realizado nuestro análisis exploratorio de los datos, vamos a tratar de confirmar las posibles conclusiones antes obtenidas en la sección anterior con un modelo de regresión múltiple.

3.3.1 AJUSTE DEL MODELO DE REGRESIÓN MÚLTIPLE

Residuals:					
Min	1Q	Median	3Q	Max	
-200.658	-62.955	2.957	63.046	268.693	
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.692e+05	4.777e+04	5.635	1.51e-07	***
Provincia[T.Cádiz]	-4.340e+02	1.905e+02	-2.278	0.024776	*
Provincia[T.Córdoba]	-8.274e+01	7.033e+01	-1.176	0.242116	
Provincia[T.Granada]	-3.163e+02	1.057e+02	-2.991	0.003472	**
Provincia[T.Huelva]	3.427e+02	6.392e+01	5.362	5.00e-07	***
Provincia[T.Jaén]	-1.555e+02	6.088e+01	-2.554	0.012105	*
Provincia[T.Málaga]	-6.564e+02	3.148e+02	-2.085	0.039540	*
Provincia[T.Sevilla]	-1.314e+03	3.850e+02	-3.411	0.000921	***
Salario.medio	7.375e-02	2.281e-02	3.233	0.001640	**
Tasa.de.paro	-8.595e+00	3.349e+00	-2.566	0.011703	*
Población	1.284e+00	3.754e-01	3.420	0.000895	***
PIB.Per.cápita	2.875e-02	1.826e-02	1.574	0.118514	
Euribor	-1.694e+04	3.115e+03	-5.438	3.59e-07	***
costes.cos	5.515e+01	9.801e+00	5.627	1.56e-07	***
compraventas	8.130e-03	4.332e-03	1.877	0.063372	.
Año	-1.369e+02	2.416e+01	-5.668	1.30e-07	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 96.11 on 104 degrees of freedom					
Multiple R-squared: 0.9501, Adjusted R-squared: 0.9429					
F-statistic: 132 on 15 and 104 DF, p-value: < 2.2e-16					

Tabla 1. Resultados del modelo de regresión con todas las variables.

Fuente. Elaboración propia

Este primer ajuste nos revela varias cosas. Primero, que los precios de los metros cuadrados de las provincias difieren significativamente de la provincia de Almería excepto la de Córdoba que en principio no parece tener diferencias significativas con la provincia de Almería. Y segundo, que las variables PIB per cápita y Compraventas son las únicas con un p-valor superior al valor de alfa (0,05), lo que las convierte en no significativas para explicar la variabilidad del precio del metro cuadrado de la vivienda.

Es un hecho que la multicolinealidad afecta en gran medida a la utilidad del modelo por eso a continuación se va a estudiar entre otros aspectos la multicolinealidad y posibles observaciones atípicas e influyentes del modelo obtenido para medir la sensibilidad del mismo.

Matriz de correlaciones

	Año	compraventas	costes.cos	Euribor	PIB.Per.cápita	Población	Salario.medio	Tasa.de.paro
Año	1.0000000	-0.45293231	0.9101330	-0.72230916	0.616058594	0.11868614	0.68064369	0.687107601
compraventas	-0.4529323	1.00000000	-0.3651903	0.39004573	0.035886095	0.55907535	0.06418137	-0.538975188
costes.cos	0.9101330	-0.36519027	1.0000000	-0.39320512	0.735662397	0.11693082	0.76788878	0.464613118
Euribor	-0.7223092	0.39004573	-0.3932051	1.00000000	-0.242411869	-0.07638229	-0.31992967	-0.684618851
PIB.Per.cápita	0.6160586	0.03588609	0.7356624	-0.24241187	1.000000000	0.19477804	0.70893828	-0.002869468
Población	0.1186861	0.55907535	0.1169308	-0.07638229	0.194778042	1.00000000	0.56601941	0.085905935
Salario.medio	0.6806437	0.06418137	0.7678888	-0.31992967	0.708938278	0.56601941	1.00000000	0.270125130
Tasa.de.paro	0.6871076	-0.53897519	0.4646131	-0.68461885	-0.002869468	0.08590593	0.27012513	1.000000000

Tabla 2. Matriz de correlaciones.

Fuente. Elaboración propia

Examinando los elementos no diagonales de esta matriz podemos observar que sólo un par de variables tiene una correlación por encima de 0,9. En concreto existe colinealidad entre las variables Año y Costes de Construcción, con una correlación de 0,910133. Esto indica de primeras que la multicolinealidad quizás no sea un problema, aunque es preciso apuntar que cuando más de dos regresores están involucrados en una dependencia lineal la matriz de correlaciones no recoge bien las correlaciones por pares.

Año	compraventas	costes.cos	Euribor	PIB.Per.cápita	Población	Salario.medio	Tasa.de.paro
124.524078	3.647005	89.180364	21.333113	4.644336	3.835806	7.590618	5.269383

Tabla 3. Factores de Inflación de la Varianza.

Fuente. Elaboración propia

Estos valores son indicios de multicolinealidad siendo la variable año la que mayor VIF tiene seguida de los Costes de construcción.

Una solución al problema de multicolinealidad es la reformulación del modelo. Una vez reformulado el modelo si sigue habiendo problemas con la multicolinealidad es mejor una adquisición adicional de datos, pero como no disponemos de más datos como ya se comentó anteriormente se va a trabajar teniendo en cuenta este problema de multicolinealidad. Más adelante se ha probado a eliminar las variables que más correlación tienen con el resto en este caso la variable Año y Coste de construcción y al eliminarlas se ha comprobado que la multicolinealidad se corrige, pero dada la significación de estas y su importancia para el estudio, los resultados se interpretan con ambas variables en el modelo. En la sección de anexos se muestran los resultados con las variables Año y Costes de construcción fuera del modelo.

Atipicidad

No Studentized residuals with Bonferonni p < 0.05			
Largest rstudent :			
	rstudent	unadjusted p-value	Bonferonni p
95	3.077629	0.0026737	0.32085

Tabla 4. Test de valores atípicos de Bonferonni.

Fuente. Elaboración propia.

El test de Bonferonni muestra la observación 95 como atípica.

Esta observación también se detecta en el gráfico de observaciones influyentes como observación influyente. Es normal en este tipo de estudio detectar alguna observación atípica, puede deberse quizás a que de un año a otro cambien mucho las cosas o tal vez a que los valores se vean afectados por la crisis. Es necesario comentar que, aunque el test de Bonferonni la detecte como observación atípica, el leverage no supera el valor de $2(k+1) / n = 0,17$ esta observación no sería considerada atípica en los regresores. Pero al comparar el valor de los residuos de esta observación con los demás valores se observa que son un poco elevados por lo que podría ser atípica en la respuesta.

Algunos resultados se muestran a continuación:

78	79	80	81	82	83	84	85	86	87	88
0.11937297	0.10562897	0.09410198	0.11182924	0.09380602	0.11610579	0.13531827	0.10594678	0.09531019	0.08789578	0.14949408
89	90	91	92	93	94	95	96	97	98	99
0.20828060	0.18434043	0.26042097	0.21555876	0.16077167	0.14067576	0.10755737	0.12693840	0.11403006	0.18977928	0.17011925
100	101	102	103	104	105	106	107	108	109	110
0.15591684	0.13552364	0.15951513	0.15275582	0.15656802	0.24009513	0.18050743	0.12486958	0.09205662	0.09442813	0.08781315
111	112	113	114	115	116	117	118	119	120	
0.12387452	0.11001209	0.19851371	0.12706109	0.12359222	0.11113713	0.10733958	0.09427219	0.10731040	0.17707217	

Tabla 5. Resultados de los leverages.

71	72	73	74	75	76	77	78	79	80
0.92853512	0.09742570	-0.49103887	-0.38147293	-0.50747300	2.08165412	1.17842740	1.15812109	0.05413774	0.01637813
81	82	83	84	85	86	87	88	89	90
-0.88827834	-0.81085129	-1.27649187	-2.12595600	-1.21115771	-0.47540825	-0.02673398	0.33427427	1.00136458	1.15986955
91	92	93	94	95	96	97	98	99	100
-2.00065847	-2.05223348	-1.03763386	0.13339216	2.95945092	1.59564254	2.53438428	2.13051696	1.90433983	0.68766268
101	102	103	104	105	106	107	108	109	110
0.06640137	-0.76168523	-2.09912185	-2.27339521	-2.21959306	1.10287220	0.69084750	0.48463460	-0.02711947	1.09236648
111	112	113	114	115	116	117	118	119	120
0.17528854	0.04249692	-0.53582134	-0.72913406	0.26199876	0.04755590	-0.08796485	-1.31856235	-1.00904323	-0.18291695

Tabla 6. Resultados de los residuos estandarizados.

Fuente. Elaboración propia

UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado
Estadística y Empresa

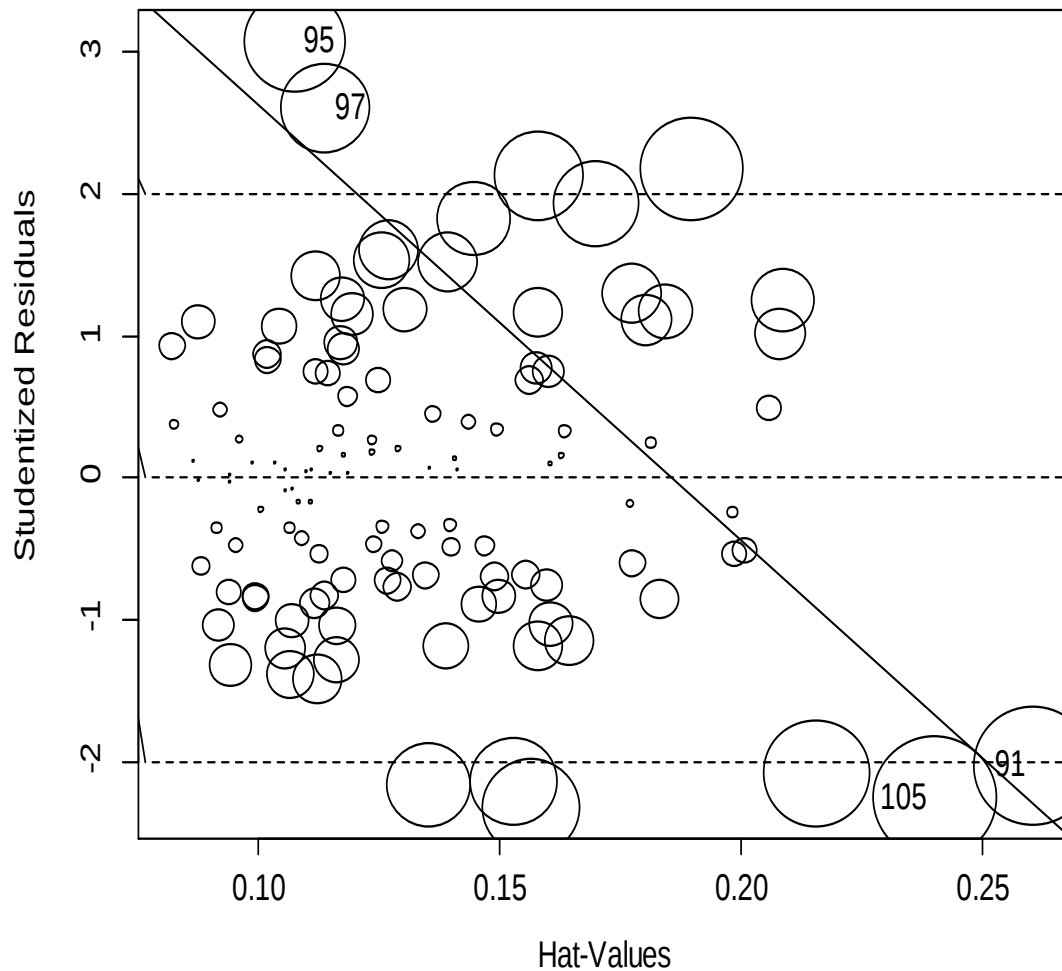


Figura 1. Gráfico de influencia.

Fuente. Elaboración propia

En este gráfico de influencia la observación 95 que antes se ha detectado como atípica con el test de Bonferonni, junto con 91, la 97 y la 105 aparecen como

observaciones influyentes, ninguna de ellas es influyente en los regresores al no superar el límite vertical.

De todas ellas la más influyente es la observación 105 al tener un diámetro mayor al de los demás, seguida de la 95 y la 91.

Comentar que su influencia no se debe a productos de error, esta se debe a los valores que tomó el precio del metro cuadrado para estas fechas. Estos valores varían mucho con respecto al resto, sobretodo la 105 que corresponde al año 2014 de la provincia de Málaga. En los datos se puede apreciar que toma un valor relativamente bajo en comparación con los años anteriores.

Realizados los análisis de influencia y multicolinealidad se da paso al estudio de la diagnosis de nuestro modelo. Comprobar que las hipótesis de normalidad, linealidad, homogeneidad de varianzas e autocorrelación se cumplan.

Diagnosis

Homocedasticidad

```
studentized Breusch-Pagan test

data: Precio.m2 ~ Provincia + Salario.medio + Tasa.de.paro + Población + PIB.Per.cápita + Euribor +
+ costes.cos + compraventas + Año

BP = 1.3087, df = 1, p-value = 0.2526
```

Tabla 7. Test de White de homocedasticidad.

Fuente. Elaboración propia

El test de White sobre valores ajustados no indica evidencias para rechazar la homocedasticidad de las varianzas al ser el p-valor superior al valor de alfa (0,05).

Normalidad

```
Title:
Shapiro - Wilk Normality Test

Test Results:
STATISTIC:
W: 0.9904
P VALUE:
0.5741
```

Tabla 8. Test de normalidad de Shapiro Wilk

Fuente. Elaboración propia

El test de normalidad de Shapiro Wilk no indica evidencias para rechazar la normalidad del modelo con un p-valor de 0.5741.

Linealidad

```
RESET test

data: Precio.m2 ~ Provincia + Salario.medio + Tasa.de.paro + Población +
+ PIB.Per.cápita + Euribor + costes.cos + compraventas + Año

RESET = 11.574, df1 = 8, df2 = 96, p-value = 2.204e-11
```

Tabla 9. Test reset de linealidad.

Fuente. Elaboración propia

El p-valor obtenido al realizar el test reset indica problemas con la linealidad del modelo, un modelo lineal no es el adecuado para explicar las relaciones entre nuestra variable respuesta y las variables regresoras. Esta no linealidad detectada alienta a transformar la variable respuesta, pues en casos de ausencia de linealidad es preferible primero transformar la variable respuesta a añadir términos cuadráticos o interactivos.

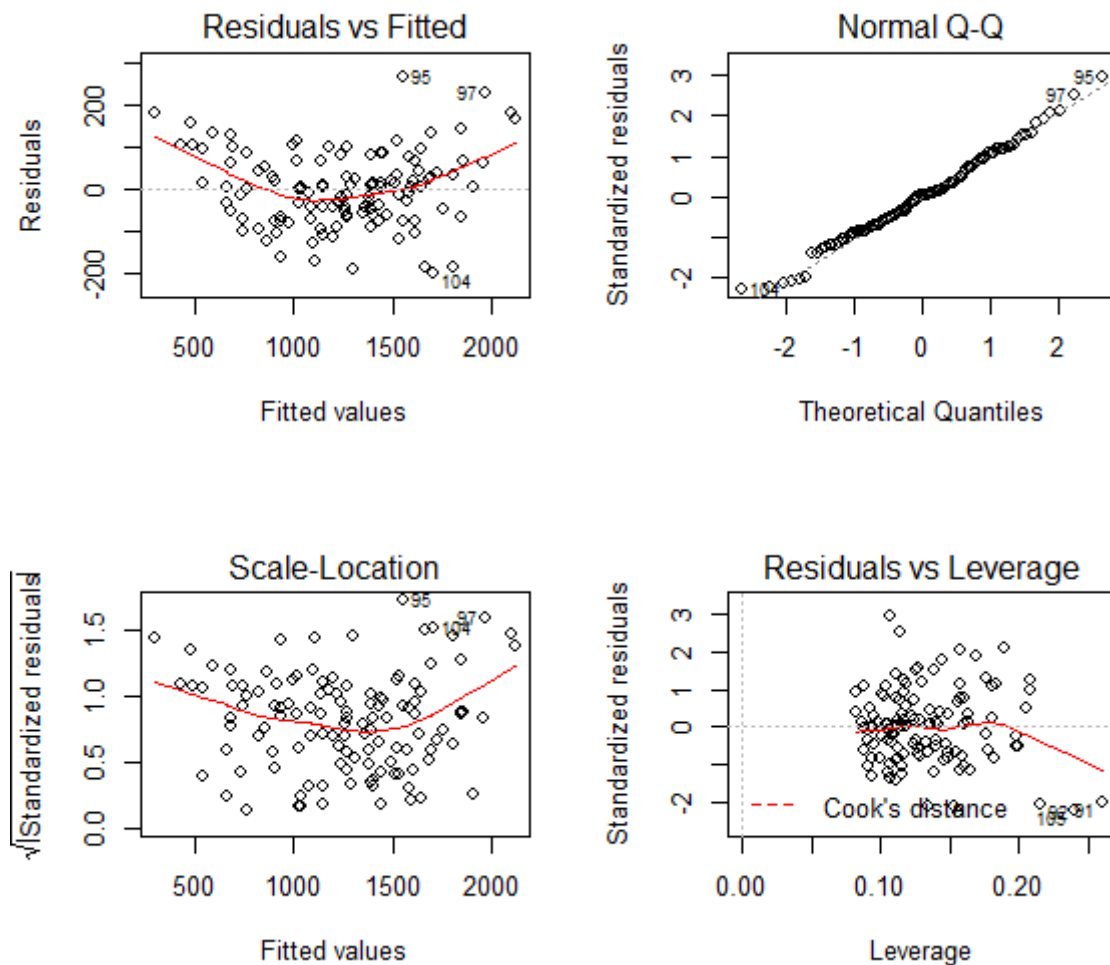


Figura 2. Gráficas básicas de diagnóstico

Fuente. Elaboración propia

El primer gráfico confirma la ausencia de linealidad antes detectada con el test reset, las hipótesis de normalidad y heterocedasticidad no parecen ser violadas según los gráficos dos y tres, y el gráfico cuatro de residuos nos muestra algunas observaciones influyentes antes detectadas, aunque parece que sólo la observación 105 supera la distancia de Cook por lo que en principio es la única observación que podría considerarse influyente.

Como se han detectado problemas con la linealidad del modelo en principio es preferible transformar la variable respuesta. Para esta transformación empleamos la herramienta Box-Cox.

Transformación

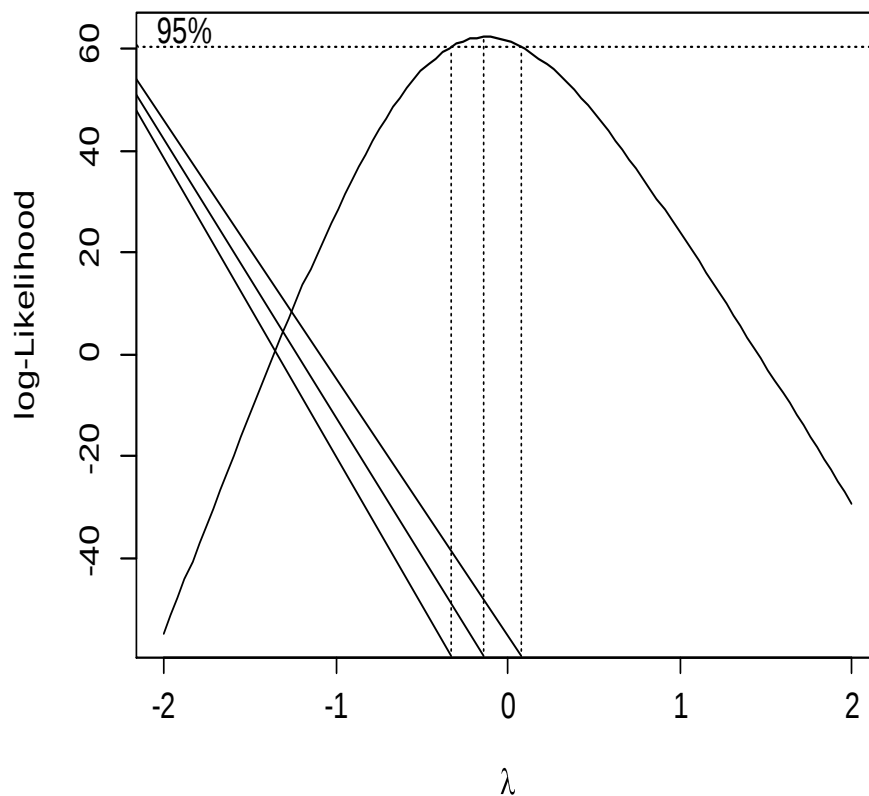


Figura 3. Intervalo de confianza para el gráfico landa de la transformación Box-Cox.

Fuente. Elaboración propia

.

Este gráfico nos muestra que es mejor una transformación logarítmica ya que indica un valor de landa aproximadamente igual al cero.

Una vez re especificado el modelo observamos que sigue habiendo problemas con la multicolinealidad (resultado no mostrado)

También se observa que la transformación ha mejorado la especificación del modelo, Se mantienen la normalidad de los residuos y la igualdad de varianzas, pero la autocorrelación sigue siendo significativa. Los P-valores que ofrecen los test de diagnosis son los mostrados en la tabla.

Test	P-valor
RESET test	0.05902
Test de White	0.06179
Shapiro - Wilk Normality Test	0.4873

Tabla 10. Test de diagnosis.

Fuente. Elaboración propia

En Las gráficas de diagnóstico (salida no mostrada) también se aprecia esta corrección de la linealidad.

Lo más llamativo en el modelo obtenido con la transformación de la variable respuesta (salida no mostrada) es el cambio en la significación de la variable Población que en el primer modelo sí que era muy significativa y en el segundo ya no.

Para solucionar este problema probamos a transformar logarítmicamente la variable Población.

Resultados del modelo con las transformaciones logarítmicas de Población.

P = Transformación logarítmica de Población

Y = Transformación logarítmica del Precio.

Residuals:					
Min	1Q	Median	3Q	Max	
-0.099241	-0.037726	-0.001692	0.035102	0.186305	
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.991e+02	2.750e+01	7.242	7.97e-11	***
Provincia[T.Cádiz]	-4.283e-01	1.304e-01	-3.285	0.00139	**
Provincia[T.Córdoba]	-1.220e-01	5.140e-02	-2.373	0.01949	*
Provincia[T.Granada]	-3.178e-01	7.516e-02	-4.229	5.06e-05	***
Provincia[T.Huelva]	3.825e-01	5.451e-02	7.017	2.38e-10	***
Provincia[T.Jaén]	-1.761e-01	3.478e-02	-5.064	1.79e-06	***
Provincia[T.Málaga]	-5.983e-01	1.845e-01	-3.243	0.00159	**
Provincia[T.Sevilla]	-9.836e-01	2.180e-01	-4.513	1.69e-05	***
Salario.medio	6.132e-05	1.291e-05	4.748	6.59e-06	***
Tasa.de.paro	-1.071e-02	1.884e-03	-5.683	1.22e-07	***
P	8.780e-01	2.005e-01	4.378	2.86e-05	***
PIB.Per.cápita	2.928e-05	1.029e-05	2.845	0.00535	**
Euribor	-1.575e+01	1.810e+00	-8.699	5.35e-14	***
costes.cos	4.196e-02	5.729e-03	7.324	5.32e-11	***
compraventas	5.800e-06	2.167e-06	2.677	0.00864	**
Año	-1.008e-01	1.379e-02	-7.307	5.77e-11	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 0.0544 on 104 degrees of freedom					
Multiple R-squared: 0.9782, Adjusted R-squared: 0.975					
F-statistic: 310.4 on 15 and 104 DF, p-value: < 2.2e-16					

Tabla 11. Resultados del modelo de regresión.

Fuente. Elaboración propia

Evidentemente al transformar la variable Población en este modelo se observa que vuelve a ser significativa para explicar la variabilidad de la respuesta Precio del metro cuadrado de la vivienda.

Veamos cómo se comporta ahora en la diagnosis.

Año	compraventas	costes.cos	Euribor	P	PIB.Per.cápita	Salario.medio	Tasa.de.paro
123.949872	4.060897	88.104331	21.169467	4.993296	5.063509	9.142497	5.240901

Tabla 12. Factores de inflación de la varianza.

Fuente. Elaboración propia

Estos valores de inflación de la varianza nos muestran que la multicolinealidad sigue presente en el modelo.

Diagnosis

Homocedasticidad

studentized Breusch-Pagan test
data: Y ~ Provincia + Salario.medio + Tasa.de.paro + P + PIB.Per.cápita + + Euribor + costes.cos + compraventas + Año
BP = 2.3791, df = 1, p-value = 0.123

Tabla 13. Test de homocedasticidad de White.

Fuente. Elaboración propia

El test de Breusch – Pagan sobre valores ajustados no indica evidencias para rechazar la homocedasticidad de las varianzas con un p-valor de 0.123 superior al valor de alfa (0,05).

Normalidad

```
Title:
Shapiro - Wilk Normality Test

Test Results:
STATISTIC:
W: 0.9807
P VALUE:
0.08238
```

Tabla 14. Test de normalidad de Shapiro Wilk.

Fuente. Elaboración propia

El test de normalidad de Shapiro Wilk no indica evidencias para rechazar la normalidad del modelo con un p-valor de 0.08238.

Linealidad

```
RESET test

data: Y ~ Provincia + Salario.medio + Tasa.de.paro + P + PIB.Per.cápita +
+ Euribor + costes.cos + compraventas + Año

RESET = 0.86024, df1 = 2, df2 = 102, p-value = 0.4261
```

Tabla 15. Test Reset de Linealidad.

Fuente. Elaboración propia

La linealidad del modelo no se ve comprometida.

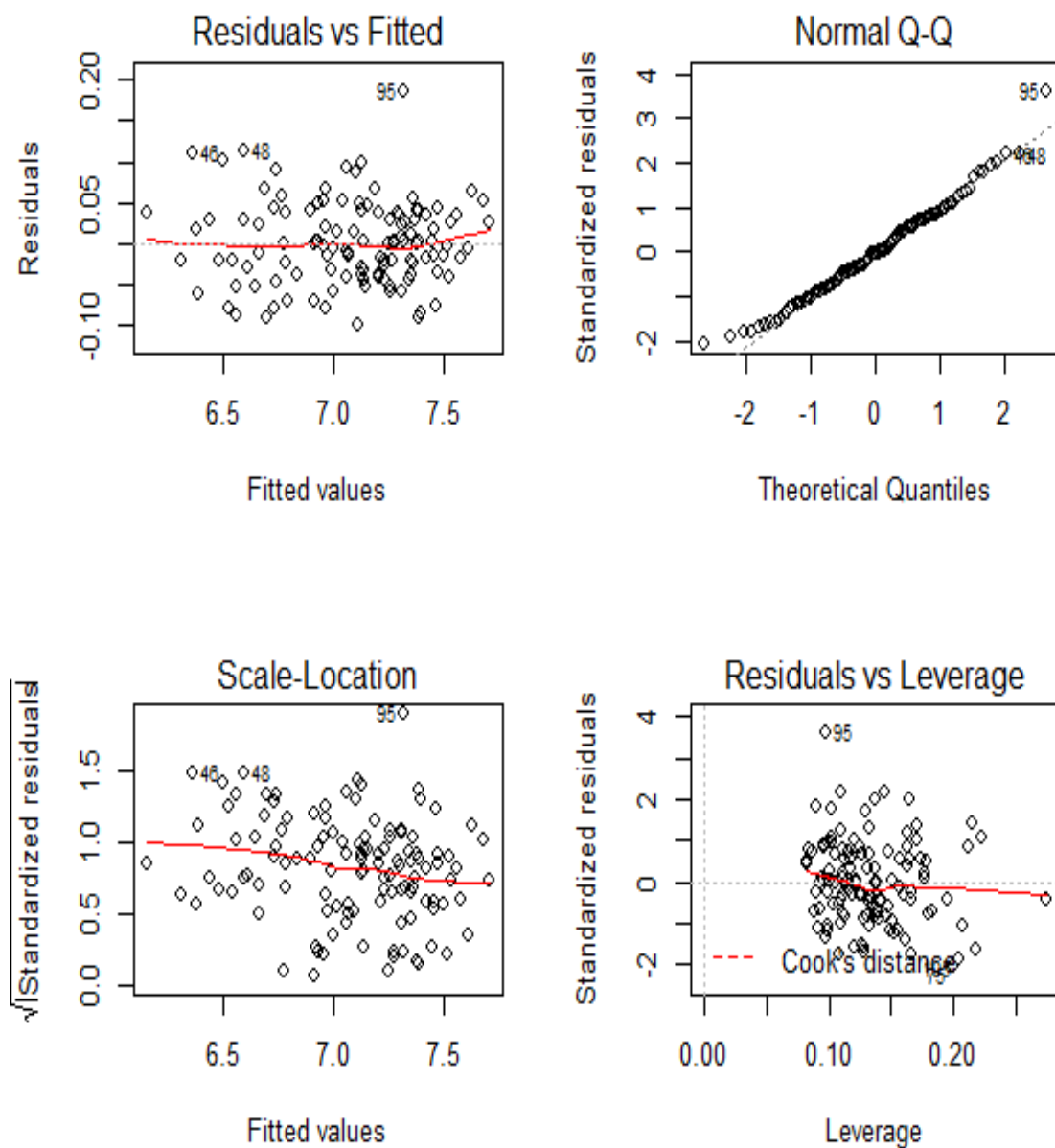


Figura 4. Gráficas básicas de diagnóstico

Fuente. Elaboración propia

El diagrama de dispersión que antes se mostraba nos evidenciaba que la relación entre el precio del metro cuadrado de la vivienda y el año podría ser cuadrática, hemos considerado la posibilidad de introducir este término en el modelo (salida no mostrada).

Al realizar tal ajuste nos modifica las condiciones favorables del modelo anterior, la linealidad en este caso supone un problema y se incrementan los factores de inflación de la varianza. Por otro lado, se ha introducido términos interactivos en el modelo (salida no mostrada) y se rechazaban las iteraciones. Por lo tanto, nuestro modelo definitivo va a ser el anteriormente mostrado.

Como hemos estado trabajando con la provincia de Almería como la categoría de referencia, nuestro modelo definitivo indica que el precio del metro cuadrado de todas ellas difiere significativamente del precio del metro cuadrado de Almería.

Todas las variables utilizadas en el estudio guardan una relación con el precio del metro cuadrado al ser estas significativas para explicar su variabilidad, aquellas que poseen un signo negativo indican que a medida que aumente su valor, decrece el de la variable respuesta y las que tengan un signo positivo, todo lo contrario. Pese a los problemas de multicolinealidad detectados vamos a proceder a la interpretación de los coeficientes, más adelante en la sección de anexos donde se corrige la multicolinealidad eliminando del modelo las variables que más problemas de multicolinealidad aportan al modelo, se observa que no se diferencian mucho los coeficientes de este modelo con los del modelo de la sección anexos, por lo que su interpretación será similar.

La interpretación de los coeficientes de regresión se haría de la siguiente manera:

Interpretación

Factores

El modelo de regresión sostiene que en condiciones iguales y dadas las variables del modelo, la provincia con precios del metro cuadrado más elevados es la de Cádiz con un 34,83% por encima de Almería. Las demás provincias tienen valores del precio del metro cuadrado por debajo de la provincia de Almería. Sevilla con un 62,60% menos, Málaga con un 45,02% menos, Cádiz con un 34,83% menos que Almería, Granada con un 27,22% menos, Jaén con un 16,14% menos y Córdoba con un 11,48% menos.

Variables numéricas

- a) Variable **Año**: El valor negativo de su coeficiente de regresión indica relación inversa siendo esta;

$$e^{-0.1008} - 1 = -0,0958.$$

Por cada año adicional que pase, el precio del metro cuadrado de las viviendas disminuye un 9,58%.

- b) Variable **Compraventas**: Tiene un signo positivo por lo tanto a medida que aumenta este, aumenta el precio del metro cuadrado de la vivienda.

$$e^{0,0000058} - 1 = 0,0000058$$

Por 1000 unidades de compras adicionales suponiendo que el resto de regresores se mantienen constantes, el precio del metro cuadrado aumenta un 0,58%

- c) Variable **Costes construcción**: Tiene un signo positivo por lo tanto a medida que aumenta este, aumenta el precio del metro cuadrado de la vivienda.

$$e^{0,04196} - 1 = 0,04285$$

Un incremento de una unidad en el índice de costes de construcción, suponiendo que el resto de regresores se mantienen constantes, implica un incremento del 4,28% en el precio del metro cuadrado de la vivienda.

Este procedimiento se realizaría de igual forma con las demás variables y observamos la lógica del modelo, porque variables como el Euribor o la tasa de paro que un incremento de ellas implica condiciones desfavorables para los habitantes tienen signo negativo en el coeficiente lo que significa que a medida que crecen estos disminuye el precio del metro cuadrado.

La variable Población tiene una interpretación distinta a las demás ya que se le ha aplicado también la transformación logarítmica. Esta variable indica que un incremento de un 1% en la Población, supone un incremento del 0,878% del precio del metro cuadrado suponiendo constantes el resto de los regresores.

3.3.2 CONTRASTES

En los modelos anteriores se ha realizado ajustes de regresión tomando como factor de referencia Almería y se han observado las diferencias del precio del metro cuadrado entre esta provincia y las demás provincias.

A continuación, contrastamos si el efecto del precio del metro cuadrado en Jaén es el mismo que en las demás provincias.

$$H_0: \beta_1 = \beta_5$$

$$H_1: \beta_1 \neq \beta_5$$

$$\text{Provincia[T.Cádiz]} - \text{Provincia[T.Jaén]} = 0$$

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
104	0.30776	1	0.011667	3.9427	0.04971 *

El p-valor indica que existen diferencias entre el precio del metro cuadrado de Jaén y el precio del metro cuadrado de la provincia de Cádiz.

$$H_0: \beta_2 = \beta_5$$

$$H_1: \beta_2 \neq \beta_5$$

$$\text{Provincia[T.Córdoba]} - \text{Provincia[T.Jaén]} = 0$$

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
104	0.30776	1	0.0048654	1.6442	0.2026

El p-valor no evidencia diferencias significativas en los precios del metro cuadrado de las provincias de Córdoba y de Jaén.

$$H_0: \beta_3 = \beta_5$$

$$H_1: \beta_3 \neq \beta_5$$

$$\text{Provincia[T.Granada]} - \text{Provincia[T.Jaén]} = 0$$

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
104	0.30776	1	0.012981	4.3867	0.03865 *

Existen diferencias significativas entre el precio del metro cuadrado de la vivienda en la provincia de Granada y el precio en la provincia de Jaén.

$$H_0: \beta_4 = \beta_5$$

$$H_1: \beta_4 \neq \beta_5$$

$$\text{Provincia[T.Huelva]} - \text{Provincia[T.Jaén]} = 0$$

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
104	0.30776	1	0.25442	85.975	2.84e-15 ***

UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado
Estadística y Empresa

Existen diferencias significativas entre los precios del metro cuadrado de la provincia de Huelva y la de Jaén.

$$H_0: \beta_6 = \beta_5$$

$$H_1: \beta_6 \neq \beta_5$$

$$\text{Provincia[T.Jaén]} - \text{Provincia[T.Málaga]} = 0$$

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
104	0.30776	1	0.016063	5.4281	0.02175 *

El p-valor indica que existen diferencias significativas entre el precio del metro cuadrado en Jaén y en Málaga.

$$H_0: \beta_7 = \beta_5$$

$$H_1: \beta_7 \neq \beta_5$$

$$\text{Provincia[T.Jaén]} - \text{Provincia[T.Sevilla]} = 0$$

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
104	0.30776	1	0.041919	14.166	0.0002773 ***

Existen diferencias significativas en el precio del metro cuadrado de la vivienda de las provincias de Jaén y Sevilla.

VALIDACIÓN

```
> MSEj  
[1] 0.003786826  
  
> MSE  
[1] 0.002772573  
  
> R2j  
[1] 0.9704484
```

Estos resultados nos muestran que el modelo de regresión ajustado se valida satisfactoriamente. El error cuadrático de Jack no es muy superior al error del modelo.

4. CONCLUSIONES

Un previo análisis del sector inmobiliario nos ha mostrado que diversos autores, en sus trabajos sobre estimación y explicación de la variabilidad del precio del metro cuadrado de la vivienda consideran que este último está sujeto a otras variables que afectan en último grado de manera positiva o negativa a su valor.

Haciendo un seguimiento de estos trabajos y en base a la información disponible en las bases de datos del Instituto Nacional de Estadística, El Instituto de Estadística y Cartografía de La Junta de Andalucía y El Ministerio de Fomento, así como información disponible en algunas páginas de registros económicos, se han considerado las variables Compraventas, Población, Salario Medio, PIB per cápita, Tasa de Paro, Euribor y Costes de Construcción para el estudio de la variabilidad del precio del metro cuadrado de las viviendas en las provincias andaluzas.

El análisis estadístico realizado con las variables indicadas evidencia las siguientes conclusiones:

UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado
Estadística y Empresa

La evolución de los precios del metro cuadrado de la vivienda es similar en las distintas provincias.

Todas tiene un patrón similar definido por una curva cóncava, donde se puede apreciar un incremento de los valores del precio desde 2000 hasta el 2007 en todas las provincias y después un descenso de estos precios a partir del año 2008. La provincia con mayores precios del metro cuadrado de la vivienda es Huelva, seguida de Almería y Córdoba. La que menos precios del metro cuadrado de la vivienda tiene es Sevilla, seguida de Málaga.

El sector inmobiliario también ha sido afectado por la crisis económica.

En todas las provincias se puede observar cómo los precios del metro cuadrado de la vivienda crecen notablemente desde el año 2000, este crecimiento alcanza su pico en el año 2007 y a partir del año 2008 se aprecia la caída de los precios. Por tanto, podemos señalar que la crisis del sector inmobiliario afectó especialmente a los precios de los inmuebles a partir del año 2008.

Las variables empleadas son todas significativas para explicar la variabilidad del precio del metro cuadrado de la vivienda.

El estudio de regresión realizado concluye que un 97,82% de la variabilidad de la respuesta se consigue explicar con nuestras variables regresoras. Las variables Año, Euribor y Tasa de paro guardan una relación inversa con el precio del metro cuadrado, cuando crece alguna de estas el precio decrece y las variables Compraventas, Costes de construcción, PIB per cápita, Población y Salario mantienen una relación directa de manera que si aumenta alguna de estas también lo hace el precio del metro cuadrado.

Los precios del metro cuadrado de las provincias difieren significativamente.

Tomando como categoría base la provincia de Almería concluimos que los precios del metro cuadrado de esta difieren significativamente de los precios de las demás provincias. Y los contrastes realizados indican que, excepto la provincia de Córdoba, todas las demás difieren significativamente del precio del metro cuadrado de la provincia de Jaén.

El problema de multicolinealidad

El resultado del trabajo no es del todo concluyente dada la presencia de multicolinealidad. Podría resolverse el problema usando componentes principales, aunque no es seguro que se vaya a solucionar empleando tal técnica. El empleo de esta técnica cambia la interpretación de las variables del modelo y el motivo fundamental era determinar si existían diferencias significativas entre los precios de las provincias.

También podría funcionar la eliminación de ciertos regresores que alcanzan valores muy elevados de FIV en la sección anexo se contempla esta posibilidad, pero dada la significación de los mismos esta solución no se va a tener en cuenta.

En última instancia lo mejor había sido disponer de más datos, por ejemplo, en vez de datos anuales disponer de datos trimestrales, al aumentar la muestra quizás se podría mitigar el problema, pero como este estudio se basa en datos disponibles en las bases de Datos, no se puede ampliar o recopilar más de los que se han conseguido.

5. BIBLIOGRAFÍA

Bernardos Domínguez G. (2009)., “Creación y destrucción de la burbuja inmobiliaria”, *ICE*, 850, 23-40. Disponible on line:

https://www.researchgate.net/publication/38292671_Creacion_y_destruccion_de_la_burbuja_inmobiliaria_en_Espana

Cuadradas, Carles M. (2014), *Nuevos Métodos de Análisis Multivariante*, Barcelona, CMC.

Hair, J. F., Anderson, R. E., Tatham, R. L., y Black, W. C. (1999), *Análisis Multivariante* 5.^a ed., Madrid, Pentice Hall Iberia.

López G. (2013)., ”La transformación del sector inmobiliario en España después de la crisis”, *Boletín de Estudios Económicos*, 68(209), 293-315. Disponible on line:

<http://0-search.proquest.com/avalos.ujaen.es/docview/1458280063/fulltextPDF/75F0CF9D1F5848EBPQ/1?accountid=14555>

Martínez Pagés J. y Maza L. A. (2003)., “Análisis del precio de la vivienda en España”, *Banco de España*, 307 Disponible on line:

<http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/03/Fic/dt0307.pdf>.

Montalvo García J. (2001)., “Un análisis empírico del crecimiento del precio de la vivienda en las comunidades autónomas españolas”. *Revista Valenciana de Economía y Hacienda*, 2, 117-138. Disponible on line:

www.econ.upf.edu/~montalvo/vivienda/rvalenciana.pdf

Montgomery, Peck y Vining. (2005), *Introducción al análisis de Regresión Lineal* 3^a ed., México, Cecs.

Pérez, J. y Gonzáles, L. (1996), “Las Estadísticas de Precios de la Vivienda en España. Un comentario a la metodología, coyuntura inmobiliaria: España y Andalucía”. *Cuadernos Económicos de Granada*, 105-114.

Ruiz, J. C. (2014), *Apuntes de la asignatura de Métodos de regresión*, Departamento de Estadística e Investigación Operativa, Universidad de Jaén.

Sánchez Carrión, J. J. (2008), *Manual de análisis estadístico de los datos 2ª ed.*, Madrid, EFCA, S.A.

Vergés, R. (2012)., “Estadística de vivienda libre en España. 1987-2011. series operativas de compraventas y precios”, *Estadística española*, 54(179), 357-420. Disponible on line:

<http://www.ricardoverges.com/pdf/ViviendaLibre.pdf>

6. ANEXO

Para mitigar el problema de la multicolinealidad una posible solución es la eliminación de variables que mantienen una correlación bastante alta entre sí. En nuestro caso son las variables Año y Costes de construcción. Cabe mencionar que también se ha realizado un modelo de regresión por pasos forward/backward (resultado no mostrado) y el resultado que ofrece es un modelo sin las variables Año y Costes de construcción en el modelo.

El modelo sin esas variables es el siguiente:

Residuals:					
Min	1Q	Median	3Q	Max	
-0.19485	-0.04140	-0.00398	0.04274	0.18045	
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-2.140e+00	1.380e+00	-1.551	0.124000	
Provincia[T.Cádiz]	-6.868e-01	1.493e-01	-4.601	1.17e-05	***
Provincia[T.Córdoba]	-1.721e-01	6.255e-02	-2.752	0.006974	**
Provincia[T.Granada]	-4.630e-01	8.882e-02	-5.212	9.26e-07	***
Provincia[T.Huelva]	4.878e-01	6.090e-02	8.010	1.59e-12	***
Provincia[T.Jaén]	-1.279e-01	3.738e-02	-3.421	0.000886	***
Provincia[T.Málaga]	-9.462e-01	2.143e-01	-4.414	2.45e-05	***
Provincia[T.Sevilla]	-1.410e+00	2.488e-01	-5.666	1.27e-07	***
Salario.medio	1.019e-04	1.418e-05	7.185	9.75e-11	***
Tasa.de.paro	-1.214e-02	1.658e-03	-7.317	5.08e-11	***
P	1.202e+00	2.321e-01	5.180	1.06e-06	***
PIB.Per.cápita	3.757e-05	1.009e-05	3.723	0.000317	***
Euribor	-2.846e+00	6.527e-01	-4.361	3.02e-05	***
compraventas	6.056e-06	2.658e-06	2.278	0.024714	*
Residual standard error: 0.06677 on 106 degrees of freedom					
Multiple R-squared: 0.9665, Adjusted R-squared: 0.9623					
F-statistic: 234.9 on 13 and 106 DF, p-value: < 2.2e-16					

Tabla 16. Resultados del modelo de regresión.

Fuente. Elaboración propia

Se observa que la interpretación de los coeficientes de regresión sería similar a la realizada con el modelo que incluía a todas las variables, pues el signo de ninguna variable es distinto y los coeficientes de regresión no son muy distintos a los obtenidos con todas las variables en el modelo por lo que podemos decir que en presencia de los otros regresores las variables año y costes de construcción no aportan mucha información extra al modelo.

	compraventas	Euribor	P	PIB.Per.cápita	Salario.medio	Tasa.de.paro
compraventas	1.00000000	0.39004573	0.55347772	0.035886095	0.06418137	-0.538975188
Euribor	0.39004573	1.00000000	-0.07812262	-0.242411869	-0.31992967	-0.684618851
P	0.55347772	-0.07812262	1.00000000	0.144883102	0.58717680	0.108057156
PIB.Per.cápita	0.03588609	-0.24241187	0.14488310	1.000000000	0.70893828	-0.002869468
Salario.medio	0.06418137	-0.31992967	0.58717680	0.708938278	1.00000000	0.270125130
Tasa.de.paro	-0.53897519	-0.68461885	0.10805716	-0.002869468	0.27012513	1.000000000

Tabla 17. Matriz de correlación

Fuente. Elaboración propia

Ningún valor fuera de la diagonal principal de esta matriz es ahora superior a 0,9.

compraventas	Euribor	P	PIB.Per.cápita	Salario.medio	Tasa.de.paro
3.583698	2.144602	4.698016	3.458155	4.638815	3.123617

Tabla 18. Factores de inflación de la varianza

Fuente. Elaboración propia

Se observa la ausencia de multicolinealidad, todos los VIFs son menores que 5.

<pre> studentized Breusch-Pagan test data: Y ~ Provincia + Salario.medio + Tasa.de.paro + P + PIB.Per.cápita + + Euribor + compraventas BP = 0.76168, df = 1, p-value = 0.3828 </pre>
<pre> RESET test data: Y ~ Provincia + Salario.medio + Tasa.de.paro + P + PIB.Per.cápita + + Euribor + compraventas RESET = 0.018839, df1 = 1, df2 = 105, p-value = 0.8911 </pre>
<pre> Shapiro - Wilk Normality Test Test Results: STATISTIC: W: 0.989 P VALUE: 0.4515 </pre>

Tabla 19. Test de homocedasticidad, linealidad y normalidad.

Fuente. Elaboración propia

```

> MSEj
[1] 0.005661819

> MSE
[1] 0.004257208

> R2j
[1] 0.9558164

```

Al igual que el modelo anterior podemos ver que este modelo también se valida satisfactoriamente.

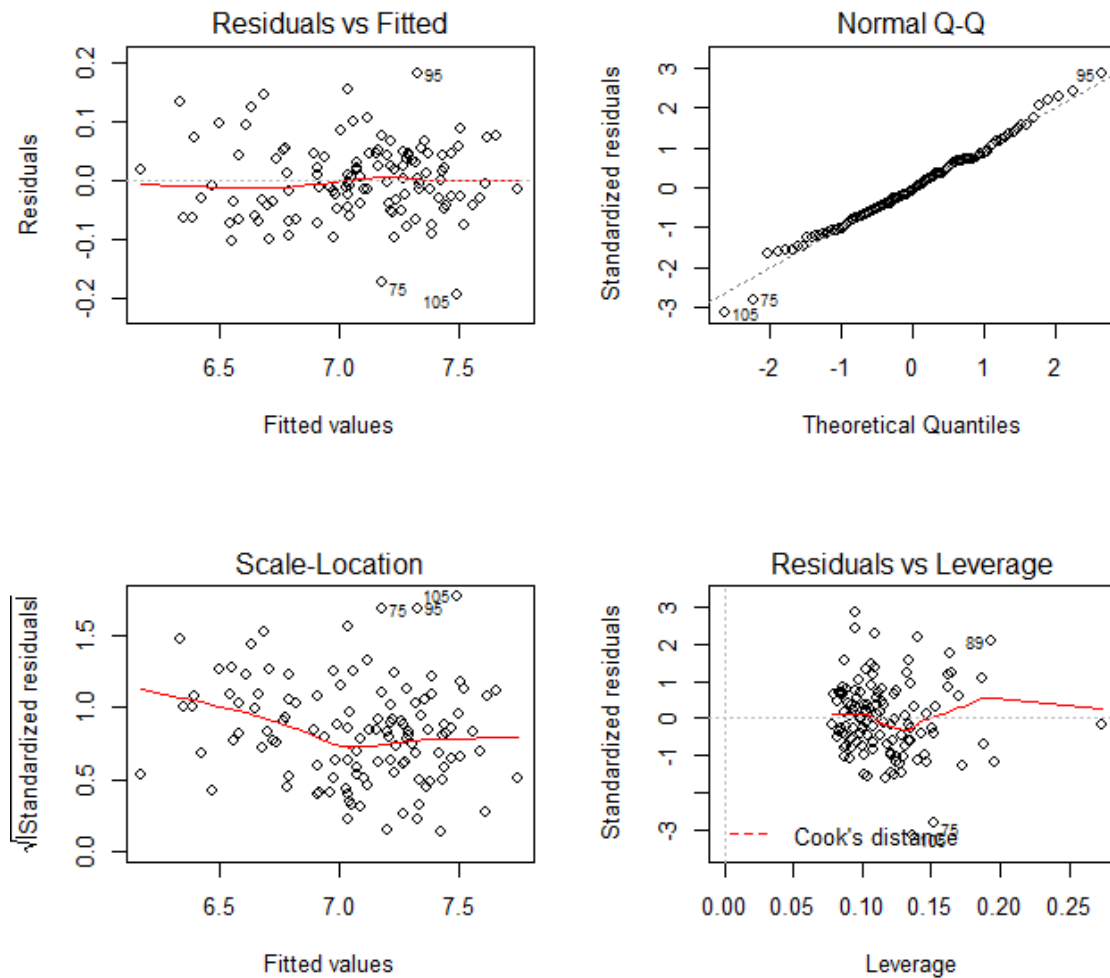


Figura 5. Gráficas básicas de diagnóstico

Fuente. Elaboración propia

Estas gráficas confirman los resultados obtenidos en los test de normalidad, linealidad y homocedasticidad.

6.1 RESULTADOS

R

Importar los datos

```
> datos <- readXL("C:/Users/Marcos/Desktop/tfg.xlsx", rownames=FALSE,  
header=TRUE, na="", sheet="Hoja 1", stringsAsFactors=TRUE)  
  
> summary(Datos)
```

Regresión Lineal con todos los datos

```
> mod <- lm(Precio.m2 ~ Provincia +Tasa.de.paro +Salario.medio +Población  
+PIB.Per.cápita +Euribor +costes.cos +compraventas +Año, data=datos)  
  
> summary(mod)
```

Diagnosis

Homocedasticidad

```
> bptest(Precio.m2 ~ Provincia + Salario.medio + Tasa.de.paro + Población +  
PIB.Per.cápita + Euribor + costes.cos +compraventas + Año, varformula = ~  
fitted.values(mod), studentize=TRUE, data=datos)  
  
library (zoo)
```

Normalidad

```
> shapiroTest(mod$residuals)  
  
library (fBasics)
```

UNIVERSIDAD DE JAÉN
Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado
Estadística y Empresa

Linealidad

```
> resettest(Precio.m2 ~ Provincia + Salario.medio + Tasa.de.paro + Población +  
PIB.Per.cápita + Euribor + costes.cos +compraventas + Año, power=2:3,  
type="regressor", data=datos)
```

Autocorrelación

```
> dwtest(Precio.m2 ~ Provincia + Tasa.de.paro + Salario.medio + Población +  
PIB.Per.cápita + Euribor + costes.cos + compraventas + Año, alternative="greater",  
data=datos)
```

Gráficos

De diagnóstico

```
> plot(mod)
```

Diagrama de dispersión

```
> scatterplot(Precio.m2~Año | Provincia, reg.line=FALSE, smooth=TRUE,  
spread=FALSE, boxplots=FALSE, span=0.5, ellipse=FALSE, levels=c(.5, .9),  
xlab="Año", ylab="Precio", by.groups=TRUE, data=datos)
```

De influencia

```
> influencePlot(mod, id.method="noteworthy", id.n=2)
```

Matriz de correlaciones

```
> cor(datos[,c("Año", "compraventas", "costes.cos", "Euribor", "PIB.Per.cápita", "Població  
n", "Salario.medio", "Tasa.de.paro")], use="complete")
```

Factores de inflación de la varianza

>

```
diag(solve(cor(datos[,c("Año", "compraventas", "costes.cos", "Euribor", "Población", "PIB",  
.Per.cápita", "Salario.medio", "Tasa.de.paro")], use="complete"))))
```

Atipicidad

```
> outlierTest(mod)
```

Influencia

```
> influence.measures(mod)
```

Transformación Box Cox

```
> boxcox(mod)
```

```
> library(MASS)
```

```
> datos$Y <- with(datos, log(Precio.m2))
```

Test de hipótesis

```
> local({.Hypothesis <- matrix(c(0,0,0,0,1,-1,0,0,0,0,0,0,0,0,0), 1, 16, byrow=TRUE)  
.RHS <- c(0)linearHypothesis(mod2, .Hypothesis, rhs=.RHS)})
```