



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**VICERRECTORADO DE INVESTIGACIÓN,
INNOVACIÓN Y TRANSFERENCIA DE
TECNOLOGÍA**

**CENTRO DE POSGRADOS
MAESTRÍA EN ENSEÑANZA DE LA MATEMÁTICA**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL
TÍTULO DE MAGISTER EN ENSEÑANZA DE LA MATEMÁTICA**

**TEMA: ANÁLISIS COMPARATIVO ENTRE UN MODELO DE
FACTORES Y REGRESIÓN HEDÓNICA PARA LA ESTIMACIÓN
DEL PRECIO DE LA VIVIENDA EN EL CANTÓN RUMIÑAHUI.**

**AUTOR: ORTEGA LOZA, FERNANDO WLADMIR
DIRECTOR: MGS. ARMIJO TORO, LIVINO MANUEL**

SANGOLQUÍ 2019



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA**

CENTRO DE POSGRADOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, *“ANÁLISIS COMPARATIVO ENTRE UN MODELO DE FACTORES Y REGRESIÓN HEDÓNICA PARA LA ESTIMACIÓN DEL PRECIO DE LA VIVIENDA EN EL CANTÓN RUMIÑAHUI”* fue realizado por el señor *Ortega Loza, Fernando Wladimir*, el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 25 de noviembre del 2019

Firma:

.....
Livino Manuel Armijos Toro

C.C.: 0916543747



TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

AUTORÍA DE RESPONSABILIDAD

Yo, *Ortega Loza, Fernando Wladimir*, con cédula de ciudadanía n° 1002635793, declaro que el contenido, ideas y criterios del trabajo de titulación: *“Análisis comparativo entre un modelo de factores y regresión hedónica para la estimación del precio de la vivienda en el cantón Rumiñahui”* es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Consecuentemente el contenido de la investigación mencionada es veraz.

Sangolquí, 25 de noviembre del 2019

Firma:

Fernando Wladimir Ortega Loza

C.C.: 1002635793



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA
CENTRO DE POSGRADOS**

AUTORIZACIÓN

Yo, *Ortega Loza, Fernando Wladimir*, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: *“Análisis comparativo entre un modelo de factores y regresión hedónica para la estimación del precio de la vivienda en el cantón Rumiñahui”* en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 25 de noviembre del 2019

Firma:



Fernando Wladimir Ortega Loza

C.C.: 1002635793

DEDICATORIA

A mi hijo Alejandro y a mi esposa Johana, por ser mi fuerza, motivación y fe en toda mi vida.

AGRADECIMIENTO

Al Mgs. Livino Armijos por su buena predisposición y tutoría académica en este trabajo.

ÍNDICE DE CONTENIDOS

CARATULA

CERTIFICADO DEL DIRECTOR i

AUTORÍA DE RESPONSABILIDAD ii

AUTORIZACIÓN iii

DEDICATORIA iv

AGRADECIMIENTO v

ÍNDICE DE CONTENIDOS vi

ÍNDICE DE TABLAS viii

ÍNDICE DE FIGURAS x

RESÚMEN xiv

ABSTRACT xv

CAPITULO I INTRODUCCIÓN 1

1.1. Antecedentes. 1

1.2. Planteamiento del Problema. 2

1.3. Objetivos. 3

1.3.1. Objetivo General. 3

1.3.2. Objetivos Específicos. 3

1.4. Justificación. 4

1.5. Hipótesis de Investigación. 5

1.6. Estado del Arte. 5

CAPITULO II MARCO TEÓRICO..... 8

2.1. Estadística Descriptiva. 8

2.2. Análisis de Componentes Principales. 9

2.3. Diseño Factorial. 13

2.3.1. Diseño Completamente al azar DCA. 13

2.3.2. Análisis de Varianza ANOVA. 13

2.3.3. Diseños Factoriales. 16

2.4. Índice de Morán. 25

2.5. Regresión Hedónica. 29

CAPITULO III METODOLOGÍA 32

3.1.	Recolección de Datos.	32
3.2.	Estudio descriptivo de los datos.....	37
3.3.	Análisis de Componentes Principales ACP.	60
3.4.	Metodología para realizar los modelos	63
3.5.	Diseño Factorial.	69
3.5.1.	Modelo de factores 1.	71
3.5.2.	Modelo de Factores 2.....	77
3.5.3.	Modelo de Factores 3.....	82
3.6.	Correlación Espacial.....	89
3.7.	Regresión Hedónica.	90
3.7.1.	Modelo de regresión hedónica 1.	92
3.7.2.	Modelo de regresión hedónica 2.	98
3.7.3.	Modelo de regresión hedónica 3.	104
3.7.4.	Modelo de regresión hedónica 4.	109
3.8.	Validación Cruzada.	114
3.9.	Selección de modelos.	115
3.9.1.	Selección de modelo de factores.	115
3.9.2.	Selección de modelo de regresión hedónica.....	116
3.9.3.	Selección de mejor modelo entre factores y regresión hedónica.....	117
3.10.	Clases de Equivalencia	118
	CAPITULO IV CONCLUSIONES Y RECOMENDACIONES.....	120
4.1.	Conclusiones.	120
4.2.	Recomendaciones.....	121
	BIBLIOGRAFÍA	122
	ANEXOS.....	128
	Anexo 1. Datos validados.....	128
	Anexo 2. Modelos de factores	147
	Anexo 3. Modelos de regresión hedónica.....	149

ÍNDICE DE TABLAS

Tabla 1 ANOVA para el diseño factorial $a \times b$	18
Tabla 2 ANOVA para el diseño factorial $a \times b \times c$	19
Tabla 3 Georreferencias de las Unidades Educativas representativas del cantón Rumiñahui.....	33
Tabla 4 Georeferenciación de los establecimientos de salud del cantón Rumiñahui... 35	
Tabla 5 Zonas de desarrollo económico del cantón Rumiñahui.....	36
Tabla 6 Instituciones de Educación Superior representativas del cantón Rumiñahui ... 36	
Tabla 7 Estadística Descriptiva de la Superficie Total (m ²)	38
Tabla 8 Estadística Descriptiva de la Superficie Cubierta (m ²)	39
Tabla 9 Estadística Descriptiva de número de cuartos	41
Tabla 10 Estadística Descriptiva de número de estacionamientos	42
Tabla 11 Estadística Descriptiva de número de baños.....	44
Tabla 12 Estadística Descriptiva de años de antigüedad.....	45
Tabla 13 Estadística descriptiva de número de pisos de las viviendas	48
Tabla 14 Estadística Descriptiva de distancia mínima a una unidad educativa	49
Tabla 15 Estadística Descriptiva de distancia mínima a un centro de salud.....	51
Tabla 16 Estadística Descriptiva de distancia mínima a centros de desarrollo económico.....	53
Tabla 17 Estadística Descriptiva de distancia mínima a una Centro de Educación Superior.....	55
Tabla 18 Estadística Descriptiva de la variable Precio por m ² total.....	56

Tabla 19 <i>Estadística Descriptiva de la variable Precio</i>	58
Tabla 22 <i>Porcentaje de variación representado por las componentes principales</i>	62
Tabla 21 <i>Representación de la variabilidad de las Componentes Principales en función en las variables</i>	62
Tabla 22 <i>Nomenclatura propuesta para Variables</i>	64
Tabla 23 <i>Correlación entre variable dependiente y variables independientes</i>	65
Tabla 24 <i>Clases de la variable Número de Habitaciones</i>	67
Tabla 25 <i>Clases de la variable Número de Baños</i>	67
Tabla 26 <i>Clases de la variable número de estacionamientos</i>	68
Tabla 27 <i>Clases de la variable Antigüedad</i>	68
Tabla 28 <i>Clases de la variable Pertinencia a Condominio</i>	69
Tabla 29 <i>Clases de la variable número de pisos</i>	69
Tabla 30 <i>Modelos corridos en función de las opciones de clases de equivalencia</i>	70
Tabla 31 <i>Asignación de variables dummy</i>	91
Tabla 32 <i>Errores de los predicho, mediante validación cruzada</i>	114
Tabla 33 <i>Prueba de igualdad de medias de los residuos mf</i>	115
Tabla 34 <i>Resumen de pruebas a modelos de factores</i>	116
Tabla 35 <i>Prueba de igualdad de medias de los residuos de losMrh</i>	116
Tabla 36 <i>Resumen de pruebas a modelos de regresión hedónica</i>	117
Tabla 37 <i>Resumen de pruebas a modelos</i>	118
Tabla 39 <i>Categorías de variables</i>	118

ÍNDICE DE FIGURAS

<i>Figura 1.</i> Gráfica alineada de probabilidad normal de los residuos	21
<i>Figura 2.</i> Grafica de predichos vs residuos.....	23
<i>Figura 3.</i> Grafica de factor vs residuos.....	23
<i>Figura 4.</i> Orden de corrida vs residuos.....	25
<i>Figura 5.</i> Contigüidad tipo torre	27
<i>Figura 6.</i> Contigüidad tipo alfil	27
<i>Figura 7.</i> Contigüidad tipo reina.....	27
<i>Figura 8.</i> Histograma de la variable Superficie Total.....	38
<i>Figura 9.</i> Diagrama de cajas de la variable Superficie Total.....	39
<i>Figura 10.</i> Histograma de la variable superficie cubierta.....	40
<i>Figura 11.</i> Diagrama de cajas de la variable Superficie Cubierta.....	40
<i>Figura 12.</i> Histograma de la variable número de cuartos	42
<i>Figura 13.</i> Diagrama de cajas de la variable número de cuartos	42
<i>Figura 14.</i> Histograma para la variable número de estacionamientos	43
<i>Figura 15.</i> Diagrama de cajas de la variable número de estacionamientos	43
<i>Figura 16.</i> Histograma de la variable número de baños	44
<i>Figura 17.</i> Diagrama de cajas de la variable número de baños.....	45
<i>Figura 18.</i> Histograma de la variable antigüedad	46
<i>Figura 19.</i> Diagrama de cajas de la variable Antigüedad	46
<i>Figura 20.</i> Diagrama de barras de la variable pertenencia a condominio.....	47
<i>Figura 21.</i> Histograma de la variable número de pisos	48

Figura 22. Histograma de la variable distancia mínima a una unidad educativa.....	50
Figura 23. Diagrama de cajas de la variable distancia mínima a una Unidad Educativa	50
Figura 24. Histograma de distancia mínima a un centro de salud	51
Figura 25. Diagrama de cajas de la variable distancia mínima a un Centro de Salud	52
Figura 26. Histograma de variable distancia mínima a centro de desarrollo económico	53
Figura 27. Diagrama de cajas de la variable distancia mínima a centro de desarrollo económico	54
Figura 28. Histograma de la variable distancia mínima a un CES	55
Figura 29. Diagrama de cajas de la variable distancia mínima a un CES	56
Figura 30. Histograma de la variable Precio por m2 total.....	57
Figura 31. Diagrama de cajas de la variable Precio por m2 total.....	57
Figura 32. Histograma de la variable Precio por m2 total.....	58
Figura 33. Diagrama de cajas de la variable Precio por m2 total.....	59
Figura 34. Circulo de correlación de las variables	60
Figura 35. Mapeo de las componentes 1 y 2 del ACP	61
Figura 36. Diagrama de Correlación entre variables independientes y precio por m2 total.....	64
Figura 37. Diagramas de pastel de las variables independientes	67
Figura 38. Pxm2T real vs. Pxm2T predicho por mf1	72
Figura 39. Histograma de los residuos de mf1	73
Figura 40. Gráfica Cuantil - Cuantil de mf1	74
Figura 41. Diagrama de variables vs Residuos mf1	75

<i>Figura 42.</i> Diagrama de orden de toma de observaciones vs. residuos	76
<i>Figura 43.</i> Pxm2T real vs. Pxm2T predicho por mf2	78
<i>Figura 43.</i> Histograma de los residuos de mf2	79
<i>Figura 45.</i> Gráfica Cuantil - Cuantil de mf2.....	79
<i>Figura 46.</i> Diagrama de variables vs Residuos mf2	80
<i>Figura 47.</i> Diagrama de orden de toma de observaciones vs. residuos	82
<i>Figura 48.</i> Pxm2T real vs. Pxm2T predicho por mf2	84
<i>Figura 49.</i> Histograma de los residuos de mf3	85
<i>Figura 50.</i> Gráfica Cuantil - Cuantil de mf3.....	86
<i>Figura 51.</i> Diagrama de variables vs Residuos mf3	87
<i>Figura 52.</i> Diagrama de orden de toma de observaciones vs. Residuos de mf3.....	88
<i>Figura 53.</i> Georeferenciaación de las viviendas	89
<i>Figura 54.</i> Pxm2T real vs. Pxm2T predicho por Mrh1.....	93
<i>Figura 55.</i> Histograma de los residuos de Mrh1	94
<i>Figura 56.</i> Gráfica Cuantil - Cuantil de Mrh1	95
<i>Figura 57.</i> Diagrama de variables vs Residuos Mrh1	96
<i>Figura 58.</i> Diagrama de orden de toma de observaciones vs. Residuos de Mrh1	97
<i>Figura 59.</i> Pxm2T real vs. Pxm2T predicho por Mrh2.....	99
<i>Figura 60.</i> Histograma de los residuos de Mrh2.....	100
<i>Figura 61.</i> Gráfica Cuantil - Cuantil de Mrh2	100
<i>Figura 62.</i> Diagrama de variables vs Residuos Mrh2.....	102
<i>Figura 63.</i> Diagrama de orden de toma de observaciones vs. Residuos de Mrh2	103

Figura 64. Pxm2T real vs. Pxm2T predicho por Mrh3.....	104
Figura 65. Histograma de los residuos de Mrh3.....	105
Figura 66. Gráfica Cuantil - Cuantil de Mrh3	106
Figura 67. Diagrama de variables vs Residuos Mrh3.....	107
Figura 68. Diagrama de orden de toma de observaciones vs. Residuos de Mrh3	108
Figura 69. Pxm2T real vs. Pxm2T predicho por Mrh4.....	109
Figura 70. Histograma de los residuos de Mrh4.....	110
Figura 71. Gráfica Cuantil - Cuantil de Mrh4	111
Figura 72. Diagrama de variables vs Residuos Mrh4.....	112
Figura 73. Diagrama de orden de toma de observaciones vs. Residuos de Mrh3	113

RESUMEN

La presente investigación tiene como objetivo comparar la precisión entre un modelo de factores y un hedónico en la estimación del precio de la vivienda en el cantón Rumiñahui, para ello se realiza una recopilación de la teoría, el estado del arte y datos del precio de la vivienda, en los portales olx.com.ec y plusvalía.com.ec, para luego de filtrarlos obtener una base de datos de 292 inmuebles, dicha base se divide en dos grupos, uno del 90% para realizar los modelos y otro de 10% para la validación cruzada. Posteriormente se realiza un estudio descriptivo y categorización de las variables intervinientes, se realiza 44 modelos de factores, tomando como variable dependiente al precio por metro cuadrado total, se incluye variables de zonificación y se realiza 20 modelos hedónicos, los cuales se evalúan con criterios de normalidad, homocedasticidad, independencia y correlación espacial mediante las pruebas de Lilliefors, Bartlett, Durbin Watson y el Índice de Morán, respectivamente, luego se escoge los mejores modelos, en los cuales se realiza la validación cruzada y se determina que el modelo de factores tiene un 25.43% y el modelo hedónico 32.93% de error de predicción, en consecuencia, se escoge como mejor al modelo de factores por su mejor bondad de ajuste.

PALABRAS CLAVE:

- **ANOVA**
- **VIVIENDA**
- **MODELADO**
- **HEDÓNICO**
- **VALIDACIÓN CRUZADA**

ABSTRACT

The purpose research is compare the precision between a factor model and a hedonic model in the price prediction of houses at Rumiñahui country, start with an issue introduction, then compile the theory, state of the art and data is related about to housing in Rumiñahui, in the portals olx.com.ec and plusvalía.com.ec, so after filtering, obtain a database of 292 properties, such a base is divided in two groups, one of 90% to make the models and the other one of 10% ,for cross-validation them. Subsequently, a descriptive study and categorized the intervening variables and forty-four factor models are made,, taking as a variable dependent , the total price per square meter, in addition zoning variables are included and twenty hedonic models are made.The models with normality, homocedasticity, independence and spatial correlation criteria are evaluated by the Lilliefors, Bartlett, Durbin Watson, and Morán Index test, this last one only for the factor models, depending on the mentioned tests is chosen the best models, in which the cross variation is performed and the prediction errors of the seven chosen models are obtained.By the error prediction, and the assumptions models test, it is determined that the factor model has 25.43% and the hedonic model 32.93% prediction error, consequently, the factor model is chosen as the best by its goodness of fit.

KEY WORDS:

- **ANOVA**
- **HOUSING**
- **MODELING**
- **HEDONIC**
- **CROSS VALIDATION**

CAPITULO I

INTRODUCCIÓN

1.1. Antecedentes.

Según resultados de la Encuesta Nacional De Empleo, Desempleo Y Subempleo (ENEMDU), en Ecuador se estima que existen 5'655.979 personas sin acceso a una vivienda adecuada¹, lo que representa que el 33.7% de la población tiene un déficit cualitativo de vivienda («ENEMDU», 2018). Este dato se refiere a personas que tienen vivienda propia, arrendada o prestada en condiciones no adecuadas. Para el año 2017 el Instituto Nacional de Estadística y Censos (INEC) en su encuesta de hogares estimó en un 35.3% los hogares sin vivienda propia, estos datos reflejan la necesidad insatisfecha de vivienda en el Ecuador, es así que el mercado inmobiliario y los estudios que se hacen alrededor de este toman especial importancia.

El costo de las viviendas y la importancia de estas en la vida de las personas, hace que la compra de una vivienda sea un acontecimiento importante que merece análisis a momento de tomar una decisión. El Código Orgánico de Organización Territorial establece lineamientos acerca de quién y cómo se hacen los avalúos catastrales de las viviendas, estos son determinados por un proceso de comparación con precios unitarios (Función Ejecutiva, Presidencia de la República del Ecuador, 2010), sin embargo estos avalúos son referenciales y el precio depende de la dinámica oferta y demanda, así como lo es en otras ciudades de Latinoamérica (Camelo Rincón & Campo Robledo, 2016).

¹ Vivienda adecuada: según la ONU es la que tiene seguridad de tenencia, disponibilidad de servicios, asequibilidad, habitabilidad, accesibilidad, ubicación y adecuación cultural.

Según datos del último censo de población y vivienda, en la Provincia de Pichincha 413.725 personas no tienen casa propia totalmente pagada, esto corresponde al 56.8% (INEC, 2010). En el cantón Rumiñahui de la provincia de Pichincha, se ha registrado un crecimiento demográfico del 24.1% del 2010 al 2015 («Gobierno Municipal Rumiñahui», 2019) y en el número de construcciones del 25% del año 2011 al 2012, cifras que se reflejan de los trámites para aprobación de planes y permisos definitivos («Más casas y plusvalía en Rumiñahui», s. f.).

1.2. Planteamiento del Problema.

Según el Código Orgánico de Organización Territorial vigente, el avalúo catastral toma en cuenta los factores: valor del suelo y valor de las edificaciones, sin embargo, existen muchos más factores alrededor del precio de una vivienda. En el mercado inmobiliario existen muchas viviendas que tienen similares características y distintos precios, es por ello que es importante disponer de modelos que expresen de buena forma el precio de la vivienda en función de los múltiples factores.

Una forma de modelar el precio de la vivienda es mediante el diseño por factores, para ello existe la necesidad de crear clases de equivalencia que puedan categorizar una vivienda con precisión, dichas clases de equivalencia deben evaluar los diversos factores (características) alrededor del precio de una vivienda.

La creación de clases de equivalencia, siguiendo aspectos de la metodología planteada por (Montero Lorenzo & Larraz, 2006) toma especial relevancia en la clasificación de la vivienda, identificando la variabilidad en cuanto al valor del inmueble y analizando ésta en función de su ubicación geográfica. El trabajo de titulación de fin de master, podrá contribuir

al análisis del valor de viviendas, así como, si la variabilidad del precio de viviendas a características similares posee una relación espacial.

La presente investigación pretende contrastar dos modelos de precio de vivienda, uno mediante el diseño de factores (características), y analizar la correlación espacial de lo no explicado por los factores, y un segundo modelo utilizando regresión hedónica, los dos modelos sobre una ubicación geográfica concentrada, tal como el cantón Rumiñahui, con un aproximado de 700 viviendas en venta. Los resultados de esta investigación se podrán extender a otras localidades de mayor dimensión, como es el caso del Distrito Metropolitano de Quito, el cual a la fecha cuenta con 13.500 publicaciones disponibles.

1.3. Objetivos.

1.3.1. Objetivo General.

Comparar la precisión entre un modelo de factores y un modelo hedónico en la estimación del precio de la vivienda en el cantón Rumiñahui.

1.3.2. Objetivos Específicos.

- Modelizar el precio de la vivienda respecto a sus factores (características), a través de análisis y diseño de experimentos (ANOVA) y caracterizar clases de equivalencia.
- Determinar los factores y sus interacciones relevantes en el modelo de factores.
- Modelizar el precio de la vivienda mediante regresión hedónica.
- Determinar que variables son relevantes en el modelo de regresión hedónica.
- Comparar el modelo por factores con el modelo de regresión hedónica.

1.4. Justificación.

La comisión Económica para América Latina cifró que 130 millones de latinoamericanos habitan en alojamientos precarios y 140 millones carecen de vivienda (Serrano, s. f.), el mercado inmobiliario de América Latina es un mercado insatisfecho y creciente , es así que las investigaciones alrededor del sistema inmobiliario toman relevancia.

Los diseños factoriales completos son la estrategia experimental óptima para estudiar simultáneamente el efecto de varios factores sobre la respuesta y sus interacciones (Medina & López, 2011), es así que los modelos utilizando diseño de factores toman especial relevancia en el ámbito inmobiliario.

Existen muchos factores alrededor del precio de las viviendas, siendo la creación de clases equivalentes de vivienda y el estudio de los factores, temas influyentes sobre el precio de las casas, ya que pueden generar modelos de estimación con gran calidad (Chica-Olmo, 2005).

El análisis descriptivo y la creación de clases equivalentes de vivienda y el Análisis de Componentes Principales son un requisito previo, para el modelamiento mediante factores, ya que las clases de equivalencia permitirán corregir distorsiones en el valor de los bienes y el Análisis de Componentes Principales es considerado como un método de reducción, esto es, un método que permite reducir la dimensión del número de variables originales que se han considerado en el análisis (De la Fuente, 2011).

La presente investigación pretende contrastar el modelo de diseño por factores con un modelo de regresión hedónica, el cual es un modelo clásico que persigue cuantificar el precio de la vivienda, involucrando las características constructivas y espaciales que más influyen sobre el precio de la vivienda (Olmo, Guervos, & Olmo, 2007). Estos modelos ayudarán a

las personas y empresas inmobiliarias a estimar el precio de venta o compra de los inmuebles con características similares.

1.5. Hipótesis de Investigación.

- El modelo de factores expresa de mejor manera el precio de la vivienda que el modelo hedónico.
- La variación del precio entre viviendas de similares características está determinada por una relación espacial.

1.6. Estado del Arte.

El diseño estadístico de experimentos fue introducido a mitad del siglo XX en Inglaterra por Ronald A. Fisher, con su libro *The Design of Experiments* (1935). Fisher también proporcionó métodos para diseñar experimentos destinados a investigar la influencia simultánea de varios factores (Pulido, 2008). Posteriormente George E. P. Box, desarrolló la metodología de superficie de respuestas, pero no es hasta 1980 con el fortalecimiento de los recursos computacionales que se promovió el uso de la estadística en calidad, los representantes en esta época fueron Deming e Ishikawa (Draper, 1992). En Japón destaca el trabajo de Genichi Taguchi, cuyos conceptos significativos en la academia en el mundo occidental (Pulido, 2008).

Los modelos de factores son parte del diseño de experimentos y se han utilizado en múltiples ámbitos de la ciencia, por ejemplo (Torres, Gamarra, & Huayana, 2019) lo utilizan para analizar y generar clases de equivalencia en el sector ganadero de la región Amazónica, en función de variables tecnológicas y productivas , (Yacuzzi, Martín, Quiñones, &

Popovsky, 2004) muestran varias aplicaciones farmacéuticas de diseño y de servicio utilizando el diseño de factores y los métodos de Taguchi, (Mg, Rc, Jm, & Vb, 2004) realizan una evaluación y optimización simultánea de papaína inmovilizada en gránulos de alginato entrecruzado mediante un diseño factorial 3x3 y la función de deseabilidad. La mayoría de autores trabajan con variables categóricas ya definidas, sin embargo, pocos muestran como pasar de variables reales a variables categóricas realizando un proceso de reducción de variables y creación de clases de equivalencia.

En el área de la vivienda (Apergis & Payne, 2012) utiliza la creación de clases de equivalencia para demostrar la convergencia entre el estado como división geográfica y los precios de las viviendas, de igual manera lo hace (Apergis, Simo-Kengne, & Gupta, 2015) en Sudáfrica, es así que la creación de clases desde un punto estadístico es una herramienta en la caracterización por grupos de tipos de vivienda.

(Montero Lorenzo, 2004) en su investigación “El precio medio del metro cuadrado de la vivienda libre: Una aproximación metodológica desde la perspectiva de la Geoestadística”, (Montero Lorenzo & Larraz, 2006) en su artículo “Estimación espacial del precio de la vivienda mediante métodos de Krigado”, utiliza diseño factorial para generar el modelo del precio de la vivienda en la ciudad de Toledo España, en donde se toma como factores superficie, garaje y estado de la vivienda, sin embargo se deja abierto el método que se utiliza para determinar la correlación positiva de factores, estas investigaciones servirán de base para la presente investigación, tomando en cuenta mayor números de factores y herramientas adicionales de estadística multivariante.

(Yarasca, 2015) utiliza el Índice de Morán para determinar la autocorrelación espacial en zonas de tránsito, así como él, muchos autores utilizan esta técnica para determinar la correlación espacial con alguna variable de interés en su investigación. (Collazos, Gamboa, Prado, & Verrardi, 2006) utilizan la I de Morán para verificar si la espacialidad (coordenadas geográficas) tiene correlación con el precio de la vivienda, sin embargo no se hace una reducción de variables ni el análisis propuesto en la presente investigación.

Los modelos hedónicos han sido ampliamente utilizados para estimar el precio de la vivienda, es así que tenemos investigaciones como las de (Cerdeira & Surhoff, 2002), (Pozo, 2007) y (Perdomo Calvo & Mendieta-Lopez, 2007), es así que se desea realizar la comparación de un modelo de factores con un modelo hedónico ampliamente difundido.

CAPITULO II

MARCO TEÓRICO

2.1. Estadística Descriptiva.

Población es el conjunto de unidades o individuos, siendo una **variable** cualquier característica de los individuos. Al observar una variable en los individuos, se obtiene **datos** u **observaciones**, cada observación de un individuo tiene un valor distinto, denotándose x_1, \dots, x_h , los h valores de una variable X (Castillo & Guijarro, 2006).

La **Frecuencia Absoluta** de un valor es el número de observaciones iguales a dicho valor. Se denota por n_i la frecuencia absoluta genérica, esto es, la frecuencia absoluta correspondiente al valor x_i . Si N es el número total de datos se tiene:

$$\sum_{i=1}^h n_i = N \quad (1)$$

La **frecuencia relativa** de un valor es la proporción de observaciones iguales a dicho valor. Se denota por f_i la frecuencia relativa del valor x_i , Así:

$$f_i = \frac{n_i}{N} \quad (2)$$

Cuando el número de valores de una variable es muy grande puede resultar aconsejable agrupar los valores en intervalos o clases.

La **media aritmética** se define como la suma de las observaciones de una variable dividida entre el número de ellas.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^h x_i n_i = \sum_{i=1}^h x_i f_i \quad (3)$$

También existen otros tipos de medias como la **media geométrica** que se emplea para promediar índices y tasas, y la **media armónica** que se usa para promediar magnitudes relativas.

La **mediana** de un conjunto de datos es el número que, en una ordenación, tiene a su derecha y a su izquierda el mismo número de observaciones.

$$Me = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} c_i \quad (4)$$

L_{i-1} y c_i son, respectivamente, el extremo inferior y la amplitud del intervalo mediano.

La **moda** de una distribución de frecuencias es el valor con mayor frecuencia

$$Mo = L_{i-1} + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} c_i \quad (5)$$

Donde c_i es la amplitud del intervalo modal y d_i es la densidad de frecuencia de dicho intervalo.

La varianza es una medida de dispersión que permite conocer que tan agrupados alrededor del centro están los datos, a partir de la distribución de frecuencias $(x_i; n_i)$, se calcula como:

$$S^2 = \frac{1}{N} \sum_{i=1}^h (x_i - \bar{x})^2 n_i = \sum_{i=1}^h (x_i - \bar{x})^2 f_i \quad (6)$$

2.2. Análisis de Componentes Principales.

El método de Análisis de Componentes Principales (ACP) sirve para transformar un conjunto de variables originales en otras llamadas Componentes Principales, como medida de la cantidad de información que tienen las componentes principales se utiliza su varianza, siendo la primera componente la que mayor varianza tiene y mayor información. Las componentes

principales están incorrelacionadas entre sí. Cuando las variables originales están muy correlacionadas entre sí, la mayor parte de su variabilidad se puede expresar con pocas componentes, es así que en la práctica se utiliza en ACP como un método de reducción de variables, cuando esas están correlacionadas entre sí.

Esta sección toma como referencia a (De la Fuente, 2011).

En el ACP se tiene una muestra de tamaño n y con p variables X_1, X_2, \dots, X_p correlacionadas entre sí, para obtener a partir de ellas un número $k \leq p$ variables incorrelacionadas Z_1, Z_2, \dots, Z_k que sean una combinación lineal de las variables iniciales y que expliquen la mayor parte de su variabilidad, la primera componente principal se expresa como $Z_{1i} = u_{11}x_{1i} + u_{12}x_{2i} + \dots + u_{1p}x_{pi}$, para n observaciones las componentes principales se expresan de la siguiente forma:

$$\mathbf{Z}_1 = \mathbf{X}\mathbf{u}_1 = \begin{bmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1n} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1p} \end{bmatrix} \quad (7)$$

La media de Z_1 es cero, significa que $E(Z_1) = E(Xu_1) = E(X)u_1 = 0$ y la varianza de Z_1 será:

$$V(Z_1) = \frac{1}{n} \sum_{i=1}^n Z_{1i}^2 = \frac{1}{n} Z_1' Z_1 = \frac{1}{n} u_1' X' X u_1 = u_1' \left[\frac{1}{n} X' X \right] u_1 = u_1' V u_1 \quad (8)$$

Si las variables están expresadas en desviaciones respecto a la media, la expresión $\frac{1}{n} X' X$ (matriz de inercia), es la matriz de covarianzas muestral a la que se denomina V en el caso general y para variables tipificadas $\frac{1}{n} X' X$ es la matriz de correlaciones R (De la Fuente, 2011).

La primer componente principal trata de hallar Z_1 maximizando $V(Z_1) = u_1'Vu_1$ con la restricción $\sum_{j=1}^p u_1'u_1 = 1$, para resolver este problema de maximización se utiliza los multiplicadores de Lagrange, mediante la función lagrangiana siguiente:

$$L = u_1'Vu_1 - \lambda(u_1'u_1 - 1) \quad (9)$$

Derivando respecto a u_1 e igualando a cero:

$$\frac{\partial L}{\partial u_1} = 2Vu_1 - 2\lambda u_1 = 0 \rightarrow (V - \lambda I)u_1 = 0 \quad (10)$$

La ecuación anterior es un sistema homogéneo en u_1 , el cual tiene solución si, y solo si, $|V - \lambda I| = 0$, la expresión anterior tiene n raíces $\lambda_1, \lambda_2, \dots, \lambda_n$, ordenadas de mayor a menor $\lambda_1 > \lambda_2 > \dots > \lambda_n$. Si la ecuación $(V - \lambda I)u_1 = 0$ se tiene:

$$u_1'(V - \lambda I)u_1 = 0 \rightarrow u_1'Vu_1 = \lambda \rightarrow V(Z_1) = u_1'Vu_1 = \lambda \quad (11)$$

Tomando λ_1 como valor propio de V y tomando u_1 como su vector propio asociado normalizado $u_1'u_1 = 1$, se define el vector de ponderaciones que se aplica a las variables originales para obtener la primera componente principal, que se expresa como sigue:

$$Z_1 = Xu_1 \quad (12)$$

Las posteriores componentes principales, se expresan como una combinación lineal de las variables iniciales, por ejemplo, la componente principal dos, para el conjunto de n observaciones se expresa así:

$$\mathbf{Z}_2 = \mathbf{X}\mathbf{u}_2 = \begin{bmatrix} Z_{21} \\ Z_{22} \\ \vdots \\ Z_{2n} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} \begin{bmatrix} u_{21} \\ u_{22} \\ \vdots \\ u_{2p} \end{bmatrix} \quad (13)$$

Luego de haber resuelto el problema de maximización de la ecuación $V(Z_2) = u_2'Vu_2 = \lambda$, se toma el segundo valor propio de λ que es λ_2 .

De esta forma se obtiene la componente h –ésima la cual se define como $\mathbf{Z}_h = \mathbf{X}\mathbf{u}_h$, donde \mathbf{u}_h es el vector propio de V , asociado a su h –ésimo valor más grande.

La varianza de cada componente es igual al valor propio de la matriz V al que va asociada y la medida de la variabilidad de las variables originales es la suma de las varianzas, las cuales aparecen en la diagonal de la matriz de varianzas-covarianzas V y se expresa de la siguiente forma:

$$\sum_{h=1}^p V(X_h) = \text{traza}(V) \quad (14)$$

Como V es una matriz diagonal simétrica y tomando en cuenta la teoría de diagonalización de matrices, existe una matriz ortogonal $P(P^{-1} = P')$ tal que $PVP' = D$, donde D es la matriz diagonal con los valores propios de V ordenados en la diagonal principal de mayor a menor.

$$\text{traza}(PVP') = \text{traza}(D) = \sum_{h=1}^p \lambda_h = \text{traza}(VI) = \text{traza}(V) = \sum_{h=1}^p V(Z_h) \quad (15)$$

Es así que, la inercia total de la nube de puntos o la suma de las varianzas de las variables es igual a la suma de las varianzas de los componentes principales e igual a la suma de los valores propios de la matriz de varianzas-covarianzas muestrales V .

La de representatividad de los datos o porcentaje de inercia explicada por la componente principal h –ésima viene dado por:

$$\frac{\lambda_h}{\sum_{h=1}^p \lambda_h} = \frac{\lambda_h}{\text{traza}(V)} \quad (16)$$

2.3. Diseño Factorial.

2.3.1. Diseño Completamente al azar DCA.

Es un diseño que se utiliza para comparar dos o más tratamientos en los cuales se consideran los tratamientos y el error aleatorio como fuentes de variabilidad, se llama completamente al azar por que las corridas experimentales son hechas aleatoriamente, de manera que los efectos ambientales o temporales se vayan distribuyendo en todas las corridas (Pulido, 2008).

En el diseño completamente al azar, las observaciones Y_{ij} se pueden representar mediante el modelo estadístico lineal:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (17)$$

Donde μ es la media global, es el parámetro de escala común a todos los tratamientos, τ_i ; es el parámetro que mide el efecto del tratamiento i y ε_{ij} es e error atribuible a la medición Y_{ij} .

2.3.2. Análisis de Varianza ANOVA.

Esta sección utiliza como referencia a (Pulido, 2008), (Montgomery, 2013) y a (Kuehl & Kuehl, 2000).

La inferencia estadística tiene como objetivo hacer afirmaciones básicas acerca de la población con base a su información poblacional o muestral. La inferencia estadística utiliza la prueba de hipótesis mediante un estadístico, los más utilizados son la distribución normal estándar, la T de Student, ji-cuadrado y F.

El **intervalo de confianza** trata de encontrar dos números L y U, tales que el parámetro m se encuentre entre ellos con una probabilidad de $(1 - \alpha)$.

$$P(L \leq \mu \leq U) = 1 - \alpha \quad (18)$$

También se puede encontrar intervalos de confianza para cualquier parámetro.

Para pruebas cuando el objetivo es comparar tratamientos se utiliza el diseño completamente al azar (DCA), el cual considera dos fuentes de variabilidad: los tratamientos y el error aleatorio.

Para representar de manera abreviada las cantidades numéricas de los experimentos Y_{ij} , con los tratamientos i y las observaciones j , se utiliza la notación de puntos, las cantidades de interés son:

$Y_{i.}$ = Suma de las observaciones del tratamiento i .

$\bar{Y}_{i.}$ = Media de las observaciones del i -ésimo tratamiento.

$Y_{..}$ = Suma total de las observaciones, en todos los tratamientos.

$\bar{Y}_{..}$ = Media de las observaciones, en todos los tratamientos.

El punto significa la sumatoria sobre el subíndice correspondiente.

El análisis de varianza (ANOVA) es la técnica central en el análisis de datos experimentales. Esta técnica separa la variación total en las partes con que contribuyen cada fuente de variación en el experimento. En el caso del DCA se separa la variabilidad debida a los tratamientos y al error. Cuando la variabilidad de los tratamientos predomina sobre la variabilidad del error, se concluye que los tratamientos tienen efecto o que las medias son diferentes y cuando las variabilidades de los tratamientos contribuyen igual o menos que el error, se concluye que las medias son iguales; Una medida de la variabilidad es la suma de los cuadrados dada por:

$$SC_T = \sum_{i=1}^k n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Y}_{ij} - \bar{Y}_{i.})^2 = SC_{TRAT} + SC_E \quad (19)$$

Donde el primer término es la suma de los cuadrados de los tratamientos que mide la variación entre tratamientos y el segundo término es la suma de los cuadrados del error, que mide la variación dentro de los tratamientos. Como hay un total de $N = \sum_{i=1}^k n_i$ observaciones, la SC_T tiene $N - 1$ grados de libertad, hay k factores de interés, así que SC_{TRAT} tiene $k - 1$ y SC_E tiene $N - k$ grados de libertad, respectivamente, al dividir la suma de cuadrados para su grados de libertad se obtiene sus cuadrados medios, que están dados por:

$$CM_{TRAT} = \frac{SC_{TRAT}}{k - 1} \quad y \quad CM_E = \frac{SC_E}{N - k} \quad (20)$$

Los valores esperados de los cuadrados medios están dados por:

$$E(CM_E) = \sigma^2 \quad y \quad E(CM_{TRAT}) = \sigma^2 + \frac{\sum_{i=1}^k n_i \tau_i^2}{N - k} \quad (21)$$

Cuando la hipótesis nula es verdadera, los cuadrados $E(CM_E)$ y $E(CM_{TRAT})$ estiman la varianza, dado que el término $\frac{\sum_{i=1}^k n_i \tau_i^2}{N - k}$ sería igual a cero; De los cuadrados medios se obtiene el operador F_0 que sirve para aceptar o rechazar la hipótesis, así:

$$F_0 = \frac{CM_{TRAT}}{CM_E} \quad (22)$$

Suponiendo que la hipótesis nula sea verdadera y siguiendo una distribución F con $(k - 1)$ grados de libertad en el numerador y $(N - k)$ grados de libertad en el denominador.

La validez de un análisis de varianza ANOVA, se ve reflejado por el cumplimiento de los supuestos del modelo, que son: normalidad, varianza constante e independencia; Los supuestos se representan por el error ε en la ecuación (17). Cuando los supuestos se cumplen,

los residuos e_{ij} se pueden ver como una distribución normal aleatoria con varianza constante y media cero. El termino residual e_{ij} se define como la diferencia entre la respuesta observada Y_{ij} y la respuesta predicha por el modelo \hat{Y}_{ij} , esto permite diagnosticar la calidad del modelo.

Cuando se realiza el ANOVA y este resulta significativo se procede a estimar el modelo ajustado, dado por:

$$\hat{Y}_{ij} = \hat{\mu} + \hat{\tau}_i \quad (23)$$

Donde \hat{Y}_{ij} es la respuesta predicha, $\hat{\mu}$ es la media global estimada y $\hat{\tau}_i$ es el efecto estimado del tratamiento i , el error desaparece, ya que se estima que sea cero. Como la media global se estima con $\bar{Y}_{..}$ y el efecto del tratamiento con $\bar{Y}_{i.} - \bar{Y}_{..}$, el modelo ajustado del DCA queda como:

$$\hat{Y}_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) = \bar{Y}_{i.} \quad (24)$$

Esta es la respuesta predicha por el modelo ara cada observación Y_{ij} y el residuo está dado por:

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i.} \quad (25)$$

Los supuestos del modelo lineal descrito en (17) son que los e_{ij} siguen una distribución normal con media cero, los e_{ij} son independientes entre sí y que para cada tratamiento los residuos tienen la misma varianza σ^2 .

2.3.3. Diseños Factoriales.

El diseño de factores, busca estudiar el efecto de varios factores, ya sean cualitativos o cuantitativos, sobre una o varias respuestas y se debe elegir al menos de niveles de prueba

ara cada factor. Con el diseño factorial completo se corren todas las posibles combinaciones aleatoriamente, con los niveles de factores a estudiarse.

El arreglo factorial o matriz de diseño, es el conjunto de posibles tratamientos que pueden formarse considerando todas las combinaciones de los niveles de los factores. Si un diseño tienen dos factores, cada uno con dos niveles, se trata de un diseño factorial $2 \times 2 = 2^2$, de esta manera un diseño factorial 2^k tiene k factores, cada uno con dos niveles.

El **efecto de un factor** es el cambio observado en l variable de respuesta debido a un cambio en el valor del factor, el **efecto principal** es la respuesta promedio observada en el nivel alto de un factor, menos la respuesta en el nivel bajo del mismo factor y el **efecto de interacción** es cuando dos factores interactúan significativamente sobre la respuesta, cuando el efecto del uno depende del nivel en el que está el otro.

Para el **diseño factorial de dos factores** A, B con $(a, b) \geq 2$ niveles de prueba, con ellos se puede construir un diseño de $a \times b$ tratamientos, los más comunes son los diseños factoriales de 2^2 y de 3^2 , el modelo estadístico par este tipo diseño es:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}; \quad (26)$$

$$i = 1, 2, \dots, a; j = 1, 2, \dots, b; k = 1, 2, \dots, n$$

donde μ es la media general, α_i es efecto debido al i -ésimo nivel del factor A , β_j es el efecto del j -ésimo nivel del factor B , $(\alpha\beta)_{ij}$ es el efecto de la interacción en la combinación ij y ε_{ijk} es el error aleatorio que sigue una distribución normal con varianza constante y media cero y l las réplicas del experimento

Para que la estimación de los parámetros del modelo en (26), sea única, se debe tomar en cuenta las siguientes restricciones:

$$\sum_{i=1}^a \alpha_i = 0, \sum_{j=1}^b \beta_j = 0, \sum_{i=1}^a \sum_{j=1}^b (\alpha\beta)_{ij} = 0 \quad (27)$$

Se puede realizar un resumen de los resultados obtenidos, que se llama tabla ANOVA, y se detalla a continuación:

Tabla 1
ANOVA para el diseño factorial $a \times b$

FV	SC	GL	CM	F0	Valor-p
Efecto A	SC_A	$a - 1$	CM_A	$\frac{CM_A}{CM_E}$	$P(F > F_0^A)$
Efecto B	SC_B	$b - 1$	CM_B	$\frac{CM_B}{CM_E}$	$P(F > F_0^B)$
Efecto AB	SC_{AB}	$(a - 1)(b - 1)$	CM_{AB}	$\frac{CM_{AB}}{CM_E}$	$P(F > F_0^{AB})$
Error	SC_E	$ab(n - 1)$	CM_E		
Total	SC_T	$abn - 1$			

Fuente: (Pulido, 2008)

El **diseño factorial de tres factores** (A, B y C), con niveles de prueba (a, b y c), corresponden a un diseño factorial de $a \times b \times c$, los más comunes de este tipo son los diseños factoriales 2^3 y 3^3 y corresponden al modelo estadístico representado a continuación:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl} \quad (28)$$

$$i = 1, 2, \dots, a; j = 1, 2, \dots, b; k = 1, 2, \dots, c; l = 1, 2, \dots, n$$

donde μ es la media general, α_i es efecto debido al i -ésimo nivel del factor A , β_j es el efecto del j -ésimo nivel del factor B , y γ_k es el efecto del nivel k -ésimo en el factor C ; $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$ y $(\beta\gamma)_{jk}$ representan efectos de interacción dobles en los niveles ij, ik, jk ,

respectivamente, y $(\alpha\beta\gamma)_{ijk}$ es el efecto de interacción triple en la combinación o punto ijk ; ε_{ijkl} representa el error aleatorio en la combinación $ijkl$ y l son las repeticiones del experimento, la tabla de resumen ANOVA se detalla a continuación:

Tabla 2
ANOVA para el diseño factorial $a \times b \times c$

FV	SC	GL	CM	F0	Valor-p
Efecto A	SC_A	$a - 1$	CM_A	$\frac{CM_A}{CM_E}$	$P(F > F_0^A)$
Efecto B	SC_B	$b - 1$	CM_B	$\frac{CM_B}{CM_E}$	$P(F > F_0^B)$
Efecto C	SC_C	$c - 1$	CM_C	$\frac{CM_C}{CM_E}$	$P(F > F_0^C)$
Efecto AB	SC_{AB}	$(a - 1)(b - 1)$	CM_{AB}	$\frac{CM_{AB}}{CM_E}$	$P(F > F_0^{AB})$
Efecto AC	SC_{AC}	$(a - 1)(c - 1)$	CM_{AC}	$\frac{CM_{AC}}{CM_E}$	$P(F > F_0^{AC})$
Efecto BC	SC_{BC}	$(b - 1)(c - 1)$	CM_{BC}	$\frac{CM_{BC}}{CM_E}$	$P(F > F_0^{BC})$
Efecto ABC	SC_{ABC}	$(a - 1)(b - 1)(c - 1)$	CM_{ABC}	$\frac{CM_{ABC}}{CM_E}$	$P(F > F_0^{ABC})$
Error	SC_E	$abc(n - 1)$	CM_E		
Total	SC_T	$abcn - 1$			

Fuente: (Pulido, 2008)

Para poder escoger el mejor modelo de regresión lineal se utiliza el ANOVA para determinar la significancia de cada uno de los factores, así como el efecto de la interacción de factores.

Una herramienta para poder escoger el mejor modelo es el criterio de Akaike (AIC), el cual es está basado en información estadística, inicialmente fué desarrollado para series temporales (Akaike, 1974) y posteriormente para análisis factorial (Wagenmakers & Farrell, 2004).

El AIC penaliza el exceso de parámetros ajustados y es un estimador de $E[\ln(f(X|\theta))]$, esperanza de la log-verosimilitud o negantropía, que se expresa como sigue:

$$AIC(k) = -2 \ln \mathcal{L}[\hat{\theta}(k)] + 2k, \quad (29)$$

Donde, $\mathcal{L}[\hat{\theta}(k)]$ es la función de verosimilitud de las observaciones, $\hat{\theta}(k)$ es la estimación máxima del vector de parámetros θ , y k es el número de parámetros estimados dentro del modelo (Caballero Díaz, 2011).

Luego de haber hecho varias iteraciones y haber escogido un modelo es importante realizar el análisis de residuos y la calidad del modelo, mediante la comprobación de los supuestos de normalidad, varianza constante, independencia y la ausencia de observaciones atípicas.

Para verificar la correlación lineal entre dos variables se puede utilizar el coeficiente de correlación de Pearson, de Spearman o de Kendall, siendo las últimas pruebas no paramétricas por lo que nos centraremos en la prueba de correlación de Spearman, la cual se define en términos variación conjunta de las variables X y Y o covarianza (Fallas, 2012), y el estadístico r , esta dado por:

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S_X} \right) \left(\frac{Y_i - \bar{Y}}{S_Y} \right) \quad (30)$$

Donde:

$\left(\frac{X_i - \bar{X}}{S_X} \right)$ son los valores estandarizados de X

$\left(\frac{Y_i - \bar{Y}}{S_Y} \right)$ son los valores estandarizados de Y

S_X es la desviación estándar de X

S_Y es la desviación estándar de Y .

El estadístico r mide la intensidad y la dirección de la correlación lineal, su valor fluctúa entre 1 y -1, siendo 1 una correlación perfecta positiva o directa, y -1 correlación perfecta negativa o inversa (Laguna, 2011).

Para verificar la normalidad se puede utilizar un procedimiento gráfico que consiste en graficar los residuos en una gráfica de probabilidad normal, los residuos tienden a quedar alineados en una línea recta, como, por ejemplo:

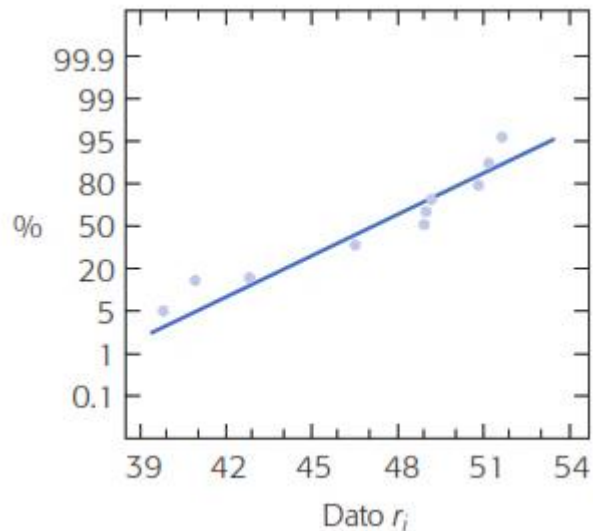


Figura 1. Gráfica alineada de probabilidad normal de los residuos

Fuente: (Pulido, 2008)

También existen varios métodos analíticos como la ji-cuadrada para bondad de ajuste, la prueba de Shapiro-Wilks, la prueba de Anderson-Darling, la prueba de Kolmogorov-Smirnov o la misma en su versión corregida llamada prueba de Lilliefors, siendo las más importantes la prueba de Shapiro-Wilks y la prueba de Lilliefors (Romero, 2016).

En la prueba de Shapiro-Wilks para la normalidad se considera una muestra aleatoria de datos x_1, x_2, \dots, x_n que son generadas por cierta distribución $F(x)$, la cual se quiere verificar si fue generada por un proceso normal, mediante las hipótesis estadísticas:

H_0 : Los datos proceden de una distribución normal (F(x) es normal).

H_A : Los datos no proceden de una distribución normal (F(x) no es normal).

Se calcula mediante el estadístico:

$$W = \frac{1}{(n-1)S^2} \left[\sum_{i=1}^k a_i (X_{(n-1+1)} - X_{(i)}) \right]^2 \quad (31)$$

Donde S^2 es la varianza muestral, tomando un nivel de significancia α de las tablas y si el valor del estadístico es mayor que su valor crítico, se rechaza la normalidad de los datos.

(Shapiro & Wilk, 1965), restringen el tamaño de la muestra a menos de 50.

El test de Lilliefors, nace a raíz de que la prueba de Kolmogrov-Smirnov (K-S) utiliza la media y la desviación estándar de la población, en la práctica la mayoría de veces se trabaja con muestras, es así que Lilliefors tabula el estadístico K-S para una muestra donde se desconoce la media y la desviación estándar. Lilliefors define el estadístico D como sigue:

$$D = \max_x |F^*(X) - S_n(X)| \quad (32)$$

Donde $S_n(X)$ es una muestra de la Función de distribución acumulativa y $F^*(X)$ es la función de distribución normal acumulativa con $\mu = \bar{X}$ (Lilliefors, 1967), y se utiliza mediante las hipótesis estadísticas:

H_0 : Los datos proceden de una distribución normal (F(x) es normal).

H_A : Los datos no proceden de una distribución normal (F(x) no es normal).

Para verificar el supuesto de varianza constante se puede graficar los predichos versus los residuos \hat{Y}_{ij} vs ε_i , siendo que \hat{Y}_{ij} va en el eje horizontal y ε_i en el vertical. Si los puntos se distribuyen dentro de una banda horizontal sin ningún patrón claro, es muestra de que los tratamientos tienen igual varianza, por ejemplo, la Figura 2, tienen un patrón de corneta, por lo que se concluye que no tiene una varianza constante.

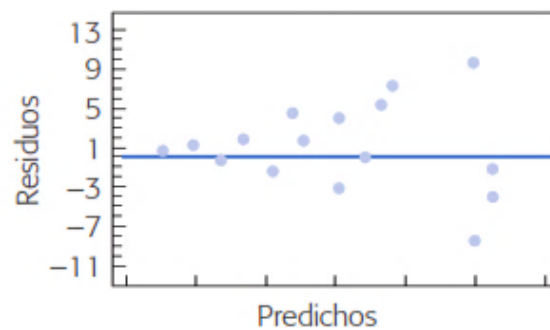


Figura 2. Grafica de predichos vs residuos

Fuente: (Pulido, 2008)

La grafica de factor en el eje horizontal versus residuos en el eje vertical, nos revela si hay varianza constante, si los puntos se distribuyen uniformemente en los niveles, por ejemplo, en la Figura 3, hay más puntos en el nivel alto que en el bajo por lo cual no tienen una varianza constante.

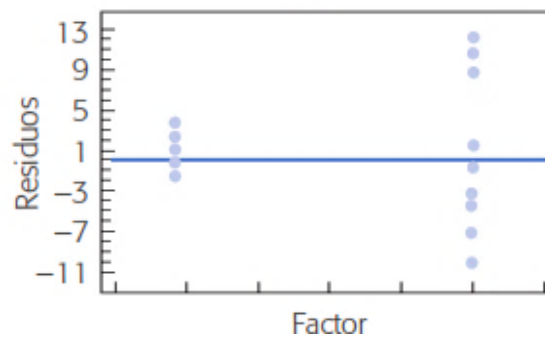


Figura 3. Grafica de factor vs residuos

Fuente: (Pulido, 2008)

La prueba analítica para verificar si la varianza es constante, se llama prueba de Bartlett para la homogeneidad de varianzas, en esta se tiene k tratamientos independientes, cada uno con distribución normal $(N(\mu_i, \sigma^2), i = 1, 2, \dots, k)$, donde las varianzas son desconocidas y se requiere verificar su igualdad, mediante las siguientes hipótesis:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2 \quad (33)$$

$$H_1: \sigma_i^2 \neq \sigma_j^2 \text{ para algún } i \neq j$$

El estadístico está dado por:

$$\chi_0^2 = 2.3036 \frac{(N - k) \log_{10} S_p^2 - \sum_{i=1}^k (n_i - k) \log_{10} S_i^2}{1 + \frac{1}{3(3k - 1)} (\sum_{i=1}^k (n_i - 1)^{-1} - (n - k)^{-1})} \quad (34)$$

$$S_p^2 = \frac{\sum_{i=1}^k (n - 1) S_i^2}{N - k}$$

donde S_i^2 es la varianza del tratamiento i , se rechaza H_0 cuando χ_0^2 es más grande que $\chi_{(\alpha, k-1)}^2$.

La independencia puede comprobarse si se grafica el orden en que se recolecto la información, estando en el eje horizontal el tiempo de recolección y los residuos en el eje vertical, por ejemplo, en la Figura 4 se muestra que hay una correlación entre los errores, por lo tanto, no existe independencia.

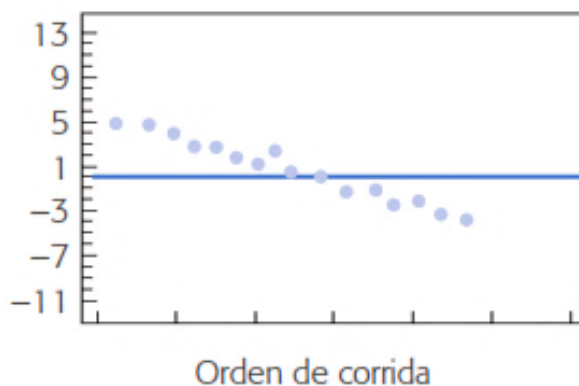


Figura 4. Orden de corrida vs residuos

Fuente: (Pulido, 2008)

Una de las pruebas analíticas para comprobar la independencia es la prueba de Durbin-Watson, en esta se define ρ como la relación que define la correlación entre parámetros consecutivos, la hipótesis será:

$$H_0: \rho = 0 \quad (35)$$

$$H_1: \rho > 0$$

El estadístico se define como:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n (e_i)^2} \quad (36)$$

Donde $e_i, i = 1, 2, \dots, n$, son los residuos ordenados cronológicamente y la decisión con respecto a la hipótesis H_0 se rechaza si $d < d_L$, H_0 no se rechaza si $d > d_U$ y no se toma una decisión si $d_L < d < d_U$, siendo d_L y d_U las cotas que se ubican en tablas.

2.4. Índice de Morán.

Esta sección hace referencia a (GeoDa, s. f.) y (Bohórquez, Álvarez, & Ceballos, 2008).

El Índice de Morán es posiblemente el indicador más utilizado para determinar la

autocorrelación espacial global y es un estadístico del producto cruzado entre la variable expresada en desviaciones de su media y la espacialidad. Para una ubicación (i, j) , donde \bar{y} es la media de la variable y , se expresa como sigue:

$$I = \frac{n}{\sum_{i=1}^n (y_i - \bar{y})^2} \frac{\sum_{i=0}^n \sum_{j=0}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=0}^n \sum_{j=0}^n w_{ij}} \quad (37)$$

donde $\sum_{i=0}^n \sum_{j=0}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})$, es la covarianza, los subíndices i y j son las diferentes unidades espaciales, y_i es el valor de la observación y w_{ij} es un elemento de la matriz de ponderaciones W . En el caso más sencillo,

$$w_{ij} = \begin{cases} 1, & \text{si la región } i \text{ y } j \text{ son adyacentes} \\ 0, & \text{caso contrario} \end{cases} \quad (38)$$

Los demás elementos de la fórmula normalizan el valor de I respecto al número de zonas en consideración, el número de adyacencias del problema y el rango de valores.

La matriz de ponderaciones o de contigüidad W donde las columnas y filas representan una región en el espacio, cada una de estas regiones se relaciona con sus regiones vecinas, una de las notaciones para construir la matriz W , es la notación binaria con 1 cuando hay presencia de contigüidad y 0 cuando no.

Existen muchas formas para definir si existe o no contigüidad, las más comunes son:

La contigüidad de torre en la cual se define $w_{ij} = 1$ para las zonas que están a la derecha, izquierda, arriba y de bajo de la zona de estudio, como se muestra en la Figura 5.

	b	
b	a	b
	b	

Figura 5. Contigüidad tipo torre

Fuente: (Bohórquez & Ceballos, 2008)

La contigüidad de alfil en la cual se define $w_{ij} = 1$ para las zonas que comparten un vértice con la zona de estudio, como se muestra en la Figura 6.

b		b
	a	
b		b

Figura 6. Contigüidad tipo alfil

Fuente: (Bohórquez & Ceballos, 2008)

La contigüidad de reina para las regiones que comparten un lado o un vértice con la zona de interés, como se muestra en la Figura 7.

b	b	b
b	a	b
b	b	b

Figura 7. Contigüidad tipo reina

Fuente: (Bohórquez & Ceballos, 2008)

Otro esquema de ponderación de pesos comúnmente utilizado se basa en la función de distancia inversa, de modo que los pesos estarán inversamente relacionados con la distancia entre las distancias i y locaciones o áreas j (Borrego, 2018), los pesos se expresan como sigue:

$$w_{ij} = \begin{cases} d_{ij}^{-\theta}, & \text{distancia } d_{ij} < d \text{ (} d > 0, \theta > 0 \text{)} \\ 0, & \text{caso contrario} \end{cases} \quad (39)$$

Moran en su artículo “Notes on Continuous Stochastic Phenomena”, propuso un estadístico de prueba para el contraste de correlación entre unidades espaciales adyacentes (Bohórquez & Ceballos, 2008), cuyo desarrollo formal presentamos a continuación:

Sea X una matriz de variables $(m \times n)$ aleatorias $x_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, n$ independientes e idénticamente distribuidas.

$$\bar{x} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n x_{ij} \quad (40)$$

Sea $z_{ij} = x_{ij} - \bar{x}$, la covarianza entre las variables z_{ij} y $z_{i,j+1}$, está dada por:

$$\text{cov}(z_{ij}, z_{i,j+1}) = E[z_{ij} - E(z_{ij})]E[z_{i,j+1} - E(z_{i,j+1})] = E[z_{ij} - E(z_{i,j+1})] \quad (41)$$

por el principio de analogía:

$$\widehat{\text{cov}}(z_{ij}, z_{i,j+1}) = \frac{\sum_i^m \sum_j^{n-1} z_{ij} z_{i,j+1}}{m(n-1)} \quad (42)$$

Siguiendo un razonamiento similar, se puede construir el coeficiente de autocorrelación espacial de primer orden, de la siguiente manera:

$$r_{11} = \left(\frac{mn}{2m - m - n} \right) \frac{\sum_i^m \sum_j^{n-1} z_{ij} z_{i,j+1} + \sum_i^{m-1} \sum_j^n z_{ij} z_{i+1,j}}{\sum_{ij} z_{ij}^2} \quad (43)$$

$$r_{11} = \left(\frac{mn}{2m - m - n} \right) \frac{A + B}{\sum_i (z_{i1}^2 + z_{i2}^2 + \dots + z_{in}^2)} \quad (44)$$

siendo:

$$A = \sum_i^m (z_{i1} z_{i2} + z_{i2} z_{i3} + \dots + z_{i,n-1} z_{in})$$

$$B = \sum_i^{m-1} (z_{i1} z_{i+1,1} + z_{i2} z_{i+1,2} + \dots + z_{in} z_{i+1,n})$$

El Índice de Morán se define a partir de la construcción anterior como:

$$I = \frac{\sum_i^m \sum_j^{n-1} z_{ij} z_{i,j+1} + \sum_i^{m-1} \sum_j^n z_{ij} z_{i+1,j}}{\sum_{ij} z_{ij}^2} \quad (45)$$

2.5. Regresión Hedónica.

Esta sección hace referencia a (Lever, s. f.) y (Olmo et al., 2007).

La Teoría de Precios Hedónicos pretende explicar el valor de un bien raíz, entendido como un conjunto de atributos o características. La forma general del modelo es la siguiente:

$$P = f(I, V, U, Z; \emptyset) \quad (46)$$

donde P es el precio del metro cuadrado construido de la vivienda, el cual está determinado por los argumentos de la función f . Se considera que esta variable dependiente corresponde al precio de la vivienda, es así que el precio del terreno se considera como una característica de las tantas que influyen sobre el precio de la vivienda.

Donde I son las características inherentes del inmueble como por ejemplo la superficie del terreno, la superficie de construcción, el número de habitaciones, equipamiento interior y exterior, calidad de los materiales de construcción, etc.

V son las características del vecindario como por ejemplo nivel socioeconómico, seguridad, etc. Hay zonas que son consideradas exclusivas o donde se asienta personas de altos ingresos económicos que afectan directamente el precio de la vivienda.

Z son las características del inmueble dentro del plano regulador (zoning) como por ejemplo la densidad de construcción. Por ejemplo, a mayor densidad poblacional aumenta los índices de delincuencia, sin embargo, existe un predominio de los factores positivos como el crecimiento económico, social y cultural, los cuales influyen en el precio (Figueroa & Lever, 1992).

U es la ubicación del bien raíz, por ejemplo, si el predio está cerca de zonas de influencia económica, social y cultural, el precio se ve afectado positivamente, y w es el conjunto de parámetros que acompaña a cada atributo y que constituyen los precios sombra de cada característica del inmueble.

La forma estimable del modelo hedónico se puede expresar de la siguiente manera:

$$P_i = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_mx_{mi} + m_i \quad (47)$$

donde $k = 1, 2, \dots, m; i = 1, 2, \dots, n$; las m variables de x_{ki} corresponden a los argumentos de la función f incorporados en los grupos de atributos en (41), los coeficientes b_k corresponden a los parámetros incluidos en el set w , y m es un error aleatorio.

La forma funcional de f no necesariamente es lineal, según la experiencia de (Lever, s. f.), la relación entre el precio y las variables explicativas, suelen responder a una forma logarítmica.

El modelo de regresión hedónica debe cumplir los supuestos de normalidad, varianza constante e independencia, las pruebas para verificar los supuestos se detallaron en la sección anterior.

CAPITULO III

METODOLOGÍA

3.1. Recolección de Datos.

La recolección de datos se hace del portal plusvalía.com.ec y de olx.com.ec, entre el 1 de enero y el 20 de marzo del 2019, no habiendo ningún impedimento legal de usar esta información (olx.com.ec, 2019) y (plusvalia.com.ec, 2019), la información se filtra para que aparezcan las publicaciones de casas que han sido hechas desde el 1 de enero del 2019 hasta el 20 de marzo del 2019 y que han sido hechas en el cantón Rumiñahui.

Entre los dos portales se tiene un aproximado de 700 publicaciones, de las cuales se excluyen los elementos que se repiten en las plataformas y se deja uno solo, se obtiene 585 publicaciones no repetidas, de las cuales se procede a verificar que tengan correcta georeferenciación, que se encuentran dentro del cantón Rumiñahui, que tengan sus datos completos y que no existen datos aberrantes, luego de realizar este filtro en todas las publicaciones se obtiene un total de 296 publicaciones validadas, la cuales tienen los siguientes variables:

- Identificador de propiedad (número entre 1 y 296)
- Superficie Total
- Superficie Cubierta
- Número de habitaciones o cuartos dormitorios
- Número de baños
- Número de estacionamientos
- Antigüedad en años

- Localización georreferenciada (longitud y latitud)
- Precio
- Número de pisos de la construcción
- Pertenencia a condominio o conjunto cerrado

La tabla de datos final, se puede observar en el **¡Error! No se encuentra el origen de la referencia.**

Paralelamente se extrae la georeferenciación dentro del cantón de las Unidades Educativas con bachillerato, más representativas tomando como referencia la información en (Montoya & Erazo, 2013), verificando la información mediante el portal www.infoescuelas.com/ecuador y mediante llamadas telefónicas, para obtener las Unidades Educativas con bachillerato, las cuales se detallan a continuación:

Tabla 3
Georreferencias de las Unidades Educativas representativas del cantón Rumiñahui

Número	Nombre	Parroquia	Ubicación
1	RUMIÑAHUI	COTOGCHOA	-0.331086, -78.445032
2	RAMON GONZALEZ ARTIGAS	SANGOLQUI	-0.345722, -78.430008
3	GREENYARD	SANGOLQUI	-0.329447, -78.452125
4	EL CAMINO	SANGOLQUI	-0.325899, -78.453554
5	LEV VYGOTSKY	SANGOLQUI	-0.336438, -78.456317
6	MAHANAYM	SANGOLQUI	-0.334374, -78.443863
7	GALILEO GALILEI	SANGOLQUI	-0.364379, -78.431919
8	FRANCISCANO DE LA INMACULADA	SANGOLQUI	-0.328397, -78.450672
9	JUAN DE SALINAS	SANGOLQUI	-0.336648, -78.438129
10	JOSE GABRIEL NAVARRO	SANGOLQUI	-0.329471, -78.428712
11	DR TELMO HIDALGO DIAZ	SANGOLQUI	-0.330482, -78.468827

CONTINUA

12	SIGLO XXI	SANGOLQUI	-0.290586, -78.528194
13	RUMIÑAHUI	SANGOLQUI	-0.331138, -78.445048
14	HELMUT WAHLMULLER	SANGOLQUI	-0.305323, -78.465106
15	NUEVO MUNDO EN LAS MANOS DE DIOS	SAN RAFAEL	-0.293305, -78.455886
16	COTOGCHOA	SANGOLQUI	-0.329039, -78.428706
17	LICEO JUAN MANTOVANI	SANGOLQUI	-0.338360, -78.450408
18	JACINTO JIJON Y CAAMAÑO	SANGOLQUI	-0.337849, -78.438175
19	TECNICO DEL VALLE	SANGOLQUI	-0.343163, -78.440925
20	JUAN MONTALVO	SANGOLQUI	-0.330989, -78.448908
21	LEOPOLDO MERCADO	SANGOLQUI	-0.329182, -78.444250
22	ALBERTO ACOSTA SOBERON	SANGOLQUI	-0.327345, -78.446104
23	MARIETTA DE VEINTIMILLA	SANGOLQUI	-0.327422, -78.446320
24	LA LLAMA DEL VALLE	SANGOLQUI	-0.328014, -78.440527
25	SANTA ANA	SANGOLQUI	-0.326268, -78.447606
26	MARQUES DE SELVA ALEGRE	SANGOLQUI	-0.342192, -78.440139
27	LA PRADERA	SANGOLQUI	-0.315106, -78.450820
28	SANTO TOMAS DE AQUINO	SAN PEDRO DE TABOADA	-0.316680, -78.471210
29	MADRE DE LA DIVINA GRACIA	SAN RAFAEL	-0.300810, -78.456373
30	JUAN PIO MONTUFAR	SANGOLQUI	-0.303645, -78.456800
31	WILLIAM HAWKING	SANGOLQUI	-0.293320, -78.455827
32	LUIS A. MARTINEZ	SAN RAFAEL	-0.310757, -78.452252
33	HENRY DUNANT	SAN RAFAEL	-0.315265, -78.448091
34	GEORGE MASON	SANGOLQUI	-0.303905, -78.454193
35	SOCIEDAD EDUCATIVA FUTURO	SANGOLQUI	-0.309138, -78.450440
36	SAN RAFAEL	SANGOLQUI	-0.306626, -78.455133
37	GIOVANNI ANTONIO FARINA	SAN RAFAEL	-0.295395, -78.461299
38	LICEO DEL VALLE	SANGOLQUI	-0.324282, -78.430511

CONTINUA

39	DARIO FIGUEROA LARCO	SANGOLQUI	-0.330646, -78.443273
40	CARLOS LARCO HIDALGO	SANGOLQUI	-0.327914, -78.445561
41	JOSE MONCADA SANCHEZ	SANGOLQUI	-0.338254, -78.450940

Para obtener el listado de los establecimientos de salud se hace referencia al documento de (Cornejo, Zorrilla, Bermúdez, Estacio, & Narváez, 2013) y mediante Google Maps se obtiene la georeferenciación.

Tabla 4
Georeferenciación de los establecimientos de salud del cantón Rumiñahui

Número	Nombre	Parroquia	Ubicación
1	Subcentro de Salud de Cotogchoa	Cotogchoa	-0.368321,-78.452686
2	Hospital cantonal de Sangolquí	Sangolquí	-0.329940, -78.443927
3	Centro de Salud de Sangolquí	Sangolquí	-0.329620, -78.450359
4	Subcentro de Salud de Jatipungo	Sangolquí	-0.361537, -78.431248
5	Centro Medico de Especialidades Los Chillos	Sangolquí	-0.330513, -78.451069
6	Hospital del IESS	Sangolquí	-0.337437, -78.441064
7	Subcentro de Salud San Fernando	Sangolquí	-0.373747, -78.422913
8	Subcentro de Salud Capelo	San Rafael	-0.308162, -78.455911
9	Subcentro de Salud de Caripungo	Cotogchoa	-0.392178, -78.433438
10	Subcentro de Salud Selva Alegre	Sangolquí	-0.341490, -78.431935
11	Subcentro de Salud de Rumipamba	Rumipamba	-0.431508, -78.418439
12	Cruz Roja	Sangolquí	-0.331940, -78.450119
13	Hospital Metropolitano	San Rafael	-0.326798, -78.462565
14	Nova Clínica del Valle	Sangolquí	-0.332801, -78.448648
15	Clínica San Rafael	San Rafael	-0.305186, -78.455338
16	Clínica de Emergencias San Francisco	Sangolquí	-0.333573, -78.445609
17	Club de Leones Quito carita de Dios	Sangolquí	-0.327371, -78.450996
18	Hospital Militar Fuerte San Jorge	Sangolquí	-0.322102, -78.450493
19	Clínica Paredes	San Rafael	-0.299974, -78.456968

Para obtener los puntos de desarrollo económico se hace referencia a (Cornejo et al., 2013) y a (Montoya & Erazo, 2013), donde se describen las zonas de mayor afluencia de personas y que son sectores comerciales, la georeferenciación se la hace mediante Google Maps, se obtiene la siguiente tabla:

Tabla 5
Zonas de desarrollo económico del cantón Rumiñahui

Número	Nombre	Ubicación
1	Centro Comercial San Luis	-0.308207,-78.450674
2	Zona El Triángulo (Esquina Norte)	-0.303757, -78.454020
3	Centro de Sangolquí (GAD Municipal)	-0.330538, -78.446845

Para obtener las Instituciones de Educación Superior representativas del cantón Rumiñahui, se consultó en Google Maps, las Universidades e Institutos en la zona y se descartó las extensiones de estudios a distancia, a continuación, se muestra los resultados:

Tabla 6
Instituciones de Educación Superior representativas del cantón Rumiñahui

Numero	Nombre	Ubicación
1	Universidad de la Fuerza Armadas ESPE	-0.314581,-78.443858
2	Instituto Superior Tecnológico Rumiñahui	-0.339461, -78.440018
3	Universidad Metropolitana del Ecuador	-0.297951, -78.455474

Levantada la información se procedió a calcular el precio del metro cuadrado total y el precio del metro cuadrado cubierto, también se calculó la distancia de cada una de las viviendas hacia cada uno de las Unidades Educativas, Establecimientos de Salud, Zonas de Desarrollo Económico e Instituciones de Educación Superior presentes en las tablas anteriores, mediante la fórmula de Haversine, que es:

$$d = 2r \arcsen \left(\sqrt{\text{sen}^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \text{sen}^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \quad (48)$$

donde φ_1, φ_2 representan las latitudes y λ_1, λ_2 representan las longitudes de los puntos 1 y 2; r representa el radio ecuatorial de la tierra que es de 6378100 metros.

A continuación, se obtiene las variables de distancia mínima desde la vivienda hacia una Unidad Educativa, distancia mínima hacia un establecimiento de salud, distancia mínima hacia una zona de desarrollo económico y distancia mínima hacia una Institución de Educación Superior, para ello se utiliza:

$$d_{min} = \min(A) \quad (49)$$

siendo d_{min} la distancia mínima y $A = \{d_1, d_2, \dots, d_n\}$ con n elementos de grupo de lugares estratégicos georeferenciados en las tablas 3,4,5 o 6.

3.2. Estudio descriptivo de los datos.

Se realiza un estudio descriptivo de los datos con el fin de conocer de mejor manera la naturaleza estadística de los mismos, para ello se analiza variable por variable.

La base de datos se la divide en dos grupos uno que corresponde al 90% de los datos o 263 predios, con los cuales se realizará el análisis descriptivo y la elaboración del modelo y el 10% o 29 datos se utilizará para realizar la validación cruzada.

La estadística descriptiva se realiza en R Studio por facilidad y la familiaridad del software. Para la superficie total se obtiene la siguiente tabla y figuras:

Tabla 7
Estadística Descriptiva de la Superficie Total (m²)

<i>Superficie Total</i>	
Media	395.8
Mediana	232.00
Moda	170.00
Desviación estándar	448.02
Varianza de la muestra	200718.9
Mínimo	60.00
Máximo	4500

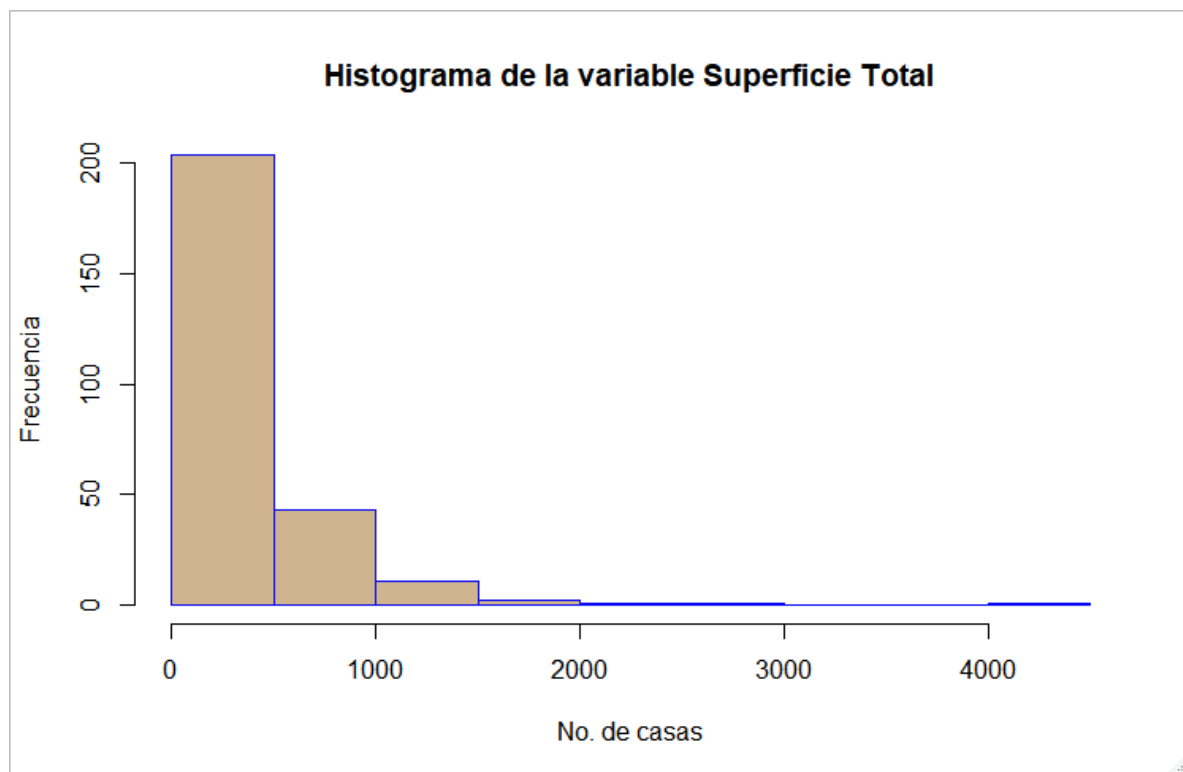


Figura 8. Histograma de la variable Superficie Total

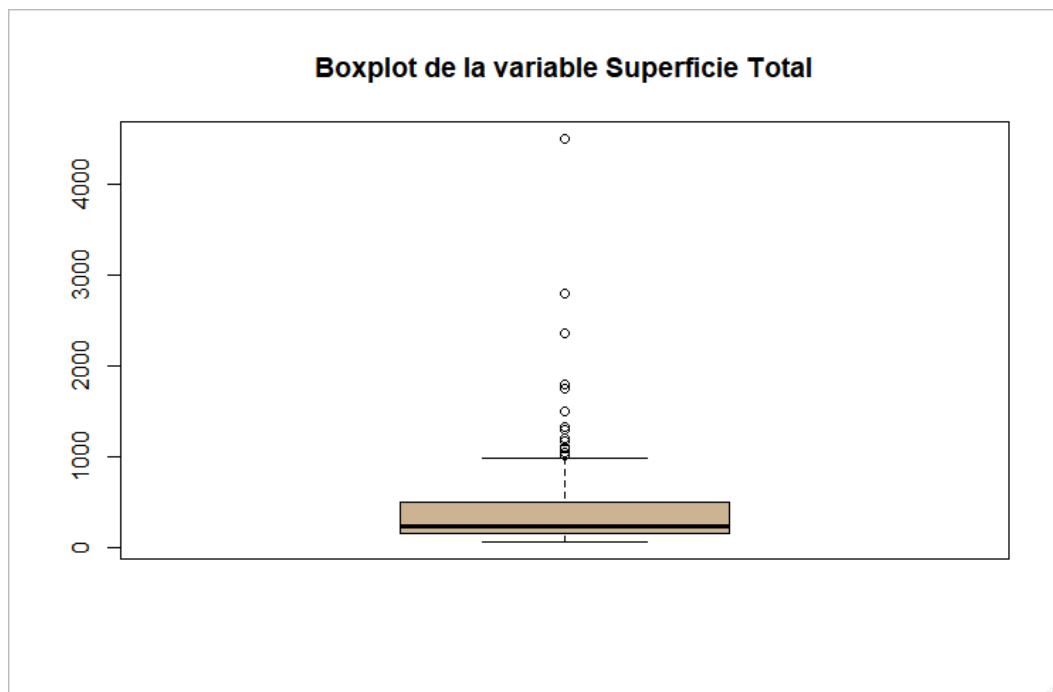


Figura 9. Diagrama de cajas de la variable Superficie Total

Se pueden observar algunos datos atípicos destacándose los tres superiores en la Figura 9. Diagrama de cajas de la variable Superficie Total, se observa que la mayoría de datos se encuentran bajo la media, el segundo cuartil se encuentra en los 232m². Es decir hay más oferta de casas con superficie de menos de 500 m², que casas grandes sobre los 5000m².

Para la superficie cubierta se obtiene la siguiente tabla y figuras:

Tabla 8
Estadística Descriptiva de la Superficie Cubierta (m²)

<i>Superficie Cubierta</i>	
<i>Media</i>	208.7
<i>Mediana</i>	170
<i>Moda</i>	120.00
<i>Desviación estándar</i>	120
<i>Varianza de la muestra</i>	14661
<i>Mínimo</i>	60.00
<i>Máximo</i>	830.00

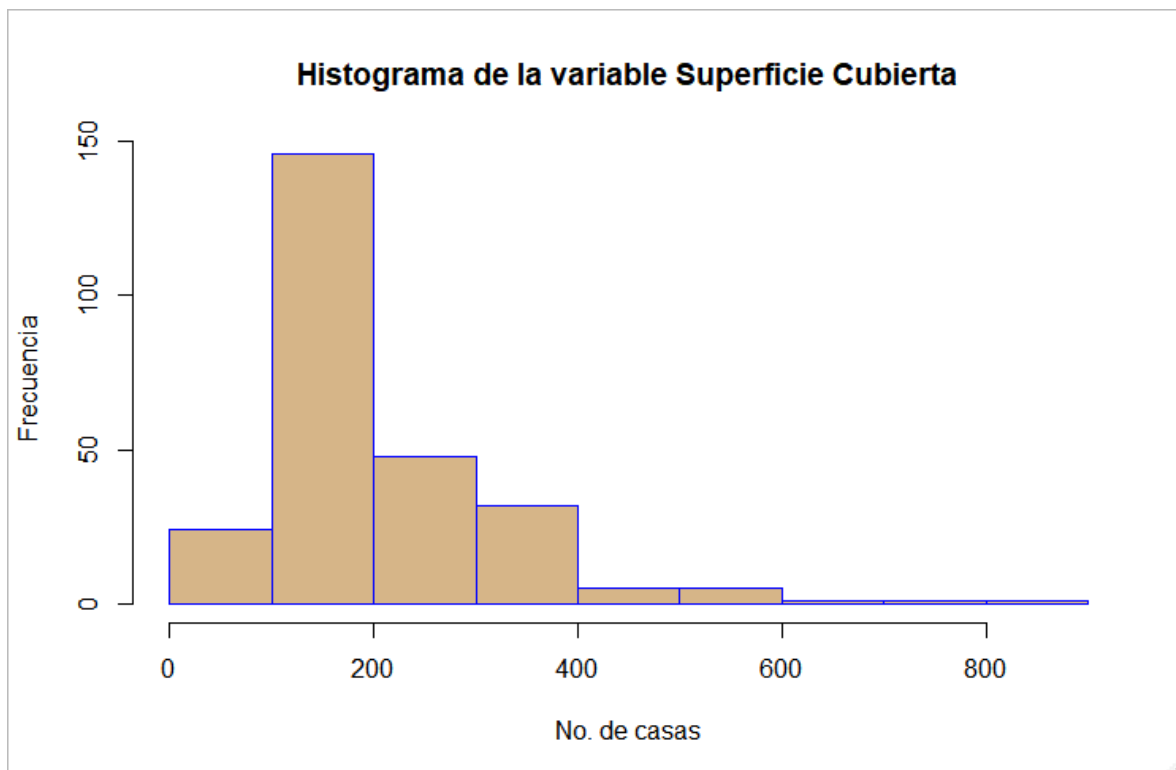


Figura 10. Histograma de la variable superficie cubierta

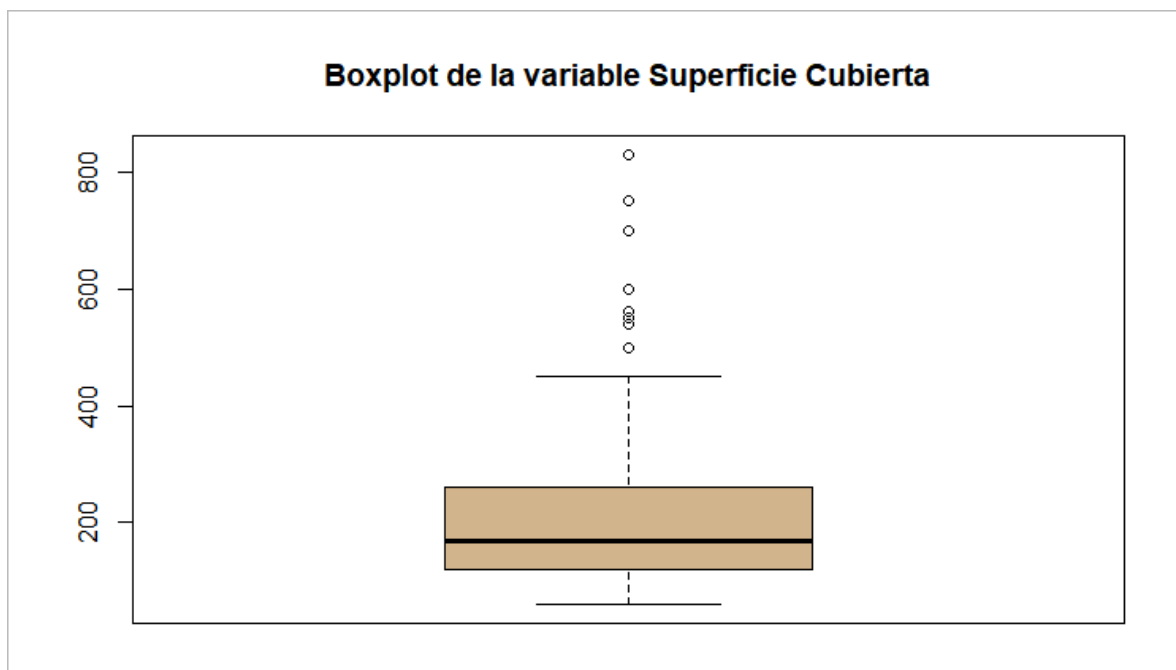


Figura 11. Diagrama de cajas de la variable Superficie Cubierta

Con respecto a la superficie cubierta se puede observar que las clases bajas tienen mayor cuenta, siendo la más popular la más popular las casas entre 100 m^2 y 200 m^2 .

Para el número de habitaciones o cuartos dormitorios, que en adelante se refiere como “cuartos”, que es una variable categórica, se obtuvieron la siguiente tabla y figuras:

Tabla 9
Estadística Descriptiva de número de cuartos

<i>Número de Habitaciones</i>	
Media	3.58
Mediana	3
Moda	3
Desviación estándar	0.98
Varianza de la muestra	0.96
Mínimo	2
Máximo	9

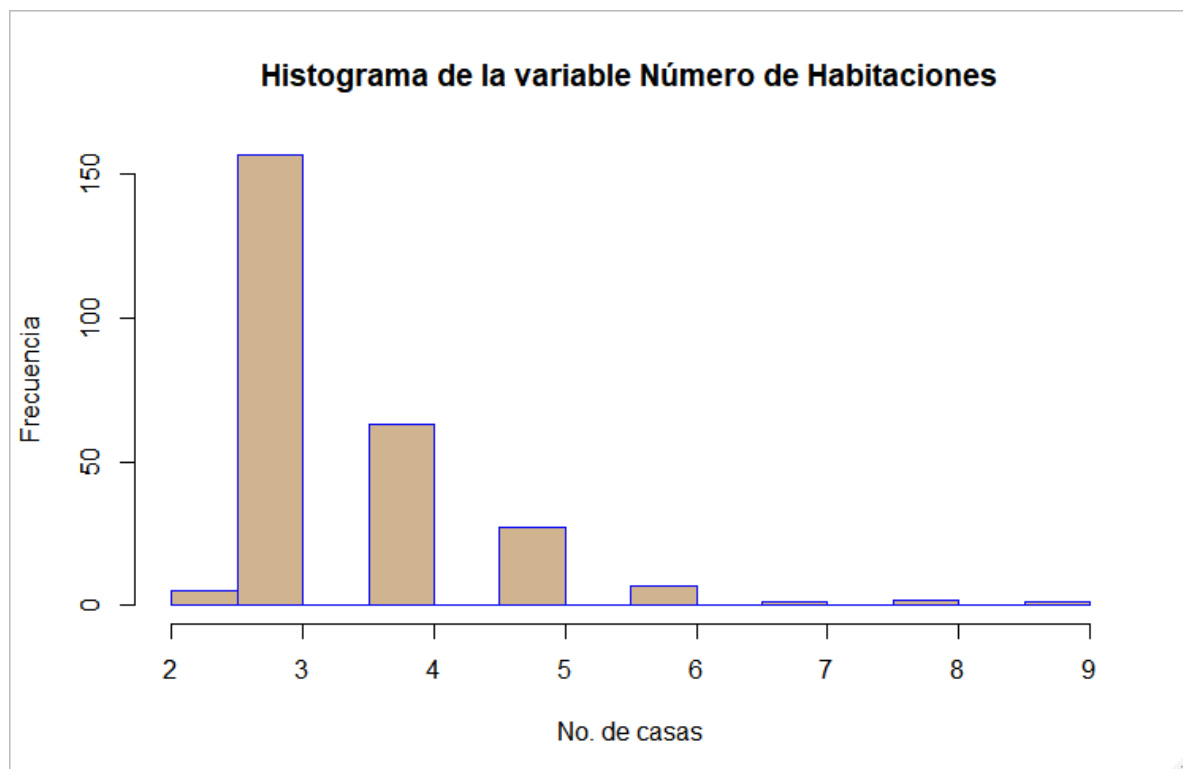


Figura 12. Histograma de la variable número de cuartos

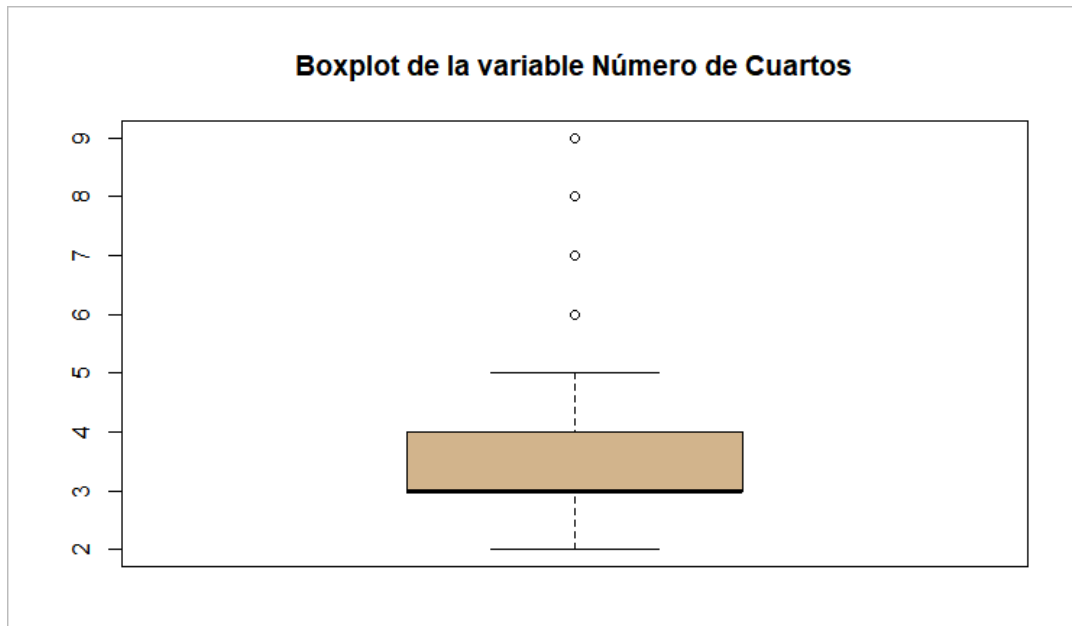


Figura 13. Diagrama de cajas de la variable número de cuartos

Se puede observar que predominan las viviendas que tienen en 3 y 4 cuartos, y se tiene 4 datos atípicos de casa con 6, 7, 8 y 9 cuartos respectivamente, ya que es muy poco común encontrar casa con 6 o más cuartos.

Para la variable categórica de número de estacionamientos, se tiene la siguiente tabla y figuras:

Tabla 10
Estadística Descriptiva de número de estacionamientos

<i>Número de Estacionamientos</i>	
Media	2.62
Mediana	2
Moda	2
Desviación estándar	1.42
Varianza de la muestra	2.02
Mínimo	0
Máximo	8

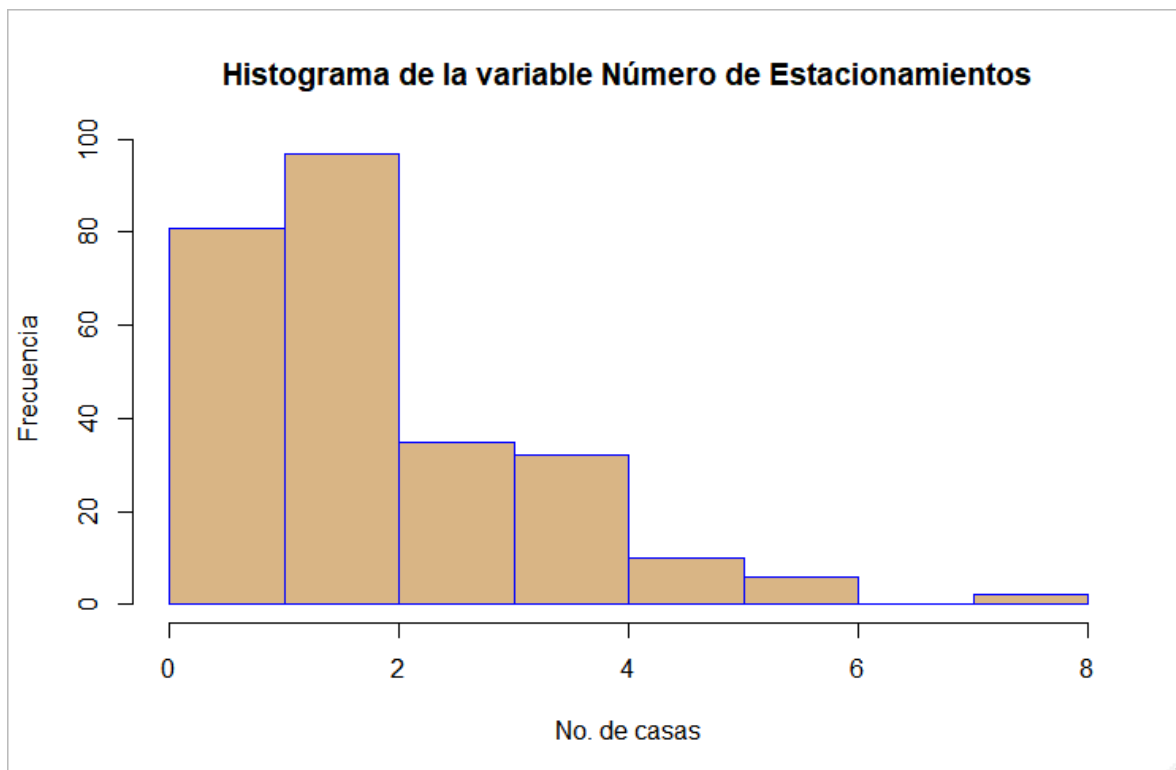


Figura 14. Histograma para la variable número de estacionamientos

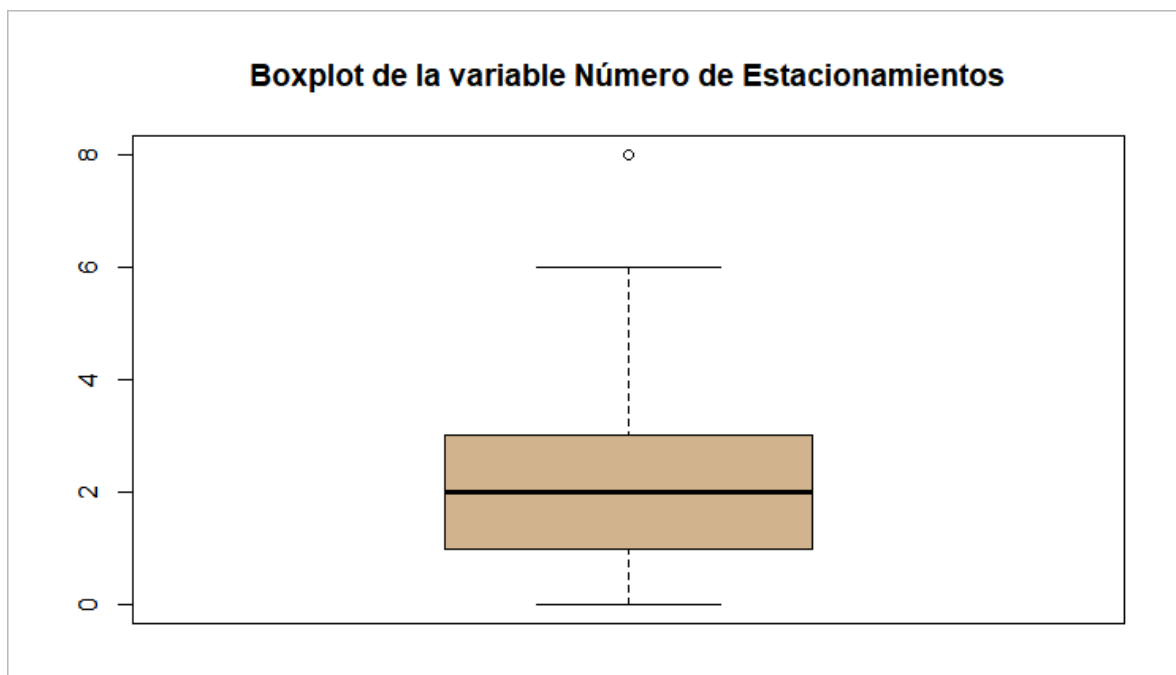


Figura 15. Diagrama de cajas de la variable número de estacionamientos

En las viviendas analizadas predomina la tenencia de 2 estacionamientos, con un dato atípico de 8 estacionamientos, del diagrama de cajas se puede observar que casi todas las casas tienen entre 1 y 3 estacionamientos.

Con respecto a la variable categórica número de baños, se obtuvo la siguiente información:

Tabla 11
Estadística Descriptiva de número de baños

<i>Número de Baños</i>	
Media	3.1
Mediana	3
Moda	2.5
Desviación estándar	1.02
Varianza de la muestra	1.05
Mínimo	1
Máximo	8

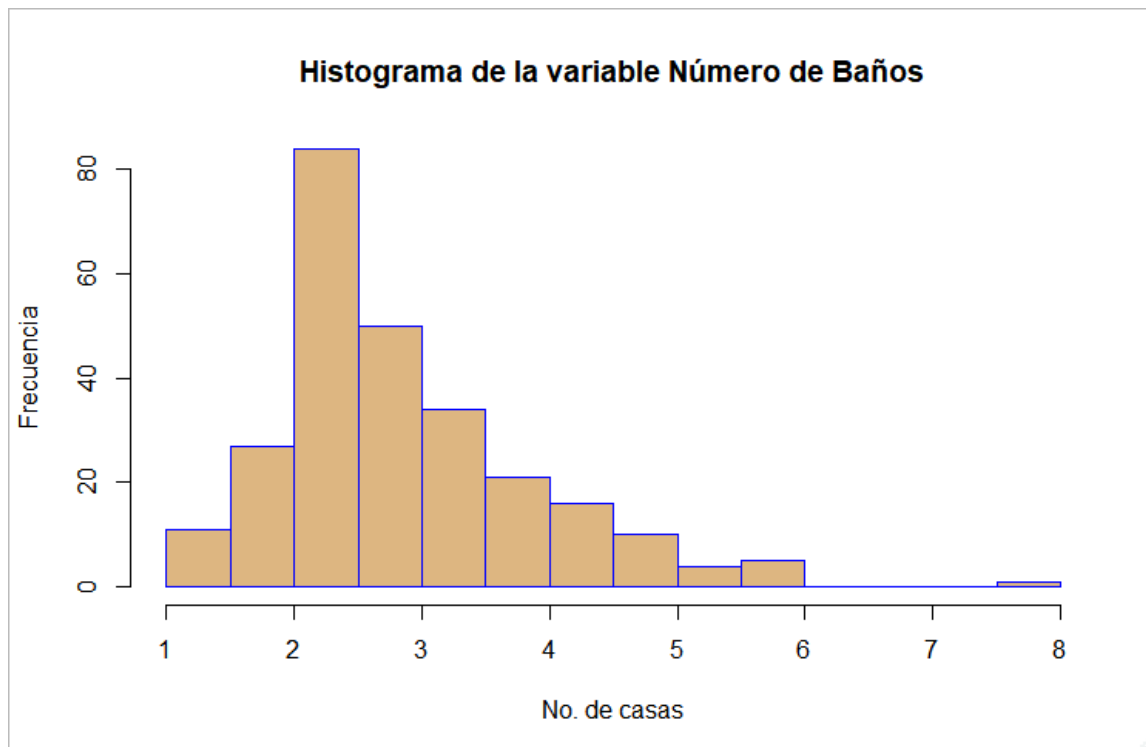


Figura 16. Histograma de la variable número de baños

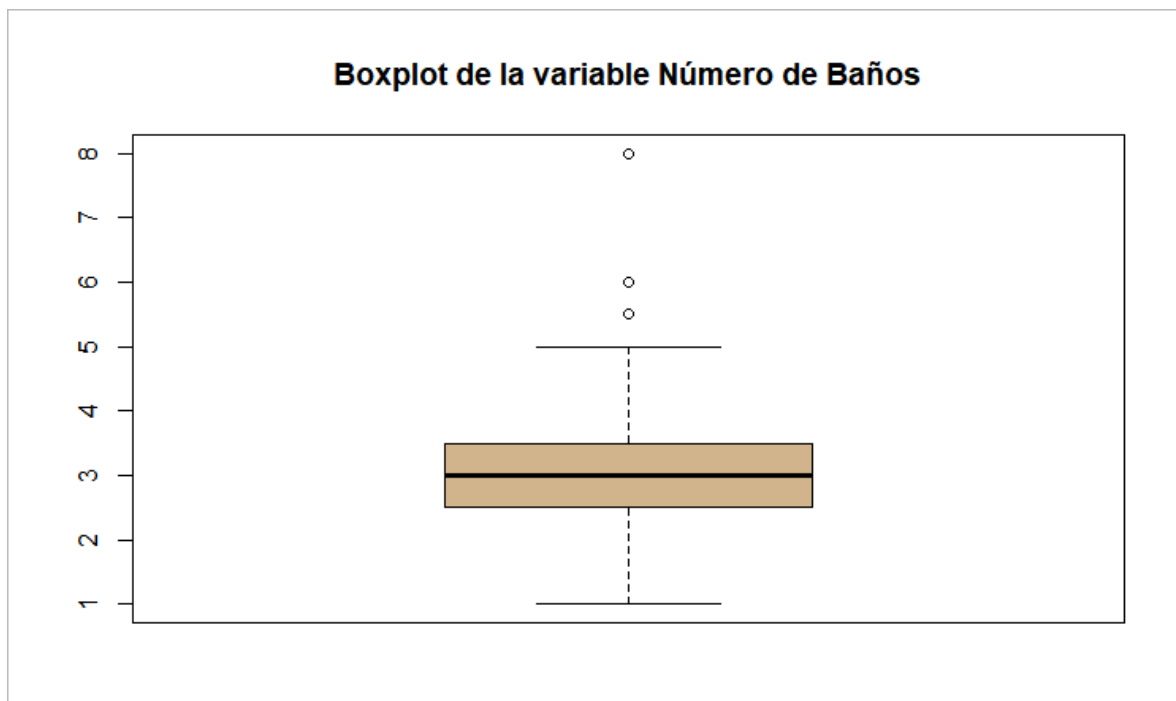


Figura 17. Diagrama de cajas de la variable número de baños

Con respecto al número de baños, las casas en su mayoría tienen de 2 a 3 y se tiene 3 datos atípicos, en esta variable se considera un baño social sin ducha como $\frac{1}{2}$ o 0.5 baño.

En la variable entera de años de antigüedad, se toma en cuenta que la antigüedad comienza a contarse desde que la casa se ocupa, esto quiere decir que, si la variable tiene el valor de cero, la casa es nueva por estrenar.

Tabla 12

Estadística Descriptiva de años de antigüedad

<i>Años de antigüedad</i>	
Media	7.53
Mediana	5
Moda	0
Desviación estándar	7.63
Varianza de la muestra	58.21
Mínimo	0
Máximo	35

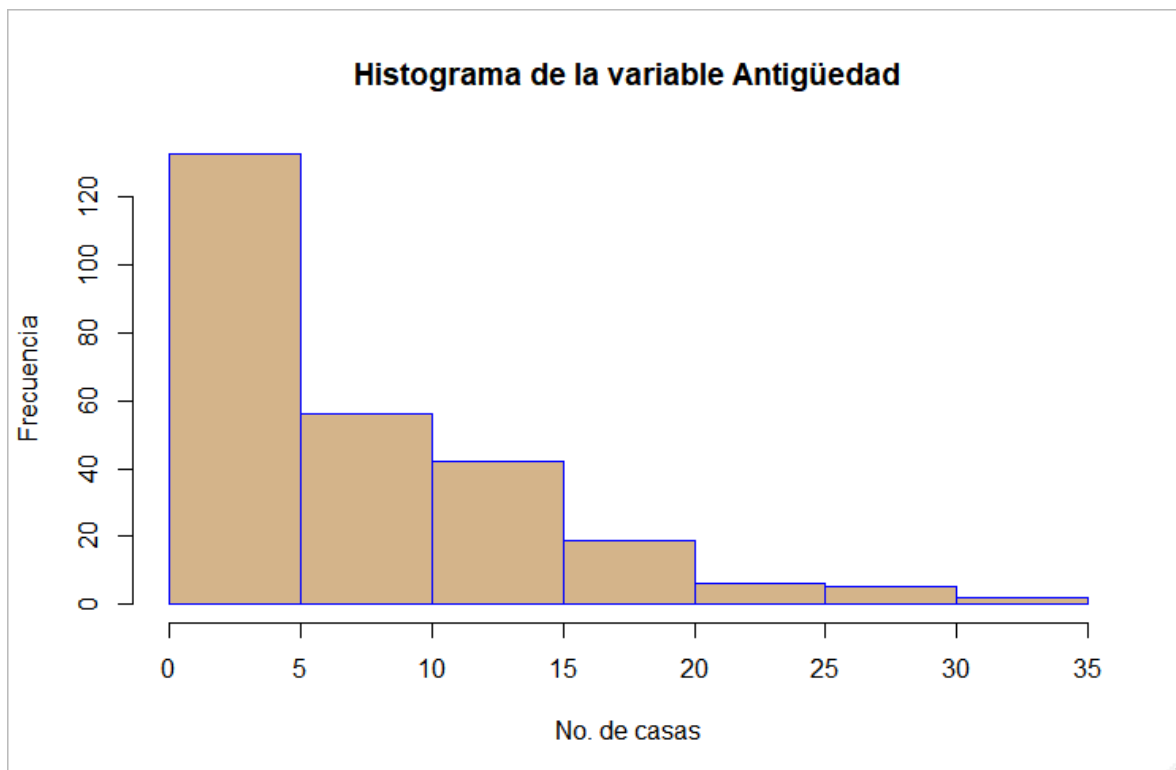


Figura 18. Histograma de la variable antigüedad

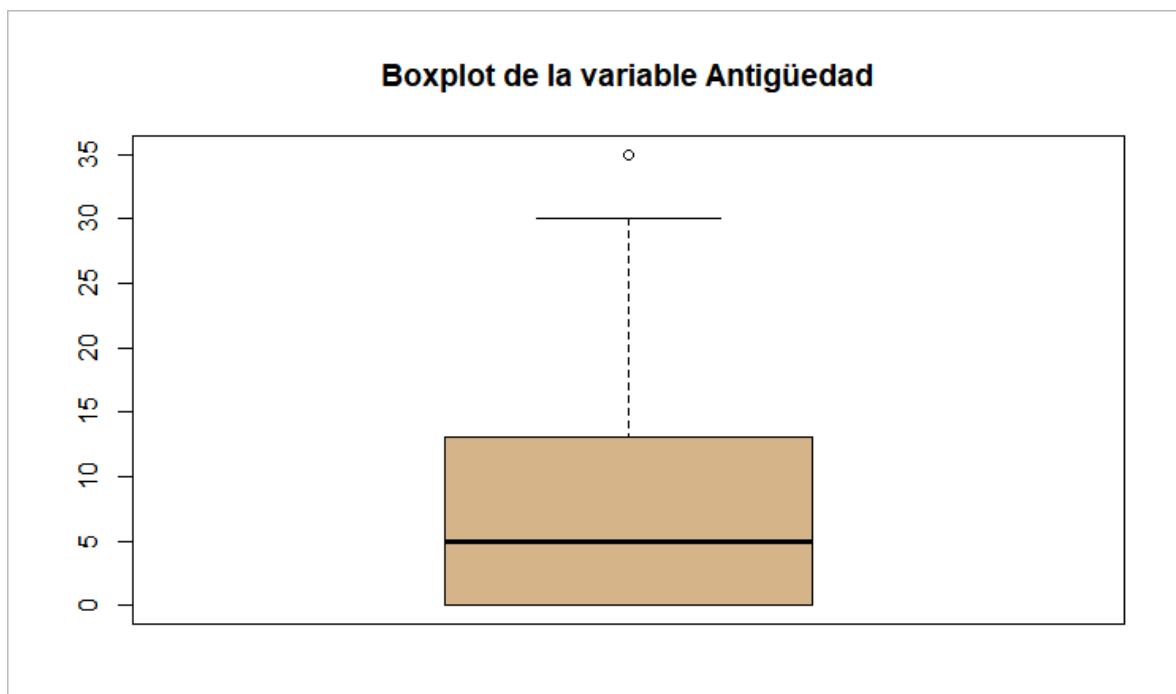


Figura 19. Diagrama de cajas de la variable Antigüedad

Dado que Rumiñahui es un cantón en expansión demográfica la oferta y demanda habitacional ha generado que se construyan muchas casas, en consecuencia, la mayoría de las casas es por estrenar, también se observa un dato atípico de una casa con 35 años de antigüedad.

Con respecto a sí las viviendas pertenecen o no a un condominio, conjunto privado o ciudadela privada cerrada, y para fines de simplificar la nomenclatura se llama pertenecía a condominio, es así que se obtuvo que 155 casas pertenecen a un condominio y que 135 no pertenecen a un condominio.

Al ser una variable discreta no se realiza el análisis de estadística descriptiva y los resultados se pueden observar en la siguiente figura:

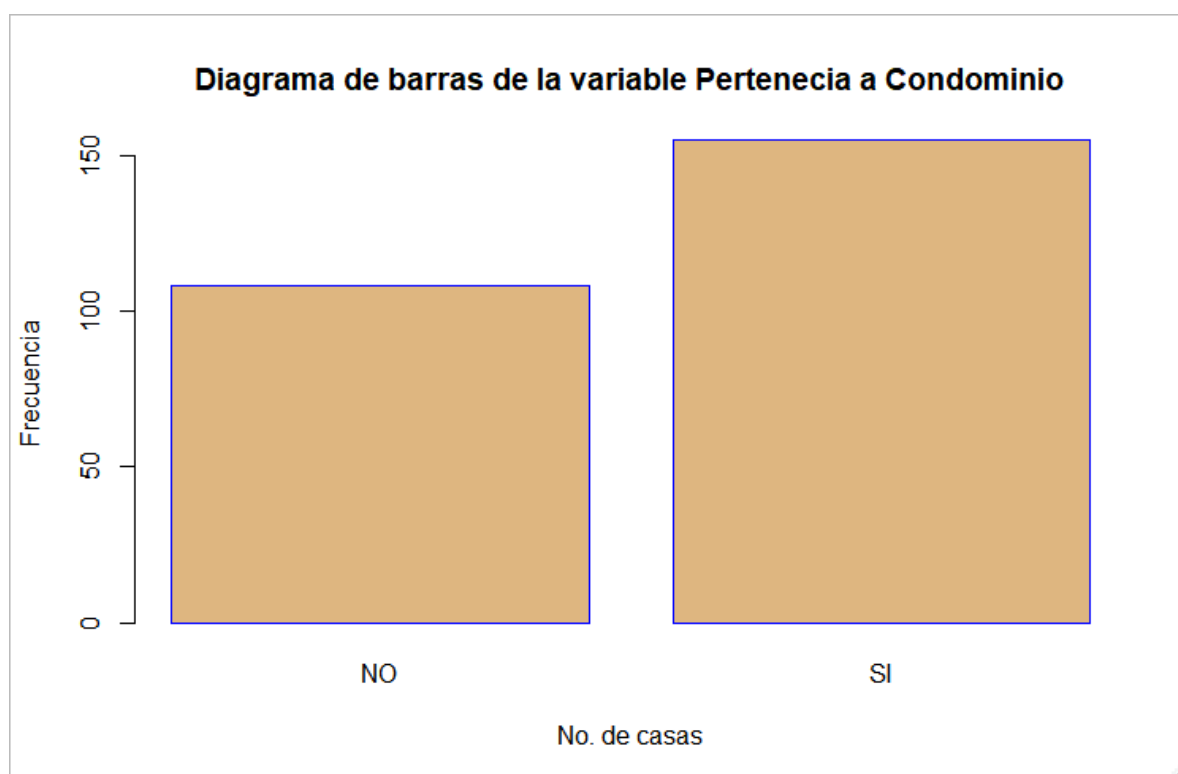


Figura 20. Diagrama de barras de la variable pertenencia a condominio

Se observa que la mayoría de casa está en condominio. Para el número de pisos de las viviendas se considera la planta baja como un piso, es así que se obtuvo la siguiente información:

Tabla 13

Estadística descriptiva de número de pisos de las viviendas

<i>Número de pisos</i>	
Media	2.02
Mediana	2
Moda	2
Desviación estándar	0.37
Varianza de la muestra	0.13
Mínimo	1
Máximo	3

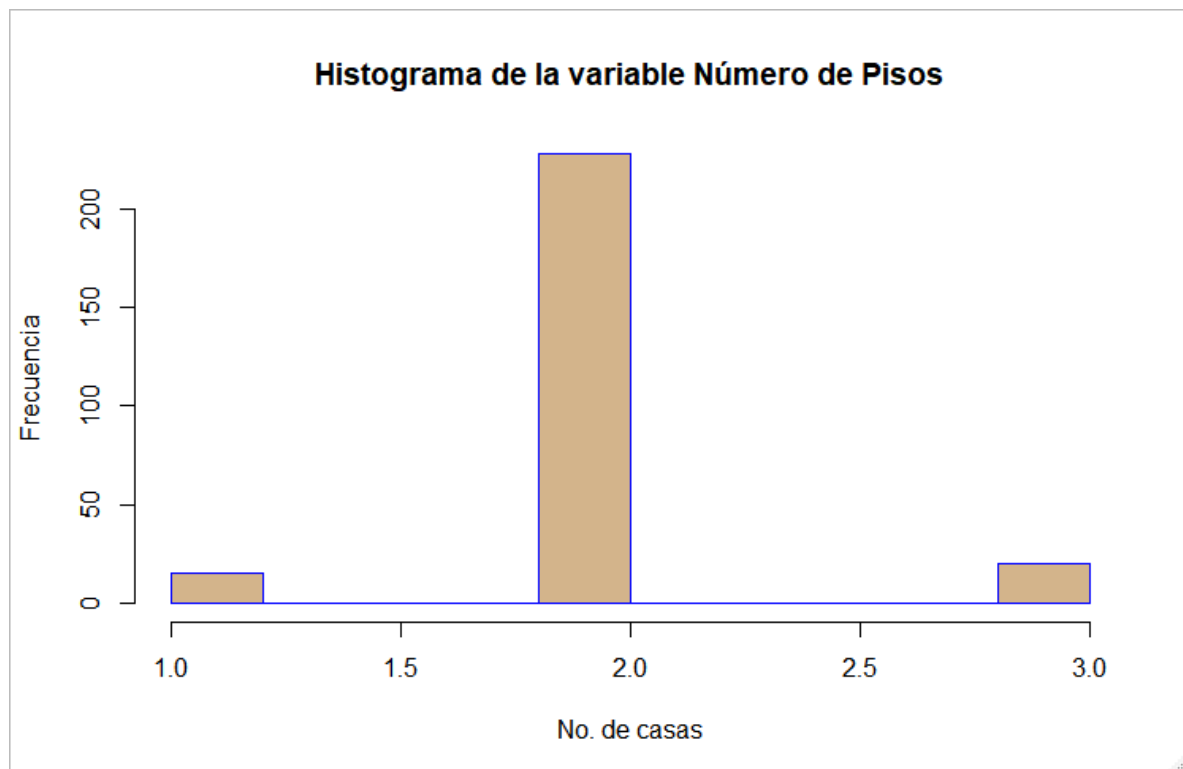


Figura 21. Histograma de la variable número de pisos

La mayoría de casas del análisis, es decir 228 casas son de dos pisos y solo se tiene inmuebles de hasta 3 pisos.

Para las variables que tienen que ver con distancia se implementó la fórmula de Haversine y de mínimo en (48) y (49), respectivamente, para cada inmueble se encontró la distancia a las otras casas y de entre ellas, se obtuvo la mínima distancia.

Para la variable distancia mínima a una unidad educativa, se obtuvo la siguiente información:

Tabla 14

Estadística Descriptiva de distancia mínima a una unidad educativa

<i>Distancia mínima a Unidad Educativa(m)</i>	
Media	616.97
Mediana	497.1
Moda	437.98
Desviación estándar	497.5
Varianza de la muestra	247505.5
Mínimo	5.84
Máximo	3543.16

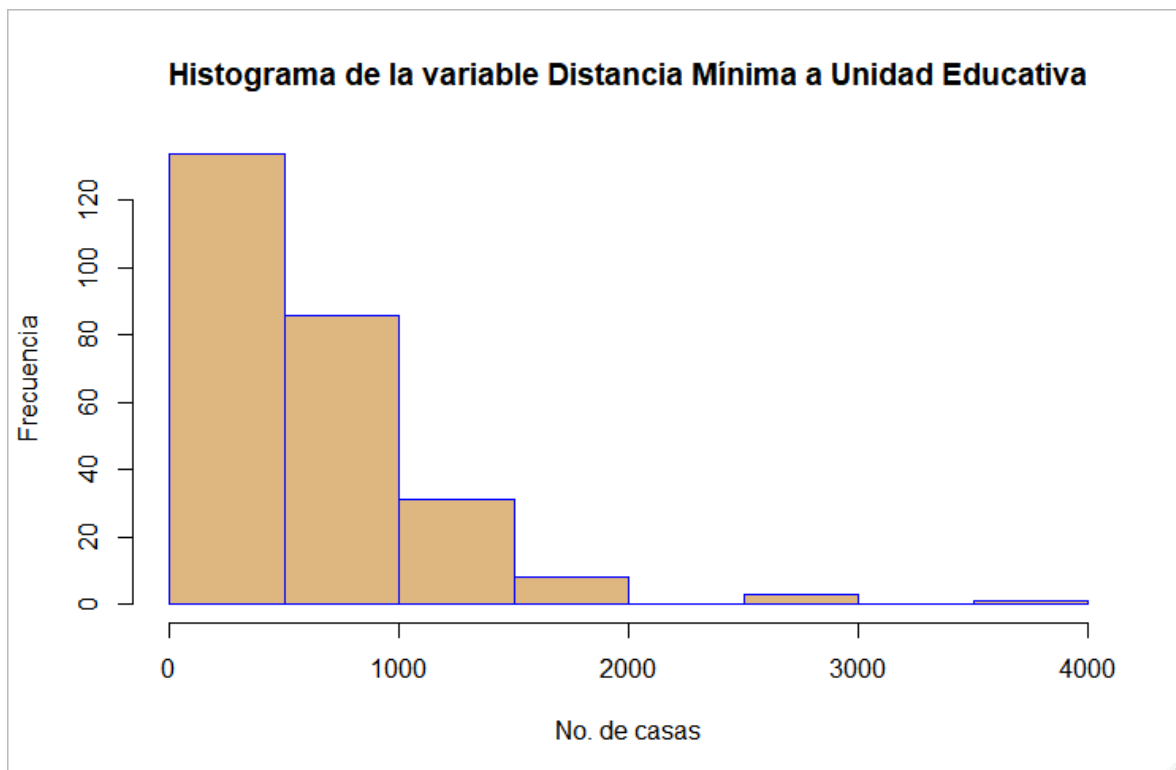


Figura 22. Histograma de la variable distancia mínima a una unidad educativa

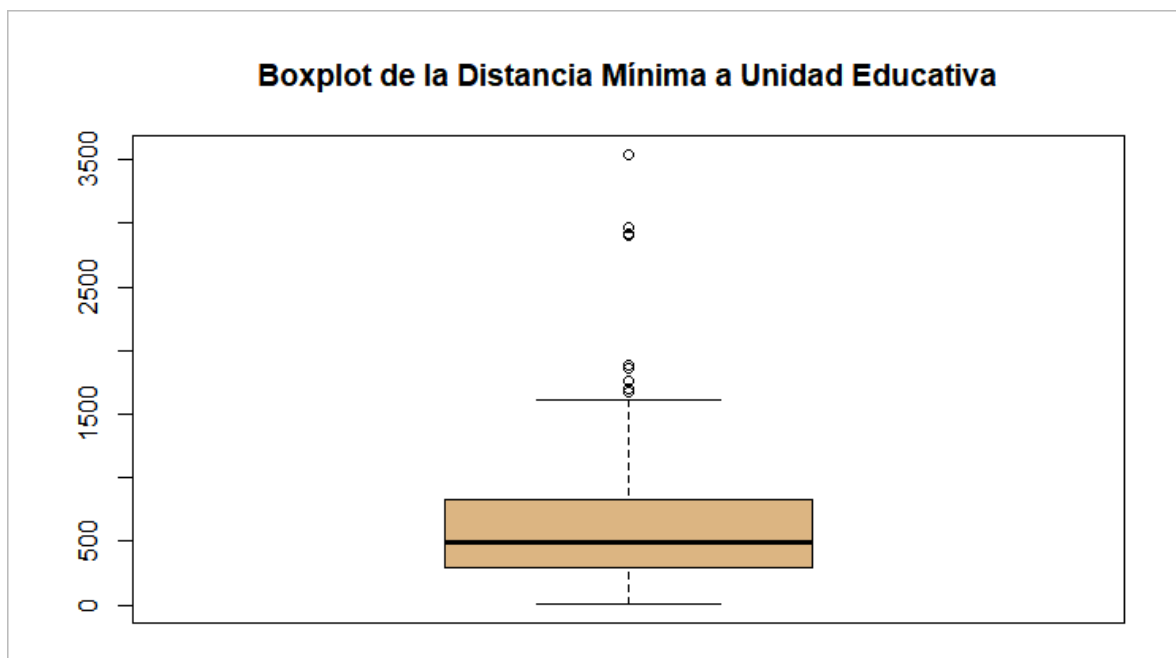


Figura 23. Diagrama de cajas de la variable distancia mínima a una Unidad Educativa

La mayoría de casas se encuentra en un rango menor a 512 metros de una Unidad Educativa, es así que la media es de 616.97 metros.

Para la variable real de distancia mínima a un centro de salud público o privado, se obtuvo los siguientes datos:

Tabla 15

Estadística Descriptiva de distancia mínima a un centro de salud

<i>Distancia mínima a un Centro de Salud (m)</i>	
Media	855.3
Mediana	842.6
Moda	1395.65
Desviación estándar	501.32
Varianza de la muestra	251318.6
Mínimo	30.40
Máximo	2375.8

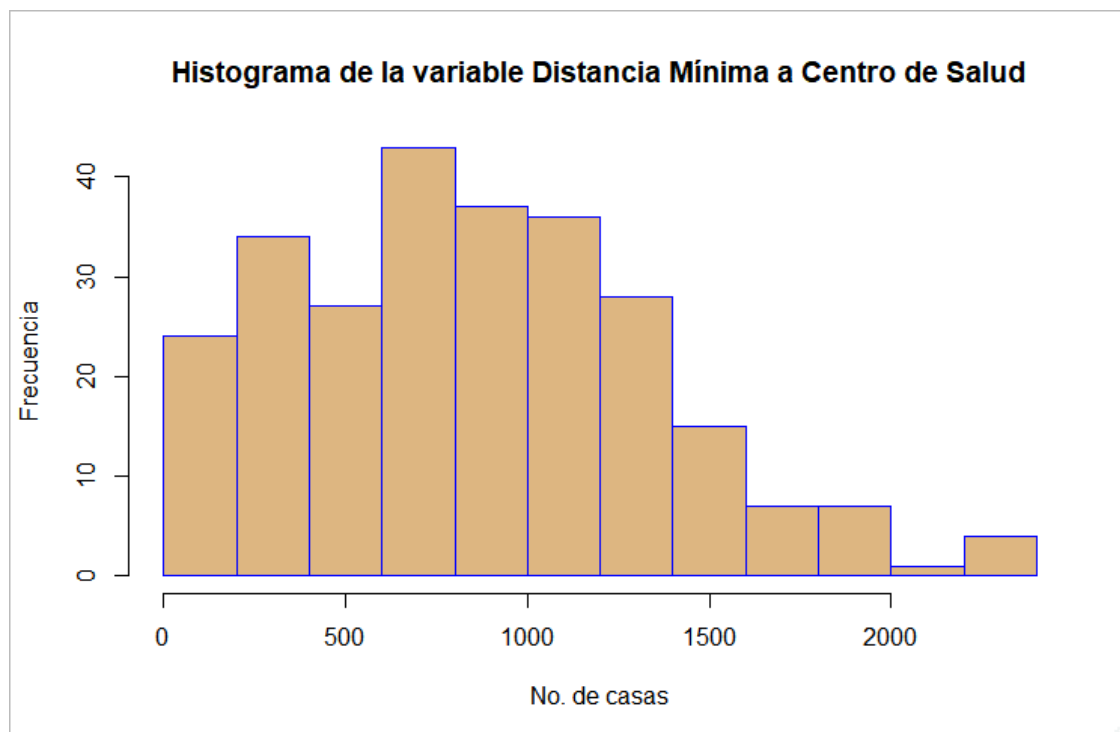


Figura 24. Histograma de distancia mínima a un centro de salud

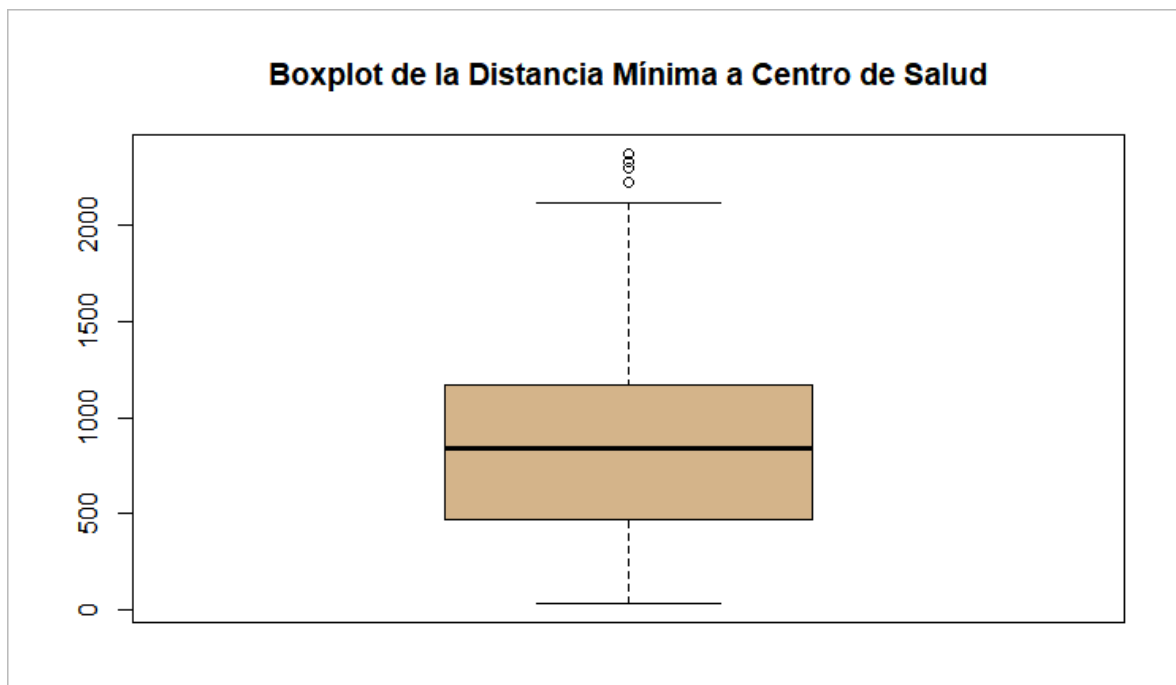


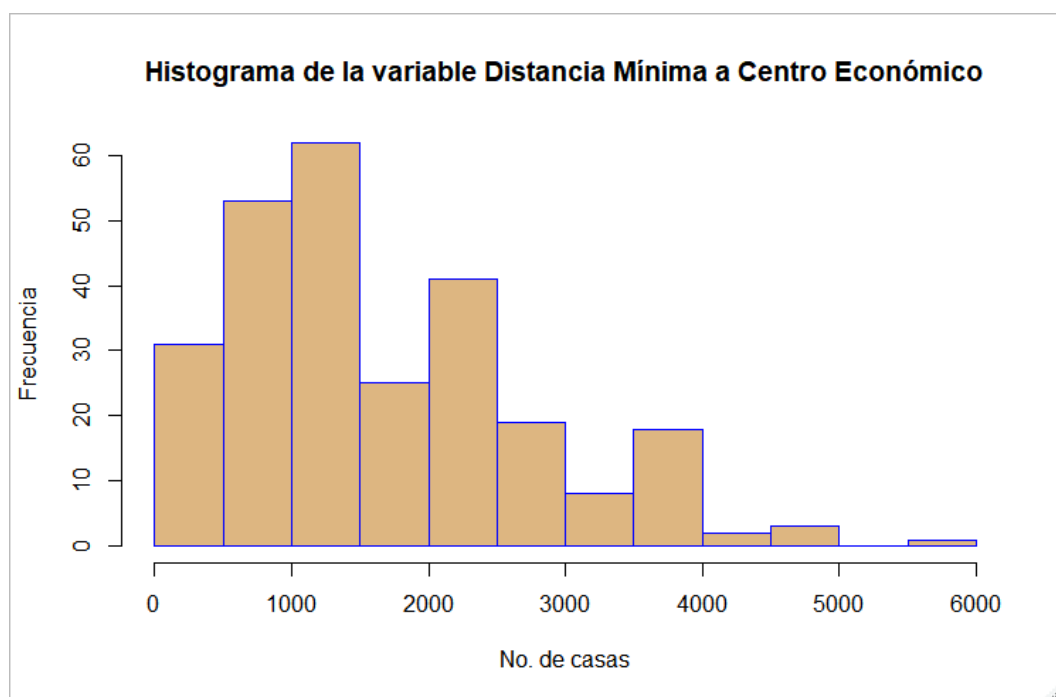
Figura 25. Diagrama de cajas de la variable distancia mínima a un Centro de Salud

La distancia media mínima hacia un centro de salud es 855.3 metros, sin embargo, se puede observar poca diferencia en frecuencia entre las primeras cuatro clases. Este es dato es lógico si se considera que los Centros de Salud están diseñados para estar accesibles a la mayor cantidad de personas posibles.

Para la variable real distancia mínima a centros de desarrollo económico, se toma tres puntos estratégicos del cantón: centro comercial San Luis, zona el Triángulo (esquina norte) y centro de Sangolquí (GAD Municipal), obteniendo los siguientes datos:

Tabla 16*Estadística Descriptiva de distancia mínima a centros de desarrollo económico*

<i>Distancia mínima a centro de desarrollo económico (m)</i>	
Media	1661.14
Mediana	1383.41
Moda	2422.74
Desviación estándar	1096.07
Varianza de la muestra	1201358
Mínimo	39.31
Máximo	5844.05

**Figura 26.** Histograma de variable distancia mínima a centro de desarrollo económico

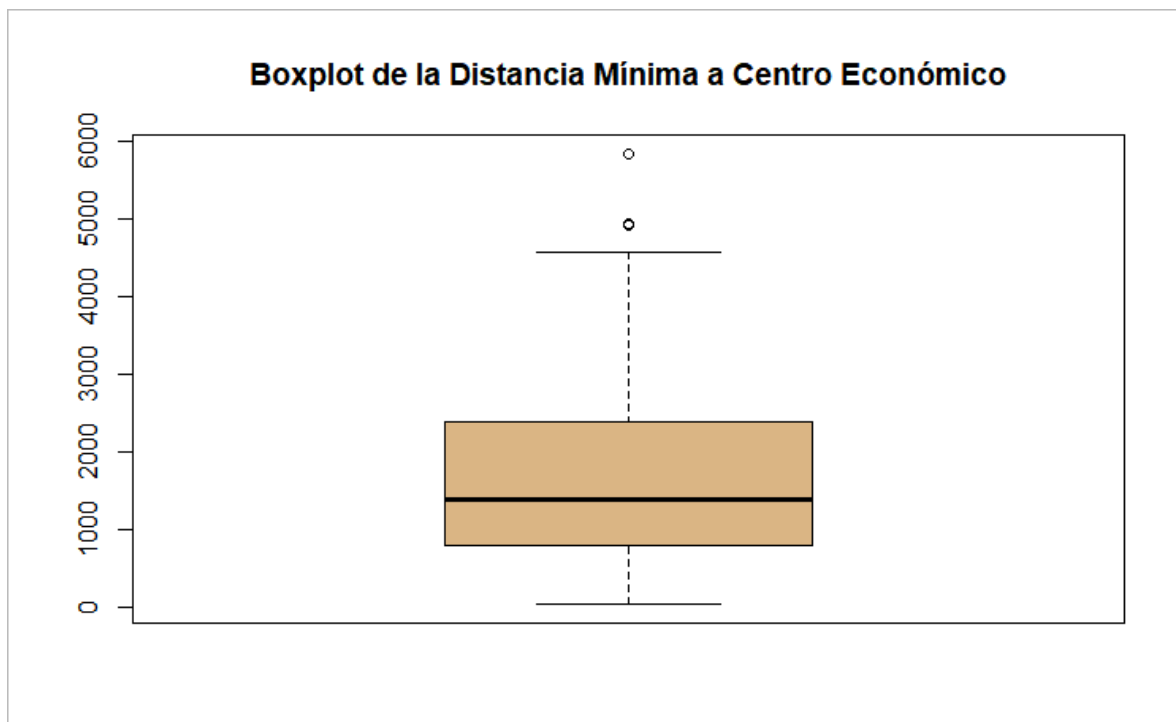


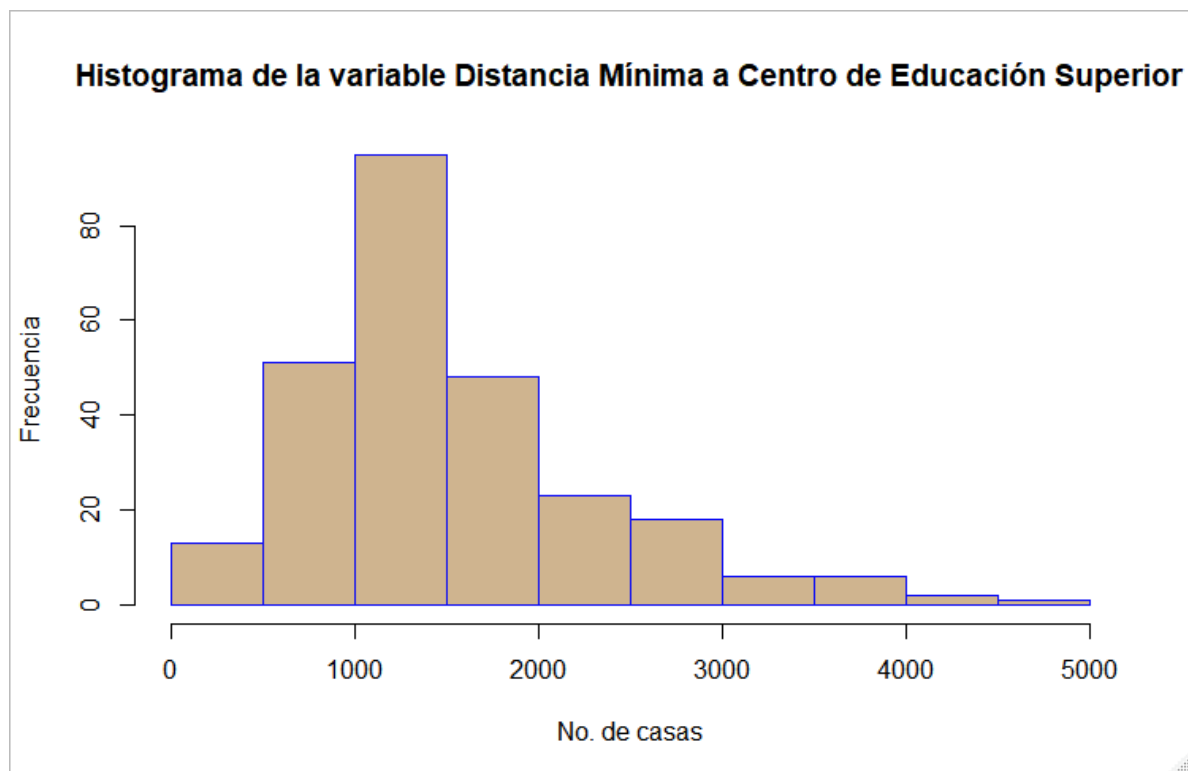
Figura 27. Diagrama de cajas de la variable distancia mínima a centro de desarrollo económico.

La distancia media de una casa a cualquiera de estos tres puntos es 1661.14 metros, se observa dos datos atípicos. El según cuartil se ubica en los 1383 m, siendo este dato menor que la media, se advierte que la mayoría de casas tiende a estar cerca de estos puntos de desarrollo económico.

Para obtener la distancia mínima a una Institución de Educación Superior (IES) se consideró a la Universidad de las Fuerzas Armadas ESPE, al Instituto Superior Rumiñahui y a la Universidad Metropolitana del Ecuador, se obtuvo los siguientes datos:

Tabla 17*Estadística Descriptiva de distancia mínima a una Centro de Educación Superior*

<i>Distancia mínima a Centro de Educación Superior</i>	
Media	1534.21
Mediana	1333.41
Moda	1910.87
Desviación estándar	804.77
Varianza de la muestra	647647.8
Mínimo	35.66
Máximo	4600.30

**Figura 28.** Histograma de la variable distancia mínima a un CES

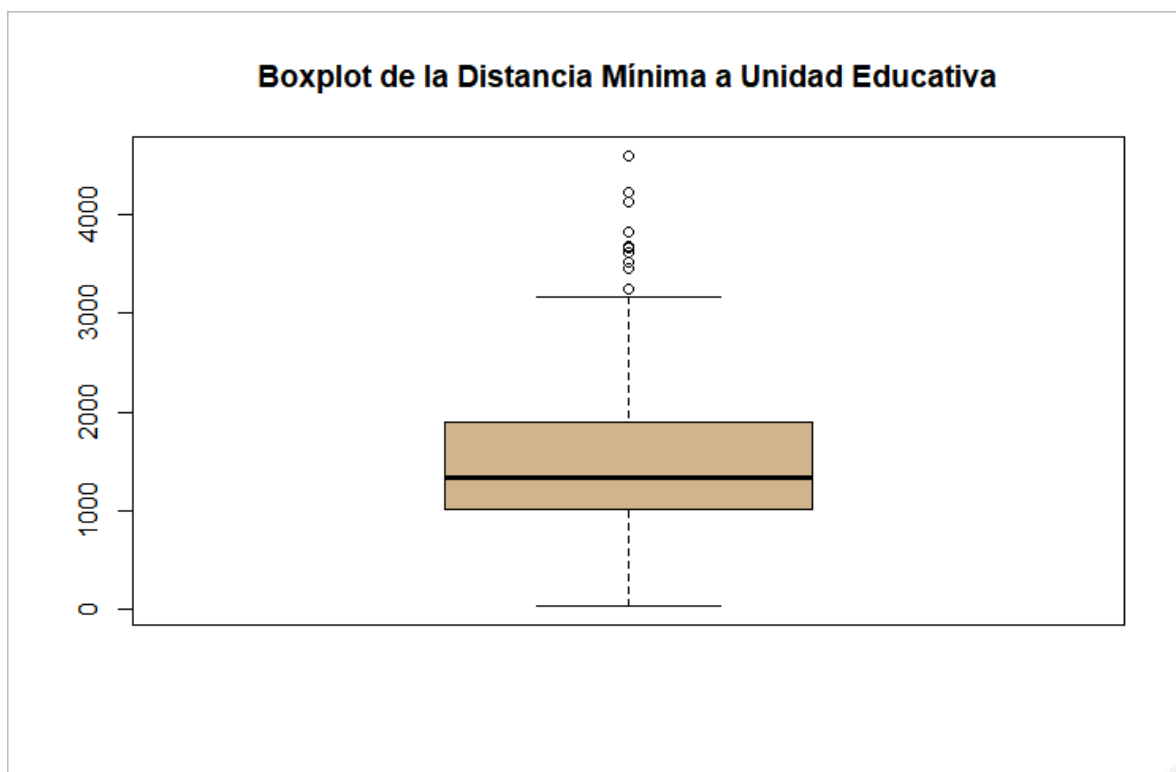


Figura 29. Diagrama de cajas de la variable distancia mínima a un CES

La distancia media mínima a un CES 1534.21 m., teniendo varios datos atípicos, el segundo cuartil se ubica en 1333 m., menor a la media, lo que sugiere que las viviendas tienden a estar lo más cerca posible de los CES, a continuación, se describe la variable precio por metro cuadrado total.

Tabla 18

Estadística Descriptiva de la variable Precio por m² total

<i>Precio por metro cuadrado total</i>	
Media	597.03
Mediana	587.71
Moda	750
Desviación estándar	218.67
Varianza de la muestra	47814.63
Mínimo	44.44
Máximo	1194.44

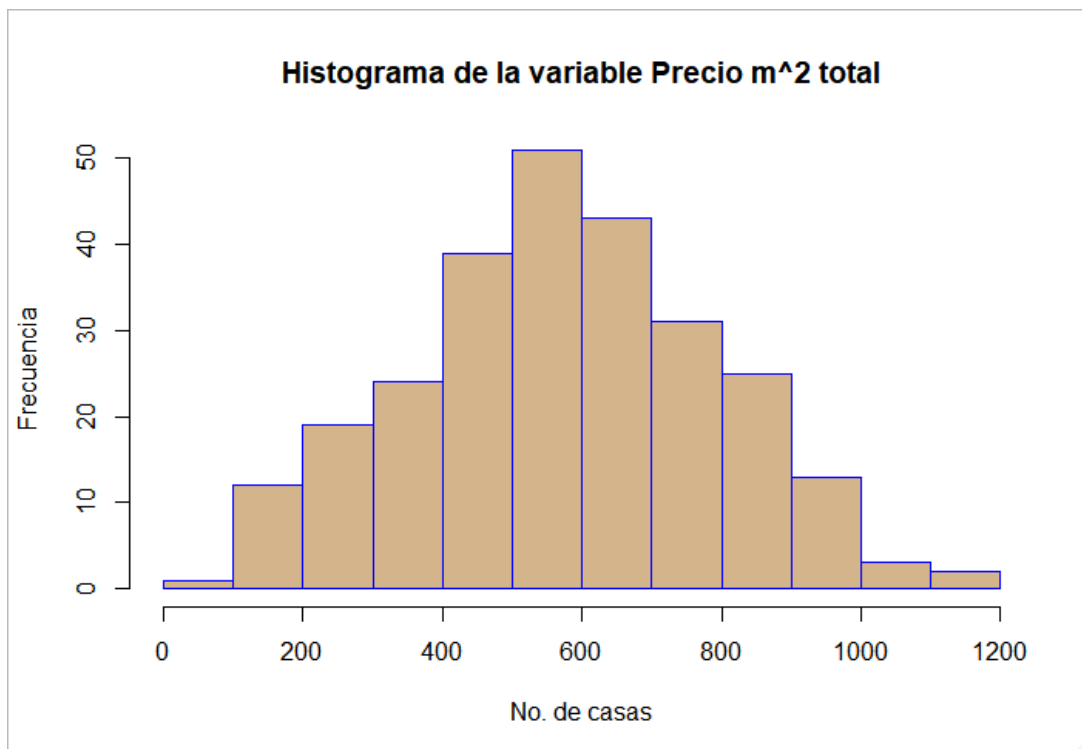


Figura 30. Histograma de la variable Precio por m^2 total

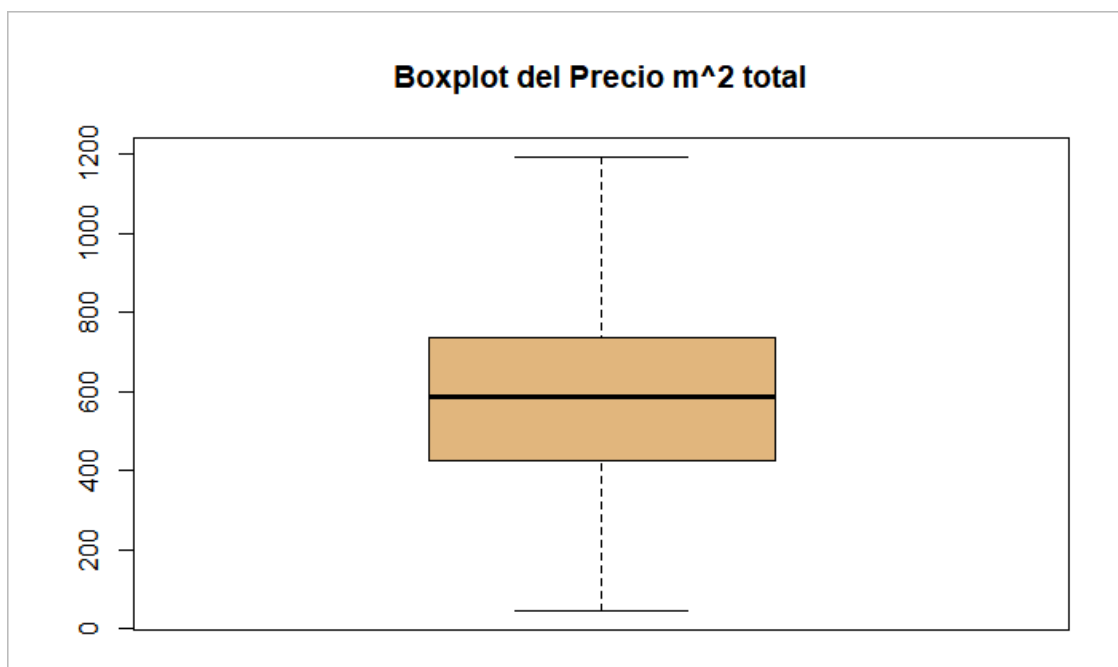


Figura 31. Diagrama de cajas de la variable Precio por m^2 total

Para el análisis de la variable precio, se tienen la siguiente información:

Tabla 19
Estadística Descriptiva de la variable Precio

<i>Distancia mínima a Centro de Educación Superior</i>	
Media	167504
Mediana	145000
Moda	185000
Desviación estándar	88627.2
Varianza de la muestra	7854780319
Mínimo	48000
Máximo	495000

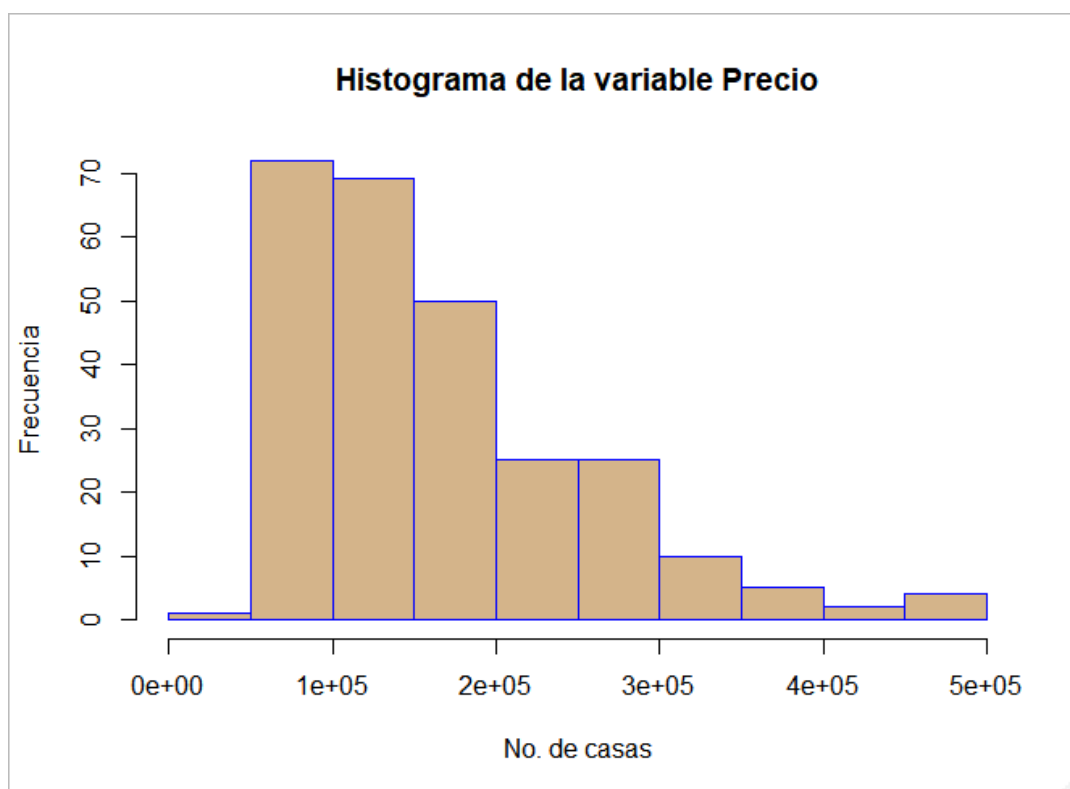


Figura 32. Histograma de la variable Precio por m^2 total

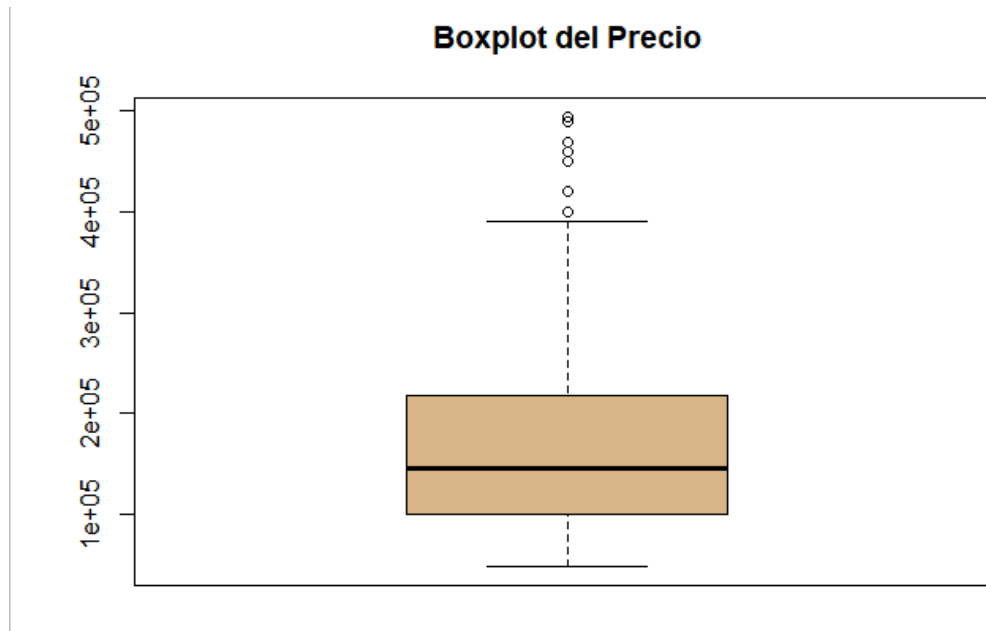


Figura 33. Diagrama de cajas de la variable Precio por m^2 total

Se observa que la variable precio por metro cuadrado total, se encuentra mejor distribuida que la variable precio, existe poca diferencia entre la media y la mediana, la mayoría de las observaciones se ubican en la segunda, tercera y cuarta clase.

Del análisis descriptivo, las variables que muestran una distribución simétrica son: superficie cubierta, número de baños, número de pisos, distancia mínima a un centro de Educación Superior y precio por metro cuadrado total, las demás variable muestran una distribución asimétrica. Las variable que no tienen relación a distancias minimas, muestran una media de 2.5 datos atípicos y un moda de 3 datos atípicos, por lo cual se filtran 3 datos ordenando las viviendas mediante la distancia mínima de Mahalanobis.

3.3. Análisis de Componentes Principales ACP.

El ACP se realiza en el software R y se decide excluir la variable precio por m^2 cubierto, ya que la variable precio por m^2 está mejor distribuida, los resultados se detallan a continuación:

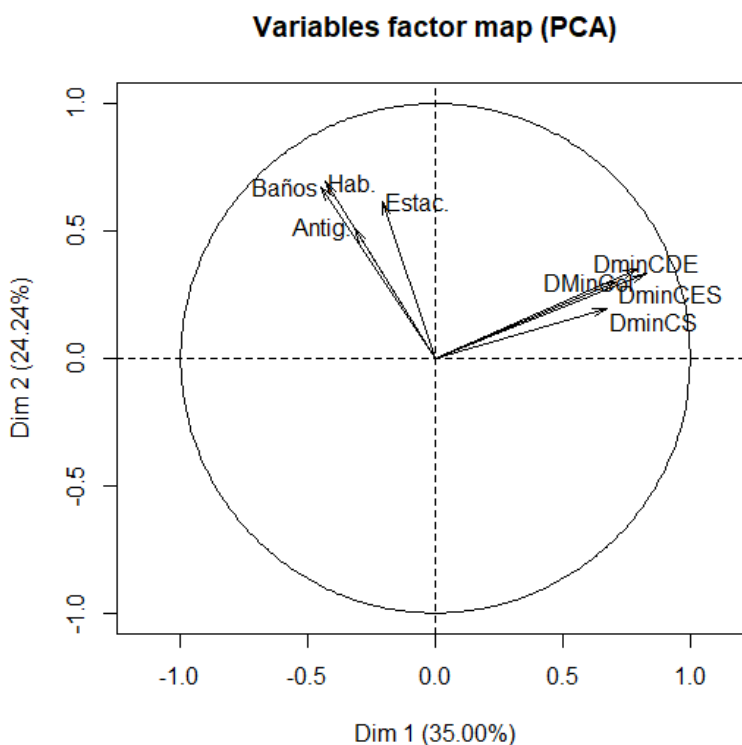


Figura 34. Circulo de correlación de las variables

Se observa cierto agrupamiento en las variables dentro del círculo de correlaciones, los agrupamientos son entre: número de habitaciones, número de baños, número de estacionamientos y antigüedad; por otro lado, se observa una correlación entre las variables espaciales: distancia mínima a una Unidad Educativa, distancia mínima a un Centro de Salud, distancia mínima a zonas de desarrollo económico y distancia mínima a Instituciones de Educación Superior.

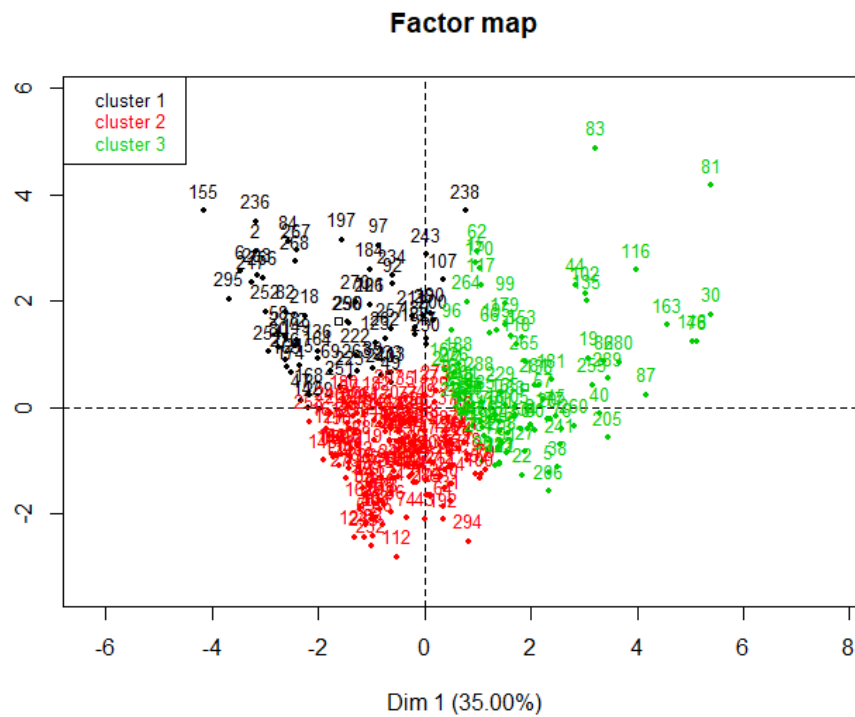


Figura 35. Mapeo de las componentes 1 y 2 del ACP

El ACP, se lo realiza de tal manera que el número de clústeres se generen automáticamente en el software, se obtiene tres clústeres y en las dos primeras componentes se representa el 59.24% de la variabilidad de los datos, al tomar las tres primeras componentes se representa el 70.04% de la variabilidad de los datos, como se puede observar en la tabla 21.

Tabla 20*Porcentaje de variación representado por las componentes principales*

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8
Varianza	2.8	1.939	0.864	0.784	0.729	0.384	0.284	0.216
% de var.	35.001	24.238	10.802	9.805	9.109	4.805	3.545	2.695
% de var. Acumulada	35.001	59.239	70.041	79.846	88.955	93.76	97.305	100

El grado de varianza representado por las variables se representa en la tabla 30.

Tabla 21*Representación de la variabilidad de las Componentes Principales en función en las variables*

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.1 + Dim.2	Dim.1+ Dim.2+ Dim.3
Habitaciones	0.187498	0.482492	5.5407E-06	0.003589	0.148163	67%	67%
Baños	0.200368	0.455948	0.02344738	0.058041	0.079329	66%	68%
Estacion.	0.044452	0.380173	0.00169056	0.097833	0.469236	42%	43%
Antigüedad	0.101410	0.259637	0.03066951	0.574444	0.023815	36%	39%
DMinCC	0.628155	0.123375	0.04968687	0.018318	0.002060	75%	80%
DminCS	0.452160	0.037495	0.40653393	0.010760	0.000138	49%	90%
DminCE	0.685509	0.109186	0.0578876	0.001969	0.000523	79%	85%
DminCES	0.500490	0.090749	0.2942255	0.019445	0.005439	59%	89%

Las dos primeras componentes principales representan el 79% de la variabilidad en la variable distancia mínima a centros de desarrollo económico, que es la que mayor variabilidad representa y la que menos representa la variabilidad es la variable antigüedad con el 36%. Tomando las 3 primeras componentes principales la variable distancia mínima a Instituciones de Educación Superior representa un 89% de la variabilidad y la variable antigüedad representa el 39% de la variabilidad.

También se puede observar en el círculo de correlación que las variables: antigüedad, baños, estacionamientos y habitaciones tienen una correlación al estar agrupadas y que las

variables georreferenciadas de distancia también se agrupan lo que quiere decir que guardan cierta correlación entre ellas.

El análisis de componentes principales establece relación entre las variables número de habitaciones, número de baños, número de estacionamientos y antigüedad, siendo las de mayor radio en el círculo de correlaciones, el número de habitaciones y de baños, que serán las variables con las que se iniciará la corrida de los modelos para luego ir incrementando otras variables en función de los resultados de los modelos.

3.4. Metodología para realizar los modelos

Para la predicción de los modelos y como variable dependiente se puede utilizar las variables: precio, precio por m² total, o precio por m² cubierto, ya que se desea predecir el precio del inmueble. La variable precio por m² cubierto no considera la superficie de terreno de un inmueble y que involucra un precio adicional en función de la cantidad de terreno que posea. Se puede observar que la variable precio por m² total tiene una distribución equitativa entre sus cuartiles y la media y la mediana se asemejan más que en la variable precio, también se observa que la variable precio posee datos atípicos a diferencia de precio por m² total, por consiguiente esta última se define como variable dependiente y como variables independientes a las variables no espaciales, para realizar los modelos de factores.

Para los modelos de regresión hedónica se utilizarán las variables independientes utilizadas para el modelo de factores más las variables espaciales, que son las distancias mínimas a puntos estratégicos.

Para poder referenciar a las variables de mejor manera se adopta la siguiente nomenclatura:

Tabla 22
Nomenclatura propuesta para Variables

<i>Variable</i>	<i>Nomenclatura</i>
Número de Baños	NumBan
Número de Estacionamientos	NumEst
Número de Cuartos	NumCua
Años de Antigüedad	Ant
Pertenece a Condominio	Cond
Número de Pisos	NumPis
Precio por m^2	Pxm2
Distancia mínima a Unidad Educativa o Colegio	DminCC
Distancia mínima a Centro de Salud	DminCS
Distancia mínima a Centro de Desarrollo Económico	DminCE
Distancia mínima a Centro de Educación Superior	DminCES

A continuación, se analiza la correlación entre las variables independientes no espaciales y la variable dependiente de forma gráfica:

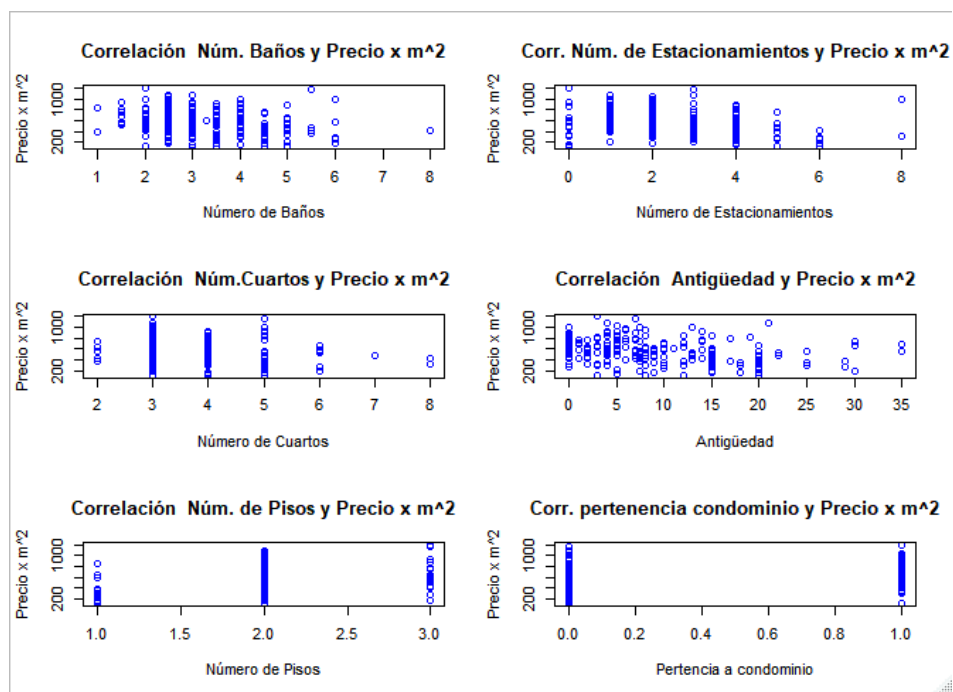


Figura 36. Diagrama de Correlación entre variables independientes y precio por m^2 total

En la anterior figura se puede advertir cierta correlación inversa en las variables: número de baños, número de estacionamientos y antigüedad, sin embargo al ser variables categóricas es difícil observar la correlación gráficamente, en consecuencia se procede a las pruebas de correlación paramétrica, para las cuales se utiliza la prueba de correlación de Pearson, el cual da como resultado un coeficiente, sin embargo la interpretación de la magnitud del coeficiente de correlación depende de la particularidad y del contexto de la aplicación, en términos generales se considera que una correlación es baja por debajo de ± 0.3 , que existe una asociación moderada o media entre ± 0.3 y ± 0.7 , y alta por encima de ± 0.7 (Laguna, 2011), de las pruebas realizadas se deriva la siguiente tabla :

Tabla 23

Correlación entre variable dependiente y variables independientes

Variable Dependiente	Pxm2	Categorización
Variables Independientes		
NumBan	-0.25	baja
NumEst	-0.32	media
NumCua	-0.29	baja
Ant	-0.35	media
Cond	0.38	media
NumPis	0.24	baja

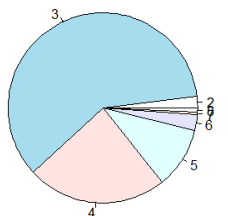
Si bien la correlación esta entre baja y media, se comprueba que existe correlación entre las variables, y por lo tanto se puede realizar los modelos mediante la regresión lineal, para los modelos de factores.

Para realizar la categorización de las variables se toma en cuenta el análisis descriptivo realizado previamente y se utiliza el software R Studio para hacer un filtro más y sacar datos atípicos.

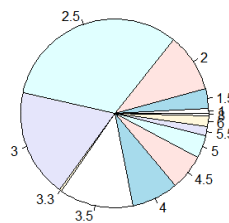
En una primera instancia se trabaja con la base de 263 viviendas, con las cuales se aplica la distancia cuadrada de Mahalanobis para determinar los datos atípicos, aplicando la técnica se obtiene que con relación a las variables dependiente e independientes, los datos atípicos son las casas número 135, 145 y 179, se extraen estas viviendas de la base, obteniendo una base de trabajo de 260 viviendas.

Como criterio de categorización de variables se toma una o dos opciones según las posibilidades de la variable y pretendiendo que las categorías queden uniformemente distribuidas en número de elementos, para dicha clasificación se hace referencia a la sección donde se hace el Estudio descriptivo de los datos y se realiza diagramas tipo pastel de las variables como se muestra a continuación:

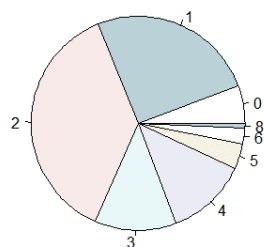
Pie de numero de cuartos



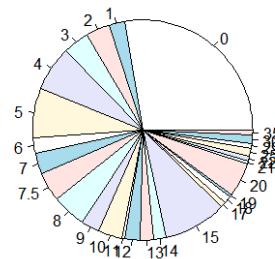
Pie de número de baños



Pie de número de estacionamientos



Pie de antigüedad del bien



CONTINUA

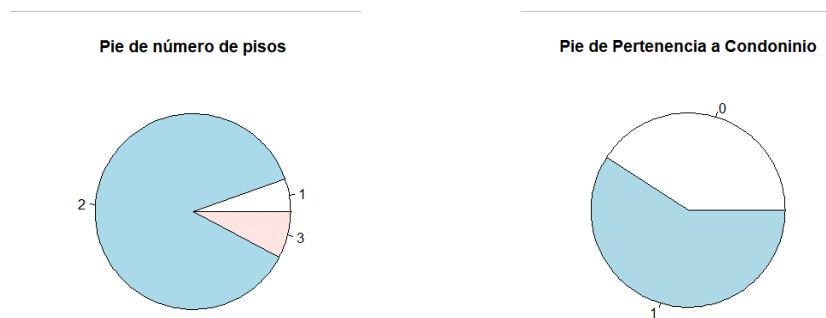


Figura 37. Diagramas de pastel de las variables independientes

Para la variable número de habitaciones o cuartos simbolizada como NumCua, se observa, que no se puede agrupar uniformemente en tres clases, así que se toma dos, de la siguiente manera:

Tabla 24
Categorización de la variable Número de Habitaciones

Número de Habitaciones (NumCua)		
Clase	Asignación	Frecuencia relativa (%)
[1,3]	1	60%
]3, 9]	2	40%

En el caso de esta variable no se pudo encontrar otra posible combinación en la cual se observe uniformidad en el número de elementos de las clases.

Para la variable número de baños, abreviada como NumBan, se agrupa en tres categorías, se obtuvo dos posibles opciones:

Tabla 25
Categorización de la variable Número de Baños

Número de Baños (NumBan)		
Opción 1		
Clase	Asignación	Frecuencia relativa (%)
[1,2]	1	13%
]2, 3]	2	52%

CONTINUA

]3, 8]	3	35%
Opción 2		
Clase	Asignación	Frecuencia relativa (%)
[1,2.5]	1	41%
[3, 4]	2	45%
[4.5, 8]	3	14%

Para la variable número de estacionamientos abreviada como NumEst y se agrupa en tres grupos, se obtuvo una sola opción en la que los datos se muestra distribuidos:

Tabla 26
Clases de la variable número de estacionamientos

Número de Estacionamientos (NumEst)		
Clase	Asignación	Frecuencia relativa (%)
[0,1]	1	29%
2	2	38%
]2, 8]	3	33%

Para la variable antigüedad, se toma tres opciones, cada una con tres categorías, se obtiene:

Tabla 27
Categorización de la variable Antigüedad

Antigüedad (Ant)		
Opción 1		
Clase	Asignación	Frecuencia relativa (%)
[0,5]	1	49%
]5, 10]	2	21%
]10, 35]	3	30%
Opción 2		
Clase	Asignación	Frecuencia relativa (%)
[1,6]	1	51%
[7, 14]	2	26%
[15, 35]	3	23%
Opción 3		
Clase	Asignación	Frecuencia relativa (%)

CONTINUA

[1,2]	1	32%
]2, 9]	2	34%
]9, 35]	3	34%

La variable pertinencia a condominio, abreviada como Cond, la cual es una variable categoría, quedaría como tal:

Tabla 28

Categorización de la variable Pertinencia a Condominio

Pertinencia a Condominio (Cond)		
Clase	Asignación	Frecuencia relativa (%)
Si	En conjunto	59%
No	No en conjunto	41%

La variable número de pisos, abreviada como NumPis, solo se la puede ordenar con tres clases, ya que las todas las casas de la base de datos tienen uno, dos o tres pisos.

Tabla 29

Categorización de la variable número de pisos

Número de Pisos (NumPis)		
Clase	Asignación	Frecuencia relativa (%)
1	1	5%
2	2	86%
3	3	9%

Se crea en la misma base de datos variables categóricas según los criterios antes señalados, están toman el mismo nombre de variable predecesora aumentando la letra C al final, que indica que la variable es o ha sido categorizado, por ejemplo, la variable NumPis, al ya estar categorizada toma el nombre de NumPisC.

3.5. Diseño Factorial.

El diseño factorial se lo realiza utilizando la base de trabajo procesada en la sección anterior y tomando en cuenta la significancia obtenida en el ANOVA para las variables y sus combinaciones.

Se utiliza el software R Studio y se corren cuarenta y cuatro modelos diferentes, de los cuales veinticuatro se generan de las seis combinaciones diferentes de las opciones de categorización de las variables, como se muestra a continuación:

Tabla 30
Modelos trabajados

Número de Habitaciones (NumCua)	Número de Baños (NumBan)	Número de Estacionamientos (NumEst)	Antigüedad (Ant)	Pertinencia a Condominio (Cond)	Número de Pisos (NumPis)	Modelo multiplicativo con todas las variables	Modelo aditivo con todas las variables	Modelo multiplicativo con Variables significativas	Modelo aditivo con Variables significativas	Corrida
Opción Única	Opción 1	Opción Única	Opción 1	Opción Única	Opción Única	mf1	mf2	mf3	mf4	1
Opción Única	Opción 2	Opción Única	Opción 1	Opción Única	Opción Única	mf1	mf2	mf3	mf4	2
Opción Única	Opción 1	Opción Única	Opción 2	Opción Única	Opción Única	mf1	mf2	mf3	mf4	3
Opción Única	Opción 2	Opción Única	Opción 2	Opción Única	Opción Única	mf1	mf2	mf3	mf4	4
Opción Única	Opción 1	Opción Única	Opción 3	Opción Única	Opción Única	mf1	mf2	mf3	mf4	5
Opción Única	Opción 2	Opción Única	Opción 3	Opción Única	Opción Única	mf1	mf2	mf3	mf4	6

Los veinte modelos no representados en la tabla anterior, se trabajan con las variables con recurrencia significativa en los veinticuatro modelos iniciales, y combinaciones de variables que tengan mayor significancia, esta última se obtiene del ANOVA.

Se utiliza la función *step* de R Studio, la cual mediante combinaciones recurrentes de factores y mediante el criterio de Akaike sugiere un “mejor modelo”, sin embargo esta función no toma en cuenta criterios de normalidad, homocedasticidad, independencia y los errores R^2 del modelo, en consecuencia se utiliza como guía sin dejar de lado los criterios antes mencionados.

Luego de haber realizado las pruebas con los cuarenta y cuatro modelos se separan tres modelos que son los que cumplen los supuestos de normalidad, homocedasticidad, independencia y que tienen mejor relación espacial, a continuación, se realiza el análisis completo de los tres mejores modelos de factores, sin embargo, este mismo análisis se realizó para los cuarenta y cuatro modelos de factores.

3.5.1. Modelo de factores 1.

El modelo de factores 1 se referencia como mf1, y las variables toman la nomenclatura de la Tabla 22, aumentando la letra C al final, que significa que la variable paso un proceso de categorización que se realizó en la sección clases de equivalencia, el mf1 tiene la siguiente ecuación:

$$Pxm2T = 354.01 + \beta_0 NumCuaC + \beta_1 AntC + \beta_2 Cond + \beta_3 NumPisC:AntC \quad (50)$$

Donde:

$$\beta_0 = \begin{cases} -61.19, & \text{si } NumCuaC > 3, \\ 0 & , \quad \text{si } NumCuaC \leq 3 \end{cases}$$

$$\beta_1 = \begin{cases} -59.29, & \text{si } AntC > 10, \\ 58.57, & \text{si } AntC > 5 \wedge AntC \leq 10, \\ 0, & \text{si } AntC \leq 5 \end{cases}$$

$$\beta_2 = \begin{cases} 85.01, & \text{si } Cond = \text{Si en conjunto}, \\ 0 & , \quad \text{si } Cond = \text{No en conjunto} \end{cases}$$

$$\beta_3 = \begin{cases} 260.09, & \text{si } AntC \leq 5 : NumPisC = 2, \\ 116.94, & \text{si } (AntC > 5 \wedge AntC) \leq 10 : NumPisC = 2, \\ 178.25, & \text{si } AntC > 10 : NumPisC = 2, \\ 299.45, & \text{si } AntC \leq 5 : NumPisC = 3, \\ 158.22, & \text{si } (AntC > 5 \wedge AntC) \leq 10 : NumPisC = 3, \\ 46.43, & \text{si } AntC > 10 : NumPisC = 3, \\ 0, & \text{cualquier otro caso} \end{cases}$$

3.5.1.1. Bondad de ajuste de mf1.

Para observar gráficamente la bondad de ajuste se gráfica los Pxm2T real y los Pxm2T predichos como sigue:

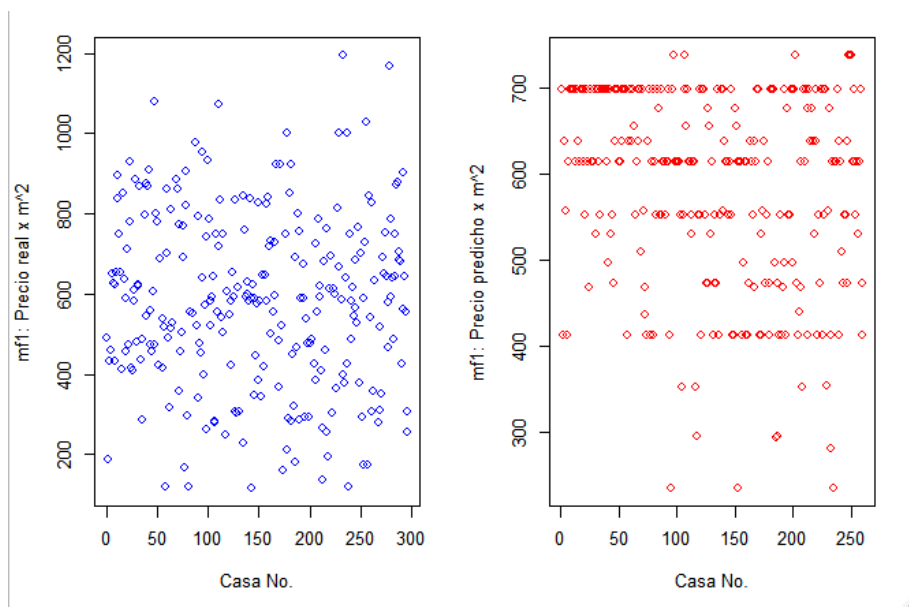


Figura 38. Pxm2T real vs. Pxm2T predicho por mf1

De la figura anterior se puede observar que el máximo predicho llega hasta menos de \$800 y el real menos de \$1200, por lo cual la prueba gráfica no es concluyente.

El error R^2 múltiple en mf1 es de 0.2765 y el error R^2 ajustado es de 0.2475, lo cual indica que la bondad de ajuste del modelo es baja, sin embargo en investigaciones como en “Uso del método de precios hedónicos, para estimar variación en precios de viviendas producto de nuevas líneas de metro” (Quesada, 2015), los modelos generados tuvieron un error R^2 de entre 0.25 y 0.45, a pesar de la baja bondad de ajuste, los modelos tuvieron una aceptable predicción debido a la significancia de las variables utilizadas y se dan como válidos. En la presente investigación la validación del modelo se la realiza mediante “Validación cruzada” que se explicará posteriormente, por lo cual se pasa por alto el R^2 bajo.

3.5.1.2. Prueba de Normalidad de mf1.

La prueba de normalidad se la realiza gráficamente mediante la superposición de la curva normal de los residuos de mf1, como se observa a continuación:

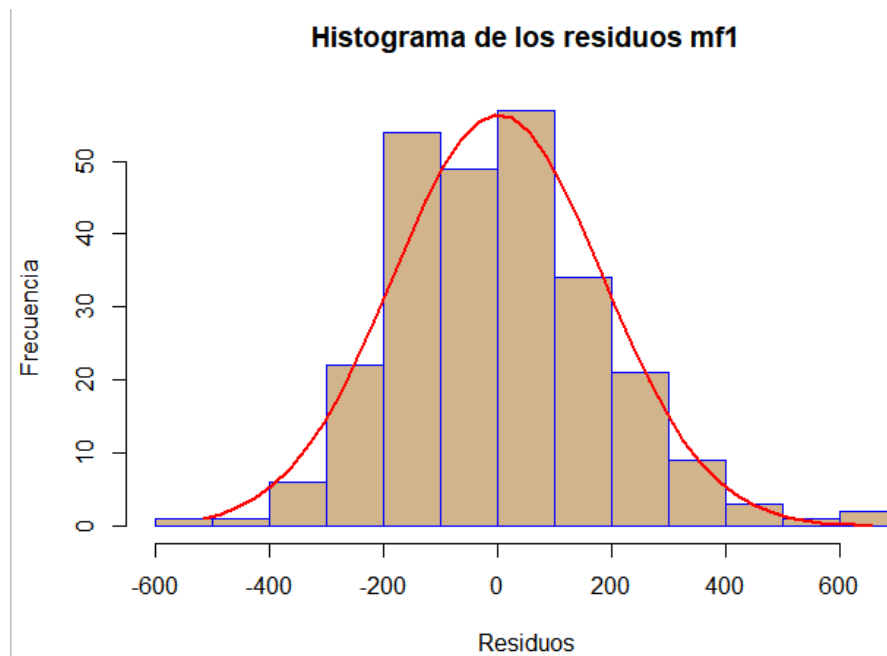


Figura 39. Histograma de los residuos de mf1

Se puede observar que los residuos, tienen una similitud a la curva normal, se realiza también la prueba gráfica de probabilidad normal o de cantiles teóricos Q-Q.

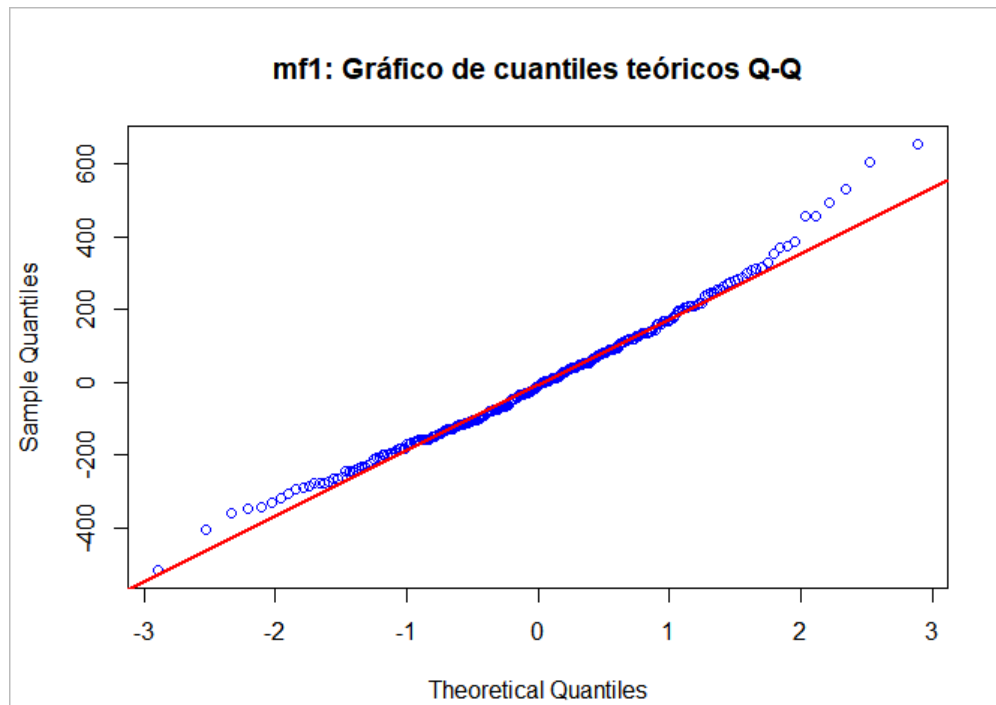


Figura 40. Gráfica Cuantil - Cuantil de mf1

Se puede observar de la Figura 40. Gráfica Cuantil - Cuantil de mf1, que los residuos se alinean en su mayoría a línea roja, esto significa que los residuos siguen una distribución normal, no obstante se realiza la prueba analítica de Lilliefors con la siguiente hipótesis:

$$H_0: p - \text{valor} \geq \alpha \text{ (Normalidad)} \quad (51)$$

$$H_1: p - \text{valor} < \alpha \text{ (No normalidad)}$$

Donde α es 0.05 es el nivel de significancia de la prueba, en adelante todas las pruebas de hipótesis tendrán este nivel de significancia y se realizan en R Studio. La prueba genera un p-valor de $0.3415 > 0.05$, por lo tanto, se acepta la hipótesis nula de que los datos siguen una distribución normal.

3.5.1.3. Pruebas de homocedasticidad de mf1.

Para la prueba gráfica de igualdad de varianza se utiliza el diagrama siguiente:

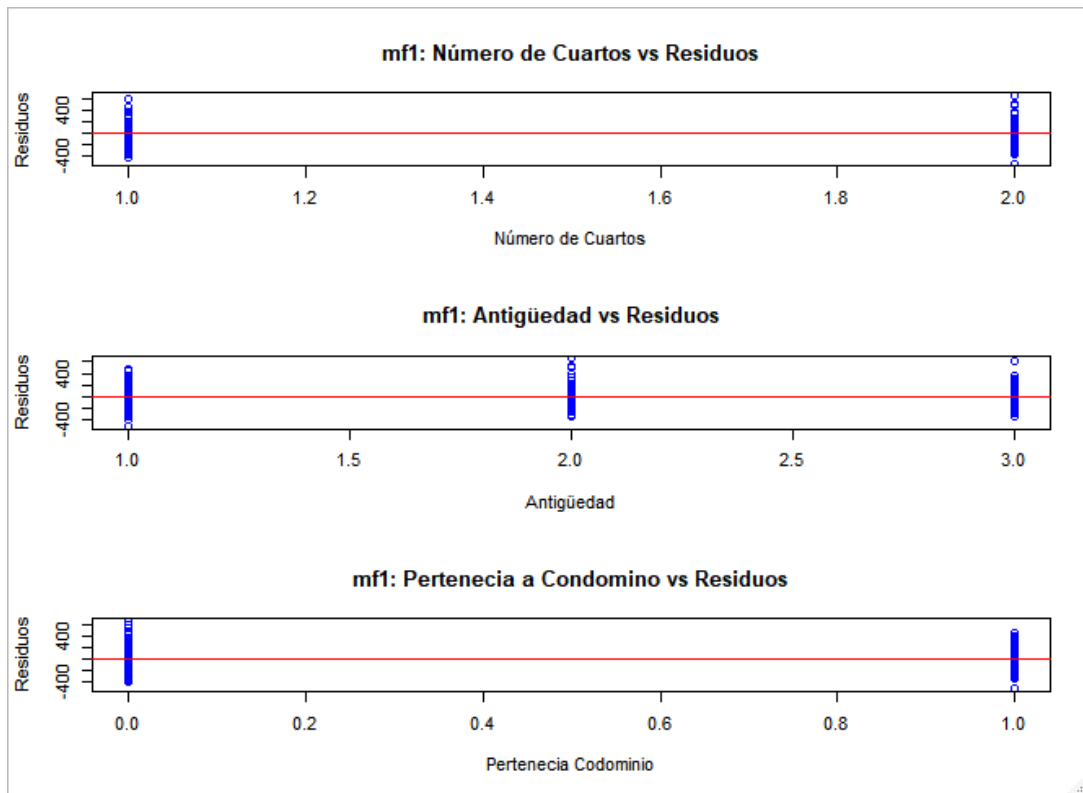


Figura 41. Diagrama de variables vs Residuos mf1

Del diagrama anterior se puede observar que las observaciones se distribuyen equitativamente sobre y bajo la línea roja, lo que sugiere que las variables utilizadas tienen iguales varianzas con los residuos del modelo, no obstante, se realiza la prueba analítica de Bartlett para las variables, con la hipótesis:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad (p - \text{valor} \geq 0.05) \quad (52)$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 \quad (p - \text{valor} < 0.05)$$

La prueba arroja un p-valor de 0.2861, 0.3635 y 0.0712 correspondiente a las variables NumCuaC, AntC y Cond, respectivamente, todos los p-valor son mayores que 0.05, lo que quiere decir que si hay igualdad de varianzas entre las variables y los residuos del modelo.

3.5.1.4. Pruebas de identidad de mf1.

Para la prueba de identidad se grafica los residuos en eje de las ordenadas y la secuencia de toma de observaciones en el eje de las abscisas, se obtiene:

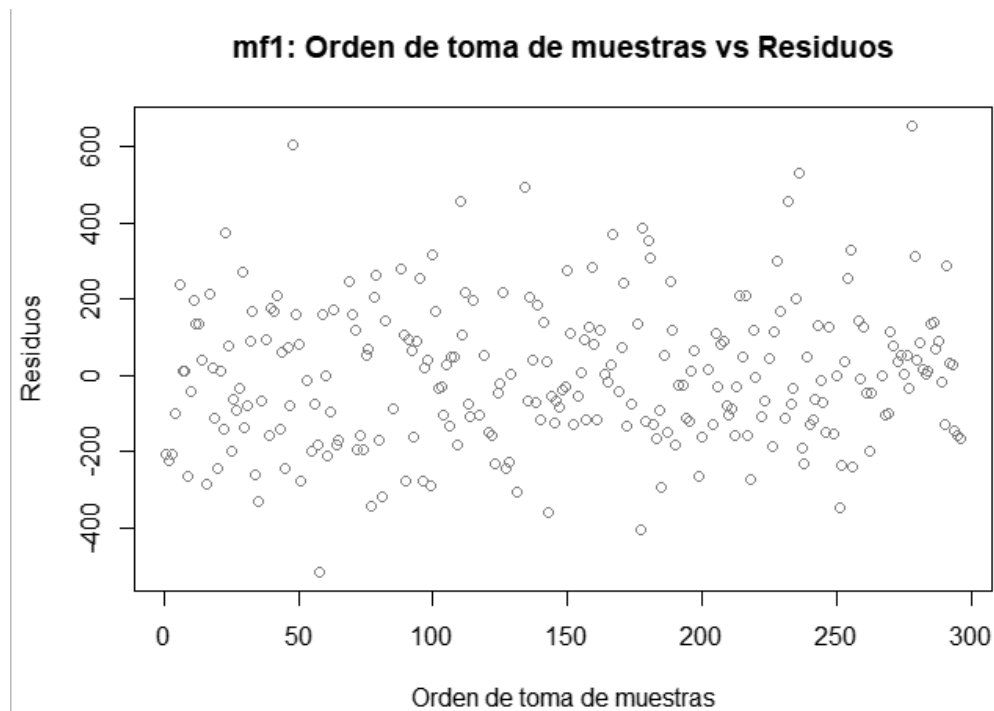


Figura 42. Diagrama de orden de toma de observaciones vs. residuos

Se puede apreciar en el grafico que no existe ningún patrón definido, sin embargo, se realiza prueba analítica de Durbin Watson, la cual define la hipótesis como sigue:

$$H_0: \rho = 0 \quad (p - \text{valor} \geq 0.05) \quad (53)$$

$$H_1: \rho > 0 \quad (p - \text{valor} < 0.05)$$

El p-valor que se obtuvo es de $0.762 > 0.05$, por lo tanto, se acepta la hipótesis nula de que existe independencia en la toma de las observaciones.

3.5.2. Modelo de Factores 2.

El modelo de factores 2 se referencia como mf2, con la nomenclatura explicada en 3.5.1., el mf2 tiene la siguiente ecuación:

$$Pxm2T = 89.94 + \beta_0 NumCuaC + \beta_1 AntC + \beta_2 Cond + \beta_3 NumEstC:NumCuaC + \beta_4 NumPisC:AntC \quad (54)$$

Donde:

$$\beta_0 = \begin{cases} -42.12, & \text{si } NumCuaC > 3, \\ 0, & \text{si } NumCuaC \leq 3 \end{cases}$$

$$\beta_1 = \begin{cases} -20.95, & \text{si } AntC > 10, \\ 52.84, & \text{si } AntC > 5 \wedge AntC \leq 10, \\ 0, & \text{si } AntC \leq 5 \end{cases}$$

$$\beta_2 = \begin{cases} 69.94, & \text{si } Cond = \text{No en conjunto}, \\ 0, & \text{si } Cond = \text{Si en conjunto} \end{cases}$$

$$\beta_3 = \begin{cases} -18.59, & \text{si } NumCuaC \leq 3 : (NumEstC = 1 \vee NumEstC = 2), \\ -14.76, & \text{si } NumCuaC > 3 : (NumEstC = 1 \vee NumEstC = 2), \\ -93.1, & \text{si } NumCuaC \leq 3 : NumEstC > 2, \\ -80.72, & \text{si } NumCuaC > 3 : NumEstC > 2, \\ 0, & \text{cualquier otro caso} \end{cases}$$

$$\beta_4 = \begin{cases} 258.36, & \text{si } AntC \leq 5 : NumPisC = 2, \\ 121.95, & \text{si } (AntC > 5 \wedge AntC \leq 10) : NumPisC = 2, \\ 138.74, & \text{si } AntC > 10 : NumPisC = 2, \\ 290.68, & \text{si } AntC \leq 5 : NumPisC = 3 \\ 163.82, & \text{si } (AntC > 5 \wedge AntC \leq 10) : NumPisC = 3, \\ 33.85, & \text{si } AntC > 10 : NumPisC = 3, \\ 0, & \text{cualquier otro caso} \end{cases}$$

3.5.2.1. Bondad de ajuste de mf2.

Para observar gráficamente la bondad de ajuste se gráfica los Pxm2T real y los Pxm2T predichos como sigue:

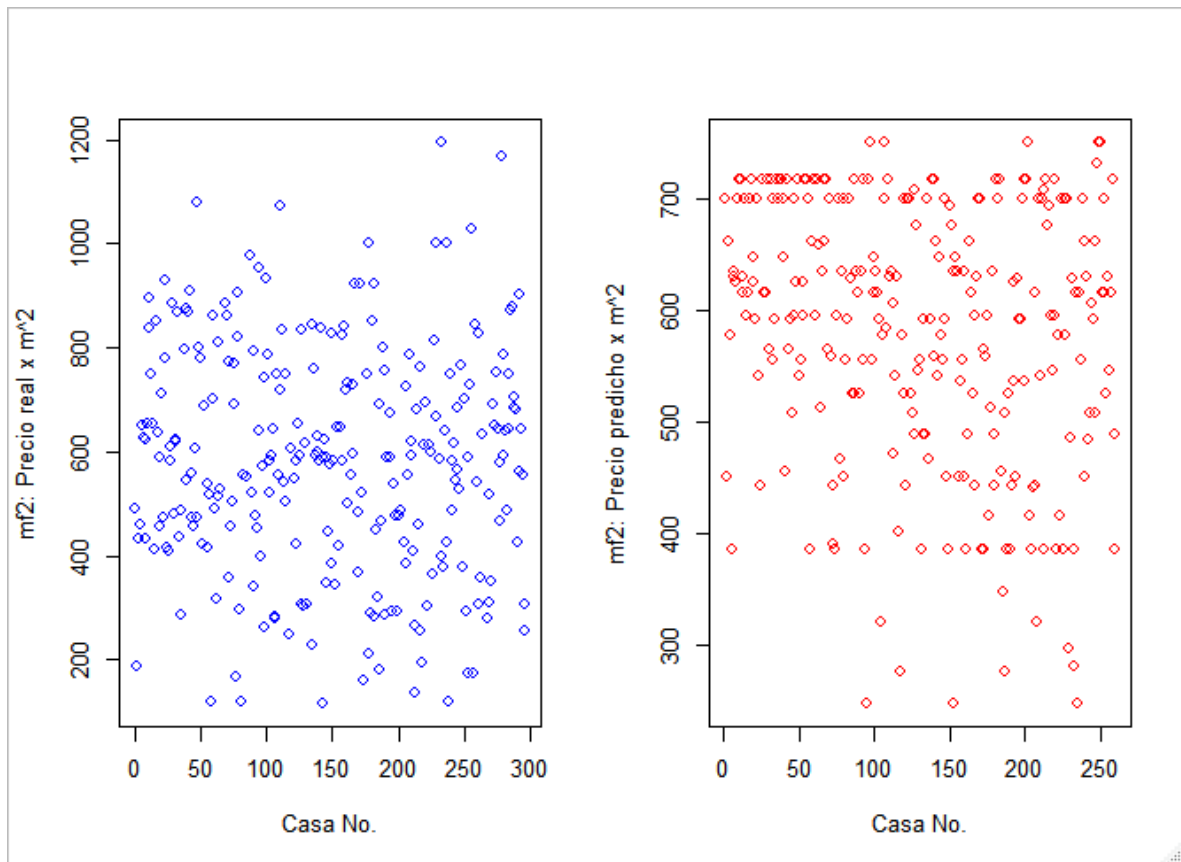


Figura 43. Pxm2T real vs. Pxm2T predicho por mf2

Se observa algunas diferencias entre los Pxm2T real y los Pxm2T predicho, por lo cual la prueba no es concluyente.

El error R^2 múltiple en mf2 es de 0.299 y el error R^2 ajustado es de 0.259, lo cual indica que la bondad de ajuste del modelo es baja y no se concluye por lo expuesto en 3.5.1.1. La bondad de ajuste se verificará posteriormente mediante validación cruzada.

3.5.2.2. Prueba de Normalidad de mf2.

La prueba de normalidad se la realiza gráficamente mediante la superposición de la curva normal de los residuos de mf2, como se observa a continuación:

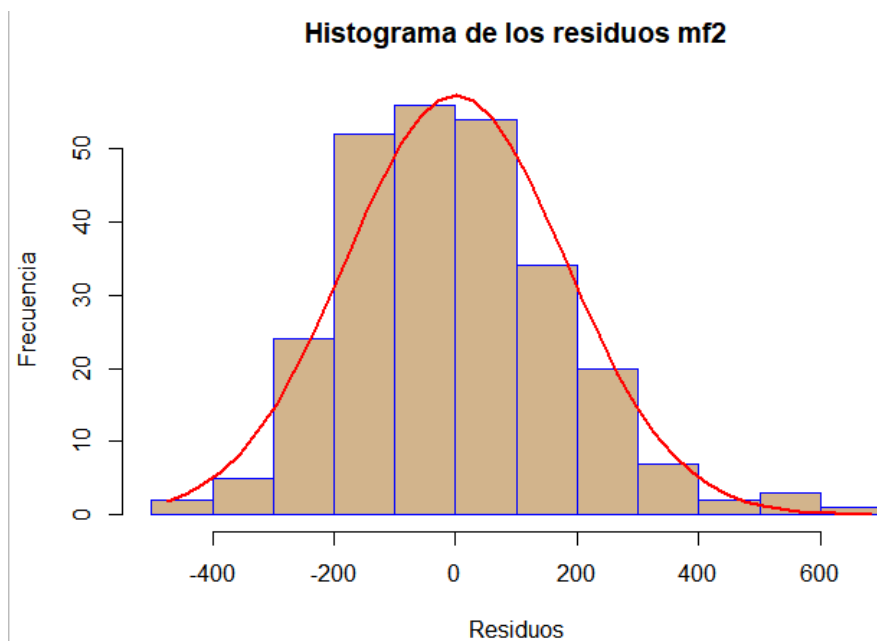


Figura 44. Histograma de los residuos de mf2

Se puede observar que los residuos, tienen una similitud a la curva normal, se realiza también la prueba gráfica de probabilidad normal o de cantiles teóricos Q-Q.

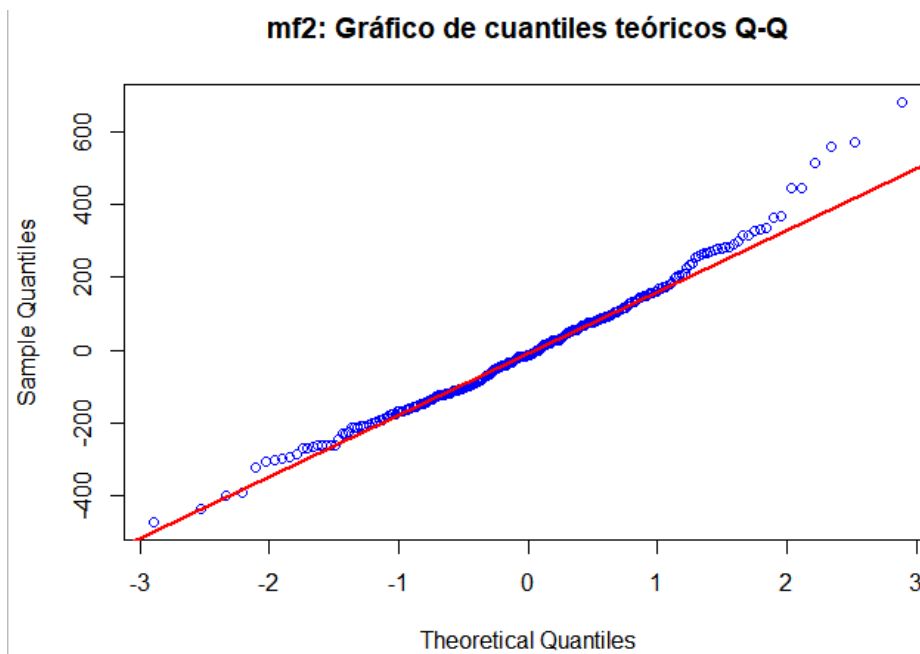


Figura 45. Gráfica Cuantil - Cuantil de mf2

Se puede observar de la Figura 40. Gráfica Cuantil - Cuantil de mf1 anterior, que los residuos se alinean en su mayoría a línea roja, esto significa que los residuos siguen una distribución normal, no obstante se realiza la prueba analítica de Lilliefors con la siguiente hipótesis planteada en (51).

La prueba genera un p-valor de $0.311 > 0.05$, por lo tanto, se acepta la hipótesis nula de que los datos siguen una distribución normal.

3.5.2.3. Pruebas de homocedasticidad de mf2.

Para la prueba gráfica de igualdad de varianza se utiliza el diagrama siguiente:

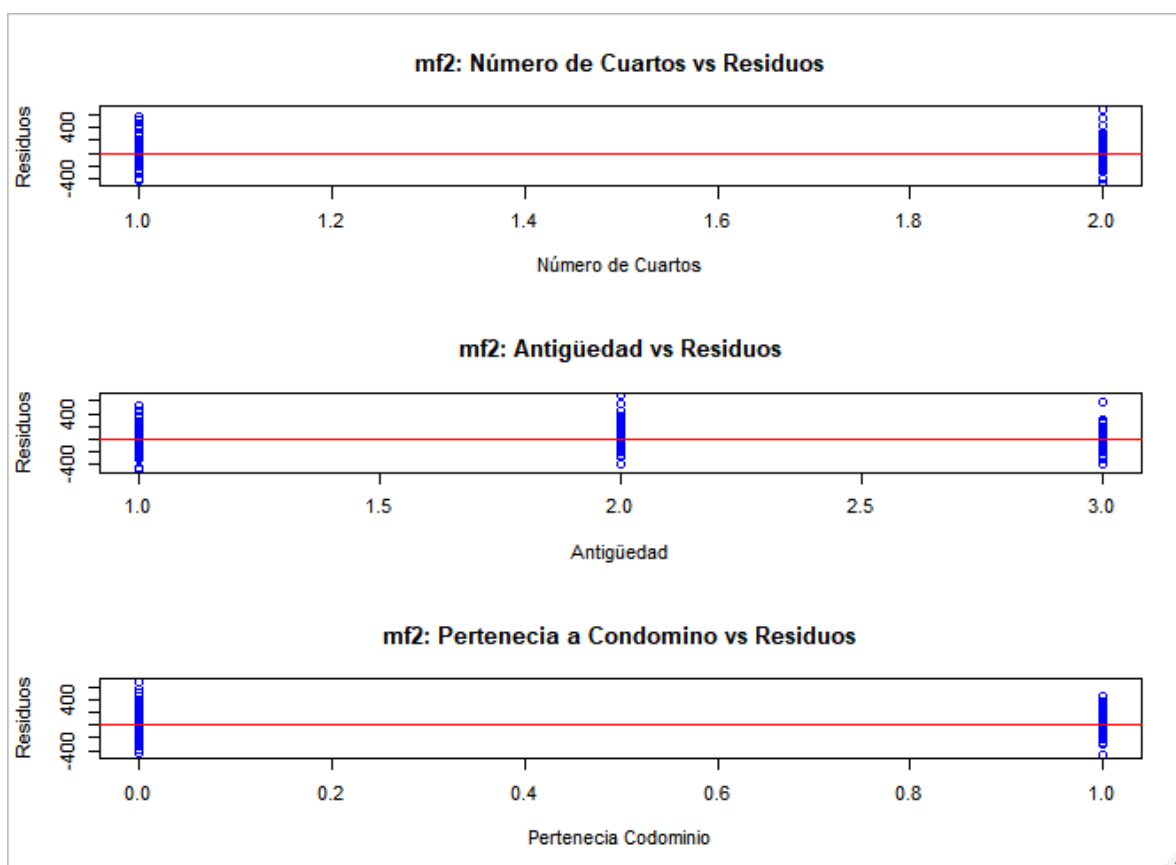


Figura 46. Diagrama de variables vs Residuos mf2

Del diagrama anterior se puede observar que las observaciones se distribuyen equitativamente sobre y bajo la línea roja, lo que sugiere que las variables utilizadas tienen iguales varianzas con los residuos del modelo, no obstante, se realiza la prueba analítica de Bartlett para las variables, con la hipótesis detallada en (52).

La prueba arroja un p-valor de 0.2977, 0.5798 y 0.017 correspondiente a las variables NumCuaC, AntC y Cond, respectivamente, los dos primeros p-valor son mayores que 0.05, lo que quiere decir que, si hay igualdad de varianzas entre las variables y los residuos del modelo, no obstante, la variable Cond no tiene igual varianza, dicha variable es discreta o binaria por lo que su varianza difícilmente coincidirá con los residuos, por tal motivo se pasa por alto la prueba de Bartlett de la variable Cond.

3.5.2.4. Pruebas de identidad de mf_2 .

Para la prueba de identidad se grafica los residuos en eje de las ordenadas y la secuencia de toma de observaciones en el eje de las abscisas, se obtiene:

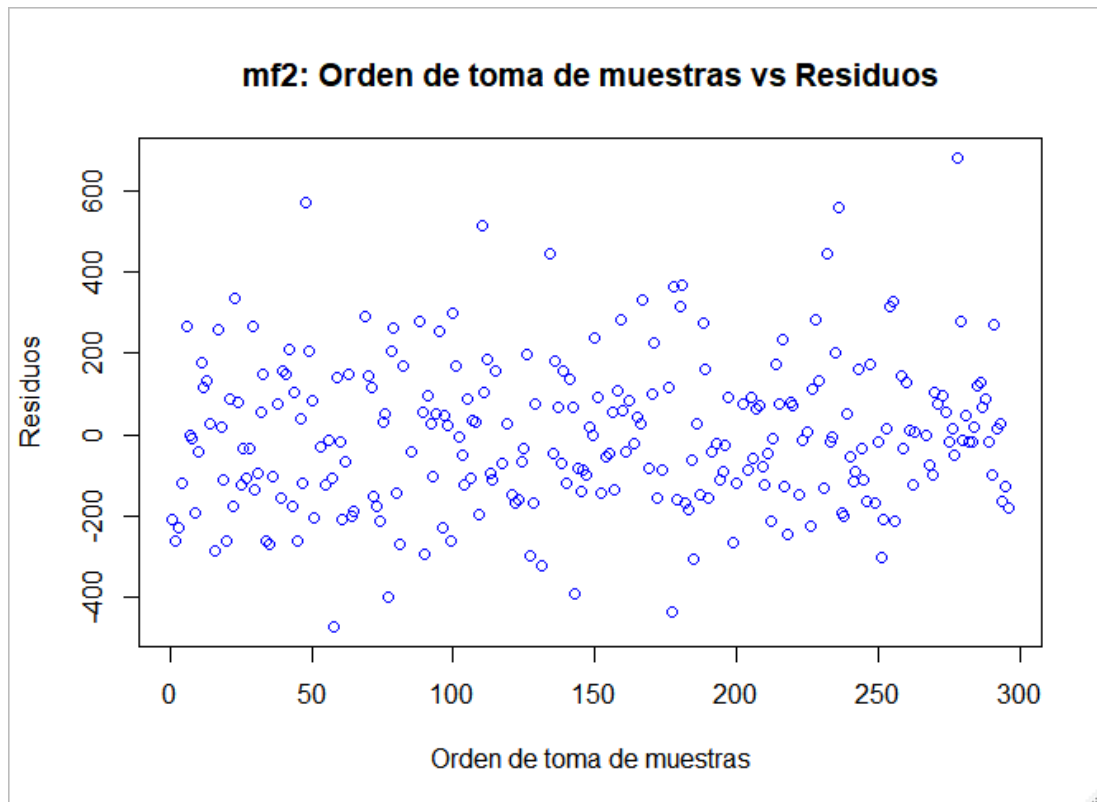


Figura 47. Diagrama de orden de toma de observaciones vs. residuos

Se puede apreciar en el grafico que no existe ningún patrón definido, sin embargo, se realiza prueba analítica de Durbin Watson, la cual define la hipótesis en **¡Error! No se encuentra el origen de la referencia..**

El p-valor que se obtuvo es de $0.562 > 0.05$, por lo tanto, se acepta la hipótesis nula de que existe independencia en la toma de las observaciones.

3.5.3. Modelo de Factores 3.

El modelo de factores 3 se referencia como mf3, con la nomenclatura explicada en 3.5.1., el mf3 tiene la siguiente ecuación:

$$Pxm2T = 381.005 + \beta_0 NumCuaC + \beta_1 AntC + \beta_2 Cond + \beta_3 NumPisC:AntC + \beta_4 NumPisC:Cond \quad (55)$$

Donde:

$$\beta_0 = \begin{cases} -65.412, & \text{si } NumCuaC > 3, \\ 0 & , \quad \text{si } NumCuaC \leq 3 \end{cases}$$

$$\beta_1 = \begin{cases} -83.71, & \text{si } AntC > 10, \\ 47.209, & \text{si } AntC > 5 \wedge AntC \leq 10, \\ 0, & \text{si } AntC \leq 5 \end{cases}$$

$$\beta_2 = \begin{cases} 8.367, & \text{si } Cond = \text{No en conjunto}, \\ 0 & , \quad \text{si } Cond = \text{Si en conjunto} \end{cases}$$

$$\beta_3 = \begin{cases} 228.309, & \text{si } AntC \leq 5 : NumPisC = 2, \\ 95.717, & \text{si } (AntC > 5 \wedge AntC \leq 10) : NumPisC = 2, \\ 176.517, & \text{si } AntC > 10 : NumPisC = 2, \\ 514.37, & \text{si } AntC \leq 5 : NumPisC = 3 \\ 303.774, & \text{si } (AntC > 5 \wedge AntC \leq 10) : NumPisC = 3, \\ 48.117, & \text{si } AntC > 10 : NumPisC = 3, \\ 0, & \text{cualquier otro caso} \end{cases}$$

$$\beta_4 = \begin{cases} 85.472, & \text{si } NumPisC = 2 : Cond = \text{No en conjunto}, \\ -162.956 & \text{si } NumPisC = 3 : Cond = \text{No en conjunto}, \\ 0, & \text{cualquier otro caso} \end{cases}$$

3.5.3.1. Bondad de ajuste de mf3.

Para observar gráficamente la bondad de ajuste se gráfica los Pxm2T real y los Pxm2T predichos como sigue:

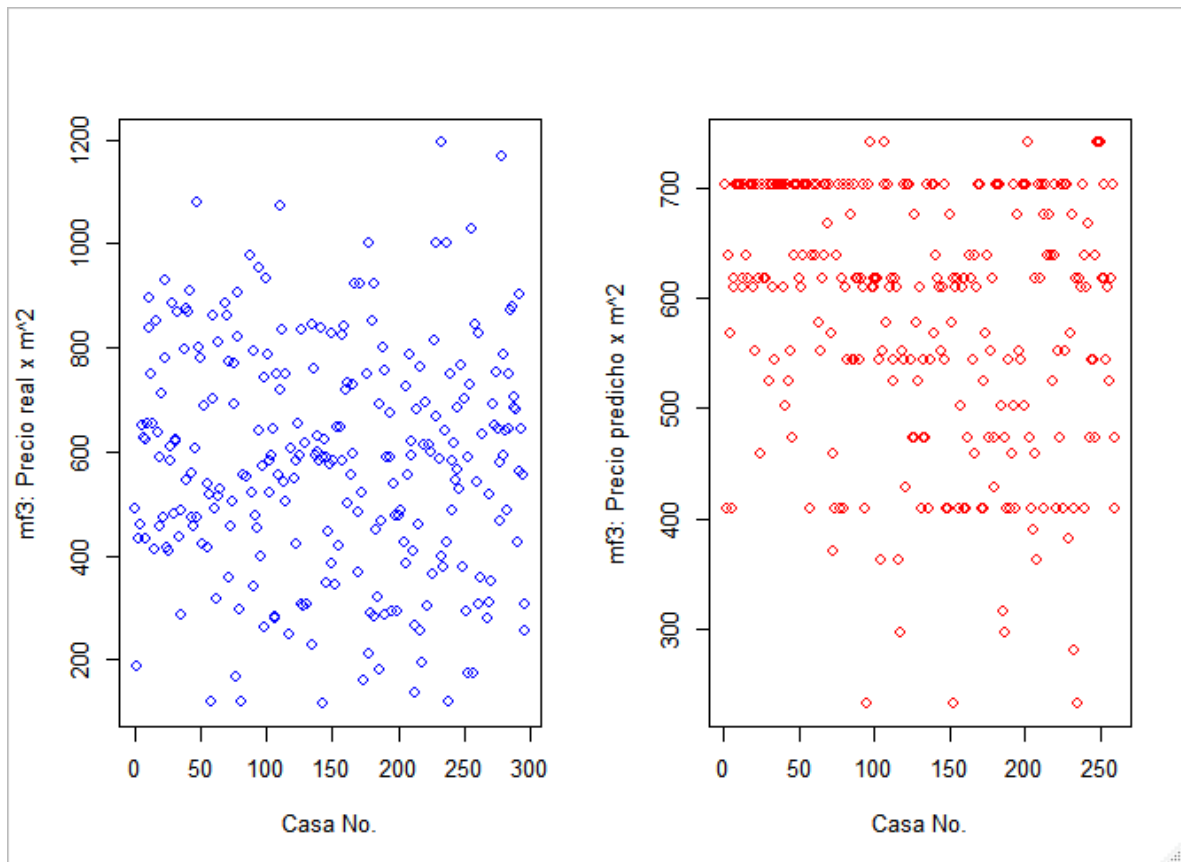


Figura 48. Pxm2T real vs. Pxm2T predicho por mf2

Se observa algunas diferencias entre los Pxm2T real y los Pxm2T predicho, por lo cual la prueba no es concluyente.

El error R^2 múltiple en mf3 es de 0.2838 y el error R^2 ajustado es de 0.249, lo cual indica que la bondad de ajuste del modelo es baja y no se concluye por lo expuesto en 3.5.1.1. La bondad de ajuste se verificará posteriormente mediante validación cruzada.

3.5.3.2. Prueba de Normalidad de mf3.

La prueba de normalidad se la realiza gráficamente mediante la superposición de la curva normal de los residuos de mf3, como se observa a continuación:

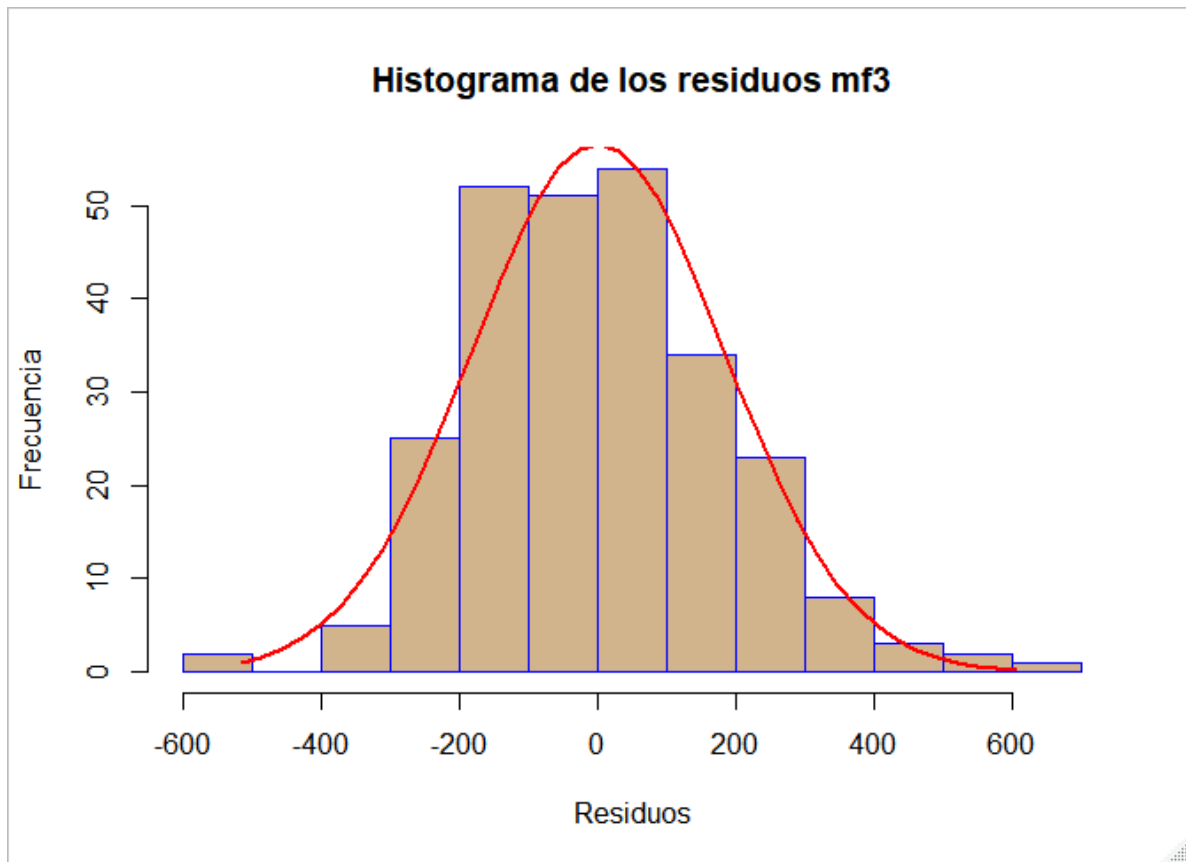


Figura 49. Histograma de los residuos de mf3

Se puede observar que los residuos, tienen una similitud a la curva normal, se realiza también la prueba gráfica de probabilidad normal o de cantiles teóricos Q-Q.

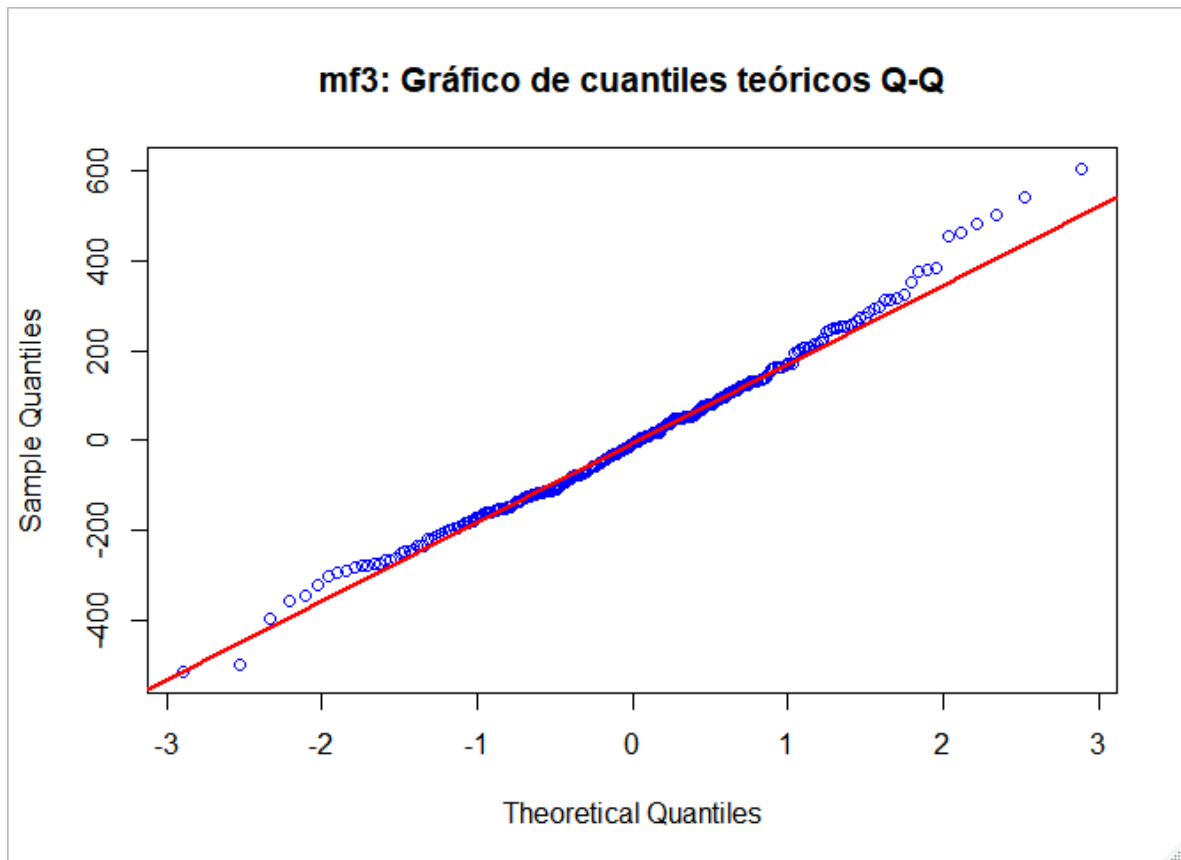


Figura 50. Gráfica Cuantil - Cuantil de mf3

Se puede observar de la Figura 40. Gráfica Cuantil - Cuantil de mf1 anterior, que los residuos se alinean en su mayoría a línea roja, esto significa que los residuos siguen una distribución normal, no obstante se realiza la prueba analítica de Lilliefors con la siguiente hipótesis planteada en **¡Error! No se encuentra el origen de la referencia.**

La prueba genera un p-valor de $0.4189 > 0.05$, por lo tanto, se acepta la hipótesis nula de que los datos siguen una distribución normal.

3.5.2.3. Pruebas de homocedasticidad de mf3.

Para la prueba gráfica de igualdad de varianza se utiliza el diagrama siguiente:

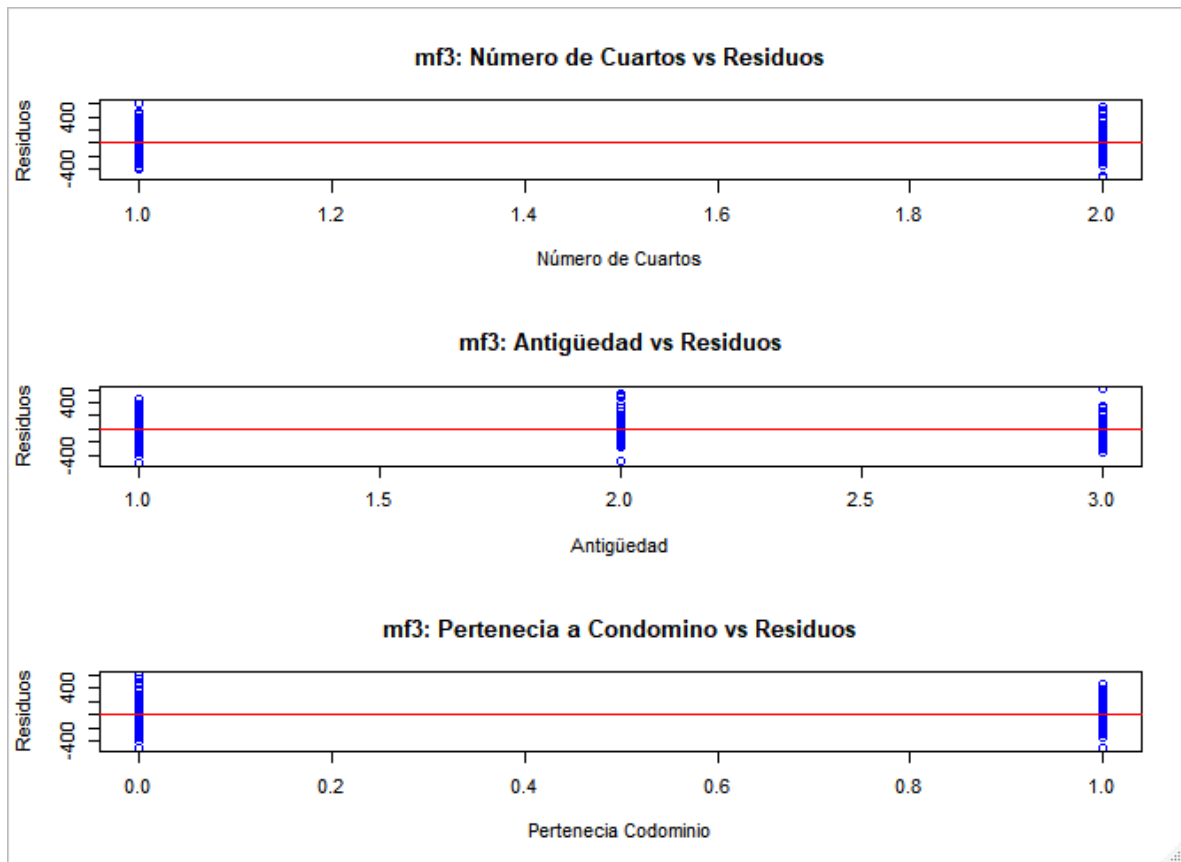


Figura 51. Diagrama de variables vs Residuos mf3

Del diagrama anterior se puede observar que las observaciones se distribuyen equitativamente sobre y bajo la línea roja, lo que sugiere que las variables utilizadas tienen iguales varianzas con los residuos del modelo, no obstante, se realiza la prueba analítica de Bartlett para las variables, con la hipótesis detallada en (52).

La prueba arroja un p-valor de 0.3446, 0.5199 y 0.0718 correspondiente a las variables NumCuaC, AntC y Cond, respectivamente, todos los p-valor son mayores que 0.05, lo que quiere decir que, si hay igualdad de varianzas entre las variables y los residuos del modelo.

3.5.3.4. Pruebas de identidad de mf3.

Para la prueba de identidad se grafica los residuos en eje de las ordenadas y la secuencia de toma de observaciones en el eje de las abscisas, se obtiene:

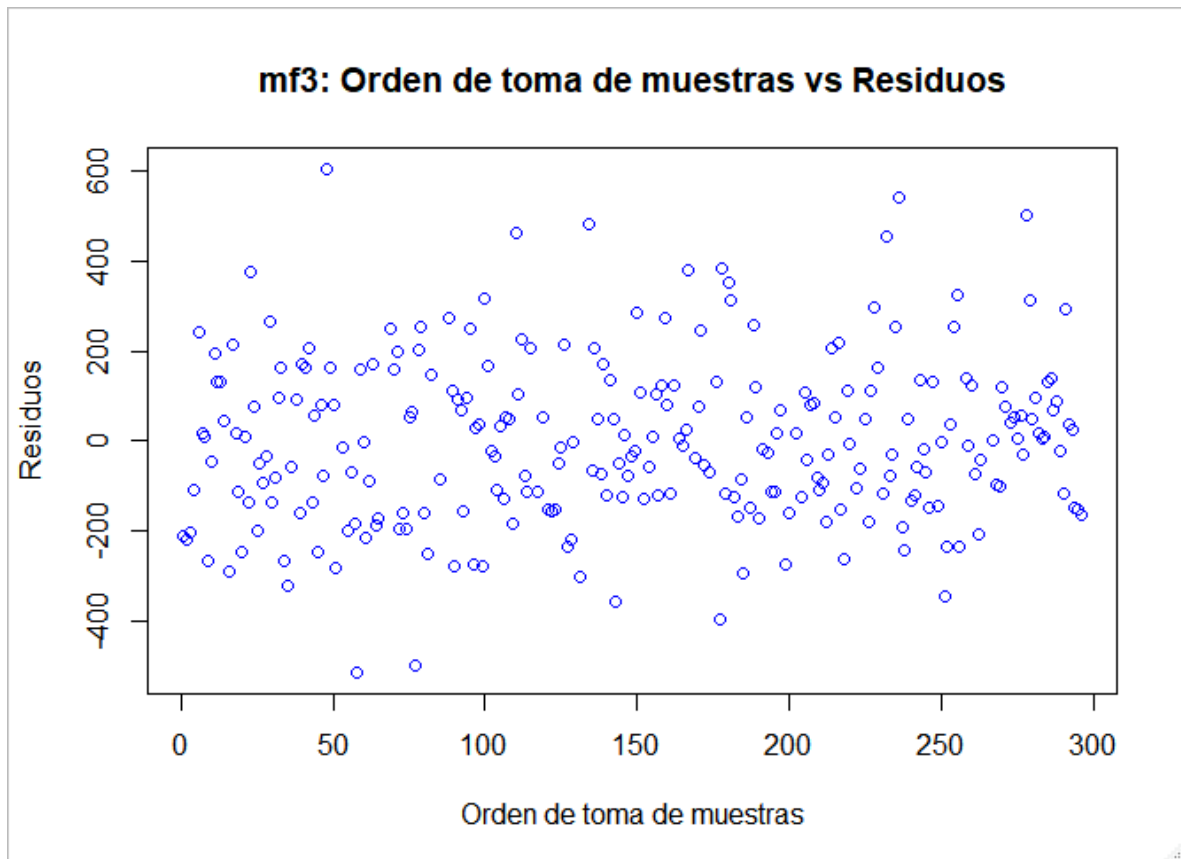


Figura 52. Diagrama de orden de toma de observaciones vs. Residuos de mf3

Se puede apreciar en el grafico que no existe ningún patrón definido, sin embargo, se realiza prueba analítica de Durbin Watson, la cual define la hipótesis en **¡Error! No se encuentra el origen de la referencia..**

El p-valor que se obtuvo es de $0.736 > 0.05$, por lo tanto, se acepta la hipótesis nula de que existe independencia en la toma de las observaciones.

3.6. Correlación Espacial.

Para verificar la correlación espacial de los datos se utiliza el Índice de Moran, para ello se utiliza el software R Studio y se lo hará en función de los errores que genere el modelo de factores, se trabaja con los 3 modelos de la sección anterior, primero se importa en la herramienta fusionables de google para poder observar cómo se distribuyen los puntos geográficamente en el mapa y poder evidenciar gráficamente ubicaciones atípicas en el mapa:

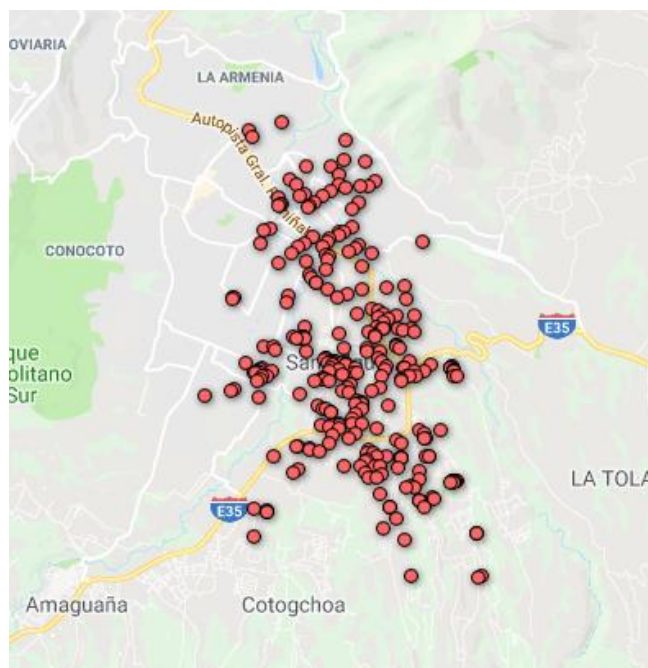


Figura 53. Georeferenciación de las viviendas

Todos los bienes inmuebles se distribuyen uniformemente en el cantón, por lo que no es necesario eliminar datos del estudio.

Para encontrar la matriz de pesos se utiliza el sistema detallado en (39), se calcula primero una matriz de distancias geográficas utilizando la ecuación de Haversine detallada en (48), luego se calcula la matriz de distancia inversa y asignado a la diagonal principal los valores

de 0, ya que la distancia de un predio a sí mismo es 0 y la inversa es infinito, esta matriz final será la matriz de pesos para el análisis del índice de Moran.

Dentro del software R Studio se tiene la función *Moran.I* dentro de la librería *ape*, este retorna un *p-valor*, el cual responde a la siguiente hipótesis:

$$H_0: p - \text{valor} \geq 0.05 \text{ (No existe correlación espacial)} \quad (56)$$

$$H_1: p - \text{valor} < 0.05 \text{ (Existe correlación espacial)}$$

Realizando la prueba para los 3 modelos se obtiene que el *p-valor* es de 0.024, 0.038 y 0.022, para mf1, mf2 y mf3 respectivamente, todos menores que 0,05, por lo tanto, se rechaza la hipótesis nula y se acepta la hipótesis alternativa de que si existe correlación espacial.

De los cuarenta y cuatro modelos de factores corridos solo los tres modelos descritos tienen obtuvieron correlación espacial.

3.7. Regresión Hedónica.

Para la regresión hedónica se realiza una zonificación en función de la ubicación de las casas y se crean las variables tipo dummy, las cuales son variables binarias de pertenecía a una zona específica, se definen las siguientes áreas: zona de la Armenia (ZArm), zona de Ilaló (ZIla), zona del Triángulo (ZTria), zona de la ESPE (ZEspe), zona del Rancho (ZRanc), zona del River mall (ZRiver), zona del Colibrí (ZColi), zona centro de Sangolquí (ZSang), zona cercana a la Panamericana (ZPana), zona de Selva Alegre (ZSelva), zona cercana al Barrio Carlos Gavilanez (ZGavil).

Para definir las zonas se ubicó un punto central de cada zona y se calculó la distancia georreferenciada a esta, el inmueble pertenece a la zona con la cual tenga menor distancia.

A las variables mencionadas se añade las variables: número de habitaciones (NumCua), número de baños (NumBan), número de estacionamientos (NumEst), número de pisos (NumPis), distancia mínima a un centro comercial (DminCC), distancia mínima a centro de salud (DminCS), distancia mínima a centro económico (DminCE) y distancia mínima a centro de educación superior (DminCES), sin embargo, las variables cuantitativas no georreferenciadas se las binariza para hacerlas de tipo dummy, y en las variables georreferenciadas se estandariza los datos para que sean independientes de las unidades de medida, la binarización responde al siguiente criterio:

Tabla 31
Asignación de variables dummy

Variable dummy	Variables Dummy	
	Asignación a variable dummy	
	1	0
DEst	NumEst >0	caso contario
DPis	NumPis >2	caso contario
DCua	NumCua >3	caso contario
DBan	NumBan >2	caso contario
Ant	Ant >0	caso contario
Variable dummy	Asignación a variable dummy	
	1	0
DEst	NumEst >2	caso contario
DPis	NumPis >1	caso contario
DCua	NumCua >2	caso contario
DBan	NumBan >3	caso contario
Ant	Ant >0	caso contario

Con estos dos tipos de asignaciones de variables tipo dummy o binarias se combina y se obtienen los mejores modelos de regresión hedónica.

Se trabaja veintisiete modelos hedónicos, en todos los modelos se hace variaciones de las variables intervinientes con la ayuda del comando *step* y sin tomar en cuenta la significancia

del ANOVA, a diferencia de los modelos de factores, luego se toma cuatro modelos que cumplen con los supuestos de normalidad, homocedasticidad e identidad y que tienen un R^2 mayor, los cuatro mejores modelos obtenidos se los denomina Mrh1, Mrh2, Mrh3 y Mrh4, cuyos resultados se analizan a continuación.

3.7.1. Modelo de regresión hedónica 1.

El modelo de hedónica 1 se referencia como Mrh1, y las variables toman la nomenclatura de la Tabla 31, el Mrh1 tiene la siguiente ecuación:

$$\begin{aligned}
 Pxm2T = & 520.7438 - 64.3901 DCua + 24.2339 DEst - 15.8533 DBan - \\
 & 50.1218 Ant + 24.0824 DPis + 94.9020 Cond - 33.8708 DMinCS + \\
 & 5.9676 DMinCES + 248.7329 ZArm + 68.4411 ZIla + 92.3022 ZTria - \\
 & 0.9232 ZEspe - 4.7622 ZColi + 43.1733 ZPana - 96.2626 ZGavil
 \end{aligned} \tag{57}$$

3.7.1.1. Bondad de ajuste de Mrh1.

Para observar gráficamente la bondad de ajuste se gráfica los Pxm2T real y los Pxm2T predichos como sigue:

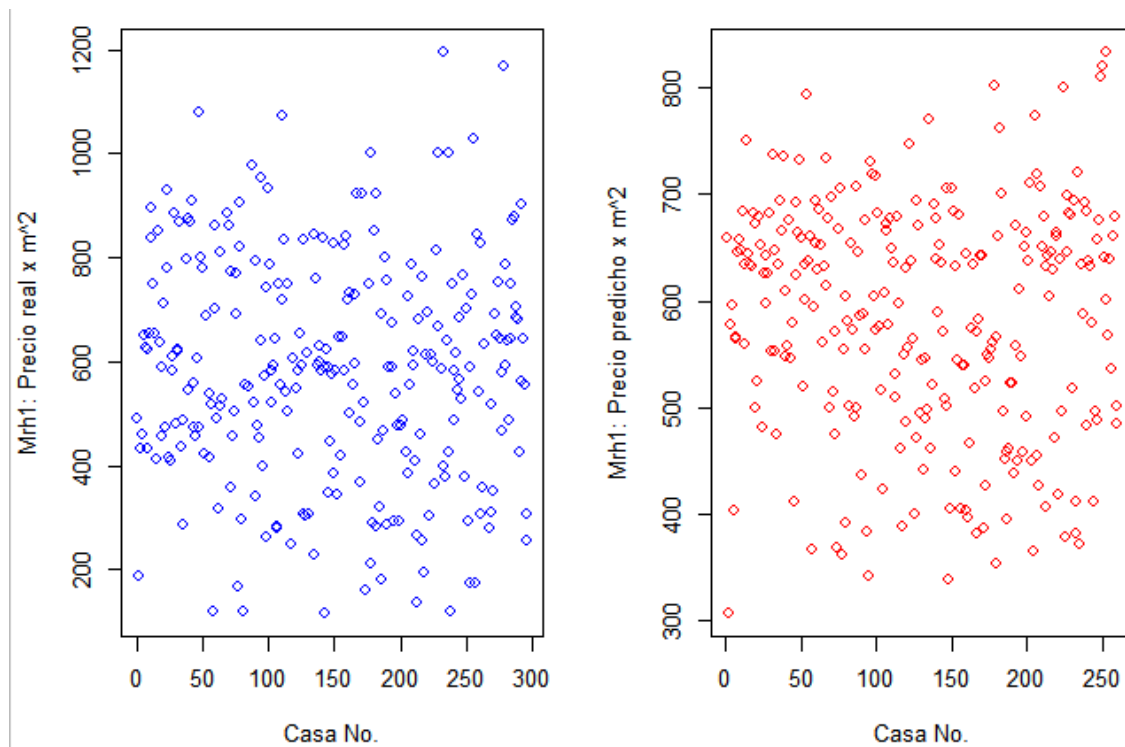


Figura 54. Pxm2T real vs. Pxm2T predicho por Mrh1

De la figura anterior se puede observar que el máximo predicho llega hasta menos de \$900 y el real menos de \$1200, por lo cual la prueba gráfica no es concluyente.

El error R^2 múltiple en Mrh1 es de 0.2539 y el error R^2 ajustado es de 0.208, lo cual indica que la bondad de ajuste del modelo es baja y no se concluye por lo expuesto en 3.5.1.1.

La bondad de ajuste se verificará posteriormente mediante validación cruzada.

3.7.1.2. Prueba de Normalidad de Mrh1.

La prueba de normalidad se la realiza gráficamente mediante la superposición de la curva normal de los residuos de Mrh1, como se observa a continuación:

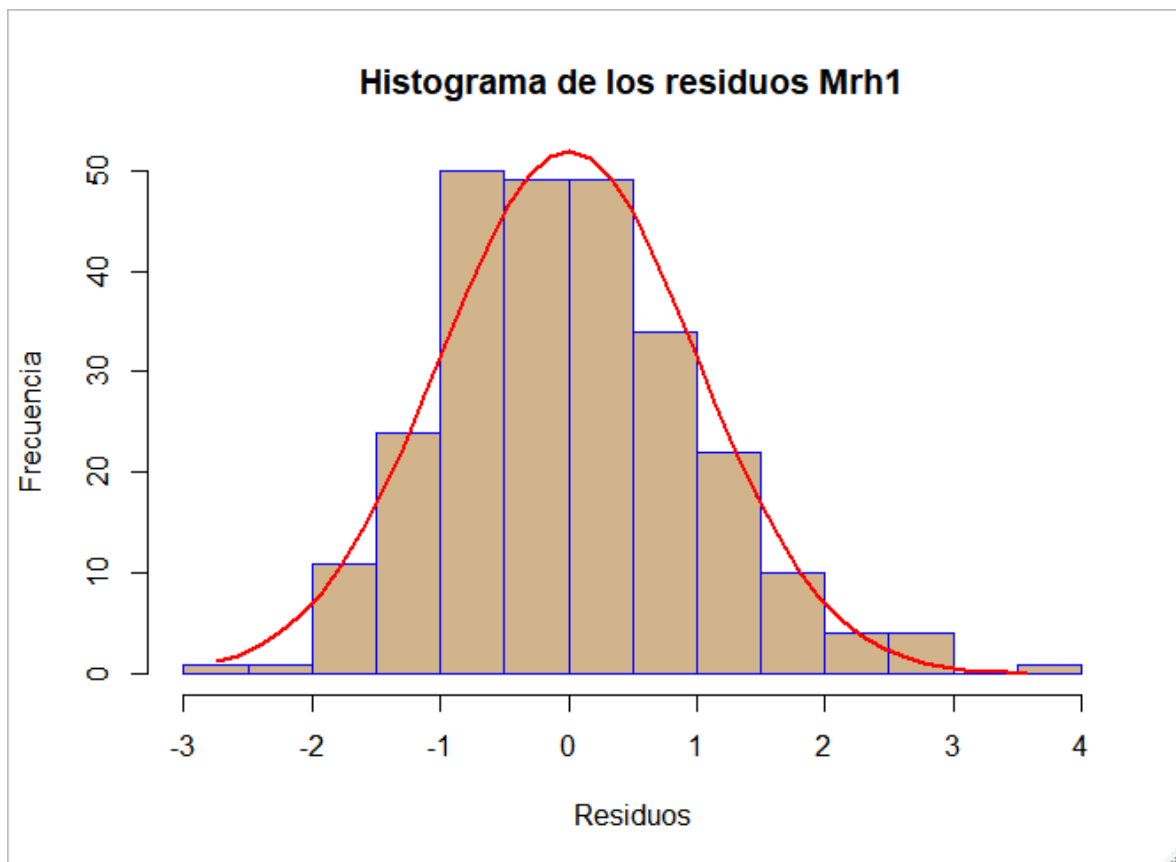


Figura 55. Histograma de los residuos de Mrh1

Se puede observar que los residuos, tienen una similitud a la curva normal, se realiza también la prueba gráfica de probabilidad normal o de cantiles teóricos Q-Q.

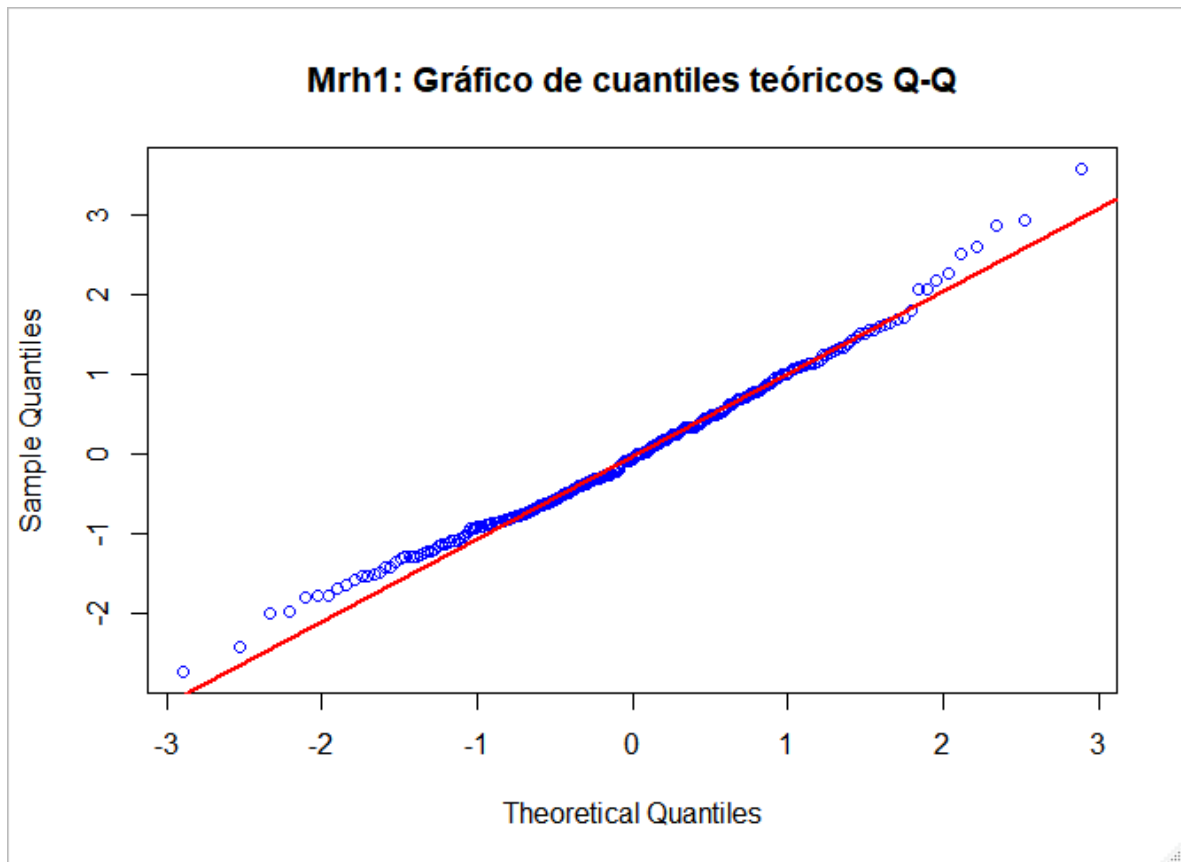


Figura 56. Gráfica Cuantil - Cuantil de Mrh1

Se puede observar de la figura anterior, que los residuos se alinean en su mayoría a línea roja, esto significa que los residuos siguen una distribución normal, no obstante, se realiza la prueba analítica de Lilliefors con la siguiente hipótesis descrita en (51)

La prueba genera un p-valor de $0.09467 > 0.05$, por lo tanto, se acepta la hipótesis nula de que los datos siguen una distribución normal.

3.7.1.3. Pruebas de homocedasticidad de Mrh1.

Para la prueba gráfica de igualdad de varianzas se utiliza el diagrama siguiente:

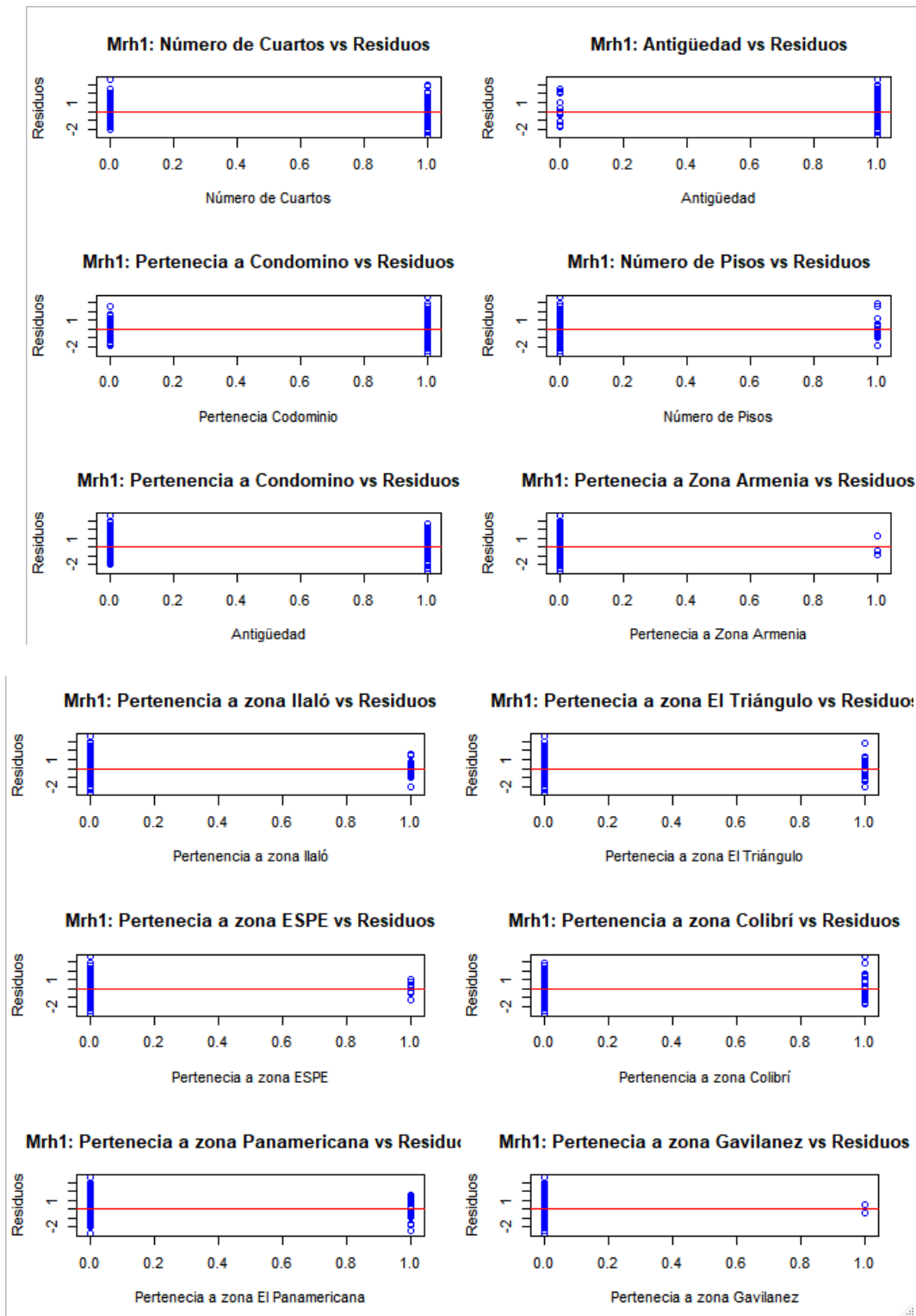


Figura 57. Diagrama de variables vs Residuos Mrh1

Del diagrama anterior se puede observar que las observaciones se distribuyen equitativamente sobre y bajo la línea roja, lo que sugiere que las variables utilizadas tienen iguales varianzas con los residuos del modelo, no obstante, se realiza la prueba analítica de Bartlett para las variables, con la hipótesis detallada en (52).

La prueba arroja un p-valor de 0.23, 0.49, 0.38, 0.05, 0.88, 0.17, 0.4, 0.25, 0.05, 0.36 y 0.62 correspondiente a las variables DCua, DBan, DPis, Cond, ZArm, ZIla, ZTria, ZEspe, ZColi, ZPana y ZGavil, respectivamente, todos los p-valor son mayores que 0.05, en consecuencia, se acepta la hipótesis nula de que si hay igualdad de varianzas entre las variables y los residuos del modelo.

3.7.1.4. Pruebas de identidad de Mrh1.

Para la prueba de identidad se grafica los residuos en eje de las ordenadas y la secuencia de toma de observaciones en el eje de las abscisas, se obtiene:

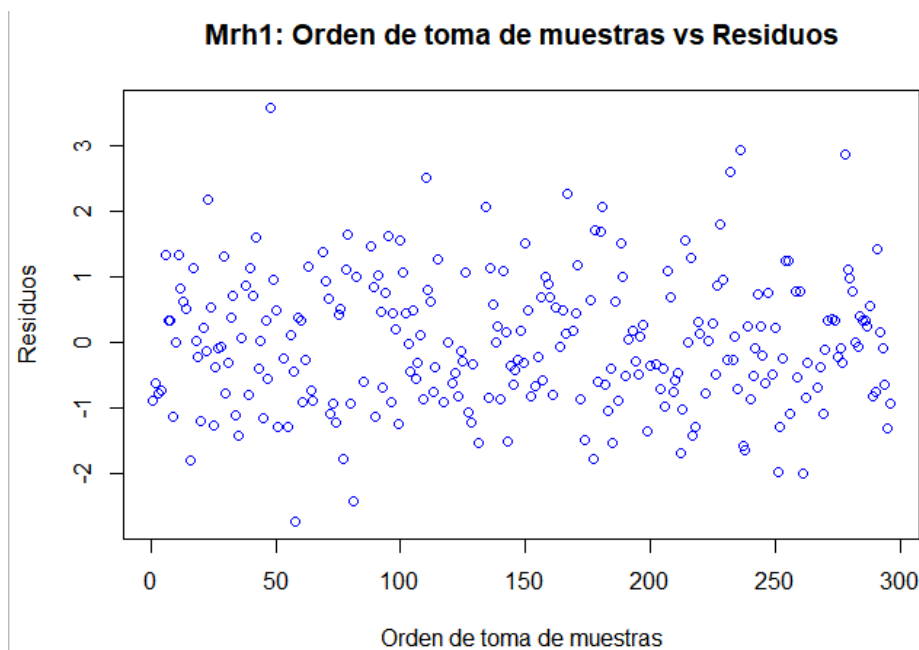


Figura 58. Diagrama de orden de toma de observaciones vs. Residuos de Mrh1

Se puede apreciar en el grafico que no existe ningún patrón definido, sin embargo, se realiza prueba analítica de Durbin Watson, la cual define la hipótesis detallada en (53).

El p-valor que se obtuvo es de $0.662 > 0.05$, por lo tanto, se acepta la hipótesis nula de que existe independencia en la toma de las observaciones.

3.7.2. Modelo de regresión hedónica 2.

El modelo de hedónica 2 se referencia como Mrh2, y tiene la siguiente ecuación:

$$\begin{aligned}
 Pxm2T = & 519.837 - 64.4701DCua + 24.5699 DEst - 16.1031 DBan - \\
 & 58.2141 Ant + 24.4022 DPis + 94.6287 Cond - 34.1585 DMinCS + \\
 & 6.3609 DMinCES + 250.1966 ZArm + 69.5699 ZIla + 93.3453 ZTria + \\
 & 0.2378 ZEspe + 44.2050 ZPana - 96.3102 ZGavil
 \end{aligned}
 \tag{58}$$

3.7.2.1. Bondad de ajuste de Mrh2.

Para observar gráficamente la bondad de ajuste se gráfica los Pxm2T real y los Pxm2T predichos como sigue:

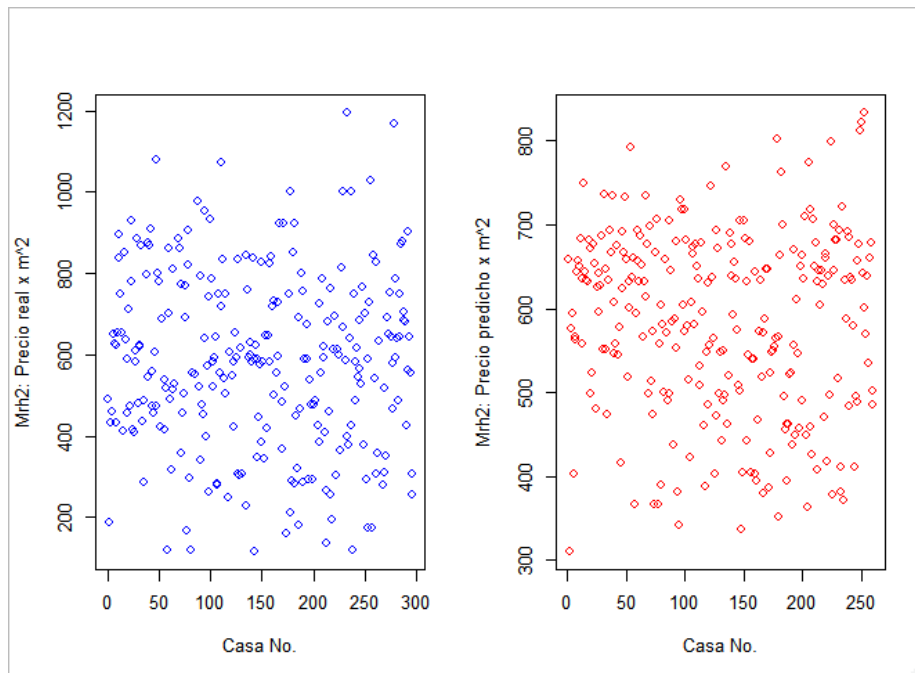


Figura 59. Pxm2T real vs. Pxm2T predicho por Mrh2

De la figura anterior se puede observar que existen diferencias, por lo cual la prueba gráfica no es concluyente.

El error R^2 múltiple en Mrh2 es de 0.2538 y el error R^2 ajustado es de 0.2112, lo cual indica que la bondad de ajuste del modelo es baja y no se concluye por lo expuesto en 3.5.1.1.

La bondad de ajuste se verificará posteriormente mediante validación cruzada.

3.7.2.2. Prueba de Normalidad de Mrh2.

La prueba de normalidad se la realiza gráficamente mediante la superposición de la curva normal de los residuos de Mrh2, como se observa a continuación:

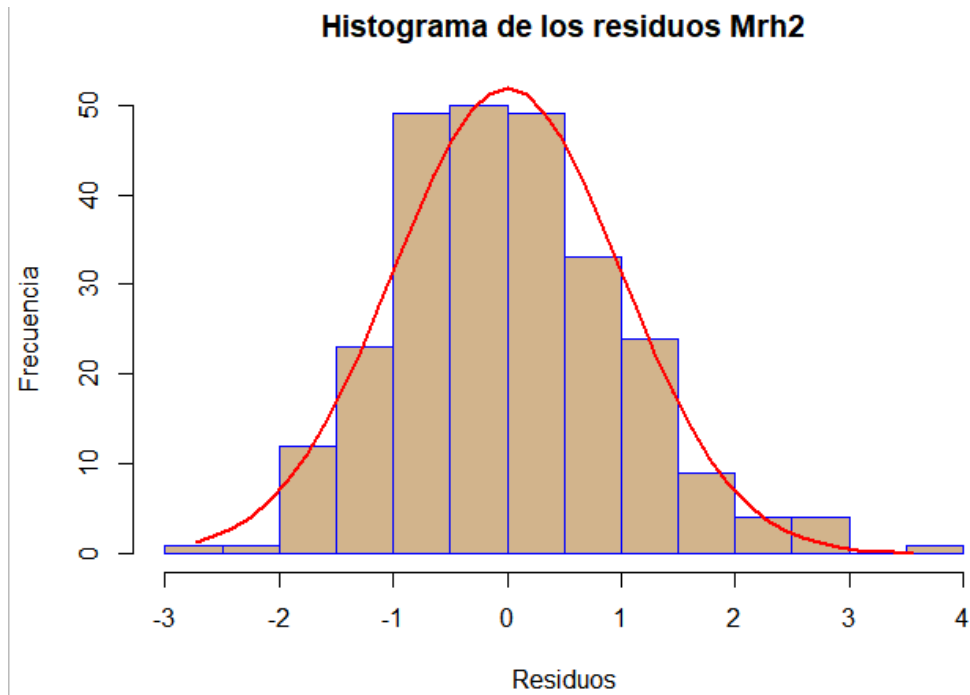


Figura 60. Histograma de los residuos de Mrh2

Se puede observar que los residuos, tienen una similitud a la curva normal, se realiza también la prueba gráfica de probabilidad normal o de cantiles teóricos Q-Q.

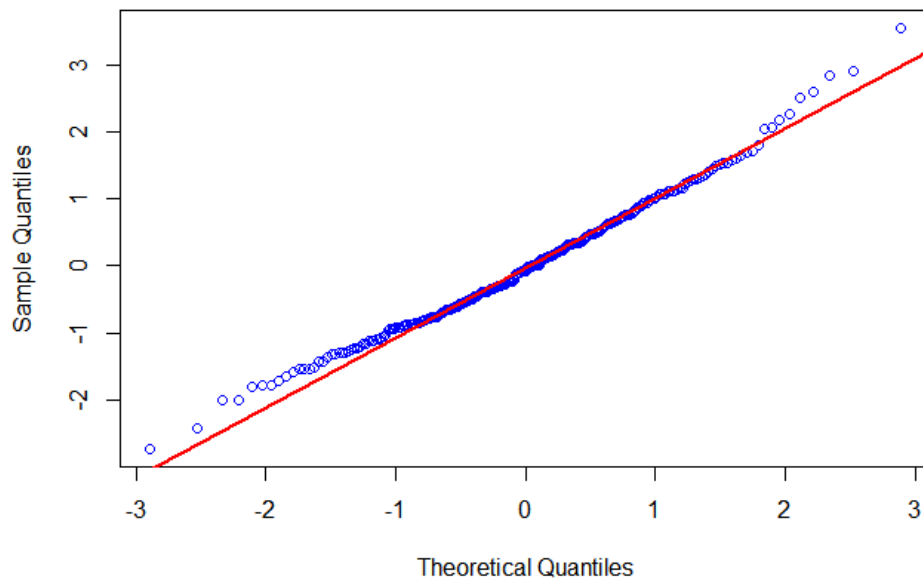


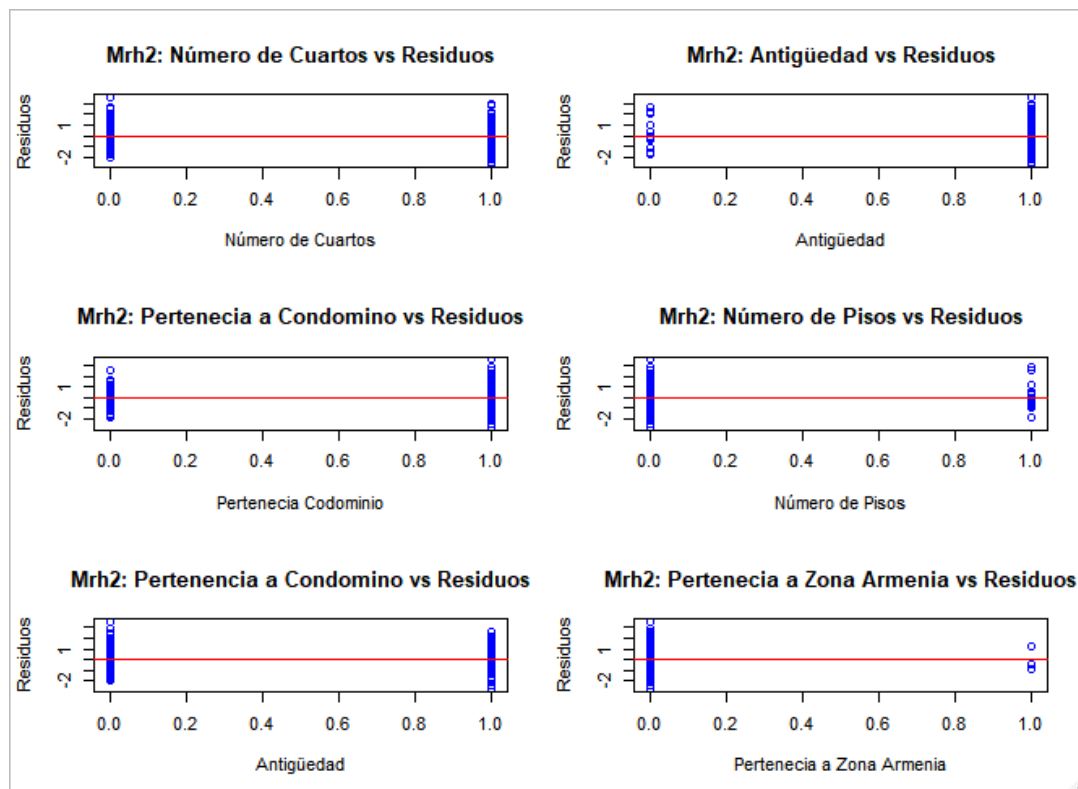
Figura 61. Gráfica Cuantil - Cuantil de Mrh2

Se puede observar de la figura anterior, que los residuos se alinean en su mayoría a línea roja, esto significa que los residuos siguen una distribución normal, no obstante, se realiza la prueba analítica de Lilliefors con la siguiente hipótesis descrita en (51)

La prueba genera un p-valor de $0.09725 > 0.05$, por lo tanto, se acepta la hipótesis nula de que los datos siguen una distribución normal.

3.7.2.3. Pruebas de homocedasticidad de Mrh2.

Para la prueba gráfica de igualdad de varianza se utiliza el diagrama siguiente:



CONTINUA

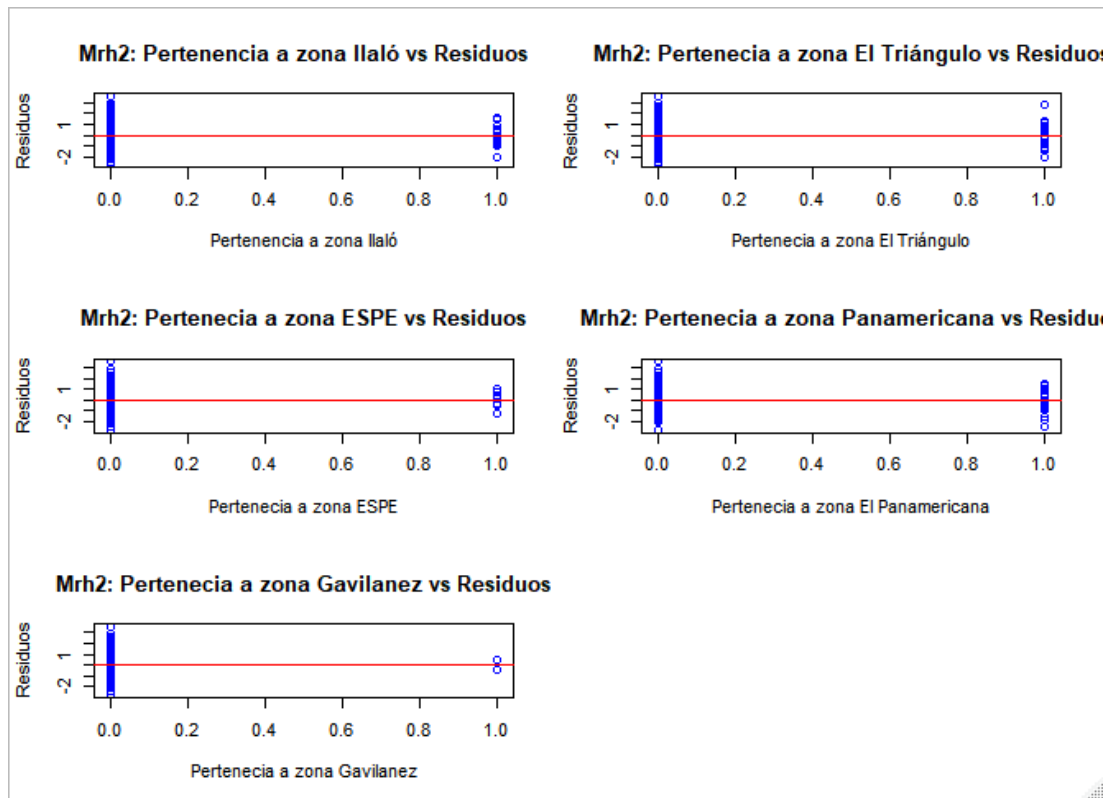


Figura 62. Diagrama de variables vs Residuos Mrh2

Del diagrama anterior se puede observar que las observaciones se distribuyen equitativamente sobre y bajo la línea roja, lo que sugiere que las variables utilizadas tienen iguales varianzas con los residuos del modelo, no obstante, se realiza la prueba analítica de Bartlett para las variables, con la hipótesis detallada en (52).

La prueba arroja un p-valor de 0.23, 0.49, 0.38, 0.05, 0.88, 0.17, 0.41, 0.26, 0.35 y 0.62 correspondiente a las variables DCua, DBan, DPis, Cond, ZArm, ZIla, ZTria, ZColi, ZPana y ZGavil, respectivamente, todos los p-valor son mayores que 0.05, en consecuencia, se acepta la hipótesis nula de que si hay igualdad de varianzas entre las variables y los residuos del modelo.

3.7.2.4. Pruebas de identidad de Mrh2.

Para la prueba de identidad se grafica los residuos en eje de las ordenadas y la secuencia de toma de observaciones en el eje de las abscisas, se obtiene:

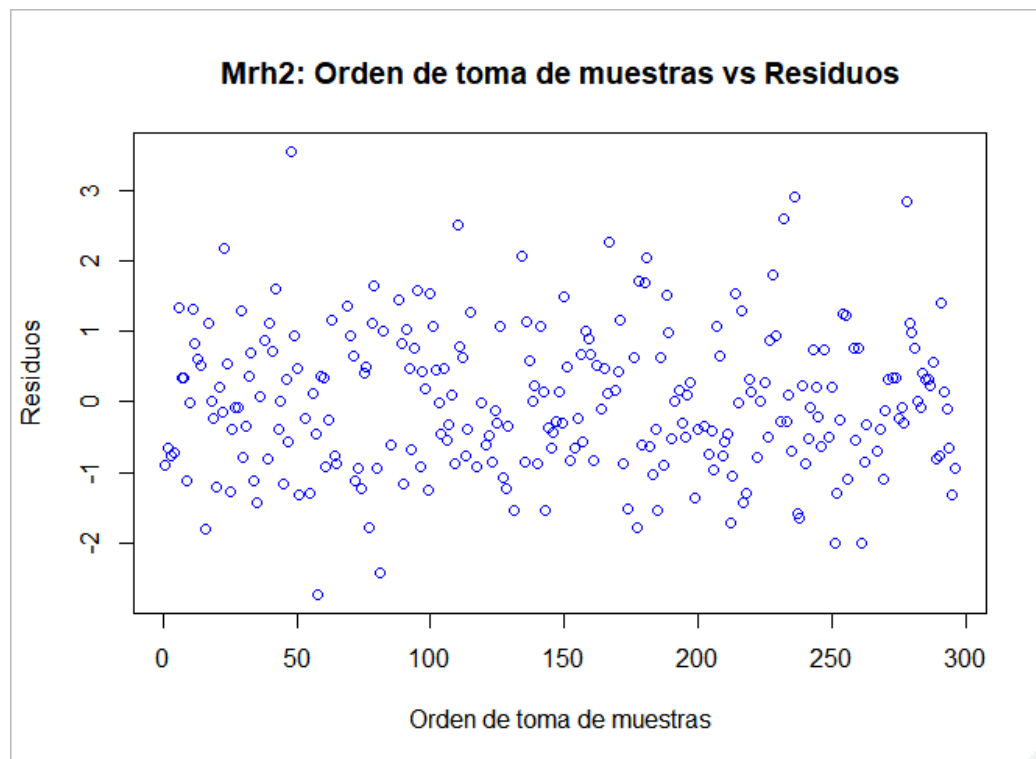


Figura 63. Diagrama de orden de toma de observaciones vs. Residuos de Mrh2

Se puede apreciar en el grafico que no existe ningún patrón definido, sin embargo, se realiza prueba analítica de Durbin Watson, la cual define la hipótesis detallada en (53).

El p-valor que se obtuvo es de $0.754 > 0.05$, por lo tanto, se acepta la hipótesis nula de que existe independencia en la toma de las observaciones.

3.7.3. Modelo de regresión hedónica 3.

El modelo de hedónica 3 se referencia como Mrh3, y tiene la siguiente ecuación:

$$\begin{aligned}
 Pxm2T = & 325.047 + 73.2DCua - 72.095 DEst - 43.248 DBan - 56.532 Ant + \\
 & 164.802 DPis + 69.642 Cond - 30.743 DMinCS + 10.208 DMinCES + \\
 & 257.942 ZArm + 81.014 ZIla + 114.625 ZTria + 5.476 ZEspe + \\
 & 57.663 ZPana - 64.862 ZGavil
 \end{aligned}
 \tag{59}$$

3.7.3.1. Bondad de ajuste de Mrh3.

Para observar gráficamente la bondad de ajuste se gráfica los Pxm2T real y los Pxm2T predichos como sigue:

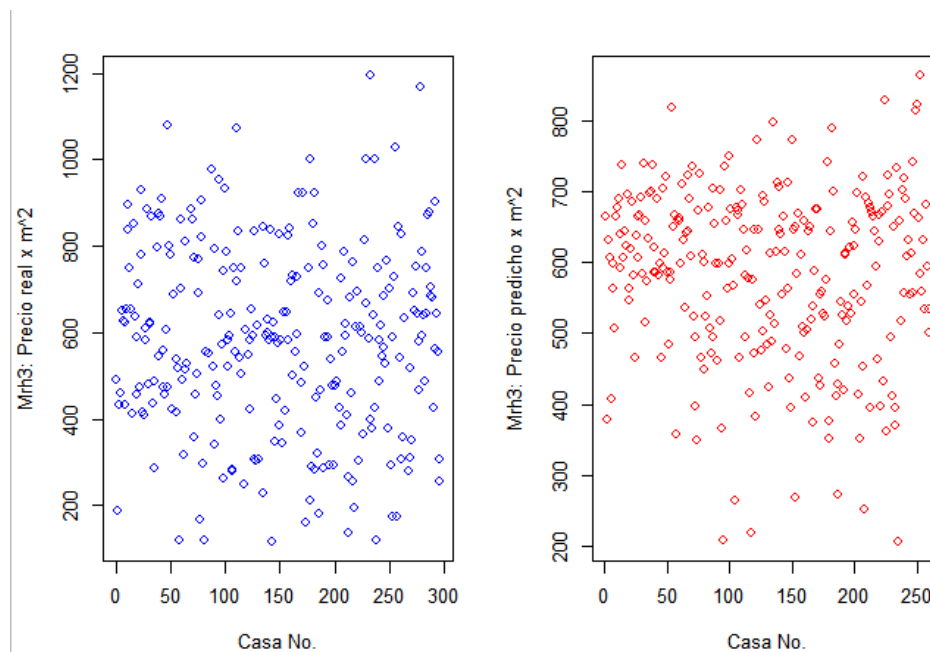


Figura 64. Pxm2T real vs. Pxm2T predicho por Mrh3

De la figura anterior se puede observar que existen diferencias, por lo cual la prueba gráfica no es concluyente.

El error R^2 múltiple en Mrh3 es de 0.3082 y el error R^2 ajustado es de 0.2656, lo cual indica que la bondad de ajuste del modelo es baja y no se concluye por lo expuesto en 3.5.1.1. La bondad de ajuste se verificará posteriormente mediante validación cruzada.

3.7.3.2. Prueba de Normalidad de Mrh3.

La prueba de normalidad se la realiza gráficamente mediante la superposición de la curva normal de los residuos de Mrh3, como se observa a continuación:

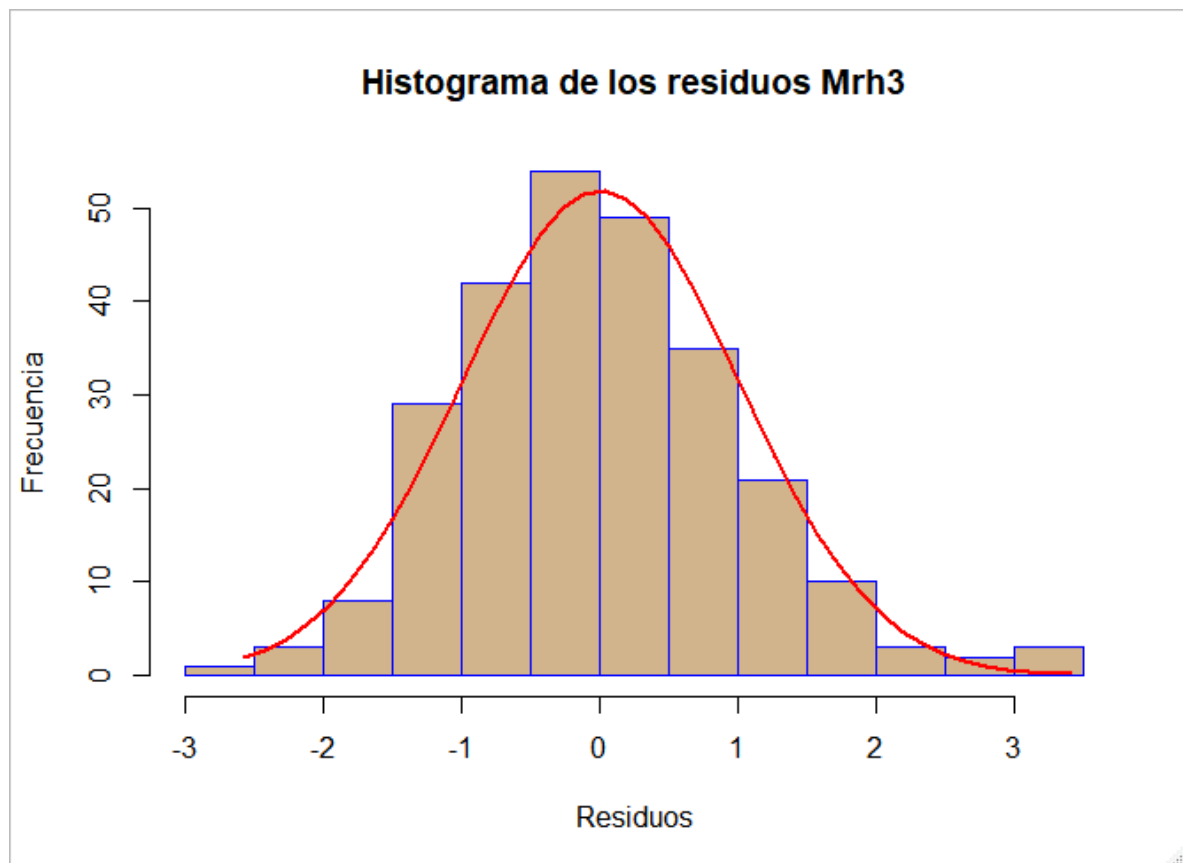


Figura 65. Histograma de los residuos de Mrh3

Se puede observar que los residuos, tienen una similitud a la curva normal, se realiza también la prueba gráfica de probabilidad normal o de cantiles teóricos Q-Q.

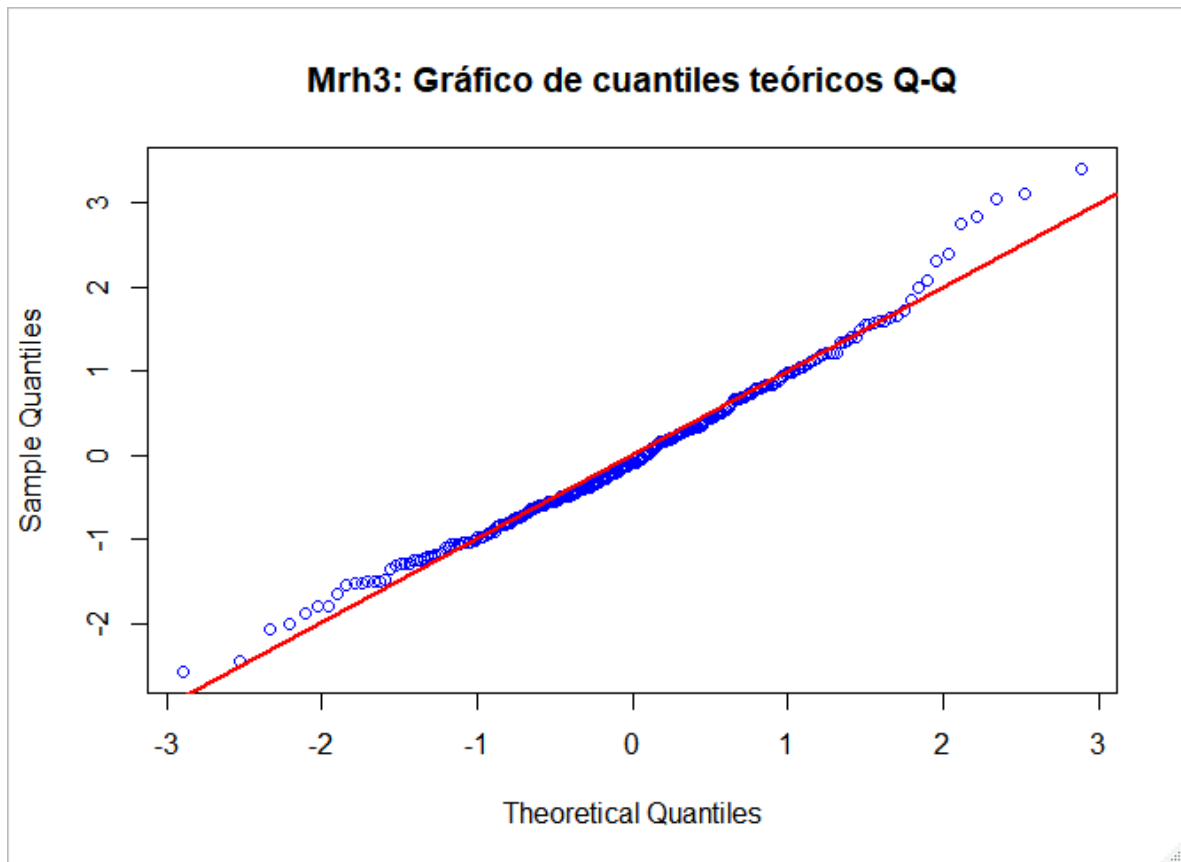


Figura 66. Gráfica Cuantil - Cuantil de Mrh3

Se puede observar de la figura anterior, que los residuos se alinean en su mayoría a línea roja, esto significa que los residuos siguen una distribución normal, no obstante, se realiza la prueba analítica de Lilliefors con la siguiente hipótesis descrita en (51)

La prueba genera un p-valor de $0.2056 > 0.05$, por lo tanto, se acepta la hipótesis nula de que los datos siguen una distribución normal.

3.7.3.3. Pruebas de homocedasticidad de Mrh3.

Para la prueba gráfica de igualdad de varianzas se utiliza el diagrama siguiente:

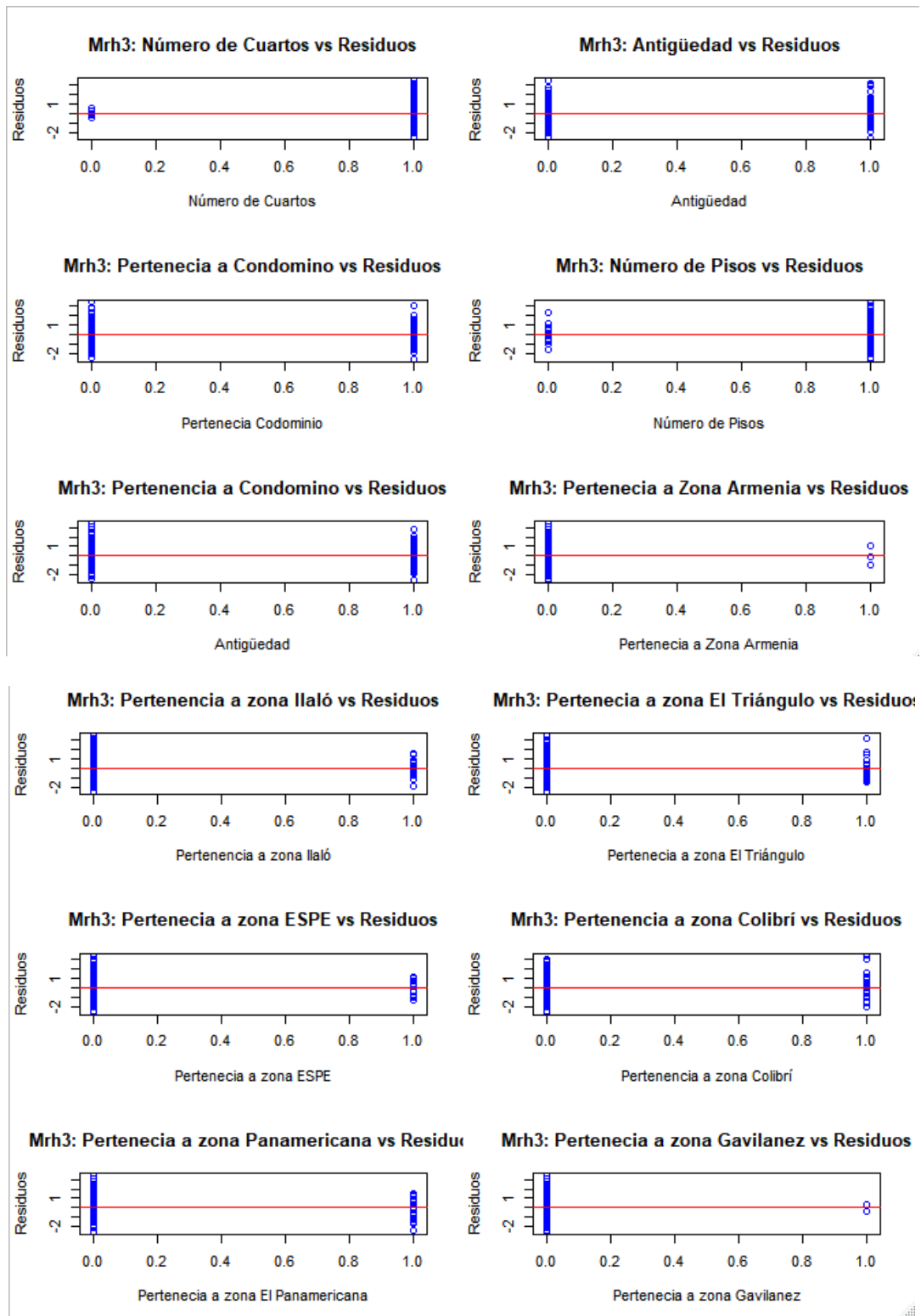


Figura 67. Diagrama de variables vs Residuos Mrh3

Del diagrama anterior se puede observar que las observaciones se distribuyen equitativamente sobre y bajo la línea roja, lo que sugiere que las variables utilizadas tienen iguales varianzas con los residuos del modelo, no obstante, se realiza la prueba analítica de Bartlett para las variables, con la hipótesis detallada en (52).

La prueba arroja un p-valor de 0.26, 0.44, 0.70, 0.9, 0.26, 0.44, 0.32, 0.09, 0.34 y 0.44 correspondiente a las variables DEst, DBan, DPis, ZArm, ZIla, ZTria, Zespe, ZColi, ZPana y ZGavil, respectivamente, todos los p-valor son mayores que 0.05, en consecuencia, se acepta la hipótesis nula de que si hay igualdad de varianzas entre las variables y los residuos del modelo.

3.7.3.4. Pruebas de identidad de Mrh3.

Para la prueba de identidad se grafica los residuos en eje de las ordenadas y la secuencia de toma de observaciones en el eje de las abscisas, se obtiene:

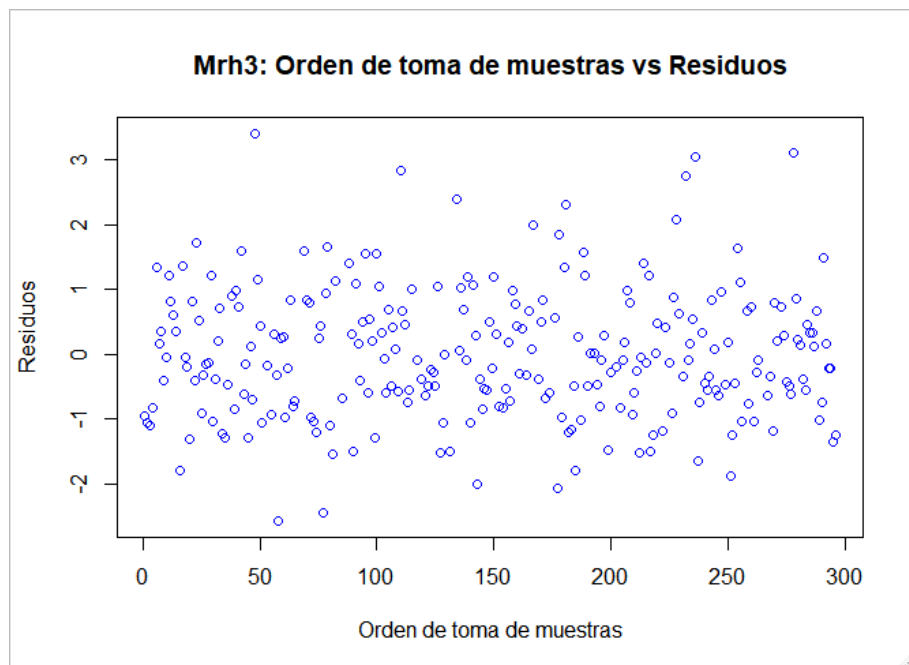


Figura 68. Diagrama de orden de toma de observaciones vs. Residuos de Mrh3

Se puede apreciar en el grafico que no existe ningún patrón definido, sin embargo, se realiza prueba analítica de Durbin Watson, la cual define la hipótesis detallada en (53).

El p-valor que se obtuvo es de $0.484 > 0.05$, por lo tanto, se acepta la hipótesis nula de que existe independencia en la toma de las observaciones.

3.7.4. Modelo de regresión hedónica 4.

El modelo de hedónica 4 se referencia como Mrh4, y tiene la siguiente ecuación:

$$\begin{aligned}
 Pxm2T = & 326.721 + 76.06DCua - 71.973 DEst - 42.333 DBan - 56.346 Ant + \\
 & 162.124 DPis + 70.695 Cond - 29.888 DMinCS + 8.985 DMinCES + \\
 & 254.642 ZArm + 76.648 ZIla + 111.181 ZTria + 1.927 ZEspe + \\
 & 54.294 ZPana - 64.677 ZGavil
 \end{aligned}
 \tag{60}$$

3.7.4.1. Bondad de ajuste de Mrh4.

Para observar gráficamente la bondad de ajuste se gráfica los Pxm2T real y los Pxm2T predichos como sigue:

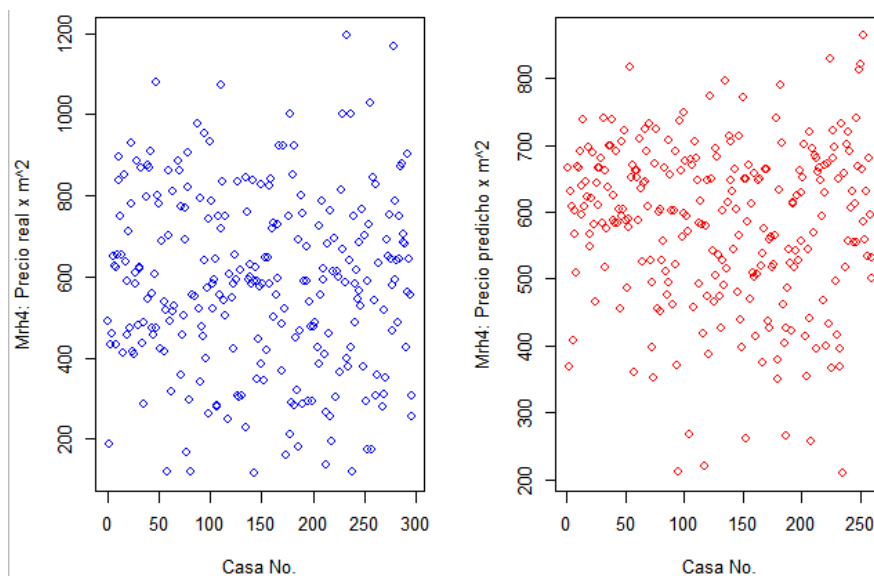


Figura 69. Pxm2T real vs. Pxm2T predicho por Mrh4

De la figura anterior se puede observar que existen diferencias, por lo cual la prueba gráfica no es concluyente.

El error R^2 múltiple en Mrh4 es de 0.3078 y el error R^2 ajustado es de 0.2682, lo cual indica que la bondad de ajuste del modelo es baja y no se concluye por lo expuesto en 3.5.1.1.

La bondad de ajuste se verificará posteriormente mediante validación cruzada.

3.7.4.2. Prueba de Normalidad de Mrh4.

La prueba de normalidad se la realiza gráficamente mediante la superposición de la curva normal de los residuos de Mrh4, como se observa a continuación:

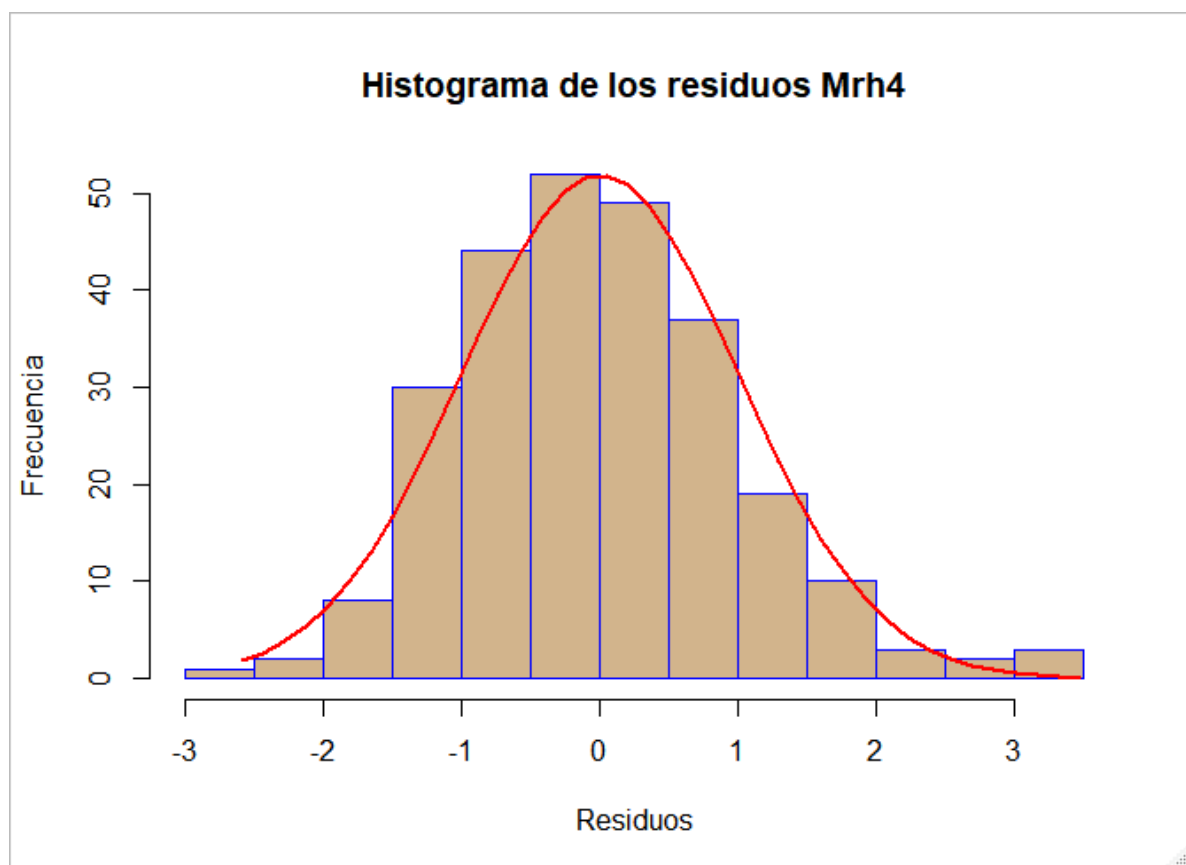


Figura 70. Histograma de los residuos de Mrh4

Se puede observar que los residuos, tienen una similitud a la curva normal, se realiza también la prueba gráfica de probabilidad normal o de cantiles teóricos Q-Q.

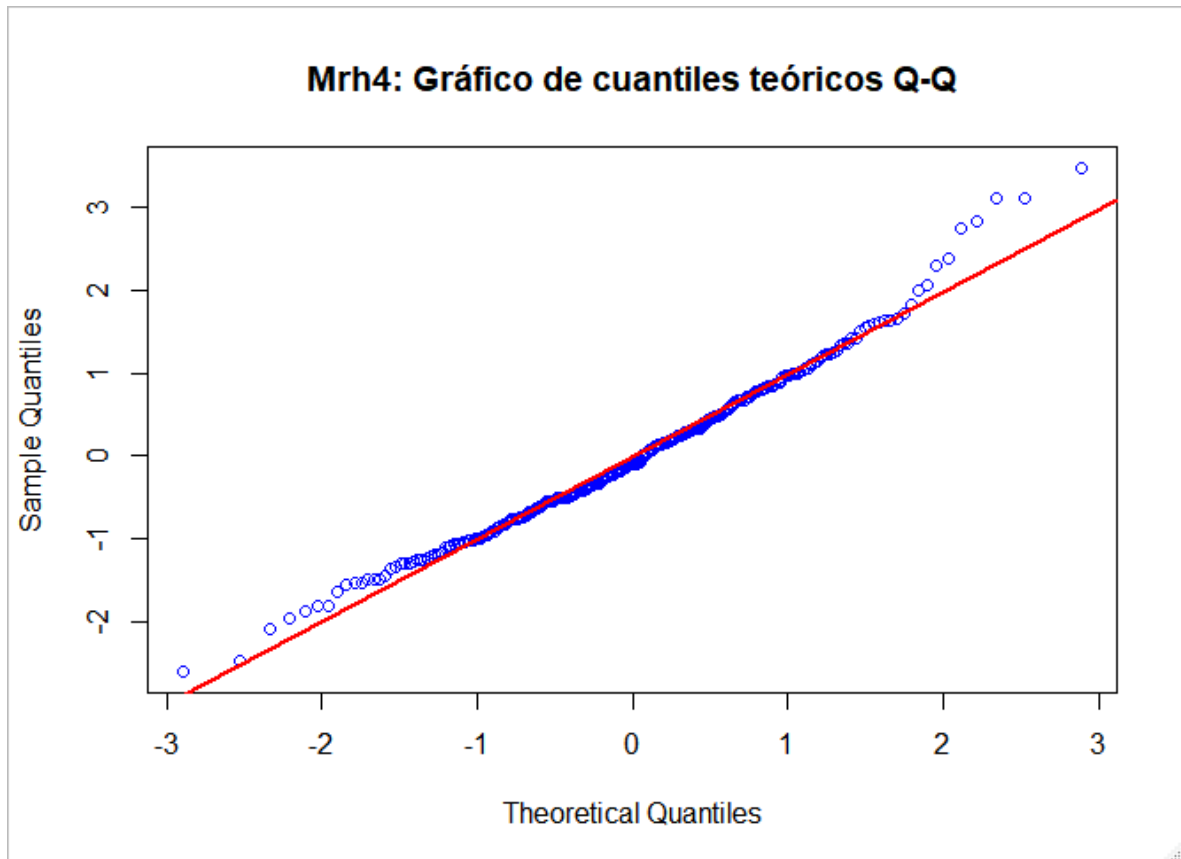


Figura 71. Gráfica Cuantil - Cuantil de Mrh4

Se puede observar de la figura anterior, que los residuos se alinean en su mayoría a línea roja, esto significa que los residuos siguen una distribución normal, no obstante, se realiza la prueba analítica de Lilliefors con la siguiente hipótesis descrita en (51)

La prueba genera un p-valor de $0.1348 > 0.05$, por lo tanto, se acepta la hipótesis nula de que los datos siguen una distribución normal.

3.7.3.3. Pruebas de homocedasticidad de Mrh4.

Para la prueba gráfica de igualdad de varianza se utiliza el diagrama siguiente:

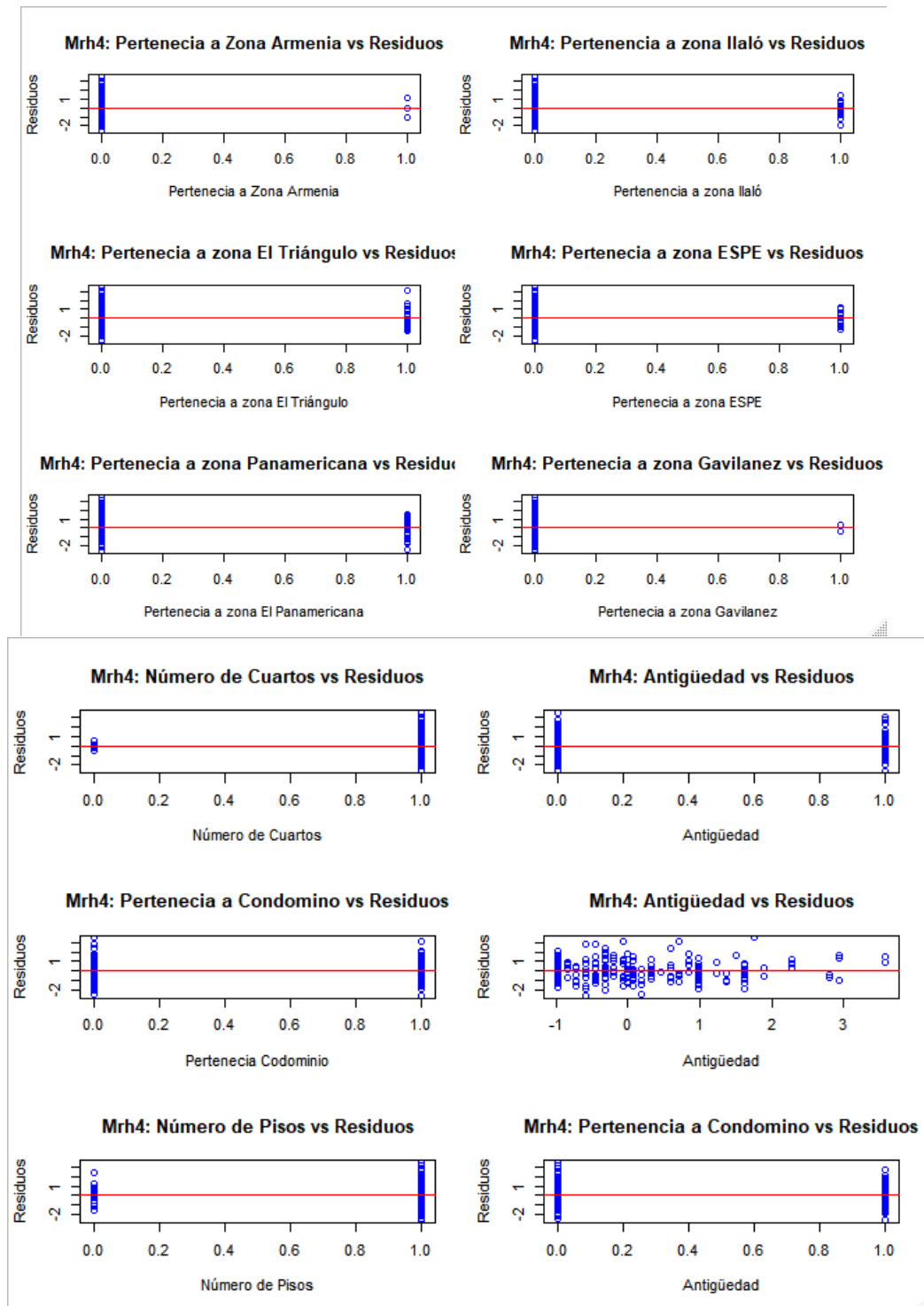


Figura 72. Diagrama de variables vs Residuos Mrh4

Del diagrama anterior se puede observar que las observaciones se distribuyen equitativamente sobre y bajo la línea roja, lo que sugiere que las variables utilizadas tienen iguales varianzas con los residuos del modelo, no obstante, se realiza la prueba analítica de Bartlett para las variables, con la hipótesis detallada en (52).

La prueba arroja un p-valor de 0.05, 0.25, 0.41, 0.74, 0.9, 0.26, 0.43, 0.32, 0.35 y 0.44 correspondiente a las variables DCua, DEst, DBan, DPis, ZArm, ZIla, ZTria, ZColi, ZPana y ZGavil, respectivamente, los p-valor son mayores que 0.05, en consecuencia, se acepta la hipótesis nula de que si hay igualdad de varianzas entre las variables y los residuos del modelo.

3.7.3.4. Pruebas de identidad de Mrh4.

Para la prueba de identidad se grafica los residuos en eje de las ordenadas y la secuencia de toma de observaciones en el eje de las abscisas, se obtiene:

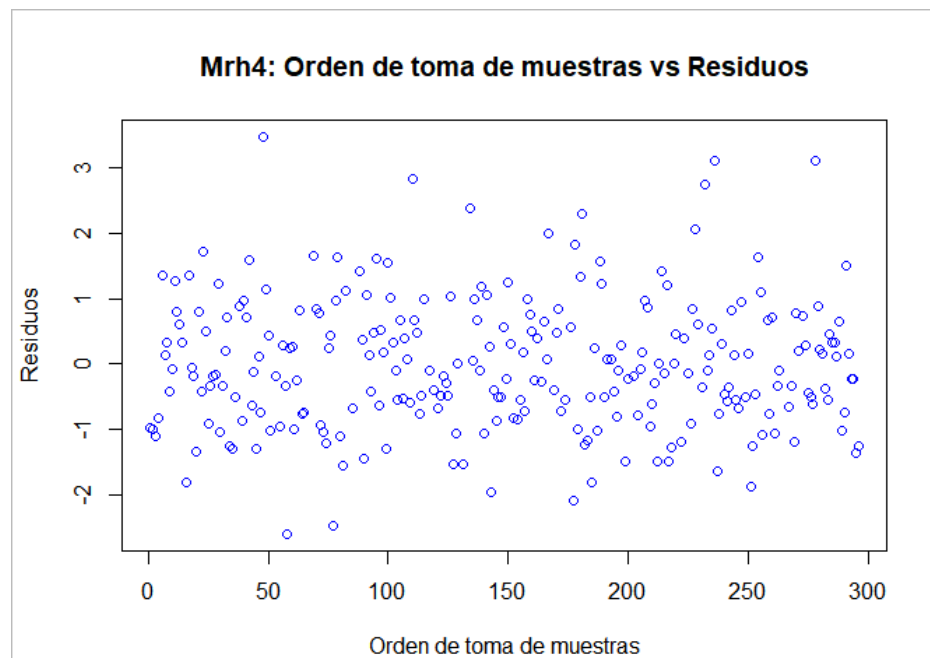


Figura 73. Diagrama de orden de toma de observaciones vs. Residuos de Mrh3

Se puede apreciar en el grafico que no existe ningún patrón definido, sin embargo, se realiza prueba analítica de Durbin Watson, la cual define la hipótesis detallada en (53).

El p-valor que se obtuvo es de $0.414 > 0.05$, por lo tanto, se acepta la hipótesis nula de que existe independencia en la toma de las observaciones.

3.8. Validación Cruzada.

Para la validación cruzada se tiene dispuesto una base de datos de 29 observaciones correspondiente al 10% de la base original, en la base en mención se analiza los datos atípicos, eliminándose 4 datos, y obteniéndose una base de trabajo de 25 observaciones, con las cuales se utiliza el comando *predict* de R, para que las 25 observaciones corran sobre los 3 modelos de factores y los 4 modelos de regresión hedónica, posteriormente se establece la media de los errores predichos y se compara con los valores reales para obtener el error en porcentaje, los resultados se detallan a continuación:

Tabla 32
Errores de los predicho, mediante validación cruzada

Modelo	Error en los predicción
mf1	27.91%
mf2	25.43%
mf3	28.61%
Mrh1	32.93%
Mrh2	32.92%
Mrh3	30.86%
Mrh4	30.37%

3.9. Selección de modelos.

Se pretende seleccionar el mejor modelo de factores y el mejor modelo de regresión hedónica, para luego hacer una comparativa entre estos dos modelos y verificar cual predice de mejor manera el precio de la vivienda.

3.9.1. Selección de modelo de factores.

Para comparar los modelos de factores, se realiza una prueba de igualdad de medias de los residuos o errores, partiendo de la siguiente hipótesis:

$$H_0: \mu_{mfi} = \mu_{mfj} \quad (i \neq j) \quad (61)$$

$$H_1: \mu_{mfi} \neq \mu_{mfj} \quad (i \neq j)$$

Al realizar la prueba con un nivel de confianza del 99%, se obtiene los siguientes resultados:

Tabla 33
Prueba de igualdad de medias de los residuos mf

Modelo	mf1	mf2	mf3
mf1			1
mf2	$\mu_{mf2} = \mu_{mf1}$		1
mf3	$\mu_{mf3} = \mu_{mf1}$	$\mu_{mf3} = \mu_{mf2}$	

De la Tabla 33, se puede observar que en la parte triangular superior el valor del estadístico, el cual es mayor a 0.01, en consecuencia, se acepta la hipótesis nula de que las medias de los errores son iguales, esto se puede ver en la parte triangular inferior de la tabla.

Al no poder llegar a una conclusión en función de las medias de los errores, se procede a observar los estadísticos de normalidad, homocedasticidad, identidad, correlación espacial y

errores de validación cruzada, con el fin de escoger el modelo que cumple con mayor nivel de confianza estos parámetros.

Tabla 34
Resumen de pruebas a modelos de factores

	Bondad de ajuste	Normalidad	Homocedasticidad	Identidad	Correlación Espacial	Validación Cruzada	
	R2	Lilliefors	Bartlett	Durbin-Watson	I. de Moran	Error %	
mf1	0.2765	0.3415		0.24	0.762	0.024	27.91%
mf2	0.299	0.311		0.2981	0.562	0.038	25.43%
mf3	0.2838	0.4189		0.3121	0.736	0.022	28.61%

De la Tabla 34, se puede observar que mf2 tiene mejor bondad de ajuste y menor error en la validación cruzada, a pesar de que mf3 tienen algunos estadísticos más altos, se escoge el mf2 por su menor porcentaje de residuos.

3.9.2. Selección de modelo de regresión hedónica.

Para comparar los modelos de regresión hedónica, se realiza una prueba de igualdad de medias de los residuos o errores, partiendo de la hipótesis planteada en (61).

Tabla 35
Prueba de igualdad de medias de los residuos de los Mrh

Modelo	Mrh1	Mrh2	Mrh3	Mrh4
Mrh1			1	1
Mrh2	$\mu_{Mrh2} = \mu_{Mrh1}$			1
Mrh3	$\mu_{Mrh3} = \mu_{Mrh1}$	$\mu_{Mrh3} = \mu_{Mrh2}$		1
Mrh4	$\mu_{Mrh4} = \mu_{Mrh1}$	$\mu_{Mrh4} = \mu_{Mrh2}$	$\mu_{Mrh4} = \mu_{Mrh3}$	

De la Tabla 35, se puede observar que en la parte triangular superior el valor del estadístico, el cual es mayor a 0.01, en consecuencia, se acepta la hipótesis nula de que las medias de los errores son iguales, esto se puede ver en la parte triangular inferior de la tabla.

Al no poder llegar a una conclusión en función de las medias de los errores, se procede a observar los estadísticos de normalidad, homocedasticidad, identidad y errores de validación cruzada, con el fin de escoger el modelo que cumple con mayor nivel de confianza estos parámetros.

Tabla 36
Resumen de pruebas a modelos de regresión hedónica

	Bondad de ajuste	Normalidad	Homocedasticidad	Identidad	Validación Cruzada	
	R^2	Lilliefors	Bartlett	Durbin-Watson	Error %	
Mrh1	0.2539	0.09467		0.33	0.662	32.92%
Mrh2	0.2538	0.09725		0.38	0.754	32.93%
Mrh3	0.3082	0.2056		0.41	0.484	30.86%
Mrh4	0.3078	0.1348		0.41	0.41	30.37%

De la Tabla 36, se puede observar que Mrh4 tiene mejor bondad de ajuste y menor error en la validación cruzada, también tiene el estadístico de normalidad y homocedasticidad mas alto, en consecuencia se escoge a Mrh4.

3.9.3. Selección de mejor modelo entre factores y regresión hedónica

Se realiza una prueba de igualdad de medias de los errores de los dos modelos, dando como resultado que las medias de los errores son iguales con un nivel de confiabilidad del 99%, lo cual no es concluyente por dicha razón se detalla una tabla resumen de los dos modelos a comparar y de sus pruebas:

Tabla 37
Resumen de pruebas a modelos

	Bondad de ajuste	Normalidad	Homocedasticidad	Identidad	Validación Cruzada	σ^2
	R^2	Lilliefors	Bartlett	Durbin-Watson	Error %	residuos
mf2	0.299	0.311	0.2981	0.562	25.43%	183
Mrh2	0.2538	0.09725	0.38	0.754	32.93%	186

Los dos modelos cumplen con las pruebas de normalidad, homocedasticidad e identidad, sin embargo, el error en validación cruzada es menor y el R^2 es mayor en el modelo de factores que en el modelo de regresión hedónica, en consecuencia se escoge al modelo de factores como mejor en comparación al modelo de regresión hedónica por su bondad de ajuste.

3.10. Clases de Equivalencia

Del modelo de factores se puede obtener clases de equivalencia de precios por m^2 de viviendas en función de los factores intervinientes, siendo número de cuartos (NumCuaC), antigüedad (AntC) y pertenecía a conjunto (Cond) las variables directas y número de pisos (NumPisC) y número de estacionamientos (NumEstC) variables de interacción, ver en la Tabla 38.

Tabla 38
Categorías de variables

Número de Cuartos	Número de Estacionamientos	Pertenencia a Conjunto	Antigüedad	Número de Pisos
NumCuaC >3	NumEstC=1 v NumEstC=2	Cond=No en conjunto	AntC>10 si AntC>5 \wedge	NumPisC=1
NumCuaC <=3	NumEstC>2 NumEstC=0	Cond=Si en conjunto	AntC≤10 si AntC≤5	NumPisC=2 NumPisC=3

De las combinaciones de las variables se derivan 108 clases de equivalencia, clases que tienen un precio determinado, en otras palabras, según sus características existen 108 conjuntos con precios fijos y una casa puede estar en uno solo de esos conjuntos.

CAPITULO IV

CONCLUSIONES Y RECOMENDACIONES

4.1. Conclusiones.

- En esta investigación el modelo de factores expresa de mejor manera el precio de la vivienda que el modelo de regresión hedónica, en función de los errores de validación cruzada y R^2 .
- La variación del precio de viviendas de similares características está dada por su relación espacial, ya que el modelo de factores tiene correlación espacial comprobada mediante el Índice de Morán.
- La variación del precio de viviendas de similares características está dada por su relación espacial, ya que el modelo de regresión hedónica contiene variables significativas de zonificación, las cuales son de naturaleza espacial.
- Se crea 108 clases de equivalencia del precio del metro cuadrado total de los bienes, a través del modelo de factores y que toma en cuenta las variables más significativas.
- Las variables de zonificación y de distancia mínima a puntos estratégicos permitieron realizar un modelo de regresión hedónica, con 32% de error en validación cruzada.

4.2. Recomendaciones.

- Se recomienda utilizar el software *R Studio* u otros softwares de similares características para realizar modelos estadísticos de predicción, el uso de esta herramienta facilita de sobremanera la y la manipulación de datos y la obtención de resultados.
- Se recomienda la utilización de la herramienta de google *fusión tables*, que permite visualizar en un mapa de google los puntos georeferenciados
- Se recomienda verificar vía telefónica o en sitio las localizaciones de portales como plusvalía.com.ec y olx.com.ec, mucha de la información con respecto a la geolocalización no es correcta.

BIBLIOGRAFÍA

- Akaike, H. (1974). A New Look at the Statistical Model Identification. En E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 215-222).
https://doi.org/10.1007/978-1-4612-1694-0_16
- Apergis, N., & Payne, J. E. (2012). Convergence in U.S. house prices by state: Evidence from the club convergence and clustering procedure. *Letters in Spatial and Resource Sciences*, 5(2), 103-111. <https://doi.org/10.1007/s12076-011-0075-7>
- Apergis, N., Simo-Kengne, B. D., & Gupta, R. (2015). Convergence in Provincial-Level South African house price: Evidence from the club convergence procedure: Convergence in Provincial House Prices. *Review of Urban & Regional Development Studies*, 27(1), 2-17. <https://doi.org/10.1111/rurd.12029>
- Bohórquez, I. A., Álvarez, N. M. G., & Ceballos, H. V. (2008). *Algunos elementos para el análisis de datos espaciales: Teoría y aplicación*. 147.
- Bohórquez, I. A., & Ceballos, E. V. (2008). Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. *Ecos de Economía*, 27, 26.
- Borrego, J. A. (2018). *Modelos de Regresión para Datos Espaciales* (Universidad de Sevilla). Recuperado de <https://idus.us.es/xmlui/bitstream/handle/11441/81660/Borrego%20S%C3%A1nchez%20Jos%C3%A9%20C%C3%81ngel%20TFG.pdf?sequence=1&isAllowed=y>
- Caballero Díaz, F. F. (2011). *Selección de modelos mediante criterios de información en análisis factorial: Aspectos teóricos y computacionales*. Editorial de la Universidad de Granada, Granada.

- Camelo Rincón, M., & Campo Robledo, J. (2016). Análisis de la política de vivienda en Bogotá: Un enfoque desde la oferta y la demanda. *Revista Finanzas y Política Económica*, 8(1), 105-122. <https://doi.org/10.14718/revfinanzpolitecon.2016.8.1.6>
- Castillo, I., & Guijarro, M. (2006). *Estadística descriptiva y cálculo de probabilidades*. Madrid: Prentice Hall.
- Cerda, F. N., & Surhoff, R. S. (2002). Estimación de un modelo hedónico para conjuntos de viviendas nuevas. *Revista Ingeniería Industrial*, 1(1). Recuperado de <http://revistas.ubiobio.cl/index.php/RI/article/view/160>
- Chica-Olmo, M. (2005). *La Geoestadística como herramienta de análisis espacial de datos de inventario forestal*. 11.
- Collazos, E., Gamboa, W., Prado, P., & Verrardi, V. (2006). Análisis espacial del precio de oferta de la vivienda en el área metropolitana de Cochabamba. *Revista Latinoamericana de Desarrollo Económico*, 6.
- Cornejo, M. del P., Zorrilla, D., Bermúdez, N., Estacio, J., & Narváez, N. (2013). *Análisis de Vulnerabilidad del cantón Rumiñahui*. Recuperado de <http://repositorio.cedia.org.ec/bitstream/123456789/857/1/Perfil%20territorial%20RUMINAHUI.pdf>
- De la Fuente, S. (2011). Análisis de Componentes Principales. *Universidad Autónoma de Madrid*, 34.
- Draper, N. R. (1992). Introduction to Box and Wilson (1951) On the Experimental Attainment of Optimum Conditions. En S. Kotz & N. L. Johnson (Eds.),

- Breakthroughs in Statistics* (pp. 267-269). https://doi.org/10.1007/978-1-4612-4380-9_22
- ENEMDU. (2018). Recuperado 19 de marzo de 2019, de <http://www.ecuadorencifras.gob.ec/enemdu-2018/>
- Fallas, J. (2012). *Correlación Lineal: Midiendo la relación entre dos variables*. Recuperado de http://www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/correlacion_lineal_2012.pdf
- Figuroa, E., & Lever, G. (1992). *Determinantes del precio de la vivienda en Santiago: Una estimación Hedónica*. Recuperado de <http://www.econ.uchile.cl/uploads/publicacion/046ec6c8-1c93-4c63-bb5f-873c4cf54d79.pdf>
- Función Ejecutiva, Presidencia de la República del Ecuador. Código Orgánico de Organización Territorial. , 303 § (2010).
- GeoDa. (s. f.). Global Spatial Autocorrelation [Workbook]. Recuperado 25 de abril de 2019, de Global Spatial Autocorrelation website: https://geodacenter.github.io/workbook/5a_global_auto/lab5a.html#introduction
- INEC. (2010). *Resultados del Censo 2010 de Población y Vivienda en el Ecuador fascículo Pichincha*. Recuperado de <http://www.ecuadorencifras.gob.ec/wp-content/descargas/Manu-lateral/Resultados-provinciales/pichincha.pdf>
- Kuehl, R. O., & Kuehl, R. O. (2000). *Design of experiments: Statistical principles of research design and analysis* (2nd ed). Pacific Grove, CA: Duxbury/Thomson Learning.

- Laguna, C. (2011). Correlación y Regresión lineal. *Instituto Aragonés de Ciencias de la Salud*, 18.
- Lever, G. (s. f.). *El Modelo de Precios Hedónicos*. 13.
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, 62(318), 399-402. <https://doi.org/10.1080/01621459.1967.10482916>
- Más casas y plusvalía en Rumiñahui. (s. f.). Recuperado 14 de febrero de 2019, de El Comercio website: <https://www.elcomercio.com/actualidad/quito/mas-casas-y-plusvalia-ruminahui.html>
- Medina, P., & López, A. (2011). Análisis Crítico del Diseño factorial sobre casos aplicados. *Scientia et Technica*, 18(47), 101-106.
- Mg, S., Rc, M., Jm, S., & Vb, S. (2004). *Evaluación y optimización simultánea de papaína inmovilizada en gránulos de alginato entrecruzado mediante un diseño factorial 3x3 y la función de deseabilidad*. 27.
- Montero Lorenzo, J. M. (2004). *El precio medio del metro cuadrado de la vivienda libre: Una aproximación metodológica desde la perspectiva de la Geoestadística*. 22, 19.
- Montero Lorenzo, J. M., & Larraz, B. (2006). *Estimación espacial del precio de la vivienda mediante métodos de Krigeado*. 48(162), 201 a 240.
- Montgomery, D. C. (2013). *Diseño y análisis de experimentos*. México: Limusa Wiley.
- Montoya, D., & Erazo, V. (2013). *Generación de Geoinformación para la gestión del Territorio a nivel nacional, escala 1:25 000*. Instituto Espacial Ecuatoriano.

- Olmo, C., Guervos, C., & Olmo, C. (2007). Modelo Hedónico Espacio-Temporal y Análisis Variográfico del Precio de la Vivienda. . . *ISSN*, 17.
- olx.com.ec. (2019). Condiciones de Uso. Recuperado 30 de septiembre de 2019, de Central de Ayuda website: <http://help.olx.com.ec/hc/es-419/articles/209282966-Condicioness-de-Uso>
- Perdomo Calvo, J. A., & Mendieta-Lopez, J. C. (2007). Specification and Estimation of a Spatial Hedonic Prices Model to Evaluate the Impact of Transmilenio on the Value of the Property in Bogotá. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.1556368>
- plusvalia.com.ec. (2019). Terminos y Condiciones de Uso. Recuperado 26 de marzo de 2019, de <https://www.plusvalia.com/terminos.bum>
- Pozo, A. G. (2007). *Una aproximación a la aplicación de la metodología hedónica: Especial referencia al caso del mercado de la vivienda*. 29.
- Pulido, H. G. (2008). *Análisis y diseño de experimentos* (Mc Graw Hill). México D.F.
- Quesada, J. H. R. (2015). *Uso del método de precios hedónicos, para estimar variación en precios de viviendas producto de nuevas líneas de metro*. Universidad de Chile, Santiago de Chile.
- Romero, M. (2016, junio 3). Pruebas de bondad de ajuste de una distribución normal. *Enfermería del Trabajo*, 105-114.
- Serrano, J. S. (s. f.). *Latinoamérica: Hambre de Vivienda*. 11.
- Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591. <https://doi.org/10.2307/2333709>

- Torres, E., Gamarra, O., & Huayana, P. (2019). Análisis factorial de las características socioeconómicas del ganadero de las cuencas lecheras de la región Amazonas, 2009. *Revista ECIPeru*, 92-98. <https://doi.org/10.33017/RevECIPeru2011.0029/>
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192-196. <https://doi.org/10.3758/BF03206482>
- Yacuzzi, E., Martín, F., Quiñones, H. M., & Popovsky, M. J. (2004). *El diseño experimental y los métodos de Taguchi: Conceptos y Aplicaciones en la Industria Farmaceuica*. 32.
- Yarasca, J. C. L. (2015). *Evaluación de autocorrelación espacial global y local para zonas de tránsito*. 14.