RESUMEN TRABAJO DE GRADO

AUTOR(ES): NOMBRES Y APELLIDOS COMPLETOS

NOMBRE(S): LINA GABRIELA APELLIDOS: GUERRERO VERJEL

NOMBRE(S): LISETH YAJAIRA APELLIDOS: VACA CAICEDO

FACULTAD: INGENIERÍA

PLAN DE ESTUDIOS: INGENIERÍA CIVIL

DIRECTOR:

NOMBRE(S): CARLOS ALEXIS APELLIDOS: BONILLA GRANADOS

TÍTULO DEL TRABAJO (TESIS): ECUACIONES DE ESTIMACIÓN DE COSTO DE DISEÑO HIDRAULICO EN URBANISMO Y EDIFICACIONES

En la actualidad no existe una forma precisa para estimar el costo de los diseños hidráulicos, el método empírico mas empleado es el “área de construcción”. El presente proyecto investigativo establece una metodología para obtener ecuaciones que estimen el costo de estos diseños y que puedan ser implementadas en proyectos de urbanismo y edificaciones; para esto, se emplean métodos estadísticos con modelos funcionales en el software de programación R, teniendo en cuenta las variables que pueden influir en un proyecto: área, número de viviendas o apartamentos y longitud de tubería, clasificadas en las tipologías: residencia unifamiliar, residencia multifamiliar y otros proyectos. Las ecuaciones obtenidas son evaluadas con proyectos hidráulicos existentes para validar su eficacia y aplicación en la ingeniería.

PALABRAS CLAVES: Ecuaciones, diseño hidráulico, costo, lenguaje de programación R, urbanismo, edificaciones, residencia unifamiliar, residencia multifamiliar.

CARACTERISTICAS:

PÁGINAS: 339 PLANOS: 0 ILUSTRACIONES: 131 CD ROOM: 0

ECUACIONES DE ESTIMACIÓN DE COSTO DE DISEÑO HIDRÁULICO EN URBANISMO Y EDIFICACIONES

LINA GABRIELA GUERRERO VERJEL

LISETH YAJAIRA VACA CAICEDO

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERÍA

PLAN DE ESTUDIOS DE INGENIERÍA CIVIL

SAN JOSÉ DE CÚCUTA

2022

ECUACIONES DE ESTIMACIÓN DE COSTO DE DISEÑO HIDRÁULICO EN URBANISMO Y EDIFICACIONES

LINA GABRIELA GUERRERO VERJEL

LISETH YAJAIRA VACA CAICEDO

Tesis para optar el titilo de ingeniero civil

Director

Carlos Alexis Bonilla Granados

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERÍA

PLAN DE ESTUDIOS DE INGENIERÍA CIVIL

SAN JOSÉ DE CÚCUTA

2022

Universidad Francisco de Paula Santander
Vigilada Mineducación

ACTA DE SUSTENTACION DE TRABAJO DE GRADO

FECHA: *19DE MAYO DE 2022* *HORA: 8:00 a. m.*

LUGAR: *SC – 302 UFPS*

PLAN DE ESTUDIOS: *INGENIERIA CIVIL*

TITULO DE LA TESIS: *"ECUACIONES DE ESTIMACION DE COSTO DE DISEÑO HIDRAULICO EN URBANISMO Y EDIFICACIONES".*

JURADOS: *ING. CLAUDIA PATRICIA CHAUSTRE SANCHEZ*
ING. JAIRO MARTIN RODRIGUEZ TENJO

DIRECTOR: *INGENIERO CARLOS ALEXIS BONILLA GRANADOS*

NOMBRE DE LOS ESTUDIANTES:	CODIGO	CALIFICACION	
		NUMERO	LETRA
LINA GABRIELA GUERRERO VERJEL	1113942	4,4	CUATRO, CUATRO
LISETH YAJAIRA VACA CAICEDO	1113940	4,4	CUATRO, CUATRO

A P R O B A D A

ING. CLAUDIA PATRICIA CHAUSTRE SANCHEZ *ING. JAIRO MARTIN RODRIGUEZ TENJO*

Vo. Bo.
JAVIER ALFONSO CARDENAS GUTIERREZ
Coordinador Comité Curricular

Betty M.

Dedicatoria

A nuestros padres y hermanos por brindarnos su apoyo incondicional en cada uno de los obstáculos y su acompañamiento en cada momento requerido; por la confianza, los sacrificios y los consejos que nos permitieron crecer como personas; por los ánimos para continuar y no dejar a un lado los sueños propuestos. Les damos gracias a ellos porque sus actos nos hicieron lo que somos hoy.

A los Docentes de la Universidad Francisco de Paula Santander, por ser los promotores de la búsqueda del conocimiento; por guiarnos hacia nuestro camino como profesionales; por su labor honesta a guiarnos en el camino como profesionales; asimismo para aquellos compañeros que nos brindaron su acompañamiento y conocimiento en este proceso educativo.

Agradecimientos

Los autores expresan sus agradecimientos a:

Al director del proyecto, Carlos Alexis Bonilla Granados, por ser el guía, acompañante, y promotor principal de la investigación.

Al docente y asesor, Gustavo Adolfo Ovalles Rodríguez, por el conocimiento, experiencia y tiempo dedicado al desarrollo estadístico de la investigación.

A la docente y asesora, Adriana Rodríguez Lizcano, por el conocimiento, experiencia y tiempo dedicado al desarrollo metodológico de la investigación.

A la Universidad Francisco de Paula Santander, por los recursos académicos y tecnológicos aportados para el crecimiento del conocimiento y el progreso en la investigación.

Al Semillero de Investigación SIRHI, por servir de guía, asesor y acompañante durante el proceso investigativo.

Tabla de contenido

Lista de tablas

Lista de figuras

Resumen

En la actualidad no existe una forma precisa para estimar el costo de los diseños hidráulicos, durante este proceso de búsqueda se ha empleado un método empírico que analiza únicamente la variable "área de construcción". El presente proyecto investigativo establece una metodología para obtener ecuaciones que estimen el costo de un diseño hidráulico y que puedan ser implementadas en proyectos de urbanismo y edificaciones; para esto, se emplean métodos estadísticos con modelos funcionales en el software de programación R, teniendo en cuenta las variables que pueden influir en un proyecto: área, número de viviendas o apartamentos y longitud de tubería, clasificadas en las tipologías: residencia unifamiliar, residencia multifamiliar y otros proyectos. Las ecuaciones obtenidas son evaluadas con proyectos hidráulicos existentes para validar su eficacia y aplicación en la ingeniería.

Palabras claves: Ecuaciones, diseño hidráulico, costo, lenguaje de programación R, edificaciones, urbanismo, residencia unifamiliar, residencia multifamiliar.

Introducción

Garantizar el óptimo costo de los diseños hidráulicos en el sector público y privado para beneficiar a gran cantidad de ingenieros civiles en sus proyectos de infraestructura, es un aspecto muy importante para favorecer el desarrollo integral de los procesos ingenieriles desde su etapa de desarrollo en donde se puede prevenir un sobrecosto para los contratistas o un déficit económico para el diseñador hidráulico.

"Muchos investigadores a nivel nacional e internacional han estudiado el comportamiento de los costos de este tipo de diseños con el objeto de obtener ecuaciones que lo describan de una manera acertada" (Peinado Calao, 2016, pág. 1). Es por esta razón que se propone la implementación de una metodología que permita ajustar ecuaciones con las que se pueda estimar el costo de un diseño hidráulico en Colombia, mediante el análisis de información presupuestal de proyectos viabilizados y ejecutados por entidades públicas y privadas.

El método consiste en formular varias ecuaciones por medio del Software R para estimar el costo de un diseño hidráulico mediante la recolección de datos de distintos proyectos del departamento de Norte de Santander realizados en los últimos 5 años (2017-2021), organizarlos según los diferentes factores que pueden afectar el costo real del diseño (el área, cantidad de pisos o niveles, cantidad de apartamentos, número de viviendas o longitud de tubería), para después, analizar los datos que han empleado diferentes ingenieros y empresas para el costo de cada diseño y las características generales de cada tipología de proyecto, en la que finalmente, se realizará un análisis estadístico de estos grupos de datos y así poder establecer distintas ecuaciones que involucren todos los factores y que permitan estimar de manera real el costo del diseño hidráulico en los proyectos de ingeniería.

1. Problema

1.1. Título

Ecuaciones de estimación de costo de diseño hidráulico en urbanismo y edificaciones.

1.2. Planteamiento del problema

En el ejercicio de la ingeniería es primordial determinar con precisión las diferentes variables que se pueden implementar para determinar el costo del diseño hidráulico de un proyecto de urbanismo o de edificaciones. No obstante, estas variables que deben definirse en el costo del diseño son numerosas y se deben valer de un análisis cuidadoso.

Hoy en día no se ha establecido un criterio definitivo para los ingenieros civiles o hidráulicos al momento de estimar el costo por la realización de un diseño, es allí donde surgen desacuerdos entre diseñadores y quien busca de sus servicios, ya que los distintos métodos empíricos conllevan costos sobrevalorados o subvalorados que no van acorde a la proporción del diseño planteado, cabe mencionar además el desgaste en tiempo que se le genera al diseñador para adecuar el costo a la realidad del proyecto.

La ingeniería civil lleva consigo gran cantidad de proyectos que van desde inmensas obras civiles hasta la construcción de una vivienda unifamiliar, sin embargo, muchos de ellos involucran diseños hidráulicos que varían según cada proyecto, lo cual hace que su costo también difiera. Es esencial en este punto de la ingeniería establecer lineamientos que permitan darle el costo correspondiente a la realización de diseños hidráulicos teniendo en cuenta los factores que puedan influir, tales como el área, cantidad de pisos o niveles, cantidad de apartamentos, número de viviendas o longitud de tubería a diseñar y/o construir, así como la tipología de las edificaciones, tales como residencia unifamiliar, multifamiliar, edificaciones de tipo institucional, comercial o industrial, o demás factores que puedan influir en el costo de diseño.

Actualmente, "el método poco confiable para estimar el costo del diseño hidráulico es por medio del área del proyecto a construir" (Baldrich Flórez & López Ceballos, 2017, pág. 6), sin embargo, este método no es aplicable a las diferentes tipologías de edificaciones y no expresa el valor real del costo que estas puedan tener. Es por esta razón que se busca una alternativa que involucre todas las variables que influyen en los distintos proyectos ingenieriles para dar una solución eficaz y certera en el proceso de costo de un diseño hidráulico.

1.3. Formulación del problema

¿Cómo determinar ecuaciones de estimación de costo de diseño hidráulico que puedan ser implementadas en proyectos de urbanismo y edificaciones?

1.4. Objetivos

1.4.1. Objetivo General

Determinar ecuaciones de estimación de costo de diseño hidráulico que puedan ser implementadas en proyectos de urbanismo y edificaciones.

1.4.2. Objetivos Específicos

1. Consultar referentes teóricos y legales que den respuesta al objeto de estudio con el fin de observar el conjunto de variables que inciden en el costo al diseñar redes hidráulicas.
2. Recolectar información de las variables en proyectos hidráulicos obtenidos de diferentes ingenieros y empresas de Norte de Santander en los últimos 5 años (2017-2021).

3. Analizar estadísticamente las variables de los proyectos hidráulicos, mediante la comparación en gráficas de cada dato obtenido.
4. Formular una serie de ecuaciones por medio del análisis estadístico que identifique variables como área, número de viviendas y longitud de tuberías en distintas tipologías de proyectos.
5. Evaluar las ecuaciones metodológicas formuladas por medio de un análisis comparativo con proyectos hidráulicos existentes para verificar su eficacia.

1.5. Justificación

Las instalaciones hidrosanitarias son uno de los ítems más importantes en cualquier tipo de proyecto constructivo, y principalmente forman una parte esencial en el saneamiento básico en proyectos de urbanismo y edificaciones para viviendas unifamiliares; en estos proyectos la edificación cuenta con un sistema para el suministro y descarga de las aguas que alimentan y evacuan a los diferentes aparatos sanitarios.

Se define entonces que el diseño de redes o instalaciones hidráulicas consiste en un sistema de tuberías, válvulas, ramales y conexiones que proveen agua para los diferentes servicios de una construcción (suministro de agua potable), de igual manera funciona para el desalojo de aguas servidas o residuales de las construcciones, (aguas residuales y aguas lluvias). (Baldrich Flórez & López Ceballos, 2017, pág. 7)

Actualmente, no existe una forma precisa para estimar el costo de este tipo de diseños, y generalmente se emplea un método empírico que analiza únicamente una sola variable “área de construcción”, es decir, este no evalúa los diferentes factores que afectan considerablemente el costo de los diseños a realizar, sino que se toma de manera generalizada a todos los proyectos y se cotizan bajo una misma categoría, por lo tanto, se desconoce si el costo asignado cumple satisfactoriamente con el tipo de proyecto establecido. Basados en la problemática e investigación, se llega a la interrogante respecto al costo de dichas instalaciones hidráulicas, pregunta que requiere rapidez y precisión en el resultado final, de manera que se busca realizar una serie ecuaciones metodológicas en las que se evalúen los diferentes factores que influyen en los proyectos, esto con el fin de obtener un costo lo más cercano a la realidad.

De acuerdo con lo anterior, para los diseños de redes hidráulicas se debe tener en cuenta diferentes variables para poder estimar un costo que se acerque lo más posible a la realidad. Por tal motivo, la idea fundamental de esta investigación es formular ecuaciones metodológicas a partir del análisis de variables que afectarán considerablemente el resultado final de una cotización, estas variables son; la longitud de tubería, el número de apartamentos o viviendas dependiendo el tipo de proyecto en el que se va a utilizar y el área total de construcción, para ello se implementarán en proyectos a partir del año 2017 que es el tiempo de recolección fijada, para poder al final evaluar la ecuación mediante un análisis comparativo de la que se deberá reflejar su eficiencia en los resultados.

1.6. Alcances y limitaciones

1.6.1. Los alcances

Los ingenieros civiles tienen dificultades de manera permanente debido a que en todo momento se realizan construcciones de obras civiles que están relacionados con el diseño hidráulico de viviendas, edificaciones u otro tipo de proyectos y que requieren de una cotización inmediata o un costo final por haber realizado el trabajo.

Es por esto, que el proyecto de investigación se enfocará específicamente en la estimación de costos para los diseños de las redes hidrosanitarias, debido a que es de gran importancia que el ingeniero civil pueda tener herramientas para calcular el costo de un diseño hidráulico de manera eficaz y que este resultado sea lo más próximo a la realidad.

1.6.2. Las limitaciones

Se estima recolectar información de los últimos 5 años (2017-2021) en proyectos realizados en distintos municipios del departamento de Norte de Santander, que puedan ser obtenidos de distintos ingenieros civiles y empresas del sector público y privado de la construcción, debido a que manejar mayor cantidad de información en otras ciudades o municipios extendería el tiempo y el análisis de datos del proyecto considerablemente.

En la elaboración de las ecuaciones para la estimación de costo de diseños hidráulicos se tendrán en cuenta las variables como; tipología (urbanismo, edificación y otros proyectos), área total construida, número de viviendas y apartamentos, y longitud de tubería. Si se desea implementar otro tipo de variables se deberá replantear las ecuaciones propuestas en este proyecto de investigación, ya sea para uso profesional o para ampliar el objeto investigativo.

1.7. Delimitaciones

1.7.1. Delimitación Espacial

El proyecto se realizará en la ciudad de Cúcuta, Norte de Santander, Colombia, pero se utilizarán datos de proyectos efectuados en distintos municipios de Norte de Santander.

1.7.2. Delimitación Temporal

El tiempo que se empleará para la realización de este proyecto de investigación es de 6 (seis) meses a partir de la aprobación del anteproyecto.

1.7.3. Delimitación Conceptual

En esta investigación se emplearán conceptos que describen el estudio de estimación de costos para el diseño hidráulico, entre los cuales encontramos:

Acueducto.

Alcantarillado.

Ecuación.

Estadística.

Urbanismo.

Edificación.

Hidráulica.

ANOVA.

Residencia unifamiliar.

Residencia multifamiliar.

2. Marco referencial

Se presentan los antecedentes tanto nacionales como internacionales más relevantes sobre los cuales se fundamenta el desarrollo de este proyecto de investigación, que nos permitirán comprender la metodología para desarrollar una serie de ecuaciones para la estimación de costos en el diseño hidráulico en urbanismo y edificaciones.

2.1. Antecedentes y estado del arte

La estimación de costos de proyectos entorno al diseño hidráulico es una de las tareas más importantes en la construcción de edificaciones y urbanismos, ya que consume tiempo y está sujeto a errores, debido a la complejidad del trabajo. Es por esto, que en la literatura está surgiendo nuevas metodologías para que este proceso de estimación sea lo más próximo a la realidad, teniendo en cuenta variables que afectan considerablemente el costo final del proyecto.

Egas, Bryan. 2020, en su trabajo de grado para optar el título profesional de ingeniero civil titulado: *Determinación de la ecuación de costo para optimizar el diseño de sistemas de alcantarillado en el cantón Santa Rosa*. Describe las técnicas e inferencias teórico prácticas para formular la ecuación que relacione los rubros más significativos con el costo de construcción de los sistemas de alcantarillado, desde el enfoque de una necesidad palpable ante el encarecimiento de obras públicas de este tipo, por la falta de optimización y análisis previo de una necesidad en el medio laboral como profesional de la construcción, aportando una ecuación con un error del 1.39% que servirá de guía para la realización de las ecuaciones que se desean desarrollar en el proyecto investigativo.

Palomino, Bach. 2018, en su trabajo de grado para optar el título profesional de ingeniero civil titulado: *Determinación de función de costo en sistemas de agua potable y alcantarillado*. Propuso la implementación de una metodología que permite ajustar funciones de costo con las que se puede estimar el costo por ítems representativos de redes de agua potable y alcantarillado en la región Ayacucho, mediante el análisis de información presupuestal de proyectos viabilizados y ejecutados por entidades del orden gubernamental, dando un aporte significativo con la implementación de métodos estadísticos y matemáticos para analizar datos similares y atípicos de varios proyectos ingenieriles.

Kindelán, Jorge. 2017, en su artículo investigativo titulado: *Metodología con enfoque a procesos para la implementación de sistemas de costos en las empresas de acueducto en Cuba*. Usando la metodología de registrar y controlar el costo de producción y servicio de acueducto, clasificándose de la siguiente manera: por su naturaleza económica (gastos por elementos económicos); por su forma de inclusión en el costo (directo o indirecto); por su relación con el volumen (variables o fijos); por su relación con la producción (primarios o de conversión), permite identificar y clasificar, a partir de las características propias de la actividad, los diferentes elementos de gastos que se incluirán en el costo de un sistema de acueducto.

Solórzano, José. 2016, en su trabajo de investigación para optar por el grado de Ingeniero en Contabilidad y Auditoría titulado: *Diseño de un sistema financiero-contable para la empresa arboleda FAINI HIDROTECNOLOGÍA CÍA. LTDA*., dedicada al servicio de ingeniería hidráulica, ubicada en la ciudad de Quito. Con su sistema financiero tiene como finalidad que la empresa pueda obtener información que sea relevante para la toma de decisiones como: las Normas Contables aplicables y Principios Generalmente Aceptados, el Plan de Cuentas con su respectivo instructivo, procedimientos contables descriptivos y gráficos, formatos de los principales documentos y concluye con un ejercicio práctico, sirviendo como modelo el sistema financiero y contable elaborado específicamente para una empresa con énfasis en hidráulica, quien requería una solución eficaz a su forma de cobrar en los proyectos.

Leiva, Carlos. 2015, en su trabajo de grado para optar el título profesional de ingeniero civil titulado: *Estudio comparativo técnico-económico de la red de alcantarillado convencional y condominal en el AA.HH.* Pamplona alta, sector las Américas. Desarrollando la toma de datos, el diseño hidráulico de los dos sistemas, la comparación técnica – económica y el desarrollo al detalle de la propuesta seleccionada aporta un presupuesto real de un alcantarillado convencional y condominal expresando un análisis económico en los procesos de ejecución del alcantarillado.

Swamee, Prabhata. Sharma, Ashok. 2013, en su artículo investigativo titulado*: Optimal design of a sewer line using Linear Programming*. El enfoque proporciona tamaños de tubería continuos, que se convierten a tamaños comerciales más cercanos para su adopción, lo que diluiría en gran medida el resultado óptimo, también se adoptan métodos de búsqueda para obtener soluciones de diseño rentables utilizando tamaños de tubería directamente comerciales, que son computacionalmente costosos, en el artículo se describe el modelo matemático para la programación de un diseño óptimo de alcantarillado, lo cual aportaría las ecuaciones planteadas, la descripción del diseño de alcantarillado, y la programación que se realizó para obtener el diseño.

Jiménez, José. 2013, en el manual elaborado como apoyo para los estudiantes de ingeniería civil de la universidad Veracruzana titulado: *Manual para el diseño de sistemas de agua potable y alcantarillado sanitario.* Con la realización de un estudio de factibilidad técnico, económico y financiero, cuyo objetivo primordial es justificar la elaboración del proyecto, garantizando que su ejecución se efectúe mediante un análisis de todos los factores técnicos, sociales, económicos, financieros, políticos y culturales que intervienen, aporta la manera más eficaz y rentable para realizar un diseño de acueducto y alcantarillado para una población necesitada de estos recursos, reduciendo costos operacionales y de diseño.

Michalus, Juan. Ibarra, María. 2006, en su artículo investigativo titulado: *Un modelo matemático para calcular el costo de manutención de un estudiante de la facultad de ingeniería*. El desarrollo del software de aplicación permite el acceso a los resultados por parte de la comunidad, especialmente a las familias de los futuros ingresantes que viven en localidades alejadas o provincias vecinas, quienes podrán tener así una base sólida para el cálculo de los gastos para mantener a su hijo estudiando en la Facultad de Ingeniería de la Universidad Nacional de Misiones, este articulo aporta el modelo matemático para calcular costos, que es la base de la investigación presentada, con diferencia del enfoque que se le da a cada análisis matemático.

Clark, Robert. Sivaganesan, Mano. Selvakumar, Ari. Sethi, Virendra. 2002, en su artículo investigativo titulado*: Cost Models for Water Supply Distribution Systems*. Con la realización de modelos que se basan en datos de un informe o de datos similares para desarrollar y modelar la propagación de contaminantes en la distribución de agua potable, muestra un análisis a la problemática de recolección de agua potable y plantea un modelo matemático para facilitar su adecuación y construcción por medio de programas.

Narváez, Paulo. Galeano, Haiver. 2002, en su artículo investigativo titulado: *Ecuación de costos y función objetivo para la optimización del diseño de redes de flujo de líquidos a presión*. El diseño de redes de flujo tiene como objetivo determinar el sistema que satisface las condiciones mínimas de presión y de flujo en los puntos de consumo, con la menor inversión y los menores costos de operación. Para dimensionar la red que cumple con las restricciones hidráulicas a un costo mínimo, se han desarrollado varios métodos de solución de los cuales han recibido especial atención en los últimos años la programación lineal, la programación no lineal, y los algoritmos genéticos, lo cual aporta el desarrollo matemático implementado para obtener una ecuación de costos para la optimización de un diseño de redes, y el análisis de los cálculos desarrollados para analizarlos y aplicarlos en el proyecto a desarrollar.

Jiménez, Claudia. Rativa, Juan. Velandia, Fabio. 2018, en su trabajo de grado de especialización en recursos hídricos titulado: *Desarrollo de una aplicación para la estimación de costos de redes hidrosanitarias en edificaciones de vivienda de interés social multifamiliar.* Permite calcular un presupuesto de instalaciones hidrosanitarias para edificios de vivienda de interés social multifamiliar, a partir de la base de datos de costos reales de insumos, mano de obra y de diseños preestablecidos, utilizando variables como son la tipología de las viviendas, el número mínimo de aparatos solicitado, las especificaciones de calidad y diámetro requeridos en función de la calidad de aparatos de cada unidad habitacional. Su aporte es importante, debido a que muestra una afirmación en la problemática de la estimación del costo en los diseños hidráulicos trayendo consigo antecedentes sobre metodologías para obtener el costo de un proyecto con la interacción de diferentes variables, en el menor tiempo posible y de la manera más cercana a la realidad.

Baldrich, Laura. López, Katherine. 2017, en su investigación de especialización en gerencia de proyectos titulado: *Modelo de cobro para la optimización de rentabilidad en el diseño de redes hidrosanitarias.* Se realiza la creación de un sistema de categorización que defina la complejidad de cada diseño y la integración de variables y procesos para un modelo de optimización y rentabilidad. Se consolida que después de analizar cada uno de los resultados obtenidos con ayuda de herramientas, fue posible generar una ecuación ajustada a la realidad en la que permitió mantener una rentabilidad fija definida por la compañía y disminuir la incertidumbre en cuanto a los costos que se producen en los diseños hidrosanitarios. Esto ayudará a comprender el efecto que causa las variables utilizadas y el enfoque en que se les da, además de que proveerán una ecuación que va enfocada con los objetivos del proyecto.

Saldarriaga, Juan. López, Laura. Páez, Diego. Luna, Daniel. González, Sebastián. 2017, en su artículo de investigación titulado: *Diseño Optimizado de redes de distribución de agua potable (Programa REDES).* Implementan un software de simulación hidráulica en donde permiten la edición, diseño y cálculo de la calidad del agua y de redes de distribución, es por esto, que el módulo de diseño optimizado cuenta con cinco metodologías diferentes con el fin de encontrar un diseño de costo mínimo, con parámetros ingresados por el usuario dependiendo de los costos de las tuberías comerciales disponibles. Debido a que el articulo tiene relación con el trabajo de investigación puesto que proporcionan diferentes ecuaciones en función a los costos para el diseño optimo y por la que será parte importante para el análisis de las ecuaciones metodológicas existentes para diseños hidráulicos.

Peinado, Carlos. 2016, En su investigación titulado: *Ecuaciones de costo para el diseño optimizado de redes de agua potable y alcantarillado.* Propone la implementación de una metodología que permite ajustar funciones de costo para redes de agua potable y alcantarillado en Colombia, la cual fue ajustada mediante una regresión lineal múltiple que permiten estimar el costo unitario de tramo construido para sistemas de alcantarillado y de abastecimiento de agua. Aunque su enfoque es hacia la estimación de costos durante la relación de construcción de los sistemas de redes de agua potable y alcantarillado, proporciona las diferentes fórmulas teóricas aplicadas en la investigación, además de ayudar en el marco teórico con la presentación de síntesis de algunos trabajos realizado en Colombia donde se desarrollaron y plantearon ecuaciones de costo para los tipos de sistemas que ocupan el objetivo de esta investigación.

Tous, Néstor. 2013, en su proyecto de grado de maestría ingeniería civil titulado: *Estructura de costos en sistemas de alcantarillado.* Se presenta un estudio completo del comportamiento de los precios en el mercado colombiano y de las actividades que son necesarias para construir un sistema de alcantarillado, a su vez buscó suministrar un criterio económico de evaluación de alternativas para la construcción de alcantarillados, en donde reconocieron mediante las variables empleadas que el componente que más peso tenía sobre la estructura de costo era el de las tuberías. De esta manera, se tuvo en cuenta los análisis de resultados y conclusiones obtenidos en la estructura de costos, además posee una serie de ecuaciones ya existentes descritas en los antecedentes sobre las variables tratadas que servirán como guía.

2.2. Marco teórico

Instalaciones hidrosanitarias

Las instalaciones hidrosanitarias comprenden de la conducción y distribución de agua potable cada una de las salidas sanitarias "hidráulica", y la evacuación de aguas residuales provenientes de la edificación "sanitarias". (Baldrich Flórez & López Ceballos, 2017, pág. ii)

Sistema de instalaciones hidráulicas

Es un conjunto de tuberías y conexiones de diferentes diámetros y diferentes materiales; para alimentar y distribuir agua dentro de la construcción, esta instalación surtirá de agua a todos los puntos y lugares de la obra arquitectónica que lo requiera, de manera que este líquido llegue en cantidad y presión adecuada a todas las zonas húmedas de esta estalación, también constará de muebles y equipos. (Gutierrez Martinez, 2008, pág. 8)

Sistema de instalaciones sanitarias

Es un conjunto de tuberías, piezas y aparatos, destinadas a dar salida a las aguas sucias de desechos o inútiles, fuera del inmueble. (Gutierrez Martinez, 2008, pág. 8)

Costos en instalaciones hidrosanitarias

Los costos en la construcción de las obras asociadas a las redes de distribución de agua potable tienen relevante importancia a la hora de realizar el diseño de la misma, debido que el objetivo del diseño es proyectar obras al mínimo costo posible cumpliendo con las restricciones hidráulicas y comerciales del sitio donde se pretenda desarrollar el proyecto, por tal motivo la solución es obtenida con la implementación de las metodologías o heurísticas de optimización, están son fuertemente influenciadas por las funciones de costo o los costos unitarios considerados para estimar el costo total de instalación de las tuberías. (Palomino Pariona , 2018, pág. 14)

Las instalaciones hidrosanitarias son uno de los aspectos más importantes en el desarrollo integral de las comunidades debido a que estás garantizan el abastecimiento del agua que es un recurso esencial para la vida humana, por otra parte, luego de ser usada es necesario su desecho por la cual es requerida las redes sanitarias (Jimenez Chicaeme, Rativa Grijalba, & Velandia Cardenas, 2018, pág. 14). Es por esto, que calcular su costo es de gran relevancia, para esto se debe entender que su resultado varía de acuerdo con las diferentes características que está pueda estar relacionada.

“En la historia los investigadores han estudiado el comportamiento de los costos de este tipo de sistemas con el objeto de obtener funciones que describan de forma acertada dichos costos” (Peinado Calao, 2016, pág. 1). Un ejemplo de esté es el proyecto Ivonne Navarro Pérez llamado *Diseño optimizado de redes de drenaje urbano* publicada en el 2009 donde se establecen tres ecuaciones para analizar los costos asociados con el diseño, en donde las dos primeras son para evaluar el costo de la tubería y la tercera para hallar el costo de la excavación.

Debido a que el objeto de estudio de la presente investigación y las ecuaciones establecidas por Navarro solo se relaciona la variable tubería es por esto, que se tendrá en cuenta solo está su primera ecuación basada en el costo de tubería en donde partió de un estudio realizado por el *Trenchless Technology Center de Lousiana Tech Univeristy* para establecer la curva de costos para tecnología de rehabilitación de alcantarillados correspondiente a tubería de PVC, la cual actualizó al año de su estudio en 2008.

$$C = 692.62d^{1.088}H^{0.303}$$

Donde:

C: Costo por metro lineal de tubería (COP/m)

d: diámetro de la tubería en milímetros (mm)

H: Profundidad de la instalación en metros (m)

La estimación de costos para licitaciones de proyectos de construcción es una de las actividades más críticas en la arquitectura, la ingeniería, la construcción y la gestión de instalaciones, en la que consume tiempo y es propenso a errores, debido a la complejidad del trabajo y la desviación causada por la representación bidimensional tradicional de los diseños. (Jiménez et al., 2018, pág. 14)

Una de estas herramientas es la tecnología Building Information Modeling (BIM), en la que permite una cercana solución a estos problemas a través del intercambio de datos, donde se tienen en cuenta las variables claves para la estimación de costos conforme a especificaciones de modelos de diseño, tomando como base la estimación de costos para los diferentes componentes de los proyectos, para esto se establece un modelo, en donde mediante una aplicación se presentan mapa de procesos y algoritmos para desarrollo de un prototipo de aplicación de software basada en BIM para la estimación del costo y los resultados de la aplicación del prototipo en un proyecto de construcción real. (Sánchez Luna, 1995, pág. 309)

Regresión Lineal

El análisis de Regresión Lineal es una metodología estadística que permite encontrar una función que representa la tendencia general de los datos. Esa función se conoce como la Función de Regresión Muestral (FRM). La idea fundamental del análisis de regresión es la dependencia estadística de una variable con respecto a una variable explicativa. El objetivo del análisis es estimar el valor promedio de la variable dependiente con base en los valores conocidos de la variable explicativa. (Gujarati & Porter, 2010, pág. 97)

y es la variable dependiente

x es la variable explicativa o regresora

La Función de Regresión Muestral (FRM) se escribe

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Donde $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores que deben calcularse, y el valor de $\hat{y}$ el cual corresponde al valor de la variable dependiente estimada por el modelo de regresión. (Gujarati & Porter, 2010, pág. 97)

Significancia estadística del modelo de regresión

Indica la validez estadística del modelo de regresión lineal construido, así como su "bondad de ajuste", es decir, qué tan bien se ajustan los datos empíricos a la recta de regresión. La significancia estadística se determina mediante el Análisis de Varianza (ANOVA) que consiste en dividir la variabilidad total del modelo de regresión en cada uno de sus componentes. (Gujarati & Porter, 2010, pág. 124)

En la siguiente figura se puede apreciar la variabilidad total y sus dos componentes: la variabilidad explicada por el modelo de regresión y la variabilidad no explicada o aleatoria:

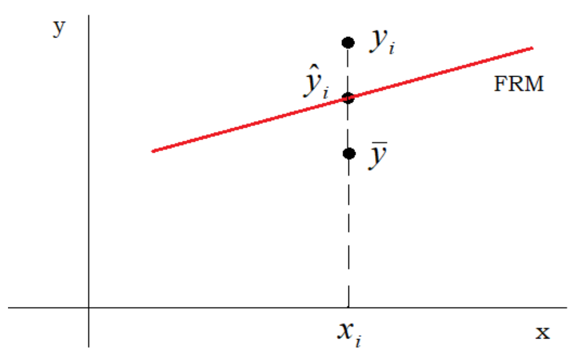


Figura 1. La variabilidad total del modelo de regresión y sus componentes

Fuente: Adaptación de (Gujarati & Porter, 2010, pág. 374)

$$\underbrace{(y_i - \bar{y})}_{\text{Variabilidad total}} = \underbrace{(\hat{y}_i - \bar{y})}_{\text{Variabilidad explicada}} + \underbrace{(y_i - \hat{y}_i)}_{\text{Variabilidad no explicada (aleatoria)}}$$

Figura 2. Ecuación de la variabilidad total

Fuente: (Gujarati & Porter, 2010, pág. 551)

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

$$STC = SEC + SRC$$

STC es la Suma Total de Cuadrados (variabilidad total)
SEC es la Suma Explicada de Cuadrados (variabilidad explicada por el modelo)
SRC es la Suma de Residuales al Cuadrado (variabilidad aleatoria)

Para que el modelo de regresión lineal sea significativo se requiere que

$$SEC >>>> SRC$$

Es decir, que la variabilidad explicada por el modelo sea muy superior a la variabilidad aleatoria o debida al azar (Gujarati & Porter, 2010, pág. 197).

Tabla 1. ANOVA resume las diferentes fuentes de variación

Fuente de variación	*Suma de cuadrados*	*Grados de libertad*	*Cuadrados Medios*	*Estadístico*
Modelo	$SEC=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	1	$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2/1$	$F=\frac{CMModelo}{CM\ \mathrm{Re}siduales}$
Residuales (Error aleatorio)	$SRC=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	$n-2$	$\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}$	
Total	$STC=\sum_{i=1}^{n}(y_i-\bar{y})^2$	$n-1$		

Fuente: Adaptación de (Gujarati & Porter, 2010, pág. 197)

Para comprobar la validez del modelo se realiza una prueba de hipótesis (prueba F)

Se prueba la hipótesis nula

$$H_0: \ \beta_1 = 0$$

La pendiente de la recta es cero, es decir, no existe relación alguna entre las variables $x \wedge y$

Contra la hipótesis alterna

$$H_1 : \beta_1 \neq 0$$

La cual indica que sí existe relación entre las variables $x \wedge y$

Para un nivel de significancia del 5%, si el p-value<0.05, entonces se rechaza la hipótesis nula y se concluye que el modelo de regresión explica significativamente la variación observada en la variable dependiente y. En caso contrario, no se rechaza la hipótesis nula y se concluye que toda la variabilidad de la variable dependiente se explica mediante perturbaciones aleatorias. (Gujarati & Porter, 2010, pág. 235)

Coeficiente de Determinación (R^2)

Es una medida de la bondad de ajuste del modelo de regresión. Mide el porcentaje de variación total de la variable dependiente y que es explicada por el modelo de regresión.

Se calcula mediante la expresión $R^2 = \dfrac{SEC}{STC}$

Propiedades:

1. $R^2 \geq 0$

2. $0 \leq R^2 \leq 1$

$R^2 = 1$ indica un ajuste perfecto mientras que $R^2 = 0$ indica que no hay relación alguna entre la variable dependiente y la variable explicativa (Gujarati & Porter, 2010, pág. 73).

Error estándar de la regresión $\hat{\sigma}$

"Es una medida de la variabilidad de los valores de la variable dependiente y en torno a la recta de regresión" (Gujarati & Porter, 2010, pág. 69). Es importante que este valor sea pequeño como se ilustra en la siguiente figura:

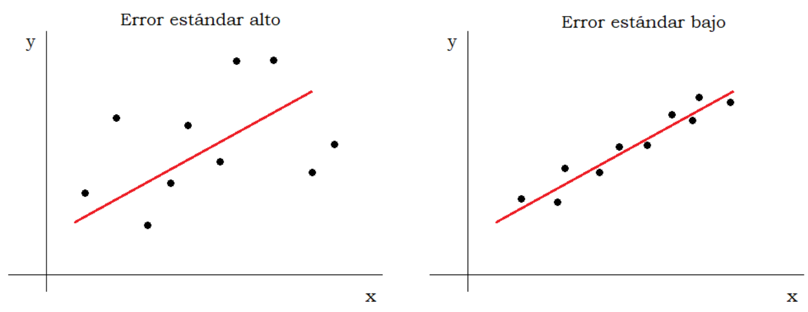


Figura 3. Error estándar de la regresión

Fuente: Adaptación de (Gujarati & Porter, 2010, págs. 116-117)

El mejor estimador del error estándar de la regresión es el cuadrado medio de los residuales

$$\hat{\sigma} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}$$

Prueba de normalidad de Shapiro – Wilk

Se emplea para contrastar la normalidad cuando el tamaño de la muestra sea máximo de 50 es decir (n < 50) y en muestras grandes es equivalente a la prueba de Kolmogórov-Smirnov. El método consiste en comenzar ordenando la muestra de menor a mayor valor, obteniendo el nuevo vector muestral. Para contrastar la normalidad con la prueba de Shapiro-Wilk, se procede a calcular la media y la varianza muestral. Se rechaza la hipótesis nula de normalidad si el estadístico Shapiro-Wilk -W- es menor que el valor crítico proporcionado por la tabla elaborada por los autores para el tamaño de la muestra y el nivel de significancia dado. (Flores Tapia & Flores Cevallos, 2021, pág. 87)

Shapiro-Wilk, como prueba de normalidad, fue introducido considerando que el gráfico de probabilidad normal que examina el ajuste de un conjunto de datos de muestra para la distribución normal es semejante a la de regresión lineal - la línea diagonal del gráfico es la recta de ajuste perfecto-, con la diferencia de que esta línea es similar a los residuos de la regresión. Mediante el análisis de la magnitud de esta variación - análisis de varianza-, la calidad del ajuste puede ser examinado. (Flores Tapia & Flores Cevallos, 2021, pág. 87)

Regresión con variables transformadas

Se utiliza cuando el comportamiento de los datos es no – lineal. El diagrama de dispersión preliminar es de gran ayuda porque permite apreciar la necesidad de realizar una transformación en los datos. Aplicando las transformaciones apropiadas se obtiene un modelo de regresión lineal en los parámetros. (Gujarati & Porter, 2010, pág. 442)

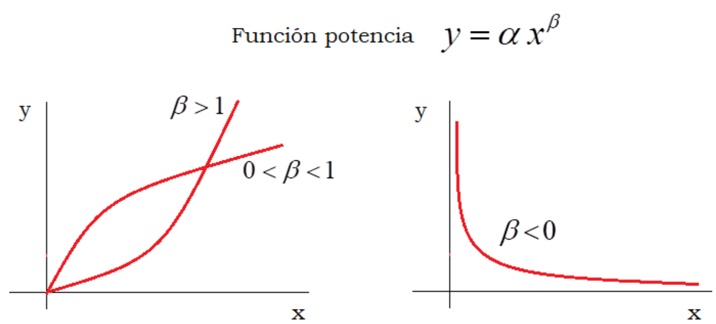


Figura 4. Diagrama no- lineal de la función potencia

Fuente: Adaptación de (Gujarati & Porter, 2010, pág. 39)

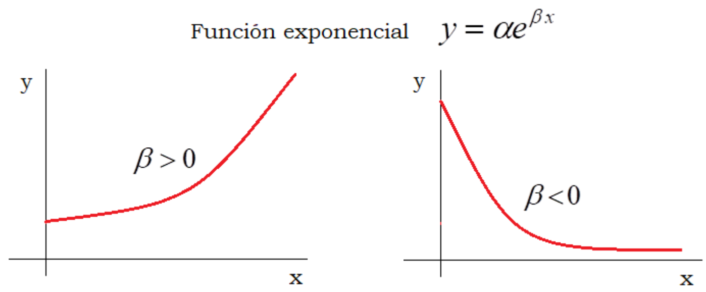


Figura 5. Diagrama no-lineal de la función exponencial

Fuente: Adaptación de (Gujarati & Porter, 2010, pág. 39)

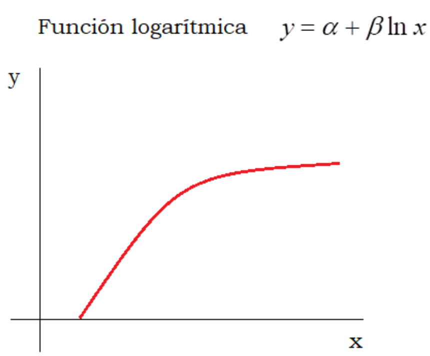


Figura 6. Diagrama no-lineal de la función logarítmica

Fuente: Adaptación de (Gujarati & Porter, 2010, pág. 172)

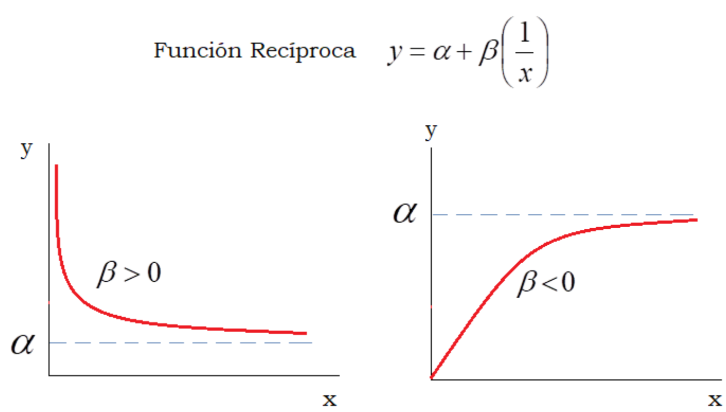


Figura 7. Diagrama no-lineal de la función recíproca

Fuente: Adaptación de (Gujarati & Porter, 2010, pág. 167)

Aplicando las transformaciones apropiadas se obtiene un modelo de regresión lineal en los parámetros α y β. La siguiente tabla presenta un resumen de las transformaciones necesarias y de la forma lineal obtenida:

Tabla 2. Transformaciones y modelo lineal obtenido de cada función

Forma funcional	*Transformación*	*Modelo lineal*
Potencial $y=\alpha x^{\beta}$	$y^{*}=\ln y,\ x^{*}=\ln x$	$y^{*}=\ln\alpha+\beta x^{*}$
Exponencial $y=\alpha e^{\beta x}$	$y^{*}=\ln y$	$y^{*}=\ln\alpha+\beta x$
Logarítmica $y=\alpha+\beta\ln x$	$x^{*}=\ln x$	$y=\alpha+\beta x^{*}$
Recíproca $y=\alpha+\beta\left(\frac{1}{x}\right)$	$x^{*}=\frac{1}{x}$	$y=\alpha+\beta x^{*}$

Fuente: Adaptación de (Walpole & Myres, 1998).

Regresión Polinomial

En la ingeniería, aunque algunos datos exhiben un patrón marcado, son pobremente representados por una línea recta, entonces, una curva podrá ser más adecuada para ajustarse a los datos. Un método para lograr este objetivo es utilizar transformaciones. Otra alternativa es ajustar polinomios a los datos mediante regresión polinomial. (Chapra, y otros, 2011, pág. 482)

Los modelos no lineales con variables transformadas se refieren a funciones de la variable independiente x que son crecientes o decrecientes en un sentido estricto. En muchas situaciones, al construir el gráfico de dispersión de los datos, se observa que la función de regresión contiene uno o más máximos o mínimos. Esto indica que la función de regresión verdadera puede aproximarse mediante un modelo polinomial. (Gujarati & Porter, 2010, pág. 210)

El modelo de regresión polinomial de grado n se escribe:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n$$

Regresión Lineal Múltiple

"El análisis de regresión lineal múltiple es un análisis de regresión condicional, donde se mantienen fijos los valores de las variables explicativas y se obtiene el valor promedio o la respuesta media de la variable dependiente o regresada". (Gujarati & Porter, 2010, pág. 188)

La función de regresión poblacional para el caso de k variables explicativas está dada por:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + \ldots.. + b_k X_{ki} + e_i$$

Donde e_i es el error o perturbación estocástica.

La función de regresión muestral (FRM) que permite estimar la función de regresión poblacional está dada por:

$$Yest_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + B_3 X_{3i} + \ldots.. + B_K X_{ki}$$

Donde $B_0, B_1, B_2, B_3, \ldots, B_k$ representan los estimadores de los parámetros del modelo, $b_0, b_1, b_2, b_3, \ldots., b_k$.

$Yest = Y$ estimada por el modelo de regresión.

2.3. Marco conceptual

Se ha considerado de gran importancia definir algunos conceptos que son utilizados en el proyecto y que permiten aclarar su significado a los lectores.

Acueducto: Se entiende por Sistemas de Acueducto, el conjunto de instalaciones que conducen el agua desde su captación en la fuente de abastecimiento hasta la acometida domiciliaria en el punto de empate con la instalación interna del predio a suministrar el servicio de agua potable. (Ramirez & De Plaza, 2016, pág. 11)

Alcantarillado: "Sistema de tuberías y buzones usado para la recolección y transporte de las aguas residuales generadas por la población desde las viviendas (conexiones domiciliarias) hasta el sitio en que se tratan". (Gobierno Regional Cusco, 2018)

Ecuación: Thomas Cerda define ecuación como: "una proposición, que afirma la igualdad de dos cantidades, o de dos sumas por medio de la señal (=) general de la Algebra para significar igualdad, el cual media entre dos partes, que se llaman Miembros de la Ecuación". (Madrid, León Mantero, Maz Machado, & López , 2019, pág. 407)

Estadística: La estadística consiste en métodos, procedimientos y fórmulas que permiten recolectar información para luego analizarla y extraer de ella conclusiones relevantes. Se puede decir que es la Ciencia de los Datos y que su principal objetivo es mejorar la comprensión de los hechos a partir de la información disponible. (Roldán, 2017)

Urbanismo: El Diccionario Jurídico Espasa define al urbanismo como la ciencia o técnica dirigida a ordenar tanto la ciudad (inicial contenido del urbanismo) como su entorno, pudiendo alcanzar sucesivamente el ámbito municipal, provincial, regional o nacional. Según Tomás Ramón Fernández, el urbanismo es, en la actualidad, "una perspectiva global e integradora de todo lo que se refiere a la relación del hombre con el medio en el que se desenvuelve y que hace de la tierra, del suelo, su eje operativo". (Sánchez Luna, 1995, pág. 307)

Edificaciones: "Son obras que diseña, planifica y ejecuta el ser humano en diferentes espacios, tamaños y formas, en la mayoría de los casos para habitarlas o usarlas como espacios de resguardo". (Definición ABC, 2009)

Hidráulica: "Es la rama de la Física que se encarga de estudiar el comportamiento y el movimiento de los fluidos". (Hernández, 2014, pág. 2)

ANOVA: Desde el punto de vista de la regresión se conoce como análisis de varianza el cual estudia los componentes de la suma de cuadrados total el cual fragmenta la suma de dos componentes; la suma de cuadrados explicada y la suma de cuadrados de residuos (Gujarati & Porter, 2010, pág. 123).

*Residencia unifamiliar: "*Es el subgrupo de ocupación residencial unifamiliar en donde se clasifican las edificaciones o espacios empleados principalmente como vivienda o dormitorio de una familia, o de menos de 20 personas" (Asociación Colombiana de la Ingeniería Sismica, 2010, pág. 10).

*Residencia multifamiliar: "*Es el subgrupo de ocupación residencial multifamiliar en donde figuran las edificaciones o espacios empleados principalmente como vivienda, o como dormitorio de tres o más familias, o de más de 20 personas" (Asociación Colombiana de la Ingeniería Sismica, 2010, pág. 11).

2.4. Marco contextual

La problemática de no saber cuánto cobrar por un diseño de cualquier tipo es una realidad que se presenta por igual para la mayoría de los ingenieros civiles en el mundo. Para el caso de los diseños hidráulicos, existen métodos comúnmente utilizados, sin embargo, depende mucho de la caracterización del proyecto.

Al ser una problemática que involucra a los ingenieros civiles e hidráulicos en el mundo, la investigación se delimitó en recolectar datos de los últimos 5 años (2017-2021) de proyectos realizados en Norte de Santander.

Norte de Santander es uno de los 32 departamentos de Colombia. Está ubicado en la zona nororiental del país, sobre la frontera con Venezuela. Se localiza geográficamente entre los

06º56’42’ y 09º18’01’’ de latitud norte y los 72º01’13’’ y 73º38’25’’ de longitud oeste. Tiene una extensión de 22.130 km2, que equivalen al 1.91% del millón ciento cincuenta y nueve mil ochocientos setenta y un kilómetros cuadrados (1.159.871,41 km2) del territorio nacional. Limita al norte y al este con Venezuela, al sur con los departamentos de Boyacá y Santander, y al oeste con Santander y Cesar. Forma parte de la Región Andina junto con los departamentos de Antioquia, Boyacá, Caldas, Cundinamarca, Huila, Santander, Quindío, Risaralda, y Tolima, la más densamente poblada del país, donde reside más del 70% de la población colombiana.

El departamento de Norte de Santander está dividido en 40 municipios y 108 corregimientos.

- Gentilicio: Nortesantandereano/a
- Superficie: 21.648 km2
- Población: 1.492 millones de Habitantes (2018)
- Densidad: 66.8 Hab/Km2
- Capital: San José de Cúcuta – 711.715 Habitantes (2018)

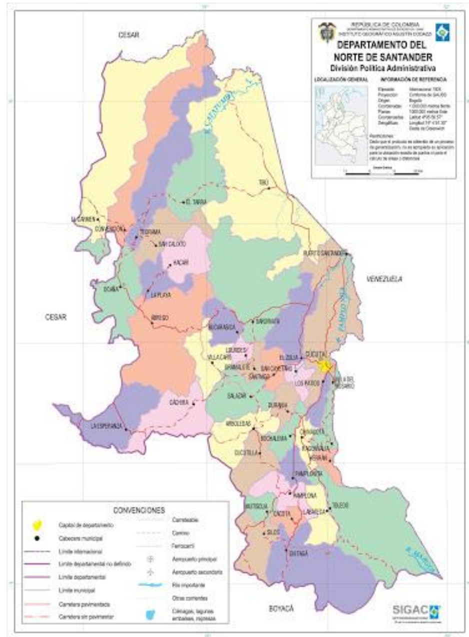


Figura 8. Departamento de Norte de Santander

Fuente: Obtenido en http://www.colombiamania.com/departamentos/nortedesantander.html

2.5. Marco legal

La propuesta de la investigación planteada se rige por los artículos del estatuto estudiantil de la universidad Francisco de Paula Santander como son el articulo 140 (Reglamentado mediante Acuerdo 069/1997) y el articulo 141.

Código colombiano de instalaciones hidráulicas y sanitarias (NTC 1500) – 2020: Esta norma presenta disposiciones aplicables al montaje, la instalación, la modificación, las reparaciones, la reubicación, el reemplazo, la ampliación, el uso o el mantenimiento de sistemas de instalaciones hidráulicas y sanitarios.

Comisión de regulación de agua potable y saneamiento (CRA)- 2008: En el capítulo referente al “Estudio de estructuración y análisis de información de inversiones de los prestadores de acueducto y alcantarillado" fue elaborada varias ecuaciones de costo referente a la tubería en los procesos de acueducto y alcantarillado.

Reglamento Técnico para el sector de agua potable y saneamiento básico. (RAS 2017) Resolución 0330 del 2017 y Resolución 799 del 2021: Con respecto al título B se definen los términos de los procesos de redes de distribución de agua potable y alcantarillado para tener en cuenta durante el proceso de elaboración de la metodología, además de definir diferentes parámetros de cálculos para tener en cuenta.

3. Diseño metodológico

Para lograr el objetivo de este proyecto de investigación, es de principal importancia consultar los referentes teóricos que tengan como enfoque las ecuaciones en función del costo de un diseño hidráulico. Para iniciar la metodología se realizará la recolección de una serie de proyectos hidráulicos suministrados por diferentes constructoras durante los últimos 5 años en el departamento Norte de Santander, los cuales se clasificarán según su tipología.

Se creará un sistema de organización de acuerdo con las diferentes variables ya definidas tales como: longitud de tubería, número de viviendas o apartamentos y el área total de construcción, con el fin de analizarlas estadísticamente para así formular una serie de ecuaciones o modelos matemáticos con fundamento en la metodología de Regresión Lineal y con ayuda del software estadístico R, de esta manera se puede obtener el costo de los diseños hidráulicos acorde a las características y variables especificadas con anterioridad. Luego se procederá a evaluar las ecuaciones metodológicas formuladas por medio de un análisis comparativo del costo real de proyectos y el que arrojan las ecuaciones para verificar su eficacia y su error estándar.

El análisis estadístico de la investigación se realiza por medio del software R, este software es un lenguaje y entorno de programación que permite efectuar cálculos numéricos, análisis estadísticos y gráficos. Fue creado por Ross Ihaka y Robert Gentleman del Departamento de Estadística de la Universidad de Auckland. Se considera un dialecto del lenguaje S, un lenguaje de manipulación de objetos creado en los Laboratorios AT & T Bell. (Rodríguez Silva, 2019, pág. 3)

R se distribuye en forma gratuita bajo licencia GNU General Public Licence. El desarrollo de este lenguaje está a cargo de un equipo de trabajo conocido como “R Development Core Team.” Los archivos necesarios para instalar R se descargan desde los sitios CRAN

(Comprehensive R Archive Network) de internet. En Colombia, se puede descargar directamente desde https://www.icesi.edu.co/CRAN/.

Al igual que S, R es un lenguaje orientado a objetos. Debe aclararse que R es un lenguaje interpretado y no compilado, esto significa que los comandos introducidos con el teclado se ejecutan directamente sin necesidad de construir ejecutables. R considera como objetos las variables, datos, funciones, resultados, etc., y los almacena en la memoria activa del computador. Estos objetos pueden ser manipulados por medio de operadores y funciones. (Santana & Farfán, 2014, pág. 7)

Las funciones están disponibles en una librería localizada en el directorio donde R está almacenado. Esta librería está conformada por una serie de paquetes (packages) de funciones que pueden ser utilizados de acuerdo con las necesidades del usuario. Por ejemplo, el paquete “base” es el núcleo de R y contiene las funciones básicas de lectura y manipulación de datos, así como funciones gráficas y estadísticas básicas (Santana & Farfán, 2014, pág. 10). Existe una gran cantidad de paquetes disponibles en la red a los que se puede acceder fácilmente desde http://www.cran.r-project.org.

Según (Rodríguez Silva, 2019, pág. 6) dentro de las características más notables de R se pueden mencionar las siguientes:

1. Contiene una amplia librería de funciones orientadas a procesos de computación numérica, manipulación y análisis de datos.
2. Una de sus mayores fortalezas es la capacidad para generar gráficas en dos y tres dimensiones, desde las más simples hasta las más sofisticadas.

3. Es un entorno de programación flexible que permite al usuario crear sus propias funciones.
4. Disponibilidad en la red de paquetes de funciones destinados a aplicaciones específicas.
5. R fue desarrollado bajo la filosofía del software libre. Por tal razón, el usuario tiene la libertad de usar, Copiar y distribuir el programa. También puede acceder al código fuente, realizar modificaciones y compartir esos cambios con la comunidad de usuarios del programa.

3.1. Tipo de investigación

La investigación es *cuantitativa* porque se recopilan y analizan datos numéricos con el fin de solucionar una problemática, además se emplea el análisis estadístico para establecer patrones en el comportamiento de los datos que permitirá la creación de un sistema de cobro.

El enfoque cuantitativo utiliza la recolección y el análisis de datos para contestar preguntas de investigación y probar hipótesis establecidas previamente, y confía en la medición numérica, el conteo y frecuentemente en el uso de la estadística para establecer con exactitud patrones de comportamiento en una población. (Sampieri Hernández, Collado Fernández, & Lucio Baptista, 2003, pág. 79)

Investigación según el propósito. La investigación es *aplicada* debido a que lleva a la práctica el desarrollo de las ecuaciones que estiman el costo de diseños hidráulicos para así satisfacer la necesidad del ingeniero civil en obtener una herramienta que le sea útil para el diseño hidráulico mediante la aplicación de conocimientos desarrollados en las ciencias puras.

Investigación según el nivel. Corresponde a una investigación *descriptiva* puesto que en su proceso de elaboración cuenta con diversos componentes que corresponden a la división de proyectos según su tipología y además de variables como longitud de tubería, número de apartamentos o viviendas y área total de construcción que serán de importante objeto de estudio.

Investigación según la estrategia. Hace parte de la investigación *documental* en donde será de importante relevancia los referentes teóricos existentes sobre la estimación de costos en los diseños hidráulicos y en donde su elaboración podrá ser escrito en la presente investigación.

3.2.Población y muestra

3.2.1. Población

La investigación va dirigida a ingenieros civiles independientes y empresas de ingeniería civil del país que realicen servicio de consultoría hidráulica o diseño de redes hidrosanitarias en construcciones civiles a quienes el uso de las ecuaciones les que garantice un costo adecuado en la ejecución de su labor.

3.2.2. Muestra

Se toma como muestra proyectos hidráulicos en urbanismo y edificaciones de municipios de Norte de Santander con el fin de analizar estadísticamente los datos obtenidos de estos proyectos y elaborar en base a los resultados, unas ecuaciones de estimación de costos que servirán como apoyo a la ingeniería civil.

Debido a que la investigación es *descriptiva* y de que el muestreo es *probabilístico estratificado* se tendrá en cuenta una muestra grande y a su vez está se dividirá por subgrupos de acuerdo a las variables del proyecto, es por esto que, se seleccionarán 40 proyectos hidráulicos en urbanismo, 40 proyectos hidráulicos en edificaciones y 40 en otros tipos (canchas, colegios, parques, etc.) dentro del departamento de Norte de Santander, obteniendo un total de 120 proyectos para la elaboración de la investigación en curso.

3.3. Instrumentos para la recolección de información

Fuentes primarias. Estas corresponden a la información que se recolectará de manera directa con el fin de aportar datos que son útiles en el desarrollo del proyecto, de esta forma, la búsqueda de constructoras que tengan alta demanda de en proyectos hidráulicos en el departamento de Norte de Santander, permitirá llegar a la obtención del material estudiado; en este sentido se lograrán ejecutar los objetivos planteados.

Fuentes secundarias. Entre las fuentes secundarias que servirán de apoyo para la realización del proyecto tenemos: consultar información relacionada con la estimación de costos en proyectos hidráulicos; especificaciones de normas que vigilan la calidad del proyecto y su forma de costo, como es el RAS 2017, Resolución 0330 del 2017; también se tendrán en cuenta tesis de grado e investigaciones similares que se relacionen con el cálculo del costo de un diseño hidráulico, lo cual permitirá analizar las ventajas y desventajas en la utilización de la metodología estimada para el planteamiento de las ecuaciones, y enriquecerá la investigación que resulte relevante para llevar a cabo los objetivos del proyecto aquí expuesto.

En la recolección de datos es necesario tener una búsqueda exhaustiva de información relevante para llevar a cabo la ejecución de la investigación, en donde se utilizarán herramientas para su obtención como base de datos en donde se encontrarán revistas, libros, tesis doctorales, normas, entre otros. Los referentes teóricos serán obtenidos de fuentes confiables que puedan corroborar su validez con la finalidad de reflejar un sólido proceso investigativo.

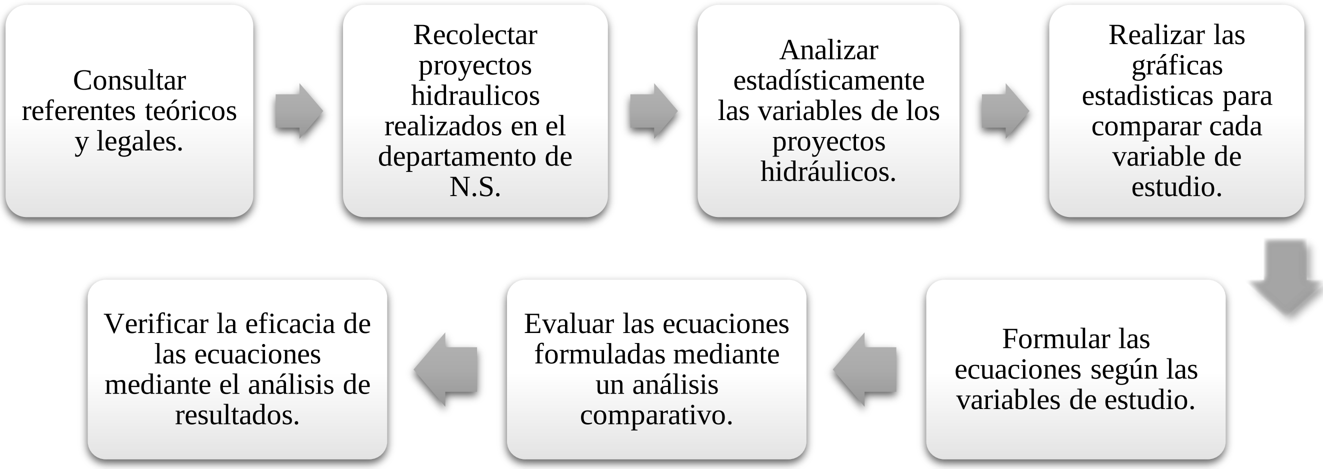
La información de los proyectos hidráulicos a analizar, serán suministrados por diferentes constructoras que han trabajado en el área de acueductos y alcantarillado durante los últimos 5 años en el departamento Norte de Santander, los cuales deben tener la información del costo para poder analizar las características a tratar en la investigación.

3.4. Técnicas de análisis y procesamiento de datos

La presentación de los resultados será de tipo *descriptiva*, en donde se realizarán gráficas y tablas estadísticas que permitirán hacer un análisis a los resultados obtenidos, además se presentarán cuadros comparativos según las variables de estudio, tablas que expongan cada proyecto con su costo en relación con cada variable y también gráficas o diagramas que muestren el comportamiento de los costos para las distintas formas de pago establecidas en el proyecto.

Los datos serán manejados con software R o lenguaje de programación para realizar análisis estadístico, el cual permite programar bucles (Loops en inglés) para analizar conjuntos sucesivos de datos, además de combinar en un solo programa diferentes funciones estadísticas para realizar análisis complejos. Este programa será de utilidad para la ejecución de gráficos, depurador de datos y análisis estadístico de todas las variables del proyecto.

3.5. Fases y actividades específicas del proyecto



Para cumplir el objetivo 1: “Consultar referentes teóricos y legales que den respuesta al objeto de estudio con el fin de observar el conjunto de variables que inciden en el costo al diseñar redes hidráulicas”.

- Consultar por medio de buscadores web o base de datos de bibliotecas, que ofrezcan información sobre el tema de estudio.
- Realizar un estado del arte que identifique los referentes con su autor, y fecha y el aporte que ofrece a la investigación.
- Organizar los referentes de los más recientes a los más antiguos según el año.

Para cumplir el objetivo 2: “Recolectar información de las variables en proyectos hidráulicos obtenidos de diferentes ingenieros y empresas de Norte de Santander en los últimos 5 años (2017-2021)”.

- Solicitar por medio digital o presencial a entidades públicas o privadas de Norte de Santander la información como planos y costos manejados en la realización de los diseños.

- Organizar los proyectos en tres categorías: edificaciones, urbanismos y conjuntos cerrados, y otros (canchas, piscinas, parques, colegios, etc.)

Para cumplir el objetivo 3: “Analizar estadísticamente las variables de los proyectos hidráulicos, mediante la comparación en gráficas de cada dato obtenido”.

- Emplear la metodología de regresión lineal y el software R para realizar el análisis estadístico de la información que permita la elección y construcción del modelo matemático que mejor ajusta los datos.

Para cumplir el objetivo 4: “Formular una serie de ecuaciones por medio del análisis estadístico que identifique variables como área, número de viviendas y longitud de tuberías en distintas tipologías de proyectos”.

- Construir los modelos matemáticos de regresión lineal para cada una de las situaciones en estudio.
- Determinar la validez o significancia estadística de los modelos de regresión lineal construidos y verificar el cumplimiento de los supuestos más importantes indicados por la teoría.

Para cumplir el objetivo 5: “Evaluar las ecuaciones metodológicas formuladas por medio de un análisis comparativo con proyectos hidráulicos existentes para verificar su eficacia”.

- Emplear proyectos elaborados en Norte de Santander para ingresar sus datos en las ecuaciones y comprobar su funcionalidad.
- Repetir el proceso para cada categoría y variable de estudio con el fin de obtener 9 ecuaciones que permitan estimar el costo por área, número de casas o apartamentos y longitud de tubería en proyectos de edificaciones, urbanismos y conjuntos y otros.

4. Contenido del proyecto

4.1. Generalidades

Una vez recolectados los planos de los diseños hidráulicos de constructoras del municipio de Norte de Santander con su debido costo, se obtienen de ellos la información que definirá las variables explicativas estudiadas:

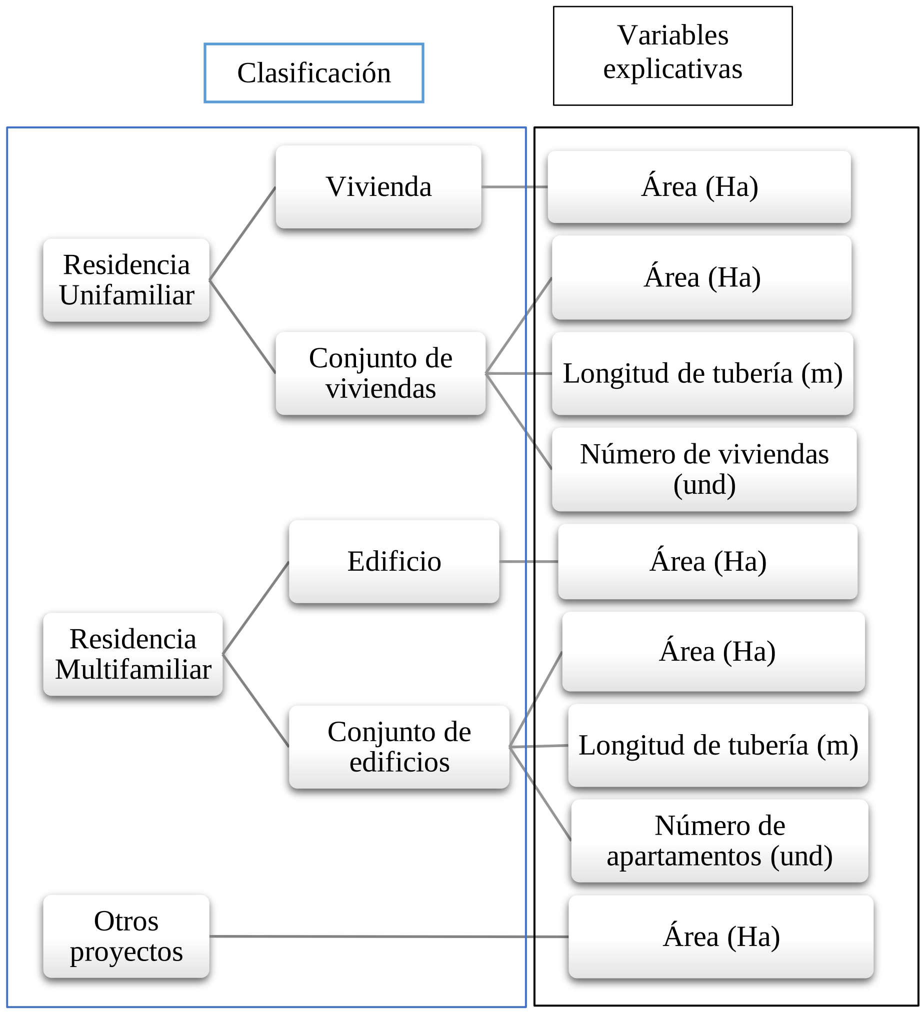


Figura 9. Clasificación de las variables explicativas

Luego de obtener los respectivos datos de los planos según su clasificación (residencia unifamiliar, multifamiliar y otros proyectos) se organizan en tablas indicando el año, nombre del proyecto, empresa responsable, ubicación, tipología y costo, con el fin de tener un mejor manejo de la información.

Con la organización de los datos se procede a realizar el análisis mediante el software R, empleando distintos modelos funcionales con el fin de evaluar el que mejor se ajuste y tenga mayor significancia estadística. El software R permite emplear los modelos de regresión lineal simple y regresión lineal múltiple, el cual permite evaluar distintas variables en un solo modelo.

Una vez se realiza la programación, se construyen las ecuaciones mediante los datos obtenidos en el modelo y se selecciona aquel cuyo coeficiente de determinación R2, y p-value sean estadísticamente significativos, de este modo se obtiene la ecuación del modelo funcional elegido con el fin de evaluarla con los proyectos de la base de datos recolectada, comparar los resultados de los costos obtenidos con la ecuación y los costos reales, y finalmente verificar su eficacia y aplicación en la ingeniería civil.

A continuación, se presentan las tablas que se emplean en la investigación con la organización de los datos de los proyectos hidráulicos suministrados por empresas consultoras especializadas de la región de Cúcuta como lo son Hidraforcis SAS, SATAR ingeniería, CABG ingeniería y el consultor Iván Hernando Ramírez Mendoza:

Tabla 3. Datos vivienda y conjunto de viviendas, residencia unifamiliar

No.	Año	Nombre proyecto	Empresa	Ubicación	Tipo	Costo proyecto ($ COP)	L. Tubería (m)	Área (Ha)	N. Viviendas (und)
1	2017	Girasoles	Paisaje Urbano	Villa del Rosario	Conjunto Vivienda	$ 5,000,000	1520,80	2,067	162
2	2017	La Macarena	Paisaje Urbano	Villa del Rosario	Conjunto Vivienda	$ 4,500,000	1162,41	1,853	132
3	2017	White Country House	Paisaje Urbano	Villa del Rosario	Conjunto Vivienda	$ 4,000,000	859,58	1,280	50
4	2017	El Recreo	Paisaje Urbano	Villa del Rosario	Conjunto Vivienda	$ 4,500,000	2173,00	1,515	114
5	2017	Altobelo	Colproyectos	Cúcuta	Conjunto Vivienda	$ 3,000,000	1754,89	1,580	150
6	2017	Firenze	Colproyectos	Cúcuta	Conjunto Vivienda	$ 3,000,000	1296,30	1,564	110
7	2017	Villa Teresa	Tour Colombia	Puerto Santander	Conjunto Vivienda	$ 3,200,000	2962,10	3,715	250
8	2017	Pinares	Colproyectos	Cúcuta	Conjunto Vivienda	$ 3,500,000	1788,33	2,159	60
9	2017	Vivienda Bellavista	Arq. Javier Bustamante	Los Patios	Vivienda	$ 800,000	-	0,023	-
10	2018	Brisas de Arkamar	Arkamar Inversiones	Cúcuta	Conjunto Vivienda	$ 5,800,000	1628,1	1,254	108
11	2018	Vivienda Niza	Arq. Javier Bustamante	Cúcuta	Vivienda	$ 700,000	-	0,012	-
12	2018	Eduviges	Ingeniero Jesús Vargas	Toledo	Vivienda	$ 1,000,000	-	0,095	-
13	2018	Altagracia	Colproyectos	Cúcuta	Conjunto Vivienda	$ 4,000,000	2042,00	1,866	176
14	2018	Acacios	Juan Maldonado	Cúcuta	Vivienda	$ 700,000	-	0,019	-
15	2018	Palujan	Constructora Andina Gómez	Los Patios	Vivienda	$ 1,000,000	-	0,078	-
16	2018	Colina Campestre	OVINCO SAS	Villa del Rosario	Vivienda	$ 700,000	-	0,010	-
17	2018	Punta del Este	Arq. Álvaro Maldonado	Villa del Rosario	Vivienda	$ 800,000	-	0,064	-
18	2018	Los naranjos	Constructora La Valenciana	Los Patios	Conjunto Vivienda	$ 3,700,000	187,4	2,600	180
19	2018	Canela	Constructora MONAPE	Cúcuta	Conjunto Vivienda	$ 4,200,000	1726,58	3,067	144
20	2019	Ebano	Paisaje Urbano	Villa del Rosario	Conjunto Vivienda	$ 5,500,000	1779,00	2,259	206

No.	Año	Nombre proyecto	Empresa	Ubicación	Tipo	Costo proyecto ($ COP)	L. Tubería (m)	Área (Ha)	N. Viviendas (und)
21	2019	Medinterrané	OVINCO SAS	Villa del Rosario	Conjunto Vivienda	$ 3,500,000	2030,69	1,149	58
22	2019	Bosque Alto	ODICCO SAS	Los Patios	Conjunto Vivienda	$ 2,500,000	234,03	0,178	11
23	2019	Molinos	EL Palustre	Cúcuta	Conjunto Vivienda	$ 2,200,000	493,17	0,261	31
24	2019	Rayuela	Paisaje Urbano	Villa del Rosario	Conjunto Vivienda	$ 4,000,000	2547,08	1,469	114
25	2019	Vallarta	Paisaje Urbano	Villa del Rosario	Conjunto Vivienda	$ 4,500,000	1897,55	2,249	197
26	2019	Villa Bolívar	Constructora San Nicolas	Cúcuta	Conjunto Vivienda	$ 5,500,000	1334,38	1,911	153
27	2019	Serrano Castellanos	Arq. Javier Bustamante	Los Patios	Vivienda	$ 800,000	-	0,042	-
28	2019	Gonzáles	Arq. Javier Bustamante	Cúcuta	Vivienda	$ 800,000	-	0,061	-
29	2019	OV1	Arq. Javier Bustamante	Villa del Rosario	Vivienda	$ 800,000	-	0,082	-
30	2019	Palma Deluxe	Confemo	Villa del Rosario	Vivienda	$ 1,000,000	-	0,025	-
31	2019	Palma Country	Confemo	Villa del Rosario	Vivienda	$ 1,000,000	-	0,050	-
32	2019	Palma Dorada	Confemo	Villa del Rosario	Vivienda	$ 1,000,000	-	0,050	-
33	2020	Carolina Campestre	Arq. Julián Jaimes	Cúcuta	Conjunto Vivienda	$ 1,500,000	308,14	0,225	19
34	2020	Carolina Deluxe	Caoba Grupo Inmob	Cúcuta	Conjunto Vivienda	$ 1,500,000	160,84	0,145	10
35	2020	El Verjel	Arq. Javier Bustamante	Los Patios	Conjunto Vivienda	$ 2,500,000	67,98	0,121	18
36	2020	Villa Verona	ODICCO SAS	Chinácota	Conjunto Vivienda	$ 2,500,000	1052,38	1,348	100
37	2020	Valena	Constructora San Nicolas	Cúcuta	Conjunto Vivienda	$ 2,500,000	467,19	0,456	24
38	2020	Cabaña Corozal	Arq. Álvaro Maldonado	Los Patios	Vivienda	$ 1,200,000	-	0,200	-
39	2020	Casagrande	OVINCO SAS	Cúcuta	Vivienda	$ 800,000	-	0,025	-
40	2020	Zafiro	Confemo	Cúcuta	Vivienda	$ 1,000,000	-	0,054	-

Tabla 4. Datos edificio y conjunto de edificios, residencia multifamiliar

No.	Año	Nombre proyecto	Empresa	Ubicación	Tipo	Costo proyecto ($ COP)	L. Tuberia (m)	Área (Ha)	N. Apartamentos (und)
1	2017	Altas Torres de Claret	Constructora NRS	Cúcuta	Conjunto Edificio	$ 4.000.000	1270,01	2,496	640
2	2017	Altos Zulia VIP	Monape	Zulia	Conjunto Edificio	$ 7.000.000	657,6	1,211	240
3	2017	Los Estoraques ciudadela	Urb. Los estoraques LTDA	Cúcuta	Edificio	$ 6.000.000	-	0,023	-
4	2017	Urb. San Rafael	ODICCO SAS	Labateca	Conjunto Edificio	$ 2.000.000	181,01	0,554	96
5	2017	Urb. Santa Eduviges	ODICCO SAS	Toledo	Conjunto Edificio	$ 5.000.000	1412,35	1,081	200
6	2018	Altos del Este	Paisaje Urbano	Cúcuta	Conjunto Edificio	$ 6.000.000	857,5	0,052	168
7	2018	Arkadia	Paisaje Urbano	Cúcuta	Conjunto Edificio	$ 6.000.000	1187,3	0,972	160
8	2017	Chibará	Constructora Yadel	Cúcuta	Conjunto Edificio	$ 9.200.000	1246,46	1,312	400
9	2018	Palma Redonda	Colproyectos	Cúcuta	Conjunto Edificio	$ 8.000.000	1093,15	1,285	288
10	2017	Nuevo Pamplona	Unión temporal nuevo Pamplona	Pamplona	Conjunto Edificio	$ 8.500.000	1142,56	0,876	160
11	2019	Valmiera	Pinzón Pacheco Investments SAS	Cúcuta	Edificio	$ 1.200.000	-	0,036	-
12	2020	Altos del Jardín	Paisaje Urbano	Cúcuta	Conjunto Edificio	$ 4.000.000	1350,91	0,589	180
13	2021	Colsag La Valle	Grupo CR Ingeniería SAS	Cúcuta	Edificio	$ 4.000.000	-	0,041	-
14	2019	Edificios La Manuela	Inversiones Arkamar	Cúcuta	Conjunto Edificio	$ 2.200.000	204,83	0,366	80
15	2017	Altos de la Candelaria	Constructora NOVATEC	Cúcuta	Conjunto Edificio	$ 5.500.000	538,81	0,554	180
16	2019	La Palmita	Constructora OASIS	Villa del Rosario	Conjunto Edificio	$ 4.000.000	507,58	0,313	64
17	2018	Ciudadela Las flores	Triada SAS	Cúcuta	Conjunto Edificio	$ 5.500.000	1963,24	2,049	328
18	2018	Terraviva	Constructora Yadel	Cúcuta	Conjunto Edificio	$ 12.000.000	2282,76	2,257	480
19	2018	Casa los faroles	Elizabeth Galindo Quiroz	Cúcuta	Edificio	$ 900.000	-	0,017	-
20	2018	DUO Condominio	Metrocol LTDA	Los Patios	Edificio	$ 10.000.000	-	0,217	-

No.	Año	Nombre proyecto	Empresa	Ubicación	Tipo	Costo proyecto ($ COP)	L. Tuberia (m)	Área (Ha)	N. Apartamentos (und)
21	2019	Torres del Norte	Torres del norte SAS	Cúcuta	Conjunto Edificio	$ 8.000.000	1004,01	0,580	196
22	2019	Urb. Santa Eduviges	ODICCO	Toledo	Conjunto Edificio	$ 4.700.000	976,05	1,081	200
23	2018	Altos del Moral	Qbico construcciones	Los Patios	Conjunto Edificio	$ 3.670.000	471,75	0,540	132
24	2021	Chantilly	Paisaje Urbano	Villa del Rosario	Conjunto Edificio	$ 7.700.000	2068,55	1,486	320
25	2020	Bosque de La Variante	Urbanizadora Andalucía SAS	Los Patios	Conjunto Edificio	$ 6.500.000	1220,42	1,026	288
26	2021	El Contento	INGPRA	Cúcuta	Edificio	$ 4.200.000	-	0,018	-
27	2021	PALLADIUM	Paisaje Urbano	Cúcuta	Conjunto Edificio	$ 4.500.000	1240,34	1,296	360
28	2021	Local comercial, El Buque	Ludy Mantilla Velazco	Pamplona	Edificio	$ 1.500.000	-	0,003	-
29	2021	Tulipanes	Paisaje Urbano	Cúcuta	Conjunto Edificio	$ 5.500.000	657,6	0,761	240
30	2021	Entre lomas	Paisaje Urbano	Villa del Rosario	Conjunto Edificio	$ 6.000.000	1970,82	2,021	440
31	2021	Brisas del jardín	Paisaje Urbano	Cúcuta	Conjunto Edificio	$ 5.000.000	657,6	0,761	240
32	2021	Iconik	Paisaje Urbano	Los Patios	Edificio	$ 4.000.000	-	0,140	-
33	2021	Ikaria	Paisaje Urbano	Cúcuta	Conjunto Edificio	$ 4.800.000	709,4	0,847	352
34	2021	Carlos Vásquez	Particular	Cúcuta	Edificio	$ 1.800.000	-	0,019	-
35	2021	Sabana	CABG constructora	Los Patios	Edificio	$ 1.300.000	-	0,007	-
36	2021	Torres de Niza	Constructor San Nicolas	Cúcuta	Edificio	$ 3.200.000	-	0,026	-
37	2020	Torres Bosques del Este	Dar Arquitectura	Cúcuta	Conjunto Edificio	$ 4.000.000	720,8	0,911	240
38	2020	Villas de Turín Barú	Constructora Monape	Los Patios	Conjunto Edificio	$ 4.200.000	1860,24	1,759	340
39	2019	Villas de San Diego	HMS constructores	Los Patios	Conjunto Edificio	$ 10.500.000	3166,49	1,921	300
40	2017	Oporto	Colproyectos	Cúcuta	Edificio	$ 4.500.000	-	0,032	-

Tabla 5. Datos otros proyectos

No.	Año	Nombre proyecto	Empresa	Ubicación	Tipo	Costo proyecto ($ COP)	Área (Ha)
1	2017	Piscina semiolímpica Villa Silvania	Confaoriente	Cúcuta	Piscina	$ 1.200.000	0,208
2	2017	Colegio Francisco Fernández	Ing. Rafael Cáceres	Ocaña	Colegio	$ 1.500.000	0,200
3	2017	Plan parcial Alameda del Este	Paisaje Urbano-Socar-Proinsa-Colproyectos	Cúcuta	Lotes	$ 18.500.000	78,934
4	2017	Acueducto Zona Norte	Acueducto Villa del Rosario	Villa del Rosario	Acueducto	$ 4.170.000	104,160
5	2018	Centro Comercial Jardín Plaza	C3 construcciones y contratos	Cúcuta	Centro comercial	$ 18.000.000	15,443
6	2017	Cancha sintética 20 de Julio	Gobernación Norte de Santander	Cúcuta	Cancha	$ 1.500.000	0,720
7	2017	Cancha sintética Carora	Gobernación Norte de Santander	Cúcuta	Cancha	$ 1.500.000	0,348
8	2017	Optimización red de agua potable	Transmateriales	Cúcuta	Red agua potable	$ 5.500.000	3,460
9	2017	Colector Campoverde	Secretaria de agua potable	Villa del Rosario	Colector	$ 2.000.000	39,880
10	2017	Colectores Alcázares Villa Silvania	Secretaria de agua potable	Villa del Rosario	Colector	$ 1.200.000	105,550
11	2018	Instituto Técnico La Sabana	Ing. Rafael Cáceres	Los Patios	Colegio	$ 1.800.000	0,065
12	2018	Colegio Km8	Ing. Rafael Cáceres	Los Patios	Colegio	$ 1.700.000	0,155
13	2018	Casa Santa Eduviges	Alcaldía de Toledo	Toledo	Colegio	$ 1.500.000	0,095
14	2018	Box Culvert Chinácota	Ing. Zulay Vásquez	Chinácota	Box Culvert	$ 1.500.000	0,059
15	2018	Casa de la cultura Pamplonita	Alcaldía de Pamplonita	Pamplonita	Casa de cultura	$ 1.300.000	0,035
16	2018	Colegio San Bartolomé Comuneros	Alcaldía de San José de Cúcuta	Cúcuta	Colegio	$ 1.700.000	0,425
17	2018	Casino Atlantis Jardín Plaza	Constructora Andina Gómez	Cúcuta	Casino	$ 1.400.000	0,075
18	2019	Edificio Ingeniería Civil UFPS	UFPS	Cúcuta	Universidad	$ 3.500.000	0,141
19	2019	Colector Vista Hermosa	Alcaldía de Villa del Rosario	Villa del Rosario	Colector	$ 2.000.000	29,556
20	2019	Alcantarillado Pluvial Las Américas	Alcaldía de San José de Cúcuta	Cúcuta	Alcantarillado pluvial	$ 3.000.000	0,041

	Año	Nombre proyecto	Empresa	Ubicación	Tipo	Costo proyecto ($ COP)	Área (Ha)
21	2019	Baterías sanitarias UNICEF	UNICEF	Tibú	Centro educativo	$ 3.000.000	0,003
22	2019	Villa Silvania	COMFAORIENTE	Cúcuta	Centro recreacional	$ 1.700.000	0,229
23	2019	Estación de Policía Torcoroma	INGPRA SAS	Cúcuta	Estación de policía	$ 1.600.000	0,110
24	2019	Restaurante MONNO	Ing. Carolina Sánchez	Cúcuta	Restaurante	$ 1.300.000	0,026
25	2019	EDS Los Patios	VIBA CONSTRUCTION SAS	Los Patios	Estación de gasolina	$ 4.500.000	0,096
26	2019	CEMCU	Alcaldía de San José de Cúcuta	Cúcuta	Centro empresarial	$ 4.750.000	1,354
27	2019	LECS Mall	LECS	Cúcuta	Centro comercial	$ 2.000.000	0,254
28	2020	CACI-VR	CMI SAS	Villa del Rosario	Centro atención integral	$ 3.000.000	0,570
29	2020	Cubierta cancha colegio de Cachira	Alcaldía de Cachira	Cáchira	Cubierta cancha	$ 1.400.000	0,058
30	2020	Alcantarillado Chitagá	Alcaldía de Chitagá	Chitagá	Alcantarillado	$ 1.300.000	3,100
31	2020	Puente Simón Bolívar CENAF	CENAF	Villa del Rosario	Puente	$ 2.000.000	2,545
32	2020	Box Culvert Cormoranes	Alcaldía de San José de Cúcuta	Cúcuta	Box Culvert	$ 2.000.000	78,270
33	2020	Plaza de mercado Cucutilla	Alcaldía de Cucutilla	Cucutilla	Plaza de mercado	$ 3.500.000	0,081
34	2019	Baterías sanitarias Nuestra Sra. de Torcoroma	Terre Des Hommes	Tibú	Comedor comunitario	$ 800.000	0,007
35	2019	Baterías sanitarias corregimiento Filo Gringo	Terre Des Hommes	El Tarra	Baterías sanitarias	$ 700.000	0,003
36	2019	Baterías sanitarias corregimiento Orú	Terre Des Hommes	El Tarra	Baterías sanitarias	$ 700.000	0,004
37	2017	Casa de gobierno e internado Tibú	Alcaldía de Tibú	Tibú	Internado	$ 1.200.000	0,042
38	2017	CDI Comuneros	Arq. German Colmenares	Cúcuta	Centro desarrollo infantil	$ 1.100.000	0,500
39	2017	Mega colegio La Frontera	Ing. Rafael Cáceres	Villa del Rosario	Colegio	$ 2.500.000	0,016
40	2020	Cubierta cancha colegio de Tibú	Ing. Carlos Pacheco	Tibú	Cubierta cancha	$ 1.600.000	0,732

La distribución del muestreo según su tipología corresponde a la cantidad de proyectos de diseños hidráulicos recolectados para el análisis de las variables explicativas, como se representa en la (figura 9). Se recolectaron en total 120 proyectos, que se dividen en tres categorías: 40 en residencias unifamiliares (16 viviendas y 24 conjunto de viviendas), 40 en residencias multifamiliares (12 edificios y 28 conjunto de edificios) y 40 en otros proyectos (comerciales e instituciones),

A continuación, se presenta la distribución según su tipología de acuerdo con los datos obtenidos en cada una de las clasificaciones:

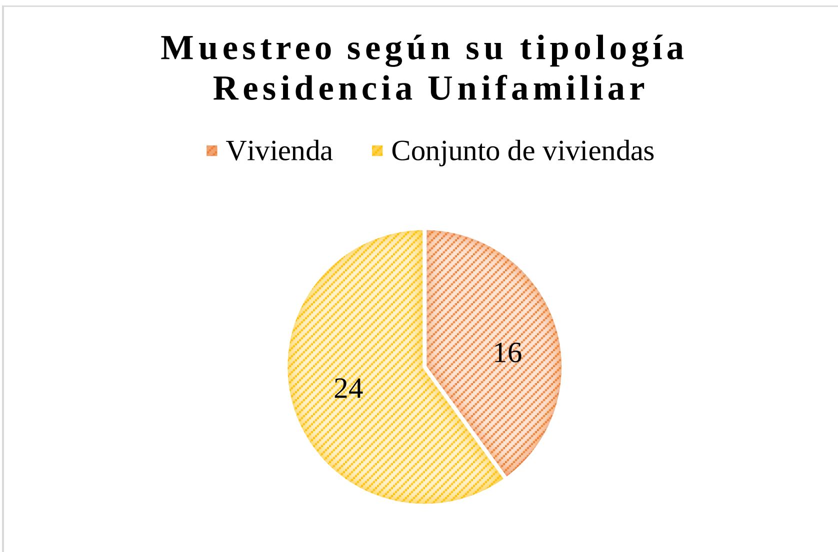


Figura 10. Distribución del muestreo según su tipología en residencia unifamiliar

Figura 11. Distribución del muestreo según su tipología en residencia multifamiliar

Figura 12. Distribución del muestreo según su tipología en otros proyectos

4.2. Programación software R

4.2.1. Residencia unifamiliar

4.2.1.1. Vivienda: Área (Ha)

<u>Modelo lineal</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo  Area
1   800000 0.023
2   700000 0.012
3  1000000 0.095
4   700000 0.019
5  1000000 0.078
6   700000 0.010
7   800000 0.064
8   800000 0.042
9   800000 0.061
10  800000 0.082
11 1000000 0.025
12 1000000 0.050
13 1000000 0.050
14 1200000 0.200
15  800000 0.025
16 1000000 0.054
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

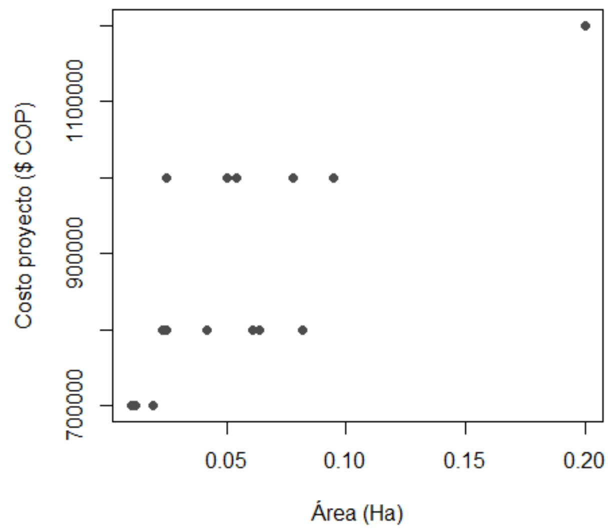


Figura 13. Gráfica de dispersión, vivienda: Área (Ha)

En la (figura 13) se observan datos en los que aumenta el área, pero no el costo del proyecto, es decir, que no son directamente proporcionales; algunos costos son elevados para el área que representan.

```
> abline(lm(Costo~Area),col="red")
```

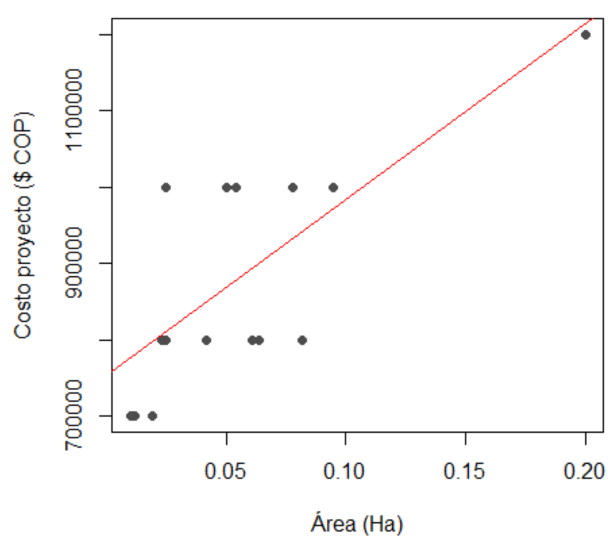


Figura 14. Gráfica de regresión lineal, modelo lineal, vivienda: Área (Ha)

En la (figura 14) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Se construye el modelo de regresión utilizando la función *lm*:

```
> modelo<-lm(Costo~Area)
> summary(modelo)

Call:
lm(formula = Costo ~ Area)

Residuals:
    Min      1Q  Median      3Q     Max 
-141839  -84175  -11903   81133  189102 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   753469      41820  18.017 4.41e-11 ***
Area         2297196     585188   3.926  0.00152 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 105000 on 14 degrees of freedom
Multiple R-squared:  0.524,	Adjusted R-squared:   0.49 
F-statistic: 15.41 on 1 and 14 DF,  p-value: 0.001523
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 49.00% de la variabilidad observada.

El modelo de regresión estimado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$Costo\ (\$\ COP) = 753469 + 2297196 * Área\ (Ha) \qquad \text{(Ecuación 1)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido:

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)   
Area       1 1.6996e+11 1.6996e+11   15.41 0.001523 **
Residuals 14 1.5441e+11 1.1029e+10                    
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales. Básicamente, los supuestos que deben verificarse son los siguientes:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.92483, p-value = 0.2017
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] -2.273737e-12
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

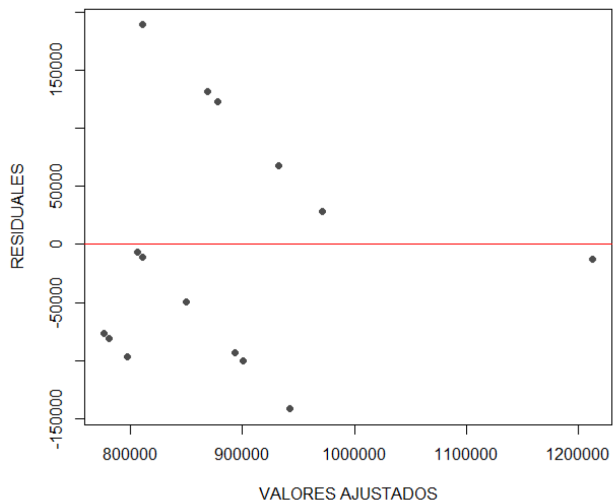


Figura 15. Gráfica de residuales, modelo lineal, vivienda: Área (Ha)

En la (figura 15) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo potencial</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo  Area
1   800000 0.023
2   700000 0.012
3  1000000 0.095
4   700000 0.019
5  1000000 0.078
6   700000 0.010
7   800000 0.064
8   800000 0.042
9   800000 0.061
10  800000 0.082
11 1000000 0.025
12 1000000 0.050
13 1000000 0.050
14 1200000 0.200
15  800000 0.025
16 1000000 0.054
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” y (x) “Área” para el modelo potencial:

```
> LNCosto<-c(log(Costo))
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
> abline(lm(LNCosto~LNArea),col="red")
```

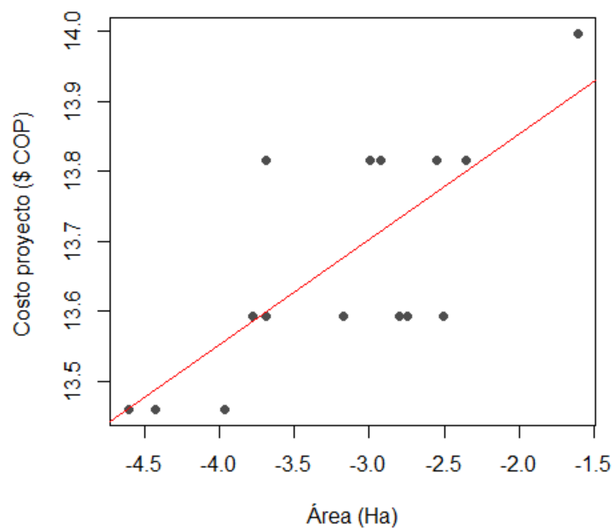


Figura 16. Gráfica de regresión lineal, modelo potencial, vivienda: Área (Ha)

En la (figura 16) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~LNArea))
> summary(modelo)
```

```
Call:
lm(formula = LNCosto ~ LNArea)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.185719 -0.087964  0.002679  0.088853  0.217017 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 14.15622    0.12192 116.114  < 2e-16 ***
LNArea       0.15119    0.03732   4.051  0.00119 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1152 on 14 degrees of freedom
Multiple R-squared:  0.5396,    Adjusted R-squared:  0.5067 
F-statistic: 16.41 on 1 and 14 DF,  p-value: 0.001191
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 50.67% de la variabilidad observada.

El modelo de regresión estimado es:

$y^* = ln\,\alpha + \beta x^*$

$LnCosto(\$\,COP) = \ln(14{,}15622) + 0{,}15119 * \acute{A}rea(Ha)$

El modelo lineal ajustado se puede expresar como la función potencial:

$\hat{y} = \alpha x^{\beta}$

$Costo\,(\$\,COP) = 1406507{,}163 * \acute{A}rea\,(Ha)^{0{,}15119}$ (Ecuación 2)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
          Df  Sum Sq  Mean Sq F value   Pr(>F)   
LNArea     1 0.21789 0.217894  16.409 0.001191 **
Residuals 14 0.18591 0.013279                    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.97165, p-value = 0.8645
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] -5.963112e-19
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

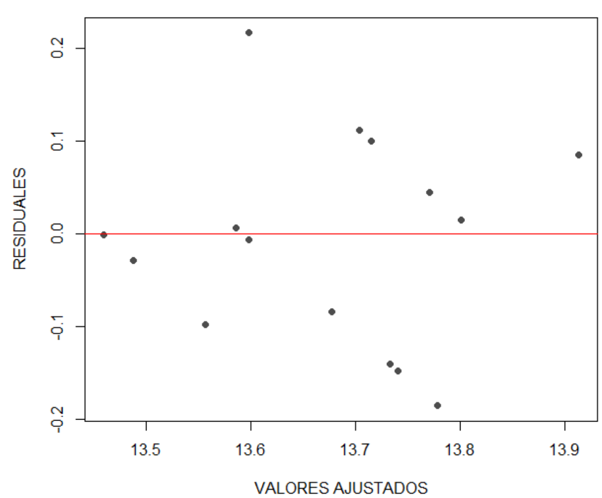


Figura 17. Gráfica de residuales, modelo potencial, vivienda: Área (Ha)

En la (figura 17) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo exponencial

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo  Area
1    800000 0.023
2    700000 0.012
3   1000000 0.095
4    700000 0.019
5   1000000 0.078
6    700000 0.010
7    800000 0.064
8    800000 0.042
9    800000 0.061
10   800000 0.082
11 1000000 0.025
12 1000000 0.050
13 1000000 0.050
14 1200000 0.200
15   800000 0.025
16 1000000 0.054
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” para el modelo exponencial:

```
> LNCosto<-c(log(Costo))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Area,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(LNCosto~Area),col="red")
```

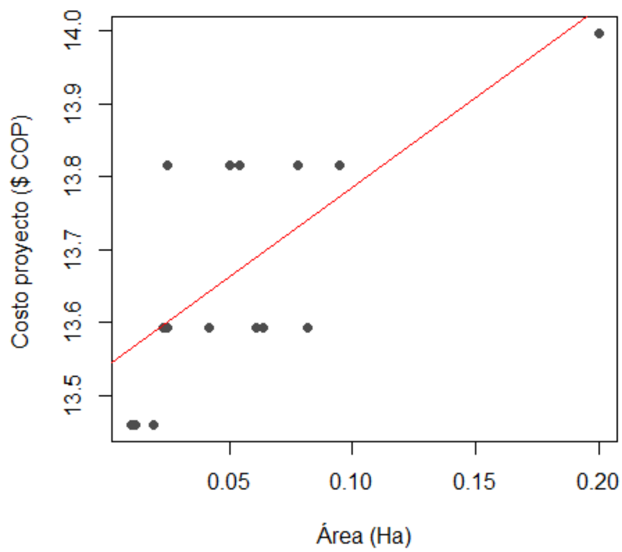


Figura 18. Gráfica de regresión lineal, modelo exponencial, vivienda: Área (Ha)

En la (figura 18) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~Area))
> summary(modelo)

Call:
lm(formula = LNCosto ~ Area)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.14902 -0.10473 -0.02150  0.09878  0.21468 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 13.53919    0.04853 278.968  < 2e-16 ***
Area         2.46586    0.67912   3.631  0.00273 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1219 on 14 degrees of freedom
Multiple R-squared:  0.485,	Adjusted R-squared:  0.4482 
F-statistic: 13.18 on 1 and 14 DF,  p-value: 0.002726
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 44.82% de la variabilidad observada.

El modelo de regresión estimado:

$$y^* = ln\,\alpha + \beta x$$

$$LnCosto\ (\$\ COP) = (13{,}53919) + (2{,}46586 * Área(Ha))$$

El modelo lineal ajustado se puede expresar como la función exponencial:

$$\hat{y} = \alpha e^{\beta x}$$

$$Costo\ (\$\ COP) = 758569{,}7262 * e^{(2{,}46586 * Área\ (Ha))} \quad \text{(Ecuación 3)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
          Df  Sum Sq  Mean Sq F value   Pr(>F)
Area       1 0.19584 0.195837  13.184 0.002726 **
Residuals 14 0.20796 0.014855
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.91024, p-value = 0.1174
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 6.830474e-18
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

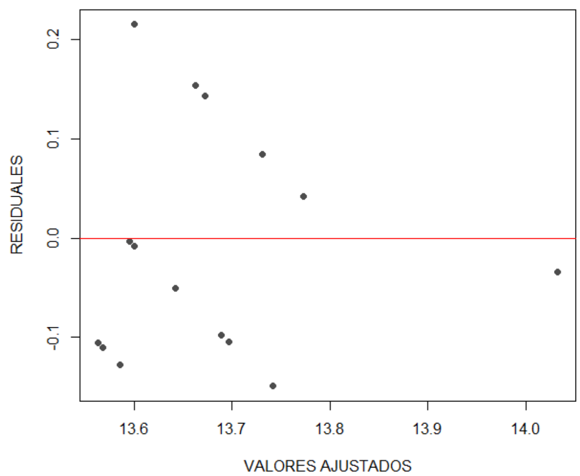


Figura 19. Gráfica de residuales, modelo exponencial, vivienda: Área (Ha)

En la (figura 19) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo logarítmico</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo  Area
1   800000 0.023
2   700000 0.012
3  1000000 0.095
```

```
4    700000 0.019
5   1000000 0.078
6    700000 0.010
7    800000 0.064
8    800000 0.042
9    800000 0.061
10   800000 0.082
11  1000000 0.025
12  1000000 0.050
13  1000000 0.050
14  1200000 0.200
15   800000 0.025
16  1000000 0.054
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (x) "Área" para el modelo logarítmico:

```
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
> abline(lm(Costo~LNArea),col="red")
```

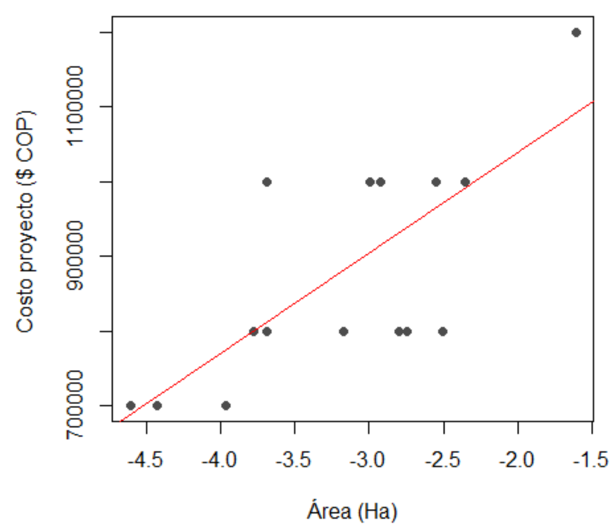


Figura 20. Gráfica de regresión lineal, modelo logarítmico, vivienda: Área (Ha)

En la (figura 20) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~LNArea))
> summary(modelo)

Call:
lm(formula = Costo ~ LNArea)

Residuals:
    Min      1Q  Median      3Q     Max
-171654  -76837    3856   87054  187935

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1307674     110354  11.850  1.1e-08 ***
LNArea        134352      33784   3.977  0.00138 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 104300 on 14 degrees of freedom
Multiple R-squared:  0.5304,    Adjusted R-squared:  0.4969
F-statistic: 15.81 on 1 and 14 DF,  p-value: 0.001377
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 49.69% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = \alpha + \beta x$

$Costo\ (\$\ COP) = 1307674 + (134352 * Área\ (Ha))$

El modelo lineal ajustado se puede expresar como la función logarítmica:

$\hat{y} = \alpha + \beta\ ln\ x$

$Costo\ (\$\ COP) = 1307674 + 134352 * LnÁrea\ (Ha)$ (Ecuación 4)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table
```

```
Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)
LNArea     1 1.7206e+11 1.7206e+11  15.815 0.001377 **
Residuals 14 1.5232e+11 1.0880e+10
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96788, p-value = 0.8032
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] -7.730705e-12
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

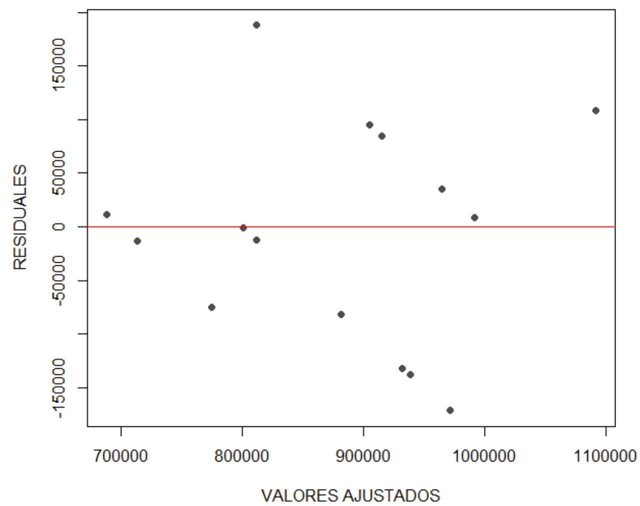


Figura 21. Gráfica de residuales, modelo logarítmico, vivienda unifamiliar: Área (Ha)

En la (figura 21) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo recíproco

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo  Area
1   800000 0.023
2   700000 0.012
3  1000000 0.095
4   700000 0.019
5  1000000 0.078
6   700000 0.010
7   800000 0.064
8   800000 0.042
9   800000 0.061
10  800000 0.082
11 1000000 0.025
12 1000000 0.050
13 1000000 0.050
14 1200000 0.200
15  800000 0.025
16 1000000 0.054
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se transforma la variable, creando un vector con la inversa de (x) "Área" para el modelo reciproco:

```
> Areainv<-c(1/Area)
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Areainv,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(Costo~Areainv),col="red")
```

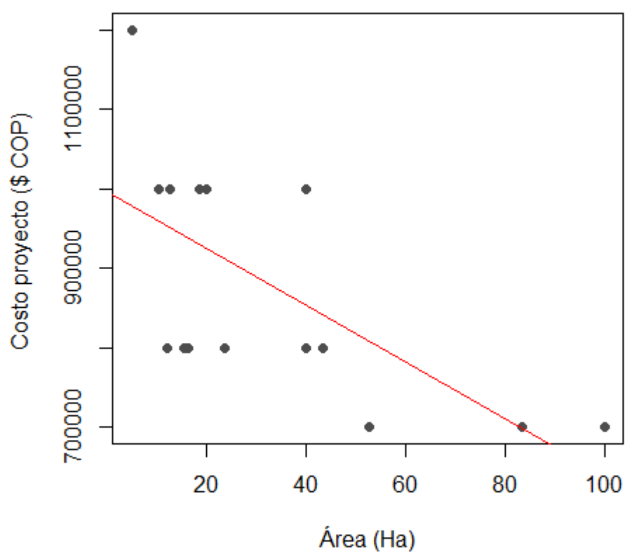


Figura 22. Gráfica de regresión lineal, modelo recíproco, vivienda: Área (Ha)

En la (figura 22) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~Areainv))
> summary(modelo)

Call:
lm(formula = Costo ~ Areainv)

Residuals:
    Min      1Q  Median      3Q     Max 
-152376 -108904   21457   71488  221972 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   995853      45574  21.851 3.23e-12 ***
Areainv        -3565       1100  -3.242  0.00591 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 115000 on 14 degrees of freedom
Multiple R-squared:  0.4288,	Adjusted R-squared:  0.388 
F-statistic: 10.51 on 1 and 14 DF,  p-value: 0.005909
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 38.80% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = \alpha + \beta x$$

$$Costo\ (\$\ COP) = 995853 - (3565 * Área\ (Ha))$$

El modelo lineal ajustado se puede expresar como la función Recíproca:

$$\hat{y} = \alpha + \beta\left(\frac{1}{x}\right)$$

$$Costo\ (\$\ COP) = 995853 - 3565\left(\frac{1}{Área\ (Ha)}\right)$$ (Ecuación 5)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)   
Areainv    1 1.3909e+11 1.3909e+11  10.509 0.005909 **
Residuals 14 1.8529e+11 1.3235e+10                    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.93819, p-value = 0.3275
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 2.50111e-12
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

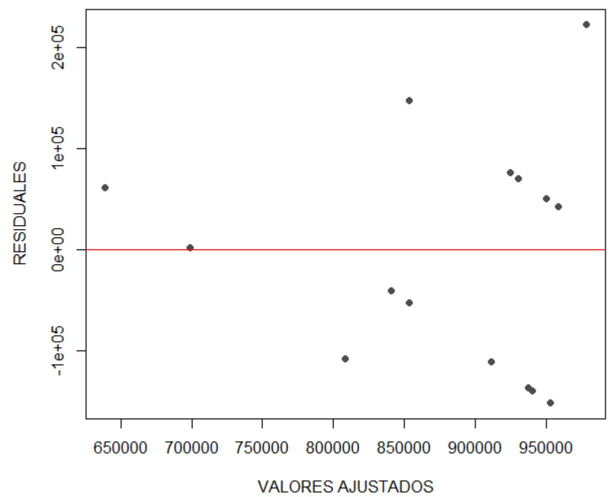


Figura 23. Gráfica de residuales, modelo recíproco, vivienda: Área (Ha)

En la (figura 23) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Polinómico

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo  Area
1   800000 0.023
2   700000 0.012
3  1000000 0.095
4   700000 0.019
5  1000000 0.078
6   700000 0.010
7   800000 0.064
8   800000 0.042
9   800000 0.061
10  800000 0.082
11 1000000 0.025
12 1000000 0.050
13 1000000 0.050
14 1200000 0.200
15  800000 0.025
16 1000000 0.054
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto($
COP)",pch=16,col="grey30")
```

Se construye el modelo de regresión utilizando la función poly (se ajusta polinomio de grado dos):

```
> fit2 <- lm(Costo ~ poly(Area, 2, raw=TRUE))
```

Se revisan los resultados con la instrucción summary:

```
> summary(fit2)

Call:
lm(formula = Costo ~ poly(Area, 2, raw = TRUE))

Residuals:
    Min      1Q  Median      3Q     Max 
-156224  -71592   -1729   68923  195376 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  729925      68039  10.728 7.91e-08 ***
poly(Area, 2, raw = TRUE)1  3088075    1869224   1.652    0.122    
poly(Area, 2, raw = TRUE)2 -4004071    8958163  -0.447    0.662    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 108200 on 13 degrees of freedom
Multiple R-squared:  0.5312,	Adjusted R-squared:  0.4591 
F-statistic: 7.365 on 2 and 13 DF,  p-value: 0.00727
```

Se aprecia que uno de los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 45.91% de la variabilidad observada.

El modelo polinomial de regresión estimado es:

$$\hat{y} = \beta + \beta_1 x - \beta_2 x^2$$

$$Costo\ (\$\ COP) = 729925 + (3088075 * Área\ (Ha)) - (4004071 * Área\ (Ha)^2)$$ *(Ecuación 6)*

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(fit2)
Analysis of Variance Table

Response: Costo
                          Df     Sum Sq    Mean Sq F value  Pr(>F)
poly(Area, 2, raw = TRUE)  2 1.7230e+11 8.6150e+10  7.3645 0.00727 **
Residuals                 13 1.5207e+11 1.1698e+10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(fit2$residuals)

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.95616, p-value = 0.5929
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(fit2$residuals)
[1] 1.875833e-12
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-fit2$residuals
> valores.ajustados<-fit2$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

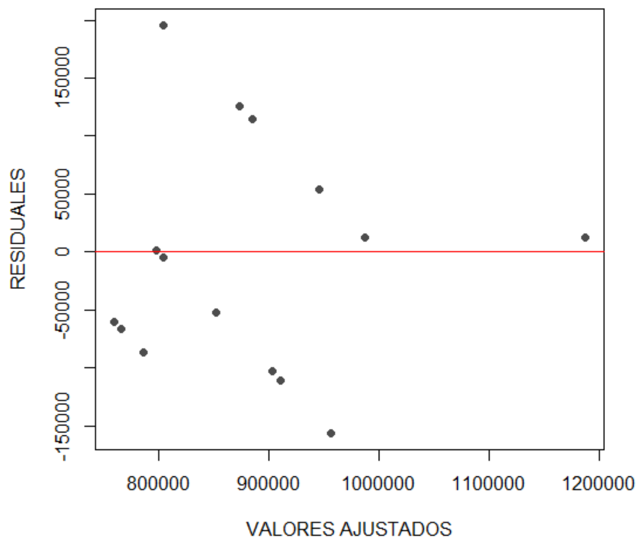


Figura 24. Gráfica de residuales, modelo polinómico, vivienda: Área (Ha)

En la (figura 24) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

De acuerdo con la realización de las ecuaciones metodológicas con los modelos funcionales de regresión lineal mediante el software de programación R, se efectúa una tabla resumen donde se encuentran cada uno de los resultados obtenidos en los diferentes modelos además de tener en cuenta los coeficientes de R^2, p-value y el error estándar de regresión en la clasificación viviendas unifamiliar con la variable área (Ha).

Tabla 6. Resultados, vivienda unifamiliar: Área (Ha)

Modelo funcional	Ecuación. Regresión Lineal	R-squared	p-value	Error estándar $\hat{\sigma}$
Lineal	(1) $Costo\ (\$\ COP) = 753469 + 2297196 * Área(Ha)$	0,49	0,001523	1,1029e+10
Potencial	(2) $Costo\ (\$\ COP) = 1406507{,}163 * Área(Ha)^{0{,}15119}$	0,5067	0,001191	0,013279
Exponencial	(3) $Costo\ (\$\ COP) = 758569{,}7262 * e^{(2{,}46586 * Área\ (Ha))}$	0,4482	0,002726	0,014855
Logarítmica	(4) $Costo\ (\$\ COP) = 1307674 + 134352 * ln\, Área\ (Ha)$	0,4969	0,001377	1,0880e+10
Recíproca	(5)$Costo\ (\$\ COP) = 995853 - 3565 \left(\frac{1}{Área\ (Ha)}\right)$	0,388	0,005909	1,3235e+10
Polinómica	(6) $Costo\ (\$\ COP) = 729925 + (3088075 * Área\ (Ha)) - (4004071 * Área\ (Ha)^2)$	0,4591	0,00727	1,1698e+10

Una vez construidos los modelos de regresión lineal para la variable explicativa “Área (Ha)” en la clasificación “vivienda”, se pueden comparar los resultados obtenidos para elegir la ecuación que presenta mayor significancia estadística; en este caso se encuentra que el modelo funcional potencial muestra un coeficiente de determinación R^2 de 0,5067, valor superior a los obtenidos en los otros modelos empleados y el que más se acerca a uno (1); también se observa en el mismo modelo que el p- value es de 0,001191 el cual presenta un valor por debajo del 5%, indicando su significancia estadística.

Los resultados también demuestran que el modelo funcional que menos se ajusta es el recíproco, dado que el coeficiente de determinación es el más bajo en comparación a los obtenidos en los otros modelos analizados y el que más se aleja de uno (1).

Teniendo en cuenta estos resultados analizados, la *(ecuación 2)* del modelo funcional potencial es la que representa mayor alcance y menor error estándar de la regresión para la variable dependiente *Costo ($ COP)*, por lo cual será empleada para esta clasificación.

4.2.1.2. Conjunto de viviendas: Área (Ha)

Modelo lineal

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo  Area
1  5000000 2.067
2  4500000 1.853
3  4000000 1.280
4  4500000 1.515
5  3000000 1.580
6  3000000 1.564
7  3200000 3.715
8  3500000 2.159
9  5800000 1.254
10 4000000 1.866
11 3700000 2.600
12 4200000 3.067
13 5500000 2.259
14 3500000 1.149
15 2500000 0.178
16 2200000 0.261
17 4000000 1.469
18 4500000 2.249
19 5500000 1.911
20 1500000 0.225
21 1500000 0.145
22 2500000 0.121
23 2500000 1.348
24 2500000 0.456
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
      > plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

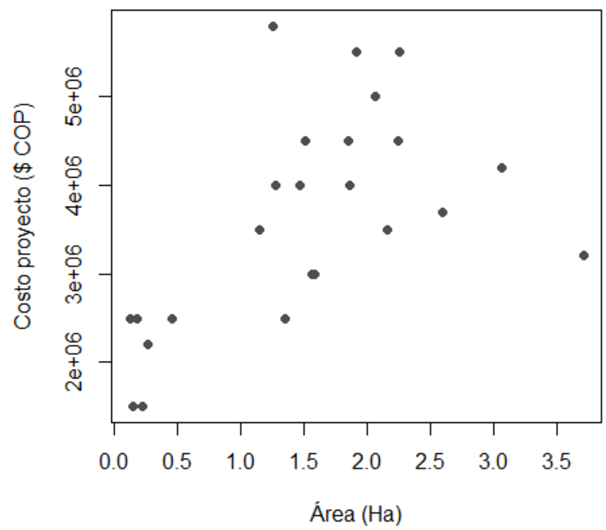


Figura 25. Gráfica de dispersión, conjunto de viviendas: Área (Ha)

En la (figura 25) se observan datos en los que aumenta el área, pero no el costo del proyecto, y también viceversa, es decir, que no son directamente proporcionales; algunos costos son elevados para el área que representan.

```
> abline(lm(Costo~Area),col="red")
```

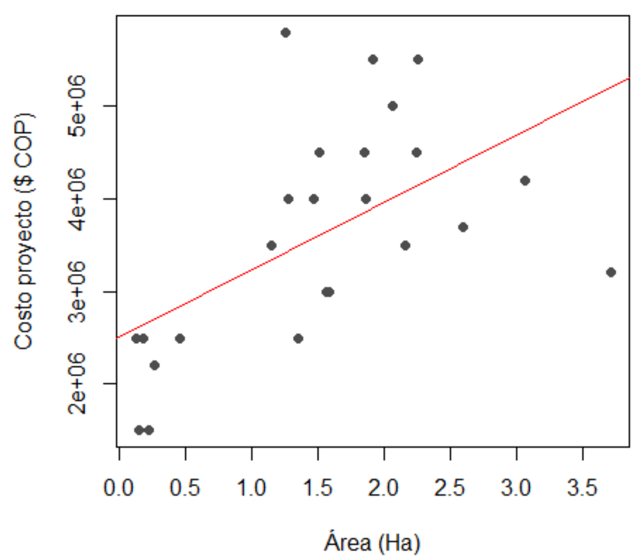


Figura 26. Gráfica de regresión lineal, modelo lineal, conjunto de viviendas: Área (Ha)

En la (figura 25) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Se construye el modelo de regresión utilizando la función *lm*:

```
> modelo<-lm(Costo~Area)
> summary(modelo)

Call:
lm(formula = Costo ~ Area)

Residuals:
     Min       1Q   Median       3Q      Max
-2002198  -648761  -122422   580971  2378430

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2514253     396446   6.342 2.21e-06 ***
Area          723539     223124   3.243  0.00374 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1020000 on 22 degrees of freedom
Multiple R-squared:  0.3234,    Adjusted R-squared:  0.2926
F-statistic: 10.52 on 1 and 22 DF,  p-value: 0.003735
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 29.26% de la variabilidad observada.

El modelo de regresión estimado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$Costo\ (\$\ COP) = 2514253 + 723539 * Área\ (Ha) \qquad \text{(Ecuación 7)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido:

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)
Area       1 1.0937e+13 1.0937e+13  10.515 0.003735 **
Residuals 22 2.2881e+13 1.0401e+12
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.98074, p-value = 0.909
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 3.757705e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

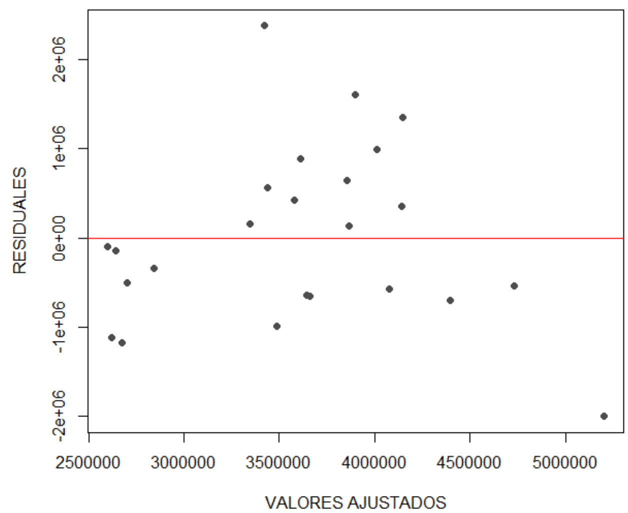


Figura 27. Gráfica de residuales, modelo lineal, conjunto de viviendas: Área (Ha)

En la (figura 27) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo potencial</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo   Area
1  5000000 2.067
2  4500000 1.853
3  4000000 1.280
4  4500000 1.515
5  3000000 1.580
6  3000000 1.564
7  3200000 3.715
8  3500000 2.159
9  5800000 1.254
10 4000000 1.866
11 3700000 2.600
12 4200000 3.067
13 5500000 2.259
14 3500000 1.149
15 2500000 0.178
16 2200000 0.261
17 4000000 1.469
18 4500000 2.249
19 5500000 1.911
```

```
20 1500000 0.225
21 1500000 0.145
22 2500000 0.121
23 2500000 1.348
24 2500000 0.456
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) "Costo" y (x) "Área" para el modelo potencial:

```
> LNCosto<-c(log(Costo))
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(LNCosto~LNArea),col="red")
```

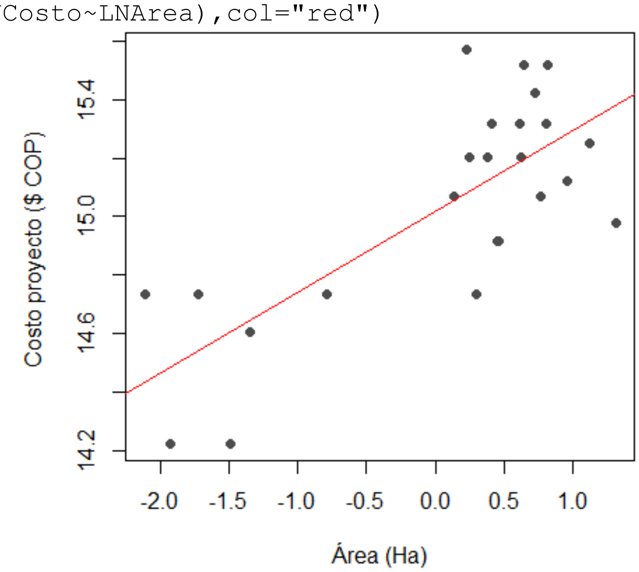


Figura 28. Gráfica de regresión lineal, modelo potencial, conjunto de viviendas: Área (Ha)

En la (figura 28) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~LNArea))
> summary(modelo)

Call:
lm(formula = LNCosto ~ LNArea)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.40327 -0.17958  0.01093  0.18734  0.49218 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 15.01851    0.05074  296.00  < 2e-16 ***
LNArea       0.27692    0.05044    5.49 1.62e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.248 on 22 degrees of freedom
Multiple R-squared:  0.5781,	Adjusted R-squared:  0.5589 
F-statistic: 30.14 on 1 and 22 DF,  p-value: 1.622e-05
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 55.89% de la variabilidad observada.

El modelo de regresión estimado es:

$y^* = ln\,\alpha + \beta x^*$

$LnCosto\ (\$\ COP) = 15{,}01851 + 0{,}27692 * Área(Ha)$

El modelo lineal ajustado se puede expresar como la función potencial:

$\hat{y} = \alpha x^{\beta}$

$Costo\ (\$\ COP) = 3330090 * Área\ (Ha)^{0{,}27692}$ (Ecuación 8)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
          Df Sum Sq Mean Sq F value    Pr(>F)    
LNArea     1 1.8539  1.8539  30.142 1.622e-05 ***
Residuals 22 1.3531  0.0615                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.97531, p-value = 0.7964
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 7.224909e-18
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

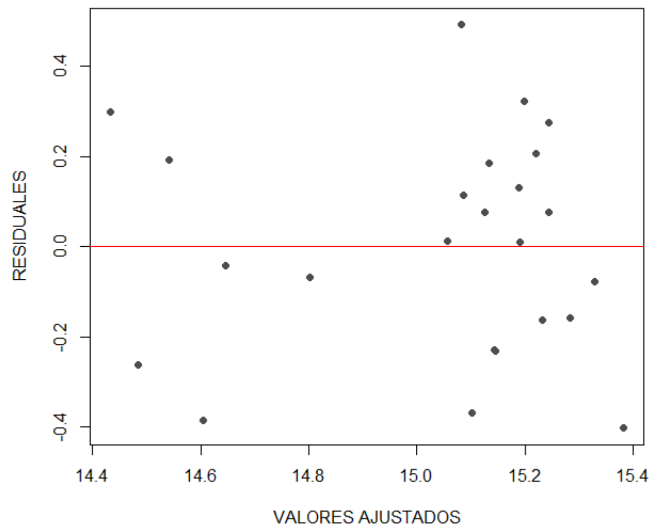


Figura 29. Gráfica de residuales, modelo potencial, conjunto de viviendas: Área (Ha)

En la (figura 29) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo exponencial

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo  Area
1  5000000 2.067
2  4500000 1.853
3  4000000 1.280
4  4500000 1.515
5  3000000 1.580
6  3000000 1.564
7  3200000 3.715
8  3500000 2.159
9  5800000 1.254
10 4000000 1.866
11 3700000 2.600
12 4200000 3.067
13 5500000 2.259
14 3500000 1.149
15 2500000 0.178
16 2200000 0.261
17 4000000 1.469
18 4500000 2.249
19 5500000 1.911
```

```
20 1500000 0.225
21 1500000 0.145
22 2500000 0.121
23 2500000 1.348
24 2500000 0.456
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” para el modelo exponencial:

```
> LNCosto<-c(log(Costo))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Area,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(LNCosto~Area),col="red")
```

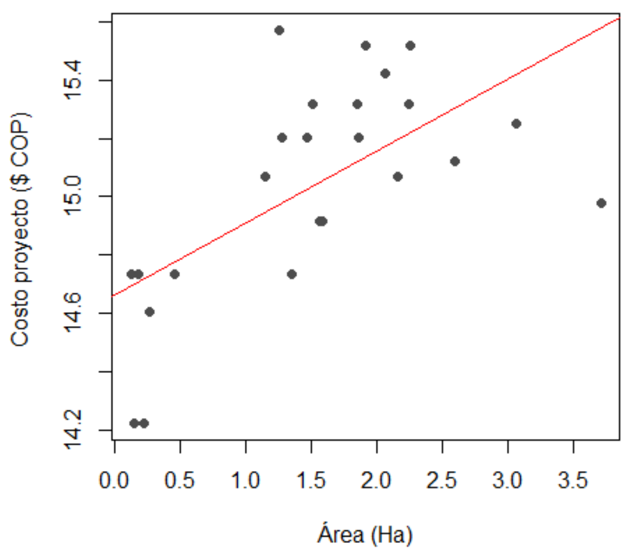


Figura 30. Gráfica de regresión lineal, modelo exponencial, conjunto de viviendas: Área (Ha)

En la (figura 30) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~Area))
> summary(modelo)

Call:
lm(formula = LNCosto ~ Area)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.60194 -0.14745  0.03065  0.20415  0.59978 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  14.6643     0.1153   127.2  < 2e-16 ***
Area          0.2467     0.0649     3.8  0.00098 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2966 on 22 degrees of freedom
Multiple R-squared:  0.3963,    Adjusted R-squared:  0.3689 
F-statistic: 14.44 on 1 and 22 DF,  p-value: 0.00098
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 36.89% de la variabilidad observada.

El modelo de regresión estimado:

$$y^{*} = ln\,\alpha + \beta x$$

$$LnCosto\ (\$\ COP) = (14{,}6643) + (0{,}2467 * Área\ (Ha))$$

El modelo lineal ajustado se puede expresar como la función exponencial:

$$\hat{y} = \alpha e^{\beta x}$$

$$Costo\ (\$\ COP) = 2336816{,}287 * e^{(0{,}2467 * Área\ (Ha))}$$ (Ecuación 9)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
          Df Sum Sq Mean Sq F value  Pr(>F)    
Area       1  1.271   1.271  14.443 0.00098 ***
Residuals 22  1.936   0.088                    
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.97738, p-value = 0.8432
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 1.214306e-17
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

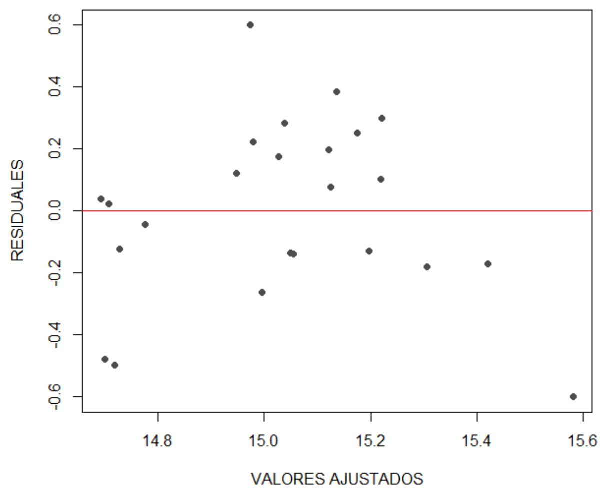


Figura 31. Gráfica de residuales, modelo exponencial, conjunto de viviendas: Área (Ha)

En la (figura 31) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo logarítmico</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo   Area
1  5000000 2.067
2  4500000 1.853
3  4000000 1.280
4  4500000 1.515
5  3000000 1.580
6  3000000 1.564
7  3200000 3.715
8  3500000 2.159
9  5800000 1.254
10 4000000 1.866
11 3700000 2.600
12 4200000 3.067
13 5500000 2.259
14 3500000 1.149
15 2500000 0.178
16 2200000 0.261
17 4000000 1.469
18 4500000 2.249
19 5500000 1.911
20 1500000 0.225
```

```
21 1500000 0.145
22 2500000 0.121
23 2500000 1.348
24 2500000 0.456
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (x) “Área” para el modelo logarítmico:

```
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(Costo~LNArea),col="red")
```

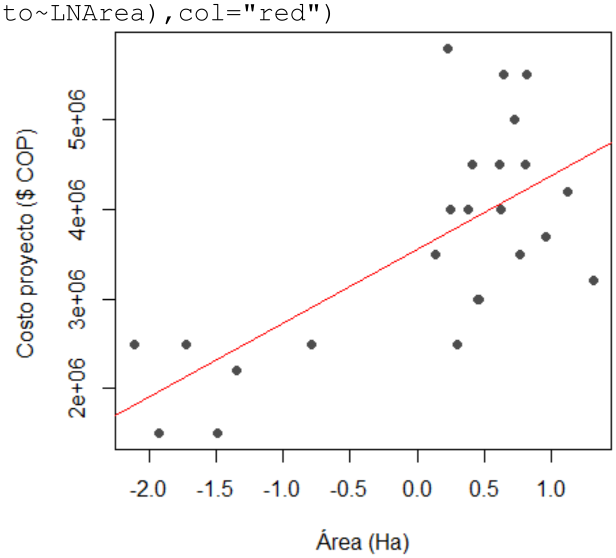


Figura 32. Gráfica de regresión lineal, modelo logarítmico, conjunto de viviendas: Área (Ha)

En la (figura 32) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~LNArea))
> summary(modelo)

Call:
lm(formula = Costo ~ LNArea)

Residuals:
     Min       1Q   Median       3Q      Max 
-1431613  -650029  -116202   481614  2061230 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3552696     182359  19.482  2.3e-15 ***
LNArea        822109     181284   4.535 0.000163 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 891300 on 22 degrees of freedom
Multiple R-squared:  0.4831,	Adjusted R-squared:  0.4597 
F-statistic: 20.57 on 1 and 22 DF,  p-value: 0.0001633
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 45.97% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = \alpha + \beta x$

$Costo\ (\$\ COP) = 3552696 + (822109 * Área\ (Ha))$

El modelo lineal ajustado se puede expresar como la función logarítmica:

$\hat{y} = \alpha + \beta\ ln\, x$

$Costo\ (\$\ COP) = 3552696 + (822109 * LnÁrea\ (Ha))$ (Ecuación 10)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value    Pr(>F)    
LNArea     1 1.6339e+13 1.6339e+13  20.566 0.0001633 ***
Residuals 22 1.7479e+13 7.9450e+11                      
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.97737, p-value = 0.843
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] -1.333807e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

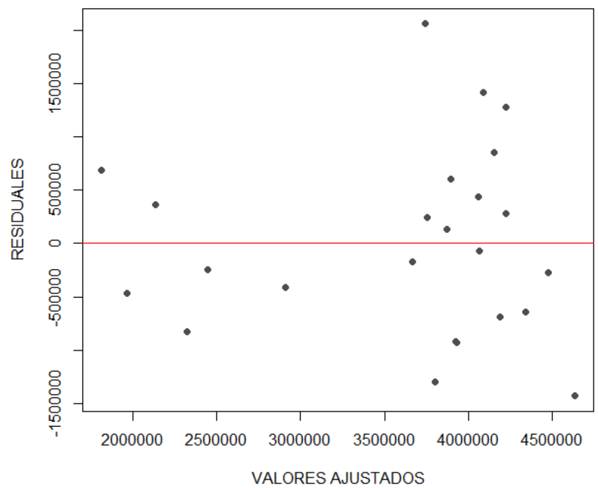


Figura 33. Gráfica de residuales, modelo logarítmico, conjunto de viviendas: Área (Ha)

En la (figura 33) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo recíproco</u>

Se importan los datos que se encuentran en el archivo Excel :

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo  Area
1  5000000 2.067
2  4500000 1.853
3  4000000 1.280
4  4500000 1.515
5  3000000 1.580
6  3000000 1.564
7  3200000 3.715
8  3500000 2.159
9  5800000 1.254
10 4000000 1.866
11 3700000 2.600
12 4200000 3.067
13 5500000 2.259
14 3500000 1.149
15 2500000 0.178
16 2200000 0.261
17 4000000 1.469
18 4500000 2.249
19 5500000 1.911
20 1500000 0.225
21 1500000 0.145
```

```
22 2500000 0.121
23 2500000 1.348
24 2500000 0.456
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se transforma la variable, creando un vector con la inversa de (x) "Área" para el modelo recíproco:

```
> Areainv<-c(1/Area)
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Areainv,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(Costo~Areainv),col="red")
```

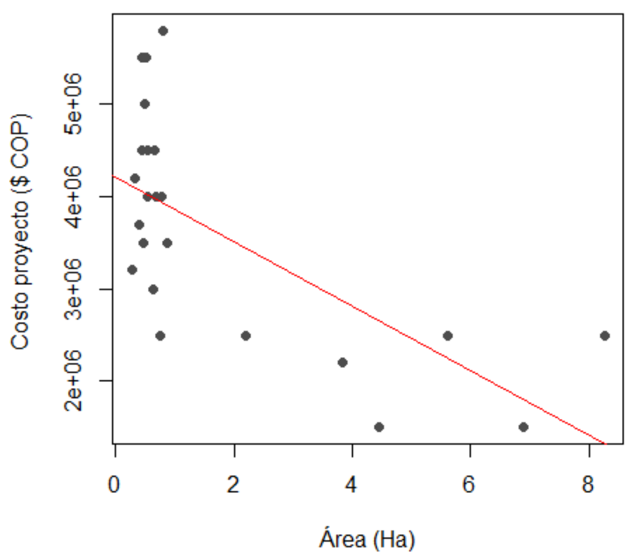


Figura 34. Gráfica de regresión lineal, modelo recíproco, conjunto de viviendas: Área (Ha)

En la (figura 34) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~Areainv))
> summary(modelo)

Call:
lm(formula = Costo ~ Areainv)

Residuals:
     Min       1Q   Median       3Q      Max
-1451673  -735834     1827   488400  1867694

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4210029     239896  17.549 2.01e-14 ***
Areainv      -348264      84607  -4.116 0.000454 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 931900 on 22 degrees of freedom
Multiple R-squared:  0.4351,    Adjusted R-squared:  0.4094
F-statistic: 16.94 on 1 and 22 DF,  p-value: 0.0004542
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 40.94% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = \alpha + \beta x$$

$$Costo\ (\$\ COP) = 4210029 - (348264 * Área\ (Ha))$$

El modelo lineal ajustado se puede expresar como la función Recíproca:

$$\hat{y} = \alpha + \beta\left(\frac{1}{x}\right)$$

$$Costo\ (\$\ COP) = 4210029 - 348264\left(\frac{1}{Área\ (Ha)}\right) \qquad \text{(Ecuación 11)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value    Pr(>F)
Areainv    1 1.4714e+13 1.4714e+13  16.944 0.0004542 ***
Residuals 22 1.9105e+13 8.6839e+11
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)
        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96092, p-value = 0.4572
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 4.06312e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

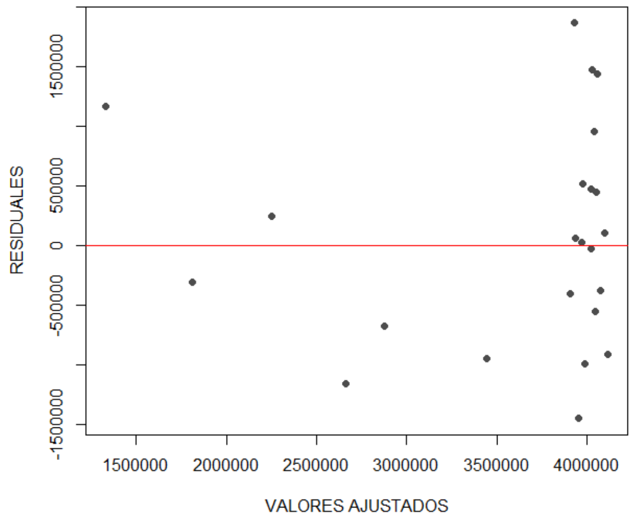


Figura 35. Gráfica de residuales, modelo recíproco, conjunto de viviendas: Área (Ha)

En la (figura 35) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo Polinómico</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo  Area
1  5000000 2.067
2  4500000 1.853
3  4000000 1.280
4  4500000 1.515
5  3000000 1.580
6  3000000 1.564
7  3200000 3.715
8  3500000 2.159
9  5800000 1.254
10 4000000 1.866
11 3700000 2.600
12 4200000 3.067
13 5500000 2.259
14 3500000 1.149
15 2500000 0.178
16 2200000 0.261
17 4000000 1.469
18 4500000 2.249
19 5500000 1.911
```

```
20 1500000 0.225
21 1500000 0.145
22 2500000 0.121
23 2500000 1.348
24 2500000 0.456
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se construye el modelo de regresión utilizando la función poly (se ajusta polinomio de grado dos):

```
> fit2 <- lm(Costo ~ poly(Area, 2, raw=TRUE))
```

Se revisan los resultados con la instrucción summary:

```
> summary(fit2)

Call:
lm(formula = Costo ~ poly(Area, 2, raw = TRUE))

Residuals:
     Min       1Q   Median       3Q      Max 
-1426968  -478549   -17312   437569  1972578 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 1587543     412954   3.844 0.000942 ***
poly(Area, 2, raw = TRUE)1  2462671     520940   4.727 0.000114 ***
poly(Area, 2, raw = TRUE)2  -539460     151595  -3.559 0.001857 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 824400 on 21 degrees of freedom
Multiple R-squared:  0.5779,    Adjusted R-squared:  0.5377 
F-statistic: 14.38 on 2 and 21 DF,  p-value: 0.0001166
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 53.77% de la variabilidad observada.

El modelo polinomial de regresión estimado es:

$\hat{y} = \beta + \beta_1 x - \beta_2 x^2$

$Costo\ (\$\ COP) = 1587543 + (2462671 * Área\ (Ha)) - (539460 * Área\ (Ha)^2)$ (Ecuación 12)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(fit2)
Analysis of Variance Table

Response: Costo
                          Df     Sum Sq    Mean Sq F value    Pr(>F)    
poly(Area, 2, raw = TRUE)  2 1.9544e+13 9.7722e+12  14.377 0.0001166 ***
Residuals                 21 1.4274e+13 6.7971e+11                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(fit2$residuals)

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.9767, p-value = 0.8282
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(fit2$residuals)
[1] 1.758504e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-fit2$residuals
```

```
> valores.ajustados<-fit2$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

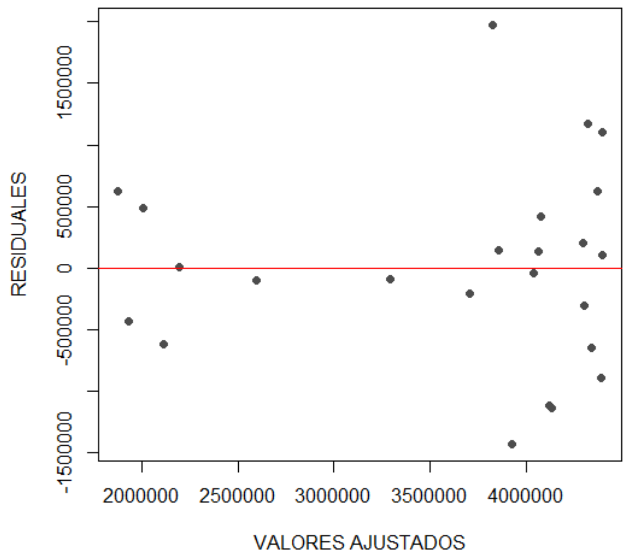


Figura 36. Gráfica de residuales, modelo polinómico, conjunto de viviendas: Área (Ha)

En la (figura 36) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

De acuerdo con la realización de las ecuaciones metodológicas con los modelos funcionales de regresión lineal mediante el software de programación R, se efectúa una tabla resumen donde se encuentran cada uno de los resultados obtenidos en los diferentes modelos además de tener en cuenta los coeficientes de R^2, p-value y el error estándar de regresión en la clasificación conjunto de viviendas unifamiliar con la variable área (Ha).

Tabla 7. Resultados, conjunto de viviendas unifamiliares: Área (Ha)

Modelo funcional	Ecuación. Regresión Lineal	R-squared	p-value	Error estándar $\hat{\sigma}$
Lineal	(7) $Costo\ (\$\ COP) = 2514253 + 723539 * Área(Ha)$	0,2926	0,003735	1,0401e+12
Potencial	(8) $Costo\ (\$\ COP) = 3330090 * Área\ (Ha)^{0,27692}$	0,5589	0,00001622	0,0615
Exponencial	(9) $Costo\ (\$\ COP) = 2336816{,}287 * e^{(0,2467 * Área\ (Ha))}$	0,3689	0,00098	0,088
Logarítmica	(10) $Costo\ (\$\ COP) = 3552696 + 822109 * LnÁrea\ (Ha)$	0,4597	0,0001633	7,9450e+11
Recíproca	(11) $Costo(\$\ COP) = 4210029 - 348264\left(\frac{1}{Área\ (Ha)}\right)$	0,4094	0,0004542	8,6839e+11
Polinómica	(12) $Costo\ (\$\ COP) = 1587543 + (2462671 * Área\ (Ha)) - (539460 * Área\ (Ha)^2)$	0,5377	0,0001166	6,7971e+11

Una vez construidos los modelos de regresión lineal para la variable explicativa "Área (Ha)" en la clasificación "conjunto de viviendas", se pueden comparar los resultados obtenidos para elegir la ecuación que presenta mayor significancia estadística; en este caso se encuentra que el modelo funcional potencial muestra un coeficiente de determinación R^2 de 0,5589, valor superior a los obtenidos en los otros modelos empleados y el que más se acerca a uno (1); también se observa en el mismo modelo que el p- value es de 0,00001622 el cual presenta un valor por debajo del 5%, indicando su significancia estadística.

Los resultados también demuestran que el modelo funcional que menos se ajusta es el lineal, dado que el coeficiente de determinación es el más bajo en comparación a los obtenidos en los otros modelos analizados y el que más se aleja de uno (1).

Teniendo en cuenta estos resultados analizados, la *(ecuación 8)* del modelo funcional potencial es la que representa mayor alcance y menor error estándar de la regresión para la variable dependiente *Costo ($ COP)*, por lo cual será empleada para esta clasificación.

4.2.1.3. Conjunto de viviendas: Longitud de tubería (m)

Modelo lineal

Se importan los datos que se encuentan en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo Longitud
1  5000000  1520.80
2  4500000  1162.41
3  4000000   859.58
4  4500000  2173.00
5  3000000  1754.89
6  3000000  1296.30
7  3200000  2962.10
8  3500000  1788.33
9  5800000  1628.10
10 4000000  2042.00
11 3700000   187.40
12 4200000  1726.58
13 5500000  1779.00
14 3500000  2030.69
15 2500000   234.03
16 2200000   493.17
17 4000000  2547.08
18 4500000  1897.55
19 5500000  1334.38
20 1500000   308.14
21 1500000   160.84
22 2500000    67.98
23 2500000  1052.38
24 2500000   467.19
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
      > plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

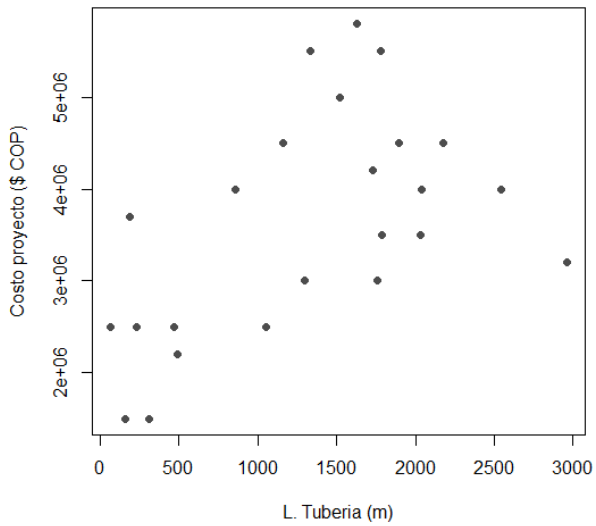


Figura 37. Gráfica de dispersión, modelo lineal, conjunto de viviendas: Longitud de tubería (m)

En la (figura 37) se observan datos en los que aumenta la longitud de tubería, pero no el costo del proyecto, y también viceversa, es decir, que no son directamente proporcionales; algunos costos son elevados para el área que representan.

```
> abline(lm(Costo~Longitud),col="red")
```

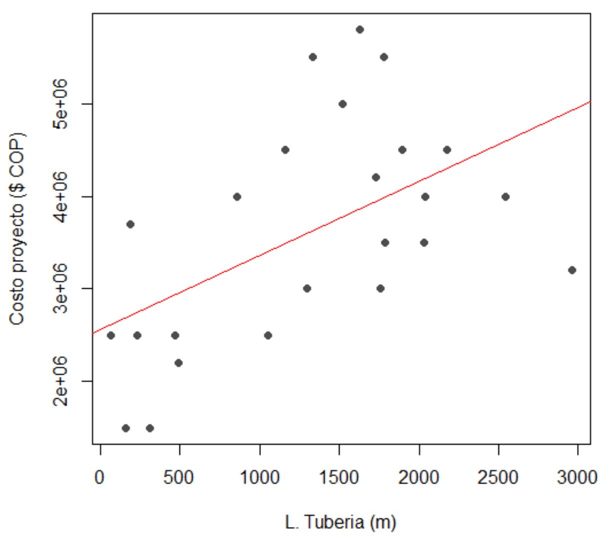


Figura 38. Gráfica de regresión lineal, modelo lineal, conjunto de viviendas: Longitud de tubería (m)

En la (figura 38) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Se construye el modelo de regresión utilizando la función lm:

```
> modelo<-lm(Costo~Longitud)
> summary(modelo)

Call:
lm(formula = Costo ~ Longitud)

Residuals:
     Min       1Q   Median       3Q      Max
-1725802  -700626  -219938   811413  1938908

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2561650.0   410208.1   6.245 2.76e-06 ***
Longitud        798.1      267.1   2.988  0.00678 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1046000 on 22 degrees of freedom
Multiple R-squared:  0.2887,    Adjusted R-squared:  0.2563
F-statistic: 8.928 on 1 and 22 DF,  p-value: 0.006781
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 25.63% de la variabilidad observada.

El modelo de regresión estimado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$Costo\ (\$\ COP) = 2561650{,}0 + 798{,}1 * Longitud\ (m) \qquad \textit{(Ecuación 13)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido:

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)
Longitud   1 9.7623e+12 9.7623e+12  8.9279 0.006781 **
Residuals 22 2.4056e+13 1.0935e+12
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.9569, p-value = 0.3793
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 6.671878e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted.

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

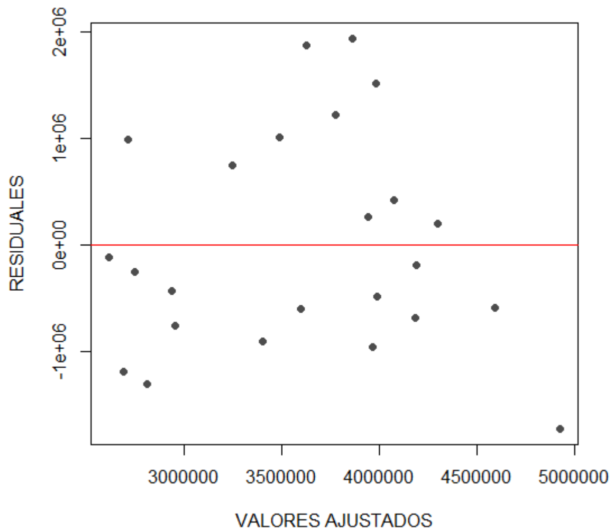


Figura 39. Gráfica de residuales, modelo lineal, conjunto de viviendas: Longitud de tubería (m)

En la (figura 39) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo potencial

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo Longitud
1  5000000  1520.80
2  4500000  1162.41
3  4000000   859.58
4  4500000  2173.00
5  3000000  1754.89
6  3000000  1296.30
7  3200000  2962.10
8  3500000  1788.33
9  5800000  1628.10
10 4000000  2042.00
11 3700000   187.40
12 4200000  1726.58
13 5500000  1779.00
14 3500000  2030.69
15 2500000   234.03
16 2200000   493.17
17 4000000  2547.08
```

```
18 4500000  1897.55
19 5500000  1334.38
20 1500000   308.14
21 1500000   160.84
22 2500000    67.98
23 2500000  1052.38
24 2500000   467.19
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) "Costo" y (x) "Longitud" para el modelo potencial:

```
> LNCosto<-c(log(Costo))
> LNLongitud<-c(log(Longitud))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNLongitud,LNCosto,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
> abline(lm(LNCosto~LNLongitud),col="red")
```

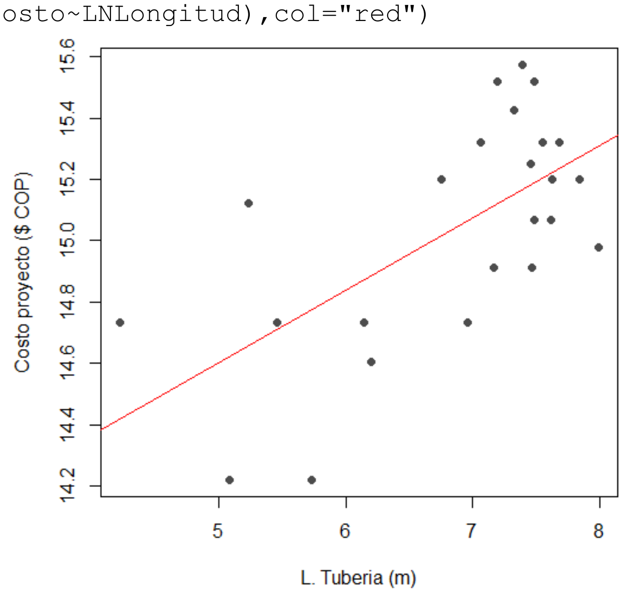


Figura 40. Gráfica de regresión lineal, modelo potencial, conjunto de viviendas: Longitud de tubería (m)

En la (figura 40) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(LNCosto~LNLongitud)
> summary(modelo)

Call:
lm(formula = LNCosto ~ LNLongitud)

Residuals:
     Min       1Q   Median       3Q      Max
-0.55477 -0.21853  0.00038  0.24101  0.46545

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.42342    0.41421  32.408  < 2e-16 ***
LNLongitud   0.23598    0.05994   3.937 0.000703 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2924 on 22 degrees of freedom
Multiple R-squared:  0.4134,    Adjusted R-squared:  0.3867
F-statistic:  15.5 on 1 and 22 DF,  p-value: 0.0007027
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 38.67% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = ln\,\alpha + \beta x^*$$

$$LnCosto\ (\$\ COP) = \ln\ (13{,}42342) + 0{,}23598 * Longitud(m)$$

El modelo lineal ajustado se puede expresar como la función potencial:

$$\hat{y} = \alpha x^{\beta}$$

$$Costo\ (\$\ COP) = 675642{,}9263 * Longitud\ (m)^{0{,}23598} \qquad \text{(Ecuación 14)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
```

```
            Df Sum Sq Mean Sq F value    Pr(>F)    
LNLongitud  1 1.3256 1.32565  15.502 0.0007027 ***
Residuals  22 1.8813 0.08551                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)
      Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96915, p-value = 0.6459
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean (modelo$residuals)
[1] 6.651467e-18
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

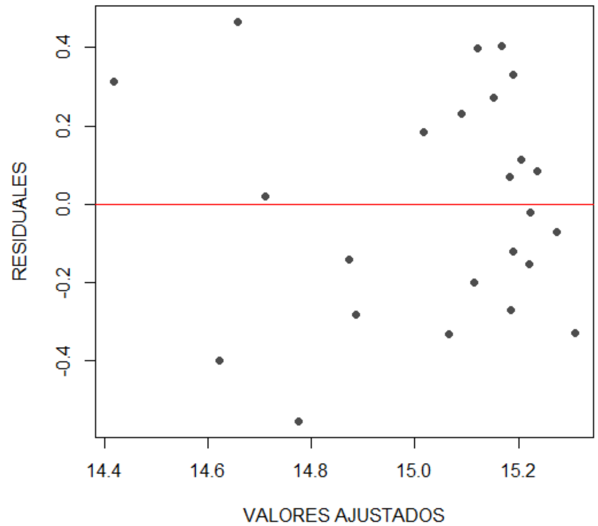


Figura 41. Gráfica de residuales, modelo potencial, conjunto de viviendas: Longitud de tubería (m)

En la (figura 41) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo exponencial</u>

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo Longitud
1  5000000  1520.80
2  4500000  1162.41
3  4000000   859.58
4  4500000  2173.00
5  3000000  1754.89
6  3000000  1296.30
7  3200000  2962.10
8  3500000  1788.33
9  5800000  1628.10
10 4000000  2042.00
11 3700000   187.40
12 4200000  1726.58
13 5500000  1779.00
14 3500000  2030.69
15 2500000   234.03
16 2200000   493.17
17 4000000  2547.08
```

```
18 4500000  1897.55
19 5500000  1334.38
20 1500000   308.14
21 1500000   160.84
22 2500000    67.98
23 2500000  1052.38
24 2500000   467.19
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” para el modelo exponencial:

```
> LNCosto<-c(log(Costo))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Longitud,LNCosto,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
> abline(lm(LNCosto~Longitud),col="red")
```

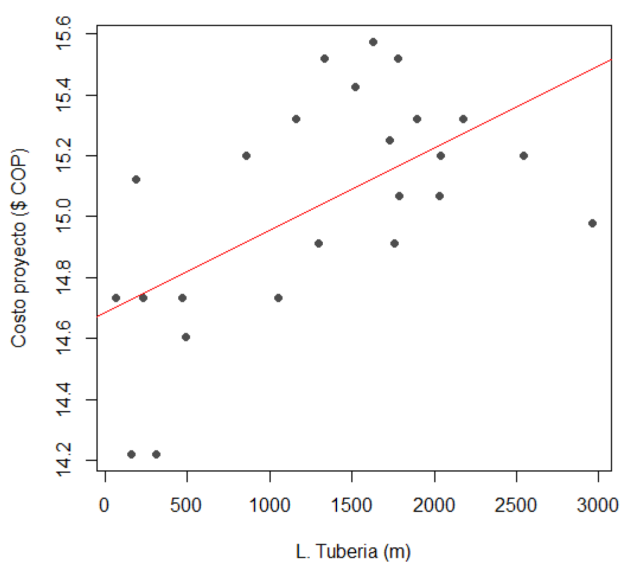


Figura 42. Gráfica de regresión lineal, modelo exponencial, conjunto de viviendas: Longitud de tubería (m)

En la (figura 42) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(LNCosto~Longitud)
> summary(modelo)

Call:
lm(formula = LNCosto ~ Longitud)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.54516 -0.18006 -0.02359  0.29564  0.47680 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.468e+01  1.208e-01 121.497  < 2e-16 ***
Longitud    2.702e-04  7.869e-05   3.434  0.00237 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3081 on 22 degrees of freedom
Multiple R-squared:  0.349,	Adjusted R-squared:  0.3194 
F-statistic: 11.79 on 1 and 22 DF,  p-value: 0.002371
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 31.94% de la variabilidad observada.

El modelo de regresión estimado:

$$y^{*} = ln\,\alpha + \beta x$$

$$LnCosto\ (\$\ COP) = (1{,}468 * 10^{01}) + \left((2{,}702 * 10^{-04}) * Longitud\ (m)\right)$$

El modelo lineal ajustado se puede expresar como la función exponencial:

$$\hat{y} = \alpha e^{\beta x}$$

$$Costo\ (\$\ COP) = 2373793{,}817 * e^{(2{,}702 * 10^{-4} * Longitud\ (m))} \qquad \text{(Ecuación 15)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table
```

```
Response: LNCosto
          Df Sum Sq Mean Sq F value   Pr(>F)
Longitud   1 1.1191  1.1191  11.792 0.002371 **
Residuals 22 2.0879  0.0949
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.9526, p-value = 0.3082
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 2.832958e-17
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

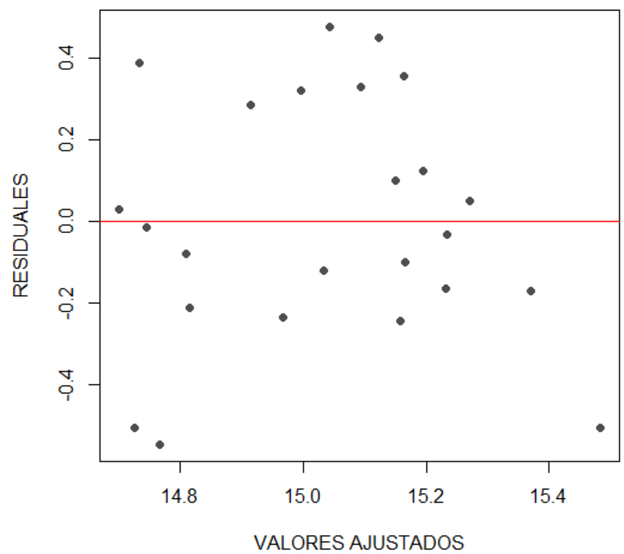


Figura 43. Gráfica de residuales, modelo exponencial, conjunto de viviendas: Longitud de tubería (m)

En la (figura 43) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo logarítmico

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo Longitud
1  5000000  1520.80
2  4500000  1162.41
3  4000000   859.58
4  4500000  2173.00
5  3000000  1754.89
6  3000000  1296.30
7  3200000  2962.10
8  3500000  1788.33
9  5800000  1628.10
10 4000000  2042.00
11 3700000   187.40
12 4200000  1726.58
13 5500000  1779.00
14 3500000  2030.69
15 2500000   234.03
16 2200000   493.17
17 4000000  2547.08
```

```
18 4500000  1897.55
19 5500000  1334.38
20 1500000   308.14
21 1500000   160.84
22 2500000    67.98
23 2500000  1052.38
24 2500000   467.19
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (x) “Longitud” para el modelo logarítmico:

```
> LNLongitud<-c(log(Longitud))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNLongitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
> abline(lm(Costo~LNLongitud),col="red")
```

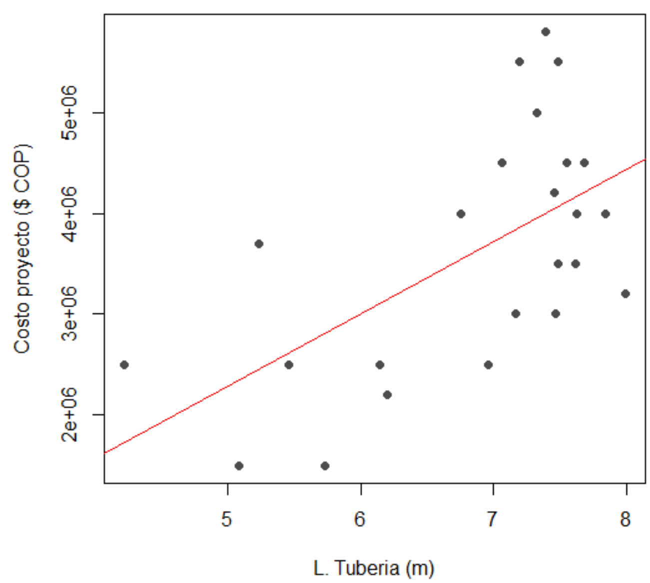


Figura 44. Gráfica de regresión lineal, modelo logarítmico, conjunto de viviendas: Longitud de tubería (m)

En la (figura 44) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(Costo~LNLongitud)
> summary(modelo)

Call:
lm(formula = Costo ~ LNLongitud)

Residuals:
     Min       1Q   Median       3Q      Max
-1313975  -844878  -143186   743067  1792773

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1293879    1403065  -0.922  0.36644
LNLongitud    716834     203024   3.531  0.00188 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 990600 on 22 degrees of freedom
Multiple R-squared:  0.3617,    Adjusted R-squared:  0.3327
F-statistic: 12.47 on 1 and 22 DF,  p-value: 0.00188
```

Se aprecia que uno de los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 49.69% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = \alpha + \beta x$

$Costo\ (\$\ COP) = -1293879 + (716834 * Longitud\ (m))$

El modelo lineal ajustado se puede expresar como la función logarítmica:

$\hat{y} = \alpha + \beta\ ln\ x$

$Costo\ (\$\ COP) = -1293879 + (716834 * LnLongitud\ (m))$ (Ecuación 16)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table
```

```
Response: Costo
           Df     Sum Sq    Mean Sq F value  Pr(>F)
LNLongitud  1 1.2232e+13 1.2232e+13  12.466 0.00188 **
Residuals  22 2.1586e+13 9.8120e+11
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.93909, p-value = 0.1556
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 2.303343e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

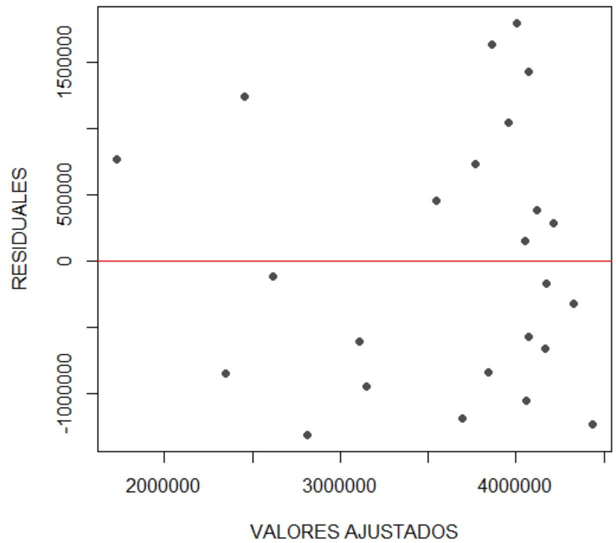


Figura 45. Gráfica de residuales, modelo logarítmico, conjunto de viviendas: Longitud de tubería (m)

En la (figura 45) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo recíproco

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo Longitud
1  5000000  1520.80
2  4500000  1162.41
3  4000000   859.58
4  4500000  2173.00
5  3000000  1754.89
6  3000000  1296.30
7  3200000  2962.10
8  3500000  1788.33
9  5800000  1628.10
10 4000000  2042.00
11 3700000   187.40
12 4200000  1726.58
13 5500000  1779.00
14 3500000  2030.69
15 2500000   234.03
16 2200000   493.17
17 4000000  2547.08
```

```
18 4500000  1897.55
19 5500000  1334.38
20 1500000   308.14
21 1500000   160.84
22 2500000    67.98
23 2500000  1052.38
24 2500000   467.19
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se transforma la variable, creando un vector con la inversa de (x) “Longitud” para el modelo recíproco:

```
> Longitudinv<-c(1/Longitud)
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Longitudinv,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP) ",pch=16,col="grey30")
> abline(lm(Costo~Longitudinv),col="red")
```

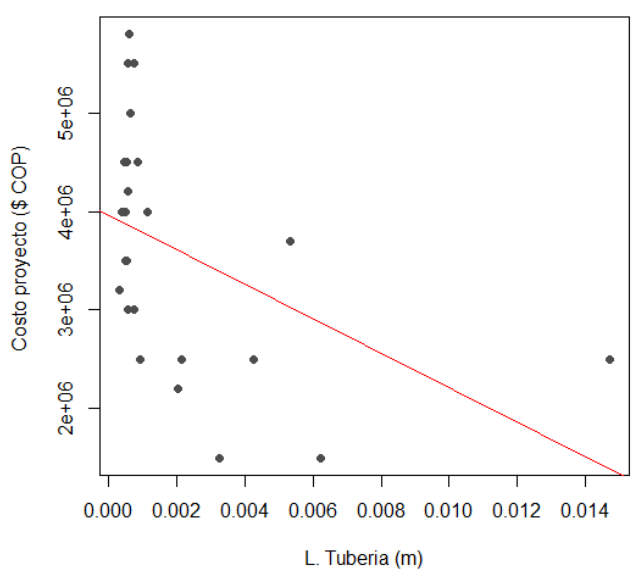


Figura 46. Gráfica de regresión lineal, modelo recíproco, conjunto de viviendas: Longitud de tubería (m)

En la (figura 46) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~Longitudinv))
> summary(modelo)

Call:
lm(formula = Costo ~ Longitudinv)

Residuals:
     Min       1Q   Median       3Q      Max
-1894813  -837806   113079   675755  1943449

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3964343     269119  14.731 7.07e-13 ***
Longitudinv -175494807   72742008  -2.413   0.0246 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1103000 on 22 degrees of freedom
Multiple R-squared:  0.2092,    Adjusted R-squared:  0.1733
F-statistic:  5.82 on 1 and 22 DF,  p-value: 0.02462
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 17.33% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = \alpha + \beta x$$

$$Costo\ (\$\ COP) = 3964343 - (175494807 * Longitud\ (m))$$

El modelo lineal ajustado se puede expresar como la función Reciproca:

$$\hat{y} = \alpha + \beta\left(\frac{1}{x}\right)$$

$$Costo\ (\$\ COP) = 3964343 - 175494807\left(\frac{1}{Longitud\ (m)}\right) \qquad \text{(Ecuación 17)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
```

```
Analysis of Variance Table

Response: Costo
            Df     Sum Sq    Mean Sq F value  Pr(>F)
Longitudinv  1 7.0753e+12 7.0753e+12  5.8205 0.02462 *
Residuals   22 2.6743e+13 1.2156e+12
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

  Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96682, p-value = 0.5895
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 1.756462e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

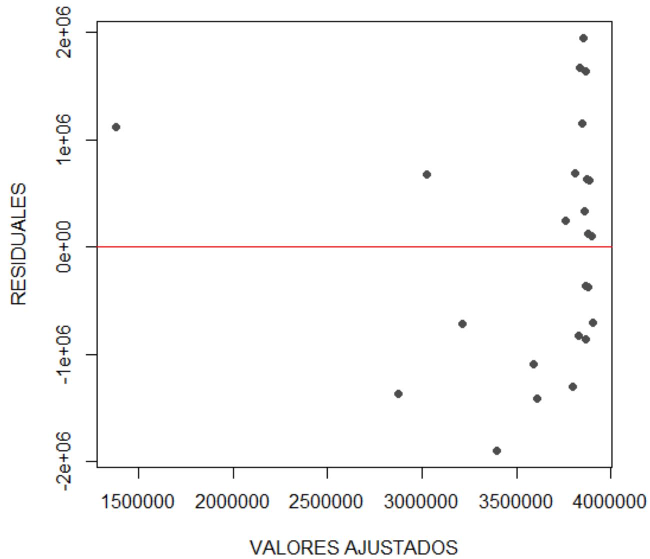


Figura 47. Gráfica de residuales, modelo recíproco, conjunto de viviendas: Longitud de tubería (m)

En la (figura 47) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo polinómico</u>

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo Longitud
1  5000000  1520.80
2  4500000  1162.41
3  4000000   859.58
4  4500000  2173.00
5  3000000  1754.89
6  3000000  1296.30
7  3200000  2962.10
8  3500000  1788.33
9  5800000  1628.10
10 4000000  2042.00
11 3700000   187.40
12 4200000  1726.58
13 5500000  1779.00
14 3500000  2030.69
15 2500000   234.03
16 2200000   493.17
17 4000000  2547.08
```

```
18 4500000  1897.55
19 5500000  1334.38
20 1500000   308.14
21 1500000   160.84
22 2500000    67.98
23 2500000  1052.38
24 2500000   467.19
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se construye el modelo de regresión utilizando la función poly (se ajusta polinomio de grado dos):

```
> fit2 <- lm(Costo ~ poly(Longitud, 2, raw=TRUE))
```

Se revisan los resultados con la instrucción summary:

```
> summary(fit2)

Call:
lm(formula = Costo ~ poly(Longitud, 2, raw = TRUE))

Residuals:
     Min       1Q   Median       3Q      Max
-1351715  -742408    10975   543726  1514662

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1.709e+06  4.712e+05   3.627  0.00158 **
poly(Longitud, 2, raw = TRUE)1  2.865e+03  7.770e+02   3.687  0.00137 **
poly(Longitud, 2, raw = TRUE)2 -7.879e-01  2.825e-01  -2.789  0.01099 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 914300 on 21 degrees of freedom
Multiple R-squared:  0.4809,    Adjusted R-squared:  0.4315
F-statistic: 9.729 on 2 and 21 DF,  p-value: 0.001023
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 43.15% de la variabilidad observada.

El modelo polinomial de regresión estimado es:

$$\hat{y} = \beta + \beta_1 x - \beta_2 x^2$$

$$Costo\ (\$\ COP) = 1{,}709 * 10^6 + (2{,}865 * 10^3) * (Longitud\ (m)) - (7{,}879 * 10^{-1}) * (Longitud\ (m)^2)$$ (Ecuación 18)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(fit2)
Analysis of Variance Table

Response: Costo
                            Df     Sum Sq    Mean Sq F value   Pr(>F)
poly(Longitud, 2, raw = TRUE)  2 1.6265e+13 8.1324e+12  9.7291 0.001023 **
Residuals                   21 1.7554e+13 8.3589e+11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(fit2$residuals)

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.95717, p-value = 0.3842
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(fit2$residuals)
[1] 6.426977e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
    > residuales<-fit2$residuals
    > valores.ajustados<-fit2$fitted
    > plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
    > abline(h=0,col="red")
```

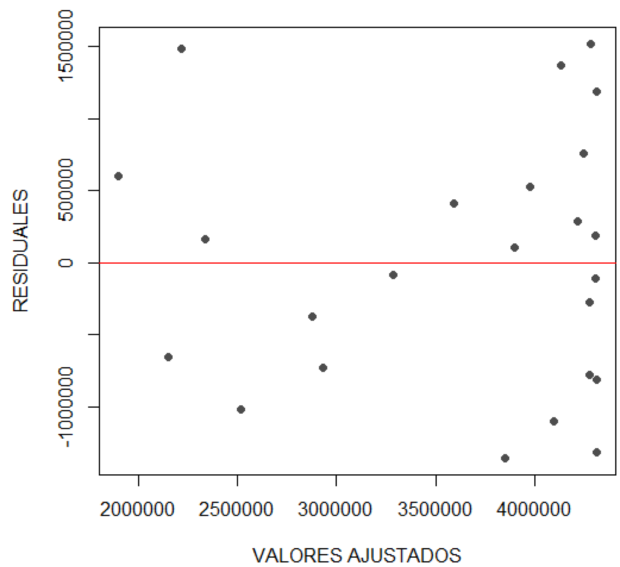


Figura 48. Gráfica de residuales, modelo polinómico, conjunto de viviendas: Longitud de tubería (m)

En la (figura 48) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

De acuerdo con la realización de las ecuaciones metodológicas con los modelos funcionales de regresión lineal mediante el software de programación R, se efectúa una tabla resumen donde se encuentran cada uno de los resultados obtenidos en los diferentes modelos además de tener en cuenta los coeficientes de R^2, p-value y el error estándar de regresión en la clasificación conjunto de viviendas unifamiliar con la variable longitud de tubería (m).

Tabla 8. Resultados, conjunto de viviendas unifamiliares: Longitud de tubería (m)

Modelo funcional	Ecuación. Regresión Lineal	R-squared	p-value	Error estándar $\hat{\sigma}$
Lineal	(13) $Costo\ (\$\ COP) = 2561650{,}0 + 798{,}1 * Longitud(m)$	0,2563	0,006781	1,0935e+12
Potencial	(14) $Costo\ (\$\ COP) = 675642{,}9263 * Longitud\ (m)^{0{,}23598}$	0,3867	0,0007027	0,08551
Exponencial	(15) $Costo\ (\$\ COP) = 2373793{,}817 *\ e^{(2{,}702*10^{-4}*Longitud\ (m))}$	0,3194	0,002371	0,0949
Logarítmica	(16) $Costo\ (\$\ COP) = -1293879 + 716834 * LnLongitud\ (m)$	0,3327	0,00188	9,8120e+11
Recíproca	(17) $Costo\ (\$\ COP) = 3964343 - 175494807\left(\frac{1}{Longitud\ (m)}\right)$	0,1733	0,02462	1,2156e+12
Polinómica	(18) $Costo\ (\$\ COP) = 1{,}709 * 10^{6} + (2{,}865 * 10^{3}) * (Longitud\ (m)) - (7{,}879 * 10^{-1}) * (Longitud\ (m)^{2})$	0,4315	0,001023	8,3589e+11

Una vez construidos los modelos de regresión lineal para la variable explicativa "Longitud de tubería (m)" en la clasificación "conjunto de viviendas", se pueden comparar los resultados obtenidos para elegir la ecuación que presenta mayor significancia estadística; en este caso se encuentra que el modelo funcional potencial muestra un coeficiente de determinación R^2 de 0,3867, aunque este no representa el valor más cercano a uno (1) es el que presenta menor error estándar de la regresión, es decir, su valor es cercano a cero (0); también se observa en el mismo modelo que el p- value es de 0,0007027 el cual presenta un valor por debajo del 5%, indicando su significancia estadística.

Los resultados también demuestran que el modelo funcional que menos se ajusta es el recíproco, dado que el coeficiente de determinación es el más bajo en comparación a los obtenidos en los otros modelos analizados y el que más se aleja de uno (1).

Teniendo en cuenta estos resultados analizados, la *(ecuación 14)* del modelo funcional potencial es la que representa un buen alcance y el menor error estándar de la regresión para la variable dependiente *Costo ($ COP)*, por lo cual será empleada para esta clasificación.

4.2.1.4. Conjuntos de viviendas: Número de viviendas (und)

Modelo lineal

Se importan los datos que se encuentan en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo Viviendas
1  5000000       162
2  4500000       132
3  4000000        50
4  4500000       114
5  3000000       150
6  3000000       110
7  3200000       250
8  3500000        60
9  5800000       108
10 4000000       176
11 3700000       180
12 4200000       144
13 5500000       206
14 3500000        58
15 2500000        11
16 2200000        31
17 4000000       114
18 4500000       197
19 5500000       153
20 1500000        19
21 1500000        10
22 2500000        18
23 2500000       100
24 2500000        24
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Viviendas,Costo,xlab="N. Viviendas (und)",ylab="Costo proyecto
($ COP)",pch=16,col="grey30")
```

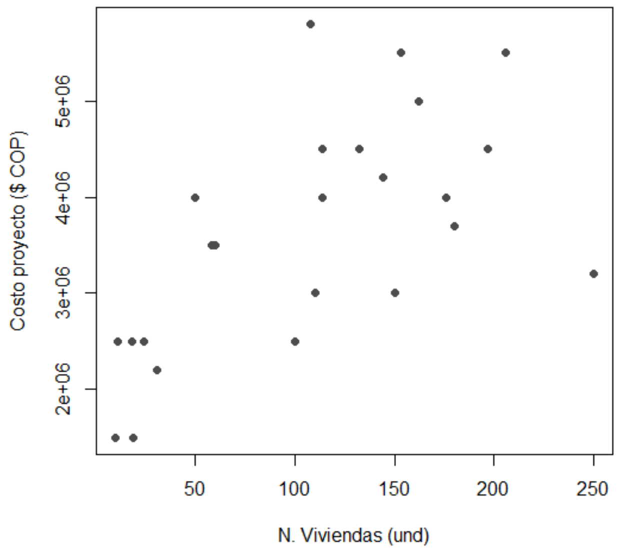


Figura 49. Gráfica de dispersión, conjunto de viviendas: Número de viviendas (und)

En la (figura 49) se observan datos en los que aumenta el número de viviendas, pero no el costo del proyecto, y también viceversa, es decir, que no son directamente proporcionales; algunos costos son elevados para el área que representan.

```
> abline(lm(Costo~Viviendas),col="red")
```

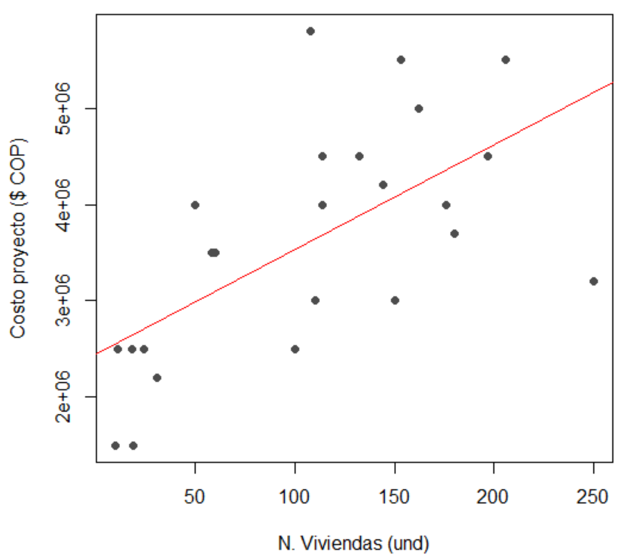


Figura 50. Gráfica de regresión lineal, modelo lineal, conjunto de viviendas: Número de viviendas (und)

En la (figura 50) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Se construye el modelo de regresión utilizando la función *lm*:

```
> modelo<-lm(Costo~Viviendas)
> summary(modelo)

Call:
lm(formula = Costo ~ Viviendas)

Residuals:
     Min       1Q   Median       3Q      Max 
-1961241  -652455   -71586   666888  2184862 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2439230     366575   6.654 1.09e-06 ***
Viviendas      10888       2879   3.782  0.00103 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 965200 on 22 degrees of freedom
Multiple R-squared:  0.394,	Adjusted R-squared:  0.3664 
F-statistic:  14.3 on 1 and 22 DF,  p-value: 0.001025
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 36.64% de la variabilidad observada.

El modelo de regresión estimado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$Costo\ (\$\ COP) = 2439230 + 10888 * Viviendas\ (und)$$ *(Ecuación 19)*

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido:

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)   
Viviendas  1 1.3324e+13 1.3324e+13  14.303 0.001025 **
Residuals 22 2.0494e+13 9.3156e+11                    
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.98701, p-value = 0.9838
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 3.760133e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

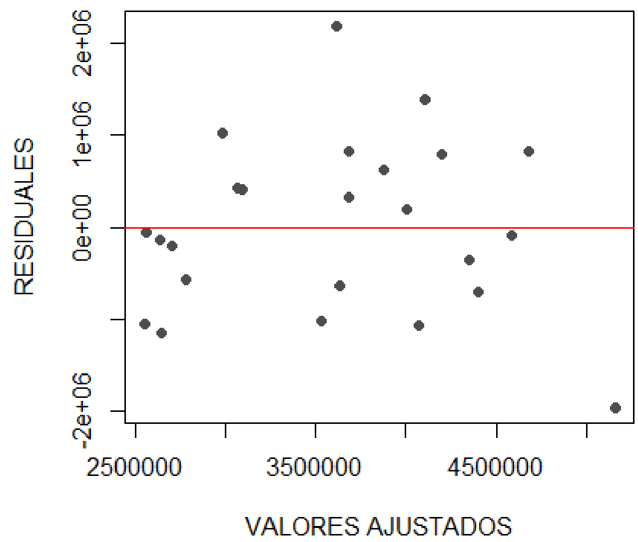


Figura 51. Gráfica de residuales, modelo lineal, conjunto de viviendas: Número de viviendas (und)

En la (figura 51) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Potencial

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo Viviendas
1  5000000       162
2  4500000       132
3  4000000        50
4  4500000       114
5  3000000       150
6  3000000       110
7  3200000       250
8  3500000        60
9  5800000       108
10 4000000       176
11 3700000       180
12 4200000       144
13 5500000       206
14 3500000        58
15 2500000        11
16 2200000        31
17 4000000       114
```

```
18 4500000       197
19 5500000       153
20 1500000        19
21 1500000        10
22 2500000        18
23 2500000       100
24 2500000        24
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Viviendas,Costo,xlab="N. Viviendas (und)",ylab="Costo proyecto
($ COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) "Costo" y (x) "Viviendas" para el modelo potencial:

```
> LNCosto<-c(log(Costo))
> LNViviendas<-c(log(Viviendas))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNViviendas,LNCosto,xlab="N. Viviendas (und)",ylab="Costo
proyecto ($ COP)",pch=16,col="grey30")
> abline(lm(LNCosto~LNViviendas),col="red")
```

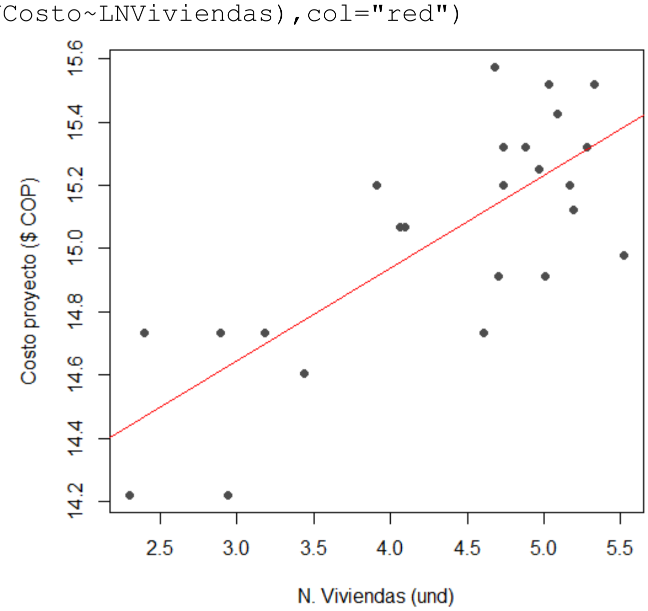


Figura 52. Gráfica de regresión lineal, modelo potencial, conjunto de viviendas: Número de viviendas (und)

En la (figura 52) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(LNCosto~LNViviendas)
> summary(modelo)

Call:
lm(formula = LNCosto ~ LNViviendas)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.4065 -0.1799  0.0419  0.1664  0.4355 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 13.76265    0.23174  59.389  < 2e-16 ***
LNViviendas  0.29371    0.05215   5.632 1.16e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2443 on 22 degrees of freedom
Multiple R-squared:  0.5905,    Adjusted R-squared:  0.5719 
F-statistic: 31.72 on 1 and 22 DF,  p-value: 1.157e-05
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 57.19% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = ln\, \alpha + \beta x^*$$

$$LnCosto\ (\$\ COP) = 13{,}76265 + 0{,}29371 * Viviendas\ (und)$$

El modelo lineal ajustado se puede expresar como la función potencial:

$$\hat{y} = \alpha x^{\beta}$$

$$Costo\ (\$\ COP) = 948512{,}2657 * Viviendas\ (und)^{0{,}29371} \qquad \text{(Ecuación 20)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
```

```
             Df Sum Sq Mean Sq F value    Pr(>F)    
LNViviendas  1 1.8936  1.8936  31.721 1.157e-05 ***
Residuals   22 1.3133  0.0597                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

      Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.95077, p-value = 0.2815
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean (modelo$residuals)

[1] 5.637851e-18
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

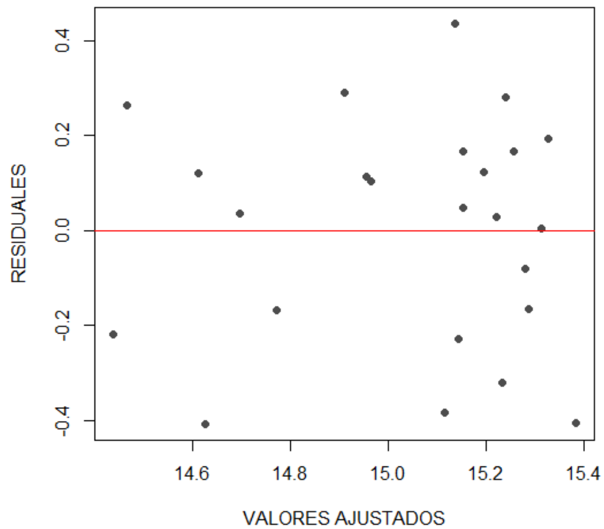


Figura 53. Gráfica de residuales, modelo potencial, conjunto de viviendas: Número de viviendas (und)

En la (figura 53) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo exponencial

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo Viviendas
1  5000000       162
2  4500000       132
3  4000000        50
4  4500000       114
5  3000000       150
6  3000000       110
7  3200000       250
8  3500000        60
9  5800000       108
10 4000000       176
11 3700000       180
12 4200000       144
13 5500000       206
14 3500000        58
15 2500000        11
16 2200000        31
17 4000000       114
```

```
18 4500000       197
19 5500000       153
20 1500000        19
21 1500000        10
22 2500000        18
23 2500000       100
24 2500000        24
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Viviendas,Costo,xlab="N. Viviendas (und)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” para el modelo exponencial:

```
> LNCosto<-c(log(Costo))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
>plot(Viviendas,LNCosto,xlab="N. Viviendas (und)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
> abline(lm(LNCosto~Viviendas),col="red")
```

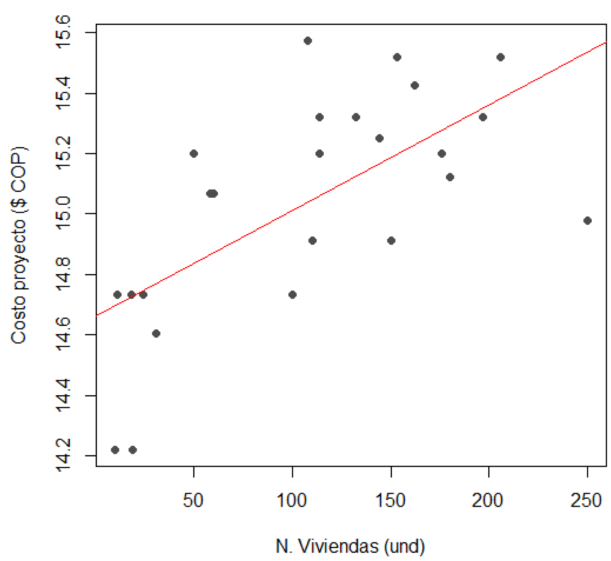


Figura 54. Gráfica de regresión lineal, modelo exponencial, conjunto de viviendas: Número de viviendas (und)

En la (figura 54) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(LNCosto~Viviendas)
> summary(modelo)
Call:
lm(formula = LNCosto ~ Viviendas)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.55934 -0.16595  0.02063  0.19627  0.53392 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.466e+01  1.093e-01 134.147  < 2e-16 ***
Viviendas   3.511e-03  8.583e-04   4.091 0.000484 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2877 on 22 degrees of freedom
Multiple R-squared:  0.432,     Adjusted R-squared:  0.4062 
F-statistic: 16.73 on 1 and 22 DF,  p-value: 0.0004836
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 40.62% de la variabilidad observada.

El modelo de regresión estimado:

$$y^* = ln\,\alpha + \beta x$$

$$LnCosto\ (\$\ COP) = (1{,}466 * 10^{01}) + \left((3{,}511 * 10^{-03}) * Viviendas\ (und)\right)$$

El modelo lineal ajustado se puede expresar como la función exponencial:

$$\hat{y} = \alpha e^{\beta x}$$

$$Costo\ (\$\ COP) = 2326789{,}55 * e^{\left((3{,}511*10^{-03})*Viviendas\ (und)\right)}$$ (Ecuación 21)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
```

```
          Df Sum Sq Mean Sq F value    Pr(>F)    
Viviendas  1 1.3854  1.3854  16.733 0.0004836 ***
Residuals 22 1.8215  0.0828                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

      Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96995, p-value = 0.6659
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 4.633094e-18
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS", ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

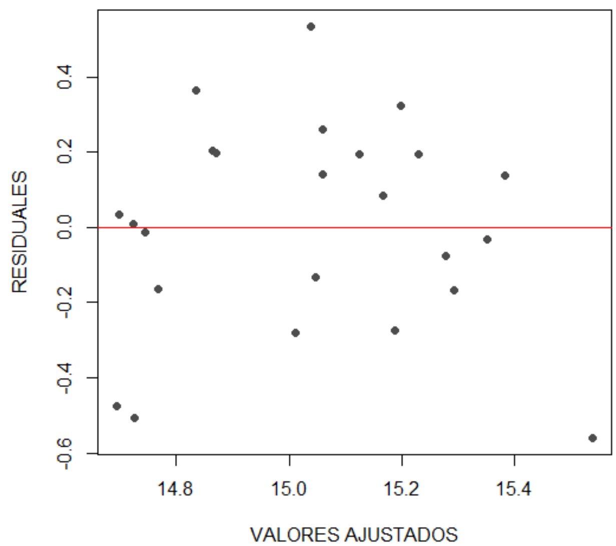


Figura 55. Gráfica de residuales, modelo exponencial, conjunto de viviendas: Número de viviendas (und)

En la (figura 55) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Logarítmico

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo Viviendas
1  5000000       162
2  4500000       132
3  4000000        50
4  4500000       114
5  3000000       150
6  3000000       110
7  3200000       250
8  3500000        60
9  5800000       108
10 4000000       176
11 3700000       180
12 4200000       144
13 5500000       206
14 3500000        58
15 2500000        11
16 2200000        31
17 4000000       114
```

```
18 4500000       197
19 5500000       153
20 1500000        19
21 1500000        10
22 2500000        18
23 2500000       100
24 2500000        24
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Viviendas,Costo,xlab="N. Viviendas (und)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (x) “Viviendas” para el modelo logarítmico:

```
> LNViviendas<-c(log(Viviendas))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNViviendas,Costo,xlab="N. Viviendas (und)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
> abline(lm(Costo~LNViviendas),col="red")
```

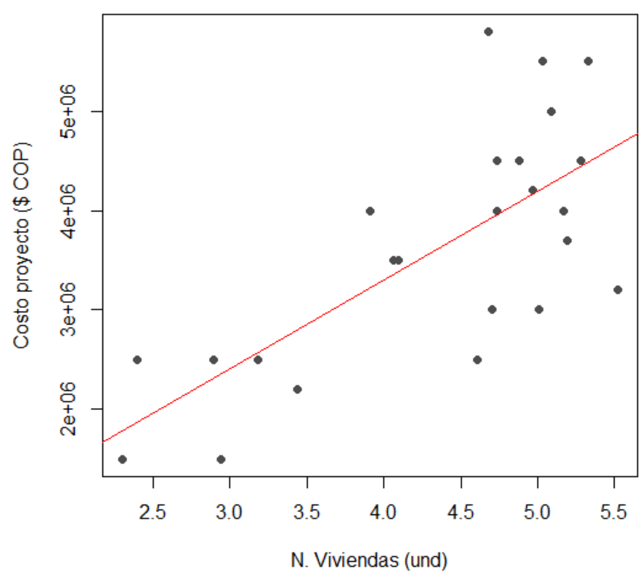


Figura 56. Gráfica de regresión lineal, modelo logarítmico, conjunto de viviendas: Número de viviendas (und)

En la (figura 56) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(Costo~LNViviendas)
> summary(modelo)

Call:
lm(formula = Costo ~ LNViviendas)

Residuals:
     Min       1Q   Median       3Q      Max
-1464707  -616866    42734   559736  1885536

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -270705     815903  -0.332    0.743
LNViviendas   893860     183606   4.868 7.25e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 860200 on 22 degrees of freedom
Multiple R-squared:  0.5186,    Adjusted R-squared:  0.4967
F-statistic:  23.7 on 1 and 22 DF,  p-value: 7.25e-05
```

Se aprecia que uno de los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 49.67% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = \alpha + \beta x$

$Costo\ (\$\ COP) = -270705 + (893860 * Viviendas\ (und))$

El modelo lineal ajustado se puede expresar como la función logarítmica:

$\hat{y} = \alpha + \beta\ ln\, x$

$Costo\ (\$\ COP) = -270705 + 893860 * LnViviendas(und)$ (Ecuación 22)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)

Analysis of Variance Table
```

```
Response: Costo
            Df     Sum Sq    Mean Sq F value   Pr(>F)    
LNViviendas  1 1.7539e+13 1.7539e+13  23.701 7.25e-05 ***
Residuals   22 1.6280e+13 7.3999e+11                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

      Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.98263, p-value = 0.9389
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] -1.757557e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS", ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

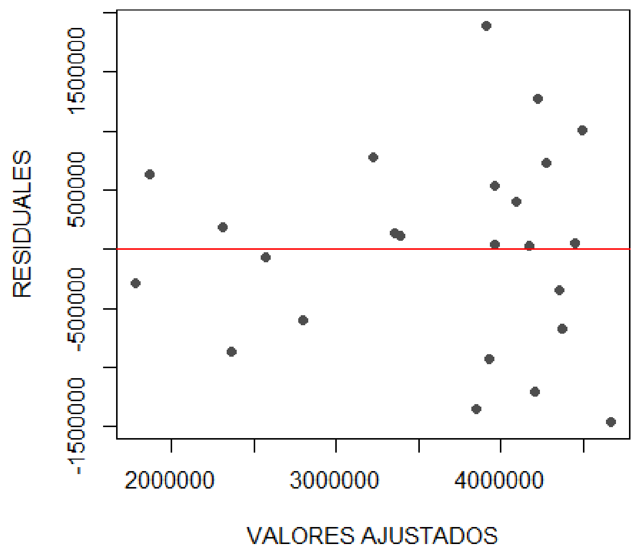


Figura 57. Gráfica de residuales, modelo logarítmico, conjunto de viviendas: Número de viviendas (und)

En la (figura 57) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Recíproco

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo Viviendas
1  5000000       162
2  4500000       132
3  4000000        50
4  4500000       114
5  3000000       150
6  3000000       110
7  3200000       250
8  3500000        60
9  5800000       108
10 4000000       176
11 3700000       180
12 4200000       144
13 5500000       206
14 3500000        58
15 2500000        11
16 2200000        31
```

```
17 4000000       114
18 4500000       197
19 5500000       153
20 1500000        19
21 1500000        10
22 2500000        18
23 2500000       100
24 2500000        24
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Viviendas,Costo,xlab="N. Viviendas (und)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
```

Se transforma la variable, creando un vector con la inversa de (x) “Viviendas” para el modelo recíproco:

```
> Viviendasinv<-c(1/Viviendas)
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Viviendasinv,Costo,xlab="N. Viviendas (und)",ylab="Costo proyecto ($ COP) ",pch=16,col="grey30")
> abline(lm(Costo~Viviendasinv),col="red")
```

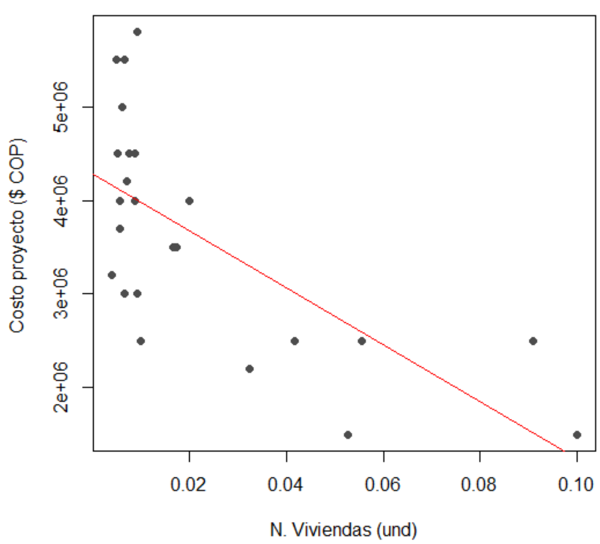


Figura 58. Gráfica de regresión lineal, modelo recíproco, conjunto de viviendas: Número de viviendas (und)

En la (figura 58) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~Viviendasinv))
> summary(modelo)

Call:
lm(formula = Costo ~ Viviendasinv)

Residuals:
     Min       1Q   Median       3Q      Max
-1479010  -625865   -53494   456217  1798413

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4283795     241660  17.727 1.63e-14 ***
Viviendasinv -30478439    6993830  -4.358 0.000252 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 908300 on 22 degrees of freedom
Multiple R-squared:  0.4633,    Adjusted R-squared:  0.4389
F-statistic: 18.99 on 1 and 22 DF,  p-value: 0.0002517
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 43.89% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = \alpha + \beta x$$

$$Costo\ (\$\ COP) = 4283795 + (-30478439 * Viviendas\ (und))$$

El modelo lineal ajustado se puede expresar como la función Recíproca:

$$\hat{y} = \alpha + \beta \left(\frac{1}{x}\right)$$

$$Costo\ (\$\ COP) = 4283795 - 30478439 \left(\frac{1}{Viviendas\ (und)}\right) \qquad \text{(Ecuación 23)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
```

```
Analysis of Variance Table

Response: Costo
             Df     Sum Sq    Mean Sq F value    Pr(>F)    
Viviendasinv  1 1.5668e+13 1.5668e+13  18.991 0.0002517 ***
Residuals    22 1.8150e+13 8.2501e+11                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

      Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96947, p-value = 0.654
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 1.757054e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

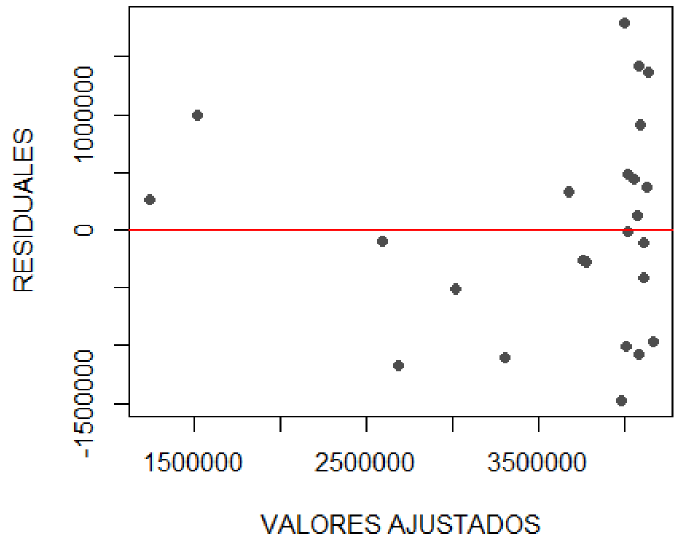


Figura 59. Gráfica de residuales, modelo recíproco, conjunto de viviendas: Número de viviendas (und)

En la (figura 59) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Polinómico

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo Viviendas
1  5000000       162
2  4500000       132
3  4000000        50
4  4500000       114
5  3000000       150
6  3000000       110
7  3200000       250
8  3500000        60
9  5800000       108
10 4000000       176
11 3700000       180
12 4200000       144
13 5500000       206
14 3500000        58
15 2500000        11
16 2200000        31
17 4000000       114
```

```
18 4500000       197
19 5500000       153
20 1500000        19
21 1500000        10
22 2500000        18
23 2500000       100
24 2500000        24
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Viviendas,Costo,xlab="N. Viviendas (und)",ylab="Costo proyecto
($ COP)",pch=16,col="grey30")
```

Se construye el modelo de regresión utilizando la función poly:

```
> fit2 <- lm(Costo ~ poly(Viviendas, 2, raw=TRUE))
```

Se revisan los resultados con la instrucción summary:

```
> summary(fit2)


Call:
lm(formula = Costo ~ poly(Viviendas, 2, raw = TRUE))

Residuals:
     Min       1Q   Median       3Q      Max
-1509814  -445272   144258   372404  1689686

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1612869.65  431238.22   3.740 0.001207 **
poly(Viviendas, 2, raw = TRUE)1   34531.41    8678.95   3.979 0.000684 ***
poly(Viviendas, 2, raw = TRUE)2    -105.62      37.12  -2.845 0.009694 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 839300 on 21 degrees of freedom
Multiple R-squared:  0.5626,    Adjusted R-squared:  0.5209
F-statistic: 13.51 on 2 and 21 DF,  p-value: 0.0001695
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 52.09% de la variabilidad observada.

El modelo polinomial de regresión estimado es:
$\hat{y} = \beta + \beta_1 x - \beta_2 x^2$
$Costo\ (\$\ COP) = 1612869{,}65 + (34531{,}41 * Viviendas\ (und)) + (-105{,}62 * Viviendas\ (und)^2)$
(Ecuación 24)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(fit2)

Analysis of Variance Table

Response: Costo
                             Df     Sum Sq   Mean Sq F value    Pr(>F)
poly(Viviendas, 2, raw = TRUE)  2 1.9026e+13 9.513e+12  13.505 0.0001695 ***
Residuals                    21 1.4792e+13 7.044e+11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(fit2$residuals)

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.98141, p-value = 0.9204
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(fit2$residuals)
[1] 2.182787e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-fit2$residuals
> valores.ajustados<-fit2$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

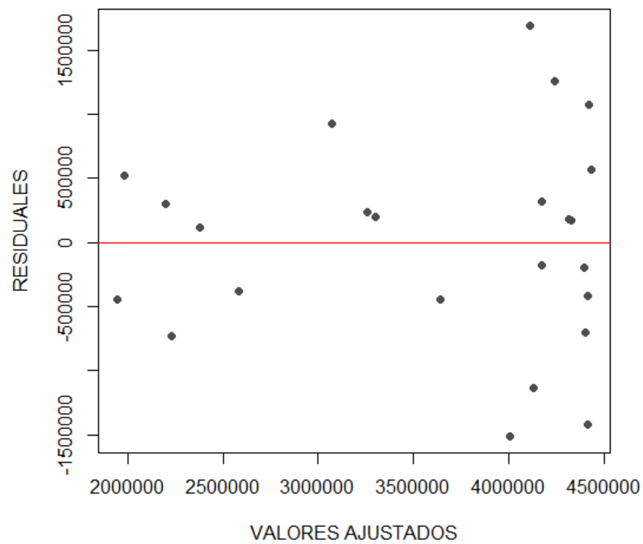


Figura 60. Gráfica de residuales, modelo polinómico, conjunto de viviendas: Número de viviendas (und)

En la (figura 60) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

De acuerdo con la realización de las ecuaciones metodológicas con los modelos funcionales de regresión lineal mediante el software de programación R, se efectúa una tabla resumen donde se encuentran cada uno de los resultados obtenidos en los diferentes modelos además de tener en cuenta los coeficientes de R^2, p-value y el error estándar de regresión en la clasificación conjunto de viviendas unifamiliar con la variable número de viviendas (und).

Tabla 9. Resultados, conjunto de viviendas unifamiliares: Número de viviendas (und)

Modelo funcional	Ecuación. Regresión Lineal	R-squared	p-value	Error estándar $\hat{\sigma}$
Lineal	(19) $Costo\ (\$\ COP) = 2439230 + 10888 * Viviendas\ (und)$	0,3664	0,001025	9,3156e+11
Potencial	(20) $Costo\ (\$\ COP) = 948512{,}2657 * Viviendas\ (und)^{0{,}29371}$	0,5719	1,157e-05	0,0597
Exponencial	(21) $Costo\ (\$\ COP) = 2326789{,}55 * e^{((3{,}511*10^{-03})*Viviendas\ (und))}$	0,4062	0,0004836	0,0828
Logarítmica	(22) $Costo\ (\$\ COP) = -270705 + 893860 * LnViviendas(und)$	0,4967	7,25e-05	7,3999e+11
Recíproca	(23) $Costo\ (\$\ COP) = 4283795 - 30478439\left(\frac{1}{Viviendas\ (und)}\right)$	0,4389	0,0002517	8,2501e+11
Polinómica	(24) $Costo\ (\$\ COP) = 1612869{,}65 + (34531{,}41 * Viviendas\ (und)) + (-105{,}62 * Viviendas\ (und)^2)$	0,5209	0,0001695	7,044e+11

Una vez construidos los modelos de regresión lineal para la variable explicativa “Número de viviendas (und)” en la clasificación “conjunto de viviendas”, se pueden comparar los resultados obtenidos para elegir la ecuación que presenta mayor significancia estadística; en este caso se encuentra que el modelo funcional potencial muestra un coeficiente de determinación R^2 de 0,5719, valor superior a los obtenidos en los otros modelos empleados y el que más se acerca a uno (1); también se observa en el mismo modelo que el p- value es de 1,157e-05 el cual presenta un valor por debajo del 5%, indicando su significancia estadística.

Los resultados también demuestran que el modelo funcional que menos se ajusta es el lineal, dado que el coeficiente de determinación es el más bajo en comparación a los obtenidos en los otros modelos analizados y el que más se aleja de uno (1).

Teniendo en cuenta estos resultados analizados, la *(ecuación 20)* del modelo funcional potencial es la que representa mayor alcance y menor error estándar de la regresión para la variable dependiente *Costo ($ COP)*, por lo cual será empleada para esta clasificación.

4.2.1.5. Conjunto de Viviendas, regresión lineal múltiple: Número de viviendas (und), Longitud de tubería (m), Área (Ha)

El modelo de regresión lineal múltiple se emplea cuando existe más de una variable de estudio con el fin de analizar su importancia, en este caso las variables explicativas a estudiar son: Número de viviendas (und), Longitud de tubería (m) y Área (Ha).

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
   Viviendas Longitud  Area   Costo
1        162  1520.80 2.067 5000000
2        132  1162.41 1.853 4500000
3         50   859.58 1.280 4000000
4        114  2173.00 1.515 4500000
5        150  1754.89 1.580 3000000
6        110  1296.30 1.564 3000000
7        250  2962.10 3.715 3200000
8         60  1788.33 2.159 3500000
9        108  1628.10 1.254 5800000
10       176  2042.00 1.866 4000000
11       180   187.40 2.600 3700000
12       144  1726.58 3.067 4200000
13       206  1779.00 2.259 5500000
14        58  2030.69 1.149 3500000
15        11   234.03 0.178 2500000
16        31   493.17 0.261 2200000
17       114  2547.08 1.469 4000000
18       197  1897.55 2.249 4500000
19       153  1334.38 1.911 5500000
20        19   308.14 0.225 1500000
21        10   160.84 0.145 1500000
22        18    67.98 0.121 2500000
23       100  1052.38 1.348 2500000
24        24   467.19 0.456 2500000
> attach(data)
```

Se construye el modelo de regresión lineal múltiple usando la función *lm*:

$$yest = B_0 + B_1x_1 + B_2x_2$$

```
> modelo<-lm(Costo~Viviendas+Longitud+Area)
> summary(modelo)

Call:
lm(formula = Costo ~ Viviendas + Longitud + Area)
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2114013  -545667   -17405   599294  2072088 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2295491.2   413834.7   5.547 1.98e-05 ***
Viviendas      8891.4     6512.4   1.365    0.187    
Longitud        324.0      352.7   0.919    0.369    
Area         -44182.6   485811.4  -0.091    0.928    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 991000 on 20 degrees of freedom
Multiple R-squared:  0.4192,    Adjusted R-squared:  0.3321 
F-statistic: 4.811 on 3 and 20 DF,  p-value: 0.01109
```

Se aprecia que solo uno de los coeficientes logra ser significativo, además el p-value obtenido indica la significancia global del método. El coeficiente de determinación (R^2) indica el 33.21% de la variabilidad observada.

La ecuación del modelo de regresión es:
$yest = B_0 + B_1 x_1 + B_2 x_2$
$Costo\ (\$\ COP) = 2295491{,}2 + (8891{,}4 * Viviendas\ (und)) + (321 * Longitud\ (m)) - (44182{,}6 * Área\ (Ha))$ (Ecuación 25)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)   
Viviendas  1 1.3324e+13 1.3324e+13 13.5667 0.001473 **
Longitud   1 8.4402e+11 8.4402e+11  0.8594 0.364962   
Area       1 8.1232e+09 8.1232e+09  0.0083 0.928440   
Residuals 20 1.9642e+13 9.8211e+11                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.98373, p-value = 0.9534
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 2.000888e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

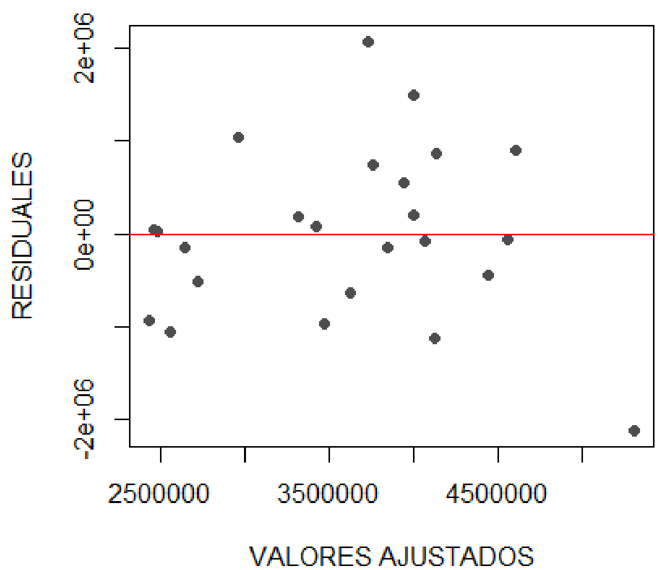


Figura 61. Gráfica de residuales, conjunto de viviendas, regresión lineal múltiple: Número de viviendas (und), Longitud de tubería (m), Área (Ha)

En la (figura 61) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Se realiza el test de contraste de homocedasticidad Breusch-Pagan. La hipótesis nula indica que los residuales tienen varianza constante.

```
> library(lmtest)
> bptest(modelo)

        studentized Breusch-Pagan test

data:  modelo
BP = 3.1697, df = 3, p-value = 0.3662
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de homocedasticidad (varianza constante) en los residuales.

De acuerdo con la realización de la ecuación metodológica con el modelo de regresión lineal múltiple mediante el software de programación R, se efectúa una tabla resumen donde se encuentran el resultado obtenido del modelo además de tener en cuenta el coeficiente de R^2, p-value y el error estándar de regresión en la clasificación conjunto de viviendas unifamiliar con las variables número de viviendas (und), longitud de tubería (m) y área (Ha).

Tabla 10. Resultados, conjunto de viviendas unifamiliares: Número de viviendas (und), Longitud de tubería (m), Área (Ha)

Método	**Ecuación. Regresión Lineal Múltiple**	**R-squared**	**p-value**	**Error estándar $\hat{\sigma}$**
Regresión Lineal Múltiple	$(25) Costo\ (\$\ COP) = 2295491{,}2 + (8891{,}4 * Viviendas\ (und)) + (321 * Longitud\ (m)) - (44182{,}6 * Área\ (Ha))$	0,3321	0,01109	9,8211e+11

De acuerdo con los resultados obtenidos en el modelo funcional y en el ANOVA, solo uno de los coeficientes logra ser significativo; esto expone que las variables explicativas tienen mayor alcance estadístico cuando son analizadas de manera independiente, es por esto, que para evaluar las ecuaciones metodológicas en la clasificación "conjunto de viviendas unifamiliares", no se tendrá en cuenta el método de regresión lineal múltiple.

4.2.2. Residencia multifamiliar

4.2.2.1. Edificio: Área (Ha)

<u>Modelo lineal</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   6000000 0.023
2   1200000 0.036
3   4000000 0.041
4    900000 0.017
5  10000000 0.217
6   4200000 0.018
7   1500000 0.003
8   4000000 0.140
9   1800000 0.019
10  1300000 0.007
11  3200000 0.026
12  4500000 0.032
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

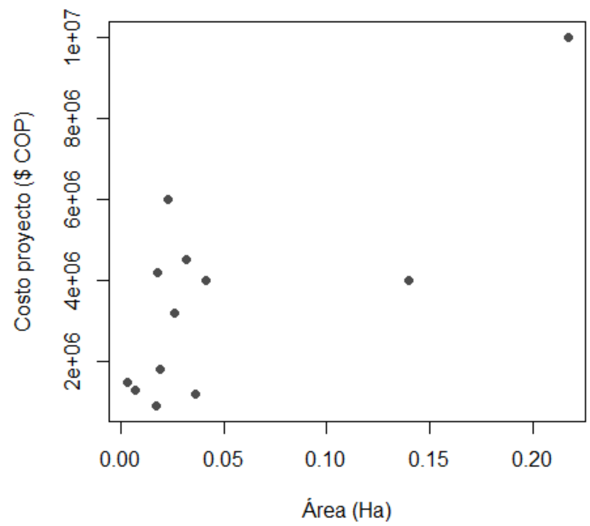


Figura 62. Gráfica de dispersión, edificio: Área (Ha)

En la (figura 62) se observan datos en los que aumenta el costo, pero no el área del proyecto, es decir, que no son directamente proporcionales; algunos costos son elevados para el área que representan.

```
> abline(lm(Costo~Area),col="red")
```

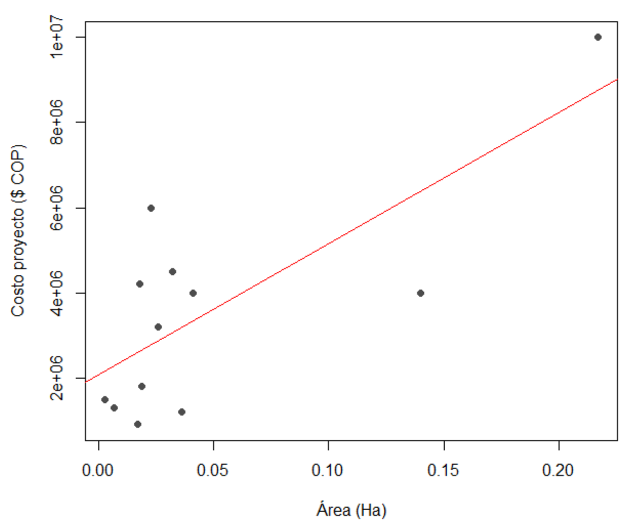


Figura 63. Gráfica de regresión lineal, modelo lineal, edificio: Área (Ha)

En la (figura 63) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Se construye el modelo de regresión utilizando la función *lm*:

```
> modelo<-lm(Costo~Area)
> summary(modelo)

Call:
lm(formula = Costo ~ Area)

Residuals:
     Min       1Q   Median       3Q      Max 
-2375799 -1156543  -160537  1302135  3227672 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  2063953     653891   3.156   0.0102 * 
Area        30798895    8387871   3.672   0.0043 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1779000 on 10 degrees of freedom
Multiple R-squared:  0.5741,    Adjusted R-squared:  0.5316 
F-statistic: 13.48 on 1 and 10 DF,  p-value: 0.004304
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 53,16% de la variabilidad observada.

El modelo de regresión estimado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$Costo\ (\$\ COP) = 2063953 + 30798895 * Área\ (Ha) \qquad \text{(Ecuación 26)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido:

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)   
Area       1 4.2676e+13 4.2676e+13  13.482 0.004304 **
Residuals 10 3.1654e+13 3.1654e+12                    
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales. Básicamente, los supuestos que deben verificarse son los siguientes:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.9595, p-value = 0.7767
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 2.667946e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

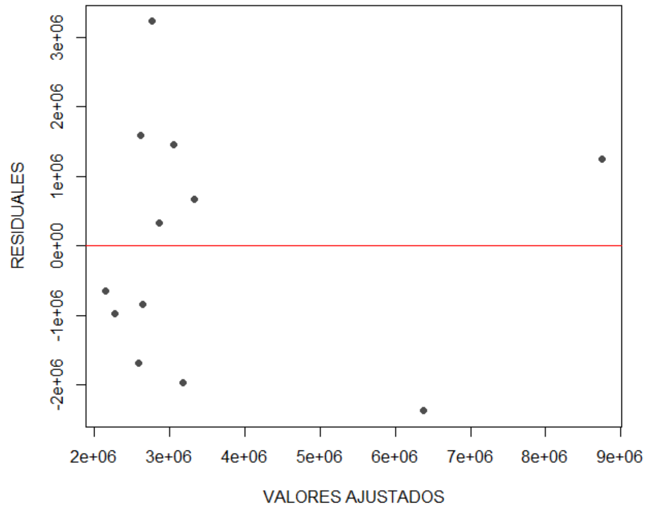


Figura 64. Gráfica de residuales, modelo lineal, edificio: Área (Ha)

En la (figura 64) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo potencial

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo  Area
1   6000000 0.023
2   1200000 0.036
3   4000000 0.041
4    900000 0.017
5  10000000 0.217
6   4200000 0.018
7   1500000 0.003
8   4000000 0.140
9   1800000 0.019
10  1300000 0.007
11  3200000 0.026
12  4500000 0.032
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto
($)COP",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) "Costo" y (x) "Área" para el modelo potencial:

```
> LNCosto<-c(log(Costo))
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($)
COP",pch=16,col="grey30")
> abline(lm(LNCosto~LNArea),col="red")
```

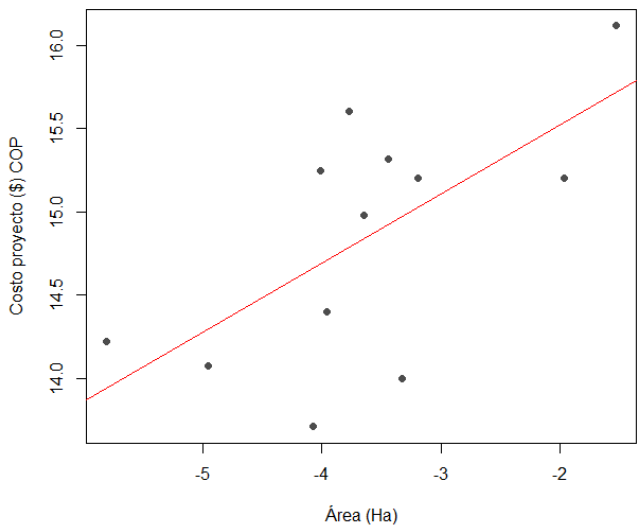


Figura 65. Gráfica de regresión lineal, modelo potencial, edificio: Área (Ha)

En la (figura 65) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~LNArea))
> summary(modelo)

Call:
lm(formula = LNCosto ~ LNArea)

Residuals:
    Min       1Q   Median       3Q      Max
```

```
-0.9747 -0.3115  0.1583  0.3970  0.8207

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.3520     0.5946  27.500 9.37e-11 ***
LNArea        0.4150     0.1563   2.655   0.0241 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5953 on 10 degrees of freedom
Multiple R-squared:  0.4135,    Adjusted R-squared:  0.3548
F-statistic: 7.049 on 1 and 10 DF,  p-value: 0.02411
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 35,48% de la variabilidad observada.

El modelo de regresión estimado es:

$y^* = ln\,\alpha + \beta x^*$

$LnCosto(\$\,COP) = \ln\,(16{,}3520) + 0{,}4150 * Área(Ha)$

El modelo lineal ajustado se puede expresar como la función potencial:

$\hat{y} = \alpha x^{\beta}$

$Costo\ (\$\,COP) = 12635236{,}29 * Área\ (Ha)^{0{,}4150}$ (Ecuación 27)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
          Df Sum Sq Mean Sq F value  Pr(>F)
LNArea     1 2.4982 2.49825  7.0489 0.02411 *
Residuals 10 3.5441 0.35441
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)
```

```
        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.93202, p-value = 0.402
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 5.782186e-17
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

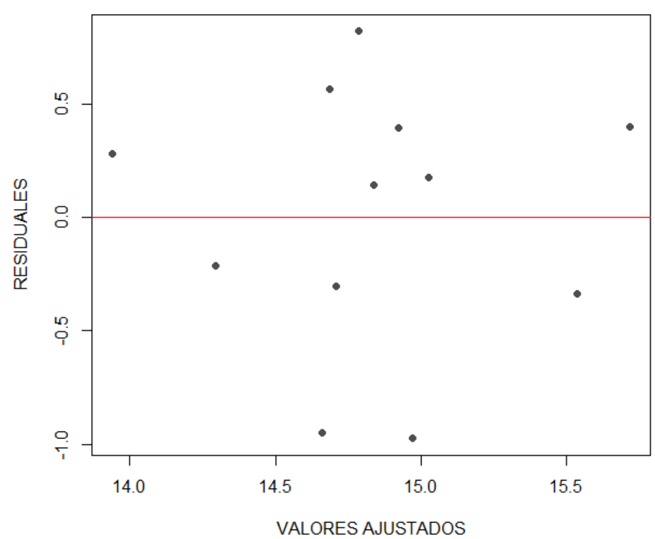


Figura 66. Gráfica de residuales, modelo potencial, edificio: Área (Ha)

En la (figura 66) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo exponencial</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   6000000 0.023
2   1200000 0.036
3   4000000 0.041
4    900000 0.017
5  10000000 0.217
6   4200000 0.018
7   1500000 0.003
8   4000000 0.140
9   1800000 0.019
10  1300000 0.007
11  3200000 0.026
12  4500000 0.032
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) "Costo" para el modelo exponencial:

```
> LNCosto<-c(log(Costo))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Area,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(LNCosto~Area),col="red")
```

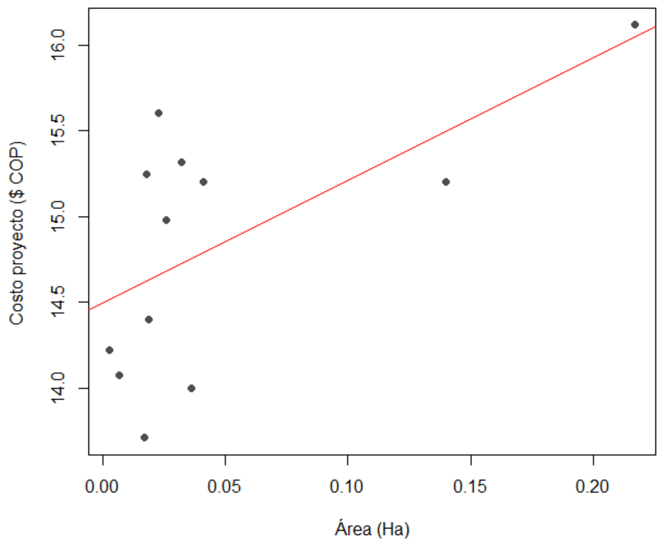


Figura 67. Gráfica de regresión lineal, modelo exponencial, edificio: Área (Ha)

En la (figura 67) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~Area))
> summary(modelo)

Call:
lm(formula = LNCosto ~ Area)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.90725 -0.33933 -0.07828  0.45846  0.94700 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  14.4959     0.2249  64.442 1.97e-14 ***
Area          7.1443     2.8855   2.476   0.0328 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.612 on 10 degrees of freedom
Multiple R-squared:   0.38,	Adjusted R-squared:  0.318 
F-statistic:  6.13 on 1 and 10 DF,  p-value: 0.03277
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 31,80% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = ln\,\alpha + \beta x$

$LnCosto\ (\$\ COP) = ln(14{,}4959) + (7{,}1443 * Área(Ha))$

El modelo lineal ajustado se puede expresar como la función exponencial:

$\hat{y} = \alpha e^{\beta x}$

$Costo\ (\$\ COP) = 1974646{,}593 * e^{(7{,}1443 * Área\ (Ha))}$ (Ecuación 28)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
          Df Sum Sq Mean Sq F value  Pr(>F)
Area       1 2.2964  2.2964  6.1301 0.03277 *
Residuals 10 3.7460  0.3746
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96397, p-value = 0.8387
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
```

```
[1] 1.017659e-16
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

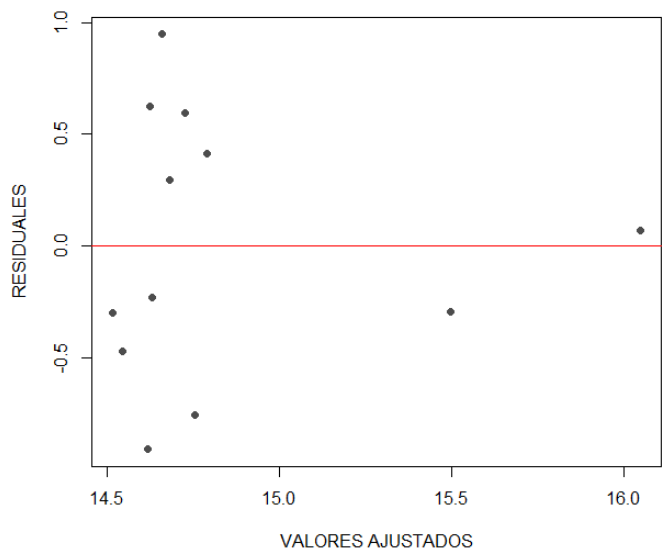


Figura 68. Gráfica de residuales, modelo exponencial, edificio: Área (Ha)

En la (figura 68) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo logarítmico</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
```

```
1   6000000 0.023
2   1200000 0.036
3   4000000 0.041
4    900000 0.017
5  10000000 0.217
6   4200000 0.018
7   1500000 0.003
8   4000000 0.140
9   1800000 0.019
10  1300000 0.007
11  3200000 0.026
12  4500000 0.032
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($)
COP",pch=16,col="grey30")
```

Se realiza la transformación del vector (x) "Área" para el modelo logarítmico:

```
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($) COP
",pch=16,col="grey30")
> abline(lm(Costo~LNArea),col="red")
```

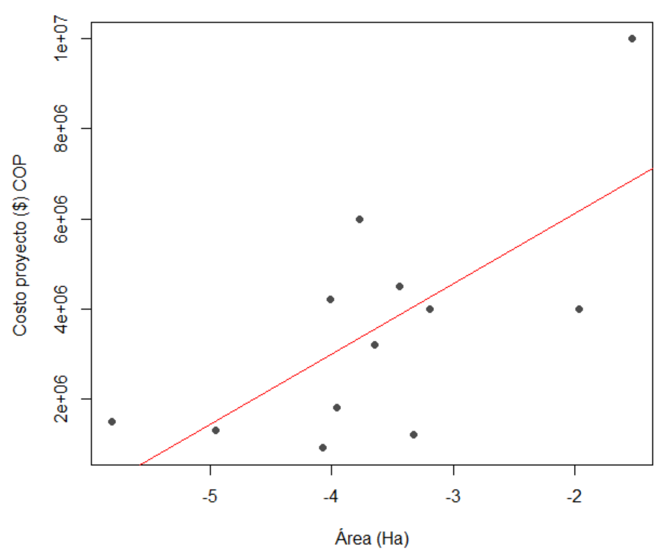


Figura 69. Gráfica de regresión lineal, modelo logarítmico, edificio: Área (Ha)

En la (figura 69) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~LNArea))
> summary(modelo)

Call:
lm(formula = Costo ~ LNArea)

Residuals:
     Min       1Q   Median       3Q      Max 
-2844766 -1431014  -220671  1257607  3157169 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9222645    1975914   4.668 0.000884 ***
LNArea       1557614     519399   2.999 0.013369 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1978000 on 10 degrees of freedom
Multiple R-squared:  0.4735,    Adjusted R-squared:  0.4208 
F-statistic: 8.993 on 1 and 10 DF,  p-value: 0.01337
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 42,08% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = \alpha + \beta x$

$Costo\ (\$\ COP) = 9222645 + (1557614 * Área\ (Ha))$

El modelo lineal ajustado se puede expresar como la función logarítmica:

$\hat{y} = \alpha + \beta \ln x$

$Costo\ (\$\ COP) = 9222645 + 1557614 * LnÁrea\ (Ha)$ (Ecuación 29)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table
```

```
Response: Costo
          Df     Sum Sq    Mean Sq F value  Pr(>F)
LNArea     1 3.5195e+13 3.5195e+13  8.9933 0.01337 *
Residuals 10 3.9135e+13 3.9135e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96564, p-value = 0.8602
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 2.328306e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

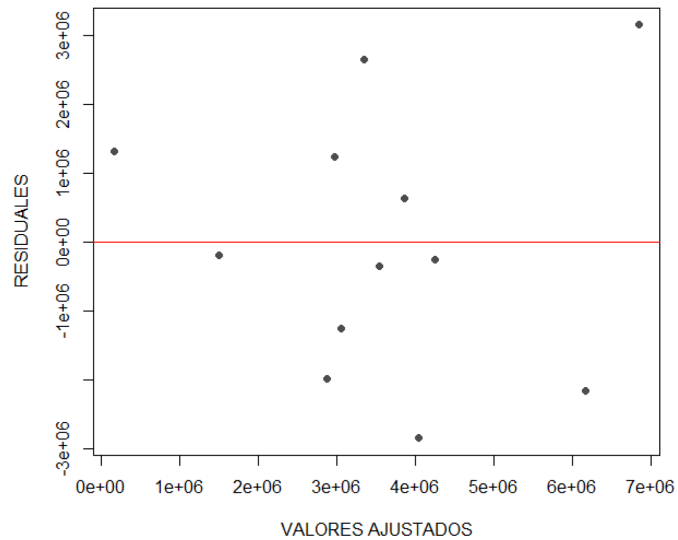


Figura 70. Gráfica de residuales, modelo logarítmico, edificio: Área (Ha)

En la (figura 70) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo recíproco

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   6000000 0.023
2   1200000 0.036
3   4000000 0.041
4    900000 0.017
5  10000000 0.217
6   4200000 0.018
7   1500000 0.003
8   4000000 0.140
9   1800000 0.019
10  1300000 0.007
11  3200000 0.026
12  4500000 0.032
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($)
COP",pch=16,col="grey30")
```

Se transforma la variable, creando un vector con la inversa de (x) "Área" para el modelo recíproco:

```
> Areainv<-c(1/Area)
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Areainv,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($) COP
",pch=16,col="grey30")
> abline(lm(Costo~Areainv),col="red")
```

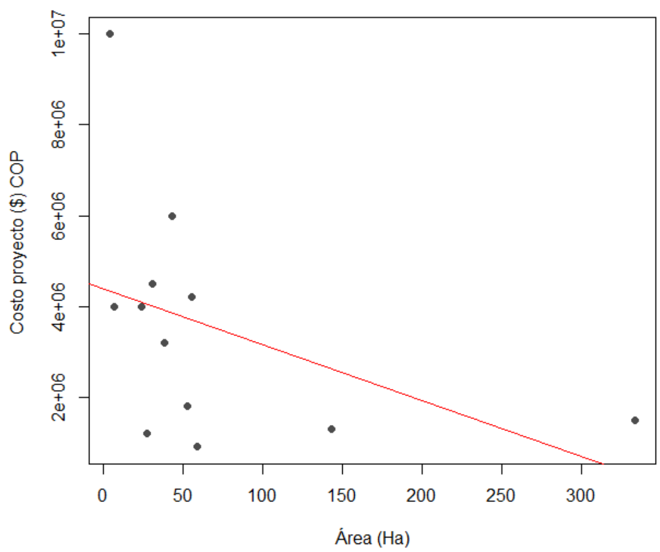


Figura 71. Gráfica de regresión lineal, modelo recíproco, edificio: Área (Ha)

En la (figura 71) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~Areainv))
> summary(modelo)

Call:
lm(formula = Costo ~ Areainv)
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2849845 -1485230  -197789   673120  5664774 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4391987     904247   4.857 0.000664 ***
Areainv       -12317       8180  -1.506 0.163068    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2462000 on 10 degrees of freedom
Multiple R-squared:  0.1848,	Adjusted R-squared:  0.1033 
F-statistic: 2.267 on 1 and 10 DF,  p-value: 0.1631
```

Se aprecia que tanto los coeficientes como el modelo de regresión no son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 10,33% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = \alpha + \beta x$$

$$Precio\ (\$\ COP) = 4391987 + (-12317 * Área\ (Ha))$$

El modelo lineal ajustado se puede expresar como la función Recíproca:

$$\hat{y} = \alpha + \beta\left(\frac{1}{x}\right)$$

$$Precio\ (\$\ COP) = 4391987 - 12317\left(\frac{1}{Área\ (Ha)}\right) \qquad \text{(Ecuación 30)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value Pr(>F)
Areainv    1 1.3737e+13 1.3737e+13   2.267 0.1631
Residuals 10 6.0593e+13 6.0593e+12               
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.91434, p-value = 0.2424
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 2.377073e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

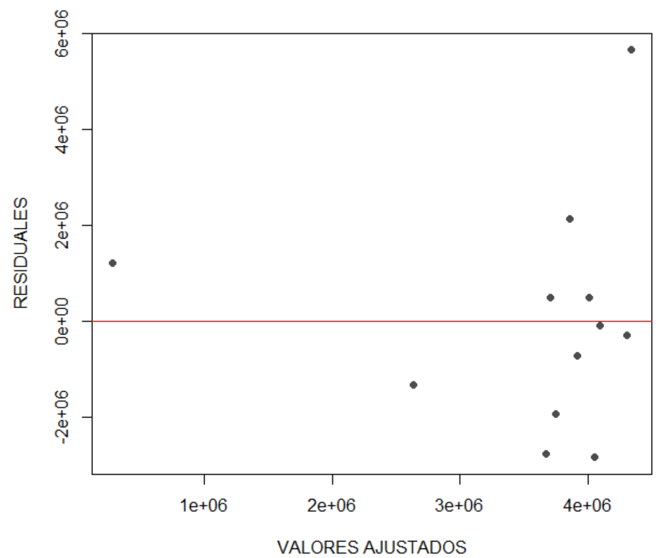


Figura 72. Gráfica de residuales, modelo recíproco, edificio: Área (Ha)

En la (figura 72) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo Polinómico</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   6000000 0.023
2   1200000 0.036
3   4000000 0.041
4    900000 0.017
5  10000000 0.217
6   4200000 0.018
7   1500000 0.003
8   4000000 0.140
9   1800000 0.019
10  1300000 0.007
11  3200000 0.026
12  4500000 0.032
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($)
COP",pch=16,col="grey30")
```

Se construye el modelo de regresión utilizando la función poly (se ajusta polinomio de grado dos):

```
> fit2 <- lm(Costo ~ poly(Area, 2, raw=TRUE))
```

Se revisan los resultados con la instrucción summary:

```
> summary(fit2)

Call:
lm(formula = Costo ~ poly(Area, 2, raw = TRUE))

Residuals:
     Min       1Q   Median       3Q      Max 
-1815880 -1409739  -251054  1176216  3247215 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)  
(Intercept)                  2666621    1022848   2.607   0.0284 *
poly(Area, 2, raw = TRUE)1    493075   39913342   0.012   0.9904  
poly(Area, 2, raw = TRUE)2 141443605  181950529   0.777   0.4569  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1815000 on 9 degrees of freedom
Multiple R-squared:  0.6009,    Adjusted R-squared:  0.5123 
F-statistic: 6.777 on 2 and 9 DF,  p-value: 0.01602
```

Se aprecia que uno de los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 51,23% de la variabilidad observada.

El modelo polinomial de regresión estimado es:

$$\hat{y} = \beta + \beta_1 x - \beta_2 x^2$$

$$Costo\ (\$\ COP) = 2666621 + \left(493075 * Área\ (Ha)\right) - \left(141443605 * Área\ (Ha)^2\right)$$

(Ecuación 31)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(fit2)
Analysis of Variance Table

Response: Costo
                          Df     Sum Sq    Mean Sq F value  Pr(>F)
poly(Area, 2, raw = TRUE)  2 4.4668e+13 2.2334e+13  6.7766 0.01602 *
Residuals                  9 2.9662e+13 3.2958e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(fit2$residuals)

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.90257, p-value = 0.1712
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(fit2$residuals)
[1] 4.753815e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-fit2$residuals
> valores.ajustados<-fit2$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

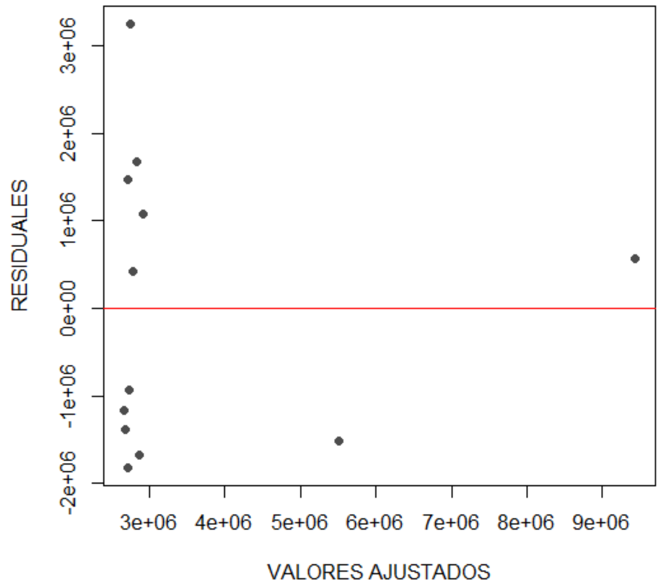


Figura 73. Gráfica de residuales, modelo polinómico, edificios: Área (Ha)

En la (figura 73) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

De acuerdo con la realización de las ecuaciones metodológicas con los modelos funcionales de regresión lineal mediante el software de programación R, se efectúa una tabla resumen donde se encuentran cada uno de los resultados obtenidos en los diferentes modelos además de tener en cuenta los coeficientes de R^2, p-value y el error estándar de regresión en la clasificación edificio multifamiliar con la variable área (Ha).

Tabla 11. Resultados, edificio multifamiliar: Área (Ha)

Modelo funcional	Ecuación. Regresión Lineal	R-squared	p-value	Error estándar $\hat{\sigma}$
Lineal	(26) $Costo\ (\$\ COP) = 2063953 + 30798895 * Área\ (Ha)$	0,5316	0,004304	3,1654e+12
Potencial	(27) $Costo\ (\$\ COP) = 12635236{,}29 * Área\ (Ha)^{0{,}4150}$	0,3548	0,02411	0,35441
Exponencial	(28) $Costo\ (\$\ COP) = 1974646{,}593 * e^{(7{,}1443 * Área\ (Ha))}$	0,318	0,03277	0,3746
Logarítmica	(29) $Costo\ (\$\ COP) = 9222645 + 1557614 * LnÁrea\ (Ha)$	0,4208	0,01337	3,9135e+12
Recíproca	(30) $Costo\ (\$\ COP) = 4391987 - 12317\left(\frac{1}{Área\ (Ha)}\right)$	0,1033	0,1631	6,0593e+12
Polinómica	(31) $Costo\ (\$\ COP) = 2666621 + \left(493075 * Área\ (Ha)\right) - \left(141443605 * Área\ (Ha)^2\right)$	0,5123	0,01602	3,2958e+12

Una vez construidos los modelos de regresión lineal para la variable explicativa "Área (Ha)" en la clasificación "edificio", se pueden comparar los resultados obtenidos para elegir la ecuación que presenta mayor significancia estadística; en este caso se encuentra que el modelo funcional lineal muestra un coeficiente de determinación R^2 de 0,5316, valor superior a los obtenidos en los otros modelos empleados y el que más se acerca a uno (1); también se observa en el mismo modelo que el p- value es de 0,004304 el cual presenta un valor por debajo del 5%, indicando su significancia estadística.

Los resultados también demuestran que el modelo funcional que menos se ajusta es el recíproco, dado que el coeficiente de determinación es el más bajo en comparación a los obtenidos en los otros modelos analizados y el que más se aleja de uno (1).

Teniendo en cuenta estos resultados analizados, la *(ecuación 26)* del modelo funcional lineal es la que representa mayor alcance y menor error estándar de la regresión para la variable dependiente *Costo ($ COP)*, por lo cual será empleada para esta clasificación.

4.2.2.2. Conjunto de edificios: Área (Ha)

Modelo lineal

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   4000000 2.496
2   7000000 1.211
3   2000000 0.554
4   5000000 1.081
5   6000000 0.052
6   6000000 0.972
7   9200000 1.312
8   8000000 1.285
9   8500000 0.876
10  4000000 0.589
11  2200000 0.366
12  5500000 0.554
13  4000000 0.313
14  5500000 2.049
15 12000000 2.257
16  8000000 0.580
17  4700000 1.081
18  3670000 0.540
19  7700000 1.486
20  6500000 1.026
21  4500000 1.296
22  5500000 0.761
23  6000000 2.021
24  5000000 0.761
25  4800000 0.847
26  4000000 0.911
27  4200000 1.759
28 10500000 1.921
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

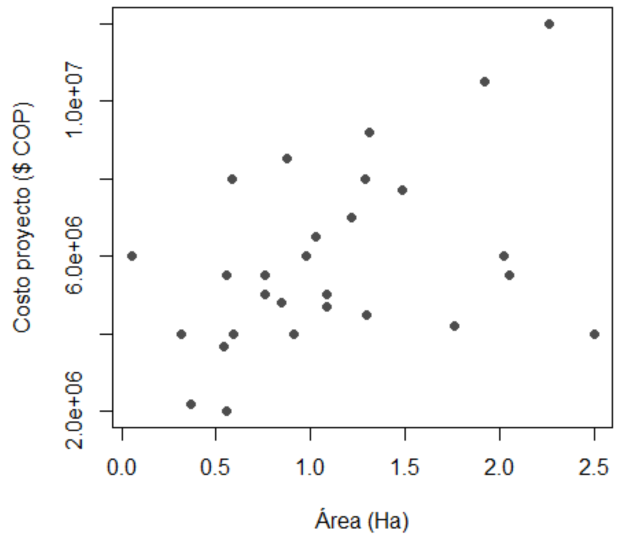


Figura 74. Gráfica de dispersión, conjunto de edificios: Área (Ha)

En la (figura 74) se observan datos en los que aumenta el área, pero no el costo del proyecto, y también viceversa, es decir, que no son directamente proporcionales; algunos costos son elevados para el área que representan.

```
> abline(lm(Costo~Area),col="red")
```

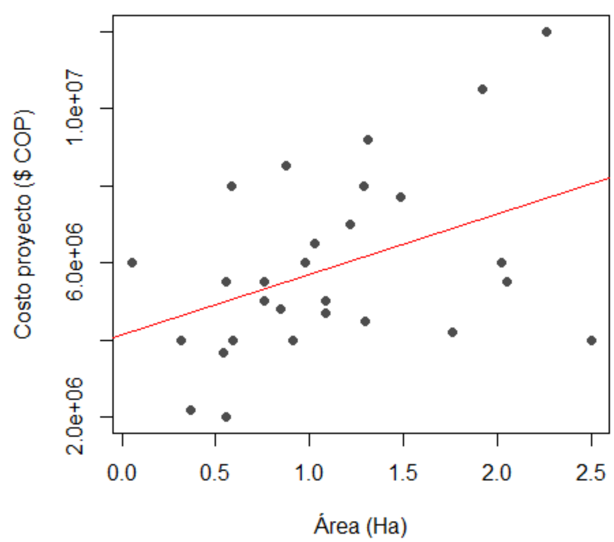


Figura 75. Gráfica de regresión lineal, modelo lineal, conjunto de edificios: Área (Ha)

En la (figura 75) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Se construye el modelo de regresión utilizando la función *lm*:

```
> modelo<-lm(Costo~Area)
> summary(modelo)

Call:
lm(formula = Costo ~ Area)

Residuals:
     Min       1Q   Median       3Q      Max
-4032522 -1363385  -466003  1384659  4341597

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4125410     843343   4.892 4.47e-05 ***
Area         1565349     667101   2.346   0.0268 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2164000 on 26 degrees of freedom
Multiple R-squared:  0.1748,    Adjusted R-squared:  0.143
F-statistic: 5.506 on 1 and 26 DF,  p-value: 0.02685
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 14,30% de la variabilidad observada.

El modelo de regresión estimado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$Costo\ (\$\ COP) = 4125410 + 1565349 * Área\ (Ha)$$ (Ecuación 32)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido:

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value  Pr(>F)
Area       1 2.5784e+13 2.5784e+13   5.506 0.02685 *
Residuals 26 1.2175e+14 4.6829e+12
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales. Básicamente, los supuestos que deben verificarse son los siguientes:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.97419, p-value = 0.6959
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 5.612882e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

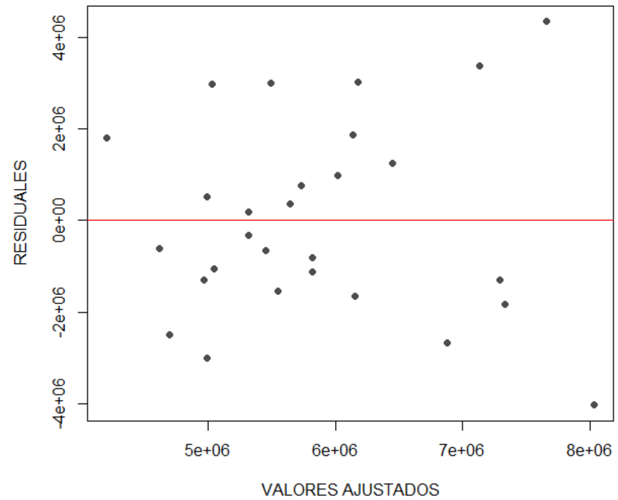


Figura 76. Gráfica de residuales, modelo lineal, conjunto de edificios: Área (Ha)

En la (figura 76) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo potencial

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   4000000 2.496
2   7000000 1.211
3   2000000 0.554
4   5000000 1.081
5   6000000 0.052
6   6000000 0.972
7   9200000 1.312
8   8000000 1.285
9   8500000 0.876
10  4000000 0.589
11  2200000 0.366
12  5500000 0.554
13  4000000 0.313
14  5500000 2.049
15 12000000 2.257
16  8000000 0.580
17  4700000 1.081
18  3670000 0.540
19  7700000 1.486
```

```
20  6500000 1.026
21  4500000 1.296
22  5500000 0.761
23  6000000 2.021
24  5000000 0.761
25  4800000 0.847
26  4000000 0.911
27  4200000 1.759
28 10500000 1.921
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto
($)COP",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” y (x) “Área” para el modelo potencial:

```
> LNCosto<-c(log(Costo))
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($)
COP",pch=16,col="grey30")
> abline(lm(LNCosto~LNArea),col="red")
```

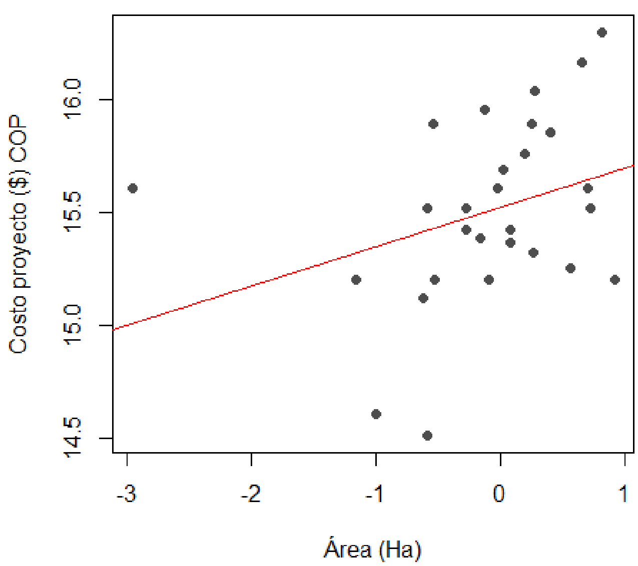


Figura 77. Gráfica de regresión lineal, modelo potencial, conjunto de edificios: Área (Ha)

En la (figura 77) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~LNArea))
> summary(modelo)

Call:
lm(formula = LNCosto ~ LNArea)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.91237 -0.23477 -0.04524  0.27977  0.63480 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 15.52387    0.07571 205.046   <2e-16 ***
LNArea       0.17413    0.09826   1.772   0.0881 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3966 on 26 degrees of freedom
Multiple R-squared:  0.1078,	Adjusted R-squared:  0.07346 
F-statistic: 3.141 on 1 and 26 DF,  p-value: 0.08808
```

Se aprecia que tanto los coeficientes como el modelo de regresión no son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 7,34% de la variabilidad observada.

El modelo de regresión estimado es:

$y^* = ln\, \alpha + \beta x^*$

$LnCosto(\$\, COP) = \ln{(15{,}52387)} + 0{,}17413 * Área(Ha)$

El modelo lineal ajustado se puede expresar como la función potencial:

$\hat{y} = \alpha x^{\beta}$

$Costo\ (\$\, COP) = 5519898 * Área\ (Ha)^{0{,}17413}$ (Ecuación 33)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table
```

```
Response: LNCosto
          Df Sum Sq Mean Sq F value  Pr(>F)
LNArea     1 0.4940 0.49402  3.1406 0.08808 .
Residuals 26 4.0898 0.15730
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.97045, p-value = 0.5927
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] -4.456845e-18
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

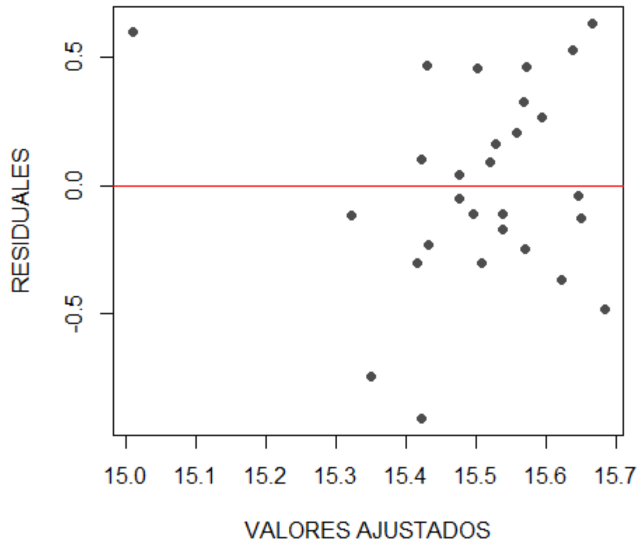


Figura 78. Gráfica de residuales, modelo potencial, conjunto de edificios: Área (Ha)

En la (figura 78) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo exponencial

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   4000000 2.496
2   7000000 1.211
3   2000000 0.554
4   5000000 1.081
5   6000000 0.052
6   6000000 0.972
7   9200000 1.312
8   8000000 1.285
9   8500000 0.876
10  4000000 0.589
11  2200000 0.366
12  5500000 0.554
13  4000000 0.313
14  5500000 2.049
15 12000000 2.257
16  8000000 0.580
17  4700000 1.081
```

```
18   3670000 0.540
19   7700000 1.486
20   6500000 1.026
21   4500000 1.296
22   5500000 0.761
23   6000000 2.021
24   5000000 0.761
25   4800000 0.847
26   4000000 0.911
27   4200000 1.759
28 10500000 1.921
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) "Costo" para el modelo exponencial:

```
> LNCosto<-c(log(Costo))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Area,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(LNCosto~Area),col="red")
```

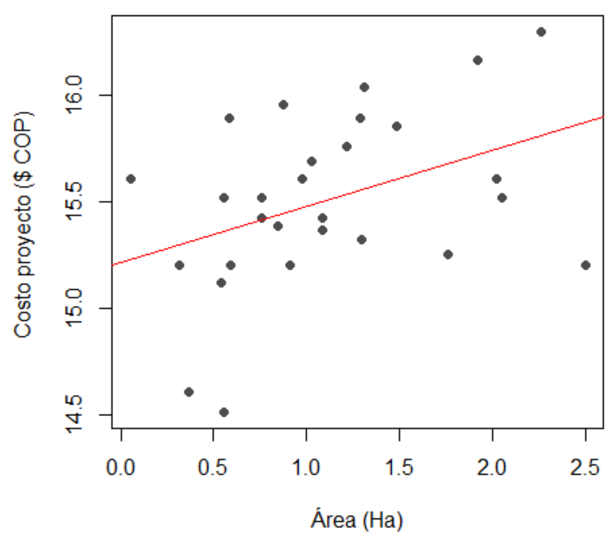


Figura 79. Gráfica de regresión lineal, modelo exponencial, conjunto de edificios: Área (Ha)

En la (figura 79) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~Area))
> summary(modelo)

Call:
lm(formula = LNCosto ~ Area)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.85061 -0.23428 -0.02075  0.27416  0.52882 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  15.2130     0.1500 101.445   <2e-16 ***
Area          0.2641     0.1186   2.226   0.0349 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3848 on 26 degrees of freedom
Multiple R-squared:  0.1601,    Adjusted R-squared:  0.1278 
F-statistic: 4.956 on 1 and 26 DF,  p-value: 0.03487
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 12,78% de la variabilidad observada.

El modelo de regresión estimado:

$$y^* = ln\,\alpha + \beta x$$

$$LnCosto\ (\$\ COP) = ln(15{,}2130) + 0{,}2641 * \acute{A}rea(Ha)$$

El modelo lineal ajustado se puede expresar como la función exponencial:

$$\hat{y} = \alpha e^{\beta x}$$

$$Costo\ (\$\ COP) = 4045031{,}921 * e^{(0{,}2641 * \acute{A}rea\ (Ha))} \quad \text{(Ecuación 34)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table
```

```
Response: LNCosto
          Df Sum Sq Mean Sq F value  Pr(>F)
Area       1 0.7339 0.73391  4.9564 0.03487 *
Residuals 26 3.8499 0.14807
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.94913, p-value = 0.1884
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 2.475272e-18
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

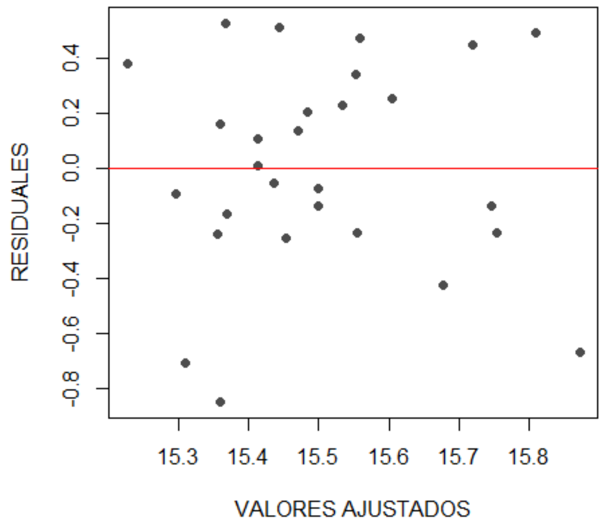


Figura 80. Gráfica de residuales, modelo exponencial, conjunto de edificios: Área (Ha)

En la (figura 80) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo logarítmico</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   4000000 2.496
2   7000000 1.211
3   2000000 0.554
4   5000000 1.081
5   6000000 0.052
6   6000000 0.972
7   9200000 1.312
8   8000000 1.285
9   8500000 0.876
10  4000000 0.589
11  2200000 0.366
12  5500000 0.554
13  4000000 0.313
14  5500000 2.049
15 12000000 2.257
16  8000000 0.580
17  4700000 1.081
18  3670000 0.540
```

```
19  7700000 1.486
20  6500000 1.026
21  4500000 1.296
22  5500000 0.761
23  6000000 2.021
24  5000000 0.761
25  4800000 0.847
26  4000000 0.911
27  4200000 1.759
28 10500000 1.921
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($)
COP",pch=16,col="grey30")
```

Se realiza la transformación del vector (x) "Área" para el modelo logarítmico:

```
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($) COP
",pch=16,col="grey30")
> abline(lm(Costo~LNArea),col="red")
```

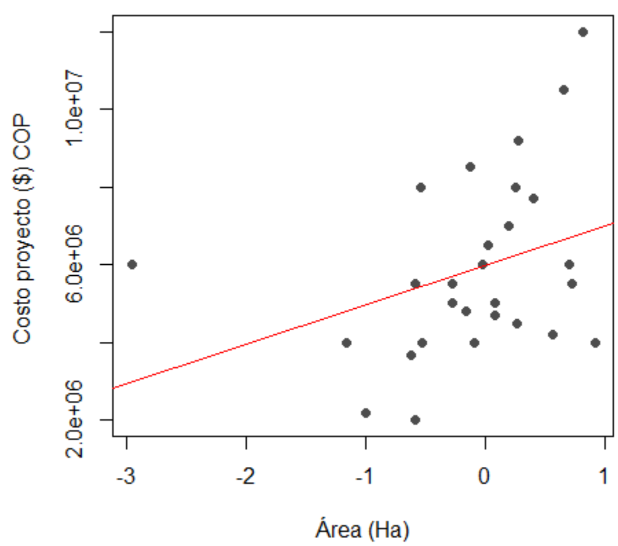


Figura 81. Gráfica de regresión lineal, modelo logarítmico, conjunto de edificios: Área (Ha)

En la (figura 81) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~LNArea))
> summary(modelo)

Call:
lm(formula = Costo ~ LNArea)

Residuals:
     Min       1Q   Median       3Q      Max
-3367150 -1489772  -684702  1443675  5207784

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5966334     428106  13.937 1.42e-13 ***
LNArea       1014552     555622   1.826   0.0794 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2243000 on 26 degrees of freedom
Multiple R-squared:  0.1137,    Adjusted R-squared:  0.07957
F-statistic: 3.334 on 1 and 26 DF,  p-value: 0.07936
```

Se aprecia que uno de los coeficientes es estadísticamente significativo, pero modelo de regresión no. Además, el coeficiente de determinación (R^2) explica aproximadamente el 7,95% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = \alpha + \beta x$

$Costo\ (\$\ COP) = 5966334 + (1014552 * Área\ (Ha))$

El modelo lineal ajustado se puede expresar como la función logarítmica:

$\hat{y} = \alpha + \beta \ln x$

$Costo\ (\$\ COP) = 5966334 + 1014552 * LnÁrea\ (Ha)$ (Ecuación 35)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table
```

```
Response: Costo
          Df     Sum Sq    Mean Sq F value  Pr(>F)
LNArea     1 1.6770e+13 1.6770e+13  3.3342 0.07936 .
Residuals 26 1.3077e+14 5.0296e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.9485, p-value = 0.1817
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 1.600711e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

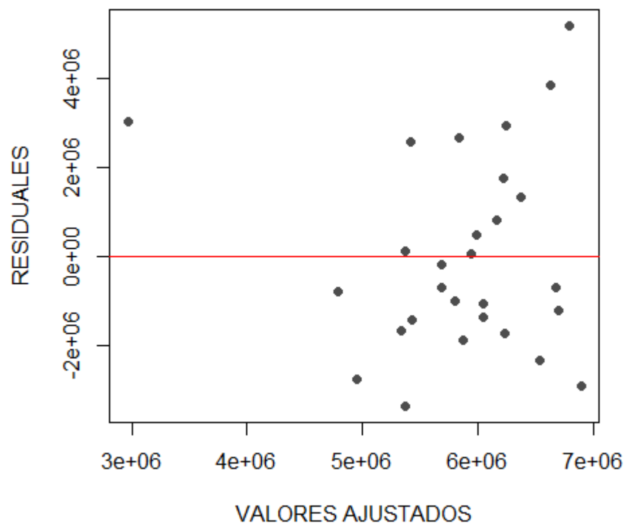


Figura 82. Gráfica de residuales, modelo logarítmico, conjunto de edificios: Área (Ha)

En la (figura 82) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo recíproco

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   4000000 2.496
2   7000000 1.211
3   2000000 0.554
4   5000000 1.081
5   6000000 0.052
6   6000000 0.972
7   9200000 1.312
8   8000000 1.285
9   8500000 0.876
10  4000000 0.589
11  2200000 0.366
12  5500000 0.554
13  4000000 0.313
14  5500000 2.049
15 12000000 2.257
16  8000000 0.580
17  4700000 1.081
18  3670000 0.540
```

```
19  7700000 1.486
20  6500000 1.026
21  4500000 1.296
22  5500000 0.761
23  6000000 2.021
24  5000000 0.761
25  4800000 0.847
26  4000000 0.911
27  4200000 1.759
28 10500000 1.921
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($)
COP",pch=16,col="grey30")
```

Se transforma la variable, creando un vector con la inversa de (x) “Área” para el modelo recíproco:

```
> Areainv<-c(1/Area)
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Areainv,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($) COP
",pch=16,col="grey30")
> abline(lm(Costo~Areainv),col="red")
```

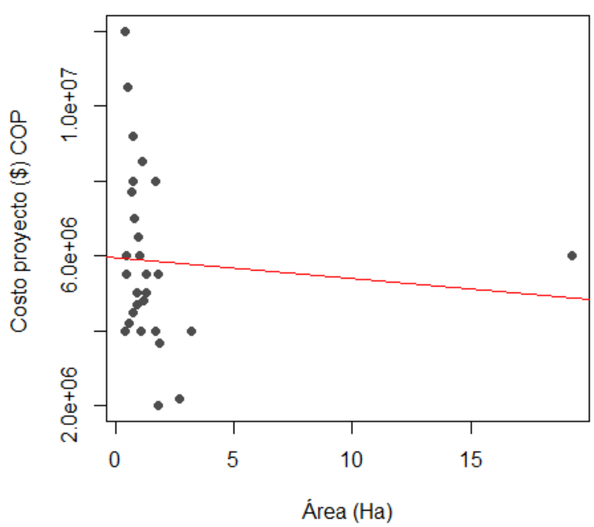


Figura 83. Gráfica de regresión lineal, modelo recíproco, conjunto de edificios: Área (Ha)

En la (figura 83) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~Areainv))
> summary(modelo)

Call:
lm(formula = Costo ~ Areainv)

Residuals:
     Min       1Q   Median       3Q      Max
-3856330 -1739128  -407453  1291750  6066647

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5958409     507590  11.739 6.82e-12 ***
Areainv       -56551     131259  -0.431     0.67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2374000 on 26 degrees of freedom
Multiple R-squared:  0.007089,  Adjusted R-squared:  -0.0311
F-statistic: 0.1856 on 1 and 26 DF,  p-value: 0.6701
```

Se aprecia que uno de los coeficientes es estadísticamente significativo, pero modelo de regresión no. Además, el coeficiente de determinación (R^2) explica aproximadamente el -3,11% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = \alpha + \beta x$$

$$Costo\ (\$\ COP) = 5958409 + (-56551 * Área\ (Ha))$$

El modelo lineal ajustado se puede expresar como la función Reciproca:

$$\hat{y} = \alpha + \beta \left(\frac{1}{x}\right)$$

$$Costo\ (\$\ COP) = 5958409 - 56551 \left(\frac{1}{Área\ (Ha)}\right) \qquad \text{(Ecuación 36)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table
```

```
Response: Costo
          Df     Sum Sq    Mean Sq F value Pr(>F)
Areainv    1 1.0459e+12 1.0459e+12  0.1856 0.6701
Residuals 26 1.4649e+14 5.6343e+12
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.95135, p-value = 0.2142
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 5.198737e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

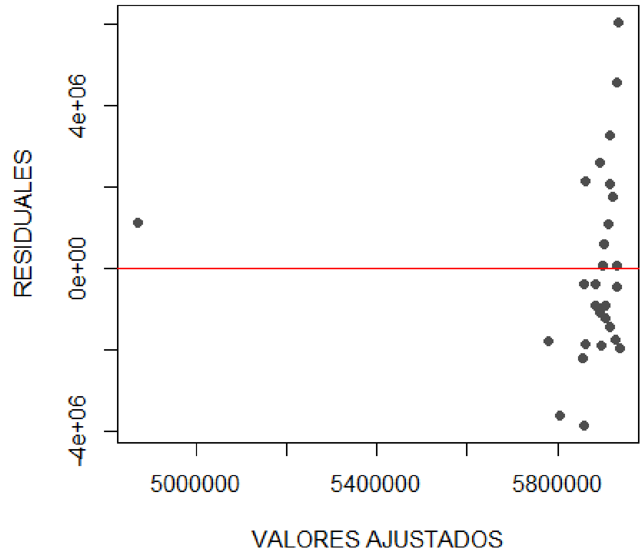


Figura 84. Gráfica de residuales, modelo recíproco, conjunto de edificios: Área (Ha)

En la (figura 84) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo Polinómico</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo  Area
1   4000000 2.496
2   7000000 1.211
3   2000000 0.554
4   5000000 1.081
5   6000000 0.052
6   6000000 0.972
7   9200000 1.312
8   8000000 1.285
9   8500000 0.876
10  4000000 0.589
11  2200000 0.366
12  5500000 0.554
13  4000000 0.313
14  5500000 2.049
15 12000000 2.257
16  8000000 0.580
17  4700000 1.081
```

```
18  3670000 0.540
19  7700000 1.486
20  6500000 1.026
21  4500000 1.296
22  5500000 0.761
23  6000000 2.021
24  5000000 0.761
25  4800000 0.847
26  4000000 0.911
27  4200000 1.759
28 10500000 1.921
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($)
COP",pch=16,col="grey30")
```

Se construye el modelo de regresión utilizando la función poly (se ajusta polinomio de grado dos):

```
> fit2 <- lm(Costo ~ poly(Area, 2, raw=TRUE))
```

Se revisan los resultados con la instrucción summary:

```
> summary(fit2)

Call:
lm(formula = Costo ~ poly(Area, 2, raw = TRUE))

Residuals:
     Min       1Q   Median       3Q      Max
-3209553 -1473977  -312201  1111471  4760540

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                 3168704    1501061   2.111   0.0449 *
poly(Area, 2, raw = TRUE)1  3547677    2650317   1.339   0.1928
poly(Area, 2, raw = TRUE)2  -772735     999331  -0.773   0.4466
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2181000 on 25 degrees of freedom
Multiple R-squared:  0.194,     Adjusted R-squared:  0.1296
F-statistic: 3.009 on 2 and 25 DF,  p-value: 0.06744
```

Se aprecia que uno de los coeficientes es estadísticamente significativo, pero modelo de regresión no. Además, el coeficiente de determinación (R^2) explica aproximadamente el 12,96% de la variabilidad observada.

El modelo polinomial de regresión estimado es:

$$\hat{y} = \beta + \beta_1 x - \beta_2 x^2$$

$$Precio\ (\$\ COP) = 3168704 + \left(3547677 * Área\ (Ha)\right) - \left(772735 * Área\ (Ha)^2\right) \quad (Ecuación\ 37)$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(fit2)
Analysis of Variance Table

Response: Costo
                          Df     Sum Sq    Mean Sq F value  Pr(>F)
poly(Area, 2, raw = TRUE)  2 2.8628e+13 1.4314e+13  3.0094 0.06744 .
Residuals                 25 1.1891e+14 4.7564e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(fit2$residuals)

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.95447, p-value = 0.256
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(fit2$residuals)
[1] 6.244656e-12
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
	> residuales<-fit2$residuals
	> valores.ajustados<-fit2$fitted
	> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
	> abline(h=0,col="red")
```

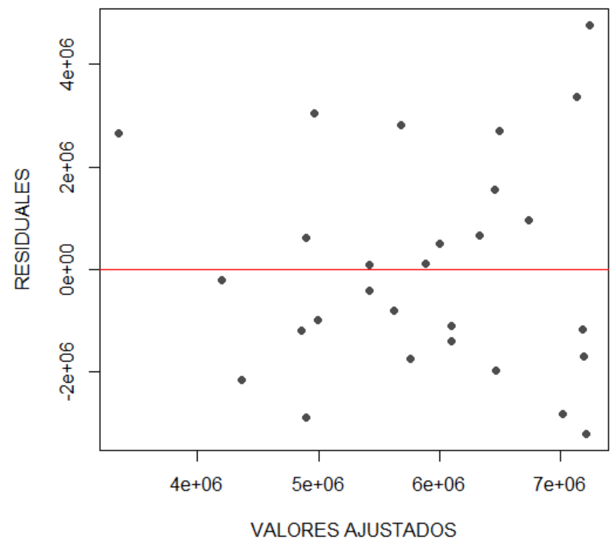


Figura 85. Gráfica de residuales, modelo polinómico, conjunto de edificios: Área (Ha)

En la (figura 85) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

De acuerdo con la realización de las ecuaciones metodológicas con los modelos funcionales de regresión lineal mediante el software de programación R, se efectúa una tabla resumen donde se encuentran cada uno de los resultados obtenidos en los diferentes modelos además de tener en cuenta los coeficientes de R^2, p-value y el error estándar de regresión en la clasificación conjunto de edificios multifamiliar con la variable área (Ha).

Tabla 12. Resultados, conjunto de edificaciones multifamiliares: Área (Ha)

Modelo funcional	Ecuación. Regresión Lineal	R-squared	p-value	Error estándar $\hat{\sigma}$
Lineal	(32) $Costo\ (\$\ COP) = 4125410 + 1565349 * Área\ (Ha)$	0,1430	0,02685	4,6829e+12
Potencial	(33) $Costo\ (\$\ COP) = 5519898 * Área\ (Ha)^{0,17413}$	0,07346	0,08808	0,15730
Exponencial	(34) $Costo\ (\$\ COP) = 4045031{,}921 * e^{(0,2641 * Área\ (Ha))}$	0,1278	0,03487	0,14807
Logarítmica	(35) $Costo\ (\$\ COP) = 5966334 + 1014552 * LnÁrea\ (Ha)$	0,07957	0,07936	5,0296e+12
Recíproca	(36) $Costo\ (\$\ COP) = 5958409 - 56551\left(\frac{1}{Área\ (Ha)}\right)$	-0,0311	0,6701	5,6343e+12
Polinómica	(37) $Costo\ (\$\ COP) = 3168704 + (3547677 * Área\ (Ha)) - (772735 * Área\ (Ha)^2)$	0,1296	0,06744	4,7564e+12

Una vez construidos los modelos de regresión lineal para la variable explicativa "Área (Ha)" en la clasificación "conjunto de edificios", se pueden comparar los resultados obtenidos para elegir la ecuación que presenta mayor significancia estadística; en este caso se encuentra que el modelo funcional lineal muestra un coeficiente de determinación R^2 de 0,1430, valor superior a los obtenidos en los otros modelos empleados y el que más se acerca a uno (1); también se observa en el mismo modelo que el p- value es de 0,02685 el cual presenta un valor por debajo del 5%, indicando su significancia estadística.

Los resultados también demuestran que el modelo funcional que menos se ajusta es el recíproco, dado que el coeficiente de determinación es el más bajo en comparación a los obtenidos en los otros modelos analizados y el que más se aleja de uno (1).

Teniendo en cuenta estos resultados analizados, la *(ecuación 32)* del modelo funcional lineal es la que representa mayor alcance y menor error estándar de la regresión para la variable dependiente *Costo ($ COP)* por lo cual será empleada para esta clasificación.

4.2.2.3. Conjunto de edificios: Longitud de tubería (m)

Modelo lineal

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Longitud
1   4000000  1270.01
2   7000000   657.60
3   2000000   181.01
4   5000000  1412.35
5   6000000   857.50
6   6000000  1187.30
7   9200000  1246.46
8   8000000  1093.15
9   8500000  1142.56
10  4000000  1350.91
11  2200000   204.83
12  5500000   538.81
13  4000000   507.58
14  5500000  1963.24
15 12000000  2282.76
16  8000000  1004.01
17  4700000   976.05
18  3670000   471.75
19  7700000  2068.55
20  6500000  1220.42
21  4500000  1240.34
22  5500000   657.60
23  6000000  1970.82
24  5000000   657.60
25  4800000   709.40
26  4000000   720.80
27  4200000  1860.24
28 10500000  3166.49
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

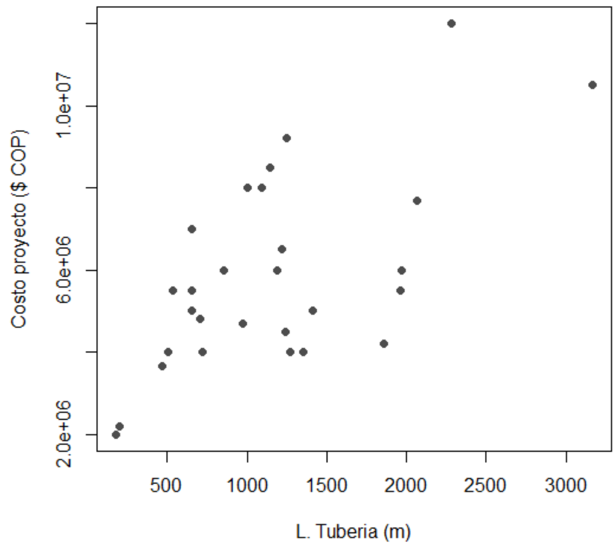


Figura 86. Gráfica de dispersión, modelo lineal, conjunto de edificios: Longitud de tubería (m)

En la (figura 86) se observan datos en los que aumenta la longitud de tubería, pero no el costo del proyecto, y también viceversa, es decir, que no son directamente proporcionales; algunos costos son elevados para el área que representan.

```
> abline(lm(Costo~Longitud),col="red")
```

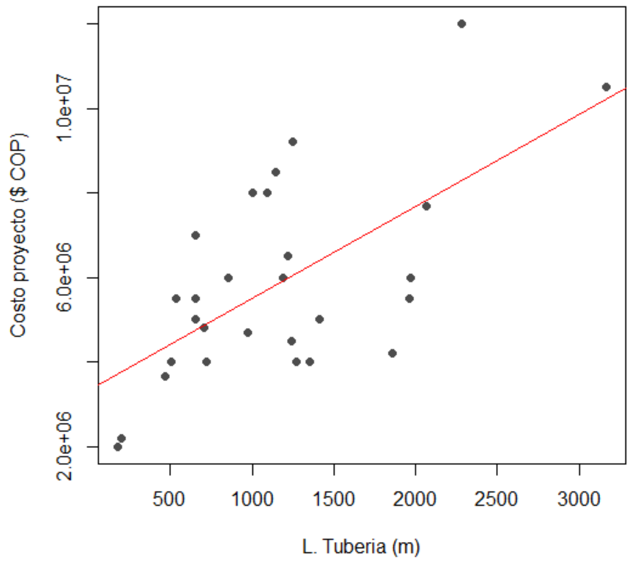


Figura 87. Gráfica de regresión lineal, modelo lineal, conjunto de edificios: Longitud de tubería (m)

En la (figura 87) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Se construye el modelo de regresión utilizando la función lm:

```
> modelo<-lm(Costo~Longitud)
> summary(modelo)

Call:
lm(formula = Costo ~ Longitud)

Residuals:
     Min       1Q   Median       3Q      Max
-3168279 -1531848   -93229   861079  3712696

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3322065.6   700489.4   4.742 6.63e-05 ***
Longitud       2175.1      521.6   4.170    3e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1844000 on 26 degrees of freedom
Multiple R-squared:  0.4008,	Adjusted R-squared:  0.3777
F-statistic: 17.39 on 1 and 26 DF,  p-value: 0.0002996
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 37,77% de la variabilidad observada.

El modelo de regresión estimado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$Costo\ (\$\ COP) = 3322065{,}6 + 2175{,}1 * Longitud\ (m) \qquad \textit{(Ecuación 38)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido:

```
> anova(modelo)

Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value    Pr(>F)
Longitud   1 5.9131e+13 5.9131e+13   17.39 0.0002996 ***
Residuals 26 8.8408e+13 3.4003e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96065, p-value = 0.3612
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 3.055902e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted.

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

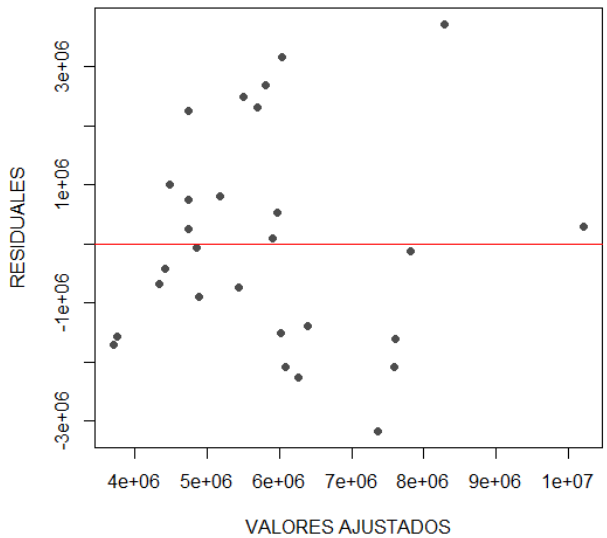


Figura 88. Gráfica de residuales, modelo lineal, conjunto de edificios: Longitud de tubería (m)

En la (figura 88) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo potencial</u>

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
     Costo Longitud
1  4000000  1270.01
2  7000000   657.60
```

```
3   2000000   181.01
4   5000000  1412.35
5   6000000   857.50
6   6000000  1187.30
7   9200000  1246.46
8   8000000  1093.15
9   8500000  1142.56
10  4000000  1350.91
11  2200000   204.83
12  5500000   538.81
13  4000000   507.58
14  5500000  1963.24
15 12000000  2282.76
16  8000000  1004.01
17  4700000   976.05
18  3670000   471.75
19  7700000  2068.55
20  6500000  1220.42
21  4500000  1240.34
22  5500000   657.60
23  6000000  1970.82
24  5000000   657.60
25  4800000   709.40
26  4000000   720.80
27  4200000  1860.24
28 10500000  3166.49
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” y (x) “Longitud” para el modelo potencial:

```
> LNCosto<-c(log(Costo))
> LNLongitud<-c(log(Longitud))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNLongitud,LNCosto,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
> abline(lm(LNCosto~LNLongitud),col="red")
```

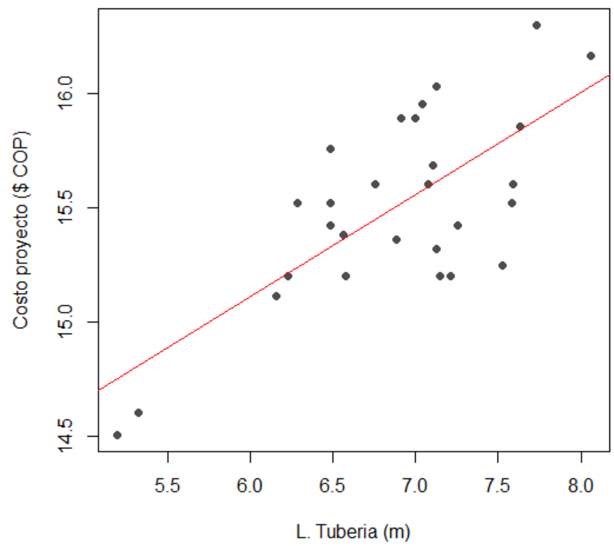


Figura 89. Gráfica de regresión lineal, modelo potencial, conjunto de edificios: Longitud de tubería (m)

En la (figura 89) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(LNCosto~LNLongitud)
> summary(modelo)

Call:
lm(formula = LNCosto ~ LNLongitud)

Residuals:
     Min       1Q   Median       3Q      Max
-0.54469 -0.22221  0.01316  0.21086  0.42981

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.43830    0.58547  21.245  < 2e-16 ***
LNLongitud   0.44591    0.08475   5.261 1.69e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.2922 on 26 degrees of freedom
Multiple R-squared:  0.5157,    Adjusted R-squared:  0.497
F-statistic: 27.68 on 1 and 26 DF,  p-value: 1.688e-05
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 49,7 % de la variabilidad observada.

El modelo de regresión estimado es:
$y^* = ln\,\alpha + \beta x^*$
$LnCosto\ (\$\ COP) = \ln\,(12{,}43830) + 0{,}44591 * Longitud(m)$

El modelo lineal ajustado se puede expresar como la función potencial:
$\hat{y} = \alpha x^{\beta}$
$Costo\ (\$\ COP) = 252281{,}2965 * Longitud\ (m)^{0{,}44591}$ (Ecuación 39)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
           Df Sum Sq Mean Sq F value    Pr(>F)
LNLongitud  1 2.3637 2.36370  27.682 1.688e-05 ***
Residuals  26 2.2201 0.08539
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

 Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.95608, p-value = 0.2805
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean (modelo$residuals)
```

```
[1] 1.437875e-17
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

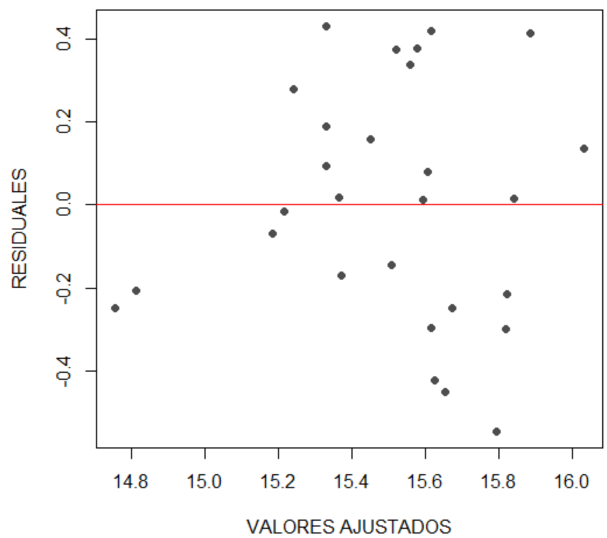


Figura 90. Gráfica de residuales, modelo potencial, conjunto de edificios: Longitud de tubería (m)

En la (figura 90) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo exponencial

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Longitud
1   4000000  1270.01
2   7000000   657.60
3   2000000   181.01
4   5000000  1412.35
5   6000000   857.50
6   6000000  1187.30
7   9200000  1246.46
8   8000000  1093.15
9   8500000  1142.56
10  4000000  1350.91
11  2200000   204.83
12  5500000   538.81
13  4000000   507.58
14  5500000  1963.24
15 12000000  2282.76
16  8000000  1004.01
17  4700000   976.05
18  3670000   471.75
19  7700000  2068.55
20  6500000  1220.42
21  4500000  1240.34
22  5500000   657.60
23  6000000  1970.82
24  5000000   657.60
25  4800000   709.40
26  4000000   720.80
27  4200000  1860.24
28 10500000  3166.49
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” para el modelo exponencial:

```
> LNCosto<-c(log(Costo))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Longitud,LNCosto,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
> abline(lm(LNCosto~Longitud),col="red")
```

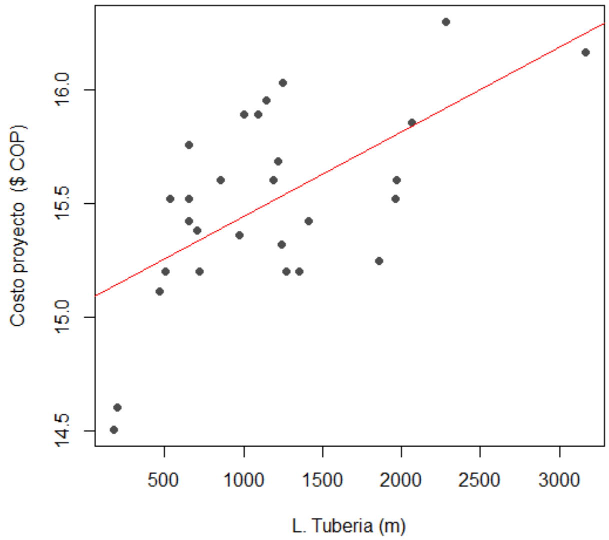


Figura 91. Gráfica de regresión lineal, modelo exponencial, conjunto de edificios: Longitud de tubería (m)

En la (figura 91) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(LNCosto~Longitud)
> summary(modelo)

Call:
lm(formula = LNCosto ~ Longitud)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.62822 -0.20270 -0.02171  0.22540  0.49930 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.507e+01  1.254e-01 120.130  < 2e-16 ***
Longitud    3.741e-04  9.340e-05   4.005 0.000462 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3302 on 26 degrees of freedom
Multiple R-squared:  0.3815,    Adjusted R-squared:  0.3577
```

```
F-statistic: 16.04 on 1 and 26 DF,  p-value: 0.0004618
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 35,77% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = ln\, \alpha + \beta x$

$LnCosto\ (\$\ COP) = ln(1{,}507 * 10^{01}) + (3{,}741 * 10^{-04}) * Longitud\ (m)$

El modelo lineal ajustado se puede expresar como la función exponencial:

$\hat{y} = \alpha e^{\beta x}$

$Costo\ (\$\ COP) = 3506047{,}877 * \ e^{(3{,}741 * 10^{-4} * Longitud\ (m))}$ (Ecuación 40)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)

Analysis of Variance Table

Response: LNCosto
          Df Sum Sq Mean Sq F value    Pr(>F)
Longitud   1 1.7487 1.74874  16.038 0.0004618 ***
Residuals 26 2.8351 0.10904
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96313, p-value = 0.4125
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 1.611565e-17
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

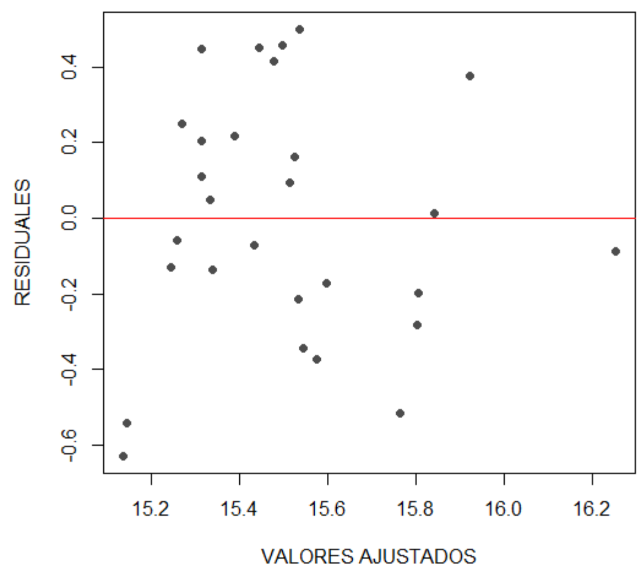


Figura 92. Gráfica de residuales, modelo exponencial, conjunto de edificios: Longitud de tubería (m)

En la (figura 92) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo logarítmico

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Longitud
1   4000000  1270.01
2   7000000   657.60
3   2000000   181.01
4   5000000  1412.35
5   6000000   857.50
6   6000000  1187.30
7   9200000  1246.46
8   8000000  1093.15
9   8500000  1142.56
10  4000000  1350.91
11  2200000   204.83
12  5500000   538.81
13  4000000   507.58
14  5500000  1963.24
15 12000000  2282.76
16  8000000  1004.01
17  4700000   976.05
18  3670000   471.75
19  7700000  2068.55
20  6500000  1220.42
21  4500000  1240.34
22  5500000   657.60
23  6000000  1970.82
24  5000000   657.60
25  4800000   709.40
26  4000000   720.80
27  4200000  1860.24
28 10500000  3166.49
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (x) “Longitud” para el modelo logarítmico:

```
> LNLongitud<-c(log(Longitud))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNLongitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
> abline(lm(Costo~LNLongitud),col="red")
```

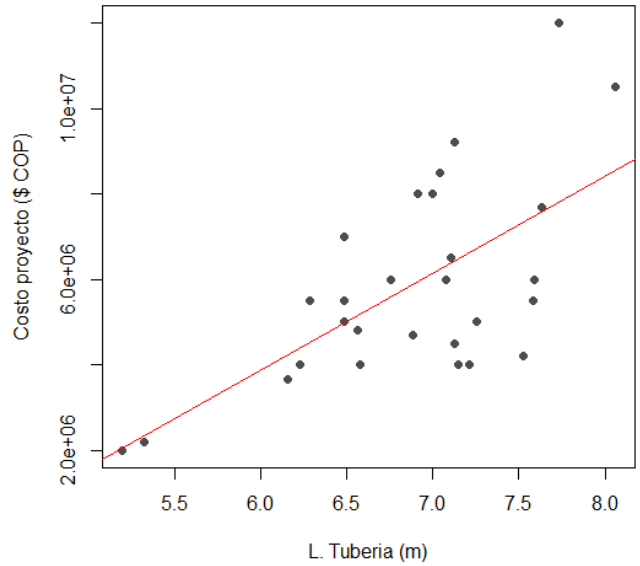


Figura 93. Gráfica de regresión lineal, modelo logarítmico, conjunto de edificios: Longitud de tubería (m)

En la (figura 93) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(Costo~LNLongitud)
> summary(modelo)

Call:
lm(formula = Costo ~ LNLongitud)

Residuals:
     Min       1Q   Median       3Q      Max 
-3132725 -1252892   -89152  1201204  4203104 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -9740280    3652407  -2.667 0.012998 *  
LNLongitud   2267795     528712   4.289 0.000219 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1823000 on 26 degrees of freedom
```

```
Multiple R-squared:  0.4144,    Adjusted R-squared:  0.3919
F-statistic:  18.4 on 1 and 26 DF,  p-value: 0.0002191
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 39,19% de la variabilidad observada.

El modelo de regresión estimado:

$$y^{*} = \alpha + \beta x$$

$$Costo\ (\$\ COP) = -9740280 + (2267795 * Longitud\ (m))$$

El modelo lineal ajustado se puede expresar como la función logarítmica:

$$\hat{y} = \alpha + \beta\, ln\, x$$

$$Costo(\$\ COP) = -9740280 + (2267795 * LnLongitud\ (m)) \quad \text{(Ecuación 41)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)

Analysis of Variance Table

Response: Costo
           Df     Sum Sq    Mean Sq F value    Pr(>F)
LNLongitud  1 6.1138e+13 6.1138e+13  18.398 0.0002191 ***
Residuals  26 8.6401e+13 3.3231e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.97356, p-value = 0.6784
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 1.309672e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

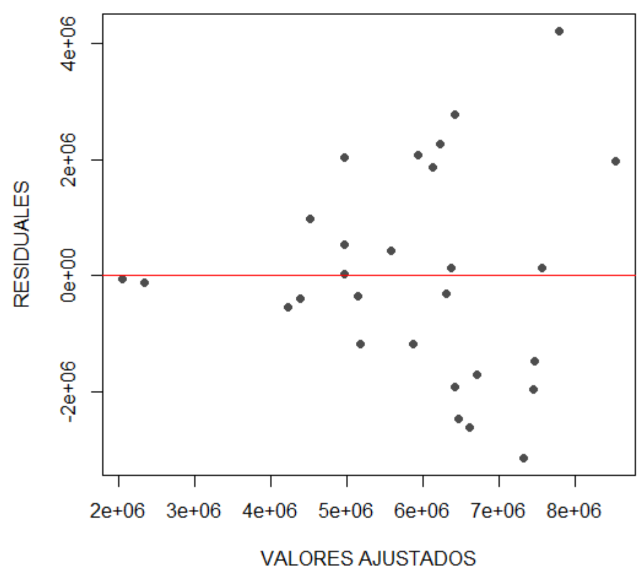


Figura 94. Gráfica de residuales, modelo logarítmico, conjunto de edificios: Longitud de tubería (m)

En la (figura 94) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo recíproco</u>

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Longitud
1   4000000  1270.01
2   7000000   657.60
3   2000000   181.01
4   5000000  1412.35
5   6000000   857.50
6   6000000  1187.30
7   9200000  1246.46
8   8000000  1093.15
9   8500000  1142.56
10  4000000  1350.91
11  2200000   204.83
12  5500000   538.81
13  4000000   507.58
14  5500000  1963.24
15 12000000  2282.76
16  8000000  1004.01
17  4700000   976.05
18  3670000   471.75
19  7700000  2068.55
20  6500000  1220.42
21  4500000  1240.34
22  5500000   657.60
23  6000000  1970.82
24  5000000   657.60
25  4800000   709.40
26  4000000   720.80
27  4200000  1860.24
28 10500000  3166.49
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se transforma la variable, creando un vector con la inversa de (x) “Longitud” para el modelo recíproco:

```
> Longitudinv<-c(1/Longitud)
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Longitudinv,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto ($
COP) ",pch=16,col="grey30")
> abline(lm(Costo~Longitudinv),col="red")
```

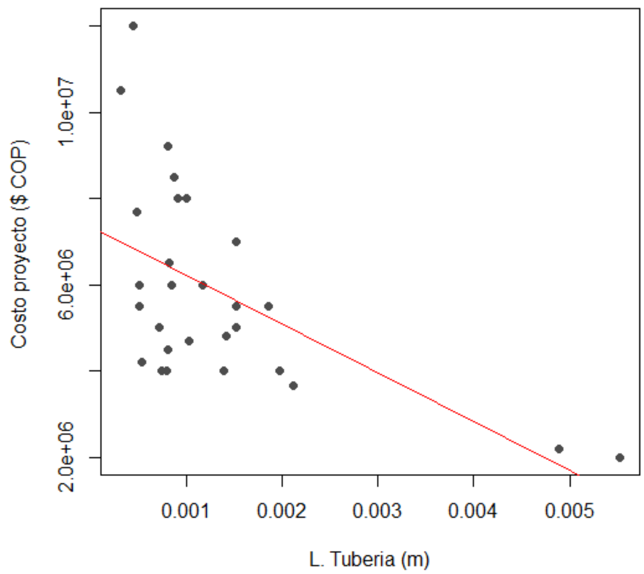


Figura 95. Gráfica de regresión lineal, modelo recíproco, conjunto de edificios: Longitud de tubería (m)

En la (figura 95) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~Longitudinv))
> summary(modelo)

Call:
lm(formula = Costo ~ Longitudinv)

Residuals:
     Min        1Q    Median       3Q      Max
-2543899 -1333507  -264124  1023730  5143332

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.353e+06  5.503e+05  13.363 3.72e-13 ***
Longitudinv -1.133e+09  3.109e+08  -3.645  0.00117 **
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1938000 on 26 degrees of freedom
Multiple R-squared:  0.3382,    Adjusted R-squared:  0.3128
F-statistic: 13.29 on 1 and 26 DF,  p-value: 0.00117
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 31,28% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = \alpha + \beta x$$

$$Costo\ (\$\ COP) = 7{,}353 * 10^6 + (-1{,}133 * 10^9 * Longitud\ (m))$$

El modelo lineal ajustado se puede expresar como la función Recíproca:

$$\hat{y} = \alpha + \beta\left(\frac{1}{x}\right)$$

$$Costo\ (\$\ COP) = 7{,}353 * 10^6 - 1{,}133 * 10^9\left(\frac{1}{Longitud\ (m)}\right) \qquad \text{(Ecuación 42)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)

Analysis of Variance Table

Response: Costo
            Df     Sum Sq    Mean Sq F value  Pr(>F)
Longitudinv  1 4.9902e+13 4.9902e+13  13.289 0.00117 **
Residuals   26 9.7637e+13 3.7553e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

       Shapiro-Wilk normality test

data:  modelo$residuals
```

```
W = 0.94344, p-value = 0.1352
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 1.14304e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

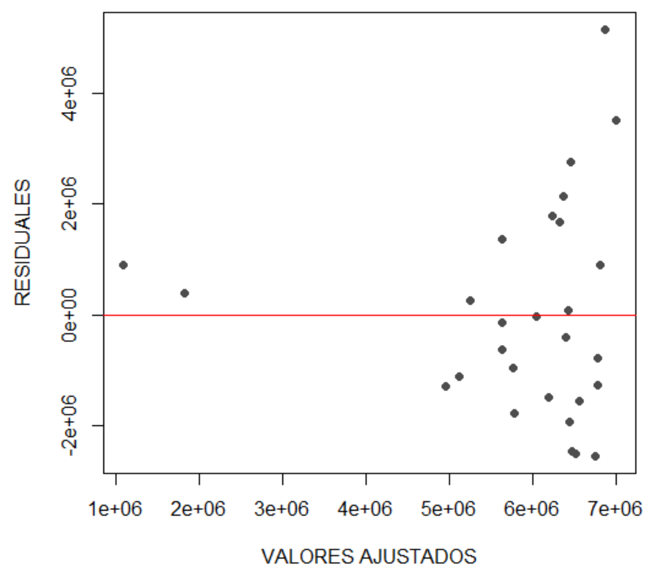


Figura 96. Gráfica de residuales, modelo recíproco, conjunto de edificios: Longitud de tubería (m)

En la (figura 96) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo polinómico

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Longitud
1   4000000  1270.01
2   7000000   657.60
3   2000000   181.01
4   5000000  1412.35
5   6000000   857.50
6   6000000  1187.30
7   9200000  1246.46
8   8000000  1093.15
9   8500000  1142.56
10  4000000  1350.91
11  2200000   204.83
12  5500000   538.81
13  4000000   507.58
14  5500000  1963.24
15 12000000  2282.76
16  8000000  1004.01
17  4700000   976.05
18  3670000   471.75
19  7700000  2068.55
20  6500000  1220.42
21  4500000  1240.34
22  5500000   657.60
23  6000000  1970.82
24  5000000   657.60
25  4800000   709.40
26  4000000   720.80
27  4200000  1860.24
28 10500000  3166.49
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Longitud,Costo,xlab="L. Tuberia (m)",ylab="Costo proyecto($
COP)",pch=16,col="grey30")
```

Se construye el modelo de regresión utilizando la función poly (se ajusta polinomio de grado dos):

```
> fit2 <- lm(Costo ~ poly(Longitud, 2, raw=TRUE))
```

Se revisan los resultados con la instrucción summary:

```
> summary(fit2)

Call:
lm(formula = Costo ~ poly(Longitud, 2, raw = TRUE))

Residuals:
     Min       1Q   Median       3Q      Max
-3199909 -1502436   -99382   859386  3717743

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                     3.196e+06  1.189e+06   2.687   0.0126 *
poly(Longitud, 2, raw = TRUE)1  2.400e+03  1.776e+03   1.352   0.1886
poly(Longitud, 2, raw = TRUE)2 -7.518e-02  5.669e-01  -0.133   0.8956
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1880000 on 25 degrees of freedom
Multiple R-squared:  0.4012,    Adjusted R-squared:  0.3533
F-statistic: 8.375 on 2 and 25 DF,  p-value: 0.001644
```

Se aprecia que uno de los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 35,33% de la variabilidad observada.

El modelo polinomial de regresión estimado es:
$\hat{y} = \beta + \beta_1 x - \beta_2 x^2$
$Costo\ (\$\ COP) = 3{,}196 * 10^6 + (2{,}400 * 10^3) * (Longitud\ (m)) - (-7{,}518 * 10^{-2}) * (Longitud\ (m)^2)$ (Ecuación 43)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(fit2)
Analysis of Variance Table

Response: Costo
                              Df     Sum Sq    Mean Sq F value   Pr(>F)
poly(Longitud, 2, raw = TRUE)  2 5.9193e+13 2.9597e+13  8.3752 0.001644 **
Residuals                     25 8.8346e+13 3.5338e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(fit2$residuals)

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.9633, p-value = 0.4163
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(fit2$residuals)
[1] 2.744086e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-fit2$residuals
> valores.ajustados<-fit2$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

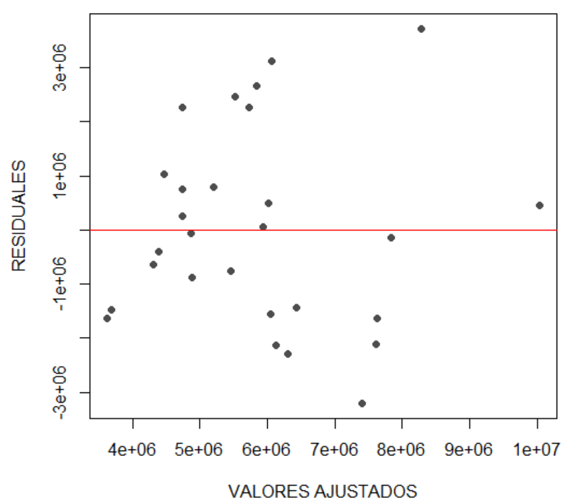


Figura 97. Gráfica de residuales, modelo polinómico, conjunto de edificios: Longitud de tubería (m)

En la (figura 97) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

De acuerdo con la realización de las ecuaciones metodológicas con los modelos funcionales de regresión lineal mediante el software de programación R, se efectúa una tabla resumen donde se encuentran cada uno de los resultados obtenidos en los diferentes modelos además de tener en cuenta los coeficientes de R^2, p-value y el error estándar de regresión en la clasificación conjunto de edificios multifamiliar con la variable longitud de tubería (m).

Tabla 13. Resultados, conjunto de edificios multifamiliares: Longitud de tubería (m)

Modelo funcional	Ecuación. Regresión Lineal	R-squared	p-value	Error estándar $\hat{\sigma}$
Lineal	(38) $Costo\ (\$\ COP) = 3322065{,}6 + 2175{,}1 * Longitud\ (m)$	0,3777	0,0002996	3,4003e+12
Potencial	(39) $Costo\ (\$\ COP) = 252281{,}2965 * Longitud\ (m)^{0{,}44591}$	0,4970	1,688e-05	0,08539
Exponencial	(40) $Costo\ (\$\ COP) = 3506047{,}877 * \ e^{(3{,}741 * 10^{-4} * Longitud\ (m))}$	0,3577	0,0004618	0,10904
Logarítmica	(41) $Costo(\$\ COP) = -9740280 + (2267795 * LnLongitud\ (m))$	0,3919	0,0002191	3,3132e+12
Recíproca	(42) $Costo\ (\$\ COP) = 7{,}353 * 10^6 - 1{,}133 * 10^9 \left(\frac{1}{Longitud\ (m)}\right)$	0,3128	0,00117	3,7553e+12
Polinómica	(43) $Costo\ (\$\ COP) = 3{,}196 * 10^6 + (2{,}400 * 10^3) * (Longitud\ (m)) - (-7{,}518 * 10^{-2}) * (Longitud\ (m)^2)$	0,3533	0,001644	3,5338e+12

Una vez construidos los modelos de regresión lineal para la variable explicativa “Longitud de tubería (m)” en la clasificación “conjunto de edificios”, se pueden comparar los resultados obtenidos para elegir la ecuación que presenta mayor significancia estadística; en este caso se encuentra que el modelo funcional potencial muestra un coeficiente de determinación R^2

de 0,4970, valor superior a los obtenidos en los otros modelos empleados y el que más se acerca a uno (1); también se observa en el mismo modelo que el p- value es de 1,688e-05 el cual presenta un valor por debajo del 5%, indicando su significancia estadística.

Los resultados también demuestran que el modelo funcional que menos se ajusta es el recíproco, dado que el coeficiente de determinación es el más bajo en comparación a los obtenidos en los otros modelos analizados y el que más se aleja de uno (1).

Teniendo en cuenta estos resultados analizados, la *(ecuación 39)* del modelo funcional potencial es la que representa mayor alcance y menor error estándar de la regresión para la variable dependiente *Costo ($ COP)* por lo cual será empleada para esta clasificación.

4.2.2.4. Conjunto de edificios: Número de apartamentos (und)

Modelo lineal

Se importan los datos que se encuentan en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Apartamentos
1   4000000          640
2   7000000          240
3   2000000           96
4   5000000          200
5   6000000          168
6   6000000          160
7   9200000          400
8   8000000          288
9   8500000          160
10  4000000          180
11  2200000           80
12  5500000          180
13  4000000           64
14  5500000          328
15 12000000          480
16  8000000          196
17  4700000          200
18  3670000          132
19  7700000          320
20  6500000          288
21  4500000          360
```

```
22  5500000         240
23  6000000         440
24  5000000         240
25  4800000         352
26  4000000         240
27  4200000         340
28 10500000         300
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Apartamentos,Costo,xlab="N. Apartamentos (und)",ylab="Costo
proyecto($)",pch=16,col="grey30")
```

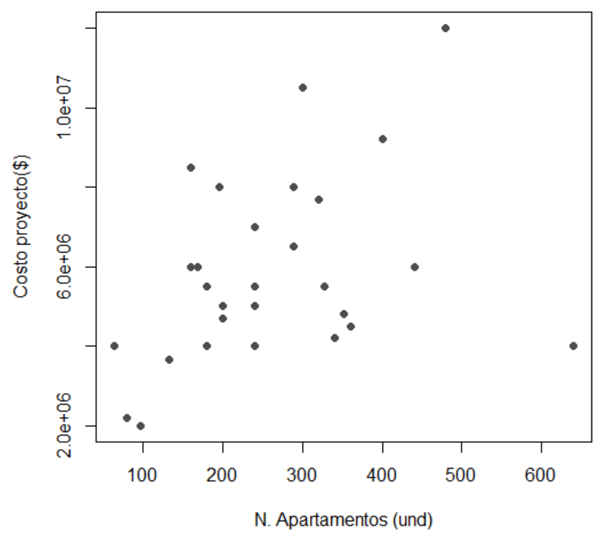


Figura 98. Gráfica de dispersión, conjunto de edificios: Número de apartamentos (und)

En la (figura 98) se observan datos en los que aumenta el número de viviendas, pero no el costo del proyecto, y también viceversa, es decir, que no son directamente proporcionales; algunos costos son elevados para el área que representan.

```
> abline(lm(Costo~Apartamentos),col="red")
```

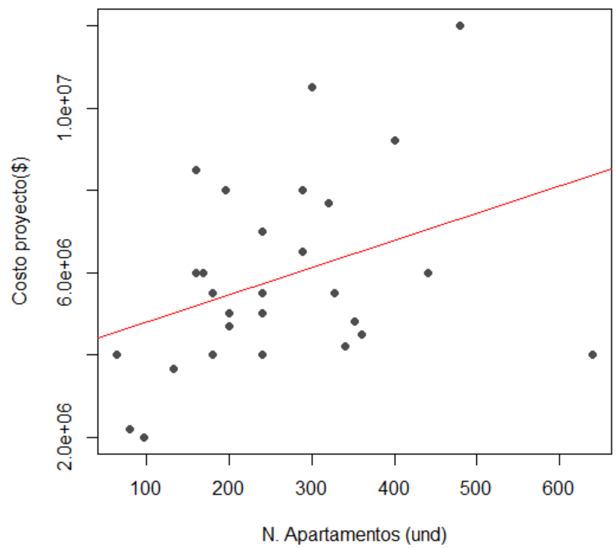


Figura 99. Gráfica de regresión lineal, modelo lineal, conjunto de edificios: Número de apartamentos (und)

En la (figura 99) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Se construye el modelo de regresión utilizando la función *lm*:

```
> modelo<-lm(Costo~Apartamentos)
> summary(modelo)

Call:
lm(formula = Costo ~ Apartamentos)

Residuals:
     Min       1Q   Median       3Q      Max
-4357564 -1414026  -503376  1326475  4698873

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   4131815     960601   4.301 0.000212 ***
Apartamentos     6603       3310   1.995 0.056623 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2218000 on 26 degrees of freedom
Multiple R-squared:  0.1328,    Adjusted R-squared:  0.0994
F-statistic:  3.98 on 1 and 26 DF,  p-value: 0.05662
```

Se aprecia que uno de los coeficientes es estadísticamente significativo, pero modelo de regresión no. Además, el coeficiente de determinación (R^2) explica aproximadamente el 9,94% de la variabilidad observada.

El modelo de regresión estimado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$Costo\ (\$\ COP) = 4131815 + 6603 * Apartamentos\ (und) \qquad \textit{(Ecuación 44)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido:

```
> anova(modelo)

Analysis of Variance Table

Response: Costo
             Df     Sum Sq    Mean Sq F value  Pr(>F)
Apartamentos  1 1.9587e+13 1.9587e+13    3.98 0.05662 .
Residuals    26 1.2795e+14 4.9212e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

      Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.97253, p-value = 0.6497
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] -6.241407e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

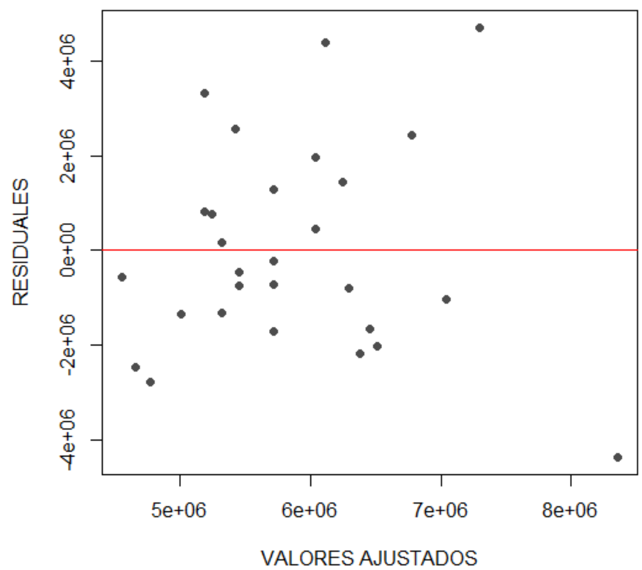


Figura 100. Gráfica de residuales, modelo lineal, conjunto de edificios: Número de apartamentos (und)

En la (figura 100) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Potencial

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
    Costo Apartamentos
1 4000000          640
2 7000000          240
3 2000000           96
4 5000000          200
5 6000000          168
```

```
6   6000000          160
7   9200000          400
8   8000000          288
9   8500000          160
10  4000000          180
11  2200000           80
12  5500000          180
13  4000000           64
14  5500000          328
15 12000000          480
16  8000000          196
17  4700000          200
18  3670000          132
19  7700000          320
20  6500000          288
21  4500000          360
22  5500000          240
23  6000000          440
24  5000000          240
25  4800000          352
26  4000000          240
27  4200000          340
28 10500000          300
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Apartamentos,Costo,xlab="N. Apartamentos (und)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” y (x) “Apartamentos” para el modelo potencial:

```
> LNCosto<-c(log(Costo))
> LNApartamentos<-c(log(Apartamentos))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNApartamentos,LNCosto,xlab="N. Apartamentos (und)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
> abline(lm(LNCosto~LNApartamentos),col="red")
```

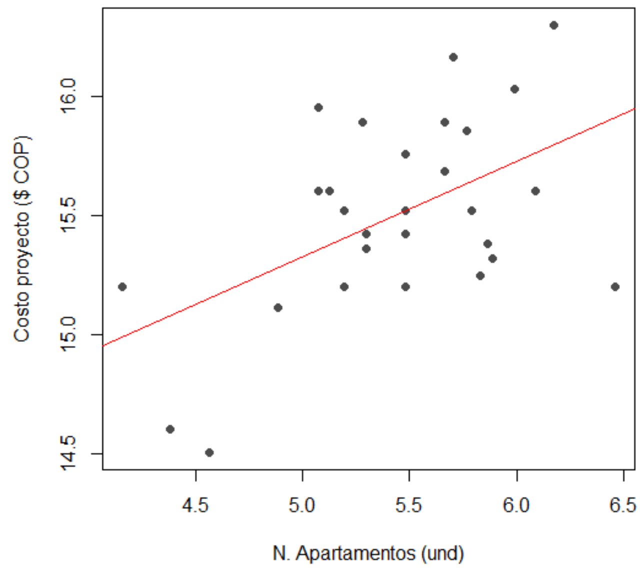


Figura 101. Gráfica de regresión lineal, modelo potencial, conjunto de edificios: Número de apartamentos (und)

En la (figura 101) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(LNCosto~LNApartamentos)
> summary(modelo)

Call:
lm(formula = LNCosto ~ LNApartamentos)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.7117 -0.2257 -0.0120  0.2424  0.5967 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)     13.3284     0.7113  18.739  < 2e-16 ***
LNApartamentos   0.4001     0.1301   3.074  0.00491 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3596 on 26 degrees of freedom
Multiple R-squared:  0.2666,    Adjusted R-squared:  0.2384 
F-statistic: 9.451 on 1 and 26 DF,  p-value: 0.004911
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 23,840% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = ln\,\alpha + \beta x^*$$

$$LnCosto\ (\$\ COP) = \ln\,(13{,}3284) + 0{,}4001 * Apartamentos\ (und)$$

El modelo lineal ajustado se puede expresar como la función potencial:

$$\hat{y} = \alpha x^{\beta}$$

$$Costo\ (\$\ COP) = 614399{,}1025 * Apartamentos\ (und)^{0{,}4001} \quad \text{(Ecuación 45)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)

Analysis of Variance Table

Response: LNCosto
               Df Sum Sq Mean Sq F value   Pr(>F)
LNApartamentos  1 1.2220  1.2220   9.451 0.004911 **
Residuals      26 3.3618  0.1293
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

      Shapiro-Wilk normality test


data:  modelo$residuals

W = 0.97613, p-value = 0.7498
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean (modelo$residuals)

[1] 7.38574e-17
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

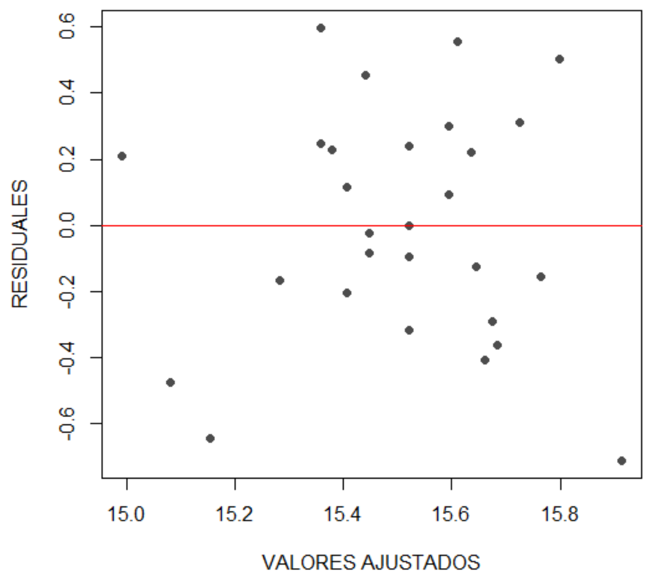


Figura 102. Gráfica de residuales, modelo potencial, conjunto de edificios: Número de apartamentos (und)

En la (figura 102) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo exponencial

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Apartamentos
1   4000000          640
2   7000000          240
3   2000000           96
4   5000000          200
5   6000000          168
6   6000000          160
7   9200000          400
8   8000000          288
9   8500000          160
10  4000000          180
11  2200000           80
12  5500000          180
13  4000000           64
14  5500000          328
15 12000000          480
16  8000000          196
17  4700000          200
18  3670000          132
19  7700000          320
20  6500000          288
21  4500000          360
22  5500000          240
23  6000000          440
24  5000000          240
25  4800000          352
26  4000000          240
27  4200000          340
28 10500000          300
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
>plot(Apartamentos,Costo,xlab="N. Apartamentos (und)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” para el modelo exponencial:

```
> LNCosto<-c(log(Costo))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
>plot(Apartamentos,LNCosto,xlab="N.    Apartamentos    (und)",ylab="Costo
proyecto ($ COP)",pch=16,col="grey30")
> abline(lm(LNCosto~Apartamentos),col="red")
```

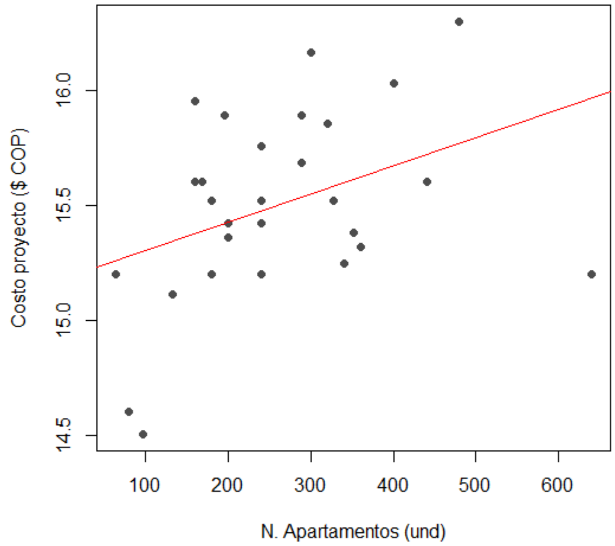


Figura 103. Gráfica de regresión lineal, modelo exponencial, conjunto de edificios: Número de apartamentos (und)

En la (figura 103) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-lm(LNCosto~Apartamentos)
> summary(modelo)

Call:
lm(formula = LNCosto ~ Apartamentos)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.79384 -0.23124 -0.02956  0.28033  0.61431 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.518e+01  1.679e-01  90.443   <2e-16 ***
Apartamentos 1.226e-03  5.785e-04   2.119   0.0438 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3877 on 26 degrees of freedom
Multiple R-squared:  0.1473,    Adjusted R-squared:  0.1145 
F-statistic: 4.491 on 1 and 26 DF,  p-value: 0.04378
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 11,45% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = ln\,\alpha + \beta x$

$LnCosto\ (\$\ COP) = ln(1{,}518 * 10^{01}) + \left((1{,}226 * 10^{-03}) * Apartamentos\ (und)\right)$

El modelo lineal ajustado se puede expresar como la función exponencial:

$\hat{y} = \alpha e^{\beta x}$

$Costo\ (\$\ COP) = 3913724{,}359 * e^{((1{,}226 * 10^{-03}) * Apartamentos\ (und))}$ (Ecuación 46)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)

Analysis of Variance Table

Response: LNCosto
             Df Sum Sq Mean Sq F value  Pr(>F)
Apartamentos  1 0.6751 0.67515   4.491 0.04378 *
Residuals    26 3.9086 0.15033
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

 Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96241, p-value = 0.397
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 6.443452e-17
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
>       plot(valores.ajustados,residuales,xlab="VALORES       AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

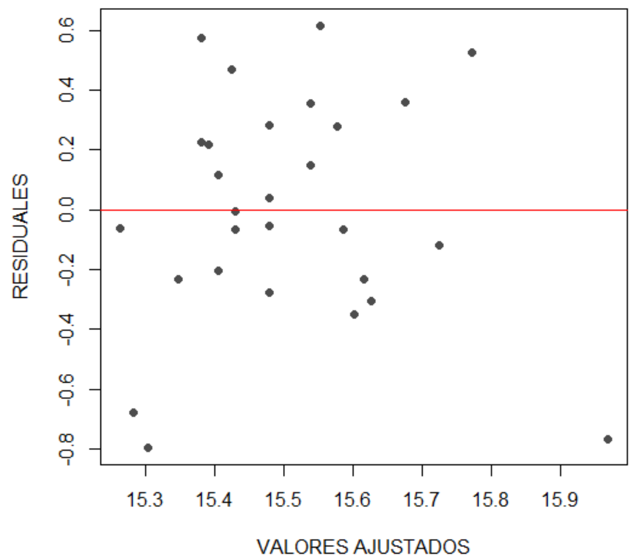


Figura 104. Gráfica de residuales, modelo exponencial, conjunto de edificios: Número de apartamentos (und)

En la (figura 104) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Logarítmico

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Apartamentos
1   4000000          640
2   7000000          240
3   2000000           96
4   5000000          200
5   6000000          168
6   6000000          160
7   9200000          400
8   8000000          288
9   8500000          160
10  4000000          180
11  2200000           80
12  5500000          180
13  4000000           64
14  5500000          328
15 12000000          480
16  8000000          196
17  4700000          200
18  3670000          132
19  7700000          320
20  6500000          288
21  4500000          360
22  5500000          240
23  6000000          440
24  5000000          240
25  4800000          352
26  4000000          240
27  4200000          340
28 10500000          300
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Apartamentos,Costo,xlab="N. Apartamentos (und)",ylab="Costo proyecto($ COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (x) “Apartamentos” para el modelo logarítmico:

```
> LNApartamentos<-c(log(Apartamentos))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNApartamentos,Costo,xlab="N. Apartamentos (und)",ylab="Costo proyecto ($ COP)",pch=16,col="grey30")
> abline(lm(Costo~LNApartamentos),col="red")
```

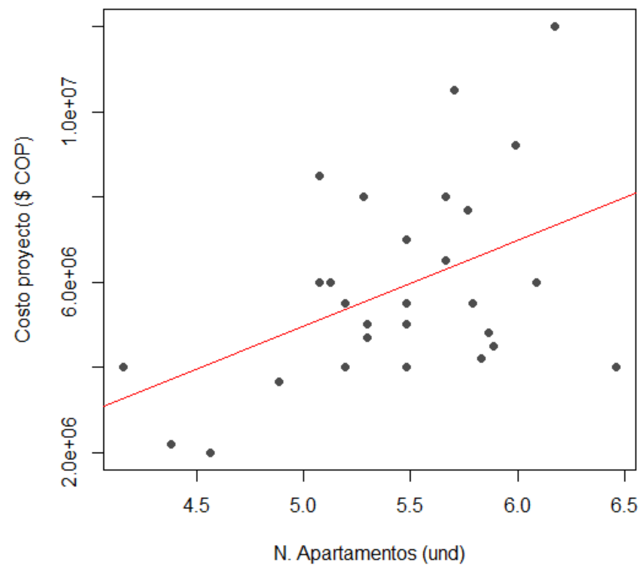


Figura 105. Gráfica de regresión lineal, modelo logarítmico, conjunto de edificios: Número de apartamentos (und)

En la (figura 105) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
>modelo<-lm(Costo~LNApartamentos)
> summary(modelo)

Call:
lm(formula = Costo ~ LNApartamentos)

Residuals:
     Min       1Q   Median       3Q      Max 
-3911081 -1400293  -503767  1093042  4667861 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)  
(Intercept)    -5092229    4189193  -1.216   0.2351  
LNApartamentos  2012439     766505   2.625   0.0143 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2118000 on 26 degrees of freedom
```

```
Multiple R-squared:  0.2096,    Adjusted R-squared:  0.1792
F-statistic: 6.893 on 1 and 26 DF,  p-value: 0.0143
```

Se aprecia que uno de los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 17,92% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = \alpha + \beta x$

$Costo\ (\$\ COP) = -5092229 + (2012439 * Apartamentos\ (und))$

El modelo lineal ajustado se puede expresar como la función logarítmica:

$\hat{y} = \alpha + \beta\, ln\, x$

$Costo\ (\$\ COP) = -5092229 + 2012439 * LnApartamentos(und)$ (Ecuación 47)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)

Analysis of Variance Table

Response: Costo
               Df     Sum Sq    Mean Sq F value Pr(>F)
LNApartamentos  1 3.0918e+13 3.0918e+13  6.8931 0.0143 *
Residuals      26 1.1662e+14 4.4854e+12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.96356, p-value = 0.4219
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 1.102194e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
>      plot(valores.ajustados,residuales,xlab="VALORES      AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

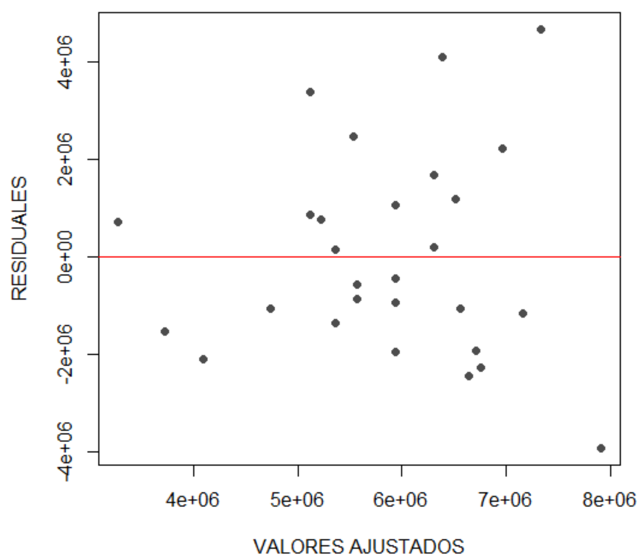


Figura 106. Gráfica de residuales, modelo logarítmico, conjunto de edificios: Número de apartamentos (und)

En la (figura 106) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Recíproco

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Apartamentos
1   4000000          640
2   7000000          240
3   2000000           96
4   5000000          200
5   6000000          168
6   6000000          160
7   9200000          400
8   8000000          288
9   8500000          160
10  4000000          180
11  2200000           80
12  5500000          180
13  4000000           64
14  5500000          328
15 12000000          480
16  8000000          196
17  4700000          200
18  3670000          132
19  7700000          320
20  6500000          288
21  4500000          360
22  5500000          240
23  6000000          440
24  5000000          240
25  4800000          352
26  4000000          240
27  4200000          340
28 10500000          300
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Apartamentos,Costo,xlab="N. Apartamentos (und)",ylab="Costo
proyecto ($ COP)",pch=16,col="grey30")
```

Se transforma la variable, creando un vector con la inversa de (x) “Apartamentos” para el modelo recíproco:

```
> Apartamentosinv<-c(1/Apartamentos)
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Apartamentosinv,Costo,xlab="N. Apartamentos (und)",ylab="Costo
proyecto ($ COP) ",pch=16,col="grey30")
> abline(lm(Costo~Apartamentosinv),col="red")
```

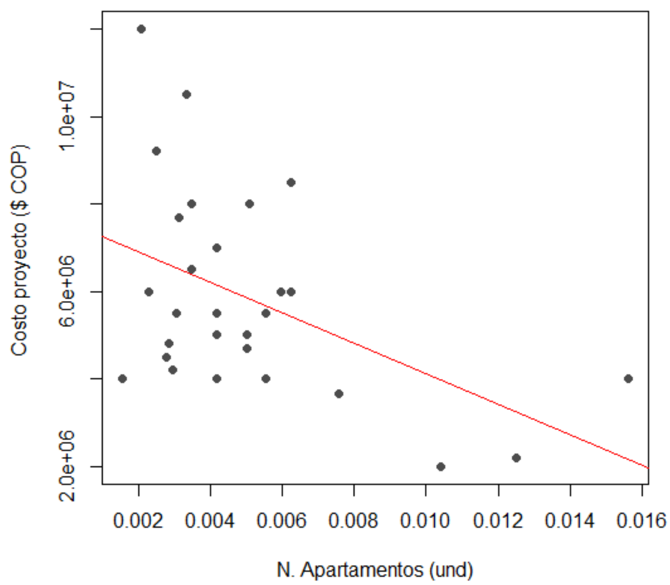


Figura 107. Gráfica de regresión lineal, modelo recíproco, conjunto de edificios: Número de apartamentos (und)

En la (figura 107) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~Apartamentosinv))
> summary(modelo)
Call:
lm(formula = Costo ~ Apartamentosinv)

Residuals:
     Min       1Q   Median       3Q      Max
-3063809 -1393571  -736839  1285413  5117524

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        7607810     748951  10.158 1.53e-10 ***
Apartamentosinv -348160172  126340233  -2.756   0.0106 *
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2096000 on 26 degrees of freedom
Multiple R-squared:  0.2261,    Adjusted R-squared:  0.1963
F-statistic: 7.594 on 1 and 26 DF,  p-value: 0.01056
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 19,63 de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = \alpha + \beta x$$

$$Costo\ (\$\ COP) = 7607810 + (-348160172 * Apartamentos\ (und))$$

El modelo lineal ajustado se puede expresar como la función Recíproca:

$$\hat{y} = \alpha + \beta\left(\frac{1}{x}\right)$$

$$Costo\ (\$\ COP) = 7607810 - 348160172\left(\frac{1}{Apartamentos\ (und)}\right)$$ (Ecuación 48)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
                Df     Sum Sq    Mean Sq F value  Pr(>F)  
Apartamentosinv  1 3.3352e+13 3.3352e+13  7.5941 0.01056 *
Residuals       26 1.1419e+14 4.3918e+12                  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

 Shapiro-Wilk normality test

data:  modelo$residuals
```

```
W = 0.93903, p-value = 0.1044
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 9.15179e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

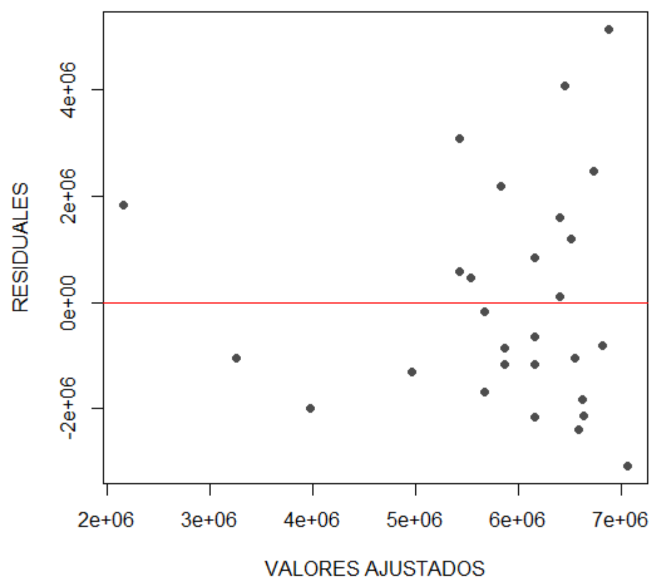


Figura 108. Gráfica de residuales, modelo recíproco, conjunto de edificios: Número de apartamentos (und)

En la (figura 108) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Polinómico

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo Apartamentos
1   4000000          640
2   7000000          240
3   2000000           96
4   5000000          200
5   6000000          168
6   6000000          160
7   9200000          400
8   8000000          288
9   8500000          160
10  4000000          180
11  2200000           80
12  5500000          180
13  4000000           64
14  5500000          328
15 12000000          480
16  8000000          196
17  4700000          200
18  3670000          132
19  7700000          320
20  6500000          288
21  4500000          360
22  5500000          240
23  6000000          440
24  5000000          240
25  4800000          352
26  4000000          240
27  4200000          340
28 10500000          300
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Apartamentos,Costo,xlab="N. Apartamentos (und)",ylab="Costo
proyecto ($ COP)",pch=16,col="grey30")
```

Se construye el modelo de regresión utilizando la función poly:

```
> fit2 <- lm(Costo ~ poly(Apartamentos, 2, raw=TRUE))
```

Se revisan los resultados con la instrucción summary:

```
> summary(fit2)

Call:
lm(formula = Costo ~ poly(Apartamentos, 2, raw = TRUE))

Residuals:
     Min       1Q   Median       3Q      Max 
-2874171 -1251173  -678590  1048115  4957419 

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)  
(Intercept)                       1094628.16 1623763.69   0.674   0.5064  
poly(Apartamentos, 2, raw = TRUE)1   30204.10   10974.54   2.752   0.0109 *
poly(Apartamentos, 2, raw = TRUE)2     -37.11      16.56  -2.241   0.0342 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2065000 on 25 degrees of freedom
Multiple R-squared:  0.2778,	Adjusted R-squared:   0.22 
F-statistic: 4.808 on 2 and 25 DF,  p-value: 0.01711
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 22% de la variabilidad observada.

El modelo polinomial de regresión estimado es:

$$\hat{y} = \beta + \beta_1 x - \beta_2 x^2$$

$$Costo\ (\$\ COP) = 1094628{,}16 + \left(30204{,}10 * Apartamentos\ (und)\right) - (-37{,}11 * Apartamentos\ (und)^2)$$ (Ecuación 49)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(fit2)
Analysis of Variance Table

Response: Costo
                                  Df     Sum Sq    Mean Sq F value  Pr(>F)  
poly(Apartamentos, 2, raw = TRUE)  2 4.0984e+13 2.0492e+13  4.8079 0.01711 *
Residuals                         25 1.0655e+14 4.2622e+12                  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(fit2$residuals)

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.93843, p-value = 0.1008
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(fit2$residuals)
[1] 7.691727e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-fit2$residuals
> valores.ajustados<-fit2$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

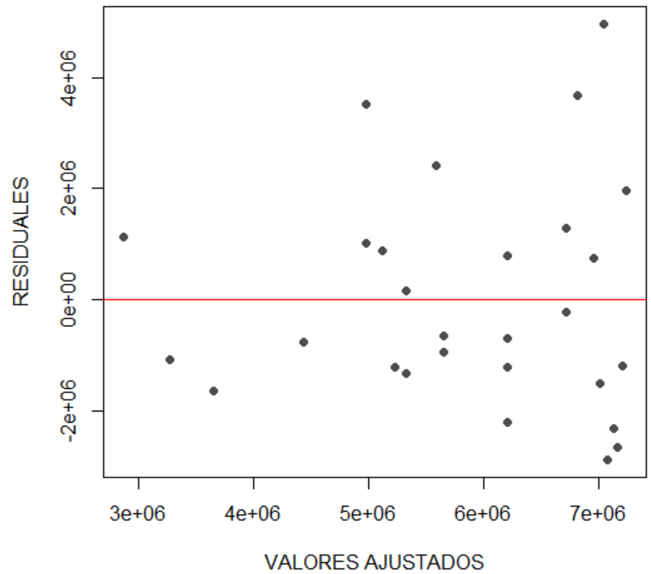


Figura 109. Gráfica de residuales, modelo polinómico, conjunto de edificios: Número de apartamentos (und)

En la (figura 109) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

De acuerdo con la realización de las ecuaciones metodológicas con los modelos funcionales de regresión lineal mediante el software de programación R, se efectúa una tabla resumen donde se encuentran cada uno de los resultados obtenidos en los diferentes modelos además de tener en cuenta los coeficientes de R^2, p-value y el error estándar de regresión en la clasificación conjunto de edificios multifamiliar con la variable número de apartamentos (und).

Tabla 14. Resultados, conjunto de edificios multifamiliares: Número de apartamentos (und)

Modelo funcional	Ecuación. Regresión Lineal	R-squared	p-value	Error estándar $\hat{\sigma}$
Lineal	(44) $Costo\ (\$\ COP) = 4131815 + 6603 * Apartamentos\ (und)$	0,0994	0,05662	4,9212e+12
Potencial	(45) $Costo\ (\$\ COP) = 614399{,}1025 * Apartamentos\ (und)^{0{,}4001}$	0,2384	0,004911	0,1293
Exponencial	(46) $Costo\ (\$\ COP) = 3913724{,}359 * e^{((1{,}226*10^{-03})*Apartamentos\ (und))}$	0,1145	0,04378	0,15033
Logarítmica	(47) $Costo\ (\$\ COP) = -5092229 + 2012439 * LnApartamentos(und)$	0,1792	0,0143	4,4854e+12
Recíproca	(48) $Costo\ (\$\ COP) = 7607810 - 348160172\left(\frac{1}{Apartamentos\ (und)}\right)$	0,1963	0,01056	4,3918e+12
Polinómica	(49) $Costo\ (\$\ COP) = 1094628{,}16 + \left(30204{,}10 * Apartamentos\ (und)\right) - \left(-37{,}11 * Apartamentos\ (und)^2\right)$	0,22	0,01711	4,2622e+12

Una vez construidos los modelos de regresión lineal para la variable explicativa “Numero de apartamentos (und)” en la clasificación “conjunto de edificios”, se pueden comparar los resultados obtenidos para elegir la ecuación que presenta mayor significancia estadística; en este caso se encuentra que el modelo funcional potencial muestra un coeficiente de determinación R^2 de 0,2384, valor superior a los obtenidos en los otros modelos empleados y el que más se acerca a uno (1); también se observa en el mismo modelo que el p- value es de 0,004911 el cual presenta un valor por debajo del 5%, indicando su significancia estadística.

Los resultados también demuestran que el modelo funcional que menos se ajusta es el lineal, dado que el coeficiente de determinación es el más bajo en comparación a los obtenidos en los otros modelos analizados y el que más se aleja de uno (1).

Teniendo en cuenta estos resultados analizados, la *(ecuación 45)* del modelo funcional potencial es la que representa mayor alcance y menor error estándar de la regresión para la variable dependiente *Costo ($ COP)* por lo cual será empleada para esta clasificación.

4.2.2.5. Conjunto de edificios, regresión lineal múltiple: Número de apartamentos (und), Longitud de tubería (m), Área (Ha)

El modelo de regresión lineal múltiple se emplea cuando existe más de una variable de estudio con el fin de analizar su importancia, en este caso las variables explicativas a estudiar son: Número de apartamentos (und), Longitud de tubería (m) y Área (Ha).

Se importan los datos que se encuentran en el archivo de Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
   Apartamentos Longitud  Area    Costo
1           640  1270.01 2.496  4000000
2           240   657.60 1.211  7000000
3            96   181.01 0.554  2000000
4           200  1412.35 1.081  5000000
5           168   857.50 0.052  6000000
6           160  1187.30 0.972  6000000
7           400  1246.46 1.312  9200000
8           288  1093.15 1.285  8000000
9           160  1142.56 0.876  8500000
10          180  1350.91 0.589  4000000
11           80   204.83 0.366  2200000
12          180   538.81 0.554  5500000
13           64   507.58 0.313  4000000
14          328  1963.24 2.049  5500000
15          480  2282.76 2.257 12000000
16          196  1004.01 0.580  8000000
17          200   976.05 1.081  4700000
18          132   471.75 0.540  3670000
19          320  2068.55 1.486  7700000
20          288  1220.42 1.026  6500000
21          360  1240.34 1.296  4500000
22          240   657.60 0.761  5500000
23          440  1970.82 2.021  6000000
24          240   657.60 0.761  5000000
25          352   709.40 0.847  4800000
26          240   720.80 0.911  4000000
27          340  1860.24 1.759  4200000
28          300  3166.49 1.921 10500000
> attach(data)
```

Se construye el modelo de regresión lineal múltiple usando la función *lm*:

$$yest = B_0 + B_1x_1 + B_2x_2$$

```
> modelo<-lm(Costo~Apartamentos+Longitud+Area)
> summary(modelo)
```

```
Call:
lm(formula = Costo ~ Apartamentos + Longitud + Area)

Residuals:
     Min       1Q   Median       3Q      Max 
-2898636 -1230120  -323289   653129  3789514 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)   
(Intercept)   3040620.6   868463.2   3.501  0.00184 **
Apartamentos     5832.8     5674.0   1.028  0.31421   
Longitud         2805.5      844.8   3.321  0.00286 **
Area         -1787348.4  1496294.0  -1.195  0.24395   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1864000 on 24 degrees of freedom
Multiple R-squared:  0.4347,	Adjusted R-squared:  0.364 
F-statistic: 6.151 on 3 and 24 DF,  p-value: 0.002967
```

Se aprecia que dos de los coeficientes logran ser significativos, además el p-value obtenido indica la significancia global del método. El coeficiente de determinación (R^2) indica el 36.40% de la variabilidad observada.

La ecuación del modelo de regresión es:

$$yest = B_0 + B_1 x_1 + B_2 x_2$$

$$Costo\ (\$\ COP) = 3040620{,}6 + \left(5832{,}8 * Apartamentos\ (und)\right) + \left(2805{,}5 * Longitud\ (m)\right) - (1787348{,}4 * Área\ (Ha))$$ (Ecuación 50)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
             Df     Sum Sq    Mean Sq F value    Pr(>F)    
Longitud      1 5.9131e+13 5.9131e+13 17.0142 0.0003842 ***
Area          1 1.3255e+12 1.3255e+12  0.3814 0.5426748    
Apartamentos  1 3.6726e+12 3.6726e+12  1.0568 0.3142070    
Residuals    24 8.3410e+13 3.4754e+12                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.94394, p-value = 0.1392
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 9.151537e-11
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

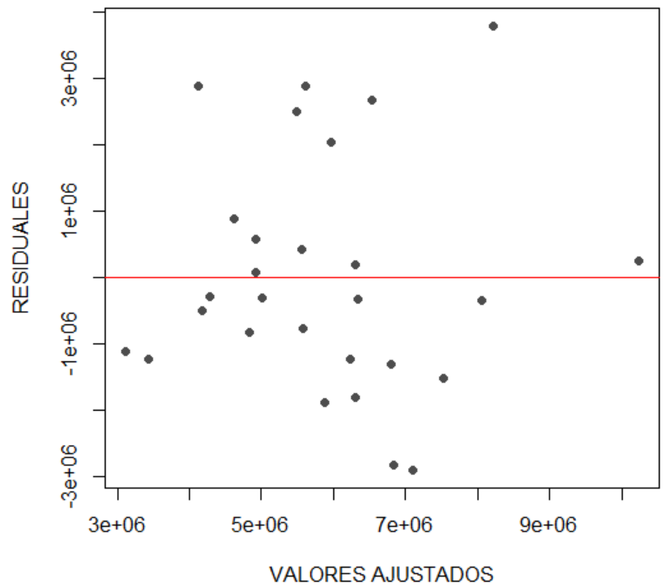


Figura 110. Gráfica de residuales, conjunto de edificios, regresión lineal múltiple: Número de apartamentos (und), Longitud de tubería (m), Área (Ha)

En la (figura 110) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Se realiza el test de contraste de homocedasticidad Breusch-Pagan con las siguientes instrucciones: La hipótesis nula indica que los residuales tienen varianza constante.

```
> library(lmtest)
> bptest(modelo)

studentized Breusch-Pagan test

data:  modelo
BP = 3.9971, df = 3, p-value = 0.2618
```

Teniendo en cuenta el p-valor obtenido, no se rechaza la hipótesis nula de homocedasticidad (varianza constante) en los residuales.

De acuerdo con la realización de la ecuación metodológica con el modelo de regresión lineal múltiple mediante el software de programación R, se efectúa una tabla resumen donde se encuentran el resultado obtenido del modelo además de tener en cuenta el coeficiente de R^2, p-value y el error estándar de regresión en la clasificación conjunto de edificios multifamiliar con las variables número de apartamentos (und), longitud de tubería (m) y área (Ha).

Tabla 15. Resultados, conjunto de edificios multifamiliares: Número de apartamentos (und), Longitud de tubería (m), Área (Ha)

Método	Ecuación. Regresión Lineal Múltiple	R-squared	p-value	Error estándar $\hat{\sigma}$
Regresión Lineal Múltiple	(50) $Costo\ (\$\ COP) = 3040620{,}6 + (5832{,}8 * Apartamentos\ (und)) + (2805{,}5 * Longitud\ (m)) + (-1787348{,}4 * Área\ (Ha))$	0,3640	0,002967	3,4754e+12

De acuerdo con los resultados obtenidos en el modelo funcional, dos de los coeficientes son significativos, sin embargo, en el ANOVA muestra que solo uno logra ser significativo; esto expone que las variables explicativas tienen mayor alcance estadístico cuando son analizadas de manera independiente, es por esto, que para evaluar las ecuaciones metodológicas en la clasificación "conjunto de edificios multifamiliares", no se tendrá en cuenta el método de regresión lineal múltiple.

4.2.3. Otros proyectos

4.2.3.1. Área (Ha)

Modelo lineal

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
```

```
      Costo    Area
1   1200000   0.208
2   1500000   0.200
3  18500000  78.934
4   4170000 104.160
5  18000000  15.443
6   1500000   0.720
7   1500000   0.348
8   5500000   3.460
9   2000000  39.880
10  1200000 105.550
11  1800000   0.065
12  1700000   0.155
13  1500000   0.095
14  1500000   0.059
15  1300000   0.035
16  1700000   0.425
17  1400000   0.075
18  3500000   0.141
19  2000000  29.556
20  3000000   0.041
21  3000000   0.003
22  1700000   0.229
23  1600000   0.110
24  1300000   0.026
25  4500000   0.096
26  4750000   1.354
27  2000000   0.254
28  3000000   0.570
29  1400000   0.058
30  1300000   3.100
31  2000000   2.545
32  2000000  78.270
33  3500000   0.081
34   800000   0.007
35   700000   0.003
36   700000   0.004
37  1200000   0.042
38  1100000   0.500
39  2500000   0.016
40  1600000   0.732
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

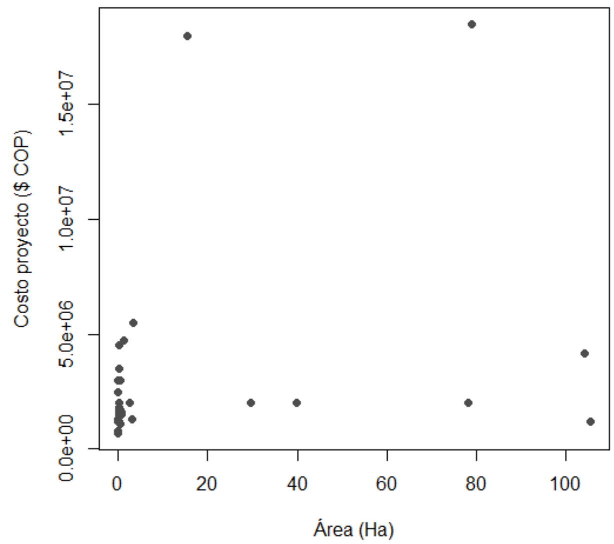


Figura 111. Gráfica de dispersión, otros proyectos: Área (Ha)

En la (figura 111) se observan datos en los que aumenta el área, pero no el costo del proyecto, y también viceversa, es decir, que no son directamente proporcionales; algunos costos son elevados para el área que representan.

```
> abline(lm(Costo~Area),col="red")
```

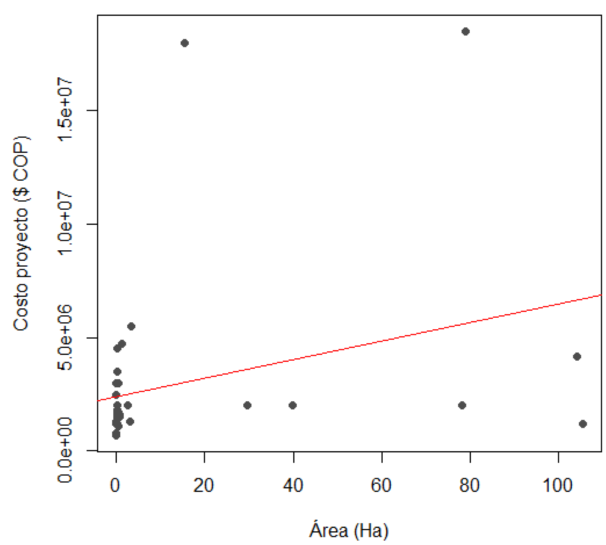


Figura 112. Gráfica de regresión lineal, modelo lineal, otros proyectos: Área (Ha)

En la (figura 112) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Se construye el modelo de regresión utilizando la función *lm*:

```
> modelo<-lm(Costo~Area)
> summary(modelo)

Call:
lm(formula = Costo ~ Area)

Residuals:
     Min       1Q   Median       3Q      Max 
-5517614 -1212929  -902994   218835 14968424 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2399844     617802   3.884 0.000397 ***
Area           40907      20291   2.016 0.050905 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3608000 on 38 degrees of freedom
Multiple R-squared:  0.09662,	Adjusted R-squared:  0.07285 
F-statistic: 4.064 on 1 and 38 DF,  p-value: 0.05091
```

Se aprecia que uno de los coeficientes es estadísticamente significativo, pero modelo de regresión no. Además, el coeficiente de determinación (R^2) explica aproximadamente el 7.28% de la variabilidad observada.

El modelo de regresión estimado es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$Costo\ (\$\ COP) = 2399844 + 40907 * Área\ (Ha) \qquad \textit{(Ecuación 51)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido:

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value  Pr(>F)  
Area       1 5.2907e+13 5.2907e+13  4.0644 0.05091 .
Residuals 38 4.9465e+14 1.3017e+13                  
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales. Básicamente, los supuestos que deben verificarse son los siguientes:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.60722, p-value = 3.931e-09
```

De acuerdo con la prueba de Shapiro-Wilks el p-value debe ser mayor al 5% esto indica que los residuales no cumplen con la normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 1.649369e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
>      plot(valores.ajustados,residuales,xlab="VALORES      AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

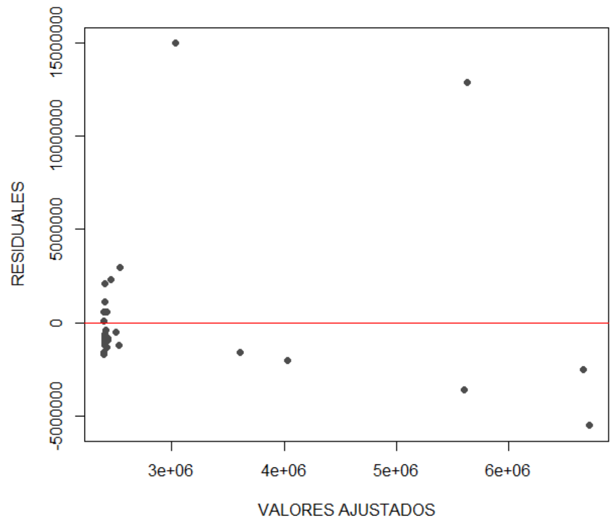


Figura 113. Gráfica de residuales, modelo lineal, otros proyectos: Área (Ha)

En la (figura 113) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo potencial</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo    Area
1   1200000   0.208
2   1500000   0.200
3  18500000  78.934
4   4170000 104.160
5  18000000  15.443
6   1500000   0.720
7   1500000   0.348
8   5500000   3.460
9   2000000  39.880
10  1200000 105.550
11  1800000   0.065
12  1700000   0.155
13  1500000   0.095
14  1500000   0.059
15  1300000   0.035
16  1700000   0.425
17  1400000   0.075
18  3500000   0.141
19  2000000  29.556
```

```
20  3000000   0.041
21  3000000   0.003
22  1700000   0.229
23  1600000   0.110
24  1300000   0.026
25  4500000   0.096
26  4750000   1.354
27  2000000   0.254
28  3000000   0.570
29  1400000   0.058
30  1300000   3.100
31  2000000   2.545
32  2000000  78.270
33  3500000   0.081
34   800000   0.007
35   700000   0.003
36   700000   0.004
37  1200000   0.042
38  1100000   0.500
39  2500000   0.016
40  1600000   0.732
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” y (x) “Área” para el modelo potencial:

```
> LNCosto<-c(log(Costo))
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(LNCosto~LNArea),col="red")
```

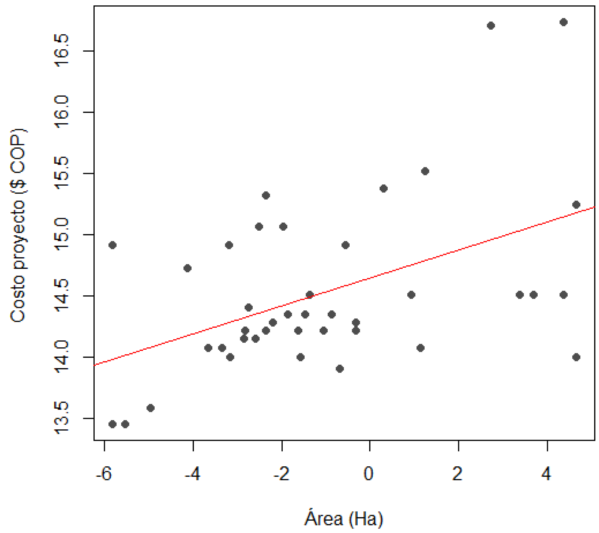


Figura 114. Gráfica de regresión lineal, modelo potencial, otros proyectos: Área (Ha)

En la (figura 114) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~LNArea))
> summary(modelo)

Call:
lm(formula = LNCosto ~ LNArea)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.1811 -0.4103 -0.1655  0.3861  1.7459 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 14.64823    0.10664 137.366  < 2e-16 ***
LNArea       0.11391    0.03487   3.266  0.00231 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.633 on 38 degrees of freedom
Multiple R-squared:  0.2192,    Adjusted R-squared:  0.1987 
F-statistic: 10.67 on 1 and 38 DF,  p-value: 0.002313
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 19.87% de la variabilidad observada.

El modelo de regresión estimado es:
$y^* = ln\,\alpha + \beta x^*$
$LnCosto\ (\$\ COP) = 14{,}64823 + 0{,}11391 * Área\ (Ha)$

El modelo lineal ajustado se puede expresar como la función potencial:
$\hat{y} = \alpha x^{\beta}$
$Costo\ (\$\ COP) = 2299563 * Area\ (Ha)^{0{,}11391}$ (Ecuación 52)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
          Df  Sum Sq Mean Sq F value   Pr(>F)
LNArea     1  4.2746  4.2746  10.669 0.002313 **
Residuals 38 15.2246  0.4006
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.90938, p-value = 0.003625
```

De acuerdo con la prueba de Shapiro-Wilks el p-value debe ser mayor al 5% esto indica que los residuales no cumplen con la normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 6.230097e-18
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

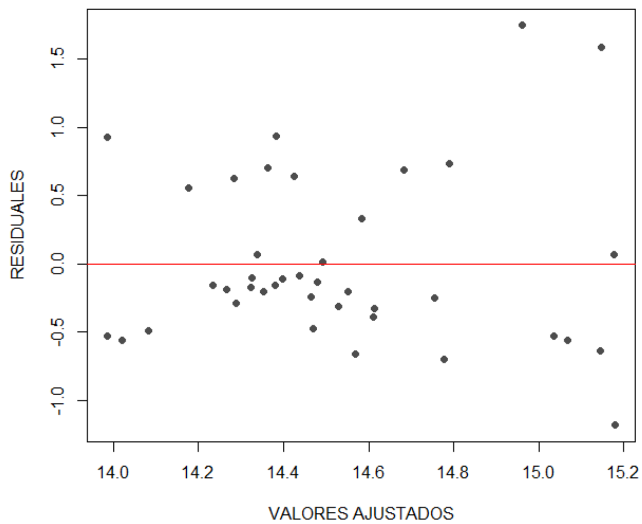


Figura 115. Gráfica de residuales, modelo potencial, otros proyectos: Área (Ha)

En la (figura 115) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

<u>Modelo exponencial</u>

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo    Area
1   1200000   0.208
```

```
2   1500000   0.200
3  18500000  78.934
4   4170000 104.160
5  18000000  15.443
6   1500000   0.720
7   1500000   0.348
8   5500000   3.460
9   2000000  39.880
10  1200000 105.550
11  1800000   0.065
12  1700000   0.155
13  1500000   0.095
14  1500000   0.059
15  1300000   0.035
16  1700000   0.425
17  1400000   0.075
18  3500000   0.141
19  2000000  29.556
20  3000000   0.041
21  3000000   0.003
22  1700000   0.229
23  1600000   0.110
24  1300000   0.026
25  4500000   0.096
26  4750000   1.354
27  2000000   0.254
28  3000000   0.570
29  1400000   0.058
30  1300000   3.100
31  2000000   2.545
32  2000000  78.270
33  3500000   0.081
34   800000   0.007
35   700000   0.003
36   700000   0.004
37  1200000   0.042
38  1100000   0.500
39  2500000   0.016
40  1600000   0.732
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (y) “Costo” para el modelo exponencial:

```
> LNCosto<-c(log(Costo))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Area,LNCosto,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(LNCosto~Area),col="red")
```

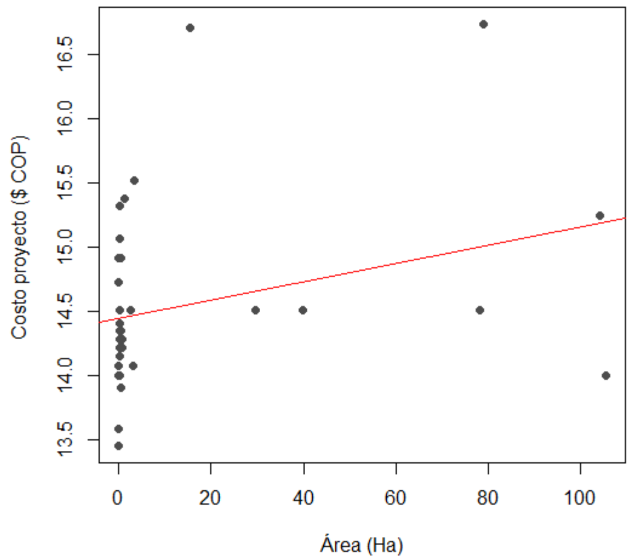


Figura 116. Gráfica de regresión lineal, modelo exponencial, otros proyectos: Área (Ha)

En la (figura 116) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(LNCosto~Area))
> summary(modelo)

Call:
lm(formula = LNCosto ~ Area)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.2001 -0.3669 -0.1621  0.3317  2.1511 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 14.444568   0.117487  122.95   <2e-16 ***
Area         0.007137   0.003859    1.85   0.0722 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6861 on 38 degrees of freedom
Multiple R-squared:  0.08259,   Adjusted R-squared:  0.05845
```

```
F-statistic: 3.421 on 1 and 38 DF,  p-value: 0.07216
```

Se aprecia que uno de los coeficientes es estadísticamente significativo, pero modelo de regresión no. Además, el coeficiente de determinación (R^2) explica aproximadamente el 7.21% de la variabilidad observada.

El modelo de regresión estimado:

$$y^* = ln\,\alpha + \beta x$$

$$LnCosto\ (\$\ COP) = (14{,}444568) + (0{,}007137 * Área\ (Ha))$$

El modelo lineal ajustado se puede expresar como la función exponencial:

$$\hat{y} = \alpha e^{\beta x}$$

$$Costo\ (\$\ COP) = 1875841 * e^{(0{,}007137 * Área\ (Ha))} \qquad \text{(Ecuación 53)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: LNCosto
          Df  Sum Sq Mean Sq F value  Pr(>F)
Area       1  1.6105 1.61046   3.421 0.07216 .
Residuals 38 17.8887 0.47075
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.90595, p-value = 0.002861
```

De acuerdo con la prueba de Shapiro-Wilks el p-value debe ser mayor al 5% esto indica que los residuales no cumplen con la normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] -9.847944e-18
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

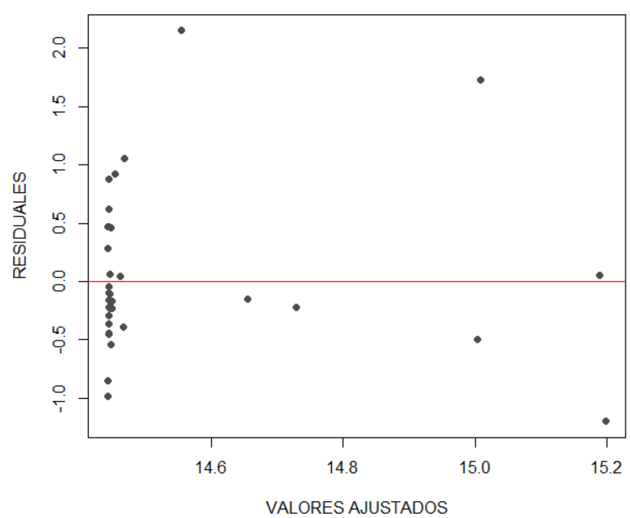


Figura 117. Gráfica de residuales, modelo exponencial, otros proyectos: Área (Ha)

En la (figura 117) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo logarítmico

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
```

```
> data
      Costo    Area
1   1200000   0.208
2   1500000   0.200
3  18500000  78.934
4   4170000 104.160
5  18000000  15.443
6   1500000   0.720
7   1500000   0.348
8   5500000   3.460
9   2000000  39.880
10  1200000 105.550
11  1800000   0.065
12  1700000   0.155
13  1500000   0.095
14  1500000   0.059
15  1300000   0.035
16  1700000   0.425
17  1400000   0.075
18  3500000   0.141
19  2000000  29.556
20  3000000   0.041
21  3000000   0.003
22  1700000   0.229
23  1600000   0.110
24  1300000   0.026
25  4500000   0.096
26  4750000   1.354
27  2000000   0.254
28  3000000   0.570
29  1400000   0.058
30  1300000   3.100
31  2000000   2.545
32  2000000  78.270
33  3500000   0.081
34   800000   0.007
35   700000   0.003
36   700000   0.004
37  1200000   0.042
38  1100000   0.500
39  2500000   0.016
40  1600000   0.732
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se realiza la transformación del vector (x) “Área” para el modelo logarítmico:

```
> LNArea<-c(log(Area))
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(LNArea, Costo,xlab="Área (Ha)",ylab=" Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(Costo~LNArea),col="red")
```

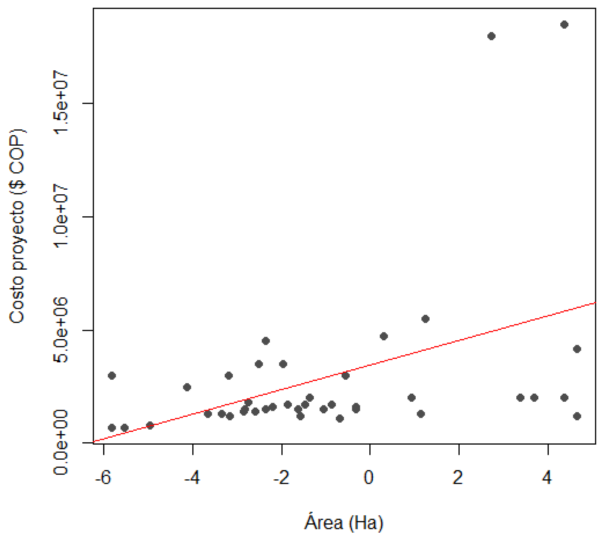


Figura 118. Gráfica de regresión lineal, modelo logarítmico, otros proyectos: Área (Ha)

En la (figura 118) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo ~LNArea))
> summary(modelo)

Call:
lm(formula = Costo ~ LNÁrea)

Residuals:
     Min       1Q   Median       3Q      Max
-4794411 -1468897  -583004   589340 13053721

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3453638     579479   5.960 6.46e-07 ***
LNÁrea        545326     189506   2.878  0.00654 **
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3440000 on 38 degrees of freedom
Multiple R-squared:  0.1789,    Adjusted R-squared:  0.1573
F-statistic: 8.281 on 1 and 38 DF,  p-value: 0.006541
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 15.73% de la variabilidad observada.

El modelo de regresión estimado:

$y^* = \alpha + \beta x$

$Costo\ (\$\ COP) = 3453638 + (545326 * \acute{A}rea\ (Ha))$

El modelo lineal ajustado se puede expresar como la función logarítmica:

$\hat{y} = \alpha + \beta\ ln\ x$

$Costo\ (\$\ COP) = 3453638 + (545326 * Ln\acute{A}rea\ (Ha))$ (Ecuación 54)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)

Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)
LNÁrea     1 9.7970e+13 9.7970e+13  8.2807 0.006541 **
Residuals 38 4.4959e+14 1.1831e+13
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.69469, p-value = 7.975e-08
```

De acuerdo con la prueba de Shapiro-Wilks el p-value debe ser mayor al 5% esto indica que los residuales no cumplen con la normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] -2.182787e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

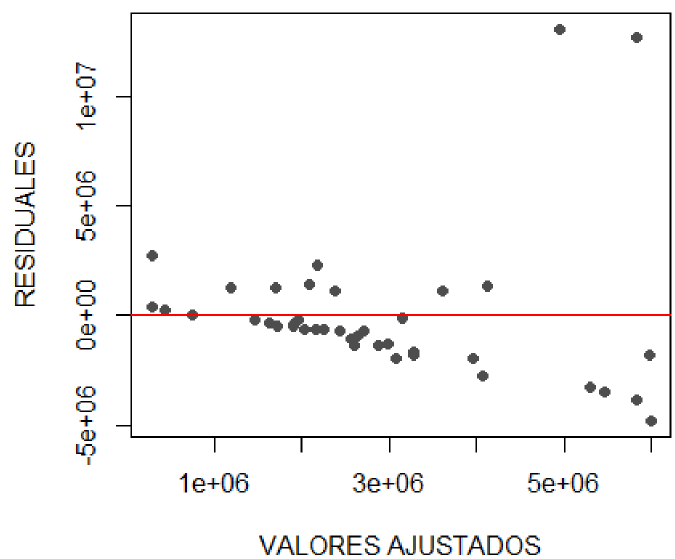


Figura 119. Gráfica de residuales, modelo logarítmico, otros proyectos: Área (Ha)

En la (figura 119) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo recíproco

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo    Area
1   1200000   0.208
2   1500000   0.200
3  18500000  78.934
4   4170000 104.160
5  18000000  15.443
6   1500000   0.720
7   1500000   0.348
8   5500000   3.460
9   2000000  39.880
10  1200000 105.550
11  1800000   0.065
12  1700000   0.155
13  1500000   0.095
14  1500000   0.059
15  1300000   0.035
16  1700000   0.425
17  1400000   0.075
18  3500000   0.141
19  2000000  29.556
20  3000000   0.041
21  3000000   0.003
22  1700000   0.229
23  1600000   0.110
24  1300000   0.026
25  4500000   0.096
26  4750000   1.354
27  2000000   0.254
28  3000000   0.570
29  1400000   0.058
30  1300000   3.100
31  2000000   2.545
32  2000000  78.270
33  3500000   0.081
34   800000   0.007
35   700000   0.003
36   700000   0.004
37  1200000   0.042
38  1100000   0.500
39  2500000   0.016
40  1600000   0.732
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se transforma la variable, creando un vector con la inversa de (x) “Área” para el modelo recíproco:

```
> Areainv<-c(1/Area)
```

Se dibuja nuevamente el diagrama de dispersión utilizando los datos transformados:

```
> plot(Areainv,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($ COP)
",pch=16,col="grey30")
> abline(lm(Costo~Areainv),col="red")
```

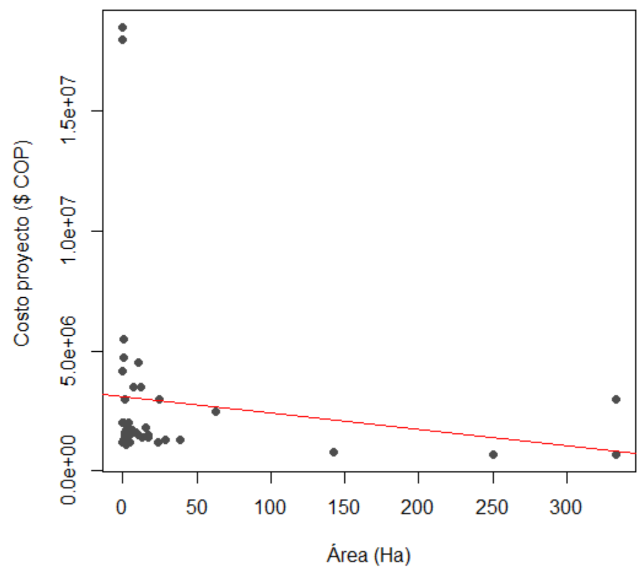


Figura 120. Gráfica de regresión lineal, modelo recíproco, otros proyectos: Área (Ha)

En la (figura 120) al tener datos que no son directamente proporcionales, la línea de regresión no se ajusta a la perfección, sin embargo, tiende a seguir la distribución de los valores.

Ahora se ajusta un modelo lineal para los datos transformados:

```
> modelo<-(lm(Costo~Areainv))
> summary(modelo)

Call:
lm(formula = Costo ~ Areainv)

Residuals:
```

```
     Min       1Q   Median       3Q      Max 
-2001449 -1584580 -1274397   -64453 15384970 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3115117     644821   4.831 2.25e-05 ***
Areainv        -6834       7276  -0.939    0.354    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3753000 on 38 degrees of freedom
Multiple R-squared:  0.02269,   Adjusted R-squared:  -0.00303 
F-statistic: 0.8822 on 1 and 38 DF,  p-value: 0.3535
```

Se aprecia que uno de los coeficientes es estadísticamente significativo, pero modelo de regresión no. Además, el coeficiente de determinación (R^2) explica aproximadamente el -0,303% de la variabilidad observada.

El modelo de regresión estimado es:

$$y^* = \alpha + \beta x$$

$$Costo\ (\$\ COP) = 3115117 + (-6834 * Área(Ha))$$

El modelo lineal ajustado se puede expresar como la función Recíproca:

$$\hat{y} = \alpha + \beta \left(\frac{1}{x}\right)$$

$$Costo\ (\$\ COP) = 3115117 - 6834 \left(\frac{1}{Área\ (Ha)}\right) \qquad \text{(Ecuación 55)}$$

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(modelo)
Analysis of Variance Table

Response: Costo
          Df     Sum Sq    Mean Sq F value   Pr(>F)   
Areainv    1 1.3909e+11 1.3909e+11  10.509 0.005909 **
Residuals 14 1.8529e+11 1.3235e+10                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(modelo$residuals)

        Shapiro-Wilk normality test

data:  modelo$residuals
W = 0.93819, p-value = 0.3275
```

Teniendo en cuenta el p-value obtenido, no se rechaza la hipótesis nula de normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(modelo$residuals)
[1] 2.50111e-12
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-modelo$residuals
> valores.ajustados<-modelo$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

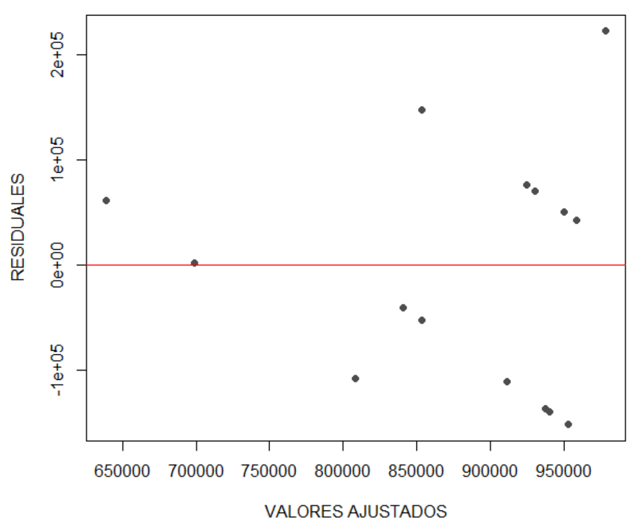


Figura 121. Gráfica de residuales, modelo recíproco, otros proyectos: Área (Ha)

En la (figura 121) se observa una variación aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales mantienen la varianza constante.

Modelo Polinómico

Se importan los datos que se encuentran en el archivo Excel:

```
> data<-read.delim("clipboard",header=T,dec=",",check.names=T)
> data
      Costo    Area
1   1200000   0.208
2   1500000   0.200
3  18500000  78.934
4   4170000 104.160
5  18000000  15.443
6   1500000   0.720
7   1500000   0.348
8   5500000   3.460
9   2000000  39.880
10  1200000 105.550
11  1800000   0.065
12  1700000   0.155
13  1500000   0.095
14  1500000   0.059
15  1300000   0.035
16  1700000   0.425
17  1400000   0.075
18  3500000   0.141
19  2000000  29.556
20  3000000   0.041
21  3000000   0.003
22  1700000   0.229
23  1600000   0.110
24  1300000   0.026
25  4500000   0.096
26  4750000   1.354
27  2000000   0.254
28  3000000   0.570
29  1400000   0.058
30  1300000   3.100
31  2000000   2.545
32  2000000  78.270
33  3500000   0.081
34   800000   0.007
35   700000   0.003
36   700000   0.004
37  1200000   0.042
38  1100000   0.500
39  2500000   0.016
40  1600000   0.732
> attach(data)
```

Se dibuja el gráfico de dispersión para tener una aproximación al modelo funcional de regresión:

```
> plot(Area,Costo,xlab="Área (Ha)",ylab="Costo proyecto ($
COP)",pch=16,col="grey30")
```

Se construye el modelo de regresión utilizando la función poly:

```
> fit2 <- lm(Costo ~ poly(Area, 2, raw=TRUE))
```

Se revisan los resultados con la instrucción summary:

```
> summary(fit2)

Call:
lm(formula = Costo ~ poly(Area, 2, raw = TRUE))

Residuals:
     Min       1Q   Median       3Q      Max
-6312759  -872635  -588243   535171 12697100

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                2052261.6   598094.8   3.431  0.00149 **
poly(Area, 2, raw = TRUE)1  244308.0    85373.3   2.862  0.00690 **
poly(Area, 2, raw = TRUE)2   -2189.7      895.8  -2.444  0.01940 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3393000 on 37 degrees of freedom
Multiple R-squared:  0.2222,    Adjusted R-squared:  0.1802
F-statistic: 5.286 on 2 and 37 DF,  p-value: 0.009569
```

Se aprecia que tanto los coeficientes como el modelo de regresión son estadísticamente significativos. Además, el coeficiente de determinación (R^2) explica aproximadamente el 18.02% de la variabilidad observada.

El modelo polinomial de regresión estimado es:

$$\hat{y} = \beta + \beta_1 x - \beta_2 x^2$$

$$Costo\ (\$\ COP) = 2052261{,}6 + \Big((244308) * Área\ (Ha)\Big) - \left(-2189{,}7 * Área\ (Ha)^2\right)$$

(Ecuación 56)

La tabla de análisis de varianza confirma la significancia global del modelo de regresión, al analizar las sumas de cuadrados y el valor p obtenido.

```
> anova(fit2)

Analysis of Variance Table

Response: Costo
                         Df     Sum Sq    Mean Sq F value   Pr(>F)
poly(Area, 2, raw = TRUE)  2 1.2168e+14 6.0839e+13  5.2856 0.009569 **
Residuals                37 4.2588e+14 1.1510e+13
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente se verifican los supuestos del modelo mediante el análisis de los residuales:

1. Los residuales siguen la distribución normal y su media es cero. La normalidad se comprueba mediante la prueba de Shapiro – Wilks, debido a que la muestra es pequeña.

```
> shapiro.test(fit2$residuals)

        Shapiro-Wilk normality test

data:  fit2$residuals
W = 0.72219, p-value = 2.296e-07
```

De acuerdo con la prueba de Shapiro-Wilks el p-value debe ser mayor al 5% esto indica que los residuales no cumplen con la normalidad.

La media de los residuales se obtiene con la siguiente instrucción:

```
> mean(fit2$residuals)
[1] 1.06192e-10
```

El resultado indica que la media es cero.

2. La varianza de los residuales es constante. Se verifica graficando los valores residuales y los ajustados. Los valores ajustados se obtienen mediante la instrucción modelo$fitted:

```
> residuales<-fit2$residuals
> valores.ajustados<-fit2$fitted
> plot(valores.ajustados,residuales,xlab="VALORES AJUSTADOS",
ylab="RESIDUALES",pch=16,col="grey30")
> abline(h=0,col="red")
```

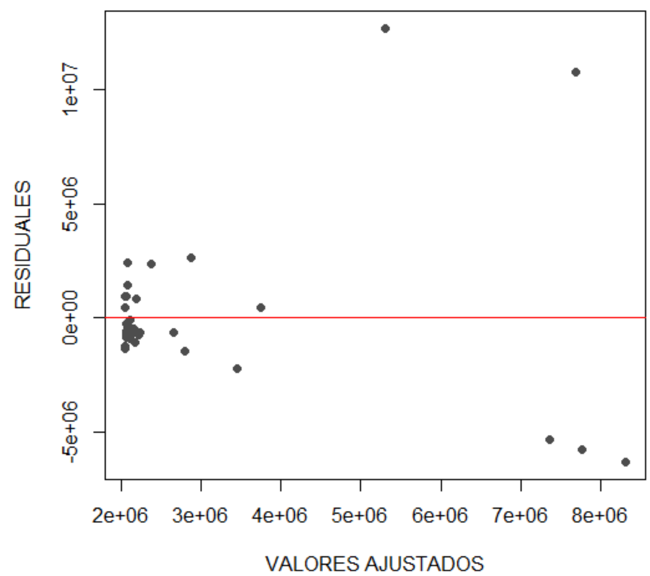


Figura 122. Gráfica de residuales, modelo polinómico, otros proyectos: Área (Ha)

En la (figura 122) se observa una variación no aleatoria de los residuales con respecto a la media, por lo tanto, se puede concluir que los residuales no mantienen una varianza constante.

De acuerdo con la realización de las ecuaciones metodológicas con los modelos funcionales de regresión lineal mediante el software de programación R, se efectúa una tabla resumen donde se encuentran cada uno de los resultados obtenidos en los diferentes modelos además de tener en cuenta los coeficientes de R^2, p-value y el error estándar de regresión en la clasificación otros proyectos con la variable área (Ha).

Tabla 16. Resultados, otros proyectos: Área (Ha)

Modelo funcional	Ecuación. Regresión Lineal	R-squared	p-value	Error estándar $\hat{\sigma}$
Lineal	(51) $Costo\ (\$\ COP) = 2399844 + 40907 * Área\ (Ha)$	0,07285	0,05091	1,3017e+13
Potencial	(52) $Costo\ (\$\ COP) = 2299563 * Área\ (Ha)^{0{,}11391}$	0,1987	0,002313	0,4006
Exponencial	(53) $Costo\ (\$\ COP) = 1875841 * e^{(0{,}007137 * Área\ (Ha))}$	0,05845	0,07216	0,47075
Logarítmica	(54) $Costo\ (\$\ COP) = 3453638 + 545326 * LnÁrea\ (Ha)$	0,1573	0,006541	1,1831e+13
Recíproca	(55) $Costo\ (\$\ COP) = 3115117 - 6834\left(\frac{1}{Área\ (Ha)}\right)$	-0,00303	0,3535	1,3235e+10
Polinómica	(56) $Costo\ (\$\ COP) = 2052261{,}6 + \left((244308) * Área\ (Ha)\right) - \left(-2189{,}7 * Área\ (Ha)^2\right)$	0,1802	0,009569	1,1510e+13

Una vez construidos los modelos de regresión lineal para la variable explicativa "Área (Ha)" en la clasificación "otros proyectos", se pueden comparar los resultados obtenidos para elegir la ecuación que presenta mayor significancia estadística; en este caso se encuentra que el modelo funcional potencial muestra un coeficiente de determinación R^2 de 0,1987, valor superior a los obtenidos en los otros modelos empleados y el que más se acerca a uno (1); también se observa en el mismo modelo que el p- value es de 0,002313 el cual presenta un valor por debajo del 5%, indicando su significancia estadística.

Los resultados también demuestran que el modelo funcional que menos se ajusta es el reciproco, dado que el coeficiente de determinación es el más bajo en comparación a los obtenidos en los otros modelos analizados y el que más se aleja de uno (1).

Teniendo en cuenta estos resultados analizados, la *(ecuación 52)* del modelo funcional potencial es la que representa mayor alcance y menor error estándar de la regresión para la variable dependiente *Costo ($ COP)*, por lo cual será empleada para esta clasificación.

4.3. Análisis de resultados

Al construir los modelos estadísticos de regresión lineal, se obtienen diferentes resultados para cada una de las variables y modelos funcionales empleados, de esta manera se hace una comparación y se elige la ecuación que presenta una mayor significancia estadística; para esto se tiene en cuenta el coeficiente de determinación R^2 y el p-value; el primero, indica la bondad de ajuste del modelo de regresión en el cual el resultado debe aproximarse a uno (1) para un mejor ajuste estadístico, y el segundo indica la significancia del modelo el cual es aplicable cuando el resultado es menor al cinco por ciento (5%).

De esta manera se realiza la comparación de los resultados obtenidos en el análisis estadístico realizado con el software R y se seleccionan aquellas ecuaciones que en el modelo ejecutado tuvieron un coeficiente de determinación R^2 y un p-value con mayor significancia estadística. Las ecuaciones de los modelos seleccionados se muestran en las tablas que se presentan a continuación con el fin de evaluarlas y verificar su eficacia y aplicación en la ingeniería civil.

En el proceso de evaluar las ecuaciones de los modelos seleccionados, se toman al azar tres proyectos hidráulicos que permiten realizar el paso a paso explicativo del uso y eficacia de la ecuación, posteriormente se muestra una tabla de los proyectos, según su clasificación, con los resultados de los costos obtenidos con la ecuación y los costos reales de los proyectos hidráulicos con el fin de hacer una comparación.

4.3.1. Residencia unifamiliar

Tabla 17. Ecuaciones seleccionadas, residencia unifamiliar

Clasificación	Variables explicativas	Modelo funcional	Ecuación. Regresión Lineal
Vivienda	Área (Ha)	Potencial	(2) $Costo\ (\$\ COP) = 1406507{,}163 * Área\ (Ha)^{0{,}15119}$
Conjunto de viviendas	Área (Ha)	Potencial	(8) $Costo\ (\$\ COP) = 3330090 * Área\ (Ha)^{0{,}27692}$
	L. de tubería (m)	Potencial	(14) $Costo\ (\$\ COP) = 675642{,}9263 * Longitud\ (m)^{0{,}23598}$
	N. de viviendas (und)	Potencial	(20) $Costo\ (\$\ COP) = 948512{,}2657 * Viviendas\ (und)^{0{,}29371}$

Para evaluar las ecuaciones se realiza el siguiente procedimiento:

<u>Vivienda, Área (Ha)</u>

Se obtiene la ecuación del modelo funcional potencial con el método de regresión lineal:

$$Costo\ (\$\ COP) = 1406507{,}163 * (Área\ (Ha)^{0{,}15119})$$

Se seleccionan de manera aleatoria tres proyectos hidráulicos para verificar la eficacia de la ecuación:

1. Proyecto Eduviges, realizado en el municipio de Toledo en el año 2018 con un área de 0,095 (Ha) y con un costo de $1.000.000 COP.

$$Costo\ (\$\ COP) = 1406507{,}163 * (0{,}095^{0{,}15119})$$

$$Costo = \$985.335{,}1601\ COP$$

2. Proyecto Bellavista, realizado en el municipio de Los Patios en el año 2017 con un área de 0,023 (Ha) y con un costo de $800.000 COP.

$$Costo\ (\$\ COP) = 1406507{,}163 * (0{,}023^{0{,}15119})$$

$$Costo = \$795.154,6075\ COP$$

3. Proyecto Corozal, realizado en el municipio de Los Patios en el año 2020 con un área de 0,200 (Ha) y con un costo de $ 1.200.000 COP.

$$Costo(\$\ COP) = 1406507,163 * (0,200^{0,15119})$$

$$Costo = \$1.102.718,532\ COP$$

$$Costo = \$1.865.318,302\ COP$$

Siguiendo este mismo procedimiento, se calcula el costo para cada proyecto con la ecuación seleccionada con el fin de hacer una comparación con el costo real:

Tabla 18. Comparación de costos, vivienda unifamiliar: Área (Ha)

Nombre proyecto	Costo real ($ COP)	Costo ecuación ($ COP)
Vivienda Bellavista	$ 800.000	$ 795.154,61
Vivienda Niza	$ 700.000	$ 720.664,88
Casa Eduviges	$ 1.000.000	$ 985.335,16
Acacios	$ 700.000	$ 772.514,62
Palujan	$ 1.000.000	$ 956.395,96
Colina Campestre	$ 700.000	$ 701.070,91
Punta del Este	$ 800.000	$ 928.214,39
Serrano Castellanos	$ 800.000	$ 870.945,56
Gonzáles	$ 800.000	$ 921.501,33
OV1	$ 800.000	$ 963.654,75
Palma Deluxe	$ 1.000.000	$ 805.242,15
Palma Country	$ 1.000.000	$ 894.209,38
Palma Dorada	$ 1.000.000	$ 894.209,38
Cabaña Corozal	$ 1.200.000	$ 1.102.718,53
Casagrande	$ 800.000	$ 805.242,15
Zafiro	$ 1.000.000	$ 904.674,94

Se realiza una gráfica que permita ver la relación entre los valores:

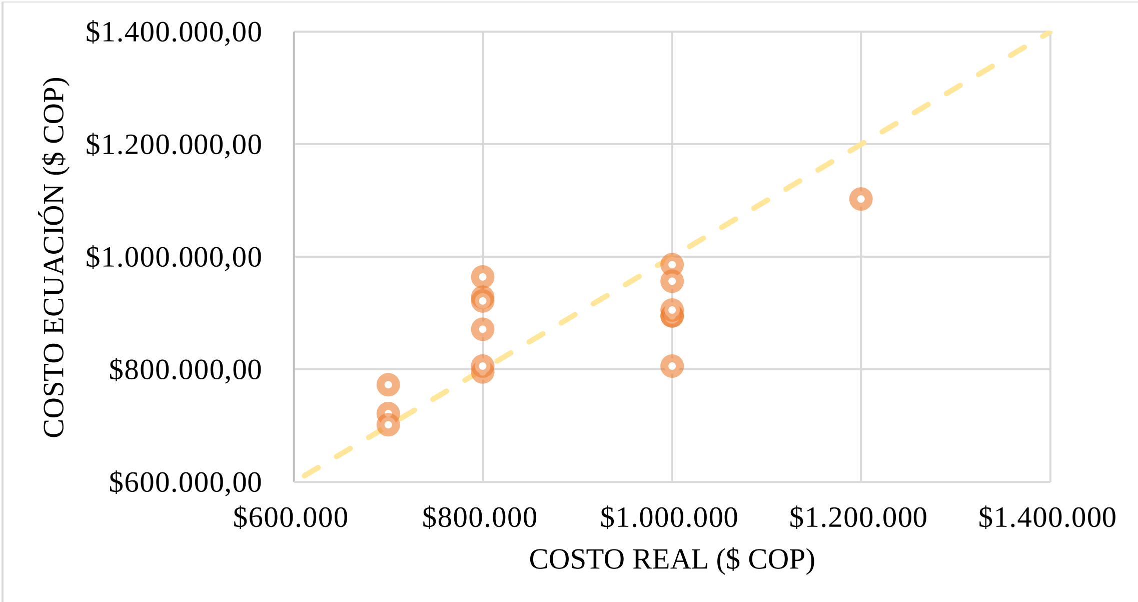


Figura 123. Relación costo real y costo ecuación, vivienda unifamiliar: Área (Ha)

La (figura 123) permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse a medida que se acercan a la diagonal, lo cual muestra que el modelo es bastante aceptable y que existe relación entre ellos.

<u>Conjunto de viviendas, Área (Ha)</u>

Se obtiene la ecuación del modelo funcional potencial con el método de regresión lineal:

$$Costo\ (\$\ COP) = 3330090 * Área\ (Ha)^{0,27692}$$

Se seleccionan de manera aleatoria tres proyectos hidráulicos para verificar la eficacia de la ecuación:

1. Proyecto Canela, realizado en el municipio de Cúcuta en el año 2018 con un área de 3,067 (Ha) y con un costo de $4.200.000 COP.

$$Costo\ (\$\ COP) = 3330090 * (3,067^{0,27692})$$

$$Costo = \$4.541.891{,}444\ COP$$

2. Proyecto Bosque Alto, realizado en el municipio de Los Patios en el año 2019 con un área de 0,178 (Ha) y con un costo de $2.500.000 COP.

$$Costo\ (\$\ COP) = 3330090 * (0{,}178^{0{,}27692})$$

$$Costo = \$2.064.820{,}38\ COP$$

3. Proyecto Carolina Deluxe, realizado en el municipio de Cúcuta en el año 2020 con un área de 0,145 (Ha) y con un costo de $ 1.500.000 COP.

$$Costo\ (\$\ COP) = 3330090 * (0{,}145^{0{,}27692})$$

$$Costo = \$1.950.841{,}556\ COP$$

Siguiendo este mismo procedimiento, se calcula el costo para cada proyecto con la ecuación seleccionada con el fin de hacer una comparación con el costo real:

Tabla 19. Comparación de costos, conjunto de viviendas unifamiliares: Área (Ha)

Nombre proyecto	Costo real ($ COP)	Costo ecuación ($ COP)
Girasoles	$ 5.000.000	$ 4.071.740,16
La Macarena	$ 4.500.000	$ 3.950.354,15
White Country House	$ 4.000.000	$ 3.565.697,99
Conjunto El Recreo	$ 4.500.000	$ 3.736.077,16
Proyectos Altobelo	$ 3.000.000	$ 3.779.793,65
Proyectos Firenze	$ 3.000.000	$ 3.769.155,12
Villa Teresa	$ 3.200.000	$ 4.789.486,65
Pinares	$ 3.500.000	$ 4.121.138,52
Brisas de Arkamar	$ 5.800.000	$ 3.545.492,12
Altagracia	$ 4.000.000	$ 3.958.009,41
Los naranjos	$ 3.700.000	$ 4.338.807,58
Canela	$ 4.200.000	$ 4.541.891,44
Conjunto Ebano	$ 5.500.000	$ 4.173.135,19
Medinterrané	$ 3.500.000	$ 3.460.666,85
Bosque Alto	$ 2.500.000	$ 2.064.820,38
Molinos	$ 2.200.000	$ 2.295.683,75

Rayuela	$ 4.000.000	$	3.704.312,75
Vallarta	$ 4.500.000	$	4.168.011,33
Villa Bolívar	$ 5.500.000	$	3.984.214,21
Carolina Campestre	$ 1.500.000	$	2.203.242,77
Carolina Deluxe	$ 1.500.000	$	1.950.841,56
Conjunto El Verjel	$ 2.500.000	$	1.855.499,73
Villa Verona	$ 2.500.000	$	3.617.176,46
Valena	$ 2.500.000	$	2.679.270,21

Se realiza una gráfica que permita ver la relación entre los valores:

Figura 124. Relación costo real y costo ecuación, conjunto de viviendas unifamiliares: Área (Ha)

La (figura 124) permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse a medida que se acercan a la diagonal, lo cual muestra que el modelo es bastante aceptable y que existe relación entre ellos.

Conjunto de viviendas, Longitud de tubería (m)

Se obtiene la ecuación del modelo funcional polinómico con el método de regresión lineal:

$$Costo\ (\$\ COP) = 675642{,}9263 * Longitud\ (m)^{0{,}23598}$$

Se seleccionan de manera aleatoria tres proyectos hidráulicos para verificar la eficacia de la ecuación:

1. Proyecto Canela, realizado en el municipio de Cúcuta en el año 2018 con una longitud de tubería de 1726,58 (m) y con un costo de $4.200.000 COP.

$$Costo\ (\$\ COP) = 675642{,}9263 * (1726{,}58)^{0{,}23598}$$

$$Costo = \$3.923.092{,}54\ COP$$

2. Proyecto Bosque Alto, realizado en el municipio de Los Patios en el año 2019 con una longitud de tubería de 234,03 (m) y con un costo de $2.500.000 COP.

$$Costo\ (\$\ COP) = 675642{,}9263 * (234{,}03)^{0{,}23598}$$

$$Costo = \$2.448.036{,}74\ COP$$

3. Proyecto Carolina Deluxe, realizado en el municipio de Cúcuta en el año 2020 con una longitud de tubería de 160,84 (m) y con un costo de $ 1.500.000 COP.

$$Costo\ (\$\ COP) = 675642{,}9263 * (160{,}84)^{0{,}23598}$$

$$Costo = \$\ 2.240{,}691.70\ COP$$

Siguiendo este mismo procedimiento, se calcula el costo para cada proyecto con la ecuación seleccionada con el fin de hacer una comparación con el costo real:

Tabla 20. Comparación de costos, conjunto de viviendas unifamiliares: Longitud de tubería (m)

Nombre proyecto	Costo real ($ COP)	Costo ecuación ($ COP)
Girasoles	$ 5,000,000	$ 3,807,348.31
La Macarena	$ 4,500,000	$ 3,573,392.58
White Country House	$ 4,000,000	$ 3,327,745.58
Conjunto El Recreo	$ 4,500,000	$ 4,141,871.19
Proyectos Altobelo	$ 3,000,000	$ 3,938,177.85
Proyectos Firenze	$ 3,000,000	$ 3,666,515.14
Villa Teresa	$ 3,200,000	$ 4,456,001.57
Pinares	$ 3,500,000	$ 3,955,759.08
Brisas de Arkamar	$ 5,800,000	$ 3,869,098.04
Altagracia	$ 4,000,000	$ 4,081,541.22
Los naranjos	$ 3,700,000	$ 2,322,979.63
Canela	$ 4,200,000	$ 3,923,092.54
Conjunto Ebano	$ 5,500,000	$ 3,950,879.23
Medinterrané	$ 3,500,000	$ 4,076,195.25
Bosque Alto	$ 2,500,000	$ 2,448,036.74
Molinos	$ 2,200,000	$ 2,918,841.98
Rayuela	$ 4,000,000	$ 4,300,066.42
Vallarta	$ 4,500,000	$ 4,011,485.92
Villa Bolívar	$ 5,500,000	$ 3,691,651.50
Carolina Campestre	$ 1,500,000	$ 2,612,233.57
Carolina Deluxe	$ 1,500,000	$ 2,240,691.70
Conjunto El Verjel	$ 2,500,000	$ 1,828,616.11
Villa Verona	$ 2,500,000	$ 3,490,515.10
Valena	$ 2,500,000	$ 2,881,803.20

Se realiza una gráfica que permita ver la relación entre los valores:

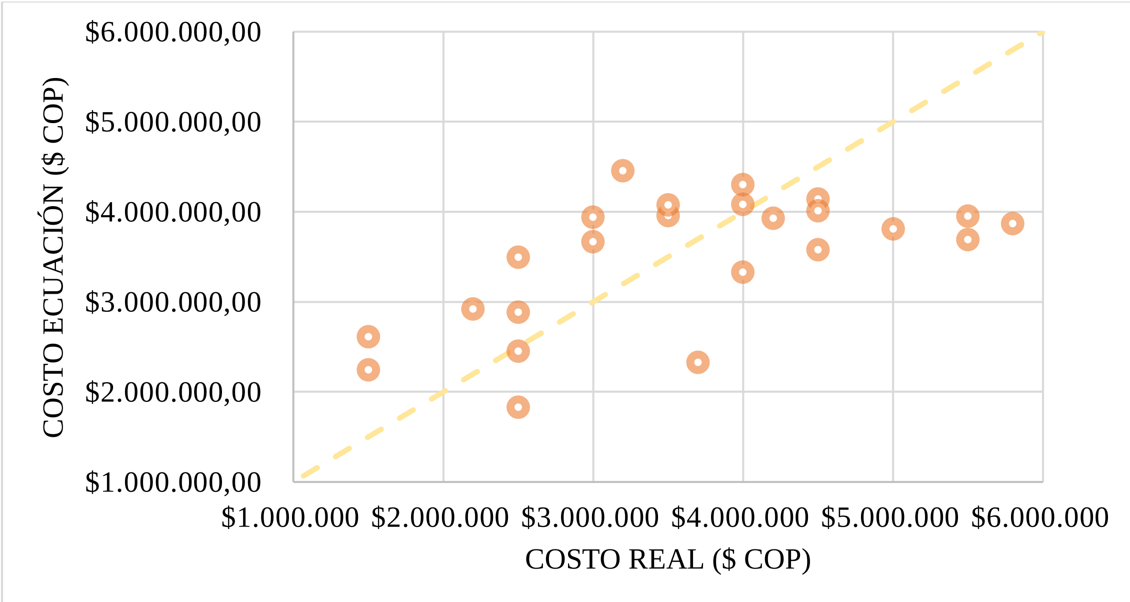


Figura 125. Relación costo real y costo ecuación, conjunto de viviendas unifamiliares: Longitud de tubería (m)

La (figura 125) permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse a medida que se acercan a la diagonal, lo cual muestra que el modelo es bastante aceptable y que existe relación entre ellos.

<u>Conjunto de viviendas, Número de viviendas (und)</u>

Se obtiene la ecuación del modelo funcional potencial con el método de regresión lineal:

$$Costo\ (\$\ COP) = 948512{,}2657 * Viviendas^{0{,}29371}$$

Se seleccionan de manera aleatoria tres proyectos hidráulicos para verificar la eficacia de la ecuación:

1. Proyecto Canela, realizado en el municipio de Cúcuta en el año 2018 con un total de 144 (und) viviendas y con un costo de $4.200.000 COP.

$$Costo\ (\$\ COP) = 948512{,}2657 * (144^{0{,}29371})$$

$$Costo = \$4.082.964{,}453\ COP$$

2. Proyecto Bosque Alto, realizado en el municipio de Los Patios en el año 2019 con un total de 11 (und) viviendas y con un costo de $2.500.000 COP.

$$Costo\ (\$\ COP) = 948512{,}2657 * (11^{0{,}29371})$$

$$Costo = \$1.918.272{,}924\ COP$$

3. Proyecto Carolina Deluxe, realizado en el municipio de Cúcuta en el año 2020 con un total de 10 (und) viviendas y con un costo de $ 1.500.000 COP.

$$Costo\ (\$\ COP) = 948512{,}2657 * (10^{0{,}29371})$$

$$Costo = \$1.865.318{,}302\ COP$$

Siguiendo este mismo procedimiento, se calcula el costo para cada proyecto con la ecuación seleccionada con el fin de hacer una comparación con el costo real:

Tabla 21. Comparación de costos, conjunto de viviendas unifamiliares: Número de viviendas (und)

Nombre proyecto	**Costo real ($ COP)**	**Costo ecuación ($ COP)**
Girasoles	$ 5.000.000	$ 4.226.682,31
La Macarena	$ 4.500.000	$ 3.979.941,79
White Country House	$ 4.000.000	$ 2.992.591,45
Conjunto El Recreo	$ 4.500.000	$ 3.812.207,01
Proyectos Altobelo	$ 3.000.000	$ 4.132.213,15
Proyectos Firenze	$ 3.000.000	$ 3.772.423,11
Villa Teresa	$ 3.200.000	$ 4.801.112,82
Pinares	$ 3.500.000	$ 3.157.212,07
Brisas de Arkamar	$ 5.800.000	$ 3.752.146,98
Altagracia	$ 4.000.000	$ 4.330.843,37
Los naranjos	$ 3.700.000	$ 4.359.523,66
Canela	$ 4.200.000	$ 4.082.964,45

Conjunto Ebano	$ 5.500.000	$ 4.535.747,74
Medinterrané	$ 3.500.000	$ 3.125.931,00
Bosque Alto	$ 2.500.000	$ 1.918.272,92
Molinos	$ 2.200.000	$ 2.600.583,45
Rayuela	$ 4.000.000	$ 3.812.207,01
Vallarta	$ 4.500.000	$ 4.476.624,08
Villa Bolívar	$ 5.500.000	$ 4.156.317,08
Carolina Campestre	$ 1.500.000	$ 2.252.297,07
Carolina Deluxe	$ 1.500.000	$ 1.865.318,30
Conjunto El Verjel	$ 2.500.000	$ 2.216.812,89
Villa Verona	$ 2.500.000	$ 3.668.284,00
Valena	$ 2.500.000	$ 2.412.263,78

Se realiza una gráfica que permita ver la relación entre los valores:

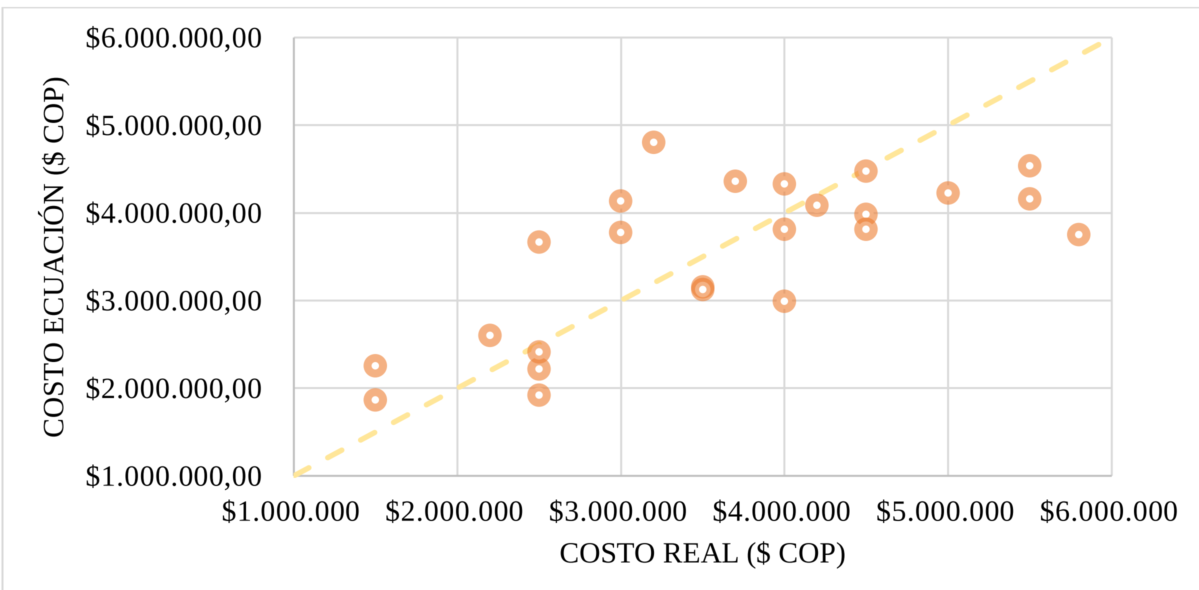


Figura 126. Relación costo real y costo ecuación, conjunto de viviendas unifamiliares: Número de viviendas (und)

La (figura 126) permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse a medida que se acercan a la diagonal, lo cual muestra que el modelo es bastante aceptable y que existe relación entre ellos.

4.3.2. Residencia Multifamiliar

Tabla 22. Ecuaciones seleccionadas, residencia multifamiliar

Clasificación	Variables explicativas	Modelo funcional	Ecuación. Regresión Lineal
Edificio	Área (Ha)	Lineal	(26) $Costo\ (\$\ COP) = 2063953 + 30798895 * Área\ (Ha)$
	Área (Ha)	Lineal	(32) $Costo\ (\$\ COP) = 4125410 + 1565349 * Área\ (Ha)$
Conjunto de edificios	L. de tubería (m)	Potencial	(39) $Costo\ (\$\ COP) = 252281{,}2965 * Longitud\ (m)^{0{,}44591}$
	N. de apartamentos (und)	Potencial	(45) $Costo\ (\$\ COP) = 614399{,}1025 * Apartamentos\ (und)^{0{,}4001}$

Para evaluar las ecuaciones se realiza el siguiente procedimiento:

<u>Edificio, Área (Ha)</u>

Se obtiene la ecuación del modelo funcional lineal con el método de regresión lineal:

$$Costo\ (\$\ COP) = 2063953 + 30798895 * Área\ (Ha)$$

Se seleccionan de manera aleatoria tres proyectos hidráulicos para verificar la eficacia de la ecuación:

1. Proyecto Valmiera, realizado en el municipio de Cúcuta en el año 2019 con un área de 0,036 (Ha) y con un costo de $1.200.000 COP.

$$Costo(\$\ COP) = 2063953 + (30798895 * 0{,}036)$$

$$Costo = \$3.172.713{,}22\ COP$$

2. Proyecto El Contento, realizado en el municipio de los Cúcuta en el año 2021 con un área de 0,018 (Ha) y con un costo de $4.200.000 COP.

$$Costo(\$\ COP) = 2063953 + (30798895 * 0{,}018)$$

$$Costo = \$2.618.333,11\ COP$$

3. Proyecto Edificio Sabana, realizado en el municipio de Los Patios en el año 2021 con un área de 0,007 (Ha) y con un costo de $ 1.300.000 COP.

$$Costo(\$\ COP) = 2063953 + (30798895 * 0,007)$$

$$Costo = \$2.279.545,265\ COP$$

Siguiendo este mismo procedimiento, se calcula el costo para cada proyecto con la ecuación seleccionada con el fin de hacer una comparación con el costo real:

Tabla 23. Comparación de costos, edificio multifamiliar: Área (Ha)

Nombre proyecto	Costo real ($ COP)	Costo ecuación ($ COP)
Los Estoraques ciudadela	$ 6.000.000	$ 2.772.327,59
Valmiera	$ 1.200.000	$ 3.172.713,22
Colsag La Valle	$ 4.000.000	$ 3.326.707,70
Casa los faroles	$ 900.000	$ 2.587.534,22
DUO Condominio	$ 10.000.000	$ 8.747.313,22
El Contento	$ 4.200.000	$ 2.618.333,11
Local comercial, El Buque	$ 1.500.000	$ 2.156.349,69
Iconik	$ 4.000.000	$ 6.375.798,30
Carlos Vásquez	$ 1.800.000	$ 2.649.132,01
Sabana	$ 1.300.000	$ 2.279.545,27
Torres de Niza	$ 3.200.000	$ 2.864.724,27
Oporto	$ 4.500.000	$ 3.049.517,64

Se realiza una gráfica que permita ver la relación entre los valores:

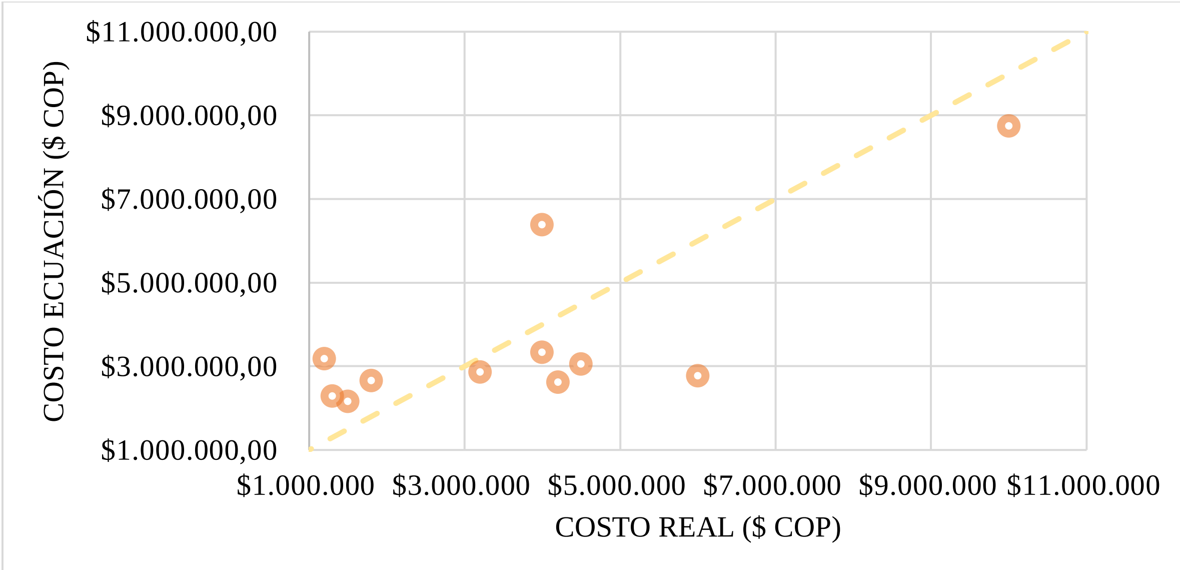


Figura 127. Relación costo real y costo ecuación, edificio multifamiliar: Área (Ha)

La (figura 127) permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse a medida que se acercan a la diagonal, lo cual muestra que el modelo es bastante aceptable y que existe relación entre ellos.

<u>Conjunto de edificios, Área (Ha)</u>

Se obtiene la ecuación del modelo funcional lineal con el método de regresión lineal:

$$Costo\ (\$\ COP) = 4125410 + 1565349 * Área\ (Ha)$$

Se seleccionan de manera aleatoria tres proyectos hidráulicos para verificar la eficacia de la ecuación:

1. Proyecto Arkadia, realizado en el municipio de Cúcuta en el año 2018 con un área de 0,972 (Ha) y con un costo de $6.000.000 COP.

$$Costo\ (\$\ COP) = 4125410 + 1565349 * 0{,}972$$

$$Costo = \$5.646.929{,}23\ COP$$

2. Proyecto Edificios la Manuela, realizado en el municipio de Cúcuta en el año 2019 con un área de 0,366 (Ha) y con un costo de $2.200.000 COP.

$$Costo\ (\$\ COP) = 4125410 + 1565349 * 0{,}366$$

$$Costo = \$4.698.327{,}73\ COP$$

3. Proyecto Bosque de la Variante, realizado en el municipio de Los Patios en el año 2020 con un área de 1,026 (Ha) y con un costo de $ 6.500.000 COP.

$$Costo\ (\$\ COP) = 4125410 + 1565349 * 1{,}026$$

$$Costo = \$5.731.458{,}07\ COP$$

Siguiendo este mismo procedimiento, se calcula el costo para cada proyecto con la ecuación seleccionada con el fin de hacer una comparación con el costo real:

Tabla 24. Comparación de costos, conjunto de edificios: Área (Ha)

Nombre proyecto	Costo real ($ COP)	Costo ecuación ($ COP)
Altas Torres de Claret	$ 4,000,000	$ 8.032.521,10
Altos Zulia VIP	$ 7,000,000	$ 6.021.047,64
Urb. San Rafael	$ 2,000,000	$ 4.992.613,35
Urb. Santa Eduviges	$ 5,000,000	$ 5.817.552,27
Altos del Este	$ 6,000,000	$ 4.206.808,15
Arkadia	$ 6,000,000	$ 5.646.929,23
Chibará	$ 9,200,000	$ 6.179.147,89
Palma Redonda	$ 8,000,000	$ 6.136.883,47
Nuevo Pamplona	$ 8,500,000	$ 5.496.655,72
Altos del Jardín	$ 4,000,000	$ 5.047.400,56
Edificios La Manuela	$ 2,200,000	$ 4.698.327,73
Altos de la Candelaria	$ 5,500,000	$ 4.992.613,35
La Palmita	$ 4,000,000	$ 4.615.364,24
Ciudadela Las flores	$ 5,500,000	$ 7.332.810,10
Terraviva	$ 12,000,000	$ 7.658.402,69

Torres del Norte	$ 8,000,000	$ 5.033.312,42
Urb Santa Eduviges	$ 4,700,000	$ 5.817.552,27
Altos del Moral	$ 3,670,000	$ 4.970.698,46
Chantilly	$ 7,700,000	$ 6.451.518,61
Bosque de La Variante	$ 6,500,000	$ 5.731.458,07
PALLADIUM	$ 4,500,000	$ 6.154.102,30
Tulipanes	$ 5,500,000	$ 5.316.640,59
Entre lomas	$ 6,000,000	$ 7.288.980,33
Brisas del jardín	$ 5,000,000	$ 5.316.640,59
Ikaria	$ 4,800,000	$ 5.451.260,60
Torres Bosques del Este	$ 4,000,000	$ 5.551.442,94
Villas de Turin Baru	$ 4,200,000	$ 6.878.858,89
Villas de San Diego	$ 10,500,000	$ 7.132.445,43

Se realiza una gráfica que permita ver la relación entre los valores:

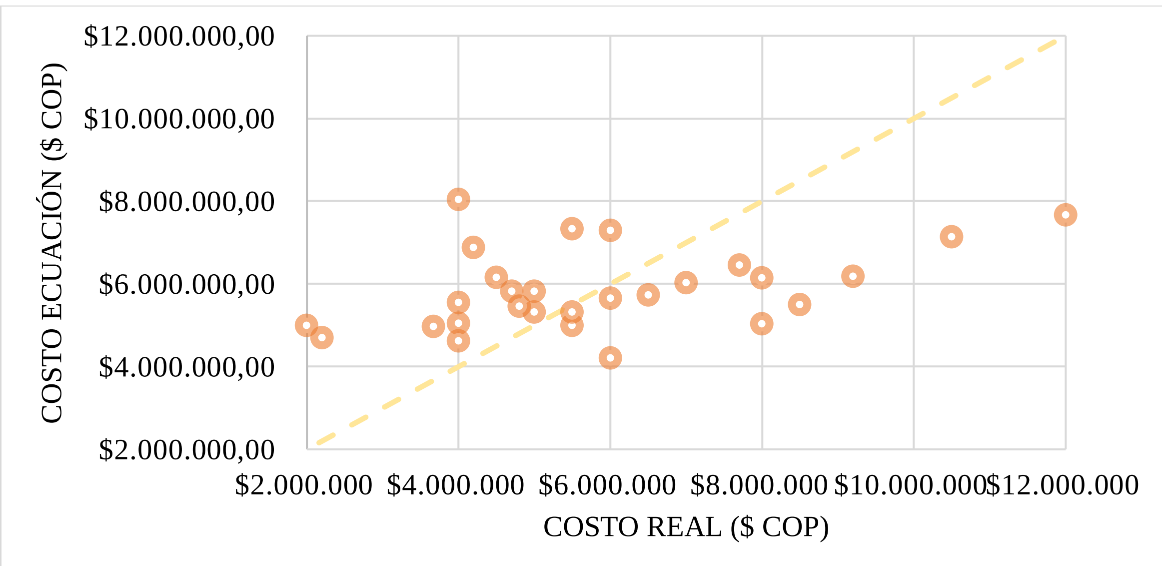


Figura 128. Relación costo real y costo ecuación, Conjunto de edificios: Área (Ha)

La (figura 128) permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse a medida que se acercan a la diagonal, lo cual muestra que el modelo es bastante aceptable y que existe relación entre ellos.

Conjunto de edificios, Longitud de tubería (m)

Se obtiene la ecuación del modelo funcional potencial con el método de regresión lineal:

$$Costo\ (\$\ COP) = 252281{,}2965 * Longitud\ (m)^{0{,}44591}$$

Se seleccionan de manera aleatoria tres proyectos hidráulicos para verificar la eficacia de la ecuación:

1. Proyecto Arkadia, realizado en el municipio de Cúcuta en el año 2018 con una longitud de tubería de 1187,3 (m) y con un costo de $6.000.000 COP.

$$Costo\ (\$\ COP) = 252281{,}2965 * (1187{,}3)^{0{,}44591}$$

$$Costo = \$5.927.373{,}75\ COP$$

2. Proyecto Edificios la Manuela, realizado en el municipio de Cúcuta en el año 2019 con una longitud de tubería de 204,83 (m) y con un costo de $2.200.000 COP.

$$Costo\ (\$\ COP) = 252281{,}2965 * 204{,}83^{0{,}44591}$$

$$Costo = \$2.707.438{,}24\ COP$$

3. Proyecto Bosque de la Variante, realizado en el municipio de Los Patios en el año 2020 con una longitud de tubería de 1220,42 (m) y con un costo de $ 6.500.000 COP.

$$Costo\ (\$\ COP) = 252281{,}2965 * 1220{,}42^{0{,}44591}$$

$$Costo = \$6.000.541{,}22\ COP$$

Siguiendo este mismo procedimiento, se calcula el costo para cada proyecto con la ecuación seleccionada con el fin de hacer una comparación con el costo real:

Tabla 25. Comparación de costos, conjunto de edificios: Longitud de tubería (m)

Nombre proyecto	Costo real ($ COP)	Costo ecuación ($ COP)
Altas Torres de Claret	$ 4,000,000	$ 6.108.065,65
Altos Zulia VIP	$ 7,000,000	$ 4.554.517,79
Urb. San Rafael	$ 2,000,000	$ 2.562.224,66
Urb. Santa Eduviges	$ 5,000,000	$ 6.404.361,54
Altos del Este	$ 6,000,000	$ 5.126.765,87
Arkadia	$ 6,000,000	$ 5.927.373,75
Chibará	$ 9,200,000	$ 6.057.298,63
Palma Redonda	$ 8,000,000	$ 5.712.980,55
Nuevo Pamplona	$ 8,500,000	$ 5.826.716,48
Altos del Jardín	$ 4,000,000	$ 6.278.598,03
Edificios La Manuela	$ 2,200,000	$ 2.707.438,24
Altos de la Candelaria	$ 5,500,000	$ 4.167.345,65
La Palmita	$ 4,000,000	$ 4.057.855,39
Ciudadela Las flores	$ 5,500,000	$ 7.417.453,29
Terraviva	$ 12,000,000	$ 7.933.339,98
Torres del Norte	$ 8,000,000	$ 5.500.346,67
Urb Santa Eduviges	$ 4,700,000	$ 5.431.509,49
Altos del Moral	$ 3,670,000	$ 3.927.533,80
Chantilly	$ 7,700,000	$ 7.592.305,77
Bosque de La Variante	$ 6,500,000	$ 6.000.541,22
PALLADIUM	$ 4,500,000	$ 6.044.018,85
Tulipanes	$ 5,500,000	$ 4.554.517,79
Entre lomas	$ 6,000,000	$ 7.430.209,87
Brisas del jardín	$ 5,000,000	$ 4.554.517,79
Ikaria	$ 4,800,000	$ 4.711.139,26
Torres Bosques del Este	$ 4,000,000	$ 4.744.748,99
Villas de Turin Baru	$ 4,200,000	$ 7.241.333,63
Villas de San Diego	$ 10,500,000	$ 9.179.686,38

Se realiza una gráfica que permita ver la relación entre los valores:

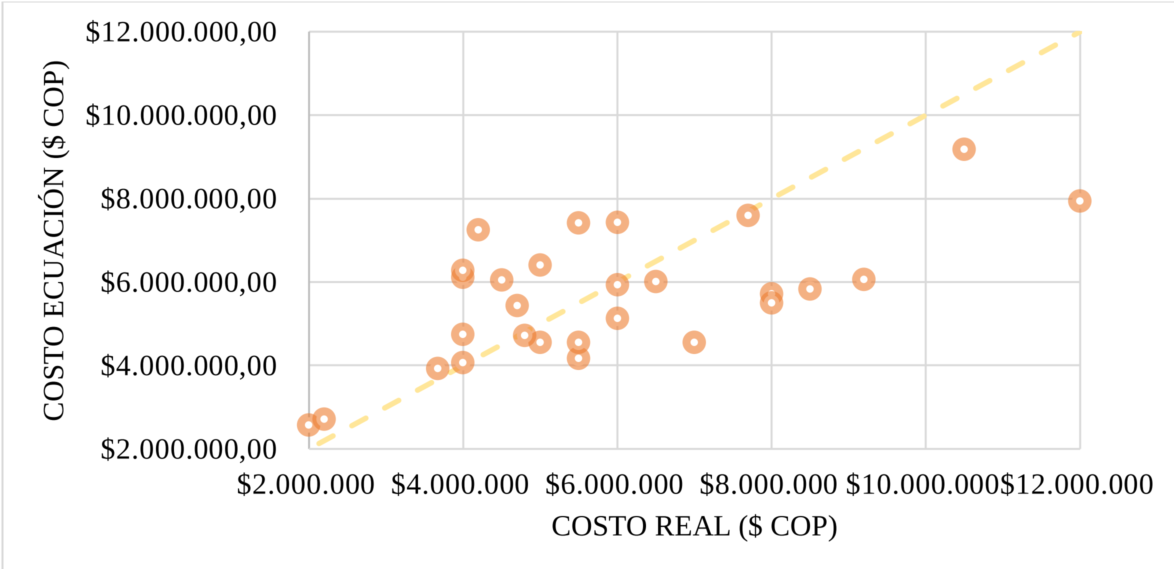


Figura 129. Relación costo real y costo ecuación, Conjunto de edificios: Longitud de tubería (m)

La (figura 129) permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse a medida que se acercan a la diagonal, lo cual muestra que el modelo es bastante aceptable y que existe relación entre ellos.

Conjunto de edificios, Número de apartamentos (und)

Se obtiene la ecuación del modelo funcional potencial con el método de regresión lineal:

$$Costo\ (\$\ COP) = 614399{,}1025 * Apartamentos\ (und)^{0{,}4001}$$

Se seleccionan de manera aleatoria tres proyectos hidráulicos para verificar la eficacia de la ecuación:

1. Proyecto Arkadia, realizado en el municipio de Cúcuta en el año 2018 con un total de 160 (und) apartamentos y con un costo de $6.000.000 COP.

$$Costo\ (\$\ COP) = 614399{,}1025 * (160)^{0{,}4001}$$

$$Costo = \$4.680.788{,}06\ COP$$

2. Proyecto Edificios la Manuela, realizado en el municipio de Cúcuta en el año 2019 con un total de 80 (und) apartamentos y con un costo de $2.200.000 COP.

$$Costo\ (\$\ COP) = 614399{,}1025 * (80)^{0{,}4001}$$

$$Costo = \$3.547.128{,}13\ COP$$

3. Proyecto Bosque de la Variante, realizado en el municipio de Los Patios en el año 2020 con un total de 288 (und) apartamentos y con un costo de $ 6.500.000 COP.

$$Costo\ (\$\ COP) = 614399{,}1025 * (288)^{0{,}4001}$$

$$Costo = \$5.921.796{,}88\ COP$$

Siguiendo este mismo procedimiento, se calcula el costo para cada proyecto con la ecuación seleccionada con el fin de hacer una comparación con el costo real:

Tabla 26. Comparación de costos, conjunto de edificios: Numero de apartamentos (und)

Nombre proyecto	Costo real ($ COP)	Costo ecuación ($ COP)
Altas Torres de Claret	$ 4,000,000	$ 8.150.855,24
Altos Zulia VIP	$ 7,000,000	$ 5.505.199,86
Urb. San Rafael	$ 2,000,000	$ 3.815.551,26
Urb. Santa Eduviges	$ 5,000,000	$ 5.117.910,35
Altos del Este	$ 6,000,000	$ 4.773.059,15
Arkadia	$ 6,000,000	$ 4.680.788,06
Chibará	$ 9,200,000	$ 6.753.591,30
Palma Redonda	$ 8,000,000	$ 5.921.796,88
Nuevo Pamplona	$ 8,500,000	$ 4.680.788,06
Altos del Jardín	$ 4,000,000	$ 4.906.650,25
Edificios La Manuela	$ 2,200,000	$ 3.547.128,13
Altos de la Candelaria	$ 5,500,000	$ 4.906.650,25
La Palmita	$ 4,000,000	$ 3.244.166,84
Ciudadela Las flores	$ 5,500,000	$ 6.238.090,88
Terraviva	$ 12,000,000	$ 7.264.658,30

Torres del Norte	$ 8,000,000	$ 5.076.708,49
Urb Santa Eduviges	$ 4,700,000	$ 5.117.910,35
Altos del Moral	$ 3,670,000	$ 4.334.032,89
Chantilly	$ 7,700,000	$ 6.176.765,01
Bosque de La Variante	$ 6,500,000	$ 5.921.796,88
PALLADIUM	$ 4,500,000	$ 6.474.812,60
Tulipanes	$ 5,500,000	$ 5.505.199,86
Entre lomas	$ 6,000,000	$ 7.016.103,52
Brisas del jardín	$ 5,000,000	$ 5.505.199,86
Ikaria	$ 4,800,000	$ 6.416.855,98
Torres Bosques del Este	$ 4,000,000	$ 5.505.199,86
Villas de Turin Baru	$ 4,200,000	$ 6.328.419,90
Villas de San Diego	$ 10,500,000	$ 6.019.311,06

Se realiza una gráfica que permita ver la relación entre los valores:

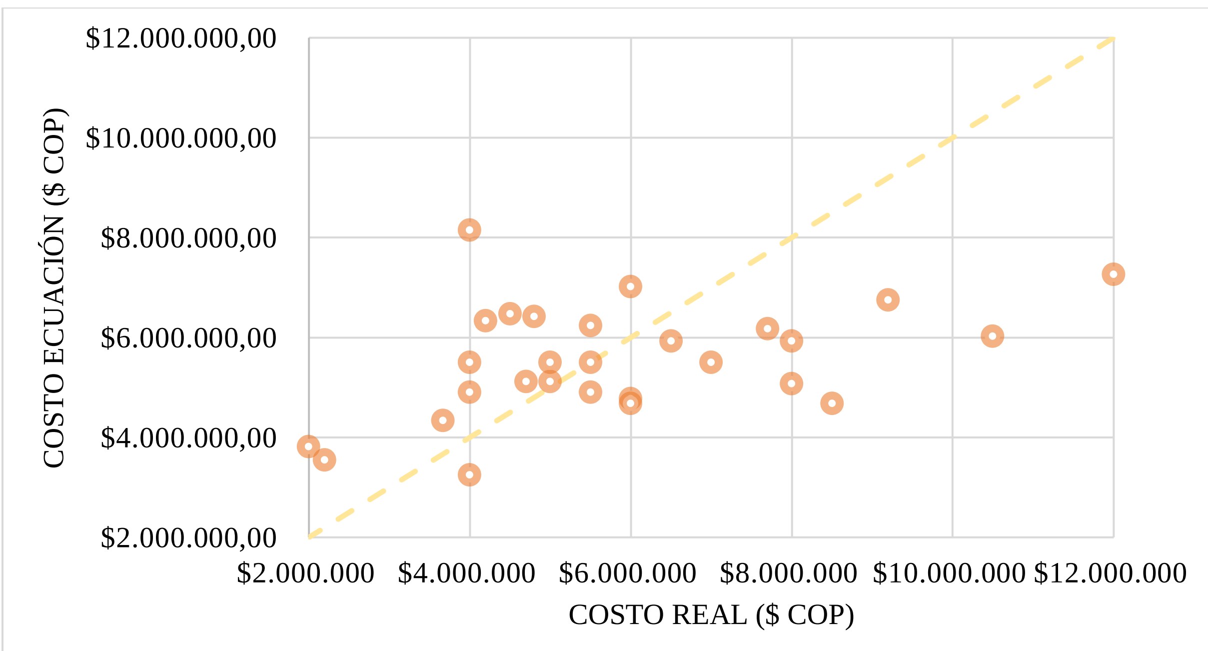


Figura 130. Relación costo real y costo ecuación, Conjunto de edificios: Número de apartamentos (und)

La (figura 130) permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse a medida que se acercan a la diagonal, lo cual muestra que el modelo es bastante aceptable y que existe relación entre ellos.

4.3.3. Otros proyectos

Al analizar todos los modelos funcionales aplicados en la clasificación "otros proyectos" se observa que los coeficientes de determinación R^2 son bastante bajos, llegando a ser el más alto el del modelo potencial de 0,1987, un valor que está muy distante a uno (1).

Al realizar la prueba Shapiro-Wilk que plantea la hipótesis nula de que una muestra proviene de una distribución normal, al tener un nivel de significancia mayor a 0,05, se observa que el p-value obtenido en el modelo funcional potencial es de 2,296e-07 siendo este menor al 5% requerido, lo cual indica que los residuales no cumplen con la distribución normal para este modelo. La prueba de normalidad tampoco cumple para los modelos funcionales: lineal, exponencial, logarítmico y polinómico, debido a que los patrones definidos o la tendencia en los residuales indican existencia de heteroscedasticidad ("hetero" distinta, "cedasticidad" dispersión), en otras palabras, su variabilidad no es constante.

Una de las razones de la existencia de heteroscedasticidad es por la presencia de datos atípicos o aberrantes. Una observación atípica es la que es muy diferente (muy pequeña o muy grande) en relación con las demás observaciones en la muestra. De manera más precisa, un dato atípico es una observación que proviene de una población distinta a la que genera las demás observaciones de la muestra. La inclusión o exclusión de una observación de este tipo, en especial si el tamaño de la muestra es pequeño, puede alterar sustancialmente los resultados del análisis de regresión (Gujarati & Porter, 2010)

Teniendo en cuenta lo dicho anteriormente, es necesario introducir medidas correctivas para una futura investigación, existen dos enfoques para remediar el problema de heteroscedasticidad: cuando se conoce σ_i^2 y cuando no se conoce σ_i^2. Cuando se conoce σ_i^2 el método más directo para corregir la heteroscedasticidad es con los mínimos cuadrados ponderados, sin embargo, existen métodos diferentes cuando no se conoce σ_i^2 se utilizan los métodos como: Varianzas y errores estándar consistentes con heteroscedasticidad de White o Supuestos razonables sobre el patrón de heteroscedasticidad. (Gujarati & Porter, 2010).

Otra forma de solucionar el problema en futuras investigaciones tener en cuenta el tipo de proyecto analizado, es decir, clasificarlos ya sea en colegios, piscinas, canchas, centros comerciales, restaurantes, etc. y realizar un análisis estadístico individual para obtener un modelo que presente menor error y mayor alcance; para este caso se recomienda una muestra de mayor tamaño con el fin de tener variabilidad en los datos analizados.

Aunque los modelos no cumplen con el supuesto de homocedasticidad, se evalúa la ecuación seleccionada del modelo funcional potencial con el fin de analizar los costos que se obtienen de ella, y compararlos con los costos reales de los proyectos.

Tabla 27. Ecuaciones seleccionadas, otros proyectos

Clasificación	Variables explicativas	Modelo funcional	Ecuación. Regresión Lineal
Otros proyectos	Área (Ha)	Potencial	(52) $Costo(\$\ COP) = 2299563 * Área\ (Ha)^{0,11391}$

Para evaluar las ecuaciones se realiza el siguiente procedimiento:

<u>Otros proyectos, Área (Ha)</u>

Se obtiene la ecuación del modelo funcional potencial con el método de regresión lineal:

$$Costo\ (\$\ COP) = 2299563 * Área\ (Ha)^{0,11391}$$

Se seleccionan de manera aleatoria tres proyectos hidráulicos para verificar la eficacia de la ecuación:

1. Proyecto Villa Silvania, realizado en el municipio de Cúcuta en el año 2019 con un área de 0,229 (Ha) y con un costo de $1.700.000 COP.

$$Costo\ (\$\ COP) = 2299563 * (0{,}229^{0,11391})$$

$$Costo = \$1.944.124{,}942\ COP$$

2. Proyecto Edificio Ingeniería Civil UFPS, realizado en el municipio de Cúcuta en el año 2019 con un área de 0,141 (Ha) y con un costo de $3.500.000 COP.

$$Costo\ (\$\ COP) = 2299563 * (0{,}141^{0,11391})$$

$$Costo = \$1.839.640{,}073\ COP$$

3. Proyecto Estación de policía Torcoroma, realizado en el municipio de Cúcuta en el año 2019 con un área de 0,110 (Ha) y con un costo de $ 1.600.000 COP.

$$Costo(\$\ COP) = 2299563 * (0{,}110^{0,11391})$$

$$Costo = \$1.788.341{,}081\ COP$$

Siguiendo este mismo procedimiento, se calcula el costo para cada proyecto con la ecuación seleccionada con el fin de compararlo con el costo real:

Tabla 28. Comparación de costos, otros proyectos: Área (Ha)

Nombre proyecto	Costo real ($ COP)	Costo ecuación ($ COP)
Piscina semiolímpica Villa Silvania	$ 1.200.000	$ 1.922.940,77
Colegio Francisco Fernández	$ 1.500.000	$ 1.914.368,94
Plan parcial Alameda del Este	$ 18.500.000	$ 3.782.358,30
Acueducto Zona Norte	$ 4.170.000	$ 3.903.746,70
Centro Comercial Jardín Plaza	$ 18.000.000	$ 3.140.897,31
Cancha sintética 20 de Julio	$ 1.500.000	$ 2.215.103,67
Cancha sintética Carora	$ 1.500.000	$ 2.039.044,03
Optimización red de agua potable	$ 5.500.000	$ 2.648.814,05
Colector Campoverde	$ 2.000.000	$ 3.499.349,59
Colectores Alcazares Villa Silvania	$ 1.200.000	$ 3.909.646,04
Instituto Técnico La Sabana	$ 1.800.000	$ 1.684.318,71
Colegio Km8	$ 1.700.000	$ 1.859.584,84
Casa Santa Eduviges	$ 1.500.000	$ 1.758.724,48
Box Culvert Chinácota	$ 1.500.000	$ 1.665.839,15
Casa de la cultura Pamplonita	$ 1.300.000	$ 1.569.640,16
Colegio San Bartolomé Comuneros	$ 1.700.000	$ 2.086.003,80
Casino Atlantis Jardín Plaza	$ 1.400.000	$ 1.711.999,13
Edificio Ingeniería Civil UFPS	$ 3.500.000	$ 1.839.640,07
Colector Vista Hermosa	$ 2.000.000	$ 3.381.945,15
Alcantarillado Pluvial Las Américas	$ 3.000.000	$ 1.598.186,73
Baterías sanitarias UNICEF	$ 3.000.000	$ 1.186.490,31
Villa Silvania	$ 1.700.000	$ 1.944.124,94
Estación de Policía Torcoroma	$ 1.600.000	$ 1.788.341,08
Restaurante MONNO	$ 1.300.000	$ 1.517.381,99
EDS Los Patios	$ 4.500.000	$ 1.760.823,51
CEMCU	$ 4.750.000	$ 2.380.334,52
LECS Mall	$ 2.000.000	$ 1.967.206,36
CACI-VR	$ 3.000.000	$ 2.156.934,80
Cubierta cancha colegio de Cáchira	$ 1.400.000	$ 1.662.598,54
Alcantarillado Chitagá	$ 1.300.000	$ 2.615.871,00
Puente Simón Bolívar CENAF	$ 2.000.000	$ 2.557.744,79
Box Culvert Cormoranes	$ 2.000.000	$ 3.778.720,39
Plaza de mercado Cucutilla	$ 3.500.000	$ 1.727.073,58
Baterías sanitarias Nuestra Sra. de Torcoroma	$ 800.000	$ 1.306.713,66
Baterías sanitarias corregimiento Filo Gringo	$ 700.000	$ 1.186.490,31
Baterías sanitarias corregimiento Oro	$ 700.000	$ 1.226.015,51
Casa de gobierno e internado Tibú	$ 1.200.000	$ 1.602.579,71

CDI Comuneros	$ 1.100.000	$ 2.124.980,68
Mega colegio La Frontera	$ 2.500.000	$ 1.435.742,71
Cubierta cancha colegio de Tibú	$ 1.600.000	$ 2.219.278,31

Se realiza una gráfica que permita ver la relación entre los valores:

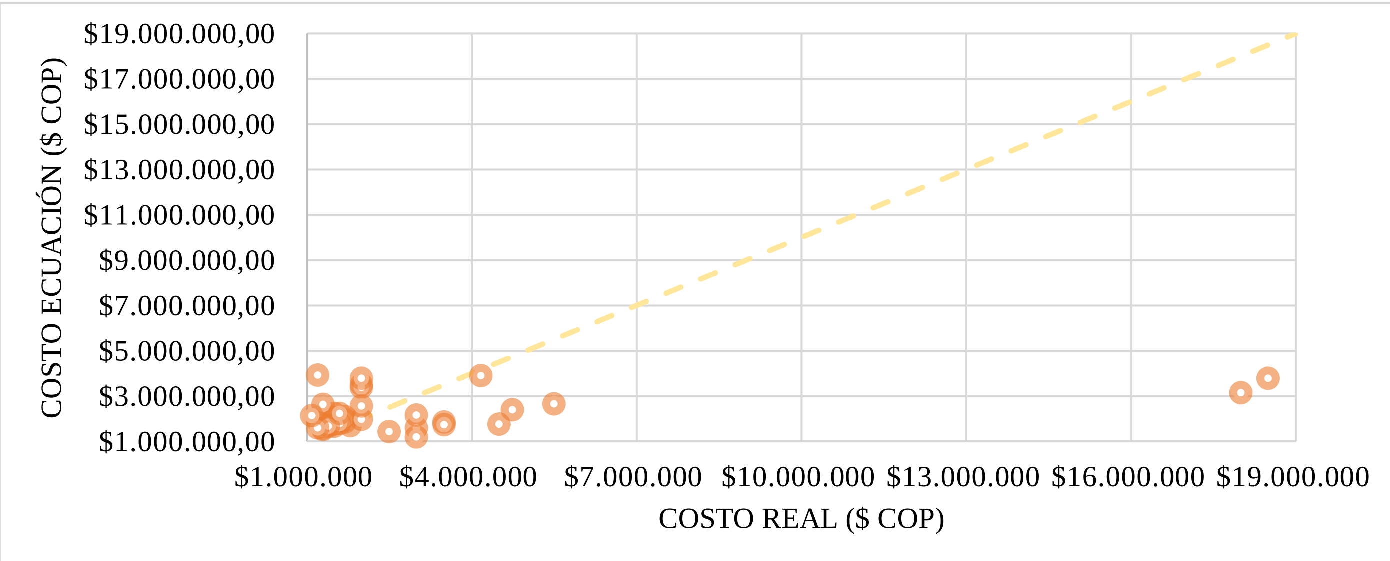


Figura 131. Relación costo real y costo ecuación, otros proyectos: Área (Ha)

La (figura 131) permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse a medida que se acercan a la diagonal, lo cual muestra que existe poca relación entre ellos debido a que el modelo no cumple con la distribución normal.

Conclusiones

- El estado del arte consta de nueve referentes teóricos internacionales y seis referentes teóricos nacionales, dentro de los cuales ningún autor da respuesta a la problemática planteada en el presente trabajo de grado, sin embargo, se encontraron autores como Nestor Enrique Tous Chimá con el proyecto *Estructura de costos en sistemas de alcantarillado* y Carlos David Peinado Calao con el proyecto *Ecuaciones de costo para el diseño optimizado de redes de agua potable y alcantarillado*, donde se asemejan a la búsqueda de ecuaciones sin embargo las emplean en el costo de la construcción de acueductos y alcantarillados, por lo tanto las variables de estudio son distintas.

- Se recolectaron en total 120 proyectos de diseños hidráulicos, que se dividieron en tres categorías: 40 en residencias unifamiliares (16 viviendas y 24 conjunto de viviendas), 40 en residencias multifamiliares (12 edificios y 28 conjunto de edificios) y 40 en otros proyectos (comerciales e instituciones), estos proyectos fueron suministrados por empresas consultoras especializadas de la región de Cúcuta como lo son Hidraforcis SAS, SATAR ingeniería, CABG ingeniería y el consultor Iván Hernando Ramírez Mendoza.

- De acuerdo con los modelos funcionales seleccionados por obtener un mayor alcance estadístico, se logró observar en las categorías “residencia unifamiliar” y “residencia multifamiliar” que las variables explicativas estudiadas de forma independiente para “conjunto de viviendas” y “conjunto de edificios” son más consistentes que al analizarlas conjuntamente en el modelo de regresión lineal múltiple.

- Dentro del análisis realizado de manera independiente a las variables de estudio, se obtuvo mayor significancia estadística en la categoría de “residencia unifamiliar” con la variable “Número de viviendas (und)”, sin embargo, para la categoría “residencia multifamiliar” tuvo mayor significancia estadística la variable “Longitud de tubería (m)”.

- El análisis estadístico efectuado en todos los modelos funcionales seleccionados muestra que el error estándar de la regresión obtenida es muy insignificante, es decir, se considera que la metodología construida es satisfactoria y aplicable para el alcance establecido.

- En la clasificación “otros proyectos” se refleja la importancia de realizar un análisis estadístico independiente según la tipología del proyecto, debido a que los resultados obtenidos conjuntamente arrojan valores pequeños para el R^2 y no cumplen la distribución normal según la prueba Shapiro-Wilk ya que no se encontraba similitud en los datos de los proyectos recolectados. Esta conclusión se hace tras realizar una pequeña prueba distribuyendo los proyectos recolectados según su tipología, y se pudo observar que el coeficiente de determinación R^2 y el p-value aumentaban su significancia estadística, sin embargo, para resultados más óptimos la muestra en cada tipología debe ser significativa.

- Durante el análisis estadístico por medio del software R, se pudo analizar que algunas variables no tenían mayor alcance en el coeficiente de determinación R^2, es decir, su valor no era próximo a uno (1), sin embargo, esto no significa que se deba sobrevaluar la medida de R^2 en el sentido de creer que mientras más alto sea mejor será el modelo. El R^2 se incrementa conforme se añaden más regresoras al modelo. Lo que reviste mayor importancia es la justificación teórica del modelo elegido, los signos de los coeficientes estimados y su importancia estadística. Si un modelo es bueno conforme a estos criterios, quizá resulte aceptable un modelo con un R^2 menor (Gujarati & Porter, 2010).

- Las ecuaciones metodológicas formuladas, seleccionadas y evaluadas que se recomiendan en el presente proyecto investigativo, son:

Tabla 29. Ecuaciones de estimación de costo de diseño hidráulico

Clasificación	Variables explicativas	Modelo funcional	Ecuación. Regresión Lineal
Vivienda	Área (Ha)	Potencial	(2) $Costo\ (\$\ COP) = 1406507{,}163 * Área\ (Ha)^{0{,}15119}$
Conjunto de viviendas	Área (Ha)	Potencial	(8) $Costo\ (\$\ COP) = 3330090 * Área\ (Ha)^{0{,}27692}$
	L. de tubería (m)	Potencial	(14) $Costo\ (\$\ COP) = 675642{,}9263 * Longitud\ (m)^{0{,}23598}$
	N. viviendas (und)	Potencial	(20) $Costo\ (\$\ COP) = 948512{,}2657 * Viviendas\ (und)^{0{,}29371}$
Edificio	Área (Ha)	Lineal	(26) $Costo\ (\$\ COP) = 2063953 + 30798895 * Área\ (Ha)$
Conjunto de edificios	Área (Ha)	Lineal	(32) $Costo\ (\$\ COP) = 4125410 + 1565349 * Área\ (Ha)$
	L. de tubería (m)	Potencial	(39) $Costo\ (\$\ COP) = 252281{,}2965 * Longitud\ (m)^{0{,}44591}$
	N. apartamentos (und)	Potencial	(45) $Costo\ (\$\ COP) = 614399{,}1025 * Apartamentos\ (und)^{0{,}4001}$
Otros proyectos	Área (Ha)	Potencial	(52) $Costo(\$\ COP) = 2299563 * Área\ (Ha)^{0{,}11391}$

- La evaluación de las ecuaciones permite observar que los costos reales vs los costos estimados con la ecuación tienden a parecerse, lo cual muestra que el modelo es bastante aceptable y que existe relación entre ellos. Aquellos costos obtenidos con la ecuación que son mayores al costo real, indica que el proyecto podía tener mayor precio; del mismo modo, aquellos costos obtenidos con la ecuación que son menores al costo real, indica que el proyecto debía ser de menor precio.

- Las series de ecuaciones formuladas y seleccionadas por los diferentes criterios estadísticos durante esta investigación servirán como base a futuros proyectos que deseen continuar, además de ser ecuaciones funcionales para los ingenieros que deseen conocer el costo de un diseño hidráulico según cumplan con variables y tipologías descritas en la investigación.

Recomendaciones

- Con la realización del análisis estadístico y la formulación de la ecuación para la categoría "otros proyectos" se recomienda que para que los modelos funcionales puedan tener significancia estadística se debe tener en cuenta la clasificación tipológica de los proyectos, es decir, realizar un análisis con proyectos que pertenezcan a la misma tipología y con una muestra de mayor tamaño para poder obtener resultados estadísticos más significativos y ecuaciones con menor margen de error. A su vez, si se desea continuar con la misma base de datos, se recomienda realizar las medidas correctivas para solucionar el problema de heteroscedasticidad del modelo estadístico.

- Durante el proceso de clasificación de proyectos no se tuvo en cuenta la zona socioeconómica de estos, sin embargo, si se quiere llegar a un alto índice de proximidad del costo del diseño en la formulación de ecuaciones, se recomienda tener en cuenta dicho factor para futuras investigaciones.

- Los proyectos hidráulicos recolectados fueron únicamente del departamento de Norte de Santander, sin embargo, para una base de datos con mayor trascendencia a nivel nacional, se recomienda usar proyectos de todos los departamentos de Colombia y ampliar la muestra de estudio.

- Debido a que el proyecto investigativo se elaboró con una base de datos de proyectos hidráulicos con costos de los años 2017 al 2021, se recomienda utilizar el IPC (índice de precios del consumidor), con el fin de mantener la vigencia de las ecuaciones propuestas. Se emplea el IPC en la *(ecuación 2)* como ejemplo.

$$(2)\ Costo\ (\$\ COP) = \left(1406507{,}163 * Área\ (Ha)^{0{,}15119}\right) * IPC$$

Referencias

Asociación Colombiana de la Ingeniería Sismica. (2010). Reglamento Colombiano de Construcciones Sismo Resistentes NSR-10. *Titulo K - Requisitos complementarios*. Bogotá, Colombia.

Baldrich Flórez, L., & López Ceballos, K. (2017). Modelo de cobro para optimización de rentabilidad en el diseño de redes hidrosanitarias (Doctoral dissertation, Posgrado). Medellín, Colombia: Corporación universitaria minuto de Dios.

Chapra, S. C., Canale, R. P., Ruiz, R. S., Mercado, V. H., Diaz, E. M., & Benites, G. E. (2011). *Métodos numéricos para ingenieros. (5ta edicción)* (Vol. 5). Trad.) México: Mc Graw Hill.

Clark, R., Sivaganesan, M., Selvakumar, A., & Sethi, V. (2002). Cost Models for water Supply Distribution Systems. *Journal of Water Resources Planning And Management, 128*(5), 312-321.

Definición ABC. (2009). *definicionabc*. Obtenido de definicionabc: https://www.definicionabc.com/tecnologia/edificacion.php

Egas Infante, B. F. (2020). Determinación de la ecuación de costo para optimizar el diseño de sistemas de alcantarillado en el Cantón Santa Rosa. Machala, Ecuador: Universidad tecnica de Machala.

Flores Tapia, C. E., & Flores Cevallos, K. L. (2021). Pruebas para comprobar la normalidad de datos en procesos productivos: ANDERSON DARLING, RYAN-JOINER, SHAPIRO-WILK Y KOLMOGÓROV-SMIRNOV. *Societas. Revista de Ciencias Sociales y Humanística, 23*(2), 83-97.

Gobierno Regional Cusco. (2018). *minos.vivienda.gob*. Obtenido de minos.vivienda.gob: http://minos.vivienda.gob.pe:8081/Documentos_SICA/modulos/FTA/SECCION%20IV/4.14/889005501_02.%20Manual%20de%20O&M%20-%20Colquepata%20V.0.pdf

Gujarati, D., & Porter, D. (2010). *Econometría (Quinta edición)*. Trad.) México: Mc Graw Hill education.

Gutierrez Martinez, A. (2008). Instructivo técnico para determinar las acciones emergentes en el sistema de instalaciones hidro-sanitarias en edificios multifamiliares. Santa Clara, Cuba: Universidad Central Marta Abreu de las Villas facultad de construcciones.

Hernández, E. D. (Noviembre de 2014). Hidraúlica. Pachuca, México: Universidad Autónoma del Estado de Hidalgo .

Jimenez Chicaeme, C. V., Rativa Grijalba, J. G., & Velandia Cardenas, F. (Junio de 2018). Desarrollo de una aplicación para la estimación de costos de redes hidrosanitarias en edificaciones de vivienda de interés social multifamiliar. Bogotá, Colombia: Universidad Católica de Colombia .

Jiménez Terán, J. M. (2013). Manual para el diseño de sistemas de agua potable y alcantarillado sanitario. Xalpaca, Mexico: Universidad Veracruzana.

Kindelan, J. W. (Abril - Junio de 2017). Metodología con enfoque a procesos para la implementacion de sistemas de costos en las empresas de acueducto en Cuba. *Revista Cubana de Finanzas y Precios* , *1*(2), 29-38.

Leiva Ucharico, C. A. (2015). Estudio comparativo tecnico-economico de la red de alcantarillado conveniconal y condominal en el AA.HH. Lima, Perú: Universidad Ricardo Palma .

Madrid, M., León Mantero, C., Maz Machado, A., & López , E. (2019). El desarrollo del concepto de ecuación en los libros españoles de matemáticas del siglo XVIII. *Investigación en Educación Matemática XXIII*, 403-412.

Michalus, J. C., & Ibarra, M. D. (2006). Un modelo matemático para calcular el costo de manutención de un estudiante de la facultad de ingeniería. *Revista Ingeniería industrial, 5*(1).

Narváez, P. C., & Galeano, H. (2002). Ecuación de costos y función objetivo para la optimización del diseño de redes de flujo de líquidos a presión. *Ingeniería e Investigación*(49), 23-29.

Palomino Pariona , B. W. (2018). Determinación de función de costo en sistemas de agua potable y alcantarillado. Ayacucho, Perú: Universidad Nacional de San Cristobal de Huamanga.

Peinado Calao, C. D. (2016). Ecuaciones de costo para el diseño optimizado de redes de agua potable y alcantarillado. Bogotá, Colombia: Universidad de los Andes.

Ramirez, C., & De Plaza, J. S. (2016). Acueducto y Alcantarillado Guia técnica. Bogotá, Colombia: Universidad Católica de Colombia.

Rodríguez Silva, J. L. (2019). ¿Qué puede hacer el software R para resolver tus problemas? *Revista Digital Universitaria*, *20*(3), 1-10. doi:http://doi.org/10.22201/codeic.16076079e.2019.v20n3.a5.

Roldán, P. N. (2017). *economipedia*. Obtenido de economipedia: https://economipedia.com/definiciones/estadistica.html

Saldarriaga, J., Lopez, L., Paez, D., Luna, D., & González, S. (2017). Diseño optimizado de redes de distribución de agua potable (programa REDES). Bogotá, Colombia : Seminario Iberoamericano de redes de agua. SEREA.

Sampieri Hernández, R., Collado Fernández, C., & Lucio Baptista, P. (2003). *Metodología de la Investigacion.* Mexico: McGraw-Hill Interamericana.

Sánchez Luna, G. (1995). El Urbanismo, la Ciudad y su Tratamiento Jurídico. *Boletín mexicano de derecho comparado, 28*(82), 307-324.

Santana, J. S., & Farfán, E. M. (2014). *El arte de programar en R: un lenguaje para la estadística.* México: Instituto Mexicano de Tecnología del Agua.

Solórzano Solórzano, J. A. (Abril de 2016). Diseño de un sistema financiero-contable para la empresa Arboleda Faini Hidrotecnología cía. ltda., dedicada al servicio de ingeniería hidráulica, ubicada en la ciudad de Quito. Ecuador: (Bachelor's thesis, Quito: UCE.).

Swamee, P. K., & Sharma, A. K. (2013). Optimal design of a sewer line using Linear Programming. *ELSEVIER, 37*(6), 4430-4439.

Tous Chimá, N. (Diciembre de 2013). Estructura de costos en sistemas de alcantarillado. Bogotá, Colombia: Universidad de los Andes .

Walpole, R., & Myres, R. (1998). *Probabilidad y Estadistica para ingenieros (Sexta edición).* México D.F: Trad.) México: Prentice Hall.