

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE POSTGRADO**

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

“MAGÍSTER EN ESTADÍSTICA APLICADA”

TEMA:

**ESTIMACIÓN ESTADÍSTICA DEL ERROR PONDERADO EN LA
MEDICIÓN DE MEDIDORES DE AGUA POTABLE DE CLIENTES
RESIDENCIALES DE GUAYAQUIL**

AUTOR:

DUNIS ELIZABETH LOPEZ CARDENAS

Guayaquil - Ecuador

2024

RESUMEN

Para una empresa que produce, distribuye y comercializa agua es de vital importancia conocer las pérdidas de agua a causa de diferentes componentes, entre ellas la pérdida ocasionada por la imprecisión o desgaste de sus equipos de medición.

El presente proyecto presenta un caso de estudio en el que se estima el error ponderado de medición a través del perfil de consumo y la curva de error del medidor, el cual infiere que la estimación de la precisión del parque de medidores es de un 93,06%, por lo que se deja de registrar el 6,94% del volumen de agua facturado por causa de la submedición que producen los medidores de agua.

Además, se evidencia la relación que existe entre el error ponderado global del parque de medidores, el consumo registrado y su antigüedad.

Palabras Claves: Error de Medición, Error ponderado, Perfil de Consumo, Curva de Error

ABSTRACT

For a company that, produces, distributes and markets water, it is very important to know the water losses due to different components, including the loss caused by the inaccuracy or wear of its measuring equipment.

This project presents a case study in which the weighted measurement error is estimated through the consumption profile and the meter error curve, which infers that the estimate of the precision of the meter park is 93.06%, so 6.94% of the volume of water billed is no longer recorded due to the submetering produced by the water meters

Furthermore, the relationship between the global weighted error of the meter fleet, the recorded consumption and its age is evident.

Keywords: Measurement Error, Weighted Error, Consumption Profile, Error Curve

DEDICATORIA

A aquellos seres maravillosos que la vida me dio, que siempre creyeron y confiaron en mí y que me impulsaron en cada momento de mi vida.

AGRADECIMIENTO

A Dios.
A mi mamá,
A mis hermanas,
A mis hermanos,

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación, me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo, corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Dunis Elizabeth López Cárdenas

TRIBUNAL DE GRADUACIÓN

Purificación Galindo Villardon, Ph.D
PRESIDENTE

Sandra Lorena García Bustos, Ph.D.
TUTOR

Francisco Antonio Moreira Villegas, Mgrt.
DOCENTE EVALUADOR

ABREVIATURAS O SIGLAS

MPE: Error máximo permitido
m³/h: metros cúbicos por hora
m³: metros cúbicos
mm: milímetros
l/h: litros por hora
Dn: diámetro nominal

TABLA DE CONTENIDO

CAPÍTULO 1	1
1. INTRODUCCIÓN.....	1
1.1. Antecedentes	2
1.2. Descripción del problema.....	2
1.3. Objetivos	3
1.3.1. Objetivo General	3
1.3.2. Objetivos específicos	3
1.4. Hipótesis	3
1.5. Alcance.....	3
CAPÍTULO 2	4
2. MARCO TEÓRICO.....	4
2.1. Conceptos y definiciones básicas.....	4
2.1.1. Medidor de agua.	4
2.1.2. Tipos de medidores.....	4
2.1.3. Características metrológicas.....	5
2.1.4. Curva de error o curva de precisión	5
2.1.5. Perfil de Consumo	5
2.1.6. Error de global de medición (error medio ponderado).....	6
2.2. Investigaciones relacionadas	7
CAPÍTULO 3	9
3. METODOLOGÍA.....	9
3.1. Definición de las variables.	9
3.2. Población, fuente de datos y diseño de muestra.	10
3.3. Tratamiento de datos.....	12
3.4. Factores que afectan la precisión de un medidor.....	13
3.5. Análisis de conglomerados.....	13
3.5.1. Medidas de distancia.	14
3.5.2. Métodos de agrupamiento.....	14
3.6. Regresión lineal simple y múltiple.	15
3.6.1. Supuestos.	15
3.7. Modelos lineales generalizados.....	16
3.7.1. Componentes de los modelos lineales generalizados.	16
3.7.2. Estimación de modelos lineales generalizados.....	16

CAPÍTULO 4	17
4. RESULTADOS.....	17
4.1. Análisis descriptivo de los datos.....	17
4.1.1. Variable Promedio histórico.....	17
4.1.2. Variable Consumo mensual.....	18
4.1.3. Edad del medidor.....	19
4.1.4. Variable m ³ registrados.....	20
4.1.5. Variable Error Promedio ponderado.....	21
4.2. Análisis estadístico bivalente.....	22
4.2.1. Correlación de las variables.....	22
4.3. Análisis de clusters.....	23
4.4. Estimación de la precisión global. (El error medio ponderado)	27
4.5. Análisis de regresión para el error ponderado con la edad.....	28
4.5.1. Comprobación de los supuestos error promedio y edad.....	30
4.6. Análisis de regresión para el error ponderado y los m ³ registrados.....	30
4.6.1. Comprobación de los supuestos error promedio y m ³ registrados.....	33
4.7. Regresión lineal múltiple.....	33
4.7.1. Comprobación de los supuestos error promedio, edad y m ³ registrados. 35	
4.8. Regresión lineal múltiple considerando los grupos	35
4.8.1. Comprobación de los supuestos error promedio, edad y m ³ registrados y grupos.....	36
CAPÍTULO 5.....	37
5. CONCLUSIONES Y RECOMENDACIONES.....	37
5.1. Conclusiones.....	37
5.2. Recomendaciones.....	37
6. Referencias	38
7. Apéndices y anexos.....	40

LISTADO DE FIGURAS

Figura 2. 1 Curva de precisión de un medidor Dn15mm Q3 2.5m ³ /h R315	6
Figura 2. 2 Ejemplo de Perfil de consumo	6
Figura 3. 1 Medidores por diámetro	10
Figura 3. 2 Antigüedad del parque de medidores	11
Figura 3. 3 Medidores por marca	11
Figura 3. 4 Medidores de media por tecnología	12
Figura 4. 1 Histograma y diagrama de caja Variable Promedio Histórico	18
Figura 4. 2 Histograma y diagrama de caja Variable consumo mensual	19
Figura 4. 3 Histograma y diagrama de caja Variable edad del medidor	20
Figura 4. 4 Histograma y diagrama de caja Variable m ³ registrados	21
Figura 4. 5 Histograma y diagrama de caja Variable Error promedio ponderado ..	22
Figura 4. 6 Correlación de variables	23
Figura 4. 7 Representación de los grupos	24
<i>Figura 4. 8 Perfil de consumo por grupos</i>	<i>25</i>
Figura 4. 9 Perfil de consumo Residencial	26
Figura 4. 10 Curva de error de medidores	26
Figura 4. 11 Recta de regresión Error ponderando y edad.....	29
Figura 4. 12 Recta de regresión Error ponderando y m ³ registrados.....	32
Figura 4. 13 Función de auto correlación Error ponderando y m ³ registrados	33

LISTADO DE TABLAS

Tabla 3. 1 Variables de estudio	9
Tabla 4. 1 Estadística Descriptiva Variable Promedio Histórico	17
Tabla 4. 2 Estadística Descriptiva Variable Consumo mensual.....	18
Tabla 4. 3 Estadística Descriptiva Variable Edad del medidor.....	19
Tabla 4. 4 Estadística Descriptiva Variable m ³ registrados	20
Tabla 4. 5 Estadística Descriptiva Variable Error promedio Ponderado.....	21
Tabla 4. 6 Calculo Error promedio ponderado	27
Tabla 4. 7 Detalle del modelo Error promedio ponderado y edad	28
Tabla 4. 8 Tabla error promedio ponderado y edad.....	28
Tabla 4. 9 Comparación de modelos	29
Tabla 4. 10 Detalle del modelo Error promedio ponderado y edad.....	29
Tabla 4. 11 Tabla ANOVA	30
Tabla 4. 12 Verificación de supuestos.....	30
Tabla 4. 13 Detalle de modelo m ³ registrados y error ponderado	31
Tabla 4. 14 Tabla m ³ registrados y error ponderado.....	31
Tabla 4. 15 Comparación de modelos.....	31
Tabla 4. 16 Error promedio ponderado y m ³ registrados	32
Tabla 4. 17 Tabla ANOVA	32
Tabla 4. 18 Verificación de supuestos.....	33
Tabla 4. 19 Modelo de regresión lineal	34
Tabla 4. 20 Modelo de regresión lineal mejorado.....	35
Tabla 4. 21 Verificación de supuestos.....	35
Tabla 4. 22 Modelo de regresión lineal	36
Tabla 4. 23 Verificación de supuestos.....	36

CAPÍTULO 1

1. INTRODUCCIÓN

La empresa de este estudio, es una empresa de servicios básicos encargada de los sistemas de agua potable, alcantarillado sanitario y alcantarillado pluvial, teniendo como desafío la mejora continua en los accesos a los servicios de agua potable mediante la operación, mantenimiento y el monitoreo constante de los servicios brindando una respuesta inmediata a la comunidad.

La empresa es responsable del suministro de agua y saneamiento a más de tres millones de habitantes entre zonas urbanas y rurales, para lo cual produce y distribuye anualmente más de 380 millones de m³ de agua potable.

La empresa tiene aproximadamente 550.000 usuarios, según base de datos proporcionada por el área comercial, que disponen de un medidor de agua para la medición de su consumo mensual. El parque de medidores está compuesto por una amplia gama de marcas, modelos y antigüedades.

Asegurar una adecuada relación entre los volúmenes consumidos y facturados de sus clientes es de gran interés para la empresa, en su compromiso de contribuir a la mejora continua de la calidad de vida de los habitantes.

Analizar la relación que existe entre el error de medición de un medidor de agua, el consumo registrado, su antigüedad y otras características de interés y la estimación del error ponderado de la medición, ofrecerá información que ayudará a establecer políticas eficientes de renovación y sustitución de medidores, pautas de dimensionamiento y mejora del conocimiento de la precisión en la medida de los usos y consumos individuales.

1.1. Antecedentes

A pesar de su importancia, la evaluación de la precisión de la medida de los medidores a lo largo de su vida útil y sus condiciones de instalación, es un tema que no se ha abordado de una manera profunda y científica en las empresas de Abastecimiento de Agua. Hay una carencia de normativa sobre la inspección y evaluación de los medidores a lo largo de su vida útil. Tampoco existen criterios homogéneamente utilizados para su renovación en las empresas del sector.

La empresa ha efectuado en los últimos años algunos estudios para valorar la precisión de los medidores residenciales, tanto en laboratorio como en los predios. Las conclusiones de estos estudios sugirieron la realización de un proyecto más detallado, en el que se incluyeran nuevos criterios no contemplados hasta la fecha.

1.2. Descripción del problema

Uno de los principales problemas a los que se enfrenta una empresa que produce y comercializa agua son los altos niveles de pérdidas de agua en los sistemas de distribución. Estas pérdidas de agua representan recursos financieros que no ingresan a la empresa.

El medidor de agua es “la caja registradora” en una empresa de agua potable, ya que se usa para medir el consumo de los clientes y basados en este consumo medido, la empresa puede facturar y recuperar los diferentes gastos que demanda cada uno de sus procesos, de allí, la necesidad de evaluar su desempeño metrológico, a fin de garantizar su eficiencia en la medición. Además, las imprecisiones de los equipos de medición representan un componente importante de las pérdidas en la distribución de un acueducto.

Considerando lo valioso que es el correcto funcionamiento de un medidor tanto para la empresa como para sus clientes es fundamental identificar los diferentes factores que puedan afectar su eficiencia en la medición, analizando sus características y diseñando el modelo ideal que ayude a tomar decisiones para mejorar la gestión global del parque de medidores.

1.3. Objetivos

1.3.1. Objetivo General

Identificar las pérdidas del parque de medidores de la ciudad de Guayaquil mediante análisis univariante y multivariante, estimando así el error ponderado en la medición de medidores de agua mejorando la gestión del mismo.

1.3.2. Objetivos específicos

- Determinar el estado del parque de medidores instalados.
- Identificar posibles factores de afectación que hacen ineficiente el desempeño de los equipos de medición.
- Diseñar un modelo de estimación del error ponderado de la medición conociendo de esta manera su desgaste durante el tiempo.
- Analizar los resultados de las estimaciones obtenidas, emitiendo comentarios y observaciones sobre ellas y sobre el comportamiento del parque de medidores.

1.4. Hipótesis

Los factores que afectan, tanto al desempeño eficiente de los medidores, como al error ponderado global del parque de medidores corresponden, entre otros, al consumo registrado y su antigüedad.

1.5. Alcance

Este estudio está orientado a identificar los factores que afectan el desempeño eficiente de los elementos de medición y estimar el error global de medición del parque de medidores de la empresa encargada de la prestación de servicios públicos de agua potable y alcantarillado de la ciudad de Guayaquil, para lo cual se utilizará la información comercial y datos sobre errores y consumos levantados entre los años 2019 y 2022.

CAPÍTULO 2

2. MARCO TEÓRICO

2.1. Conceptos y definiciones básicas

Para propósitos técnicos, es preciso definir algunos términos que serán empleados a lo largo del documento.

2.1.1. Medidor de agua.

Instrumento destinado a medir continuamente, memorizar y mostrar el volumen de agua que pasa a través de su transductor de medida dentro de las condiciones de medición. Un medidor de agua incluye por lo menos un transductor de medida, un calculador (incluyendo dispositivos de corrección o ajuste, si los hay) y un dispositivo indicador. Estos tres elementos pueden estar en diferentes alojamientos. (International Organization for Standardization [ISO 4064], 2014)

Dispositivo o artefacto perteneciente a una conexión, que mide el consumo de agua proveída por la Empresa. (Registro oficial del Ecuador N° 522, 2018, Artículo 8)

2.1.2. Tipos de medidores.

Hay varios tipos de medidores de agua. Se pueden clasificar según la tecnología que utilizan para medir el caudal o el volumen de agua que fluye por él.

- **Volumétrico.** Medidor compuesto por cámaras de volumen conocido accionadas por el flujo de agua, estas cámaras se llenan y vacían sucesivamente, de acuerdo al número de llenadas y vaciadas este indicara el volumen que atraviesa. (ISO 4064, 2014)
- **Inferencial o de velocidad.** Instrumento instalado en un conducto cerrado, compuesto por un dispositivo móvil (hélice o turbina) accionado directamente por la velocidad del flujo de agua, cuyo movimiento es transmitido por medios mecánicos u otros al dispositivo indicador que totaliza el volumen. Estos a la vez se dividen en chorro único, chorro múltiple y chorro Axial o tipo Woltmann.

2.1.3. Características metrológicas.

- **Error.** Diferencia cuantificada entre un valor medido y un valor de referencia (ISO 4064, 2014)
- **Volumen real.** Volumen total de agua que pasa a través del medidor, sin tener en cuenta el tiempo transcurrido (ISO 4064, 2014)
- **Error máximo permitido (MPE).** Valor extremo del error de medida, con respecto al valor de referencia conocido, permitido por especificaciones o reglamentaciones para un medidor determinado. (ISO 4064, 2014)
- **Caudal.** Cociente del volumen de agua que pasa por el medidor (Volumen registrado en el indicador) y el tiempo que demora en hacerlo. (ISO 4064, 2014)

2.1.4. Curva de error o curva de precisión

La curva de error del medidor representa el error de medición en valor porcentual que presenta un medidor para cada valor de caudal que circula por el mismo. El error de medición se determina mediante un ensayo, comparando el volumen contabilizado por el medidor con el volumen registrado por un elemento de control (medidor patrón, tanque calibrado, etc.), se representa mediante un gráfico logarítmico (Caudal vs. Error%) que muestra los errores de indicación del medidor en función de sus caudales.

El objetivo básico de la curva de errores para la observación y análisis de un medidor consiste en determinar si su grado de exactitud se encuentra en los rangos de tolerancia o márgenes de error permitidos.

En la figura 2.1 se puede observar un ejemplo de una curva de error para un medidor de 15mm de diámetro.

2.1.5. Perfil de Consumo

Característica proveniente del consumo de agua potable en un determinado usuario, relacionado al caudal instantáneo de operación y los volúmenes que fluyen en cada rango de caudal por un periodo significativo, se representa mediante un histograma de porcentajes de volumen consumido para cada rango de caudales, informa de la distribución porcentual del consumo de un cliente por rangos de caudales.

La figura 2.2 muestra un ejemplo de un perfil de consumo.

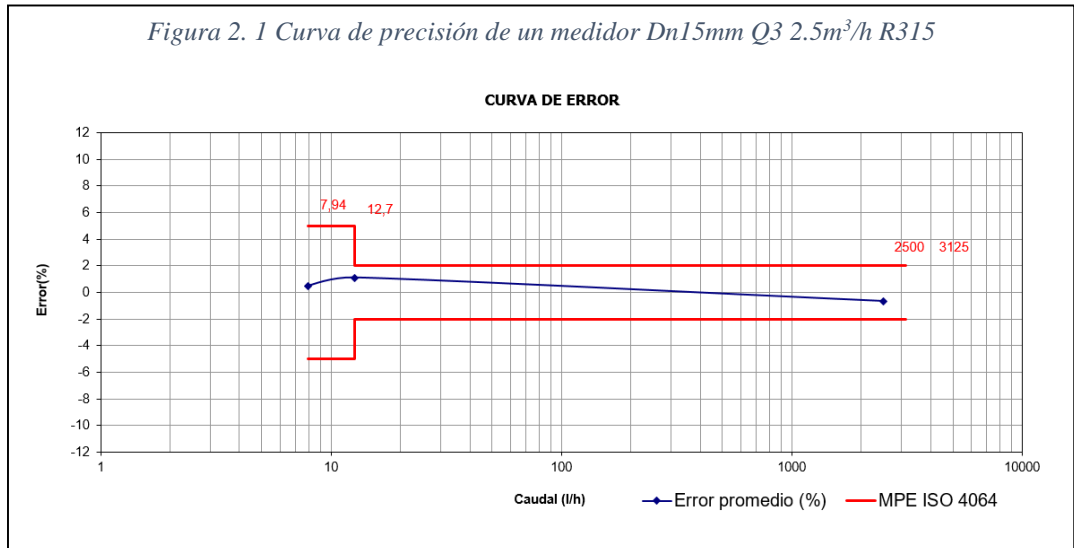
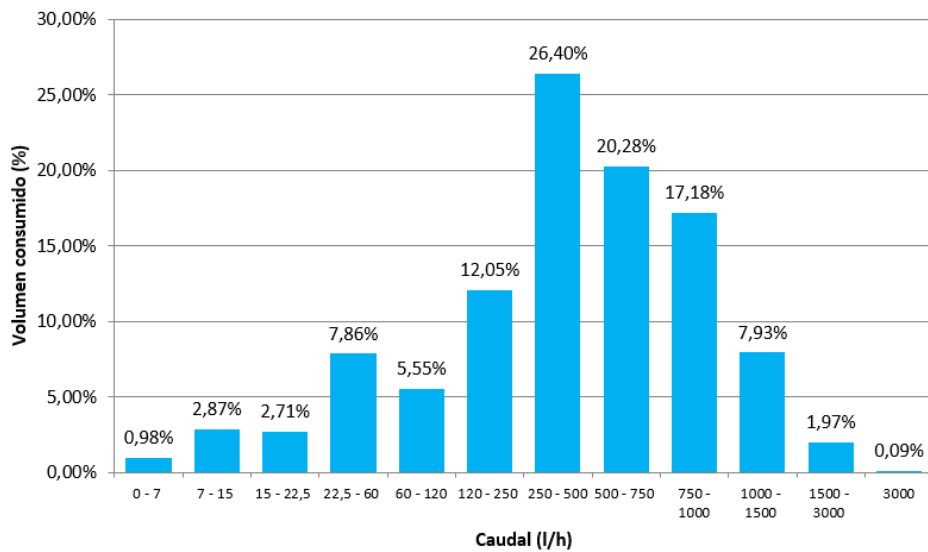


Figura 2.2 Ejemplo de Perfil de consumo



2.1.6. Error de global de medición (error promedio ponderado)

El error de medición global de un medidor es la diferencia entre el volumen contabilizado por este y el realmente consumido.

Como el error de medición de un medidor varía dependiendo del caudal que circula por el mismo, la diferencia entre el volumen contabilizado y el realmente consumido se debe calcular ponderando el error que se produce en cada valor de caudal de consumo con el porcentaje de volumen que el cliente usa en ese mismo valor.

Por tanto, el error de medición global del parque de medidores o error ponderado depende tanto de su curva de error como de la forma en que utilizan el agua los usuarios, es decir, el perfil de consumo y es representado por la siguiente ecuación:

$$Error\ ponderado(\%) = \sum_{i=1}^n errorQ_i * PesoQ_i$$

Donde:

Error Q_i es el error del medidor en cada rango de caudal

Peso Q_i es el volumen consumido en cada rango de caudal /volumen total consumido. (Arriegui, 1998)

2.2. Investigaciones relacionadas

Gonza et al. (2021) realizaron un estudio para identificar los principales factores que afectan la inexactitud de los medidores de agua. Se utilizaron regresiones múltiples para analizar estadísticamente los efectos y grados de relación entre los factores y los errores del medidor de agua. Los hallazgos de esta investigación indican que los principales factores que contribuyeron a la inexactitud del medidor de agua fueron el caudal, antigüedad del servicio del medidor.

Sánchez (2015), en su tesis “Evaluación del error global de contadores de agua domésticos y análisis de su degradación a partir de una muestra real. Análisis económico como criterio de renovación” realiza un análisis del error de medición considerando la edad del medidor mediante regresión lineal simple, cuadrado de edad y raíz cuadrada de edad, de estos modelos el mejor ajuste lo da el modelo de regresión lineal simple. Adicional, realiza un análisis del error de medición considerando el volumen acumulado del medidor, elabora varios modelos concluyendo que el que más se ajusta es el modelo de raíz cuadrada de volumen acumulado. Por último, realiza el análisis del error ponderado considerando la edad y el volumen registrado, mediante modelos de regresión no lineal y regresión múltiple, entre las dos alternativas, a pesar de que la regresión no lineal explica el 77.43% de la variabilidad del error ponderado frente al 75.73% de la regresión múltiple, se puede preferir la segunda alternativa por simplicidad y menos constantes a determinar.

En el artículo, “Flow measurement accuracies of in-service residential water meters” Stoker et al. (2012) investigan la relación entre la degradación de la precisión del medidor y factores como la edad, el desgaste y el rendimiento de los medidores de agua mediante regresión lineal para lo cual se realizaron pruebas en laboratorio de una muestra de medidores y realizaron regresión

lineal entre el error de medidor en caudal mínimo, intermedio y máximo obteniendo como resultado que no existe correlación considerable entre el volumen, edad y el rendimiento del medidor.

Arregui de la Cruz, et al. (1998) realizaron un análisis estadístico de un parque de medidores en función de la forma de consumir de los abonados a fin de calcular la precisión media del mismo y, en consecuencia, el volumen no registrado en los medidores domiciliarios. Este estudio se realizó sobre una muestra, por tanto, los errores ponderados que se calculan son solo estimaciones de los valores reales.

Tortes et al. (2017). En función de la tecnología de medición y del modelo analizado, se aportan modelos combinados que estiman la degradación del error de medición en función de la edad y del volumen acumulado del medidor. Para obtener estos modelos combinados, se han realizado diferentes regresiones múltiples e iteraciones, tratando de buscar los parámetros más óptimos que mejor se adapten a los valores obtenidos consiguiendo como resultado que la degradación del error en función de la edad, excepto para uno de los modelos, ha sido lineal y que la degradación del error en función del volumen acumulado depende directamente del modelo de medidor analizado.

Palau et al (2018). El periodo de funcionamiento metrológico de los medidores se estimó, mediante un gráfico de regresión lineal simple entre el error de medida y el tiempo de uso.

CAPÍTULO 3

3. METODOLOGÍA

En este apartado, se detalla la metodología empleada para lograr los objetivos planteados en este proyecto, encaminados a estimar estadísticamente el error de la medición de medidores de agua.

La metodología implementada en este proyecto tiene un enfoque netamente cuantitativo, ya que se utiliza mediciones y cálculos numéricos, recolección de datos, bases de datos y el análisis estadístico para llegar a responder las preguntas de la investigación.

3.1. Definición de las variables.

Para el análisis, se evaluarán las siguientes variables de estudio:

Tabla 3. 1 Variables de estudio

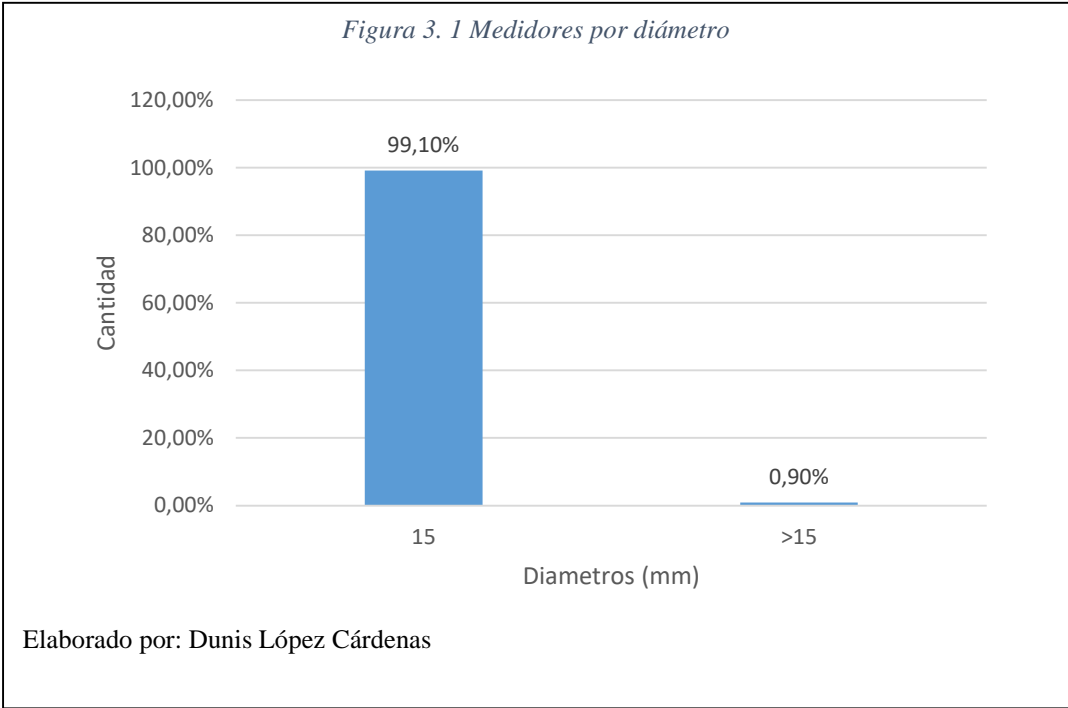
VARIABLES	DESCRIPCION
MARCA DE MEDIDOR	Variable nominal que corresponde al fabricante del medidor
MODELO DE MEDIDOR	Variable nominal que corresponde al modelo de determinada marca
TIPO DE MEDIDOR	Variable nominal que corresponde a la tecnología que utiliza el medidor para contabilizar el volumen de agua que fluye por él.
RANGO/CLASE	Variable nominal que define el rango de trabajo metrológico de un medidor
M ³ REGISTRADOS	Variable numérica que indica la cantidad de metros cúbicos que ha registrado el medidor desde su instalación.
EDAD	Variable numérica que indica los años de antigüedad del medidor.
ERRORES DEL MEDIDOR (12 caudales)	Variables numéricas que corresponden a la diferencia entre el valor medido y el valor de referencia.
CONSUMO	Variable numérica que indica los metros cúbicos consumidos en el último mes.
CONSUMO PROMEDIO	Variable numérica que corresponde al consumo promedio en metros cúbicos de los últimos 6 meses.
VOLUMEN CONSUMO (12 rangos de caudal)	Variable numérica que corresponde al volumen consumido en cada rango de caudal.
CICLO	Variable nominal que corresponde a una división realizada por el área comercial de la empresa.
ZONA	Variable nominal. Está compuesta por los ciclos

ERROR PONDERADO	Variable numérica que resulta de la suma total de porcentaje de volumen que consume un cliente por el error del medidor en el mismo rango de caudal.
-----------------	--

3.2. Población, fuente de datos y diseño de muestra.

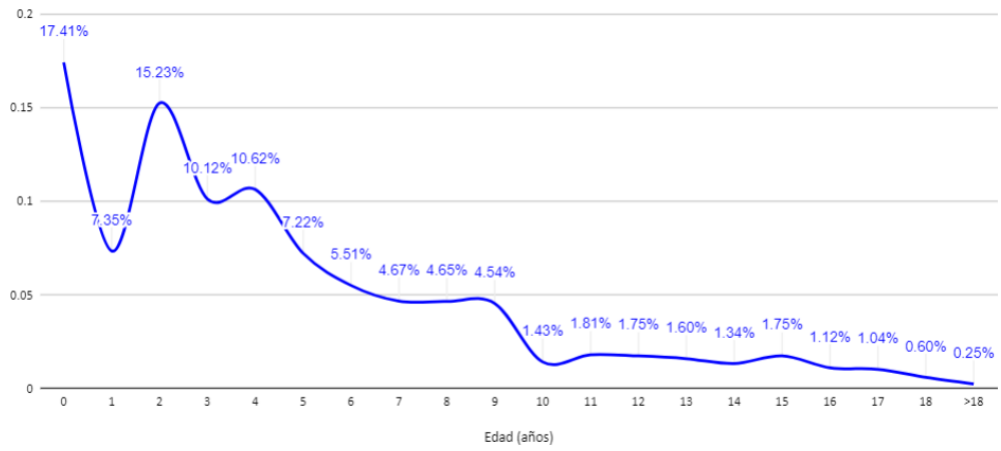
El parque de medidores de la empresa está conformado por aproximadamente 561,981 usuarios activos entre residenciales, comerciales, industriales y cantones al mes de junio 2022.

La figura 3.1, muestra la distribución del parque de medidores, por diámetros de medidor, en ella se puede observar que el 99.10% de medidores corresponden a medidores de ½ pulgada o 15mm, esto es, consumo residencial, que serán la población de estudio de este proyecto.



La edad promedio del parque de medidores es 4.63 años, el 67.95% del parque de medidores reporta una edad menor a 5 años (Figura 3.2).

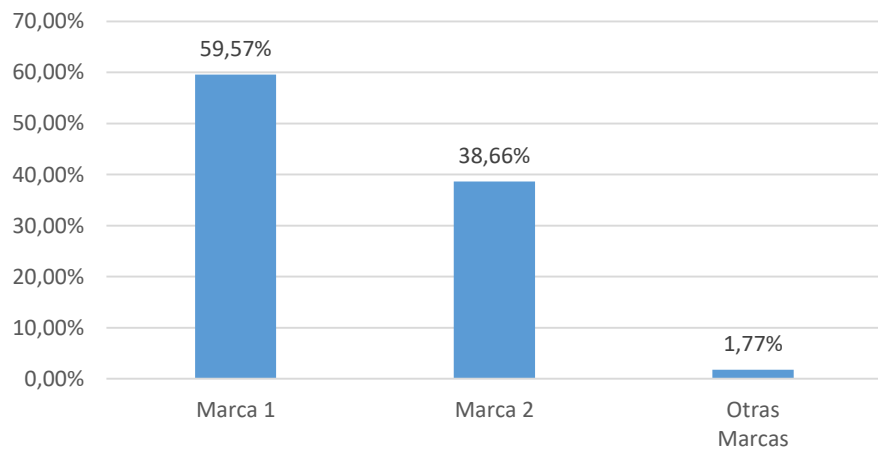
Figura 3. 2 Antigüedad del parque de medidores



Elaborado por: Dunis López Cárdenas

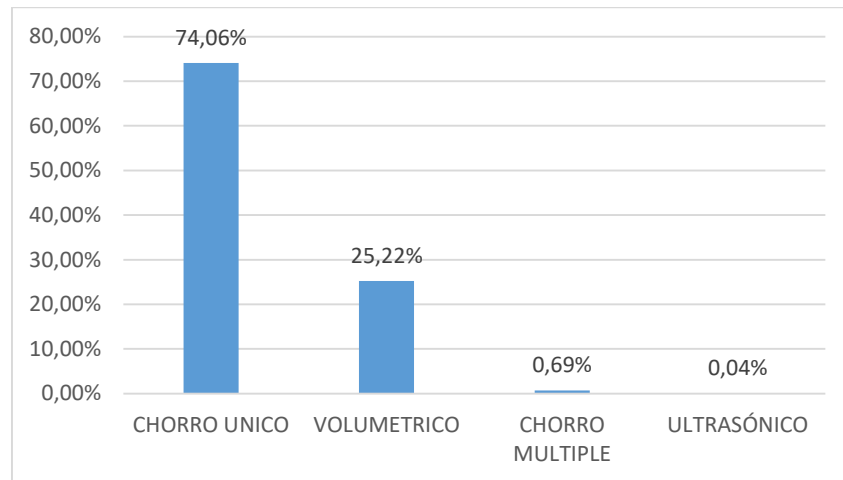
El parque de medidores está conformado en un 98.23% por dos marcas representativas. Figura 3.3

Figura 3. 3 Medidores por marca



Elaborado por: Dunis López Cárdenas

Figura 3. 4 Medidores de media por tecnología



Los datos para este estudio provienen de la base de datos del área comercial y errores de medición que son los resultados de ensayos realizados en el laboratorio de ensayo de medidores y volúmenes de consumo recopilados mediante trabajo de campo realizada a una muestra representativa determinada a través de muestreo probabilístico dirigidos a clientes activos, residenciales y que tengan instalados un medidor de 15mm de diámetro.

Dado que la precisión del medidor depende de varios factores como la calidad del agua, las condiciones ambientales, el volumen registrado, etc. La muestra se definió a través del empleo de muestreo aleatorio simple en base a un error de muestreo del 0.5 para un nivel de confianza del 95%. Para el cálculo del tamaño total de la muestra, se empleó la siguiente fórmula:

$$n = \frac{z_{\alpha}^2 N \sigma^2}{e^2(N - 1) + z_{\alpha}^2 \sigma^2}$$

Donde:

$$N = 421269$$

$$Z_{\alpha} = 1.96$$

$$\sigma^2 = 209.57$$

$$e = 0.5m^3$$

Al reemplazar estos valores en la formula se obtuvo una muestra total de 3196 clientes.

3.3. Tratamiento de datos.

Una vez obtenidos los datos se realizó una exploración de los mismos, para validarlos y depurarlos. Se verificó la estructura de las variables, sean estas cualitativas o cuantitativas.

Posteriormente, las observaciones recopiladas fueron sometidas a un análisis estadístico univariante y multivariante.

3.4. Factores que afectan la precisión de un medidor.

Entre los factores que afectan el desempeño de un medidor están: el manejo e instalación del medidor, las características propias de fabricación del equipo, las condiciones ambientales, la calidad y velocidad del agua, el deterioro del equipo por la edad, el volumen registrado del equipo y la manipulación del equipo.

La correcta instalación del equipo es muy importante de acuerdo al modelo de medidor, ya que su precisión se verá afectada si la posición del equipo es horizontal o vertical.

Dentro de las características propias de fabricación del medidor esta la tecnología, sea esta de chorro único, chorro múltiple, pistón rotativo, tangencial o Woltman. En consumos residenciales se usa en su mayoría medidores de velocidad (chorro único y los de chorro múltiple) y medidores volumétricos (medidores de pistón rotativo) (Arregui et al. 2006b).

Existen maneras en las que se puede cometer fraude para evitar el correcto registro del equipo, entre los que se puede mencionar la introducción de elementos que bloqueen el giro de la hélice, retirar el medidor o realizar un by pass, golpes que dañen su capacidad metrológica, perforación del equipo, etc. (Arregui et al. 2006b).

3.5. Análisis de conglomerados.

El análisis de conglomerados (clusters) tiene por objetivo agrupar elementos en grupos homogéneos en función de las similitudes o similaridades entre ellos (Peña, 2002).

También conocido como clasificación no supervisada, en donde a partir de las observaciones obtenidas para un conjunto de variables se desea agrupar de forma homogénea sin conocer a priori los grupos a los que pertenecen (Mañas & Ángel, 2017).

Tiene las siguientes características:

- Cada elemento pertenece a un solo grupo
- Todo el elemento queda clasificados
- Cada grupo internamente es homogéneo.

3.5.1. Medidas de distancia.

Para la medición de la diferencia entre los elementos, se aplican medidas de distancia como:

- *Distancia euclídea:* La distancia euclídea, nos dice que los conglomerados o clusters deben constituirse de tal manera que, al juntarse dos elementos, la pérdida de información resultante debe ser mínima. Por lo tanto, la cantidad de información se cuantifica como la suma de las distancias al cuadrado de cada elemento respecto al elemento del centro del conglomerado al que pertenece.

La distancia euclídea está dada por la siguiente notación:

$$d^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2$$

- *Distancia de Minkowski:* Se define por:

$$\sqrt[q]{\sum_{k=1}^p (x_{ik} - x_{jk})^q}, q > 0$$

- *Distancia de Manhattan:* Es una variante de la distancia de Minkowski con $q=1$.

Cuanto mayor es q más énfasis se les da a las diferencias en cada variable. Todas estas distancias no son invariantes a cambios de escala por lo que se aconseja estandarizar los datos si las unidades de medida de las variables no son comparables.

3.5.2. Métodos de agrupamiento.

- *Vecino más cercano:* También es conocido como método simple, mide la proximidad entre dos grupos calculando la distancia entre sus objetos más próximos o la similitud entre sus objetos más semejantes. Este método se basa en la distancia mínima, encuentra dos elementos separados por la distancia más corta y la coloca en el primer conglomerado. Luego encuentra la distancia más corta, y o bien un tercer elemento se une a los primeros para formar un conglomerado o se forma un nuevo conglomerado de dos miembros.
- *Promedio:* Determina la proximidad entre dos grupos calculando la media de las distancias entre objetos de ambos grupos o la media de las similitudes entre objetos de ambos grupos.

- *Método de Ward*: El atributo más distintivo del método de Ward es que no emplea las distancias entre conglomerados para realizar la aproximación, sino que trata de hacer mínima la variabilidad entre conglomerados, esto es de hacer que cada conglomerado sea lo más homogéneo posible.

3.6. Regresión lineal simple y múltiple.

El caso más simple de regresión lineal ajusta a la ecuación de la recta los valores de la variable independiente X a la variable dependiente Y, es decir:

$$Y = b_0 + b_1X$$

donde b_0 es la ordenada en el origen y b_1 es la pendiente de la recta. El ajuste a esta ecuación (mediante mínimos cuadrados) se caracteriza por la obtención de b_0 , b_1 y el coeficiente de correlación r .

La regresión lineal múltiple se basa en obtener una relación lineal entre un conjunto de variables independientes x_1, x_2, \dots, x_n con una variable dependiente Y, es decir:

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + error$$

El éxito de determinar una correlación lineal múltiple es que exista una correlación lineal simple de cada variable independiente con la variable dependiente. (Mendenhall, 2015)

3.6.1. Supuestos.

Para realizar un análisis de regresión lineal múltiple se hacen las siguientes consideraciones sobre los datos:

- *Linealidad*: los valores de la variable dependiente están generados por el siguiente modelo lineal:

$$Y = X * B + U$$

- *Homocedasticidad*: todas las perturbaciones tienen la misma varianza.

$$V(u_i) = \sigma^2$$

- *Independencia*: las perturbaciones aleatorias son independientes entre sí.

$$E(u_i \cdot u_j) = 0; \text{ para } i \neq j$$

- *Normalidad*: la distribución de la perturbación aleatoria tiene distribución normal.

$$U \approx N(0, \sigma^2)$$

- Las variables explicativas X_n toman valores distintos no relacionados entre sí.

Si los datos presentan estas hipótesis, entonces el teorema de Gauss-Markov establece que el método de estimación de mínimos cuadrados va a producir estimadores óptimos, en el sentido que los parámetros estimados van a estar centrados y van a ser de mínima varianza.

3.7. Modelos lineales generalizados.

El modelo lineal generalizado (GLM) amplía el modelo lineal general, incorporando la posibilidad de modelar variables respuestas continuas o categóricas con distribuciones del error no necesariamente homocedásticos. Los modelos log lineales, logit, probit, logístico y de regresión lineal son algunos de los modelos que forman parte de esta familia. (Nelder y Wedderburn, 1972)

3.7.1. Componentes de los modelos lineales generalizados.

- *Componente aleatoria*: formada por el vector aleatorio observable Y , tal que sus elementos son independientes e idénticamente distribuidos con función de distribución perteneciente a la familia exponencial.
- *Componente sistemática*: Al igual que en la regresión lineal está dada por el predictor lineal.
- *Función de enlace*: Función que relaciona el predictor lineal, con el valor esperado de la variable respuesta.

3.7.2. Estimación de modelos lineales generalizados.

Los dos métodos clásicos para estimar los parámetros desconocidos de un modelo lineal generalizados son:

- *Método de máxima verosimilitud*: se selecciona los valores de los parámetros aquellos que tienen más probabilidad de ocurrir o son más compatibles con los datos observados.
- *Método de mínimos cuadrados generalizados*: método en el cual, los parámetros se obtienen minimizando la suma cuadrática de los errores.

CAPÍTULO 4

4. RESULTADOS

4.1. Análisis descriptivo de los datos.

Se realiza un estudio estadístico univariante de las variables relevantes expuestas en el capítulo anterior, se presentarán parámetros como media, mediana, moda, varianza, sesgo, kurtosis, cuartiles y gráficos que ayuden a visualizar mejor los datos.

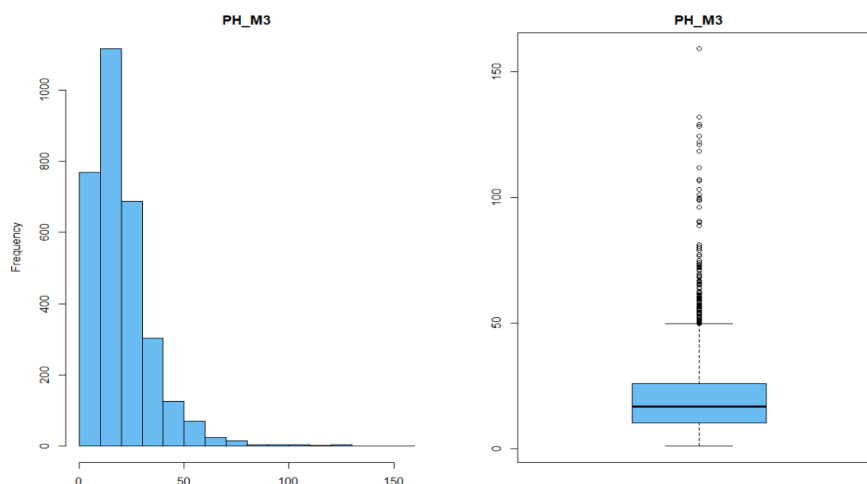
4.1.1. Variable Promedio histórico.

Tabla 4. 1 Estadística Descriptiva Variable Promedio Histórico

Media	20,18
Mediana	16,83
Moda	11,00
Desviación Est.	15,38
Varianza	236,43
Asimetría	2,37
Kurtosis	10,46
Q ₁	10,17
Q ₂	16,83
Q ₃	26,00
Mínimo	1,00
Máximo	159,17

La tabla anterior nos muestra la estadística descriptiva de la variable promedio histórico donde se tiene que el promedio del promedio histórico de consumo es 20.8 m³, el valor que más se repite es 11 m³; la distribución es asimétrica positiva lo que significa que los datos se encuentran concentrados hacia la derecha del promedio, Con respecto a los cuartiles, tenemos el primer cuartil Q₁ es 10.17 m³, que significa que por debajo de este valor se sitúa el 25% de la frecuencia. El segundo cuartil es 16.83 m³, es decir que entre este valor y el primer cuartil se sitúa otro 25% de la frecuencia. Y el tercer cuartil es 26, ya que entre este valor y el segundo cuartil se sitúa otro 25% de la frecuencia. A continuación, se presenta el histograma y el diagrama de caja.

Figura 4. 1 Histograma y diagrama de caja Variable Promedio Histórico



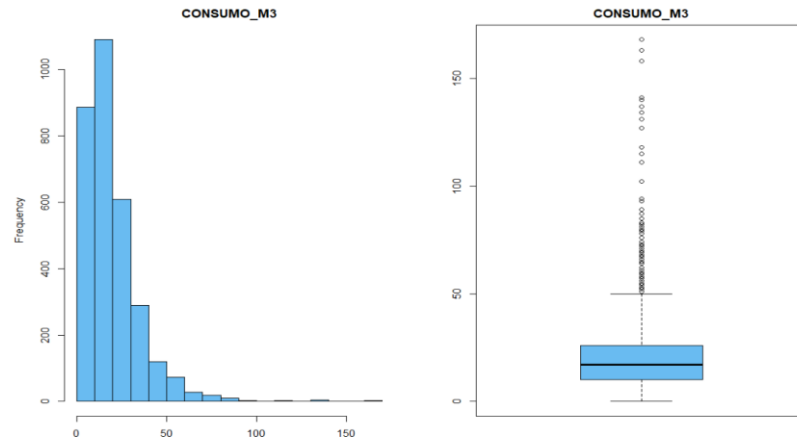
4.1.2. Variable Consumo mensual.

La tabla 4.2 nos muestra que el promedio de consumo es 19.82 m^3 . La desviación estándar es 16.11 m^3 . En esta distribución los datos se encuentran sesgados a la derecha con respecto a la media debido a que el coeficiente de sesgo es positivo (2.55), en lo que respecta al coeficiente de kurtosis (13.03) es leptocúrtica. Es decir, más picuda que una distribución normal estándar.

Tabla 4. 2 Estadística Descriptiva Variable Consumo mensual

Media	19,82
Mediana	17,00
Moda	12,00
Desviación Est.	16,11
Varianza	259,49
Asimetría	2,55
Kurtosis	13,03
Q1	10,00
Q2	17,00
Q3	26,00
Mínimo	0,00
Máximo	168,00

Figura 4. 2 Histograma y diagrama de caja Variable consumo mensual



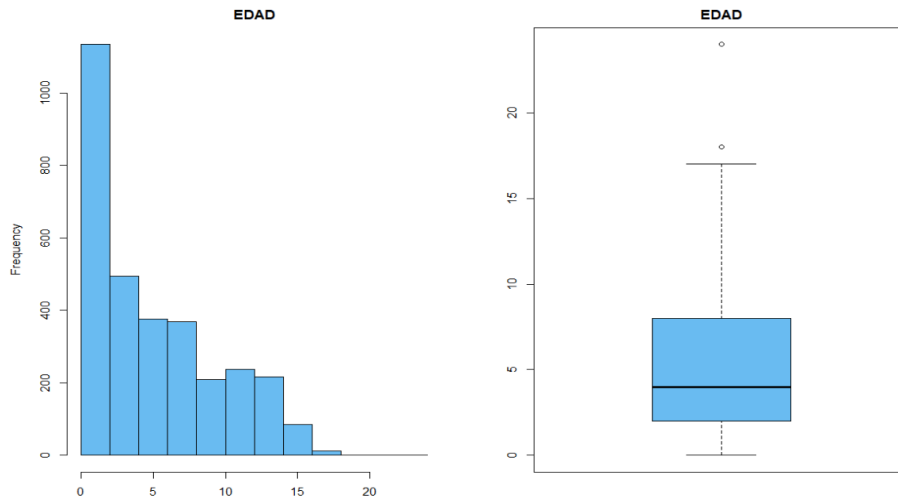
4.1.3. Edad del medidor.

En la siguiente tabla 4.3 podemos ver la estadística descriptiva de la variable Edad del medidor, en la cual la media es 5.39, la mediana 4, tiene una desviación estándar de 4.42, el coeficiente de asimetría es positivo lo que significa que los datos se encuentran concentrados a la derecha de promedio, esto lo podemos observar en el histograma.

Tabla 4. 3 Estadística Descriptiva Variable Edad del medidor

Media	5,39
Mediana	4,00
Moda	1,00
Desviación Est.	4,42
Varianza	19,57
Asimetría	0,70
Kurtosis	-0,56
Q1	2,00
Q2	4,00
Q3	8,00
Mínimo	0,00
Máximo	24,00

Figura 4. 3 Histograma y diagrama de caja Variable edad del medidor



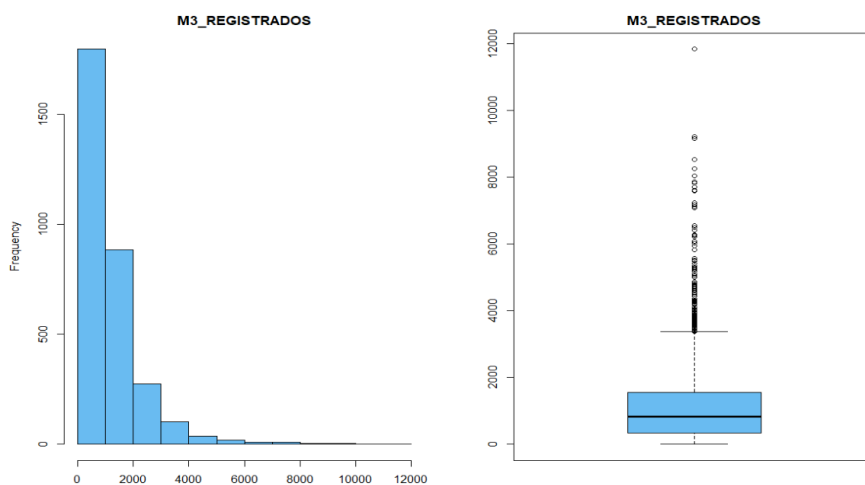
4.1.4. Variable m³ registrados.

Para la variable m³ registrados observamos que la media es 1122.96, además vemos que el valor mínimo es cero y el valor máximo es 11848m³.

Tabla 4. 4 Estadística Descriptiva Variable m³ registrados

Media	1122,96
Mediana	832,00
Moda	12,00
Desviación Est.	1127,43
Varianza	2711,09
Asimetría	2,53
Kurtosis	10,93
Q1	334,00
Q2	832,00
Q3	1555,00
Mínimo	0,00
Máximo	11848,00

Figura 4. 4 Histograma y diagrama de caja Variable m³ registrados



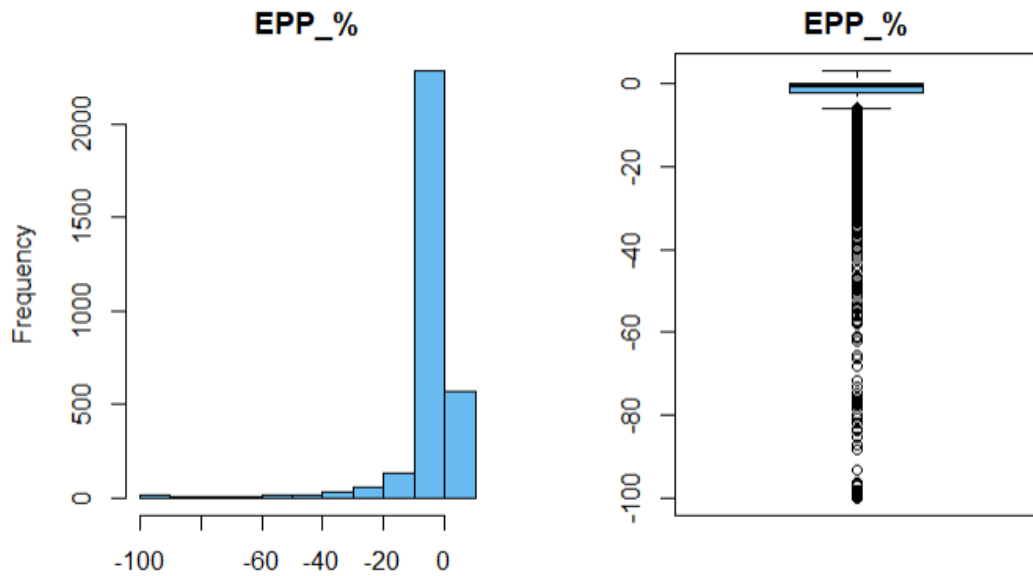
4.1.5. Variable Error Promedio ponderado.

Podemos ver en la tabla 4.5, que el promedio del error promedio ponderado es -3.82% con una desviación estándar de 11.72.

Tabla 4. 5 Estadística Descriptiva Variable Error promedio Ponderado

Media	-3,82
Mediana	-0,18
Moda	-1,00
Desviación Est.	11,72
Varianza	137,28
Asimetría	-5,32
Kurtosis	32,49
Q1	-2,30
Q2	-0,18
Q3	-0,01
Mínimo	-100,00
Máximo	3,30

Figura 4. 5 Histograma y diagrama de caja Variable Error promedio ponderado



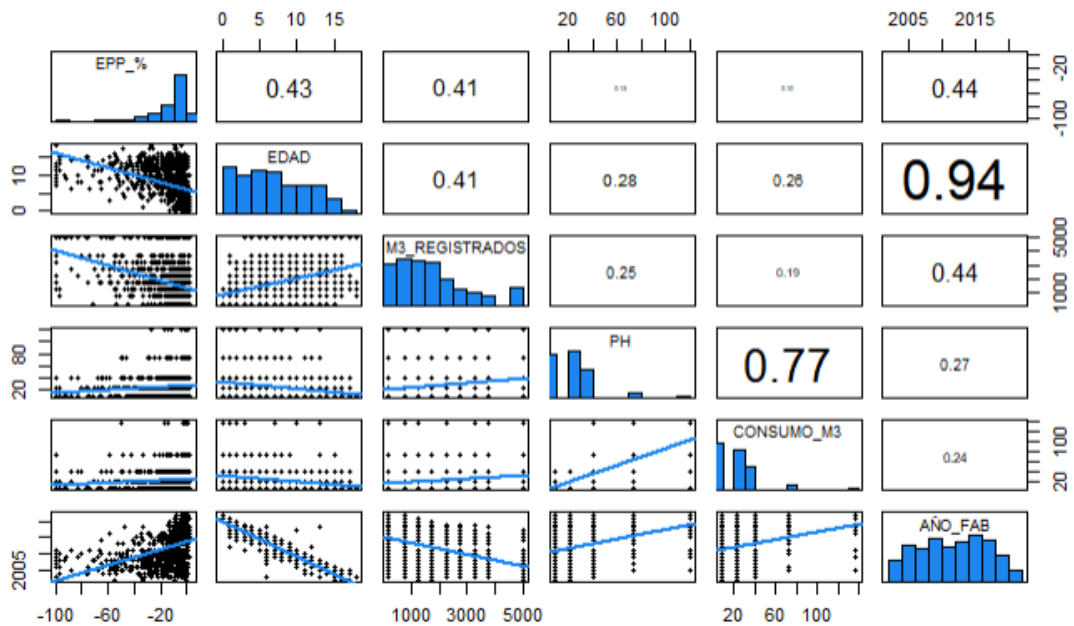
4.2. Análisis estadístico bivalente.

En este apartado se realizará el análisis conjunto de dos características con el propósito de detectar posibles relaciones entre ellas.

4.2.1. Correlación de las variables.

En el siguiente gráfico se presentan los coeficientes de correlación de las variables cuantitativas que forman partes de este proyecto, esto nos permitirá conocer que tan fuerte están relacionadas las variables independientes frente al error promedio ponderado. Se puede observar que existe una correlación de más del 0.4 entre error promedio ponderado y las variables edad, m³ registrados y año de fabricación, sin embargo, la correlación entre el error medio ponderado y las variables promedio histórico y consumo es muy pequeña. Hay una correlación de 0.77 entre las variables promedio histórico y el consumo facturado, 0.94 entre edad del medidor y año de fabricación y 0.41 entre edad del medidor y m³ registrados.

Figura 4. 6 Correlación de variables



4.3. Análisis de clusters.

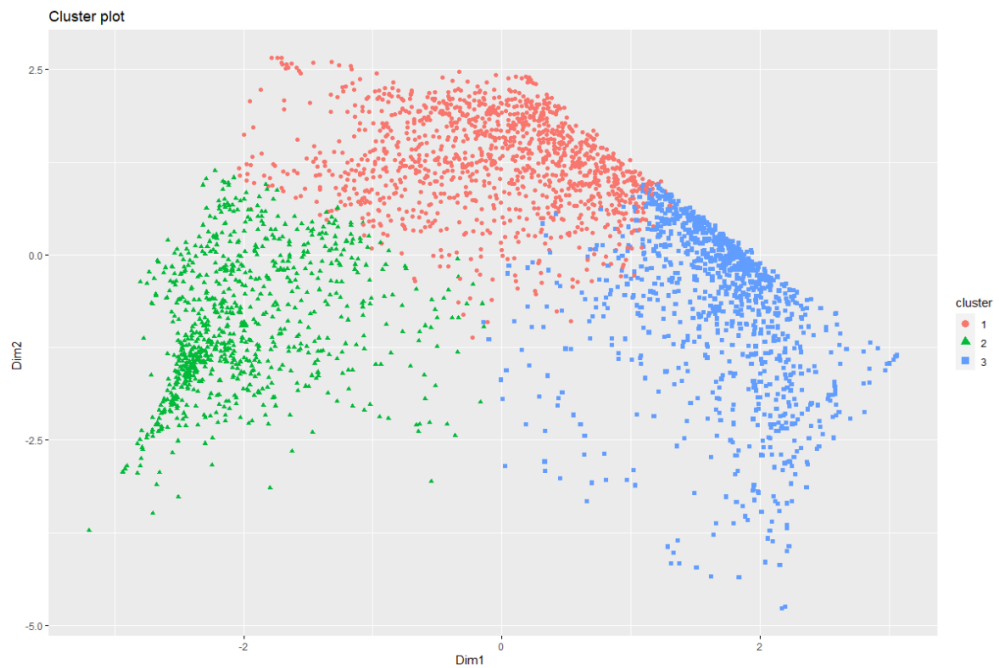
Se aplicó este método con la finalidad de agrupar o clasificar el perfil de consumo de los usuarios, para lo cual se usarán las variables de porcentaje de consumo en rangos de caudal, consumo promedio y consumo del mes.

Para realizar este análisis, se utilizó el software estadístico R versión 4.1.0 (R Core Team, 2021) y los paquetes factoextra versión 1.0.7 (Alboukadel Kassambara and Fabian Mundt, 2020), cluster versión 2.1.2 (Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2021), tidyverse (Wickham, 2019) y NbClust (Malika Charrad, Nadia Ghazzali, Veronique Boiteau, Azam Niknafs, 2014).

Para estimar el número óptimo de clúster se utilizó el método de la silueta, el método de codo y el método de la brecha. El método de codo indicaba trabajar con más de 10 grupos, tanto el método de la silueta y el método de la brecha indicaba trabajar con tres grupos.

En la Figura 4.7 se obtuvo una representación gráfica de los clústeres, donde se aprecian la conformación de 3 grupos o patrones de consumo. El grupo 1 conformado por 1246 clientes, el grupo 2 por 837 y el grupo 3 por 1052 individuos.

Figura 4. 7 Representación de los grupos



En la Figura 4.8 se muestra el perfil de consumo de cada uno de los grupos formado de acuerdo a la forma de consumir el agua potable. Se observa que el grupo 1 está conformado por individuos que consumen entre los rangos de caudal 250 l/h y 750 l/h, el grupo 2 entre 750 l/h y 3000l/h y el grupo 3 a menos de 250l/h estos pueden ser clientes que usan cisterna.

Figura 4. 8 Perfil de consumo por grupos

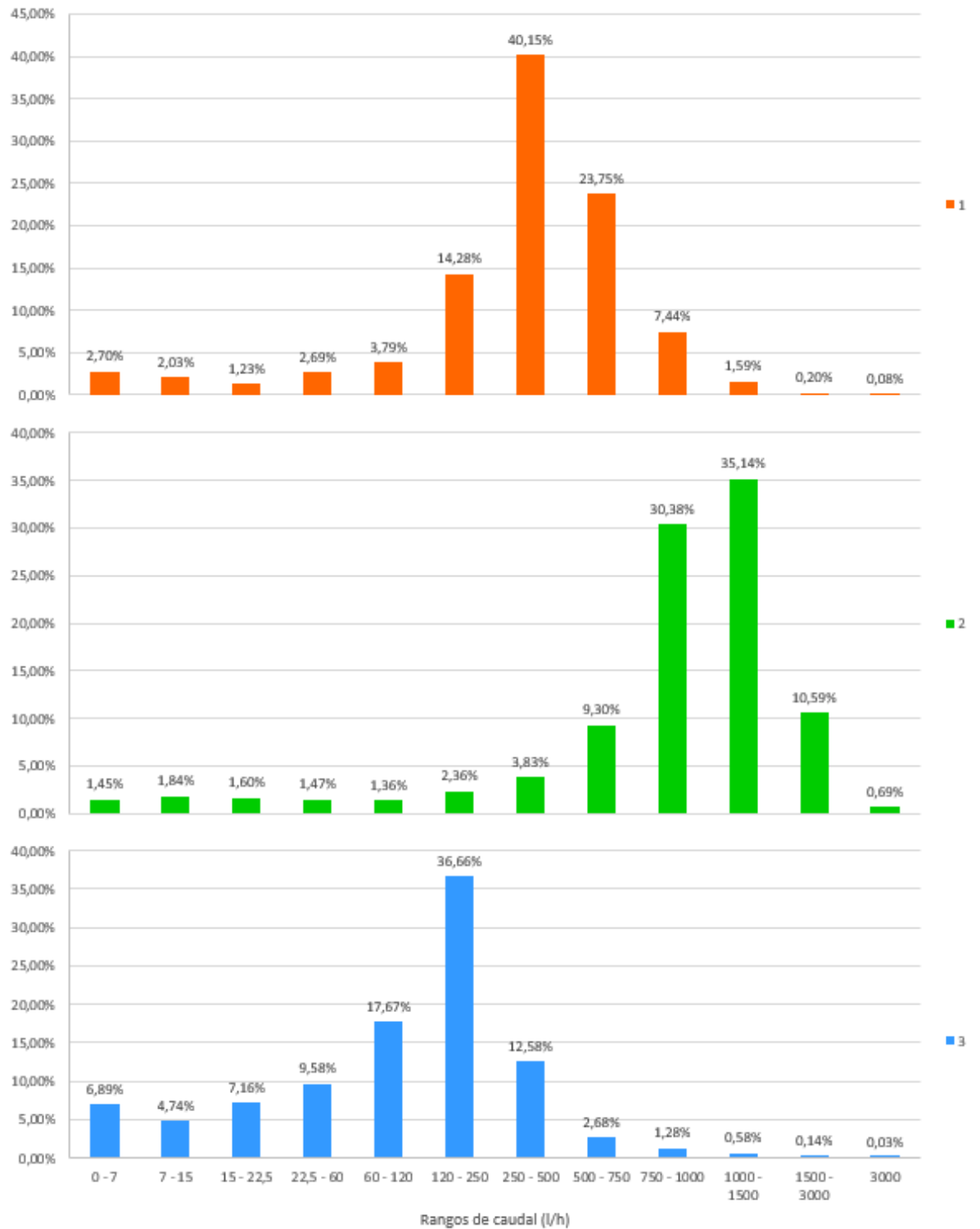
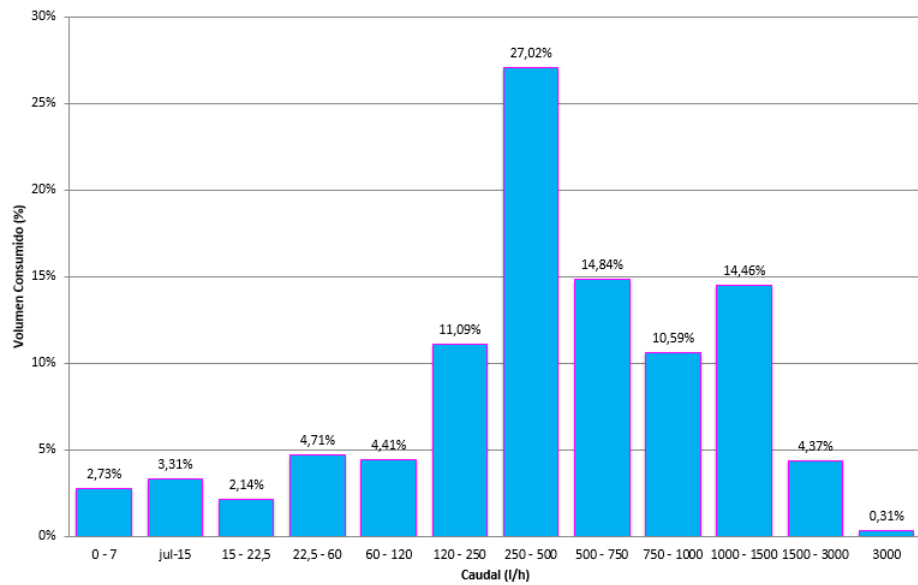
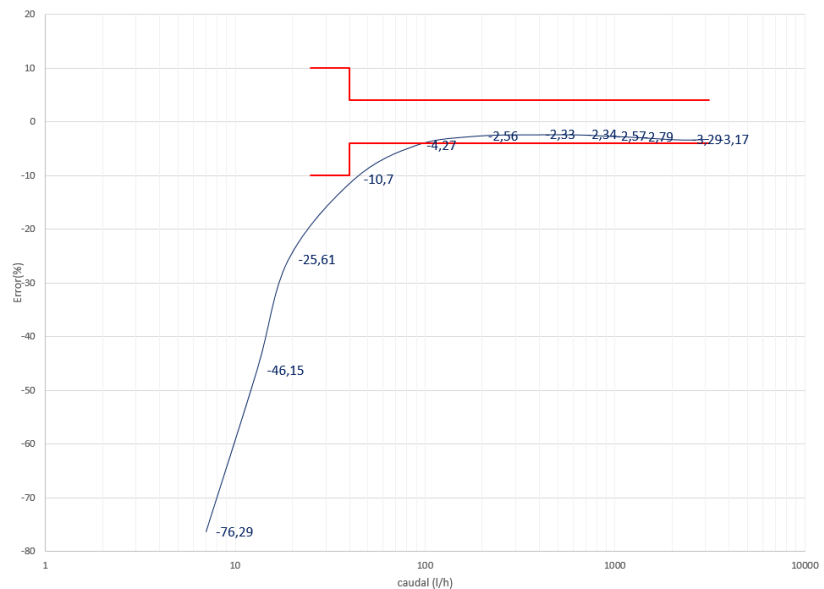


Figura 4. 9 Perfil de consumo Residencial



La figura 4.9 representa el perfil de consumo de todos los clientes que conforman la muestra, en donde se observa que el mayor consumo se da entre caudales de 120 l/h y 1500 l/h con un 78% del consumo registrado, el 14,57% se consume entre 7 l/h y 120 l/h y el 4,68% se consume entre 1500 l/h y 3000 l/h. Existen también consumos bajo los 7 l/h lo cual podría suponer fugas en el interior de la residencia de 2,73%.

Figura 4. 10 Curva de error de medidores



En la figura 4.10 se muestra la curva de error a cada caudal, así como también el túnel que identifica los límites de error permitidos por la norma ISO 4064 vigente.

4.4. Estimación de la precisión global. (El error medio ponderado global)

Conociendo el perfil de consumo de los usuarios (figura 4.9) y la curva de error de los medidores (figura 4.10), ya existen las condiciones para estimar el error de medición de todo el parque de medidores (apartado 2.1.6). Dado que el perfil de consumo asigna un peso a cada rango de consumos podemos ponderar estos errores y obtener el error de medición global (tabla 4.6)

Para cada rango de caudales se extrae el porcentaje que indica el perfil de consumo y el error de medición correspondiente. Por ejemplo, para el rango de 60-120 l/h el peso que indica el perfil es del 4,41% y el error de medición es -4,27%; el error promedio ponderado sería $4.41\% \cdot (-4.27\%) = -0.19\%$. Procediendo de esta forma para cada rango de caudales obtenemos de acuerdo a la suma ponderada, que el error promedio ponderado es del -6,94%, o lo que es lo mismo, que vamos a registrar el 93,06% del consumo real.

Tabla 4. 6 Calculo Error promedio ponderado

Caudales de ensayo (l/h)	Error de medición (%)	Rango de Caudales de consumo (l/h)	Porcentaje de consumo	Error promedio ponderado
7	-76,29%	0 – 7	2,73%	-2,08%
13	-46,15%	7-15	3,31%	-1,53%
19	-25,61%	15 - 22,5	2,14%	-0,55%
42	-10,70%	22,5 - 60	4,71%	-0,50%
90	-4,27%	60 - 120	4,41%	-0,19%
190	-2,56%	120 - 250	11,09%	-0,28%
380	-2,33%	250 - 500	27,02%	-0,63%
630	-2,34%	500 - 750	14,84%	-0,35%
900	-2,57%	750 - 1000	10,59%	-0,27%
1.250	-2,79%	1000 - 1500	14,46%	-0,40%
2.250	-3,29%	1500 - 3000	4,37%	-0,14%
3.125	-3,17%	>3000	0,33%	-0,01%
				-6,94%

4.5. Análisis de regresión para el error ponderado con la edad.

En primera instancia se realizó análisis de regresión lineal entre las variable edad y error ponderado con todos los datos, sin embargo, el valor de R2 es muy pequeño 18% (tabla 4.7); por lo cual se agrupó el error ponderado en intervalos de edad para realizar el análisis de regresión. (Sánchez, 2015).

La tabla 4.8 muestra el detalle de la agrupación realizada

Tabla 4. 7 Detalle del modelo Error promedio ponderado y edad

```
Call:
lm(formula = EPP_ ~ EDAD, data = Base)

Residuals:
    Min       1Q   Median       3Q      Max
-99.55  -1.19    2.83    7.50   29.12

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.45093    0.45200  -0.998   0.319
EDAD        -1.72742    0.06493 -26.605 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.04 on 3133 degrees of freedom
Multiple R-squared:  0.1843,    Adjusted R-squared:  0.184
F-statistic: 707.8 on 1 and 3133 DF,  p-value: < 2.2e-16
```

Tabla 4. 8 Tabla error promedio ponderado y edad

EDAD	Promedio de error ponderado	Número de individuos	Promedio de m ³ registrados
0	-0,98	258	162,33
1	-1,23	472	278,01
2	-0,13	405	747,42
3	-8,26	285	908,54
4	-11,22	209	1057,85
5	-9,80	185	1418,70
6	-11,47	192	1302,39
7	-9,59	205	1453,78
8	-9,32	164	1478,09
9	-17,69	109	1896,73
10	-24,56	99	1783,94
11	-25,40	139	1832,28
12	-17,25	98	2238,57
13	-25,34	152	2053,50
14	-18,48	65	2009,18
15	-23,60	60	1907,52
16	-33,24	26	2153,19
17	-41,85	12	3411,25

Se compararon varios modelos para determinar cual tiene el mejor ajuste. En la siguiente tabla se puede observar que el mejor ajuste en el lineal

Tabla 4. 9 Comparación de modelos

Modelo	Error residual	R cuadrado
Lineal	0,0469	83,98%
Cuadrado de Edad	0,4905	82,55%
Raíz cuadrada de Edad	0,5797	75,62%

La ecuación que se ajusta de forma lineal es:

$$\text{Error promedio ponderado} = 0.0054 - 0.0195 * \text{Edad}$$

Tabla 4. 10 Detalle del modelo Error promedio ponderado y edad

```
Call:
lm(formula = EPP ~ EDAD, data = BaseREdad)

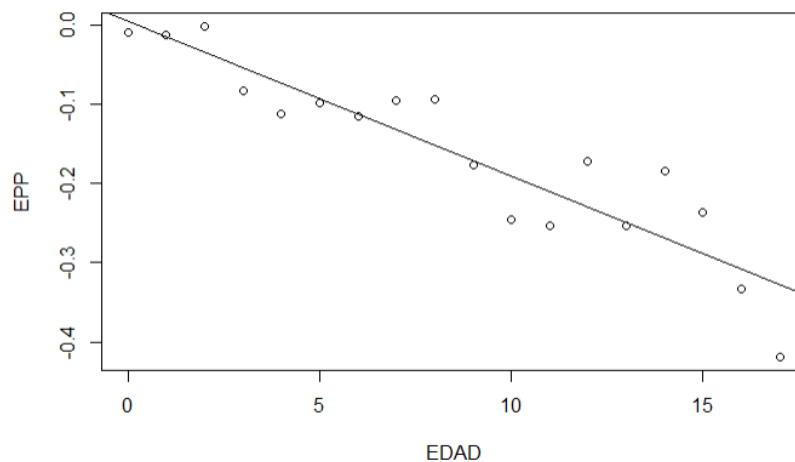
Residuals:
    Min       1Q   Median       3Q      Max
-0.09152 -0.02826 -0.00514  0.03476  0.08352

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.005412   0.021260   0.255   0.802
EDAD        -0.019552   0.002135  -9.158 9.2e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04699 on 16 degrees of freedom
Multiple R-squared:  0.8398,    Adjusted R-squared:  0.8298
F-statistic: 83.87 on 1 and 16 DF,  p-value: 9.201e-08
```

La figura 4.11 muestra la curva ajustada con las observaciones

Figura 4. 11 Recta de regresión Error ponderando y edad



A continuación, se observa el análisis de varianza para el modelo seleccionado con el objetivo de validar el modelo encontrado.

Tabla 4. 11 Tabla ANOVA

Analysis of Variance Table						
Response: EPP						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
EDAD	1	0.185223	0.185223	83.874	9.201e-08	***
Residuals	16	0.035333	0.002208			

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

Con el resultado mostrado en la tabla 4.11, se puede analizar que el valor de F es alto siendo mayor a $F(\alpha, 1, n-2)$, por lo cual se rechaza la hipótesis nula concluyendo que se tiene un b_1 significativo.

4.5.1. Comprobación de los supuestos error promedio y edad.

En la tabla de 4.12 se muestra la comprobación de los supuestos para el modelo lineal de error promedio y la edad.

Tabla 4. 12 Verificación de supuestos.

Supuesto	Test	Valor p	Resultado
Normalidad	shapiro.test	0,8570	No rechazo Ho. Existe evidencia de la normalidad
Homocedasticidad	Breusch-Pagan	0,6401	No rechazo Ho. Los residuos son homocedásticos
AutoCorrelación	Durbin-Watson	0,3277	No rechazo Ho. No existe auto correlación

4.6. Análisis de regresión para el error ponderado y los m^3 registrados.

Siguiendo con el análisis se realizó regresión lineal entre el error ponderado y los m^3 registrados obteniendo los siguientes resultados:

Tabla 4. 13 Detalle de modelo m³ registrados y error ponderado

```
Call:
lm(formula = EPP_ ~ M3_REGISTRADOS, data = Base)

Residuals:
    Min       1Q   Median       3Q      Max
-96.491  -0.374   4.109   7.153  52.472

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.035832   0.414483  -7.324 3.04e-13 ***
M3_REGISTRADOS -0.005991   0.000261 -22.949 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.43 on 3133 degrees of freedom
Multiple R-squared:  0.1439,    Adjusted R-squared:  0.1436
F-statistic: 526.6 on 1 and 3133 DF,  p-value: < 2.2e-16
```

Se puede observar que el R2 es muy bajo por lo que no es un modelo adecuado.

Al igual que la edad se agruparon los datos de acuerdo en intervalos de m³ registrados. La tabla a continuación muestra los datos agrupados.

Tabla 4. 14 Tabla m³ registrados y error ponderado

m ³ registrados	Promedio de error ponderado	Número de individuos	Promedio de m ³ registrados
200	-3,88	1026	210,02
750	-6,98	770	736,88
1250	-9,79	503	1238,61
1750	-15,57	380	1728,40
2250	-17,09	176	2222,52
2750	-24,18	100	2751,84
3250	-20,71	62	3241,85
3750	-29,31	40	3706,88
5000	-32,010	78	5606,91

En la siguiente tabla se puede observar la comparación de los modelos y el que mejor ajuste tiene es el lineal

Tabla 4. 15 Comparación de modelos

Modelo	Error residual	R cuadrado
Lineal	0,02393	94,75%
Cuadrado de m ³ registrados	0,04468	81,71%
Raíz cuadrada de m ³ registrados	0,02591	93,85%

La ecuación que se ajusta de forma lineal es:

*Error promedio ponderado = $-0.0326 - 0.000622 * m^3$ registrados*

Tabla 4. 16 Error promedio ponderado y m^3 registrados

```
Call:
lm(formula = EPP ~ M3, data = BaseRm3)

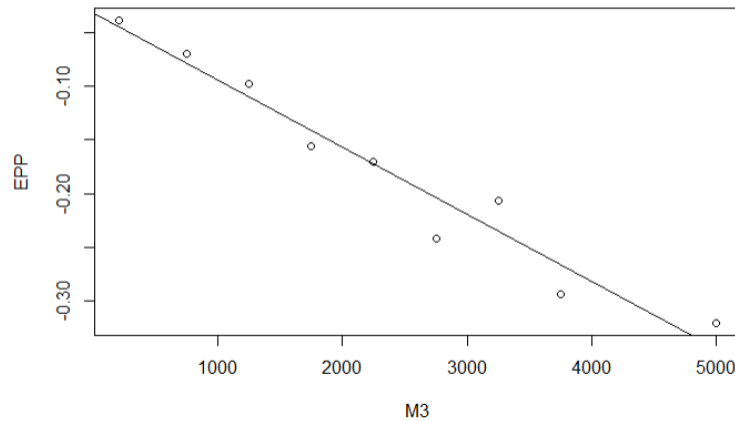
Residuals:
    Min       1Q   Median       3Q      Max
-0.03831 -0.01439  0.00608  0.01227  0.02756

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.246e-02  1.515e-02  -2.143  0.0693 .
M3           -6.220e-05  5.532e-06 -11.244 9.83e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02393 on 7 degrees of freedom
Multiple R-squared:  0.9475,    Adjusted R-squared:  0.94
F-statistic: 126.4 on 1 and 7 DF,  p-value: 9.827e-06
```

La figura 4.12 muestra la curva ajustada con las observaciones

Figura 4. 12 Recta de regresión Error ponderando y m^3 registrados



A continuación, se observa el análisis de varianza para el modelo seleccionado con el objetivo de validar el modelo encontrado.

Tabla 4. 17 Tabla ANOVA

```
Analysis of Variance Table

Response: EPP
      Df Sum Sq Mean Sq F value    Pr(>F)
EDAD   1  0.185223  0.185223  83.874 9.201e-08 ***
Residuals 16  0.035333  0.002208
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En la tabla de análisis de varianza mostrada en la tabla 4.17, se puede analizar que el valor de F es alto (83,87) siendo mayor a $F(\alpha, 1, n-2)$, por

lo cual se rechaza la hipótesis nula concluyendo que se tiene un b_1 significativo.

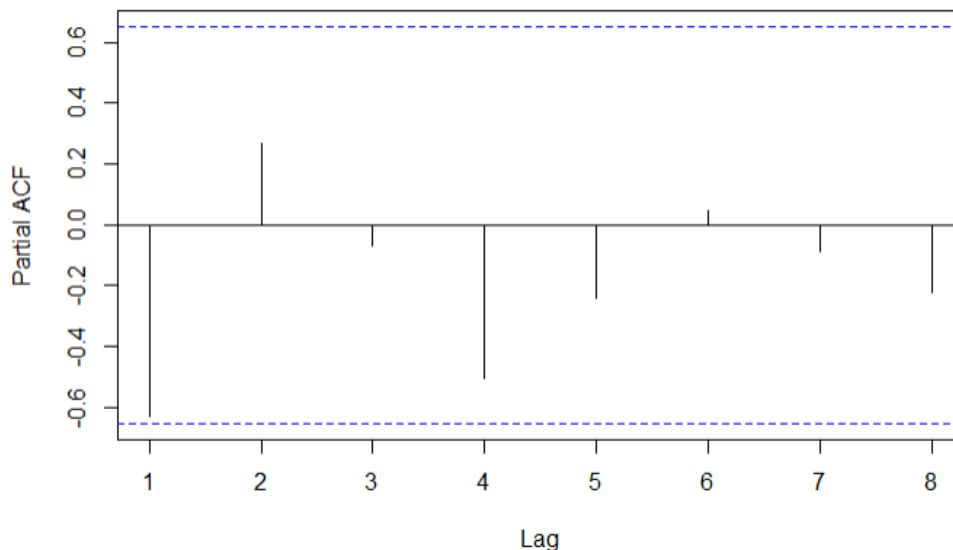
4.6.1. Comprobación de los supuestos error promedio y m^3 registrados.

En la tabla de 4.18 se muestra la comprobación de los supuestos para el modelo lineal de error promedio y m^3 registrados.

Tabla 4. 18 Verificación de supuestos.

Supuesto	Test	Valor p	Resultado
Normalidad	shapiro.test	0,5453	No rechazo Ho. Existe evidencia de la normalidad
Homocedasticidad	Breusch-Pagan	0,0904	No rechazo Ho. Los residuos son homocedásticos
AutoCorrelación	Durbin-Watson	0,1246	No rechazo Ho. No existe auto correlación

Figura 4. 13 Función de auto correlación Error ponderando y m^3 registrados



4.7. Regresión lineal múltiple.

En este apartado se analizará la relación entre la variable respuesta error ponderado y las variables edad, m^3 registrados, consumo, promedio histórico y año de fabricación.

Tabla 4. 19 Modelo de regresión lineal

```

Call:
lm(formula = BaseMM$EPP ~ BaseMM$EDAD + BaseMM$m3_REGISTRADOS +
    BaseMM$CONSUMO_M3 + BaseMM$PH + BaseMM$AÑO_FAB)

Residuals:
    Min       1Q   Median       3Q      Max
-0.076505 -0.017130  0.003008  0.017980  0.085103

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.279e+00  8.500e+00   0.739  0.46418
BaseMM$EDAD  -1.982e-02  5.658e-03  -3.503  0.00111 **
BaseMM$m3_REGISTRADOS  2.976e-05  1.580e-05   1.883  0.06662 .
BaseMM$CONSUMO_M3   4.897e-04  1.408e-03   0.348  0.72972
BaseMM$PH        -1.423e-04  1.531e-03  -0.093  0.92638
BaseMM$AÑO_FAB   -3.120e-03  4.207e-03  -0.742  0.46238
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03156 on 42 degrees of freedom
Multiple R-squared:  0.7533,    Adjusted R-squared:  0.7239
F-statistic: 25.65 on 5 and 42 DF,  p-value: 9.086e-12

```

El modelo con todas las variables introducidas como predictores tiene un R2 alto (0,7239), es capaz de explicar el 72,39% de la variabilidad observada en error promedio ponderado. El p-value del modelo es significativo (9,086e-12) por lo que se puede aceptar que el modelo no es por azar, al menos uno de los coeficientes parciales de regresión es distinto de 0. Muchos de ellos no son significativos, lo que es un indicativo de que podrían no contribuir al modelo.

Para seleccionar las mejores variables predictoras se van a emplear la estrategia de stepwise mixto. El valor matemático empleado para determinar la calidad del modelo va a ser Akaike(AIC).

El mejor modelo ha sido

$$\begin{aligned}
 & \text{Error promedio ponderado} \\
 & = -0.02044 - 0.0168 * \text{Edad} - 0.0000348 * m3 \text{ registrados}
 \end{aligned}$$

Tabla 4. 20 Modelo de regresión lineal mejorado

```

Call:
lm(formula = BaseMM$EPP ~ BaseMM$EDAD + BaseMM$m3_REGISTRADOS)

Residuals:
    Min       1Q   Median       3Q      Max
-0.081228 -0.016192  0.001841  0.019712  0.087245

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.045e-02  7.454e-03  -2.743  0.00870 **
BaseMM$EDAD    -1.680e-02  1.603e-03 -10.482 1.17e-13 ***
BaseMM$m3_REGISTRADOS  3.480e-05  1.241e-05   2.804  0.00742 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0309 on 45 degrees of freedom
Multiple R-squared:  0.7467,    Adjusted R-squared:  0.7354
F-statistic: 66.33 on 2 and 45 DF,  p-value: 3.818e-14
    
```

4.7.1. Comprobación de los supuestos error promedio, edad y m³ registrados.

En la tabla de 4.21 se muestra la comprobación de los supuestos para el modelo lineal de error promedio, edad y m³ registrados.

Tabla 4. 21 Verificación de supuestos.

Supuesto	Test	Valor p	Resultado
Normalidad	shapiro.test	0,5487	No rechazo Ho. Existe evidencia de la normalidad
Homocedasticidad	Breusch-Pagan	0,28	No rechazo Ho. Los residuos son homocedásticos
AutoCorrelación	Durbin-Watson	0,3635	No rechazo Ho. No existe auto correlación

El modelo lineal múltiple

Error promedio ponderado

$$= -0.02044 - 0.0168 * Edad - 0.0000348 * m3 \text{ registrados}$$

puede explicar el 73,54% de la variabilidad observada en el error promedio ponderado. El test F muestra que es significativo (p-value: 3,81e-14) y se satisfacen todas las condiciones para este tipo de regresión múltiple.

4.8. Regresión lineal múltiple considerando los grupos

Se realizará un análisis entre la relación entre la variable respuesta error ponderado y las variables edad, m³ registrados, y los grupos formados.

Tabla 4. 22 Modelo de regresión lineal

```
Call:
lm(formula = BaseDA$EPP ~ BaseDA$EDAD + BaseDA$m3_REGISTRADOS +
  BaseDA$cluster)

Residuals:
    Min       1Q   Median       3Q      Max
-0.301877 -0.034906  0.003859  0.038215  0.219782

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.474e-02  1.558e-02   4.156 5.84e-05 ***
BaseDA$EDAD  -1.431e-02  1.715e-03  -8.347 8.96e-14 ***
BaseDA$m3_REGISTRADOS -3.750e-05  8.949e-06  -4.190 5.11e-05 ***
BaseDA$cluster2 -4.508e-03  1.484e-02  -0.304  0.762
BaseDA$cluster3 -1.168e-01  1.430e-02  -8.164 2.45e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06949 on 130 degrees of freedom
Multiple R-squared:  0.6532,    Adjusted R-squared:  0.6425
F-statistic: 61.21 on 4 and 130 DF,  p-value: < 2.2e-16
```

El modelo con todas las variables tiene un R2 de 64,25% y está dado por:

$$\begin{aligned} & \text{Error promedio ponderado} \\ & = 0.06474 - 0.0143 * \text{Edad} - 0.0000375 * m3 \text{ registrados} \\ & - 0.0045077 * \text{grupo2} - 0.1167515 * \text{grupo3} \end{aligned}$$

En este modelo se puede observar que el grupo 2 y el grupo 1 tienen una afectación similar al error promedio ponderado.

4.8.1. Comprobación de los supuestos error promedio, edad y m³ registrados y grupos.

En la tabla de 4.23 se muestra la comprobación de los supuestos para este modelo de regresión

Tabla 4. 23 Verificación de supuestos.

Supuesto	Test	Valor p	Resultado
Normalidad	shapiro.test	0,0008147	Se rechaza Ho. No Existe evidencia de la normalidad
Homocedasticidad	Breusch-Pagan	0,007003	Se rechaza Ho. Los residuos no son homocedásticos
AutoCorrelación	Durbin-Watson	0,9674	No rechazo Ho. No existe auto correlación

En este modelo no se cumple los supuestos de normalidad y homecedastidad, además el grupo 2 es semejante al grupo 1 para explicar el error promedio ponderado.

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- Se pudo determinar que el 99.10% de medidores corresponden a medidores de ½ pulgada o 15mm, que la edad promedio del parque de medidores es 4.63 años. El 98.23% corresponden a dos marcas específicas. Con respecto a la tecnología de construcción tenemos que el 74.06% de medidores son chorro único y el 25.22% son volumétricos.
- Él estudio arrojó que la estimación de la precisión del parque de medidores residenciales es de un 93.06%, concluyendo que se deja de registrar el 6.94% del volumen de agua facturado por causa de la submedición que producen los medidores de agua.
- Varios son los factores que afectan el desempeño de un medidor entre estas el deterioro del equipo por la edad y el volumen registrado.
- De los resultados obtenidos, se puede observar que entre las variables analizadas las que más influyen en la evolución del error promedio ponderado son la edad del medidor y el volumen registrado. La forma en que se ha ido degradando el error en función de la edad se ajusta a un modelo lineal con un $R^2=83.98\%$. La degradación con respecto al volumen consumido también se ajusta a una función lineal con un $R^2=94.75\%$. El modelo lineal múltiple puede explicar el 73,54% de la variabilidad observada en el error promedio ponderado con respecto a la edad y volumen registrado.

5.2. Recomendaciones

- Las pérdidas por agua no facturada debido a las imprecisiones de los equipos de medición es un escenario con el que la empresa tendrá que enfrentarse a lo largo de su existencia, por lo cual debe mantener una gestión inteligente de renovación de medidores residenciales.
- Realizar seguimiento de nuevos modelos de medidores a través de ejercicios de perfiles de consumo y ensayos con la finalidad de comprobar la calidad y eficiencia de estos equipos.

6. Referencias

- D. Stoker, S. Barfuss and M. Johnson, (2012). Flow measurement accuracies of in-service residential water meters, *Journal AWWA*, vol. 104, pp. 637-642.
- D. Gonza, F. Tekleweld, T. Hishe, et al. (2021). Performance analysis of water meters for measuring domestic water consumption, the case of Dire Dawa, Ethiopia, *Int J Energ Water Res*, Vol. 5, pp. 405-412.
- A. Sánchez. (2015). Evaluación del error global de contadores de agua domésticos y análisis de su degradación a partir de una muestra real. Análisis económico como criterio de renovación. Trabajo Fin de Master, Universidad Politécnica de Valencia, Valencia, España.
- F. Arregui de la Cruz, G. García, G. López, & J. Martínez. (1998) Metodología para la evaluación de error de medición, *Ingeniería del Agua*, Vol. 5(4), pp. 55-66.
- F. Tortes, F. Aparici, & F. Negre, (2017). Estudio del comportamiento metrológico de los contadores en abastecimientos de agua.
- LÓPEZ-GONZÁLEZ, E., & RUIZ-SOLER, M. (2011). Análisis de datos con el Modelo Lineal Generalizado. Una aplicación con R. *Revista Española de Pedagogía*, 69(248), 59–80. <http://www.jstor.org/stable/23766383>
- Mañas, A., & Ángel, M. (2017). Clustering en redes bayesianas. Implementación en R.
- Mendenhall, W., Beaver, R. J., Beaver, B. M., & Velázquez Arellano, J. A. (2015). *Introducción a la probabilidad y estadística* (14a. ed. --.). México, D.F.: Cengage Learning.
- Montgomery, D., (2008) *Diseño y análisis de experimentos*, 2da edición, Limusa-Wiley, México
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370–384. <https://doi.org/10.2307/2344614>.
-
- Peña, D. (2002). *Análisis de Datos Multivariantes*. Mc Graw - Hill. España.
- International Organization for Standardization. (2014). *Medidores de agua para agua potable fría y caliente - Parte 1: Requisitos metrológicos y técnicos (ISO 4064-1)*. <https://www.iso.org/standard/55371.html>
- Registro Oficial – Edición especial nº 522, Reglamento interno de manejo de los servicios de agua potable, alcantarillado sanitario y drenaje pluvial de la ciudad de Guayaquil. 22 de agosto de 2018.

<https://www.registroficial.gob.ec/index.php/registro-oficial-web/publicaciones/ediciones-especiales/item/10840-edicion-especial-no-522>

Organización Internacional de Normalización e Comisión Electrotécnica Internacional (2007). International vocabulary of metrology — Basic and general concepts and associated terms (VIM) (ISO/IEC Guide 99). <https://www.iso.org/standard/45324.html>

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

7. Apéndices y anexos

ANEXO A

Tabla A.1 Tabla Error, edad, m³ registrados y grupos

Edad	m ³ registrados	Promedio de m ³ registrados	Promedio de Error ponderado
0	200	138,01	-3,77
0	750	613,44	-1,50
0	1250	1135,00	-1,00
1	200	186,45	-1,72
1	750	624,13	-3,37
1	1250	1269,00	-0,42
1	1750	1777,50	-1,76
1	2250	2427,00	-1,74
1	3250	3298,00	-3,22
1	3750	3535,00	2,55
2	200	302,50	-3,63
2	750	716,64	-2,58
2	1250	1215,49	-1,57
2	1750	1658,71	-0,87
2	2250	2287,33	-4,56
2	2750	2982,00	-0,73
2	3250	3246,00	0,30
3	200	315,95	-5,43
3	750	726,13	-4,80
3	1250	1187,61	-4,74
3	1750	1684,00	-6,30
3	2250	2200,44	-4,63
3	2750	2842,25	-0,74
3	3250	3172,50	-7,00
3	5000	4861,00	-1,51
4	200	319,46	-5,89
4	750	749,96	-5,45
4	1250	1233,16	-4,01
4	1750	1685,63	-6,39
4	2250	2191,30	-1,42
4	2750	2695,50	-17,91
4	3250	3185,00	-0,91
4	3750	3847,00	-27,47
4	5000	5574,00	1,49
5	200	273,54	-7,94
5	750	757,44	-10,89
5	1250	1240,78	-8,56
5	1750	1716,60	-4,44

5	2250	2233,94	-5,23
5	2750	2690,50	-5,81
5	3250	3264,00	-15,85
5	3750	3851,00	-25,58
5	5000	5317,40	-13,87
6	200	264,72	-4,88
6	750	791,86	-6,49
6	1250	1241,00	-13,83
6	1750	1776,79	-10,65
6	2250	2221,07	-4,46
6	2750	2773,50	-10,26
6	3250	3389,00	-5,82
6	3750	3822,00	-39,13
6	5000	5063,50	-31,85
7	200	277,10	-15,87
7	750	783,74	-10,78
7	1250	1267,74	-10,84
7	1750	1730,80	-12,87
7	2250	2157,42	-13,84
7	2750	2761,22	-16,80
7	3250	3322,14	-7,75
7	3750	3647,50	-1,57
7	5000	6579,40	-12,20
8	200	291,95	-4,84
8	750	794,67	-15,39
8	1250	1231,72	-6,63
8	1750	1687,82	-24,50
8	2250	2186,00	-24,12
8	2750	2789,80	-21,35
8	3250	3201,75	-17,52
8	3750	3627,33	-37,86
8	5000	5114,83	-43,05
9	200	364,50	-21,61
9	750	709,25	-12,16
9	1250	1237,25	-17,38
9	1750	1762,42	-22,96
9	2250	2268,82	-19,11
9	2750	2754,50	-20,91
9	3250	3265,25	-23,54
9	3750	3664,00	-15,85
9	5000	6028,25	-41,61
10	200	131,45	-17,32
10	750	805,28	-11,87
10	1250	1280,57	-10,13
10	1750	1787,33	-27,22
10	2250	2160,38	-20,11
10	2750	2771,33	-17,43
10	3250	3254,40	-41,14

10	3750	3642,00	-20,81
10	5000	5727,67	-46,47
11	200	263,14	-7,28
11	750	760,45	-24,60
11	1250	1263,81	-18,31
11	1750	1725,78	-21,72
11	2250	2348,81	-24,43
11	2750	2688,46	-28,43
11	3250	3182,50	-27,45
11	3750	3679,80	-40,27
11	5000	4894,44	-23,93
12	200	271,00	-9,61
12	750	864,36	-21,67
12	1250	1256,13	-20,03
12	1750	1742,20	-24,00
12	2250	2193,50	-27,07
12	2750	2707,00	-27,66
12	3250	3272,20	-16,79
12	3750	3829,25	-46,23
12	5000	5313,10	-36,68
13	200	160,41	-2,05
13	750	802,00	-8,25
13	1250	1272,30	-19,97
13	1750	1713,43	-21,44
13	2250	2205,67	-27,41
13	2750	2736,93	-31,90
13	3250	3256,60	-34,83
13	3750	3666,57	-30,38
13	5000	5918,20	-26,83
14	200	238,67	-13,42
14	750	690,55	-14,06
14	1250	1248,09	-29,03
14	1750	1749,00	-15,37
14	2250	2276,67	-18,65
14	2750	2770,00	-50,16
14	3250	3207,60	-18,06
14	3750	3550,00	-10,32
14	5000	5268,33	-41,17
15	200	244,78	-3,74
15	750	777,40	-28,53
15	1250	1263,77	-10,91
15	1750	1793,91	-30,38
15	2250	2167,00	-24,61
15	2750	2849,25	-26,24
15	3250	3182,67	-8,66
15	3750	3768,50	-39,65
15	5000	6439,33	-57,98
16	750	603,00	-19,65

16	1250	1242,50	-18,85
16	1750	1718,44	-12,93
16	2250	2192,33	-10,19
16	2750	2714,00	-29,41
16	3250	3099,50	-32,90
16	3750	3651,50	-23,43
16	5000	4009,00	-14,16
17	1250	1313,50	-8,38
17	1750	1714,67	-26,25
17	2250	2425,00	-49,18
17	5000	8547,00	-28,40
18	750	662,00	-4,84
18	1250	1071,00	-9,28
18	1750	1770,00	-99,37
18	5000	5491,00	-26,59
