



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE MECÁNICA
CARRERA MANTENIMIENTO INDUSTRIAL

**ANÁLISIS ESTADÍSTICO DE LOS FACTORES QUE
INTERVIENEN EN EL DESGASTE DE LA TUBERÍA DE
REGADÍO DE LA COMUNIDAD SAN JOSÉ LA SILVERIA,
PARROQUIA SAN ANDRÉS, PROVINCIA DE CHIMBORAZO**

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERO EN MANTENIMIENTO INDUSTRIAL

AUTOR:

JOAQUÍN DANIEL ZAMBRANO PINEDA

Riobamba - Ecuador

2022



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE MECÁNICA
CARRERA MANTENIMIENTO INDUSTRIAL

**ANÁLISIS ESTADÍSTICO DE LOS FACTORES QUE
INTERVIENEN EN EL DESGASTE DE LA TUBERÍA DE
REGADÍO DE LA COMUNIDAD SAN JOSÉ LA SILVERIA,
PARROQUIA SAN ANDRÉS, PROVINCIA DE CHIMBORAZO**

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERO EN MANTENIMIENTO INDUSTRIAL

AUTOR: JOAQUÍN DANIEL ZAMBRANO PINEDA

DIRECTOR: Ing. EDISSON FERNANDO CALDERÓN FREIRE

Riobamba - Ecuador

2022

© 2022, Joaquín Daniel Zambrano Pineda

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Yo, Joaquín Daniel Zambrano Pineda, declaro que el presente trabajo de titulación es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor asumo la responsabilidad legal y académica de los contenidos de este trabajo de titulación; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 31 de mayo de 2022

A handwritten signature in blue ink, consisting of stylized, overlapping loops and lines, likely representing the initials 'JDP'.

Joaquín Daniel Zambrano Pineda
180434904-9

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE MECÁNICA

CARRERA MANTENIMIENTO INDUSTRIAL

El Tribunal del Trabajo de Integración Curricular certifica que: El Trabajo de Integración Curricular; Tipo: Proyecto de Investigación; **ANÁLISIS ESTADÍSTICO DE LOS FACTORES QUE INTERVIENEN EN EL DESGASTE DE LA TUBERÍA DE REGADÍO DE LA COMUNIDAD SAN JOSÉ LA SILVERIA, PARROQUIA SAN ANDRÉS, PROVINCIA DE CHIMBORAZO**, realizado por el señor: **JOAQUÍN DANIEL ZAMBRANO PINEDA**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Integración Curricular, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal Autoriza su presentación.

	FIRMA	FECHA
Ing. Marco Antonio Ordoñez Viñán PRESIDENTE DEL TRIBUNAL		2022-05-31
Ing. Edison Fernando Calderón Freire DIRECTOR DE TRABAJO DE INTEGRACIÓN CURRICULAR		2022-05-31
Ing. Eduardo Segundo Hernández Dávila MIEMBRO DE TRIBUNAL		2022-05-31

DEDICATORIA

Quiero dedicar esta tesis de grado a Dios por permitirme culminar con éxito mi tan anhelada Carrera de Mantenimiento Industrial, darme buena salud y fortaleza en todo momento. A mis padres, por ser pilar fundamental en mi vida, inculcarme buenos valores y darme la mejor educación. A mi novia Karen Morales, por darme su fortaleza e inspiración. Por creer siempre en mí y decirme a diario que si podía lograrlo.

Joaquín

AGRADECIMIENTO

A la Escuela Superior Politécnica de Chimborazo, en especial a la Carrera de Mantenimiento Industrial por formarme como Ingeniero en Mantenimiento Industrial y dentro de ello a mis maestros por sus conocimientos y experiencias otorgadas en cada semestre.

Al Ing. Edison Fernando Calderón Freire por sus valiosos comentarios, asesoramiento y apoyo dentro y fuera de este Trabajo de Integración Curricular.

Al Ing. Eduardo Segundo Hernández Dávila por su valiosa participación en el asesoramiento de este Trabajo de Integración Curricular.

Joaquín

TABLA DE CONTENIDO

ÍNDICE DE TABLAS.....	ix
ÍNDICE DE FIGURAS.....	xi
ÍNDICE DE GRÁFICOS.....	xiii
ÍNDICE DE ECUACIONES	xiv
ÍNDICE DE ABREVIATURAS.....	xv
ÍNDICE DE ANEXOS	xvi
RESUMEN.....	xvii
SUMMARY.....	xviii
INTRODUCCIÓN	1

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL.....	5
1.1. Definiciones básicas	5
1.2. Mecanismos de desgaste.....	5
1.2.1. <i>Clasificación del proceso de desgaste</i>	6
1.2.2. <i>Abrasión</i>	7
1.2.3. <i>Erosión</i>	7
1.2.4. <i>Corrosión</i>	8
1.2.4.1. <i>Factores de la corrosión por erosión en tuberías internas</i>	8
1.3. Estadística.....	5
1.3.1. <i>Importancia de la estadística</i>	9
1.3.2. <i>Estadística descriptiva</i>	10
1.3.3. <i>Medidas de tendencia central</i>	11
1.3.3.1. <i>Media</i>	11
1.3.3.2. <i>Mediana</i>	12
1.3.3.3. <i>Moda</i>	12
1.3.4. <i>Medidas de dispersión</i>	12
1.3.4.1. <i>Rango</i>	12
1.3.4.2. <i>Varianza</i>	13
1.3.4.3. <i>Desviación estándar</i>	13
1.3.4.4. <i>Coficiente de variación</i>	13
1.3.5. <i>Análisis Exploratorio de Datos (EDA)</i>	14
1.3.6. <i>Análisis de datos atípicos</i>	14

1.3.7.	<i>Prueba de normalidad</i>	15
1.3.7.1.	<i>Prueba Shapiro Wilk</i>	15
1.3.7.2.	<i>Prueba Kolmogorov-Smirnov</i>	16
1.3.7.3.	<i>Estadístico t de Student</i>	16
1.3.7.4.	<i>Hipótesis estadística</i>	17
1.3.8.	<i>Teorema del límite central</i>	17
1.3.8.1.	<i>Métodos Bootstrap</i>	18
1.4.	Ciencia de datos	18
1.5.	R	19
1.5.1.	<i>Ventajas del uso de R</i>	20
1.6.	RStudio	20
1.6.1.	<i>Paquetes en RStudio</i>	21
1.6.2.	<i>Descripciones de gráficas de los datos</i>	21
1.6.2.1.	<i>Histograma</i>	21
1.6.2.2.	<i>Diagrama de caja</i>	22
1.6.2.3.	<i>Estimadores Kernel de la densidad</i>	22
1.6.2.4.	<i>Gráficos cuantil-cuantil (Q-Q plots)</i>	23

CAPÍTULO II

2.	MARCO METODOLÓGICO	24
2.1.	Tipo y diseño de investigación	25
2.2.	Población y muestra	26
2.3.	Obtención y preparación de datos de la tubería de regadío	26
2.3.1.	<i>Descripción de la tubería de agua de regadío</i>	27
2.3.2.	<i>Importar datos de Excel a RStudio</i>	27
2.3.3.	<i>Descripción de la base de datos de la tubería de regadío en RStudio</i>	28
2.4.	Instalación de librerías para el análisis de datos	28
2.5.	Análisis exploratorio de los datos	29
2.5.1.	<i>Prueba de normalidad de los datos</i>	30
2.5.2.	<i>Remuestreo de Bootstrap</i>	31
2.5.3.	<i>Reemplazo de los valores atípicos (Outliers)</i>	31
2.5.4.	<i>Prueba t de Student</i>	34

CAPÍTULO III

3.	MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS	36
----	--	----

3.1.	Análisis estadístico de cada tramo de la tubería de regadío	36
3.1.1.	<i>Tramo N° 01</i>	36
3.1.2.	<i>Tramo N° 02</i>	37
3.1.3.	<i>Tramo N° 03</i>	38
3.1.4.	<i>Tramo N° 04</i>	39
3.1.5.	<i>Tramo N° 05</i>	40
3.1.6.	<i>Tramo N° 06</i>	42
3.1.7.	<i>Tramo N° 07</i>	43
3.1.8.	<i>Tramo N° 08</i>	44
3.1.9.	<i>Tramo N° 09</i>	45
3.1.10.	<i>Tramo N° 10</i>	46
3.1.11.	<i>Tramo N° 11</i>	47
3.1.12.	<i>Tramo N° 12</i>	48
3.1.13.	<i>Tramo N° 13</i>	49
3.1.14.	<i>Tramo N° 14</i>	50
3.1.15.	<i>Tramo N° 15</i>	51
3.1.16.	<i>Punto Antes de la Soldadura (PAS)</i>	52
3.1.17.	<i>Punto Antes de la Soldadura (PDS)</i>	53
3.2.	Resumen estadístico de cada tramo de la tubería de regadío	55
3.2.1.	<i>Análisis comparativo y descriptivo de mayor espesor</i>	55
3.2.2.	<i>Análisis comparativo y descriptivo de menor espesor</i>	56
3.2.3.	<i>Análisis comparativo y descriptivo de mayor dispersión de datos</i>	58
3.3.	Análisis de resultados	59
	CONCLUSIONES	62
	RECOMENDACIONES	63
	BIBLIOGRAFÍA	
	ANEXOS	

ÍNDICE DE TABLAS

Tabla 1-1:	Medidas descriptivas	11
Tabla 1-2:	Conjunto de datos de la tubería de agua de regadío	27
Tabla 2-2:	Estadísticos para la aplicación del análisis de datos.....	28
Tabla 3-2:	Análisis descriptivo de la tubería de regadío.....	29
Tabla 4-2:	Resultado del análisis estadístico de los datos.....	33
Tabla 1-3:	Resultado del análisis estadístico del tramo N° 1	36
Tabla 2-3:	Prueba t-Student para el tramo N° 1	36
Tabla 3-3:	Resultado del análisis estadístico del tramo N° 2.....	37
Tabla 4-3:	Prueba t-Student para el tramo N° 2	37
Tabla 5-3:	Resultado del análisis estadístico del tramo N° 3.....	38
Tabla 6-3:	Prueba t-Student para el tramo N° 3	39
Tabla 7-3:	Resultado del análisis estadístico del tramo N° 4.....	39
Tabla 8-3:	Prueba t-Student para el tramo N° 4.....	40
Tabla 9-3:	Resultado del análisis estadístico del tramo N° 5.....	41
Tabla 10-3:	Prueba t-Student para el tramo N° 5	41
Tabla 11-3:	Resultado del análisis estadístico del tramo N° 6.....	42
Tabla 12-3:	Prueba t-Student para el tramo N° 6.....	42
Tabla 13-3:	Resultado del análisis estadístico del tramo N° 7.....	43
Tabla 14-3:	Prueba t-Student para el tramo N° 7.....	43
Tabla 15-3:	Resultado del análisis estadístico del tramo N° 8.....	44
Tabla 16-3:	Prueba t-Student para el tramo N° 8.....	44
Tabla 17-3:	Resultado del análisis estadístico del tramo N° 9.....	45
Tabla 18-3:	Prueba t-Student para el tramo N° 9.....	45
Tabla 19-3:	Resultado del análisis estadístico del tramo N° 10.....	46
Tabla 20-3:	Prueba t-Student para el tramo N° 10.....	46
Tabla 21-3:	Resultado del análisis estadístico del tramo N° 11.....	47
Tabla 22-3:	Prueba t-Student para el tramo N° 11.....	47
Tabla 23-3:	Resultado del análisis estadístico del tramo N° 12.....	48
Tabla 24-3:	Prueba t-Student para el tramo N° 12.....	48
Tabla 25-3:	Resultado del análisis estadístico del tramo N° 13.....	49
Tabla 26-3:	Prueba t-Student para el tramo N° 13.....	49
Tabla 27-3:	Resultado del análisis estadístico del tramo N° 14.....	50
Tabla 28-3:	Prueba t-Student para el tramo N° 14.....	50
Tabla 29-3:	Resultado del análisis estadístico del tramo N° 15.....	51

Tabla 30-3:	Prueba t-Student para el tramo N° 15	51
Tabla 31-3:	Resultado del análisis estadístico de los PAS.....	52
Tabla 32-3:	Prueba t-Student de los PAS.....	52
Tabla 33-3:	Resultado del análisis estadístico del tramo de los PDS	53
Tabla 34-3:	Prueba t-Student de los PDS.....	54
Tabla 35-3:	Resumen de mayor espesor en función de la mediana	55
Tabla 36-3:	Resumen de menor espesor en función de la mediana	56
Tabla 37-3:	Resumen de la mayor desviación estándar	58
Tabla 38-3:	Síntesis de resultados.....	59

ÍNDICE DE FIGURAS

Figura 1-1:	Tipos de Desgaste	6
Figura 2-1:	Clasificación general del desgaste	7
Figura 3-1:	Distribución normal de los datos	15
Figura 4-1:	Diagrama de proceso de Bootstrap	18
Figura 5-1:	Etapas en la aplicación de ciencia de datos.....	19
Figura 6-1:	Interfaz del software R.....	19
Figura 7-1:	Interfaz del software RStudio	20
Figura 8-1:	Histograma.....	21
Figura 9-1:	Diagrama de caja y bigote.....	22
Figura 10-1:	Gráfico de densidad	23
Figura 11-1:	Gráficos cuantil-cuantil (QQplot)	23
Figura 1-2:	Fase metodológica	24
Figura 2-2:	Sistema de riego de la tubería de regadío	25
Figura 3-2:	Bases de datos de la tubería de regadío.....	26
Figura 4-2:	Importación de la base de datos de Excel a RStudio	27
Figura 5-2:	Base de datos en RStudio de la tubería de regadío	28
Figura 6-2:	Histograma de los datos	29
Figura 7-2:	Diagrama de caja de normalidad de los datos.....	30
Figura 8-2:	Diagrama de caja de la distribución de Bootstrap.....	31
Figura 9-2:	Diagrama de caja del reemplazo de los valores atípicos.....	32
Figura 10-2:	Gráficos de los datos normalizados	33
Figura 11-2:	Prueba t-Student entre la referencia A y C	34
Figura 12-2:	Prueba t-Student entre la referencia B y C.....	34
Figura 13-2:	Prueba t-Student entre la referencia D y C	35
Figura 14-2:	Diagrama de caja de la prueba t de Student	35
Figura 1-3:	Diagrama de caja y bigote del tramo N° 1	37
Figura 2-3:	Diagrama de caja y bigote del tramo N° 2	38
Figura 3-3:	Diagrama de caja y bigote del tramo N° 3	39
Figura 4-3:	Diagrama de caja y bigote del tramo N° 4	40
Figura 5-3:	Diagrama de caja y bigote del tramo N° 5	41
Figura 6-3:	Diagrama de caja y bigote del tramo N° 6.....	42
Figura 7-3:	Diagrama de caja y bigote del tramo N° 7	43
Figura 8-3:	Diagrama de caja y bigote del tramo N° 8.....	44
Figura 9-3:	Diagrama de caja y bigote del tramo N° 9	45

Figura 10-3:	Diagrama de caja y bigote del tramo N° 10	46
Figura 11-3:	Diagrama de caja y bigote del tramo N° 11	47
Figura 12-3:	Diagrama de caja y bigote del tramo N° 12	48
Figura 14-3:	Diagrama de caja y bigote del tramo N° 13	49
Figura 14-3:	Diagrama de caja y bigote del tramo N° 14	50
Figura 15-3:	Diagrama de caja y bigote del tramo N° 15	52
Figura 16-3:	Diagrama de caja y bigote de los PAS	53
Figura 17-3:	Diagrama de caja y bigote de los PDS	54
Figura 18-3:	PAS y PDS	54
Figura 18-3:	Ubicación geográfica de las referencias (A, B y D).....	56
Figura 19-3:	Ubicación geográfica de la referencia C	57
Figura 20-3:	Tramo N° 06 con mayor desgaste	60
Figura 21-3:	Ubicación geográfica de los PAS y PDS	61

ÍNDICE DE GRÁFICOS

Gráfico 1-3:	Curva de tendencia de mayor espesor en función de la mediana	55
Gráfico 2-3:	Curva de tendencia de menor espesor en función de la mediana	57
Gráfico 3-3:	Curva de tendencia de mayor desviación estándar	58

ÍNDICE DE ECUACIONES

Ecuación 1-1:	Media	11
Ecuación 2-1:	Mediana	12
Ecuación 3-1:	Moda	12
Ecuación 4-1:	Rango	12
Ecuación 5-1:	Varianza	13
Ecuación 6-1:	Varianza de la población	13
Ecuación 7-1:	Desviación estandar	13
Ecuación 8-1:	Coefficiente de variación	13
Ecuación 9-1:	Shapiro-Wilk	16
Ecuación 10-1:	Kolmogorov-Smirnov	16
Ecuación 11-1:	T de Student	16

ÍNDICE DE ABREVIATURAS

AWWA:	Asociación Americana de Obras Hidráulicas
ASTM:	Sociedad Americana para Pruebas y Materiales
API:	Instituto Americano del Petróleo
API 570:	Código de inspección de tuberías
IDE:	Entorno Integrado de Desarrollo
EDA:	Análisis Exploratorio de Datos
PAS:	Puntos Antes de la Soldadura
PDS:	Puntos Después de la Soldadura
IQR:	Rango intercuartílico o intercuartil

ÍNDICE DE ANEXOS

Anexo A: Código en RStudio

Anexo B: Trayectoria de la tubería de agua de regadío

RESUMEN

El objetivo de la presente investigación fue analizar los factores que intervienen en el desgaste de la tubería de regadío de la Comunidad San José la Silveria de la parroquia San Andrés, provincia de Chimborazo. Se realizó un Análisis Exploratorio de Datos (EDA), la prueba de Kolmogorov-Smirnov, el método de Bootstrap y la prueba T de Student, partiendo de los datos secundarios que tuvieron su respectiva medición de espesores en el año 2019. Como segundo pilar de la investigación se utilizó el software RStudio para realizar el análisis estadístico de los datos en función de las medidas descriptivas que exponen cuatro variables como puntos de referencia (A, B, C, D), que son ubicaciones geométricas de la tubería de regadío, seguido de la comparación entre la referencia (A-C, B-C, D-C), para contrastar el menor desgaste en los 15 tramos de la tubería y, además, en los Puntos Antes de la Soldadura (PAS) y Puntos Después de la Soldadura (PDS). Con el software RStudio, se evidenció que los espesores en los puntos de referencia A, B y D presentan un desgaste mínimo de 25,4%, y el punto de referencia C presenta el mayor desgaste, con una pérdida progresiva de 74,6%, ocasionando cortes, rayones, astillado y disminuyendo así, la vida útil de la tubería. Concluyendo que la tendencia que tiene menor espesor se encuentra en la referencia C, ya que se muestra a través de sus interpretaciones el tipo de desgaste más predominante, que, en esta investigación fue el desgaste por abrasión. Finalmente, se recomienda a la Comunidad San José de la Silveria y sus alrededores, realizar mingas de mantenimiento preventivo, aplicando pintura anticorrosiva en las superficies externas de la tubería de regadío, para propiciar la prolongación de la vida útil del sistema de riego y aportar a la producción de la zona.

Palabras clave: <DESGASTE DE TUBERÍA> <ANÁLISIS ESTADÍSTICO> <ANÁLISIS EXPLORATORIO DE DATOS (EDA)> <ABRASIÓN> <RSTUDIO (SOFTWARE)>.



Firmado electrónicamente por:
**HOLGER GERMAN
RAMOS UVIDIA**

1278-DBRA-UTP-2022

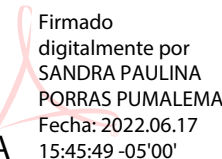
2022-06-29

SUMMARY

The objective of this investigation was to analyze the factors to consider in the irrigation pipe wear at San José la Silveria Community, San Andrés parish, Chimborazo province. An Exploratory Data Analysis (EDA), the Kolmogorov-Smirnov test, the Bootstrap method and the Student's T-test were carried out, based on the secondary data that had their respective thickness measurement in the year 2019. As a second pillar of investigation, the RStudio software was used to perform the statistical analysis of the data based on the descriptive measures that expose four variables as reference points (A, B, C, D), which are geometric locations of the irrigation pipe, followed of the comparison between the reference (A-C, B-C, D-C), to contrast the least wear in the 15 sections of the pipe. In addition, in the Points Before Welding (PAS) and Points After Welding (PDS). With the RStudio software, it was shown that the thicknesses at reference points A, B and D present a minimum wear of 25,4%, and reference point C presents the greatest wear, with a progressive loss of 74,6%. , causing cuts, scratches, chipping and thus reducing the useful life of the pipe. It is concluded that the trend that has less thickness is found in reference C, since the most predominant type of wear is shown through its interpretations, which, in this investigation, was wear by abrasion. Finally, it is recommended that San José de la Silveria Community and its surroundings carry out preventive maintenance mingas, applying anticorrosive paint on the external surfaces of the irrigation pipe, to promote the extension of the useful life of the irrigation system and contribute to the production of the area.

Keywords: <PIPE WEAR> <STATISTICS ANALYSIS> < EXPLORATORY DATA ANALYSIS (EDA)> <ABRASION> <RSTUDIO (SOFTWARE)>.

SANDRA
PAULINA
PORRAS
PUMALEMA



Firmado
digitalmente por
SANDRA PAULINA
PORRAS PUMALEMA
Fecha: 2022.06.17
15:45:49 -05'00'

Sandra Paulina Porrás Pumalema

C.I. 0603357062

INTRODUCCIÓN

El desgaste es uno de los temas de investigación más importante hoy en día, ya que, en la utilización de tuberías para transporte de agua, gases, oleoductos y derivados, se presentan una serie de causas, relacionadas con la pérdida de material, cambios estructurales, deformación plástica, agrietamiento, fisuras superficiales, entre otros, que involucran abrasión, erosión y corrosión, lo que puede ser perjudicial en la economía agrícola e industrial.

El desgaste se define como un desprendimiento de material de una superficie, mediante la interacción con otras superficies. Casi todas las máquinas pierden su durabilidad capacidad y fiabilidad debido al desgaste. Por lo tanto, el control de este debe ser inspeccionado periódicamente, apoyándose de las bondades tecnológicas existentes en el mercado industrial.

El Análisis Exploratorio de Datos (EDA), se utiliza como un análisis previo visual y numérico para encontrar medidas, descripciones y comparaciones, que permitan explorar los datos en búsqueda de patrones inesperados. Escritores clásicos como John Tukey, Frederick Hartwig y Brian Dearing lo clasificaron como un "estado mental". Es decir, un reconocimiento de que el ojo en el horizonte es la mejor herramienta que podemos utilizar para observar lo completamente impredecible.

Por su parte, Hartwig y Dearing sostienen que los investigadores deben aprender tanto como sea posible sobre una variable o grupo de variables, antes de usar los datos para probar hipótesis o teorías sobre las relaciones sociales. Más recientemente, Eugene Horper y Dominique Ladiray han sugerido que el "pensamiento exploratorio" es un plan de análisis que enriquece la capacidad del investigador para encontrar nuevas respuestas a los problemas emergentes.

La distribución normal, también conocida como distribución *gaussiana*, es una de las distribuciones teóricas que utilizan en la práctica, e investigaciones un contraste de hipótesis por medio de pruebas paramétricas y no paramétricas.

Planteamiento del problema

En los últimos 49 años, la tubería de regadío de la Comunidad San José la Silveria ha presentado cierto grado de desgaste, el cual puede tener varias causas como son: abrasión, erosión, y corrosión, por lo que es vital una investigación de análisis estadístico de los factores que intervienen en el desgaste de la tubería de regadío. Es así como el tipo de desgaste dominante puede modificar de uno a otra forma las características de la superficie de la tubería, causadas por el tamaño de partícula, forma, dureza, concentración, densidad y velocidad, y sumando a esto cuenta con el respectivo análisis de medición de espesores y toma de datos que se realizó en la tubería de regadío mediante ensayo no destructivos empleando ultrasonido.

La presente investigación permitirá resolver preguntas como: ¿Existen factores que intervienen en el desgaste de la tubería de regadío de la Comunidad San José la Silveria?, ¿Los factores que intervienen en el desgaste de la tubería es condición directa de la abrasión, erosión, y corrosión?

Justificación

(Calderón Andrés et al, 2005) proponen un “Análisis estadístico de los factores que intervienen en la corrosión de tuberías enterradas”. En este artículo se describen las técnicas experimentales empleadas y el análisis estadístico llevado a cabo sobre 4 variables de mayor incidencia en la estimación de la agresividad del suelo, según la norma AWWA C-105 (Asociación Americana de Obras Hidráulicas): resistividad, potencial redox, pH y temperatura. A partir del análisis multivariado y la clasificación por grupos pudo obtenerse un conjunto de 3 variables, las cuales permiten hacer una clasificación de los diferentes tipos de suelo según su corrosividad, con confiabilidad del 90% y error del 20%.

Actualmente la tubería de regadío cuenta un total de 7548 m de longitud la misma que cuenta con su respectiva medición de espesores y toma de datos realizado en el año 2019, dejando por realizar un análisis estadístico de los factores que intervienen en el desgaste de la tubería. Se estima que en la mayor parte se encuentra con desgaste lo cual hasta el momento no se establece en dar una solución a este problema.

OBJETIVOS

Objetivo general

Analizar los factores que intervienen en el desgaste de la tubería de regadío de la Comunidad San José la Silveria, parroquia San Andrés, provincia de Chimborazo.

Objetivos específicos

Aplicar los conceptos estadísticos sobre los datos obtenidos a partir del ensayo no destructivo con ultrasonido.

Procesar estadísticamente los datos obtenidos de la tubería de regadío.

Utilizar el software RStudio, para determinar que factor intervienen en el desgaste de la tubería de regadío.

Determinar el desgaste más importante que se presenta en la tubería de regadío.

HIPÓTESIS

La abrasión, erosión, y corrosión influyen en el desgaste de la tubería de regadío de la Comunidad San José la Silveria, parroquia San Andrés, provincia de Chimborazo.

Variable dependiente

Desgaste de la tubería

Variables independientes

Abrasión

Erosión

Corrosión

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL

1.1. Definiciones básicas

La norma ASTM G40-02 define la tribología como la disciplina técnica y tecnológica relacionada con la interacción de superficies móviles y la interacción entre fricción, lubricación y desgaste. Según (Tristancho Reyes et al., 2008) existen distintos tipos de desgaste:

- Desgaste adhesivo, causado por la fricción entre dos superficies.
- Desgaste abrasivo, causado por la presencia de partículas en el punto de contacto.
- Desgaste erosivo, causado por el choque de partículas en las superficies.
- Desgaste por cavitación, causado por la formación de burbujas a alta presión.

La norma ASTM G40-2 define el desgaste como resultado de la pérdida o separación de material, que es causada por la fricción a través de la superficie de contacto. De igual forma, señala que la fricción como la pérdida gradual de energía cinética se da cuando los objetos o fluidos se mueven entre sí, es decir, la resistencia del material de la superficie. (Tristancho Reyes et al., 2008)

La fricción puede ocurrir en: sólidos y gases, como fricción aerodinámica, en sólido y líquido, como fricción líquida y en procesos de disipación de energía, como fricción interna.

1.2. Mecanismos de desgaste

Basado en el análisis paramétrico del desgaste de las partes mecánicas, alrededor del 50% actúa sobre el desgaste abrasivo de forma notable. Además, ha existido un enorme conjunto de procesos de desgaste que se hallan en la práctica y en la investigación. El daño por desgaste se ve concedido por la observación, pues cada proceso de desgaste es definido por la acción de un mecanismo de desgaste predominante, entre los cuales tenemos: (Ali et al., 2013)

- 15% es desgaste adhesivo.
- 8% es desgaste por erosión.
- 5% es desgaste por corrosión.
- 14% es solo una combinación de desgaste abrasivo, erosivo y corrosivo.

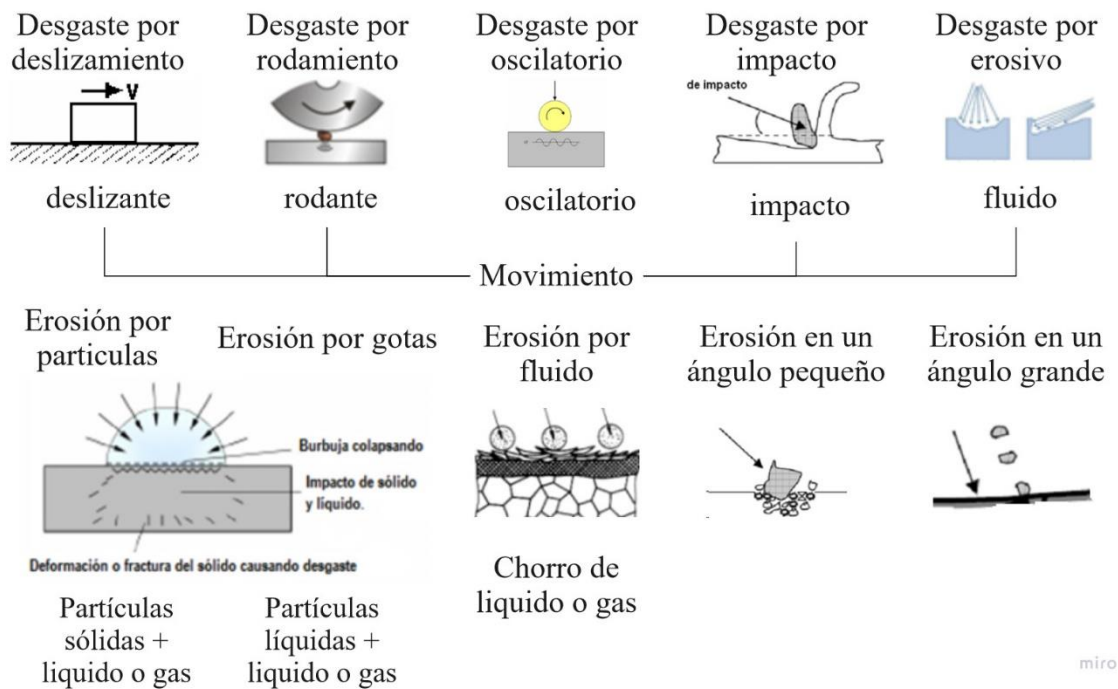


Figura 1-1: Tipos de Desgaste

Fuente: (Tristancho Reyes et al., 2008)

Realizado por: Zambrano J, 2022

1.2.1. Clasificación del proceso de desgaste

Comprender los cuatro mecanismos de desgaste es esencial para un control adecuado para cada proceso de desgaste. Esta comprobación se puede hacer durante la fase de diseño, en el cual, los mecanismos de desgaste operativos se pueden identificar y rediseñar el sistema de fricción, después de una avería provocada por el desgaste.

Para lograr esta tarea, se requiere un análisis de falla adecuado, y este solo se puede realizar si el mecanismo son características intrínsecas de los materiales, debido a que están sujetas a los materiales de acoplamiento, el tipo de contacto, la superficie característica de los cuerpos de acoplamiento. (Straffelini, 2015, p. 85)

En el análisis de las fallas por desgaste, es usual tener en cuenta los mecanismos de procesos de desgaste definidos con referencia al tipo y geometría del movimiento relativo entre dos superficies. La comprensión de dicho proceso es fundamental al instante de intervenir adecuadamente el fenómeno de desgaste. En la Figura 2-1, se muestran la clasificación general de los procesos de desgaste. (Straffelini, 2015, p. 86)

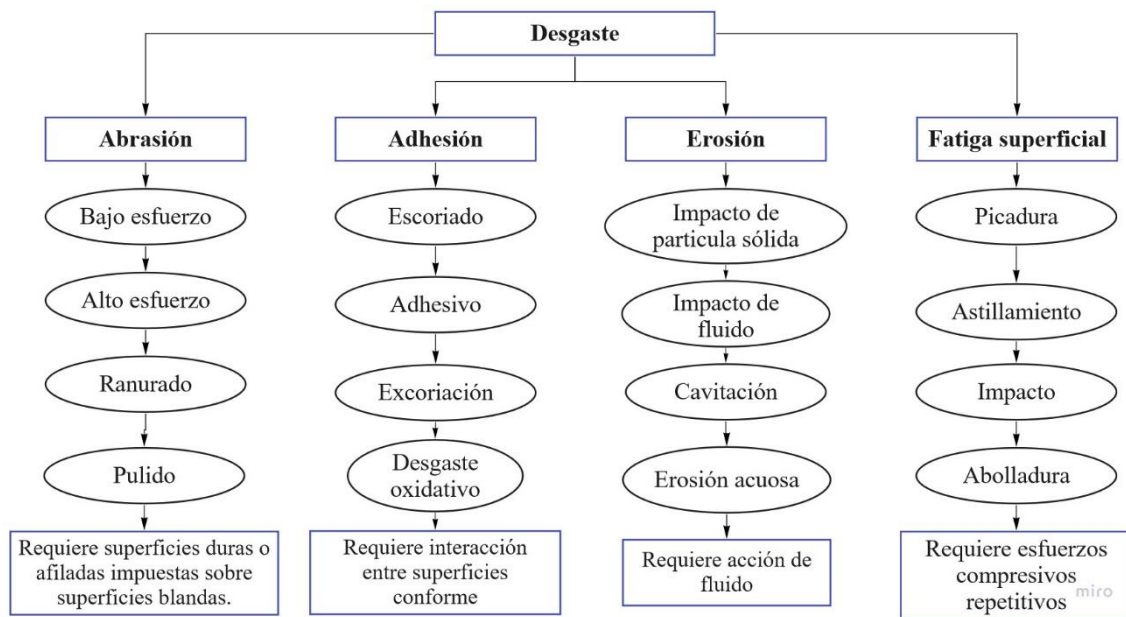


Figura 2-1: Clasificación general del desgaste

Fuente: (Davis, 2001, p. 56)

Realizado por: Zambrano J, 2022

1.2.2. *Abrasión*

El desgaste por abrasión, según la norma ASTM se produce por la pérdida de material debido a partículas duras o protuberancias del fluido, al moverse contra una superficie sólida, causando daño y pérdida progresiva del material. El 50% de los problemas de desgaste en las industrias y agricultura se deben a la abrasión, por exceso de trabajo o altas presiones. Los laboratorios han comprobado y tratado de racionalizar el comportamiento de desgaste abrasivo, al combinar una serie de elementos para evitar la pérdida de material. (Ali et al., 2013)

El desgaste abrasivo se produce cuando las partículas ásperas y duras se deslizan sobre una superficie sólida eventualmente la aparición de micropartículas o ranurado, creación de huecos, corte, fractura superficial o deformación plástica. El desgaste abrasivo generalmente ocurre por cortes, astillado o grietas por fatiga, esto puede rayar en ambas superficies llevando a formar un ligero desgaste abrasivo. (Jeyaprakash et al., 2020)

1.2.3. *Erosión*

El desgaste por erosión puede deberse a impactos repetidos de partículas sólidas, a una alta velocidad de fluido, causando la eliminación de material de la superficie. La tensión de contacto es generada por la energía cinética de las partículas en el agua o en una corriente de aire cuando se encuentra con una superficie.

En el desgaste por erosión, el ángulo de impacto y la velocidad. se combinan con la energía cinética de las partículas que chocan, formando residuos de desgaste. (Jeyaprakash et al., 2020). La erosión se debe a factores repetitivos, las variables más importantes son cuatro:

- **Ángulo de impacto:** depende del ángulo de incidencia y la velocidad del fluido de las partículas sólidas, el ángulo comprendido de 45° a 90°.
- **Dureza de materiales y partículas:** La dureza de partículas sólidas es generalmente más grande que el material, por lo que habrá desgaste.
- **Velocidad de partícula:** el tamaño de desgaste es proporcional a la velocidad del fluido y a la densidad del fluido.
- **Tamaño y forma de las partículas:** partículas de forma angular son más abrasivas que las de forma redonda.

1.2.4. Corrosión

La corrosión es un proceso natural en el que los metales en su estado de fabricación o refinado vuelven a su estado de oxidación protegiéndole de las reacciones electroquímicas que conducen a su deterioro natural. Para que acontezca la corrosión es: metal, oxígeno y electrolito.

La incrustación es la partícula que se aloja en el fondo de la tubería, llevando a la cristalización de sales sólidas generalmente del agua o salmuera. La escala depende de tres condiciones físicas: temperatura, presión y pH. Asimismo, corrosión y las incrustaciones pueden ocurrir en pozos de producción, intercambiadores de calor, calderas y tuberías. (Royani, Ahmad et al., 2021)

Según Reinchanschuss für Metallschutz la corrosión “es la devastación de un material sólido causada por un ataque de naturaleza química o electroquímica, que se inicia en el área metálica”. Fundamentalmente, se muestra una actitud espontánea entre el medio ambiente y un material metálico, generando productos que tienen la posibilidad de ser firmes, líquidos, sólidos o gaseoso, aunque en algunos casos dichos no son apreciables ni generan una pérdida de volumen o peso significativa. (Pancorbo, 2010, p. 7)

1.2.4.1. Factores de la corrosión por erosión en tuberías internas

La erosión por corrosión es una representación del daño que ocurre cuando la erosión de partículas y el flujo presentan una alta velocidad, a esto se suma la eliminación de capa protectora u otra incrustación protectora en la superficie metálica, a esto se denomina corrosión asistida por velocidad. (API 571, 2020, p. 144)

La corrosión por erosión se debe a varios factores críticos como:

- Con la erosión mecánica de partículas sólidas, la tasa de pérdida del metal dependerá de la velocidad y el número de impactos. Partículas, así como tamaño, forma, dureza y densidad. El número de partículas de impacto, la dureza del material erosionado y el ángulo de impacto.
- Para líquidos químicos, la tasa de pérdida de metal dependerá de la velocidad y la cantidad o el grado de impacto de corrosividad del líquido.
- La corrosión asistida por velocidad ocurre en muchos ambientes líquidos corrosivos. En algunos casos, hay una velocidad umbral por debajo de la cual la corrosión es mínima, pero por encima de la cual la corrosión se vuelve significativa. Y para otro caso no hay umbral específico, la tasa de corrosión es más alta solo ocurren cuando la velocidad es mayor.
- En un ambiente corrosivo para un material en particular, mayor es la erosión. Efectos de corrosión, especialmente cuando desestabilizan películas protectoras, incrustaciones u otras sustancias. Una barrera en la que se basan los metales para resistir la corrosión. cualquier factor que conduzca al aumento, la corrosividad del medio ambiente, como la temperatura, el pH, entre otros, puede aumentar la susceptibilidad a la erosión y la remoción del metal por corrosión. (API 571, 2020, pp. 144-145)

1.3. Estadística

La estadística es una ciencia que se encarga de recopilar, ordenar, representar, analizar e interpretar los datos generados en la investigación sobre hechos, individuos o grupos con el fin de extraer conclusiones congruentes, para la toma de decisiones, y predicciones futuras. (Salazar Pinto et al., 2017, p. 13)

1.3.1. Importancia de la estadística

(Salazar Pinto et al., 2017, p. 11) Mencionaron cinco razones importantes, por lo que tanto estudiantes como profesionales necesitan contar con aportes estadísticos en el espacio laboral, ya que la estadística aporta de distintas maneras, por ejemplo:

- La estadística como herramienta de trabajo, aporta métodos para sintetizar, expresar y obtener conclusiones congruentes sobre el análisis de los datos.

- Durante la investigación, la estadística es la base, para dar respuesta a las preguntas: ¿Cómo mejorar el estudio de dicha investigación?, ¿Cómo desarrollar una entre variables de los datos?, entre otras.
- En la investigación teórica, servirá para generar predicciones de los datos y el comportamiento en determinadas situaciones.
- Aporta en la investigación ya sea en lo teórica o aplicada de estudiantes y profesionales, generando información cuantitativa.
- La recopilación y el análisis de datos es muy importante para el desarrollo investigativo de los estudiantes, hoy en día la tecnología brinda una serie de técnicas estadísticas, que se ejecutan en pocos segundos y que ayudan a alcanzar resultados confiables.

De acuerdo con (Julio et al., 2013), “el punto de partida de cualquier investigación radica en la identificación y selección de problemas en donde la estadística y la investigación, conjuntamente garantice distinguir las variables analizadas, tomar decisiones, sino que también se puede utilizar como una herramienta metodológica para el contraste de la hipótesis de la investigación mediante algunos softwares estadísticos que son gratuitos”.

Básicamente, la estadística se clasifica en dos partes principales: estadística descriptiva o matemática y estadística inferencial, las cuales se expresan completamente a través de la probabilidad.

1.3.2. Estadística descriptiva

Se caracteriza por sus medidas de posición, dispersión y forma, estas se aplican a investigaciones descriptivas, para conceptualizar la muestra, como primer paso para hacer un estudio de los datos, por medio de los cuales se obtienen conclusiones verídicas que son de enorme ayuda para la toma de decisiones.

Para la ejecución de dicho estudio, los datos son logrados por medio de un respectivo procedimiento de recolección, por el cual se analizan y se convierten en datos de información válida. Luego, por medio de análisis de gráficos, se desarrolla una exploración más amplia de los datos, que servirá como eje inicial en la investigación. (Salazar Pinto et al., 2017, p. 14)

Además de las medidas de tendencia central, las medidas de variabilidad o dispersión también se utilizan en la estadística descriptiva. Estas medidas se usan para determinar la homogeneidad o heterogeneidad de los datos de una población o muestra que se está analizando en la investigación. (Arias, 2010, p. 15)

En el análisis de datos las medidas descriptivas más utilizadas, y que proporcionan resultados fiables y óptimos para la toma de decisiones. En la Tabla 1-1, se evidencian las medidas descriptivas:

Tabla 1-1: Medidas descriptivas

Descripción	Comando en RStudio
Media	mean(x)
Mediana	median(x)
Desviación estándar	sd(x)
Varianza	var(x)
Percentil	quantile (x)
Rango	rango(x)
Máximo	max(x)
Mínimo	min(x)
Rango	diff(rango(x))
Rango intercuartílico	IQR(x)
Resumen	Summary(x)

Fuente: (Ayala, 2019, pp. 15-20)

Realizado por: Zambrano J, 2022.

1.3.3. *Medidas de tendencia central*

Son llamadas también como medidas de posición, estas informan sobre el comportamiento de los valores en el centro de la distribución de datos y se denomina métrica o parámetro de tendencia central o de centralización, cuando se menciona la posición de estos parámetros en un conjunto de datos, y si estos se distribuyen normalmente. Las medidas de tendencia central son: la media, la mediana, y la moda. (Salazar Pinto et al., 2017, p. 49-54)

De igual forma, se debe tener en cuenta las variables cualitativas y variables cuantitativas, ya que se utiliza las medidas de posición o de tendencia según el tipo de variable que se está estudiando, en este caso se observa la variable cuantitativa.

1.3.3.1. *Media*

La media aritmética o promedio es una medida muy sensible a los valores que se alejan mucho de la mayor parte de los valores de una distribución. Si conocemos la media aritmética de un conjunto de números, entonces la suma de sus desviaciones en relación con su media es cero. (Arias, 2010, p. 173)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1-1)$$

1.3.3.2. Mediana

La mediana es una medida de tendencia central robusta, propia de las escalas ordinales son escalas de rangos de mayor a menor, las cuales comprende una relación de mayor que, es decir de una secuencia de números ordenados. Se utiliza para deducir la mayor tendencia. (Arias, 2010, p. 173)

$$x=L+\left(\frac{\frac{N}{2}-\sum f_i}{f_{\text{mediana}}}\right) \quad (2-1)$$

1.3.3.3. Moda

Es el valor de la variable que se muestra con más frecuencia. Se utiliza para distribuciones sumamente sesgadas, una vez que aparecen dos picos o para eliminar el efecto de los valores extremos, en una secuencia de números la moda puede no existir o la serie podría ser bimodal.

$$\text{Moda}=L+\frac{\Delta_1}{\Delta_1+\Delta_2} \times C \quad (3-1)$$

En una curva de distribución de frecuencias la moda es el valor máximo correspondiente a la curva sobre el eje de las ordenadas.

1.3.4. Medidas de dispersión

Las medidas de la dispersión tienen como propósito determinar el nivel de similitud que presenta un conjunto de datos. Dado que la tarea fundamental de la estadística es estudiar variables aleatorias, y por tanto la variabilidad que muestran los datos, dependiendo de la muestra, indicarán una menor o mayor dispersión. Estas medidas pueden ser más relevantes que las medidas de tendencia central. (Contento Rubio, 2019, p. 61)

1.3.4.1. Rango

Es el entorno estadístico de la diferencia entre el máximo y el mínimo valor del conjunto de observaciones, se define por la siguiente ecuación 4-1.

$$R = \max\{x_1, x_2, \dots, x_n\} - \min\{x_1, x_2, \dots, x_n\} \quad (4-1)$$

1.3.4.2. Varianza

Es la medida tradicional para evaluar la variabilidad de un conjunto de datos. Se dispone de notación para discriminar entre varianza poblacional (σ^2) y varianza de la muestra (s^2). Al utilizar rangos para verificar la variabilidad, solo considera los dos valores más extremos. La varianza se basa en la diferencia de cada observación desde un punto de referencia, en este caso la media. La varianza muestral, se define por la siguiente ecuación 5-1. (Contento Rubio, 2019, p. 62)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5-1)$$

De manera similar, la varianza poblacional se define en términos de la media de la población (μ) y el tamaño de la población (N), se define por la siguiente ecuación 6-1.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (6-1)$$

1.3.4.3. Desviación estándar

También llamada como desviación típica, es la raíz cuadrada de la varianza, esta medida se utilizada para manifestar el grado de dispersión de los datos de una distribución de frecuencias respecto a su media aritmética se define por la siguiente ecuación 7-1. (Contento Rubio, 2019, p. 64)

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (7-1)$$

1.3.4.4. Coeficiente de variación

Esta medida evalúa qué tan grande es la dispersión en relación con la magnitud general de las observaciones. se define por la siguiente expresión matemática.

$$CV = \frac{s}{\bar{x}} \times 100\% \quad (8-1)$$

Si los datos presentan un coeficiente de variación mayor, los datos son heterogéneos, esto significa que la desviación estándar es mayor que la media. (Contento Rubio, 2019, pp. 61-63)

1.3.5. Análisis Exploratorio de Datos (EDA)

El análisis exploratorio de datos es un paso esencial para realizar análisis de investigación sobre los datos. El propósito principal del análisis exploratorio es proporcionar un resumen e interpretar los datos, detectar valores atípicos, anomalías de los datos, y validación de hipótesis. También proporciona metodologías para generar estimación al visualizar y comprender datos, a menudo a través de gráficos exploratorios.

Según Howard Seltman, “cualquier método de análisis de datos que no incluya modelos estadísticos descriptivo e inferencial nos es un análisis exploratorio de datos”. (Komorowski et al., 2016, p. 185)

El EDA es un instrumento estadístico que examina las propiedades y patrones presentes en el conjunto de los datos. Los gráficos de visualización de los datos como, gráfico de barra, diagramas de caja y bigote, histogramas, gráfico de dispersión y gráfico cuantil-cuantil. Además, el EDA es primordial en el desarrollo de series de tiempo, y modelos de pronóstico, debido a que muestran y explora la interacción de variables de entrada y salida. (Javed, Umar et al., 2021)

1.3.6. Análisis de datos atípicos

Los valores atípicos son puntos de datos que difieren del resto de los datos, es decir son las anomalías, discordancias, desviaciones atípicas. En una observación se pueden determinar no solo errores sino también datos inconsistentes que pueden resultar de la variación natural en una población o medición. También, suelen contener información valiosa y útil sobre la investigación de los datos. (Salgado et al., 2016, pp. 163-164)

En el campo industrial, las principales fuentes de valores atípicos son fallas del equipo, errores humanos, anomalías causadas por el comportamiento específico de las vibraciones naturales de la maquinaria, transportación de fluido, entre otras. Los datos atípicos son muy valiosos, pues ayudan a contrastar conclusiones congruentes e información útil, misma que servirá para la predicción y toma de decisiones. (Salgado et al., 2016, p. 164)

Las consecuencias de no filtrar la toma de datos, conlleva a la aparición de valores atípicos que pueden minimizar e incluso desaparecer errores como:

- Aumento de la varianza del error y depreciación de los valores descriptivos.
- Normalidad mínima donde los valores atípicos no se distribuyen aleatoriamente.

1.3.7. Prueba de normalidad

La aplicación de las pruebas de normalidad pretende asegurar la solidez de los estudios estadísticos, más aún cuando en las empresas se dedica tiempo y recursos para eso, razón por la cual es deseable llegar a conclusiones necesarias. En este sentido, resulta clave revisar que, una vez que se aplica una cierta herramienta estadística al estudio de cambiantes sucesivas o cuantitativas, la información obtenida a lo largo del proceso se distribuya de forma óptima; ya que, todos los análisis paramétricos necesitan el cumplimiento de la normalidad. (Correa et al., 2006)

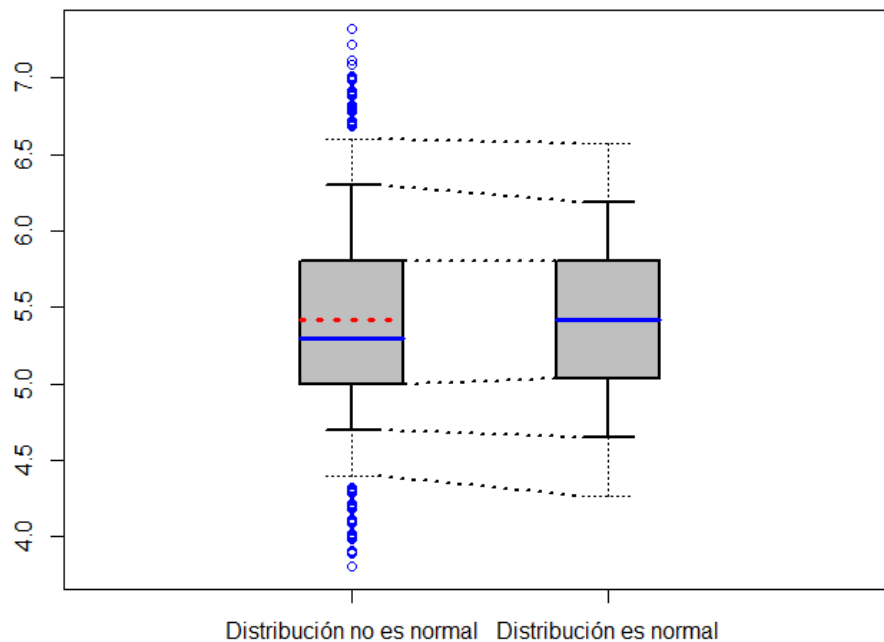


Figura 3-1: Distribución normal de los datos

Realizado por: Zambrano J, 2022.

1.3.7.1. Prueba Shapiro Wilk

Shapiro Wilk es una prueba de normalidad, para contrastar la normalidad cuando el tamaño de la muestra es inferior a 50 datos. Este método consiste en ordenar las muestras de menor a mayor valor, calculando la media y la varianza de la muestra. Se acepta la hipótesis nula si el estadístico es mayor al nivel de significancia de 0,05 y se rechaza la hipótesis nula si el estadístico es menor al nivel de significancia de 0,05. (Flores, Tapia et al., 2021)

Shapiro-Wilk, ha sido introducido tomando en cuenta que el gráfico de posibilidad usual que revisa el ajuste de un grupo de datos de muestra para el reparto usual es parecido a la de regresión lineal. La línea diagonal del gráfico es la recta de ajuste perfecto, con la diferencia de que esta línea es parecida a los residuos de la regresión.

Por medio de la exploración del tamaño de esta alteración estudio de varianza, la calidad del ajuste podría ser analizado. El estadístico matemático para la prueba se muestra en la ecuación 9-1.

$$W = \frac{\sum_{i=1}^n \alpha_i Y_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (9-1)$$

1.3.7.2. Prueba Kolmogorov-Smirnov

La prueba de Kolmogorov-Smirnov es una prueba de bondad de ajuste extensamente usada para probar la normalidad de los datos cuando el tamaño de la muestra es mayor a 50 datos. Siendo especialmente eficaz en procesos físicos no lineales e interactivos, por cuanto éstos conducen, principalmente, a distribuciones no gaussianas y, por consiguiente, el mecanismo generador de los procesos puede entenderse mejor al analizar el reparto de las cambiantes seleccionadas.

Además, para llevar a cabo la prueba de normalidad varias pruebas estadísticas necesitan estar bajo el supuesto de normalidad y, por consiguiente, constituye un prerequisite decidir si este supuesto cumple la normalidad de una muestra de una población. (Flores, Tapia et al., 2021)

La prueba de Kolmogorov-Smirnov se aplica para constatar la hipótesis si una distribución es normal o no, el estadístico matemático para la prueba se muestra en la ecuación 10-1.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n . \quad (10-1)$$

1.3.7.3. Estadístico t de Student

La prueba t de Student fue descubierta por William Sealy Gosset (1876-1937), como una distribución de probabilidad relacionada con la distribución normal. Este problema surge cuando se desea contrastar la comparación entre dos variables a partir de una distribución normal. El estadístico matemático para la prueba se muestra en la ecuación 11-1. (Contento Rubio, 2019, pp. 227-228)

$$t = \frac{Z}{\sqrt{\frac{x^2}{v}}} \quad (11-1)$$

Donde Z es la variable normal estándar, x^2 es una distribución ji cuadrado y v grados de libertad.

La prueba t de Student pueden ser de dos categorías de una o dos colas si la hipótesis alternativa utiliza (" $>$ " o " $<$ "). Es una prueba de una cola que analiza la desviación en una sola dirección. Si la hipótesis alternativa es del tipo "diferente de", se trata de una prueba de dos colas en la que se analizan los posibles sesgos en ambas direcciones. Las pruebas de una cola solo se utilizan cuando se determina que el sesgo de interés es en una dirección.

La prueba t-Student permite realizar la comparación en tres principales que son clave para la toma de decisiones y conclusiones congruentes. La prueba de comparación de dos variables del mismo tamaño muestral, y la prueba en función de la media para dos muestras permiten determinar si la media de una muestra sigue una distribución normal, es igual al valor (u_0). Al probar la diferencia de media de dos muestras, se encuentra si son iguales, es decir si no hay diferencias significativas entre las medias. (Lugo et al., 2022)

1.3.7.4. Hipótesis estadística

La prueba de hipótesis se basa en la suposición de que existe una realidad desconocida y que el investigador desea probar una hipótesis que está debidamente respaldada por alguna teoría. La hipótesis que contiene el investigador es la hipótesis alternativa (H_1 o H_a) el rechazo de la hipótesis alternativa se denomina hipótesis nula (H_0). Suponiendo que el p-valor, para un nivel de confianza del 95%, como resultado sea mayor se acepta la hipótesis nula y sea menor se rechaza la hipótesis nula, concluyendo la toma de decisión y predicciones del conjunto de datos. (Contento Rubio, 2019, pp. 303-304)

Existen pruebas de hipótesis para contrastar la normalidad del conjunto de datos si cumplen con una distribución normal; la más utilizada es la prueba Kolmogorov-Smirnov para muestras grandes, y para una muestra pequeña la prueba de Kolmogorov-Smirnov pierde efectividad, y es idóneo usar la prueba de Shapiro-Wilk. (Sheldom M. Ross, 2018, pp. 385-387)

1.3.8. Teorema del límite central

El Teorema del Límite Central (CLT) establece que, para variable aleatoria de una observación, los datos a medida que aumenta el tamaño de la media se aproximan a una distribución normal. El CLT simplifica los problemas estadísticos porque es válido incluso con la distribución no es normal. La aplicación del CLT se utiliza generalmente para una precisión para muestras menores de 30 datos, exclusivamente para la estadística inferencial tales como es el método de Bootstrap que es para muestra mayores con resultados confiables y eficaz. (Velasco et al., 2021)

1.3.8.1. Métodos Bootstrap

El método de Bootstrap es un proceso de remuestreo mediante la sustitución de la muestra original para estimar la distribución de muestreo de una estadística. Muestreo con reemplazo significa que, después de tomar una observación de la muestra, la reemplaza de tal manera que la probabilidad de elegir la observación sigue siendo la misma en cada remuestreo. Bootstrap es una poderosa herramienta para evaluar variabilidad estadística de la muestra. (Velasco et al., 2021)

El método Bootstrap por Bradley Efron 1979 y Efron y Tibshirani 1993, permite realizar el remuestreo aleatorio de una muestra, la estimación de la varianza de varias muestras, y métodos de inferencia de cálculo estadísticos. Especialmente en la simulación de Monte Carlo, parte de la metodología en realizar un remuestreo de la muestra original y generar nuevas muestras del mismo tamaño, con el objetivo de estimar una distribución normal. (Navarro, 2016, p. 13)

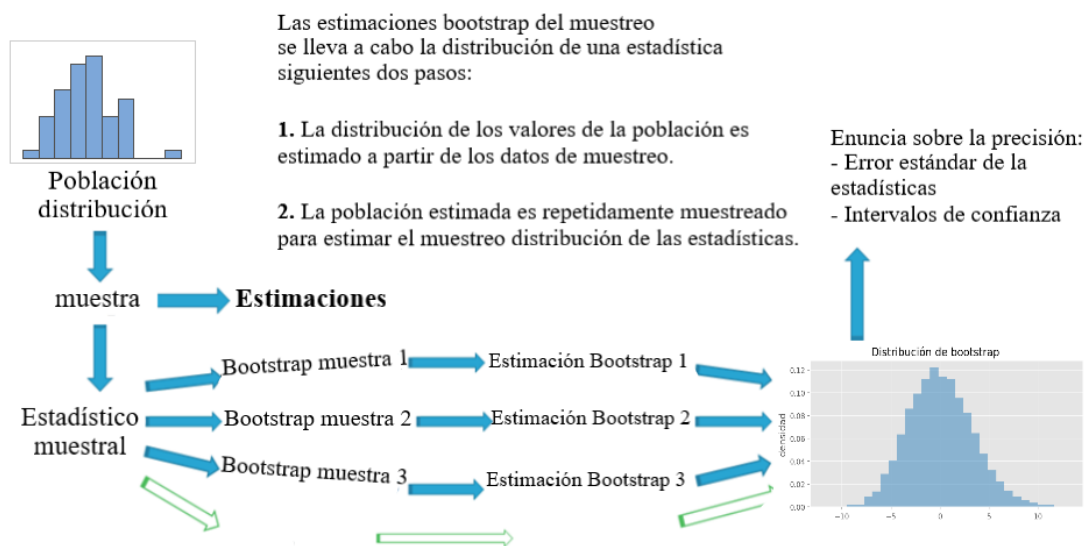


Figura 4-1: Diagrama de proceso de Bootstrap

Fuente: (Navarro, 2016, p. 16)

Realizado por: Zambrano J, 2022.

1.4. Ciencia de datos

En la ciencia de datos el propósito principal es adquirir conocimiento del análisis de datos para la toma de decisiones y predicciones de la investigación. Al analizar los datos permiten sustraer patrones inusuales de las bases de datos para explicar problemas o formular hipótesis. Los patrones identificados en el análisis de datos ayudan a conseguir información valiosa que un investigador o científico, para obtener resultados y conclusiones futuras. (Saura José, 2021)

Las cuatro capacidades acceden en acción en cada plan que implica ciencia de datos, en más grande o menor medida según la fase de estudio. Hablando de fases, Hadley Wickham, uno de los referentes recientes en el campo, las grafica de esta forma: (Hadley Wickham et al., 2017)

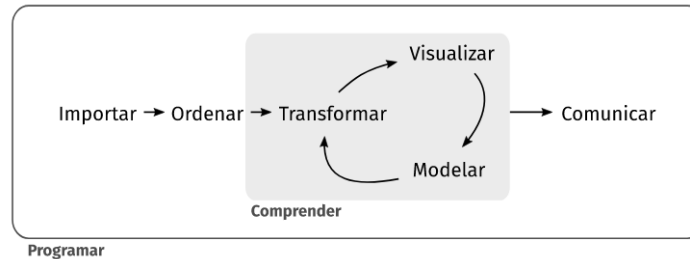


Figura 5-1: Etapas en la aplicación de ciencia de datos

Fuente: (Hadley Wickham et al., 2017)

1.5. Software R

Es un programa creado en 1993 por los profesores estadísticos Robert Gentleman y Ross Ihaka, desarrollado por “*The R Foundation for Statistical Computing*”. R es un lenguaje y entorno para computación estadística y gráficos, utilizado para el análisis de datos, cuyo hincapié de usarlo, modificar el software, con el objetivo de proporcionar un lenguaje de programación abierto y gratuito, antes una interfaz amigable y fiable. (Bosoni et al., 2019, p. 13)

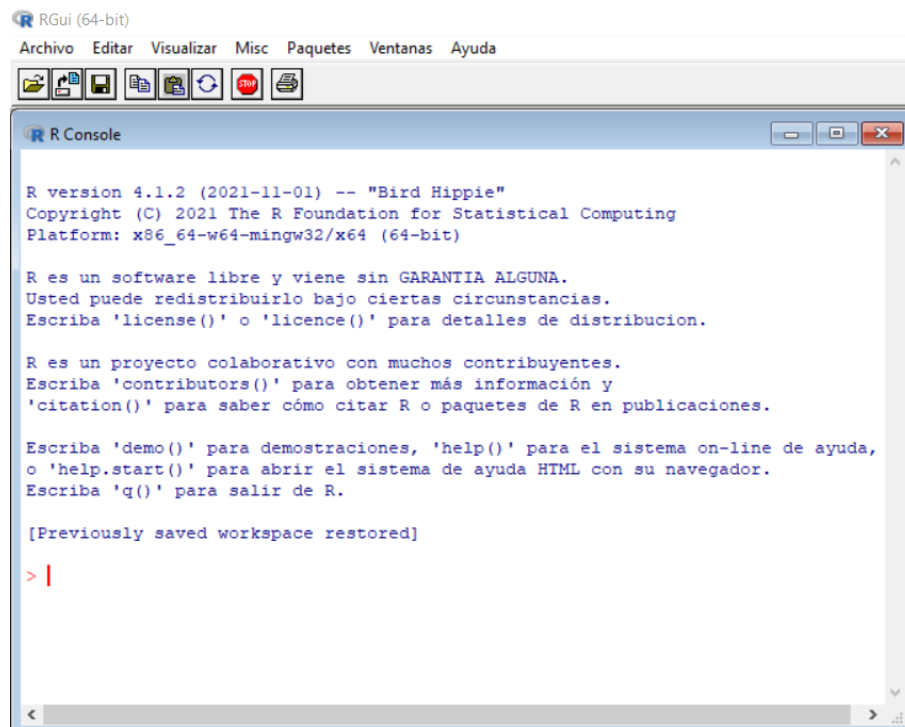


Figura 6-1: Interfaz del software R

Realizado por: Zambrano J, 2022.

1.5.1. Ventajas del uso de R

Replicabilidad: Componente importante en la indagación científica y cada vez más en las revistas académicas donde se exigen los archivos de sintaxis para la publicación de resultados.

Eficiencia: En condiciones reales de trabajo continuo, la utilización de sintaxis representa un aumento exponencial de la efectividad; por ejemplo, sirve para hacer un solo cálculo como calcular una media aritmética entre otros.

Control: Posibilita un control casi total en el trabajo de estudio, esto además posibilita identificar errores y potencia el trabajo colaborativo, debido a que el lenguaje que diferentes estudios emplearán para el análisis de datos. (Bosoni et al., 2019, p. 15)

1.6. RStudio

RStudio es una interfaz para usar R, que permite un Entorno de Desarrollo Integrado (IDE). El software RStudio es un lenguaje de programación utilizado para el análisis de datos, análisis estadístico y la manipulación de datos, a partir de la visualización gráfica y dinámica del comportamiento de los datos. En la Figura 7-1, se muestra cuatro ventanas de la interfaz gráfica de RStudio: (Estefanía Vargas et al., 2021, p. 11)

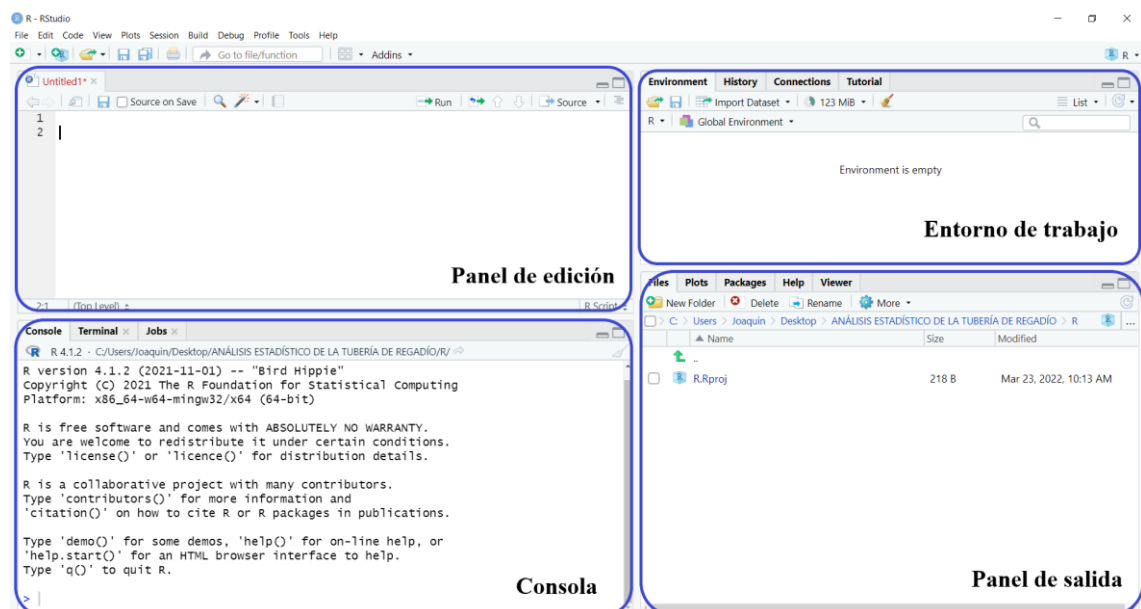


Figura 7-1: Interfaz del software RStudio

Realizado por: Zambrano J, 2022.

1.6.1. Paquetes en RStudio

Los paquetes de RStudio permiten instalar librerías y crear funciones para el análisis de datos y predicciones del conjunto de datos, a continuación, se representa el significado de los paquetes que se utilizaron para la investigación.

- **"Readxl"**: paquete que permite leer archivos de Excel e importar a RStudio.
- **"Tidyverse"**: paquete que permite el análisis de datos y visualización de gráficos. Este paquete incluye las librerías ggplot2, tibble, tidyr, readr, purrr, y dplyr; las cuales permiten realizar histogramas, diagramas de caja, y gráficos de dispersión.
- **"PASWR"**: paquete que permite realizar la probabilidad y la estadística, generando un histograma, un diagrama de densidad, un diagrama de caja y un diagrama Q-Q.
- **"Nortest"**: paquete que permite realizar la prueba de Kolmogorov-Smirnov para contrastar la normalidad de los datos.

1.6.2. Descripciones de gráficas de los datos

1.6.2.1. Histograma

Para una variable cuantitativa una buena elección para visualizar los datos es un histograma. La iniciativa de un histograma es bastante sencilla, son los datos de los cuales deseamos edificar el histograma consideramos el intervalo que va del mínimo al mayor, o sea, el intervalo. (Ayala, 2019, p. 23)

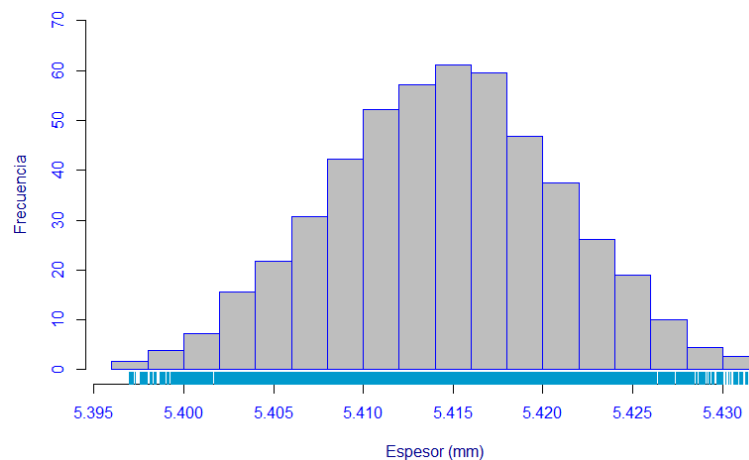


Figura 8-1: Histograma

Realizado por: Zambrano J, 2022.

1.6.2.2. Diagrama de caja

El EDA se debe a los aportes de John Tukey (1915-2000), fundamentados en lo cual él nombró "la vivencia de los inconvenientes reales y los datos reales". El diagrama de cajas (boxplot) es una herramienta para explorar un grupo de datos y aporta información para identificar la simetría, analizar la variabilidad e identificar valores atípicos u outliers.

El diagrama de caja de una variable de observación es una representación gráfica fundamentado en el primer y tercer cuartil, así como en sus valores máximos y mínimo. Este grafico proporciona una forma visual del conjunto de datos. Entre sus características son: (Ayala, 2019, pp. 23-24)

- Valores mínimo y máximo de la distribución
- Representa la mediana de los datos, los extremos del bigote limitan el intervalo que tiene por extremo inferior el primer cuartil (Q_1) del 25 % y por extremo superior el tercer cuartil (Q_3) del 75 %.

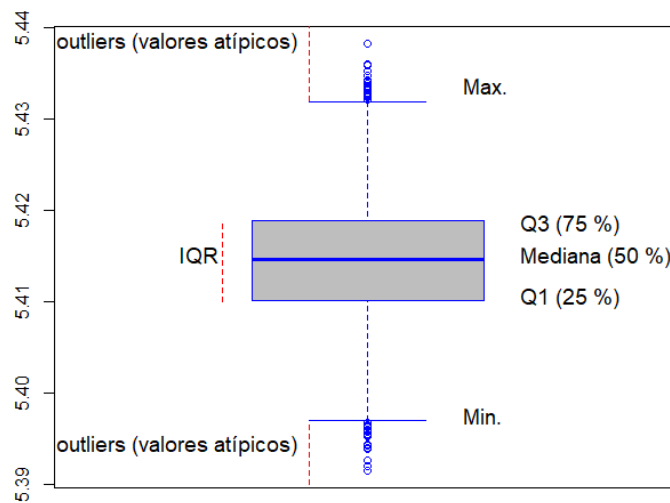


Figura 9-1: Diagrama de caja y bigote

Realizado por: Zambrano J, 2022.

La interpretación de cuartiles, percentiles y deciles es para determinar cuántos datos están por encima o por debajo de la mediana.

1.6.2.3. Estimadores Kernel de la densidad

Un gráfico alternativo y de mucha más calidad, es el estimador kernel de la densidad cuya finalidad es interpretar si los datos siguen una distribución normal o distribución *gaussiana*.

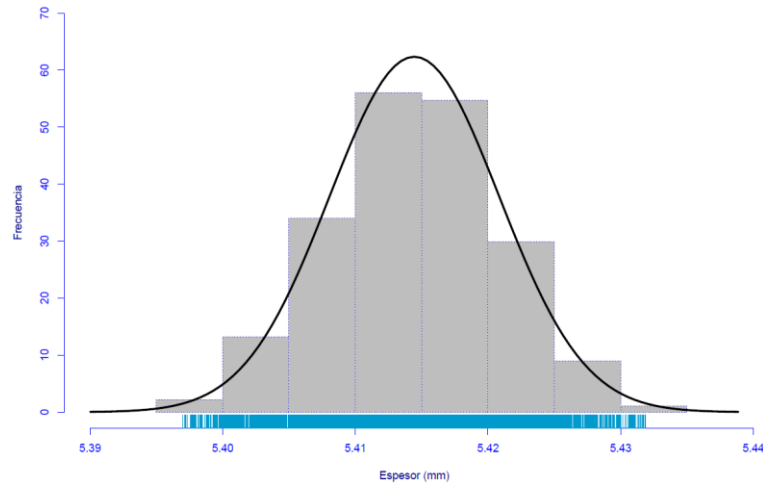


Figura 10-1: Gráfico de densidad

Realizado por: Zambrano J, 2022.

1.6.2.4. Gráficos cuantil-cuantil (Q-Q plots)

Son un instrumento gráfico para explorar si un grupo de datos o muestra procede de una población con cierta repartición, en especial aquí nos interesan para aprender la normalidad de un grupo de datos. La funcionalidad qqnorm permite identificar si se ajustan a la recta y la funcionalidad qqline añade una línea de referencia que ayuda el contraste de la normalidad. (Ayala, 2019, pp. 109-111)

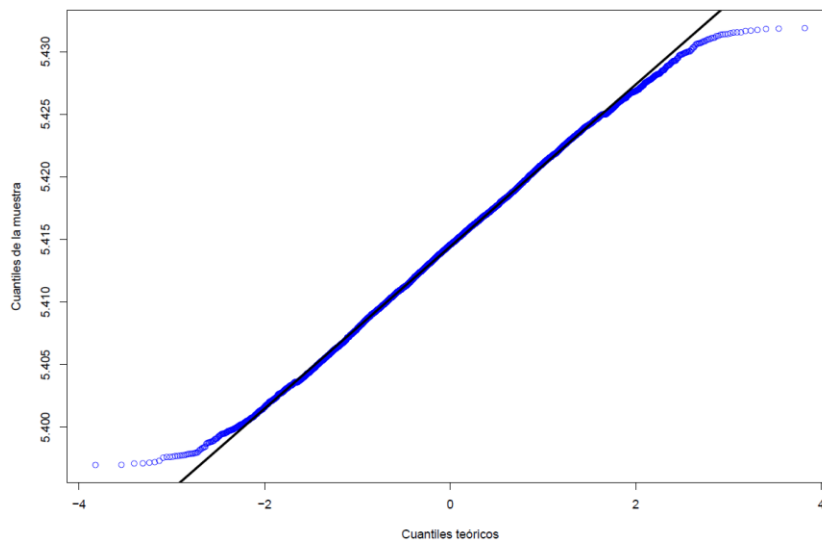


Figura 11-1: Gráficos cuantil-cuantil (QQplot)

Realizado por: Zambrano J, 2022.

CAPÍTULO II

2. MARCO METODOLÓGICO

En la presente investigación se realizó un estudio descriptivo de datos, que emplea el software RStudio. Además, un análisis estadístico de los factores que intervienen en el desgaste de la tubería de riego, por medio del Análisis Exploratorio de Datos (EDA), prueba de normalidad con la prueba de Kolmogorov-Smirnov, Método de Bootstrap y la prueba T de Student, aplicando la metodología de Box-Jenkins. (Marrugo Petro et al, 2020)

Las fases metodológicas propuestas son:

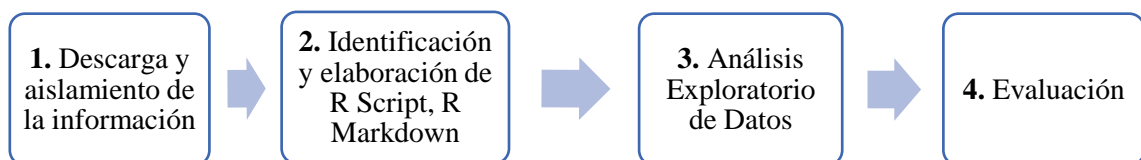


Figura 1-2: Fase metodológica

Fuente: (Marrugo Petro et al, 2020)

Realizado por: Zambrano J, 2022.

Fase 1: Descarga y aislamiento de la información

Esta etapa tiene como propósito:

- Descargar la información alfanumérica, obtenida por "Jorge Leonel Quinchuela Pomagualli", quien desarrolló la toma de datos en el año 2019, en la Comunidad San José la Silveria, parroquia San Andrés, provincia de Chimborazo.
- Estandarizar y depurar los datos desplegando la información alfanumérica y ordenar de mayor a menor de forma tal que, corresponda los datos en numérico para el análisis estadístico.

Fase 2: Identificación y elaboración de R Script, R Markdown

Aquí se identificaron los paquetes estadísticos y análisis de gráficos según la programación para cada tramo de la tubería, se realizó una hoja de "Script", es decir una serie de comandos y la instalación de librerías a integrar en el software RStudio, según las necesidades en el análisis estadístico de los datos.

Fase 3: Análisis Exploratorio de Datos

Esta etapa, se realizó:

- **Descripción:** Paso en el que se busca llamar la base de datos, llevado a cabo la depuración y estableciendo el tipo de variables.
- **Análisis Exploratorio:** Es la consulta de valores máximos y mínimos, la mediana, desviación estándar y finalización la visualización de los gráficos de histogramas de frecuencias, diagramas de caja y bigote, gráfico cuantil-cuantil, verificación de normalidad de los datos por medio de la prueba de Kolmogorov-Smirnov.
- **Prueba de normalidad:** Con la prueba Kolmogorov-Smirnov se contrastó en cada tramo los datos, si estos siguen una distribución normal, se aplica la prueba T-Student, y si no son normales se aplica el método de Bootstrap para normalizar los datos

Fase 4: Evaluación

Aquí se usó la prueba T-Student, para comparar las referencias (A-C, B-C, y D-C) y para contrastar el tipo de hipótesis, en donde si “p-valor” es mayor al 0,05, se considera hipótesis nula, por el contrario, si no cumple con esta precisión es una hipótesis alternativa.

2.1. Tipo y diseño de investigación

La investigación fue de tipo descriptivo y cuantitativo, cuyo propósito fue pronosticar los factores que intervienen en el desgaste de la tubería de regadío, que es la encargada de conducir el agua natural a los diferentes campos de cultivo de la Comunidad San José la Silveria desde el glaciar en el volcán Chimborazo, con una temperatura promedio de 11, 19° C, en dependencia de la época del año y una velocidad de desgaste promedio entre 0,0634 mm/año, con una longitud de 7548 m. El material de construcción de la tubería es acero al carbono.



Figura 2-2: Sistema de riego de la tubería de regadío

Realizado por: Zambrano J, 2022.

2.2. Población y muestra

Población: La presente investigación se llevó a cabo en la Comunidad San José la Silveria, que es parte de la parroquia San Andrés, provincia de Chimborazo, la tubería principal de regadío tiene una longitud de 7548 m y es de suma importancia, pues es la encargada de conducir el agua a los diferentes campos de cultivo de la Comunidad.

Muestra: Se obtuvo mediante datos secundarios, información que fue medida y obtenida por "Jorge Leonel Quinchuela Pomagualli" realizado en la Comunidad San José la Silveria, a través de ensayos no destructivos empleando ultrasonido, con un total de 15 fichas de inspección y un total de 7548 m de longitud medido.

2.3. Obtención y preparación de datos de la tubería de regadío

Los datos en el presente trabajo investigativo están conformados por un formato de inspección, el cual contiene los datos de cada tramo medido, estableciendo las mediciones a una distancia de 1m a lo largo de la tubería y a 90° en cada uno de los cuadrantes de la sección transversal tal como se especifica en norma API 570. Se lo hizo por tramos con una distancia de 500 m para cada uno de ellos dentro de los cuales se obtuvo 15 fichas de inspección. Además, la medición los Puntos Antes de la Soldadura (PAS) y Puntos Después de la Soldadura (PDS) sirvieron para realizar el posterior análisis.

METROS	(A)0	(B)90	(C)180	(D)270	OBSERVACIONES
0	5.3	5.1	5.3	5.3	
1	5.1	4.7	4.7	5	
2	5	5.2	4.9	5.1	
3	5.2	5.1	5	5.3	
4	5.3	5.2	5.1	5.2	
5	5.4	5.2	4.9	5.3	
6	4.8	5.1	4.8	5.2	
PAS	4.8	4.5	4.6	5	
PDS	5.2	4.7	4.3	4.9	
7	4.9	5	4.6	5	
8	5.2	5.2	5.1	5.3	
9	5.1	5.1	5	5.4	
10	5.4	5.1	4.9	5.1	

Figura 3-2: Bases de datos de la tubería de regadío

Realizado por: Zambrano J, 2022.

2.3.1. Descripción de la tubería de agua de regadío

Tabla 1-2: Conjunto de datos de la tubería de agua de regadío

Ítem	Distancia [m]	Diámetro [in]	Espesor nominal [mm]	Tiempo transcurrido [año]
Tramo 1	0-500	6	7,11	49
Tramo 2	501-1000	6	7,11	49
Tramo 3	1001-1500	6	7,11	49
Tramo 4	1501-2000	6	7,11	49
Tramo 5	2001-2500	6	7,11	49
Tramo 6	2501-3000	6	7,11	49
Tramo 7	3001-3500	6	7,11	49
Tramo 8	3501-4000	6	7,11	49
Tramo 9	4001-4500	6	7,11	49
Tramo 10	4501-5000	6	7,11	49
Tramo 11	5001-5500	6	7,11	49
Tramo 12	5501-6000	6	7,11	49
Tramo 13	6001-6500	6	7,11	49
Tramo 14	6501-7000	6	7,11	49
Tramo 15	7000-7548	6	7,11	49
PAS	0-7548	6	7,11	49
PDS	0-7548	6	7,11	49

Fuente: (Quinchuela, 2021)

Realizado por: Zambrano J, 2022.

2.3.2. Importar datos de Excel a RStudio

Para importar los datos se utilizó la “librería readxl”, con el comando “read_excel ()”, este código ayuda a importar archivos de Microsoft Excel a RStudio. También sirvió para exportar las hojas de cálculo desde Excel.

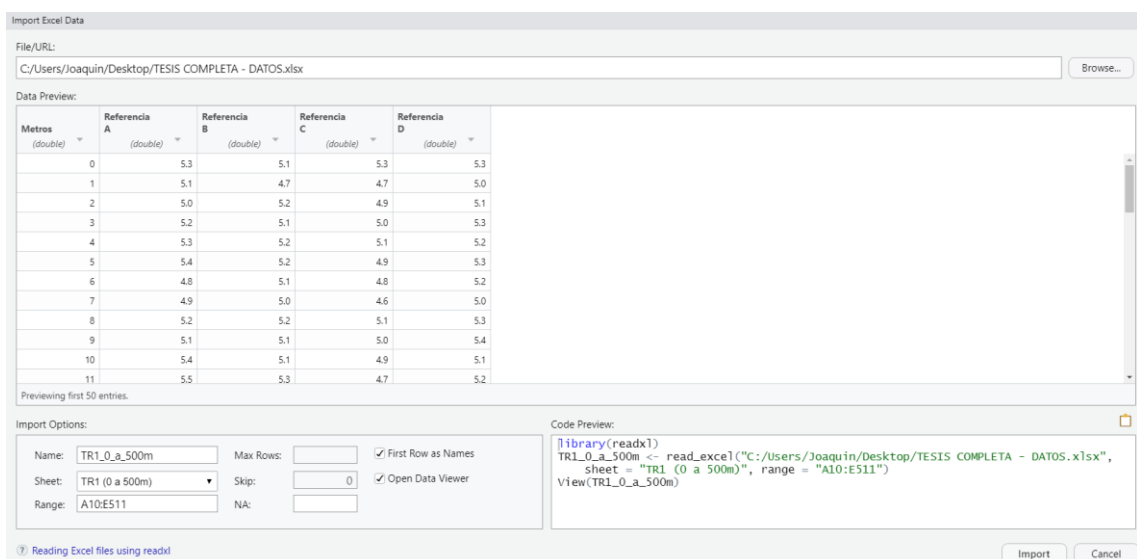


Figura 4-2: Importación de la base de datos de Excel a RStudio

Realizado por: Zambrano J, 2022.

2.3.3. Descripción de la base de datos de la tubería de regadío en RStudio

La base de datos consta de 15 fichas de inspección, cada tramo está dividido en cuatro referencias (A, B, C, D). Además, se tiene la medición de los PAS y PDS. Esto permitió realizar un análisis estadístico de los datos adecuado y realizar un análisis exploratorio de cada tramo de los datos medidos en la tubería de regadío, donde contamos con una interacción más fluida con el programa RStudio. Además, se ordenó de mayor a menor cada tramo de tubería en las que se realizaron las mediciones de los espesores.

Nombre	Observaciones	Variabes
PAS	652 obs.	4 variables
PDS	680 obs.	4 variables
TR1_0_a_500m	501 obs.	4 variables
TR10_4501m_a_5000m	500 obs.	4 variables
TR11_5001m_a_5500m	500 obs.	4 variables
TR12_5501m_a_6000m	500 obs.	4 variables
TR13_6001m_a_6500m	500 obs.	4 variables
TR14_6501m_a_7000m	500 obs.	4 variables
TR15_7001_a_7548m	548 obs.	4 variables
TR2_501m_a_1000m	499 obs.	4 variables
TR3_1001m_a_1500m	500 obs.	4 variables
TR4_1500m_a_2000m	500 obs.	4 variables
TR5_2001m_a_2500m	500 obs.	4 variables
TR6_2501m_a_3000m	500 obs.	4 variables
TR7_3001m_a_3500m	500 obs.	4 variables
TR8_3501m_a_4000m	500 obs.	4 variables
TR9_4001m_a_4500m	500 obs.	4 variables

Figura 5-2: Base de datos en RStudio de la tubería de regadío

Realizado por: Zambrano J, 2022.

2.4. Instalación de librerías para el análisis de datos

En esta etapa, se identificaron las necesidades de la información para realizar el script y se incorporaron para el análisis y pronóstico de los datos de la tubería de regadío.

Tabla 2-2: Librerías para la aplicación del análisis de datos

Nombre	Librería	Características
Readxl	library(readxl)	Importar de Excel a RStudio read_excel("datasets.xlsx")
Tidyverse	library(tidyverse)	Para análisis de datos y visualización de datos
Pruebas de Normalidad	library(nortest)	Prueba de Kolmogorov-Smirnov lillie.test(`Referencia A`)
Análisis Exploratorio de Datos	library(PASWR)	Probabilidad y Estadística EDA(`Referencia A`)

Realizado por: Zambrano J, 2022.

2.5. Análisis exploratorio de los datos

Con el programa “RStudio”, se realizó un análisis general, con una muestra de 7548 m de la tubería de regadío, mediante la interpretación descriptiva y la visualización de gráficas, las cuales, ayudaron a saber si las mediciones se distribuyen normalmente, en la Tabla 3-2 se presenta un resumen descriptivo de las cuatro referencias.

Tabla 3-2: Análisis descriptivo de la tubería de regadío

Análisis descriptivo	Referencia A	Referencia B	Referencia C	Referencia D
Media	5,414375	5,421317	5,35367	5,449894
Mediana	5,30	5,30	5,20	5,30
Desviación estándar	0,5700848	0,5760767	0,584297	0,5798354
Curtosis	-0,417353	-0,4700263	-0,5361127	-0,4463234
Asimetría	0,3603162	0,40478	0,3329418	0,3591662
Mínimo	3,80	3,70	3,60	3,80
Máximo	7,30	7,30	8,00	7,40
Rango	3,5	3,6	4,4	3,6
IQR	0,80	0,90	0,90	0,90

Realizado por: Zambrano J, 2022.

En la Figura 6-2, se exponen los resultados del análisis de gráficos a partir del histograma, el contraste de normalidad de los datos en las cuatro referencias de la tubería de regadío.

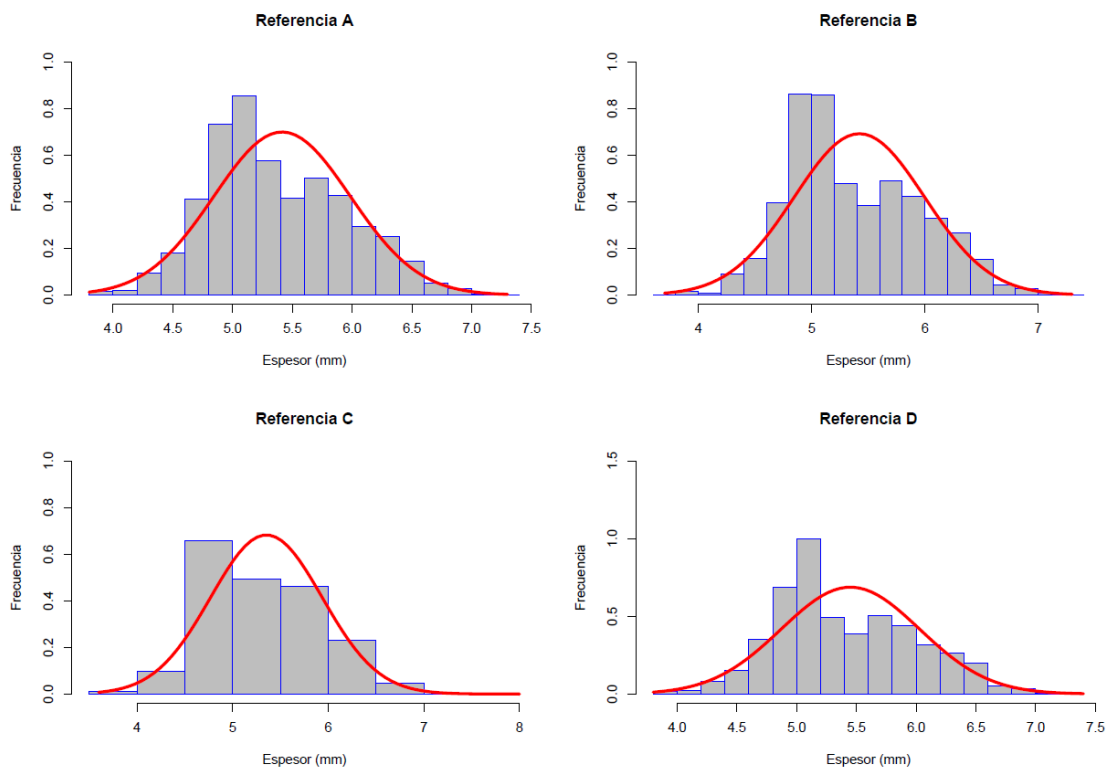


Figura 6-2: Histograma de los datos

Realizado por: Zambrano J, 2022.

2.5.1. Prueba de normalidad de los datos

Para contrastar la normalidad de los cuatro referencias se tomó en cuenta las siguientes consideraciones:

- Test de Shapiro-Wilk, para muestras ($n < 50$)
- Test Kolmogorov-Smirnov, para muestras ($n > 50$)

Se analizó en RStudio la normalidad de los datos y se utilizó la “librería nortest”, con el comando “`lillie.test ()`” para contrastar si cada referencia sigue una distribución normal o no, por medio de una prueba de hipótesis que se plantea de la siguiente forma:

- $H_0: \mu \geq 0,05$; se ajusta a una distribución normal
- $H_1: \mu < 0,05$; no se ajusta a una distribución normal

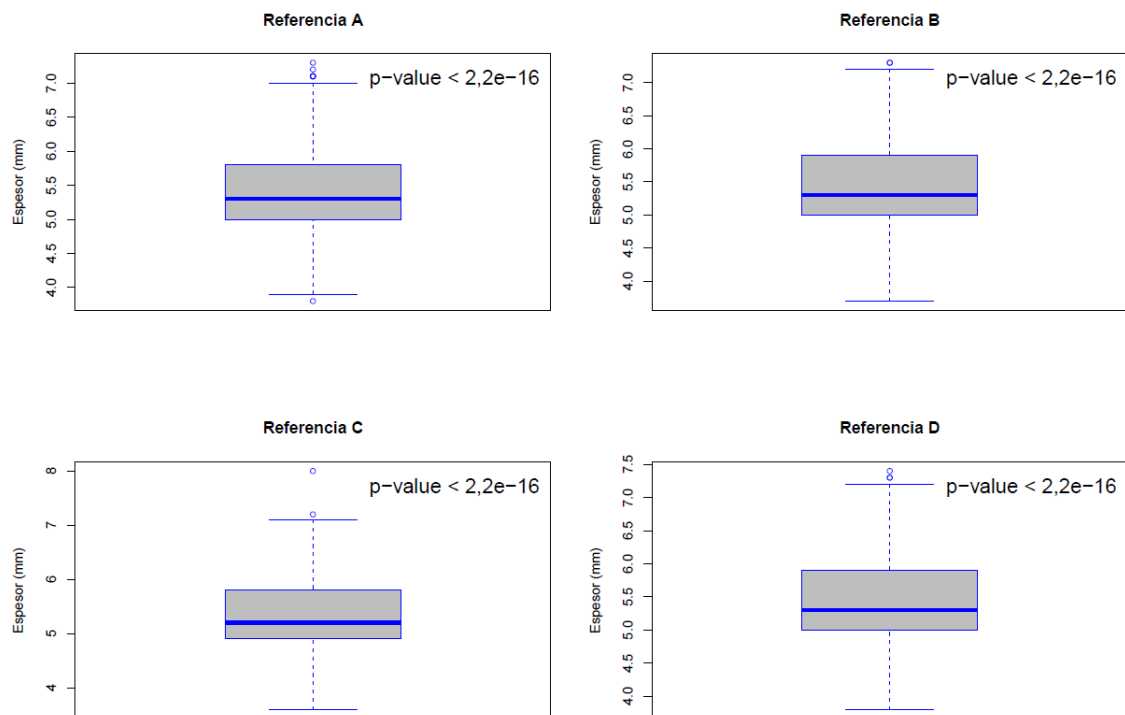


Figura 7-2: Diagrama de caja de normalidad de los datos

Realizado por: Zambrano J, 2022.

Como resultado del análisis, p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula, por lo que se concluye que los datos no se distribuyen normalmente.

2.5.2. Remuestreo de Bootstrap

Se realizó el método de Bootstrap, para contrastar en cada tramo que los datos sigan una distribución normal. Además, se utilizó dos comandos la función “replicate ()”, que permite crear un remuestreo de la muestra originaria, y la función “apply ()”, que permite manipular el remuestreo de los datos, en función del argumento de la media, obteniendo como resultado una muestra de las medias del mismo tamaño de la muestra originaria.

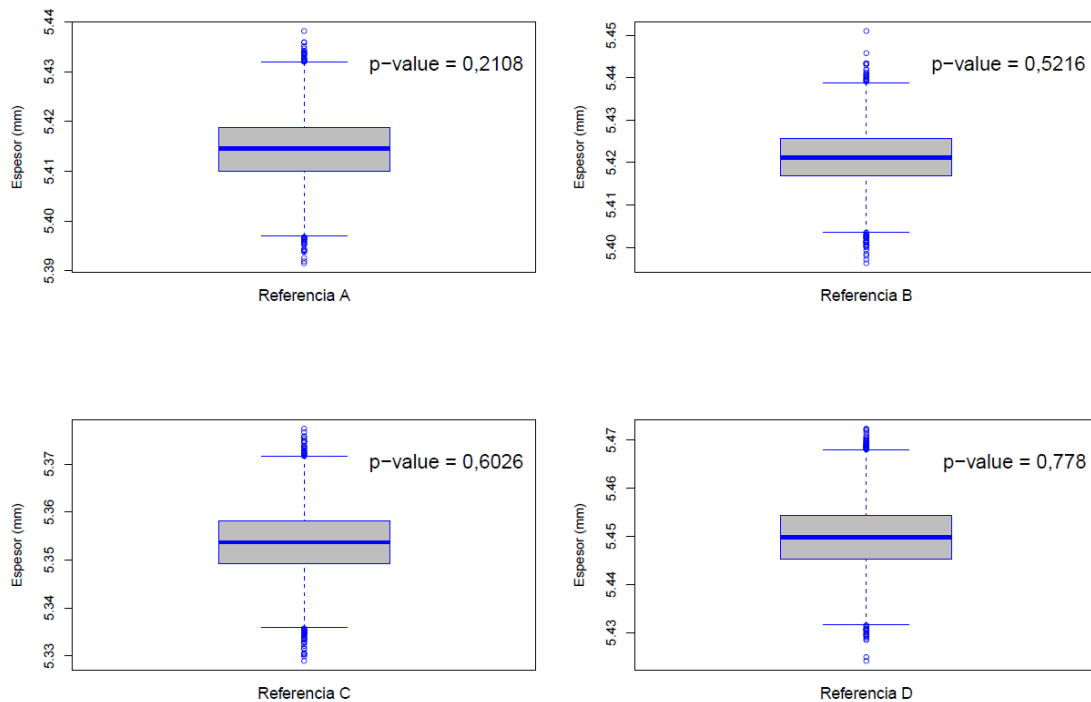


Figura 8-2: Diagrama de caja de la distribución de Bootstrap

Realizado por: Zambrano J, 2022.

En la Figura 8-2 se detalla la fiabilidad del método de Bootstrap, aplicado en las 4 referencias de la tubería de regadío, dando como resultado el p-valor $> 0,05$ por lo tanto, se acepta la hipótesis nula, concluyendo que los datos se distribuyen normalmente.

2.5.3. Reemplazo de los valores atípicos (Outliers)

Los diagramas de caja y bigotes están constituidos por los cuartiles y permiten observar los valores atípicos en cada tramo de la tubería, a través de los siguientes pasos:

- Visualización de valores atípicos, utilizando diagramas de caja y bigotes.
- Reemplazar los valores atípicos, con el objetivo de mantener la muestra originaria, en cada tramo de la tubería, generando resultados fiables para la toma de decisiones.

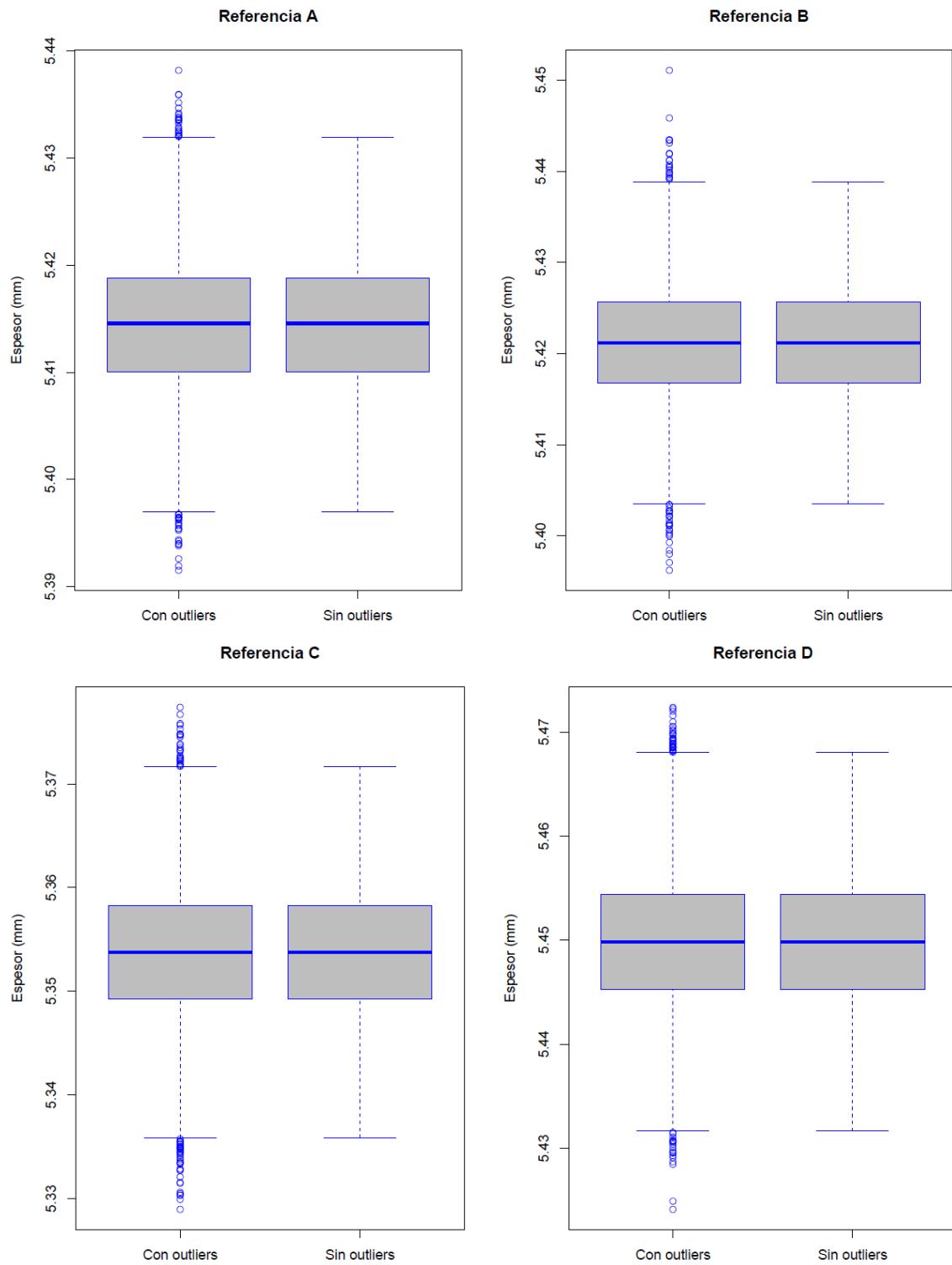


Figura 9-2: Diagrama de caja del reemplazo de los valores atípicos

Realizado por: Zambrano J, 2022.

En la Figura 9-2 se realizó el reemplazo de los valores atípicos con el primer cuartil para los valores inferiores del bigote minino y los del tercer cuartil para los valores superiores del bigote mayor, con un nivel de confianza del 95%, sin afectar la muestra originaria.

En la Tabla 4-2, se muestra los resultados del análisis estadístico en función de la mediana y la desviación estándar, ordenados de mayor a menor espesor, reflejando la referencia C con un desgaste mayor en la parte inferior de la tubería, y acotando que este desgaste se presenta en los 7548m de la tubería de regadío.

Tabla 4-2: Resultado del análisis estadístico de los datos

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,778	5,449	0,00652
Referencia B	p-valor > 0,5216	5,421	0,00648
Referencia A	p-valor > 0,2108	5,415	0,00639
Referencia C	p-valor > 0,6026	5,354	0,00652

Realizado por: Zambrano J, 2022.

En la Figura 10-2 se muestran los resultados obtenidos al aplicar el método de Bootstrap y el reemplazo de los valores atípicos. Observando así la normalidad de los datos en las cuatro referencias de la tubería de regadío.

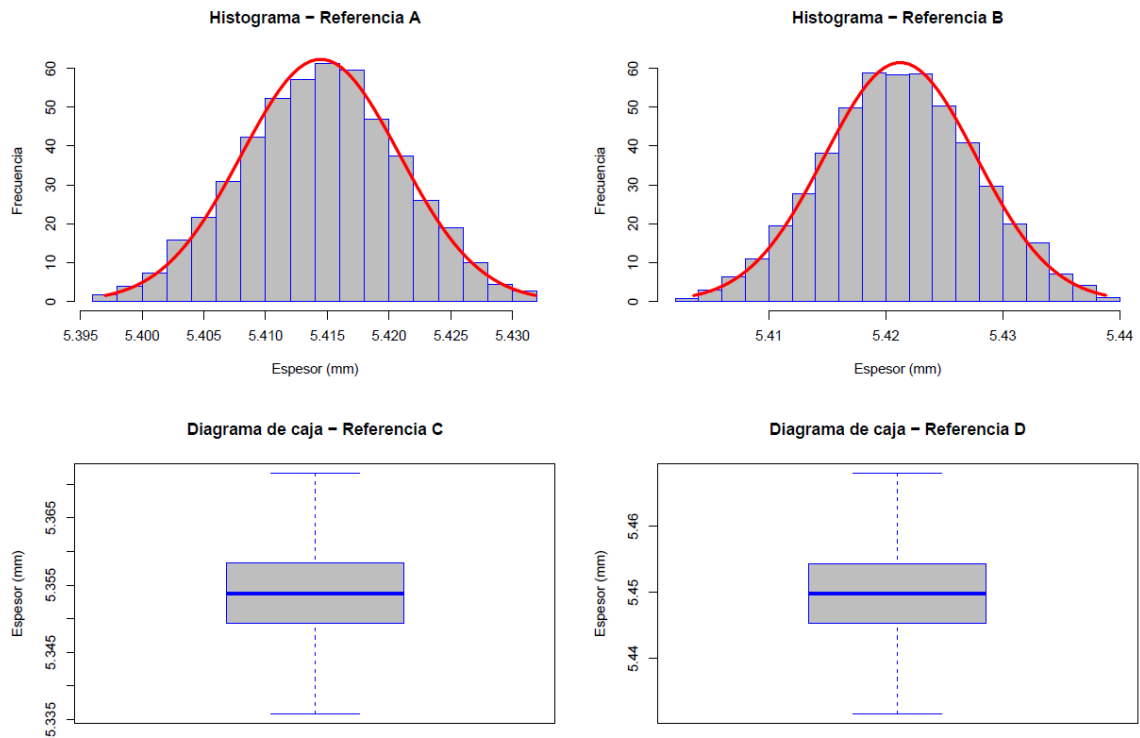


Figura 10-2: Gráficos de los datos normalizados

Realizado por: Zambrano J, 2022.

2.5.4. Prueba t de Student

Se realizó la prueba t de Student, para comparar entre la referencia (A-C, B-C, y D-C), en donde se utilizó el comando `t.test(x, y, paired = FALSE, alternative = c("two.sided", "less", "greater"))`, especificando una hipótesis alternativa “greater = mayor”, es decir que el punto A tiene una media mayor y que punto C tiene una media menor.

- **Referencia A – C**

$H_0: \mu_A \leq \mu_C$; los espesores del punto A son menores que los del punto C.

$H_1: \mu_A > \mu_C$; los espesores del punto A son mayores que los del punto C.

```
> t.test(Referencia_A, Referencia_C, paired = T, alternative = "greater")  
  
Paired t-test  
  
data: Referencia_A and Referencia_C  
t = 575.15, df = 7547, p-value < 2.2e-16
```

Figura 11-2: Prueba t-Student entre la referencia A y C

Realizado por: Zambrano J, 2022.

En la Figura 11-2, se realizó la comparación, dando como el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que el espesor de la referencia A son mayores que los de la referencia C

- **Referencia B – C**

$H_0: \mu_B \leq \mu_C$; los espesores del punto A son menores que los del punto C.

$H_1: \mu_B > \mu_C$; los espesores del punto A son mayores que los del punto C.

```
> t.test(Referencia_B, Referencia_C, paired = T, alternative = "greater")  
  
Paired t-test  
  
data: Referencia_B and Referencia_C  
t = 636.95, df = 7547, p-value < 2.2e-16
```

Figura 12-2: Prueba t-Student entre la referencia B y C

Realizado por: Zambrano J, 2022.

En la Figura 12-2, se realizó la comparación, dando como el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que el espesor de la referencia B son mayores que los de la referencia C

- **Referencia D - C**

$H_0: \mu_D \leq \mu_C$; los espesores del punto A son menores que los del punto C.

$H_1: \mu_D > \mu_C$; los espesores del punto A son mayores que los del punto C.

```
> t.test(Referencia_D, Referencia_C, paired = T, alternative = "greater")
```

Paired t-test

```
data: Referencia_D and Referencia_C
t = 914.92, df = 7547, p-value < 2.2e-16
```

Figura 13-2: Prueba t-Student entre la referencia D y C

Realizado por: Zambrano J, 2022.

En la Figura 13-2, se realizó la comparación, dando como el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que el espesor de la referencia B son mayores que los de la referencia C

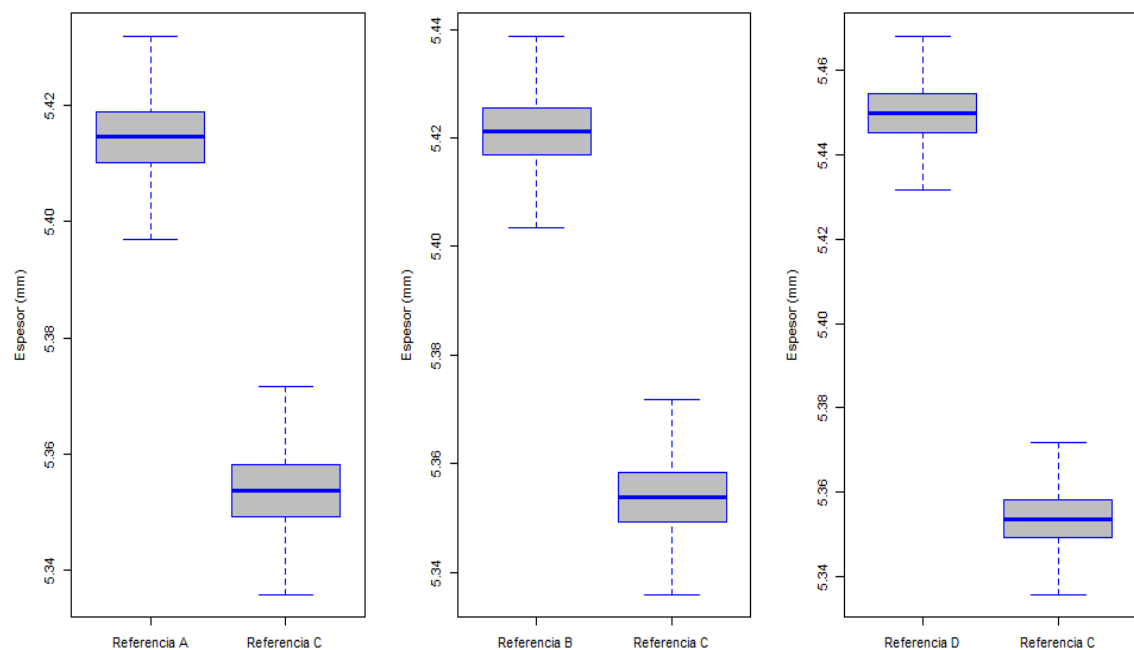


Figura 14-2: Diagrama de caja de la prueba t de Student

Realizado por: Zambrano J, 2022.

CAPÍTULO III

3. MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS

3.1. Análisis estadístico de cada tramo de la tubería de regadío

El análisis se basó en función de la mediana (ordenando de mayor a menor) y la desviación estándar. Asimismo, se aplicó la prueba de normalidad para contrastar la hipótesis de que el “p-valor” sea mayor al 0,05, cumpliendo con una distribución normal. Además, se desarrolló el análisis estadístico en el software RStudio por cada tramo de la tubería de regadío, con el cual se obtuvieron los siguientes resultados:

3.1.1. Tramo N° 1

Tabla 1-3: Resultado del análisis estadístico del tramo N° 1

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,3139	5,079	0,0078
Referencia A	p-valor > 0,9415	5,068	0,0094
Referencia B	p-valor > 0,221	5,015	0,0079
Referencia C	p-valor > 0,1659	4,942	0,0100

Realizado por: Zambrano J, 2022.

En la Tabla 1-3, se obtuvo que, el p-valor > 0,05, es decir que los datos se ajustan a una distribución normal, se acepta la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 2-3: Prueba t-Student para el tramo N° 1

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 212,39	df = 500	p-valor < 2,2e-16
Referencia B-C	t = 126,47	df = 500	p-valor < 2,2e-16
Referencia D-C	t = 236,65	df = 500	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 2-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

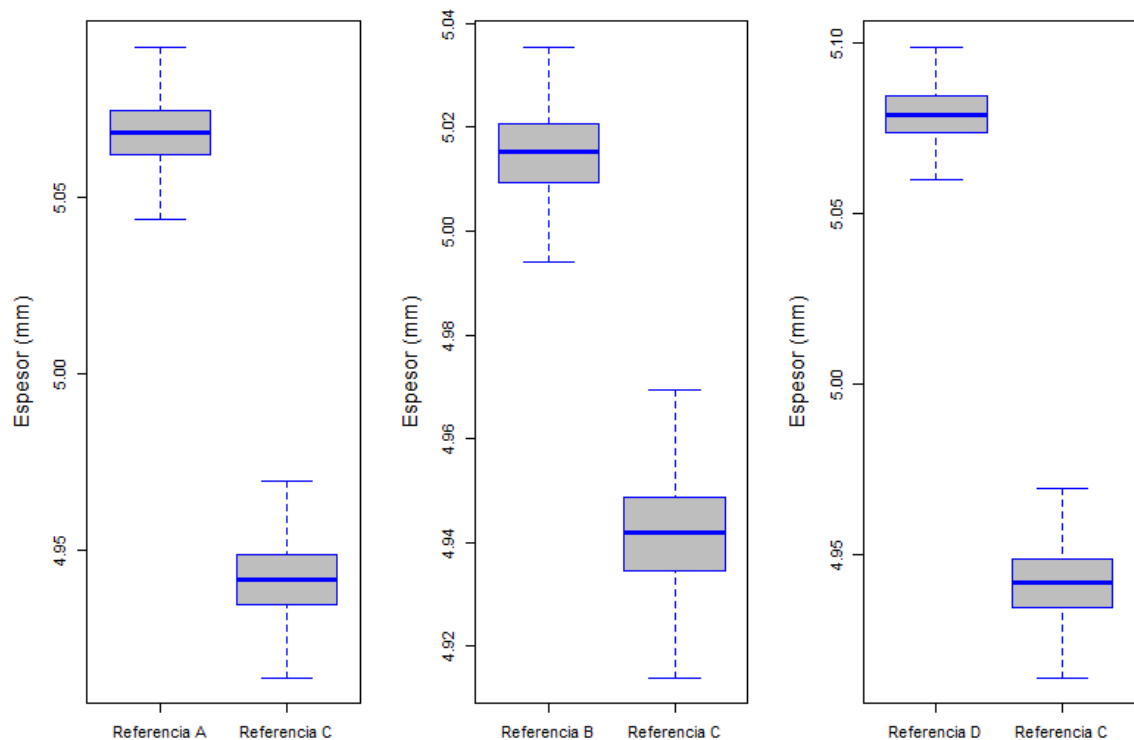


Figura 1-3: Diagrama de caja y bigote del tramo N° 1

Realizado por: Zambrano J, 2022.

3.1.2. Tramo N° 2

Tabla 3-3: Resultado del análisis estadístico del tramo N° 2

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia A	p-valor > 0,8065	5,148	0,0081
Referencia D	p-valor > 0,1837	5,131	0,0064
Referencia B	p-valor > 0,1631	5,076	0,0060
Referencia C	p-value > 0,1638	4,970	0,0066

Realizado por: Zambrano J, 2022.

En la Tabla 3-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 4-3: Prueba t-Student para el tramo N° 2

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 359,9	df = 498	p-valor < 2,2e-16
Referencia B-C	t = 254,18	df = 498	p-valor < 2,2e-16
Referencia D-C	t = 388,95	df = 498	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 4-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

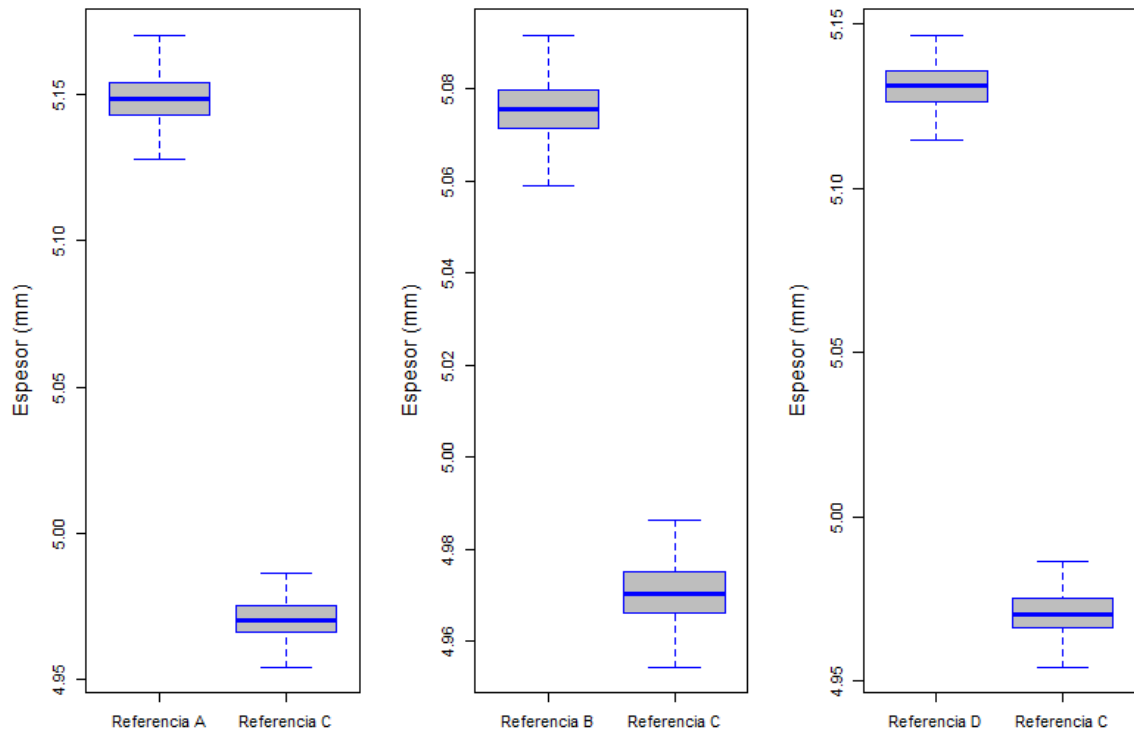


Figura 2-3: Diagrama de caja y bigote del tramo N° 2

Realizado por: Zambrano J, 2022.

3.1.3. Tramo N° 3

Tabla 5-3: Resultado del análisis estadístico del tramo N° 3

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,1652	5,079	0,0074
Referencia B	p-valor > 0,4068	4,991	0,0072
Referencia A	p-valor > 0,1277	5,024	0,0085
Referencia C	p-value > 0,5517	4,915	0,0079

Realizado por: Zambrano J, 2022.

En la Tabla 5-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 6-3: Prueba t-Student para el tramo N° 3

Punto de referencia	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 211,79	df = 499	p-valor < 2.2e-16
Referencia B-C	t = 154,27	df = 499	p-valor < 2.2e-16
Referencia D-C	t = 352,46	df = 499	p-valor < 2.2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 6-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

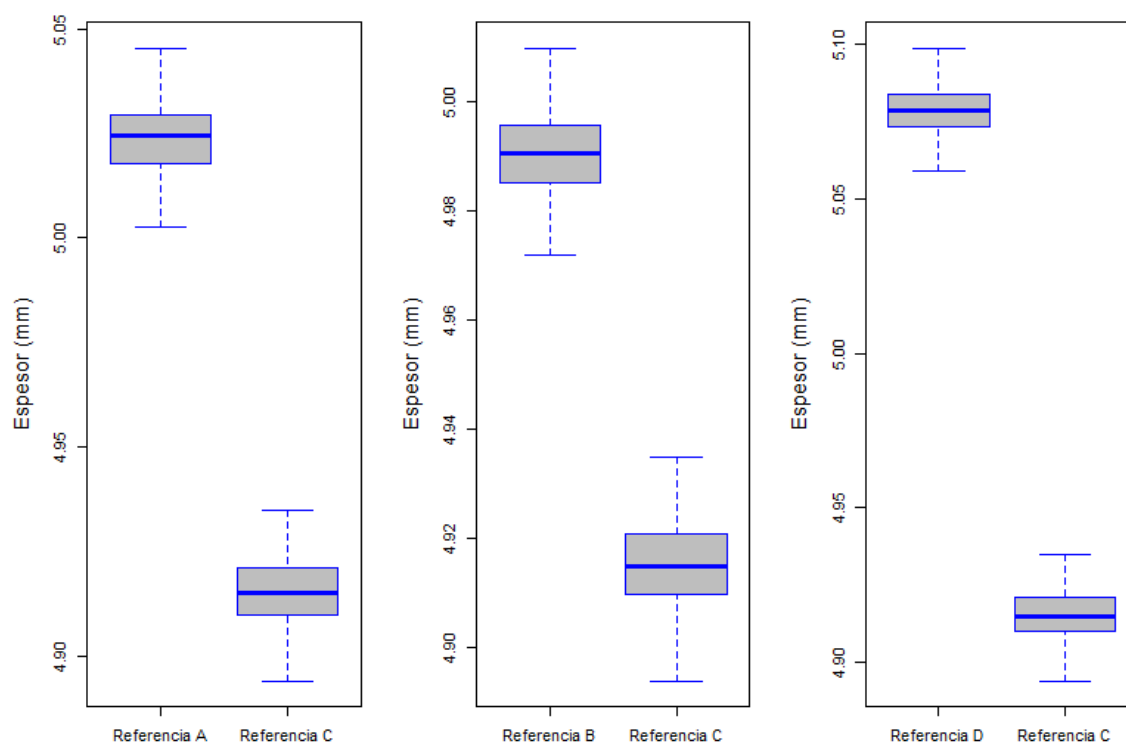


Figura 3-3: Diagrama de caja y bigote del tramo N° 3

Realizado por: Zambrano J, 2022.

3.1.4. Tramo N° 4

Tabla 7-3: Resultado del análisis estadístico del tramo N° 4

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,7413	5,065	0,0155
Referencia B	p-valor > 0,1878	5,060	0,0150
Referencia A	p-valor > 0,7811	5,048	0,0151
Referencia C	p-valor > 0,3732	4,999	0,0161

Realizado por: Zambrano J, 2022.

En la Tabla 7-3, se obtuvo que, el p -valor $> 0,05$, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 8-3: Prueba t-Student para el tramo N° 4

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	$t = 46,87$	$df = 499$	$p\text{-valor} < 2,2e-16$
Referencia B-C	$t = 60,695$	$df = 499$	$p\text{-valor} < 2,2e-16$
Referencia D-C	$t = 67,238$	$df = 499$	$p\text{-valor} < 2,2e-16$

Realizado por: Zambrano J, 2022.

En la Tabla 8-3, se obtuvo que, el p -valor es menor a $0,05$, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

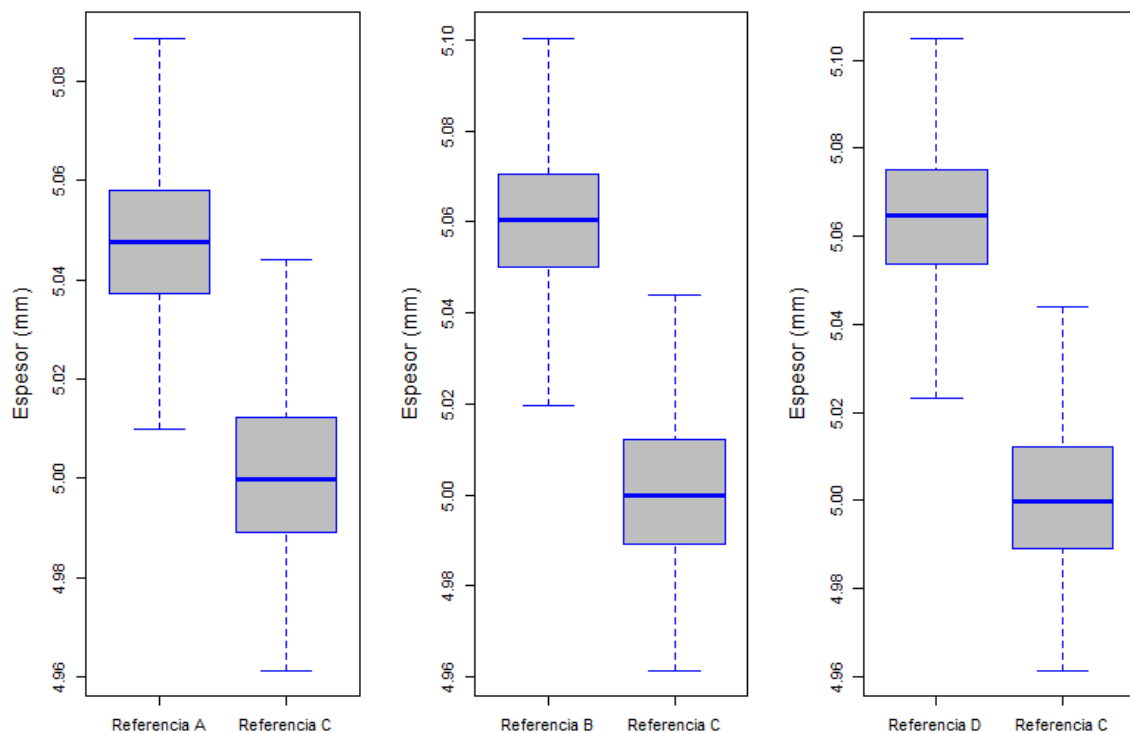


Figura 4-3: Diagrama de caja y bigote del tramo N° 4

Realizado por: Zambrano J, 2022.

3.1.5. Tramo N° 5

En la Tabla 9-3, se obtuvo que, el p -valor $> 0,05$, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 9-3: Resultado del análisis estadístico del tramo N° 5

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,837	5,066	0,0151
Referencia B	p-valor > 0,6076	5,045	0,0156
Referencia A	p-valor > 0,531	5,016	0,0151
Referencia C	p-valor > 0,4153	5,001	0,0177

Realizado por: Zambrano J, 2022.

En la Tabla 10-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

Tabla 10-3: Prueba t-Student para el tramo N° 5

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 14,104	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 42,158	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 63,994	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

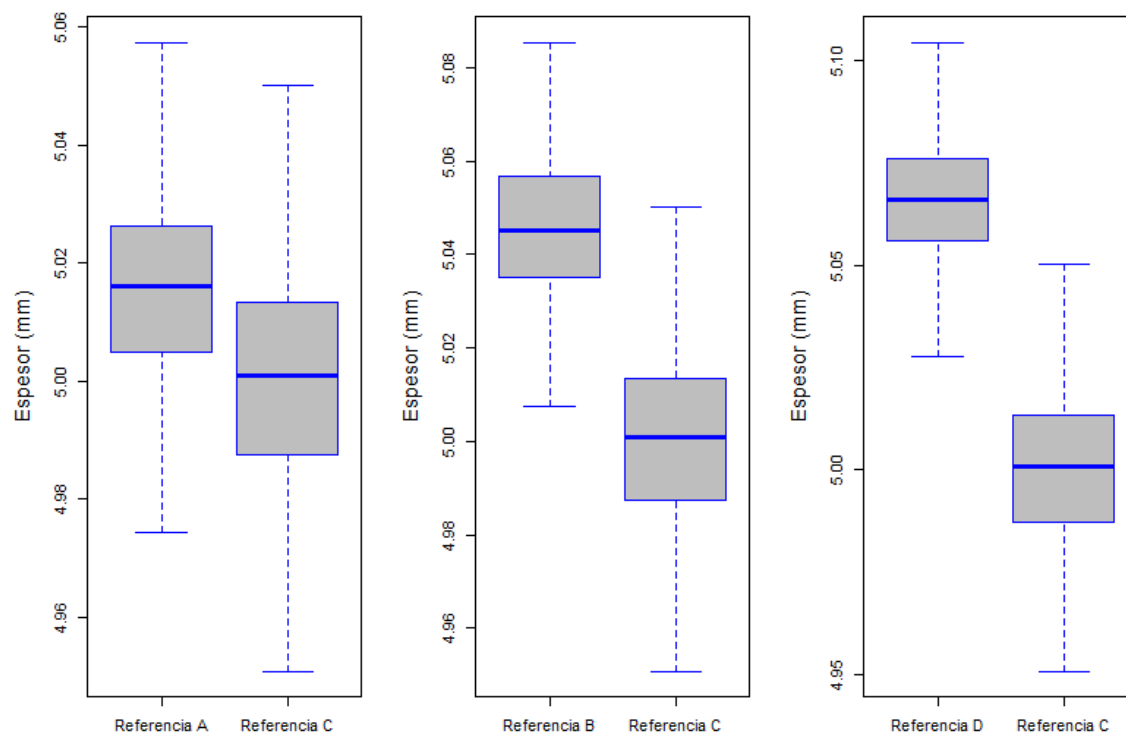


Figura 5-3: Diagrama de caja y bigote del tramo N° 5

Realizado por: Zambrano J, 2022.

3.1.6. Tramo N° 6

Tabla 11-3: Resultado del análisis estadístico del tramo N° 6

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia B	p-valor > 0,1859	4,972	0,0178
Referencia D	p-valor > 0,389	4,968	0,0157
Referencia A	p-valor > 0,6492	4,936	0,0171
Referencia C	p-valor > 0,2316	4,885	0,0180

Realizado por: Zambrano J, 2022.

En la Tabla 11-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 12-3: Prueba t-Student para el tramo N° 6

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 46,352	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 75,558	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 77,661	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 12-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

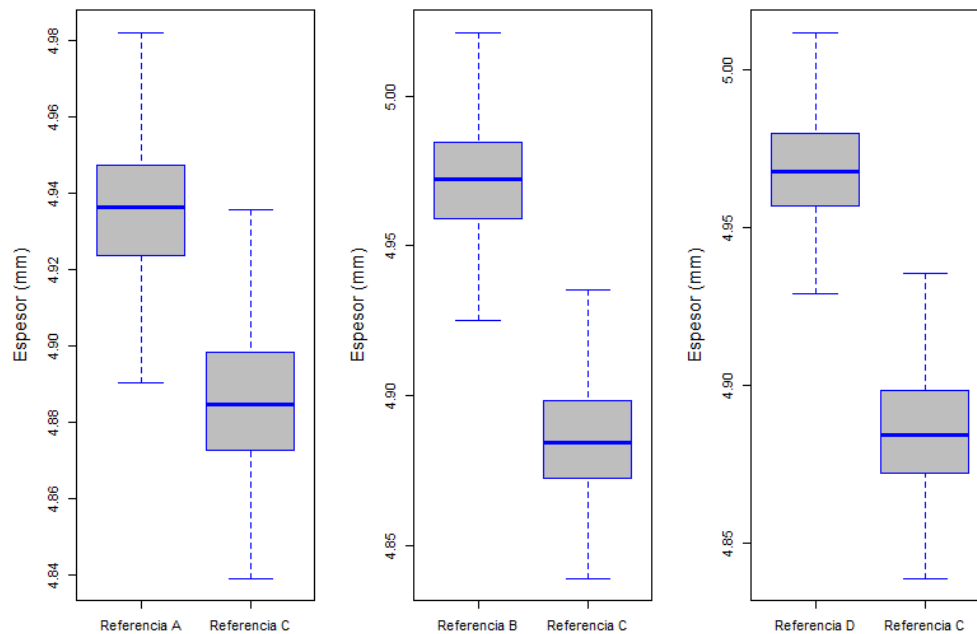


Figura 6-3: Diagrama de caja y bigote del tramo N° 6

Realizado por: Zambrano J, 2022.

3.1.7. Tramo N° 7

Tabla 13-3: Resultado del análisis estadístico del tramo N° 7

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,2912	5,020	0,0149
Referencia B	p-valor > 0,8082	5,006	0,0141
Referencia A	p-valor > 0,3953	4,965	0,0156
Referencia C	p-valor > 0,2045	4,932	0,0148

Realizado por: Zambrano J, 2022.

En la Tabla 13-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 14-3: Prueba t-Student para el tramo N° 7

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 35,43	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 83,401	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 99,638	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 14-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

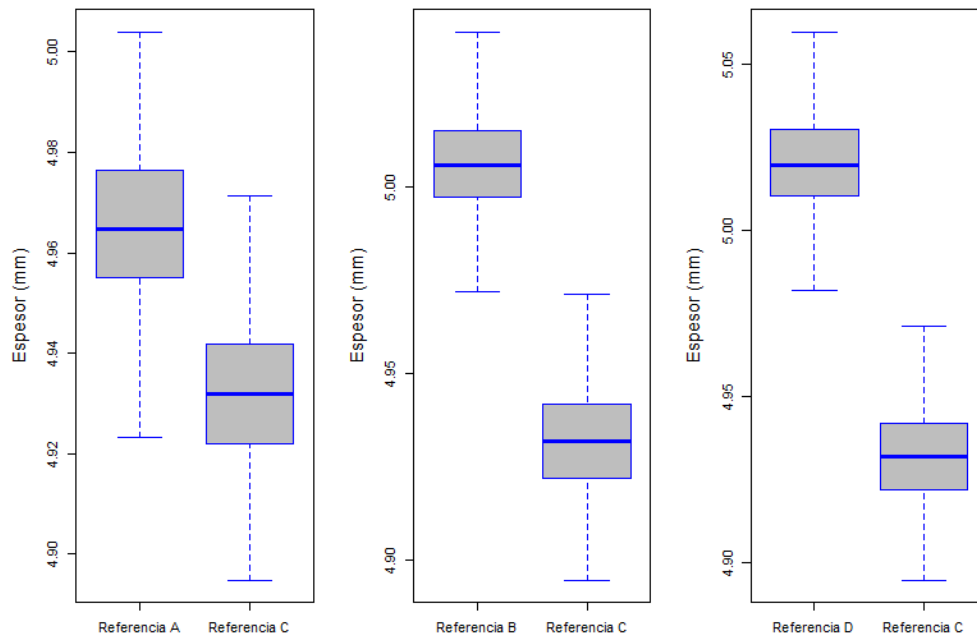


Figura 7-3: Diagrama de caja y bigote del tramo N° 7

Realizado por: Zambrano J, 2022.

3.1.8. Tramo N° 8

Tabla 15-3: Resultado del análisis estadístico del tramo N° 8

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia B	p-valor > 0,6166	5,004	0,0149
Referencia A	p-valor > 0,3112	4,990	0,0145
Referencia D	p-valor > 0,7243	4,985	0,0154
Referencia C	p-valor > 0,0810	4,935	0,0163

Realizado por: Zambrano J, 2022.

En la Tabla 15-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 16-3: Prueba t-Student para el tramo N° 8

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 55,684	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 69,768	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 48,205	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 16-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

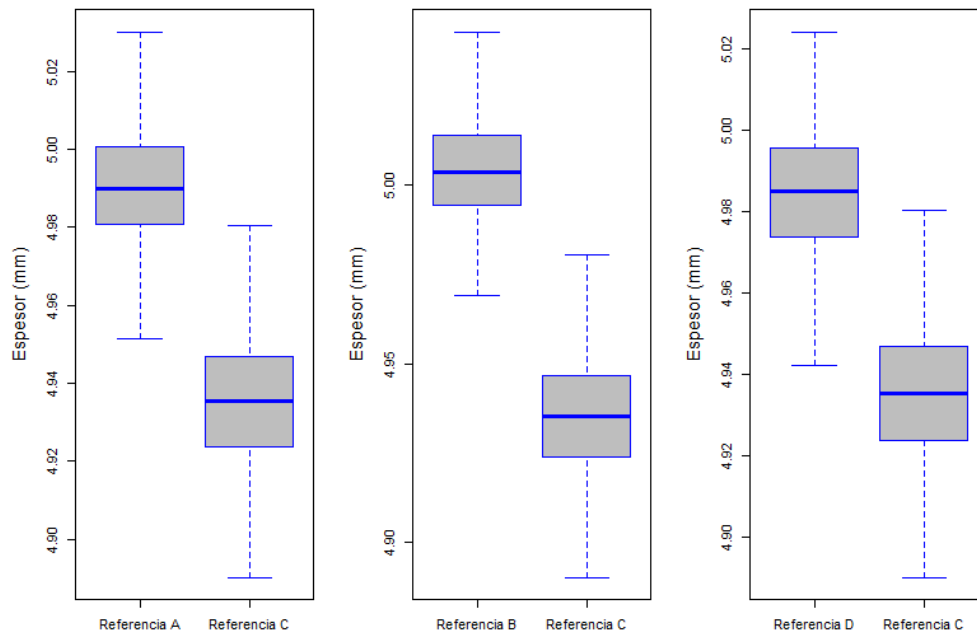


Figura 8-3: Diagrama de caja y bigote del tramo N° 8

Realizado por: Zambrano J, 2022.

3.1.9. Tramo N° 9

Tabla 17-3: Resultado del análisis estadístico del tramo N° 9

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia B	p-value > 0,16	5,505	0,0242
Referencia D	p-value > 0,5447	5,495	0,0265
Referencia A	p-value > 0,515	5,475	0,0246
Referencia C	p-value > 0,7208	5,402	0,0253

Realizado por: Zambrano J, 2022.

En la Tabla 17-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 18-3: Prueba t-Student para el tramo N° 9

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 46.912	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 65.661	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 56.429	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 18-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

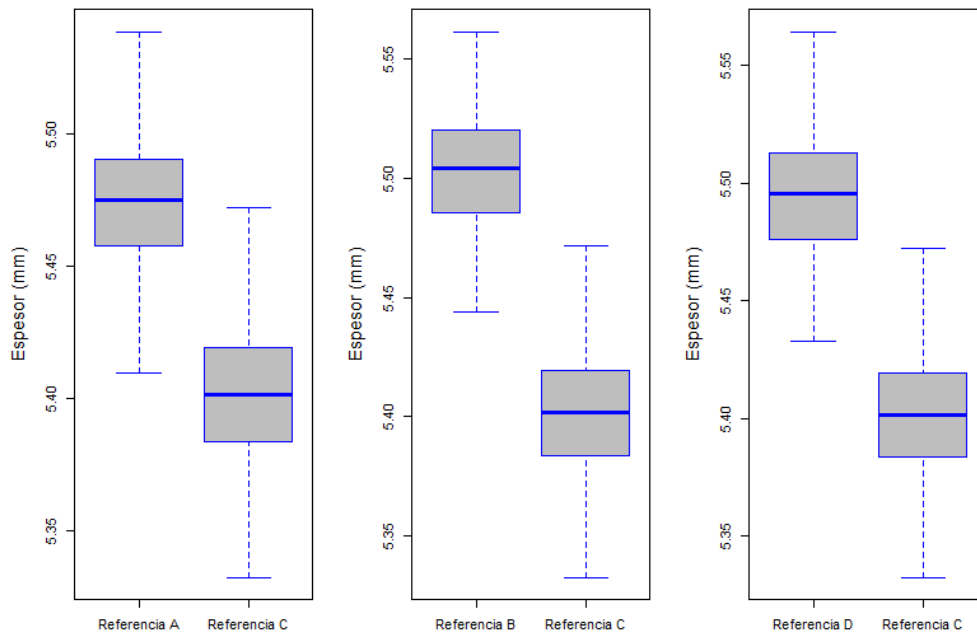


Figura 9-3: Diagrama de caja y bigote del tramo N° 9

Realizado por: Zambrano J, 2022.

3.1.10. Tramo N° 10

Tabla 19-3: Resultado del análisis estadístico del tramo N° 10

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,4365	5,896	0,0182
Referencia B	p-valor > 0,2293	5,872	0,0189
Referencia A	p-valor > 0,2708	5,852	0,0181
Referencia C	p-valor > 0,8634	5,789	0,0190

Realizado por: Zambrano J, 2022.

En la Tabla 19-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 20-3: Prueba t-Student para el tramo N° 10

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 55.383	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 69.046	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 93.455	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 20-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

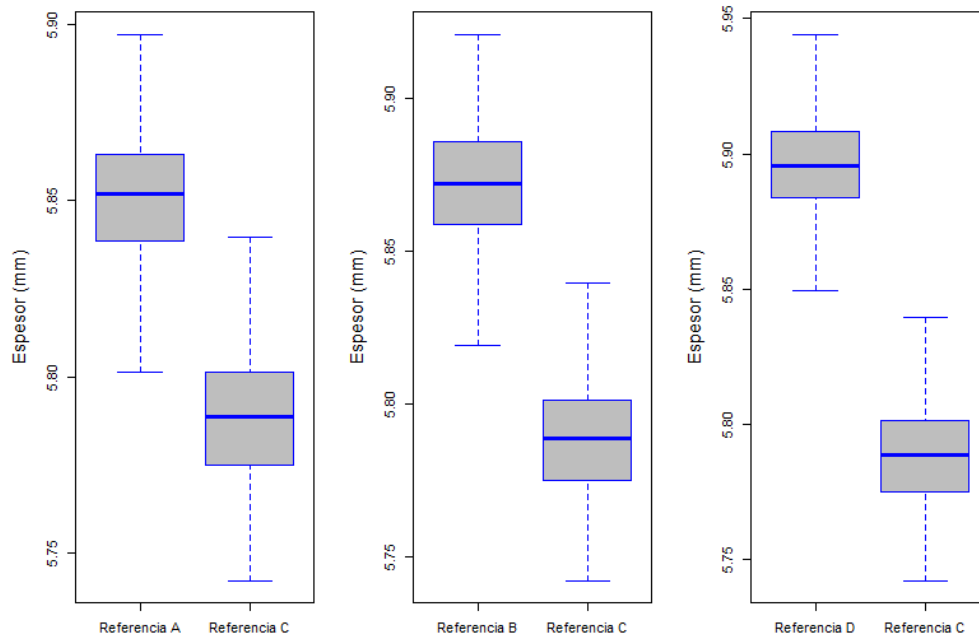


Figura 10-3: Diagrama de caja y bigote del tramo N° 10

Realizado por: Zambrano J, 2022.

3.1.11. Tramo N° 11

Tabla 21-3: Resultado del análisis estadístico del tramo N° 11

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,6648	5,936	0,0189
Referencia B	p-valor > 0,0859	5,902	0,0186
Referencia A	p-valor > 0,2625	5,901	0,0202
Referencia C	p-valor > 0,2223	5,787	0,0183

Realizado por: Zambrano J, 2022.

En la Tabla 21-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 22-3: Prueba t-Student para el tramo N° 11

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 89.474	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 97.866	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 122.19	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 22-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

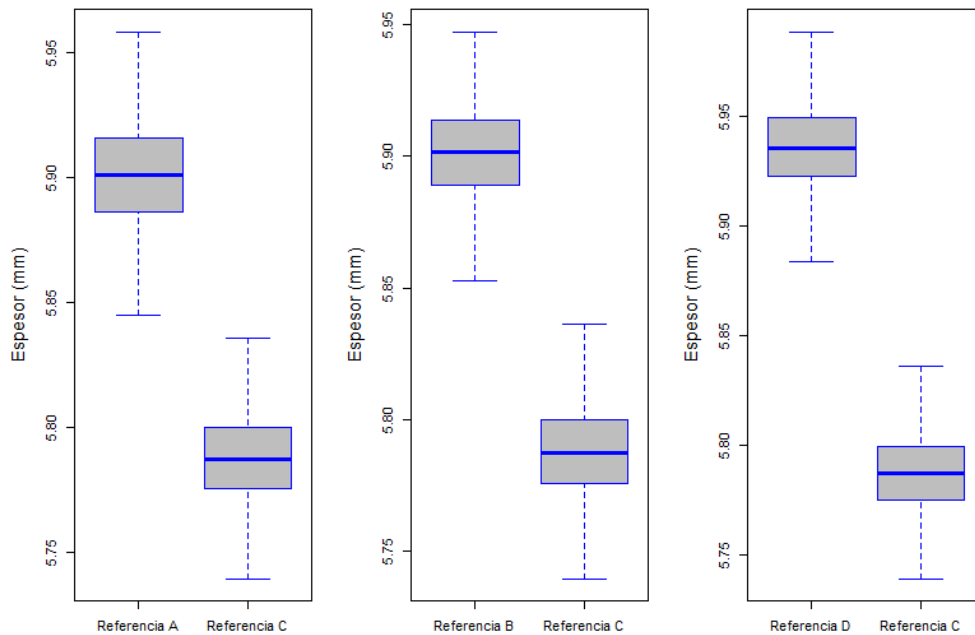


Figura 11-3: Diagrama de caja y bigote del tramo N° 11

Realizado por: Zambrano J, 2022.

3.1.12. Tramo N° 12

Tabla 23-3: Resultado del análisis estadístico del tramo N° 12

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,8467	6,012	0,0176
Referencia B	p-valor > 0,6208	5,956	0,0171
Referencia A	p-valor > 0,2978	5,939	0,0179
Referencia C	p-valor > 0,3635	5,914	0,0173

Realizado por: Zambrano J, 2022.

En la Tabla 23-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 24-3: Prueba t-Student para el tramo N° 12

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 22.706	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 38.804	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 89.64	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 24-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

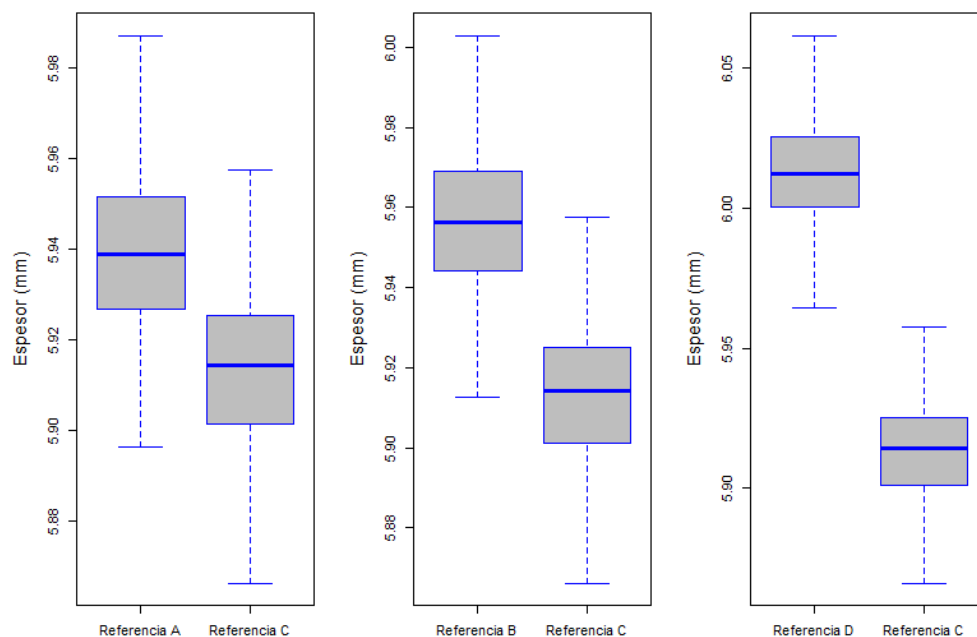


Figura 12-3: Diagrama de caja y bigote del tramo N° 12

Realizado por: Zambrano J, 2022.

3.1.13. Tramo N° 13

Tabla 25-3: Resultado del análisis estadístico del tramo N° 13

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,5685	6,006	0,0172
Referencia B	p-valor > 0,6992	5,989	0,0174
Referencia A	p-valor > 0,4411	5,971	0,0164
Referencia C	p-valor > 0,1239	5,919	0,0163

Realizado por: Zambrano J, 2022.

En la Tabla 25-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 26-3: Prueba t-Student para el tramo N° 13

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 48.8	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 64.378	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 78.878	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 26-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

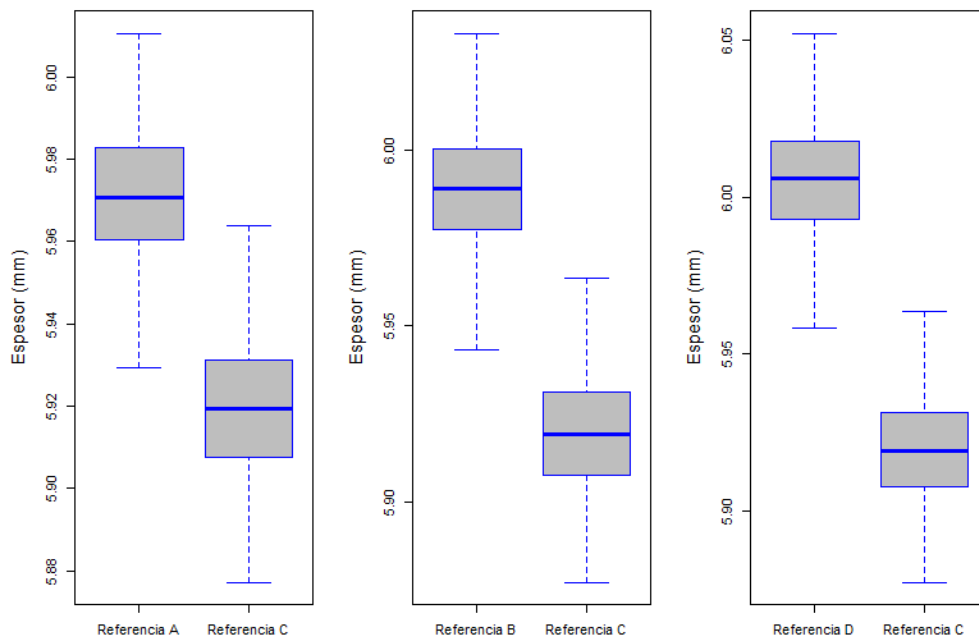


Figura 14-3: Diagrama de caja y bigote del tramo N° 13

Realizado por: Zambrano J, 2022.

3.1.14. Tramo N° 14

Tabla 27-3: Resultado del análisis estadístico del tramo N° 14

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,0689	5,965	0,0183
Referencia B	p-valor > 0,6204	5,915	0,0192
Referencia A	p-valor > 0,7871	5,895	0,0176
Referencia C	p-valor > 0,8218	5,884	0,0165

Realizado por: Zambrano J, 2022.

En la Tabla 27-3, se obtuvo que, el p-valor > 0,05, no rechazamos la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 28-3: Prueba t-Student para el tramo N° 14

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 9.927	df = 499	p-valor < 2,2e-16
Referencia B-C	t = 27.886	df = 499	p-valor < 2,2e-16
Referencia D-C	t = 72.70	df = 499	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 28-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

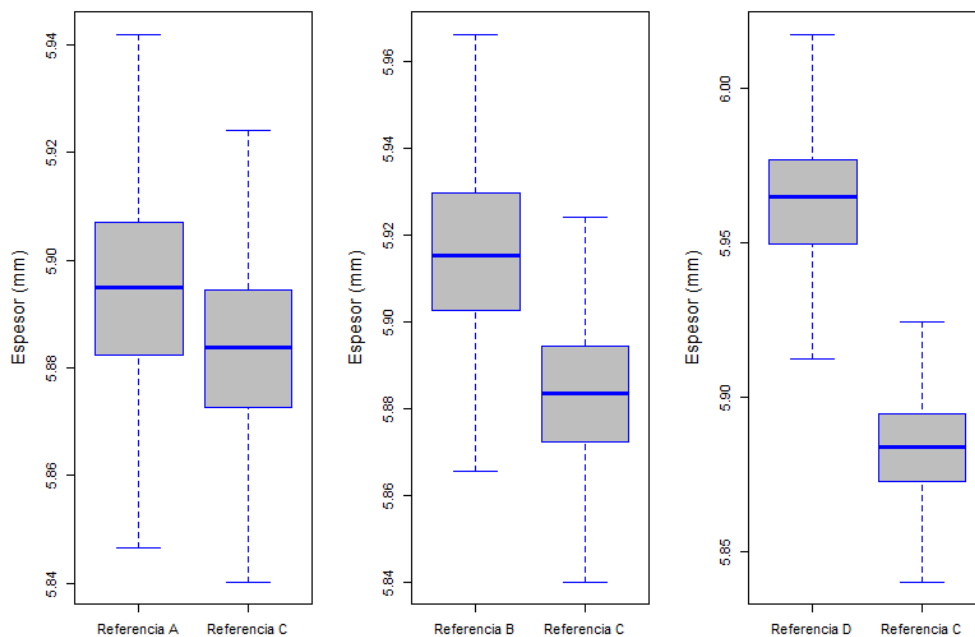


Figura 14-3: Diagrama de caja y bigote del tramo N° 14

Realizado por: Zambrano J, 2022.

3.1.15. Tramo N° 15

Tabla 29-3: Resultado del análisis estadístico del tramo N° 15

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,7027	5,993	0,0172
Referencia C	p-valor > 0,289	5,966	0,0170
Referencia B	p-valor > 0,6361	5,960	0,0168
Referencia A	p-valor > 0,4928	5,936	0,0174

Realizado por: Zambrano J, 2022.

En la Tabla 29-3, se obtuvo que, el p-valor > 0,05, es decir que los datos se ajustan a una distribución normal, se acepta la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia A.

Tabla 30-3: Prueba t-Student para el tramo N° 15

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 28.164	df = 547	p-valor = 1
Referencia B-C	t = 6.4697	df = 547	p-valor = 1
Referencia D-C	t = 24.736	df = 547	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

- H0: $\mu_A \leq \mu_C$; los espesores del punto A son menores que los del punto C.
- H1: $\mu_A > \mu_C$; los espesores del punto A son mayores que los del punto C.

En la Tabla 30-3, se obtuvo que, en la referencia D, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que el espesor del punto D es mayor que el espesor del punto C.

Mientras que en los espesores de la referencia A y B el p-valor = 1 > 0,05, se acepta la hipótesis nula y se rechaza la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B son menores que el espesor del punto C.

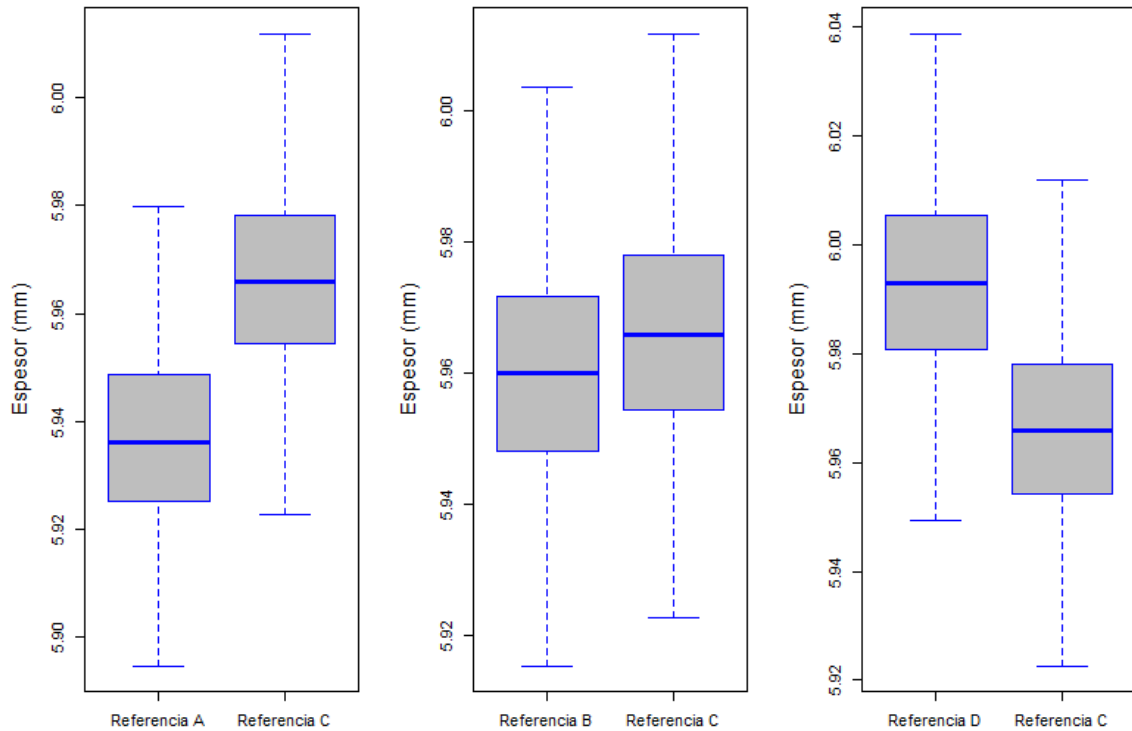


Figura 15-3: Diagrama de caja y bigote del tramo N° 15

Realizado por: Zambrano J, 2022.

3.1.16. Punto Antes de la Soldadura (PAS)

Tabla 31-3: Resultado del análisis estadístico de los PAS

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,9768	5,508	0,0217
Referencia B	p-valor > 0,1782	5,472	0,0231
Referencia A	p-valor > 0,3552	5,438	0,0218
Referencia C	p-valor > 0,8674	5,378	0,0202

Realizado por: Zambrano J, 2022.

En la Tabla 31-3, se obtuvo que, el p-valor > 0,05, es decir que los datos se ajustan a una distribución normal, se acepta la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 32-3: Prueba t-Student de los PAS

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 51.869	df = 651	p-valor < 2,2e-16
Referencia B-C	t = 80.251	df = 651	p-valor < 2,2e-16
Referencia D-C	t = 113.53	df = 651	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 32-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

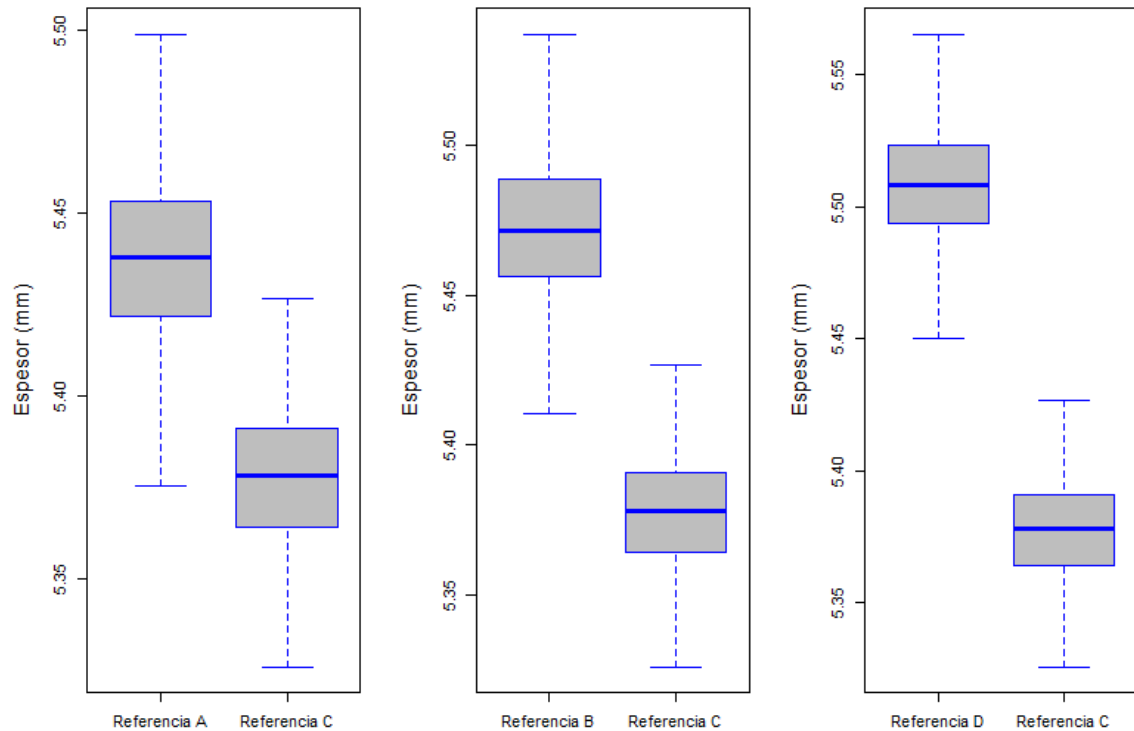


Figura 16-3: Diagrama de caja y bigote de los PAS

Realizado por: Zambrano J, 2022.

3.1.17. Punto Antes de la Soldadura (PDS)

Tabla 33-3: Resultado del análisis estadístico del tramo de los PDS

Punto de referencia	Prueba de normalidad	Mediana [mm]	Desviación estándar [mm]
Referencia D	p-valor > 0,6922	5,496	0,0210
Referencia B	p-valor > 0,7414	5,457	0,0214
Referencia A	p-valor > 0,755	5,452	0,0212
Referencia C	p-valor > 0,2453	5,393	0,0216

Realizado por: Zambrano J, 2022.

En la Tabla 33-3, se obtuvo que, el p-valor > 0,05, es decir que los datos se ajustan a una distribución normal, se acepta la hipótesis nula y en función a la mediana, el menor desgaste está en la referencia C.

Tabla 34-3: Prueba t-Student de los PDS

Punto de comparación	Valor estadístico	Grados de libertad	Valor p
Referencia A-C	t = 51.217	df = 679	p-valor < 2,2e-16
Referencia B-C	t = 55.379	df = 679	p-valor < 2,2e-16
Referencia D-C	t = 88.074	df = 679	p-valor < 2,2e-16

Realizado por: Zambrano J, 2022.

En la Tabla 34-3, se obtuvo que, el p-valor es menor a 0,05, es decir ($2,2e-16 < 0,05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que los espesores del punto A, B y D son mayores que los del punto C.

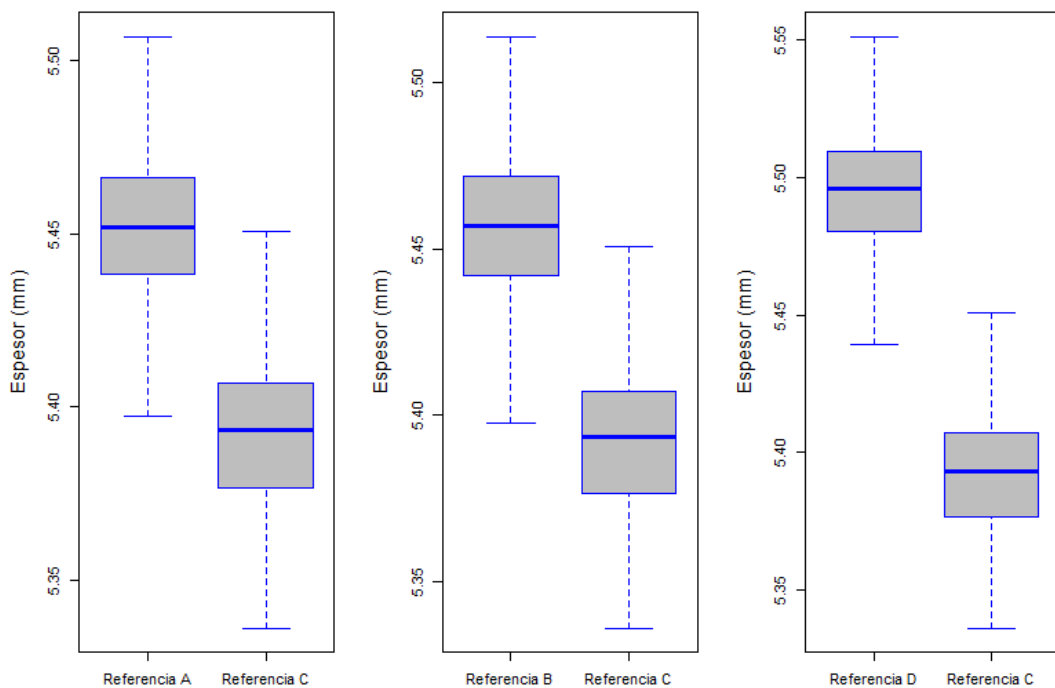


Figura 17-3: Diagrama de caja y bigote de los PDS

Realizado por: Zambrano J, 2022.

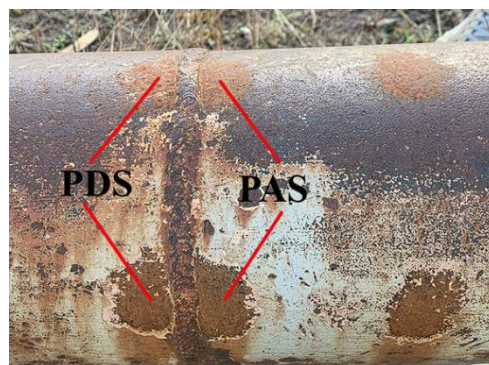


Figura 18-3: PAS y PDS

Realizado por: Zambrano J, 2022.

3.2. Resumen estadístico de cada tramo de la tubería de regadío

3.2.1. Análisis comparativo y descriptivo de mayor espesor

Tabla 35-3: Resumen de mayor espesor en función de la mediana

Tramos	Distancia [m]	Muestra [n]	Punto de referencia	Mediana Espesor [mm]
Tramo 1	0-500	501	Referencia D	5,079
Tramo 2	501-1000	499	Referencia A	5,148
Tramo 3	1001-1500	500	Referencia D	5,079
Tramo 4	1501-2000	500	Referencia D	5,065
Tramo 5	2001-2500	500	Referencia D	5,066
Tramo 6	2501-3000	500	Referencia B	4,972
Tramo 7	3001-3500	500	Referencia D	5,020
Tramo 8	3501-4000	500	Referencia B	5,004
Tramo 9	4001-4500	500	Referencia B	5,505
Tramo 10	4501-5000	500	Referencia D	5,896
Tramo 11	5001-5500	500	Referencia D	5,936
Tramo 12	5501-6000	500	Referencia D	6,012
Tramo 13	6001-6500	500	Referencia D	6,006
Tramo 14	6501-7000	500	Referencia D	5,965
Tramo 15	7000-7548	548	Referencia D	5,993
PAS	0-7548	652	Referencia D	5,508
PDS	0-7548	680	Referencia D	5,496

Realizado por: Zambrano J, 2022.

En la Tabla 35-3, se presentan las medidas de los 15 tramos en función de la mediana, obteniendo que la mayor tendencia de espesor está el tramo N° 12, en el punto de referencia D, y existe un menor espesor en el tramo N° 6, en el punto de referencia B.

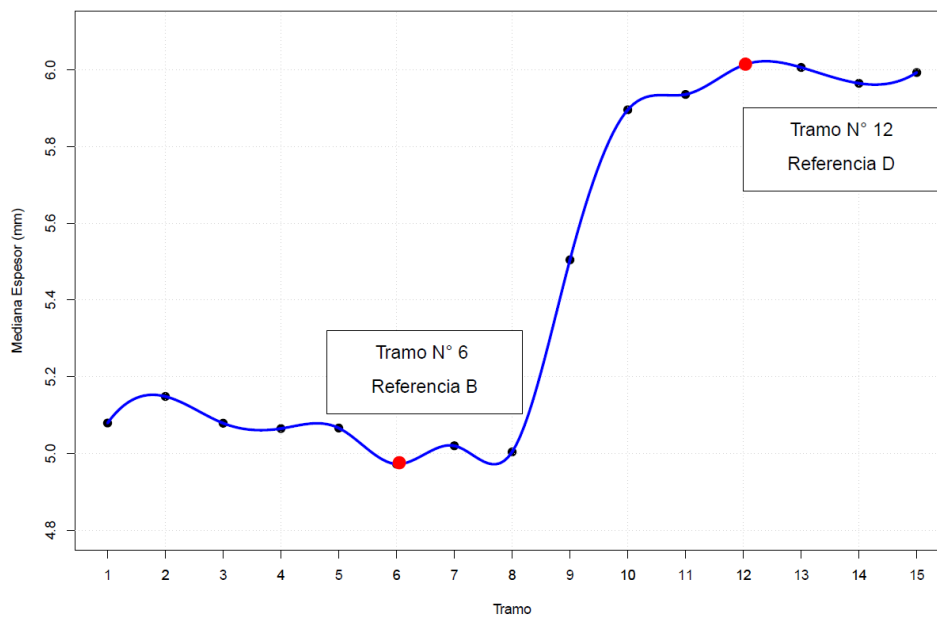


Gráfico 1-3: Curva de tendencia de mayor espesor en función de la mediana

Realizado por: Zambrano J, 2022

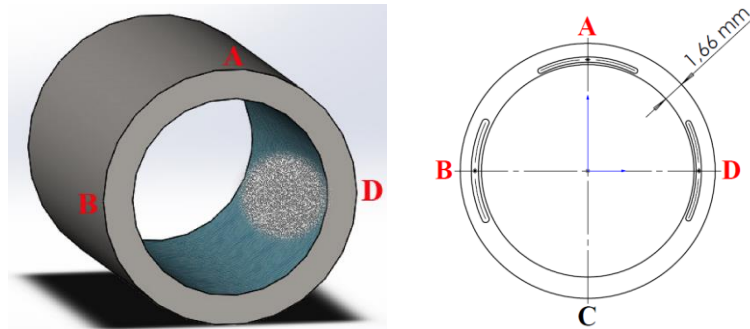


Figura 18-3: Ubicación geográfica de las referencias (A, B y D)

Realizado por: Zambrano J, 2022.

En comparación del espesor nominal que es de 7,11 mm, existe un desgaste promedio de 1,66 mm entre las referencias A, B y D, presentan un desgaste mínimo en un 25,4%, puesto que se encuentran en las partes laterales y superior de la tubería como se muestra en la Figura, teniendo menor alteración de fluido y menor velocidad de desgaste causado por la pérdida paulatina del material.

3.2.2. *Análisis comparativo y descriptivo de menor espesor*

Tabla 36-3: Resumen de menor espesor en función de la mediana

Tramos	Distancia [m]	Muestra [n]	Punto de referencia	Mediana Espesor [mm]
Tramo 1	0-500	501	Referencia C	4,942
Tramo 2	501-1000	499	Referencia C	4,970
Tramo 3	1001-1500	500	Referencia C	4,915
Tramo 4	1501-2000	500	Referencia C	5,000
Tramo 5	2001-2500	500	Referencia C	5,001
Tramo 6	2501-3000	500	Referencia C	4,885
Tramo 7	3001-3500	500	Referencia C	4,932
Tramo 8	3501-4000	500	Referencia C	4,935
Tramo 9	4001-4500	500	Referencia C	5,402
Tramo 10	4501-5000	500	Referencia C	5,789
Tramo 11	5001-5500	500	Referencia C	5,787
Tramo 12	5501-6000	500	Referencia C	5,914
Tramo 13	6001-6500	500	Referencia C	5,919
Tramo 14	6501-7000	500	Referencia C	5,884
Tramo 15	7000-7548	548	Referencia A	5,936
PAS	0-7548	652	Referencia C	5,378
PDS	0-7548	680	Referencia C	5,393

Realizado por: Zambrano J, 2022.

Se observa en la Tabla 36-3, se presenta las medidas en función de la mediana, el mayor espesor es el tramo N° 15 y la menor tendencia que tenga menor espesor es el tramo N° 6, los puntos de referencia con mayor aleatoriedad en los 15 tramos es la referencia C.

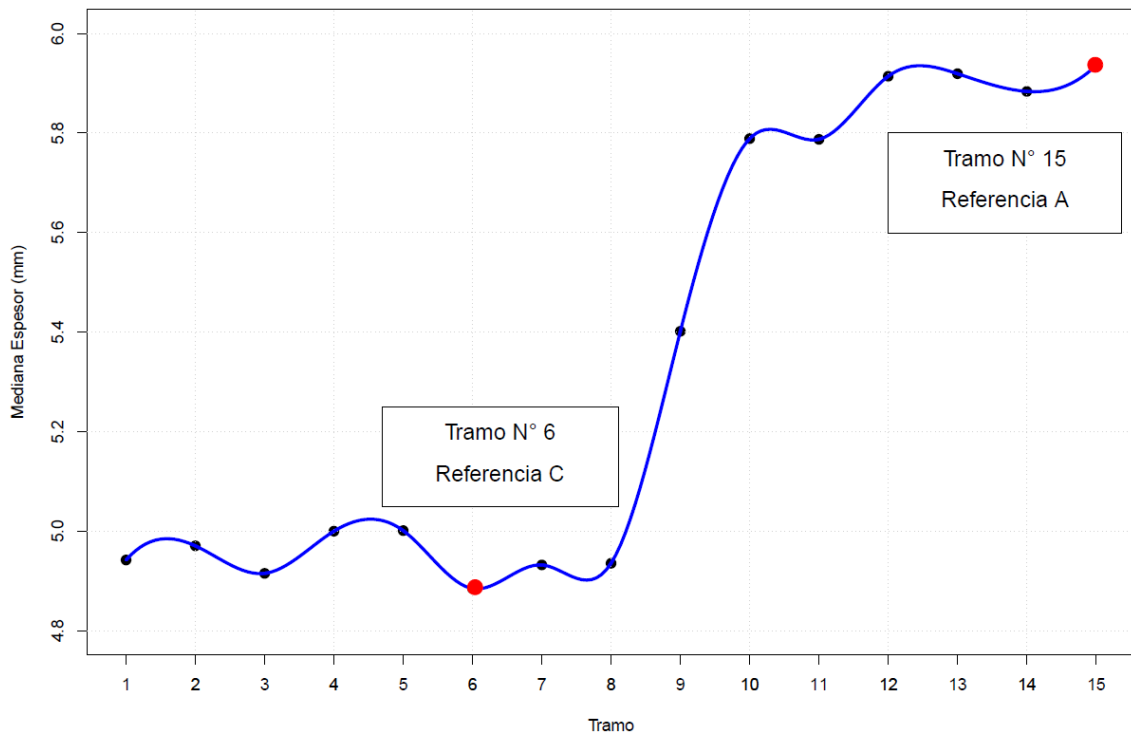


Gráfico 2-3: Curva de tendencia de menor espesor en función de la mediana

Realizado por: Zambrano J, 2022.

En comparación del espesor nominal que es de 7,11 mm existe un desgaste promedio de 1,76 mm, la referencia C, presenta el mayor desgaste, con una pérdida progresiva en un 74,6%, puesto que se encuentra en la parte inferior y tiene mayor alteración de fluido y mayor velocidad de desgaste, producido por partículas que contiene el agua de riego, que provienen de las vertientes, estos sedimentos se acumulan en la parte inferior de la tubería, como se muestra en la Figura 19-3, causando cortes, rayones, astillado, pérdidas progresiva del material y disminuyendo la vida útil de la tubería.

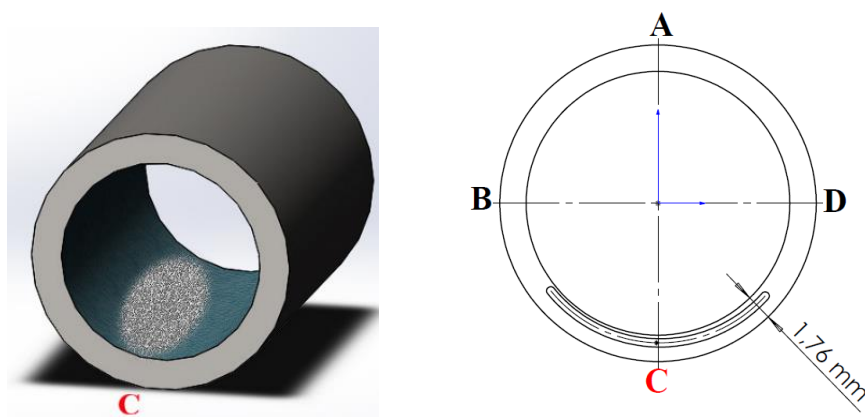


Figura 19-3: Ubicación geográfica de la referencia C

Realizado por: Zambrano J, 2022.

3.2.3. Análisis comparativo y descriptivo de mayor dispersión de datos

Tabla 37-3: Resumen de la mayor desviación estándar

Tramos	Distancia [m]	Muestra [n]	Punto de referencia	Desviación estándar [mm]
Tramo 1	0-500	501	Referencia C	0,0101
Tramo 2	501-1000	499	Referencia A	0,0081
Tramo 3	1001-1500	500	Referencia A	0,0085
Tramo 4	1501-2000	500	Referencia C	0,0161
Tramo 5	2001-2500	500	Referencia C	0,0177
Tramo 6	2501-3000	500	Referencia C	0,0180
Tramo 7	3001-3500	500	Referencia A	0,0156
Tramo 8	3501-4000	500	Referencia C	0,0163
Tramo 9	4001-4500	500	Referencia D	0,0265
Tramo 10	4501-5000	500	Referencia C	0,0190
Tramo 11	5001-5500	500	Referencia A	0,0202
Tramo 12	5501-6000	500	Referencia A	0,0179
Tramo 13	6001-6500	500	Referencia B	0,0174
Tramo 14	6501-7000	500	Referencia B	0,0192
Tramo 15	7000-7548	548	Referencia A	0,0174
PAS	0-7548	652	Referencia B	0,0231
PDS	0-7548	680	Referencia C	0,0216

Realizado por: Zambrano J, 2022.

En la Tabla 37-3, se presentan las medidas de los 15 tramos en función de la desviación estándar, obteniendo que la mayor dispersión se encuentra en el tramo N° 9, en el punto de referencia D, ya que se encuentra en la parte lateral de la tubería, esto indica una mayor dispersión ya que los datos son más heterogéneos. Esto es causado por la pérdida paulatina del material.

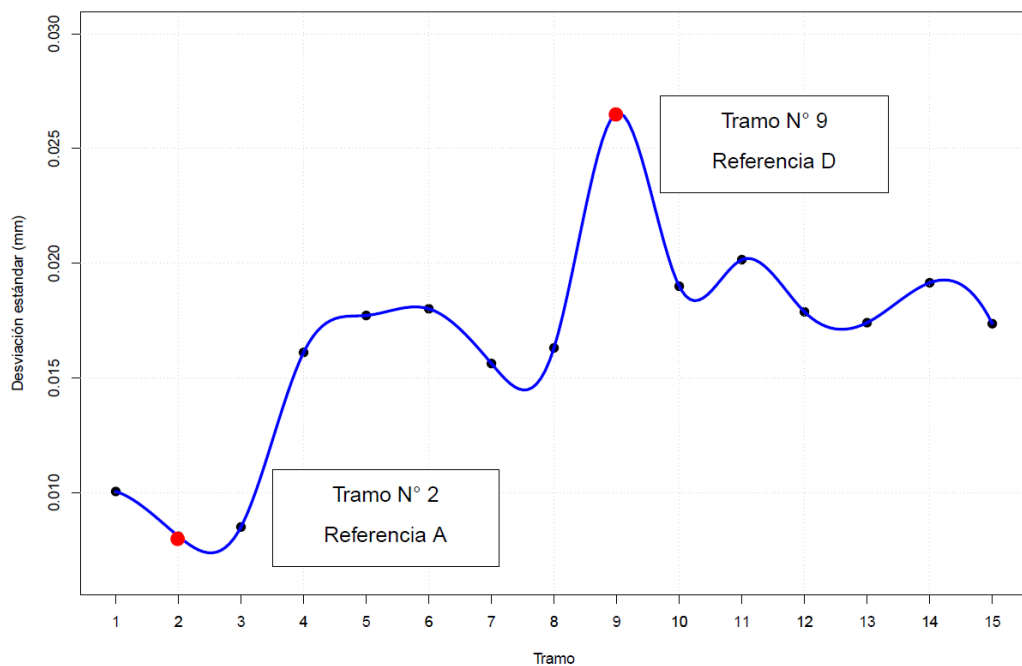


Gráfico 3-3: Curva de tendencia de mayor desviación estándar

Realizado por: Zambrano J, 2022.

De igual forma, se encuentra una menor dispersión en el tramo N° 2, en el punto de referencia A, puesto que se encuentra en la parte superior de la tubería, esto indica una menor dispersión ya que los datos son más homogéneos. Esto generalmente se da por los sedimentos y partículas que se acumulan en la parte superior de la tubería.

3.3. Análisis de resultados

Tabla 38-3: Síntesis de resultados

Tipo de función	Punto de referencia	Tipo de desgaste	Causa	Observación
Mayor espesor	Referencia D	Abrasión	Pérdida paulatina del material	Gráfico 1-3
Menor espesor	Referencia C	Abrasión	Partículas contaminantes por el agua de regadío	Gráfico 2-3
Mayor desviación estándar	Referencia D	Abrasión	Pérdida paulatina del material	Gráfico 3-3
Menor desviación estándar	Referencia B	Abrasión	Partículas contaminantes por el agua de regadío	Gráfico 3-3
Tramo N° 6	Referencia B Referencia C	Abrasión	Acumulación de partículas, por estar en una zona baja y alta	Gráfico 1-3 Gráfico 2-3

Realizado por: Zambrano J, 2022.

La tubería de regadío de la Comunidad San José la Silveria, consta de 7548 metros, divididos en 15 tramos, en los que se realizó un análisis estadístico. A estos se suman los PAS y los PDS, que también fueron parte de la investigación, obtenido los siguientes resultados:

Los resultados evidencian que cada tramo de la tubería a nivel estadístico indica que el mayor espesor está en la referencia D, esto significa que en dichos tramos existe menor alteración de fluido y menor velocidad de desgaste.

En función de la mayor desviación estándar, se evidencia que en la referencia D, los datos son más heterogéneos, concluyendo que, el desgaste que se presenta con mayor frecuencia es de carácter abrasivo mínimo, causado por la pérdida paulatina del material.

En cuanto al menor espesor y desviación estándar, la mayor aleatoriedad en los 15 tramos es la referencia C, ya que se encuentra en la parte inferior de la tubería, concluyendo que el desgaste que se presenta es el desgaste abrasivo, que es causando por cortes, rayones, astillado y pérdida progresiva del material, disminuyendo la vida útil de la tubería de regadío.

El tramo N° 6 que se encuentra a una distancia de 2500-3000 m, existe un desgaste promedio de 4,94 mm, con una pérdida progresiva del material entre las referencias B, C y D, como se muestra en la Figura 20-3, existiendo en esta parte de la tubería un cambio de alteración de fluido ya sea por las curvas o pendientes que puedan existir en dicho tramo.

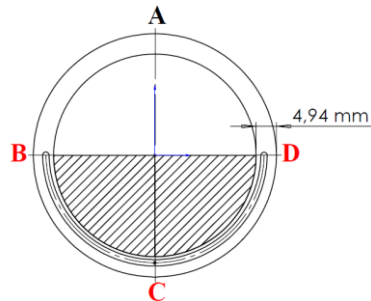


Figura 20-3: Tramo N° 06 con mayor desgaste

Realizado por: Zambrano J, 2022.

A continuación, se realizó un análisis en función del caudal, partiendo de la fórmula que es igual a la velocidad del flujo por el área de la sección, que se define por la siguiente ecuación:

$$Q=V \times A$$

De aquí se despejó la velocidad del flujo:

$$V=\frac{Q}{A}$$

Así mismo el área de la sección se define por la siguiente ecuación:

$$A=\frac{\pi \times r^2}{2}$$

Finalmente se realizó el cálculo del área de la sección y la velocidad del fluido.

Datos:

- $\varnothing_{\text{Nominal}} = 6 \text{ pulgadas} = 152,4 \text{ mm} = 15,24 \text{ cm}$
- Espesor nominal= 7,11 mm
- Espesor real= 4,94 mm
- Espesor desgastado= 2,17 mm

Cálculo del área de la sección

- $\varnothing_{\text{Real}} = 152,4 \text{ mm} + 2,1703 \text{ mm} = 154,57 \text{ mm} = 15,457 \text{ cm}$
- $r = 7,7285 \text{ cm}$

$$A_{\text{Real}} = \frac{\pi \times (7,7285 \text{ cm})^2}{2}$$

$$A_{\text{Real}} = 93,82 \text{ cm}^2$$

Calculo de la velocidad del fluido

$$V = \frac{Q}{A_{\text{Real}}}$$

$$V = \frac{Q}{93,82 \text{ cm}^2}$$

“Q”, representa la velocidad del fluido cuyo valor depende del flujo volumétrico que está circulando por la sección de la tubería, pues si la tubería se encuentra en estado horizontal la velocidad del fluido se ralentiza es decir se mantiene constante, mientras que, cuando concurren pendientes existen aumentos de velocidad del fluido, concluyendo que, si aumenta el caudal, aumenta la velocidad, de igual forma si disminuye el caudal, también disminuye la velocidad.

Por último, se analizó los PAS y PDS, evidenciando el menor desgaste en la referencia C, causado por el exceso de soldadura o cordón defectuoso. De igual manera, en función a la desviación estándar, la mayor dispersión se encuentra en la referencia C.

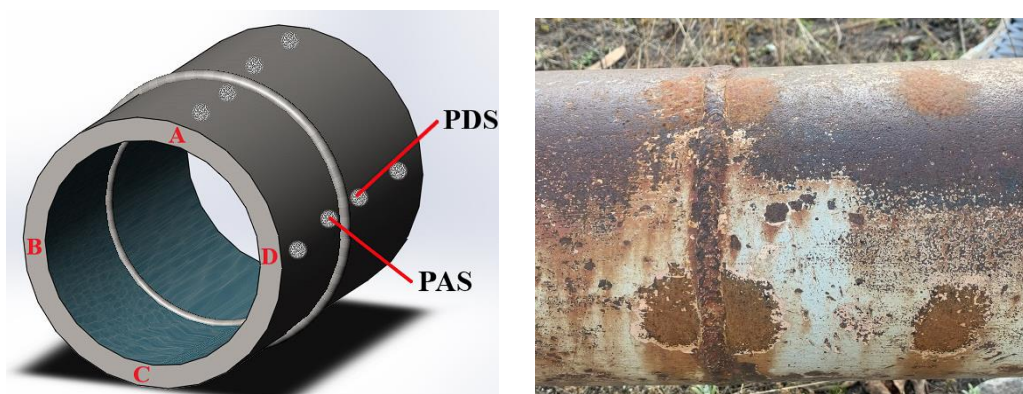


Figura 21-3: Ubicación geográfica de los PAS y PDS

Realizado por: Zambrano J, 2022.

CONCLUSIONES

La aplicación de los conceptos estadísticos en función de la mediana y la desviación estándar a partir del ensayo no destructivo con ultrasonido sirve como principal instrumento para la pronosticación de desgaste de la tubería y permite analizar profundamente los factores incidentes.

El análisis de datos permite realizar un estudio más específico y una depuración tomando en cuenta los Puntos Antes de la Soldadura (PAS) y Puntos Después de la Soldadura (PDS), esta estrategia sugiere el estudio por separado de los tramos dentro del sistema de riego y la obtención de datos de forma ágil.

El procesamiento de los datos y la elaboración de script en el software RStudio demostraron una gran capacidad y fiabilidad para el análisis estadístico descriptivo para cada tramo de la tubería mediante diagrama de caja.

La prueba T-Student aporta significativamente en el análisis del tipo y nivel de desgaste de la tubería, ya que muestra a través de sus interpretaciones el tipo de desgaste más predominante, que en esta investigación es el desgaste por abrasión.

RECOMENDACIONES

Es importante manejar con responsabilidad los datos recolectados y brindarles un estudio de carácter profesional, pues los resultados obtenidos servirán en la toma de decisiones y el establecimiento de estrategias.

Se debe tener en cuenta el valor de la desviación estándar en los casos de que este sea alto, debido a que se considera volátil y conlleva más riesgo en tomar decisiones, lo cual provoca una mayor incertidumbre.

La depuración de los datos se debe realizar de una manera precisa en la tabulación y con una rigurosa revisión para evitar inconvenientes y contratiempos durante el análisis ya que a través de ello se podrá garantizar veracidad en la información.

Se recomienda a la Comunidad San José la Silveria y sus alrededores, realizar mingas de mantenimiento, aplicando pintura anticorrosiva en las superficies externas de la tubería de regadío, para propiciar la prolongación de la vida útil del sistema de riego y aportar a la producción de la zona.

Dentro de la malla curricular de la Carrera de Mantenimiento Industrial, se debe incluir más horas orientadas hacia una investigación científica y análisis de datos que profundice en la enseñanza de programas estadísticos en el manejo y manipulación de gran cantidad de datos como RStudio y Python, ya que son herramientas muy útiles para el desarrollo de investigaciones de campo.

BIBLIOGRAFÍA

ALI, AHMED ABDUL HUSSEIN, et al. “An Investigation to the Abrasive Wear in Pipes Used for Oil Industry”. *Journal of Engineering* [en línea], 2013, (Irak) 19(11), pp. 1382-1394. [Consulta: 9 febrero 2022]. ISSN 17264073-25203339. Disponible en: <https://www.iasj.net/iasj/article/79279>

CALDERÓN, JORGE ANDRÉS, et al. “Análisis estadístico de los factores que intervienen en la corrosión de tuberías enterradas”. *Revista Facultad de Ingeniería Universidad de Antioquia*. [en línea], 2005, (Colombia) 34(1), pp. 23-34. [Consulta: 10 febrero 2022]. ISSN 0120-6230. Disponible en: <https://www.redalyc.org/pdf/430/43003402.pdf>

API 571. *Damage Mechanisms Affecting Fixed Equipment in the Refining Industry*. Mecanismos de Daño que Afectan a Equipos Fijos en la Industria de Refinación

ARIAS, ENERIO RODRÍGUEZ. “Estadística: Medición, descripción e inferencia”. *Perspectivas Psicológicas*, Santo Domingo. [en línea], 2010, (República Dominicana) 6(1), pp. 126-172. [Consulta: 6 febrero 2022]. ISSN 1992-5131. Disponible en: <http://pepsic.bvsalud.org/pdf/pp/v6e7n10/a23.pdf>

AYALA, GUILLERMO. *Estadística Básica* [en línea]. Valencia-España, 2019. [Consulta: 7 febrero 2022]. Disponible en: <https://www.uv.es/ayala/docencia/nmr/nmr13.pdf>

BOSONI, GIORGIO BOCCARDO; BRUZZONE, FELIPE RUIZ. *RStudio para Estadística Descriptiva en Ciencias Sociales*. [en línea]. Santiago-Chile, 2019. [Consulta: 17 febrero 2022]. Disponible en: https://www.researchgate.net/profile/Felipe-Ruiz-13/publication/334308025_RStudio_para_Estadistica_Descriptiva_en_Ciencias_Sociales_Manual_de_apoyo_docente_para_la_asignatura_Estadistica_Descriptiva_Carrera_de_Sociologia_Universidad_de_Chile_segunda_edicion/links/5d23aafa92851cf4407274cf/RStudio-para-Estadistica-Descriptiva-en-Ciencias-Sociales-Manual-de-apoyo-docente-para-la-asignatura-Estadistica-Descriptiva-Carrera-de-Sociologia-Universidad-de-Chile-segunda-edicion.pdf

CONTENTO RUBIO, MANUEL RICARDO. *Estadística con aplicaciones en R* [en línea]. Bogotá-Colombia: Tadeo Lozano, 2019. [Consulta: 6 febrero 2022]. Disponible en: <https://expeditiorepositorio.utadeo.edu.co/handle/20.500.12010/21660>

CORREA, CORREA, JUAN CARLOS; IRAL, RENÉ; ROJAS, LUCINIA. “Estudio de potencia de pruebas de homogeneidad de varianza”. *Revista colombiana de estadística*. [en línea], 2006, (Colombia) 29(1), p. 57-76. [Consulta: 6 febrero 2022]. ISSN 0120-175. Disponible en: <https://www.redalyc.org/pdf/899/89929104.pdf>

DAVIS, JOSEPH R, et al. *Surface engineering for corrosion and wear resistance, ASM International*. [en línea]. USA, 2001. [Consulta: 9 febrero 2022]. Disponible en: <https://fdocuments.in/document/surface-engineering-for-corrosion-and-wear-resistancepdf.html?page=1>

VARGAS, LISETH E., ET AL. *Introducción al análisis de datos con RStudio* [en línea]. Bogotá-Colombia, 2021. [Consulta: 6 febrero 2022]. Disponible en: <http://52.200.198.20/bitstream/handle/123456789/141281/Manual%20Estudio%20IMw.pdf?sequence=6&isAllowed=y>

WICKHAM, Hadley; GROLEMUND, Garrett. *R for Data Science: import, tidy, transform, visualize, and model data*. [en línea]. Canadá: O'Reilly Media, 2017. [Consulta: 6 febrero 2022]. Disponible en: <http://barbra-coco.dyndns.org/yuri/R/R%20for%20Data%20Science-%20Import,%20Tidy,%20Transform,%20Visualize,%20and%20Model%20Data.pdf>

JAVED, UMAR, ET AL. “Exploratory Data Analysis Based Short-Term Electrical Load Forecasting: A Comprehensive Analysis”. *Energies* [en línea], 2021, (Pakistán) 14(17), pp. 5510. [Consulta: 7 febrero 2022]. ISSN 19961073. Disponible en: <https://www.mdpi.com/1996-1073/14/17/5510/htm>

JEYAPRAKASH, NATARAJAN; YANG, CHE-HUA. *Friction, Lubrication, and Wear. Tribology in Materials and Manufacturing Wear, Friction and Lubrication* [en línea] Croacia: IntechOpen, 2020, [Consulta: 9 febrero 2022]. Disponible en: <https://www.intechopen.com/chapters/73357>

JULIO, GABRIEL. “Validación de una investigación usando la estadística como herramienta Research validation using statistics as tool” *Journal of the Selva Andina Biosphere*. [en línea], 2013, (Bolivia) 1(1), pp. 55-57. [Consulta: 6 febrero 2022]. ISSN 2308-386. Disponible en: <http://portal.amelica.org/ameli/jatsRepo/71/711335006/711335006.pdf>

KOMOROWSKI, MATTHIEU, et al. *Exploratory Data Analysis*. Secondary analysis of electronic health records [en línea]. Massachusetts-USA, 2016. [Consulta: 16 febrero 2022]. Disponible en: https://link.springer.com/chapter/10.1007/978-3-319-43742-2_15

LUGO, ARMENTA, JESÚS GUADALUPE; PINO-FAN, LUIS R. “Niveles de Razonamiento Inferencial para el Estadístico t-Student”. *Bolema: Boletim de Educação Matemática*, [en línea], 2022, (Brasil) 35(71), pp. 1776-1802. [Consulta: 24 febrero 2022]. ISSN 1980-4415. Disponible en: <http://www.scielo.br/j/bolema/a/KwHRZWjFLHt4CwvHdKp5C5f/abstract/?lang=es>

MARRUGO PETRO, HORACIO MIGUEL; MELO MAYORGA, JORGE ANDRÉS. “Propuesta análisis exploratorio de datos georreferenciables y temporales, caso de estudio: producción nacional del oro 2012 a 2020”. *Fundación Universitaria Los Libertadores*. [en línea], 2020, (Colombia) pp. 1-23 [Consulta: 7 febrero 2022]. Disponible en: <https://repository.libertadores.edu.co/handle/11371/3596>

NAVARRO, ALBERTO, J. A. “The bootstrap: a 35 years old young very useful for analyzing biological data”. *Cuban Journal of Agricultural Science*. [en línea], 2016, (Cuba) 50(1), pp. 11-23. [Consulta: 16 febrero 2022]. ISSN 0034-7485. Disponible en: <http://www.redalyc.org/articulo.oa?id=193045926004>

PANCORBO FLORISTAN, FRANCISCO J. *Corrosión, degradación y envejecimiento de los materiales empleados en la edificación*. [en línea]. Barcelona-España: MARCOMBO SA Retrieved, 2010. [Consulta: 9 febrero 2022]. Disponible en: <https://books.google.com.ec/books?id=ULYPNbQISxoC&pg=PA7&lpg=PA7&dq=Seg%20%20%20Reinchanschuss+f%20%20%20BCr+Metallschutz%20%20%20D&source=bl&ots=Kd5xRBsZ-d&sig=ACfU3U2b9VKPwoUQJyOzv115v9TjU2NowQ&hl=es&sa=X&ved=2ahUKEwjB0JiOzvb1AhX3QzABHeIrDmUQ6AF6BAgCEAM#v=onepage&q=Seg%20%20%20Reinchanschuss%20%20%20BCr%20%20%20Metallschutz%20%20%20D&f=false>

ROYANI, AHMAD, et al. “A Preliminary Study of Internal Corrosion in Condensate Pipelines at Geothermal Power Plants”. *International Journal of Integrated Engineering*. [en línea], 2021, (Indonesia) 13(4), pp. 122-132. [Consulta: 20 febrero 2022]. ISSN 2229838X. Disponible en: https://www.researchgate.net/publication/351820139_A_Preliminary_Study_of_Internal_Corrosion_in_Condensate_Pipelines_at_Geothermal_Power_Plants

SALAZAR PINTO, CECILIA; CASTILLO GALARZA, SANTIAGO DEL; DEL CASTILLO GALARZA, SANTIAGO. *Fundamentos básicos de estadística* [en línea]. Quito-Ecuador, 2017. [Consulta: 6 febrero 2022]. Disponible en: <http://www.dspace.uce.edu.ec/bitstream/25000/13720/3/Fundamentos%20B%c3%a1sicos%20de%20Estad%c3%adstica-Libro.pdf>

SALGADO, CÁTIA M., et al. *Noise Versus Outliers*. Secondary analysis of electronic health records [en línea]. Massachusetts-USA, 2016. [Consulta: 16 febrero 2022]. Disponible en: https://link.springer.com/chapter/10.1007/978-3-319-43742-2_14

SAURA, JOSÉ RAMON. “Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics”. *Journal of Innovation & Knowledge*. [en línea], 2021, (España) 6(2), pp. 92-102. [Consulta: 16 febrero 2022]. ISSN 2444-569X. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2444569X20300329>

ROSS, SHELDON M. *Introducción a la estadística*. [en línea]. Barcelona-España: Reverté, 2018. [Consulta: 6 febrero 2022]. Disponible en: <https://books.google.es/books?hl=es&lr=&id=Ed3eDwAAQBAJ&oi=fnd&pg=PR5&dq=Introducci%C3%B3n+a+la+estad%C3%ADstica&ots=U839qm2Dhg&sig=KnuNgDHntI43A-xaOasGuFcmmlM#v=onepage&q=Introducci%C3%B3n%20a%20la%20estad%C3%ADstica&f=false>

STRAFFELINI, GIOVANNI. *Friction and wear*. Springer Tracts in Mechanical Engineering. [en línea]. Trento-Italia: Springer, 2015. [Consulta: 3 febrero 2022]. Disponible en: <http://link.springer.com/10.1007/978-3-319-05894-8>

TAPIA, CARLOS ERNESTO FLORES; CEVALLOS, KARLA LISSETTE FLORES. “Pruebas para comprobar la normalidad de datos en procesos productivos: ANDERSON-DARLING, RYAN-JOINER, SHAPIRO-WILK Y KOLMOGÓROV-SMIRNOV”. *Societas* [en línea], 2021, (Ecuador) 23(2), pp. 83-106. [Consulta: 16 febrero 2022]. ISSN 1560-0408. Disponible en: <https://www.revistas.up.ac.pa/index.php/societas/article/view/2302>

TRISTANCHO, JOSÉ LUIS; CAÑAS, LUZ ADRIANA. “Diseño y construcción de un tribómetro de cilindros cruzados, bajo norma técnica ASTM G83 96”. *Scientia et technica* [en línea], 2008, (Colombia) 14(38), p. 171-176. [Consulta: 3 febrero 2022]. ISSN 0122-1701. Disponible en: <https://www.redalyc.org/pdf/849/84903830.pdf>

VELASCO, MARÍA CAROLINA; ZHAO, ISAAC. “Introducción a la estimación”. *Metro Ciencia* [en línea], 2021, (Chile) 29(4), pp. 65-72. [Consulta: 16 febrero 2022]. ISSN 2737-6303.

Disponible

en:

<https://www.revistametrociencia.com.ec/index.php/revista/article/view/302>

ANEXOS

Anexo A: Código en RStudio

ANÁLISIS ESTADÍSTICO

JOAQUÍN DANIEL ZAMBRANO PINEDA

2022

Instalación de paquetes y librerías

```
install.packages("readxl")
library(readxl)

install.packages("tidyverse")
library(tidyverse)

install.packages("nortest")
library(nortest)
```

Análisis Exploratorio de Datos

Cargando la base de datos

```
TR_TOTAL_7548m <- read_excel("C:/Users/Joaquin/Desktop/Base de datos 2022 - Análisis Estadístico.xlsx", sheet = "TR TOTAL (7548m)", range = "B2:E7550")

View(TR_TOTAL_7548m)

attach(TR_TOTAL_7548m)
```

Representación gráfica

Métodos gráficos: Histograma, Gráfico cuantil-cuantil y Diagrama de caja

Histogramas de los datos

```
par(mfrow=c(2,2))

hist(`Referencia A`, prob = TRUE,
main = "Referencia A", xlab = "Espesor (mm)", ylab = "Frecuencia", ylim = c(0, 1), border = "blue", col = "gray")
x <- seq(min(`Referencia A`), max(`Referencia A`), length = 70)
f <- dnorm(x, mean = mean(`Referencia A`), sd = sd(`Referencia A`))
lines(x, f, col = "red", lwd = 3)

hist(`Referencia B`, prob = TRUE,
main = "Referencia B", xlab = "Espesor (mm)", ylab = "Frecuencia", ylim = c(0, 1), border = "blue", col = "gray")
x <- seq(min(`Referencia B`), max(`Referencia B`), length = 70)
f <- dnorm(x, mean = mean(`Referencia B`), sd = sd(`Referencia B`))
lines(x, f, col = "red", lwd = 3)

hist(`Referencia C`, prob = TRUE,
main = "Referencia C", xlab = "Espesor (mm)", ylab = "Frecuencia", ylim = c(0, 1), border = "blue", col = "gray")
x <- seq(min(`Referencia C`), max(`Referencia C`), length = 70)
f <- dnorm(x, mean = mean(`Referencia C`), sd = sd(`Referencia C`))
lines(x, f, col = "red", lwd = 3)

hist(`Referencia D`, prob = TRUE,
main = "Referencia D", xlab = "Espesor (mm)", ylab = "Frecuencia", ylim = c(0, 1.5), border = "blue", col = "gray")
x <- seq(min(`Referencia D`), max(`Referencia D`), length = 70)
f <- dnorm(x, mean = mean(`Referencia D`), sd = sd(`Referencia D`))
lines(x, f, col = "red", lwd = 3)
```

Gráfico cuantil-cuantil de los datos

```
par(mfrow=c(2,2))

qqnorm(`Referencia A`,main = "Referencia A", xlab = "", ylab = "Espesor (mm)",)
qqline(`Referencia A`, col = "red", lwd = 3)

qqnorm(`Referencia B`, main = "Referencia B", xlab = "", ylab = "Espesor (mm)")
qqline(`Referencia B`, col = "red", lwd = 3)

qqnorm(`Referencia C`,main = "Referencia C", xlab = "", ylab = "Espesor (mm)")
qqline(`Referencia C`, col = "red", lwd = 3)

qqnorm(`Referencia D`,main = "Referencia D", xlab = "", ylab = "Espesor (mm)")
qqline(`Referencia D`, col = "red", lwd = 3)
```

Comprobación de los datos si siguen una distribución normal

Prueba de Normalidad de los datos

- Para muestras ($n < 50$) Prueba de Shapiro-Wilk
- Para muestra ($n > 50$) Prueba de Kolmogorov-Smirnov
 - El p-valor $> 0,05$; siguen una distribución normal
 - El p-valor $< 0,05$; no siguen una distribución normal

Prueba de Kolmogorov-Smirnov

```
lillie.test(`Referencia A`)
lillie.test(`Referencia B`)
lillie.test(`Referencia C`)
lillie.test(`Referencia D`)
```

Diagrama de caja de normalidad de los datos

```
par(mfrow=c(2,2))

boxplot(`Referencia A`, ylab="Espesor (mm)",main="Referencia A",border = "blue", col = "gray")

boxplot(`Referencia B`,ylab="Espesor (mm)",main="Referencia B",border = "blue", col = "gray")

boxplot(`Referencia C`,ylab="Espesor (mm)",main="Referencia C",border = "blue", col = "gray")

boxplot(`Referencia D`,ylab="Espesor (mm)",main="Referencia D",border = "blue", col = "gray")
```

Método de Bootstrap

Referencia A

```
Bootstrap_A <- replicate(n=7548,sample(`Referencia A`,replace = TRUE))
Media_Referencia_A <- apply(Bootstrap_A,MARGIN=2,FUN= mean)
```

Referencia B

```
Bootstrap_B <- replicate(n=7548,sample(`Referencia B`,replace = TRUE))
Media_Referencia_B <- apply(Bootstrap_B,MARGIN=2,FUN= mean)
```

Referencia C

```
Bootstrap_C <- replicate(n=7548,sample(`Referencia C`,replace = TRUE))
Media_Referencia_C <- apply(Bootstrap_C,MARGIN=2,FUN= mean)
```

Referencia D

```
Bootstrap_D <- replicate(n=7548,sample(`Referencia D`,replace = TRUE))
Media_Referencia_D <- apply(Bootstrap_D,MARGIN=2,FUN= mean)
```

Prueba de Kolmogorov-Smirnov de la distribución de Bootstrap

```
lillie.test(Media_Referencia_A)
lillie.test(Media_Referencia_B)
lillie.test(Media_Referencia_C)
lillie.test(Media_Referencia_D)
```

Diagrama de caja de la distribución de Bootstrap

```
par(mfrow=c(2,2))

boxplot(Media_Referencia_A,ylab="Espesor (mm)",main="Referencia A",border = "blue", col = "gray")

boxplot(Media_Referencia_B,ylab="Espesor (mm)",main="Referencia B",border = "blue", col = "gray")

boxplot(Media_Referencia_C,ylab="Espesor (mm)",main="Referencia C",border = "blue", col = "gray")

boxplot(Media_Referencia_D,ylab="Espesor (mm)",main="Referencia D",border = "blue", col = "gray")
```

Reemplazo de los valores atípicos (Outliers)

```
replace_outliers <- function(x, removeNA = TRUE){
  qrts <- quantile(x, probs = c(0.25, 0.75), na.rm = removeNA)
  caps <- quantile(x, probs = c(.05, .95), na.rm = removeNA)
  iqr <- qrts[2]-qrts[1]
  h <- 1.5 * iqr
  x[x<qrts[1]-h] <- caps[1]
  x[x>qrts[2]+h] <- caps[2]
  x
}
```

```
Referencia_A <- replace_outliers(Media_Referencia_A)
Referencia_B <- replace_outliers(Media_Referencia_B)
Referencia_C <- replace_outliers(Media_Referencia_C)
Referencia_D <- replace_outliers(Media_Referencia_D)
```

```
Data.A <- data.frame(Media_Referencia_A, Referencia_A)
Data.B <- data.frame(Media_Referencia_B, Referencia_B)
Data.C <- data.frame(Media_Referencia_C, Referencia_C)
Data.D <- data.frame(Media_Referencia_D, Referencia_D)
```

Diagrama de caja del reemplazo de los valores atípicos

```
par(mfrow = c(2,2))

with(Data.A, boxplot(Data.A, names = c("Con outliers","Sin outliers"),
  ylab="Espesor (mm)", main="Referencia A", border = "blue", col = "gray"))

with(Data.B, boxplot(Data.B, names = c("Con outliers","Sin outliers"),
  ylab="Espesor (mm)", main="Referencia B", border = "blue", col = "gray"))

with(Data.C, boxplot(Data.C, names = c("Con outliers","Sin outliers"),
  ylab="Espesor (mm)", main="Referencia C", border = "blue", col = "gray"))

with(Data.D, boxplot(Data.D, names = c("Con outliers","Sin outliers"),
  ylab="Espesor (mm)", main="Referencia D", border = "blue", col = "gray"))
```

Histograma de los datos normalizados

```
par(mfrow = c(2,2))

hist(Referencia_A, prob = TRUE,
  main = "Referencia A",xlab = "Espesor (mm)",ylab = "Frecuencia",ylim = c(0, 70),border = "blue", col = "gray")
x <- seq(min(Referencia_A), max(Referencia_A), length = 70)
f <- dnorm(x, mean = mean(Referencia_A), sd = sd(Referencia_A))
lines(x, f, col = "red", lwd = 3)

hist(Referencia_B, prob = TRUE,
  main = "Referencia B",xlab = "Espesor (mm)",ylab = "Frecuencia",ylim = c(0, 70),border = "blue", col = "gray")
x <- seq(min(Referencia_B), max(Referencia_B), length = 70)
f <- dnorm(x, mean = mean(Referencia_B), sd = sd(Referencia_B))
lines(x, f, col = "red", lwd = 3)
```

```

hist(Referencia_C, prob = TRUE,
main = "Referencia C",xlab = "Espesor (mm)",ylab = "Frecuencia",ylim = c(0, 70),border = "blue", col = "gray")
x <- seq(min(Referencia_C), max(Referencia_C), length = 70)
f <- dnorm(x, mean = mean(Referencia_C), sd = sd(Referencia_C))
lines(x, f, col = "red", lwd = 3)

hist(Referencia_D, prob = TRUE,
main = "Referencia D",xlab = "Espesor (mm)",ylab = "Frecuencia",ylim = c(0, 70),border = "blue", col = "gray")
x <- seq(min(Referencia_D), max(Referencia_D), length = 70)
f <- dnorm(x, mean = mean(Referencia_D), sd = sd(Referencia_D))
lines(x, f, col = "red", lwd = 3)

```

Gráfico cuantil-cuantil normalizado

```

par(mfrow=c(2,2))

qqnorm(Referencia_A,main = "", xlab= "Referencia A", ylab = "Espesor (mm)")
qqline(Referencia_A, col = "red", lwd = 3)

qqnorm(Referencia_B,main = "",xlab= "Referencia B", ylab = "Espesor (mm)")
qqline(Referencia_B, col = "red", lwd = 3)

qqnorm(Referencia_C,main = "",xlab= "Referencia C", ylab = "Espesor (mm)")
qqline(Referencia_C, col = "red", lwd = 3)

qqnorm(Referencia_D,main = "",xlab= "Referencia D", ylab = "Espesor (mm)")
qqline(Referencia_D, col = "red", lwd = 3)

```

Prueba t-Student de los datos

Referencia A-C

H0: $\mu_A \leq \mu_C$; los espesores del punto A son mayores que los del punto C.

H1: $\mu_A > \mu_C$; los espesores del punto A son menores que los del punto C.

```
t.test(Referencia_A,Referencia_C, paired = TRUE, alternative = "greater")
```

Como resultado el p-valor $< 0,05$, se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que el espesor de la referencia A son mayores que los de la referencia C.

Referencia B-C

H0: $\mu_B \leq \mu_C$; los espesores del punto B son mayores que los del punto C.

H1: $\mu_B > \mu_C$; los espesores del punto B son menores que los del punto C.

```
t.test(Referencia_B,Referencia_C, paired = TRUE, alternative = "greater")
```

Como resultado el p-valor $< 0,05$, se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que el espesor de la referencia B son mayores que los de la referencia C.

Referencia D-C

H0: $\mu_D \leq \mu_C$; los espesores del punto D son mayores que los del punto C.

H1: $\mu_D > \mu_C$; los espesores del punto D son menores que los del punto C.

```
t.test(Referencia_D,Referencia_C, paired = TRUE, alternative = "greater")
```

Como resultado el p-valor $< 0,05$, se rechaza la hipótesis nula y se acepta la hipótesis alternativa, con esto se concluye que existe suficiente evidencia que respalda la observación que el espesor de la referencia D son mayores que los de la referencia C.

Diagrama de caja de la prueba t-Student

```

par(mfrow=c(1,3))

boxplot(Referencia_A,Referencia_C, ylab="Espesor (mm)", names=c("Referencia A","Referencia C"), border = "blue",col = "gray")

boxplot(Referencia_B,Referencia_C, ylab="Espesor (mm)", names=c("Referencia B","Referencia C"), border = "blue",col = "gray")

boxplot(Referencia_D,Referencia_C, ylab="Espesor (mm)", names=c("Referencia D","Referencia C"), border = "blue",col = "gray")

```


Anexo B: Trayectoria de la tubería de agua de riego

- **Comunidad de Urbina: Tramo N° 1**



- **Comunidad de Santa Rosa: Tramo N° 5**



- **Comunidad de Santa Rosa: Tramo N° 6**



- **Comunidad de Santa Lucía: Tramo N° 12**



- **Comunidad de Cuatro Esquinas: Tramo N° 15**



