



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

**“COMPARACIÓN ESTADÍSTICA DE LAS TÉCNICAS UTILIZADAS
EN LA PRODUCCIÓN DE HIDRÓGENO A PARTIR DE LODOS
RESIDUALES IRRADIADOS DE LA ZONA MINERA EN PONCE
ENRIQUEZ”**

Trabajo de Titulación

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERA ESTADÍSTICA INFORMÁTICA

AUTORA:

BEATRIZ MERCEDES OLMEDO PEREA

Riobamba – Ecuador

2022



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

**“COMPARACIÓN ESTADÍSTICA DE LAS TÉCNICAS UTILIZADAS
EN LA PRODUCCIÓN DE HIDRÓGENO A PARTIR DE LODOS
RESIDUALES IRRADIADOS DE LA ZONA MINERA EN PONCE
ENRIQUEZ”**

Trabajo de titulación

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERA ESTADÍSTICA INFORMÁTICA

AUTORA: BEATRIZ MERCEDES OLMEDO PEREA

DIRECTORA: Ing. CRISTINA ESTEFANIA RAMOS ARAUJO.

Riobamba – Ecuador

2022

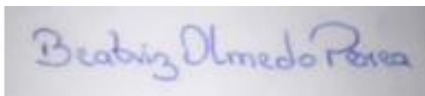
© 2022, **Beatriz Mercedes Olmedo Perea**

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

Yo, BEATRIZ MERCEDES OLMEDO PEREA, 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 autora 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, 24 de noviembre del 2022



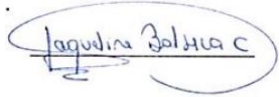
A rectangular box containing a handwritten signature in blue ink that reads "Beatriz Olmedo Perea".

Beatriz Mercede Olmedo Perea

080325800-3

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

El Tribunal del Trabajo de Titulación, certifica que: El Trabajo de Titulación.; Tipo: Proyecto de Investigación “**COMPARACIÓN ESTADÍSTICA DE LAS TÉCNICAS UTILIZADAS EN LA PRODUCCIÓN DE HIDRÓGENO A PARTIR DE LODOS RESIDUALES IRRADIADOS DE LA ZONA MINERA EN PONCE ENRIQUEZ**”, realizado por la señorita: **BEATRIZ MERCEDES OLMEDO PEREA**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación. El mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal autoriza su presentación.

	FIRMA	FECHA
Ing. Natalia Alexandra Pérez Londo, MSc. PRESIDENTE DEL TRIBUNAL		2022 - 11 - 24
Ing. Cristina Estefanía Ramos Araujo, MGs. DIRECTOR DEL TRABAJO DE TITULACION		2022 - 11 - 24
Dra. Jaqueline Elizabeth Balseca Castro, MGs. MIEMBRO DE TRIBUNAL		2022 - 11 - 24

DEDICATORIA

El presente trabajo de investigación, está dedicada en primer lugar a dios grande y todo poderoso, a mi madre Mercedes Perea Ballesteros que siempre me inculco el estudio, apoyándome en todos mis sueños, metas y que a pesar de su perdida me sigue dando fuerzas desde el cielo. A mi padre Genaro Olmedo Obando, el cual me ha ayudado económica pero más importante aun emocionalmente después de la gran pérdida te tuvimos como familia, siempre con buenos consejos ha sabido guiarme para salir adelante y no rendirme a pesar de las adversidades y miedos. A mis hermanas Pamela y Cinthia que son el motor fundamental para seguir adelante y no rendirme. Y a mis amigas que siempre han estado conmigo en los buenos y malos momentos escuchándome y siempre presentes en cada paso de mi vida.

Beatriz

AGRADECIMIENTO

Agradecer a los docentes de la carrera que han sabido enseñarme de la mejor manera y con gran paciencia, agradecer inmensamente a la Ingeniera Cristina Estefanía Ramos Araujo tutora de mi investigación que ha sabido guiarme en el transcurso de todo este proceso. A la docente a la Dra. Jaqueline Balseca miembro del trabajo de titulación y maestra en el transcurso de la carrera siempre me dio ánimo para seguir enseñándome que todo en la vida tiene solución y no que hay que decaer ante cualquier adversidad y la ingeniera Natalia Pérez presidente de tribunal por su colaboración y buena predisposición.

Beatriz

TABLA DE CONTENIDO

ÍNDICE DE FIGURAS.....	x
ÍNDICE DE GRÁFICOS.....	xi
ÍNDICE DE ECUACIONES	xiii
ÍNDICE DE ANEXOS	xiv
RESUMEN.....	xv
ABSTRACT	xvi
INTRODUCCIÓN	1

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL.....	6
1.1. Bases conceptuales	6
1.1.1. <i>Análisis estadísticos exploratorio</i>	6
1.1.1.1. <i>Medidas de tendencia central</i>	6
1.1.1.2. <i>Medidas de dispersión</i>	7
1.1.1.3. <i>Medida de forma</i>	8
1.1.2. <i>Análisis inferencial</i>	10
1.1.3. <i>Análisis de regresión</i>	15
1.1.4. <i>Depuración de datos</i>	17
1.2. Bases teóricas.	21

CAPITULO II

2. MARCO METODOLÓGICO	23
2.1. Tipo de investigación	23
2.2. Localización de estudio.....	24
2.2.1. <i>Población de estudio</i>	25
2.2.2. <i>Tamaño de la muestra</i>	25
2.2.3. <i>Identificaciones variables</i>	25
2.2.4. <i>Planteamiento de la hipótesis</i>	26
2.3. Variables en estudio.....	26
2.3.1. <i>Operacionalización de las variables</i>	26

CAPÍTULO III

3.	MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS.....	27
3.1.	Depuración de la base de datos.....	28
3.2.	Imputación de la base datos.....	29
3.2.1.	<i>Imputación múltiple</i>	<i>29</i>
3.3.	Análisis de datos Atípicos.....	31
3.3.1.	<i>Prueba analítica Tukey.....</i>	<i>32</i>
3.3.1.1.	<i>Reactor 1</i>	<i>32</i>
3.3.1.2.	<i>Reactor 2</i>	<i>33</i>
3.3.1.3.	<i>Reactor 3</i>	<i>34</i>
3.3.1.4.	<i>Reactor 4</i>	<i>35</i>
3.4.	Análisis descriptivo de las variables.....	36
3.5.	Análisis Estadístico Inferencial	48
3.5.1.	<i>Aplicación del Test Mann Whitney Wilcoxon.....</i>	<i>48</i>
3.5.1.1.	<i>Prueba de Normalidad.....</i>	<i>48</i>
3.5.2.	<i>Test de Varianza Constante.</i>	<i>49</i>
3.5.3.	<i>Test de Mann Whitney</i>	<i>49</i>
3.6.	Series Temporales.....	50
3.6.1.	<i>Reactor 1_ Variable Hidrógeno 1.....</i>	<i>50</i>
3.6.2.	<i>Reactor 2_ Variable Hidrógeno 2.....</i>	<i>51</i>
3.6.3.	<i>Reactor 3_ Variable Hidrógeno 3.....</i>	<i>51</i>
3.6.4.	<i>Reactor 4_ Variable Hidrógeno 4.....</i>	<i>52</i>
3.7.	Comparación Estadística y Análisis grafico de Bland-Altman.....	53
3.7.1.	<i>Prueba de Normalidad.....</i>	<i>53</i>
3.7.1.1.	<i>Gráficas de Normalidad.....</i>	<i>53</i>
3.7.2.	<i>Prueba de Normalidad.....</i>	<i>55</i>
3.7.2.1.	<i>Gráficas de Distribución Normal.....</i>	<i>56</i>
3.7.3.	<i>Resumen de los Datos</i>	<i>58</i>
3.7.4.	<i>Diagrama de dispersión y Modelo lineal de los datos.....</i>	<i>59</i>
3.7.5.	<i>Intervalo de confianza.....</i>	<i>60</i>
3.7.6.	<i>Gráfico de Bland-Altman.....</i>	<i>61</i>
	CONCLUSIONES.....	65
	RECOMENDACIONES.....	66

BIBLIOGRAFÍA
ANEXOS

ÍNDICE DE TABLAS

Tabla 1-1:	Criterios para la interpretación de los valores de P.	11
Tabla 1-2:	Operacionalización de variables.	26
Tabla 1-3:	Diseño de las técnicas FO y EM.....	27
Tabla 2-3:	Valores Omitidos de la base por el método de (listwise)	28
Tabla 3-3:	Porcentaje de datos faltantes.....	30
Tabla 4-3:	Hidrógeno 1.....	36
Tabla 5-3:	Hidrógeno 2.....	37
Tabla 6-3:	Hidrógeno 3.....	38
Tabla 7-3:	Hidrógeno 4.....	39
Tabla 8-3:	Temperatura 1.....	40
Tabla 9-3:	Temperatura 2.....	41
Tabla 11-3:	Temperatura 4.....	43
Tabla 12-3:	Humedad 1.	44
Tabla 13-3:	Humedad 2.	45
Tabla 14-3:	Humedad 3.	46
Tabla 15-3:	Humedad 4.	47
Tabla 16-3:	Test de Normalidad (FO).....	48
Tabla 17-3:	Test de Normalidad (EM).....	48
Tabla 18-3:	Test de Varianza Constante.	49
Tabla 19-3:	Test de Mann Whitney.	49
Tabla 20-3:	Normalidad de las mediciones (FO y EM).....	53
Tabla 21-3:	Prueba de normalidad en las diferencias de las medias (FO y EM).	55
Tabla 22-3:	Resumen de datos.....	58
Tabla 23-3:	Intervalo de Confianza de los parámetros del modelo lineal.....	60

ÍNDICE DE FIGURAS

Figura 1-1: Distribución simétrica.....	8
Figura 2-1: Asimétrica a la derecha (positiva).....	9
Figura 3-1: Asimétrica a la izquierda (negativa).	9
Figura 4-1: Hipótesis nula para $N = 25$	12
Figura 5-1: Región crítica de dos colas para un nivel de significancia de 0.05.....	13
Figura 6-1: Región crítica de dos colas para un nivel de significancia de 0,01	13
Figura 7-1: Valor del índice de correlación.	16
Figura 8-1: Gráfico de Bland – Altman.	19
Figura 1-2: Proceso cuantitativo.	23
Figura 2-2: Cantón Ponce Enríquez.....	25

ÍNDICE DE GRÁFICOS

Gráfico 1-3:	Variables con valores faltantes.....	29
Gráfico 2-3:	Porcentajes de valores faltantes.....	31
Gráfico 3-3:	Identificación de atípicos Reactor 1 FO.....	32
Gráfico 4-3:	Identificación de atípicos Reactor 1 EM.....	32
Gráfico 5-3:	Identificación de atípicos Reactor 2 FO.....	33
Gráfico 6-3:	Identificación de atípicos Reactor 2 EM.....	33
Gráfico 7-3:	Identificación de atípicos Reactor 3 FO.....	34
Gráfico 8-3:	Identificación de atípicos Reactor 3 EM.....	34
Gráfico 9-3:	Identificación de atípicos Reactor 4 FO.....	35
Gráfico 10-3:	Identificación de atípicos Reactor 4 EM.....	35
Gráfico 11-3:	Histograma variable Hidrógeno 1.....	36
Gráfico 12-3:	Histograma variable Hidrógeno 2.....	37
Gráfico 13-3:	Histograma variable Hidrógeno 3.....	38
Gráfico 14-3:	Histograma variable Hidrógeno 4.....	39
Gráfico 15-3:	Histograma variable Temperatura 1.....	40
Gráfico 16-3:	Histograma variable Temperatura 2.....	41
Gráfico 17-3:	Histograma variable Temperatura 3.....	42
Gráfico 18-3:	Histograma variable Temperatura 4.....	43
Gráfico 19-3:	Histograma variable Humedad 1.....	44
Gráfico 20-3:	Histograma variable Humedad 2.....	45
Gráfico 21-3:	Histograma variable Humedad 3.....	46
Gráfico 22-3:	Histograma variable Humedad 4.....	47
Gráfico 23-3:	Serie temporal de la variable hidrógeno 1.....	50
Gráfico 24-3:	Serie temporal de la variable hidrógeno 2.....	51
Gráfico 25-3:	Serie temporal de la variable hidrógeno 3.....	51
Gráfico 26-3:	Serie temporal de la variable hidrógeno 4.....	52
Gráfico 27-3:	Histograma variable Hidrógeno 1.....	53
Gráfico 28-3:	Histograma variable Hidrógeno 2.....	54
Gráfico 29-3:	Histograma variable Hidrógeno 3.....	54
Gráfico 30-3:	Histograma variables Hidrógeno 4.....	55
Gráfico 31-3:	Distribución de las diferencias de las medias variable H1.....	56
Gráfico 32-3:	Distribución de las diferencias de las medias variable H2.....	56

Gráfico 33-3:	Distribución de las diferencias de las medias variable H3.	57
Gráfico 34-3:	Distribución de las diferencias de las medias variable H4.	57
Gráfico 35-3:	Dispersión con bisectriz y modelo lineal de los datos.	59
Gráfico 36-3:	Bland – Altman Plot Hidrógeno 1.	61
Gráfico 37-3:	Bland – Altman Plot Hidrógeno 2.	62
Gráfico 38-3:	Bland – Altman Plot Hidrógeno 3.	63
Gráfico 39-3:	Bland – Altman Plot Hidrógeno 4.	64

ÍNDICE DE ECUACIONES

Ecuación 1-1:	Media Aritmética	7
Ecuación 2-1:	Mediana	7
Ecuación 3-1:	Desviación Estándar	7
Ecuación 4-1:	Varianza.....	7
Ecuación 5-1:	Asimetría.....	8
Ecuación 6-1:	Ho y H1 del test de Kolmogorv- Smirnov	14
Ecuación 7-1:	Estadístico de prueba del test de Kolmogorov-Smirnov.....	14
Ecuación 8-1:	Mann- Whitney - Wilcoxon.....	15
Ecuación 9-1:	Correlación entre variables	16
Ecuación 10-1:	Modelo de regresión lineal	17
Ecuación 11- 1:	Regresión Múltiple	17
Ecuación 12-1:	Prueba de Tukey identificación de atípicos	18

ÍNDICE DE ANEXOS

ANEXO A: AVAL DEL GRUPO GEAA

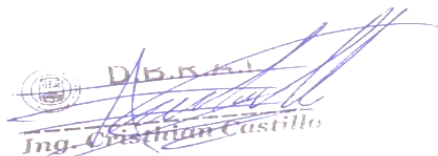
ANEXO B: CÓDIGO USADO PARA EL CÁLCULO EN R STUDIO

ANEXO C: BASE DE DATOS

RESUMEN

Esta investigación se ha basado en realizar una comparación estadística de las técnicas ya implementadas en la obtención de hidrógeno mediante lodos residuales. Teniendo un diseño de investigación experimental, variables en estudio cuantitativas, y con un nivel de profundización exploratorio transversal. En la población se emplearon en cuatro reactores las técnicas fermentación oscura y electrólisis microbiana, donde se monitoreó la generación de hidrógeno de los sedimentos con lodos residuales en el cantón Ponce Enríquez. Para la muestra se trabaja con la población total. En base a los resultados se inició con la depuración de datos utilizando las técnicas de datos completos (listwise), e imputación múltiple; teniendo así, 965 individuos imputados que corresponden al 96% de la información. En el análisis de atípico, los datos encontrados expresaron el 24% del conjunto total, mientras que para el análisis descriptivo se ha realizado gráficas y estadísticas. Las pruebas de normalidad se calcularon con la prueba basada en Kolmogorov- Smirnov y se concluyó que los datos no se distribuyen de forma normal para ambas técnicas. Según el modelo lineal de los datos el coeficiente Pearson muestra una asociación negativa entre los valores obtenidos de las técnicas, y se ha obtenido un coeficiente de correlación bajo negativo. Así mismo al comparar ambas técnicas mediante el test no paramétrica de Mann Whitney, se ha demostrado que a un nivel de 95% de confianza existe una diferencia significativa de la producción de hidrógeno según la técnica utilizada, por ende, en las series temporales han verificado el comportamiento de las técnicas en el tiempo. Finalmente, el análisis comparativo mediante la gráfica de Bland Altman indicó que no existe concordancia entre las técnicas Fermentación Oscura y Electrólisis Microbiana. Los datos no distribuyen normalmente por lo que se recomienda pruebas no paramétricas.

Palabras clave: <COMPARACIÓN ESTADÍSTICA> <DATOS ATÍPICOS > <ESTADÍSTICA INFERENCIAL> <FERMENTACIÓN OSCURA > <SERIES TEMPORALES>.

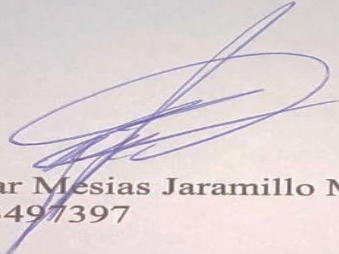

Ing. Cristian Castillo



ABSTRACT

This research has been based on performing a statistical comparison of the techniques already implemented in obtaining hydrogen through residual sludge. Having an experimental research design, quantitative study variables, with a cross-sectional exploratory level of depth. In the population, the dark fermentation and microbial electrolysis techniques were used in four reactors, where the generation of hydrogen from the sediments with residual sludge in the Ponce Enríquez canton was monitored. For the sample, the total population was considered. Based on the results, data purification has begun using the complete data techniques (listwise), and multiple imputation, thus having 965 attributed individuals that correspond to 96% of the information. In the atypical analysis, the data found expressed 24% of the total set, while for the descriptive analysis graphs and statistics have been performed. The normality tests, the test based on Kolmogorov-Smirnov have been calculated and it was concluded that the data is not distributed normally for both techniques. According to the linear model of the data, the Pearson coefficient shows a negative association between the values obtained from the techniques, and a low negative correlation coefficient has been obtained. Likewise, when comparing both techniques using the non-parametric Mann Whitney test, it has been shown that at a 95% confidence level there is a significant difference in hydrogen production according to the technique used, therefore, in the time series they have verified the behavior of techniques over time. Finally, the comparative analysis using the Bland Altman graph indicating that there is no concordance between the dark fermentation and microbial electrolysis techniques. The data is not normally distributed, so nonparametric tests are recommended.

Keywords: <STATISTICAL COMPARISON>, <ATYPICAL DATA>, <INFERENCEAL STATISTICS>, <DARK FERMENTATION>, <TIME SERIES>.



Edgar Mesias Jaramillo Moyano
0603497397

INTRODUCCIÓN

Se ha evidenciado el incremento de la población en las zonas urbanas y los ciertos cambios climáticos, por ello el daño en los ríos es más evidente. La adquisición de agua bebible ha escaseado en la actualidad para casi 2000 millones de personas y esto ha aumentado, por lo que el ahorro de agua ha sido de vital importancia. Si se considera reciclar el agua, esta puede ser accesible para más población (López Cedeño, 2019, p. 1).

La minería ha ido incrementando en el Ecuador en la última década, de la débil gobernabilidad en 2017 al retorno del pro mercado en 2019, reduciendo la figura gubernamental en la gestión de minería y fortaleciendo la participación del sector privado como eje esencial de la economía del país, esto ha influenciado para que se extraiga más recursos naturales como: oro, plata, cobre, hierro, etc (Almeida, 2019).

En el cantón Ponce Enríquez de la provincia del Azuay ha venido surgiendo la minería desde los años 80, inspirados por el aumento internacional de los precios del oro, así también, por la crisis climática ha existido un incremento en las oscilaciones de la precipitación durante El Niño Southern Oscillations (ENSO) en la región costa ecuatoriana. Este acontecimiento ha dejado al descubierto grandes vetas de oro y asentamientos descontrolados en secciones poco profundos, ha generado un impacto negativo por contaminante de metales pesados en agua, aire, suelo, flora y fauna (Sandoval, 2001, p. 3).

En las nuevas investigaciones científicas se ha considerado nuevas fuentes de energía alternativas para mitigar los efectos de la contaminación. Se ha estudiado el proceso del hidrógeno mediante electrólisis para generar energía (Velásquez & Quiceno , 2013).

Las muestras de lodos derivados de la minería de Ponce Enríquez han sido utilizadas para aplicar dos técnicas comparativas: la técnica fermentación oscura que consiste: en colocar estas muestras de lodos en los reactores buscando saber si tenían carga molecular o bacteriana, la otra técnica es electrólisis microbiana que consiste: en colocar un ánodo y cátodo de un voltio. Estas se monitorean mediante sensores de hidrógeno, sensor de metano y sensor de temperatura y humedad que han sido registrados durante 42 días (Méndez Cuzco, 2020, p. 104).

La analítica de datos es importante para describir, explorar, inferir y entender el comportamiento de variables en estudio. Para la obtención de los datos se identificó como variable dependiente el hidrógeno y los factores que se consideraron son: cuatro reactores, las técnicas fermentación oscura y electrólisis microbiana y el tiempo, posteriormente se aplicó el experimento para obtener los datos (Méndez Cuzco, 2020, p. 34).

Esta investigación se basa en comparar la producción de hidrógeno a partir de las dos técnicas. El estudio estadístico planteado busca identificar cuál de las técnicas comparadas es la más óptima en la producción de hidrógeno. Se ha realizado un análisis estadístico de la investigación, partiendo de la limpieza de la base de datos, obteniendo resultados eficientes, se ha empleado la estadística inferencial comparando las dos técnicas que son: fermentación oscura y electrólisis microbiana. Se ha estudiado el comportamiento de la variable respuesta en el tiempo, así mismo, se ha buscado conocer el grado de concordancia mediante un análisis comparativo utilizando la gráfica de Bland-Altman que cuantifica la diferencia entre ambos métodos.

Antecedentes

A partir de la producción de hidrógeno los lodos residuales irradiados tienen un proceso incorporando la fermentación oscura y electrólisis microbiana. Sin embargo, no se ha realizado suficientes estudios biotecnológicos para la purificación de lodos residuales, por consiguiente, no se ha fomentado tecnologías que permitan conseguir energía o gases no contaminantes para el aprovechamiento de los residuos derivados de la acción humana como la minería. Estos procesos son considerados significativos para los avances en tecnologías básicas y para la producción de energía, como también se consideran las represas hidráulicas y los paneles solares (Méndez Cuzco, 2020, p. 24).

Se han encontrado las siguientes investigaciones comparativas: El análisis estadístico comparativo de contenido de óxidos presentes en los lodos sin calcinar, lodos calcinados y cemento portland en donde llegan a la conclusión de que existe un efecto del tipo de lodos sobre el óxido presente, utilizando una técnica no paramétrica Kruskal Wallis (Pacheco & Varela, 2019, p. 72).

La concentración de espermatozoides es uno de los parámetros más importantes en la evaluación de la fertilidad masculina, al comparar el uso de la cámara de Makler con el uso de la cámara de Neubauer dio como resultado que la concentración de espermatozoides es uno de los parámetros más importantes en la evaluación de la fertilidad masculina. Las medias de las concentraciones espermáticas fueron $107,8 \pm 73,64$ y $106,2 \pm 74,24 \times 10^6$ espermatozoides/mL utilizando las cámaras de Makler® y Neubauer® respectivamente. El gráfico de Bland Altman indicó que los valores obtenidos por cualquier método eran similares, con un sesgo de -1,55 y un intervalo de confianza del 95% entre -46,8 y 43,7 (Cardona, et al., 2008, p. 2).

El estado nutricional actual de dos muestras de jóvenes adultos, de ambos sexos, con edades comprendidas entre los 18 y los 29 años, de dos niveles socioeconómicos diferentes (clase media, SES1 y clase desfavorecida, SES2). Para evaluar trastornos alimentarios en población chilena en hombres y mujeres entre 13 y 18 años, a través de análisis factoriales y pruebas T, resultando que la gran mayoría de las dimensiones, así como la escala total, presentaron alfas superiores a 0,70. Se observaron diferencias estadísticamente significativas en las dimensiones obsesión por la delgadez ($t(g.1=1399)=-12,286$; $p<0.001$), insatisfacción personal ($t(g.1=1364)=-6,818$; $p<0.001$), conciencia introceptiva ($t(g.1=1380)=-4,430$; $p<0.001$) e ineficacia ($t(g.1)(1363)=-3,604$; $p<0.001$), obtenido así que las mujeres obtienen puntuaciones mucho más altas que los hombres (Martínez, et al., 2019).

Para esta investigación se ha considerado la técnica comparativa Bland Altman, sin embargo, en la literatura científica no se ha encontrado aplicaciones de análisis de comparación para el área de minería. Por ello esta investigación aporta a la comparativa de dos técnicas para producción de

hidrógeno como energía alternativa y una posible solución al tratamiento de aguas contaminadas en la minería.

Planteamiento del Problema

Enunciado del problema

Se considera que una de las problemáticas más grandes para el cambio climático es el combustible procedente de fuentes no renovables, en el Ecuador se intenta mitigar estos efectos buscando energías alternativas, ya que estas no contaminan y trabajan a través del reciclaje de residuos. Se ha determinado que en ciertos puntos de ríos que están conectados con el río Chico en Camilo Ponce Enríquez contienen material contaminante debido a que abarcan un sin número de lagunas de cianuro provenientes de todos los molinos de las mineras. Se ha realizado un experimento que produce hidrógeno a partir de lodos residuales irradiados integrando un proceso de fermentación oscura y electrólisis microbiana. Para que el estudio sea relevante se ha considerado una muestra significativa de datos, que ayudará a comparar la producción de hidrógeno según cada proceso y tiempo (Méndez Cuzco, 2020, p. 21).

Es por ello que este estudio aportará a la explicación de la variable dependiente (producción de hidrógeno) y el efecto que producirá las variables independientes (fermentación oscura y electrólisis microbiana).

Formulación

¿Existe una variación significativa de la producción de hidrógeno en función de la técnica?

¿Cuáles son los factores que influyen sobre la producción de hidrógeno?

¿Existe una variación significativa de la producción de hidrógeno en función del tiempo?

OBJETIVOS

Objetivo General

- Realizar una comparación estadística en las técnicas utilizadas para la producción de hidrógeno a partir de lodos residuales.

Objetivos Específicos

- Realizar una depuración de los datos.
- Realizar un análisis estadístico inferencial para comparar la producción de hidrógeno en base a las técnicas: fermentación oscura y electrólisis microbiana.
- Realizar un análisis comparativo con el gráfico de Bland-Altman que cuantifica la diferencia entre ambos métodos.

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL.

1.1. Bases conceptuales

1.1.1. Análisis estadísticos exploratorio

Estadística Descriptiva

Se entiende la estadística descriptiva como una manera de recolectar grupos de datos, analizarlos y poder presentarlos de forma correcta o a quien los valla a utilizar, la estadística como es bien sabido se divide en descriptiva e inferencial siendo la primera de forma tabulada y la siguiente en análisis o contrastes de hipótesis (Puente Viedma, 2018, p. 17).

La estadística descriptiva es una metodología para describir numéricamente conjuntos numéricos (Vargas Sabadías, 1995, p. 33).

Estadística Descriptiva es un método para describir numéricamente conjuntos numéricos.

variables aleatorias

Las variables aleatorias es un experimento que da resultado a lo que se quiere decir, son cantidades inciertas que están asociadas, el interés está en que la mayor cantidad de experimentos no está interesada en el resultado, si no en las cantidades asociadas a dichos resultados (Lladser, 2010, p. 54).

1.1.1.1. Medidas de tendencia central

Las medidas de tendencia central en estadísticas resumen los valores de datos de un valor, son el centro en donde están ubicados el conjunto de datos, los más usados son: moda, media y mediana (Quevedo, 2011).

Media Aritmética

La media aritmética es conocida como promedio, su cálculo se obtiene realizando la sumatoria de cada valor, y luego vamos a realizar una división de estas.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \qquad 1-1$$

Mediana

La mediana es la más importante dentro de la estadística ya que su posición central divide los valores dejando cantidades iguales a ambos lados.

$$\bar{x} = \begin{cases} x(n+1)/2 \\ \frac{1}{2} \left(\frac{x_n}{2} + \frac{x_{n+1}}{2} + 1 \right) \end{cases} \qquad 2-1$$

Variables Continuas

Las variables continuas permiten diferenciar valores, teniendo como características a las mediciones de los procesos y sus fenómenos y aceptan número decimales (Martínez, 2020, p. 17).

1.1.1.2. Medidas de dispersión

Desviación Estándar

Es la medida común más utilizada la cual indica si los datos se encuentran dispersos con respecto a la media y se utiliza para las variaciones generales en algún proceso a estimar.

$$S = \sqrt{\frac{\sum_{i=1}^n d_i}{n-1}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \qquad 3-1$$

Varianza

La varianza representa la variabilidad de la media de los datos (Vargas Sabadías, 1995, p. 56).

$$S^2 = \frac{\sum_{i=1}^n d_i}{n-1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \qquad 4-1$$

1.1.1.3. Medida de forma

Las medidas de forma son utilizadas para describir las medidas de centralizado: moda, mediana, media. y al igual mide el grado de dispersión mediante desviación estándar y varianza, así mismo el grado de asimetría y curtosis (Requena Serra, 2014, p. 106).

Asimetría

Asimetría Cuando $A=0$,son valores simétricos perfectamente, presentan los coincidirán los cuando los valores de la media, mediana y moda (Requena Serra, 2014).

$$A = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{S^3} \quad 5-1$$

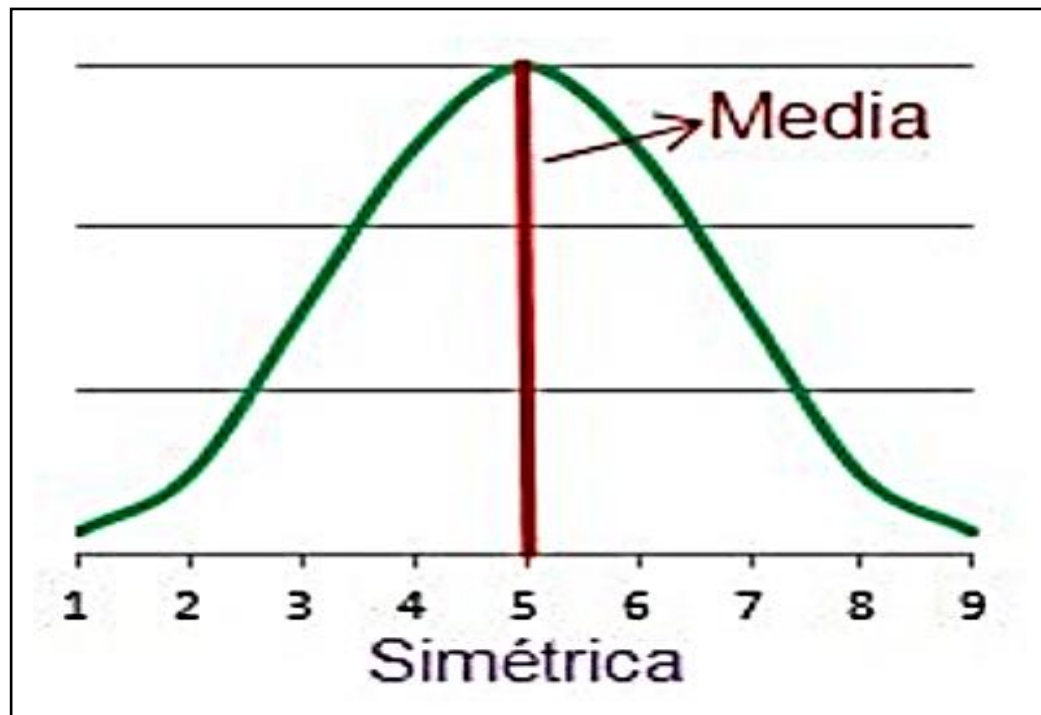


Figura 1-1: Distribución simétrica.
Fuente: (Requena Serra, 2014).

Cuando $A > 0$ es asimetría positiva: La asimetría es positiva en dirección a la derecha, donde media es mayor a la mediana (Requena Serra, 2014).

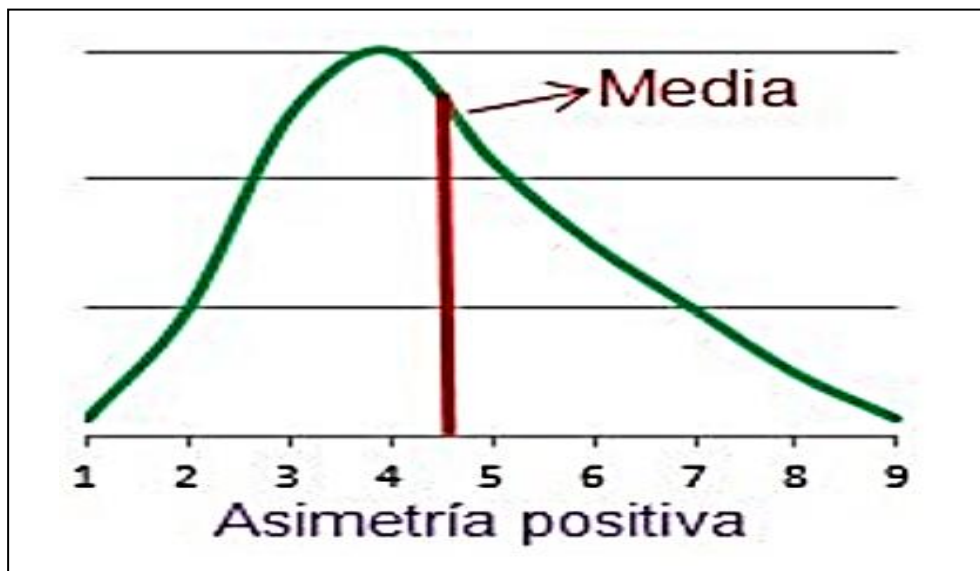


Figura 2-1: Asimétrica a la derecha (positiva).

Fuente: (Requena Serra, 2014).

Cuando $A < 0$ es asimetría negativa: La asimetría negativa en dirección a la izquierda, donde la mediana es mayor a la media (Requena Serra, 2014).

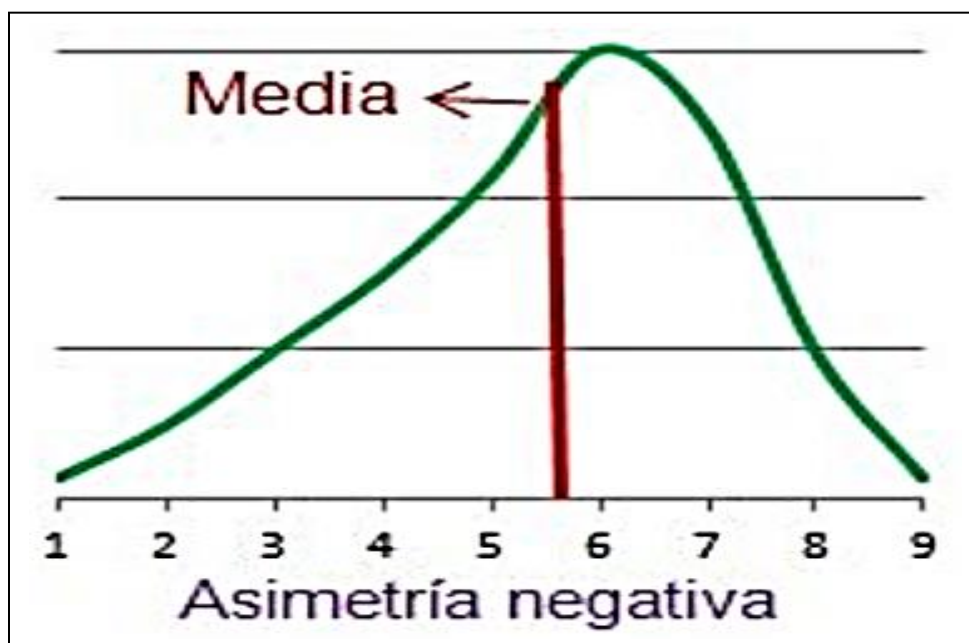


Figura 3-1: Asimétrica a la izquierda (negativa).

Fuente: (Requena Serra, 2014).

1.1.2. Análisis inferencial

Estadística Inferencial

La estadística inferencial interpreta métodos y procedimientos que deduce las probabilidades haciendo inferencia, mediante generalizaciones se toman decisiones de la información obtenida de la muestra (Rodríguez, 2015, p. 4).

Inferencia Estadística

La inferencia estadística tiene objetivo estimar los parámetros de la población partiendo de las muestras (Vázquez, 2011, p. 5).

Población

La población en estudio tiene variable de interés sobre la población estudiando la ley de probabilidad desconocida de la variable aleatoria o sus parámetros (García, 2003, p. 26).

Muestra

La muestra es la parte representativa de la población formadas por variables aleatorias simples e independientes (García, 2003, p. 12).

Intervalos de confianza

Se presenta como un modo de estimación que se realiza en inferencia estadística, que hace acotaciones a un uno o varios valores alrededor de la media muestral, esto dentro de una muestra superior y otra inferior la cual determinará la probabilidad y encontrar el parámetro poblacional. Define que el intervalo de confianza es un modo de estimación que se realiza en inferencia estadística, que hace acotaciones a un uno o varios valores alrededor de la media muestral, esto dentro de una muestra superior y otra inferior la cual determinará la probabilidad y encontrar el parámetro poblacional (Llinás Solano, 2017, p. 78).

$$\text{Intervalo de confianza} = \text{media} \pm \text{margen de error}$$

Contraste De Hipótesis

El contraste de hipótesis estadístico es un test de significación, que indican el proceso decidiendo la proposición de la población, debe ser aceptada o no (García, 2003, p. 3).

Tabla 1-1: Criterios para la interpretación de los valores de P.

Valor P	Criterio de análisis del valor P
$0,1 < P < 1,0$	La hipótesis nula es evidentemente débil
$0,01 < P < 0,1$	Cuando P disminuye, es evidente el aumento de la hipótesis nula
$0,001 < P < 0,01$	
$0,0001 < P < 0,001$	Cuando P disminuye, hay fuerte evidencia contra la hipótesis nula
$P < 0,0001$	

Fuente: (Puente Viedma, 2018).

Elaborado por: Olmedo, Beatriz, 2022.

Hipótesis Nula (Ho) e Hipótesis Alternativa (H1)

Esta hipótesis toma el supuesto de que el parámetro va a obtener toda la carga de la prueba, en donde el contraste clásico permite rechazar la hipótesis nula a priori y la probabilidad de cometer el error tipo I. Y saber si esta hipótesis nula es cierta o no, usualmente se fija en el 1%, 5% o 10%.

Esta hipótesis recibe su nombre al ser lo contrario de la hipótesis nula presentando cierto grado que se encuentra indefinido, siendo esta no cierta puede ser contraste bilateral o a dos colas, (Curso, 2010, p. 8).

Tipos de Errores

Error tipo I- rechaza la hipótesis nula cuando esta es verdadera, y sabiendo que su probabilidad está dada por la cual es más conocido con el nombre de nivel de significancia.

Error tipo II.- rechaza la hipótesis nula cuando esta es falso representando el tamaño de la muestra, (R, 2010, p. 66).

Toma de decisiones

analizar estas alternativas y seleccionar la mejor. Las etapas en la toma de decisiones son procesos que toman las decisiones a un bien conducido, entre ellas se encuentran. Definir el problema, identificar los criterios existentes para la decisión, distribuir ponderaciones a los criterios, desarrollar una serie de posibles alternativas, analizar estas alternativas y seleccionar la mejor, (Vélez Evans, 2006, p. 6).

Etapas de la toma de decisiones

Analizar estas alternativas y seleccionar la mejor. Las etapas en la toma de decisiones son procesos que toman las decisiones a un bien conducido, entre ellas se encuentran. Definir el problema, identificar los criterios existentes para la decisión, distribuir ponderaciones a los criterios, desarrollar una serie de posibles alternativas, analizar estas alternativas y seleccionar la mejor (Vélez Evans, 2006, p. 18).

Nivel de significancia

El nivel de significancia es uno de los elementos importantes de la estadística, siendo el eje central del proceso de debate. El p - value, incluso la propia regla de aceptación o rechazo (decisión dicotómica tradicionalmente asociada al valor 0.05 - correspondiente al nivel de confianza del 95%). Está vinculado directamente a la prueba estadística y se podría definir como la probabilidad (Editor, 2019, p. 3).

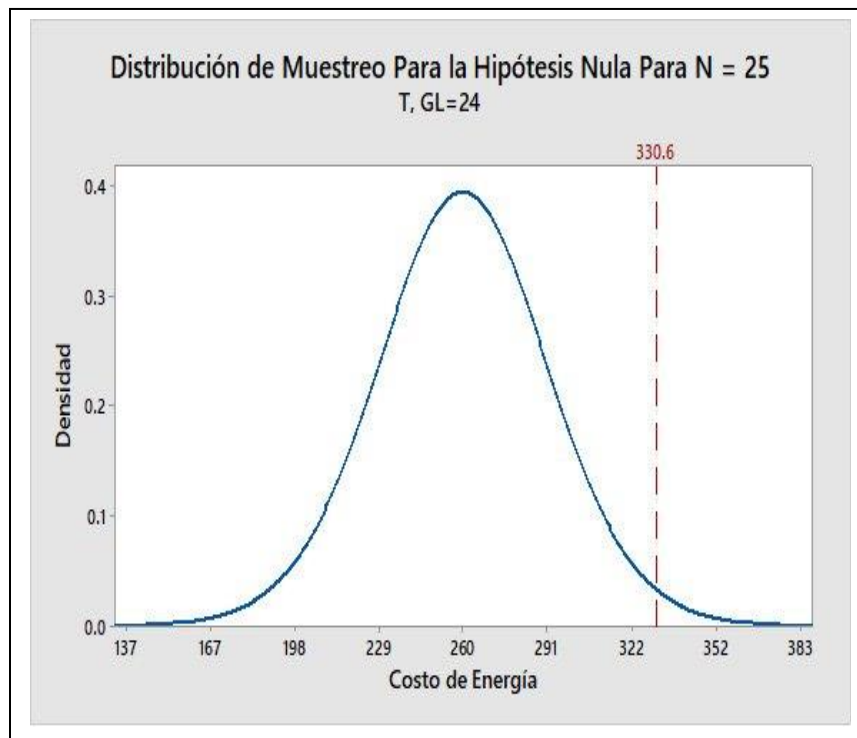


Figura 4-1: Hipótesis nula para N = 25

Fuente: (Editor, 2019)



Figura 5-1: Región crítica de dos colas para un nivel de significancia de 0.05
Fuente: (Editor, 2019).

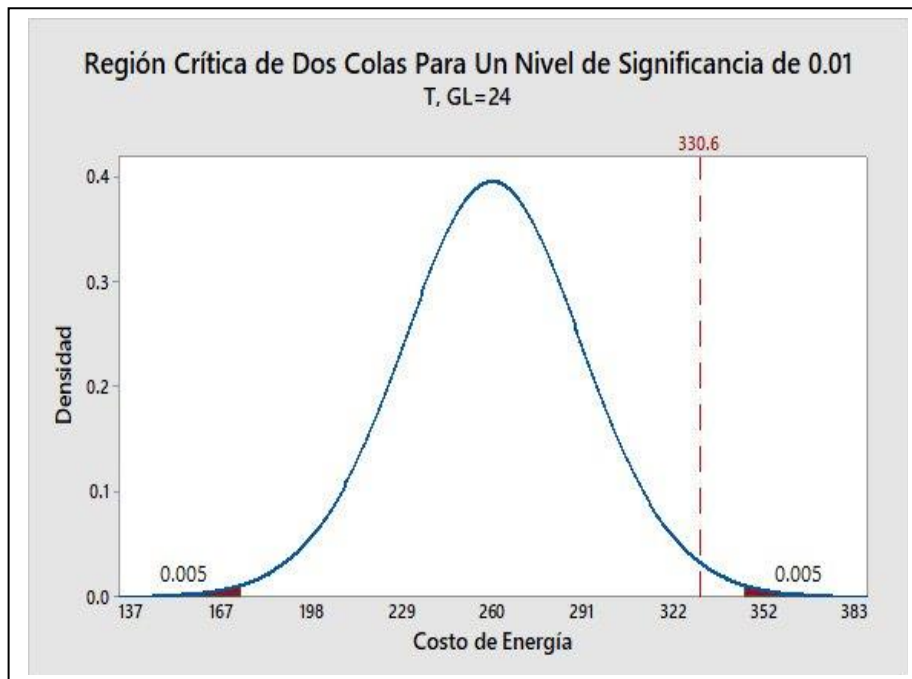


Figura 6-1: Región crítica de dos colas para un nivel de significancia de 0,01
Fuente: (Editor, 2019).

Estadísticos paramétricas y no paramétricas

La estadística paramétrica especifica la forma de distribución de las variables aleatorias, además estas realizan intervalos de confianza y prueba de hipótesis. Por otro parte los estadísticos no paramétricos no se tienen parámetros para estimar solo distribuciones que comparar y las variables en estudio tendrán que ser nominal u ordinal (Olaya Ochoa, 2012, p. 21).

Kolmogorov – Smirnov

La prueba de Kolmogorov- Smirnov más conocidas como K-S, se aplica para contrastar normalidad en los datos procedentes de una distribución normal, usadas en muestras mayores a 50 y se utilizan en variables cuantitativas continuas. Es aplicada para realizar hipótesis de normalidad, es la máxima diferencia, siendo que es poco potente para mejorar se deberá realizar una modificación mediante el test Lilliefors (Fajardo, 2007).

Definir el problema

H₀: Los datos proceden de una distribución normal.

H₁: Los datos no proceden de una distribución normal.

$$\mathbf{H}_0: F_1(x) = F_2(x) \quad 6-1$$

$$\mathbf{H}_1: F_1(x) \neq F_2(x)$$

Nivel de significancia

$$\alpha = 0.05$$

Estadístico de Prueba

$$\mathbf{D} = \sup_{1 \leq i \leq n} |\hat{F}_n(x_i) - F_0(x_i)| \quad 7-1$$

Mann-Whitney–Wilcoxon

Este tipo de pruebas no paramétricas se utiliza principalmente para comparar dos grupos diferentes e independientes de poblaciones diferentes, es semejante a la prueba T de student (Turcios Sánchez, 2015, p. 2).

$$\begin{aligned} H_0: P(X > Y) = P(Y > X) & ; & H_0: P(X > Y) = 0.5 & \quad 8-1 \\ H_\alpha: P(X > Y) \neq P(Y > X) & ; & H_\alpha: P(X > Y) \neq 0.5 \end{aligned}$$

Condiciones necesarias del test de Mann–Whitney–Wilcoxon

La Mann-Whitney-Wilcoxon tiene datos independientes ordinales, además debe asumir normalidad ya que busca comparar la mediana e igualdad de varianza y asimetría. Para realizar el test se necesitan las siguientes condiciones antes planteada (Amat , 2017) .

Coefficiente r de Pearson

El coeficiente de correlación r de Pearson mide el grado de asociación lineal entre dos variables, el valor de r puede situarse entre -1 y +1. La prueba de significación se hace con la hipótesis nula de que no hay asociación, $r = 0$, deben calcularse y comunicar los intervalos de confianza de r. Antes de decidir la aplicabilidad de una correlación lineal se debe siempre graficar en una “nube de puntos” la relación entre las dos variables (Dagnino, 2014).

Test Fligner – Killeen

La prueba de Fligner-Killeen es una prueba no paramétrica para la homogeneidad de las variaciones de grupo basada en rangos. Es útil cuando los datos no están distribuidos normalmente o cuando los problemas relacionados con valores atípicos en el conjunto de datos no se pueden resolver. También es una de las muchas pruebas de homogeneidad de varianzas que es más robusta frente a desviaciones de la normalidad (Amat , 2016).

1.1.3. *Análisis de regresión*

Coefficiente de correlación

El Coeficiente de Correlación son valores que relacionan más de dos variables cuantitativas, estas varían entre -1 y 1 siendo -1 los valores proporcionales a una correlación inversa negativa y el otro 1 los valores proporcionales a una correlación directa o positiva (Requena Serra, 2014).

Correlación entre variables

El Coeficiente de Correlación son valores que relacionan más de dos variables cuantitativas, estas varían entre -1 y 1 siendo -1 los valores proporcionales a una correlación inversa negativa y el otro 1 los valores proporcionales a una correlación directa o positiva (Peña, 2013).

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{N \sum X^2 - (\sum X)^2} * \sqrt{N \sum Y^2 - (\sum Y)^2}} \quad 9-1$$

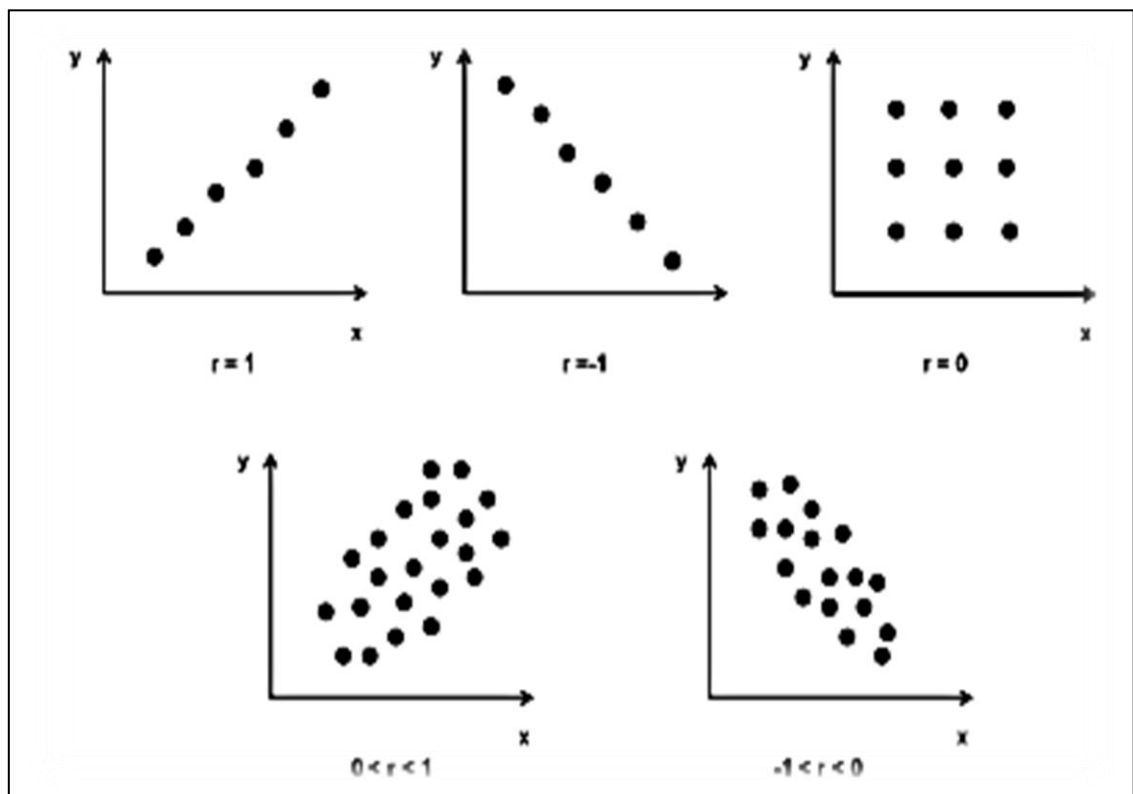


Figura 7-1: Valor del índice de correlación.

Fuente: (Becerra, 2020, p. 34).

Modelo de regresión lineal

Los modelos de regresión lineal representan la relación de dos o más variables formando un modelo que utiliza expresión lógica matemática, permitiendo realizar predicciones de valores tomándola de una de las dos (Universidad de Pamplona, et al., 2017, p. 12).

$$\widehat{Y} = \beta_0 + \beta_1 * X_1 \quad 10-1$$

Regresión Múltiple

La regresión múltiple es usada comúnmente para la explicación a la relación de las variables dependiente e independiente (Peña, 2013, p. 259).

11- 1

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

1.1.4. Depuración de datos

Los datos se pueden depurar mediante algunos métodos estadísticos estos se los puede se utilizan cuándo en la base de datos existan datos faltantes, incompletos, erróneos, e inexactos, con el propósito de organizarlos y tabularlos correctamente una vez completada la depuración de los mismo. Datos faltantes Son aquellos datos donde frecuentemente los archivos presentan observaciones "aberrantes" o faltantes (Galván & Medina, 2007, p. 15).

Datos faltantes

Son aquellos datos donde frecuentemente los archivos presentan observaciones “aberrantes” o faltantes, y existen veces en que el objeto de investigación deliberadamente se omite de la base de datos o de la información que se está tomando respecto a la población en estudio (Galván & Medina, 2007, p. 11).

Método de depuración datos completos (listwise)

El método de depuración de datos completos utiliza mayor frecuencia asumiendo datos faltantes en un patrón definido, es una manera de proceder únicamente con observaciones que disponen de información completa de las variables (Galván & Medina, 2007, p. 21).

Reglas de Rubín

La regla de Rubín propone que los datos omitidos que son remplazados por múltiples procedimientos aleatorios utilizando información que muestra realizar inferencia basado en los parámetros (Rodrigo & Marcelo , 2009).

Imputación Múltiple

La imputación múltiple tiene como significado el suplantar a las observaciones que no cuenta con información en alguna data set, la imputación múltiple crea $m > 1$ conjuntos de datos completos. Estos conjuntos de datos utilizan un software estándar y se analiza los m datos resultantes los cuales se agrupan en una estimación puntual final más un error estándar al agrupar basado en las reglas de Rubin (Galván & Medina, 2007, p. 31).

Cómo funciona la imputación múltiple

El análisis comienza con datos observados, incompletos. La imputación múltiple crea varias versiones completas de los datos al reemplazar los valores faltantes por valores de datos plausibles. Estos valores plausibles provienen de una distribución modelada específicamente para cada entrada faltante. El segundo paso es estimar los parámetros de interés los grupo de datos imputado. El último paso es juntar la m . Las estimaciones de los parámetros en una estimación, y para estimar su varianza. La varianza combina la varianza de muestreo convencional (varianza dentro de la imputación) y la varianza adicional causada por la varianza extra de los datos faltantes causada por los datos faltantes (varianza entre la imputación). Bajo las condiciones apropiadas, las estimaciones agrupadas son insesgadas y tienen las propiedades estadísticas correctas. Nota: La imputación múltiple (Rubin 1987b; Rubin 1996) resuelve el problema de errores estándar "demasiado pequeños" (Galván & Medina, 2007, p. 33).

Datos atípicos

Los datos atípicos son aquellas observaciones que se presentan de forma distinta al resto de los datos, se caracterizan por ser valores alejados del resto y generalmente aparecen entre un 1 y un 3 por 100 observaciones resto al resto de las muestras (Peña, 2013, p. 137).

Prueba de Tukey identificación de atípicos

Para la identificación de atípicos mediante la prueba de Tukey busca presentar un rango mediante el rango intercuartílico siendo que no depende de valores enormes o excesivos, así permitiendo obtener datos atípicos mediante los datos que sobrepasen el intervalo establecido (Ramírez Mordán, 2018).

$$\text{IQR} = q_3 - q_1$$

12- 1

Gráfico de Bland-Altman

El método gráfico Bland-Altman es un método que permite comparar dos técnicas de medición sobre una misma variable cuantitativa, es útil para verificar la variabilidad respecto al tamaño que se está midiendo, y se representan así: (Hernández , et al., 2014).

- **Límite de acuerdo superior:** diferencia media + 1,96·SD.
- **Diferencia media:** valor medio determinado por A – valor medio determinado por B. Refleja el error sistemático, en contraposición con la precisión (SD y CI) que refleja el error aleatorio.
- **Límite de acuerdo inferior:** diferencia media - 1,96·SD.

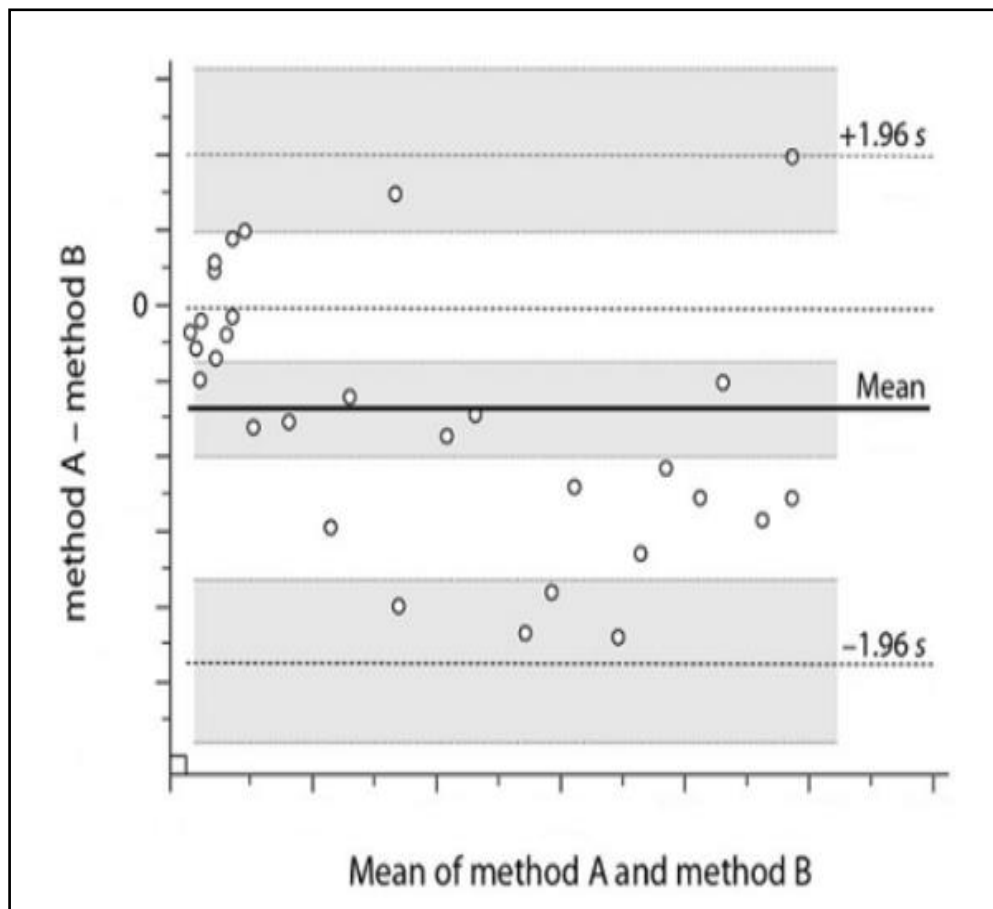


Figura 8-1: Gráfico de Bland – Altman.
Fuente: (Martínez, 2021).

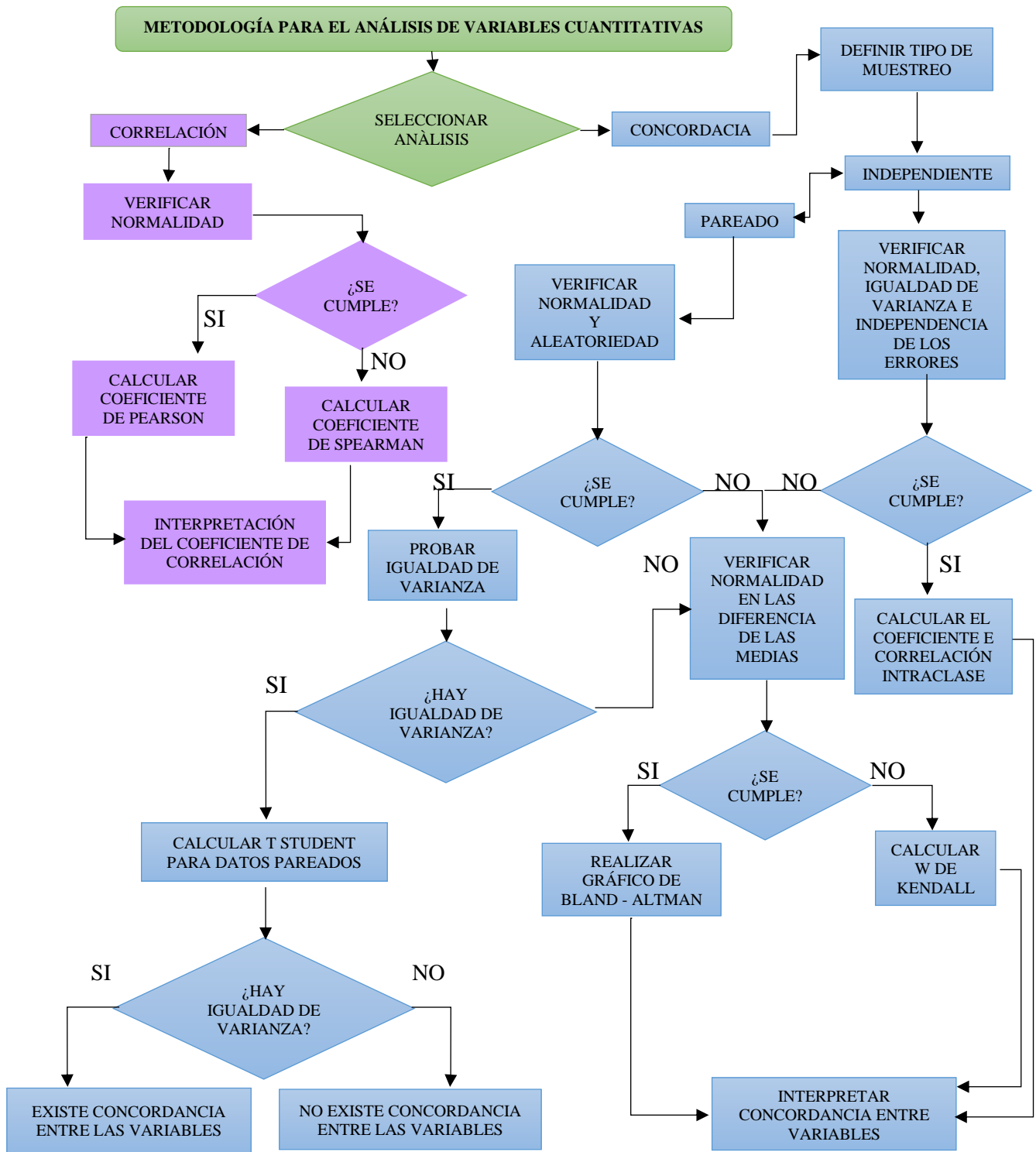


Figura 9-1: Análisis de correlación y concordancia en variables cuantitativas.

Fuente: (Hernández , et al., 2014).

Serie temporal

Las series temporales se pueden definir como sucesiones de los valores que una variable adquiere al pasar del tiempo o un tiempo determinado de forma que conserven la misma distancia entre elementos, precisando la utilización del plano cartesiano donde en el eje de las x se pondrá el tiempo y en el eje de las ordenadas las series observadas (García Díaz, 2016, p. 16).

Componentes de una serie temporal

Los componentes de las series temporales es el fruto del comportamiento de cuatro componentes fundamentales, la tendencia, estacionalidad, aleatoriedad y las variaciones cíclicas (García Díaz, 2016, p. 17).

Tendencia (T). - Estas componentes de las series reflejan su comportamiento a largo plazo, pueden ser estacionarias y de naturaleza lineal, exponencial.

Componente cíclica (C). – Son componentes periódicas de amplitud de un año, no son regulares y se dan de forma alternativa.

Componente estacional (S). - Son serie que se dan en periodos iguales o menos a un año, en series mensuales, semanales, diarias u horarias.

Componente aleatorio o irregular (I). – Son componentes de series temporales de eventos imprevisibles, siendo que son previsibles estas actúan en cualquier serie ya sea en mayor o menor medida.

1.2. Bases teóricas.

Hidrógeno

El hidrógeno es un elemento de que relaciona directamente con la energía, tanto en la física como en la biología. Al aplicar una pizca de fuego este arde con total facilidad, denotando la energía de la potencia del hidrógeno (Mendoza Medellín, 2006, p. 2).

Lodos Residuales

Los lodos residuales se les llama en un proceso de tratamiento de aguas como sólidos, estos microorganismos son materia de organismos residuales, compuestos no biodegradables y/o potencialmente tóxicos que han cambiado en el tratamiento (Ortiz, et al., 1995, p. 3).

Fermentación oscura

La fermentación oscura son residuos que se refieren a la obtención de este gas a través de microorganismos los cuales son capaces de convertir la materia orgánica a ácidos y alcoholes con una liberación simultanea de hidrógeno molecular (Blanco Londoño, 2013, p. 116).

Electrólisis

Los microorganismos no asimilan la materia orgánica particulada, la primera fase consiste (polímeros), (moléculas menores), las cuales atraviesan las paredes fermentativas. Esta conversión se consigue a través de la adición de enzimas excretadas por las bacterias hidrolíticas (Blanco Londoño, 2013, p. 118).

Sensor de Metano

Los sensores nos permiten medir y cuantificar las magnitudes físicas. Cada vez se desarrollan más sensores que nos permiten medir con mayor exactitud, o detectar componentes que serían letales para las personas. Los sensores en el sector industrial, se utilizan para el monitoreo en procesos, reduciendo pérdidas y contribuyendo a su automatización (Blanco Londoño, 2013, p. 72).

CAPITULO II

2. MARCO METODOLÓGICO

2.1. Tipo de investigación

Esta investigación es de tipo cuantitativo que se caracteriza por un proceso ordenado en etapas:

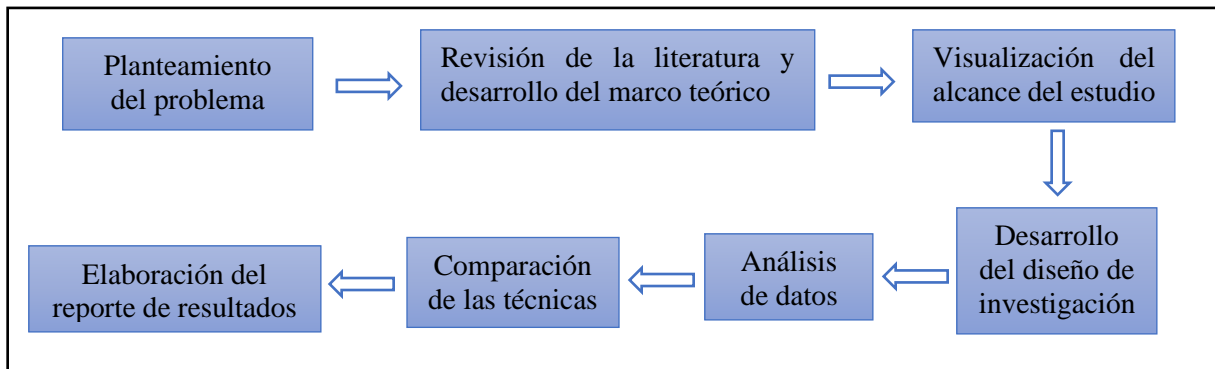


Figura 1-2: Proceso cuantitativo.

Fuente: (Hernández, et al., 2014).

Planteamiento del problema

El Ecuador intenta atenuar los efectos contaminantes que genera la minería dentro de zonas como Ponce Enríquez, sin embargo, no se ha considerado el análisis estadístico para la toma de decisiones. Se desconoce el poder de los datos y por ello no se recolecta información de las áreas mineras que ayudarían a mitigar la contaminación mediante los procesos: cuantificar, analizar, comparar e interpretar las variables de interés.

Revisión de la literatura y desarrollo del marco teórico

En este estudio se ha considerado fuentes primarias como: artículos científicos de alto, medio y bajo impacto, de la misma manera fuentes secundarias como: libros de Estadísticas, Matemática, Programación en RStudio y Metodologías de la Investigación.

Visualización del alcance del estudio

El nivel de alcance de esta investigación es exploratorio, ya que se busca realizar una comparación estadística en las técnicas fermentación oscura y electrolisis microbiana, para la producción de hidrógeno a partir de lodos residuales.

Desarrollo del diseño de investigación

El grado de manipulación de las variables en la investigación es experimental, ya que las variables fueron manipuladas en base a un tratamiento con dos niveles (fermentación oscura y electrólisis microbiana). Esta investigación es transversal debido a que la información fue obtenida en un lapso corto de 42 días.

Análisis de datos

Para analizar los datos en esta investigación se utilizó el lenguaje de programación RStudio. En el pre procesamiento de los datos se ha realizado una estadística descriptiva mediante gráficas y métricas que ha permitido identificar datos anómalos y faltantes, se ha realizado la imputación múltiple para completar los datos. Para la comparativa de técnica se ha utilizado las gráficas Bland Altman y series temporales.

Comparación de las técnicas

Se compara las técnicas mediante gráficas y métricas estadísticas, de la misma manera, se emplea estadísticos de prueba como test de Mann–Whitney.

Elaboración del reporte de resultados

Para el reporte de resultados se realizó un análisis crítico de las gráficas y métricas estadísticas para redactar los resultados obtenidos.

Diseño de la investigación

El diseño de la investigación es experimental debido a que los datos han sido recolectados mediante sensores que han medido y guardado la información por horas de un experimento en donde se han considerado cuatro reactores y se aplicado el tratamiento: fermentación oscura y electrólisis microbiana para medir la producción de hidrógeno.

2.2. Localización de estudio

Croquis de ubicación / Dirección

La investigación se desarrolló de dos formas: In situ: Se recogió muestras de lodos residuales se recogió del río Chico en un punto de coordenadas UTM 9661525 N y 0643750 E, a 15 km aguas abajo de una zona minera Barranco Colorado ubicada en la vía a Shumiral perteneciente al cantón Camilo Ponce Enríquez, ciudad del Azuay. Con una estación meteorológica-portátil de la marca Kestrel se determinó la altitud (43 m.s.n.m) y los siguientes parámetros ambientales: temperatura

(25,2 °C), velocidad máxima del viento (1,48 km/h), presión (1 atm) y humedad (100%). En estas fuentes hídricas acumulando una gran cantidad de lodos residuales contaminados con metales pesados, derivados de la actividad minera cuyo principal objetivo es la extracción de oro, (Méndez Cuzco, 2020, p. 22).

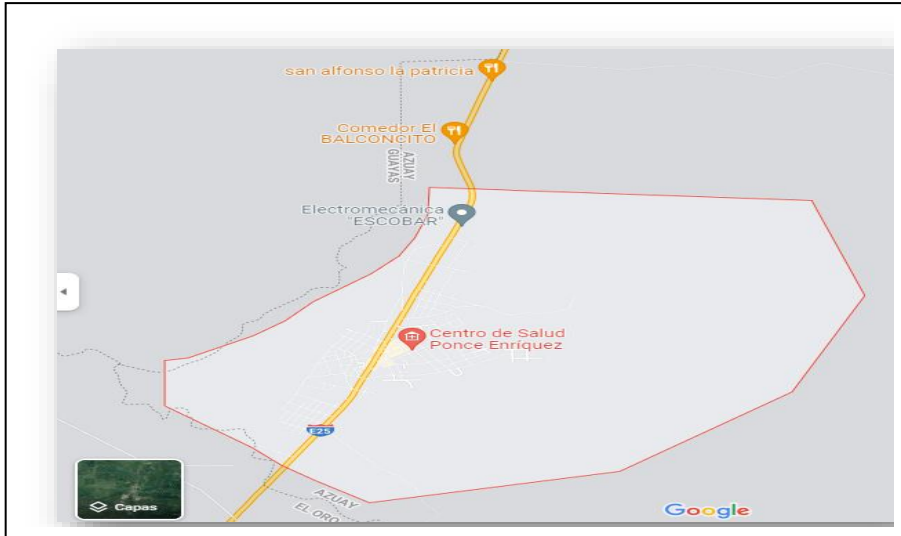


Figura 2-2: Cantón Ponce Enríquez.

Fuente: Google maps.

2.2.1. Población de estudio

Los datos fueron obtenidos por (Méndez Cuzco, 2020, p. 23), los cuales arrojaron que los sedimentos con lodos residuales del río Chico muestreados de una zona cercana a una mina en la parroquia Shumiral del cantón Ponce Enríquez de la ciudad del Azuay, que fueron empleados como sustrato en reactores de fermentación oscura y electrólisis microbiana donde se monitoreó la generación de hidrógeno desarrollados en el GEAA de la ESPOCH.

2.2.2. Tamaño de la muestra

No se usará ningún método de muestreo ya que se trabajará con toda la población.

2.2.3. Identificaciones variables

Variable Independiente:

- Técnica (fermentación oscura y electrólisis microbiana)

Variable Dependiente:

- Hidrógeno

2.2.4. Planteamiento de la hipótesis

Las hipótesis de la investigación se plantearán a continuación:

Existe una diferencia significativa de la producción de hidrógeno según la técnica.

Existe una diferencia significativa de la producción de hidrógeno según el tiempo.

2.3. Variables en estudio

2.3.1. Operacionalización de las variables

(conceptualiza la variable, colocar indicadores y los instrumentos de medición)

Tabla 1- 2: Operacionalización de variables.

VARIABLE	CONCEPTO	INDICADOR	INSTRUMENTO
VARIABLE INDEPENDIENTE:	Técnica (fermentación oscura y electrólisis microbiana)	- Estadística Descriptiva	- Software RStudio - Computador
VARIABLE DEPENDIENTE:	Hidrógeno	- Estadística Descriptiva - Análisis comparativo gráfico de Bland-Altman - Análisis de Estadístico Inferencial paramétrico o no paramétrico	- Software RStudio - Computador

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

CAPÍTULO III

3. MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS

Se describe el experimento previo a la toma se ha empleado el primer tratamiento, la técnica fermentación oscura (FO) en los 4 reactores por tiempos cortos de (10 y 11 días) y largos (21 días), luego se sometieron a electrolisis microbiana (EM) con los mismos periodos de tiempos, con un tiempo total de 42 días de experimentación, (Méndez Cuzco, 2020, p. 53).

Tabla1-3: Diseño de las técnicas FO y EM.

Tiempo (días)	Reactor 1	Reactor 2	Reactor 3	Reactor 4
10	FO	FO	FO	FO
11	FO	FO	EM	EM
10	EM	EM	FO	FO
11	EM	EM	EM	EM

Fuente: Grupo de Alternativas y Ambiente (GEAA), (Méndez Cuzco, 2020).

Elaborado por: Olmedo, Beatriz, 2022.

3.1. Depuración de la base de datos

Patrón general (listwise): Este método ha ayudado a identificar los valores faltantes de forma sistemática. Los cuales ha sido omitidos de la base de datos.

Tabla 2-3: Valores Omitidos de la base por el método de (listwise)

Fecha	Observación
29/11/2019	Del día 1 se omitieron las horas 13 a la 18 por no existir información.
2/12/2019	Del día 4 se omitieron las horas 15 a la 17 por no existir información.
8/12/2019	Del día 10 se omitieron las horas 17 a la 23 por no existir información.
9/12/2019	Del día 11 se omitieron las horas 0 a la 7 por no existir información.
11/12/2019	Del día 13 se omitieron las horas 15 a la 23 por no existir información.
12/12/2019	Del día 14 se omitieron las horas 0 a la 7 y la hora 13 igualmente por no existir información.
19/12/2019	Del día 21 se omitieron la hora 9 por no existir información.

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

De un total de 1008 muestras recolectadas en el transcurso de los 42 días, se redujo a 965 muestras que corresponden al 96% de la información, el 4% restante fueron retiradas de posteriores tablas.

3.2. Imputación de la base datos

3.2.1. Imputación múltiple

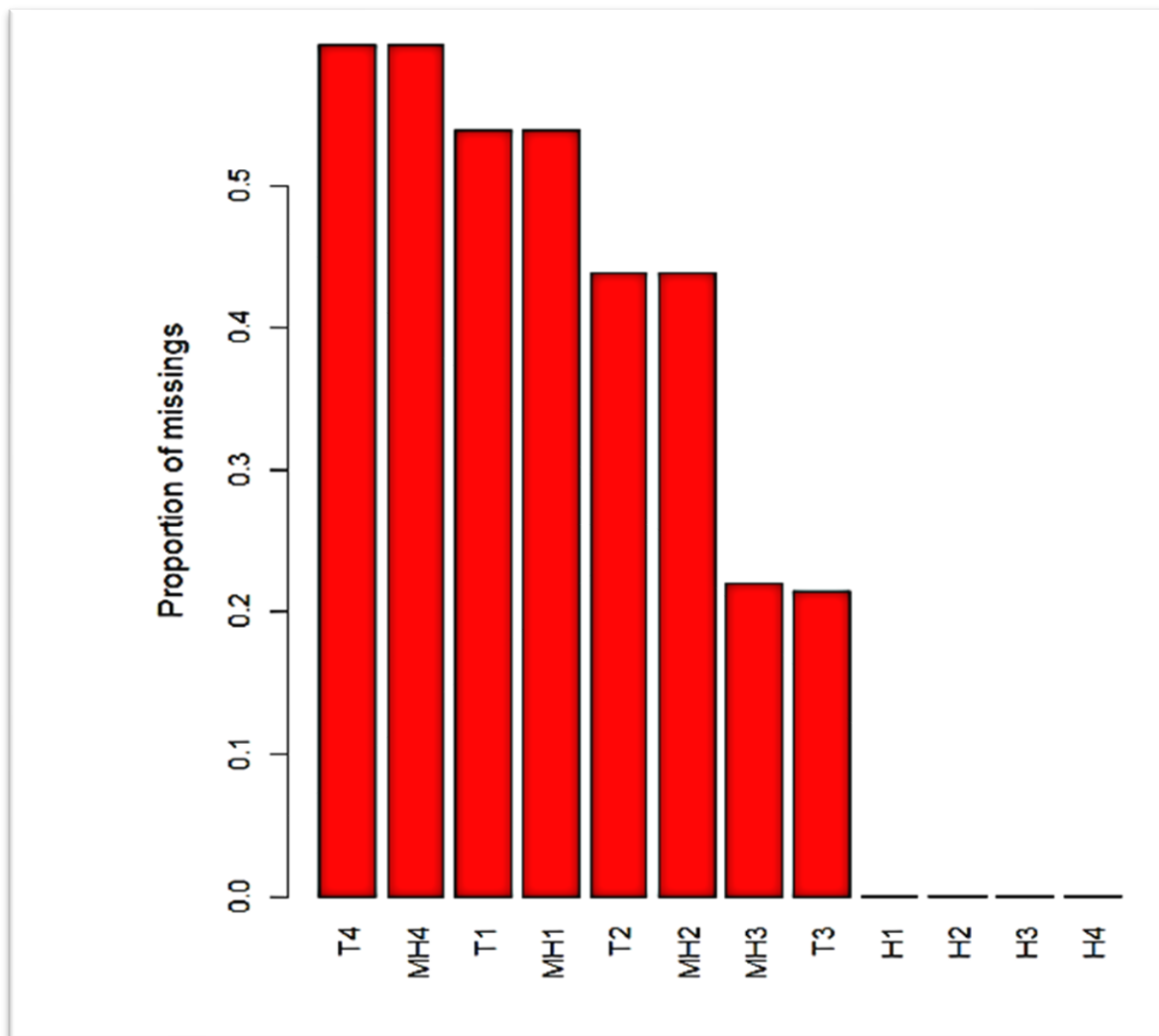


Gráfico 1-3: Variables con valores faltantes.

Elaborado por: Olmedo, Beatriz, 2022.

Tabla 3-3: Porcentaje de datos faltantes.

Variable	Count
T4	0.5989637
MH4	0.5989637
T1	0.5388601
MH1	0.5388601
T2	0.4383420
MH2	0.4383420
MH3	0.2196891
T3	0.2145078
H1	0.0000000
H2	0.0000000
H3	0.0000000
H4	0.0000000

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

En la gráfica 1-3 se observa el porcentaje de los valores faltantes de cada variable. La variable T4 (temperatura 4) y MH4 (humedad 4) tienen 59.89% de datos faltantes, así mismo, la variable T1 (temperatura 1) con el 53.88%, variable MH1 (humedad 1) tiene el 53.89%, T2 y MH2 (temperatura 2 y humedad 2) tienen el 43.83%, MH3 (humedad 3) tiene el 21.97%, T3 (temperatura 3) tiene el 21.45%. Las variables H1, H2, H3, y H4 (hidrógeno) no tienen datos faltantes.

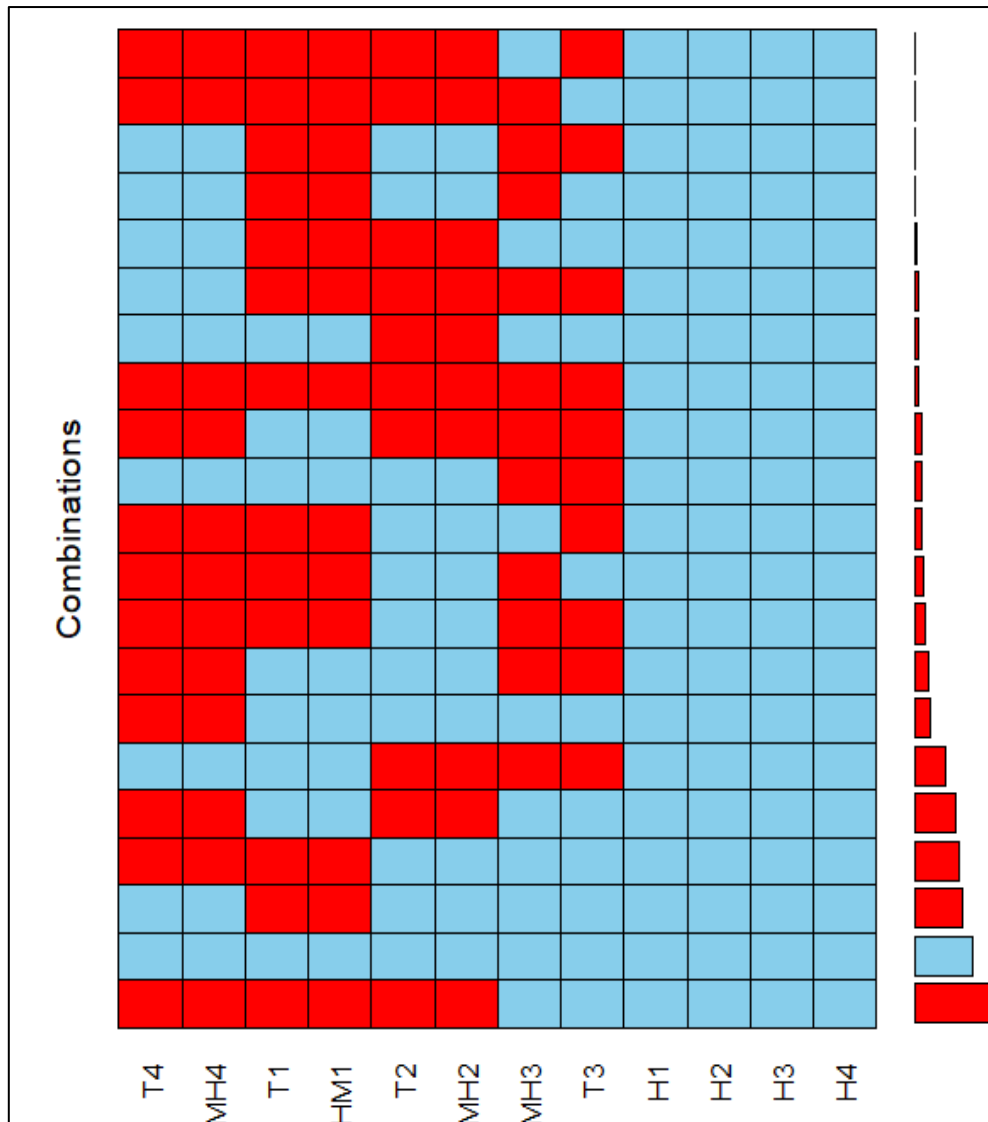


Gráfico 2-3: Porcentajes de valores faltantes.

Elaborado por: Olmedo, Beatriz, 2022

3.3. Análisis de datos Atípicos

Las variables H1, H2, H3, y H4 (hidrógeno) no obtuvieron datos faltantes por lo cual se procedió a realizar el análisis de datos atípicos univariante mediante la prueba Tukey ajustada que identifico 241 datos atípicos que fueron separados del conjunto de datos.

3.3.1. Prueba analítica Tukey

3.3.1.1. Reactor 1

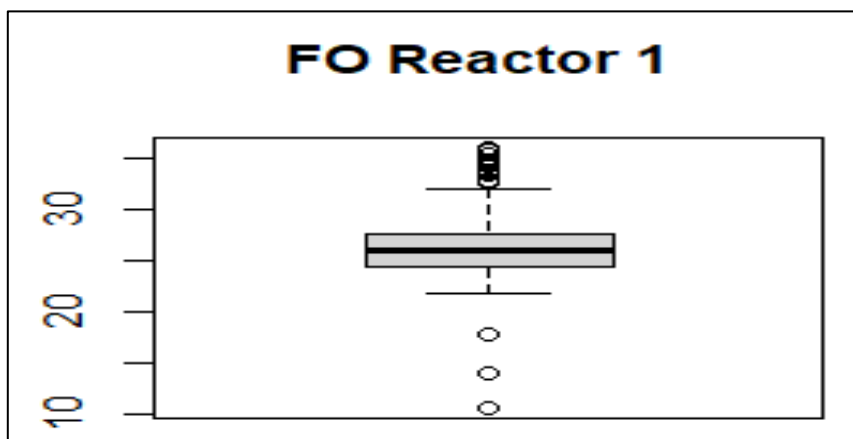


Gráfico 3-3: Identificación de atípicos Reactor 1 FO.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Mediante la prueba analítica de Tukey se han detectado 29 datos atípicos en la variable hidrógeno 1, dentro de la técnica fermentación oscura (FO), se corrobora esta información en el diagrama de cajas.

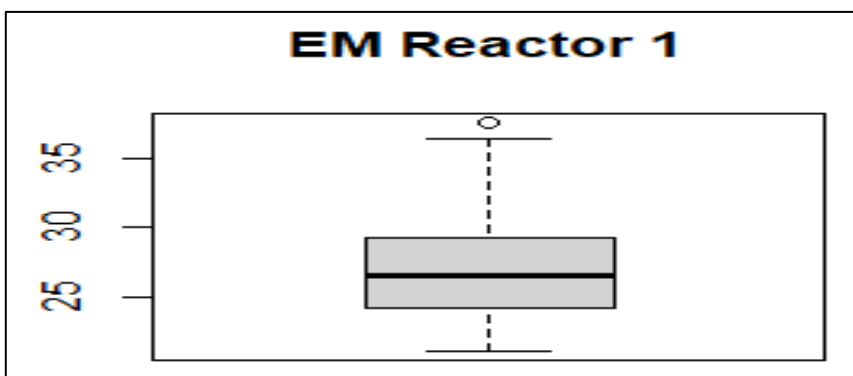


Gráfico 4-3: Identificación de atípicos Reactor 1 EM.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Mediante la prueba analítica de Tukey se han detectado 5 datos atípicos en la variable hidrógeno 1, dentro de la técnica electrolisis microbiana (EM), se corrobora esta información en el diagrama de cajas.

3.3.1.2. Reactor 2

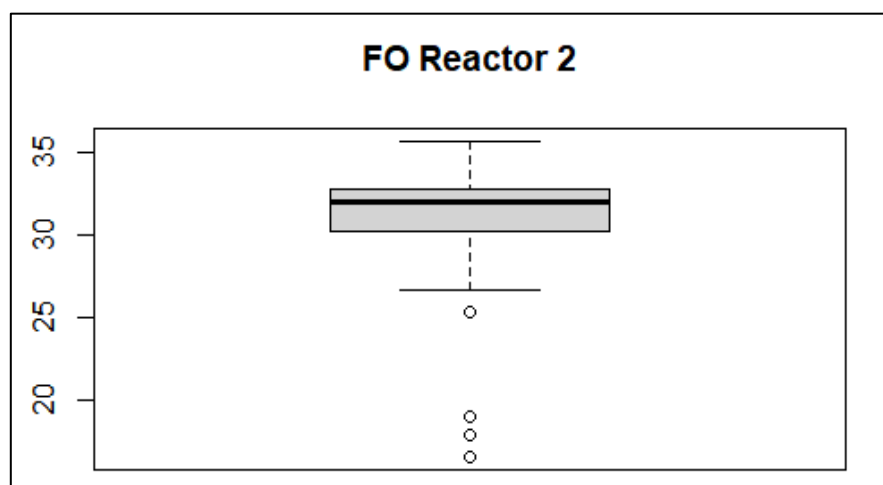


Gráfico 5-3: Identificación de atípicos Reactor 2 FO.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Mediante la prueba analítica de Tukey se han detectado 21 datos atípicos en la variable hidrógeno 2, dentro de la en la técnica fermentación oscura (FO), se corrobora esta información en el diagrama de cajas.

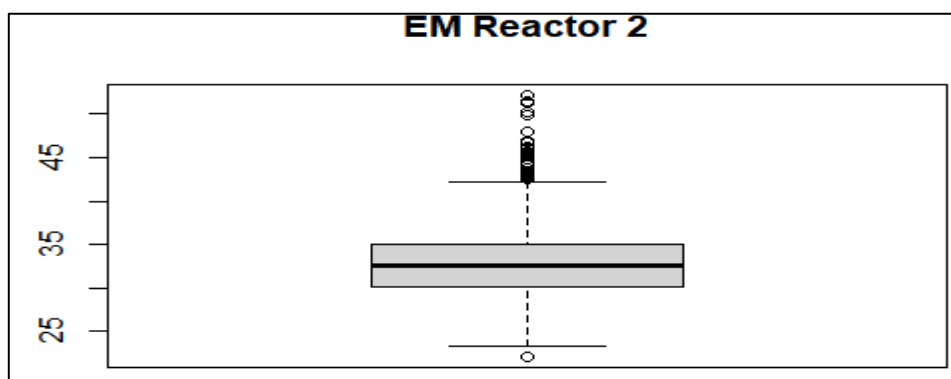


Gráfico 6-3: Identificación de atípicos Reactor 2 EM.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Mediante la prueba analítica de Tukey se han detectado 96 datos atípicos en la variable hidrógeno 2, dentro de la técnica electrolisis microbiana (EM), se corrobora esta información en el diagrama de cajas.

3.3.1.3. Reactor 3

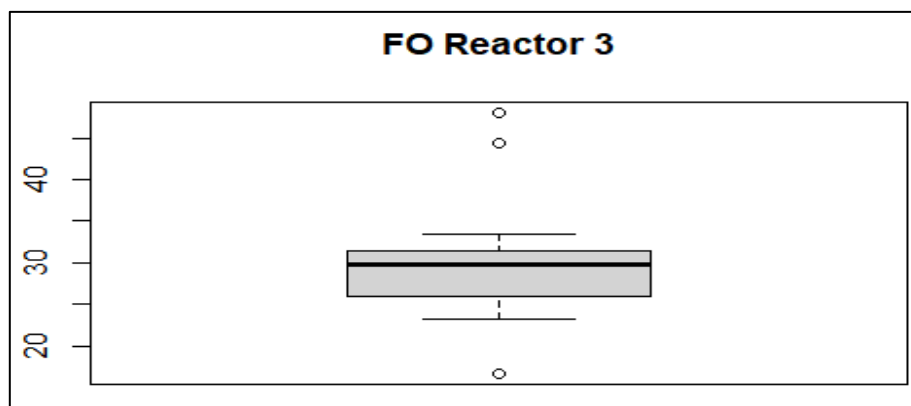


Gráfico7-3: Identificación de atípicos Reactor 3 FO.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Mediante la prueba analítica de Tukey se han detectado 3 datos atípicos en la variable hidrógeno 3, dentro de la técnica fermentación oscura (FO), se corrobora esta información en el diagrama de cajas.

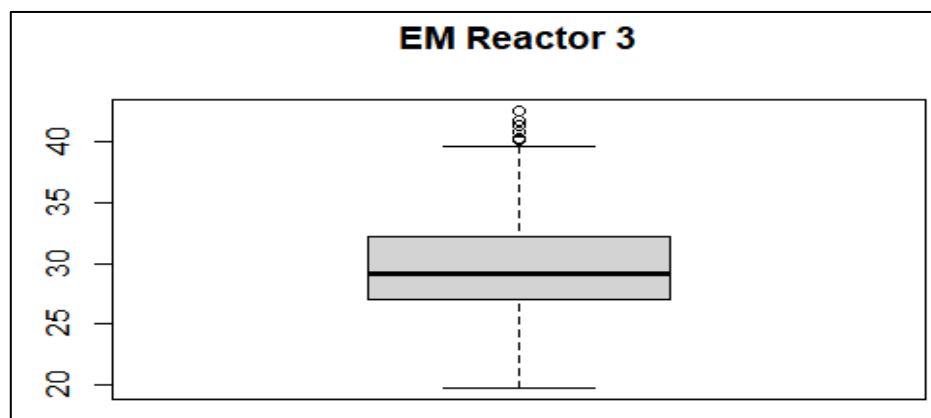


Gráfico 8-3: Identificación de atípicos Reactor 3 EM.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Mediante la prueba analítica de Tukey se han detectado 18 datos atípicos en la variable hidrógeno 3, dentro de la técnica electrolisis microbiana (EM), se corrobora esta información en el diagrama de cajas.

3.3.1.4. Reactor 4

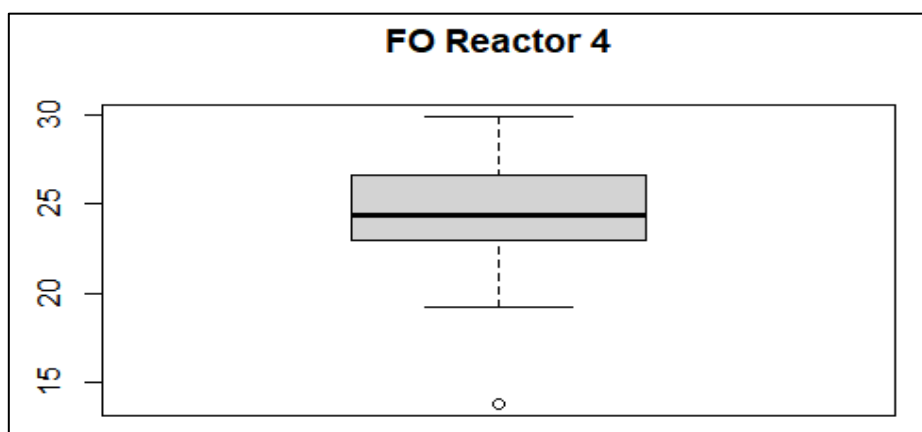


Gráfico 9-3: Identificación de atípicos Reactor 4 FO.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Mediante la prueba analítica de Tukey se han detectado 2 datos atípicos en la variable hidrógeno 4, dentro de la técnica fermentación oscura (FO), se corrobora esta información en el diagrama de cajas.

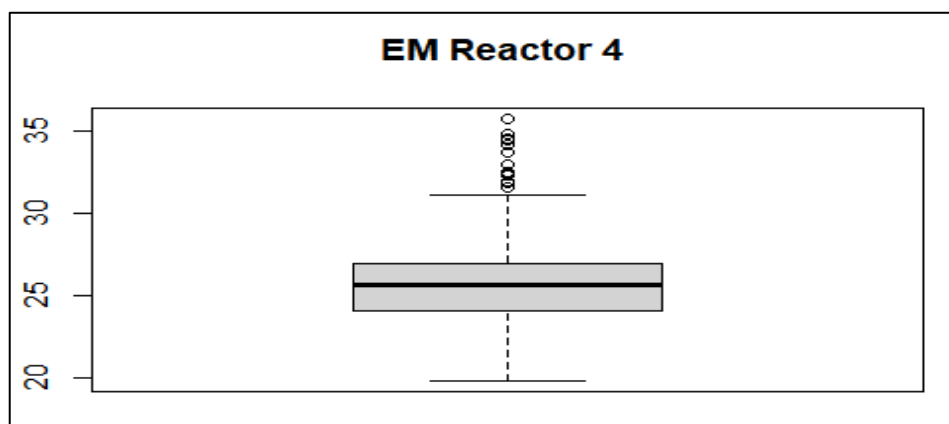


Gráfico 10-3: Identificación de atípicos Reactor 4 EM.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Mediante la prueba analítica de Tukey se han detectado 67 datos atípicos en la variable hidrógeno 4, dentro de la técnica electrolisis microbiana (EM), se corrobora esta información en el diagrama de cajas.

Los datos atípicos encontrados que corresponden al 24% del conjunto total de datos.

3.4. Análisis descriptivo de las variables

Para el análisis descriptivo se ha realizado gráficas y estadísticas:

Tabla 4-3: Hidrógeno 1.

<i>H1</i>	
Media	26,2976814
Error típico	0,08242275
Mediana	26,0899302
Desviación estándar	2,42272419
Varianza de la muestra	5,86959248
Curtosis	-0,09329086
Coefficiente de asimetría	0,58619685
Rango	13,6633495
Mínimo	21,1645287
Máximo	34,8278783
Suma	22721,1967
Cuenta	864

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

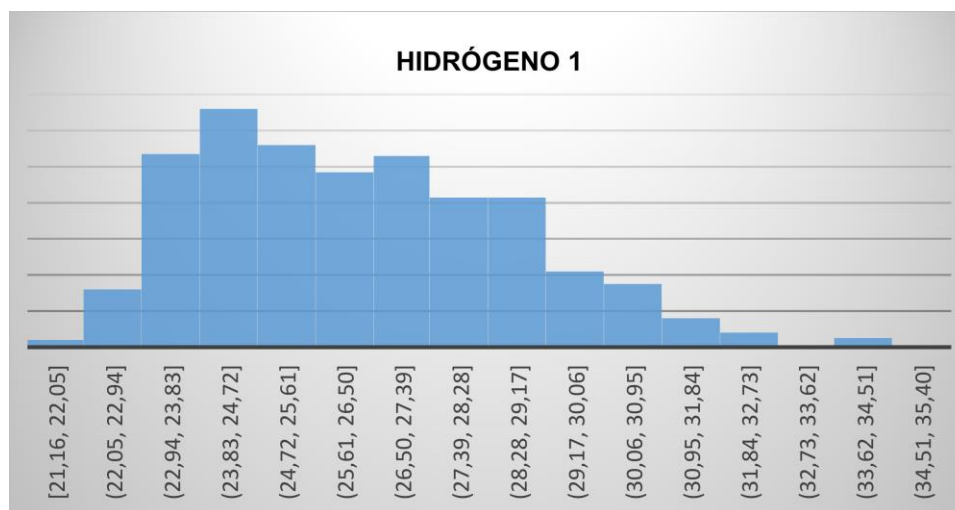


Gráfico 11-3: Histograma variable Hidrógeno 1.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Hidrógeno 1, tiene mediana de 26.08 y varianza de 5.86, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [23.83; 24.72] de un total de 864 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 5-3: Hidrógeno 2.

H2	
Media	32,495638
Error típico	0,09652813
Mediana	32,4771348
Desviación estándar	2,43243234
Varianza de la muestra	5,91672707
Curtosis	1,1438215
Coefficiente de asimetría	0,06145978
Rango	13,700941
Mínimo	25,7989607
Máximo	39,4999017
Suma	20634,7302
Cuenta	635

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

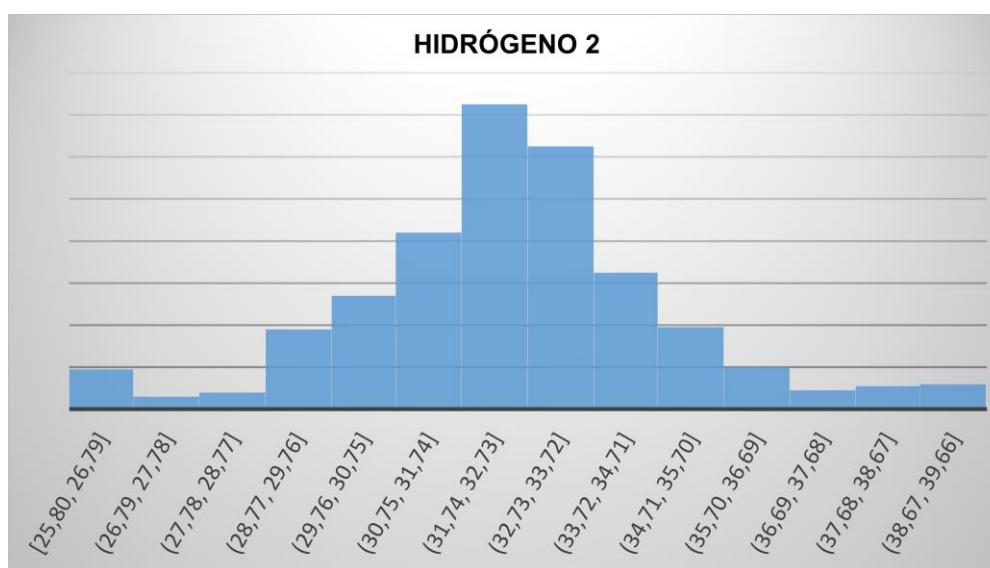


Gráfico 12-3: Histograma variable Hidrógeno 2.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Hidrógeno 2, tiene mediana de 32.47 y varianza de 5.91, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [31.74; 33.72], de un total de 635 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 6-3: Hidrógeno 3.

H3	
Media	30,0054267
Error típico	0,11305437
Mediana	30,3755164
Desviación estándar	2,88897619
Varianza de la muestra	8,34618345
Curtosis	1,67471518
Coefficiente de asimetría	0,17105278
Rango	24,8347981
Mínimo	23,1652019
Máximo	48
Suma	19593,5437
Cuenta	653

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

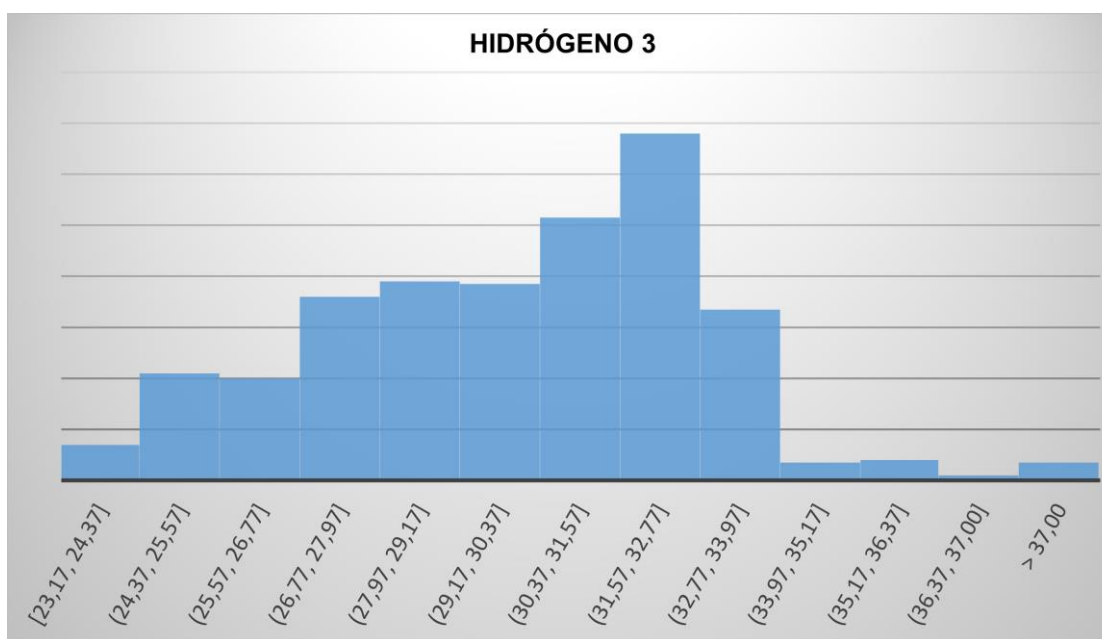


Gráfico 13-3: Histograma variable Hidrógeno 3.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Hidrógeno 3, tiene mediana de 30.37 y varianza de 2.88, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [31.57; 32.77] de un total de 653 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 7-3: Hidrógeno 4.

H4	
Media	24,7871776
Error típico	0,08007781
Mediana	24,7453367
Desviación estándar	2,04786383
Varianza de la muestra	4,19374628
Curtosis	-0,64225176
Coefficiente de asimetría	0,00817218
Rango	10,5291711
Mínimo	19,1942425
Máximo	29,7234136
Suma	16210,8141
Cuenta	654

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

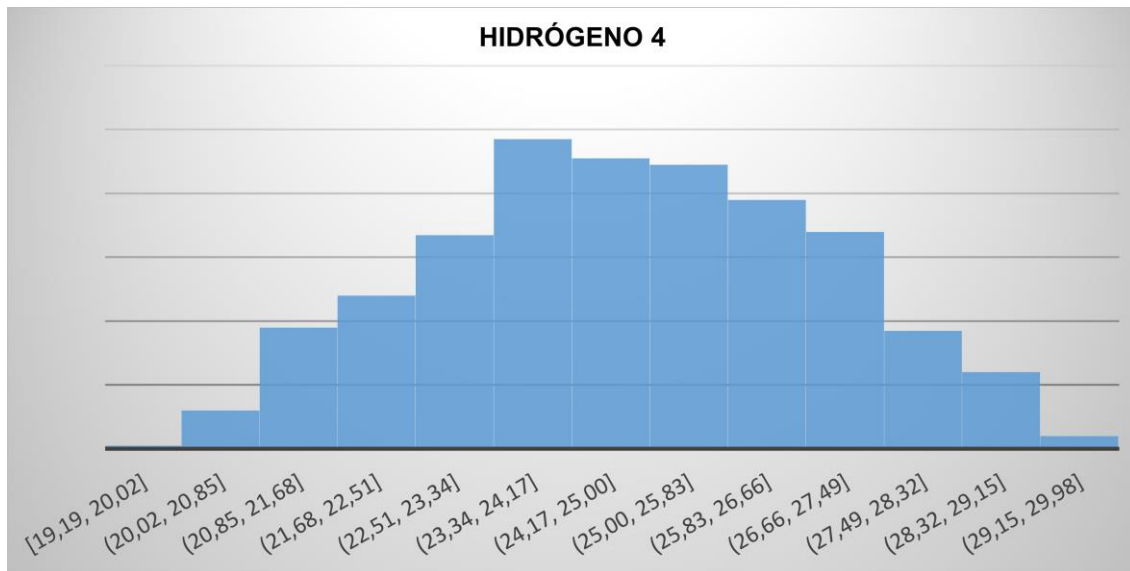


Gráfico 14-3: Histograma variable Hidrógeno 4.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Hidrógeno 4, tiene mediana de 24.74 y varianza de 4.19, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [23.34; 24.17] de un total de 654 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 8-3: Temperatura 1.

T1	
Media	24,7724013
Error típico	0,18214346
Mediana	22,9220339
Moda	20,4722222
Desviación estándar	5,65818624
Varianza de la muestra	32,0150715
Curtosis	1,66588039
Coefficiente de asimetría	1,32243069
Rango	31,3435088
Mínimo	17,16
Máximo	48,5035088
Suma	23905,3673
Cuenta	965

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

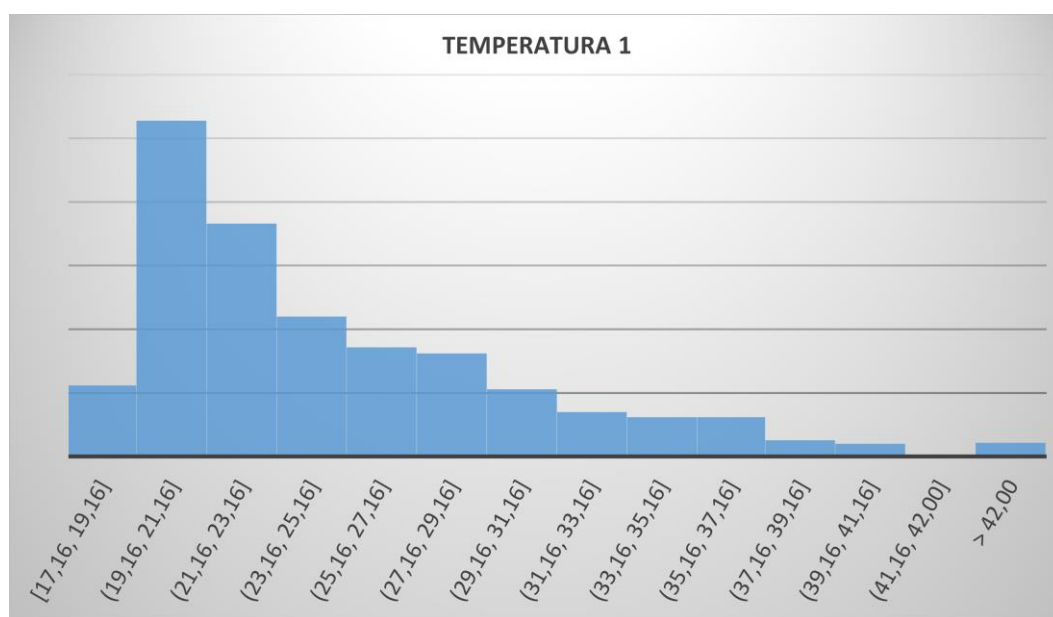


Gráfico 15-3: Histograma variable Temperatura 1.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Temperatura 1, tiene mediana de 22.92 y varianza de 32.01, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [19,16; 21,16] de un total de 965 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 9-3: Temperatura 2.

T2	
Media	24,177211
Error típico	0,18844524
Mediana	23
Moda	19
Desviación estándar	5,85394767
Varianza de la muestra	34,2687033
Curtosis	0,67696925
Coefficiente de asimetría	0,99412546
Rango	29,2220339
Mínimo	14
Máximo	43,2220339
Suma	23331,0086
Cuenta	965

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

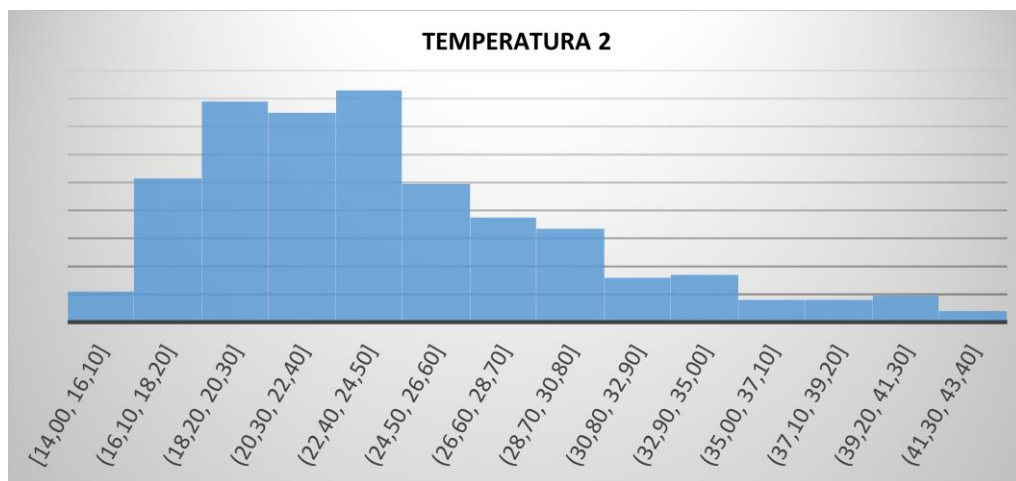


Gráfico 16-3: Histograma variable Temperatura 2.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Temperatura 2, tiene mediana de 23 y varianza de 34.26, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [22,30; 22.40] de un total de 965 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 10-3: Temperatura 3.

T3	
Media	25,348658
Error típico	0,17299112
Mediana	23,7016667
Moda	30,3254237
Desviación estándar	5,37387397
Varianza de la muestra	28,8785215
Curtosis	1,37855953
Coefficiente de asimetría	1,27139415
Rango	27,1522775
Mínimo	18,1508475
Máximo	45,303125
Suma	24461,455
Cuenta	965

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

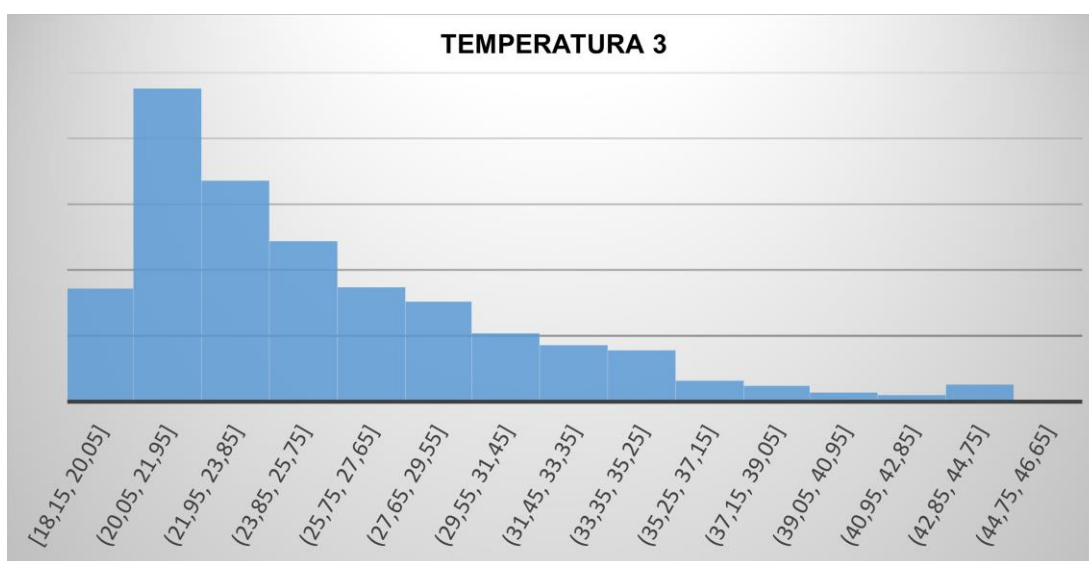


Gráfico 17-3: Histograma variable Temperatura 3.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Temperatura 3, tiene mediana de 23.79 y varianza de 28,76, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [20,05; 21.95] de un total de 965 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 11-3: Temperatura 4.

T4	
Media	24,9154499
Error típico	0,18223087
Mediana	23,2372881
Moda	19,1101695
Desviación estándar	5,66090153
Varianza de la muestra	32,0458061
Curtosis	1,37169653
Coefficiente de asimetría	1,23434351
Rango	31,9444209
Mínimo	15,8708333
Máximo	47,8152542
Suma	24043,4091
Cuenta	965

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

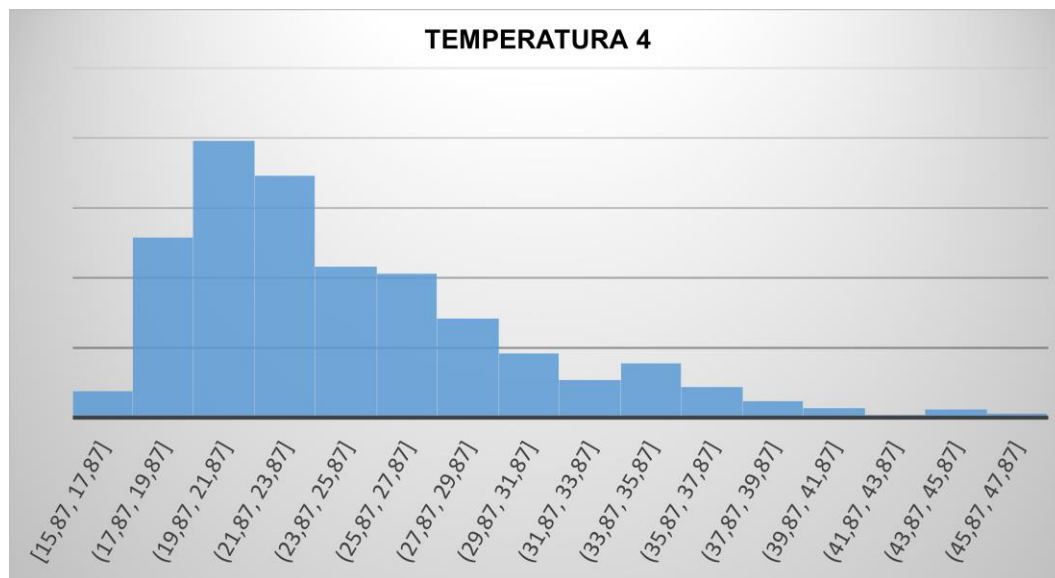


Gráfico 18-3: Histograma variable Temperatura 4.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Temperatura 4, tiene mediana de 23.23 y varianza de 32.04, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [19.87; 21.87] de un total de 965 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 12-3: Humedad 1.

MH1	
Media	93,3980981
Error típico	0,06942142
Mediana	94
Moda	95
Desviación estándar	2,15653822
Varianza de la muestra	4,65065709
Curtosis	5,08497644
Coficiente de asimetría	-2,01763888
Rango	13,5333333
Mínimo	81,4666667
Máximo	95
Suma	90129,1646
Cuenta	965

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

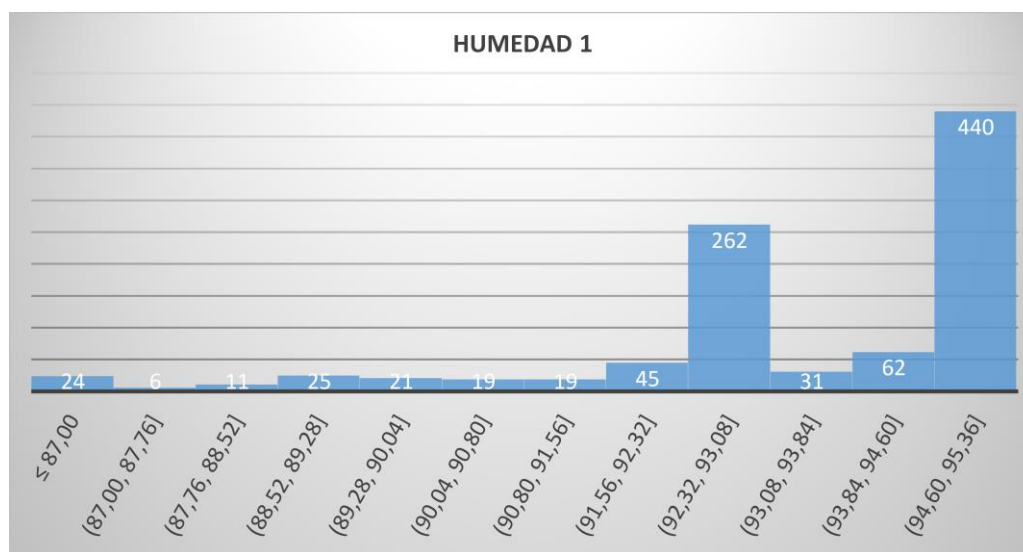


Gráfico 19-3: Histograma variable Humedad 1.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Humedad 1, tiene mediana de 94 y varianza de 4.65, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [94,60; 95,36] de un total de 965 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 13-3: Humedad 2.

MH2	
Media	93,0940687
Error típico	0,08437967
Mediana	94
Moda	95
Desviación estándar	2,62120811
Varianza de la muestra	6,87073194
Curtosis	5,78843293
Coficiente de asimetría	-2,13507838
Rango	15,65
Mínimo	79,35
Máximo	95
Suma	89835,7763
Cuenta	965

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

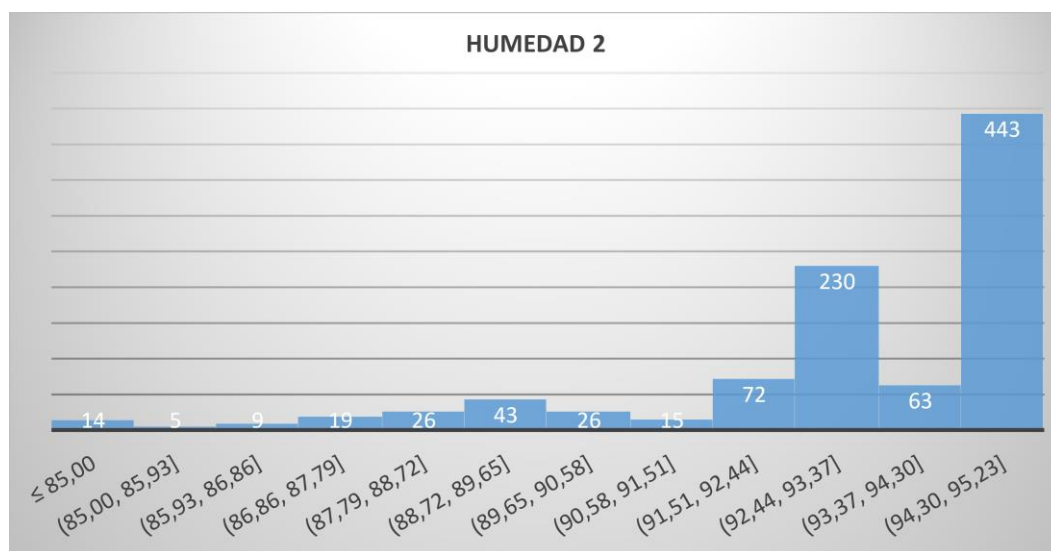


Gráfico 20-3: Histograma variable Humedad 2.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Humedad 2, tiene mediana de 94 y varianza de 6.87, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [94,30; 95.23] de un total de 965 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 14-3: Humedad 3.

MH3	
Media	91,5240353
Error típico	0,1674009
Mediana	94
Moda	95
Desviación estándar	5,20021685
Varianza de la muestra	27,0422553
Curtosis	1,19812002
Coficiente de asimetría	-1,51823456
Rango	26,6333333
Mínimo	68,3666667
Máximo	95
Suma	88320,6941
Cuenta	965

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

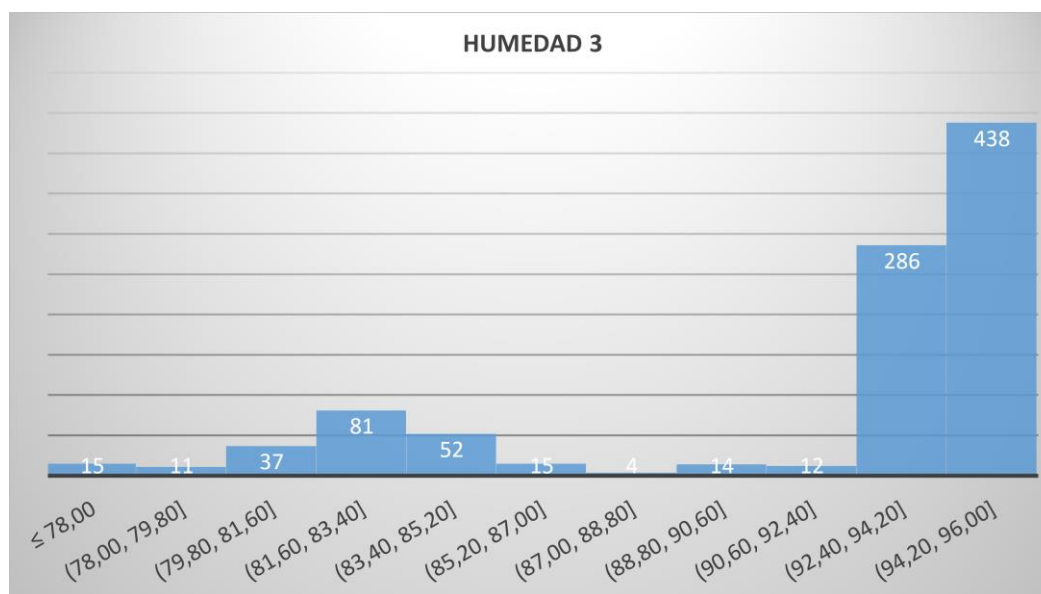


Gráfico 21-3: Histograma variable Humedad 3.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Humedad 3, tiene mediana de 94 y varianza de 27.04 se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [94,20; 96,00] de un total de 965 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

Tabla 15-3: Humedad 4.

MH4	
Media	93,0504568
Error típico	0,078179
Mediana	93
Moda	93
Desviación estándar	2,42858769
Varianza de la muestra	5,89803817
Curtosis	16,6879109
Coficiente de asimetría	-3,53946368
Rango	18,2068966
Mínimo	76,7931034
Máximo	95
Suma	89793,6908
Cuenta	965

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

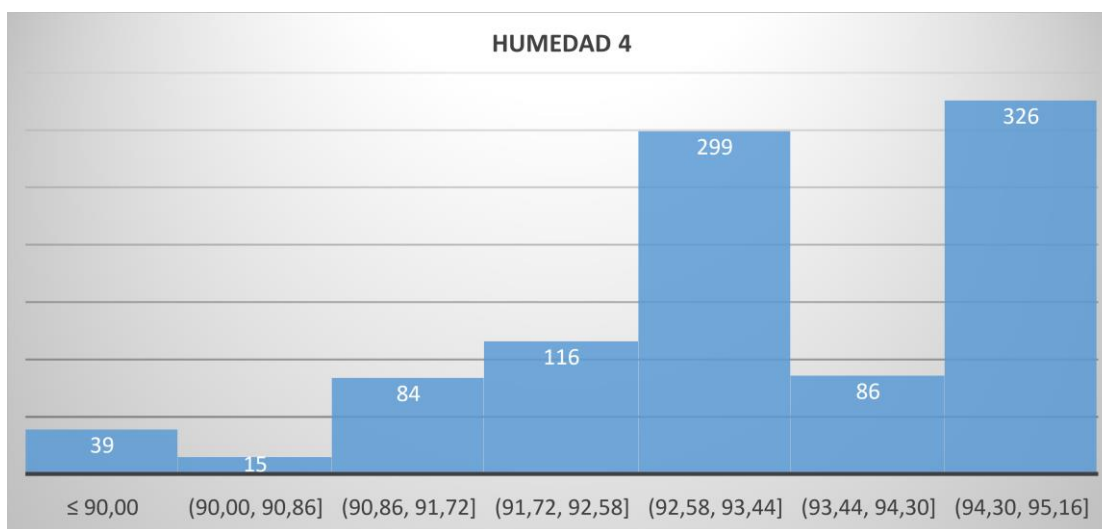


Gráfico22-3: Histograma variable Humedad 4.

Elaborado por: Olmedo, Beatriz, 2022.

La variable Humedad 4, tiene mediana de 93 y varianza de 5.89, se ha observado en la gráfica que la mayor frecuencia obtuvo un intervalo [94,30; 95.16] de un total de 965 datos de los lodos residuales irradiados tomadas de la zona minera en Ponce Enríquez, el histograma de frecuencia muestra que los datos no tienen una Distribución Normal.

3.5. Análisis Estadístico Inferencial

3.5.1. Aplicación del Test Mann Whitney Wilcoxon

El análisis estadístico inferencial utilizó la prueba no paramétrica de Mann Whitney para comparar las técnicas FO y EM.

3.5.1.1. Prueba de Normalidad

Tabla 16-3: Test de Normalidad (FO).

Lilliefors (Kolmogorov-Smirnov) normality test				
data:	FO-H1	FO-H2	FO-H3	FO-H4
D =	0.12053	0.12782	0.19468	0.14908
p-value =	2.217e-06	2.692e-07	1.15e-09	0.003328

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Los valores p del estadístico Kolmogorov Smirnov con la corrección de Lilliefors son menores a $\alpha = 0.05$; y se rechaza H_0 , por lo que se concluye que los datos no siguen una Distribución Normal para la técnica FO.

Tabla 17-3: Test de Normalidad (EM).

Lilliefors (Kolmogorov-Smirnov) normality test				
data:	EM-H1	EM-H2	EM-H3	EM-H4
D =	0.27746	0.22292	0.2943	0.23875
p-value =	2.576e-07	6.96e-05	6.929e-05	0.1956

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Los valores p del estadístico Kolmogorov Smirnov con la corrección de Lilliefors son menores a $\alpha = 0.05$; y se rechaza H_0 , por lo que se concluye que los datos no siguen una Distribución Normal para la técnica EM.

3.5.2. Test de Varianza Constante.

Tabla 18-3: Test de Varianza Constante.

Fligner-Killeen test of homogeneity of variances (FO, EM)				
data:	H1	H2	H3	H4
Fligner-Killeen:	Med	Med	Med	Med
chi-squared =	58.061	8.7877	4.8579	27.017
Df	1	1	1	1
p-value =	2.542e-14	0.003033	0.02752	2.017e-07

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Se aplica el test Fligner-Killeen para determinar varianzas constantes, el p value < 0.05 por lo que se rechaza la hipótesis nula y se concluye que las varianzas de las muestras no son iguales.

3.5.3. Test de Mann Whitney

Tabla 19-3: Test de Mann Whitney.

Wilcoxon rank sum test with continuity correction FO and EM				
data:	H1	H2	H3	H4
W =	76197	10462	17810	28148
p-value =	3.068e-06	2.2e-16	8.539e-07	0.009513
alternative hypothesis:	0	0	0	0

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Al 95% de confianza se evidencia una diferencia significativa de la producción de hidrógeno según la técnica utilizada.

3.6. Series Temporales

En las técnicas FO y EM, se describe el comportamiento del hidrógeno según las técnicas: fermentación oscura y electrolisis microbiana, buscando posibles patrones temporales y como estas actúan en función del tiempo.

3.6.1. Reactor 1_ Variable Hidrógeno 1

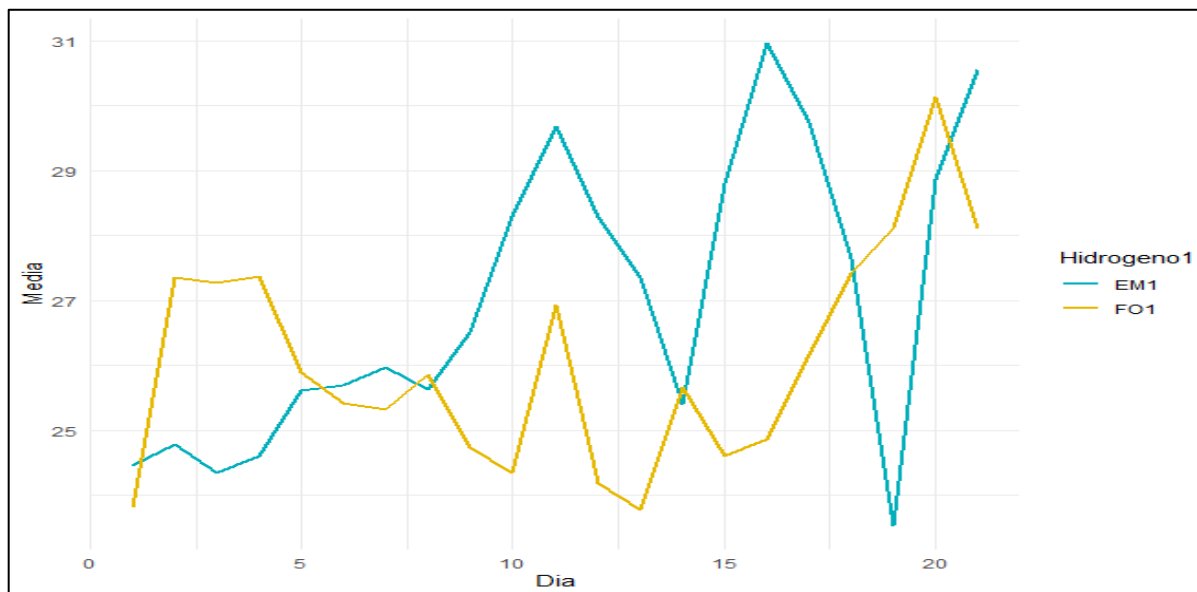


Gráfico 23-3: Serie temporal de la variable hidrógeno 1.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Para la serie temporal de la variable hidrógeno uno del reactor 1, se observa que en la técnica FO existe una tendencia creciente pronunciada, mientras que en la técnica EM la variación del hidrógeno es mayor a medida que pasa el tiempo.

3.6.2. Reactor 2_ Variable Hidrógeno 2

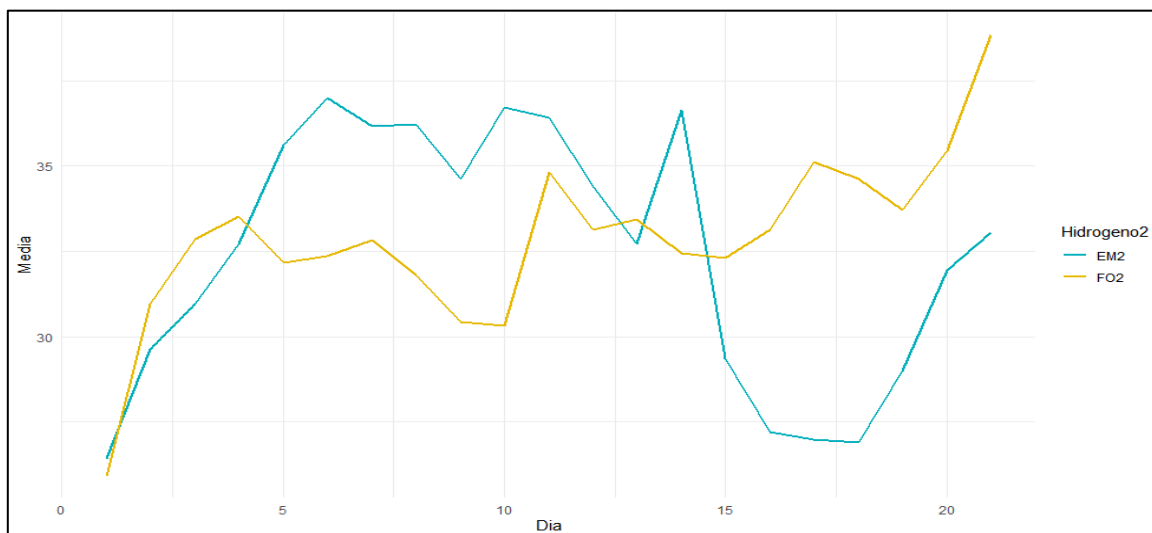


Gráfico 24-3: Serie temporal de la variable hidrógeno 2.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Para la serie temporal de la variable hidrógeno dos del reactor 2, se ha observado que en la técnica FO existe una tendencia creciente pronunciada, mientras que en la técnica EM la variación del hidrógeno fluctúa en menor medida con el paso del tiempo.

3.6.3. Reactor 3_ Variable Hidrógeno 3

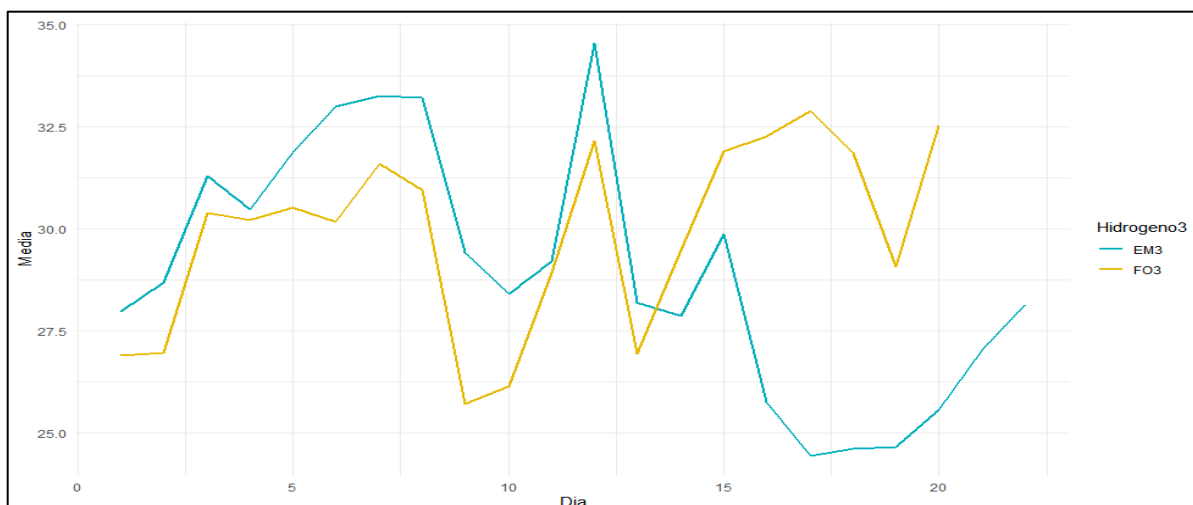


Gráfico 25-3: Serie temporal de la variable hidrógeno 3.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Se ha observado que para la serie temporal de la variable hidrógeno tres del reactor 3, en ambas técnicas FO y EM existe tendencia creciente, y la variación de hidrogeno fluctúa para ambas a medida que pasa el tiempo.

3.6.4. Reactor 4_ Variable Hidrógeno 4

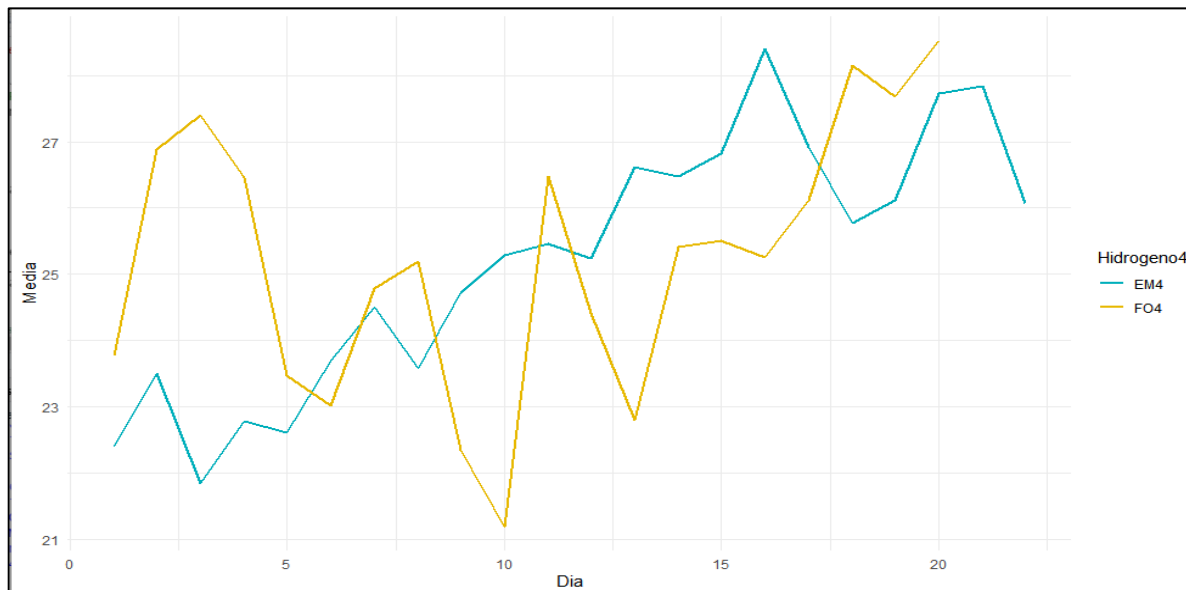


Gráfico 26-3: Serie temporal de la variable hidrógeno 4.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Para la serie temporal de la variable hidrógeno cuatro del reactor 4, se ha observado que en la técnica FO aumenta su oscilación, mientras que en la técnica EM la variación de hidrógeno tiene una tendencia creciente en mayor medida con el paso del tiempo.

3.7. Comparación Estadística y Análisis gráfico de Bland-Altman

3.7.1. Prueba de Normalidad

Para el análisis de Bland Altman se ha realizado un análisis en las mediciones, obteniendo así el siguiente resultado.

Tabla 20-3: Normalidad de las mediciones (FO y EM).

Normalidad de las mediciones	H1	H2	H3	H4
p-value FO =	2.217e-06	2.692e-07	1.15e-09	0.003328
p-value EM=	2.576e-07	6.96e-05	6.929e-05	0.1956

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

3.7.1.1. Gráficas de Normalidad

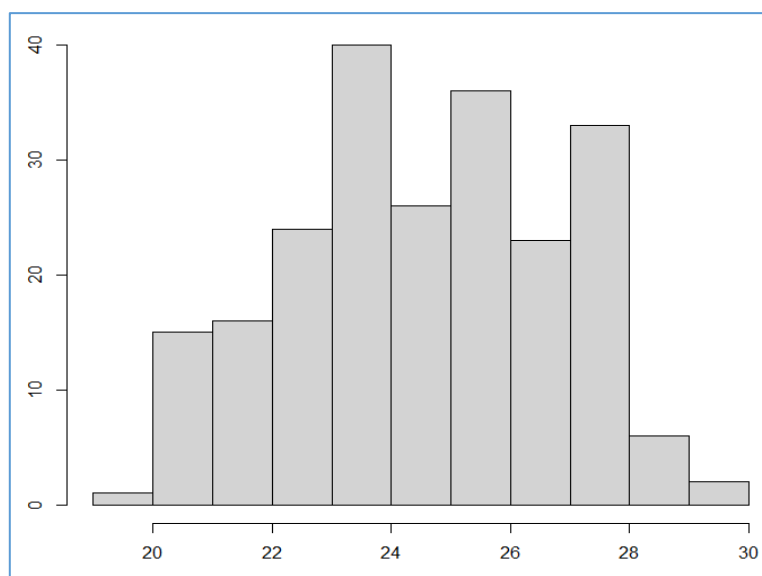


Gráfico 27-3: Histograma variable Hidrógeno 1.

Elaborado por: Olmedo, Beatriz, 2022.

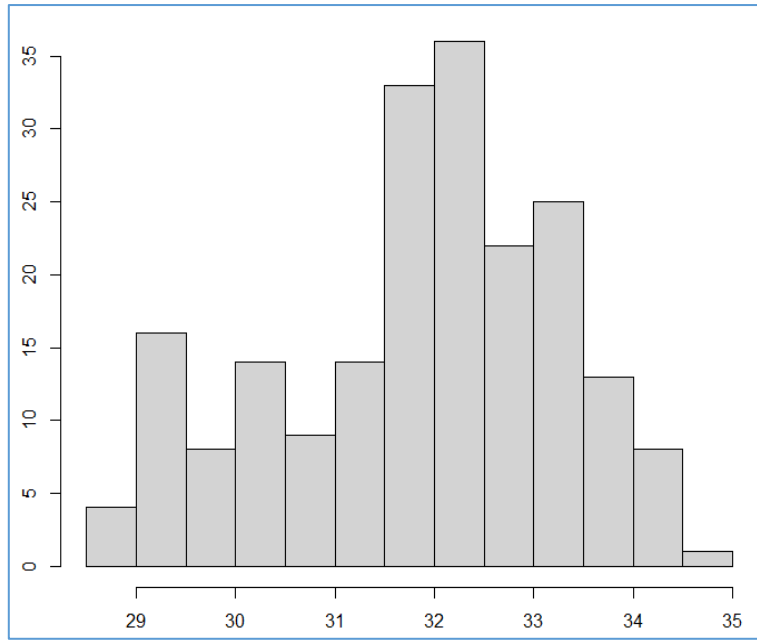


Gráfico 28-3: Histograma variable Hidrógeno 2.

Elaborado por: Olmedo, Beatriz, 2022.

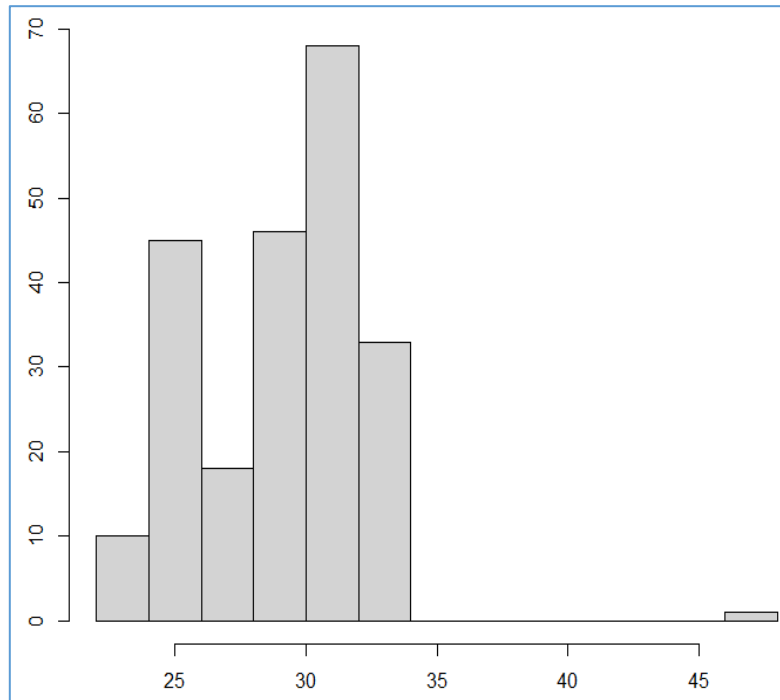


Gráfico 29-3: Histograma variable Hidrógeno 3.

Elaborado por: Olmedo, Beatriz, 2022.

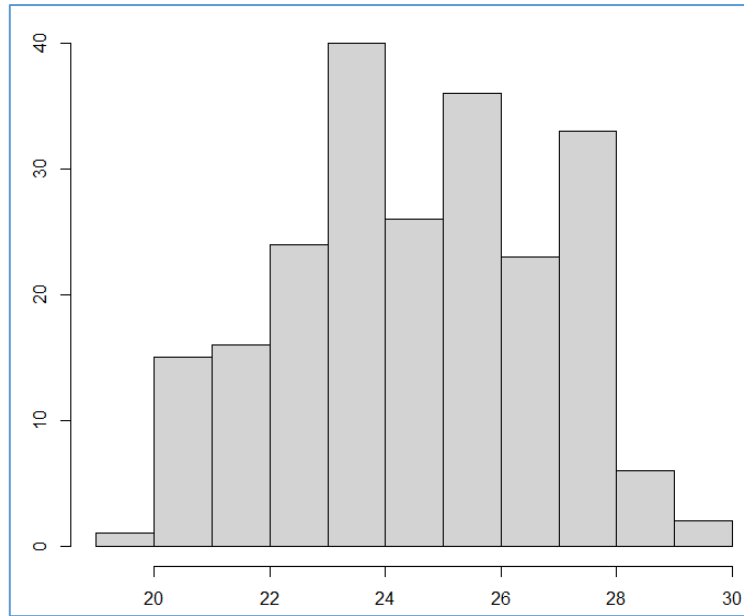


Gráfico 30-3: Histograma variables Hidrógeno 4.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Los valores p del estadístico Kolmogorov Smirnov es menor al alfa = 0.05; y se rechaza H0, y se ha podido concluir que datos no siguen una Distribución Normal en ambas técnicas (FO y EM).

3.7.2. Prueba de Normalidad

Para el análisis de Bland Altman se ha realizado un análisis en la diferencia de las medias, obteniendo así el siguiente resultado.

Tabla 21-3: Prueba de normalidad en las diferencias de las medias (FO y EM).

Lilliefors (Kolmogorov-Smirnov) normality test - data2\$diferencia				
data:	H1	H2	H3	H4
D =	0.048758	0.096129	0.079839	0.059377
p-value =	0.01556	0.0001003	0.001621	0.05534

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

3.7.2.1. Gráficas de Distribución Normal

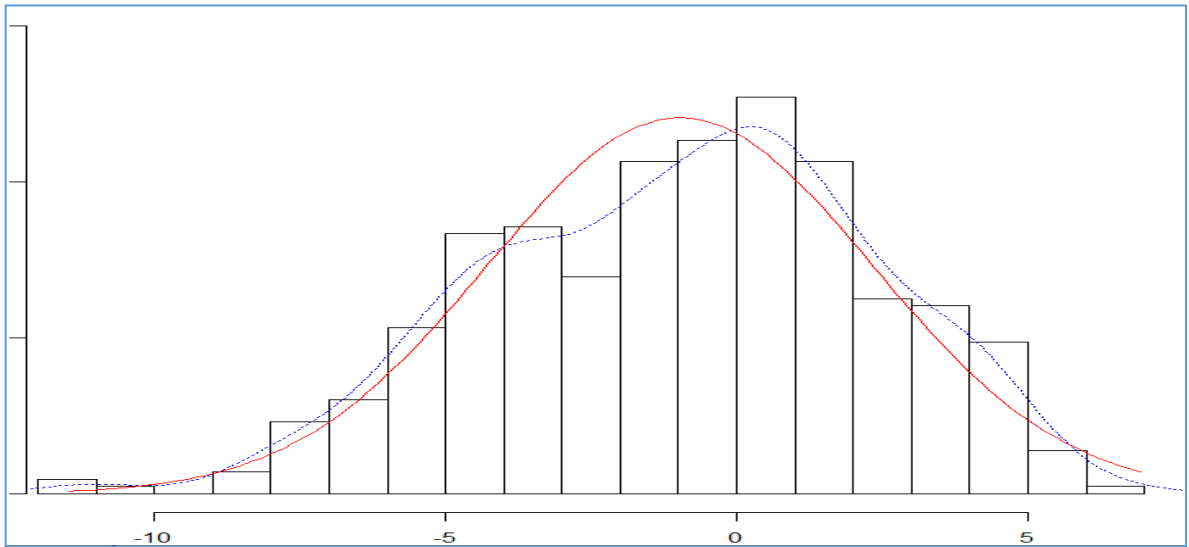


Gráfico 31-3: Distribución de las diferencias de las medias variable H1.

Elaborado por: Olmedo, Beatriz, 2022.

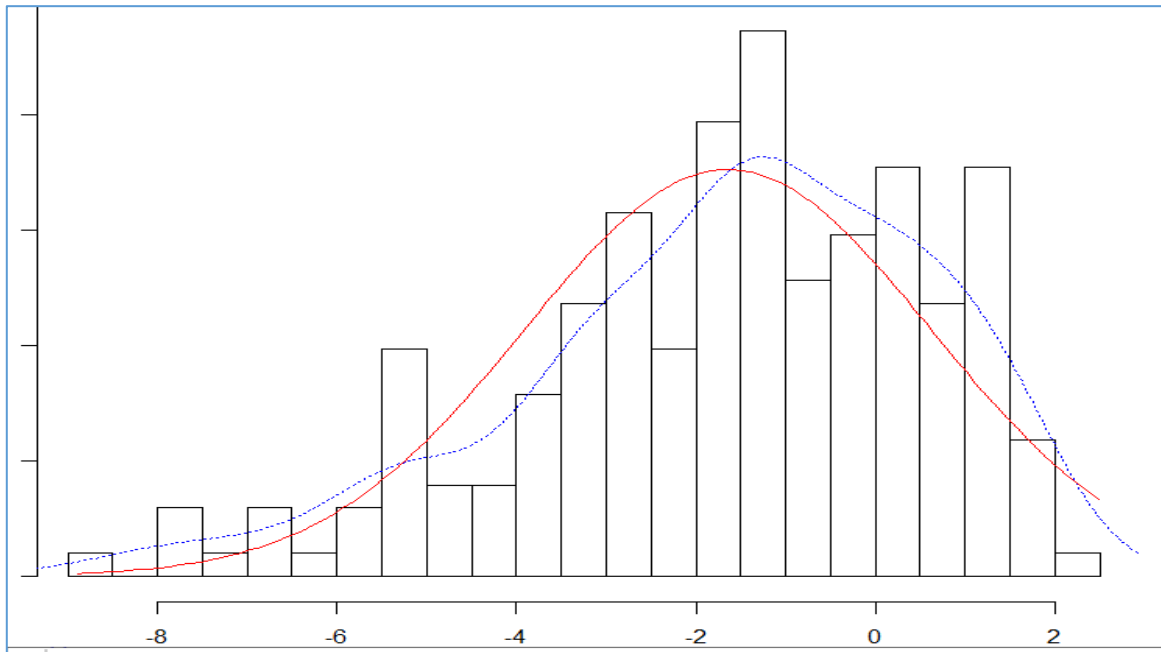


Gráfico 32-3: Distribución de las diferencias de las medias variable H2.

Elaborado por: Olmedo, Beatriz, 2022.

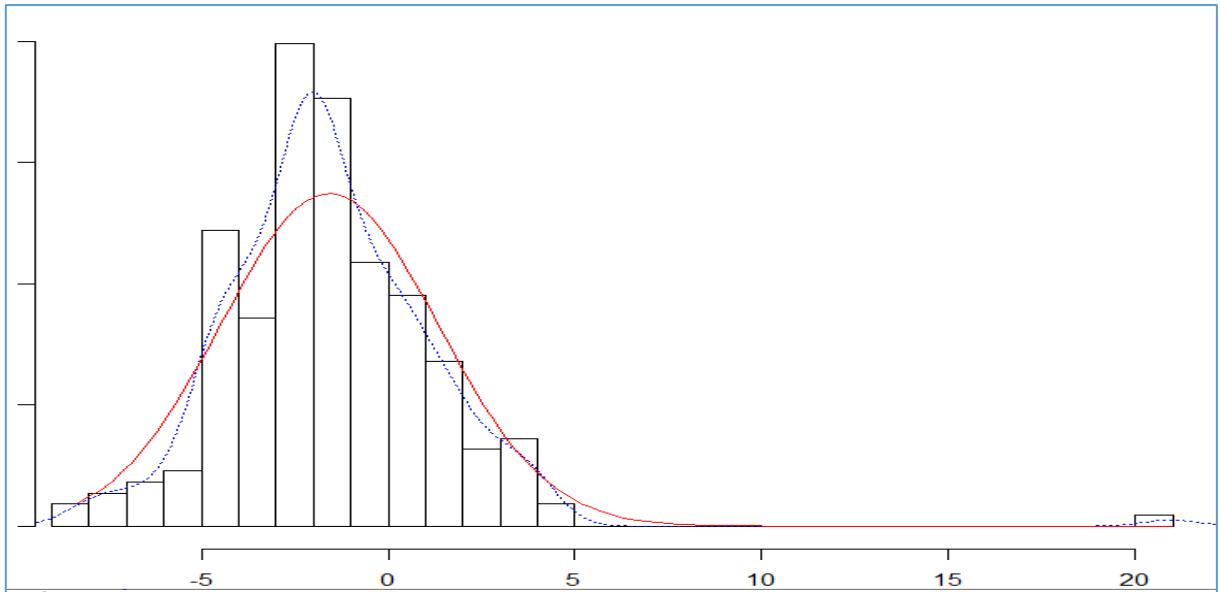


Gráfico 33-3: Distribución de las diferencias de las medias variable H3.

Elaborado por: Olmedo, Beatriz, 2022.

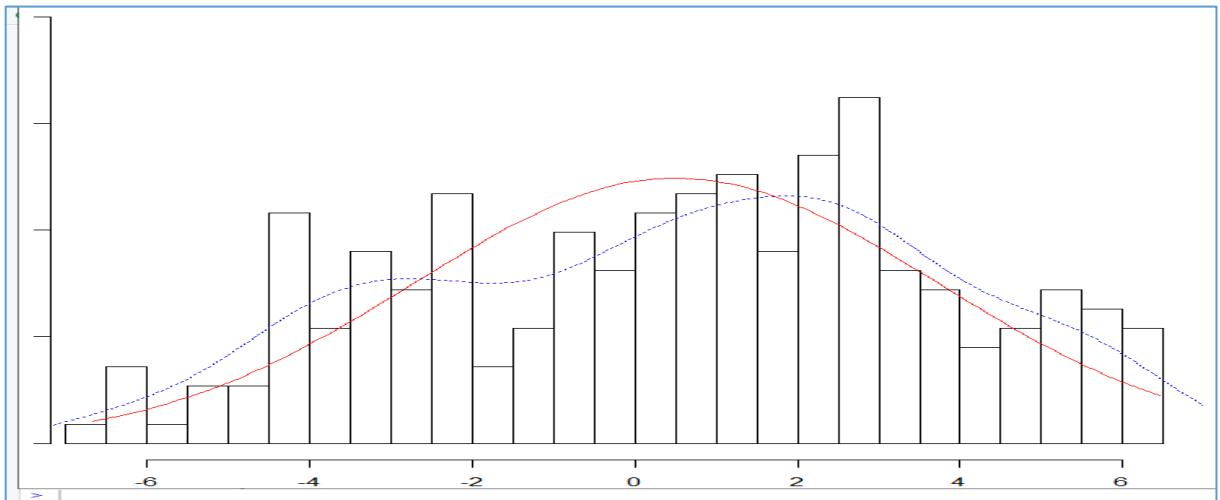


Gráfico34-3: Distribución de las diferencias de las medias variable H4.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Los valores p del estadístico Kolmogorov Smirnov son menores al $\alpha = 0.05$; y se rechaza H_0 , por lo que se ha podido concluir que, en la diferencia de los promedios de hidrógeno en ambas técnicas FO y EM no siguen una Distribución Normal.

3.7.3. Resumen de los Datos

Se ha realizado un resumen de datos de las técnicas (FO y EM), y se calculó la diferencia, media y porcentaje de dichas técnicas.

Tabla 22-3: Resumen de datos.

Reactor 1	FO	EM	Diferencia	Media	Porcentaje
Min	21.73	21.16	-11.42	22.75	-39.78
Median	25.65	26.44	-0.67	26.06	-2.58
Mean	25.81	26.78	-0.97	26.30	-3.45
Max.	30.88	34.83	6.89	31.96	25.43
Reactor 2	FO	EM	Diferencia	Media	Porcentaje
Min	28.57	30.30	-8.84	30.18	-25.19
Median	32.02	33.41	-1.35	32.68	-4.21
Mean	31.87	33.52	-1.64	32.69	-4.96
Max	34.70	39.50	2.44	36.68	7.38
Reactor 3	FO	EM	Diferencia	Media	Porcentaje
Min	23.17	26.37	-8.24	24.86	-27.73
Median	29.86	30.74	-1.82	30.07	-6.11
Mean	29.06	30.64	-1.57	29.85	-5.52
Max	48.00	37.96	20.93	37.54	55.75
Reactor 4	FO	EM	Diferencia	Media	Porcentaje
Min	19.19	21.30	-6.62	21.78	-27.23
Median	24.36	23.98	0.84	24.36	3.49
Mean	24.57	24.06	0.51	24.32	1.89
Max	29.20	27.68	6.40	27.30	26.15

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

3.7.4. Diagrama de dispersión y Modelo lineal de los datos

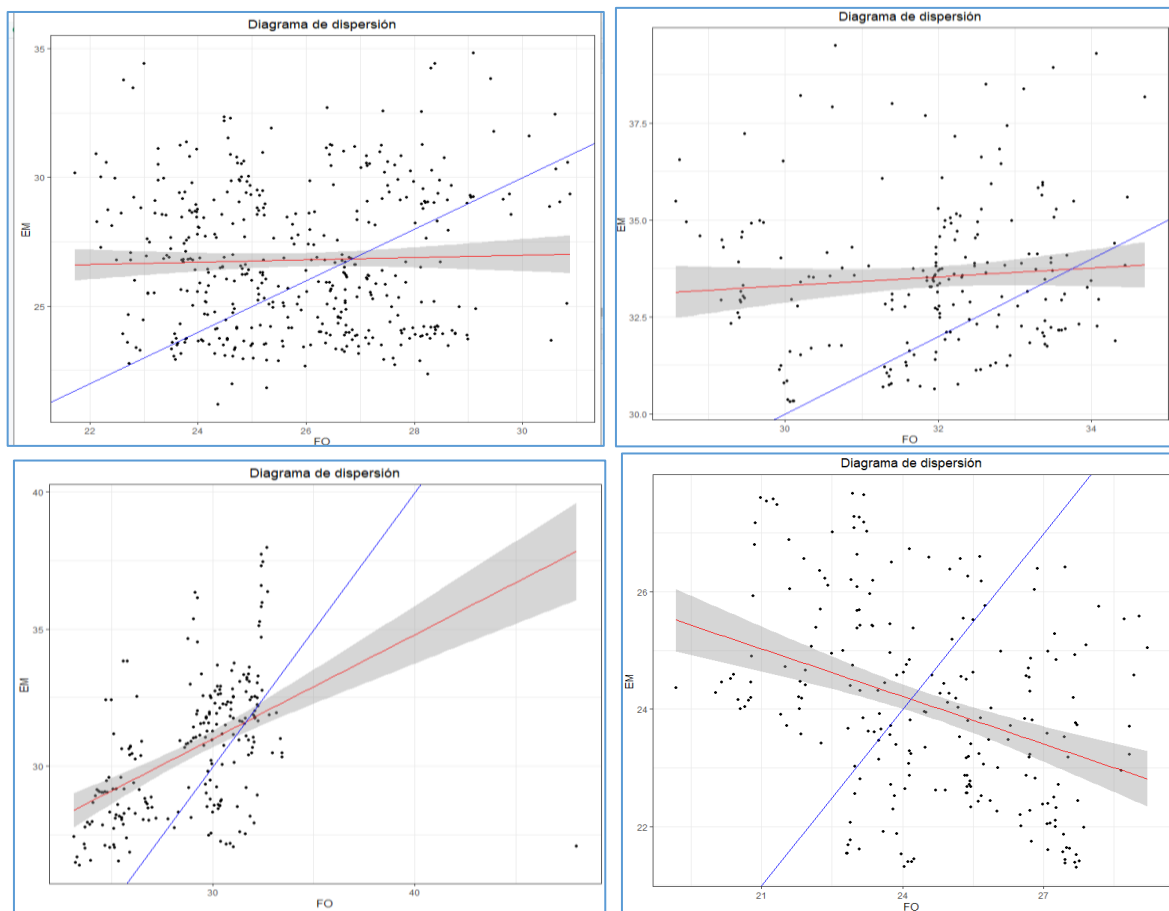


Gráfico 35-3: Dispersión con bisectriz y modelo lineal de los datos.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

El coeficiente Pearson muestra una asociación negativa entre los valores obtenidos de las técnicas FO y EM en los cuatro reactores, y se ha obtenido un coeficiente de correlación bajo negativo.

3.7.5. Intervalo de confianza

Tabla 23-3: Intervalo de Confianza de los parámetros del modelo lineal

Reactor 1	2.5 %	97.5 %
(Intercept)	22.13	29.11
FO	-0.089	0.179
Reactor 2	2.5 %	97.5 %
(Intercept)	23.91	35.84
FO	-0.07	0.30
Reactor 3	2.5 %	97.5 %
(Intercept)	16.88	22.30
FO	0.287	0.472
Reactor 4	2.5 %	97.5 %
(Intercept)	28.44	32.95
FO	-0.36	-0.17

Fuente: Grupo de Alternativas y Ambiente (GEAA).

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Para la recta de regresión lineal, se ha evaluado las pendientes de las variables en los cuatro reactores, obteniendo una mala concordancia entre las técnicas. En este proceso, la pendiente muestra que la técnica EM no sobreestima a la técnica FO. Así mismo, los datos han presentado una dispersión muy alejada de la bisectriz, lo que ha sugerido que existe un alto nivel de discrepancia entre las técnicas Fermentación Oscura y Electrólisis Microbiana.

3.7.6. Gráfico de Bland-Altman

En el método gráfico de Bland Altman se ha evaluado las observaciones de las técnicas utilizadas en la producción de hidrógeno a partir de lodos residuales irradiados de la zona minera en Ponce Enríquez, y se ha contrastado si existe concordancia entre dichas técnicas, siendo estas Fermentación Oscura (FO) y Electrólisis Microbiana (EM).

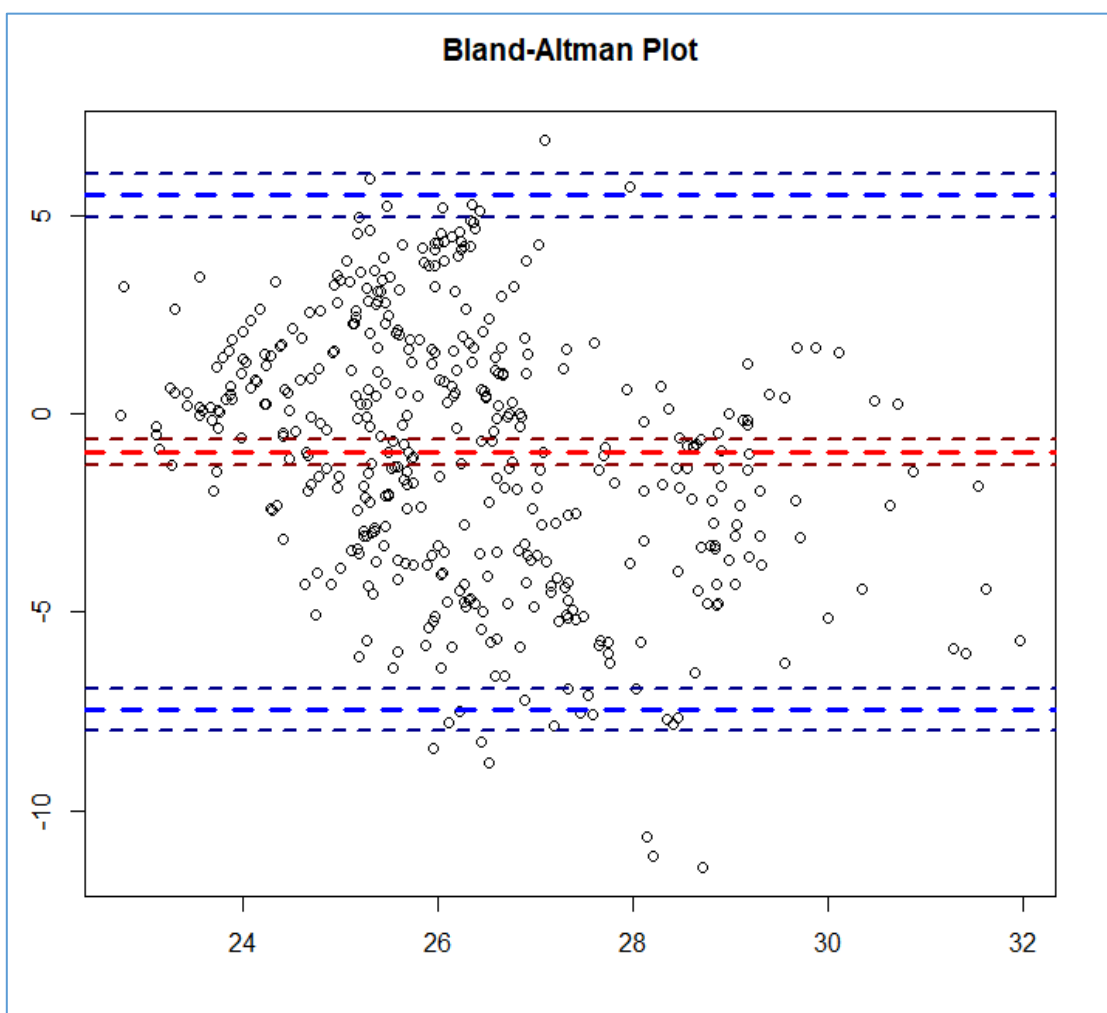


Gráfico 36-3: Bland – Altman Plot Hidrógeno 1.

Elaborado por: Olmedo, Beatriz, 2022.

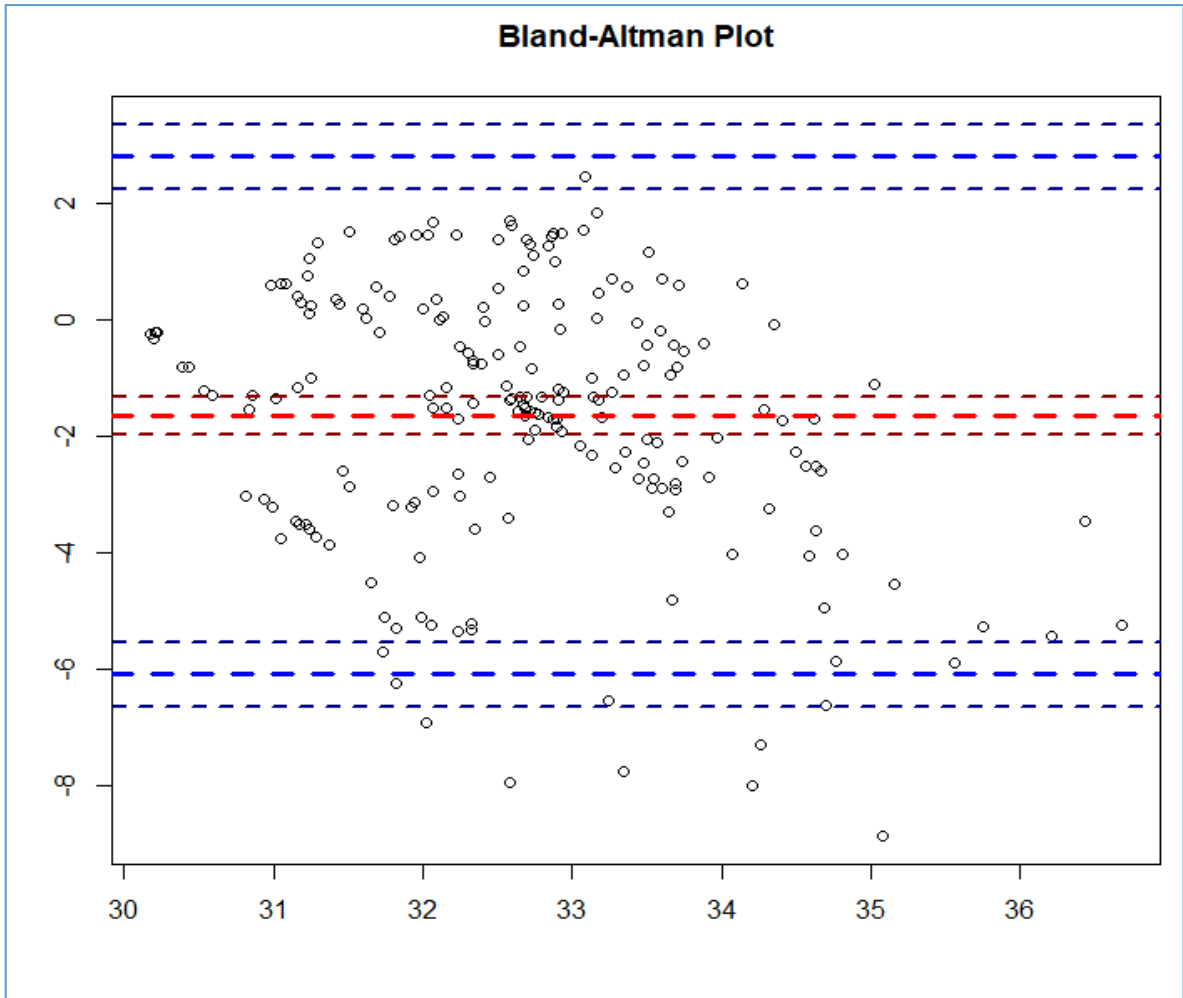


Gráfico 37-3: Bland – Altman Plot Hidrógeno 2.

Elaborado por: Olmedo, Beatriz, 2022.

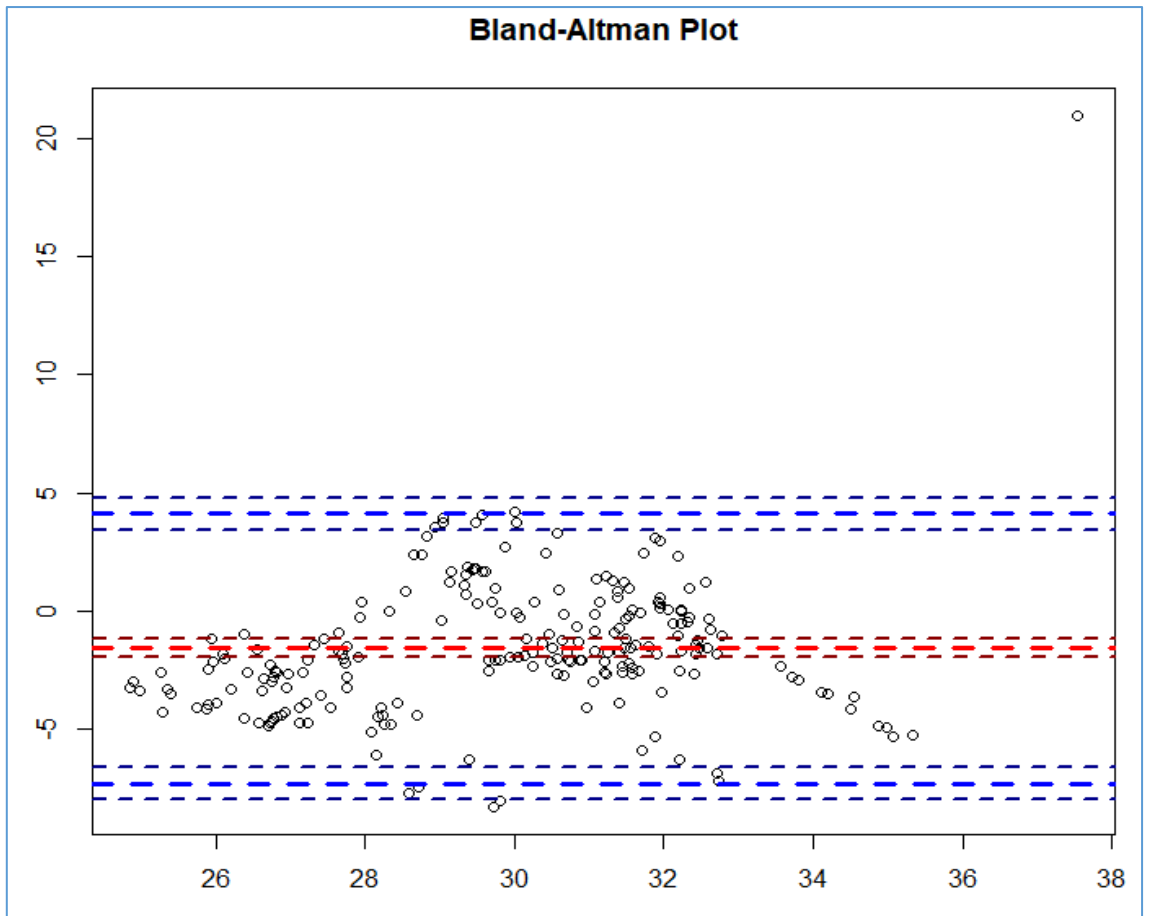


Gráfico 38-3: Bland – Altman Plot Hidrógeno 3.

Elaborado por: Olmedo, Beatriz, 2022.

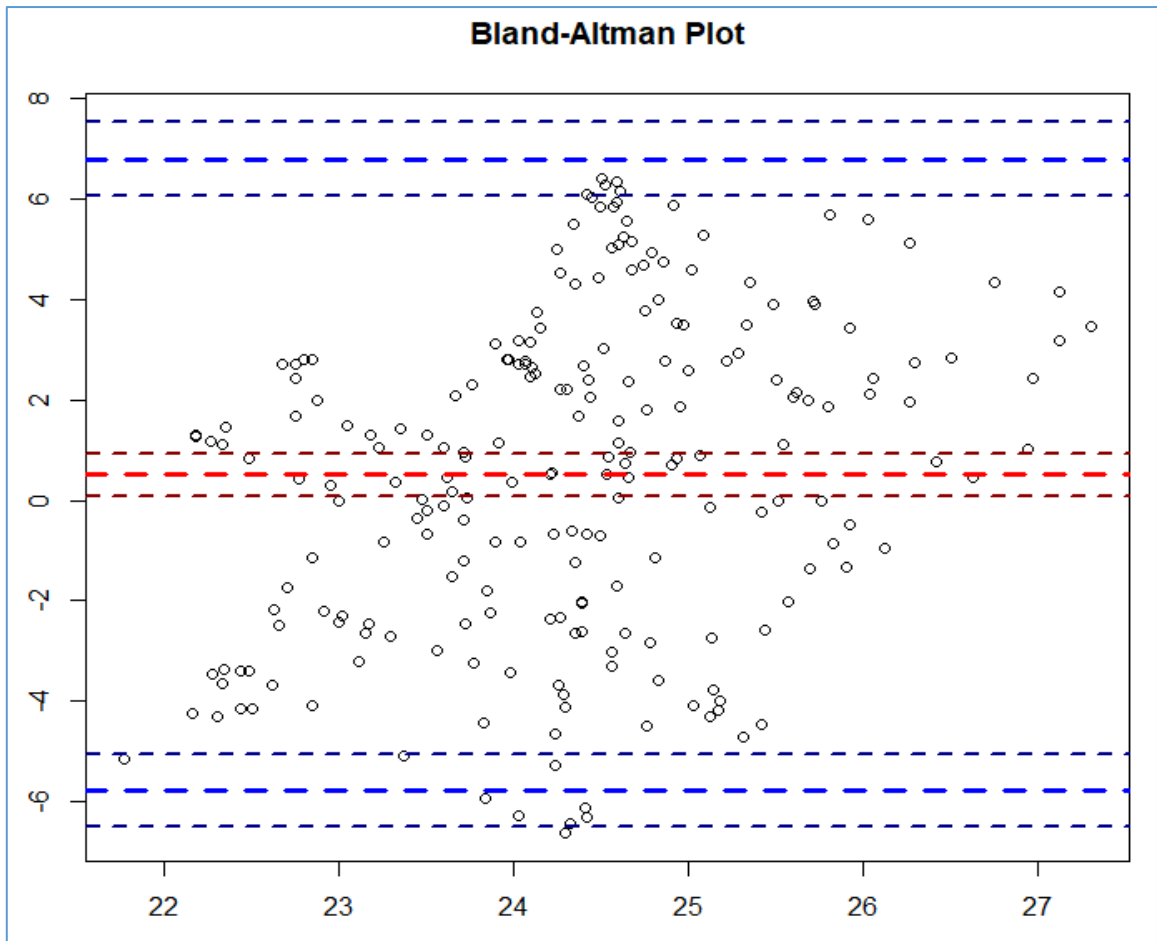


Gráfico 39-3: Bland – Altman Plot Hidrógeno 4.

Elaborado por: Olmedo, Beatriz, 2022.

Interpretación:

Se ha representado el gráfico de Bland Altman, donde en el eje **Y** corresponde a la diferencia entre las técnicas, mientras que el eje **X** representa el valor de la media de ambas. Se ha observado que las dos técnicas Fermentación oscura y Electrólisis microbiana obtuvieron valores promedios parecidos, y que los datos han presentado una dispersión muy alejada de la bisectriz, de igual manera la amplitud del intervalo nos dio una idea de que tan equivalentes son ambas técnicas. Se ha tenido como resultado que las técnicas utilizadas no arrojaron efectos parecidos, indicando así que no existe concordancia entre las técnicas FO y EM.

CONCLUSIONES

- Se ha realizado una comparación estadística en las técnicas utilizadas FO y EM, para la producción de hidrógeno a partir de lodos residuales. En ella se ha podido recolectar en el transcurso de 42 días un total de 1008 datos, se ha identificado y completado datos faltantes mediante las técnicas “Listwise e Imputación Múltiple” obteniendo el 96% de la información. Así mismo, se han encontrado datos atípicos mediante la prueba ajustada de Tukey y diagramas de cajas, donde se obtuvieron 241 datos dentro de los cuatro reactores, que corresponden al 24% del conjunto total de datos. Se ha trabajado con el 72% de la información para este estudio.
- Se ha realizado un análisis estadístico inferencial para comparar la producción de hidrógeno en base a las técnicas: fermentación oscura y electrólisis microbiana. Para ello se ha utilizado la prueba de del estadístico Kolmogorov Smirnov con la corrección de Lilliefors, que indican que los datos no siguen una Distribución Normal para ambas técnicas. Así mismo se ha aplicado el test Fligner-Killeen que indica que las varianzas de los grupos (FO y EM) no son constantes. Por consiguiente, se aplica la prueba de Mann Whitney, donde se demuestra que existe una diferencia significativa de la producción de hidrógeno según la técnica utilizada.
- Se describe el comportamiento de la producción de hidrógeno según la técnica mediante series temporales, se corrobora que existe una diferencia significativa de la variable en estudio en función del tiempo. La producción de hidrógeno según la técnica EM mantuvo una tendencia creciente pronunciada y una mejor variación a diferencia de la técnica FO.
- Se ha realizado el análisis comparativo con el gráfico de Bland-Altman, en donde se ha podido cuantificar la diferencia entre ambas técnicas. Se ha observado que las dos técnicas Fermentación oscura y Electrólisis microbiana obtuvieron valores centrales similares y una dispersión muy alejada de la bisectriz, los intervalos de confianza han evidenciado que la pendiente de la técnica EM no sobreestima a la técnica FO, existiendo discrepancia y nula concordancia entre ellas.

RECOMENDACIONES

- Se recomienda utilizar una técnica de imputación adecuada de acuerdo a la cantidad de datos faltantes y atípicos.
- Los datos no presentan una distribución normal por lo que se recomienda pruebas no paramétricas.
- Se recomienda el uso del análisis gráfico de Bland Altman para determinar si la variabilidad o precisión entre dos grupos de técnicas.

BIBLIOGRAFÍA

ALMEIDA, M. D., repositorio.cepal.org. 2019. Available at: https://repositorio.cepal.org/bitstream/handle/11362/44876/S1900582_es.pdf?sequence=1&isAllowed=y

AMAT, J. R., Análisis de la homogeneidad de varianza (homocedasticidad), 2016. Available at: https://rpubs.com/Joaquin_AR/218466.

AMAT, J. R., Test de Wilcoxon Mann Whitney como alternativa al t-test, 2017. Available at: https://www.cienciadedatos.net/documentos/17_mann%E2%80%93whitney_u_test.

BECERRA, J. M., " Estadística Descriptiva". apuntes_matematicas, 2020, p. 38. Available at: http://132.248.164.227/publicaciones/docs/apuntes_matematicas/34.%20Estadistica%20Descriptiva.pdf

BLANCO LONDOÑO, S. A., " Producción de biohidrógeno a través de la fermentación oscura de residuos, revisión crítica". Colombia. 2013. Available at: <https://repositorio.unal.edu.co/handle/unal/52055>.

CURBELO, G. M., "Para el análisis de correlación y concordancia en equipos de mediciones similares". *scielo*, 2016, 8(4), p. 6.

CARDONA, M., BERDUGO, J. & CADAVID, A., Comparación de la concentración espermática usando la cámara de Makler y la cámara de Neubauer. 2008. Available at: <https://scielo.isciii.es/pdf/ae/v32n4/v32n4a10.pdf>

CURSO, "Conceptos básicos en el contraste de hipótesis". *blog.minitab*, 2010. Available at: <http://halweb.uc3m.es/esp/Personal/personas/aarribas/eng/docs/estII/tema2.pdf>

DAGNINO, J., Correlacion. [En línea] 2014. Available at: https://www.sachile.cl/upfiles/revistas/54e63a1a778ff_15_correlacion-2-2014_edit.pdf.

EDITOR, M. B., "Entendiendo las Pruebas de Hipótesis: niveles de Significancia (Alfa) y Valores P en Estadística". 2019. [En línea], [Último acceso: 09 01 2022]. Available at: <https://blog.minitab.com/es/entendiendo-las-pruebas-de-hipotesis-niveles-de-significancia-alfa-y-valores-p-en-estadistic>

FAJARDO, A., "Contrastes de Normalidad".[blog], 2007. [En línea], [Último acceso: 09 01 2022]. Available at: http://www.ub.edu/aplica_infor/spss/cap5-6.htm

GALVÁN, M. Y MEDINA, F., *Imputación de datos: teórica y práctica. En: Estudios estadísticos prospectivos*. Santiago: Naciones Unidas, CEPAL, Div. de Estadística y Proyecciones Económicas. 2007, 978-92-1-323101-2.

GARCÍA DÍAZ, J. C., *Predicción en el dominio del tiempo: análisis de series temporales para ingenieros*. Valencia: Editorial de la Universidad Politécnica de Valencia, 2016.

GARCÍA, F. M., "*Popper, el contraste de hipótesis y el método crítico*". scielo [En línea], 2003. [Último acceso: 09 02 2022]. Available at: <https://www.redalyc.org/pdf/214/21429108.pdf>

HERNÁNDEZ, S., FERNÁNDEZ, C. & BAPTISTA, L., *Metodología de la Investigación*. sexta edición ed. México D.F: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V. 2014.

LLADSER, M. *Variables Aleatorias y Simulación Estocástica*. Santiago de Chile: J.C. Sáez Editor, 2010.

LLADSER, M., *Variables Aleatorias y Simulación Estocástica*. Santiago de Chile: J.C. Sáez Editor. 2010, 978-956-306-068-3.

LLINÁS SOLANO, H., *Estadística Inferencial*. Barranquilla : Universidad del Norte., 2017. 978-958-741-916-0, P.444.

LÓPEZ CEDEÑO, D. A., "Evaluación al proceso técnico, operativo y ambiental de la planta de tratamiento de aguas residuales de la ciudad de jipijapa". *unesum*, [En línea], 2019, P. 109, [Último acceso: 9 1 2022]. Available at: <http://repositorio.unesum.edu.ec/handle/53000/1877>

MARTÍNEZ, C. G., *Data Science con R*, Madrid: github, 2021.

MARTÍNEZ, E., *Estadística.. República dominicana: Universidad Abierta para Adultos (UAPA)*, 2020, 978-9945-580-89-1, P. 221.

MARTÍNEZ, & OTROS, *Estudio comparativo del estado nutricional de dos muestras de jóvenes adultos*. [En línea] 2019. Available at: <http://www.didac.ehu.es/antropo/10/10-2/Martinez.pdf>

MÉNDEZ CUZCO, P. A., Producción Producción de biohidrógeno a partir de lodos residuales irradiados mediante un proceso integrado de fermentación biohidrógeno a partir de lodos residuales irradiados mediante un proceso integrado de fermentación oscura y electrólisis microbiana, Trabajo de titulación, (Ingeniería), Espoch, Riobamba - Ecuador, 2020, P.104, [En línea] Available at: <http://dspace.espoch.edu.ec/handle/123456789/14271>

MENDOZA MEDELLÍN, A., El hidrógeno y la energía. *redalyc*, marzo-junio, vol. 13(1405-0269), 2006, p. 7.

OLAYA OCHOA, J., "*Métodos de regresión no paramétrica*". Cali: Programa Editorial Universidad del Valle, 2012.

ORTIZ, L., ET AL.,, Propuestas de manejo de los lodos residuales de la planta de tratamiento de la ciudad industrial del valle de cuernavaca, estado de Morelos, Mexico", 1995." *Revista Internacional de Contaminación*, septiembre-noviembre, 11(0188-4999), p. 12.

PACHECO, D. & VARELA, K., "Análisis estadístico comparativo del contenido de óxidos presentes en lodos sin calcinar, lodos calcinados y cemento portland". [En línea], 2019, [Último acceso: 09 01 2022], Available at: <http://hdl.handle.net/11323/4724>.

PEÑA, D., *Análisis de datos multivariantes*. Arvaca Madrid: McGraw-Hill España, 2013, 978-84-481-9184-9, P.558.

PUENTE VIEDMA, C., *Estadística descriptiva e inferencial..* Madrid: Ediciones IDT, 2018, 978-1-5129-5292-6 978-84-943724-7-6, P.258.

QUEVEDO, F., Medidas de tendencia central y dispersión. [En línea], 2011, [Último acceso: 02 02 2022]. Available at: <https://www.medwave.cl/link.cgi/Medwave/Series/MBE04/4934>.

RAMÍREZ MORDÁN, N., Economía Aplicada. [En línea] 2018. Available at: <https://betaeconomia.blogspot.com/2018/10/metodos-para-la-identificacion-de.html>.

REQUENA SERRA, B., Asimetría Y Curtosis. *universoformulas* [En línea], 2014, [Último acceso: 26 02 2022], Available at: <https://www.universoformulas.com/estadistica/descriptiva/asimetria-curtosis/>

R, M., *Estadísticas*. España: Instituto Politécnico Nacional, 2010.

RODRIGO , A. & MARCELO , F., Imputación Múltiple en Encuestas Microeconómicas. *scielo*, 46(134), 2009. p. 288.

RODRÍGUEZ, T., Ensayo de estadística inferencial. Guayama: slideshare, 2015.

SANDOVAL, F., La Pequeña Minería en el Ecuador. [En línea] 2001. Available at: https://intranetua.uantof.cl/crea/cguerra/pdf/files/otros/075_ecuador_esp.pdf.

TURCIOS SÁNCHEZ, R. A., Prueba de Wilcoxon-Mann-Whitney: mitos y realidades. [En línea], 2015, [Último acceso: 04 02 2022], P.4, Available at: <https://biblat.unam.mx/hevila/Revistamexicanadeendocrinologiametabolismo&nutricion/2015/vol2/no1/3.pdf>

UNIVERSIDAD DE PAMPLONA, ET AL ,"Modelos de regresión para estimar el peso seco de órganos y área del limbo del duraznero, variedad jarillo". *Rev. U.D.C.A Act. & Div. Cient.*, 20(01234226, 26192551), 2017, p. 12.

VARGAS SABADÍAS, . A.,. Estadística descriptiva e inferencial. España: Univ de Castilla La Mancha, 1995, 84-88255-87-X, P. 576.

VÁZQUEZ, L. P., Alternativas didácticas de la Estadística Inferencial en el pregrado de las ciencias de la salud. *Revista Ciencias Estratégicas*, octubre-diciembre, 14(1794-8347), 2011, p. 18.

VELÁSQUEZ , J. D. & QUICENO , J. F., Diseño de un sistema de generación de hidrógeno por electrólisis. [En línea] 2013. Available at: <https://hdl.handle.net/11059/3655>

VÉLEZ EVANS, M. I., "El proceso de toma de decisiones como un espacio para el aprendizaje en las organizaciones". *web y empresas*, 16 julio, 14(1794-8347), 2006. p. 18.

ANEXOS

ANEXO A



AVAL

EL DIRECTOR DEL GRUPO DE ENERGÍAS ALTERNATIVAS Y AMBIENTE
(GEAA)

DOCTOR CELSO GUILLERMO RECALDE MORENO, SUSCRIBE

AVALA:

Que la señorita **Beatriz Mercedes Olmedo Perea** con CI: 080325800-3, estudiante de la Carrera de Ingeniería en Estadística se encuentra desarrollando su trabajo de integración Curricular como modalidad; con el tema denominado: **COMPARACIÓN ESTADÍSTICA DE DOS TÉCNICAS UTILIZADAS EN LA PRODUCCIÓN DE BIOHIDRÓGENO A PARTIR DE LODOS RESIDUALES IRRADIADOS DE LA ZONA MINERA EN PONCE ENRIQUEZ.**

Tipo: PROYECTO DE INVESTIGACIÓN en el GRUPO DE ENERGÍAS ALTERNATIVAS Y AMBIENTE PERTENECIENTE A LA FACULTAD DE CIENCIAS EN LA ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO.

Se extiende el presente documento a solicitud del interesado, para los fines que estime convenientes.

Riobamba, 15 de noviembre del 2021

Dr. Celso Guillermo Recalde
DIRECTOR GEAA-FAC. CIENCIAS

Teléfono: 593(03) 2998-200 Ext.2221

geaesPOCH@gmail.com

ANEXO B

```
# TEMA #  
  
# BASE DE DATOS DE COMPARACIÓN ESTADÍSTICA DE LAS TÉCNICAS UTILIZADAS #  
# EN LA PRODUCCIÓN DE HIDRÓGENO A PARTIR DE LODOS RESIDUALES #  
# IRRADIADOS DE LA ZONA MINERA EN PONCE ENRIQUEZ #  
  
library(readxl)  
setwd("C:/Users/Personal/Desktop/BEATRIZ/Tesis_Completa/Tesis/Tesis_BEATRIZ")  
data <- read_excel("Base de Datos_Faltantes.xlsx")  
  
data_a_imputar <- data[,-c(1,2,3,4,5,6,7,8)]  
  
# Numero de valores faltantes a nivel general  
n <- sum(is.na(data_a_imputar))  
n  
  
# valores faltantes NAs por variable  
fa <- colsums(is.na(data_a_imputar))  
fa  
  
# fr = ni/n, porcentaje de NAs por variable  
fr <- fa/965  
fr  
  
# Clasificacion datos perdidos  
#install.packages("VIM")  
library(VIM)  
  
  
# VARIABLES Temperatura  
columnst <- c("T1", "T2", "T3", "T4")  
  
imputed_dataT <- mice(data[,names(data) %in% columnst], seed = 2022, print = F,  
m = 30)  
|  
# Data completa añadido los datos imputados mediante IM  
  
complete_dataT <- complete(imputed_dataT)  
  
# VARIABLES MH  
columnsmh <- c("MH1", "MH2", "MH3", "MH4")  
  
imputed_dataMH <- mice(data[,names(data) %in% columnsmh], seed = 2022, print = F,  
m = 30)  
  
# Data completa añadido los datos imputados mediante IM  
  
complete_dataMH <- complete(imputed_dataMH)  
  
# Data en estudio imputada  
  
data_imputada <- data.frame(data[,c(1,2,3,4,5,6,7,8)], complete_dataT, complete_dataMH)  
view(data_imputada)  
  
# Exportacion de datos de R.data a .xlsx  
  
library(openxlsx)  
write.xlsx(data_imputada, file = "Base de Datos_Faltantes_Imputados.xlsx")
```

```

# TEMA #

# BASE DE DATOS DE COMPARACIÓN ESTADÍSTICA DE LAS TÉCNICAS UTILIZADAS #
# EN LA PRODUCCIÓN DE HIDRÓGENO A PARTIR DE LODOS RESIDUALES #
# IRRADIADOS DE LA ZONA MINERA EN PONCE ENRIQUEZ #

## ANALISIS DE DATOS ATIPIICOS

library(readxl)
setwd("C:/Users/Personal/Desktop/BEATRIZ/Tesis_Completa/Tesis/Tesis_BEATRIZ")
data <- read_excel("Base de Datos_Faltantes_Imputados (1).xlsx", sheet = "4H")

attach(data)

#####

# Con datos atípicos

library(BlandAltmanLeh)

bland.altman.plot(group1 = data$F01, group2 = data$EM1, main = "Reactor 1 - Bland-Altman",
                  xlab = "", ylab = "", conf.int = 0.95)

bland.altman.plot(group1 = data$F02, group2 = data$EM2, main = "Reactor 2 - Bland-Altman",
                  xlab = "", ylab = "", conf.int = 0.95)

bland.altman.plot(group1 = data$F03, group2 = data$EM3, main = "Reactor 3 - Bland-Altman",
                  xlab = "", ylab = "", conf.int = 0.95)

bland.altman.plot(group1 = data$F04, group2 = data$EM4, main = "Reactor 4 - Bland-Altman",
                  xlab = "", ylab = "", conf.int = 0.95)

#####

# Prueba analítica

EM4 <- na.omit(EM4)
EM4_95 <- quantile(EM4, c(0.1, 0.90))
EMrecortada<- (EM4[EM4 > EM4_95[1] & EM4 < EM4_95[2]])
q1 <- quantile(EMrecortada, 0.25)
q3 <- quantile(EMrecortada, 0.75)
iqr <- q3-q1
t_tukeyR <- (EM4 < (q1-1.5*iqr) | EM4 > (q3+1.5*iqr))
which(t_tukeyR)
ndata_atipEM4 <- c(5, 7, 8, 46, 47, 48, 49, 52, 53, 54, 55, 56, 57, 58, 71, 72, 73, 96, 277, 279,
                 280, 281, 282, 283, 299, 300, 301, 302, 303, 304, 398, 421, 422, 423, 424, 425,
                 445, 446, 447, 448, 449, 450, 451, 468, 469, 470, 471, 472, 473, 588, 589, 590,
                 591, 592, 594, 596, 597, 598, 614, 615, 684, 685, 686, 687, 688, 710, 711)

length(ndata_atipEM4)
data_atipEM4 <- EM4[ndata_atipEM4]
data_sin_atipEM4 <- EM4[-ndata_atipEM4]

# Exportación de datos EM1
library(openxlsx)
write.xlsx(x = as.data.frame(data_sin_atipEM4), file = "ReactorEM4H.xlsx")

# Análisis de Bland-Altman sin datos atípicos F04 vs EM4
library(BlandAltmanLeh)
bland.altman.plot(group1 = data_sin_atipF04, group2 = data_sin_atipEM4[1:222],
                  main = "FO vs EM Recator 1 Bland-Altman", xlab = "", ylab = "",
                  conf.int = 0.95)
length(data_sin_atipF04)
length(data_sin_atipEM4)

# Boxplots sin datos atípicos

boxplot(data_sin_atipF04)
boxplot(data_sin_atipEM4)

```

```

# TEMA #

# BASE DE DATOS DE COMPARACIÓN ESTADÍSTICA DE LAS TÉCNICAS UTILIZADAS #
# EN LA PRODUCCIÓN DE HIDRÓGENO A PARTIR DE LODOS RESIDUALES #
# IRRADIADOS DE LA ZONA MINERA EN PONCE ENRIQUEZ #

## SERIE DE TIEMPO HIDRÓGENO FO y EM

#####3

# Datos Reactor 1

library(readxl)

setwd("C:/Users/Personal/Desktop/BEATRIZ/Tesis_Completa/Tesis/Tesis_BEATRIZ")
data_sin_atipicos_H1 <- read_excel("Base de Datos_Faltantes_Imputados (1).xlsx", sheet = "H1 (2)")
attach(data_sin_atipicos_H1)

library(tidy)
library(dplyr)
mediasF01 <- data_sin_atipicos_H1 %>%
  group_by(DiaF01) %>%
  summarise("Promedio" = mean(F01))
mediasF01 <- na.omit(mediasF01)

mediasEM1 <- data_sin_atipicos_H1 %>%
  group_by(DiaEM1) %>%
  summarise("Promedio" = mean(EM1))
mediasEM1 <- na.omit(mediasEM1)

tF01 <- rep(x = "F01", 21)
tEM1 <- rep(x = "EM1", 21)

dataH1 <- data.frame("Dia" = 1:21, "Hidrogeno1" = c(tF01, tEM1),
  "Media" = c(mediasF01$Promedio,mediasEM1$Promedio))

# Datos Reactor 4

library(readxl)
setwd("C:/Users/Personal/Desktop/BEATRIZ/Tesis_Completa/Tesis/Tesis_BEATRIZ")

data_sin_atipicos_H4 <- read_excel("Base de Datos_Faltantes_Imputados (1).xlsx", sheet = "H4 (2)")
attach(data_sin_atipicos_H4)

mediasF04 <- data_sin_atipicos_H4 %>%
  group_by(DiaF04) %>%
  summarise("Promedio" = mean(F04))

mediasF04 <- na.omit(mediasF04)

mediasEM4 <- data_sin_atipicos_H4 %>%
  group_by(DiaEM4) %>%
  summarise("Promedio" = mean(EM4))
mediasEM4 <- na.omit(mediasEM4)

tF04 <- rep(x = "F04", 20)
tEM4 <- rep(x = "EM4", 22)

dataH4 <- data.frame("Dia" = c(1:20,1:22), "Hidrogeno4" = c(tF04, tEM4),
  "Media" = c(mediasF04$Promedio,mediasEM4$Promedio))

ggplot(dataH4, aes(x = Dia, y = Media)) +
  geom_line(aes(color = Hidrogeno4), size = 1) +
  scale_color_manual(values = c("#00AFBB", "#E7B800")) +
  theme_minimal()

```

```

# TEMA #

# BASE DE DATOS DE COMPARACIÓN ESTADÍSTICA DE LAS TÉCNICAS UTILIZADAS #
# EN LA PRODUCCIÓN DE HIDRÓGENO A PARTIR DE LODOS RESIDUALES #
# IRRADIADOS DE LA ZONA MINERA EN PONCE ENRIQUEZ #

## GRÁFICA DE BLAND ALTMAN

library(readxl)
setwd("C:/Users/Personal/Desktop/BEATRIZ/Tesis_Completa/Tesis/Tesis_BEATRIZ")
data <- read_excel("Base de Datos_Faltantes_Imputados (1).xlsx", sheet = 4)
data <- na.omit(data[,])

##### VARIABLE HIDROGENO 1 #####
#####
# NORMALIDAD DE LOS MEDICIONES DE AMBAS TECNICAS

FO <- as.data.frame(data$FO)
EM <- as.data.frame(data$EM)
class(EM)

library(nortest)
lillie.test(FO$`data$FO`)
lillie.test(EM$`data$EM`)

hist(FO$`data$FO`)

```

```

library(ggplot2)
# Diagrama de dispersión con bisectriz y modelo lineal de los datos

ggplot(data = data2, mapping = aes(x = FO, y = EM)) +
  geom_point(color = "black", size = 1) +
  labs(title = "Diagrama de dispersión", FO = "Fermentación Oscura", EM = "Electrolisis Microbiana")
  geom_smooth(method = "lm", se = TRUE, color = "red", lwd = 0.5) +
  geom_abline(intercept = 0, slope = 1, lwd = 0.7, col = "blue") +
  theme(axis.line = element_line(colour = "black"),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank()) +

  theme_bw() + theme(plot.title = element_text(hjust = 0.5))

#####

# Paso 2: Análisis gráfico de los datos
library(ggplot2)
ggplot(data = data2, mapping = aes(x = FO, y = EM)) +
  geom_point(color = "black", size = 1) +
  labs(title = "Diagrama de dispersión", x = "FERMENTACION OSCURA", y = "ELECTROLISIS MICROBIANA") +
  cor.test(data2$FO, data2$EM, method = "pearson")

# Coeficiente de correlacion baja negativa

# Modelo lineal de los datos
modelo_lineal <- lm(formula = EM ~ FO, data = data2)
summary(modelo_lineal)

# Intervalos de confianza (CI) de los parámetros del modelo lineal
confint(modelo_lineal, level = 0.95)

#####

```

```
#####
sd <- sd(data2$diferencia)
bias <- mean(data2$diferencia)
LoA_superior <- bias + 1.96 * sd
LoA_inferior <- bias - 1.96 * sd

error <- (LoA_superior - LoA_inferior)/mean(FO) * 100
error

# install.packages("BlandAltmanLeh")
library(BlandAltmanLeh)

bland.altman.plot(group1 = FO, group2 = EM, main = "Bland-Altman Plot", xlab =
  ylab = "", conf.int = 0.95)

# Paso 4: verificar la distribución normal de los datos
# install.packages("psych")
library(psych)
multi.hist(x = data2$diferencia, dcol = c("blue", "red"),
  dltty = c("dotted", "solid"), main = "")

qqnorm(data2$diferencia, main = "FO-EM")
qqline(data2$diferencia)

library(nortest)
lillie.test(data2$diferencia)

# la diferencia de Los datos son normales

bland.altman.stats(FO, EM)$CI.lines
```

```
#####
sd <- sd(data2$diferencia)
bias <- mean(data2$diferencia)
LoA_superior <- bias + 1.96 * sd
LoA_inferior <- bias - 1.96 * sd

error <- (LoA_superior - LoA_inferior)/mean(FO) * 100
error

# install.packages("BlandAltmanLeh")
library(BlandAltmanLeh)

bland.altman.plot(group1 = FO, group2 = EM, main = "Bland-Altman Plot", xlab = "",
  ylab = "", conf.int = 0.95)

# Paso 4: verificar la distribución normal de los datos
# install.packages("psych")
library(psych)
multi.hist(x = data2$diferencia, dcol = c("blue", "red"),
  dltty = c("dotted", "solid"), main = "")

qqnorm(data2$diferencia, main = "FO-EM")
qqline(data2$diferencia)

library(nortest)
lillie.test(data2$diferencia)

# la diferencia de Los datos son normales

bland.altman.stats(FO, EM)$CI.lines
```

ANEXO C

N	Fecha	Día	Hora	H1	H2	H3	H4	T1	T2	T3	T4	MH1	MH2	MH3	MH4
1	29/11/2019	1	0	25,99	28,48	30,37	24,25	23,95	22,17	26,41	24,02	91,67	88,83	93,00	80,00
2		1	1	25,49	28,32	30,30	24,20	23,59	24,88	19,75	23,64	91,56	79,35	77,32	80,00
3		1	2	25,27	26,90	29,73	24,12	23,28	21,12	20,63	23,31	91,92	87,92	85,02	80,00
4		1	3	25,02	27,84	30,31	24,03	22,92	19,67	33,69	22,99	92,00	88,78	93,00	80,02
5		1	4	24,83	27,10	30,38	23,97	22,59	23,08	27,91	22,72	91,61	80,37	76,79	80,00
6		1	5	24,68	27,59	30,46	23,88	22,27	23,00	30,33	22,44	92,00	85,10	82,02	80,00
7		1	6	24,50	27,50	30,41	23,80	22,02	19,97	30,12	22,14	92,00	93,00	95,00	80,00
8		1	7	24,32	27,18	30,10	23,86	22,08	22,27	23,27	22,20	92,00	89,10	94,32	80,77
9		1	8	24,36	28,29	30,00	24,06	22,45	22,55	27,28	22,63	92,08	88,63	93,00	80,42
10		1	9	24,13	28,62	29,83	24,50	23,51	24,04	21,86	23,92	92,03	80,86	79,38	79,73
11		1	10	10,65	17,80	23,17	21,51	24,35	24,53	21,25	25,23	91,57	80,18	77,32	78,70
12		1	11	17,76	18,97	23,41	19,19	27,30	25,00	27,73	27,90	93,03	80,18	79,77	76,79
13		1	12	13,91	16,54	16,73	13,80	26,66	29,50	23,94	28,34	84,41	80,37	72,03	79,44
14		1	19	29,15	25,36	23,98	28,92	24,07	24,44	24,14	24,17	95,00	95,00	94,66	95,00
15		1	20	27,71	26,68	23,64	27,09	24,38	25,01	24,31	24,77	95,00	95,00	86,67	95,00
16		1	21	27,12	27,36	23,70	26,10	24,12	24,55	23,26	24,54	95,00	95,00	83,63	95,00
17		1	22	26,84	27,80	23,81	25,26	23,51	24,00	23,62	24,07	93,00	95,00	83,00	95,00
18		1	23	26,75	28,08	23,93	25,39	21,05	23,47	24,88	23,68	93,00	95,00	83,00	95,00
19	30/11/2019	2	0	26,71	28,57	24,06	25,36	22,71	23,10	24,30	23,24	95,00	94,75	83,00	94,88
945		42	3	29,07	31,41	28,42	25,00	18,12	17,00	18,42	17,91	92,81	93,00	85,00	95,00

946		42	4	29,19	31,57	28,43	25,03	17,90	16,49	18,15	17,91	92,00	94,95	95,00	93,00
947		42	5	29,63	31,66	28,41	25,33	17,95	16,90	19,84	17,91	92,52	93,00	94,00	93,00
948		42	6	29,73	30,65	28,01	25,45	17,64	14,69	22,07	15,87	92,00	91,00	93,00	93,85
949		42	7	29,49	29,76	27,52	25,26	17,16	14,00	20,96	16,57	92,00	91,00	85,00	95,00
950		42	8	28,89	29,01	26,65	24,79	17,82	16,90	18,42	17,70	92,41	95,00	94,63	95,00
951		42	9	23,68	29,59	24,60	24,93	19,53	17,00	18,15	18,80	93,00	95,00	93,00	94,59
952		42	10	30,81	30,94	26,56	25,88	20,90	19,97	19,99	20,58	93,00	93,00	93,00	92,84
953		42	11	31,33	31,42	26,26	26,23	24,94	22,15	20,19	23,88	95,00	94,72	95,00	94,60
954		42	12	31,14	32,01	26,09	26,15	26,94	24,67	32,27	27,25	94,90	95,00	91,15	93,06
955		42	13	31,84	34,29	27,03	26,69	28,86	27,00	27,77	28,28	95,00	95,00	93,00	93,00
956		42	14	34,42	36,95	28,26	28,66	29,89	28,00	22,40	32,18	94,86	95,00	94,63	91,22
957		42	15	33,21	36,23	28,36	27,78	29,92	27,78	31,33	28,95	95,00	95,00	95,00	92,39
958		42	16	31,99	36,76	28,99	26,91	29,13	26,80	30,43	28,65	95,00	95,00	93,00	93,58
959		42	17	31,95	37,38	29,78	26,83	27,79	25,41	36,98	27,15	95,00	95,00	95,00	95,00
960		42	18	31,95	37,68	30,35	26,75	25,75	23,08	30,87	25,67	95,00	95,00	95,00	93,00
961		42	19	32,08	36,91	30,25	26,81	24,28	22,11	25,20	23,92	95,00	95,00	89,41	94,64
962		42	20	31,65	34,58	29,27	26,50	23,70	25,75	19,41	25,23	95,00	91,00	93,00	91,19
963		42	21	31,41	32,99	28,78	26,37	22,07	21,91	30,29	22,20	94,57	92,00	93,05	92,98
964		42	22	31,05	32,19	28,57	26,21	21,16	23,76	25,56	23,67	93,71	91,00	93,00	91,19
965		42	23	30,79	31,46	28,36	26,10	20,22	17,91	21,09	19,11	93,00	92,89	85,23	94,67

FO1	EM1	FO2	EM2	FO3	EM3	FO4	EM4
25,49	22,86	28,32	36,55	30,30	28,58	24,20	21,39
25,27	22,89	26,90	35,48	29,73	28,52	24,12	21,39
25,02	22,95	27,84	34,94	30,31	28,44	24,03	21,31
24,83	22,95	27,10	34,58	30,38	28,57	23,97	21,23
24,68	23,31	27,59	34,47	30,46	28,75	23,88	21,53
24,50	23,07	27,50	34,30	30,41	28,70	23,80	21,28
24,32	23,14	27,18	33,91	30,10	28,57	23,86	21,22
24,36	24,12	28,29	37,21	30,00	28,31	24,06	21,88
24,13	24,68	28,62	34,98	29,83	27,47	24,50	22,29
10,65	24,89	17,80	34,01	23,17	27,42	21,51	23,39
17,76	24,63	18,97	33,54	23,41	26,37	19,19	22,64
13,91	26,10	16,54	33,54	16,73	26,56	13,80	23,94
29,15	26,84	25,36	34,28	23,98	27,15	28,92	24,72
27,71	26,19	26,68	32,92	23,64	27,78	27,09	24,36
27,12	26,27	27,36	32,58	23,70	27,93	26,10	24,57
26,84	26,50	27,80	32,43	23,81	27,87	25,26	24,98
26,75	26,29	28,08	32,24	23,93	27,96	25,39	24,98
26,71	25,64	28,57	32,78	24,06	28,66	25,36	24,56
26,71	24,85	28,71	33,15	24,11	28,92	25,38	23,80
26,69	24,24	28,89	33,01	24,21	29,13	25,43	23,16
26,72	23,97	29,17	32,87	24,29	29,09	25,37	22,87
26,74	23,96	29,20	32,81	24,37	29,07	25,42	22,78
26,74	23,86	29,39	32,77	24,42	29,02	25,37	22,57
26,74	24,04	29,47	32,68	24,52	29,03	25,38	22,71
26,80	24,10	29,67	32,72	24,60	29,07	25,46	22,56
26,95	24,20	29,94	32,48	24,70	29,05	25,45	22,67
27,16	24,21	30,31	33,49	24,81	29,15	25,62	22,68
27,82	24,14	30,59	33,58	25,07	29,15	25,87	22,33
28,19	24,06	30,87	33,88	25,25	29,16	26,01	22,43
28,43	24,03	31,40	34,98	25,63	29,22	27,26	22,44
27,98	23,90	32,03	35,04	28,81	29,01	28,65	22,26
28,40	24,16	32,40	35,15	29,69	28,82	28,82	22,32
28,71	24,96	32,78	34,70	26,65	28,47	28,83	22,96
28,76	25,27	33,04	33,15	26,81	28,01	29,03	23,23
28,32	25,70	33,17	32,30	26,87	27,54	29,20	23,70
27,59	26,60	32,83	32,54	29,95	27,52	28,71	25,58
27,24	26,89	32,42	32,25	31,62	27,91	28,18	25,05
26,98	27,55	32,20	32,18	32,09	28,51	27,91	25,53

24,72	24,15	30,02	33,73	25,59	33,83	23,10	26,42
24,74	23,52	30,04	32,94	25,80	33,82	23,04	26,03
24,77	25,76	30,06	32,76	26,27	32,52	23,03	25,77
24,81	25,06	30,11	32,93	25,95	30,73	22,98	25,38
24,81	24,29	30,11	32,33	26,12	29,37	22,94	24,43
25,00	23,99	30,20	32,98	26,36	29,12	22,97	24,27
25,25	25,35	30,38	32,93	26,96	28,87	23,08	24,58
25,49	25,93	30,55	32,60	27,02	28,49	23,18	24,27
25,84	25,56	30,75	32,48	27,19	28,10	23,17	23,95
26,31	24,24	31,09	32,87	27,79	28,06	23,36	23,81
26,25	22,93	31,27	33,04	26,59	28,04	23,35	23,71
25,87	23,32	31,83	33,14	25,52	28,04	22,74	23,65
26,24	23,90	32,63	33,31	26,20	28,28	22,94	23,61
26,01	24,66	32,29	34,54	25,59	27,88	22,49	23,62
25,15	23,27	31,68	34,68	25,13	27,73	21,93	23,85
23,58	24,46	30,08	34,90	25,75	27,37	20,78	24,31
23,51	27,60	30,16	34,93	25,91	26,85	20,73	25,43
22,62	26,93	29,16	36,51	23,30	26,67	20,03	26,08
22,72	27,46	29,29	38,20	23,24	26,48	20,15	26,63
22,94	28,59	29,47	39,95	23,71	27,00	20,35	27,08
22,99	29,08	29,42	39,50	24,93	28,30	20,42	27,68
23,06	30,49	29,39	39,89	24,78	29,46	20,50	27,28
23,05	31,11	29,40	41,60	25,92	30,40	20,55	27,26
23,11	28,56	29,43	38,00	25,52	30,63	20,65	27,65
23,21	27,22	29,45	36,44	26,17	30,25	20,78	27,18
23,21	27,65	29,42	35,96	24,88	29,57	20,80	26,19
23,27	26,41	29,45	34,95	25,32	28,58	20,82	25,42
23,26	26,87	29,44	34,92	25,21	28,08	20,86	24,99
23,34	26,79	29,44	34,67	25,50	28,12	20,88	24,75
23,50	27,95	29,56	34,56	25,25	28,23	20,99	24,95
23,83	29,07	29,72	34,46	25,39	28,16	21,11	24,41
24,18	30,59	29,97	34,76	25,65	28,27	21,26	24,19
24,48	30,43	30,20	35,29	25,87	28,44	21,34	24,14
25,21	28,54	30,66	35,45	26,71	28,70	21,59	24,28
26,07	26,49	31,40	35,26	26,79	28,60	21,91	24,45
27,13	26,18	32,82	35,39	28,32	28,31	22,24	24,52
27,51	25,87	33,36	35,56	28,96	28,13	22,42	24,58
26,47	25,83	32,51	35,77	28,14	27,75	21,60	24,16