



PERÚ

Ministerio del Ambiente

Organismo de Evaluación y Fiscalización Ambiental - OEFA

"Año del Buen Servicio al Ciudadano"



INFORME N° 029 -2017-OEFA/DE-SDLB-CEAPIQ

A : LÁZARO WALTHER FAJARDO VARGAS
Subdirector de la Línea de Base y Agentes Contaminantes

DE : JUAN GERARDO SANTOYO FIGUEROA
Coordinador de Evaluaciones Ambientales en Pesquería, Industria y Otros

EDGAR LUIS CÁCERES ANGULO
Tercero Locador

ASUNTO : Informe del cálculo de niveles de fondo y niveles de referencia de metales en suelos para la Evaluación Ambiental en el ámbito de la zona Industrial de Ventanilla - Mi Perú, de la Provincia Constitucional del Callao, durante el año 2017.

FECHA : Lima, 21 AGO. 2017 2017-101-25971

Tenemos el agrado de dirigimos a usted con relación al asunto de la referencia, a fin de informar lo siguiente:

I. INFORMACIÓN GENERAL

1. Detalles de la evaluación ambiental:

a.	Zona evaluada	Zona Industrial de Ventanilla – Mi Perú		
b.	Ámbito de influencia	Distritos de Ventanilla y Mi Perú, de la Provincia Constitucional del Callao		
c.	Problemática de la zona evaluada	Presencia de metales pesado en suelo		
d.	¿A pedido de qué se realizó la actividad?	Planefa 2017		
e.	¿Se realizó en el marco de un espacio de diálogo, mesa de diálogo o mesa de desarrollo?	SI	NO	X

2. Detalles del diagnóstico ambiental:

a.	Componentes y puntos de monitoreo evaluados	Suelos	20 muestras compuestas
----	---	--------	------------------------

3. Equipo profesional:

El profesional que participó en el análisis de los resultados en la siguiente tabla:

Nombre	Profesión
Edgar Luis Cáceres Angulo	Lic. en Estadística



el



PERÚ

Ministerio
del Ambiente

Organismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

II. OBJETIVO

4. Calcular los niveles de fondo y niveles referencia para 30 metales evaluados en los suelos de la zona Industrial Ventanilla - Mi Perú, desarrollado en los Anexos, que se adjunta y forma parte del presente informe.

III. CONCLUSIÓN

5. En vista que el informe del cálculo de niveles de fondo y niveles de referencia de metales en suelos para la Evaluación Ambiental en el ámbito de la zona Industrial de Ventanilla - Mi Perú cuenta con el sustento técnico requerido, se sirve elevar dicho documento a la Coordinación de Evaluaciones Ambientales en Pesquería Industria y Otros, así como a la Subdirección de la Línea de Base y Agentes Contaminantes a efectos que se proponga su aprobación ante la Dirección de Evaluación.

Atentamente,

Juan Gerardo Santoyo Figueroa
Coordinador de Evaluaciones
Ambientales en Pesquería, Industria y Otros
Dirección de Evaluación
Organismo de Evaluación y Fiscalización
Ambiental - OEFA

Edgar Luis Cáceres Angulo
Tercero Evaluador
Dirección de Evaluación
Organismo de Evaluación y Fiscalización
Ambiental - OEFA



PERÚ

Ministerio
del Ambiente

Organismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Lima, 21 AGO. 2017

Visto el Informe N° 029-2017-OEFA/DE-SDLB-CEAPIO y habiéndose verificado que su contenido se encuentra enmarcado en el ejercicio de la función evaluadora, así como su coherencia normativa; el Subdirector de la Línea de Base y Agentes Contaminantes recomienda su APROBACIÓN a la Dirección de Evaluación, razón por la cual se TRASLADA el presente Informe.

Atentamente,

LÁZARO WALTHER FAJARDO VARGAS

Subdirector de la Línea de Base y Agentes Contaminantes
Dirección de Evaluación
Organismo de Evaluación y Fiscalización Ambiental - OEFA

21 AGO. 2017

Visto el Informe N° 029-2017-OEFA/DE-SDLB-CEAPIO y en atención a la recomendación de la Coordinación de Evaluaciones Ambientales en Pesquería, Industria y otros, así como de la Subdirección de la Línea de Base y Agentes Contaminantes, la Dirección de Evaluación ha dispuesto aprobar el presente Informe.

Atentamente,

FRANCISCO GARCÍA ARAGÓN

Director de Evaluación
Dirección de Evaluación
Organismo de Evaluación y Fiscalización Ambiental – OEFA

100

100



PERÚ

Ministerio
del Ambiente

Organismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"



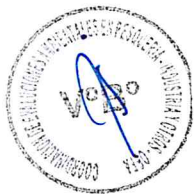
Organismo
de Evaluación
y Fiscalización
Ambiental

**INFORME DEL CÁLCULO DE NIVELES DE FONDO Y NIVELES
DE REFERENCIA DE METALES EN SUELOS PARA LA
EVALUACIÓN AMBIENTAL EN EL ÁMBITO DE LA ZONA
INDUSTRIAL DE VENTANILLA - MI PERÚ, DE LA PROVINCIA
CONSTITUCIONAL DEL CALLAO, DURANTE EL AÑO 2017**



**COORDINACIÓN DE EVALUACIONES AMBIENTALES EN
PESQUERÍA, INDUSTRIA Y OTROS**

DIRECCIÓN DE EVALUACIÓN



2017

ÍNDICE DEL CONTENIDO

1.	INTRODUCCIÓN	1
2.	ANTECEDENTES	1
3.	JUSTIFICACIÓN	2
4.	OBJETIVOS	2
4.1.	Objetivo general.....	2
4.2.	Objetivos específicos.....	2
5.	MARCO CONCEPTUAL.....	3
5.1.	Niveles de fondo y referencia en suelos.....	3
5.1.1.	Niveles de fondo.....	3
5.1.2.	Nivel de referencia.....	3
5.2.	Estimación de parámetros	4
5.2.1.	La estimación por Máxima Verosimilitud.....	4
5.2.2.	El Estimador Insesgado de Mínima Varianza	5
5.2.3.	Estimación por Método de Momentos (MME)	6
5.3.	Métodos no paramétrico Kaplan-Meier (KM) para la estimación de parámetros	6
5.4.	Distribución de probabilidad de los modelos para el cálculo del nivel de fondo y referencia.....	6
5.4.1.	Distribución de probabilidad	6
5.5.	Prueba para la detección de valores atípicos.....	9
5.5.1.	Prueba de Dixon.....	9
5.6.	Pruebas de bondad de ajuste	10
5.6.1.	Prueba de Anderson-Darling (A-D).....	10
5.6.2.	Prueba de Shapiro-Wilk (S-W).....	11
5.6.3.	Prueba de Kolmogorov-Smirnov (K-S).....	12
5.6.4.	Prueba de Lilliefors.....	12
5.6.5.	Gráfico Q-Q.....	13
5.7.	Métodos ROS para imputar valores no detectados (NDs) usados para la bondad de ajuste	13
5.8.	Teorema del Límite Central.....	13
5.9.	Desigualdad de Chebyshev	13
5.10.	Métodos de estimación de parámetros para distribuciones no discernibles: (Métodos de remuestreos: Bootstrap o Jackknife)	14
5.10.1.	Método de Jackknife (Quenouille, 1949 – Tukey, 1958)	14
5.10.2.	Método de Bootstrap (Efron, 1979).....	14
5.11.	Coeficiente de Correlación.....	15
5.12.	Coeficiente de variación.....	15
5.13.	Principio de parsimonia.....	16
6.	METODOLOGÍA.....	16
7.	RESULTADOS.....	21
8.	DISCUSIONES.....	22
9.	CONCLUSIONES.....	23
10.	REFERENCIAS BIBLIOGRÁFICAS	23
11.	ANEXOS	24





ÍNDICE DE FIGURAS

Figura 5-1. Representaciones graficas del UCL y UTL 4

Figura 5-2. Funciones de densidad de probabilidad de dos estimadores insesgados diferentes. Fuente: Devore, J. L. ,2008..... 5

Figura 6-1. Esquema del tratamiento de datos en ProUCL 5.1 17

Figura 6-2. Diagrama Q-Q que muestra potenciales valores atípicos 18

Figura 6-3. Resumen las recomendaciones brindadas en la guía técnica de ProUCL 5.1. 20

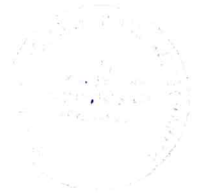


ÍNDICE DE TABLAS

Tabla 7-1. Resultados de los niveles de fondo y niveles de referencia de 30 metales evaluados.....21



Handwritten signature



I. INTRODUCCIÓN

1. La calidad del suelo en la Zona Industrial de los distritos de Ventanilla y Mi Perú, de la Provincia Constitucional del Callao se presenta afectada por la presencia de metales producto de las actividades industriales y otras.
2. La determinación de los niveles naturales (nivel de fondo) de las concentraciones de los metales en suelos y de máxima concentración (nivel de referencia) es imprescindible para poder determinar el grado de enriquecimiento o contaminación del medio (Ezpeleta, A. M. ,2003).
3. En este informe se empleó el software ProUCL versión 5.1 para calcular los niveles de fondo y referencia, este software fue desarrollado inicialmente por la EPA, el cual posee un conjunto de herramientas estadísticas para obtención de estos niveles a través de los límites superiores de intervalos estadísticos.
4. Son tres puntos importantes a considerar al momento de realizar el cálculo de nivel de fondo y de referencia:
5. El primero, la detección de valores de concentración por debajo del límite de cuantificación, valores perdidos o valores atípicos, en cada una de las variables.
6. El segundo, la bondad de ajuste de cada una de las variables; es decir; la forma aproximada de la distribución de las variables, y al ser continuas estas variables, los modelos que se disponen son la distribución normal, lognormal y gamma, los cuales están implementados en ProUCL 5.1.
7. El tercero, el cálculo del nivel de fondo a través del límite superior del intervalo de confianza para la media de la distribución evaluada, para un nivel de confianza de 95% (UCL). Y el nivel de referencia a través del límite superior del intervalo de confianza para el percentil 95 de la distribución evaluada, para un nivel de confianza de 95% (UTL) o a través de métodos no paramétricos en el caso de tener distribuciones desconocidas.



II. ANTECEDENTES

8. Mediante la investigación de Daniel, D.L. (2015) denominada: *A Case Study Perspective on Working with ProUCL and a State Environmental Agency in Determining Background Threshold Values* propone la aplicación del ProUCL, el cual es un software puesto a disposición por la Agencia de Protección Ambiental (EPA) de los Estados Unidos de América para proporcionar a los científicos ambientales las mejores herramientas para realizar análisis estadísticos. ProUCL ha estado en elaboración durante más de diez años y actualmente está en su quinta versión. Con el tiempo, se ha incluido herramientas para análisis más sofisticados y adecuados.
9. La investigación de Hernández-Crespo, M. Martín (2015) denominada; *Determination of background levels and pollution assessment for seven metals (Cd, Cu, Ni, Pb, Zn, Fe, Mn) in sediments of a Mediterranean coastal lagoon*; el cual expone que la determinación de los niveles de fondo de los metales traza en sedimentos son un punto clave para la correcta evaluación del grado de





contaminación. Este estudio demuestra la conveniencia de integrar métodos geológicos y estadísticos para la determinación adecuada de los niveles de fondo, aplicándolo a los sedimentos del lago Albufera, un humedal de importancia que es altamente eutrofizada. Los límites superiores de las poblaciones de fondo, definidos como la media + 2σ y expresados en mg/kg, son Cd (0,38), Cu (28,8), Ni (25,9), Pb (25,5), Zn (88,6), Fe (2,2) and Mn (345,7).

10. La guía del Estado de Connecticut, EEUU, elaborado por el Departamento de Energía y Protección Ambiental y el Departamento de Salud Pública (DPH, 2014), junto con profesionales ambientales del sector privado, desarrollaron la guía para calcular el límite superior del intervalo de confianza para la media al 95% (UCL) para la evidencia del cumplimiento con las normas ambientales y guiar a los administrados en la realización del cálculo del (UCL) en el conjuntos de datos de suelo y agua subterránea para demostrar cumplimiento con ciertos criterios del *Connecticut Remediation Standard Regulations* (RSR).

III. JUSTIFICACIÓN

11. La evaluación ambiental en el ámbito de la zona industrial de Ventanilla – Mi Perú se desarrolló en marco de la función evaluadora establecida en el Plan Anual de Evaluación y Fiscalización Ambiental del OEFA (Planefa) correspondiente al año 2017, el cual fue aprobado por resolución de Consejo Directivo N° 004-2017-OEFA/CD del 30 de enero de 2017.
12. De acuerdo al Planefa, la función evaluadora consiste en analizar los factores externos que inciden en la calidad del ambiente de las áreas de influencia de las actividades fiscalizables de competencia del OEFA. En ese sentido, el presente informe propone el cálculo de niveles de fondo y niveles de referencia de metales, como una herramienta estratégica dirigida hacia la evaluación de la calidad de los suelos en el marco de la evaluación ambiental en el ámbito de la zona industrial de Ventanilla-Mi Perú, Provincia Constitucional Del Callao, en el año 2017.
13. La metodología empleada en el presente informe, se encuentra recomendada por la Agencia de Protección Ambiental de los Estados Unidos de América (U.S EPA) y en diversas investigaciones, que le brinda una base y un soporte para su aplicación.

IV. OBJETIVOS

4.1. Objetivo general

14. Calcular los niveles de fondo y niveles referencia para 30 metales evaluados en los suelos de la zona de Ventanilla y Mi Perú.

4.2. Objetivos específicos

15. Hallar las estadísticas descriptivas para conocer el nivel de concentración de cada una de las variables.
16. Evaluar la bondad de ajuste de la distribución de cada uno de los metales.
17. Calcular el nivel de fondo y el nivel de referencia de cada uno de los metales.

V. MARCO CONCEPTUAL

5.1. Niveles de fondo y referencia en suelos

18. El Estándar de Calidad Ambiental (ECA) para suelo aprobado mediante Decreto Supremo N° 002-2013-MINAM, define al nivel de fondo como la concentración en el suelo de los químicos regulados que no fueron generados por la actividad objeto de análisis y que se encuentran en el suelo de manera natural o fueron generados por alguna fuente antropogénica ajena a la considerada.
19. El *background threshold value* (BTV) o valor de umbral de fondo es la estimación del límite superior de la concentración de contaminantes de fondo (ya sea natural o antropogénicos) utilizado para representar contaminantes ambientales no relacionados específicamente con el sitio bajo investigación (Schumacher, B., Singh, A., & Singh, A. K., 2015).
20. Asimismo, la U.S. EPA (2002) define dos tipos de fondo, el primero corresponde al fondo de origen natural definido como las sustancias presentes en el medio ambiente de formas que no han sido influenciadas por la actividad humana. El segundo corresponde al fondo antropogénico el cual se define como las sustancias naturales y humanas presentes en el medio ambiente como resultado de actividades humanas no específicamente relacionadas con el sitio en cuestión.

5.1.1. Niveles de fondo

21. Según Ramos (2002), no existe un criterio común universalmente aceptado para la definición de los niveles de fondo. Sin embargo, diferentes autores como Chen *et al.* (1999), definen al nivel de fondo como el rango de concentración alrededor de la media, en suelos no contaminados. Asimismo, el *Model Toxics Control Act* (MTCA, Ch. 173-340-200 WAC), lo define como la concentración de sustancias peligrosas, presentes de forma sistemática en el medio natural que no han sido influenciadas por actividades humanas localizadas.
22. En ese mismo sentido, De Miguel, E. (2002) manifiesta que los valores de fondo caracterizan la situación real de los suelos naturales de una región en el momento de su estudio. En consecuencia, dichos valores de fondo describen la distribución de valores de concentración de los diferentes elementos químicos, en los suelos de una región no afectada por actividades antrópicas.

5.1.2. Nivel de referencia

23. Para el presente informe, se entiende por valor de referencia a aquel valor tal que su superación por una unidad muestral de la población actual caracterizada sea poco probable y que, por tanto, si no ha habido modificaciones en las condiciones de muestreo, sea también poco probable su superación por una unidad muestral obtenida posteriormente (De Miguel, E., 2002).
24. El *Upper Confidence Limit* (UCL 95%) o límite superior de confianza (es la cota superior del intervalo de confianza para la media poblacional), el cual comúnmente se considera a un 95% de confianza. Este valor frecuentemente se usa como un umbral fijo para comparar con los datos muestreados de un sitio y determinar que este se encuentre limpio.



25. El *Upper Tolerance Limit* (UTL 95-95%) o límite superior de tolerancia (es la cota superior del intervalo de confianza para el percentil 95), el cual comúnmente se considera a un 95% de confianza. Un límite de tolerancia superior (UTL) está diseñado para contener, un porcentaje de la población (es decir, 95%) de las posibles concentraciones de fondo, se utiliza para comparar con los datos muestreados de un sitio y establecer que este se encuentre contaminado.
26. Los límites de confianza pueden ser paramétricos o no paramétricos. Para el cálculo de los límites de confianza paramétricos, se debe conocer la distribución estadística subyacente para seleccionar el límite de confianza apropiado.

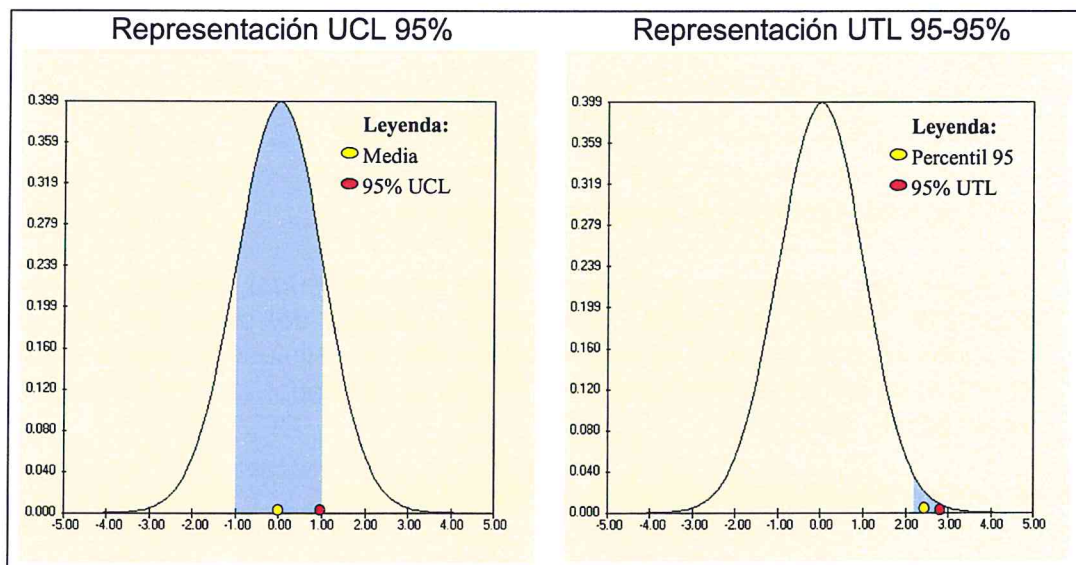


Figura 5-1. Representaciones del UCL y UTL

5.2. Estimación de parámetros

27. La estimación de parámetros es brindar métodos que permitan determinar con cierta precisión, aspectos de una población cuyo valor concreto describe la forma de la distribución de la variable.

5.2.1. La estimación por Máxima Verosimilitud

28. La estimación por Máxima Verosimilitud o *Maximum Likelihood Estimation* (MLE) es un método de optimización que supone que la distribución de probabilidad de las observaciones es conocida. El principio de MLE es la siguiente:
29. El método consiste en encontrar el valor del parámetro que maximiza el valor de la Función de Verosimilitud. Para una muestra aleatoria simple y_1, y_2, \dots, y_n de observaciones independientes e idénticamente distribuidas (i.i.d.) de una distribución con función de probabilidad o de densidad $f(y_i; \theta)$, la función L , se denomina Función de Verosimilitud:

$$y_i \sim f(y_i; \theta) \quad (1)$$

$$L(\theta) = f(y_1, y_2, \dots, y_n; \theta) = \prod_{i=1}^n f(y_i; \theta) \quad (2)$$

30. Si, para un determinado valor de θ , la verosimilitud es pequeña, es poco probable que ese θ sea el valor correcto que ha generado los datos que observamos.
31. Si, para un determinado valor de θ , la verosimilitud es grande, es bastante probable que ese θ sea el valor correcto que ha generado los datos que observamos.
32. Por tanto, tenemos que elegir el θ que maximizar $L(\theta)$. Es decir, el estimador MLE será el que satisfaga la condición de primer orden:

$$\left. \frac{\partial L(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0 \quad (3)$$

33. O bien:

$$\left. \frac{\partial \ln L(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0 \quad (4)$$

5.2.2. El Estimador Insesgado de Mínima Varianza

34. El Estimador Insesgado de Mínima Varianza o *Minimum Variance Unbiased Estimator* (MVUE), se define como aquel estimador del parámetro que entre todos los insesgados, es el de menor varianza (dicho estimador no existe siempre, si existe, es único); en otras palabras, $\hat{\theta}$ será un estimador insesgado de mínima varianza, si y solo si:

- El estimados $\hat{\theta}$ es insesgado, es decir; $E(\hat{\theta}) = \theta$ (5)
- $V(\hat{\theta}) < V(\tilde{\theta})$ donde $\tilde{\theta}$ es cualquier estimador insesgado. (6)

35. La Figura 5-2, ilustra las funciones de densidad de probabilidad de los dos estimadores insesgados, donde $\hat{\theta}_1$ tiene una varianza más pequeña que $\hat{\theta}_2$. Entonces es más probable que produzca una estimación próxima al valor verdadero θ que $\hat{\theta}_2$. El estimador insesgado con varianza mínima es, en cierto sentido, el que tiene más probabilidades entre todos los estimadores insesgados de producir una estimación cercana al verdadero θ .

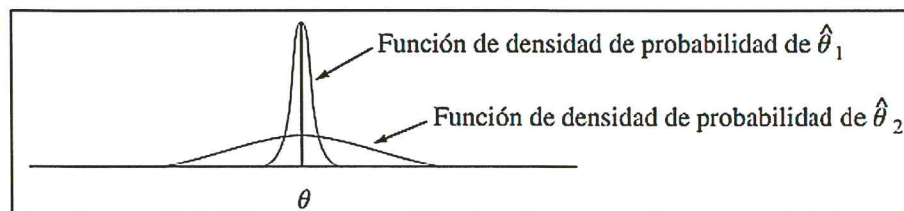


Figura 5-2. Funciones de densidad de probabilidad de dos estimadores insesgados diferentes. Fuente: Devore, J. L., 2008

36. Cuando existe este estimador, puede optarse por dos técnicas para encontrarlo:
 - Técnica basada en la cota inferior de Cramér-Rao.

- Técnica basada en la noción de estadísticos suficientes.

37. Cuando no existe, se pueden agregar restricciones. Por ejemplo, imponer que el estimador, además de insesgado, sea lineal en los datos (estimador lineal).

5.2.3. Estimación por Método de Momentos (MME)

38. Este método fue propuesto por Pearson (1857-1936) y consiste en igualar un determinado número de momentos teóricos de la distribución de la población con los correspondientes momentos muestrales, para obtener una o varias ecuaciones que, resueltas, permitan estimar los parámetros desconocidos de la distribución poblacional (Freund, J. E., Miller, I., & Miller, M., 2000).

39. Sea una muestra aleatoria simple y_1, y_2, \dots, y_n , de una distribución con función de densidad $f(y_i; \theta)$, el k -ésimo momento sobre este conjunto de observaciones, es la media de sus k -ésimas potencias y se denota por m_k ; simbólicamente:

$$m'_k = \frac{\sum_{i=1}^n y_i^k}{n} = \int_{-\infty}^{\infty} y_i^k f(y_i; \theta) dy \quad (7)$$

5.3. Métodos no paramétrico Kaplan-Meier (KM) para la estimación de parámetros

40. El método de estimación de KM (Kaplan y Meier, 1958) también conocido como método de Estimación de Límite de Producto (PLE), se basa en una estimación de la función de distribución, como la función de distribución de la muestra, excepto que el método KM se ajusta para censurar.

41. El método KM se utiliza comúnmente en el análisis de supervivencia; asumen que los sujetos que se van del estudio antes de su finalización (censurados a la derecha) se habrían comportado del mismo modo que los que han sido seguidos hasta el final.

42. El método KM estima la media poblacional, desviación estándar (sd) y el error estándar (SE) de la media para los conjuntos de datos censurados a la izquierda. Puede consultarse; Kaplan y Meier (1958) y el informe preparado por Bechtel Jacobs Company para el DOE (2000). Las propiedades del método KM están bien investigadas (Gillespie, Chen *et al.*, 2010). Específicamente, el estimador de KM representa un estimador consistente y para grandes conjuntos de datos, el estimador de KM es asintóticamente eficiente y normalmente distribuido (Gu, Zhang 1993).

5.4. Distribución de probabilidad de los modelos para el cálculo del nivel de fondo y referencia

5.4.1. Distribución de probabilidad

43. Si X es una variable aleatoria (discreta), la función dada por $f(x)=P(X=x)$ para cada x dentro del intervalo de X se llama la distribución de probabilidad de X .

44. Para el caso continuo, la función dada por $f(x) = P(a \leq x \leq b) = \int_a^b f(x)dx$ para cualquier constante real a y b con $a \leq b$, se llama una función de densidad de probabilidad o sencillamente distribución de probabilidad de X .

45. Además, sin importar si la variable es discreta o continua, debe satisfacer las siguientes condiciones:

$$f(x) \geq 0, \forall x \in \mathbb{R} \quad (8)$$

$$\int_{-\infty}^{\infty} f(x) dx = \sum f(x) = 1 \quad (9)$$

46. El software ProUCL 5.1; emplea tres diferentes distribuciones de probabilidad los cuales son; la distribución Normal, Lognormal y Gamma.

a. Distribución Normal

47. Se dice que una variable aleatoria X sigue una distribución normal si su función de densidad es:

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ para } -\infty \leq x \leq +\infty \quad (10)$$

48. Donde su media (μ_x) y varianza (σ_x^2) es;

$$\mu_x = E(X) = \mu \quad (11)$$

$$\sigma_x^2 = V(X) = \sigma^2 \quad (12)$$

b. Distribución Lognormal

49. Dada una variable aleatoria $X \sim N(\mu, \sigma^2)$, se dice que la variable aleatoria $Y = e^X$, sigue una distribución lognormal. La función de densidad de dicha distribución es:

$$f_y(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln(y)-\mu)^2}, \text{ para } y > 0 \quad (13)$$

50. Donde su media (μ_y) y varianza (σ_y^2) es;

$$\mu_y = E(Y) = e^{\mu + \frac{\sigma^2}{2}} \quad (14)$$

$$\sigma_y^2 = V(Y) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) \quad (15)$$

c. Distribución Gamma

51. La variable aleatoria X con función de densidad de probabilidad:

$$f_x(x, \lambda, r) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \text{ para } x > 0 \quad (16)$$



52. Tiene una distribución gamma con parámetros $\lambda > 0$ y $r > 0$. Si r es un entero, entonces X tiene una distribución Erlang.

53. Una equivalencia es;

θ (Parámetro de escala)

k (Parámetro de forma)

$$f_x(x, k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}, \text{ para } x > 0 \quad (17)$$

= 0, caso contrario

Donde $k > 0$ y $\theta > 0$

54. La distribución de gamma produce valores más fiables y estables del 95% de UCL en la práctica. Por lo tanto, es ideal que un conjunto de datos ambientales siga una distribución gamma. Pues si el conjunto está sesgado, entonces un 95% UCL de la media de la población debe calcularse utilizando una distribución gamma. Para los conjuntos de datos que siguen una distribución gamma, el UCL ajustado del 95% de la media, es óptima (Bain y Engelhardt 1991) y proporciona aproximadamente una confianza del 95% de la media de la población, $\mu_1 = k\theta$ (Singh, Singh y Laci, 2002).

k hat = MLE del parámetro de forma de una distribución gamma.

k star = Corrección de sesgo MLE del parámetro de forma de una distribución gamma.

θ hat = MLE del parámetro de escala de una distribución gamma.

55. Donde su media (μ_x) y varianza (σ_x^2) es;

$$\mu_x = E(X) = \frac{r}{\lambda} = \theta k \quad (18)$$

$$\sigma_x^2 = V(X) = \frac{r}{\lambda^2} = \theta^2 k \quad (19)$$

56. La distribución ji-cuadrada es un caso especial de la distribución gamma en la que $\lambda = 1/2$ y r es igual a uno de los valores $1/2, 1, 3/2, 2, \dots$

57. La función gama es:

$$\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx, \text{ para } r > 0 \quad (20)$$

58. Propiedades:

$$\Gamma(r) = (r-1)\Gamma(r-1) \quad (21)$$

59. Si r es un entero positivo;

$$\Gamma(r) = (r-1)!, \text{ si } r \text{ es un entero positivo} \quad (22)$$



60. Además;

$$\Gamma(1) = 0! \quad (23)$$

$$\Gamma\left(\frac{1}{2}\right) = \pi^{1/2} \quad (24)$$

d. Distribución Erlang

61. La variable aleatoria X que es igual al intervalo que se presenta r fallas en un proceso de Poisson con media $\lambda > 0$, tiene una distribución Erlang con parámetros λ y r . La función de densidad de probabilidad de X es:

$$f_x(x, \lambda, r) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}, \text{ para } x > 0 \text{ y } r = 1, 2, \dots \quad (25)$$

62. Donde su media (μ_x) y varianza (σ_x^2) es;

$$\mu_x = E(X) = \frac{r}{\lambda} \quad (26)$$

$$\sigma_x^2 = V(X) = \frac{r}{\lambda^2} \quad (27)$$

63. Una variable aleatoria Erlang con $r = 1$ es una variable aleatoria exponencial.

5.5. Prueba para la detección de valores atípicos

64. Son observaciones cuyos valores son muy diferentes a las otras observaciones del mismo grupo de datos. Los datos atípicos son ocasionados por:

- Errores de procedimiento.
- Acontecimientos extraordinarios.
- Causas no conocidas.

65. Los datos atípicos distorsionan los resultados de los análisis, y por esta razón hay que identificarlas y tratarlos de manera adecuada, o incluso excluirlos del análisis.

5.5.1. Prueba de Dixon

66. La prueba de Dixon, se emplea cuando el tamaño muestra es menor o igual a 25. También se puede realizar una prueba de normalidad en los datos sin la sospecha de valores atípicos. Debido a que la prueba solo reconoce un valor atípico potencial, es necesario ir retirando uno a uno estos valores, hasta quedarnos sin ningún valor atípico. Para la aplicación de esta prueba, se debe evitar el uso de datos con transformación, tal como una transformación logarítmica antes de aplicar la prueba de Dixon. Si no se cumple el supuesto de normalidad, aplíquenos un método robusto de identificación de valores atípicos. El software Scout 2008 (EPA 2009d) incorpora este tipo de técnica.

67. Los pasos para aplicar la prueba de Dixon son:

68. Paso 1; Ordenar los valores en forma creciente.
69. Paso 2; X_1 es un valor atípico potencial (Caso 1): Calcular la estadística de prueba, (C), donde;

$$C = \frac{X_3 - X_1}{X_{n-2} - X_1} \text{ para } 14 \leq n \leq 25 \quad (28)$$

70. Paso 3; Si (C) excede el valor crítico para el nivel de significancia especificado α , entonces X_1 es un valor atípico y debe investigarse más a fondo.
71. Paso 4; X_n es un valor atípico potencial (Caso 2): Calcular la estadística de prueba, (C), donde;

$$C = \frac{X_n - X_{n-2}}{X_n - X_3} \text{ para } 14 \leq n \leq 25 \quad (29)$$

72. Paso 5: Si (C) excede el valor crítico para el nivel de significación especificado α , entonces X_n es un valor atípico y debe investigarse más a fondo.

5.6. Pruebas de bondad de ajuste

73. Las pruebas de bondad de ajuste se utilizan para determinar si una población tiene una distribución de los datos teórica específica.
74. La distribución de probabilidad de una variable puede ajustarse a más de una distribución de probabilidad específica (como una distribución Normal, Lognormal y Gamma) o no, es decir, puede que no sea posible discernir la distribución que presenta.

75. Cuando una variable se ajusta a más de una distribución específica, aplica el principio de parsimonia en la mayor medida posible, pero con la debida moderación para evitar la sobre simplificación del modelo.

76. Es así que a continuación se abordan las pruebas de ajuste empleadas como parte del tratamiento estadístico de los datos del presente informe:

5.6.1. Prueba de Anderson-Darling (A-D)

77. Se emplea cuando el tamaño muestra es menor o igual a 25 (Abreu, O. V., Alvear, D., Cuesta, A., & Alonso, V.). La prueba Anderson-Darling (Stephens, 1974) se utiliza para probar si una muestra de datos proviene de una población con una distribución específica. Es una modificación de la prueba de Kolmogorov-Smirnov (K-S) y da más peso a los extremos de las distribuciones en comparación a la prueba de K-S.

78. La prueba Anderson-Darling se define como:

H_0 : Los datos siguen una distribución específica.

H_1 : Los datos no siguen la distribución específica.

Estadística:

$$A^2 = - N - S \quad (30)$$

Donde:

$$S = \sum_{i=1}^N \frac{(2i-1)}{N} [\ln F(Y_i) + \ln(1 - F(Y_{N+1-i}))] \quad (31)$$

79. N es el número de datos.
80. F es la función de distribución acumulativa de la distribución especificada. Los Y_i son datos ordenados.
81. F es la función de distribución acumulativa de la distribución especificada. Además, Y_i son los datos ordenados.
82. Es una prueba unilateral y la hipótesis H_0 , de que la distribución es de una forma específica, se rechaza si la estadística de prueba, (A), es mayor que el valor crítico.

5.6.2. Prueba de Shapiro-Wilk (S-W)

83. La prueba Shapiro-Wilk es una prueba usada para probar la normalidad o lognormalidad de un conjunto de datos. ProUCL 5.1 realiza esta prueba para muestras de tamaño hasta 2000 (Royston 1982, 1982a).
84. Para tamaños de muestra menores o iguales a 50, además de una estadística de prueba y un valor crítico, también se muestra un valor p aproximado asociado con la prueba S-W. Para muestras de tamaño mayores a 50, sólo se muestran valores p aproximados. Basado en el nivel de significancia seleccionado y en el estadístico de prueba calculado, ProUCL 5.1 informa al usuario si el conjunto de datos es normalmente distribuido (o lognormalmente).
85. La prueba Shapiro-Wilk se define como:

H_0 : Los datos siguen una distribución normal.

H_1 : Los datos no siguen la distribución normal.

86. El método consiste en comenzar ordenando la muestra de menor a mayor, a continuación, se calcula el estadístico de contraste;

$$W = \frac{1}{ns^2} \left(\sum_{j=1}^h a_{in} (x_{n-i+1} - x_i) \right)^2 \quad (32)$$

87. Siendo s^2 la varianza muestral y los coeficientes a_{in} tabulados,

$$h = \begin{cases} \frac{n}{2}, & \text{Si } n \text{ es par} \\ \frac{n-1}{2}, & \text{Si } n \text{ es impar} \end{cases} \quad (33)$$

88. H_0 será rechazada si la estadística de prueba, (W), es menor al valor crítico.



5.6.3. Prueba de Kolmogorov-Smirnov (K-S)

89. La prueba de Kolmogorov-Smirnov se utiliza para decidir si un conjunto de datos proviene de una población con una distribución específica. La prueba de Kolmogorov-Smirnov se basa en la función de distribución empírica (EDF). ProUCL 5.1 utiliza la prueba K-S para probar la hipótesis nula si un conjunto de datos sigue una distribución gamma.
90. El estadístico de contraste es:

$$D = \sup_{1 \leq i \leq n} |\hat{F}_n(x_i) - F_0(x_i)| \quad (34)$$

Donde:

- x_i es el i -ésimo valor observado en la muestra (cuyos valores se han ordenado previamente de menor a mayor).
 - $\hat{F}_n(x_i)$ es un estimador de la probabilidad de observar valores menores o iguales que x_i .
 - $F_0(x_i)$ es la probabilidad de observar valores menores o iguales que x_i cuando H_0 es cierta.
91. Así pues, (D) es la mayor diferencia absoluta observada entre la frecuencia acumulada observada $\hat{F}_n(x_i)$ y la frecuencia acumulada teórica $F_0(x_i)$, obtenida a partir de la distribución de probabilidad que se especifica como hipótesis nula.
92. Si los valores observados $\hat{F}_n(x_i)$ son similares a los esperados $F_0(x_i)$, el valor de (D) será pequeño. Cuanto mayor sea la discrepancia entre la distribución empírica $\hat{F}_n(x_i)$ y la distribución teórica, mayor será el valor de (D).
93. Por tanto, el criterio para la toma de la decisión entre las dos hipótesis será de la forma:
- Si $D \leq D_\alpha \Rightarrow$ Aceptar H_0
 - Si $D > D_\alpha \Rightarrow$ Rechazar H_0

94. Donde el valor D_α se elige de tal manera que:

- $P(\text{Rechazar } H_0 / H_0 \text{ es cierta}) = P(D > D_\alpha / \text{Los datos siguen la distribución } M) = \alpha$

Siendo α el nivel de significación del contraste.

5.6.4. Prueba de Lilliefors

95. En la versión ProUCL 5.1, se han utilizado e incorporado al programa valores críticos de la prueba de Lilliefors desarrollados por Moling y Abdi en *Encyclopedia of Measurement and Statistics* (Salkind, N. Editor 2007). Los valores críticos descritos en Salkind (2007) se utilizan para tamaños de muestras hasta 50 y para tamaños mayores de 50, valores críticos aproximados, se calculan utilizando la siguiente fórmula:

$$\text{Valor crítico} = \frac{\text{Factor}}{f(n)}; \text{ donde } f(n) = \frac{0.83+n}{\sqrt{n}} - 0.01 \quad (35)$$

96. Presenta la misma hipótesis que la prueba de S-W, H_0 será rechazada si la estadística de prueba, es mayor al valor crítico.

5.6.5. Gráfico Q-Q

97. Un gráfico Cuantil-Cuantil permite observar que tan cerca está la distribución de un conjunto de datos a alguna distribución ideal o comparar la distribución de dos conjuntos de datos.
98. Se forman pares $p_1 =$ (Observaciones teóricas, observaciones hipotéticas) y se grafican en un plano, lo ideal es que estos puntos caigan sobre la recta diagonal ($y = x$), es decir; las observaciones teóricas son aproximadamente similares a las observaciones hipotéticas.

5.7. Métodos ROS para imputar valores no detectados (NDs) usados para la bondad de ajuste

99. El método ROS (también conocido como ROS robusto) basado en regresión de estadísticas de orden. Los métodos ROS imputan NDs (*Non detects* - valores debajo del límite de cuantificación) basado en una distribución hipotética como una distribución gamma o una distribución Lognormal.
100. En ProUCL5.1, GROS representa el método ROS para una distribución gamma.

5.8. Teorema del Límite Central

101. El teorema del límite central o teorema central del límite indica que, en condiciones muy generales, si S_n es la suma de n variables aleatorias independientes y de varianza no nula pero finita, entonces la función de distribución de S_n se aproxima a una distribución normal (también llamada distribución gaussiana, curva de Gauss o campana de Gauss). Así pues, el teorema asegura que esto ocurre cuando la suma de estas variables aleatorias e independientes es lo suficientemente grande (Filmus, Y., 2010).

5.9. Desigualdad de Chebyshev

102. En probabilidad, la desigualdad de Chebyshev (también escrito como Chebychev) es un resultado que ofrece una cota inferior a la probabilidad de que el valor de una variable aleatoria con varianza finita esté a una cierta distancia de su esperanza matemática. La desigualdad recibe su nombre del matemático ruso Pafnuti Chebyshev.
103. En la literatura, a este tipo de desigualdades, cuya característica es la comparación de la probabilidad de la cola de la distribución y su valor esperado, se le conoce como desigualdades tipo Chebyshev.



104. Estas desigualdades son la herramienta básica para demostrar resultados como la ley de los grandes números, entre otros. Además de que tienen aplicaciones en estadística, así como en otras áreas de las matemáticas (Knuth, D. E. ,1997).

5.10. Métodos de estimación de parámetros para distribuciones no discernibles: (Métodos de remuestreos: Bootstrap o Jackknife)

105. Los métodos de remuestreo permiten obtener conclusiones acerca de una población usando muestras extraídas de una muestra aleatoria, que fue extraída previamente de esa población.

106. Se caracterizan por que no requieren suposiciones sobre la población modelo, y que el número de remuestreos es muy elevado, por lo que se hace imprescindible el uso del ordenador.

107. El propósito de estos métodos es de verificación de los datos obtenidos en un muestreo.

108. Los métodos más populares son los de Jackknife y de Bootstrap.

5.10.1. Método de Jackknife (Quenouille, 1949 – Tukey, 1958)

109. El método obtiene muestras artificiales mediante sustitución de uno o de un grupo de valores de la muestra original.

110. Su denominación alude a la navaja multiuso, que pueden ser tan eficaces y de gran utilidad como una herramienta unitaria específica, es decir, describe su versatilidad al ser aplicable en una pluralidad de situaciones y problemas diferentes.

111. En un principio Quenouille (1949) dividió la muestra en dos mitades para la estimación de un coeficiente de correlación serial.

112. Posteriormente se amplía la idea a la división de la muestra en (N) grupos de tamaño (K), y la aplica a diferentes problemas de estimación.

113. El método Jackknife no es un método de estimación en el sentido clásico del término, sino una técnica operativa de utilizar dichos métodos clásicos con ciertas ventajas.

5.10.2. Método de Bootstrap (Efron, 1979)

114. El Bootstrap es un método que permite mediante remuestreo aleatorio, y algunos casos, prescindiendo u obviando algunas hipótesis sobre el modelo planteado, calcular medidas de precisión de un estimador, intervalos de confianza y en general, medir el grado de incertidumbre en una cantidad de problemas. Es una herramienta simple pero potente, es insesgado y sobretodo, puede aplicarse sin tener en cuenta los parámetros y naturaleza del modelo.

115. El método Bootstrap consiste, si tenemos una muestra de tamaño (N), en generar un gran número de muestras de tamaño (N) efectuando un muestreo con reemplazamiento de esos valores.

116. Es como si metiésemos los valores en una urna, extraemos una papeleta, anotamos el resultado, y volvemos a colocarlo en la urna, y así hasta obtener (N) valores.
117. En esa muestra calculamos el valor del parámetro que estamos estimando. Y así repetimos el proceso un gran número (B) de veces (por ejemplo 10 000 o más), con lo que obtenemos una distribución de valores para el parámetro en la que podemos calcular su dispersión (análogo del error estándar) y determinar unos límites de confianza utilizando esa distribución.
118. El Bootstrap se diferencia del enfoque tradicional paramétrico en que emplea un gran número de cálculos repetitivos para estimar la forma de la distribución muestral del estadístico, en lugar de fuertes asunciones distribucionales y fórmulas analíticas. Esto permite al investigador hacer inferencias en casos donde tales soluciones analíticas son inviables, y donde tales asunciones son insostenibles.
119. El Bootstrap no es, por tanto, un estadístico per se. Más bien, es una aproximación al uso de la estadística para hacer inferencias sobre los parámetros poblacionales.
120. La idea básica del Bootstrap es tratar la muestra como si fuera la población.

5.11. Coeficiente de Correlación

121. La fortaleza de la relación lineal que existe entre dos variables se mide mediante la correlación que existe entre ellas.
122. El coeficiente de correlación lineal de Pearson permite medir la fuerza de asociación, teniendo como condición para su aplicabilidad, la presencia de normalidad bivariada o al menos linealidad entre las variables. Esta última condición puede ser verificada en un gráfico Q-Q, visualizando que los puntos sigan una tendencia lineal (Privitera, G. J. ,2011).

$$\rho = \frac{\text{cov}(x,y)}{\sqrt{v(x)}\sqrt{v(y)}} \quad (36)$$

$$-1 \leq \rho \leq 1 \quad (37)$$

123. Donde; la covarianza entre las variables x e y se representa por (cov) y la varianza por (v).
124. Cuando el coeficiente de correlación lineal de Pearson (ρ), tiende a uno; existirá una fuerte relación lineal directa entre las variables x e y. Pero, si tiende a menos uno; existirá una fuerte relación lineal inversa entre las variables x e y. En cambio, si tiende a cero; no existirá relación lineal, puede que exista otro tipo de relación.

5.12. Coeficiente de variación

125. El coeficiente de variación es una calificación que permite a los usuarios evaluar la calidad estadística de las estimaciones (Reed, G. F., Lynn, F., & Meade, B. D., 2002).



- Hasta del 7%, es precisa;
- Entre el 8 y el 14% significa que existe una precisión aceptable;
- Entre el 15% y 20% precisión regular y por lo tanto se debe utilizar con precaución.
- Mayor del 20% indica que la estimación es poco precisa y por lo tanto se recomienda utilizarla sólo con fines descriptivos.

5.13. Principio de parsimonia

126. El principio de la parsimonia es común a toda teoría científica, y establece que a igualdad de condiciones, se elige el modelo más simple, aquél que contiene menor número de variables de más fácil determinación.

VI. METODOLOGÍA



127. El tamaño de la muestra de suelos considerado es de 20 muestras compuestas bajo un diseño aleatorio simple y con la condición que población de interés es relativamente homogénea; es decir, no se esperan patrones mayores de contaminación.

128. Dichas muestras de suelo fueron analizadas mediante la técnica de análisis denominada espectrometría de masas con plasma de acoplamiento inductivo (ICP-MS), la cual, determinar de forma cuantitativa la mayoría de los elementos de la tabla periódica.

129. En el presente informe se ha considerado utilizar el software ProUCL 5.1, y a continuación se presenta el siguiente esquema para el tratamiento de datos:



[Handwritten mark]

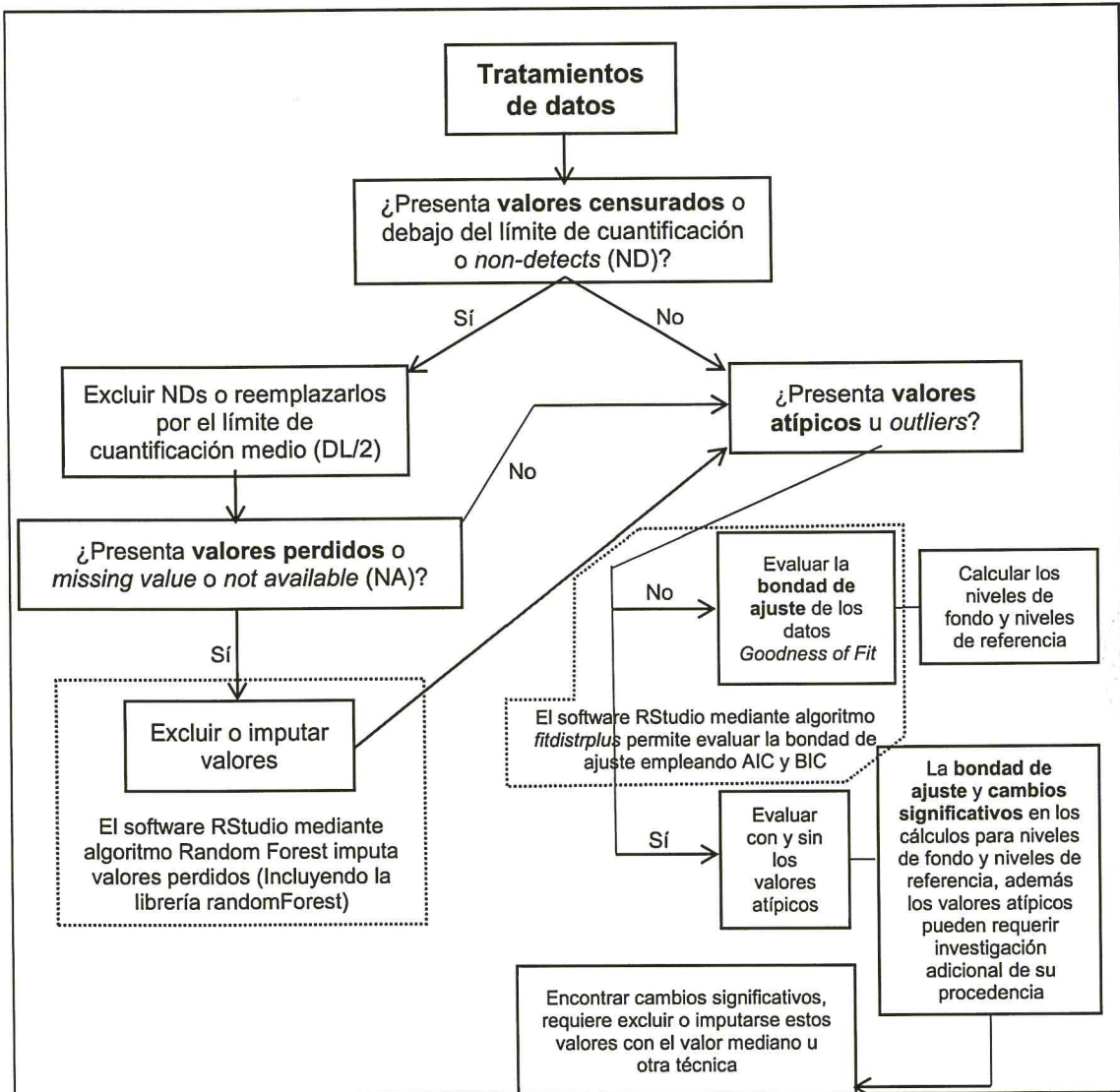


Figura 6-3. Esquema del tratamiento de datos en ProUCL 5.1

130. La metodología para determinar los niveles de fondo y niveles de referencia es a través de las herramientas estadísticas, el cálculo de estos valores requiere previamente un tratamiento y una verificación de los supuestos de las herramientas a utilizar. Existen algunos softwares estadísticos para realizar estos cálculos, tales como, RStudio (incluyendo la librería EnvStats) y ProUCL 5.1 de la Agencia de Protección Ambiental de los Estados Unidos (USEPA).

131. Los resultados de un laboratorio de ensayo, ocasionalmente reportan valores de la concentración de metales por debajo del límite de cuantificación, dependiendo del tipo de análisis empleado para cada elemento. Lo cual representa una dificultad al momento de tratar los datos en conjunto. Por estos motivos, más adelante se describen los criterios para el tratamiento estadístico tanto de los valores atípicos como aquellos conjuntos de muestras que contengan valores debajo del límite de cuantificación.

132. Cuando se presenta valores censurados y atípicos a la vez, podemos excluir los valores censurados o reemplazarlos por el límite de detección medio (DL/2), y debe realizarse ambos procedimientos con la finalidad de identificarlos correctamente.
133. Las pruebas clásicas para valores atípicos (EPA 2006b, 2009), son la prueba Dixon y la prueba Rosner. Estas pruebas a menudo sufren "efectos de enmascaramiento" en presencia de múltiples valores atípicos. Se sugiere que los procedimientos clásicos deben ir acompañados de gráficos, como el diagrama de cajas y el gráfico Q-Q.
134. Las pruebas de Dixon y de Rosner, suponen que el conjunto de datos sin los valores atípicos sigue una distribución normal.
135. La prueba de Rosner, puede utilizarse para identificar hasta 10 valores atípicos en conjuntos de datos de tamaños mayores a 25. Motivo por el cual, no se aplicará esta prueba debido a que el tamaño de la muestra utilizada en el presente informe es de 20.
136. Los valores atípicos distorsionan la mayoría de las estadísticas (por ejemplo, media, UCLs, UPLs, estadísticas de prueba) de interés. Por lo tanto, es deseable calcular las estadísticas de decisiones basadas en conjuntos de datos que representan la población principal dominante y no calcular estadísticas distorsionadas.
137. Se considerarán dos casos: valores atípicos no influyentes y valores atípicos de apalancamiento los cuales se encuentran representados en la Figura 6-2.

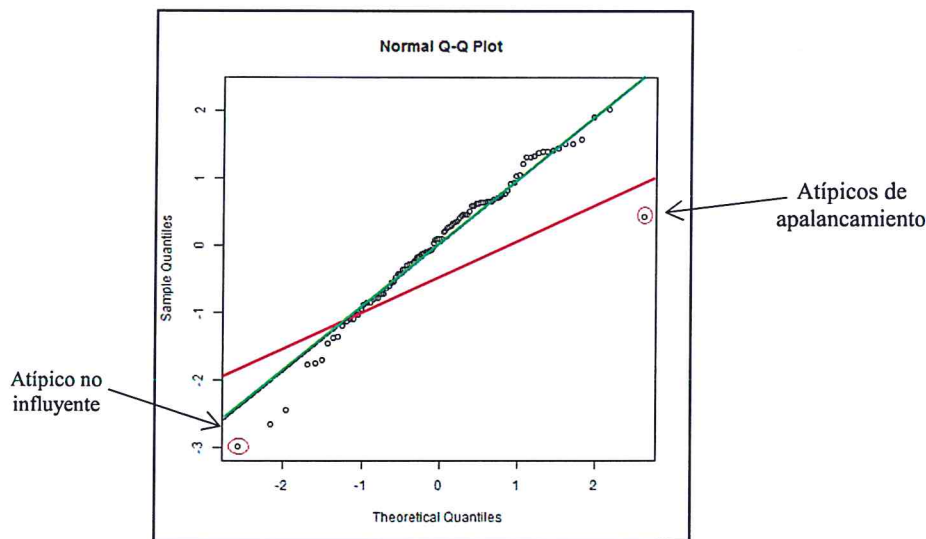


Figura 6-4. Diagrama Q-Q que muestra potenciales valores atípicos

138. Se puede observar en la Figura 6-2 que si se ajusta una recta esta se inclina muy levemente ante la presencia de un punto atípico no influyente (línea segmentada en verde), en cambio, la recta se inclina demasiado cuando hay un atípico de apalancamiento (línea punteada en rojo).

139. A través de las pruebas de bondad y ajuste o goodness of fit test (GOF); evaluamos el tipo de distribución de cada variable, y además nos podemos ayudar para la identificación de la distribución de las variables, observando el gráfico Q-Q, y considerando los cuantiles hipotéticos mostrados a lo largo del eje x, ajustados para los cuantiles asociados con valores no NDs, esto permitirá una mejor visualización del tipo de distribución que presenta.
140. Existen varios métodos diferentes para calcular Límite Superior de Confianza o Upper Confidence Limit (UCL). Antes de seleccionar un método apropiado, los datos deben caracterizarse mediante análisis exploratorio.
141. El ajuste de la distribución los datos es una parte crucial en el análisis exploratorio de datos (Schulz y Griffin 1999). Como se recomienda por la EPA (1992), "cuando hay una incertidumbre sobre la distribución del conjunto de datos, se debe usar una prueba estadística para identificar el mejor ajuste de la distribución para el conjunto de datos". Esto es necesario porque ningún tipo de distribución única se ajusta a todos los datos ambientales.
142. Cuando la concentración de la distribución es normal, se emplea el enfoque clásico basado en el estadístico T de Student. Cuando la distribución es Lognormal, se ha utilizado el método Land basado en el estadístico H. Cuando la distribución es Gamma se emplea el método de Wong.
143. Existen métodos libres de distribución o llamados no paramétricos y se emplean cuando no se puede discernir si los datos siguen una distribución conocida. La EPA describe varios métodos (EPA, 2000c). Para grandes conjuntos de datos, se puede utilizar un enfoque basado en el teorema de límite central con una corrección para la asimetría positiva. Para conjuntos de datos que no son lo suficientemente grandes para este enfoque, hay más de un enfoque disponible, aunque ninguno es ideal en todas las circunstancias. Los métodos generales incluyen un enfoque basado en la desigualdad de Chebyshev y un enfoque basado en el procedimiento de remuestreo bootstrap. Éstos se describen en EPA (2000c) y en Schulz y Griffin (1999). Ambos documentos dan ejemplos y comparaciones de las UCLs calculadas por varios métodos.



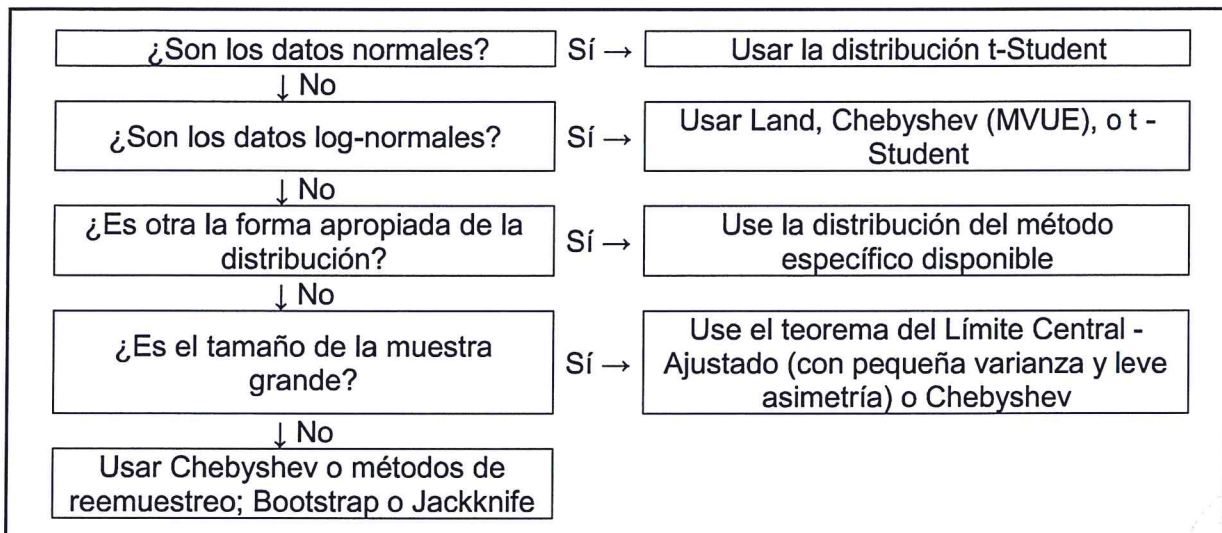


Figura 6-5. Resumen las recomendaciones brindadas en la guía técnica de ProUCL 5.1. Fuente: U.S. EPA Washington, D.C. ,2002

144. Identificada la distribución de las variables, se continúa calculando el UCL y el UTL basado en la distribución de los datos obtenida, se puede usar la Figura 6-3, para apoyarse en la distribución de las variables.
145. El software ProUCL 5.1, presenta dos tipos de prueba de bondad de ajuste GOF, la prueba Shapiro Wilk y Lilliefors para el ajuste de una distribución Normal, o la prueba de Anderson-Darling y Kolmogorov-Smirnov para el ajuste de una distribución Gamma o la prueba de Shapiro Wilk y Lilliefors para el ajuste de una distribución Lognormal.
146. Cuando una de las dos pruebas, para el ajuste de una distribución específica se contradice, es necesario buscar otra distribución, si no se encuentra distribución específica, se considera adecuado los métodos no paramétricos.
147. Si la variable presenta valores atípicos potenciales, lo cual puede visualizarse en el gráfico Q-Q y además no se ajusta a ninguna de las distribuciones continuas (Normal, Lognormal y Gamma) y el tamaño de la muestra es pequeña ($N < 30$) y al calcularse el UCL y UTL con y sin valores atípicos potenciales, no se observa diferencias significativas en la bondad de ajuste, se recomienda usar Chebyshev o métodos no paramétricos como los de remuestreo; Bootstrap o Jackknife.
148. Si se observa cambios significativos en la bondad de ajuste (es decir; realizando los cálculos del UCL y UTL, previamente habiendo excluido los valores atípicos potenciales), es probable que estos valores atípicos potenciales sean atípicos de apalancamiento, y por lo tanto, deberían excluirse o imputarse, con la finalidad de calcular estadísticas de decisiones basadas en conjuntos de datos que representan la población principal dominante y no calcular estadísticas distorsionadas.



e

VII. RESULTADOS

149. En la Tabla 7-1 se presenta los resultados del nivel de fondo y referencia para cada una de los 30 metales evaluados. El tamaño de la muestra para cada uno de los metales es de $N = 20$ y el nivel de confianza usado para el cálculo de los intervalos es del 95%. Sin embargo, la estadística descriptiva completa de los metales en estudio se encuentra en el Anexo, juntamente con las gráficas de diagrama de cajas, histograma y Q-Q

Tabla 7-1. Resultados de los niveles de fondo y niveles de referencia de 30 metales evaluados.

Parámetro	Nivel de fondo (mg/kg PS)	Nivel de referencia (mg/kg PS)
Aluminio (Al)	6 757	7 501
Arsénico (As)	11,32	15,99
Bario (Ba)	33,55	44,4
Bismuto (Bi)	0,164	0,251
Boro (B)	30,87	46,02
Cadmio (Cd)	0,286	0,377
Calcio (Ca)	13 809	19 507
Cerio (Ce)	20,55	25,77
Cobalto (Co)	7,312	8,483
Cobre (Cu)	19,99	29,89
Cromo (Cr)	15,91	20,66
Estroncio (Sr)	69,17	105,3
Fósforo (P)	885,10	1 018
Hierro (Fe)	23 099	28 793
Litio (Li)	9,839	10,86
Magnesio (Mg)	6 392	7 477
Manganeso (Mn)	306,10	355,60
Molibdeno (Mo)	1,446	1,86
Níquel (Ni)	5,701	7,266
Plomo (Pb)	19,21	26,51
Antimonio (Sb)	1,695	3,458



e



Parámetro	Nivel de fondo (mg/kg PS)	Nivel de referencia (mg/kg PS)
Potasio (K)	1 294	1 660
Sodio (Na)	8 522	15 448
Talio (Tl)	0,0513	0,0636
Titanio (Ti)	578,6	629
Torio (Th)	7,278	15,57
Uranio (U)	0,113	0,179
Vanadio (V)	88,46	115,50
Wolframio (W)	0,139	0,168
Zinc (Zn)	85,99	112,50

VIII. DISCUSIONES

150. Para el cálculo del nivel de fondo y referencia, es necesario tener las siguientes consideraciones. Las especies químicas pueden presentar valores potencialmente atípicos, es recomendable realizar una inspección sobre estos valores a través métodos exploratorios, como el gráfico de Q-Q, en este gráfico se observan los cuantiles observados versus los cuantiles esperados, cada uno de los puntos mostrados son las mediciones o puntos muestrales o valores de la variable, si hay un adecuado ajuste para una distribución específica, es de esperar que estos puntos caigan sobre la línea horizontal, si en cambio, se observa que este punto se aleja del resto pero sigue la tendencia de la línea diagonal, diremos que es un valor atípico no influyente, pero si este punto se aleja del resto en cualquier otra dirección diferente al de la tendencia, diremos que es un valor atípico de apalancamiento.

151. Si este último es el caso y además ninguna de las pruebas de bondad y ajuste (GOF) ha determinado una distribución discernible, se procederá a calcular el nivel de fondo y el nivel de referencia, excluyendo estos valores atípicos potenciales, generando así un nuevo conjunto de valores o nueva variable. En esta nueva variable, puede que se ajuste o no a una distribución específica, de no ajustarse a ninguna distribución, se calculará el nivel de fondo y el nivel de referencia a través de los métodos no paramétricos.

IX. CONCLUSIONES

152. El enfoque y los datos muestreados permiten determinar los niveles de fondo y de referencia, para cada una de las treinta variables o especies químicas.
153. En los resultados se reporta la concentración media, acompañado del nivel de fondo para evidenciar el rango de variación de esta concentración, y el nivel referencia, como un valor aproximado de máxima concentración, debido a la necesidad de conocer la calidad de un suelo afectado por la actividad antrópica, la cual resulta particularmente importante para valorar tal contaminación y las consecuencias administrativas que de ella se puedan derivar.
154. Los niveles de fondo y de referencia fueron calculados empleando la metodología recomendada por U.S. EPA, con la finalidad de obtener resultados óptimos.

X. REFERENCIAS BIBLIOGRÁFICAS

Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

Abreu, O. V., Alvear, D., Cuesta, A., & Alonso, V. Métodos de Ajuste y Estimación de las Distribuciones de las Variables Aleatorias de Entrada de los Modelos de Evacuación. Sea, 1(2), 1.

Berenson, M. L. L., Berenson, D. M. M. L., & Levine, D. M. (1992). Estadística Básica en Administración: conceptos y aplicaciones. Prentice-Hall Hispanoamericana, 899.

De Miguel, E. (2002). Determinación de niveles de fondo y niveles de referencia de metales pesados y otros elementos traza en suelos de la Comunidad de Madrid (No. 2). IGME.

Daniel, DL (2015). Una perspectiva Estudio de caso sobre el trabajo con ProUCL y una Agencia de Medio Ambiente del Estado en la determinación de valores de umbral de fondo. Revista Internacional de Investigación del Medio Ambiente y Salud Pública, 12 (10), 12905-12923. <http://doi.org/10.3390/ijerph121012905>

Devore, J. L. (2008). Probabilidad y estadística para ingenierías y ciencias. Cengage Learning Editores.

Ezpeleta, A. M. (Ed.). (2003). Clima y calidad ambiental (Vol. 136). Univ Santiago de Compostela.7

Filmus, Y. (2010). Two proofs of the central limit theorem.

Freund, J. E., Miller, I., & Miller, M. (2000). Estadística matemática con aplicaciones. Pearson Educación.

U.S. EPA Washington, D.C. (2002). Calculating upper confidence limits for exposure point concentrations at hazardous waste sites.





Knuth, D. E. (1997). The Art of Computer Programming. Volume 1: Fundamental Algorithms. Volume 2: Seminumerical Algorithms. Bull. Amer. Math. Soc.

Montgomery, D. C., Runger, G. C., & Medal, E. G. U. (1996). Probabilidad y estadística aplicadas a la ingeniería (No. 968-18-5914-6. 01-A1 LU. AL-PyE. 1.). McGraw Hill.

Privitera, G. J. (2011). Statistics for the behavioral sciences. Sage.

Reed, G. F., Lynn, F., & Meade, B. D. (2002). Use of coefficient of variation in assessing variability of quantitative assays. Clinical and diagnostic laboratory immunology, 9(6), 1235-1239.

Sánchez, A., Fernández, F., Valero, C., Muñoz, M., Rodríguez, A., López, M., & Espejo, I. (2009). Estadística descriptiva y probabilidad:(Teoría y problemas). Cádiz: Universidad de Cádiz, 2009.

Schumacher, B., Singh, A., & Singh, A. K. (2015). ProUCL Version 5.1 Technical Guide.



XI. ANEXOS

Anexo A: Pasos para calcular el nivel de fondo y el nivel de referencia

Anexo B: Cálculo del nivel de fondo y nivel de referencia de las variables

Anexo C: Análisis exploratorio de las variables



Handwritten signature

ANEXOS



[Handwritten signature]



PERÚ

Ministerio del Ambiente

Organismo de Evaluación y Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

ÍNDICE DEL CONTENIDO DEL ANEXO

Anexo A: Pasos para calcular el nivel de fondo y el nivel de referencia 4
 Anexo B: Cálculo del nivel de fondo y nivel de referencia de las variables..... 1
 Anexo C: Análisis exploratorio de las variables..... 16

ÍNDICE DE TABLAS

Tabla A - 1: Pasos para calcular la UCL para la media de una distribución Normal – T de Student 1
 Tabla A - 2: Pasos para calcular la UCL para la media de una distribución Lognormal – Método Land 2
 Tabla A - 3: Pasos para el cálculo de UCL basado en la desigualdad Chebyshev - MVUE Enfoque para distribuciones Lognormal 3
 Tabla A - 4: Condiciones favorables para establecer el uso de la desigualdad de Chebyshev (MVUE) sobre el método Land 4
 Tabla A - 5: Pasos para calcular la UCL usando el Teorema del Límite Central (Ajustado) 4
 Tabla A - 6: Pasos para calcular la UCL usando Hall's Bootstrap 5
 Tabla A - 7: Pasos para calcular UCL basado en la desigualdad de Chebyshev - no paramétrico 6
 Tabla A - 8: Resumen de los métodos de cálculo de la UCL 7
 Tabla B - 1: Detección de valores atípicos potencialmente altos 8
 Tabla B - 2: Análisis descriptivos 9
 Tabla B - 3: Percentiles 10
 Tabla B - 4: Bondad de ajuste 11
 Tabla B - 5: Nivel de fondo y de referencia 12





ÍNDICE DE FIGURAS

Figura C - 1: Datos de aluminio (Al); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 16

Figura C - 2: Datos de arsénico (As); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 17

Figura C - 3: Datos de bario (Ba); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 18

Figura C - 4: Datos de boro (B); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 19

Figura C - 5: Datos de bismuto (Bi); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 20

Figura C - 6: Datos de calcio (Ca); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 21

Figura C - 7: Datos de cadmio (Cd); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 22

Figura C - 8: Datos de cerio (Ce); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 23

Figura C - 9: Datos de cobalto (Co); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 24

Figura C - 10: Datos de cromo (Cr); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 25

Figura C - 11: Datos de cobre (Cu); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 26

Figura C - 12: Datos de hierro (Fe); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 27

Figura C - 13: Datos de potasio (K); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 28

Figura C - 14: Datos de litio (Li); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 29

Figura C - 15: Datos de magnesio (Mg); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 30

Figura C - 16: Datos de manganeso (Mn); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 31

Figura C - 17: Datos de molibdeno (Mo); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 32

Figura C - 18: Datos de sodio (Na); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 33

Figura C - 19: Datos de níquel (Ni); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 34

Figura C - 20: Datos de plomo (Pb); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 35



Handwritten signature



Figura C - 21: Datos de fósforo (P); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 36

Figura C - 22: Datos de antimonio (Sb); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 37

Figura C - 23: Datos de estroncio (Sr); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 38

Figura C - 24: Datos de torio (Th); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 39

Figura C - 25: Datos de titanio (Ti); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 40

Figura C - 26: Datos de talio (Tl) a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 41

Figura C - 27: Datos de uranio (U); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 42

Figura C - 28: Datos de vanadio (V); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 43

Figura C - 29: Datos de wolframio (W); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 44

Figura C - 30: Datos de zinc (Zn); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad 45



Handwritten signature



PERÚ

Ministerio
del Ambiente

Organismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Anexo A: Pasos para calcular el nivel de fondo y el nivel de referencia



el



Tabla A - 1: Pasos para calcular la UCL para la media de una distribución Normal – T de Student

Sean X_1, X_2, \dots, X_n representan las n concentraciones aleatoriamente muestreadas.

PASO 1: Calcular la media muestral

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

PASO 2: Calcular la desviación estándar de la muestra

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (2)$$

PASO 3: Utilice una tabla de cuantiles de la distribución t de Student para encontrar el cuantil $(1-\alpha)$ avo de la distribución t de Student con $n-1$ grados de libertad. Por ejemplo, el valor en el nivel 0,05 con 40 grados de libertad es 1,684. En Gilbert (1987, página 255, donde los valores están indexados por $p = 1-\alpha$, en lugar de un nivel α) se puede encontrar una tabla de valores t de Student. El valor t apropiado para calcular el 95% UCL se puede obtener en Microsoft Excel® con la fórmula TINV $((1-0.95) * 2, n-1)$.

PASO 4: Calcule el límite de confianza superior unilateral $(1-\alpha)$ sobre la media

$$UCL = \bar{X} \pm t_{\alpha, n-1} \frac{S}{\sqrt{n}} \quad (3)$$

Fuente: U.S. EPA Washington, D.C., 2002



Handwritten signature



PERÚ

Ministerio
del Ambiente

Organismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Tabla A - 2: Pasos para calcular la UCL para la media de una distribución Lognormal – Método Land

Sean X_1, X_2, \dots, X_n representan las n concentraciones aleatoriamente muestreadas.

PASO 1: Calcular la media aritmética de los datos log-transformados

$$\overline{\ln X} = \frac{1}{n} \sum_{i=1}^n \ln(X_i) \quad (4)$$

PASO 2: Calcular la desviación estándar asociada.

$$S_{\ln X} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\ln(X_i) - \overline{\ln(X)})^2} \quad (5)$$

PASO 3: Busque la estadística $H_{1-\alpha}$ para el tamaño de la muestra n y la desviación estándar Observada de los datos log-transformados. Tablas de estos valores están dadas por Gilbert (1987, Tablas A-10 y A-12) y Land (1975).

PASO 4: Calcule el límite de confianza superior unilateral $(1-\alpha)$ sobre la media

$$UCL_{1-\alpha} = \exp\left(\overline{\ln X} + \frac{S_{\ln X}^2}{2} + H_{1-\alpha} S_{\ln X} / \sqrt{n-1}\right) \quad (6)$$

Fuente: U.S. EPA Washington, D.C., 2002



Handwritten signature

Tabla A - 3: Pasos para el cálculo de UCL basado en la desigualdad Chebyshev - MVUE
Enfoque para distribuciones Lognormal

Sean X_1, X_2, \dots, X_n representan las n concentraciones aleatoriamente muestreadas.

PASO 1: Calcular la media aritmética de los datos log-transformados

$$\overline{\ln X} = \frac{1}{n} \sum_{i=1}^n \ln(X_i) \quad (7)$$

PASO 2: Calcular la varianza asociada

$$S_{\ln X}^2 = \frac{1}{n-1} \sum_{i=1}^n (\ln(X_i) - \bar{y})^2 \quad (8)$$

PASO 3: Calcular el estimador no sesgado de varianza mínima (MVUE) de la media poblacional para una distribución lognormal $\mu_{LN} = \exp(\overline{\ln X}) g_n(S_{\ln X}^2/2)$, donde g_n indica una función para la cual se dispone de tablas (Aitchison y Brown, 1969, Tabla A2, Koch y Link 1980, Tabla A7).

PASO 4: Calcular el MVUE de la varianza asociada de esta media.

$$\sigma_{\mu}^2 = \exp(2 \ln X) \left(\left(g_n(S_{\ln X}^2/2) \right)^2 - g_n\left(\frac{n-2}{n-1} S_{\ln X}^2\right) \right) \quad (9)$$

PASO 5: Calcule el límite de confianza superior unilateral $(1-\alpha)$ sobre la media.

$$UCL_{1-\alpha} = \hat{\mu}_{LN} + \sqrt{\left(\frac{1}{\alpha} - 1\right) \sigma_{\mu}^2} \quad (10)$$

Fuente: U.S. EPA Washington, D.C., 2002



Tabla A - 4: Condiciones favorables para establecer el uso de la desigualdad de Chebyshev (MVUE) sobre el método Land

Desviación estándar de datos log-transformados	Tamaño de la muestra	Recomendaciones
1 – 1,5	<25	95% Chebyshev (MVUE) UCL
1,5 - 2	<20 20 - <50	99% Chebyshev (MVUE) UCL 95% Chebyshev (MVUE) UCL
2 – 2,5	<25 25 - 70	99% Chebyshev (MVUE) UCL 95% Chebyshev (MVUE) UCL
2,5 – 3,0	<30 30 - <70	99% Chebyshev (MVUE) UCL 95% Chebyshev (MVUE) UCL

Fuente: U.S. EPA Washington, D.C., 2002

Tabla A - 5: Pasos para calcular la UCL usando el Teorema del Límite Central (Ajustado)

Sean X_1, X_2, \dots, X_n representan las n concentraciones aleatoriamente muestreadas.

PASO 1: calcula la media muestral.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (11)$$

PASO 2: Calcular la desviación estándar de la muestra.

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (12)$$

PASO 3: Calcular la asimetría de la muestra

$$\beta = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s} \right)^3 \quad (13)$$

Este puede ser calculado con Microsoft Excel con la función SKEW.

PASO 4: Sea z_α el $(1-\alpha)$ avo cuantil de la distribución normal estándar. Para el nivel de confianza del 95%, $z_\alpha = 1,645$.

PASO 5: Calcule el límite de confianza superior unilateral $(1-\alpha)$ sobre la media

$$UCL_{1-\alpha} = \bar{X} + \left(z_\alpha + \frac{\beta}{6\sqrt{n}} (1 + 2z_\alpha^2) \right) s / \sqrt{n} \quad (14)$$

Fuente: U.S. EPA Washington, D.C., 2002



Tabla A - 6: Pasos para calcular la UCL usando Hall's Bootstrap

Sean X_1, X_2, \dots, X_n representan las n concentraciones aleatoriamente muestreadas.

PASO 1: Calcular el promedio muestral.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (15)$$

PASO 2: Calcular la desviación estándar de la muestra.

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (16)$$

PASO 3: Calcular la asimetría de la muestra

PASO 4: Para $b = 1$ a B (un número muy grande) haga lo siguiente:

- 4.1: Genere un conjunto de datos de muestra de Bootstrap; es decir, para $i = 1$ a n , sea j un número entero aleatorio entre 1 y n y añada la observación X_j al conjunto de datos de muestra de bootstrap.
- 4.2: Calcular la media aritmética \bar{X}_b del conjunto de datos construido en el paso 4.1.
- 4.3: Calcular la desviación estándar asociada S_b del conjunto de datos construido.
- 4.4: Calcular la asimetría k_b de los datos construidos usando la fórmula en el Paso 3.
- 4.5: Calcular la media estudiada

$$W = (\bar{X}_b - \bar{X}) / s_b \quad (17)$$

4.6: Calcule la estadística de Hall

$$Q = W + k_b W^2 / 3 + k_b^2 W^3 / 27 + k_b / (6n) \quad (18)$$

PASO 5: Clasifique todos los valores de Q calculados en el paso 4 y seleccione el cuantil inferior (α) avo de estos valores de B . Es el (αB) avo valor en una lista ascendente de Q 's.

Este valor es de la cola izquierda de la distribución.

PASO 6: Calcular

$$W(Q) = \frac{3}{k} \left(\left(1 + k \left(Q_\alpha - \frac{k}{6n} \right) \right)^{1/3} - 1 \right) \quad (19)$$

PASO 7: Calcule el límite de confianza unilateral ($1-\alpha$) en la media.

$$UCL_{1-\alpha} = \bar{X} - W(Q_\alpha) s \quad (20)$$

Fuente: U.S. EPA Washington, D.C., 2002





Tabla A - 7: Pasos para calcular UCL basado en la desigualdad de Chebyshev - no paramétrico

Sean X_1, X_2, \dots, X_n representan las n concentraciones aleatoriamente muestreadas.

PASO 1: Calcula la media aritmética de los datos

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (21)$$

PASO 2: Calcular la desviación estándar de la muestra.

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad (22)$$

PASO 3: Calcule el límite de confianza superior unilateral $(1-\alpha)$ sobre la media

$$UCL_{1-\alpha} = \bar{X} + \sqrt{\frac{1}{\alpha} - 1} (s / \sqrt{n}) \quad (23)$$

Fuente: U.S. EPA Washington, D.C., 2002



Handwritten signature





PERÚ

Ministerio
del AmbienteOrganismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Tabla A - 8: Resumen de los métodos de cálculo de la UCL

Métodos	Aplicabilidad	Ventajas	Desventajas	Referencia
Para distribuciones Normal o Lognormal				
T de estudiantes	Medias distribuidas normalmente, muestras aleatorias	Simple, si es robusto; n es grande	La distribución de las medias debe ser normal	Gilbert 1987; EPA 1992
Land's H	Datos lognormal, pequeña varianza, n grande, muestras aleatorias	Buena cobertura ¹	Sensible a las desviaciones de la log normalidad, produce valores muy altos para la varianza grande o n pequeño	Gilbert 1987; EPA 1992
Desigualdad de Chebyshev (MVUE)	Asimetría y variación pequeña o moderada, muestras aleatorias	A menudo más pequeño que Land	Necesidad de recurrir a niveles de confianza más altos para una cobertura	Singh et al. 1997
Wong	Distribución gamma	Precisión de segundo orden ²	Requiere solución numérica de una integral impropia	Schulz and Griffin 1999; Wong 1993
Métodos no paramétricos / sin distribución				
Teorema del límite central - ajustado	Grandes n, muestras aleatorias	Simple, robusto	El tamaño de la muestra puede no ser suficiente	Gilbert 1987; Singh et al. 1997
Bootstrap t Resampling	El muestreo es aleatorio y representativo	Útil cuando la distribución no puede ser identificada	Cobertura inadecuada para algunas distribuciones; computacionalmente intensiva	Singh et al. 1997; Efronw 1982
Procedimiento Bootstrap de Hall	El muestreo es aleatorio y representativo	Útil cuando no se puede identificar la distribución; toma el sesgo y la asimetría en cuenta	Cobertura inadecuada para algunas distribuciones; computacionalmente intensiva	Hall 1988; Hall 1992; Manly 1997; Schultz and Griffin 1999
Procedimiento Jackknife	El muestreo es aleatorio y representativo	Útil cuando la distribución no puede ser identificada	Cobertura inadecuada para algunas distribuciones; computacionalmente intensiva	Singh et al. 1997
Chebyshev Desigualdad	Asimetría y variación pequeña o moderada, muestras aleatorias	Útil cuando la distribución no puede ser identificada	Inadecuado para tamaños pequeños de muestra cuando la asimetría o la varianza es grande	Singh et al. 1997; EPA 2000c

¹ Cobertura se refiere a si un método UCL realiza de acuerdo con su definición.² A diferencia de la estimación de máxima verosimilitud, que ofrece una precisión de primer orden.

Fuente: U.S. EPA Washington, D.C., 2002





PERÚ

Ministerio
del Ambiente

Organismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Anexo B: Cálculo del nivel de fondo y nivel de referencia de las variables



le



PERÚ

Ministerio
del Ambiente

Organismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Tabla B - 1: Detección de valores atípicos potencialmente altos

Variables	Potencial valor atípico alto
Bi	0,2762; 0,2323
Cd	0,8653; 0,7222
Ce	25,77; 23,63
Cu	28
Mn	359
Pb	51,4; 47,5
Sb ¹	4,792
Sr	102; 102
Th	15,62
Tl	0,0635; 0,0618
U	0,3282; 0,2043
W	0,591; 0,3529; 0,2873; 0,2092
Zn	155; 126



¹Presenta valores censurados, se realizó el cálculo excluyendo los NDs y reemplazando por LD/2



Handwritten signature or mark.





PERÚ

Ministerio
del AmbienteOrganismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Tabla B - 2: Análisis descriptivos

Variables	N	Mínimo	Máximo	Media	Desviación estándar	Error estándar de la media	Asimetría	Curtosis	CV%
Al	20	6013	7329	6614	369,9	82,72	0,449	-0,274	5,59
As	20	5,5	15	10,43	2,323	0,519	-0,172	-0,031	22,3
B	20	17,3	41,9	27,96	7,539	1,686	0,426	-0,444	27
Ba	20	24,8	41,3	31,47	5,397	1,207	0,654	-0,955	17,2
Bi	20	0,0983	0,276	0,147	0,0434	0,00971	1,753	3,415	29,5
Ca	20	7434	18526	12713	2836	634,1	-0,0588	0,010	22,3
Cd	20	0,199	0,865	0,32	0,169	0,0378	2,657	6,667	52,8
Ce	20	17,19	25,77	19,76	2,025	0,453	1,691	3,441	10,2
Co	20	6,14	8,59	7,087	0,583	0,13	0,888	0,991	8,22
Cr	20	12	20	15	2,362	0,528	0,453	-0,647	15,7
Cu	20	13	28	18,3	4,156	0,929	1,015	0,587	22,7
Fe	20	18258	28230	22003	2834	633,7	0,405	-0,626	12,9
K	20	825	1505	1224	182	40,7	-0,472	-0,201	14,9
Li	20	8,85	10,7	9,643	0,508	0,114	0,208	-0,650	5,26
Mg	20	5102	6902	6183	540,2	120,8	-0,577	-0,701	8,74
Mn	20	248	359	296,6	24,62	5,506	0,492	1,085	8,3
Mo	20	0,954	1,86	1,35	0,249	0,0556	0,65	-0,497	18,4
Na	20	2754	14434	7189	3447	770,789	0,806	-0,355	47,9
Ni	20	4,39	7,3	5,4	0,779	0,174	0,746	0,381	14,4
P	20	725	976	859,5	66,2	14,8	-0,278	-0,511	7,7
Pb	20	12,5	51,4	20,92	10,34	2,313	2,352	5,181	49,4
Sb ¹	20	0,0017	4,792	1,177	1,295	0,29	1,295 ³	1,529 ³	110
Sb ²	15	0,0076	4,792	1,569	1,318	0,295	1,1	1,268	84
Sr	20	42,8	102	63,06	15,68	3,506	1,519	2,364	24,9
Th	20	2,838	15,62	5,989	3,1	0,693	1,829	3,923	51,8
Ti	20	522	629	566,8	30,64	6,851	0,565	-0,953	5,41
Tl	20	0,0381	0,0635	0,0489	0,00614	0,00137	0,89	1,160	12,6
U	20	0,0687	0,328	0,117	0,062	0,0139	2,412	6,693	53
V	20	65	109	83,25	13,47	3,012	0,198	-1,156	16,2
W	20	0,112	0,591	0,178	0,115	0,0256	2,914	9,091	64,3
Zn	20	66,2	155	87,03	21,7	4,852	1,947	4,396	24,9

¹ Presenta valores censurados, se realizó los cálculos mediante el método de Kaplan Meier.

² Presenta valores censurados, se realizó los cálculos excluyendo los NDs.

² Existen cinco valores no censurados (son los mismos e iguales a 0,0017)



e



PERÚ

Ministerio
del AmbienteOrganismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

³ El coeficiente de asimetría y curtosis, se realizó los cálculos mediante LD en vez de NDs.

Tabla B - 3: Percentiles

Variable	N	10%ile	20%ile	25%ile(Q1)	50%ile(Q2)	75%ile(Q3)	80%ile	90%ile	95%ile	99%ile
Al	20	6278	6309	6381	6545	6841	6887	7159	7303	7324
As	20	8,14	8,84	8,975	10,5	12	12,2	13	13,1	14,62
B	20	18,21	21,74	24,7	26,8	31,4	33,98	40,53	40,86	41,69
Ba	20	26,04	26,78	27,13	29,6	35,25	36,96	39,72	40,83	41,21
Bi	20	0,112	0,115	0,117	0,138	0,16	0,164	0,183	0,234	0,268
Ca	20	8732	10911	11817	12702	14166	14684	15973	16897	18200
Cd	20	0,213	0,232	0,234	0,272	0,31	0,32	0,387	0,729	0,838
Ce	20	18,04	18,34	18,49	19,31	20,21	20,56	21,53	23,74	25,36
Co	20	6,505	6,574	6,76	6,975	7,385	7,464	7,86	7,897	8,451
Cr	20	12	12,8	13,75	14,5	17	17,2	18	18,1	19,62
Cu	20	14	15	15	18	21	21,2	22,5	27,05	27,81
Fe	20	18775	19061	19374	21618	24361	24471	25389	25622	27708
K	20	962,3	1112	1146	1233	1354	1372	1414	1491	1502
Li	20	8,998	9,132	9,223	9,715	9,985	10,02	10,13	10,42	10,64
Mg	20	5398	5774	5870	6358	6569	6660	6758	6892	6900
Mn	20	269,9	277,6	282,5	294,5	312	315,6	323	324,8	352,2
Mo	20	1,134	1,14	1,185	1,26	1,558	1,592	1,664	1,794	1,847
Na	20	3487	4591	4820	6036	8924	10010	12013	13594	14266
Ni	20	4,408	4,776	4,85	5,26	5,783	5,924	6,494	6,569	7,154
P	20	780,2	814,2	823,8	864,5	905	906,6	918,8	954,2	971,6
Pb	20	13,18	13,86	14,9	18,75	20,58	21,14	27,16	47,7	50,66
Sb ¹	20	0,0017	0,0017	0,00613	0,777	1,946	2,196	2,381	3,58	4,55
Sr	20	48,29	52,34	52,95	59,05	67,48	69,9	76,89	102	102
Th	20	3,204	3,997	4,052	5,072	6,777	7,18	9,844	10,73	14,64
Ti	20	541,9	544	544	551,5	595,5	603	607	608,1	624,8
Tl	20	0,0429	0,0451	0,0452	0,0472	0,0512	0,0521	0,0549	0,0619	0,0632
U	20	0,0724	0,079	0,0822	0,0924	0,135	0,141	0,182	0,21	0,305
V	20	66,8	69,8	70,75	83	95,25	96	99,2	101,4	107,5
W	20	0,119	0,123	0,126	0,134	0,152	0,176	0,294	0,365	0,546
Zn	20	70,12	71,56	72,53	80,1	95,1	97,9	101,3	127,5	149,5

¹ Presenta valores censurados, se realizó los cálculos mediante LD en vez de NDs.



[Handwritten signature]



PERÚ

Ministerio del Ambiente

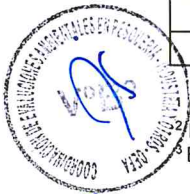
Organismo de Evaluación y Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Tabla B - 4: Bondad de ajuste

Variable	N	Akaike's Information Criterion (AIC)			Bayesian Information Criterion (BIC)			Ajuste de distribución
		Normal	Lognormal	Gamma	Normal	Lognormal	Gamma	
Al	20	296,263688	295,86762	295,984315	298,255152	297,859085	297,97578	Lognormal
As	20	93,443566	95,531727	94,499413	95,435031	97,523192	96,490877	Normal
Ba	20	127,166162	125,330115	125,846005	129,157626	127,32158	127,837469	Lognormal
Bi	20	-65,738883	-72,484573	-70,613446	-63,747418	-70,493108	-68,621982	Lognormal
B	20	140,532953	139,581011	139,545964	142,524418	141,572475	141,537429	Gamma
Cd	20	-11,373916	-28,470397	-23,458349	-9,382451	-26,478933	-21,466884	Lognormal
Ca ³	20	377,7324	379,3894	378,4365	379,7239	381,3809	380,428	Normal
Ce	20	87,948019	85,246868	86,089363	89,939484	87,238333	88,080828	Lognormal
Co	20	38,1227	36,995755	37,333589	40,114164	38,987219	39,325054	Lognormal
Cu	20	116,715359	113,480276	114,340436	118,706824	115,471741	116,331901	Lognormal
Cr	20	94,111678	93,118103	93,346801	96,103142	95,109568	95,338266	Lognormal
Sr	20	169,827989	164,74788	166,149627	171,819454	166,739344	168,141092	Lognormal
P	20	227,440008	227,925763	227,734671	229,431473	229,917227	229,726136	Normal
Fe	20	377,710051	376,966277	377,146014	379,701515	378,957741	379,137479	Lognormal
Li	20	32,608897	32,450144	32,490644	34,600362	34,441608	34,482108	Lognormal
Mg	20	311,408554	312,412808	312,045902	313,400019	314,404272	314,037367	Normal
Mn	20	187,881083	187,381684	187,501051	189,872548	189,373149	189,492516	Lognormal
Mo	20	4,099132	2,373131	2,806318	6,090597	4,364595	4,797783	Lognormal
Ni	20	49,732573	48,223054	48,624707	51,724038	50,214518	50,616172	Lognormal
Pb	20	153,189551	139,050871	142,993083	155,181015	141,042336	144,984548	Lognormal
Sb ¹	20	71,0936	30,42255	21,45491	73,08507	32,41401	23,44637	Gamma
Sb ²	15	71,08764	34,00476	26,07349	73,07911	35,99623	28,06496	Gamma
K	20	267,89242	269,659575	268,944659	269,883884	271,65104	270,936124	Normal
Na ³	20	385,54	380,9485	381,5211	387,5344	382,9399	383,5126	Lognormal
Tl	20	-144,00194	-145,539104	-145,121626	-142,010475	-143,547639	-143,130161	Lognormal
Ti	20	196,622099	196,095299	196,260554	198,613563	198,086764	198,252018	Lognormal
Th	20	104,988255	94,529553	96,956054	106,97972	96,521018	98,947518	Lognormal
U	20	-51,497332	-65,955967	-62,044106	-49,505867	-63,964503	-60,052641	Lognormal
V	20	163,752438	163,353033	163,3898	165,743903	165,344498	165,381265	Lognormal
W	20	-26,91814	-47,662041	-41,659092	-24,926675	-45,670576	-39,667627	Lognormal
Zn	20	182,818872	176,238737	178,158069	184,810336	178,230201	180,149534	Lognormal



1 Presenta valores censurados, se realizó los cálculos reemplazando los NDs por LD/2.
 2 Presenta valores censurados, se realizó los cálculos excluyendo los NDs.
 3 Empleando la estimación de Método de Momentos (MME).



PERÚ

Ministerio
del AmbienteOrganismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Tabla B - 5: Nivel de fondo y de referencia

Variable	Nivel de fondo	Nivel de referencia	Comentarios	Conclusiones
Al	6 757	7 501	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,107) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
As	11,32	15,99	Según AIC y BIC es más adecuado la Normalidad, además el estadístico (0,101) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Ba	33,55	44,4	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,178) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Bi	0,164	0,251	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ y al no encontrar cambios significativos en la bondad de ajuste (se mantiene la Normalidad), se consideró reportar el UCL y UTL con valores potencialmente atípicos . Para este último, el estadístico (0,174) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
B	30,87	46,02	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,153) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Cd	0,286	0,377	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ (atípico de apalancamientos) y se encontró cambios significativos en la bondad de ajuste, se consideró reportar el UCL y UTL excluyendo los valores potencialmente atípicos . Pues representa la población principal dominante. Para este último, el estadístico (0,174) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Ca	13 809	19 507	Según AIC y BIC es más adecuado la Normalidad, además el estadístico (0,146) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Ce	20,55	25,77	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ, además no cumple ninguna de las pruebas GOF (prueba Shapiro Wilk, Lilliefors, Anderson-Darling Gamma, Kolmogorov-Smirnov Gamma, Shapiro Wilk Lognormal y Lilliefors Lognormal) y el tamaño de la muestra no es grande ($n < 30$), se determina considerar al 95% Jackknife UCL, dado que este valor también coincide con 95% Student's-t UCL y para nivel de referencia se utilizó el 95% Percentile Bootstrap UTL con 95% Coverage con valores potencialmente atípicos .	Distribución libre no paramétrica
Co	7,312	8,483	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,166) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Cu	19,99	29,89	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ, según AIC y BIC es más adecuado la Lognormal y su UCL es bastante similar con y sin valores potencialmente atípicos, además cumple las pruebas GOF Log (Shapiro Wilk Lognormal y Lilliefors Lognormal), se reporta 95% H-UCL y 95% UTL con 95% Coverage ambas para Lognormal con valores potencialmente atípicos .	Lognormal
Cr	15,91	20,66	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,164) y valor crítico (0,192)	Normal





PERÚ

Ministerio
del AmbienteOrganismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Variable	Nivel de fondo	Nivel de referencia	Comentarios	Conclusiones
			de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	
Sr	69,17	105,3	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ, según AIC y BIC es más adecuado la lognormal, además cumple las pruebas GOF Log (Shapiro Wilk Lognormal y Lilliefors Lognormal), se reporta 95% H-UCL y 95% UTL con 95% Coverage ambas para Lognormal con valores potencialmente atípicos .	Lognormal
P	885,10	1 018	Según AIC y BIC es más adecuado la Normalidad, además el estadístico (0,159) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Fe	23 099	28 793	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,136) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Li	9,839	10,86	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,121) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Mg	6 392	7 477	Según AIC y BIC es más adecuado la Normalidad, además el estadístico (0,173) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Mn	306,10	355,60	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ y al no encontrar cambios significativos en la bondad de ajuste (se mantiene la Normalidad), se consideró reportar el UCL y UTL con valores potencialmente atípicos . Para este último, el estadístico (0,0918) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Mo	1,446	1,86	Como solo se cumple una de las pruebas GOF (prueba Shapiro Wilk, Lilliefors, Anderson-Darling Gamma, Kolmogorov-Smirnov Gamma, Shapiro Wilk Lognormal y Lilliefors Lognormal) y el tamaño de la muestra no es grande ($n < 30$), se determina considerar al 95% Jackknife UCL, dado que este valor también coincide con 95% Student's-t UCL y 95% Hall's Bootstrap UCL y para nivel de referencia se utilizó el 95% Percentile Bootstrap UTL con 95% Coverage.	Distribución libre no paramétrica
Ni	5,701	7,266	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,131) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Pb	19,21	26,51	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ (atípico de apalancamientos) y se encontró cambios significativos en la bondad de ajuste, se consideró reportar el UCL y UTL excluyendo los valores potencialmente atípicos . Pues representa la población principal dominante. Para este último, el estadístico (0,139) y valor crítico (0,202) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Sb	1,695	3,458	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,168) y valor crítico (0,22) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de normalidad solo para valores detectados, también puede corroborarse observando la gráfica de QQ, además se mantiene la bondad de ajuste (presencia de Normalidad) con y sin valores potencialmente atípicos, se reporta con valores potencialmente atípicos ; el 95% KM (t) UCL y KM	Normal



b



PERÚ

Ministerio
del AmbienteOrganismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Variable	Nivel de fondo	Nivel de referencia	Comentarios	Conclusiones
			95% UTL 95% Coverage debido a que presenta valores censurados.	
K	1 294	1 660	Según AIC y BIC es más adecuado la Normalidad, además el estadístico (0,126) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de normalidad.	Normal
Na	8 522	15 448	Se ha decidido considerar Normalidad, debido al principio de parsimonia*, además el estadístico (0,17) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
TI	0,0513	0,0636	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ y al no encontrar cambios significativos en la bondad de ajuste (se mantiene la Normalidad), se consideró reportar el UCL y UTL con valores potencialmente atípicos. Para este último, el estadístico (0,146) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	Normal
Ti	578,6	629	No cumple ninguna de las pruebas GOF (prueba Shapiro Wilk, Lilliefors, Anderson-Darling Gamma, Kolmogorov-Smirnov Gamma, Shapiro Wilk Lognormal y Lilliefors Lognormal) y el tamaño de la muestra no es grande ($n < 30$), se determina considerar al 95% Jackknife UCL, dado que este valor también coincide con 95% Student's-t UCL y para nivel de referencia se utilizó el 95% Percentile Bootstrap UTL con 95% Coverage.	Distribución libre no paramétrica
Th	7,278	15,57	Según AIC y BIC es más adecuado la lognormal, además cumple las pruebas GOF (Shapiro Wilk Lognormal y Lilliefors Lognormal), se reporta 95% H-UCL y 95% UTL con 95% Coverage ambas para Lognormal.	Lognormal
U	0,113	0,179	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ (atípico de apalancamientos), se consideró reportar el UCL y UTL excluyendo los valores potencialmente atípicos . Pues representa la población principal dominante. No cumple ninguna de las pruebas GOF (prueba Shapiro Wilk, Lilliefors, Anderson-Darling Gamma, Kolmogorov-Smirnov Gamma, Shapiro Wilk Lognormal y Lilliefors Lognormal) y el tamaño de la muestra no es grande ($n < 30$), se determina considerar al 95% Jackknife UCL, dado que este valor también coincide con 95% Student's-t UCL y para nivel de referencia se utilizó el 95% Percentile Bootstrap UTL con 95% Coverage.	Distribución libre no paramétrica
V	88,46	115,50	Se ha decidido considerar normalidad, debido al principio de parsimonia*, además el estadístico (0,148) y valor crítico (0,192) de la prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de normalidad.	Normal
W	0,139	0,168	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ (atípico de apalancamientos) y se encontró cambios significativos en la bondad de ajuste, se consideró reportar el UCL y UTL excluyendo los valores potencialmente atípicos . Pues representa la población principal dominante. Para este último, el estadístico (0,163) y valor crítico (0,213) de la	Normal





PERÚ

Ministerio
del Ambiente

Organismo de Evaluación y
Fiscalización Ambiental - OEFA

Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

Variable	Nivel de fondo	Nivel de referencia	Comentarios	Conclusiones
			prueba Lilliefors al ser el primero menor que el segundo, no se rechaza la H_0 de Normalidad.	
Zn	85,99	112,50	Se calculó con y sin los valores potencialmente atípicos, se observó la gráfica de QQ (atípico de apalancamientos) y se encontró cambios significativos en la bondad de ajuste, se consideró reportar el UCL y UTL excluyendo los valores potencialmente atípicos . Pues representa la población principal dominante. Para este último, cumple las pruebas GOF (Shapiro Wilk Lognormal y Lilliefors Lognormal), se reporta 95% H-UCL y 95% UTL con 95% Coverage ambas para Lognormal.	Lognormal

* Aplicando el principio de parsimonia, debido que los datos presentan varios de distribución, se eligió el modelo más simple.





PERÚ

Ministerio
del Ambiente

Organismo de Evaluación y
Fiscalización Ambiental - OEFA

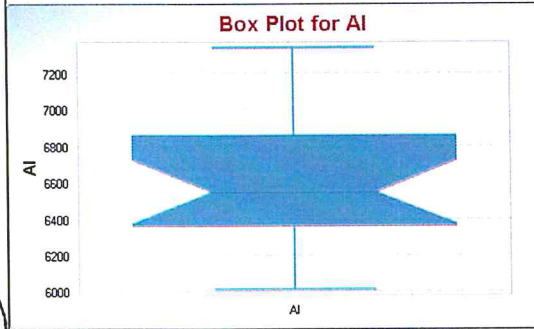
Dirección de Evaluación

"Año del Buen Servicio al Ciudadano"

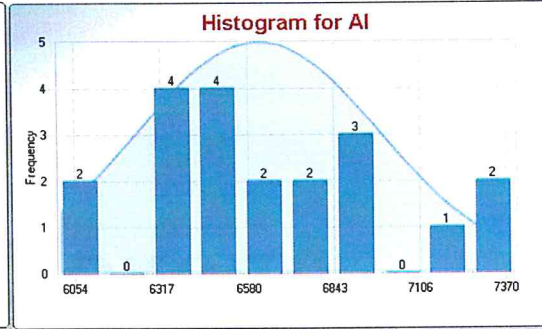
Anexo C: Análisis exploratorio de las variables



a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

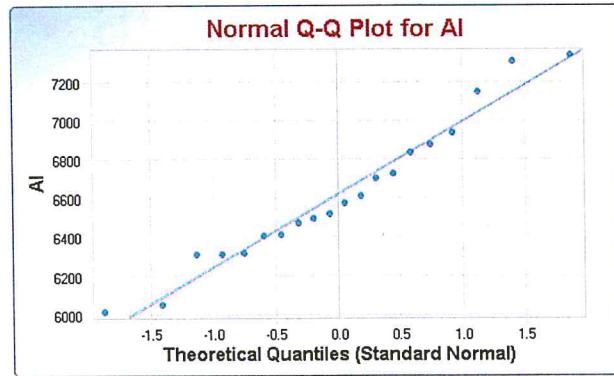


Figura C - 1: Datos de aluminio (Al); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



Handwritten signature

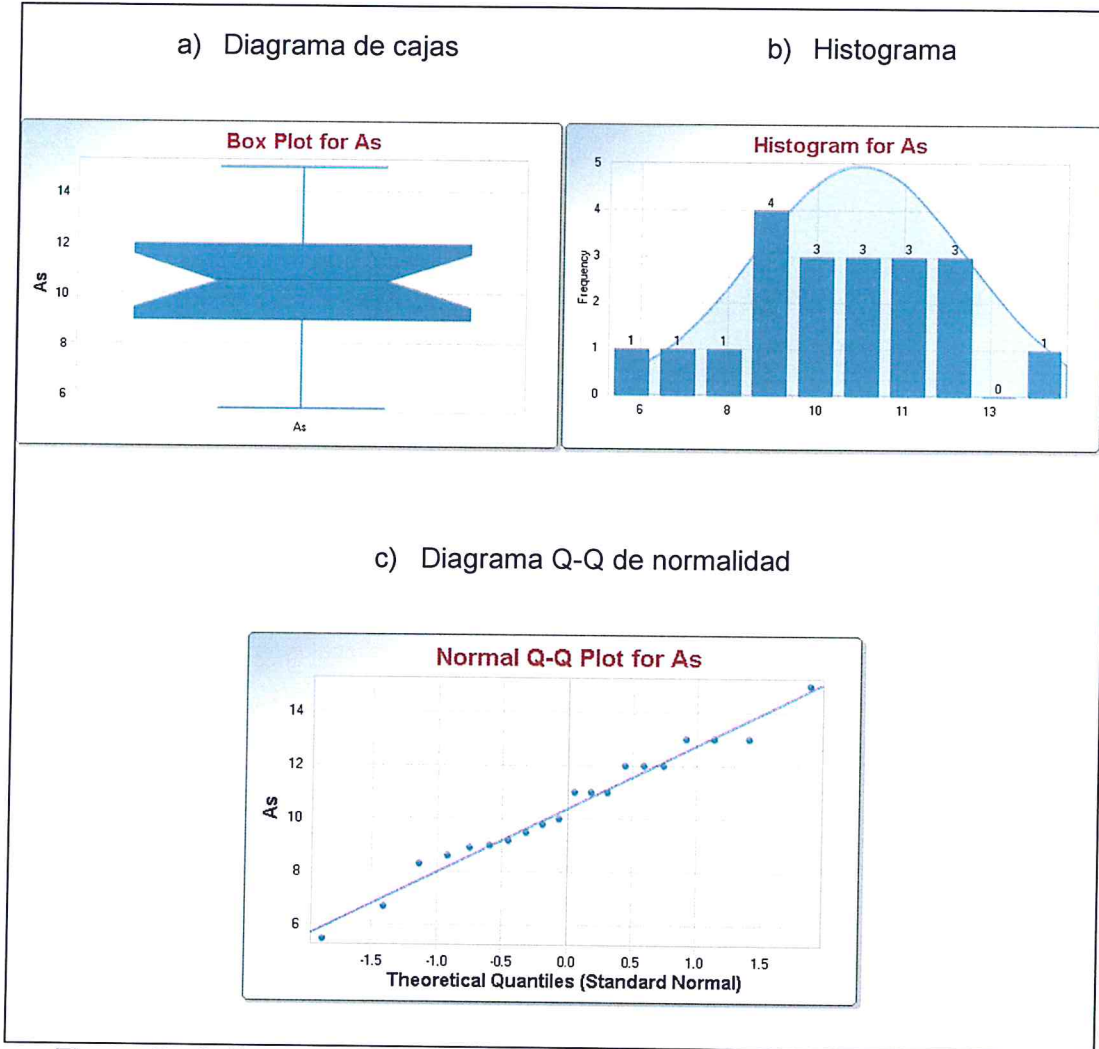


Figura C - 2: Datos de arsénico (As); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



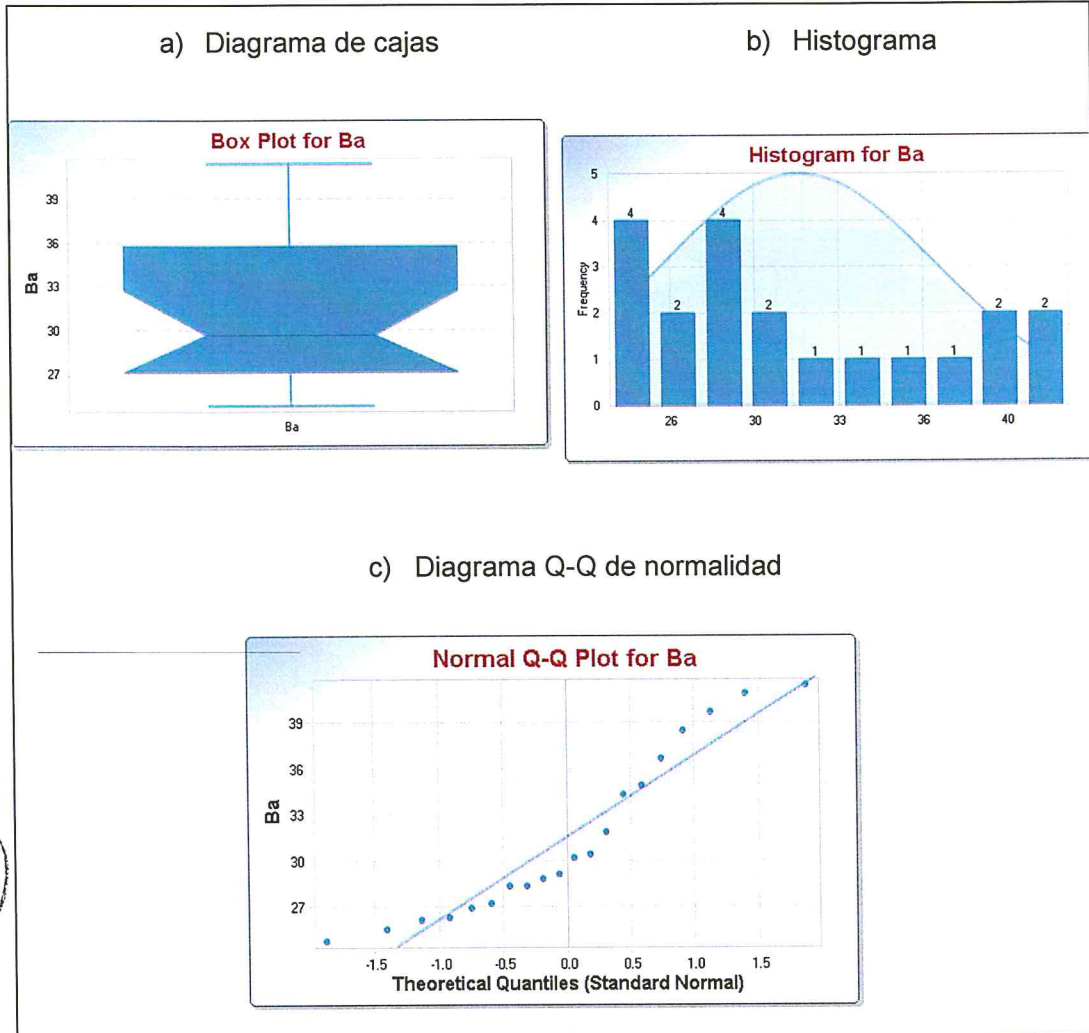
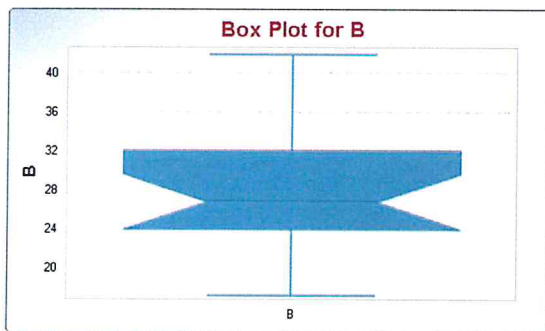
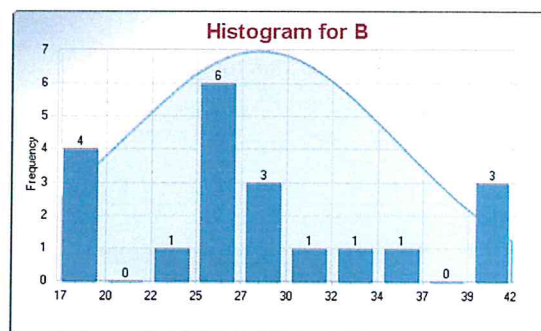


Figura C - 3: Datos de bario (Ba); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

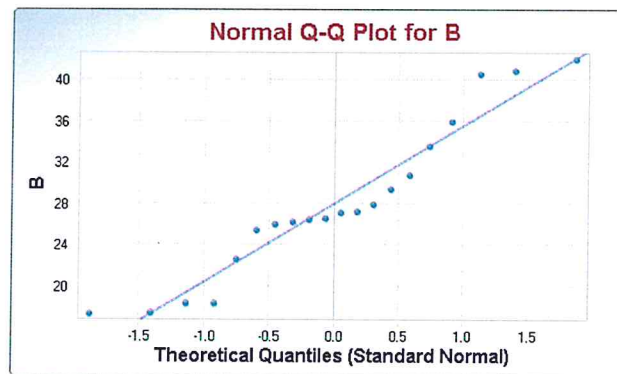
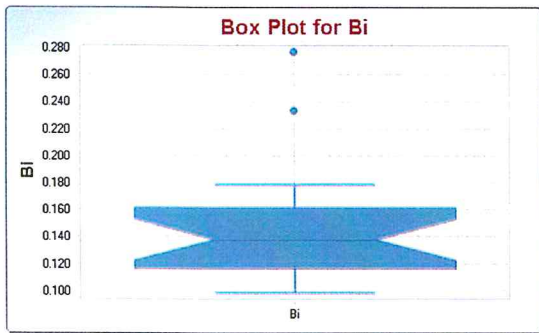


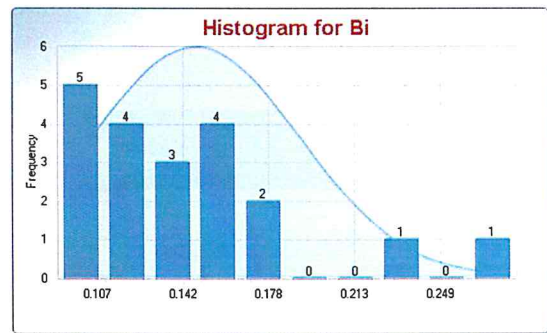
Figura C - 4: Datos de boro (B); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

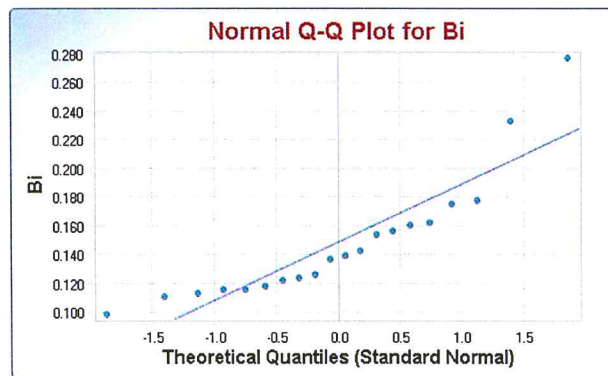
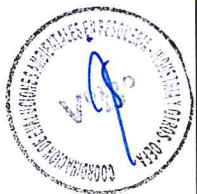


Figura C - 5: Datos de bismuto (Bi); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



Handwritten signature

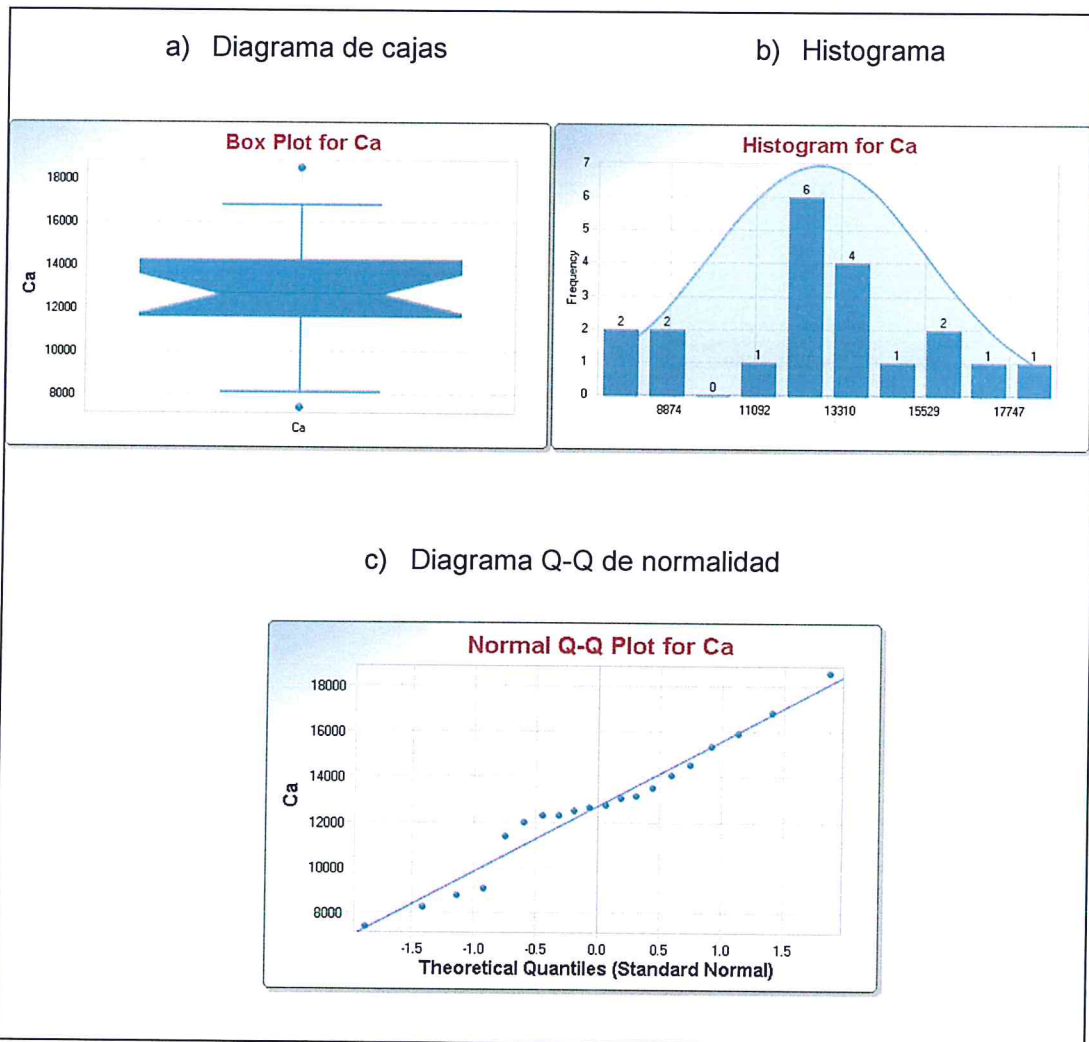
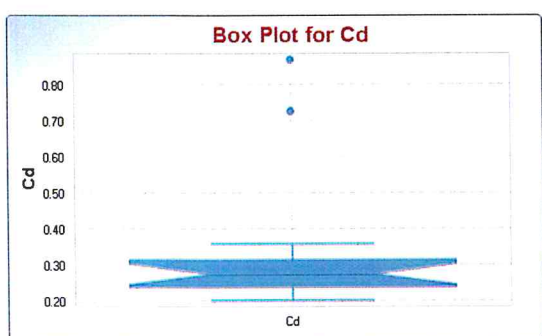


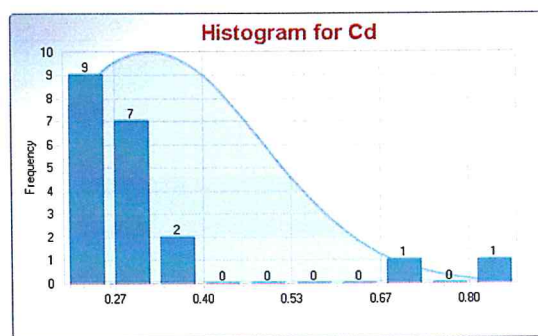
Figura C - 6: Datos de calcio (Ca); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

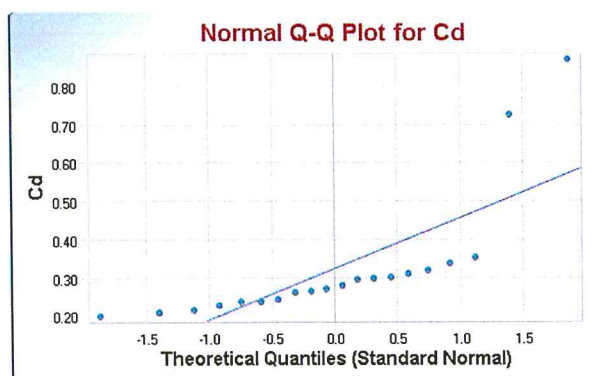


Figura C - 7: Datos de cadmio (Cd); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

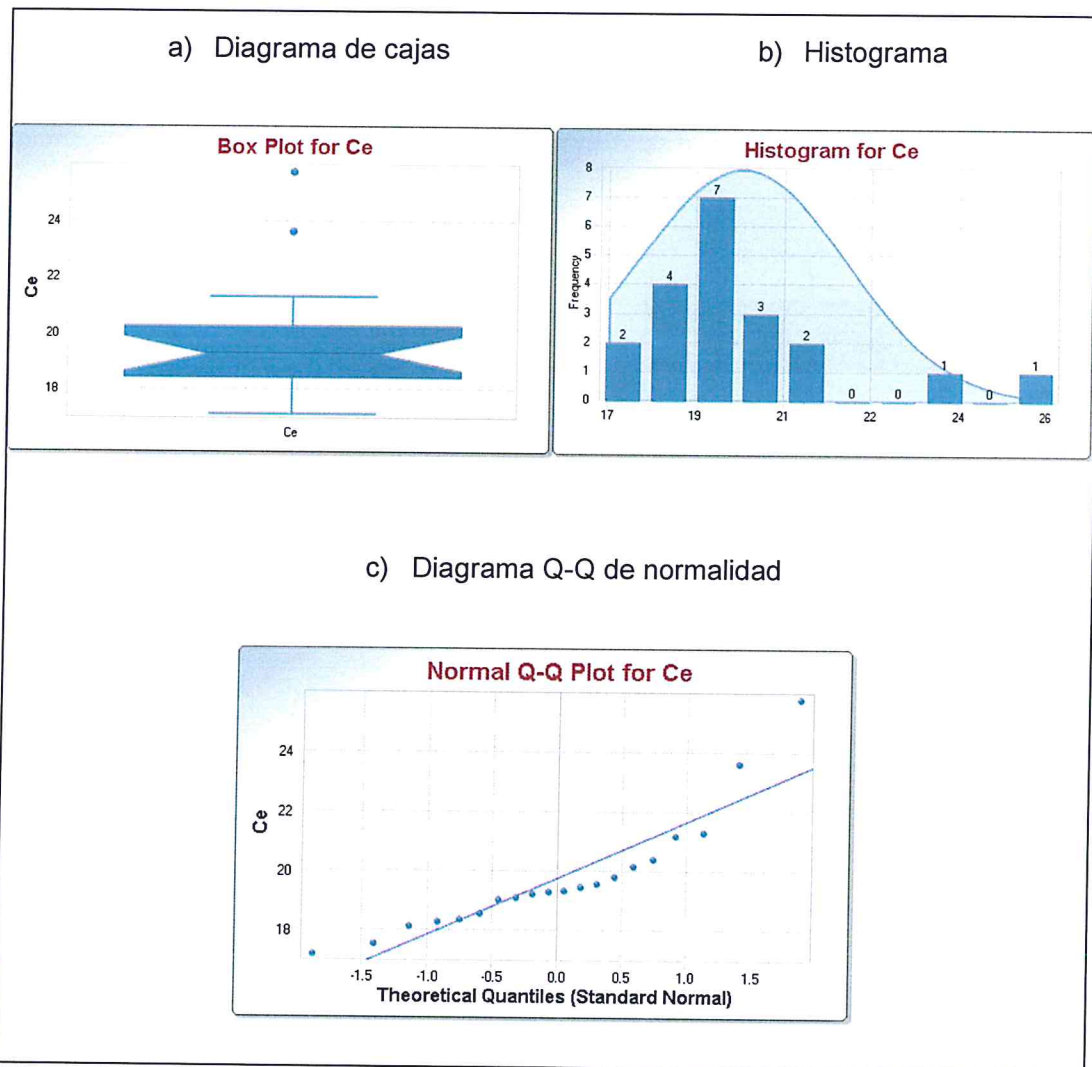
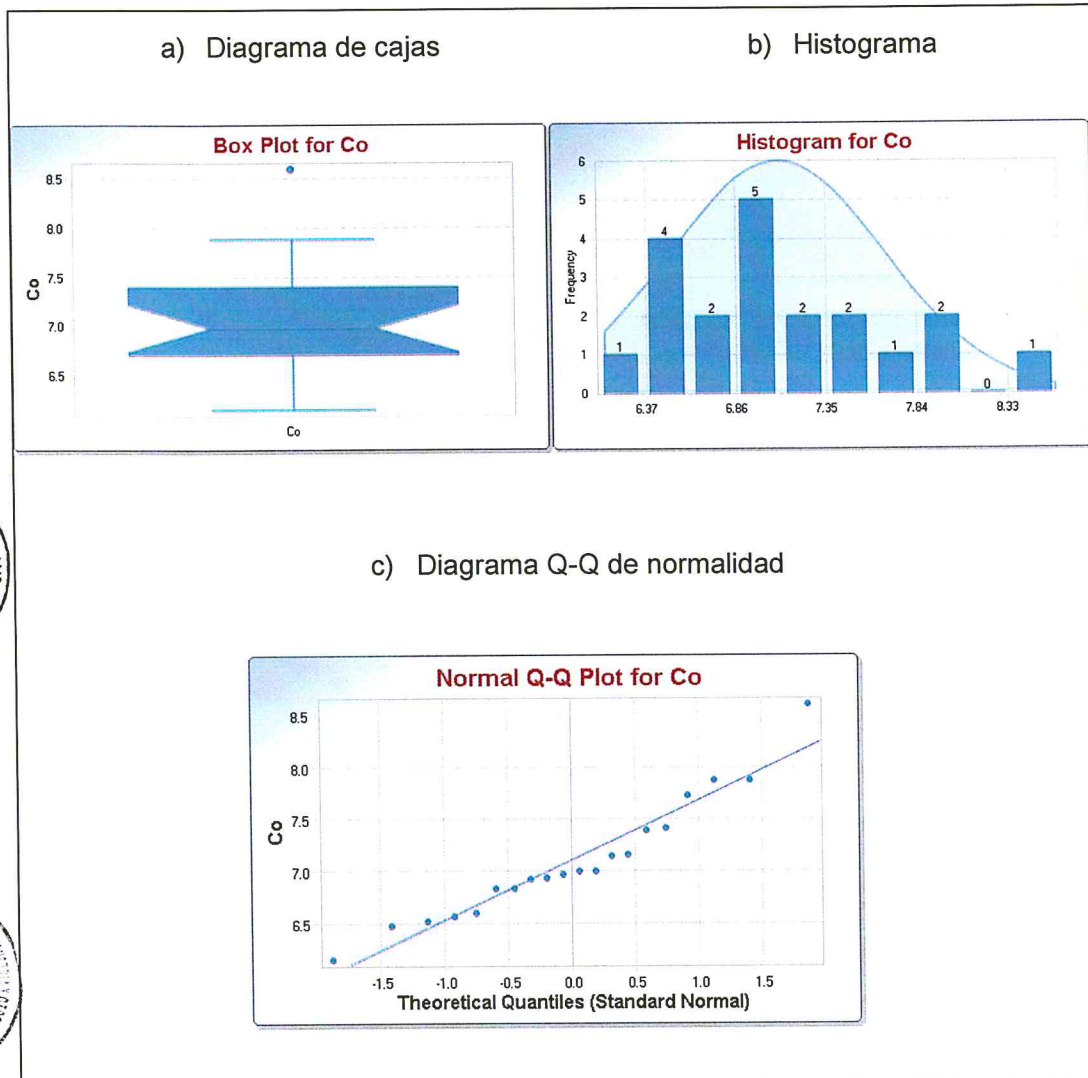


Figura C - 8: Datos de cerio (Ce); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



Handwritten signature



Handwritten signature

Figura C - 9: Datos de cobalto (Co); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

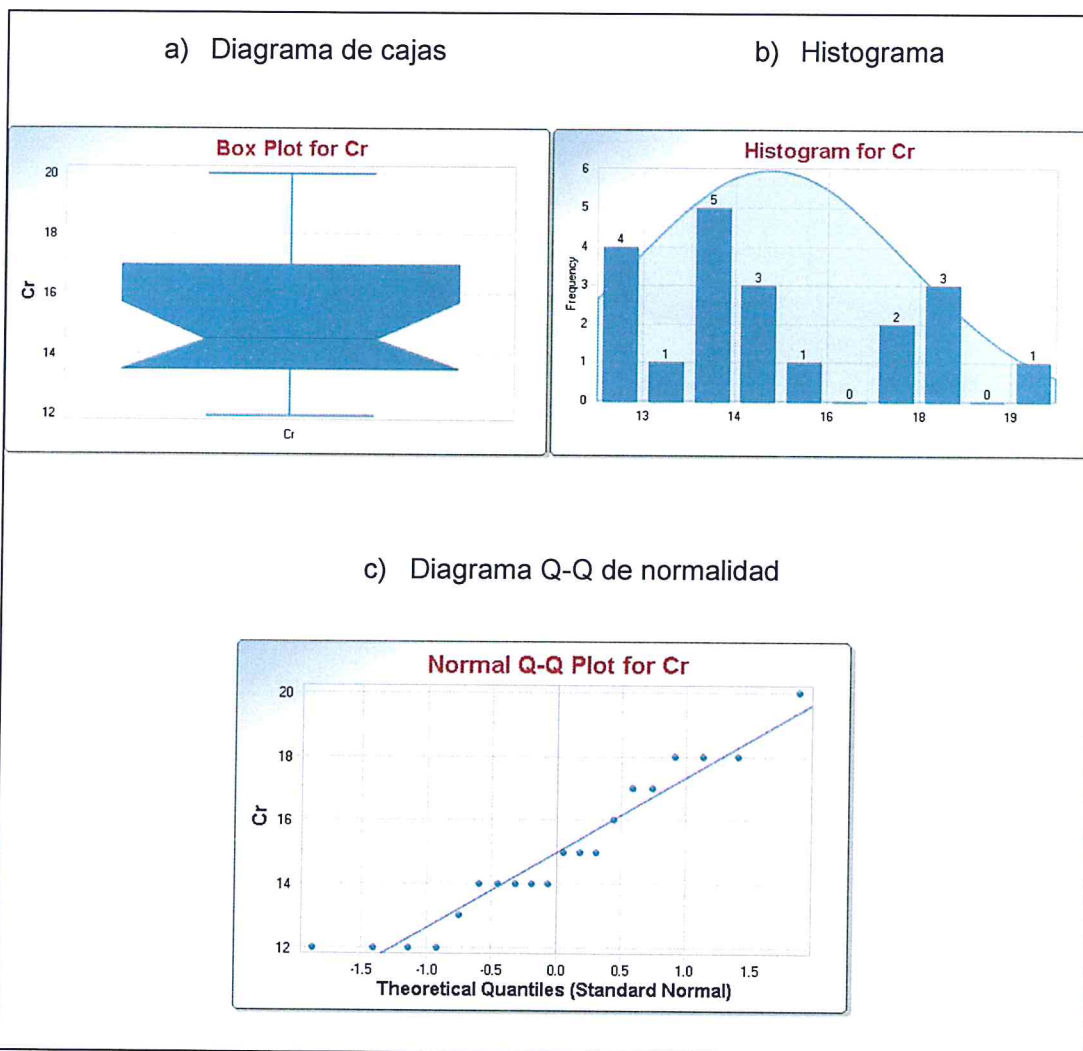


Figura C - 10: Datos de cromo (Cr); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

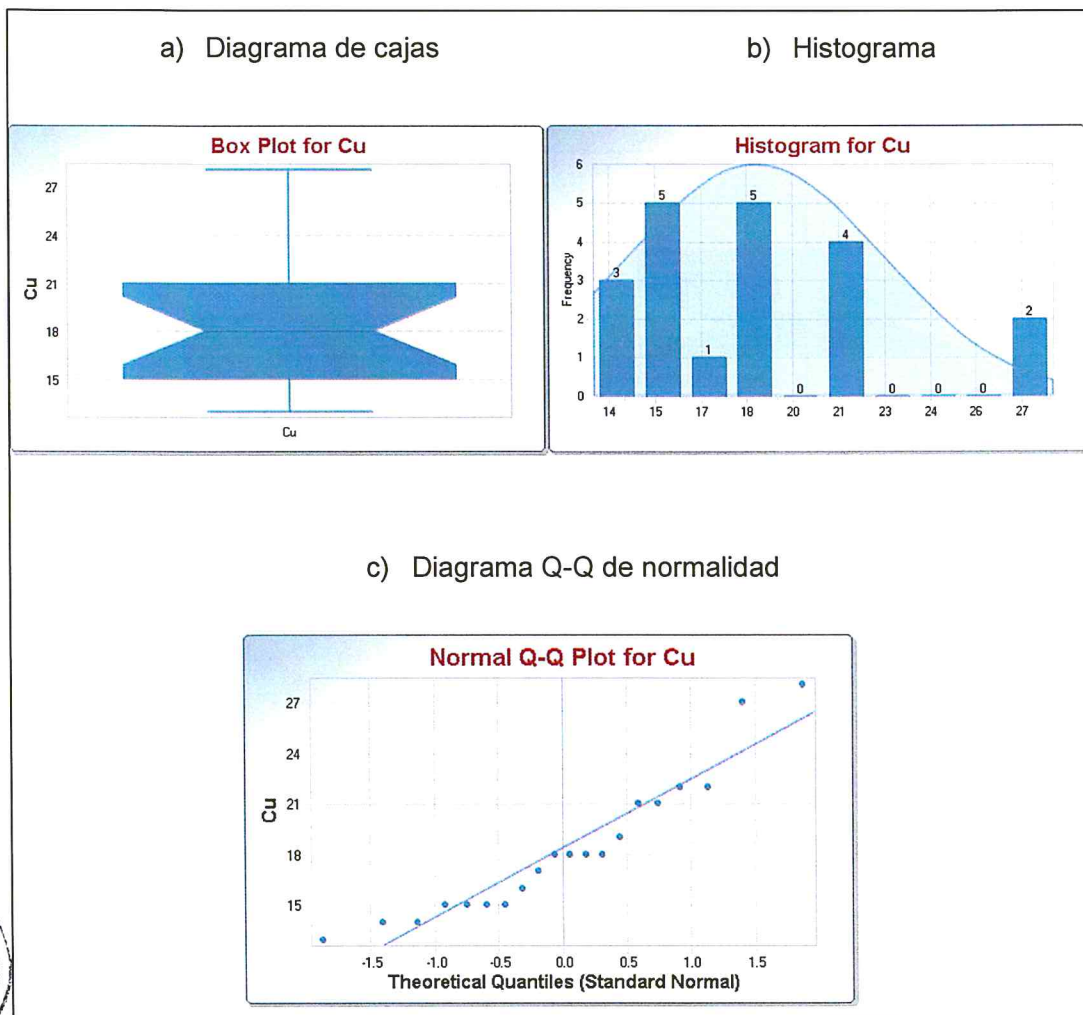
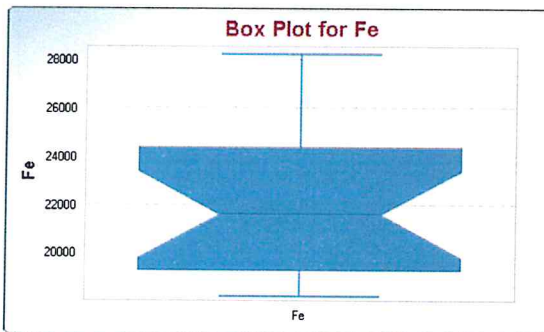
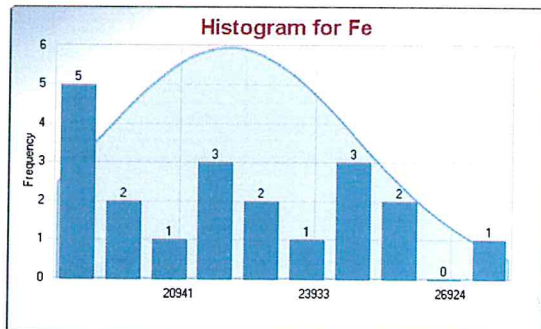


Figura C - 11: Datos de cobre (Cu); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

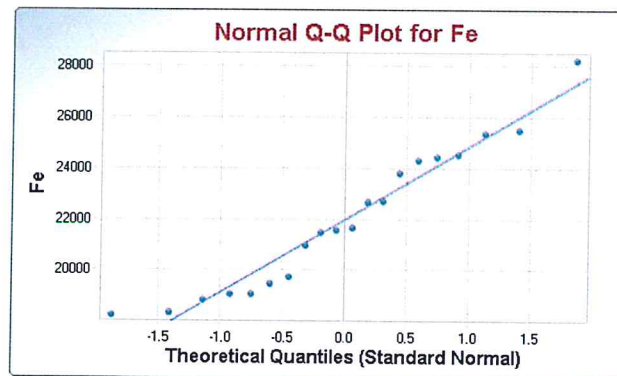
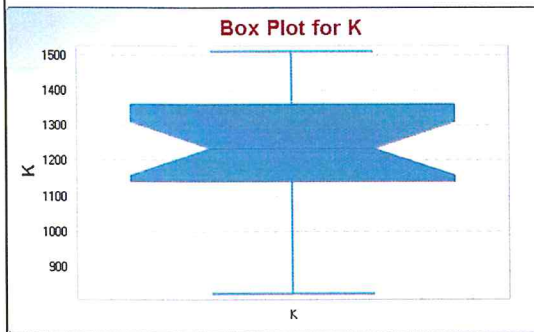
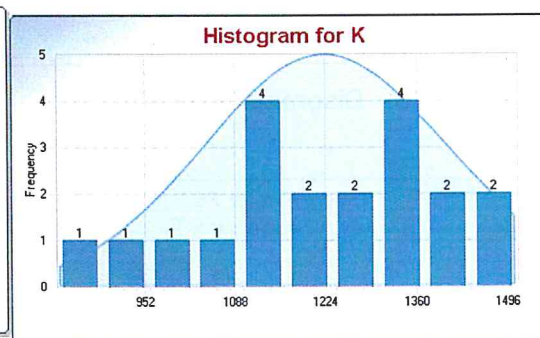


Figura C - 12: Datos de hierro (Fe); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

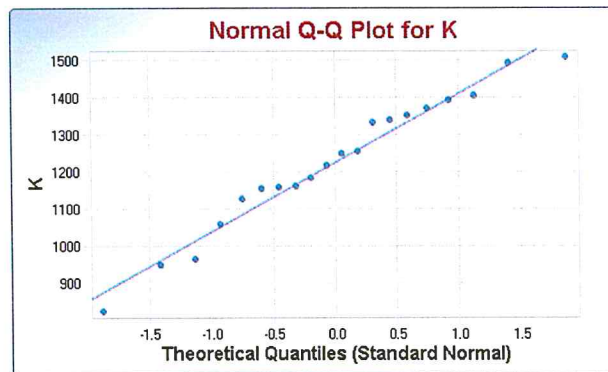
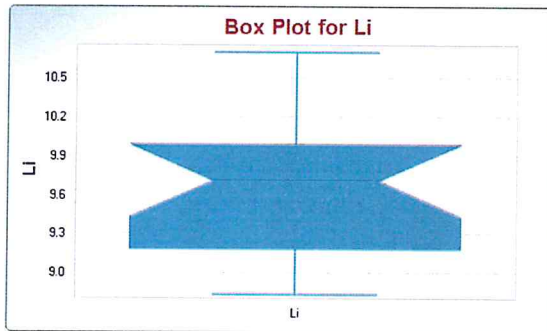


Figura C - 13: Datos de potasio (K); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

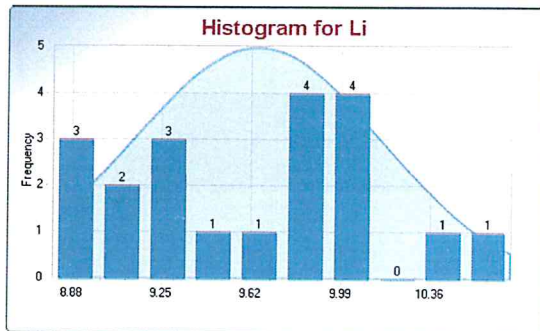


Handwritten signature

a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

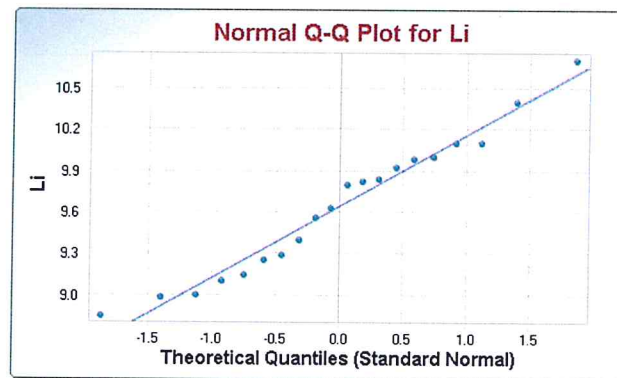
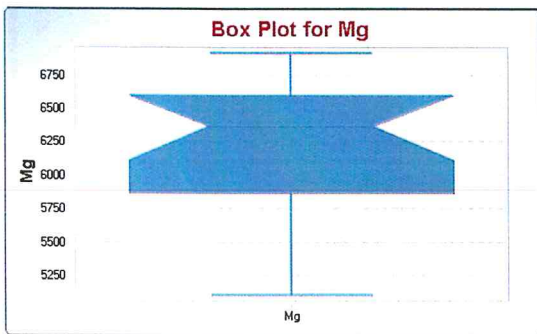


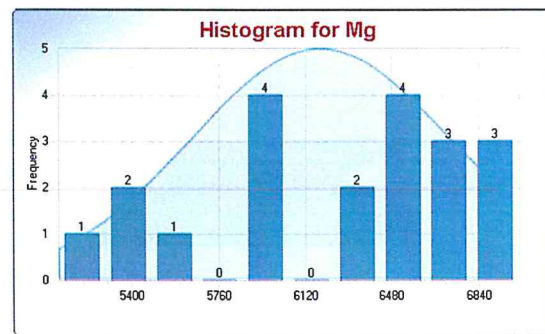
Figura C - 14: Datos de litio (Li); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

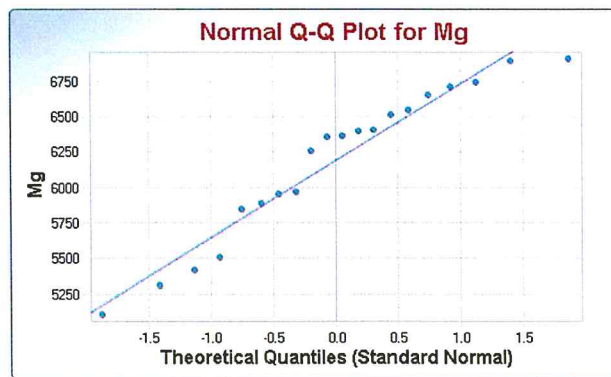


Figura C - 15: Datos de magnesio (Mg); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



ll

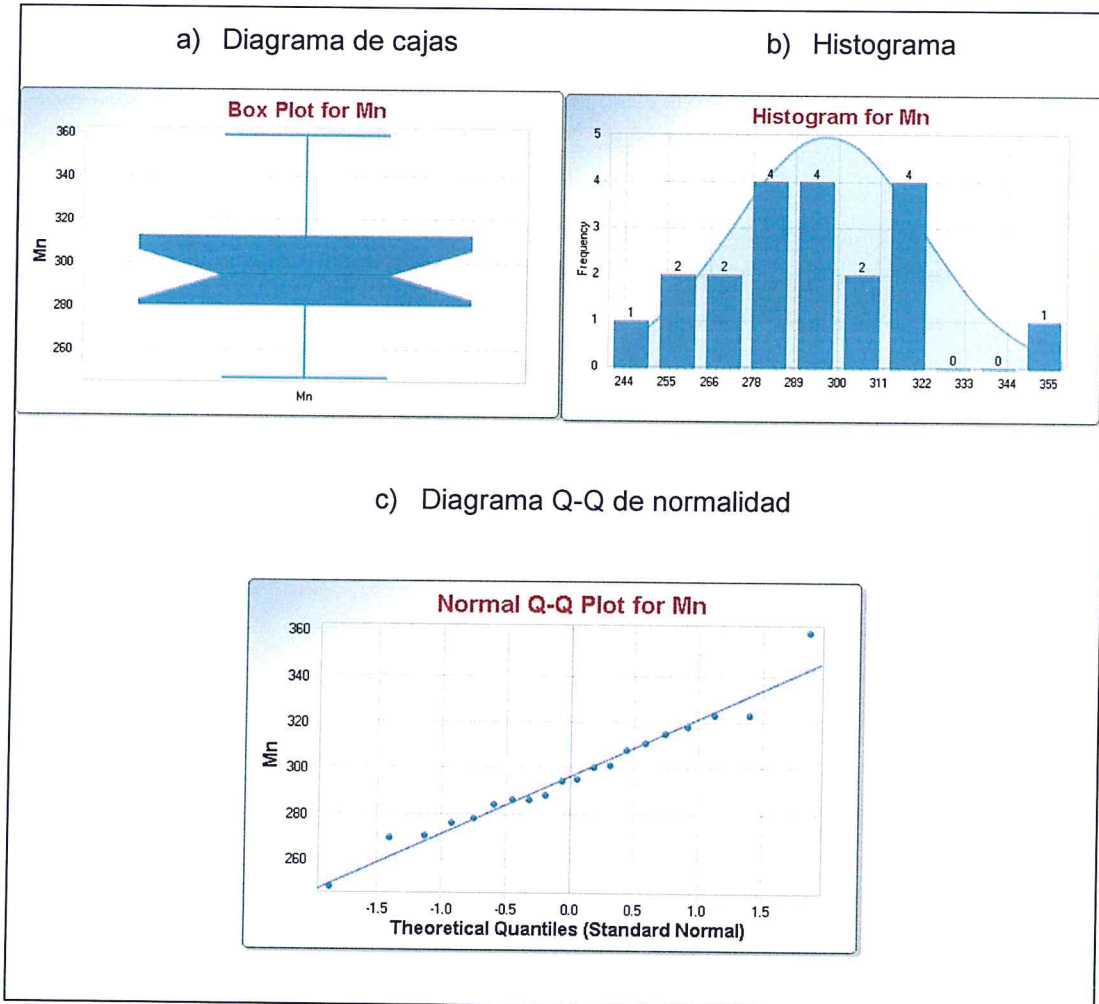
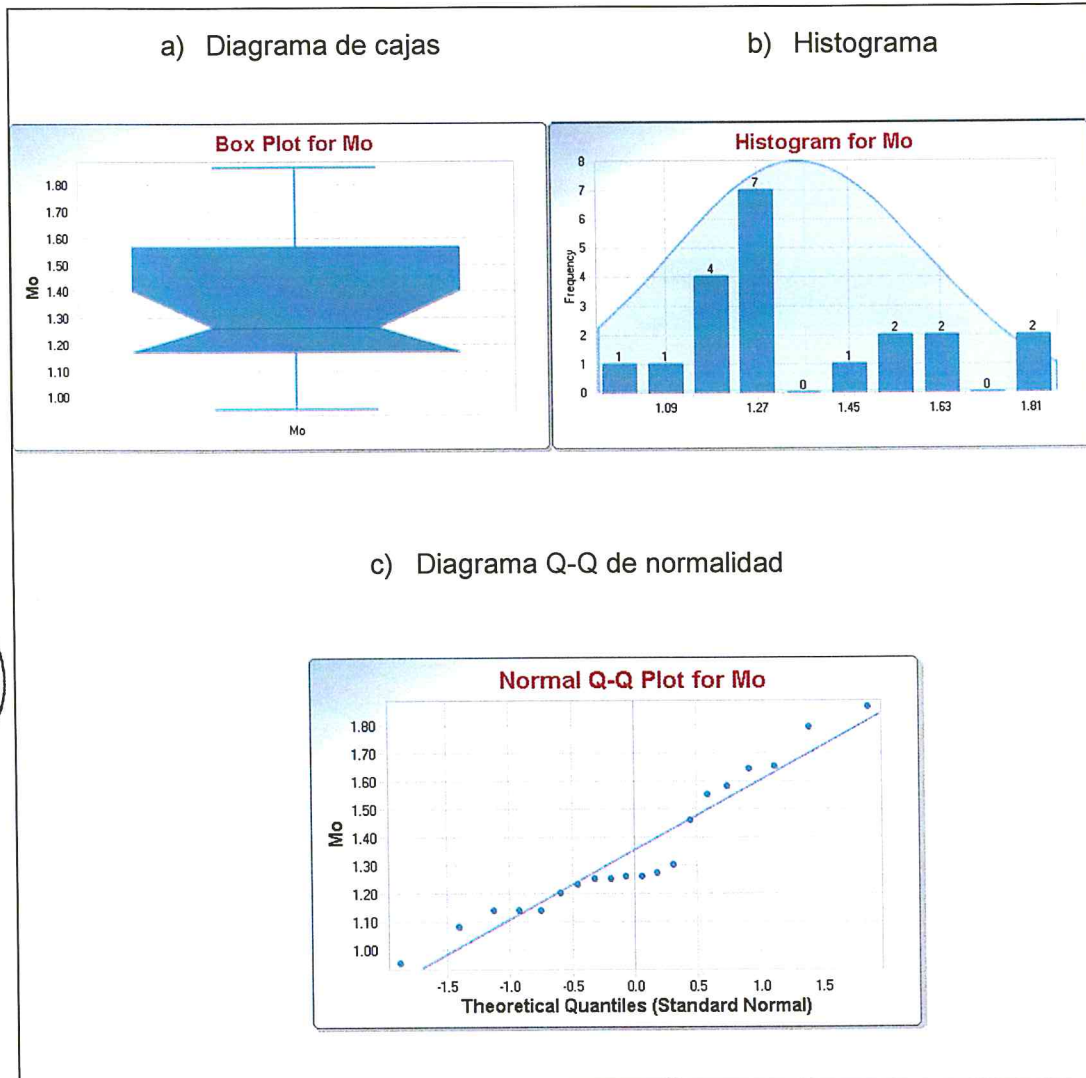


Figura C - 16: Datos de manganeso (Mn); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



[Handwritten signature]



ll

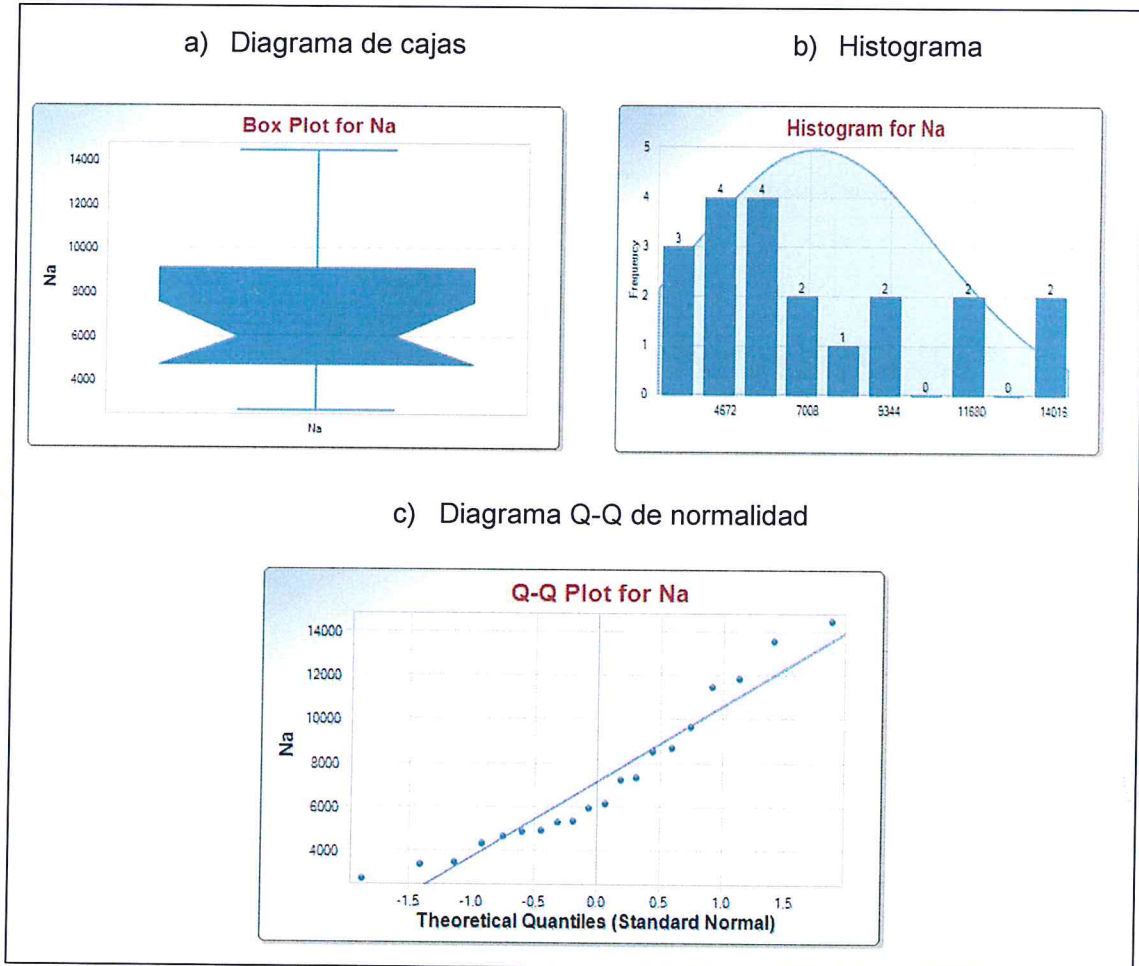


Figura C - 18: Datos de sodio (Na); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



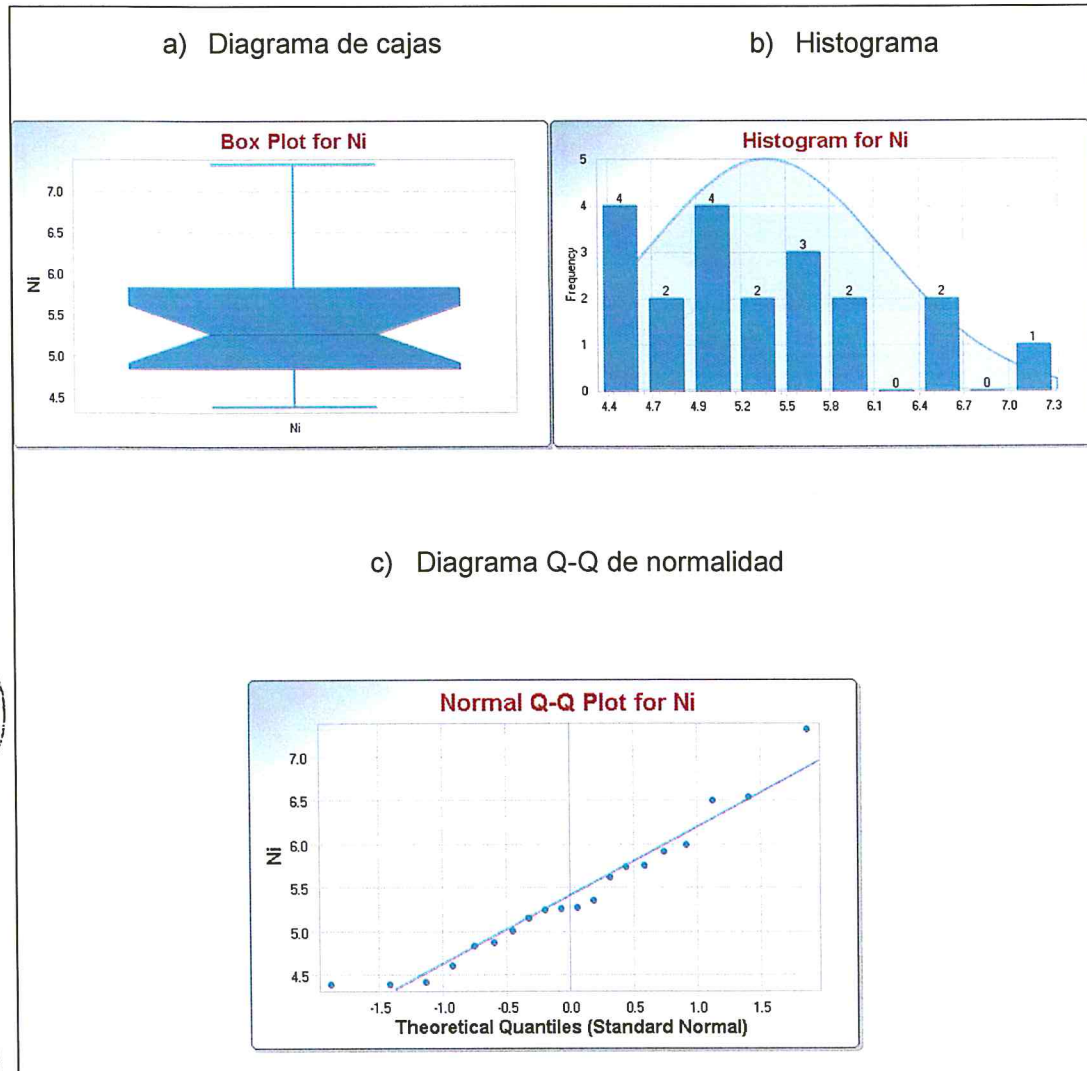


Figura C - 19: Datos de níquel (Ni); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

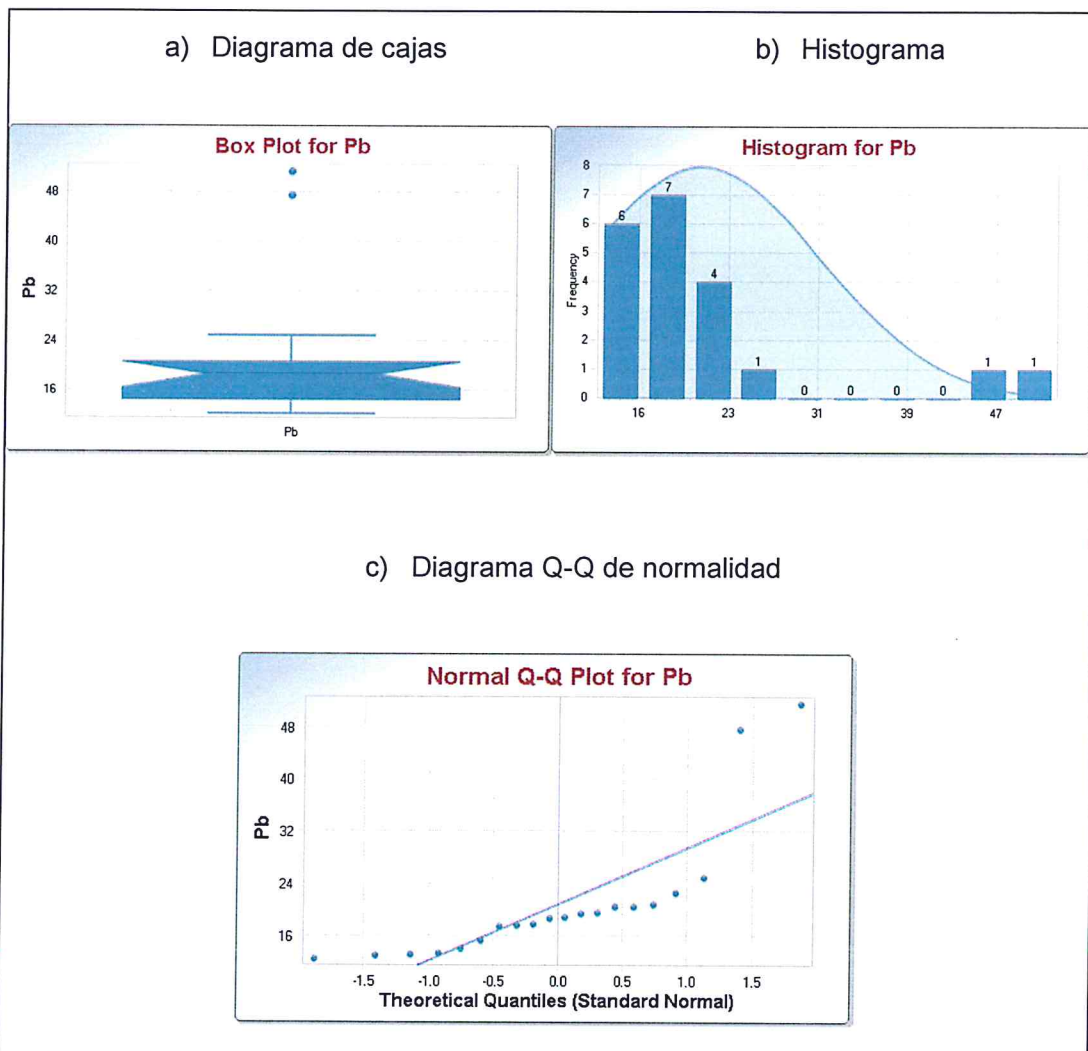


Figura C - 20: Datos de plomo (Pb); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

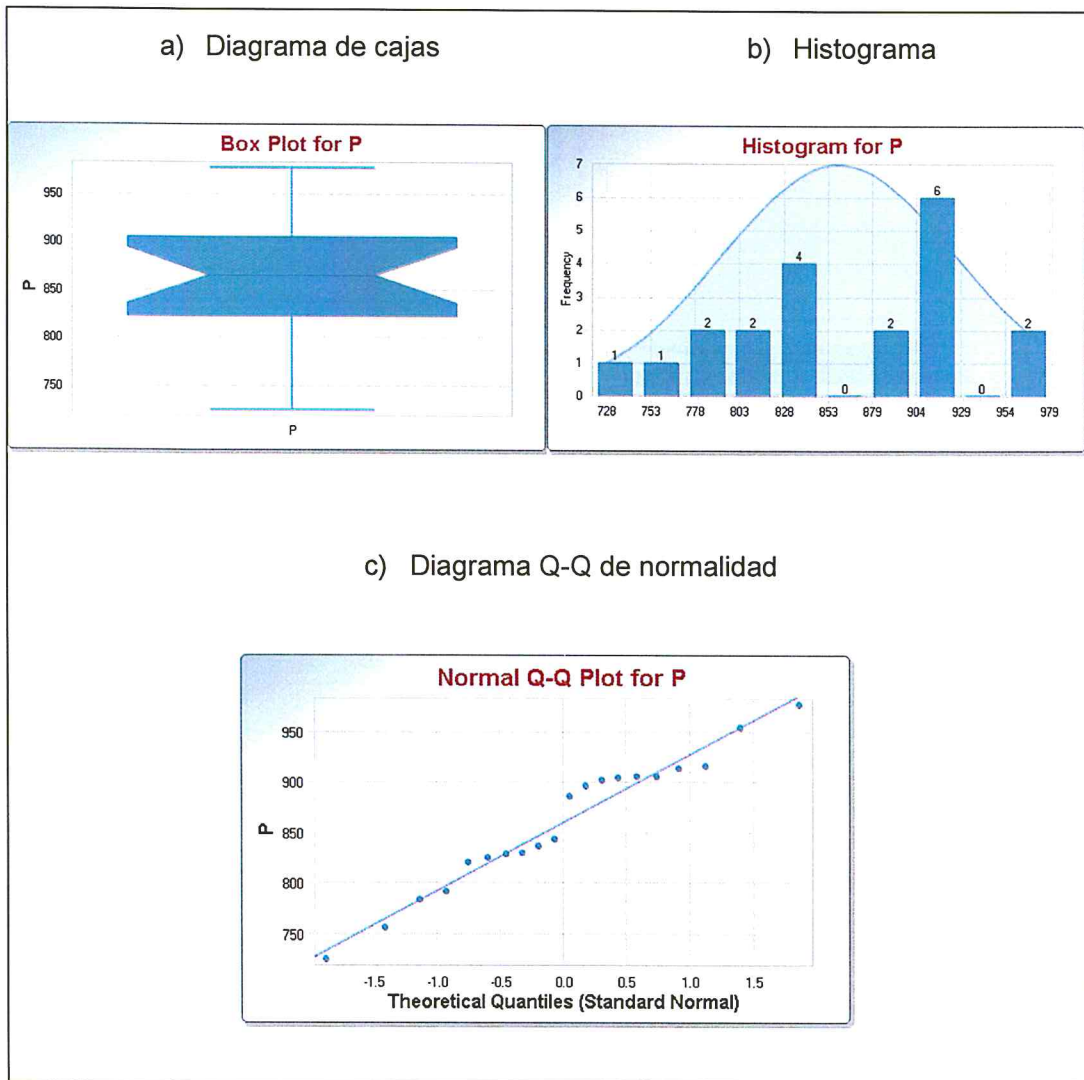


Figura C - 21: Datos de fósforo (P); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



Handwritten signature

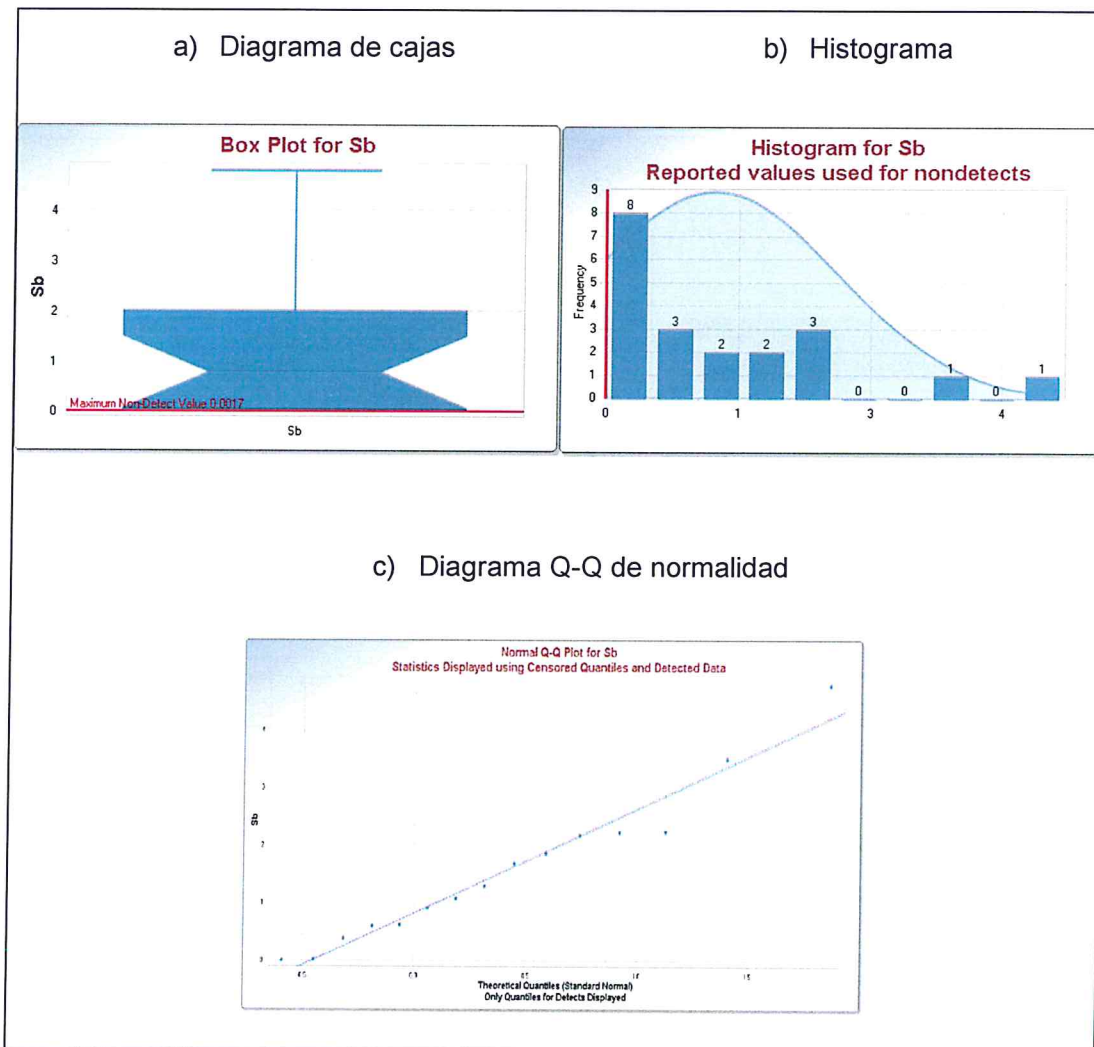
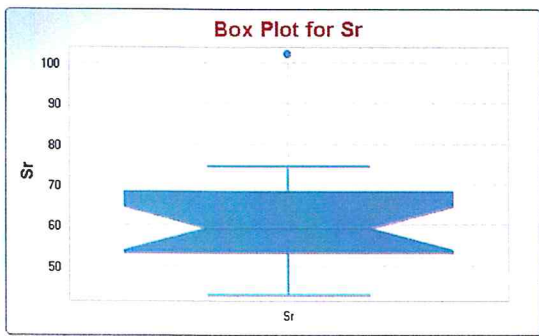
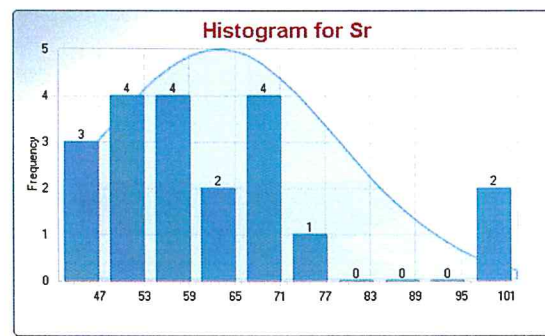


Figura C - 22: Datos de antimonio (Sb); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

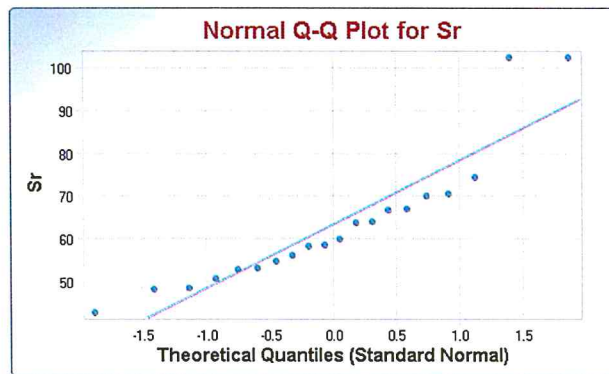
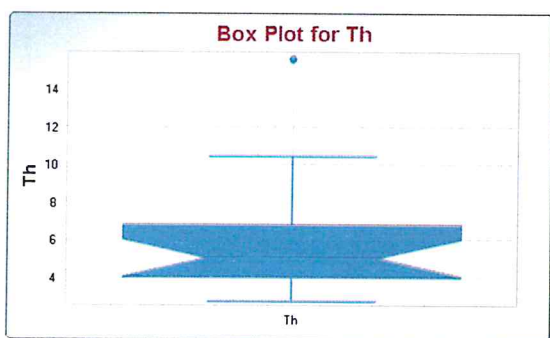


Figura C - 23: Datos de estroncio (Sr); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

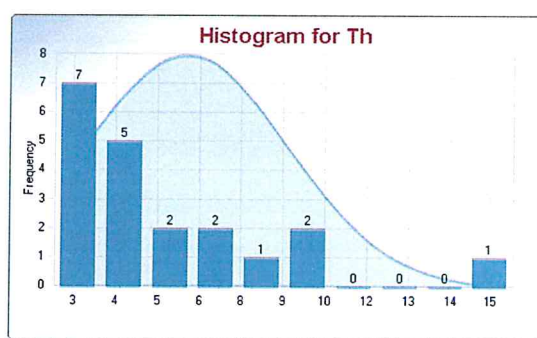


le

a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

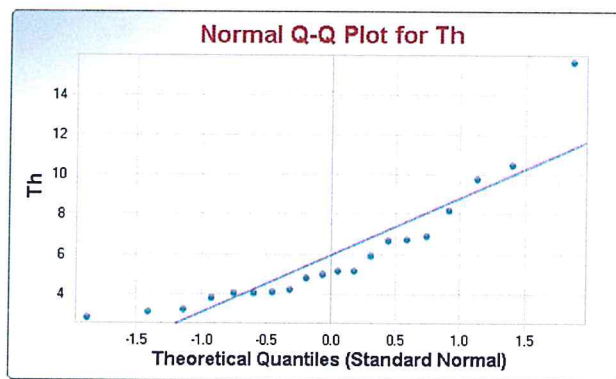


Figura C - 24: Datos de torio (Th); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

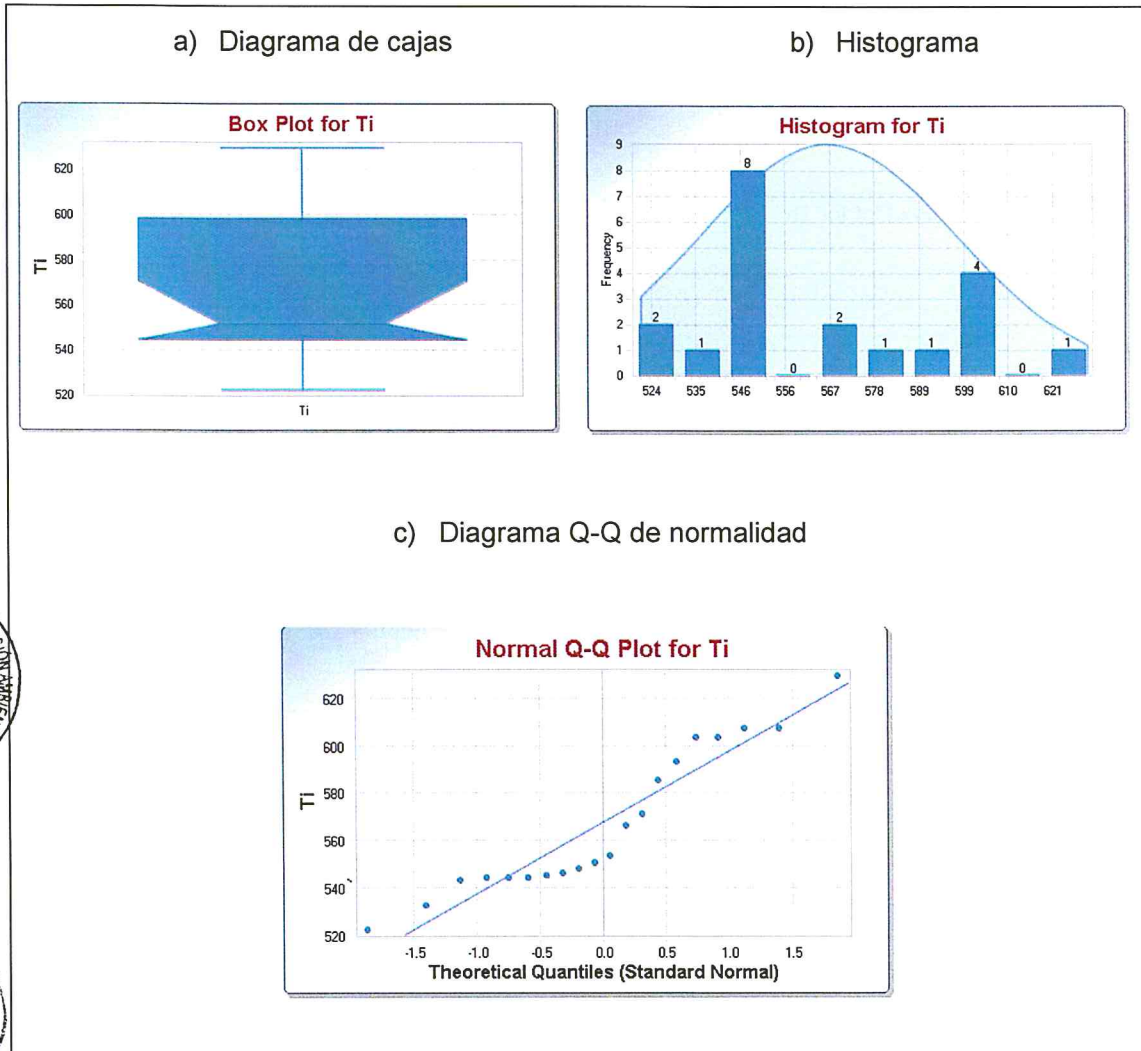
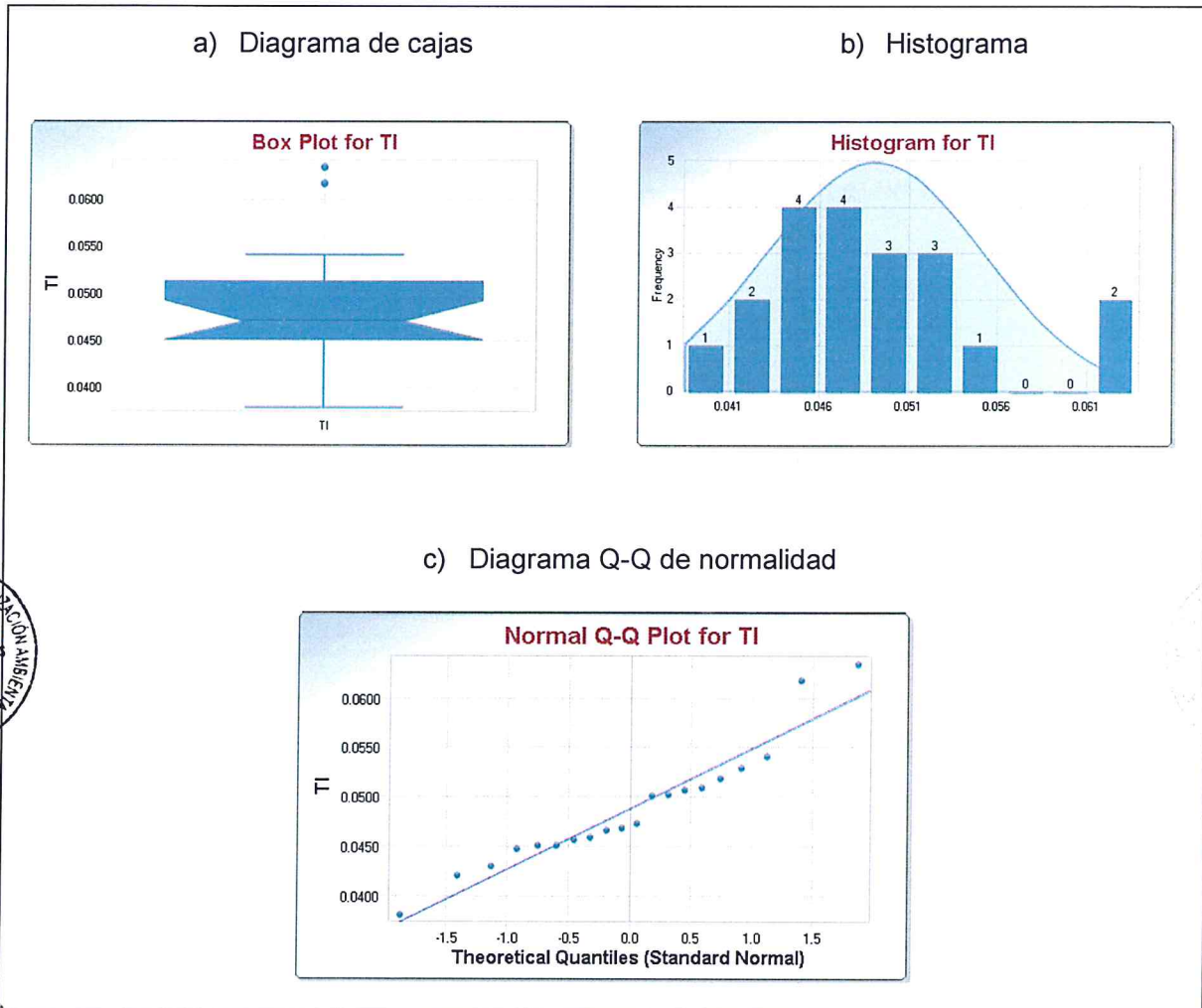


Figura C - 25: Datos de titanio (Ti); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



Handwritten signature



Handwritten signature

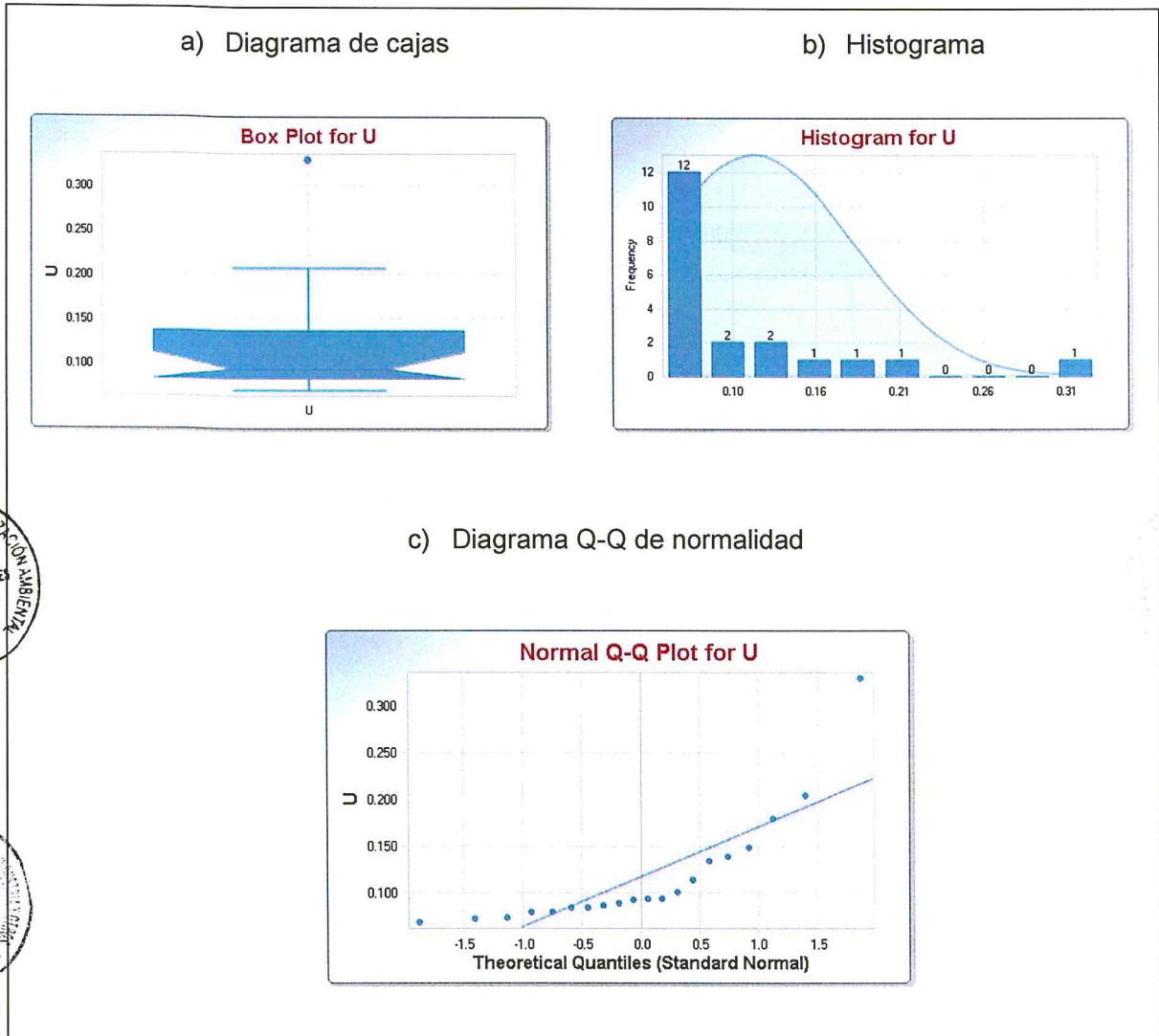
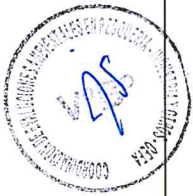
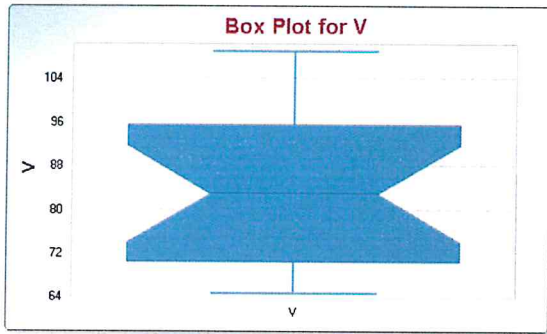


Figura C - 27: Datos de uranio (U); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

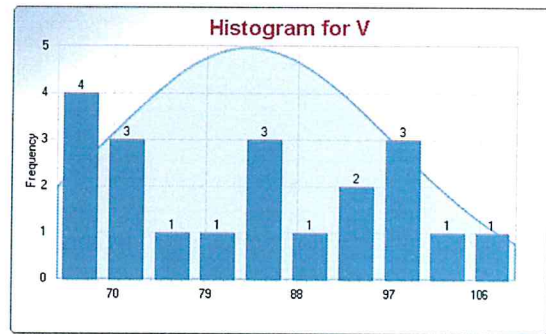


le

a) Diagrama de cajas



b) Histograma



c) Diagrama Q-Q de normalidad

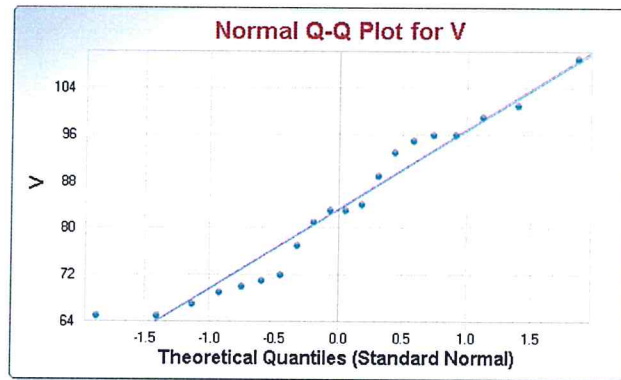


Figura C - 28: Datos de vanadio (V); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad

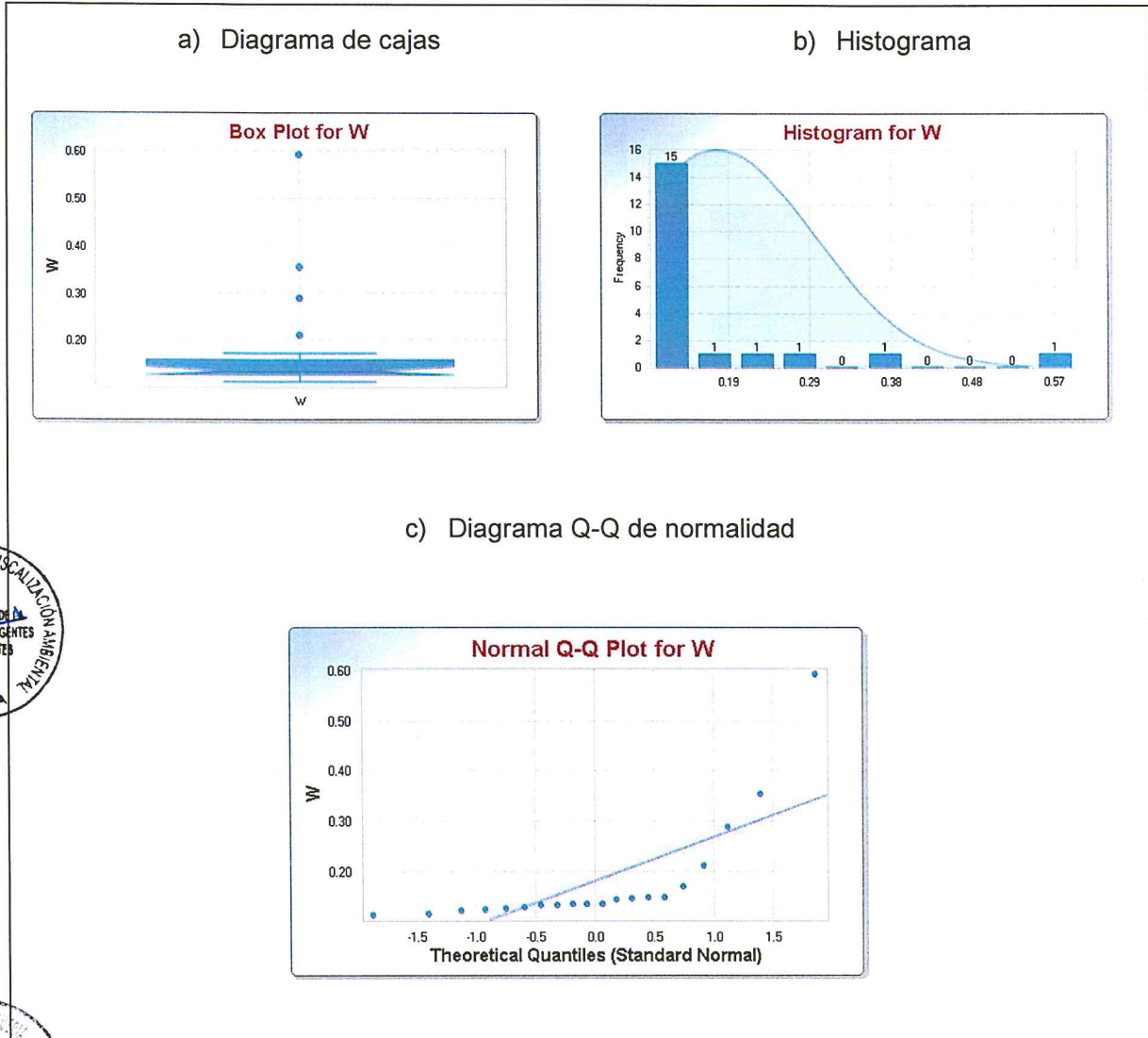
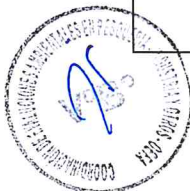


Figura C - 29: Datos de wolframio (W); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



ll

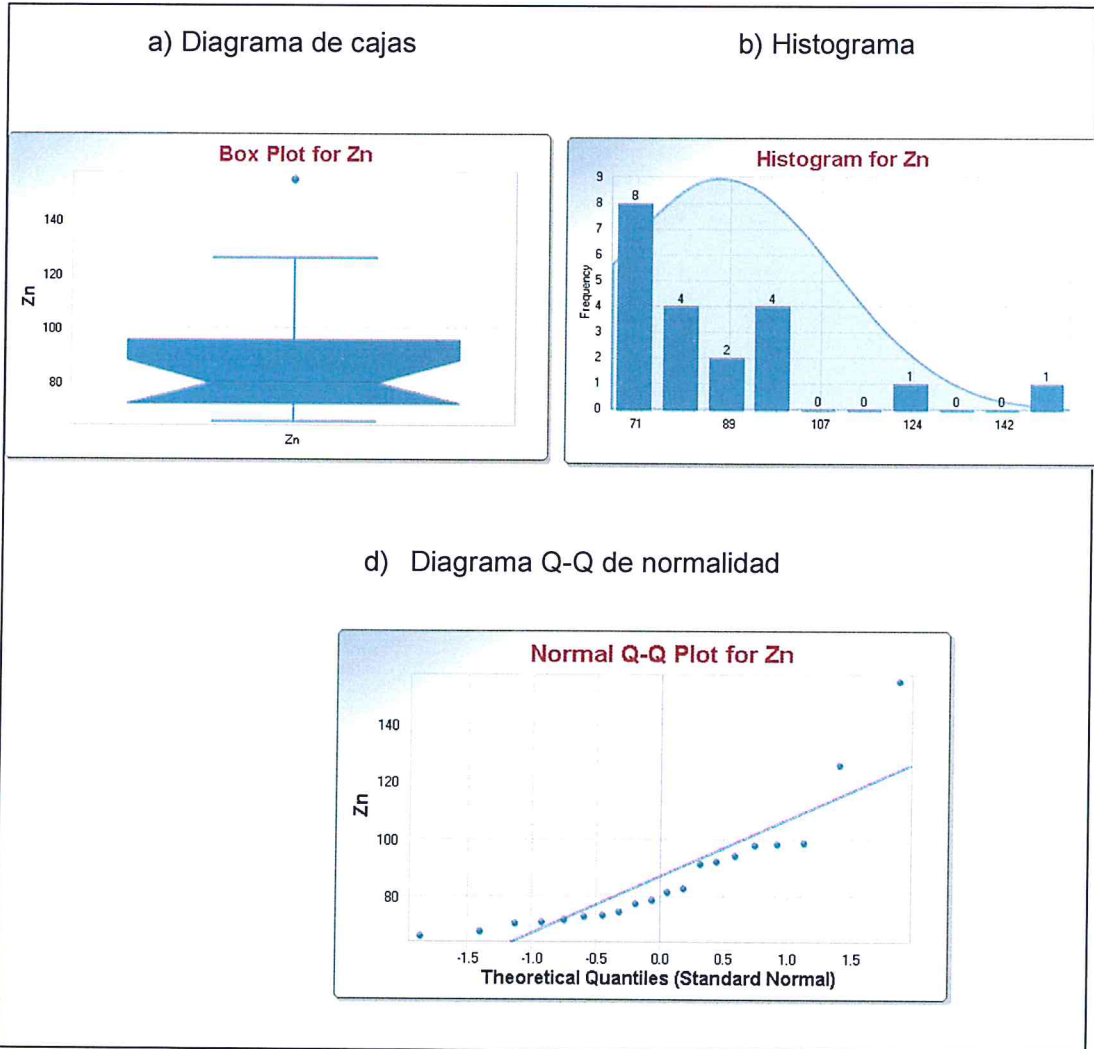


Figura C - 30: Datos de zinc (Zn); a) Diagrama de cajas, b) Histograma y c) Diagrama Q-Q de normalidad



