

Curso de Geoestadística

Posgrado UNAM - IMP

Tarea1: AED

ANÁLISIS EXPLORATORIO DE DATOS DE POZOS DE AGUA: RELACIÓN Li VS. F

CONTENIDO:

1. INTRODUCCIÓN
2. UBICACIÓN DE LA ZONA DE ESTUDIO
3. ESTADÍSTICA UNIVARIADA:
 - i. DATOS DE LITIO
 - ii. DATOS DE FLUORURO
4. ESTADÍSTICA BIVARIADA
5. CONCLUSIONES

1. INTRODUCCIÓN

En algunas partes del mundo, tales como ciertas áreas de la India, Kenya y Tanzania, los suministros de agua natural contienen niveles elevados de flúor ~ 4 ppm. Estas concentraciones de flúor suelen generar importantes efectos en la salud de la población: fluorosis dental y esquelética.

Al padecer de fluorosis dental, la dentadura de la persona es invadida por parches blancos, para posteriormente volverse de color café con áreas decoloradas. (Foto 1). La fluorosis esquelética tiene consecuencias más serias y puede resultar de un prolongado consumo de agua con mayores niveles de flúor de 4 a 15 ppm. Un estudio realizado en el norte de Tanzania reveló una alta incidencia de anormalidades fluoróticas en los huesos de sujetos mayores que usualmente consumían agua con altos niveles de flúor. Los exámenes radiológicos demostraron que los huesos son muy densos o escleróticos y que la calcificación anormal es común en los ligamentos intervertebrales, donde los tendones unen los músculos a los huesos. La foto 2 muestra un ejemplo en el antebrazo. La fluorosis esquelética puede causar dolor de espalda y rigidez, así como deformidades neurológicas.



FOTO 1. Efectos de la fluorosis dental.

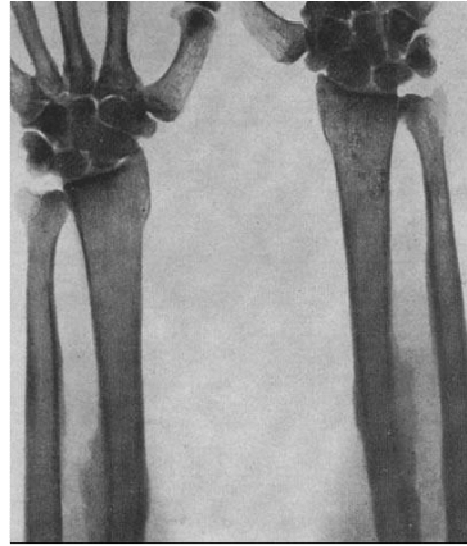


FOTO 2. Radiografía de los antebrazos de una persona con fluorosis esquelética, se evidencia una mayor densidad ósea y calcificación entre el radio y el cubito

En el área escogida para realizar el análisis estadístico los valores de Fluoruro tienen un rango de 0.08-9.8 ppm, por lo que es importante conocer el origen y el comportamiento de dicho ion. La Norma oficial Mexicana No 127 establece un límite máximo permisible de 1.5 ppm. Sin embargo, como se ha observado a nivel mundial, las variaciones en las concentraciones afectan de forma diferente, por lo que este trabajo podría servir de base para establecer límites máximos por regiones.

2. UBICACIÓN DE LA ZONA DE ESTUDIO

El estudio se enfoca en las baterías que suministran agua potable a la Ciudad de León Guanajuato. La ciudad cuenta con un millón de habitantes y sus actividades económicas industriales y agrícolas demandan grandes cantidades de agua. Además, la zona posee ~1200 pozos activos, de los cuales el 70% es utilizado para riego, el 21% para el suministro municipal, y el 9% restante para uso industrial. Las baterías con la que la Ciudad de León (Guanajuato) se abastece de agua potable son: Batería Ciudad, Turbio, Sur, Poniente y Oriente con un total de 119 pozos (Figura 1).

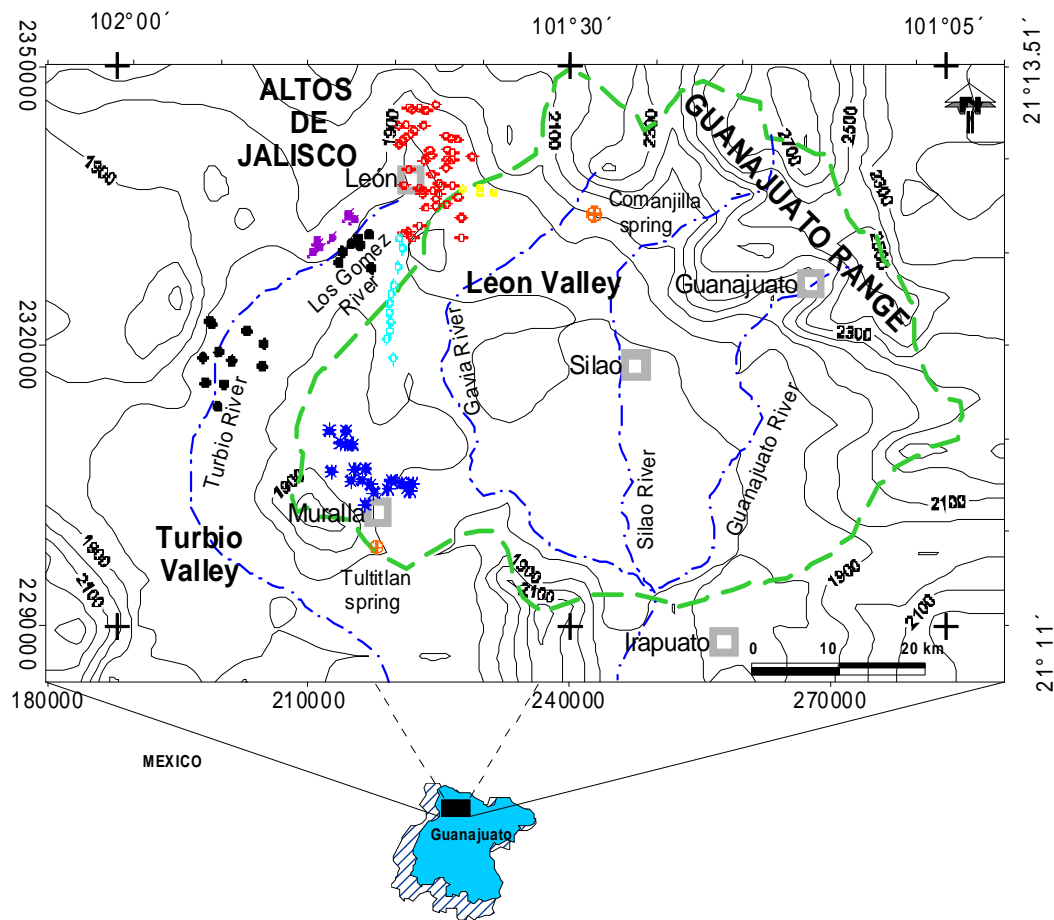


Figura 1. Distribución de baterías de pozos que suministran agua potable a la ciudad de León, Guanajuato. En Azul, la batería Muralla; en negro, la batería Turbio; en azul claro, la batería Sur; en violeta la batería Poniente y en amarillo, la batería Oriente.

Para el análisis exploratorio de los datos, además del Fluoruro (F), se eligió el ion de Litio (Li). Estas variables fueron colectados en 50 pozos de agua potable y tres manantiales, localizados en un área de 40 km x 40 km (Figura 2). En total son 53 datos, cuyos valores se encuentran listados en la Tabla 1. Las diferentes familias de cuerpos de agua presentes en el área de estudio se especifican la figura 3.

Tanto el muestreo como los análisis de laboratorio se realizaron bajo estrictas normas de calidad. Las concentraciones de Li y F se encuentran en ppm. Seguidamente se presentan los resultados del análisis exploratorio de los datos por separado.

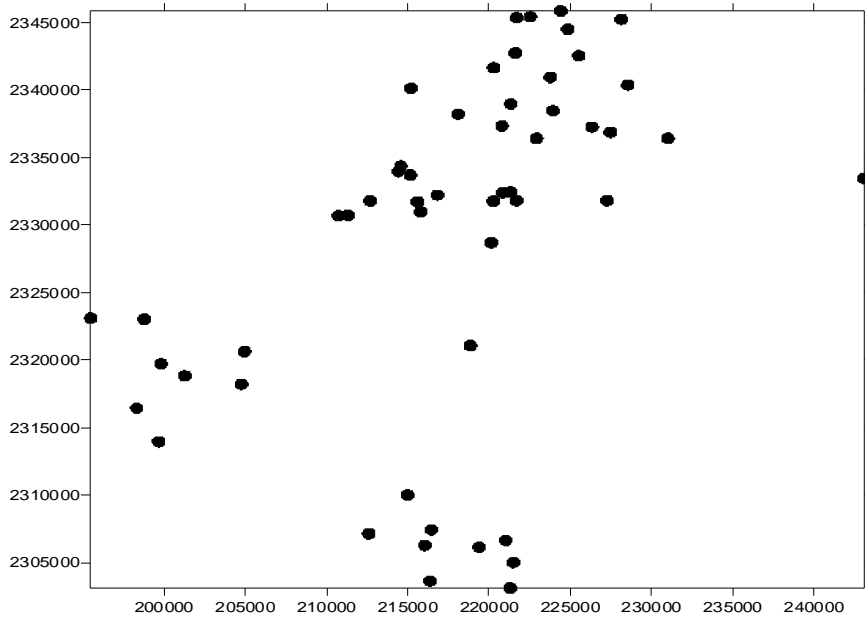


Figura 2. Acercamiento de la distribución de pozos, las coordenadas están dadas en UTM. El área es de ~40 km x 40 km.

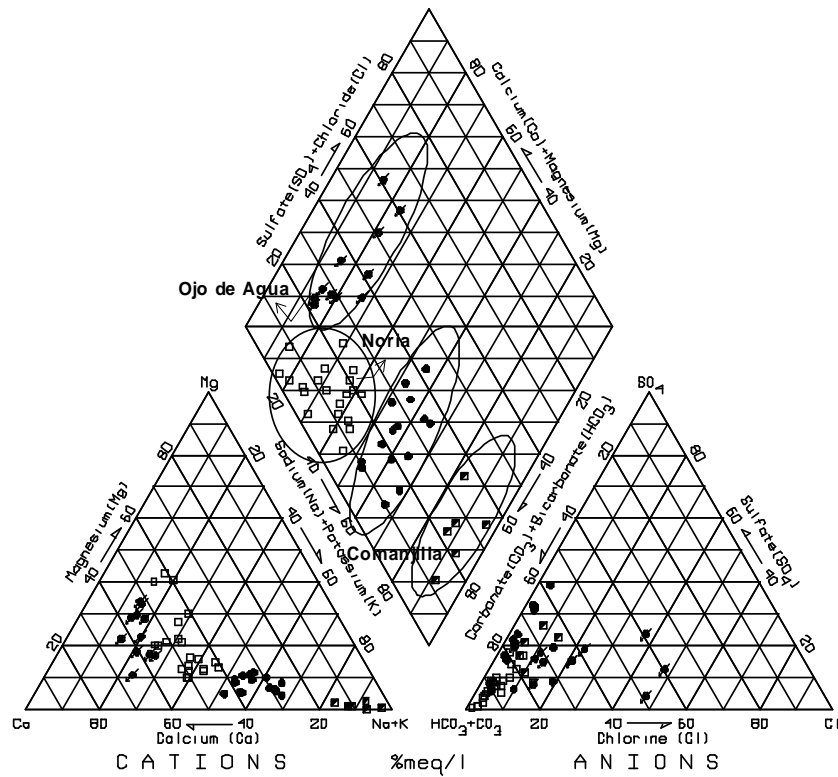


Figura 3. Diagrama de Piper que muestra el comportamiento de cuatro familias de agua. En la construcción del diagrama se utilizan variables (iones) Li y F.

No	Li	F
Fom	0.010	0.10
	(mg/L)	(mg/L)
1	0.020	0.17
2	0.000	0.45
3	0.077	0.41
4	0.112	0.94
5	0.140	1.16
6	0.100	0.71
7	0.100	0.75
8	0.150	0.91
9	0.140	0.91
10	0.060	1.27
11	0.140	1.20
12	0.220	1.45
13	0.000	0.80
14	0.150	0.88
15	0.140	0.95
16	0.147	1.09
17	0.130	1.02
18	0.230	1.38
19	0.060	0.60
20	0.180	1.41
21	0.040	0.26
22	0.030	0.21
23	0.030	0.22
24	0.010	0.10
25	0.020	0.17
26	0.016	0.03
27	0.014	0.33
28	0.019	0.31
29	0.146	0.63
30	0.079	0.39
31	0.023	0.41
32	0.038	0.43
33	0.036	0.65
34	0.026	0.43
35	0.040	0.75
36	0.029	0.71
37	0.038	0.13
38	0.171	0.57
39	0.029	0.15
40	0.023	0.24
41	0.030	0.32
42	0.044	0.34
43	0.040	0.23
44	0.058	0.23
45	0.040	0.03
46	0.019	0.22
47	0.029	0.13
48	0.030	0.03
49	0.023	0.13
50	0.015	0.07
51	0.022	0.71
62	1.077	0.80
53	0.010	0.30

Tabla 1. Datos de F y LI usados para el AED.

3. ESTADÍSTICA UNIVARIADA:

Para realizar el AED se usó una mezcla de programas u hojas de cálculos, entre ellos: Statistica, Gnumeric y Excel.

DATOS DE LITIO

El histograma de frecuencia para esta variable mostró una distribución lejana a la normal o gaussiana, como se muestra en la figura 4. Es leptocurtica y con asimetría positiva. Se observa con claridad un valor atípico, el cual resalta aún más en el diagrama Box Plot, pues está muy alejado del intervalo intercuartil (25%-75%), escapando incluso del rango non-outlier. La mediana es menor que el valor medio y este a su vez es menor que la desviación estándar de la muestra. En la tabla adjunta a la figura 4 se muestran todos los valores estadísticos de la variable Li.

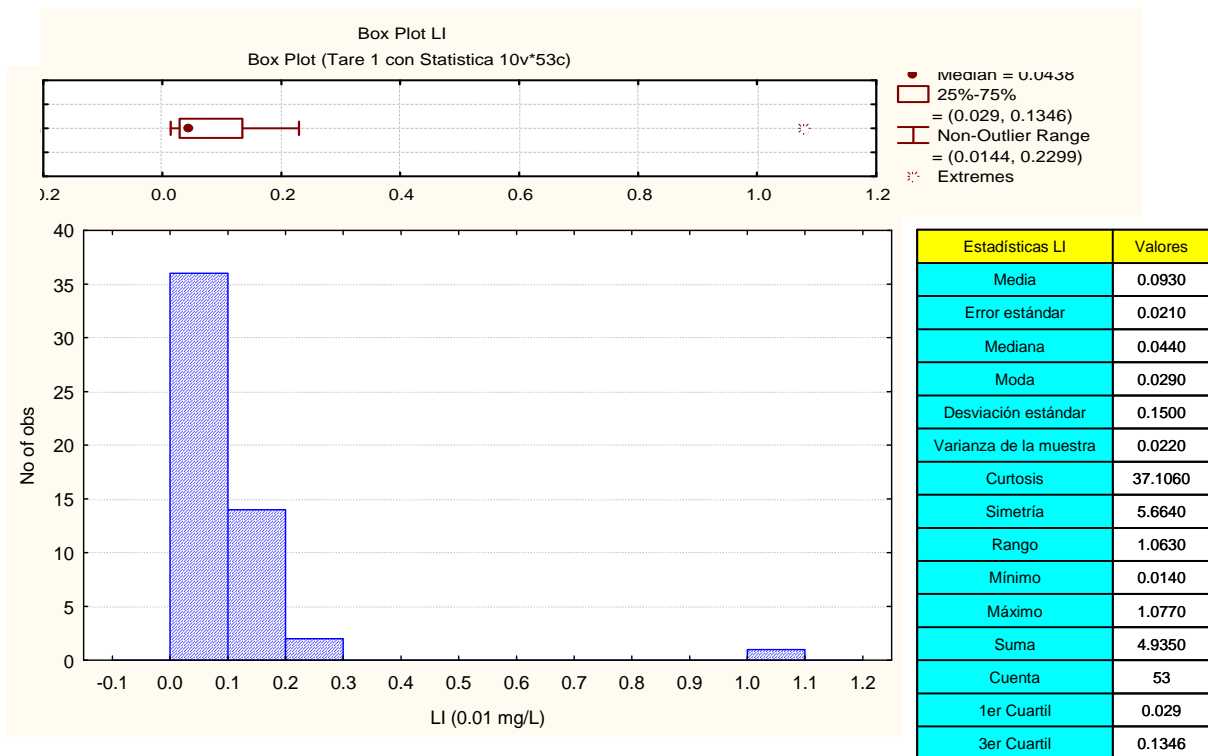


Figura 4. Histograma y Box Plot de Datos LI.

De igual forma, el gráfico Q-Q refleja una distribución normal para la mayoría de los datos, exceptuando el valor atípico antes mencionado (figura 5).

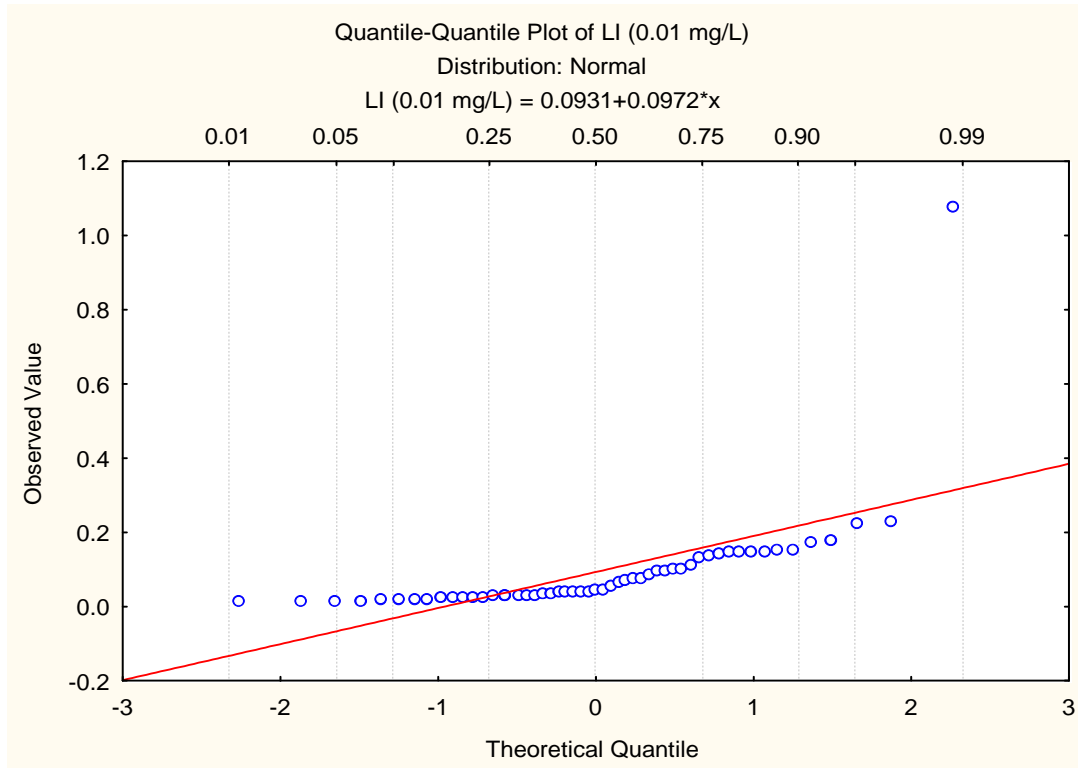


Figura 5. Q-Q plot de datos LI.

Al eliminar el valor atípico y graficar el histograma de los otros 52 datos, se obtiene una distribución de frecuencias bastante más homogénea (figura 5). Sin embargo aun se conserva, en menor proporción, el rasgo característico de simetría positiva, por lo cual la media continúa siendo superior que la mediana. La curtosis en este caso se hace planicúrtica. El diagrama de Box Plot refleja mayor paridad entre el grupo de datos, ampliándose el intervalo intercuartil. En la tabla adjunta a la figura 5 se muestran todos los valores estadísticos de la variable Li excluyendo el valor atípico.

La correspondiente gráfica Q-Q muestra una mayor variabilidad de los residuos en escala lineal (figura 6). Esto se debe precisamente a la mayor dispersión de los datos. Al realizar pruebas de ajuste en la escala se observó que al tomar una escala exponencial en el eje X mejoraba la aproximación lineal del gráfico Q-Q (figura 7).

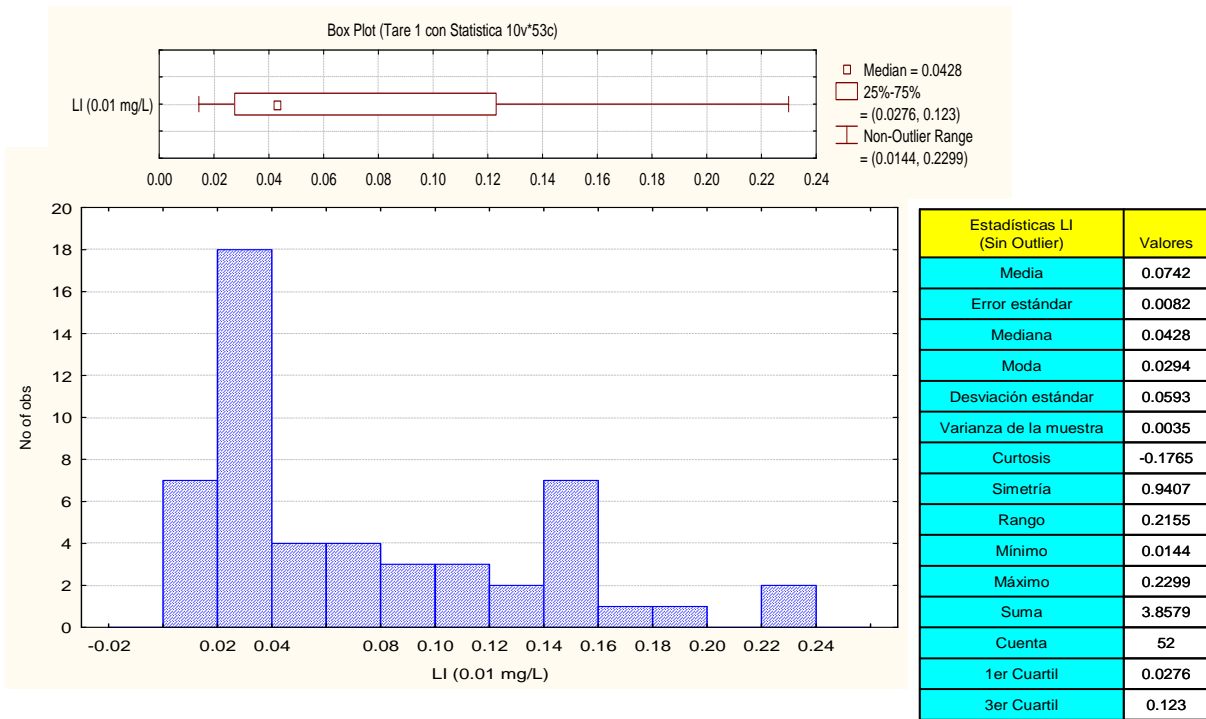


Figura 5. Histograma y Box Plot de Datos LI (sin Outlier)

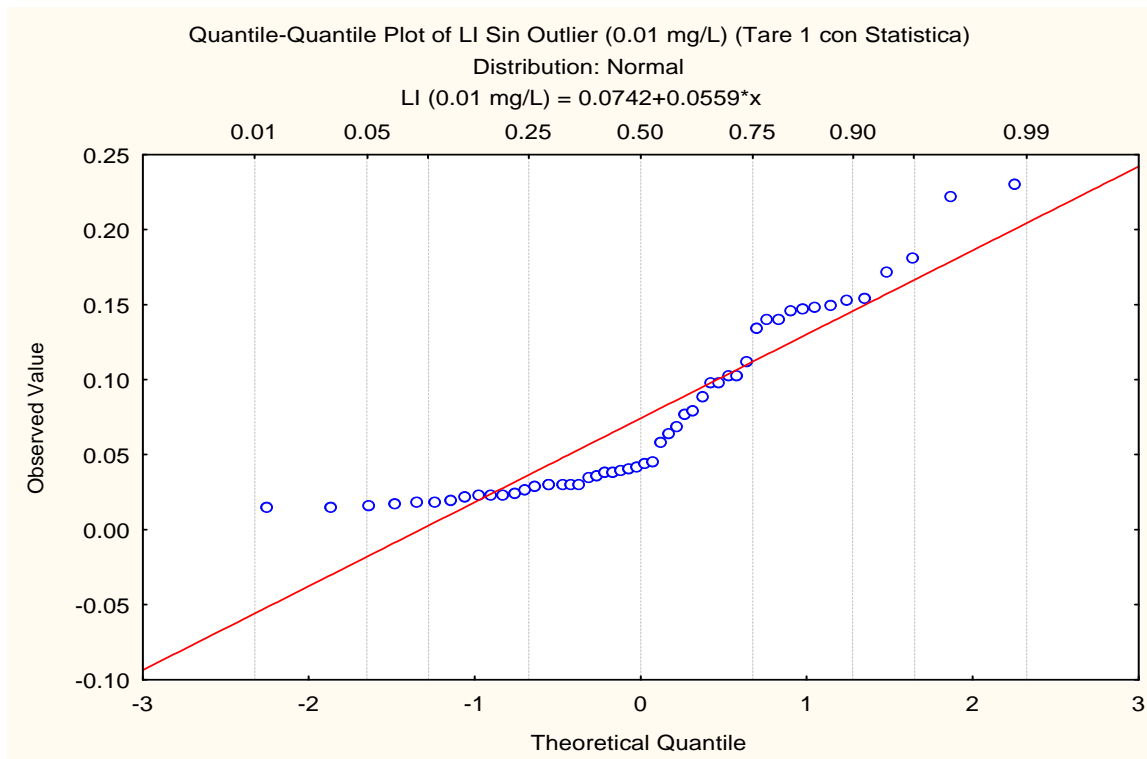


Figura 6. Q-Q Plot de datos LI. Ajuste Lineal.

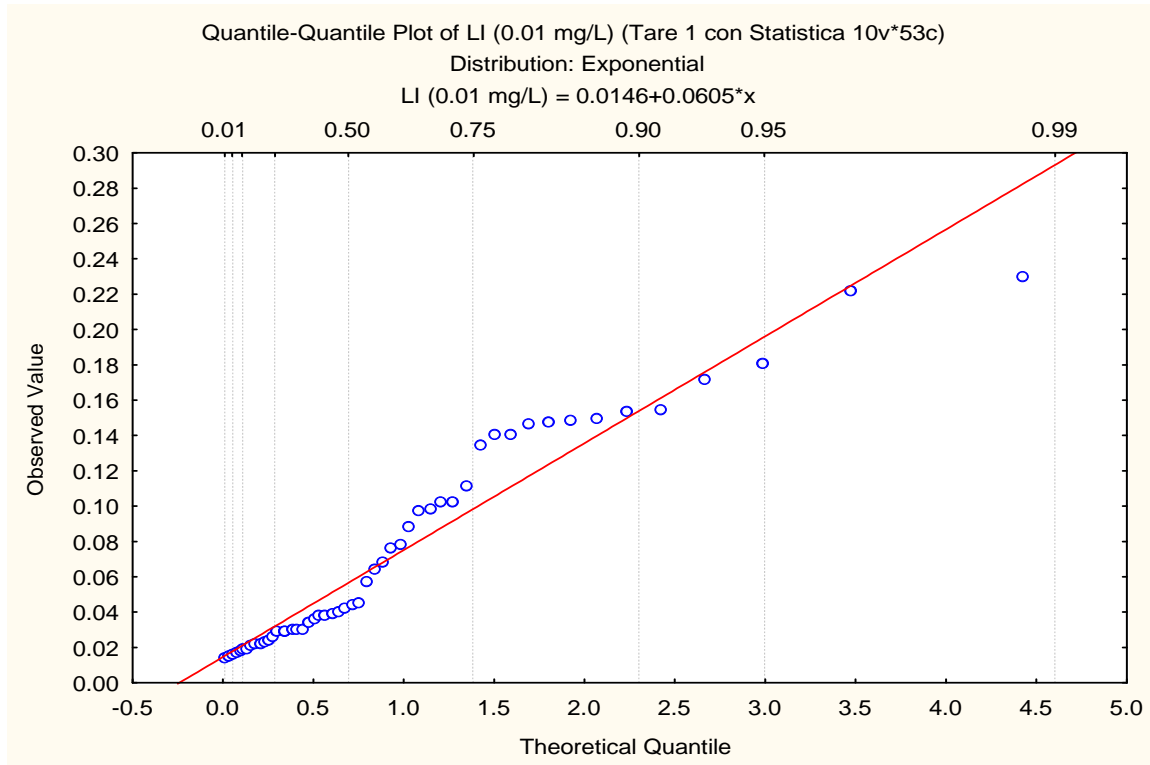


Figura 7. Q-Q Plot de datos LI. Ajuste Exponencial.

DATOS DE FLUORURO

El comportamiento de este grupo de datos resultó similar a los de Litio. Muestra de ello se observa en el histograma de frecuencia para esta variable, el cual tampoco presenta una distribución normal o gaussiana (figura 8). Es de tipo leptocurtica y con asimetría positiva. Al igual que en el caso anterior, se observa un valor atípico, el cual resalta aún más en el diagrama Box Plot, pues está muy alejado del intervalo intercuartil. La mediana es menor que el valor medio y este a su vez es menor que la desviación estándar de la muestra. En la tabla adjunta a la figura 8 se muestran todos los valores estadísticos de la variable F. Por otra parte, el grafico Q-Q refleja una distribución normal para la mayoría de los datos, destacándose el valor atípico al extremo superior derecho del mismo (figura 9).

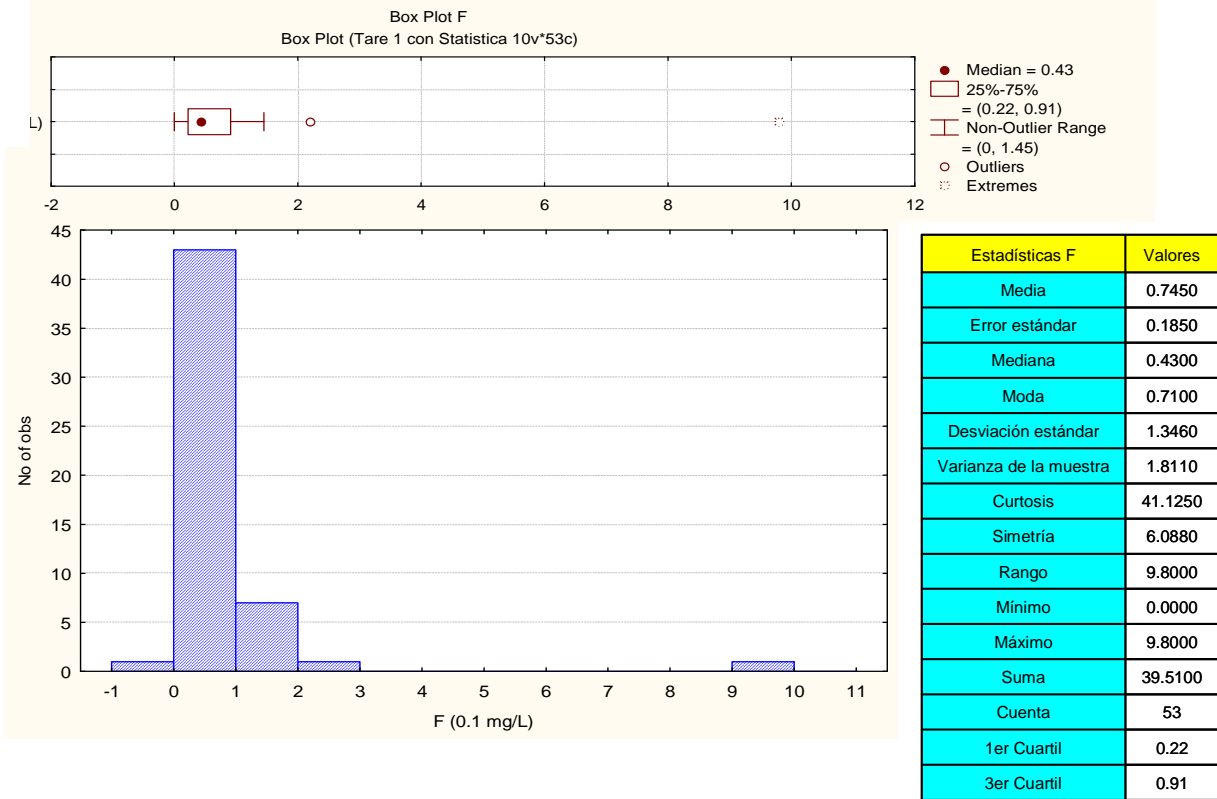


Figura 8. Histograma y Box Plot de datos F.

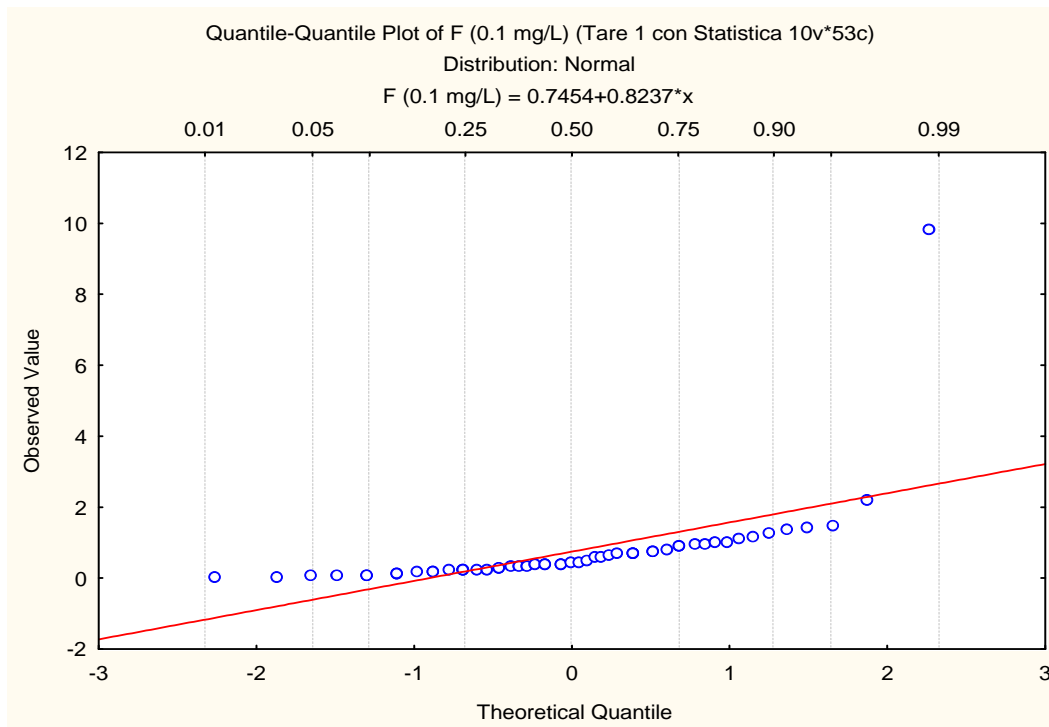


Figura 9. Q-Q Plot de datos F. Ajuste lineal.

Análogo al caso anterior, al eliminar el valor atípico y graficar el histograma de los restantes 52 datos, se obtiene una distribución de frecuencias bastante más homogénea (figura 10). Se conserva la simetría positiva, por lo cual la media continúa siendo superior que la mediana. La curtosis en este caso se hace planicúrtica. El diagrama de Box Plot refleja mayor paridad entre el grupo de datos, ampliándose el intervalo intercuartil. A pesar de que aun aparece un valor atípico, la magnitud de éste es muy inferior a la del discriminado. En la tabla adjunta a la figura 10 se muestran todos los valores estadísticos de la variable F excluyendo el valor atípico principal.

La correspondiente grafica Q-Q muestra una mayor variabilidad de los residuos en escala lineal (figura 11). Esto se debe precisamente a la mayor dispersión de los datos. Al realizar pruebas de ajuste en la escala se observó que al tomar una escala exponencial en el eje X mejoraba la aproximación lineal del grafico Q-Q (figura 12).

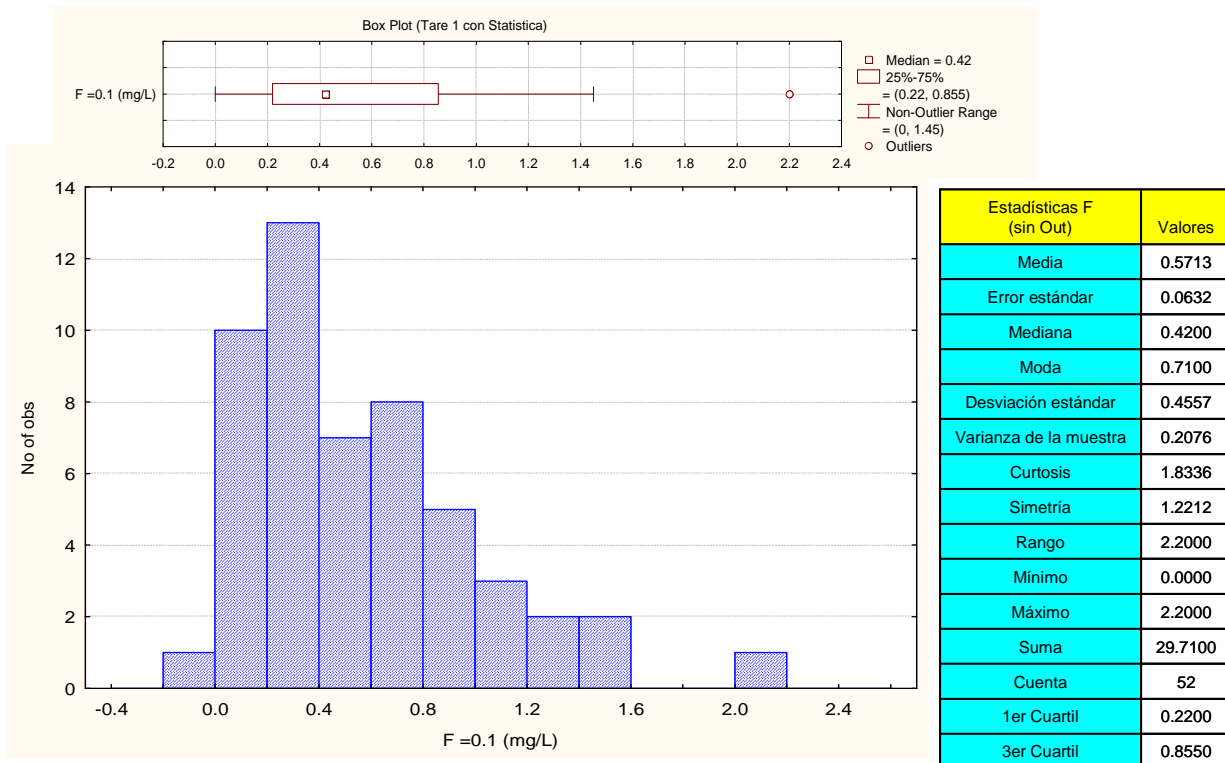


Figura 10. Histograma y Box Plot de datos F (sin Outlier).

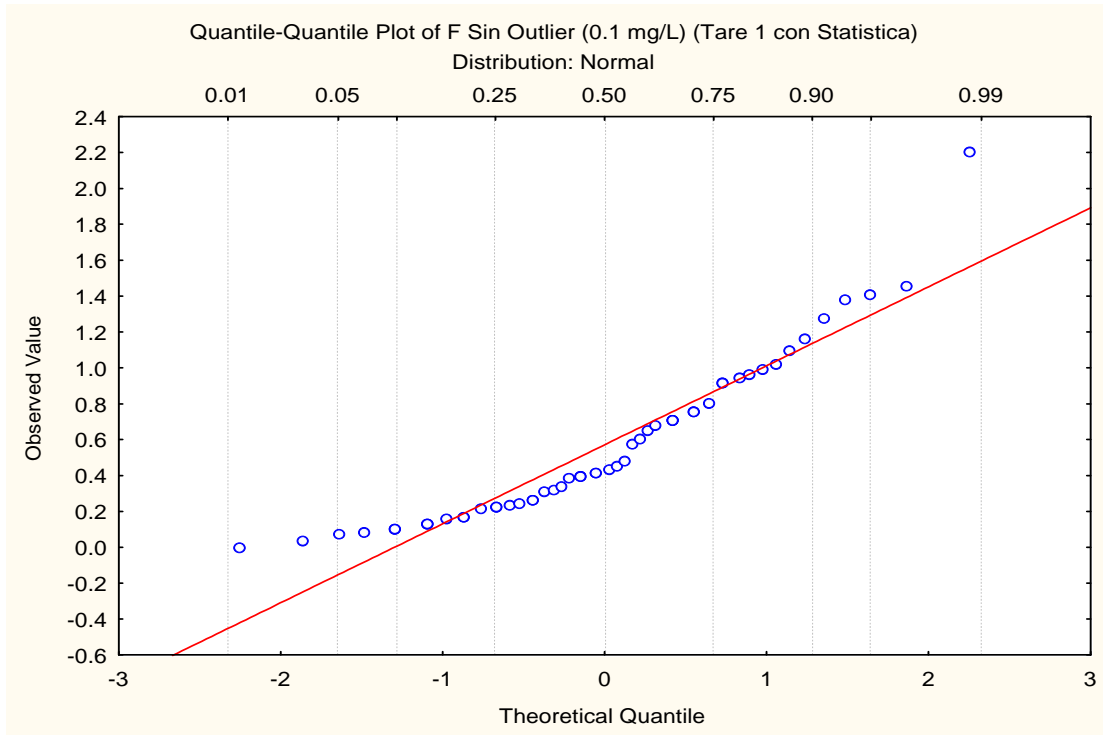


Figura 11. Q-Q Plot de datos F (sin Outlier). Ajuste lineal.

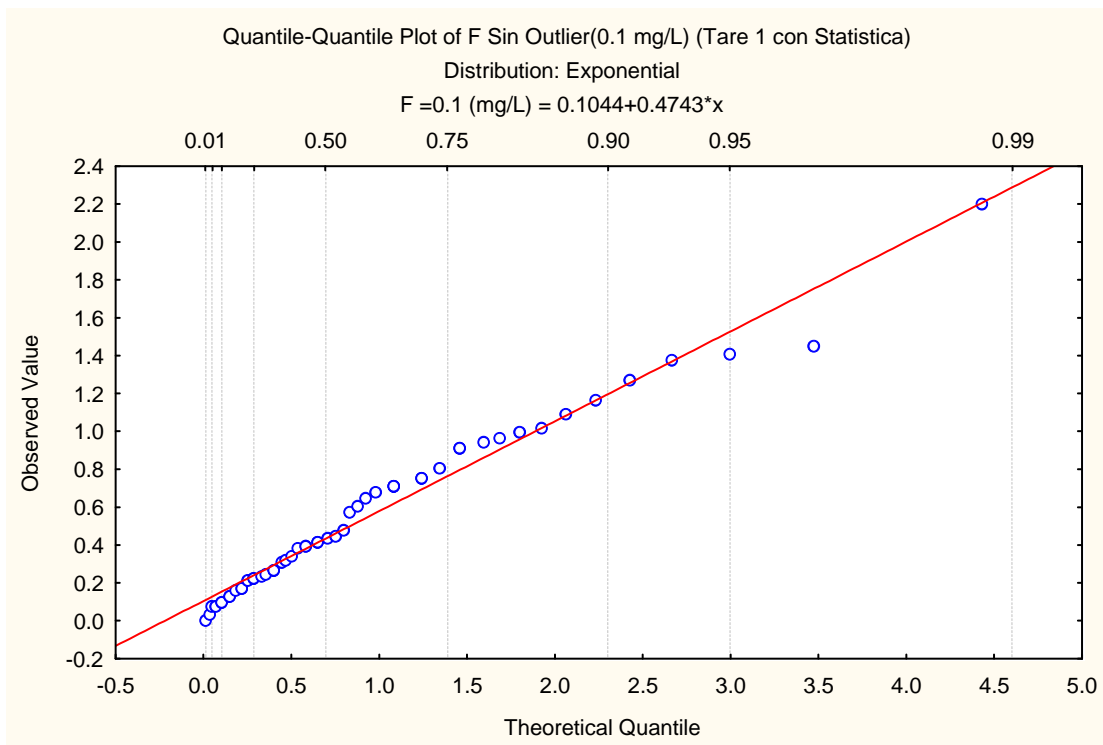


Figura 12. Q-Q Plot de datos F (sin Outlier). Ajuste exponencial.

4. ESTADÍSTICA BIVARIADA

El diagrama de dispersión de los datos de Litio en función de los de Fluoruro muestra una tendencia proporcional (figura 13). De hecho al realizar un ajuste lineal el correspondiente coeficiente de correlación es muy cercano a 1 ($r=0.9684$), lo cual implica una alta linealidad. En la figura 13 se muestra además los histogramas de frecuencia y Box plots ya vistos en el análisis univariado. La figura 14 sólo expresa una modificación de la figura 13, convirtiendo las escalas lineales a logarítmicas para obtener una mejor visualización de la nube de datos concentrada en el rango de valores pequeños.

En la figura 15 se presenta el diagrama de dispersión de los datos sin incluir el par de valores atípicos. En este caso se aprecia una ligera disminución del coeficiente de correlación ($r=0.7740$), debido a la eliminación del efecto de ajuste que producen los valores atípicos. Sin embargo, aun la correlación lineal es considerada como buena.

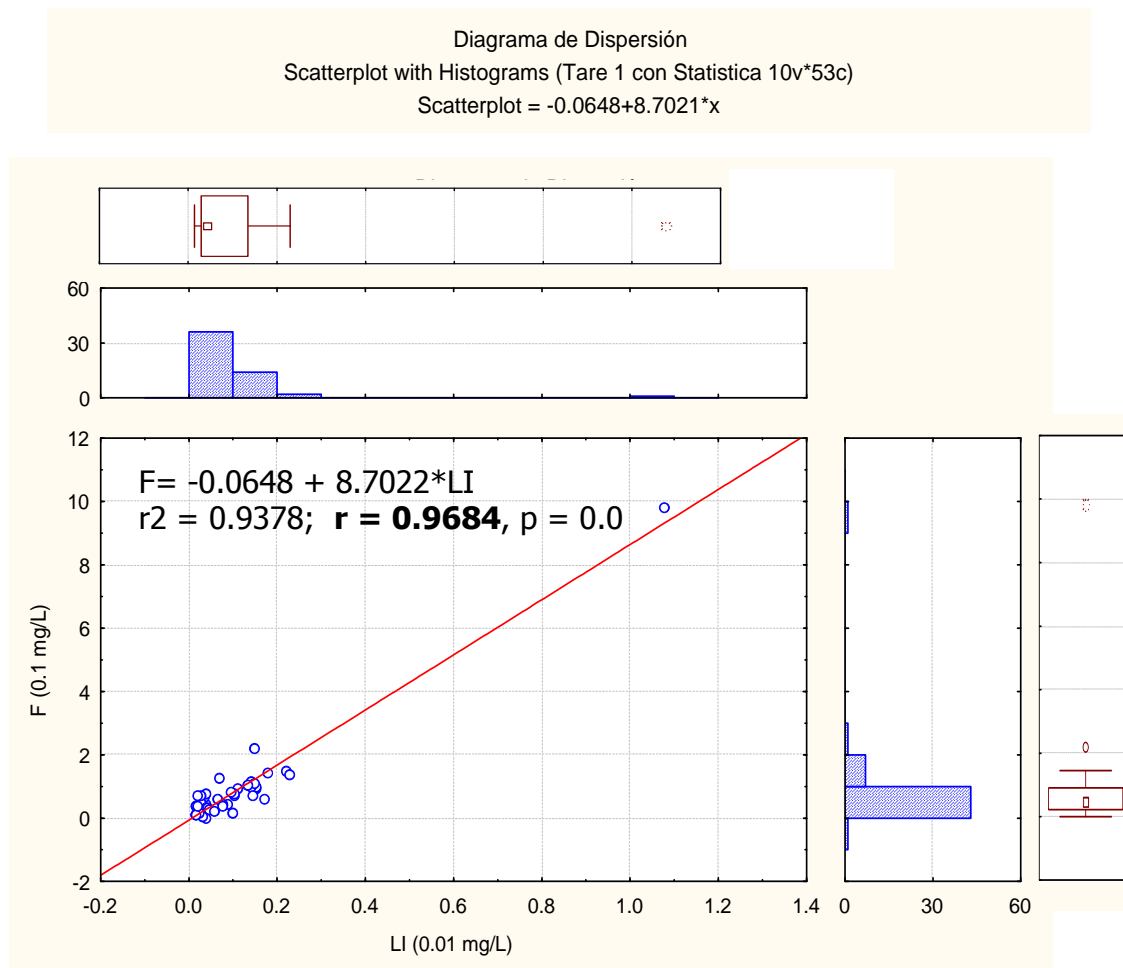


Figura 13. Diagrama de dispersión F vs. LI.

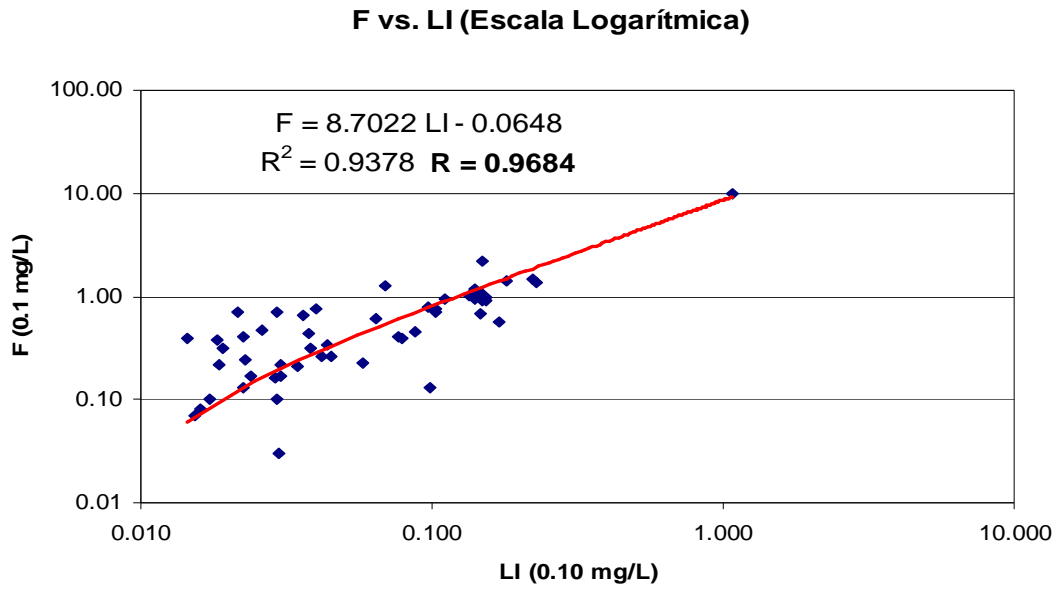


Figura 14. Dispersión F vs. LI. Escalas logarítmicas.

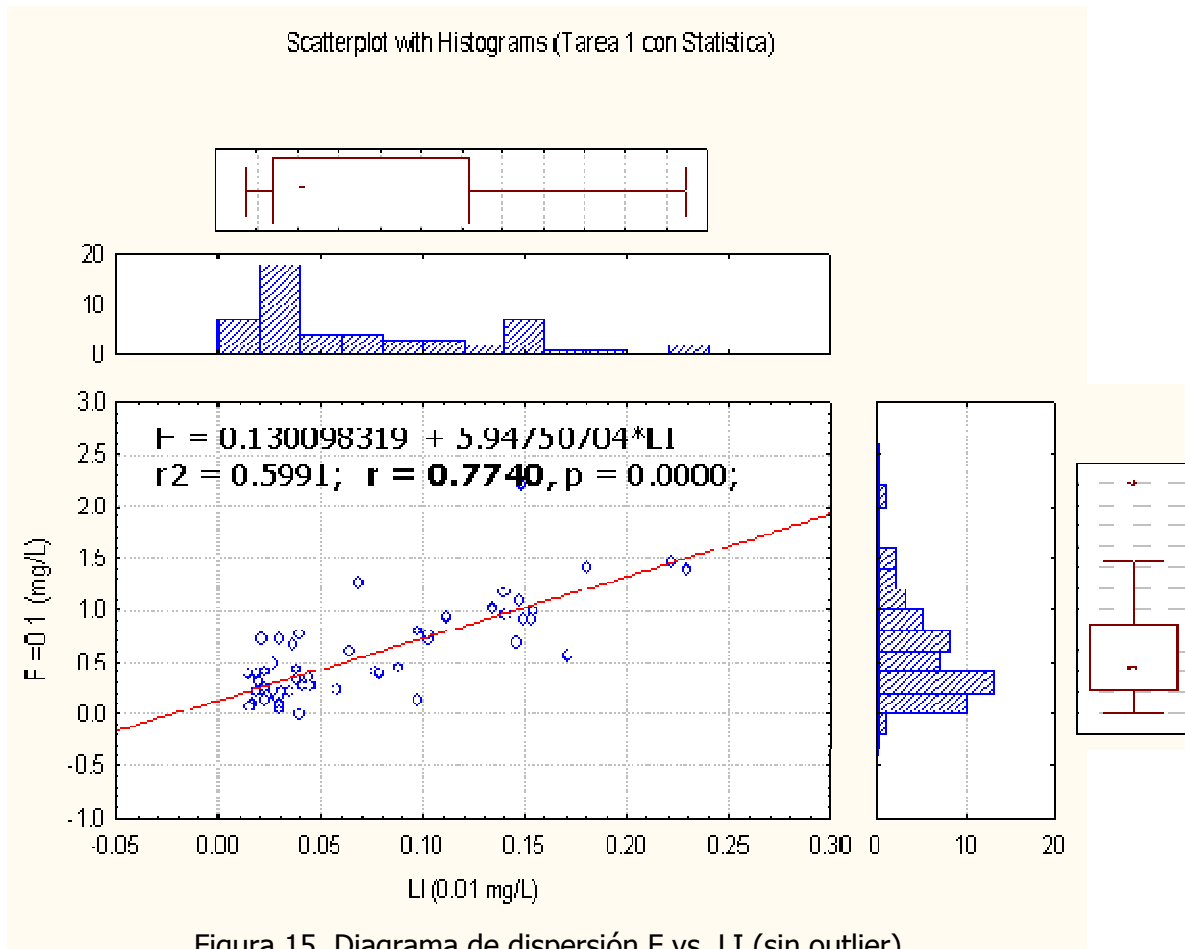


Figura 15. Diagrama de dispersión F vs. LI (sin outlier).

5. CONCLUSIONES

Se realizó el análisis exploratorio de datos de Litio y Fluoruro. Los valores presentados en este estudio corresponden a pozos que incluyen diferentes familias o facies, localizadas en subcuencas hidrológicas adyacentes; las cuales poseen características que las hacen diferenciables. Al respecto, vale destacar que los valores atípicos estudiados corresponden a un manantial de origen volcánico.

La presencia de dicho par de valores atípicos influye de manera importante en el comportamiento de los datos; por lo que los cálculos de los estimadores estadísticos no presentan distribuciones normales; sin embargo, el estudio completo de los datos incluyendo dichos valores, evidencia un comportamiento de proporcionalidad entre los iones F y Li.