

Calidad de datos en el monitoreo de aguas mediante sensores remotos estáticos aplicados en la cuenca del Lago Ypacaraí, Paraguay

Regina León-Ovelar⁽¹⁾, Mariella B. Galeano⁽²⁾, José A. Aguero⁽³⁾, Kevin Cikel⁽⁴⁾, Mario Arzamendia⁽⁵⁾, Derlis Gregor⁽⁶⁾

(1; 3) Laboratorio de Química y Saneamiento, Facultad de Ingeniería, Universidad Nacional de Asunción, rleon@ing.una.py, jose_ag93@hotmail.com

(2) Departamento de Medio Ambiente, Facultad de Ingeniería, Universidad Nacional de Asunción, mgaleano@ing.una.py

(4; 5; 6) Laboratorio de Sistemas Distribuidos, Facultad de Ingeniería, Universidad Nacional de Asunción, dgregor@ing.una.py

RESUMEN

El presente trabajo trata del monitoreo de calidad de aguas de forma remota mediante sensores estáticos en la cuenca del Lago Ypacaraí, la cual se encuentra ubicada entre los departamentos de Central, Cordillera y Paraguarí del Paraguay. Uno de los problemas asociados a la transmisión de datos de calidad de aguas de manera remota, trata sobre la cantidad de mediciones que el sensor pudiera realizar y la calidad final de los mismos; con el fin de optimizar el registro de datos con sensores móviles programados en un vehículo autónomo de superficie para el estudio de calidad del agua en lagos y lagunas, se ha estudiado el comportamiento de un sensor estático que ha considerado los datos obtenidos en el Río Salado dentro de la campaña de monitoreo de la cuenca del Lago realizado por la entidad hidroeléctrica ITAIPU Binacional. El periodo de estudio considerado ha sido junio de 2016 y los parámetros analizados fueron pH, conductividad y temperatura. El sensor registraba datos de calidad cada 20 minutos diariamente, lo cual proporcionaba 2160 datos de partida, sin embargo, mediante InfoStat V.2018 no se observaron diferencias significativas entre los promedios de las mediciones diarias y el promedio de dos mediciones diarias para los parámetros en estudio. Según la prueba de Shapiro – Wilks, los datos de temperatura y pH se distribuyen con normalidad ($p\text{-value} > 0.05$) no así la variable de conductividad. Los parámetros de temperatura y conductividad muestran correlación positiva, en tanto, los parámetros de conductividad y temperatura muestran correlación con el pH según el análisis de correlaciones de Pearson.

Palabras Clave: calidad de aguas, sensores remotos, promedio de datos, Lago Ypacaraí

ABSTRACT

The present work deals with water quality monitoring remotely using static sensors in the Ypacaraí lake basin, which is located between the departments of Central, Cordillera and Paraguarí of Paraguay. One of the problems associated with the transmission of water quality data remotely, is about the number of measurements that

the static sensor could make and the final quality of them. In order to optimize data recording with programmed mobile sensors in an autonomous surface vehicle for the study of water quality in lakes and lagoons, the behavior of a static sensor that has considered the data obtained in the Salado River has been studied. The data is the result of the monitoring campaign of the Ypacaraí lake basin carried out by the hydroelectric entity ITAIPU Binacional. The study period considered was June 2016 and the parameters analyzed were pH, conductivity and temperature. The sensor recorded quality data every 20 minutes daily, which provided 2160 starting data, however, through InfoStat V.2018 no significant differences were observed between the averages of the daily measurements and the average of two daily measurements for the parameters in study. According to the Shapiro - Wilks test, the temperature and pH data are distributed normally ($p\text{-value} > 0.05$), but the conductivity variable it was not distributed normally. The temperature and conductivity show positive correlation, while the conductivity and temperature parameters show correlation with the pH according to the Pearson correlation coefficient.

Keywords: water quality, remote sensing, average data, Yparacai lake

1. INTRODUCCIÓN

La vigilancia hidroquímica es de vital importancia por una cuestión de cuidado medio ambiental pero también de salud pública, si a esto se agrega el factor económico vinculado al crecimiento urbano o el turismo, se encuentra que la problemática actual con la contaminación de cauces hídricos, debería ser uno de los ejes principales de políticas públicas en cualquier gobierno. Chung, W. Y., y Yoo, J. H. (2015) mencionan que el aumento de la contaminación de cauces hídricos plantea una urgencia en cuanto a investigación e implementación de políticas de conservación, y en particular, el crecimiento de la contaminación hídrica evidencia la importancia del monitoreo ambiental. Se presume que el monitoreo continuo en aguas es la mejor técnica para la evaluación de promedios, percentiles, fluctuaciones y tendencias, entre otros, sin embargo, este tipo de mediciones también presenta errores en el muestreo y registro de datos (Bulucea, C. A., et. al., 2009). Los avances en tecnología nos llevan a la búsqueda de mejores metodologías para vigilancia medio ambiental, y se han logrado grandes progresos en cuanto a economía de recursos y técnicas de ensayos en agua, sin embargo, el monitoreo de aguas in situ, automatizado con equipos de sensores y la transmisión de datos a distancia, sigue siendo un área de debates por la complejidad que conlleva utilizar un dispositivo que aunque sea programado y calibrado, realiza mediciones en un ambiente abierto, sometido a los cambios climáticos y con las propias características de cada ecosistema, en general, el monitoreo ambiental es muy complejo y con desafíos a contemplar, como volumen de datos, la velocidad, variedad de formatos según los sensores, veracidad y utilidad real de los datos obtenidos (Gupta, S., et. al., 2018). La medición continua y remota es el futuro del monitoreo ambiental aunque cada sonda debe contar con un sistema de autolimpieza con el fin de garantizar la fiabilidad de las mediciones realizadas. Por ejemplo, la vigilancia constante de clorofila, revela patrones diurnos (Absalon, D., et. al., 2014), así, este tipo de mediciones podrían convertirse en una potente herramienta en el área de investigación. La contaminación en cuerpos de agua, puede monitorearse utilizando una red de sensores inalámbricos (WSN), consistente en sensores inalámbricos distribuidos

especialmente para transmitir los datos monitoreados a la ubicación principal a través de una red (Chung, W. Y. y Yoo, J. H. 2015) y un análisis estadístico proporcionaría la base de la evaluación de resultados considerando las incertidumbres inherentes, e implica considerar el desarrollo de procedimientos de recopilación de datos, contemplando las diferentes fuentes de datos. Algunos métodos no siempre se pueden implementar debido al costo, tiempo y recursos involucrados (Gupta, S., et. al., 2018).

La distribución normal de los datos se considera un supuesto básico del Análisis de Varianza (ANAVA) (Echeverri, L. et al. 2015). El ANAVA permite contrastar la hipótesis nula de que las medias de poblaciones son iguales, frente a la hipótesis alternativa de que por lo menos una de las poblaciones difiere de las demás tomando en cuenta su valor esperado (Alea Riera, V. et al. 2005). Para comprobar la distribución normal de los datos se realiza la prueba de Shapiro-Wilk (modificada), donde los valores de $p\text{-value} > 0.05$ indican que no se rechaza la hipótesis nula, es decir, los datos podrían cumplir con el supuesto de distribución normal. Los valores del ANAVA que demuestren que no existen diferencias significativas entre las medias de los parámetros bajo un nivel de confianza del 95% utilizando el estadístico F_0 , por tanto, si F_0 es pequeño ($p\text{-value} > 0.05$) se confirma la validez de la H_0 y viceversa (Gutiérrez, H. y De la Vara, R. 2008). En caso de no cumplir con los supuestos básicos se podrá optar por análisis de varianza no paramétricos como la del estadístico de Kruskal-Wallis (1952) utilizado por el software y; el coeficiente de variación (CV), adimensional, entendido como la razón entre dos estadísticos, refleja las diferentes características de las muestras, donde un CV con valor mayor a 30 % podría ser descartado por la baja precisión obtenida (Vázquez, E.R. y Caballero, A. 2011; Gordón-Mendoza, R. y Camargo-Buitrago, I. 2015). Este coeficiente es una medida de asociación lineal adecuada para variables cuantitativas donde se toman valores entre -1 y 1 (Chorro Gascó, J.I. 2012), y si este alcanza los valores absolutos de 0 y 1 es que existe concordancia o discordancia perfecta (Alea Riera, V. et al. 2005).

2. METODOLOGÍA

Selección de datos y sensores

El dispositivo de monitoreo de aguas se encuentra ubicado en la zona del Río Salado, este río sirve como canal de descarga hacia el Río Paraguay. Los sensores se encuentran en un sistema tipo tablero IP65 alimentado con paneles solares y baterías con entradas para los mismos y la transmisión de datos se realiza por acceso a una Virtual Private Network (VPN) utilizando la red de telefonía móvil 2G (GPRS). Los sensores analógicos instalados son de la marca Campbell Scientific para monitoreo de aguas, programados para medición cada 20 minutos, las 24 hs del día, todo el mes. El periodo de toma de datos ha sido de un mes, junio de 2016 y se ha realizado considerando la disponibilidad de los mismos y la homogeneidad de datos ya que en otros meses existían saltos en las mediciones.

Tabla 1. Características técnicas de los sensores Marca Campbell utilizados.

<i>Sensor/ parámetro</i>	<i>Modelo</i>	<i>Características</i>
Ph	CSIM11-L	Rango: 0 to 14 Exactitud: $\pm 0.1\%$

Conductividad	CS547A-L	Rango: 0.005 to 7.0 mS/cm Exactitud: $\pm 5\%$
Temperatura	109-L	Rango: -50° to $+70^{\circ}\text{C}$ Exactitud: $\pm 0.2^{\circ}\text{C}$ Tolerancia: 0° to $+70^{\circ}\text{C}$

Estudio estadístico de datos

En total se obtuvieron 2160 datos de partida registrados por el sensor diariamente, cada 20 minutos durante 30 días. Estos datos fueron promediados considerando las mediciones diarias del sensor ($n_1=72$) y dos mediciones principales del turno mañana y noche ($n_2=2$) (ver Tabla 2). Posteriormente, los promedios obtenidos de las variables estudiadas de temperatura ($^{\circ}\text{C}$), conductividad ($\text{mS}\cdot\text{cm}^{-1}$) y pH, fueron procesados a través del programa estadístico InfoStat V.2018 (Universidad Nacional de Córdoba, 2018).

Tabla 2. Codificación y descripción de variables considerando los diferentes promedios diarios realizados.

<i>Variable</i>	<i>Denominación</i>	<i>Descripción</i>
Temperatura	T1	Temperatura promedio diaria ($n_1=72$)
	T2	Temperatura promedio diaria ($n_2=2$)
Conductividad	C1	Conductividad promedio diaria ($n_1=72$)
	C2	Conductividad promedio diaria ($n_2=2$)
pH	P1	Potencial hidrógeno promedio diario ($n_1=72$)
	P2	Potencial hidrógeno promedio diario ($n_2=2$)

Fuente: elaboración propia

Inicialmente, se procedió a la prueba de Shapiro-Wilks (modificado), a fin de observar la distribución normal de los datos, un supuesto básico para la prueba de ANAVA (Echeverri, L.I. et al. 2015). La prueba se realiza con un estadístico modificado por Manibbur y Govindarajulu (1997) utilizado por el software, bajo las hipótesis estadísticas de:

H₀: los datos proceden de una distribución normal ($F(x)$ es normal)

H₁: los datos no proceden de una distribución normal ($F(x)$ no es normal)
(Gutiérrez, H. y De la Vara, R., 2008).

Revisada la distribución normal de los datos ($p\text{-value}>0.05$) se realizó el Análisis de Varianza (ANAVA) bajo un nivel de confianza del 95 %, donde se encuentra la información necesaria para calcular el estadístico de prueba F_0 que permite llegar un valor de significancia o $p\text{-value}$. La tabla de análisis de varianza expone la FV, fuente de variación (efecto); SC, suma de cuadrados; GL, grado de libertad; CM, cuadrado medio; F_0 : estadístico de prueba y $p\text{-value}$ o significancia observada (Gutiérrez, H. y De la Vara, R. 2008). Las suma de cuadrados que se consideran son del *Tipo III*, llamadas parciales, ya que reflejan la contribución de cada término del modelo, dado que todos los otros términos están también presentes en el modelo (Balzarini, M. et al., 2008).

Fueron evaluados los valores del coeficiente de variación (CV), tradicionalmente utilizados para deducir la confiabilidad de la experiencia, considerando un valor mayor al 30 %, sujeto a descarte (Gordón-Mendoza, R. y Camargo-Buitrago, I. 2015). Así mismo, se ha realizado un análisis de correlación de Pearson para observar el grado de

relación existente entre las variables donde el coeficiente fluctúa entre -1 y 1 (Restrepo, L. y González, J., 2007), considerando los siguientes criterios:

Tabla 3. Interpretación del coeficiente de correlación de Pearson
Fuente: Munar, E. (2014).

<i>Coeficiente</i>	<i>Interpretación</i>
r=-1	Negativa perfecta
-0.90	Negativa muy fuerte
-0.75	Negativa considerable
-0.50	Negativa media
-0.25	Negativa débil
-0.10	Muy débil
0	Nula
0.10	Muy débil
0.25	Positiva débil
0.50	Positiva media
0.75	Positiva considerable
0.90	Positiva muy fuerte
r = 1	Correlación perfecta

3. RESULTADOS

Para la comprobación de los supuestos de ANAVA, bajo la prueba de normalidad de Shapiro-Wilks (modificado), se observaron valores distribuidos con normalidad tanto para los datos de temperatura y pH, sin embargo para la variable de conductividad bajo los diferentes promedios ($p\text{-value} < 0.05$), se encuentran evidencias para rechazar la distribución normal de esta variable (Figura 1; Tabla 4) y resulta conveniente realizar la prueba no paramétrica de Kruskal-Wallis. La variable de conductividad no posee una distribución que permita su análisis estadístico bajo modelos lineales, por lo que se recomienda aplicar previamente las transformaciones logarítmicas o de potencia, con el objeto de no afectar el nivel de significancia ni la sensibilidad del estadístico F_0 respecto a la hipótesis nula en el ANAVA (Avanza, M. et al. 2002).

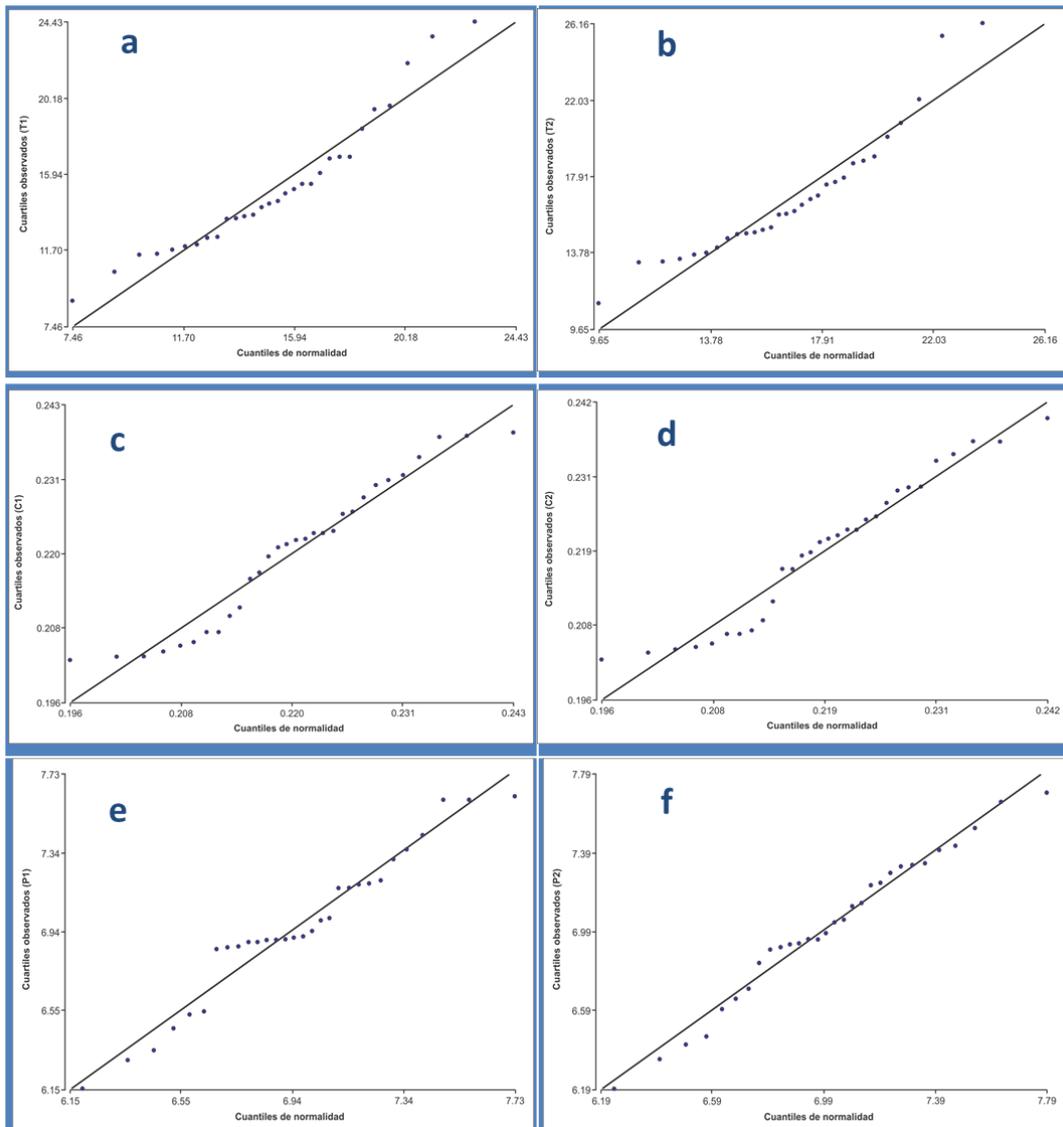


Figura 1. Grado de ajuste de los diferentes parámetros de temperatura, conductividad y pH con una distribución normal bajo los diferentes promedios. Fuente: elaboración propia

En la figura 1, se muestran los cuantiles observados de los parámetros versus los cuantiles de normalidad hallados. Se observa una distribución normal de los datos de los parámetros de pH y temperatura (a,b,e,f), sin embargo, el grado de ajuste de los datos de conductividad reflejan un distribución alejada de la normal (c,d). En el caso de la variable temperatura para el promedio diario de dos mediciones (diurna y nocturna) (T2) se observa un $p\text{-value} < 0.05$, por lo tanto los datos podrían no distribuirse bajo el supuesto de normalidad, en contraparte a lo observado en los valores promedios de temperatura con tomas cada 20 minutos (T1) (Tabla 4).

Tabla 4. Prueba Shapiro-Wilks (modificada) para los parámetros de temperatura, conductividad y pH

<i>Variable</i>	<i>Media</i>	<i>D.E. (3)</i>	<i>W</i>	<i>p</i> <i>(Unilateral D)</i>
T1 (1)	15,16	3,78	0,92	0,0880
C1(1)	0,22	0,01	0,90	0,0178
P1 (1)	6,96	0,38	0,93	0,2031
T2 (2)	16,76	3,48	0,91	0,0350
C2 (2)	0,22	0,01	0,91	0,0376
P2 (2)	7,01	0,38	0,96	0,5800
(1) Temperatura, Conductividad y pH, promedios diarios de tomas de 20 minutos				
(2) Temperatura, Conductividad y pH promedios diarios de dos tomas diarias de datos				
(3) D.E: Desviación estándar				

Bajo el supuesto de normalidad y homocedasticidad de las varianzas, se realiza el ANAVA de las variables de pH y temperatura bajo de los diferentes promedios (ver Tabla 5 y 6), donde no se han observado diferencias significativas entre los datos promedio de ambas variables del estudio. Asimismo, los datos de las medianas de la variable de conductividad no presentaron diferencias significativas bajo el estadístico no paramétrico de Kruskal-Wallis con $H=0.04$ ($p\text{-value}>0.05$) (Tabla 7). Por tanto, no se encuentran diferencias significativas bajo los diferentes promedios diarios de las variables de temperatura, conductividad y pH, lo que podría significar que una frecuencia menor de mediciones ($n=2$) podría resultar conveniente ante una frecuencia de mediciones mayor ($n=72$), dependiendo de los resultados esperados. Los coeficientes de variación podrían ser aceptados debido que no superan el 30 % (Gordón-Mendoza, R. y Camargo-Buitrago, I., 2015).

Tabla 5. Resultados de ANAVA considerando la variable pH

<i>Factor de variación</i>	<i>S.C.</i>	<i>G.L</i>	<i>Test F</i> <i>(5%)</i>	<i>p-value</i>
Modelo	3,00E-02	1	0,24	0,6256
Promedios diarios	3,00E-02	1	0,24	0,6256
Error	8,32	58		
Total	8,36	59		
Coeficiente de Variación	5,442			

Tabla 6. Resultados de ANAVA considerando la variable temperatura

<i>Factor de variación</i>	<i>S.C.</i>	<i>G.L.</i>	<i>Test F (5%)</i>	<i>p-value</i>
Modelo	38,35	1	2,91	0,0936
Promedios diarios	38,35	1	2,91	0,0936
Error	765,57	58		
Total	803,92	59		
Coeficiente de Variación	22, 76			

Tabla 7. Resultados del ANAVA no paramétrico de Kruskal-Wallis para el parámetro de conductividad

<i>Variable</i>	<i>Medias</i>	<i>D.E.</i>	<i>Medianas</i>	<i>H</i>	<i>p-value</i>
Conductividad (C1)	30	0.22	0.01	0.04	0.8476
Conductividad (C2)	30	0.22	0.01		

Con respecto al coeficiente de Pearson, se han observado valores de correlación estadísticamente significativos ($p\text{-value} < 0,05$). El coeficiente de correlación de Pearson, indicaría una fuerte asociación lineal positiva entre las variables temperatura y conductividad. Sin embargo, se observó una moderada asociación lineal negativa entre la variable de pH y las correspondientes a conductividad y temperatura (Tabla 8).

Tabla 8. Resultados del análisis de correlación del Pearson

<i>Variable 1</i>	<i>Variable 2</i>	<i>Pearson</i>	<i>p-value</i>
Temperatura	Temperatura	1,00	<0,0001
Temperatura	Conductividad	0,60	0,0004
Temperatura	pH	-0,43	0,0170
Conductividad	Temperatura	0,60	0,0004
Conductividad	Conductividad	1,00	<0,0001
Conductividad	pH	-0,52	0,0033
pH	Temperatura	-0,43	0,0170
pH	Conductividad	-0,52	0,0033
pH	pH	1,00	<0,0001

Las mediciones de pH o temperatura no brindan una respuesta instantánea, poseen un tiempo T necesario para obtener aproximadamente el 63 % de la respuesta total (Bulucea, C. A., et. al., 2009). Wan-Young Chung y Jae-Ho Yoo (2015), realizaron la transmisión de datos desde los servidores de campo mediante protocolos de enrutamiento al destino final (estación base conectada al PC terminal), y con un

algoritmo de promedios se han reducido los datos, aumentando el rendimiento de transferencia de datos resultante entre los nodos del sensor y la estación base en el campo, red de sensores inalámbricos. Las medidas de iones se corrigen electrónicamente, en la fuente. Según la aritmética simple, el cálculo del promedio general se realiza solo con variables normales, pero muchas variables ambientales siguen distribución tipo logarítmica, lo que impone una media geométrica. Finalmente, una base de datos se reduce en gran cantidad para la evaluación de resultados (Bulucea, C. A., et. al., 2009).

4. CONCLUSIONES

En este estudio no se ha determinado la presencia o no de alteraciones naturales ya que el objetivo ha sido demostrar la necesidad del tratamiento previo de los datos para la evaluación de la hidroquímica, sin embargo, ante una red de sensores con transmisión de datos, será necesario el control del estado físico de los mismos para descartar interferencias que pudieran afectar el análisis matemático.

Cada vigilancia ambiental, mediante sensores remotos deberá ser diseñada de forma exclusiva ya que aunque se demuestra la no necesidad de mediciones por hora, cada parámetro posee un comportamiento diferente por lo que se debe considerar, la hora del día de particular interés o si se trata de redes de vigilancia con sistemas de alertas ante catástrofes.

Se ha expuesto brevemente la complejidad de la toma de datos, la cantidad de los mismos a considerar y la movilidad de estos para una evaluación con datos representativos para cierta investigación, el análisis estadístico será determinante para la obtención de resultados fiables del trabajo de campo.

5. REFERENCIAS

Absalon, D., Ruman, M., Matysik, M., Koziół, K., & Polkowska, Ż. (2014). Innovative solutions in surface water quality monitoring. *APCBEE procedia*, 10, 26-30.

Alea Riera, V.; Guillén, M; Muñoz, C; Maqueda de Anta, I; Torrelles, E; Viladomiu, N. (2005). *Manual de SPPS*. Barcelona, ES.

Avanza, M; Mazza, S; Martínez, G; Gimenez, L. (2002). *Uso de transformaciones para el cumplimiento de los supuestos de normalidad y homocedasticidad para modelos lineales*. Facultad de Ciencias Agrarias UNNE, Corrientes, AR.

Balzarini, M; Gonzalez, L; Tablada, E; Casanoves, F; Di Rienzo, J; Robledo, C. (2008). *Manual del Usuario*. Grupo InfoStat, FCA, Universidad Nacional de Córdoba. Primera Edición, Editorial Brujas Argentina

Bulucea, C. A., Popescu, M. C., Bulucea, C. A., Manolea, G., & Patrascu, A. (2009, February). Interest and difficulty in continuous analysis of water quality. In *Proceedings of the 4th IASME/WSEAS international conference on Energy & environment* (pp. 220-225). World Scientific and Engineering Academy and Society (WSEAS).

Chorro Gascó, J.I. (2012). Estadística aplicada a la Psicología. Universitat de Valencia, ES.

Echeverri, L.I; Acevedo, E; Hernandez, F. (2015). Comparación de pruebas de normalidad. XXV Simposio Internacional de Estadística 2015. Armenia, CO. pp 1-4

Chung, W. Y., & Yoo, J. H. (2015). Remote water quality monitoring in wide area. Sensors and Actuators B: Chemical, 217, 51-57.

Gordón-Mendoza, R; Camargo-Buitrago, I. (2015). Selección de estadísticos para la estimación de la precisión experimental en ensayos de maíz. Agron. Mesoam. 26(1):55-63.

Gupta, S., Mateu, J., Degbelo, A., & Pebesma, E. (2018). Quality of life, big data and the power of statistics. Statistics & Probability Letters, 136, 101-104.

Gutiérrez, H.; De la Vara, R. Análisis y diseño de experimentos. (2008). McGraw-Hill/Interamerica Editores. México, D.F. 2 ed. pp 65-545

Munar, É. F. (2014). Analizando y modelando la violencia interpersonal en la ciudad de Bogotá durante el año 2011, con apoyo en información geográfica. Revista Criminalidad, 56(3), 103-117

Restrepo, L.F; Gonzalez, J. (2007). De Pearson a Spearman. Revista Colombiana de Ciencias Pecuarias; 20 (2) :183-192

Vásquez, E.R., y A. Caballero. (2011). Inconsistencia del coeficiente de variación para expresar la variabilidad de un experimento en un modelo de análisis de varianza. Cultivos Tropicales 32(3):42-45.

6. Agradecimientos

CONACYT financia este proyecto a través del Programa PROCENCIA con recursos del Fondo para la Excelencia de la Educación y la Investigación (FEEI). Proyecto PINV15- 177