

Aplicación de Técnicas Geoespaciales para el Análisis de Contaminantes Relacionados a la Prevalencia de Desórdenes del Espectro del Autismo

*Eneida Luz Ayala Cartagena
Maestría en Ciencia y Tecnología Geoespacial
Prof. Raúl Matos Flores
Departamento de Agrimensura y Ciencia Geoespacial
Universidad Politécnica de Puerto Rico*

Resumen — *El autismo es un desorden para el que se conocen posibles causas, entre ellas el factor ambiental. En estudios previos se ha encontrado un nexo entre la exposición a diferentes emisiones tóxicas al ambiente y la cercanía a los mismos como detonante de condiciones de deficiencias en el desarrollo humano, incluyendo los desórdenes del espectro del autismo. Los Sistemas de Información Geográfica pueden ofrecer alternativas para representar, analizar y predecir el comportamiento de los fenómenos que se manifiestan espacialmente. Se utilizó la Regresión Exploratoria y el método de Mínimos Cuadrados Ordinarios con el fin de encontrar un modelo aceptable para explicar la influencia de distintos contaminantes sobre la cantidad de niños con autismo por municipio en Puerto Rico. El propósito de esta investigación es conocer si la tecnología geoespacial puede ser efectiva para dar luz en el análisis de este fenómeno como apoyo a investigaciones en otras disciplinas relacionadas a este tipo de estudio.*

Palabras Clave — *Autismo, Contaminantes, Regresión, Sistemas de Información Geográfica.*

INTRODUCCIÓN

En Puerto Rico, la prevalencia de autismo es de 1 de cada 62 niños según un estudio realizado por la Escuela Graduada de Salud Pública de la Universidad de Puerto Rico en el 2011. En los Estados Unidos es de 1 de cada 68 niños según Centro de Control de Enfermedades (CDC) para el 2010.

El autismo es una deficiencia en el desarrollo que puede ocasionar retos importantes en las áreas sociales, en comunicación y el comportamiento. Es más común en niños que en niñas. Entre sus posibles causas se encuentran factores genéticos y

ambientales [1]. En estudios previos se ha encontrado un nexo entre la exposición a diferentes emisiones tóxicas al ambiente y la cercanía a los mismos como detonante de condiciones de deficiencias en el desarrollo humano, incluyendo los desórdenes del espectro del autismo [2].

Una de las posibles causas de autismo es por exposición a contaminantes en el medio ambiente de mujeres en estado de gestación y niños. Casi cada 20 minutos se diagnóstica un caso nuevo. Por tal motivo, se están realizando estudios que de alguna manera puedan ofrecer alternativas en tratamiento que aún se encuentran en investigación y fomentar la prevención.

Según investigaciones de Mount Sinai School of Medicine en una alianza con la Agencia de Protección Ambiental (EPA), se ha podido establecer una relación entre las concentraciones y exposición a ciertos químicos tóxicos en el ambiente como sospechosos de causar autismo y otros problemas en el desarrollo. Existen otros estudios que determinan mayor incidencia [de autismo] en lugares contaminados, en zona de fábricas o donde esté afectado el medio ambiente [3].

En Texas se realizó un estudio de Palmer [4] sobre la proximidad a plantas de energía y facilidades industriales que generan emisiones de mercurio y la prevalencia de autismo. Palmer pudo establecer que por cada 1,000 libras de mercurio liberadas existe un 3.7% de aumento en la tasa de autismo y que por cada 10 millas que se aleja de la fuente de emisión hay una disminución en la tasa de autismo. En esta investigación utilizó Sistemas de Información Geográfica para modelar los datos de autismo por distrito escolar y datos con coordenadas XY sobre las emisiones de mercurio de las facilidades desde el 1998 hasta el 2002.

Por otro lado, Windham [5] estudió la posible asociación entre los trastornos del espectro autista y las exposiciones ambientales con datos obtenidos de las concentraciones estimadas de contaminantes en el aire reconocidos como sospechosos de aumentar el riesgo de autismo. Utilizó datos de los contaminantes de la Agencia de Protección Ambiental (EPA), la selección de un grupo de niños diagnosticados con autismo y un grupo control nacidos en 1994 en la Bahía de San Francisco para conocer el nivel de exposición del lugar de residencia por sector censal. Los resultados sugieren un vínculo más fuerte entre la incidencia de autismo y la exposición a metales pesados.

Roberts [6] estudió la relación entre la contaminación del aire y sus efectos sobre la función neurológica del feto en el útero. Se pudo validar su hipótesis de que la exposición prenatal a los contaminantes en el aire aumenta el riesgo de padecer de algún desorden dentro del espectro del autismo.

Las tecnologías geoespaciales han demostrado ser de gran utilidad para modelar la contaminación ambiental. Los Sistemas de Información Geográfica son capaces de capturar, guardar, analizar estadísticamente datos y desplegar información georreferenciada. Por este medio se puede analizar la distribución de las enfermedades, identificar regiones y tendencias para encontrar relaciones entre condiciones de salud y factores ambientales [7].

El modelo aditivo generalizado fue utilizado en el estudio de Hoffman [8] para analizar como la distribución espacial de los predictores conocidos de los trastornos del espectro autista afectan la variabilidad geográfica de la prevalencia en el centro de Carolina del Norte. Para este estudio se identificaron en los registros de los 8 condados del estado niños que a sus 8 años de edad cumplen con la definición estandarizada para diagnóstico de Trastorno del Espectro Autista desde el 2002 hasta el 2008 y se comparó con los registro vitales de los mismos años de nacimiento que los niños con autismo. Los registros le permitieron tener acceso a la información de residencia de nacimiento, lo que

les permitió geocodificar cada caso para poder aplicar el modelo. Se ajustaron factores de riesgo de autismo que pueden causar confusión espacial como género, año de nacimiento del niño, edad de la madre al momento de dar a luz, raza de la madre, nivel de escolaridad materno y uso de tabaco durante el embarazo con el fin de mostrar una variabilidad geográfica más acertada. Luego de realizar este ajuste, se observó que la prevalencia se atenuó en comparación con los resultados del modelo sin ajuste. Se pudo concluir en el estudio la importancia de tomar en cuenta otros factores que pueden afectar la distribución geográfica tales como el nivel de escolaridad materna en estudios de exposición a tóxicos ambientales.

Los Sistemas de Información Geográfica pueden ofrecer alternativas para representar, analizar y predecir el comportamiento de los fenómenos que se manifiestan espacialmente. Partiendo de esta premisa y estudios realizados que han comprobado que existe un vínculo entre la incidencia y prevalencia de Desórdenes del Espectro del Autismo en niños y la exposición a fuentes de tóxicos que afectan el neurodesarrollo como el Arsénico, Cadmio, Mercurio, Benceno, Tolueno, Xileno y PCBs [5] se justifica el que se realicen estudios para observar su comportamiento espacial.

El propósito de esta investigación es conocer si la tecnología geoespacial puede ser efectiva para dar luz en el análisis de este fenómeno como apoyo en investigaciones en otras disciplinas relacionadas a este tipo de estudio.

METODOLOGÍA

La primera regla de la geografía (Tobler, 1970) establece que todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas entre sí que las lejanas.

Históricamente el análisis espacial ha sido fundamental en el entendimiento de la ocurrencia de diversas enfermedades y su relación con el entorno geográfico. Por ejemplo, en tiempos de la Antigua Grecia, Hipócrates estudió las variaciones

en enfermedades de personas que viven en lugares altos en comparación a las que viven en lugares de menor elevación. En Londres para el 1854, el Dr. John Snow trazó en un mapa la distribución de las muertes por Cólera en el cual encontró una relación entre un pozo de agua y la mayor cantidad de muertes alrededor del mismo. Una vez los habitantes del área dejaron de consumir agua del pozo, mermaron las muertes por esta condición de salud.

Para esta investigación, se utilizó la herramienta de Regresión Exploratoria del paquete de herramientas de Estadísticas Espaciales del programa ArcGIS Desktop 10.2.2. La misma evalúa todas las posibles combinaciones de las variables explicativas en busca de los modelos que mejor explican la variable dependiente en el contexto de los criterios especificados por el usuario [9]. Se estudió el comportamiento de la cantidad de estudiantes con autismo que se encuentran registrados en el Departamento de Educación de Puerto Rico para el año escolar 2009-2010 por municipio con relación a contaminantes en el aire, identificados en investigaciones previas que tienen un nexo entre su exposición y la incidencia de autismo en la población. Se recopiló por medio del software TRI.net de la Agencia de Protección Ambiental (EPA, por sus siglas en inglés) información sobre la cantidad de contaminantes liberados al aire por libra para el año 2010. Se definió como variable dependiente la cantidad de estudiantes con autismo registrados bajo el programa de educación especial y como variables explicativas los contaminantes antimonio, compuestos de antimonio, arsénico, compuestos de arsénico, benceno, cadmio, cromo, compuestos de cromo, etilbenceno, plomo, compuestos de plomo, manganeso, compuestos de manganeso, mercurio, compuestos de mercurio, níquel, compuestos de níquel, estireno, tolueno y xileno.

Los datos por municipio del Departamento de Educación de Puerto Rico fueron descargados en formato Excel Workbook (.xlsx) del web site PR

ED Facts del Departamento de Educación y unidos a la tabla de atributos del nivel de información de Municipios de Puerto Rico en formato ESRI shapefile (.shp). Los datos de EPA fueron descargados por contaminante para las facilidades registradas por municipio de Puerto Rico con coordenadas geográficas. Se levantó cada tabla de contaminante liberado en el software ArcGIS Desktop 10.2.2 y se exportó como un .shp de puntos. Se realizó la transformación del sistema de coordenadas geográficas WGS 84 a State Plane NAD 1983 (rev. 2011) para Puerto Rico e Islas Vírgenes. Esta información se unió espacialmente al nivel de información de Municipios con la cantidad de estudiantes con autismo resultando en uno con las sumatorias de las cantidades liberadas por libra, es decir, la información de los datos de partida puntuales ahora están contenidas en los polígonos que representan cada municipio.

Una vez preparada la información, se utilizó la herramienta de Regresión Exploratoria para escoger la combinación de variables explicativas (independientes) que mejor puedan predecir el comportamiento espacial de la dependiente. Luego de estudiar los resultados se utiliza la mejor combinación de variables para utilizarlas como modelo aceptable para evaluar con Mínimos Cuadrados Ordinarios.

Resultará de esto un mapa que representará la significancia estadística de los valores en una distribución normal lo cual nos permitirá reconocer las áreas geográficas donde existe una relación o influencia más fuerte de las variables explicativas sobre la dependiente.

Para analizar la auto correlación espacial de los residuales se utilizó el índice Global de Moran I.

ANÁLISIS

El reporte que se genera luego de utilizar la herramienta de Regresión Exploratoria tiene cinco secciones que nos ayudan a la selección de modelos que mejor explican la variable dependiente [10].

Mejores Modelos por el Número de Variables Explicativas

El primer conjunto de resúmenes en el informe de salida se agrupa por el número de variables explicativas en los modelos probados. Se especificó 1 para el número mínimo de variables explicativas, y un 20 para el número máximo. De las 20 secciones de resumen, 12 tenían posibles modelos que pudiesen predecir la variable dependiente.

Cada sección de resumen también incluye los valores de diagnóstico para cada modelo enumerado: Criterio de Información de Akaike Corregido (AICc), Jarque-Bera (JB), Breusch-Pagan (BP), factor de inflación de la varianza (VIF), y el índice Global de Moran (MI). Estos resúmenes que dan una idea de lo bien que sus modelos están prediciendo (R2). Si cualquiera de los modelos pasa todos los criterios de diagnóstico

pueden ser modelos a utilizarse para la regresión lineal de Mínimos Cuadrados Ordinarios.

Resumen Global de la Regresión Exploratoria

En esta sección se enumeran las 5 pruebas de diagnóstico y la cantidad de modelos que pasó cada una de las pruebas.

Según se puede observar en la figura 1, ningún modelo pasó la prueba de Jarque-Bera. Cuando esto sucede se está teniendo un problema con el sesgo del modelo. Las fuentes comunes de sesgo son:

- Las relaciones de las variables no son lineales
- Existen valores extremos en los datos [10].

Resumen de Significancia de las Variables

Ayuda a determinar cuáles son las variables explicativas con más fuerza y las más débiles para un modelo.

```

***** Exploratory Regression Global Summary (CANTIDAD) ****
Percentage of Search Criteria Passed
Search Criterion Cutoff Trials # Passed % Passed
Min Adjusted R-Squared > 0.50 8157 1536 18.83
Max Coefficient p-value < 0.05 8157 9 0.11
Max VIF Value < 7.50 8157 4670 57.25
Min Jarque-Bera p-value > 0.10 8157 0 0.00
Min Spatial Autocorrelation p-value > 0.10 39 34 87.18

```

Figura 1
Resumen Global de la Regresión Exploratoria

```

summary of variable significance
Variable % significant % Negative % Positive
SUM_NICK_C 71.88 75.00 25.00
SUM_MERC_C 60.21 3.13 96.87
SUM_ANTIM_ 47.78 92.36 7.64
SUM_MANG_C 47.62 92.44 7.56
SUM_CHROM_ 47.41 92.37 7.63
SUM_MANGAN 40.39 0.00 100.00
SUM_LEAD_C 29.27 22.09 77.91
SUM_ETHYLB 22.29 0.00 100.00
SUM_ANTIMO 18.81 16.58 83.42
SUM_LEAD 18.80 30.13 69.87
SUM_TOLUEN 10.27 62.45 37.55
SUM_BENZEN 0.59 94.13 5.87
SUM_XYLENE 0.42 50.07 49.93
SUM_STYREN 0.05 64.49 35.51
SUM_ARSENI -----
SUM_ARSE_C -----
SUM_CADMIU -----
SUM_CHROMI -----
SUM_MERCUR -----
SUM_NICKEL -----

```

Figura 2
Resumen de Significancia de las Variables

Resumen de Multicolinealidad

Este resumen puede ser utilizado en conjunto con el resumen de significancia de las variables para entender cuáles variables explicativas pueden ser removidas del análisis. En este caso, las variables con un Factor de Inflación de la Varianza (VIF) mayor a 7.50 fueron antimonio (120.51), Plomo (118.36), compuestos de plomo (20.02) y compuestos de mercurio (22.42). Las variables arsénico, compuestos de arsénico, cadmio, cromo, mercurio y níquel se eliminaron por ser su único valor para todos los Municipios cero. A pesar de que los compuestos de mercurio aparecen en la lista de las variables que deben ser consideradas para ser removidas, es una de las que cuenta con un mayor porcentaje de significancia.

Por tal motivo, se eligieron las combinaciones con el valor del Criterio de Información de Akaike Corregido (AICc) menor según aparece en la tabla

de los resultados para analizarlas con la herramienta de Mínimos Cuadrados Ordinarios.

Las mejores combinaciones de variables fueron las siguientes:

- Compuestos de Antimonio, Compuestos de Mercurio y Compuestos de Níquel.
- Compuestos de Cromo, Compuestos de Mercurio y Compuestos de Níquel.
- Compuestos de Manganeseo, Compuestos de Mercurio y Compuestos de Níquel.

Para estas tres combinaciones, el valor del AICc es 700.773459, por lo cual se decidió utilizar las tres con Mínimos Cuadrados Ordinarios. Se puede notar que en las tres combinaciones seleccionadas las variables más consistentes son los Compuestos de Mercurio y los Compuestos de Níquel. Las áreas rojas en el mapa representan la mayor cantidad de estudiantes con autismo de las que predijo el modelo y las azules el número menor [11].

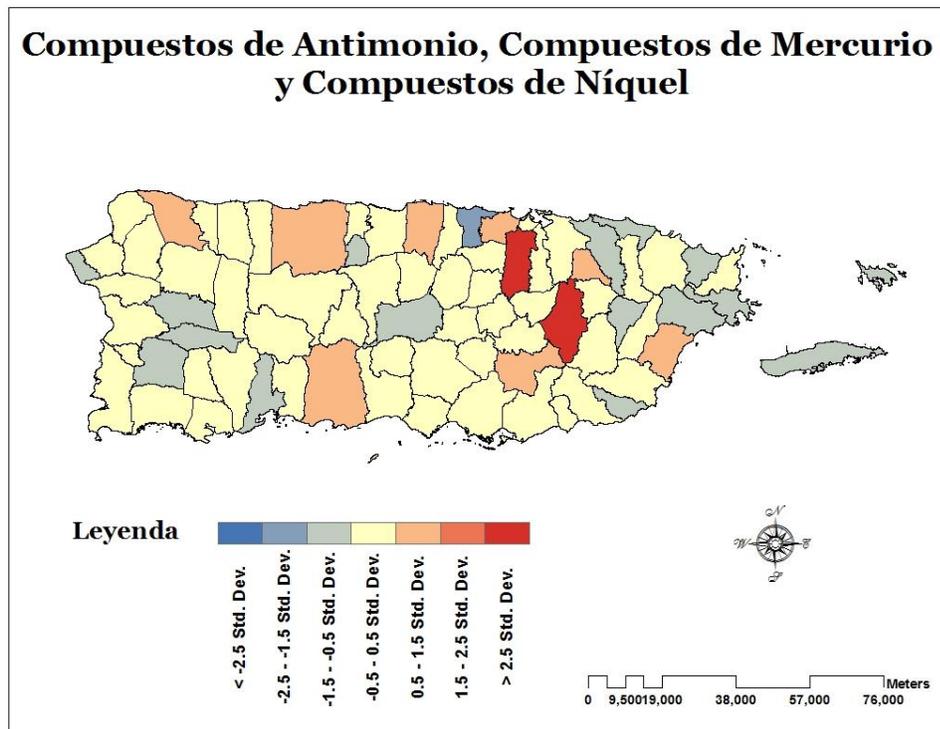


Figura 3
Mapa de Residuales para Compuestos de Antimonio, Compuestos de Mercurio y Compuestos de Níquel

Compuestos de Cromo, Compuestos de Mercurio y Compuestos de Níquel

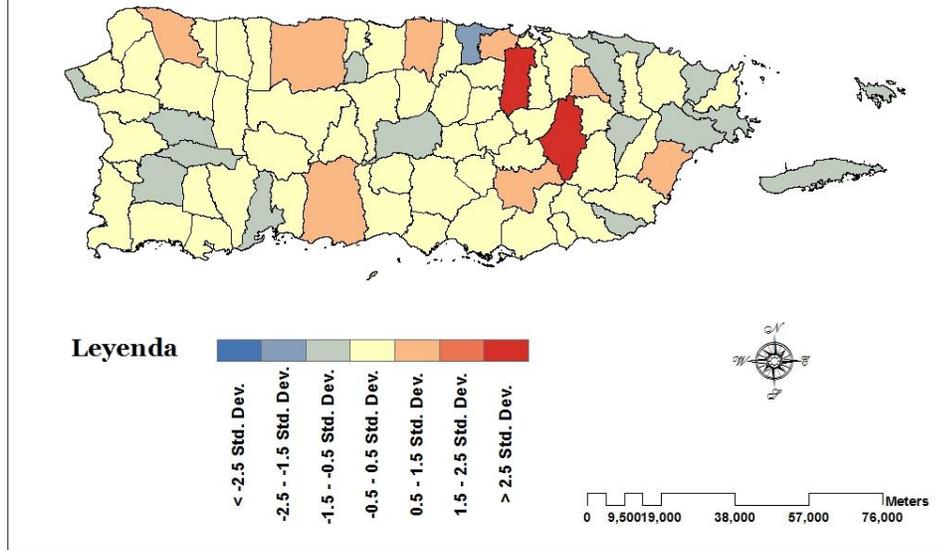


Figura 4

Mapa de Residuales para Compuestos de Cromo, Compuestos de Mercurio y Compuestos de Níquel

Compuestos de Manganeso, Compuestos de Mercurio y Compuestos de Níquel

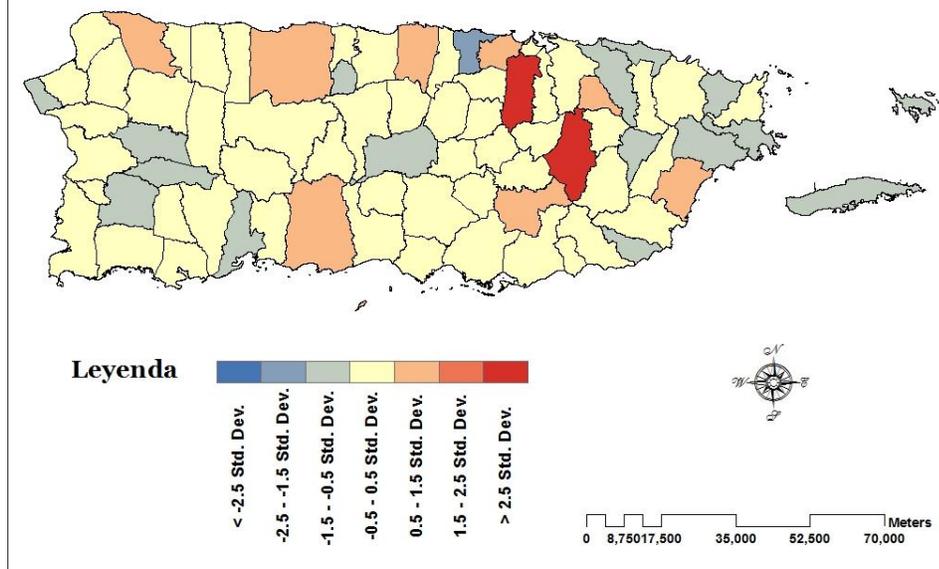


Figura 5

Mapa de Residuales para Compuestos de Manganeso, Compuestos de Mercurio y Compuestos de Níquel

Se realizó un análisis de auto correlación espacial (Global Moran I) para los residuales y verificar cuán agrupados espacialmente se encuentran. Como se puede observar en la Figura 6, la regresión indica que los residuales para los tres modelos están distribuidos aleatoriamente.

Para el análisis de los resultados con la herramienta de Mínimos Cuadrados Ordinarios se utilizaron los reportes que se generaron para cada modelo. Los tres modelos tuvieron valores iguales a

lo largo de los reportes al igual que para el índice global de Moran.

Los coeficientes para las variables de los tres modelos son negativos excepto para los compuestos de mercurio, lo que podría sugerir que a mayor cantidad de compuestos de mercurio liberados al aire, mayor la cantidad de estudiantes con autismo.

No se observó redundancia entre las variables explicativas por ser el valor de inflación de la varianza menor a 7.50 para todas las variables.

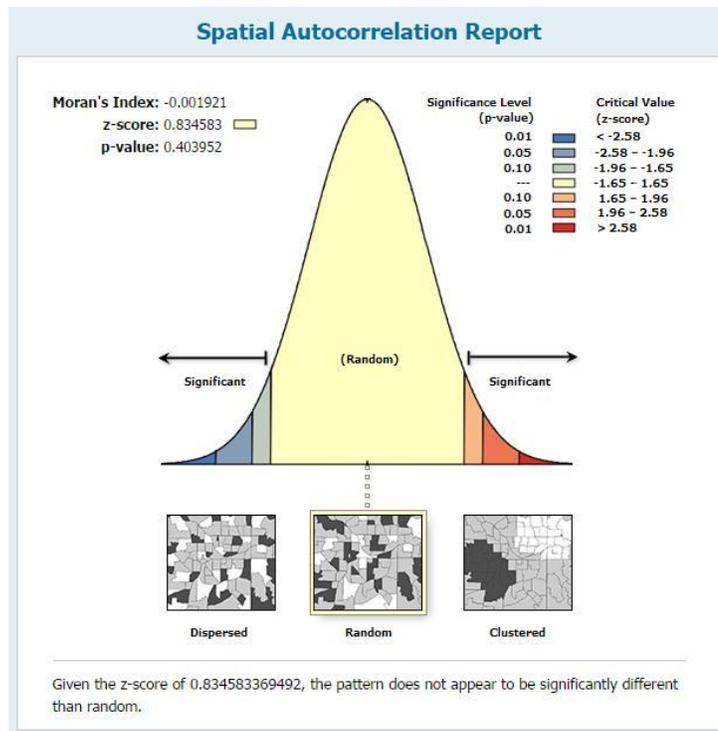


Figura 6
Reporte de Auto Correlación Espacial de los Residuales

Variable	Coefficient	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	16.376343	2.435524	6.723951	0.000000*	2.385298	6.865533	0.000000*	-----
SUM_ANTIM_	-0.002284	0.000327	-6.993076	0.000000*	0.000144	-15.864277	0.000000*	1.990753
SUM_MERC_C	0.001208	0.000122	9.863936	0.000000*	0.000076	15.903142	0.000000*	4.013036
SUM_NICK_C	-0.000000	0.000000	-7.808091	0.000000*	0.000000	-11.449119	0.000000*	3.065960

Input Features:		OLS Diagnostics		Dependent Variable:
Number of Observations:	78	Join_Final1222	78	CANTIDAD
Multiple R-Squared [d]:	0.568653	Akaike's Information Criterion (AICc) [d]:	700.773459	
Joint F-Statistic [e]:	32.518568	Adjusted R-Squared [d]:	0.551166	
Joint Wald Statistic [e]:	254.906698	Prob(>F), (3,74) degrees of freedom:	0.000000*	
Koenker (BP) Statistic [f]:	0.062196	Prob(>chi-squared), (3) degrees of freedom:	0.000000*	
Jarque-Bera Statistic [g]:	1394.772157	Prob(>chi-squared), (3) degrees of freedom:	0.995951	
		Prob(>chi-squared), (2) degrees of freedom:	0.000000*	

Figure 7
Resumen de Resultados de Mínimos Cuadrados Ordinarios

Las variables explicativas tienen valores estadísticamente significantes, lo cual apoya los modelos. Sin embargo, los valores de la prueba de Jarque-Bera (JB) si son estadísticamente significantes. Este resultado afirma que los residuales no están normalmente distribuidos o sesgados. El valor de R al cuadrado ajustado es de 0.551166 lo cual nos indica que sólo la mitad (55%) de la variación en la cantidad de estudiantes con autismo puede ser explicada con los tres modelos.

CONCLUSIÓN

Luego de analizar los resultados podemos concluir que los tres modelos que se eligieron por medio de la herramienta de Regresión Exploratoria están sesgados. El modelo no fue concluyente ya que sólo explica el 55 % de la ocurrencia de los casos. Los contaminantes que tuvieron más significancia como resultado de este estudio fueron los Compuestos de Mercurio y los Compuestos de Níquel. Por tal motivo, esto podría deberse a que es posible que se necesiten otras variables explicativas clave para predecir la cantidad de estudiantes con autismo por Municipio de Puerto Rico. Otro motivo podría ser la necesidad de datos más precisos y más estudios en Puerto Rico sobre el tema.

AGRADECIMIENTOS

A Dios, a mis padres José Ayala y Neida Cartagena, a Eliud Reyes y en especial a mi hijo Kaled por ser mi inspiración y mi fuerza en todo momento.

También le doy gracias a Ileana García, Karen Rivera, Aurora Calderón, Itzel Santiago y Elan Echevarría por su amistad incondicional y apoyo.

REFERENCIAS

- [1] Centers for Disease Control and Prevention, "Facts About ASD", Recuperado el 5 de junio de 2013 de <http://www.cdc.gov/ncbddd/autism/facts.html>.
- [2] Landrigan, P., "A Research Strategy to Discover the Environmental Causes of Autism and Neurodevelopmental Disabilities", *Environmental Health Perspectives*, 2012,

Recuperado el 17 de junio de 2013 de [10.1289/ehp.1104285](http://dx.doi.org/10.1289/ehp.1104285).

- [3] Galenus, "Entrevista a María L. Santaella, MD", *Galenus: Revista para los Médicos de Puerto Rico*, 33(5), 5, 2012, pp. 50-53.
- [4] Palmer, R., F., *et al.*, "Proximity to point sources of environmental mercury release as a predictor of autism prevalence", *Health & Place*, Recuperado el 11 de octubre de 2013 de [10.1016/j.healthplace.2008.02.001](http://dx.doi.org/10.1016/j.healthplace.2008.02.001), 2008.
- [5] Windham, G. C., *et al.*, "Autism Spectrum Disorders in Relation to Distribution of Hazardous Air Pollutants in the San Francisco Bay Area", *Environmental Health Perspectives*, 114(9), 2006, pp. 1438-1444.
- [6] Roberts, A. L., *et al.*, "Perinatal Air Pollutant Exposures and Autism Spectrum Disorder in the Children of Nurses' Health Study II Participants", *Environmental Health Perspectives*, 121(8), 2013, pp. 978-984. Recuperado el 11 de octubre de 2013 de <http://dx.doi.org/10.1289/ehp.1206187>.
- [7] Pantaleoni, E., "Using a GIS approach to better understand risk among individuals with developmental disabilities", Recuperado el 8 de julio de 2013 de <http://kc.vanderbilt.edu/quant/seminar/PantaleoniPresentati onDec10.pdf>.
- [8] Hoffman, J. L., *et al.*, "The spatial distribution of known predictors of autism spectrum disorders impacts geographic variability in prevalence in central North Carolina" *Environmental Health*, 11(80), 2012, pp. 1-10.
- [9] Rosenshein, L., "Exploratory Regression Basic Tool Help", *ArcGIS Desktop 10.2.2 Help*, 2014.
- [10] Rosenshein, L., "Interpreting Exploratory Regression Results", *ArcGIS Desktop 10.2.2 Help*, 2014.
- [11] Childs, C., *et al.*, "Tutorial: Regression Analysis in ArcGIS (ArcGIS10.0)", Recuperado el 12 de agosto de 2013 de <http://www.arcgis.com/home/item.html?id=71a65d35688a4502b123cbdfc99afdee>.