

# Aplicación de gráficos de control para detectar anomalías en la mortalidad de niños en Ecuador

## Application of control charts to detect anomalies in child mortality in Ecuador

Sandra García-Bustos\*

Eva Mera-Intriago\*

Ana Debón\*\*

Iris Bustamante \*

Illych Álvarez\*

\* *Escuela Superior Politécnica del Litoral, Facultad de Ciencias Naturales y Matemáticas, Ecuador*

\*\* *Universitat Politècnica de València, Departamento de Estadística e I.O. Aplicadas y Calidad, España*

### Resumen

En este trabajo se han diseñado gráficos de control con el objetivo de detectar anomalías en los datos de mortalidad infantil del área urbana en Ecuador. A fin de diseñar estos gráficos, se construyeron modelos de regresión binomial negativa para la tasa de mortalidad anual de niños y niñas con edades comprendidas entre uno y menos de cinco años. Con los modelos construidos se elaboraron dos tipos de gráficos de control: el primero, una versión EWMA (*Exponentially Weighted Moving Average*) de los residuos de desviación de los modelos construidos y el segundo, un gráfico que monitorea la media de muertes a través de la estimación obtenida por el modelo de regresión. Para ambos gráficos se ha considerado un nivel de confianza aproximado de 95 por ciento. El gráfico EWMA fue más eficiente en detectar los sub-registros que suelen presentarse en las muertes de niños del área urbana de Ecuador.

*Palabras clave:* Gráfico de control, residuos de desviación, regresión binomial negativa, EWMA

### Abstract

In this work, control charts have been designed in order to detect anomalies in the infant mortality data of the urban area in Ecuador. In order to design these charts, negative binomial regression models were constructed for the annual mortality rate of boys and girls aged between one and less than five years. With the constructed models, two types of control charts were prepared: the first one, an EWMA (*Exponentially Weighted Moving Average*) version of the deviance residues of the constructed models and the second, a chart that monitors the average number of deaths through the estimate obtained by the regression model. For both charts, an approximate 95 per cent confidence level has been considered. The EWMA chart was more efficient in detecting the sub-records that usually occur in the deaths of urban children in Ecuador.

*Keywords:* Control chart, deviance residuals, Binomial Negative Regression, EWMA.

## INTRODUCCIÓN

Las mediciones de mortalidad en un país son vitales a nivel epidemiológico y demográfico, por lo que monitorear la mortalidad es importante para identificar posibles crecimientos excesivos o problemas de subregistros, como suele ocurrir en muchos países en vías de desarrollo.

Aunque los gráficos de control han sido ampliamente utilizados en la industria, recientemente se está prestando gran atención a su utilidad en el área de cuidado de la salud pública (Woodall, 2006; Marshall y Mohammed, 2007; Rolka *et al.*, 2007; Neuburger *et al.*, 2017).

La tasa de mortalidad, calculada como el cociente entre el número de personas que fallecen respecto al total de la población en un período de tiempo, usa una variable de tipo conteo en el numerador. En procesos o fenómenos de estudio donde la variable de interés que se quiere controlar es de conteo, suele usarse el conocido gráfico de control *c* de Shewhart (Montgomery, 2009) basado en la distribución Poisson. Sin embargo, debido a que esa variable de conteo suele estar influenciada por distintos factores o variables explicativas, varios autores han diseñado gráficos de control considerando ajustes de regresión.

Skinner *et al.* (2003) trabajaron suponiendo que la variable de respuesta sigue una distribución Poisson, para luego obtener los residuos de desviación a través de la *ratio* de la verosimilitud generalizada (GLR). El diseño del gráfico de control consistió en usar los residuos de desviación como estadístico de control. A partir de dichos residuos, concluyeron que el rendimiento de un gráfico de control basado en modelos lineales generalizados (GLM), es mejor que el de un gráfico *c* de Shewhart en detectar cambios en la media de la variable de respuesta. Además, Pierce y Schafer (1986) sugirieron en su trabajo el uso de estos residuos, ya que tienen un mejor desempeño y su distribución es asintóticamente normal estándar.

Sin embargo, a menudo en este tipo de procesos el supuesto del modelo Poisson no se cumple, es decir, la media y la varianza no son iguales dado que la varianza puede ser mayor a la media según Agresti (2002), este fenómeno se conoce como sobredispersión. Así pues, Navarro *et al.* (2001) concluyeron que el uso de la binomial negativa es más apropiado que el de la regresión Poisson para esos casos. En este sentido, Yoshioka y Esparza Ochoa (2009) propusieron usar un modelo de distribución binomial negativa para el estudio de la mortalidad infantil en Nicaragua, debido a la calidad de los datos que suele presentarse en sus registros de mortalidad.

Gráficos de control tipo CUSUM (*Cumulative Sum*) y EWMA (*Exponentially Weighted Moving Average*) basados en los residuos de desviación de un modelo de regresión binomial negativa han sido propuestos por Alencar *et al.* (2017) y Urbieta *et al.* (2017) a fin de monitorear el número de hospitalizaciones a causa de enfermedades respiratorias para personas mayores de 65 años en Sao Paulo, Brasil. De manera general el gráfico EWMA presentó mejores resultados.

En concordancia con los resultados anteriores, en este trabajo se muestra el desarrollo de un gráfico EWMA basado en los residuos de desviación de un GLM donde la variable de respuesta sigue una distribución binomial negativa. De acuerdo con Skinner *et al.* (2003) los estadísticos de control basados en los residuos de un modelo bien ajustado generalmente no están correlacionados con el tiempo.

Siendo Ecuador un país en vías de desarrollo, suelen presentarse problemas en los registros de muertes (Vera *et al.*, 2019). Monitorear cada grupo de edad es indispensable, sin embargo, siendo la mortalidad en la niñez un indicador importante y estrechamente relacionado a la pobreza de un país, colocado como una prioridad por la Cumbre Mundial en favor de la Infancia (Naciones Unidas, 2015), en este trabajo se utiliza la metodología de gráficos de control para monitorear el registro de las defunciones de niños de área urbana, ya sea para detectar incrementos excesivos o posibles subregistros.

El presente estudio se ha llevado a cabo tomando los datos proporcionados por el Instituto Nacional de Estadística y Censos (INEC, 2018) de las defunciones de niños del área urbana con edades entre uno y menores de cinco años en Ecuador, por sexo y año en que se produjo el fallecimiento. No se ha considerado en el análisis las defunciones de menores de un año, debido a que en el primer año de vida la probabilidad de muerte es mayor, en especial los primeros 28 días de vida. Según la Organización Mundial de la Salud (2018), en 2016, 46 por ciento de las muertes en menores de cinco años ocurrieron en el primer mes de vida. Adicionalmente, de acuerdo con Becker (1992) y Behm (2011), las principales causas de muerte en menores de un año corresponden a causas congénitas, mientras que, para los niños entre uno y cinco años en su mayoría son causadas por infecciones respiratorias, causas externas y otros procesos infecciosos relacionados con el ambiente físico y social. Por esta razón, dado el comportamiento diferencial de los menores de un año con otros grupos de edad, para el estudio de la mortalidad en menores de cinco años, varios investigadores estudian la mortalidad de menores de un año y de menores con edades

comprendidas entre uno y menos de cinco años por separado (Krzysztof, 2009; Wolleswinkel-van den Bosch *et al.*, 2000).

Este trabajo se organiza de la siguiente manera: En la sección 2, se presenta una breve descripción de la teoría estadística usada en esta investigación. En la sección 3, se consideran los datos de las defunciones de menores entre uno y cinco años en Ecuador, ajustando modelos adecuados para los datos, que en este caso resulta ser binomial negativa. Los residuos de desviación de estos modelos son analizados y utilizados como estadístico para el monitoreo de las defunciones; además, se plantean como otra opción gráficos de control basados en la media de muertes. En la sección 4, se presenta una comparación entre el rendimiento de los gráficos de control usados en la sección 3. Conclusiones y observaciones finales se dan en la sección 5.

### **MODELOS LINEALES GENERALIZADOS Y GRÁFICO EWMA DE RESIDUOS DE DESVIACIÓN**

Es común encontrarse con aplicaciones donde los datos no siguen una distribución normal. En esos casos no se deben utilizar los modelos de regresión clásica basados en el método de mínimos cuadrados ordinarios (OLS), sino los modelos lineales generalizados (GLM) que utilizan la metodología de estimación de máxima verosimilitud (Skinner *et al.* 2003).

Los GLM proveen una conexión entre los modelos lineales y no lineales a través de una función enlace, donde la variable de respuesta puede provenir de una distribución de la familia exponencial. De esta manera, los parámetros del modelo enlazan la respuesta a la distribución de las variables regresoras a través de una función lineal dada por:

$$x_i' \beta = s(\mu_i) \quad (1)$$

#### **Regresión Binomial Negativa**

El modelo de regresión Binomial Negativa es un GLM que combina las distribuciones Gamma y Poisson (Agresti, 2002). Supongamos que  $Y$  tiene una distribución Poisson con media  $\lambda$ , y a su vez  $\lambda$  tiene una distribución Gamma  $G(k, \mu)$ , entonces  $\lambda$  tiene una densidad dada por:

$$f(\lambda; k, \mu) = \frac{(k/\mu)^k}{\Gamma(k)} \exp(-k\lambda/\mu) \lambda^{k-1}, \quad \lambda \geq 0 \quad (2)$$

donde

$$E(\lambda) = \mu, \quad \text{var}(\lambda) = \mu^2 / k$$

A partir de (2), se obtiene la función de probabilidad de  $Y$  como una binomial negativa de la siguiente forma:

$$p(y; k, \mu) = \frac{\Gamma(y + k)}{\Gamma(k)\Gamma(y + 1)} \left(\frac{k}{\mu + k}\right)^k \left(1 - \frac{k}{\mu + k}\right)^y, \quad y = 0, 1, 2, \dots \quad (3)$$

donde

$$E(Y) = \mu, \quad \text{var}(Y) = \mu + \mu^2 / k$$

y  $k^{-1}$  es llamado el parámetro de dispersión.

El modelo de regresión Binomial Negativa se obtiene a partir de la siguiente relación,

$$\begin{aligned} s(\mu) &= \eta = x\beta \\ \log(\mu) &= \eta = x\beta + \varepsilon \end{aligned} \quad (4)$$

### Residuos de desviación para un modelo de regresión Binomial Negativa

La adecuación de estos modelos puede ser analizada a través de los residuos, los cuales indican cuan próximos se encuentran el valor observado y el estimado. Un tipo particular, son los residuos de desviación, que, para el caso de un modelo de regresión binomial negativa, estaría definido como (Urbieta *et al.*, 2017):

$$dr_i = \text{sign}(n_i - \hat{\mu}_i) * \sqrt{d_i^2} \quad (5)$$

$$d_i = \left\{ \begin{array}{ll} 2\phi * \ln(1 + \hat{\mu}_i/\phi) & \text{if } n_i = 0 \\ 2n_i \ln\left(\frac{n_i}{\hat{\mu}_i}\right) - 2\phi \left(1 + \frac{n_i}{\phi}\right) \ln\left(\frac{1 + \frac{n_i}{\phi}}{1 + \frac{\hat{\mu}_i}{\phi}}\right) & \text{if } n_i > 0 \end{array} \right\} \quad (6)$$

donde  $\phi$  es el parámetro de dispersión,  $n_i$  es el valor de conteo observado y  $\hat{\mu}_i$  es el valor de conteo estimado por el modelo de Regresión Binomial Negativa.

### Gráfico EWMA para los residuos de desviación

El gráfico EWMA fue introducido por Roberts (1959), el cual se utiliza si el interés está en detectar cambios pequeños en la media del proceso, ya que es el promedio ponderado de las observaciones pasadas y presentes. Es especialmente útil cuando se tienen observaciones individuales por su insensibilidad al supuesto de normalidad (Montgomery, 2009).

El estadístico de control se define como:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1} \quad (7)$$

donde  $0 < \lambda \leq 1$  y  $z_0 = \mu_0$  (también suele usarse el promedio de los datos preliminares  $z_0 = \bar{x}$ ).

Si se sustituye de manera recursiva  $z_{i-1}$  en (7), se puede observar cómo los pesos decaen geométricamente con el número de muestras.

$$z_i = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j x_{i-j} + (1 - \lambda)^i z_0$$

$$\lambda \sum_{j=0}^{i-1} (1 - \lambda)^j = \lambda \left[ \frac{1 - (1 - \lambda)^i}{1 - (1 - \lambda)} \right] = 1 - (1 - \lambda)^i. \quad (8)$$

Para cada variable independiente  $x_i$  se tiene una varianza  $\sigma^2$ , entonces la varianza de  $z_i$  está dada por:

$$\sigma_{z_i}^2 = \sigma^2 \left( \frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}] \quad (9)$$

De esta manera, los límites de control vienen determinados por:

$$UCL = \mu_0 + L\sigma \sqrt{\left( \frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}]} \quad (10)$$

$$CL = \mu_0$$

$$LCL = \mu_0 - L\sigma \sqrt{\left( \frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}]} \quad (11)$$

Frecuentemente se omite el término  $[1 - (1 - \lambda)^{2i}]$  en (10) y (11), cuando el gráfico se ha ejecutado para varios periodos de tiempo, puesto que tenderá a la unidad conforme  $i$  aumente. Mientras que  $L$  es el factor que determina el ancho de los límites.

Si consideramos que la variable aleatoria a monitorear es el residuo de desviación, entonces el estadístico EWMA quedaría de la siguiente forma:

$$Zdr_i = \lambda dr_i + (1 - \lambda)Zdr_{i-1} \quad (12)$$

Siendo el valor inicial  $Zdr_0 = 0$ , dado que se espera que la media de los residuos de desviación sea igual a cero cuando los datos no se han desviado del modelo ajustado.

Y los límites de control estarán dados por:

$$LCS = L\hat{\sigma}_e \sqrt{\left(\frac{\lambda}{2 - \lambda}\right) [1 - (1 - \lambda)^{2i}]} \quad (13)$$

$$LCI = -L\hat{\sigma}_e \sqrt{\left(\frac{\lambda}{2 - \lambda}\right) [1 - (1 - \lambda)^{2i}]} \quad (14)$$

donde  $\hat{\sigma}_e$  es la desviación estándar estimada de los residuos de desviación. En este caso se supone como valor de  $\hat{\sigma}_e = 1$ , debido a que los residuos de desviación tienen aproximadamente una distribución normal estándar (McCulloch y Searle, 2001; McCullagh y Nelder, 1989).

### Gráfico de control tipo Shewhart para la media de muertes

Los gráficos de control tipo Shewhart consideran estadísticos cuyas fórmulas se basan en la información actual de la muestra y no consideran los valores de las muestras pasadas (García-Bustos *et al.*, 2015). Como parte de este análisis se incluye un gráfico de control de este tipo, cuyo valor a graficar o estadístico estaría dado por el conteo de número de muertes de la población a analizar. Por otra parte, los límites de control superior e inferior estarían determinados por el modelo estimado de regresión binomial negativa que presente un mejor ajuste a los datos. Considerando que la metodología de estimación del modelo es de máxima verosimilitud, y que esta metodología produce estimaciones asintóticamente normales según Agresti (2002) y que, bajo una distribución normal, 95 por ciento de confianza se obtiene a  $\pm 1.96$  desviaciones estándar de la media, los límites de control para este gráfico estarían dados por:

$$LCS = \hat{\mu} + 1.96\hat{\sigma} \qquad LCI = \hat{\mu} - 1.96\hat{\sigma}$$

donde  $\hat{\mu} = Población * e^{x\hat{\beta}}$ ,  $\hat{\sigma} = (\hat{\mu} + \hat{\mu}^2/k)^{1/2}$ , *Población* representa el tamaño de la población del cual se desea estimar el número medio de muertes,  $\beta_{est}$  es el vector de estimadores de máxima verosimilitud de los coeficientes del modelo y  $x$  representa a las variables usadas en el modelo de regresión binomial negativa.

### Metodología para calcular el ARL

El ARL (*Average Run Length*) es la medida de rendimiento más usada en el diseño de gráficos de control y está definido como el número promedio de muestras necesarias hasta que haya un punto fuera de control. El diseño adecuado de gráficos de control es el que tiene un valor de ARL alto cuando el proceso está bajo control ( $ARL_0$ ), mientras que el valor de ARL debe ser lo más pequeño cuando el proceso está fuera de control ( $ARL_1$ ).

Los valores de  $ARL_1$  son usados para comparar rendimientos de gráficos de control. Si las muestras son independientes, el ARL se puede definir como  $ARL = 1/p$ , donde  $p$  es la probabilidad de que un valor del estadístico salga de los límites de control, es decir, que el gráfico presente una señal indicando un descontrol en el proceso (Montgomery, 2009). En estos casos de independencia de las observaciones del estadístico, se usan las probabilidades de los errores tipo 1 y tipo 2 ( $\alpha$  y  $\beta$ ), para calcular los valores de ARL bajo control y fuera de control:  $ARL_0 = 1/\alpha$  y el  $ARL_1 = 1/(1 - \beta)$ , respectivamente. Los valores de  $\alpha$  y  $\beta$  pueden ser calculados considerando la distribución de probabilidades del estadístico usado en el gráfico de control, o también mediante simulaciones.

Sin embargo, debido a que un gráfico EWMA no presenta independencia entre los valores del estadístico del gráfico de control, los valores de ARL deben ser calculados usando procedimientos como cadenas de Markov, ecuaciones integrales o mediante una simulación del proceso (García-Bustos *et al.*, 2015). En este trabajo se usaron simulaciones para obtener los valores de ARL.

### MODELOS AJUSTADOS AL NÚMERO DE MUERTES POR AÑO PARA NIÑOS DE UNO A CINCO años en Ecuador

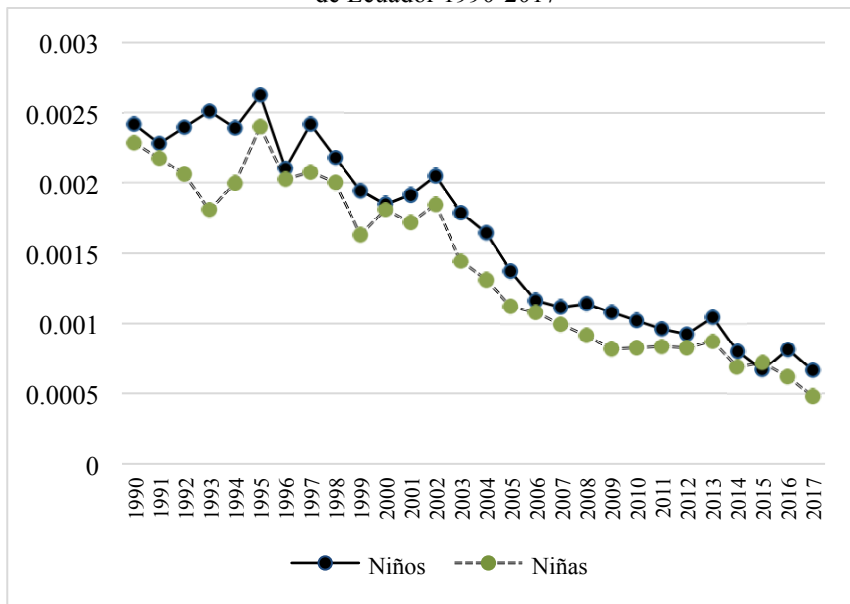
Los datos considerados en este estudio están constituidos por las defunciones anuales registradas en el Ecuador para niños de las áreas urbanas con edades mayores o iguales a uno y menores de cinco años (desde 1990



hasta 2017). Como variable explicativa se considera el tiempo. El total poblacional por año son las estimaciones del total de menores entre uno y cinco años para cada caso realizadas por el INEC (2018) con base al censo 2010.

En la Figura 1 se muestra el cociente de las defunciones y la población en esa edad, separados por sexo. Observándose un descenso en la ratio de muerte a medida que pasan los años.

Figura 1: Tasa de muertes para niños y niñas del área urbana de Ecuador 1990-2017



Fuente: elaboración propia a partir de los datos del INEC (2018).

Las tablas 1 y 2 resumen los resultados obtenidos para los modelos de regresión binomial negativa ajustados a los datos de conteo de muertes para las poblaciones de niños y niñas del área urbana de Ecuador. La estimación del parámetro de sobredispersión  $k$  se obtuvo mediante la función `glm.nb` del paquete MASS (Venables y Ripley, 2002), donde  $R$  (R Core Team, 2008) realiza una estimación previa del parámetro utilizando el método de máxima verosimilitud, para luego incluirlo en el algoritmo.

Tabla 1: Resultados del modelo de Regresión binomial negativa para las muertes de niños urbanos

Coefficientes	Estimación	Error Std	z	Pr (>  z  )
(Intercepto)	97.59	5.5922	17.45	< 2e-16 ***
Año	-0.0519	0.0028	-18.61	< 2e-16 ***
Inversa del parámetro de dispersión (k)	82.01			
Devianza	27.688 con 26 grados de libertad			
AIC	317.39			
Pseudo-R <sup>2</sup>	0.9224			
Valor p de Shapiro test	0.5218			
Valor p de devianza	0.3749			

Fuente: Estimación propia a partir de los datos del INEC (2018).

Tabla 2: Resultados del modelo de Regresión binomial negativa para las muertes de niñas urbanas

Coefficientes	Estimación	Error Std	z value	Pr(> z )
(Intercepto)	102.868	6.129	16.78	< 2e - 16 ***
Año	-0.05467	0.0031	-17.87	< 2e - 16 ***
Inversa del parámetro de dispersión (k)	68.5			
Devianza	27.666 con 26 grados de libertad			
AIC	311.69			
Pseudo-R <sup>2</sup>	0.9157			
Valor p de Shapiro test	0.6052			
Valor p devianza	0.3751			

Fuente: estimación propia a partir de los datos del INEC (2018).

Las estimaciones de los coeficientes del modelo  $\log(\text{Deaths}) = \beta_0 + \beta_i \cdot \text{Year} + \log(\text{Population}) + \varepsilon$  para el caso de niños y niñas se presentan en las columnas “Estimación” de las tablas 1 y 2. Los coeficientes que corresponden a ambos modelos indican que, a medida que pasa el tiempo la media de muertes en los niños y niñas del área urbana se reduce en cinco por ciento con respecto al año anterior.

La prueba de normalidad de “Shapiro-Wilk” da indicio que los residuos de desviación se pueden considerar aproximadamente normales. Además, el valor  $p$  de 0.37 de la devianza da indicio que los modelos propuestos son cercanos al modelo saturado. Los valores de Pseudo- $R^2$  mayores a 91 por ciento ratifican el buen ajuste de los modelos.

Generalmente, para el diseño de gráficos de control en la industria se hace con  $ARL_0$  de más de 200 ( $\alpha < ((1/200) = 0.005)$ ) porque el monitoreo es continuo o se lo hace a intervalo de tiempos cortos. Sin embargo, por la naturaleza de los datos de este trabajo, donde se hace el seguimiento anual, se hará el diseño del gráfico con un  $ARL_0$  de 20, equivalente a una  $\alpha$  de 0.05. Previamente, se probaron valores de  $\alpha$  más pequeños similares a los usados en el diseño de gráficos de control para industria, sin embargo, con esos valores no se detectaron problemas en la mortalidad, lo que no es del todo cierto porque en países como Ecuador existen problemas en los registros de la mortalidad.

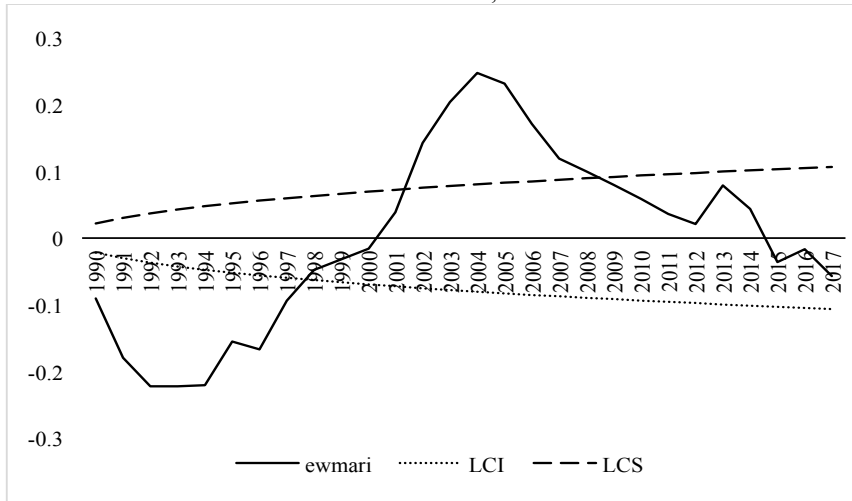
El valor de  $\lambda$  para el estadístico EWMA es importante, Montgomery (2009) recomienda valores de  $0.05 \leq \lambda \leq 0.25$ . Urbietta *et al.* (2017) realizan un análisis comparativo para diversos valores de  $\lambda$  para un gráfico de control EWMA para residuos de desviación de una regresión binomial negativa, siendo los valores de  $\lambda$  más pequeños los que presentan mejor rendimiento. Por tanto, se escogió de valor de  $\lambda = 0.05$ . El valor de  $L = 1.36$  fue obtenido mediante simulación de un gráfico EWMA para una variable normalmente distribuida con media 0 y desviación estándar estimada de 1. Por consiguiente, los límites de control para los gráficos serían:

$$LCS = 1.36 * \sqrt{\left(\frac{0.05}{2-0.05}\right) [1 - (1 - 0.05)^{2i}]}, LCI = -1.36 * \sqrt{\left(\frac{0.05}{2-0.05}\right) [1 - (1 - 0.05)^{2i}]}$$

En las Figuras 2 y 3 se puede apreciar varios puntos fuera de los límites.

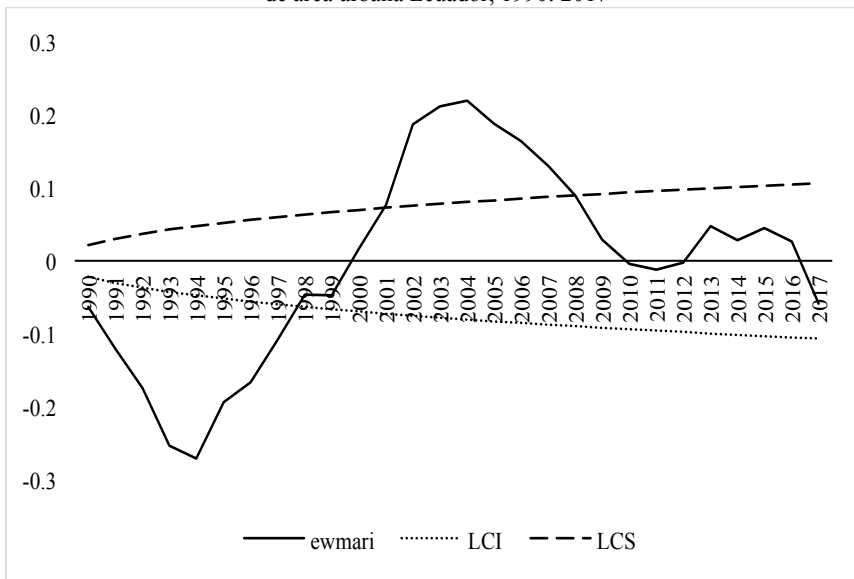
Los puntos fuera del límite inferior indican que en esos años podría haber ocurrido una eventualidad como subregistros en la mortalidad, lo cual es un problema que es probable que exista en Ecuador. Mientras que los valores sobre el límite superior indican que en esos años hubo un crecimiento en la media esperada de muerte, específicamente durante los años de 2002 a 2004. Según Romero-Sandoval *et al.* (2019), esto puede explicarse debido a que en ese período hubo “un aumento en el registro de nacimientos y defunciones, motivado por un mayor monitoreo de la población, así como la estandarización de los registros de nacimientos y muertes infantiles, lo cual tiene un efecto inicial de mayor mortalidad, debido a la visualización de los problemas”.

Figura 2: Gráfico EWMA de residuos de desviación para niños de área urbana Ecuador, 1990: 2017



Fuente: elaboración propia a partir de los datos del INEC (2018).

Figura 3: Gráfico EWMA de residuos de desviación para niñas de área urbana Ecuador, 1990: 2017



Fuente: elaboración propia a partir de los datos del INEC (2018).

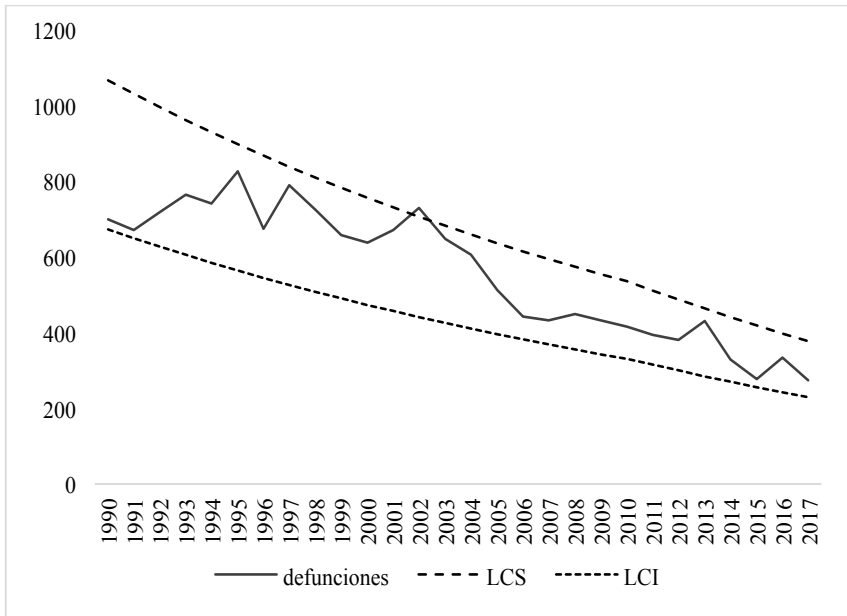
A fin de destacar la ventaja de los gráficos tipo EWMA, se presenta como contraparte, los gráficos tipos Shewhart, los cuales tienen como estadístico el conteo de muertes. A fin de obtener los límites de control superior e inferior se requiere de acuerdo a (15) las estimaciones de las medias y desviación estándar:

$$\hat{\mu} = Población * e^{97.59 - 0.0519Año} \quad y \quad \hat{\sigma} = (\widehat{\mu} + \hat{\mu}^2 / 82.01)^{\frac{1}{2}} \text{ para niños urbanos;}$$

$$\hat{\mu} = Población * e^{102.868 - 0.05467Año} \quad y \quad \hat{\sigma} = (\widehat{\mu} + \hat{\mu}^2 / 68.5)^{\frac{1}{2}} \text{ para niñas}$$

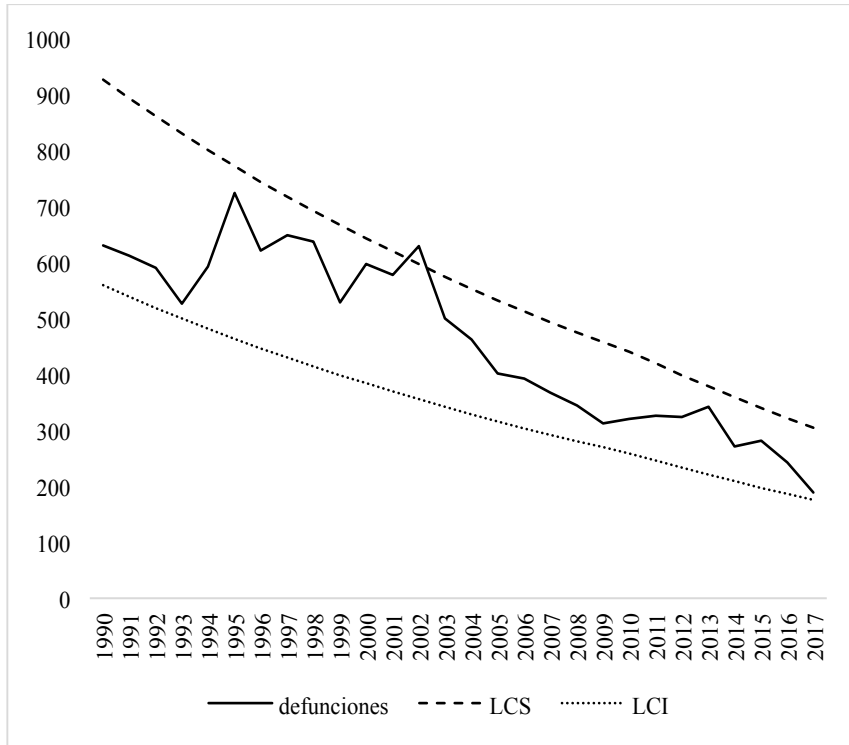
Las figuras 4 y 5 presentan los gráficos de control.

Figura 4: Gráfico para la media estimada de muertes, niños urbanos de Ecuador, 1990: 2017



Fuente: elaboración propia a partir de los datos del INEC (2018).

Figura 5: Gráfico para la media estimada de muertes, niñas urbanas de Ecuador, 1990: 2017



Fuente: elaboración propia a partir de los datos del INEC (2018).

Se puede apreciar que estos gráficos sólo presentan un valor atípico, el cual corresponde al año 2002, y no detecta los problemas de subregistros, muy común en países como Ecuador. Aquí, se destaca la importancia de usar gráficos con efecto memoria como el EWMA a fin de acelerar la detección de problemas en los registros de muerte.

### RENDIMIENTOS DE LOS GRÁFICOS PROPUESTOS

En esta sección se evalúan los rendimientos de los gráficos propuestos a través del ARL. Se hace un análisis considerando cambios aditivos en la media, para comparar las respuestas de los gráficos EWMA y Shewhart para detectar esos cambios. Para este estudio los valores de ARL han sido calculados usando 20 mil simulaciones considerando una población de infantes fija de 100 mil. Cuando se habla de cambios aditivos se hace refe-

rencia a los cambios en la media directamente en unidades de desviaciones estándar.

Las Tablas 3 y 4 presentan el comportamiento del ARL para los gráficos propuestos para controlar la mortalidad en niños y niñas en el área urbana del Ecuador. Estos valores de ARL han sido obtenidos mediante simulaciones de variables binomial negativa con medias  $\mu$  para la zona urbana y varianzas  $\mu + \mu^2 / k$ , considerando la expresión (5) para el cálculo de los residuos de desviación. Para estos residuos, los valores de  $n_i$  son los valores simulados con distribución binomial negativa mientras que los valores de  $n_j$  son obtenidos mediante los modelos presentados en las Tablas 1 y 2.

Tabla 3: Comparación de rendimientos para incrementos de la media

Incrementos en $\mu$ en unidades de desviación estándar	Valores de ARL			
	Niños		Niñas	
	EWMA	Shewhart	EWMA	Shewhart
0	20	20	20	20
0.25	12.88	15.23	13.13	15.58
0.5	6.92	10.55	6.9	10.79
1	3.04	5.12	3.08	5.22
1.5	1.93	2.98	1.96	3.03
2	1.46	2	1.48	2.01

Fuente: estimación propia a partir de los datos del INEC (2018).

Tabla 4: Comparación de rendimientos para decrementos de la media

Decrementos en $\mu$ en unidades de des- viación estándar	Valores de ARL			
	Niños		Niñas	
	EWMA	Shewhart	EWMA	Shewhart
0	20	20	20	20
0.25	10.79	21.3	10.71	20.93
0.5	5.69	16.79	5.68	16.49
1	2.54	7.37	2.54	7.22
1.5	1.57	3.39	1.57	3.33
2	1.2	1.87	1.2	1.86

Fuente: estimación propia a partir de los datos del INEC (2018).

Las Tablas 3 y 4 muestran al gráfico EWMA de los residuos de desviación más eficiente que el gráfico Shewhart en detectar cambios en las medias, puesto que presenta menores valores de ARL, es decir, una detección más rápida de los cambios. Por ejemplo, un incremento de 0.5 desviaciones estándar, si se considera el modelo de niñas urbanas, implicaría un cambio en la media de 267.17 a 285.26 muertes cada 100 mil niños, el cual sería detectado en 6.9 períodos de tiempo para el EWMA, mientras que el gráfico Shewhart, lo detectaría en 10.55 períodos de tiempo. Las mejoras del gráfico EWMA resultan más significativas cuando se producen decrementos en la media de muertes, esto explica el hecho que el gráfico Shewhart en la sección anterior no detectara posibles subregistros en la mortalidad de niños, cuando se conoce que Ecuador adolece de este problema.

## CONCLUSIONES

En este trabajo se presenta el uso de dos gráficos de control: el primero es un EWMA para los residuos de desviación de modelos de Regresión Binomial Negativa para la tasa de mortalidad de niños con edades comprendidas entre uno y menos de cinco años separados por género en el área urbana del Ecuador; y el segundo gráfico de control está basado en las estimaciones de la media de muerte del modelo de Regresión Binomial negativa. Con el fin de obtener un modelo adecuado, no se incluye en el análisis las defunciones de menores a un año, dado el comportamiento diferencial de los menores de un año con otros grupos de edad.

La estimación del modelo de regresión binomial negativa reflejó que a medida que pasa el tiempo la media de muertes en los niños y niñas del área urbana se reduce en cinco por ciento con respecto al año anterior. Resultado positivo para la población urbana infantil de Ecuador en términos del control de enfermedades infecciosas que son las que más afectan a los niños de esta edad.

En el diseño del gráfico se consideró un  $ARL_0$  de 20, similar a un  $\alpha$  de 0.05, y un valor de  $\lambda$  de 0.05 para el gráfico EWMA, así se da una mayor ponderación a la información pasada para agilizar la detección de cambios en la media de muertes.

Con el gráfico EWMA, se pudo detectar ciertos valores atípicos en las poblaciones de niños y niñas del área urbana en década de los noventas y 2002-2004, lo cual podría evidenciar que hay subregistros o crecimiento excesivo en la mortalidad.



De la evaluación de rendimientos de los gráficos propuestos cuando hay cambios aditivos en la media se puede decir que el gráfico tipo EWMA es más eficiente que el gráfico tipo Shewhart. El gráfico Shewhart resulta lento en detectar decrementos en la media, lo cual no lo hace aplicable en países subdesarrollados donde los problemas de subregistros pueden estar presentes.

Aunque el control de la mortalidad se ha hecho en este trabajo de manera anual, los gráficos de control pueden ser aplicados para menores intervalos de tiempo entre registros.

## REFERENCIAS BIBLIOGRÁFICAS

- Agresti, A., 2002, *Categorical Data Analysis*. Second edition ed. Gainesville, Florida, John Wiley and Sons, Inc., p. 131.
- Alencar, A. P., Ho, L. L. and Esparza Albarracin, O. Y., 2017, "CUSUM control charts to monitor series of Negative Binomial count data", en *Statistical Methods in Medical Research*, 26(4), pp. 1925-1935.
- Becker, R., 1992, *Análisis de la mortalidad. Lineamientos básicos. Programa de Análisis de la Situación de Salud y sus Tendencias*, Washington DC: Organización Panamericana de la Salud - Organización Mundial de la Salud.
- Behm, H., 2011, "Social and economic determinants of mortality in Latin America", en *Salud Colectiva (English edition)*, 7(2), pp. 231-253.
- García-Bustos, S., Aparisi, F. and Epprecht, E. K., 2015, "Optimal EWMA of linear combination of Poisson variables for multivariate statistical process control", en *International Journal of Production Research*, 54(14), pp. 4141-4159.
- Instituto Nacional de Estadísticas y Censos, INEC, 2018, *Sistema Nacional de Información*. Available at: [www.sni.gob.ec](http://www.sni.gob.ec). Último acceso: 20/10/2018.
- Krzysztof, T., 2009, "Correlates of infant and childhood mortality: A theoretical overview and new evidence from the analysis of longitudinal data of the Bejsce (Poland) parish register reconstitution study of the 18th-20th centuries", en *Demographic Research*, 20(23), p. 559-594.
- Marshall, T. & Mohammed, A. M., 2007, "Case-mix and the use of control charts in monitoring mortality rates after coronary artery bypass", en *BMC Health Services Research*, 7(63).
- McCullagh, P. and Nelder, J. A., 1989, *Generalized linear models*, London: Chapman and Hall/CRC.
- McCulloch, C.E. and Searle, S.R., 2001, *Linear and generalized linear mixed models*, New York: Wiley, 2001.
- Montgomery, D. C., 2009, *Introduction to Statistical Quality Control*, Arizona: John Wiley and Sons, Inc.

Naciones Unidas, 2015, *Transforming our world: the 2030 Agenda for Sustainable Development*. <https://sustainabledevelopment.un.org/post2015/transformingourworld>

Navarro, A., Utzet, F., Puig, P., Caminal, J. and Martín, M., 2001, “La distribución binomial negativa frente a la de Poisson en el análisis de fenómenos recurrentes”, en *Gaceta Sanitaria*, vol. 15, pp. 447-452.

Neuburger, J., Walker K, Sherlaw-Johnson C, van der Meulen, J., Cromwell, D., 2017, “Comparison of control charts for monitoring clinical performance using binary data”, en *BMJ quality & safety*, vol. 26, pp. 919-928.

Organización Mundial de la Salud, 2018, *Organización Mundial de la Salud*. Available at <https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality>. Último acceso: 15/03/2019.

Pierce, D. A. and Schafer, D. W., 1986, “Residuals in Generalized Linear Models”, en *Journal of the American Statistical Association*, 81(396), pp. 977-986.

R Core Team, 2008, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna (Austria): <https://www.R-project.org/>.

Roberts, S. W., 1959, “Control Chart Tests Based on Geometric Moving Averages”, en *Techonometrics*, 42(1), pp. 97-102.

Rolka, H. Burkom, H., Cooper, G.F., Kulldorff, M., Madigan, D., Wong, W.K., 2007, “Issues in Applied Statistics for Public Health Bioterrorism Surveillance Using Multiple Data Streams: Some Research Needs”, en *Statistics in Medicine*, vol. 26, pp. 1834-1856.

Romero-Sandoval, N., Alcázar, D., Pastor, J., and Martín, M., 2019, “Ecuadorian infant mortality linked to socioeconomic factors during the last 30 years”, en *Revista Brasileira de Saúde Materno Infantil*, 19(2), 295-301. Epub July 22, 2019. <https://dx.doi.org/10.1590/1806-93042019000200003>

Skinner, K. R., Montgomery, D. C. and Runger, G. C., 2003, “Process monitoring for multiple count data using generalized linear model-based control charts”, en *International Journal of Production Research*, 41(6), pp. 1167-1180.

Urbietta, P., Lee Ho, L. and Alencar, A., 2017, “CUSUM and EWMA Control Charts for Negative Binomial Distribution”, en *Quality and Reliability Engineering International*, 33(4), pp. 793-801.

Venables, W. N. and Ripley, B. D., 2002, *Modern Applied Statistics with S*, New York: Fourth Edition. Springer.

Vera, F., Solórzano, M., Ochoa, G., García Bustos, S. Cevallos, S., 2019, “Tablas de Mortalidad de Ecuador continental mediante un análisis de supervivencia”, en *Papeles de Población*, 24. ISSN 2448-7147. Disponible en <https://rppoblacion.uaemex.mx/article/view/9061>. Fecha de acceso: 20/03/2019

Wolleswinkel-van den Bosch, J. H., van Poppel, F. W., Looman, C. W. and Mackenbach, J. P., 2000, “Determinants of infant and early childhood mortality levels

and their decline in The Netherlands in the late nineteenth century”, en *International Journal of Epidemiology*, 29(6), p. 1031–1040.

Woodall, W. H., 2006, “The Use of Control Charts in Health-Care and Public-Health Surveillance”, en *Journal of Quality Technology*, 38(2), pp. 89-104.

Yoshioka, H. and Esparza Ochoa, J. C., 2009, “Desigualdad socioeconómica y mortalidad infantil en Nicaragua: ¿Una cuestión étnica?”, en *Latin American Research Review*, 44(1), pp. 199-217.

## RESUMEN CURRICULAR DE LAS AUTORAS

### *Sandra García Bustos*

Es profesora de estadística en la Escuela Superior Politécnica del Litoral (ESPOL) en Guayaquil, Ecuador. En 2014, recibió su doctorado en Estadística y Optimización en la Universidad Politécnica de Valencia, España. Su investigación doctoral se centró en el estudio de gráficos de control para variables de Poisson correlacionadas. Tiene estudios de tercer nivel en Estadística y una Maestría en Administración de Empresas. Sus intereses de investigación actuales son el control estadístico de procesos y estadísticas sociales.

Dirección electrónica: [slgarcia@espol.edu.ec](mailto:slgarcia@espol.edu.ec)

### *Eva Mera Intriago*

Es profesora de estadística en la Escuela Superior Politécnica del Litoral (ESPOL) en Guayaquil, Ecuador. Actualmente está estudiando su doctorado en Demografía en la Universidad Autónoma de Barcelona.

Dirección electrónica: [emera@espol.edu.ec](mailto:emera@espol.edu.ec)

### *Ana Debón*

Actualmente es Vicedecana de Calidad, Innovación e Infraestructuras de la Facultad de ADE. Respecto a su investigación ha realizado estancias en Concordia University (Canadá) y en Cass Business School (United Kingdom). La misma se centra en el campo de la estadística (modelación y pronóstico de mortalidad) con aplicaciones actuariales. Su trabajo ha aparecido en revistas líderes como *Insurance, Mathematics and Economics* o *European Journal Operation Research*. Además, participa en proyectos de I + D relacionados desde 2004 financiados por entidades regionales y nacionales.

Dirección electrónica: [andean@eio.upv.es](mailto:andean@eio.upv.es)

*Iris Bustamante*

Es Ingeniera en Estadística en la Universidad ESPOL. Trabaja en el Departamento de Matemáticas de ESPOL. Sus intereses de investigación actuales son el control estadístico de procesos y estadísticas sociales.

Dirección electrónica: [ijbustam@espol.edu.ec](mailto:ijbustam@espol.edu.ec)

*Illych Álvarez*

Es Profesor Asistente de la Facultad de Ciencias Naturales y Matemáticas de la Escuela Superior Politécnica de Litoral. En el ámbito docente, imparte clases en el Ciclo Básico de Carrera y las asignaturas de Álgebra Lineal, Ecuaciones Diferenciales y Cálculo. Actualmente es Coordinador de Aprendizaje Activo en Matemáticas en el curso de nivelación de la Universidad. Respecto a su investigación ha realizado estudios en Educación en Matemáticas, estudios Estadísticos de comportamiento demográfico y en Ecuaciones diferenciales relacionadas con la Ecuación del Transporte.

Dirección electrónica: [ialvarez@espol.edu.ec](mailto:ialvarez@espol.edu.ec)