



UNIVERSIDAD NACIONAL DE ROSARIO  
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA  
SECRETARIA DE CIENCIA Y TECNOLOGIA E INSTITUTOS DE INVESTIGACIONES

# **ACTAS**

***Jornadas Anuales***

***“Investigaciones en la Facultad”  
de Ciencias Económicas y  
Estadística***



**Anselmi Natalia**  
**Harvey Guillermina Beatriz**  
**Boggio Gabriela Susana**

*Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística*

## **MODELOS DE REGRESIÓN PARA DATOS DE CONTEO CON ALTA ASIMETRÍA<sup>1</sup>**

### **Resumen:**

En estudios que involucran el análisis de datos de conteo es común encontrar una variabilidad mayor que la esperada bajo el modelo paradigmático para este caso, el modelo Poisson. La sobredispersión provocada por una alta asimetría en los datos ha sido observada al estudiar la demanda de consultas pediátricas en centros de salud de la ciudad de Rosario, Argentina. Específicamente, para el estudio del número de consultas realizadas en los servicios de guardia considerando la edad de los niños y la asistencia a los consultorios pediátricos se consideraron diferentes alternativas de modelización. Las mismas, se enmarcan en la clase de modelos lineales generalizados con diferentes componentes aleatorias y sus versiones truncadas en cero debido a que sólo se consideraron los niños que requirieron atención. Se sumó, además, la propuesta desarrollada por Bonat *et al.* basada en los modelos con dispersión Poisson-Tweedie, los cuales se adaptan en forma automática a diferentes grados de sobredispersión en datos de conteo. Resultó satisfactorio el ajuste del modelo binomial negativo truncado y el Poisson-Tweedie con parámetro de potencia estimado igual a 4,09, lo que confirma que dicho modelo se adaptó a la alta asimetría observada. Si bien los resultados en términos de la significación de los efectos son los mismos bajo ambos ajustes, la simplificación que otorga el uso de la familia Poisson-Tweedie en el proceso de búsqueda del mejor modelo la hace sumamente atractiva.

Palabras claves: Datos de conteo, Asimetría, Modelos truncados, Familia Poisson-Tweedie

### **Abstract:**

In studies involving count data it is common to find greater variability than the expected under the paradigmatic model for this case, the Poisson model. The overdispersion caused by the high skewness has been observed when analysing the demand for consultations of a group of children who receive health care in a public health center in Rosario, Argentina. Specifically, to assess the effect of children's age and the number of visits carried out in the pediatric offices on the number of pediatric consultations in the emergency room of a health center, different modeling alternatives for the treatment of overdispersed data were considered. These alternatives are framed in the generalized linear models class, with different random components and their zero-truncated versions since only children who required care were considered. It was also included a new type of model, based on the Poisson-Tweedie dispersion models, which was proposed by Bonat *et al.* This model can automatically adapt to different degrees of overdispersion in count data. The zero-truncated negative binomial model and the Poisson-Tweedie model showed satisfactory fit. The estimated power parameter of the Poisson-Tweedie model was 4.09, which confirms that this model adapted to the high skewness observed. The results obtained in both models were similar. However, the flexibility of the Poisson-Tweedie model makes it an attractive

---

<sup>1</sup> Trabajo elaborado en el marco del Proyecto 80020220700095UR, titulado: "Modelización estadística para estructuras complejas de datos", dirigido por Gabriela S. Boggio.



choice since the estimation of the power parameter can work as an automatic distribution selection.

Keywords: Count data, Skewness, Truncated models, Poisson-Tweedie family

## Introducción

En el proceso de ajuste de un modelo de regresión para el caso de respuestas de tipo conteo la distribución paradigmática es la distribución Poisson. La misma se caracteriza por la igualdad entre su media y variancia, supuesto difícil de verificar ya que en la práctica las observaciones de conteo frecuentemente exhiben una variabilidad que excede la esperada. El fenómeno por el cual un modelo tiene mayor variancia que la supuesta por su componente aleatoria se denomina sobredispersión.

La sobredispersión tiene importantes implicancias, ya que puede llevar a una subestimación de los errores estándares asociados a los coeficientes estimados, lo que, a su vez, puede conducir a inferencias erróneas durante el análisis. Detectar este fenómeno y utilizar metodología adecuada se vuelve crucial. La literatura que lo aborda es muy extensa y las propuestas para tratarlo son diversas (Hinde y Demétrio, 1998; Hilbe, 2011; Hilbe, 2014; Bonat *et al.*, 2018).

Son muchas las posibles causas de la presencia de sobredispersión y en el proceso de modelar datos, varias de ellas pueden intervenir simultáneamente. A veces puede corregirse mediante la incorporación de un predictor apropiado, la inclusión de interacciones significativas, o bien utilizando una función de enlace diferente. Otras veces la sobredispersión puede provenir de dos características inherentes a la variable respuesta: la presencia excesiva de ceros o bien una alta asimetría (Hinde y Demétrio, 1998; Hilbe, 2011).

Esta última característica se observó al analizar el número de consultas pediátricas en centros de salud de la ciudad de Rosario, lo que llevó a explorar diversas estrategias de modelado para abordar su tratamiento. En particular, el objetivo del estudio fue modelar la demanda de consultas en los servicios de guardia de estos centros de salud, teniendo en cuenta la edad de los niños y el número de consultas realizadas en los consultorios de pediatría de dichos centros.

Es así que este trabajo presenta, mediante la aplicación al problema recién mencionado, opciones para identificar y tratar la sobredispersión presente en los datos debida a alta asimetría.



### Descripción de los datos

Se cuenta con información referida al número de consultas pediátricas realizadas por 1122 niños en los servicios de guardia de cinco centros de salud de la ciudad de Rosario, Argentina, durante el año 2019. Los centros considerados son Casals Mauricio, Dr. Salvador Mazza, Ceferino Namuncurá, Luis Pasteur e Itatí ubicados en los distritos oeste, norte, noroeste, sur y sudoeste, respectivamente. Cada centro de salud corresponde a un distrito de la ciudad, excluyendo el distrito centro debido a que presenta características particulares. Su centro de salud no cuenta con una población infantil adscripta al mismo, es decir, con un equipo médico de referencia en dicha institución.

Para cada niño, se tiene información sobre su edad (entre 0 y 4 años) y el número de consultas realizadas en los consultorios del servicio de pediatría de dichos centros. Como ya fuera mencionado, se observa una distribución asimétrica hacia la derecha en la variable número de visitas al servicio de guardia (Figura 1). Para estos datos, el número medio de visitas es 2,7 y la varianza 5,9. Tanto la asimetría observada, como la relación entre la media y la varianza muestral, sugieren la presencia de sobredispersión. Por lo tanto, no sería apropiado el uso del modelo Poisson para este análisis.

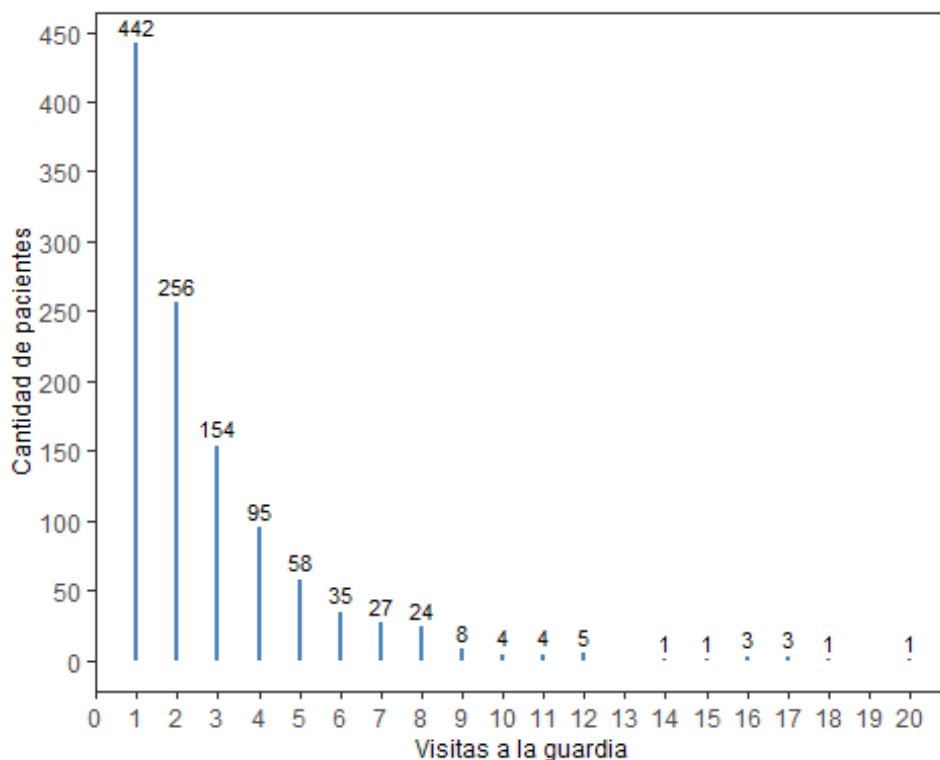


Figura 1: Distribución del número de visitas al servicio de guardia

Al analizar las distribuciones condicionales, parece notarse un menor número de visitas en aquellos niños de 3 y 4 años, en comparación con los de menor edad así como una leve correlación positiva entre el número de visitas al servicio de guardia y en consultorio (Figura 2).

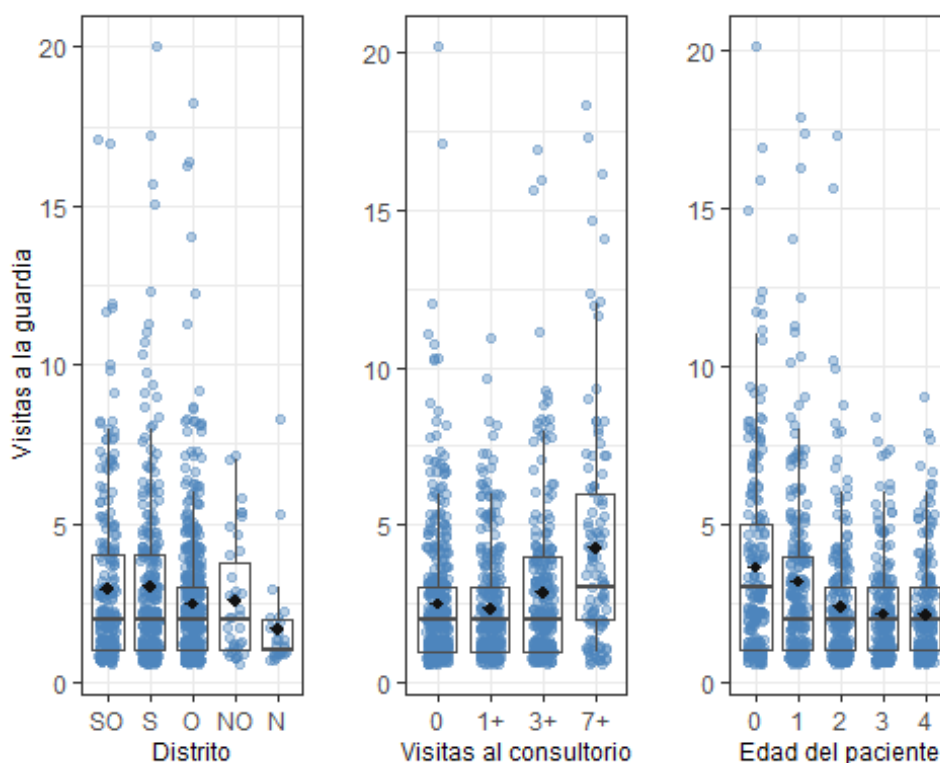


Figura 2: Distribución del número visitas al servicio de guardia en función del distrito, de las visitas al consultorio externo y de la edad

A continuación se presenta una breve descripción de las posibilidades analíticas para tener en cuenta la alta asimetría observada en los datos.

### Alternativas de modelización

Para el modelado de los datos de conteo se utilizan habitualmente los modelos lineales generalizados (MLGs) definidos por Nelder y Wedderburn (1972), los cuales se describen brevemente a continuación.

#### Modelos lineales generalizados

Los MLGs extienden los modelos de regresión lineal para abarcar variables respuesta con distribuciones no normales y funciones no lineales de la media. Se definen mediante tres componentes:

- **Componente aleatoria:** identifica la variable respuesta  $Y$ , para la cual asume una distribución de probabilidad perteneciente a la familia exponencial con dispersión (Jørgensen, 1987). La función de densidad de esta familia es de la forma:

$$f(y_i; \theta_i, \phi) = a(y_i, \phi) \exp \left[ \frac{1}{\phi} \{y_i \theta_i - \kappa(\theta_i)\} \right],$$

donde las observaciones  $y_i$ , con  $i = 1, \dots, n$  son independientes,  $a(\cdot)$  es el término de normalización que garantiza que la probabilidad sume uno,  $\kappa(\cdot)$  es la función acumulante que determina los momentos de  $y_i$ ,  $\theta_i$  es el parámetro canónico o natural y  $\phi$  es el parámetro de dispersión. La elección de  $a(\cdot)$  y  $\kappa(\cdot)$  determina la distribución.

La media de la distribución está dada por  $\mu_i = E(y_i) = \kappa'(\theta_i)$  y la variancia por  $\text{Var}(y_i) = \phi v(\mu_i)$  siendo  $v(\mu_i) = \kappa''(\theta_i)$  la denominada función variancia.

Un caso especial ocurre cuando  $v(\mu) = \mu^p$  para cualquier  $p$  fuera del intervalo  $(0,1)$ . Estos modelos pertenecen a una subclase de la familia exponencial con dispersión, los modelos Tweedie (Jørgensen, 1987). El parámetro  $p$  determina la distribución. La distribución normal ( $p = 0$ ), Poisson ( $p = 1$ ), gamma ( $p = 2$ ) y gaussiana inversa (IG) ( $p = 3$ ) pertenecen a la familia Tweedie. Fuera de los casos particulares, las distribuciones Tweedie no tienen funciones de densidad que se puedan escribir en forma cerrada (Dunn y Smyth, 2005).

- **Componente sistemática:** se define como el predictor lineal  $x_i\beta$ , donde  $\beta = (\beta_1, \dots, \beta_p)^T$  es el vector de parámetros y  $x_i$  es el vector de covariables asociado a la  $i$ -ésima observación con  $i = 1, \dots, n$ .
- **Función de enlace:**  $g(\cdot)$  es una función monótona que establece la relación funcional entre la componente sistemática y el valor esperado de la componente aleatoria, es decir  $g(\mu_i) = x_i\beta$ . La función  $g$  para la cual  $g(\mu_i) = \theta_i$ , se denomina enlace canónico o natural.

En cuanto a la estimación de los MLGs, en los casos donde la función de densidad puede especificarse, el vector de parámetros se estima por medio de máxima verosimilitud mediante métodos iterativos (Agresti, 2015).

La distribución más simple para datos de conteo es la distribución Poisson. Sus probabilidades dependen de un único parámetro, la media  $\mu > 0$ . La función de probabilidad queda expresada como sigue:

$$f(y, \mu) = \frac{e^{-\mu} \mu^y}{y!} = \exp[y \log(\mu) - \mu - \log(y!)]$$

El parámetro natural es  $\theta = \log \mu$  y la esperanza es igual a la variancia  $E(y) = \text{Var}(y) = \mu$ . Mientras mayor sea la media, mayor será la variabilidad de los datos. Una alta asimetría en los datos de conteo producirá una media menor que la variancia observada.

El supuesto de que la variancia de los datos es igual a la media también implica que la variabilidad de los resultados que comparten los mismos valores de covariables es igual a la correspondiente media. Si no es cierto, la sobredispersión presente en los datos provoca que las estimaciones de los errores estándar sean sesgadas. Por lo general, tienden a ser demasiado pequeños y, por lo tanto, los tests resultan liberales seleccionando modelos sobrep parametrizados conduciendo a interpretaciones incorrectas (Hilbe, 2014; Yang *et al.*, 2017).

Un modelo que puede ser lo suficientemente flexible para solucionar el problema de la sobredispersión es el modelo Binomial negativo (BN), en el cual la variable respuesta se distribuye Poisson con una media  $\lambda$  que varía siguiendo una distribución gamma:

$$y \mid \lambda \sim \text{Poisson}(\lambda) \\ \lambda \sim \text{gamma}(\mu, \phi)$$

donde  $\mu$  es la media y  $\phi$  es el parámetro de dispersión.

La función de probabilidad sigue la forma:

$$f(y; \mu, \phi) = \frac{\Gamma(1/\phi + y)}{\Gamma(1/\phi)\Gamma(1 + y)} \left( \frac{1}{1 + \phi\mu} \right)^{1/\phi} \left( \frac{\phi\mu}{1 + \phi\mu} \right)^y$$

En este caso, la variancia es  $\text{Var}(y) = \mu + \phi\mu^2$ . A medida que  $\phi$  disminuye, la distribución BN tiende a la distribución Poisson. Este modelo puede capturar una parte de la sobredispersión.

Otro modelo que se propone es el Poisson gaussiano inverso (PIG). Este modelo tiene la capacidad de capturar picos pronunciados y asimetría hacia la derecha en la distribución de datos de conteo. En consecuencia puede comportarse mejor que la BN cuando los datos están muy dispersos (Hilbe, 2014). En este caso, la variable respuesta se distribuye Poisson con media  $\lambda$  que sigue, a su vez, una distribución IG:

$$\begin{aligned} y | \lambda &\sim \text{Poisson}(\lambda) \\ \lambda &\sim \text{IG}(\mu, \phi) \end{aligned}$$

donde  $\mu$  es la media y  $\phi$  es el parámetro de dispersión ya mencionados.

La función de probabilidad es:

$$f(y; \mu, \phi) = \left(\frac{2\alpha}{\phi}\right)^{\frac{1}{2}} \mu^y \frac{e^{\frac{1}{\phi}} K_{y-(1/2)}(\alpha)}{(\alpha\phi)^y y!}$$

donde  $\alpha^2 = \frac{1}{\phi^2} + \frac{2\mu}{\phi}$  y  $K_\lambda(\cdot)$  es la función de Bessel modificada de tercer tipo (Johnson *et al.*, 2005). La variancia está dada por  $\text{Var}(y) = \mu + \phi\mu^3$  y al igual que con la distribución BN, a medida que  $\phi$  aumenta, la distribución tiende a Poisson.

Una situación que puede presentarse es que no haya posibilidad de que la variable de conteo tome el valor cero, nulo. Cuando los datos estructuralmente excluyen los conteos nulos, la distribución de probabilidad subyacente no debería admitir ese resultado entre los posibles. Es decir, las distribuciones de probabilidad deben ser corregidas para excluir los ceros. Esta corrección implica ajustar la escala de la distribución de modo que la suma de todas las probabilidades sea equivalente a 1, esto es:

$$P(Y = y | Y > 0) = \frac{P(Y = y)}{1 - P(Y = 0)}$$

A modo de ejemplo, en el caso de la distribución Poisson, la probabilidad de obtener un conteo igual a cero es  $e^{-\mu}$ . Entonces, la función de probabilidad truncada queda expresada como sigue:

$$f(y, \mu) = \frac{e^{-\mu} \mu^y}{(1 - e^{-\mu}) y!} = \exp[y \log(\mu) - \mu - \log(1 - e^{-\mu}) - \log(y!)]$$

Análogamente, las distribuciones de probabilidad BN y PIG se modifican teniendo en cuenta la correspondiente probabilidad de obtener un conteo nulo. En la distribución BN esta probabilidad es  $(1 + \phi\mu)^{-1/\phi}$  y en la PIG es  $(\pi/\phi)^{1/2} e^{(1/\phi) - \alpha}$ .

Si bien estas distribuciones truncadas no son las consideradas por McCullagh y Nelder (1989) al definir los MLGs, pueden ser utilizadas como componente aleatoria de estos modelos ya que pertenecen a la familia exponencial natural.

Es importante considerar la utilidad de las distribuciones truncadas, ya que su aplicación puede no ser necesaria a menos que la media de la distribución de conteos a modelar sea extremadamente baja. Por ejemplo, en un modelo Poisson, si la media es igual o mayor a 5, la probabilidad de tener un conteo nulo es menor a 0,01. En estos casos, dado que esta probabilidad es tan baja, resulta irrelevante considerar la versión truncada.

### Modelo Poisson-Tweedie

Una alternativa de modelización que se suma a las mencionadas, es el denominado modelo Poisson-Tweedie introducido por Kokonendji *et al.* (2004) y luego desarrollado por Jørgensen y Kokonendji (2016). Éste es un modelo que puede adaptarse automáticamente a la dispersión subyacente y ser implementado fácilmente en la práctica (Bonat *et al.*, 2018). La variable respuesta se distribuye Poisson con una media  $\lambda$  que varía siguiendo una distribución Tweedie de parámetro de potencia  $p$  y parámetro de dispersión  $\phi$ :

$$y \mid \lambda \sim \text{Poisson}(\lambda) \\ \lambda \sim \text{Tw}_p(\mu, \phi)$$

La variancia tiene la forma  $\text{Var}(y) = \mu + \phi\mu^p$ . Tanto la distribución BN como la distribución PIG son casos particulares del modelo Poisson-Tweedie cuando  $p = 2$  y  $p = 3$  respectivamente. Puede notarse que para estos casos, resultan  $\text{Var}(y) = \mu + \phi\mu^2$  y  $\text{Var}(y) = \mu + \phi\mu^3$ , como ya fuera mencionado previamente.

Esta clase provee una estructura unificada para manejar la sobredispersión ya sea originada por alta asimetría o por exceso de ceros, así como también equidispersión y subdispersión.

En la estimación de estos modelos se utiliza un enfoque de función de estimación introducido por Bonat y Jørgensen (2016) que se basa en una estimación cuasi-verosímil para los parámetros de la regresión y funciones de estimación de Pearson para los parámetros de dispersión  $\phi$  y  $p$ .

La importancia de esta clase de modelos es que el parámetro de potencia juega un rol fundamental permitiendo abarcar un abanico de posibilidades para captar diferentes niveles de asimetría en los datos así como exceso de ceros. Bonat *et al.* (2018), mediante un estudio empírico, muestran por ejemplo que los modelos Poisson-Tweedie con  $p = 3$  son adecuados en situaciones en las que la sobredispersión se debe a una alta asimetría a la derecha mientras que cuando el parámetro  $p$  se encuentra entre 1 y 2 se corresponde con distribuciones que presentan exceso de ceros. El caso de  $p = 2$  representa situaciones en las cuales hay un punto de inflexión entre distribuciones con un alto exceso de ceros y aquellas con una alta asimetría. Es así que el parámetro de potencia funciona, de cierta forma, como un método de selección de distribución automático.

Es importante señalar que el cálculo del valor del AIC no es posible en el contexto del modelo Poisson-Tweedie, dado que este enfoque no implica el uso del método de estimación de máxima verosimilitud. Por lo tanto, la comparación basada en el AIC no está disponible para este modelo específico.

## Resultados

Para modelar el número de consultas de los niños que requirieron atención en los servicios de guardia de los centros de salud se ajustan los modelos planteados en la sección previa.

En primer lugar, se ajustan los MLGs clásicos con componente aleatoria Poisson, BN y PIG. Cabe señalar que, a pesar de la sobredispersión detectada, se considera de todos modos el ajuste Poisson con fines comparativos. Al confrontar dichos modelos se observa que los coeficientes estimados exhiben una notoria similitud, mientras que los errores estándares presentan diferencias. Conforme a lo previsto, resultan menores en el modelo Poisson, fenómeno atribuible a la sobredispersión que caracteriza los datos. De acuerdo a los valores de AIC, parecería que el modelo PIG es el más adecuado (Tabla 1).





Tabla 1: Resultados obtenidos bajo los modelos Poisson, BN y PIG

Parámetro	Poisson		BN		PIG	
	Estimación (EE)	Prob. asoc.	Estimación (EE)	Prob. asoc.	Estimación (EE)	Prob. asoc.
Intercepto	1,26 (0,05)	<0,001	1,24 (0,06)	<0,001	1,23 (0,06)	<0,001
Visitas al consultorio	0,04 (0,00)	<0,001	0,04 (0,01)	<0,001	0,04 (0,01)	<0,001
Edad	-0,12 (0,01)	<0,001	-0,12 (0,02)	<0,001	-0,11 (0,02)	<0,001
Distrito (S vs SO)	0,02 (0,05)	0,752	0,01 (0,06)	0,852	0,01 (0,07)	0,830
Distrito (O vs SO)	-0,26 (0,05)	<0,001	-0,26 (0,06)	<0,001	-0,25 (0,06)	<0,001
Distrito (NO vs SO)	-0,18 (0,11)	0,096	-0,16 (0,13)	0,238	-0,14 (0,14)	0,290
Distrito (N vs SO)	-0,67 (0,15)	<0,001	-0,65 (0,17)	<0,001	-0,63 (0,18)	<0,001
$\phi$			-1,68 (0,11)	<0,001	-1,49 (0,12)	<0,001
AIC	4567		4396		4368	

EE: error estándar

Dado que la variable bajo análisis sólo puede tomar valores positivos se evalúa si es necesaria la consideración de las versiones truncadas de los modelos recién presentados. Así, se calculó la probabilidad de un conteo nulo bajo cada distribución. Se obtuvo un valor de 0,065 bajo la distribución Poisson; 0,110 bajo la BN y 0,112 para la PIG. Dada la magnitud de estas probabilidades, se consideró pertinente incluir el ajuste de los modelos truncados. El ajuste del modelo PIG truncado (PIG-T) presentó problemas de convergencia, por lo que en Tabla 2 se presenta la comparación de los modelos Poisson y BN con su correspondiente versión truncada (Poisson-T y BN-T).

La consideración de las distribuciones truncadas no muestran diferencias en cuanto la significación de los efectos, tanto en lo que respecta al modelo Poisson como al BN. No obstante, en ambos casos se notan diferencias en la magnitud de las estimaciones de los coeficientes y de sus errores estándares. Los valores son mayores en las versiones truncadas y las diferencias más marcadas se presentan en la BN. Las versiones truncadas presentan valores de AIC menores, por lo que el modelo BN truncado (BN-T) podría ser el que mejor representa la complejidad inherente a los datos en comparación con las alternativas consideradas. Notar que su valor de AIC es incluso menor al del modelo PIG señalado como el más adecuado entre los clásicos.

Por último, se agrega a la comparación el ajuste de la familia Poisson-Tweedie la que, mediante la estimación del parámetro de potencia óptimo, permite captar la asimetría presente en los datos (Tabla 3). Los efectos significativos resultan ser los mismos que bajo el modelo BN-T, sin embargo, se observa un valor para el parámetro de potencia estimado mayor a 4 captando la alta asimetría observada en la Figura 1. Es de notar que este valor es mayor que el correspondiente a la distribución paradigmática para el manejo de datos asimétricos, la PIG, cuyo valor de parámetro de potencia en la familia Poisson-Tweedie es 3.



Tabla 2: Resultados obtenidos bajo los modelos Poisson, Poisson-T, BN y BN-T

Parámetro	Poisson		Poisson-T		BN		BN-T	
	Estimación (EE)	Prob. asoc.	Estimación (EE)	Prob. asoc.	Estimación (EE)	Prob. asoc.	Estimación (EE)	Prob. asoc.
Intercepto	1,26 (0,05)	<0,001	1,25 (0,05)	<0,001	1,24 (0,06)	<0,001	0,91 (0,11)	<0,001
Visitas al consultorio	0,04 (0,00)	<0,001	0,04 (0,00)	<0,001	0,04 (0,01)	<0,001	0,06 (0,01)	<0,001
Edad	-0,12 (0,01)	<0,001	-0,16 (0,02)	<0,001	-0,12 (0,02)	<0,001	-0,19 (0,03)	<0,001
Distrito (S vs SO)	0,02 (0,05)	0,752	0,02 (0,06)	0,762	0,01 (0,06)	0,852	0,01 (0,11)	0,954
Distrito (O vs SO)	-0,26 (0,05)	<0,001	-0,34 (0,06)	<0,001	-0,26 (0,06)	<0,001	-0,42 (0,10)	<0,001
Distrito (NO vs SO)	-0,18 (0,11)	0,096	-0,24 (0,12)	0,056	-0,16 (0,13)	0,238	-0,19 (0,22)	0,403
Distrito (N vs SO)	-0,67 (0,15)	<0,001	-1,00 (0,21)	<0,001	-0,65 (0,17)	<0,001	-1,13 (0,31)	<0,001
$\phi$					-1,68 (0,11)	<0,001	-0,02 (0,16)	0,456
AIC	4567		4332		4396		3903	

EE: error estándar; Poisson-T: Poisson truncado; BN-T: BN truncado

En relación al efecto que tienen las covariables consideradas puede mencionarse que los niños que mayor demanda generan en el servicio de guardia pediátrica son quienes realizan un mayor número de visitas a los consultorios pediátricos y los de menor edad. Con respecto a los distritos, la mayor demanda parece presentarse en el sudoeste, con diferencias significativas entre este distrito y el norte y el sur. Estos resultados están en concordancia con el análisis marginal descripto oportunamente.

Tabla 3: Resultados obtenidos bajo los distintos modelos BN-T y Poisson-Tweedie

Parámetro	BN-T		Poisson-Tweedie	
	Estimación (EE)	Prob. asoc.	Estimación (EE)	Prob. asoc.
Intercepto	0,91 (0,11)	<0,001	1,22 (0,07)	<0,001
Visitas al consultorio	0,06 (0,01)	<0,001	0,04 (0,01)	<0,001
Edad	-0,19 (0,03)	<0,001	-0,12 (0,02)	<0,001
Distrito (S vs SO)	0,01 (0,11)	0,954	0,01 (0,07)	0,997
Distrito (O vs SO)	-0,42 (0,10)	<0,001	-0,25 (0,06)	<0,001
Distrito (NO vs SO)	-0,19 (0,22)	0,403	-0,10 (0,14)	0,493
Distrito (N vs SO)	-1,13 (0,31)	<0,001	-0,61 (0,17)	<0,001
$\phi$	-0,02 (0,16)	0,456	0,03 (0,02)	0,043
$p$			4,09 (0,38)	<0,001

EE: error estándar; BN-T: BN truncado

## Consideraciones finales

La gran variabilidad detectada por la alta asimetría de los datos sobre la demanda de consultas en el servicio de guardia pediátrica pudo ser captada al considerar MLGs con componentes aleatorias que presentan variancias superiores al valor medio.

En este trabajo, se exploraron los modelos tradicionales para abordar el problema de la sobredispersión y los modelos truncados, resultando satisfactorio el ajuste del MLG BN-T. Asimismo, se consideró la innovadora clase de modelos, la familia Poisson-Tweedie, la cual proporciona un marco unificado para manejar datos de conteos con distintas peculiaridades, como el exceso de ceros y la alta asimetría que se observa en el problema abordado. El ajuste de este modelo resultó en un parámetro de potencia estimado igual a 4,09, lo que confirma que el modelo se adaptó a la alta asimetría observada. Si bien los resultados en términos de la significación de los efectos son los mismos bajo ambos modelos, la simplificación que otorga el uso de la familia Poisson-Tweedie en el proceso de búsqueda del mejor modelo la hace sumamente atractiva.

Vale comentar que, si bien la versión truncada de los modelos Poisson-Tweedie ha sido desarrollada por Valero *et al.* (2012), no fue implementada en software estadístico.

Queda pendiente la incorporación de efectos aleatorios en el ajuste de la familia Poisson-Tweedie lo cual permitiría capturar las variaciones debidas al factor de distrito. Ello se debe a que los centros analizados pueden considerarse como una muestra aleatoria de la amplia red de centros de salud pertenecientes a la Municipalidad de Rosario. Este enfoque permitiría modelar las complejas interacciones dentro de la red, ofreciendo una comprensión más rica y detallada de los datos observados.

## Referencias Bibliográficas

- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons. <https://books.google.com.ar/books?id=jllqBgAAQBAJ>.
- Bonat, W. H.; Jørgensen, B. (2016). Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 65(5), 649–675. <https://doi.org/10.1111/rssc.12145>.
- Bonat, W. H.; Jørgensen, B. ; Kokonendji, C. C. ; Hinde, J.; Demétrio, C. G. B. (2018). Extended Poisson–Tweedie: Properties and regression models for count data. *Statistical Modelling*, 18(1), 24–49. <https://doi.org/10.1177/1471082X17715718>.
- Dunn, P. K.; Smyth, G. K. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, 15(4), 267–280. <https://doi.org/10.1007/s11222-005-4070-y>.
- Hilbe, J. M. (2011). *Negative binomial regression, second edition*. <https://doi.org/10.1017/CBO9780511973420>.
- Hilbe, J. M. (2014). *Modeling Count Data*. Cambridge University Press. <https://books.google.com.ar/books?id=aZLfAwAAQBAJ>.
- Hinde, J.; Demétrio, C. G. B. (1998). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, 27(2), 151–170. [https://doi.org/10.1016/S0167-9473\(98\)00007-3](https://doi.org/10.1016/S0167-9473(98)00007-3).
- Johnson, N. L. ; Kemp, A. W.; Kotz, S. (2005). *Univariate Discrete Distributions*. John Wiley & Sons. <https://books.google.com.ar/books?id=JchiadWLnykC>.
- Jørgensen, B. (1987). Exponential Dispersion Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(2), 127–162. <https://www.jstor.org/sTabla/2345415>.



UNR

- Jørgensen, B.; Kokonendji, C. C. (2016). Discrete dispersion models and their Tweedie asymptotics. *AStA Advances in Statistical Analysis*, 100(1), 43–78. <https://doi.org/10.1007/s10182-015-0250-z>.
- Kokonendji, C.; Demétrio, C. G. B.; Dossou Gbete, S. (2004). Overdispersion and Poisson-Tweedie exponential dispersion models. *VIII Journées Zaragoza-Pau de Mathématiques Appliquées Et de Statistiques :Jaca, Spain, September 15-17, 2003, 2003-01-01, ISBN 84-7733-720-9, Pags. 365-374, 31*.
- McCullagh, P.; Nelder, J. (1989). *Generalized Linear Models*, Second Edition. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Nelder, J. A.; Wedderburn, R. W. M. (1972). *Generalized linear models*. Journal of the Royal Statistical Society. Series A (General), 135(3):370.
- Valero, J.; Ginebra, J.; Pérez-Casany, M. (2012). *Extended truncated Tweedie-Poisson model*. <https://doi.org/10.1007/s11009-012-9277-8>.
- Yang, S.; Puggioni, G.; Harlow, L. L.; Redding, C. A. (2017). A comparison of different methods of zero - inflated data analysis and an application in health surveys. *Journal of Modern Applied Statistical Methods*, 16(1), 518–543. <https://doi.org/10.22237/jmasm/1493598600>.

## Fuentes

Los datos analizados fueron provistos por la Secretaría de Salud Pública de la Municipalidad de Rosario.