



Distribución de Poisson en el análisis de defunciones en menores de un año

Poisson distribution in the analysis of deaths in children under one year of age

Distribuição de Poisson na análise de óbitos em menores de um ano

Verónica Janeth Argüello-Pazmiño ^I
veronica.arguello@ueb.edu.ec
<http://orcid.org/0000-0002-5508-9538>

Salomón Rodrigo Cargua-Suárez ^{II}
scargua@ueb.edu.ec
<https://orcid.org/0000-0002-8011-4598>

Andrés Miguel Argüello-Pazmiño ^{III}
aarguello@umet.edu.ec
<https://orcid.org/0000-0003-3661-2925>

Alexandra Maribel Argüello-Pazmiño ^{IV}
amarguello@ueb.edu.ec
<https://orcid.org/0000-0002-1409-6360>

Correspondencia: veronica.arguello@ueb.edu.ec

Ciencias Matemáticas
Artículo de Investigación

* **Recibido:** 15 de febrero de 2023 * **Aceptado:** 20 de abril de 2023 * **Publicado:** 04 de mayo de 2023

- I. Magíster en Matemática Mención Modelación y Docencia, Universidad Estatal de Bolívar, Ecuador.
- II. Magíster en Pedagogía de la Matemática, Universidad Estatal de Bolívar, Ecuador.
- III. Magíster en Dirección de Operaciones y Seguridad Industrial, Universidad Metropolitana del Ecuador, Ecuador.
- IV. Magíster en Estadística Mención en Gestión de la Calidad y Productividad, Universidad Estatal de Bolívar, Ecuador.

Resumen

En la Universidad hemos sido abrumados con mucha información dejando de lado la importancia que tiene dicha información en la aplicación de los conocimientos y arremetiendo al docente en la clase, es así que invitamos en la presente investigación a caracterizar y modelar las muertes en menores de un año, mediante modelos lineales generalizados a través de la distribución de Poisson. Este estudio tuvo una investigación descriptiva puesto que se aplicó un análisis exploratorio que permitió conocer la tasa de mortalidad y que causas de muerte son las más influyentes, se contó con una investigación correlacional, con el número de eventos en menores de un año que va a permitir modelar los datos de conteo y se verificó la medida de calidad del modelo mediante AIC. Las estimaciones de los coeficientes del modelo, se verificó que todas las variables son significativas debido a que tiene el valor p menor que el nivel de significancia de 0.05. Finalmente la regresión de Poisson es la que mejor se ajusta a los datos, además tiene un R^2 de 0.858 y un AIC de 650.72, se concluye que el modelo puede utilizar datos de mortalidad infantil de menores de un año de todo el Ecuador.

Palabras Clave: Modelo Matemático; Modelo Lineal Generalizado; Regresión de Poisson.

Abstract

At the University we have been overwhelmed with a lot of information, leaving aside the importance of said information in the application of knowledge and attacking the teacher in the class, so we invite in the present investigation to characterize and model deaths in children under one year, using generalized linear models through the Poisson distribution. This study had a descriptive investigation since an exploratory analysis was applied that allowed to know the mortality rate and which causes of death are the most influential, there was a correlational investigation, with the number of events in children under one year of age that will allow modeling the count data and the quality measure of the model was verified through AIC. The estimates of the model coefficients, it was verified that all the variables are significant because it has a p value less than the significance level of 0.05. Finally, the Poisson regression is the one that best fits the data, it also has an R^2 of 0.858 and an AIC of 650.72, it is concluded that the model can use infant mortality data for children under one year of age from all over Ecuador.

Keywords: Mathematical model; Generalized Linear Model; Poisson regression.

Resumo

Na Universidade temos ficado sobrecarregados com muitas informações, deixando de lado a importância dessas informações na aplicação do conhecimento e atacando o professor na aula, por isso convidamos na presente investigação a caracterizar e modelar mortes em crianças menores de um ano, utilizando modelos lineares generalizados através da distribuição de Poisson. Este estudo teve uma investigação descritiva pois foi aplicada uma análise exploratória que permitiu conhecer a taxa de mortalidade e quais causas de morte são mais influentes, houve uma investigação correlacional, com o número de eventos em menores de um ano que permitirá a modelagem dos dados de contagem e a medida de qualidade do modelo foram verificadas por meio do AIC. Pelas estimativas dos coeficientes do modelo, verificou-se que todas as variáveis são significativas porque possui um valor de p menor que o nível de significância de 0,05. Por fim, a regressão de Poisson é a que melhor se ajusta aos dados, também possui um R^2 de 0,858 e um AIC de 650,72, conclui-se que o modelo pode utilizar dados de mortalidade infantil de crianças menores de um ano de todo o Equador.

Palavras-chave: Modelo matemático; Modelo Linear Generalizado; Regressão de Poisson.

Introducción

Cada ciclo se amplía el desarrollo de modelos matemáticos que conllevan a la aplicación y manipulación de datos a partir de la utilización de modelos estadísticos que permiten analizar el comportamiento de los mismos. Por lo tanto, un modelo es un proceso de abstracción desde la realidad al sistema matemático con el objetivo de facilitar la comprensión del suceso que se está estudiando (Calzada, 2017). Por otra parte, el objetivo de la modelización estadística es, por tanto, a través de la observación o experimentación, explicar el comportamiento de una o más variables en los individuos de una población en base a la diferencia entre las características asociadas al objeto de estudio. Así mismo Calzada (2017) menciona que en el planteamiento de un modelo es muy importante distinguir el tipo de variable que intervienen y la clase de relaciones funcionales que se admiten para analizar la relación entre la variable objetivo y las variables explicativas además señala que un modelo debe pasar por las siguientes etapas: especificación del modelo teórico, estimación de parámetros, selección del modelo, evaluación del modelo e interpretación del modelo (p.13). Los modelos lineales (GLM) amplían los modelos de regresión lineal para abarcar distribuciones de respuesta no normales como es el caso de la distribución de Poisson

(Arguello, 2022). Con respecto a la importancia que engloba la distribución de poisson, no obstante, algunos autores como Reading y Reid (2013) han reportado que existe evidencias sobre la posibilidad de investigar sobre el desarrollo de las concepciones de los estudiantes sobre la variabilidad y distribución. En relación con la teoría de probabilidad y estadística la distribución de Poisson es una distribución de probabilidad discreta que expresa la probabilidad de un número de eventos, esta distribución fue descubierta por Siméon Denis Poisson, que fue conocida a partir de 1838 (Rodríguez, 2022). Habría que decir también que resulta atractivo el trabajo de Ramírez et al., (2013) mencionan la enseñanza de la distribución de Poisson resulta verla como un proceso en el cual ocurren sucesos discretos en un intervalo continuo de tiempo que tiene tres lugares condicionantes esenciales primero, que la probabilidad de observar exactamente un suceso en el intervalo sea estable, segundo, que la probabilidad de observar dos o más sucesos en el intervalo sea cero y tercero, que la ocurrencia de un suceso en cualquier intervalo sea estadísticamente independiente de la ocurrencia de un suceso en cualquier otro intervalo de igual magnitud. Hay que mencionar que la distribución de Poisson esta estrechamente relacionada con la distribución binomial y que juega un papel importante en la solución de problemas (Ramírez et al., 2013). En relación a lo que caracteriza a tales procesos es su dependencia del tiempo, o sea, el hecho de que ciertos eventos suceden o no (por azar) a intervalos regulares de tiempo o en un intervalo continuo de tiempo (Freund et al., 2006). Con respecto a lograr una mejor comprensión de la distribución de Poisson, es conveniente complementar el análisis con dos consideraciones importantes. Por una parte, la probabilidad de ocurrencia de un suceso en el intervalo es proporcional a la longitud del intervalo, lo que equivale a plantear la relación $P = \theta \Delta t$, donde Δt indica un intervalo muy pequeño y θ una constante que representa el promedio de sucesos que ocurren en el intervalo. Por otra parte, la probabilidad de que ocurra más de un suceso en el intervalo pequeño es despreciable, lo cual se puede expresar mediante la relación $P \approx 0$ (Ramírez et al., 2013).

El siguiente aspecto trata de la devianza que es una medida de la bondad de ajuste de los modelos lineales generalizados R nos da medidas de devianza, la devianza nula y la desviación residual. La devianza nula muestra qué tan bien la variable de respuesta se predice mediante un modelo que incluye solo la intersección (AE, 2019). Los supuestos de un modelo de Poisson corresponde a la equidispersión Rstudio cuenta con la función glm que permite aplicar modelos lineales generalizados especificando en el argumento family de acuerdo a la distribución a aplicar poisson (link= "log") y quasipoisson (link= "log"), la distribución de Poisson tiene un único parámetro a

estimar, μ , la media, el cual a veces es llamado parámetro de localización. La principal característica de esta distribución es que la media y la varianza son iguales. Luego cuanto mayor sea la media mayor será la varianza o variabilidad en los datos, esta propiedad en la distribución de Poisson es llamada propiedad de equidispersión (Calzada, 2017).

Se debe agregar que un proceso de Poisson es un experimento aleatorio que consiste en observar la ocurrencia de eventos específicos sobre un soporte, tal que el proceso es estable el número de ocurrencias, λ es constante a largo plazo y los eventos ocurren de forma aleatoria e independiente. La distribución de Poisson se usa para modelar el número de eventos que ocurren en un proceso de Poisson. Sea $X \sim P(\lambda)$, esto es, una variable aleatoria con distribución de Poisson donde el número medio de eventos que ocurren en un determinado intervalo es λ (R Coder, 2020). El interés de la investigación es caracterizar y modelar las muertes en menores de un año, mediante modelos lineales generalizados a través de la distribución de Poisson, así como la importancia que radica el conocimiento con su aplicación.

Metodología

Esta investigación es descriptiva puesto que se aplicó un análisis exploratorio que permitió conocer la tasa de mortalidad y que causas de muerte son las más influyentes, también contará con una investigación correlacional, es decir, se realizará un análisis de correlación se utilizó una regresión de Poisson con número de eventos en menores de un año que va a permitir modelar estos datos de conteo y se verificará la medida de calidad del modelo mediante AIC, para verificar que tan adecuado es con estas bases simuladas y dar una conclusión sobre si aplicar o no este modelo a los datos de mortalidad infantil a nivel Ecuador. La presente investigación inició con el método cuantitativo de los datos de las defunciones en menores de un año proporcionados por el Instituto Nacional de Estadísticas y Censos (INEC), iniciando con un análisis descriptivo, luego un inferencial y por último multivariante tuvo un enfoque cuantitativo puesto que es una investigación empírica-analista, ya que utiliza datos para probar las hipótesis utilizando análisis estadístico descriptivo y analítico que permita establecer patrones de comportamiento para predecir la media de defunciones en menores de un año objeto de estudio. Los datos se obtuvieron del INEC dado que es una base de datos se consideró los registros del INEC del año 2019, considerando las 11 provincias de la región sierra que son: Azuay, Bolívar, Cañar, Carchi, Cotopaxi, Chimborazo Imbabura, Loja, Pichincha, Tungurahua y Santo Domingo de los Tsáchilas con un total de 119840

nacidos vivos y 1517 defunciones en menores de un año, no se aplicó ninguna técnica de muestreo. La base datos que se manejó para el estudio de modelos GLM son las defunciones de menores de un año del 2019 recolectadas por el Instituto Nacional de Estadísticas y Censos la información conto: provincias, causa de muerte, número de muertes, número de nacidos vivos; todo esto correspondiente a la región sierra del Ecuador.

Resultados

En esta sección se hace referencia los aspectos que se consideraron en la investigación para desarrollar el modelo generalizado Poisson.

Modelo Lineal Generalizado (Distribución Poisson)

$$Y_i = e^0 * e^{9.467 \text{ ProvinciaBolívar}} * e^{5.154 \text{ ProvinciaCañar}} * e^{8.248 \text{ ProvinciaCarchi}} * e^{2.884 \text{ ProvinciaChimborazo}} * e^{3.716 \text{ ProvinciaCotopaxí}} * e^{2.640 \text{ Provincialmbabura}} * e^{2.499 \text{ ProvinciaLoja}} * e^{0.002 \text{ ProvinciaPichincha}} * e^{1.611 \text{ ProvinciaSanto Domingo de los Tsáchilas}} * e^{1.925 \text{ ProvinciaTungurahua}} * e^{0.059 \text{ CM Causas mal definidas}} * e^{0.118 \text{ CM Influenza y Neumonía}} * e^{0.543 \text{ CM Malformaciones}} * e^{0.129 \text{ CM Obstrucción respiración}} * e^{0.073 \text{ CM Resros de causas}}$$

Tablas

Tabla 1. Tabla de contingencia variables Género y Causas de muerte número de muertes en menores de un año.

Causas de Muerte							
Género	Afecciónes	Obstrucción	Influenza	Resto	Causas	Total	
periodo	Malformaciones	y	de	mal			
respiratoria prenatal	neumonía	causas	definidas				
Hombre	440	210	48	53	22	24	797
Mujer	291	186	46	33	31	19	606
Total	731	396	94	86	53	43	1403

Se identificó que en hombres se presenta con mayor porcentaje la principal causa de muerte que es afecciones en periodo prenatal, seguida por malformaciones y se concluye que los hombres son más propensos a morir.

Tabla 2.

Variables Independientes	Estimaciones	Error Estándar	Valor Z	Valor p
		0.838		
I. Escalas a considerar en los ítems				
Intercepto	-7.702 -9.186 0.000			
ProvinciaBolívar	2.248	0.671	3.351	0.001
ProvinciaCañar	1.640	0.565	2.902	0.004
ProvinciaCarchi	2.110	0.698	3.022	0.003
ProvinciaChimborazo	1.059	0.377	2.810	0.005
ProvinciaCotopaxi	1.313	0.380	3.458	0.001
ProvinciaImbabura	0.971	0.391	2.479	0.013
ProvinciaLoja	0.916	0.355	2.579	0.010
ProvinciaPichincha	-6.343	2.136	-2.970	0.003
ProvinciaSanto Domingo de los Tsáchilas	0.477	0.262	1.818	0.069
ProvinciaTungurahua	0.655	0.298	2.200	0.028
CMuerteCausas mal definidas	-2.827	0.157	-18.017	0.000
CmuerteInfluenza y Neumonía	-2.135	0.114	-18.724	0.000
CmuerteMalformaciones	-0.610	0.062	-9.777	0.000
CmuerteObstrucción y Respiración	-2.046	0.110	-18.669	0.000
CmuerteResto de Causas	-2.618	0.142	-18.406	0.000
Nacidos	0.000	0.000	3.052	0.002

Se visualiza las estimaciones de los coeficientes del modelo de Poisson y se verificó que todas las variables son significativas debido a que tiene el valor p menor que el nivel de significancia de 0.05.

Tabla 3.

II. *Distribución de Poisson a partir de la estimación exponencial*

Variables Independientes	Estimaciones Exp(Estimación)	Interpretación	
Intercepto	-7.702	0.000	
ProvinciaBolívar	2.248	9.467	La media de muertes en menores de un año en Bolívar es 90.5% menos que la media de Azuay.
ProvinciaCañar	1.640	5.154	La media de muertes en menores de un año en Cañar es 94.8% menos que la media de Azuay.
ProvinciaCarchi	2.110	8.248	La media de muertes en menores de un año en Carchi es 91.8% menos que la media de Azuay.
ProvinciaChimborazo	1.059	2.884	La media de muertes en menores de un año en Chimborazo es 97.1% menos que la media de Azuay.
ProvinciaCotopaxi	1.313	3.716	La media de muertes en menores de un año en Cotopaxi es 96.3% menos que la media de Azuay.
ProvinciaImbabura	0.971	2.640	La media de muertes en menores de un año en Imbabura es 97.4% menos que la media de Azuay.
ProvinciaLoja	0.916	2.499	La media de muertes en menores de un año Loja es 97.5% menos que la media de Azuay.

ProvinciaPichincha	-6.343	0.002	La media de muertes en menores de un año en Pichincha es 99.998% más que la media de muertes de Azuay.
ProvinciaSanto Domingo de los Tsáchilas	0.477	1.611	La media de muertes en menores de un año en Santo Domingo de los Tsáchilas es 98.4% menos que la media de Azuay.
ProvinciaTungurahua	0.655	1.925	La media de muertes en menores de un año en Tungurahua es 98.1% menos que la media de Azuay.
CmuerteCausas mal definidas	-2.827	0.059	La media de muertes en menores de un año por muertes mal definidas aumenta 6% que la media de muertes por afecciones en el periodo prenatal.
CmuerteInfluenza y Neumonía	-2.135	0.118	La media de muertes en menores de un año por influenza y neumonía aumenta 11.8% que la media de muertes por afecciones en el periodo prenatal.
CmuerteMalformaciones	-0.610	0.543	La media de muertes en menores de un año por malformaciones aumenta 54.3% que la media de muertes por afecciones en el periodo prenatal.
CmuerteObstrucción y Respiración	-2.046	0.129	La media de muertes en menores de un año por obstrucción y respiración aumenta 12.9% que la media de muertes por afecciones en el periodo prenatal.

CmuerteResto de Causas	-2.618	0.073	La media de muertes en menores de un año por otras causas de muerte aumenta 7.3% que la media de muertes por afecciones en el periodo prenatal.
Nacidos	0.000	1.000	

Se observa que el modelo a partir de la estimación exponencial una comparación de medias entre las muertes en menores de un año de las regiones objeto de estudio.

Tabla 4.

III. *Análisis de Devianza para la variable muertes*

	Grados de libertad	Devianza	Devianza media
Modelo Nulo	131	1676.990	12.801
Modelo Residual	115	208.930	1.817
Total Corregido	210	4776.6	

En la tabla 3 se sintetiza la devianza de la bondad de ajuste del modelo generalizado aplicado en la investigación. Es así que el análisis de devianza, reportan que el modelo es estadísticamente significativo y concluyó lo siguiente: El modelo de regresión de Poisson tiene un pseudo- R^2 de 0,858 en comparación con el de regresión múltiple que tenía un R^2 de 0.966. la hipótesis referente al pseudo- R^2 , es que $H_0: R^2 = 0$ vs $H_1: R^2 \neq 0$, y H_0 es rechazada lo que involucra que los datos observados se ajustan al modelo de regresión de Poisson, además se cumple con la equidispersión ya que se obtuvo un valor de $\phi = 1.002$.

Etapas del modelo

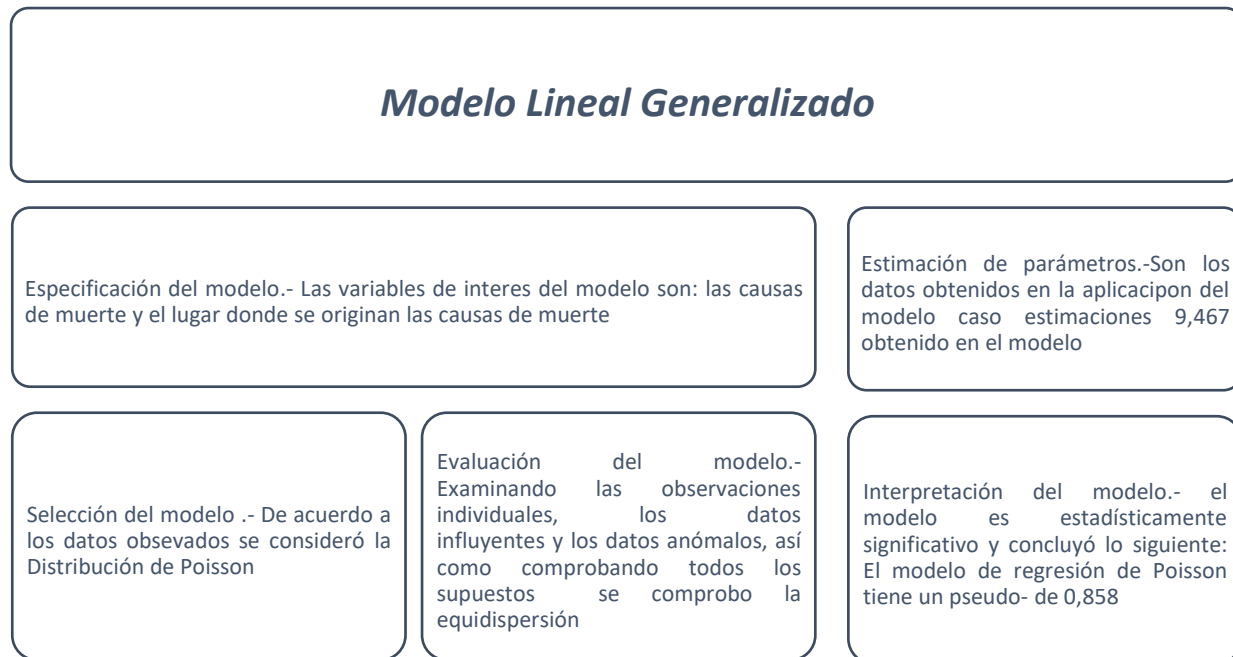


Figura 1. Etapas de la construcción del modelo

Discusión

Con el objetivo de dar continuidad a la presente investigación, es conveniente diseñar nuevos estudios, de modo que se capacite el aprendizaje de la distribución de Poisson, el hecho radica en la comparación que tome como referente el aprendizaje desarrollador de conceptos probabilísticos en el marco de la formación de los estudiantes. La enseñanza de contenidos estadísticos constituye un aspecto importante en la formación universitaria de los profesionales de pregrado. La modelización de cualquier situación real no es un proceso particular, sino que depende de muchos factores tales como calidad de datos, modelo teórico subyacente entre otras. Con esta concentración se han mostrado las posibilidades y limitaciones de la aplicación de técnicas estadísticas para datos de conteo en el campo de la vida real.

Conclusiones

Las simulaciones comparativas permiten verificar si el estudio realizado se puede utilizar a nivel de todo el país y se concluye que el modelo puede utilizar datos de mortalidad infantil de menores de un año de todo el Ecuador con regresión de Poisson. Los modelos lineales generalizados son una amplificación de los modelos lineales que permiten modelar variable que pertenecen a la familia de distribución de probabilidad exponencial, la misma que contiene distribuciones normal,

binomial, poisson y binomial negativa sin embargo se debe considerar el tipo de datos que se va a manipular para emplear el mejor modelo que contribuirá al análisis que deseamos obtener con el propósito de que el modelo propuesto es el indicado para la modelación de los datos y evitar uno de los supuestos que es la equidispersión.

Finalmente, las simulaciones comparativas son rápidas de realizar, permiten tomar decisiones adecuadas, razonadas, y en el tiempo que se requieren.

Referencias

- Rodríguez Rodríguez, M. (2022). Importancia de la distribución binomial y de Poisson. Colombia: Corporación Universitaria del Meta .
- AE. (9 de febrero de 2019). rstudio-pubs-static.s3.amazonaws.com. Obtenido de https://rstudio-pubs-static.s3.amazonaws.com/465671_ee66d8ea51e84ba79554f6dd807d2189.html#predicci%C3%B3n-de-probabilidades
- Arguello, V. (2022). Modelos lineales generalizados para el análisis de defunciones en menores de un año en la región Sierra del Ecuador. Riobamba : Escuela Superior Politécnica de Chimborazo .
- Calzada, P. A. (2017). Modelos de regresión con datos de conteo. Aplicación a competiciones deportivas . Sevilla: Universidad de Sevilla .
- Freund , J. E., Miller, I. R., & Johnson, R. (2006). Probabilidades y Estadística para Ingenieros. Habana : Félix Varela.
- R coder. (2020). R CODER . Obtenido de https://r-coder.com/distribucion-poissonr/#La_distribucion_de_Poisson
- Ramírez, M. C., Álvarez Reyes, S. E., & Pérez Santos, F. J. (2013). Sobre la Enseñanza de la Distribución de Poisson en Carreras de Ingeniería. Bolema, 1117-1134.
- READING, C., & REID, J. (2013). An emerging hierarchy of reasoning about distribution: From a variation perspective. Statistics Education Research Journal, 46-68.