

APLICACIÓN DEL MODELO LINEAL GENERALIZADO (GLM) EN LA BIOLOGÍA: UN ENFOQUE VERSÁTIL PARA EL ANÁLISIS DE DATOS COMPLEJOS

JULIO TRUJILLO GONZÁLEZ
Universidad de Panamá,
julio.trujillo@up.ac.pa
orcid.org/0000-0002-3664-8058
DOI: 10.61209/re.v2i1.39

ABRAHAM DE SEDAS
Universidad de Panamá,
abraham.desedas@up.ac.pa
orcid.org/0009-0006-7170-8649

Resumen.

El Modelo Lineal Generalizado (GLM) es una extensión de los modelos lineales que permite modelar relaciones entre variables con distribuciones no necesariamente normales. Los GLM son útiles en matemáticas aplicadas, permitiendo modelar y entender relaciones entre variables en situaciones complejas. En Latinoamérica, los GLM se han aplicado en campos como ecología, epidemiología y economía. Los GLM se componen de tres elementos: la distribución de la variable de respuesta, el predictor lineal y la función de enlace. Sin embargo, la interpretación de los resultados puede ser compleja y si los supuestos del modelo no se cumplen, las estimaciones pueden ser sesgadas. Un estudio de Juan Carlos Di Trani utilizó GLM para analizar el comportamiento de forrajeo de las abejas en relación con varios factores ambientales, proporcionando información valiosa para la gestión de los cultivos y la protección de las abejas.

Palabras Clave: Modelo lineal generalizado, matemáticas aplicadas, comportamiento de forrajeo

ABSTRACT.

The Generalized Linear Model (GLM) is an extension of ordinary linear models that allows for modeling relationships between variables where the dependent variable does not necessarily follow a normal distribution. GLMs are useful in applied mathematics, enabling the modeling and understanding of relationships between variables in complex situations. In Latin America, GLMs have been applied in fields such as ecology, epidemiology, and economics. GLMs consist of three components: the distribution of the response variable, the linear predictor, and the link function. However, the interpretation of results can be complex and if the assumptions of the model are not met, estimates can be biased. A study by Juan Carlos Di Trani used GLMs to analyze the foraging behavior of bees in relation to various environmental factors, providing valuable information for crop management and bee protection.

Key Words: Generalized linear model, applied mathematics, foraging behavior

Introducción.

El Modelo Lineal Generalizado (GLM, por sus siglas en inglés) es una herramienta fundamental en el campo de la matemática aplicada, que ha encontrado una amplia gama de aplicaciones en diversas disciplinas científicas (Wang et al., 2022). Los GLM son una extensión de los modelos lineales ordinarios que permiten la modelización de relaciones entre variables donde la variable dependiente no sigue necesariamente una distribución normal (Paderewski & Rodrigues, 2018). Este logro se realiza a través de la introducción de una función de enlace que transforma la variable dependiente, permitiendo la modelización de una variedad de distribuciones de datos (Kohyama et al., 2020).

Los GLM son útiles para una variedad de propósitos en matemáticas aplicadas. Permiten a los matemáticos y científicos modelar y entender las relaciones entre variables en situaciones complejas. Su verdadera fuerza radica en su capacidad para manejar una variedad de tipos de datos y distribuciones, lo que los hace particularmente útiles en situaciones donde los datos no se ajustan a las suposiciones de los modelos lineales ordinarios (Dunlap & Willis, 2022).

En Latinoamérica, los GLM han encontrado aplicaciones en una variedad de campos. Por ejemplo, en la ecología, se han utilizado para modelar la distribución de especies en respuesta a factores ambientales (Wang et al., 2022). En la epidemiología, se han aplicado para entender la propagación de enfermedades infecciosas. En la economía, se han utilizado para modelar la relación entre variables económicas.

En las ciencias naturales y la ingeniería, los GLM también han demostrado ser valiosos. En la biología, se han utilizado para modelar la relación entre rasgos genéticos y fenotípicos. En la ingeniería, se han aplicado en áreas como la optimización de procesos y el control de calidad.

Los Modelos Lineales Generalizados son una herramienta matemática versátil y poderosa que ha encontrado aplicaciones en una amplia gama de disciplinas y contextos. A lo largo de este artículo, exploraremos en profundidad la teoría matemática detrás de los GLM, así como sus aplicaciones en Latinoamérica y en las ciencias naturales e ingeniería (Kohyama et al., 2020).

Método

El Modelo Lineal Generalizado (GLM) se basa en la regresión lineal, pero es más flexible ya que permite que la variable dependiente tenga una distribución que no sea normal y que la relación entre la variable dependiente y las variables independientes sea no lineal (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989). Un GLM se compone de tres elementos principales: la distribución de la variable de respuesta, el predictor lineal y la función de enlace. Estos se combinan en la siguiente ecuación general: Distribución de la variable de respuesta (Y): En un GLM, la variable de respuesta Y sigue alguna distribución de la familia exponencial, que incluye, pero no se limita a, la distribución normal, binomial, Poisson, gamma, entre otras.

Predictor lineal (η): El predictor lineal es una combinación lineal de los predictores o variables independientes (X) y sus coeficientes asociados (β). Se calcula como:

$$\eta = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + e$$

$$E(e) = 0$$

$$e \sim N(0, \sigma^2) \quad \eta \sim N(\mu, \sigma^2)$$

donde X es la matriz de predictores, e es el error que se distribuye normalmente, β es el vector de coeficientes, $\mu = E(\eta)$ es la esperanza matemática de η , y p es el número de predictores.

Función de enlace (g): La función de enlace proporciona la relación entre el predictor lineal y la media de la distribución de la variable de respuesta. La ecuación de la función de enlace es:

$$g(\mu) = \eta$$

La elección de la función de enlace depende de la distribución de la variable de respuesta. Por ejemplo, para una distribución binomial (como en la regresión logística), se utiliza la función logit; para una distribución de Poisson (como en la regresión de Poisson), se utiliza la función logarítmica.

Las ventajas de los GLM incluyen su flexibilidad para manejar diferentes tipos de variables de respuesta y relaciones no lineales, así como su capacidad para proporcionar estimaciones de los efectos de las variables predictoras en la variable de respuesta en la escala original de medición (Wang et al., 2017; Xiang & Zhu, 2011; Rodríguez & Goldman, 1995).

Sin embargo, los GLM también tienen desventajas. Por ejemplo, la interpretación de los resultados puede ser más compleja que en una regresión lineal, especialmente cuando se utiliza una función de enlace no lineal. Además, si los supuestos del modelo no se cumplen, las estimaciones pueden ser sesgadas o ineficientes (Wang et al., 2017; Xiang & Zhu, 2011; Rodríguez & Goldman, 1995).

Resultados

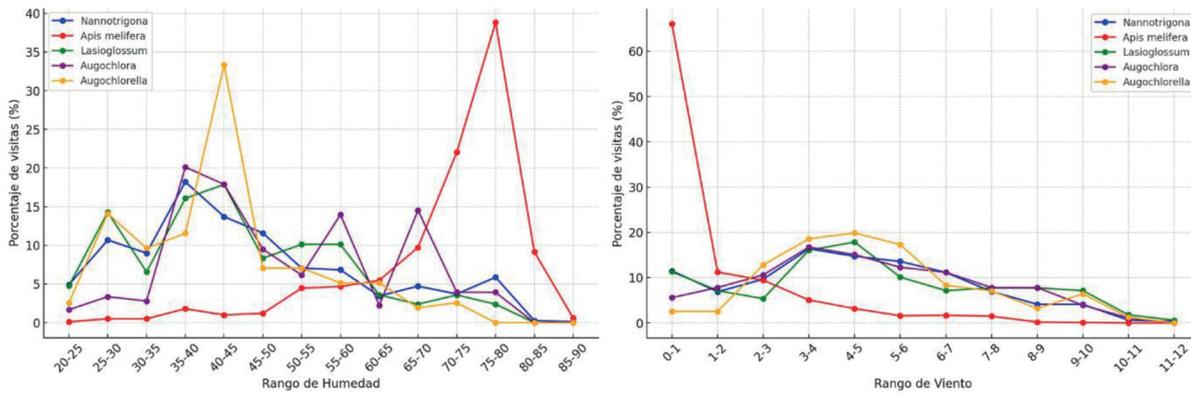
Los estudios realizados por Juan Carlos Di Trani se centran en el comportamiento de forrajeo de las abejas, específicamente en relación con las flores de melón (*Cucumis melo*) y sandía (*Citrullus lanatus*) en Panamá. Los patrones de forrajeo de las abejas se registraron en función de varios factores ambientales, incluyendo el día, la hora, la temperatura, la iluminación, la humedad y la velocidad del viento.

En su estudio, Di Trani observó cuatro especies de abejas: *Nannotrigona*, *Apis mellifera*, *Lasioglossum*, *Augochlora* y *Augochlorella*. Descubrió que las abejas recolectaban principalmente néctar, con la *Apis mellifera* dedicando casi la mitad de sus visitas a la recolección de polen. Los patrones de forrajeo variaban según la especie de abeja y las condiciones ambientales, con las abejas mostrando preferencias por ciertas condiciones.

Para analizar estos datos, determinaron que un modelo lineal generalizado (GLM) podría ser útil. Este modelo podría incluir las variables del día, la hora, la temperatura, la iluminación, la humedad y la velocidad del viento como predictores de la actividad de forrajeo de las abejas. Las especies de abejas podrían considerarse como una variable categórica en el modelo. Este enfoque permitiría examinar cómo cada uno de estos factores influye en el comportamiento de forrajeo de las abejas y podría proporcionar información valiosa para la gestión de los cultivos y la protección de las abejas (Di Trani, 2007; Di Trani et al., 2022; 2023).

Figura 2.

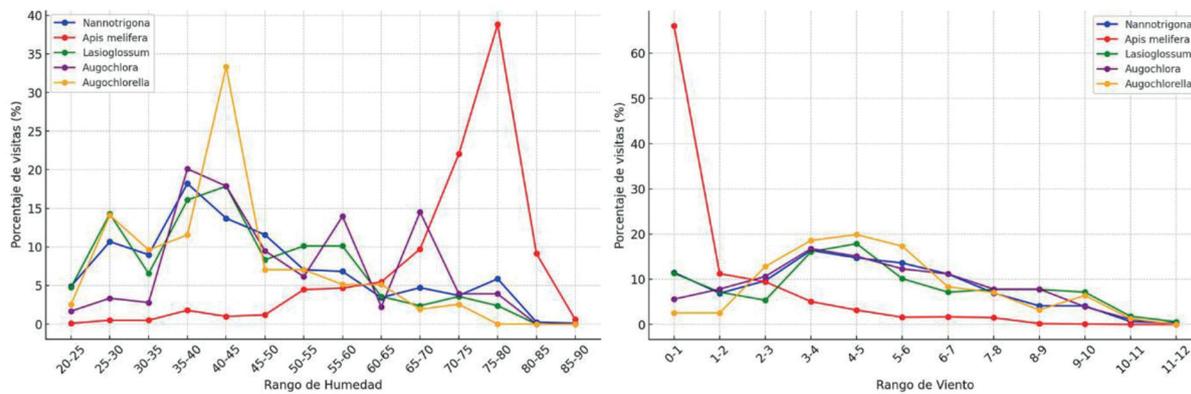
Porcentaje de visitas para las condiciones ambientales: Humedad y Viento



Fuente. (Di Trani et al.,2022)

Figura 2.

Porcentaje de visitas para las condiciones ambientales: Humedad y Viento



Fuente. (Di Trani et al.,2022)

Los gráficos que hemos creado muestran cómo las visitas de cada especie de abeja varían con diferentes condiciones ambientales. Aquí hay un resumen de las tendencias observadas para cada especie de abeja:

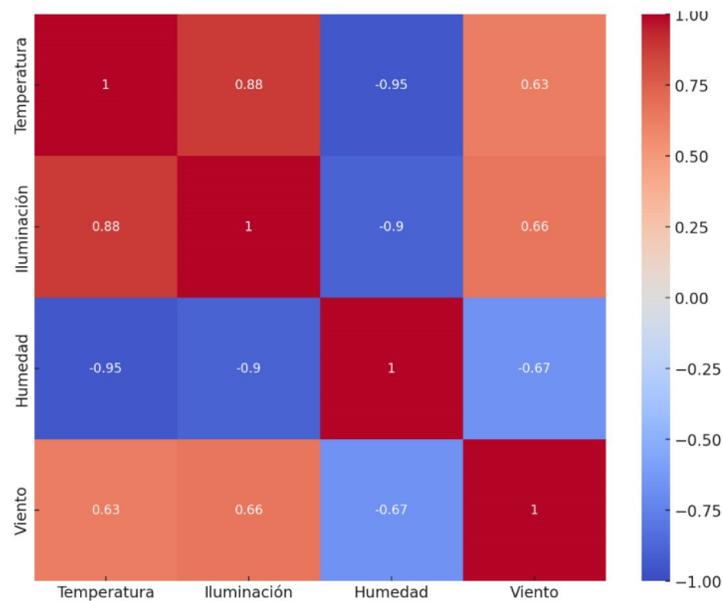
- **Nannotrigona:** Esta especie tiende a visitar más a medida que aumenta la temperatura, con un pico en el rango de 40-42 grados. En cuanto a la iluminación, las visitas de Nannotrigona aumentan significativamente a medida que aumenta la iluminación, con un pico en el rango de 102500-115000. En cuanto a la humedad, las visitas parecen ser bastante consistentes a través de los rangos de humedad, con un ligero pico en el rango de 55-60. En cuanto al viento, las visitas aumentan con la velocidad del viento, con un pico en el rango de 8-9.
- **Apis melifera:** Esta especie prefiere temperaturas más bajas, con la mayoría de las visitas ocurriendo en el rango de 24-26 grados. En cuanto a la iluminación, las visitas de Apis melifera disminuyen a medida que aumenta la iluminación. En cuanto a la humedad, las visitas aumentan significativamente a medida que aumenta la humedad, con un pico en el rango de 75-80. En cuanto al viento, la mayoría de las visitas ocurren en el rango de 0-1, y disminuyen a medida que aumenta la velocidad del viento.



- **Lasioglossum:** Esta especie tiende a visitar más a medida que aumenta la temperatura, con un pico en el rango de 38-40 grados. En cuanto a la iluminación, las visitas de *Lasioglossum* aumentan significativamente a medida que aumenta la iluminación. En cuanto a la humedad, las visitas parecen ser bastante consistentes a través de los rangos de humedad. En cuanto al viento, las visitas aumentan con la velocidad del viento, con un pico en el rango de 8-9.
 - **Augochlora:** Esta especie tiende a visitar más a medida que aumenta la temperatura, con un pico en el rango de 38-40 grados. En cuanto a la iluminación, las visitas de *Augochlora* aumentan significativamente a medida que aumenta la iluminación. En cuanto a la humedad, las visitas aumentan a medida que aumenta la humedad, con un pico en el rango de 65-70. En cuanto al viento, las visitas aumentan con la velocidad del viento, con un pico en el rango de 7-8.
 - **Augochlorella:** Esta especie tiende a visitar más a medida que aumenta la temperatura, con un pico en el rango de 38-40 grados. En cuanto a la iluminación, las visitas de *Augochlorella* aumentan significativamente a medida que aumenta la iluminación. En cuanto a la humedad, las visitas parecen ser bastante consistentes a través de los rangos de humedad. En cuanto al viento, las visitas aumentan con la velocidad del viento, con un pico en el rango de 8-9.
- Por lo tanto, podemos concluir que cada especie de abeja tiene diferentes preferencias en cuanto a las condiciones ambientales. Algunas especies, como *Nannotrigona* y *Augochlorella*, parecen preferir condiciones más extremas (altas temperaturas, alta iluminación, vientos fuertes), mientras que otras especies, como *Apis mellifera*, parecen preferir condiciones más moderadas (temperaturas más bajas, baja iluminación, vientos suaves).

Figura 3.

Correlación de las variables



La matriz de correlación de las condiciones ambientales, representada como un mapa de calor. Los valores en el mapa de calor representan los coeficientes de correlación de Pearson, que miden la relación lineal entre dos conjuntos de datos. Los valores pueden variar entre -1 y 1, donde 1 significa una correlación positiva perfecta, -1 una correlación negativa perfecta, y 0 ninguna correlación.



Observaciones clave:

Temperatura e Iluminación: Hay una fuerte correlación positiva (0.88) entre la temperatura y la iluminación. Esto sugiere que cuando la temperatura es alta, la iluminación también tiende a ser alta.

Temperatura y Humedad: Hay una fuerte correlación negativa (-0.95) entre la temperatura y la humedad. Esto sugiere que cuando la temperatura es alta, la humedad tiende a ser baja.

Viento y las demás variables: La velocidad del viento muestra una correlación moderada con las demás variables, siendo positiva con la temperatura y la iluminación, y negativa con la humedad. Estos resultados tienen sentido dado que las condiciones ambientales a menudo están interrelacionadas. Por ejemplo, durante un día caluroso y soleado (alta temperatura e iluminación), es probable que la humedad sea baja. Además, es común que los días ventosos estén asociados con temperaturas y niveles de iluminación más altos.

Es importante tener en cuenta que estas correlaciones son puramente estadísticas y no implican causalidad. Además, la correlación entre dos variables no significa necesariamente que una variable afecte directamente a la otra, ya que puede haber otras variables no observadas que estén influyendo en ambas.

Teniendo en consideración lo anterior Di Trani et al. (2022) consideraron las variables: temperatura, iluminación, viento en el modelo GLM, obteniendo lo siguiente:

Tabla 1.

Resultados del GLM

	Nannotrigona_ coef	Nannotrigona_p_v alue	Apis_melifera_ coef	Apis_melifera_p _value
Intercept	2.108299421	2.65332E-49	6.153097503	1.64262E-62
Temperatura	-0.015463657	0.003210769	-0.147992404	1.94383E-24
Iluminación	1.02613E-06	0.149813879	-1.2013E-05	5.01501E-09
Viento	0.045841369	2.34689E-08	-0.135547891	7.9168E-09

Fuente. (Di Trani et al., 2022)

Tabla 2.

Resultados del GLM

	Lasioglossum_ coef	Lasioglossum _p_value	Augochlora_ coef	Augochlora_p_v alue	Augochloraella_ coef	Augochloraella_ p_value
Intercept	0.409835456	0.517287968	1.113969918	0.079796947	-1.01709884	0.11366787
Temperatura	-0.071597825	0.002345432	0.085378848	0.000343229	-0.019899416	0.396580384
Iluminación	7.54771E-06	0.015480736	-2.22679E-07	0.942686875	4.34608E-07	0.891521831
Viento	0.098203029	0.004753713	0.204898746	7.999E-10	0.118241351	0.00100657

Fuente. (Di Trani et al., 2022)

En general, estos resultados muestran cómo las visitas de cada especie de abeja están asociadas con la temperatura, la iluminación y el viento. Por ejemplo, para la especie *Apis melifera*, los coeficientes negativos para la temperatura y el viento sugieren que las visitas de esta especie disminuyen a medida que la temperatura y la velocidad del viento aumentan. En cambio, para la especie *Nannotrigona*, el coeficiente positivo para el viento sugiere que las visitas de esta especie aumentan a medida que la velocidad del viento aumenta.



Conclusiones

El Modelo Lineal Generalizado (GLM) es una herramienta versátil que ha encontrado aplicaciones en una amplia gama de disciplinas científicas. Su capacidad para modelar relaciones entre variables donde la variable dependiente no sigue necesariamente una distribución normal lo hace particularmente útil en situaciones complejas (Bower, 2009).

En el campo de las matemáticas aplicadas, los GLM se han utilizado para resolver problemas en la mecánica de sólidos, demostrando su amplia aplicabilidad (Bower, 2009). En el ámbito de la informática, los GLM se han empleado en el procesamiento de imágenes y sistemas de visión inteligente, destacando su utilidad en el manejo de diversos tipos de datos (Blanc-Talon, n.d.).

En el campo de la hidrología, los GLM se han utilizado para analizar tendencias en los datos, demostrando su potencial para manejar grandes y complejos conjuntos de datos (Clarke, 2002).

En biología, los GLM se han aplicado en la predicción del número de ovocitos necesarios para una fertilización in vitro exitosa, mostrando su potencial en el campo de la biología reproductiva (Esteves et al., 2020).

Referencias

- Blanc-Talon, J. (n.d.). *Advanced Concepts for Intelligent Vision Systems*. This book includes a discussion on the use of GLMs in the field of image processing and vision systems.
- Bower, A. (2009). *Applied Mechanics of Solids*. This work discusses the application of mathematical models, including GLMs, in the field of solid mechanics.
- Clarke, R. (2002). Estimating trends in data from the Weibull and a generalized extreme value distribution. This paper discusses the use of GLMs for analyzing trends in hydrological data.
- Di Trani de la Hoz, J. (2007). Visita de abejas (*Apis mellifera*, Hymenoptera: Apoidea) a flores de melón *Cucumis melo* (Cucurvitaceae) en Panamá. *Revista de Biología Tropical*, 55(2), 677-680.
- Di Trani, J. C., Ramírez, V. M., Añino, Y., & Barba, A. (2022). Environmental conditions and bee foraging on watermelon crops in Panama. *Journal of Animal Behaviour and Biometeorology*, 10(4), 2234-2234.
- Di Trani, J. C., Ramírez, V. M., Barba, A., & Añino, Y. (2023). Foraging patterns of bees on watermelon (*Citrullus lanatus* Thunb.) flowers in Panama. *Journal of Animal Behaviour and Biometeorology*, 11(3), 2023022-2023022.
- Dunlap, D., & Willis, R. (2002). *Understanding And Utilizing Adjunct Professors For Non Traditional Engineering And Technology Graduate Education*.
- Esteves, S., Yarali, H., Ubaldi, F., Carvalho, J. F., Bento, F. C., Vaiarelli, A., Cimadomo, D., Özbek, I. Y., Polat, M., Bozdog, G., Rienzi, L., & Alviggi, C. (2020). Validation of ART Calculator for Predicting the Number of Metaphase II Oocytes Required for Obtaining at Least One Euploid Blastocyst for Transfer in Couples Undergoing in vitro Fertilization/Intracytoplasmic Sperm Injection. This paper discusses the use of GLMs in predicting the number of oocytes required for successful in vitro fertilization.
- Kohyama, M., Tanaka, S., & Shiihara, Y. (2020). *Ab Initio Local-Energy and Local-Stress Calculations for Materials Science and Engineering*.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. CRC press.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370-384.
- Paderewski, J., & Rodrigues, P. (2018). *Constrained AMMI Model: Application to Polish Winter Wheat Post-Registration Data*.
- Rodríguez, G., & Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 158(1), 73-89.
- Wang, N., Lombardo, L., Cheng, W., Marconcini, M., Bachofer, F., Guo, L., & Xiong, J. (2022). Space-time susceptibility modeling of hydro-morphological processes at the Chinese national scale.
- Xiang, R., & Zhu, J. (2011). Generalized linear mixed models for longitudinal data with informative dropouts. *Journal of Statistical Planning and Inference*, 141(1), 423-437. on.

