

ANÁLISIS DE INCIDENTES DE ACTIVIDAD CRIMINAL EN COLOMBIA (2023) USANDO MODELOS DE REGRESIÓN PARA DATOS DE CONTEO

Analysis of criminal activity incidents in Colombia (2023) using count data regression models

Laura Valentina Montes Montes.^a
laura.montes@usantotomas.edu.co

Wilmer Darío Pineda Ríos Ph.D.^b
wpinedar@unal.edu.co

RESUMEN

El presente trabajo aborda un análisis estadístico de los hurtos a comercio en Colombia, utilizando datos de los 32 departamentos del país. El objetivo principal es identificar y modelar los factores económicos, sociales y espaciales que explican la incidencia de este tipo de delito, con el fin de proporcionar herramientas analíticas para la formulación de políticas públicas más efectivas.

El análisis parte de una exploración inicial de los datos, en la que se identificaron patrones espaciales significativos de los hurtos, confirmados mediante el índice de Moran positivo y estadísticamente significativo, lo que sugiere dependencia espacial entre departamentos. A partir de este diagnóstico, se implementaron Modelos de Regresión Poisson y Binomial Negativo, ajustados por población como variable de exposición, para modelar tasas de hurtos en lugar de conteos absolutos, facilitando la comparabilidad entre regiones con diferentes tamaños poblacionales.

El Modelo Poisson parsimonioso demostró ser una herramienta sólida para el análisis, pero la presencia de sobre-dispersión en los datos justificó la implementación del modelo binomial negativo, que incluye un parámetro adicional para capturar la variabilidad excedente. Los resultados de ambos modelos identificaron como factores clave la reducción de hurtos a comercio: el PIB, los niveles de pobreza monetaria, la tasa de criminalidad, y un efecto espacial medido por el lag de hurtos en departamentos vecinos. En particular, el PIB mostró un efecto negativo y estadísticamente significativo, destacando el papel del desarrollo económico en la mitigación de delitos. El término de Lag Hurtos resaltó la importancia de los efectos espaciales en la propagación o contención de los hurtos, indicando que dinámicas departamentales compartidas influyen significativamente en los resultados observados.

La evaluación de los residuos de ambos modelos, mediante gráficas y análisis del índice de Moran, confirmó la ausencia de autocorrelación espacial en los residuos, validando la especificación estadística y espacial de los modelos ajustados. Además, el modelo binomial negativo mostró un ajuste superior según el *AIC* y la inclusión de sobre-dispersión ($\alpha = 0.1149$).

En conclusión, el estudio evidencia la importancia de factores económicos y espaciales en la explicación de los hurtos a comercio, destacando la necesidad de políticas regionales coordinadas y enfoques basados en el desarrollo económico para abordar la problemática. Se recomienda la implementación de estrategias

^aEstudiante de Estadística Universidad Santo Tomás Bogotá

^bDirector de Trabajo de grado, Facultad de Estadística Universidad Santo Tomás Bogotá

conjuntas entre departamentos vecinos y el uso continuo de modelos avanzados para monitorear y evaluar los patrones delictivos. Este trabajo aporta un marco analítico robusto y replicable para el análisis de fenómenos criminológicos en contextos espaciales.

ABSTRACT

This study presents a statistical analysis of thefts from businesses in Colombia, using data from the country's 32 departments. The primary objective is to identify and model the economic, social, and spatial factors that explain the incidence of this type of crime, with the aim of providing analytical tools for the formulation of more effective public policies.

The analysis begins with an initial exploration of the data, during which significant spatial patterns of thefts were identified, confirmed by a positive and statistically significant Moran's Index. This finding suggests spatial dependence between departments. Based on this diagnostic, Poisson and Negative Binomial Regression Models were implemented, adjusted for population as an exposure variable, to model theft rates instead of absolute counts, thereby facilitating comparability across regions with different population sizes.

The parsimonious Poisson Model proved to be a robust tool for analysis; however, the presence of over-dispersion in the data justified the implementation of the Negative Binomial Model, which includes an additional parameter to capture excess variability. The results of both models identified key factors influencing the reduction of thefts from businesses: GDP, levels of monetary poverty, crime rates, and a spatial effect measured by the lag of thefts in neighboring departments. In particular, GDP showed a negative and statistically significant effect, underscoring the role of economic development in crime mitigation. The lagged thefts term highlighted the importance of spatial effects in the propagation or containment of thefts, indicating that shared departmental dynamics significantly influence the observed outcomes.

The evaluation of residuals from both models, through graphs and Moran's Index analysis, confirmed the absence of spatial autocorrelation in the residuals, validating the statistical and spatial specifications of the adjusted models. Moreover, the Negative Binomial Model demonstrated superior fit according to the *AIC* and the inclusion of over-dispersion ($\alpha = 0.1149$).

In conclusion, the study underscores the importance of economic and spatial factors in explaining thefts from businesses, highlighting the need for coordinated regional policies and approaches based on economic development to address this issue. The implementation of joint strategies between neighboring departments and the continuous use of advanced models to monitor and evaluate crime patterns are recommended. This work provides a robust and replicable analytical framework for the study of criminological phenomena in spatial contexts.

Índice

RESUMEN	1
ABSTRACT	3
OBJETIVOS	6
Objetivo General	6
Objetivos Específicos	6
JUSTIFICACIÓN Y ANTECEDENTES	7
INTRODUCCIÓN	9
1. MODELOS LINEALES GENERALIZADOS	10
1.1. Componentes Fundamentales de los GLM	10
1.1.1. Distribución de la variable dependiente	10
1.1.2. Función de enlace	11
1.1.3. Predictor Lineal	11
1.2. Estimación de Parámetros	11
1.2.1. Log-Verosimilud	12
1.2.2. Estimación iterativa	12
1.3. Varianza de la Variable Dependiente	12
1.4. Diagnóstico y Bondad de Ajuste	12
1.4.1. Residuales deviance	12
1.4.2. Pseudo- R^2	13
1.4.3. Criterio de información AIC	14
1.4.4. Criterio de información BIC	15
2. MODELOS DE REGRESIÓN PENALIZADA	17
2.1. Modelo Lasso	17
3. MODELOS PREDICTIVOS PARA DATOS DE CONTEO	18
3.1. Modelo de Regresión Poisson	18
3.1.1. Distribución de Probabilidad Discreta Poisson	18
3.1.2. La Distribución Poisson como un caso particular de la Familia Exponencial	19
3.1.3. Modelo de Regresión Poisson	20
3.1.4. Equidispersión	21
3.2. Modelo de Regresión Binomial Negativa	22

Análisis de Incidentes Criminales en Colombia (2023)	5
3.2.1. Distribución de Probabilidad Binomial Negativa	23
3.2.2. La Distribución Binomial Negativa como un caso particular de la Familia Exponencial	23
3.2.3. Modelo de Regresión Binomial Negativa	25
4. APLICACIÓN	28
4.1. Datos y Análisis Preliminar de Variables	28
4.1.1. Tratamiento de Datos Faltantes	29
4.1.2. Análisis Descriptivo y Exploratorio	31
4.2. Validación de Correlación Espacial	34
4.3. Selección de Variables Predictoras	36
4.4. Modelo Poisson con Efectos Espaciales	37
4.4.1. Implementación	37
4.4.2. Validación del Supuesto de Equisdispersión	40
4.5. Modelo Binomial Negativa	42
CONCLUSIÓN Y RECOMENDACIÓN	44
Conclusiones	44
Recomendaciones	44
APÉNDICE	45
Apéndice A	45
Apéndice B	46
REFERENCIAS BIBLIOGRÁFICAS	53

OBJETIVOS

Estudiar y presentar las propiedades de los modelos de regresión para datos de conteo, como los modelos Poisson y Binomial Negativa, y aplicarlos al análisis de los factores asociados a los hurtos a comercio en Colombia, incorporando estudios de diagnóstico y dependencia espacial.

1. **Revisar la literatura** sobre modelos de regresión para datos de conteo, incluyendo los modelos Poisson y Binomial Negativa, con un enfoque en su aplicación a fenómenos de sobre-dispersión y dependencia espacial.
2. **Evaluar y presentar las propiedades** de los modelos de regresión para datos de conteo:
 - Comparar el ajuste y los supuestos de los modelos Poisson y Binomial Negativa.
 - Incorporar herramientas de diagnóstico para identificar problemas de sobre-dispersión y dependencia espacial en los datos.
3. **Implementar los métodos de estimación clásica** para los modelos Poisson y Binomial Negativa, evaluando su desempeño en la identificación de los factores asociados a los hurtos a comercio.
4. **Estudiar la dependencia espacial** de los hurtos a comercio mediante la inclusión de términos de lag espacial en los modelos, evaluando su contribución al ajuste y la interpretación de los resultados.
5. **Realizar una aplicación** del modelo de regresión para datos de conteo al análisis de los hurtos a comercio en Colombia, incluyendo estudios de diagnóstico para validar los supuestos del modelo.

JUSTIFICACIÓN Y ANTECEDENTES

La utilización de clases de modelos lineales generalizados (GLM), como la regresión Poisson y la Binomial Negativa, en el análisis de incidentes de actividad criminal en Colombia proporciona una herramienta estadística robusta para comprender y modelar estos fenómenos. Su relevancia radica en los siguientes aspectos:

1. Modelado estadístico de eventos discretos:

Los incidentes de actividad criminal son variables de naturaleza discreta, y los GLM permiten modelar directamente la relación entre estos conteos y factores explicativos, capturando la estructura subyacente de los datos (Agresti, 2015).

2. Manejo de sobredispersión:

Mientras que el modelo Poisson asume una igualdad entre la media y la varianza, la regresión Binomial Negativa extiende este enfoque al abordar casos donde la varianza excede la media, lo cual es común en datos de criminalidad (Cameron & Trivedi, 2013).

3. Evaluación de factores explicativos:

Estos modelos permiten identificar las variables asociadas con la frecuencia de los incidentes, como características demográficas, económicas y espaciales, proporcionando una base para comprender las dinámicas delictivas (Hilbe, 2014).

4. Predicción y diagnóstico:

Los GLM facilitan predicciones ajustadas al contexto y permiten evaluar la efectividad de intervenciones pasadas o diseñar estrategias futuras con base en patrones observados (Dobson & Barnett, 2018).

5. Flexibilidad y generalización:

Al pertenecer a una clase amplia de modelos, los GLM pueden adaptarse a diferentes distribuciones de datos y contextos, ampliando su aplicabilidad más allá del caso específico de los incidentes de actividad criminal (McCullagh & Nelder, 1989).

6. Apoyo a la toma de decisiones:

Estos modelos proporcionan información cuantitativa fundamentada para priorizar intervenciones, asignar recursos y diseñar políticas públicas orientadas a la prevención del delito (Agresti, 2015).

7. Contribución metodológica:

La integración de diferentes clases de GLM, como Poisson y binomial negativa, ofrece una perspectiva robusta que combina rigor estadístico y aplicabilidad práctica, permitiendo un análisis más profundo y adaptado a las características de los datos (Cameron & Trivedi, 2013).

La aplicación de modelos estadísticos avanzados como los GLM no solo permite entender las dinámicas subyacentes de los eventos criminales, sino que también abre camino para analizar problemas más específicos que afectan la seguridad ciudadana y el desarrollo económico. Un caso particular que requiere atención es la inseguridad en locales comerciales, un fenómeno que refleja y agrava los desafíos más amplios de la criminalidad en Colombia, como lo son:

1. Generar un impacto positivo en los ámbitos Económico y Social:

El impacto económico de los robos y actos de violencia en locales comerciales es significativo. Los costos directos e indirectos incluyen pérdidas de inventario, daños a la propiedad, aumento en las primas de seguros y costos adicionales en medidas de seguridad. Además, estos incidentes pueden afectar negativamente la percepción de seguridad entre los consumidores, disminuyendo la afluencia de clientes y afectando las ventas y la rentabilidad de los negocios (Rincón & Sánchez, 2020).

2. Desarrollar Estrategias Eficaces de Mitigación:

La identificación y evaluación cuantitativa de los riesgos son pasos fundamentales para desarrollar es-

estrategias de mitigación efectivas. Un análisis integral del riesgo permite a los propietarios de locales comerciales y a las autoridades locales comprender mejor la naturaleza y magnitud de los riesgos a los que se enfrentan, facilitando la implementación de medidas preventivas y correctivas adecuadas. Estrategias como la mejora de la infraestructura de seguridad, la capacitación del personal en manejo de crisis y la colaboración con fuerzas de seguridad pueden reducir significativamente el riesgo de incidentes (Gómez et al., 2019).

3. Contribuir al Desarrollo Sostenible:

La seguridad es un componente esencial para el desarrollo sostenible de las ciudades. Garantizar la seguridad de los locales comerciales no solo protege los activos y la integridad de los comerciantes y clientes, sino que también fomenta un entorno propicio para el crecimiento económico y social. La implementación de estrategias de mitigación del riesgo contribuye a la estabilidad y prosperidad de las comunidades, incentivando el emprendimiento y la inversión (Montoya & Ruiz, 2021).

La inseguridad en locales comerciales en Colombia ha alcanzado niveles preocupantes, impactando no sólo a los dueños de negocios, sino también a empleados y clientes. Este problema tiene profundas implicaciones económicas y sociales, y su comprensión es fundamental para desarrollar estrategias efectivas de mitigación. La teoría de los vidrios rotos ofrece una perspectiva útil para entender y abordar esta problemática.

En los últimos años, los robos a locales comerciales han mostrado un aumento significativo en Colombia. Según datos de la Policía Nacional, en 2022 se registraron más de 25,000 incidentes de robo a locales comerciales, representando un aumento del 15 % en comparación con el año anterior (Policía Nacional de Colombia, 2022). Este incremento no solo genera pérdidas económicas directas por el robo de mercancías y daños a la propiedad, sino que también afecta la confianza de los empresarios y consumidores en la seguridad del entorno comercial.

La teoría de los vidrios rotos, propuesta por Wilson y Kelling (1982), sostiene que los signos visibles de desorden y deterioro, como ventanas rotas, grafitis y basura, fomentan comportamientos antisociales y criminales. Aplicada a la inseguridad en locales comerciales, esta teoría sugiere que la presencia de delitos menores y desorden puede generar una atmósfera de impunidad y descontrol, aumentando la probabilidad de delitos más graves. En otras palabras, si no se abordan adecuadamente los pequeños delitos y el desorden, se crea un entorno propicio para la escalada de criminalidad (Wilson & Kelling, 1982).

En muchas ciudades colombianas, el desorden y la falta de control sobre pequeños delitos son evidentes en las zonas comerciales. La falta de una respuesta rápida y efectiva a estos problemas puede perpetuar un ciclo de inseguridad. Estudios han demostrado que la percepción de inseguridad aumenta en áreas donde el desorden y la criminalidad menor no son controlados, afectando negativamente el comportamiento de los ciudadanos y empresarios (Montoya & Ruiz, 2021).

La ocurrencia de incidentes de actividad criminal en Colombia está influenciada por una variedad de factores socioeconómicos, demográficos y de disuasión. Un estudio de Arango & Restrepo (2020) que aplica el modelo de Becker-Ehrlich a datos panel de municipios colombianos entre 2010 y 2017 identifica que variables como la tasa de desempleo, la desigualdad social y la eficiencia policial tienen un impacto significativo en las tasas de criminalidad.

Además, Gómez & Morales (2016) encuentran que la percepción de inseguridad entre la población está estrechamente relacionada con la actividad criminal. Factores como la presencia policial, las capturas realizadas y la incautación de drogas influyen en cómo los ciudadanos perciben su seguridad, lo que a su vez puede afectar la incidencia de delitos.

Por otro lado, Nuñez et al. (2023) afirman que las características sociodemográficas, como la densidad poblacional y la estructura etaria, también juegan un papel en la dinámica del comportamiento delictivo. Un análisis descriptivo de la criminalidad en Colombia durante 2022 destaca cómo estas variables están asociadas con la ocurrencia de delitos en diferentes regiones del país.

Estos estudios subrayan la importancia de abordar la criminalidad desde una perspectiva multidimen-

sional, considerando tanto los factores estructurales como las percepciones ciudadanas para desarrollar estrategias efectivas de prevención y control del delito.

INTRODUCCIÓN

En la mayoría de los casos, cuando la variable de respuesta sigue una distribución Normal, es habitual emplear modelos de regresión lineal clásica. Sin embargo, los modelos lineales generalizados (MLG) ofrecen una ventaja significativa al permitir trabajar con variables de respuesta que no necesariamente cumplen con la restricción de normalidad, ampliando su aplicabilidad a distribuciones pertenecientes a la familia exponencial. Esta flexibilidad es especialmente útil en situaciones donde las características de los datos no se ajustan a las suposiciones de la regresión lineal tradicional (Winkelmann, 2000).

Las variables de recuento, definidas como la cantidad de eventos o sucesos que ocurren dentro de una unidad de observación específica y durante un intervalo de tiempo determinado, presentan propiedades únicas que las distinguen de las variables continuas: son de naturaleza discreta y no negativas. Estas características hacen que el uso de modelos estadísticos adecuados para su análisis sea crucial.

En este contexto, la distribución Poisson, introducida formalmente por el matemático francés Simeon Denis Poisson en 1837, juega un papel fundamental al modelar recuentos (Poisson, 1837). Aunque Poisson desarrolló su trabajo en el ámbito penal, los fundamentos de esta distribución ya habían sido abordados previamente por Abraham De Moivre en 1718, quien la presentó como un caso límite de la distribución binomial al considerar eventos raros tras un número elevado de repeticiones (De Moivre, 1718). Por otro lado, la distribución Binomial Negativa, inicialmente estudiada por Pierre Rémond de Montmort en el contexto de juegos de azar y posteriormente definida por Blaise Pascal, ha demostrado ser una alternativa valiosa para modelar datos de recuento cuando existe sobredispersión, es decir, cuando la varianza supera a la media (Montmort, 1714; Pascal, 1665).

Además de los modelos Poisson y Binomial Negativa, existen alternativas diseñadas para manejar escenarios más complejos. Por ejemplo, los modelos Zero-Inflated Poisson (ZIP) y Zero-Inflated Binomial Negativa (ZINB) se utilizan cuando los datos presentan un exceso de ceros, es decir, cuando hay una frecuencia elevada de observaciones donde no ocurre ningún evento. Estos modelos permiten capturar con mayor precisión la estructura de los datos, evitando sesgos en las estimaciones (Lambert, 1992). Asimismo, los modelos hurdle se han empleado exitosamente en situaciones donde los eventos tienen una alta probabilidad de no ocurrir, pero presentan un comportamiento distinto una vez que se supera el umbral inicial (Mullahy, 1986).

Un aspecto crucial en el análisis de datos de recuento es la identificación de la sobredispersión, que ocurre cuando la varianza excede a la media. Este fenómeno, común en muchas aplicaciones, puede deberse a factores como heterogeneidad en la población o dependencia no modelada entre los eventos. Para detectarla, se utilizan métricas como la devianza y la relación entre la varianza observada y la media (Cameron & Trivedi, 2013). Por otro lado, aunque menos frecuente, la subdispersión (cuando la varianza es menor que la media) también puede aparecer y requerir modelos específicos que ajusten esta relación. La correcta identificación de estas propiedades es esencial para elegir el modelo adecuado.

Tanto el modelo de regresión Poisson como el modelo de regresión Binomial Negativa han sido aplicados con éxito en diversas áreas de investigación, como ciencias sociales, ciencias de la salud, farmacología, criminología, relaciones laborales y economía (Lindsey, Jones & Jarvis, 2001; Osgoode, 2000; Sturman, 1999). Por ejemplo, en medicina, estos modelos son ampliamente utilizados para predecir la incidencia de enfermedades; en logística, para modelar la frecuencia de llegada de pedidos en tiempo real; y en análisis de redes sociales, para estudiar el número de interacciones o publicaciones realizadas en un período determinado. Este enfoque práctico ha hecho que estos modelos sean fundamentales en el análisis moderno de datos.

El avance en las herramientas computacionales ha impulsado significativamente el uso de modelos para

datos de recuento. Lenguajes como Python y R han desarrollado bibliotecas especializadas, como statsmodels, GLM y pscl, que permiten implementar modelos Poisson, Binomial Negativa y Zero-Inflated de manera eficiente, incluso con grandes volúmenes de datos (Zeileis, Kleiber & Jackman, 2008). Esto ha facilitado su integración en disciplinas emergentes, como la minería de datos y la inteligencia artificial, donde la capacidad de procesar datos masivos es esencial.

Aunque los modelos de recuento son herramientas robustas y ampliamente utilizadas, existen enfoques alternativos que también pueden ser considerados, como las redes neuronales y los modelos bayesianos (McElreath, 2020). Estos últimos son especialmente útiles cuando se dispone de grandes volúmenes de datos o se requiere una mayor flexibilidad en las estimaciones. Sin embargo, los modelos de recuento ofrecen una estructura matemática más directa y están específicamente diseñados para abordar las particularidades de las variables discretas, lo que los hace ideales en contextos donde la interpretación lógica de los resultados es primordial.

Finalmente, cabe destacar que el análisis de datos de recuento no solo permite describir fenómenos observados, sino que también contribuye al desarrollo metodológico en estadística. Esto resalta la importancia de elegir modelos adecuados para garantizar conclusiones válidas y útiles en cada contexto de estudio.

1. MODELOS LINEALES GENERALIZADOS

Los Modelos Lineales Generalizados (GLM) son una extensión de los modelos lineales tradicionales que permiten modelar una variable dependiente $\mathbf{Y}$ que sigue distribuciones no necesariamente normales. Introducidos por Nelder & Wedderburn (1972), los GLM se fundamentan en tres componentes esenciales: la distribución de la variable respuesta, la función de enlace y la estructura lineal de predictores.

1.1. Componentes Fundamentales de los GLM

1.1.1. Distribución de la variable dependiente

En los GLM, la variable respuesta $\mathbf{Y}$ pertenece a la familia exponencial de distribuciones, que incluye:

- Normal
- Binomial
- Poisson
- Gamma
- Inversa Gaussiana

La Función de Masa de Probabilidad para la familia exponencial tiene la forma general:

$$f_Y(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}$$

Donde:

- θ : Parámetro Canónico.
- $b(\theta)$: Función de normalización que asegura que la distribución sea válida.
- $a(\phi)$: Escala (frecuentemente dependiente de un parámetro de dispersión ϕ).
- $c(y, \phi)$: Término que incluye y y θ .

1.1.2. Función de enlace

La relación entre el predictor lineal η y el valor esperado de la respuesta $E[Y] = \mu$ se establece mediante una función de enlace $g(\cdot)$

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

Donde:

- η es el predictor lineal.
- X es la matriz de diseño.
- β es el vector de coeficientes.

La elección de $g(\cdot)$ depende de la distribución de Y y asegura que los valores predichos sean válidos para la naturaleza de los datos.

Las funciones de enlace más conocidas para los GLM son:

Función	Enlace
Logaritmo	$\log \mu_i = \eta_i$
Identidad	$\mu_i = \eta_i$
Raíz Cuadrada	$\sqrt{\mu_i} = \eta_i$
Logit	$\log \left(\frac{\mu_i}{n - \mu_i} \right) = \eta_i$
Recíproca	$\frac{1}{\mu_i} = \eta_i$
Exponencial	$\mu_i^n = \eta_i$
Inverso	$\frac{1}{\mu_i} = \eta_i$
Normal Inversa	$\phi^{-1}(\mu_i)$

Tabla 1: Funciones de enlace comunes en GLM

1.1.3. Predictor Lineal

El predictor lineal η es una combinación lineal de los predictores:

$$\eta = X\beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

Donde:

- X es la matriz de tamaño $n \times p$ de variables explicativas.
- β es el vector de tamaño $p \times 1$ de coeficientes.

1.2. Estimación de Parámetros

Los GLM utilizan el método de máxima verosimilitud para estimar los coeficientes β

1.2.1. Log-Verosimilitud

La función de log-verosimilitud para un GML es:

$$\ell(\beta; Y) = \sum_{i=1}^n \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right]$$

1.2.2. Estimación iterativa

Debido a la no linealidad en los GLM, la estimación se realiza mediante algoritmos iterativos como:

- Algoritmo de Fisher Scoring
- Newton-Raphson

El procedimiento de iteración puede expresarse como:

$$\beta^{(t+1)} = \beta^{(t)} + (X^T W C)^{-1} X^T W z$$

Donde:

- W es una matrix de pesos basada en la varianza de Y
- z es el vector de respuestas ajustadas (pseudo-residuales)

1.3. Varianza de la Variable Dependiente

La varianza de la variable respuesta Y en los GLM se relaciona con su valor esperado $E[Y] = \mu$ a través de la función de varianza $V(\mu)$:

$$Var(Y) = a(\phi)V(\mu)$$

Donde $V(\mu)$ es la función de varianza de la Distribución de la Y .

1.4. Diagnóstico y Bondad de Ajuste

Para evaluar el ajuste de un modelo GLM, se emplean herramientas específicas como:

1.4.1. Residuales deviance

Los residuales de deviance y la deviance en general son conceptos fundamentales en los modelos lineales generalizados (GLM). La deviance mide qué tan bien el modelo ajustado explica los datos observados en comparación con el modelo saturado (el modelo con máxima verosimilitud). Los residuales de deviance proporcionan una herramienta para evaluar individualmente las discrepancias.

McCullagh y Nelder (1989) presentan una descripción completa de la deviance y los residuales de deviance, incluyendo cómo estas métricas derivan de la verosimilitud y cómo se utilizan para diagnosticar el ajuste del modelo.

Según Dobson y Barnett (2008), los residuales de deviance son útiles para identificar observaciones atípicas y posibles problemas en el ajuste del modelo.

La deviance es una medida basada en la función de verosimilitud que evalúa qué tan bien el modelo ajusta los datos observados en comparación con un modelo saturado (modelo que predice perfectamente los datos).

La fórmula de la deviance:

$$D = 2 \cdot [\log L_{saturado} - \log L_{modelo}]$$

Donde:

- $L_{saturado}$ es la función de verosimilitud del modelo saturado.
- L_{modelo} es la función de verosimilitud del modelo ajustado.

Cálculo de los residuales de deviance:

Los residuales de deviance para cada observación i se definen como:

$$d_i = \text{sign}(y_i - \hat{y}_i) \cdot \sqrt{2 \cdot [\ell(y_i; y_i) - \ell(y_i; \hat{y}_i)]}$$

Donde:

- y_i es el valor observado
- $\hat{y}_i$ es el valor predicho
- $\ell(y_i; \hat{y}_i)$ es el logaritmo de la verosimilitud de la observación.

Interpretación:

- Los residuales de deviance miden la discrepancia entre los valores observados y los predichos en términos de verosimilitud.
- Valores cercanos a 0 indican que el modelo predice bien para esa observación.
- Valores altos (positivos o negativos) indican posibles observaciones atípicas o que el modelo no ajusta bien.

Aplicación:

- **Diagnóstico de modelos:** Los residuales de deviance pueden ayudar a identificar valores atípicos o patrones no modelados.
- **Comparación de modelos:** La suma de los residuales de deviance totales se utiliza para calcular la deviance del modelo, que puede compararse entre modelos para evaluar ajustes relativos.

1.4.2. Pseudo- R^2

Se ha desarrollado para ofrecer una medida intuitiva de la proporción de variabilidad explicada por un modelo, adaptada a contextos donde el R^2 tradicional no aplica, como en los GLM.

Según McFadden (1974), el Pseudo- R^2 basado en la log-verosimilitud es ampliamente utilizado en modelos logit y probit.

$$R_{McFadden}^2 = 1 - \frac{\log L_{modelo}}{\log L_{nulo}}$$

Cox & Snell (1989) introducen una versión que utiliza la relación entre la verosimilitud del modelo ajustado y el modelo nulo. Limita a un valor menor que 1, dependiendo de la escala de la verosimilitud.

$$R_{Cox-Snell}^2 = 1 - \left(\frac{L_{modelo}}{L_{nulo}} \right)^{\frac{2}{n}}$$

Nagelkerke (1991) propone una variante que ajusta el Pseudo- R^2 de Cox y Snell para que alcance un máximo de 1, facilitando la interpretación y comparación entre modelos.

$$R_{Nagelkerke}^2 = \frac{R_{Cox-Snell}^2}{1 - L_{nulo}^{\frac{2}{n}}}$$

Donde:

- L_{modelo} : Verosimilitud del modelo ajustado.
- L_{nulo} : Verosimilitud del modelo nulo (sin predictores).
- n : Tamaño de la muestra.

Interpretación:

- Valores cercanos a 1 indican que el modelo explica una mayor proporción de la variabilidad.
- Es importante notar que los valores de Pseudo- R^2 tienden a ser más bajos que los de R^2 en regresión lineal, lo cual es normal y no necesariamente indica un modelo deficiente.

Limitación:

- No tiene una interpretación única o estándar como el R^2 de la regresión lineal.

Aplicaciones:

- Comparar modelos: Es útil para comparar la capacidad explicativa de diferentes modelos.
- Evaluar ajuste: Proporciona una medida intuitiva para interpretar el desempeño del modelo, aunque no tiene una escala directamente comparable con el R^2 tradicional.

1.4.3. Criterio de información AIC

El Criterio de Información de Akaike (AIC), propuesto por Hirotugu Akaike en 1974, es una herramienta estadística utilizada para evaluar y comparar modelos. Su propósito es seleccionar el modelo que mejor balancea entre el ajuste a los datos y la simplicidad estructural, evitando el sobreajuste que puede ocurrir al aumentar innecesariamente el número de parámetros.

El AIC se basa en la combinación de dos conceptos fundamentales:

1. **Teoría de Máxima Verosimilitud:** Evalúa qué tan bien un modelo explica los datos observados.
2. **Teoría de la información y entropía:** Busca medir la pérdida de información al aproximar los datos reales mediante un modelo estadístico.

La clave del AIC es penalizar la complejidad del modelo, incorporando el número de parámetros para evitar favorecer modelos que incluyan demasiados ajustes superfluos.

Se define como:

$$AIC = -2 \cdot \log L + 2K$$

Donde:

- $\log L$: Logaritmo de la función de verosimilitud evaluada en el modelo ajustado.

- K : Número de parámetros estimados en el modelo.

Interpretación:

1. **Primer término** ($-2 \cdot \log L$): Representa la bondad de ajuste del modelo, Un mayor valor de log-verosimilitud implica un mejor ajuste.
2. **Segundo término** ($2K$): Introduce una penalización proporcional al número de parámetros en el modelo, incentivando la simplicidad y evitando el sobreajuste.

Comparación entre Modelos:

El AIC se utiliza para comparar varios modelos ajustados al mismo conjunto de datos. El modelo con el menor valor de AIC se considera preferible, ya que logra un mejor balance entre ajuste y complejidad.

Sin embargo, el AIC no proporciona una medida absoluta de la calidad del modelo, sino que debe interpretarse de manera relativa al comparar diferentes opciones.

Supuestos del AIC:

1. Los modelos evaluados deben estar anidados o ser comparables en términos de los datos y la función de verosimilitud.
2. Se asume que el tamaño de muestra es suficientemente grande para que las estimaciones de máxima verosimilitud sean válidas.

Limitaciones del AIC:

1. **Dependencia del tamaño de muestra:** El AIC puede sobrevalorar modelos más complejos en muestras pequeñas. Para estos casos se debe emplear el AIC corregido (AIC_c) :

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1}$$

Donde n es el tamaño de la muestra.

2. **Comparación restringida:** El AIC sólo es válido para comparar modelos ajustados al mismo conjunto de datos.
3. **No mide la calidad absoluta:** No permite evaluar si el modelo en sí es bueno, sólo cuál es mejor relativo a los demás.

Aplicaciones del AIC:

El AIC se utiliza ampliamente en:

- **Modelos lineales y no lineales:** Incluyendo regresiones lineales, logísticas y modelos generalizados.
- **Series temporales:** Selección de modelos ARIMA.
- **Machine Learning:** En problemas de selección de características y ajuste de hiperparámetros

1.4.4. Criterio de información BIC

El Criterio de Información Bayesiano (BIC), también conocido como Schwarz Information Criterion (SIC), fue introducido por Gideon Schwarz en 1978. Este criterio es una herramienta estadística utilizada para la selección de modelos, similar al AIC, pero con una penalización más fuerte para la complejidad

del modelo. El BIC es particularmente útil cuando el tamaño de la muestra es grande, ya que prioriza modelos más simples.

El BIC se fundamenta en la teoría de la máxima verosimilitud, pero incorpora un término basado en la probabilidad a posteriori dentro de un contexto bayesiano. Evalúa la calidad de un modelo ajustado mediante una penalización que crece con el tamaño de la muestra y el número de parámetros.

El BIC se calcula como:

$$BIC = -2 \cdot \log L + K \cdot \log(n)$$

Donde:

- $\log L$: Logaritmo de la función de verosimilitud evaluada en el modelo ajustado.
- K : Número de parámetros estimados en el modelo.
- n : Tamaño de la muestra.

Interpretación:

1. **Primer término** ($-2 \cdot \log L$): Similar al AIC, este término mide qué tan bien el modelo ajusta los datos.
2. **Segundo término** ($K \cdot \log(n)$): Penaliza la complejidad del modelo, pero esta penalización aumenta con el tamaño de la muestra (n), lo que favorece modelos más simples en conjuntos de datos grandes.

Supuestos del BIC:

1. Los modelos evaluados deben estar anidados o ser comparables.
2. Se asume que el tamaño de muestra es suficientemente grande para que las estimaciones sean consistentes.

Ventajas del BIC:

- Penaliza más fuertemente la complejidad del modelo, lo que ayuda a evitar el sobreajuste.
- Es consistente: selecciona el modelo verdadero si el tamaño de la muestra tiende a infinito.

Limitaciones del BIC:

- En muestras pequeñas, la penalización puede ser demasiado severa, descartando modelos adecuados.
- Al igual que el AIC, el BIC solo permite comparar modelos ajustados al mismo conjunto de datos.

Aplicaciones del BIC:

- **Modelos lineales y no lineales:** Regresiones, modelos generalizados, etc.
- **Series temporales:** Selección de modelos ARIMA.
- **Clusterización:** Para determinar el número óptimo de grupos.

Diferencias entre AIC y BIC

1. **Penalización:**

- En el AIC, la penalización es constante ($2K$)
- En el BIC, la penalización depende del tamaño de la muestra ($K \cdot \log(n)$)

2. Enfoque:

- El AIC busca minimizar la pérdida de información.
- El BIC incorpora un enfoque más bayesiano, considerando la probabilidad a posteriori.

3. Selección de modelos:

- El AIC es más flexible y tiende a seleccionar modelos más complejos.
- El BIC tiende a preferir modelos más simples, especialmente en muestras grandes.

2. MODELOS DE REGRESIÓN PENALIZADA

Hastie, T., Tibshirani, R., & Friedman, J. (2009) definen os modelos de regresión penalizada como técnicas diseñadas para manejar problemas de sobreajuste y selección de variables en presencia de multicolinealidad o conjuntos de datos con muchas variables. Estas técnicas añaden un término de penalización a la función de pérdida de la regresión lineal estándar, las más comunes son:

1. Ridge Regression (Tikhonov Regularization):

- Penaliza la suma de los cuadrados de los coeficientes ($\lambda \sum \beta_j^2$).
- Mantiene todos los coeficientes, pero los reduce hacia cero.

2. Lasso (Least Absolute Shrinkage and Selection Operator):

- Penaliza la suma de los valores absolutos de los coeficientes ($\lambda \sum |\beta_j|$).
- Puede forzar algunos coeficientes a ser exactamente cero, seleccionando automáticamente variables relevantes.

3. Elastic Net:

- Combina las penalizaciones de Ridge y Lasso ($\lambda_1 \sum |\beta_j| + \lambda_2 \sum \beta_j^2$).
- Es útil cuando hay alta colinealidad entre variables.

2.1. Modelo Lasso

Tibshirani, R. (1996) afirma que el modelo Lasso puede eliminar automáticamente variables irrelevantes al asignarles un coeficiente de cero. Esto lo hace útil en escenarios donde hay muchas variables explicativas, pero no todas son significativas.

Función de Pérdida:

La función de pérdida en el Lasso incluye una penalización L_1

$$\text{Min} \left\{ \frac{1}{2n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

λ es un hiperparámetro que controla el grado de penalización. Cuando $\lambda = 0$, el modelo es una regresión lineal estándar; cuando λ es muy grande, todos los coeficientes tienden a cero.

Ventajas:

Según Hastie, T., Tibshirani, R., & Friedman, J. (2009) este modelo rreviene el sobreajuste al regularizar los coeficientes y facilita interpretaciones al reducir el número de variables seleccionadas.

Limitaciones:

Zou, H., & Hastie, T. (2005) exponen que en casos de alta correlación entre variables predictoras, el Lasso seleccionará una entre las variables correlacionadas y excluirá a las demás y que puede ser inestable si no se elige adecuadamente el parámetro λ

Interpretación de coeficientes:

Los coeficientes seleccionados son sensibles al valor de λ . Coeficientes cercanos a cero indican variables con poca o ninguna relación con la respuesta.

3. MODELOS PREDICTIVOS PARA DATOS DE CONTEO

El acervo teórico de esta sección se desarrollará tomando como base las ideas y planteamientos descritos en el trabajo de (Contreras Vilca 2013)

3.1. Modelo de Regresión Poisson

Cuando la variable dependiente representa datos de conteo, es recomendable emplear la distribución de Poisson. En el Modelo de Regresión de Poisson (MRP), se modela la media μ en función de las variables explicativas mediante una función de enlace logarítmica, que asegura que los valores predichos sean positivos.

Este modelo asume que la media y la varianza de los datos son iguales ($\text{Var}(Y) = \mathbb{E}(Y) = \mu$), lo que es conocido como equiprobabilidad. Además, el MRP es particularmente útil para modelar eventos raros y para situaciones en las que los datos están compuestos por valores enteros no negativos.

La flexibilidad del MRP permite incorporar efectos de diferentes factores y su interacción, haciéndolo una herramienta valiosa para analizar fenómenos en campos como epidemiología, economía, ciencias sociales, entre otros.

3.1.1. Distribución de Probabilidad Discreta Poisson

Sea Y_i una variable aleatoria discreta que representa la cantidad de veces que ocurre un determinado evento, cuya función de probabilidad está definida como:

$$f(y_i) = P(Y_i = y_i) = \frac{e^{-\mu} \mu^{y_i}}{y_i!}, \text{ para } y = 1, 2, 3...$$

Donde:

- y_i : Número de veces que ocurre un evento.
- μ : Parámetro positivo que indica la cantidad esperada de veces que el evento ocurrirá en un intervalo de tiempo dado.

La Función de Probabilidad Acumulada de la Distribución Poisson es:

$$F(y|\mu) = \frac{\Gamma([y_i + 1], \mu)}{[y_i]!}, \text{ para todo } y_i \geq 0$$

Donde $\Gamma(x, y)$ es la función Gamma Incompleta.

Propiedades:

1. A medida que μ aumenta, la distribución se desplaza hacia la derecha. En este caso, la esperanza matemática de y_i está dada por $E(y_i) = \mu$. El parámetro μ se denomina "tasa" porque representa el número promedio de veces que un evento ocurre por unidad de tiempo.
2. Equidispersión: $E(y_i) = Var(y_i) = \mu$, la varianza es a la esperanza matemática.
3. A medida que μ aumenta, la probabilidad de que $Y_i = 0$ disminuye. La Figura 1 muestra la función de masa de probabilidad para diferentes valores de λ , ilustrando cómo cambia la forma de la distribución.

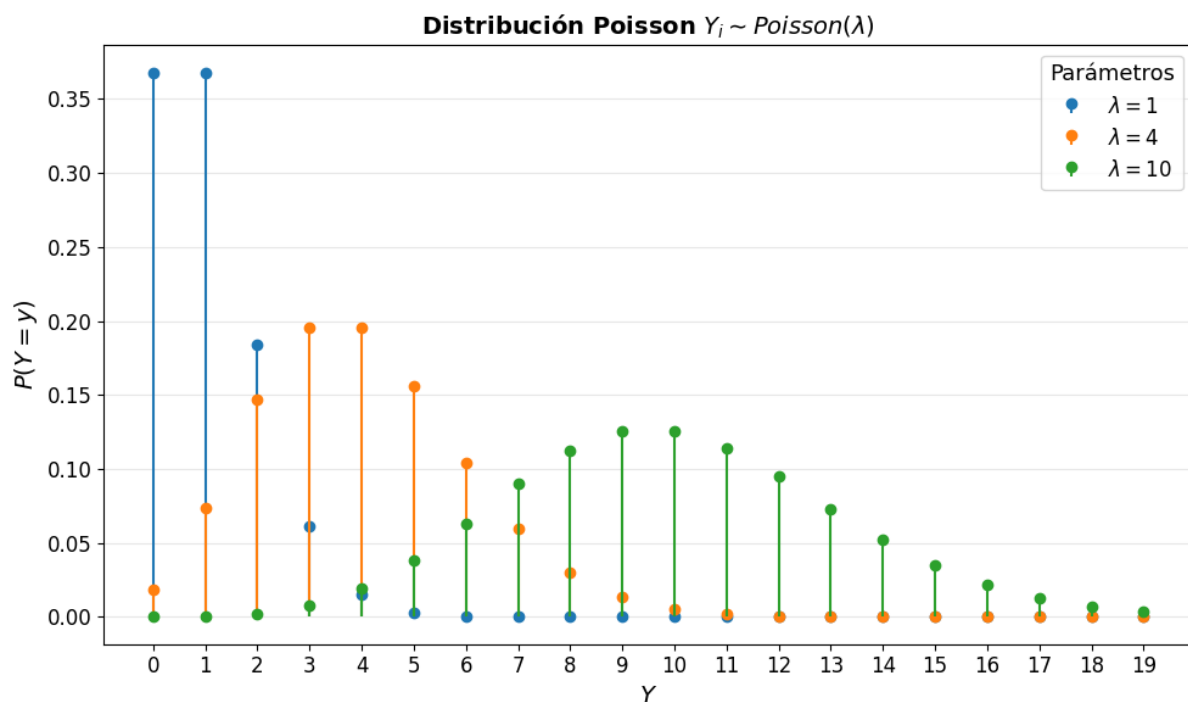


Figura 1: Distribución Poisson para $\lambda = 1$, $\lambda = 4$ y $\lambda = 10$.

4. A medida que el valor de μ aumenta, la Distribución Poisson se aproxima a la Distribución Normal.

3.1.2. La Distribución Poisson como un caso particular de la Familia Exponencial

Sean $Y_1, \dots, Y_n$ V.A. i.i.d donde su función de probabilidad pertenece a la familia exponencial y está dado por:

$$\begin{aligned}
 \ln f(y_i) &= \exp\{y_i \ln(\mu_i) - \mu_i - \ln(y_i!)\} \\
 &= -\ln(y_i!) + \frac{y_i \theta_i - b(\theta_i)}{\phi}
 \end{aligned}$$

Donde:

- $\theta = 1$, parámetro de escala.
- $\theta_i = \ln(\mu_i)$

- $b(\theta_i) = e^{\theta_i}$
- $c(y_i, \phi) = -\ln(y_i!)$

Por lo tanto, la Distribución de Poisson pertenece a la familia exponencial con Media y Varianza:

$$b'(\theta_i) = e^{\theta_i} = \mu_i = E(y_i)$$

$$b''(\theta_i) = \mu_i = \text{Var}(y_i)$$

3.1.3. Modelo de Regresión Poisson

Y_i sigue un Modelo de Regresión Poisson si cumple lo siguiente:

$$Y_i \sim P(\mu_i), i = 1, 2, 3, \dots, n$$

$$g(\mu_i) = \eta_i = x_i^T \beta$$

Donde

- $x_i = (x_{i1}, \dots, x_{ip})^T$, vector de covariantes explicativas.
- $\beta = (\beta_0, \dots, \beta_p)^T$, vector de parámetros desconocidos.

Componentes del Modelo de Regresión Poisson:

- **Componente Aleatorio:**

Sea $Y_1, \dots, Y_n$ un vector que contiene la variable dependiente positiva y $x_i = (x_{i1}, \dots, x_{ip})^T$ un vector de covariantes explicativas con parámetro μ_i , su componente aleatorio está dado por:

$$Y_i \sim P(\mu_i), i = 1, 2, 3, \dots, n$$

- **Componente Sistemático:**

Considerando μ_i y el denominado predictor lineal, representado mediante:

$$\begin{aligned} \eta_i &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \\ &= \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \\ &= x_i^T \beta \end{aligned}$$

- **Función Enlace:**

Los componentes expuestos anteriormente se integran en el modelo a través de la selección de una función de enlace:

Función	Enlace
Logaritmo	$\log \mu_i = \eta_i$
Identidad	$\mu_i = \eta_i$
Raíz Cuadrada	$\sqrt{\mu_i} = \eta_i$

Tabla 2: Funciones de Enlace para el Modelo de Regresión Poisson

Si se utiliza el enlace logarítmico, $\hat{\mu}_i = \exp(x_i^T \hat{\beta})$, este garantiza que el resultado sea positivo.

■ **Función Desvío:**

La teoría de la Función Desvío se plasmará de acuerdo al trabajo de Paula (2010), Si $\theta_i = \log(\mu_i)$, implica que $\tilde{\theta}_i = \log(y_i)$ para $y_i > 0$ y $\hat{\theta}_i = \log(\hat{\mu}_i)$. Por lo tanto,

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left\{ y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right\}$$

Cuando $y_i = 0$, el i -ésimo término de $D(y_i; \hat{\mu}) = 2\hat{\mu}_i$. Por lo tanto, la Función de Desvío para el Modelo Poisson es:

$$d^2(y_i, \hat{\mu}_i) = \begin{cases} 2 \left\{ y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right\}, & \text{si } y_i > 0, \\ 2\hat{\mu}_i, & \text{si } y_i = 0. \end{cases}$$

Estimación por Máxima Verosimilitud en el Modelo de Regresión Poisson

Considerando $Y_1, \dots, Y_n$ como un conjunto de n observaciones aleatorias e independientes, donde el predictor está representado por x . La función de verosimilitud se define como:

$$\frac{\prod_{i=1}^n \mu_i^{y_i} \exp \left(- \sum_{i=1}^n \mu_i \right)}{\prod_{i=1}^n y_i!}$$

donde $\mu_i = g^{-1}(x^T, \beta)$.

Después de seleccionar la función de enlace, es posible proceder a su maximización. El logaritmo de la función de verosimilitud se expresa como:

$$L(\beta) = \sum_{i=1}^n y_i \ln(\mu_i) - \sum_{i=1}^n \ln(\mu_i) - \sum_{i=1}^n \ln(y_i!)$$

El valor que maximiza $L(\beta)$, es el vector de coeficientes estimado $\hat{\beta}$

3.1.4. Equidispersión

El Modelo de Regresión Poisson se considera una herramienta destacada para modelar datos de conteo, debido a sus ventajas evidentes. Sin embargo, su aplicación puede ser inadecuada cuando no se cumplen ciertos supuestos, cuya causa puede ser variada, como lo señala Winkelmann (2000).

Una característica fundamental de la distribución de Poisson es la propiedad de **equidispersión**, lo que implica que:

$$\text{Var}(y) = \mathbb{E}(y) = \mu$$

Esta propiedad es un supuesto esencial en muchos Modelos Lineales Generalizados (MLG). No obstante, la violación de este supuesto sobre la varianza conlleva automáticamente la invalidez de la distribución de Poisson. Un problema recurrente en este modelo es que la relación entre la media y la varianza no sea igualitaria, dando lugar a:

- **Sobredispersión:** Cuando $\text{Var}(y) > \mathbb{E}(y)$, es decir, si $\sigma^2 > 1$.
- **Infradispersión o subdispersión:** Cuando $\text{Var}(y) < \mathbb{E}(y)$, es decir, si $\sigma^2 < 1$.

Estas desviaciones respecto a la equidispersión pueden afectar significativamente los resultados y la interpretación del modelo, haciéndolo inapropiado para ciertos conjuntos de datos.

De acuerdo con lo señalado por Krzanowski (1998) y Winkelmann (2000), las situaciones de sobredispersión son significativamente más comunes que las de infradispersión.

Cuando existe una variabilidad excesiva en los datos, las estimaciones de los errores estándar pueden estar sesgadas, lo que puede llevar a errores en las inferencias realizadas a partir de los parámetros del modelo de regresión (Krzanowski, 1998). Este fenómeno es típico en aplicaciones que utilizan distribuciones con una varianza rígida, como la Poisson o la Binomial.

Entre las principales causas de sobredispersión se incluyen las siguientes:

- Alta variabilidad inherente en los datos.
- Los datos no se ajustan a una distribución Poisson.
- Los eventos no ocurren de forma independiente en el tiempo.
- Falta de estabilidad en los datos, es decir, la probabilidad de que ocurra un evento no está necesariamente relacionada con el promedio μ . Esto puede deberse a la omisión de variables explicativas o al uso de transformaciones incorrectas en lugar de relaciones lineales en el modelo (Winkelmann, 2000).
- Elecciones inadecuadas de la función de enlace.
- Heterogeneidad en la muestra, que podría deberse a la variabilidad entre experimentos.

Existen varias metodologías propuestas para identificar la presencia de sobredispersión. Una de ellas es el **Índice de Dispersión (In)**, sugerido por Lindsey (1995), el cual se utiliza como indicador para evaluar el cumplimiento del supuesto de equidispersión. Este índice se define como la relación entre la varianza y la esperanza matemática:

$$In = \frac{\text{Var}(y)}{\mathbb{E}(y)}$$

Teóricamente, si $\text{Var}(y) = \mathbb{E}(y)$, el Índice de Dispersión debería ser igual a 1. Con base en este valor, se pueden identificar las siguientes situaciones:

- Sobredispersión: cuando $In > 1$.
- Infradispersión: cuando $In < 1$.

La magnitud del Índice de Dispersión será determinante para establecer la presencia de sobredispersión o infradispersión en los datos.

Otro método sencillo para detectar sobredispersión es evaluar si la varianza estimada excede el doble de la media estimada. Si esto ocurre, es probable que los datos presenten sobredispersión, incluso después de incluir las variables explicativas en el modelo. Este criterio fue planteado por Cameron y Trivedi (1986).

3.2. Modelo de Regresión Binomial Negativa

El Modelo de Regresión Binomial Negativa (MRBN) se considera generalmente como la principal alternativa al Modelo de Regresión Poisson, ya que no exige la igualdad entre la media y la varianza, lo que lo hace adecuado para manejar situaciones de sobredispersión en los datos.

3.2.1. Distribución de Probabilidad Binomial Negativa

La función de probabilidad está definida como:

$$f(y_i) = \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} (1 - \mu_i)^{y_i} \mu_i^\phi$$

Sus propiedades son:

- $\mathbb{E}(Y) = \phi \frac{1-\mu}{\mu}$
- $Var(Y) = \phi \frac{1-\mu}{\mu^2}$

La función de probabilidad de la distribución Binomial Negativa puede adoptar diferentes formas dependiendo de los valores de los parámetros que la definen. En la siguiente Figura, se muestran estas funciones como si fueran densidades, aunque en realidad corresponden a valores discretos, con

$$Y_i \sim BN(r, p)$$

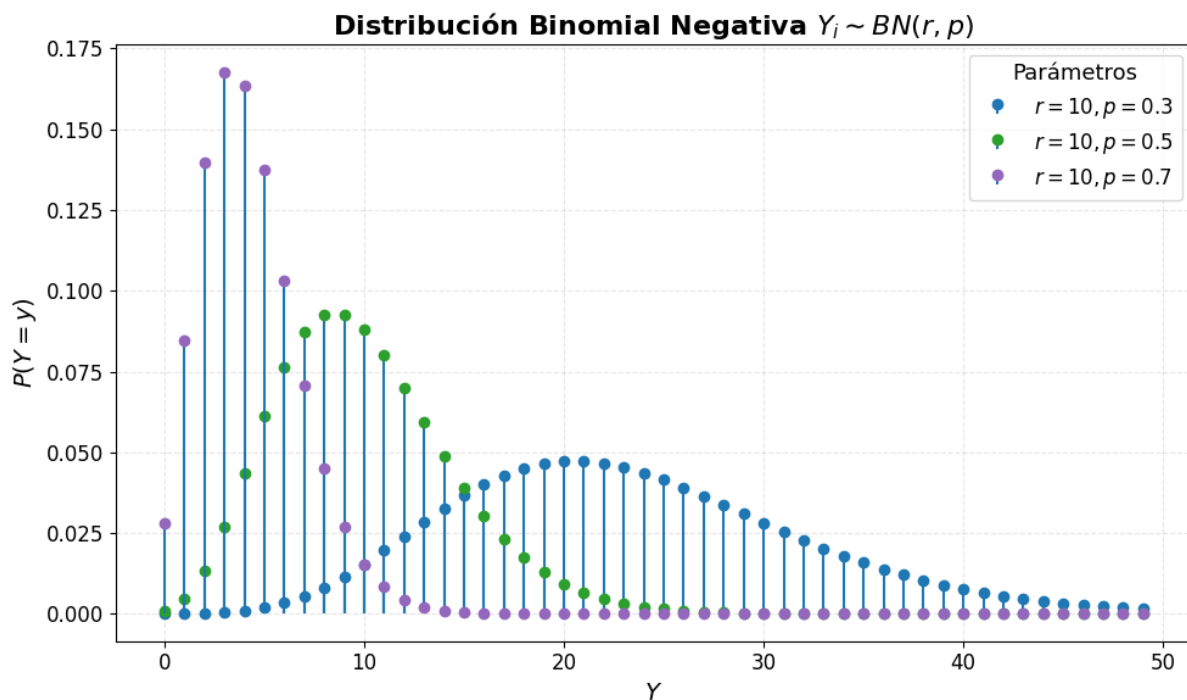


Figura 2: Distribución Poisson para $\lambda = 1$, $\lambda = 4$ y $\lambda = 10$.

3.2.2. La Distribución Binomial Negativa como un caso particular de la Familia Exponencial

La función de probabilidad de la Binomial Negativa está dada por:

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

donde, $y = 0, 1, \dots, n$ con parámetros μ_i y ϕ , con $\mu_i > 0$ y $\phi > 0$.

Si $1/\phi \rightarrow 0$, entonces la varianza $\text{Var}(Y_i) \rightarrow \mu_i$, lo que implica que la distribución binomial negativa converge hacia una distribución de Poisson.

Cuando ϕ es constante, la densidad se incluye dentro de la familia exponencial, permitiendo que el modelo se considere un modelo lineal generalizado (MLG) con distribución binomial negativa.

Por lo tanto, si representamos $Y|z \sim P(z)$ y $Z \sim G(\mu, \phi)$, donde ϕ no depende de μ , según Paula (2010).

Para este caso:

$$\mathbb{E}(Z) = \mu; \text{Var}(Z) = \frac{\mu^2}{\phi}$$

Entonces,

$$f(y|z) = \frac{e^{-z} z^y}{y!}$$

$$g(z; \mu, \phi) = \frac{1}{\Gamma(\phi)} \left(\frac{z\phi}{\mu} \right)^\phi e^{-\frac{\phi z}{\mu}} \frac{1}{z}$$

La Función de Probabilidad de Y está dada por:

$$P(Y = y) = \int_0^\infty f(y|z)g(z; \mu, \phi) dz$$

$$= \frac{1}{y!\phi} \left(\frac{\phi}{\mu} \right)^\phi \int_0^\infty e^{-z(1+\phi/\mu)} z^{\phi+y-1} dz$$

Transformando la variable,

$$t = z \left(1 + \frac{\phi}{\mu} \right)$$

Luego,

$$\frac{dz}{dt} = \left(1 + \frac{\phi}{\mu} \right)$$

Entonces,

$$P(Y = y) = \frac{1}{y!\Gamma(\phi)} \left(\frac{\phi}{\mu} \right)^\phi \left(1 + \frac{\phi}{\mu} \right)^{-(\phi+y)} \int_0^\infty e^{-t} t^{\phi+y-1} dt$$

$$= \frac{\Gamma(\phi + y) \mu^y \phi^\phi}{\Gamma(\phi) \Gamma(y + 1) (\mu + \phi)^{\phi+y}}$$

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

$$= \frac{\Gamma(\phi + y)}{\Gamma(y + 1) \Gamma(\phi)} (1 - \pi)^\phi \pi^y$$

donde, $\pi = \mu/(\mu + \phi)$.

Por consiguiente, Y sigue una distribución Binomial Negativa con media μ y parámetro de dispersión ϕ .

La función de probabilidad de esta distribución puede representarse de la siguiente forma:

$$\log f(y) = \exp \left\{ y \log \left(\frac{\mu}{\mu + \phi} \right) - \phi \log \left(\frac{\mu + \phi}{\phi} \right) + \log \left(\frac{\Gamma(y + \phi)}{\Gamma(y + 1)\Gamma(\phi)} \right) \right\}$$

donde:

- $\phi = 1$ parámetro de escala.
- $\theta = \log \left(\frac{\mu}{\mu + \phi} \right)$
- $b(\theta) = \phi \log \left(\frac{\mu + \phi}{\phi} \right)$
- $b(\theta) = \phi \log (1 - e^\theta)$
- $c(y_i, \phi) = \log \left[\frac{\Gamma(y + \phi)}{\Gamma(y + 1)\Gamma(\phi)} \right]$

Con Media y Varianza:

$$\mathbb{E}(Y_i) = \mu_i; \quad \text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\phi}$$

3.2.3. Modelo de Regresión Binomial Negativa

Una variable Y_i sigue un modelo de Regresión Binomial Negativa si:

$$Y_i \sim BN(\mu_i, \phi), i = 0, 1, 2, 3, ..$$

$$g(\mu_i) = x_i^T \beta$$

Donde,

- $x_i = (x_{i1}, ..., x_{ip})^T$, vector de covariables explicativas.
- $\beta = (\beta_0, ..., \beta_p)^T$, vector de parámetros desconocidos,

Componentes del Modelo de Regresión Binomial Negativo:

■ Componente Aleatorio:

Considerando $Y_1, \dots, Y_n$ como variables aleatorias independientes que representan el número de eventos necesarios para alcanzar r éxitos. En otras palabras, el número de éxitos está fijado, mientras que la aleatoriedad radica en la cantidad de eventos requeridos, de forma que:

$$Y_i \sim BN(\mu_i, \phi), i = 0, 1, 2, 3, ...$$

■ Componente Sistemático:

Dado μ_i y el denominado predictor lineal, representado mediante:

$$\eta_i = x_i^T \beta$$

■ **Función Enlace:**

Los componentes expuestos anteriormente se integran en el modelo a través de la selección de una función de enlace:

$$g(\mu_i) = \eta_i$$

Funciones de enlaces usadas para MRBN

Función	Enlace
Logaritmo	$\log \mu_i = \eta_i$
Identidad	$\mu_i = \eta_i$
Raíz Cuadrada	$\sqrt{\mu_i} = \eta_i$

Tabla 3: Funciones de Enlace para el Modelo de Regresión Binomial Negativa

- **Función Desvío:** En el libro de Paula (2010) se asume ϕ constante, y la Función Desvío, está dada por:

$$D^*(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[\phi \log \left(\frac{\hat{\mu}_i + \phi}{y_i + \phi} \right) + y_i \log \left(\frac{y_i(\hat{\mu}_i + \phi)}{\hat{\mu}_i(y_i + \phi)} \right) \right]$$

Dado que $\hat{\mu}_i = g^{-1}(x_i^T \beta)$, bajo la hipótesis de que el modelo ajustado es adecuado, el estadístico $D^*(y; \hat{\mu})$ sigue una distribución $\chi^2_{(n-p)}$ con $(n-p)$ grados de libertad, siempre que ϕ y μ_i sean grandes.

Estimación por Máxima Verosimilitud en el Modelo de Regresión Poisson

De acuerdo con Paula (2010) se considera la partición $\theta = (\beta^T, \phi)^T$, la cual representa el logaritmo de la función de verosimilitud, así:

$$L(\theta) = \sum_{i=1}^n \left[\log \left(\frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \right) + \phi \log \phi + y_i \log \mu_i - (\phi + y_i) \log(\mu_i + \phi) \right]$$

donde $\mu_i = g^{-1}(x_i^T \beta)$, es una Función Score para β .

Calculando la derivada de la Función Score para β , se tiene que:

$$\begin{aligned} \frac{\partial L(\theta)}{\partial \beta_j} &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{\beta_j} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} \frac{d\mu_i}{d\eta_i} x_{ij} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} \right\} \\ &= \sum_{i=1}^n \left\{ \frac{\phi \frac{d\mu_i}{d\eta_i}}{\mu_i(\phi + \mu_i)} (y_i - \mu_i) x_{ij} \right\} \\ &= \sum_{i=1}^n w_i f_i^{-1} (y_i - \mu_i) x_{ij}. \end{aligned}$$

donde,

$$w_i = \frac{(d\mu_i/d\eta_i)^2}{(\mu_i^2 \phi^{-1} + \mu_i)}$$

$$f_i = \frac{d\mu_i}{\eta_i}$$

Entonces, la Función Score se puede expresar de forma matricial, así:

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{X}^T \mathbf{W} \mathbf{F}^{-1}(\mathbf{y} - \boldsymbol{\mu})$$

Donde:

- $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T]^T$: Cada fila $\mathbf{x}_i^T$ representa el vector de variables independientes para la observación i , con $i = 1, \dots, n$.
- $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$
- $\mathbf{F} = \text{diag}\{f_1, \dots, f_n\}$
- $\mathbf{y} = (y_1, \dots, y_n)^T$
- $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$

De manera similar, la función score para ϕ se puede expresar como:

$$U_\phi(\boldsymbol{\theta}) = \sum_{i=1}^n \left[\psi(\phi + y_i) - \psi(\phi) - \frac{(y_i + \phi)}{(\phi + \mu_i)} + \log \left(\frac{\phi}{\phi + \mu_i} \right) + 1 \right]$$

donde $\psi(\cdot)$ es la función digamma.

Para calcular la matriz de información de Fisher, se calculan las segundas derivadas parciales de la función de verosimilitud:

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n \left\{ \frac{(\phi + y_i)}{(\phi + \mu_i)^2} \frac{y_i}{\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{il} \right\} + \sum_{i=1}^n \left\{ \frac{y_i}{\mu_i} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \right\} \frac{d^2 \mu_i}{d\eta_i^2} x_{ij} x_{il}.$$

Los valores esperados de las derivadas segundas están dados por:

$$E \left\{ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} \right\} = - \sum_{i=1}^n \left\{ \frac{\phi \left(\frac{d\mu_i}{d\eta_i} \right)^2}{(\phi + \mu_i)} x_{ij} x_{il} \right\} = - \sum_{i=1}^n w_i x_{ij} x_{il}.$$

Luego, la información de Fisher para β puede representarse en forma matricial como:

$$K_\beta(\boldsymbol{\theta}) = E \left\{ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta \partial \beta^T} \right\} = \mathbf{X}^T \mathbf{W} \mathbf{X}.$$

Por otro lado, *Lawless* (1982) establece que la información de Fisher para ϕ se puede expresar como:

$$K_\phi(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^{\infty} (\phi + j)^2 P_r(Y_i \geq j) - \phi^{-1} \mu_i / (\mu_i + \phi).$$

Aquí, β y ϕ son parámetros ortogonales, por lo que la matriz de información de Fisher para $\boldsymbol{\theta}$ tiene una estructura diagonal por bloques, como sigue:

$$K_{\boldsymbol{\theta}} = \begin{bmatrix} K_\beta & 0 \\ 0 & K_\phi \end{bmatrix}.$$

La estimación de máxima verosimilitud para θ y ϕ puede calcularse mediante un algoritmo de mínimos cuadrados ponderados para obtener $\hat{\beta}$ y mediante el método de Newton-Raphson para estimar $\hat{\phi}$, como se detalla a continuación:

$$\beta^{(m+1)} = \left(\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{y}^{*(m)}$$

$$\phi^{(m+1)} = \phi^{(m)} - \frac{U_{\phi}^{(m)}}{\ddot{L}_{\phi\phi}^{(m)}}$$

donde $m = 0, 1, 2, \dots$, y:

$$\mathbf{y}^* = \mathbf{X}\beta + \mathbf{F}^{-1}(\mathbf{y} - \mu)$$

4. APLICACIÓN

Para el análisis de eventos de actividad criminal, se utilizan los datos recopilados y publicados por el Sistema de Información Estadístico, Delincuencial Contravencional y Operativo de la Policía Nacional – SIEDCO en el 2023. Este estudio busca identificar, mediante de los Modelos Lineales Generalizados, los factores económicos, sociales y demográficos que influyen en el volumen de hurtos a comercio que se presentan en los diferentes departamentos del país.

Los datos utilizados incluyen información detallada sobre los eventos de actividad criminal, clasificados según su ubicación, así como variables auxiliares que podrían estar relacionadas con la ocurrencia de estos eventos. Entre las covariables que se consideran relevantes para el estudio se encuentran el Producto Interno Bruto (PIB), el PIB per cápita, la Tasa de Desocupación (TD), el Coeficiente de Gini, el número de noticias criminales reportadas, la densidad poblacional, la tasa de criminalidad, y otros factores que reflejan el contexto socioeconómico y demográfico de los departamentos analizados.

El estudio se centra en analizar cómo estas variables explican el comportamiento del volumen de hurtos en los locales comerciales, permitiendo identificar patrones espaciales en las actividades criminales. Este análisis no solo proporcionará una comprensión más profunda de los factores que afectan a los eventos criminales en Colombia, sino que también podrá aportar información clave para el diseño de estrategias de prevención y mitigación del crimen en el sector retail.

4.1. Datos y Análisis Preliminar de Variables

Los datos utilizados en este análisis corresponden al año 2023 y fueron recolectados de diversas fuentes confiables. Estas variables representan factores económicos, sociales y demográficos que influyen en los eventos de actividad criminal en Colombia. La información fue procesada para garantizar su consistencia y fiabilidad, permitiendo un análisis más robusto de los patrones observados.

Variable	Descripción	Fuente
Hurto a Comercio	Número de eventos criminales registrados en comercios	Policía Nacional de Colombia
PIB	Producto Interno Bruto de cada departamento (Miles de Millones de Pesos COP)	DANE
PIB per cápita	PIB dividido entre la población de cada departamento (Pesos COP)	DANE
Tasa de Desocupación (TD)	Número de personas desocupadas sobre la población activa	DANE
Coefficiente Gini	Medida de desigualdad en la distribución del ingreso	DANE
Pobreza Monetaria	Porcentaje de la población en condición de pobreza	DANE
Noticias Criminales	Número de denuncias registradas	Corporación Excelencia en la Justicia
Población	Total de habitantes por departamento	Corporación Excelencia en la Justicia
Tasa de Criminalidad	Número de delitos registrados por cada 100.000 habitantes	Corporación Excelencia en la Justicia

Tabla 4: Descripción de las variables incluidas en el análisis.

4.1.1. Tratamiento de Datos Faltantes

Las variables Tasa de Desocupación (TD), Coeficiente de Gini y Pobreza Monetaria presentaron un 27 % de datos faltantes, afectando principalmente a departamentos periféricos o con menor disponibilidad de información estadística, como Amazonas, Arauca, Casanare, Guainía, Guaviare, Putumayo, San Andrés y Providencia, Vaupés y Vichada. Para abordar este problema, se realizó una imputación de datos basada en el promedio de los departamentos vecinos, detectados mediante una matriz de pesos espaciales construida bajo el criterio de contigüidad geográfica.

La decisión de imputar datos faltantes utilizando promedios de vecinos, basada en una matriz de contigüidad, está respaldada por la literatura en econometría espacial (LeSage & Pace 2009) y análisis geográfico (Anselin 1995). Estas técnicas han sido aplicadas con éxito en estudios relacionados con pobreza y criminalidad (Elbers et al. 2003), destacando la validez de los patrones locales como una aproximación razonable en contextos con datos incompletos.

La justificación para esta técnica radica en que los departamentos vecinos suelen compartir características socioeconómicas y demográficas similares, lo que hace que el promedio ponderado sea una aproximación razonable para los valores faltantes. Esta estrategia garantiza que la estructura espacial y las relaciones entre departamentos se preserven, evitando introducir sesgos significativos en el análisis. Además, al utilizar una matriz de contigüidad, se asegura que los valores imputados reflejen relaciones espaciales reales, fortaleciendo la validez del análisis posterior.

A continuación, se presentan Mapas Cloropléticos que permiten visualizar los patrones espaciales de las variables antes y después del proceso de imputación de datos.

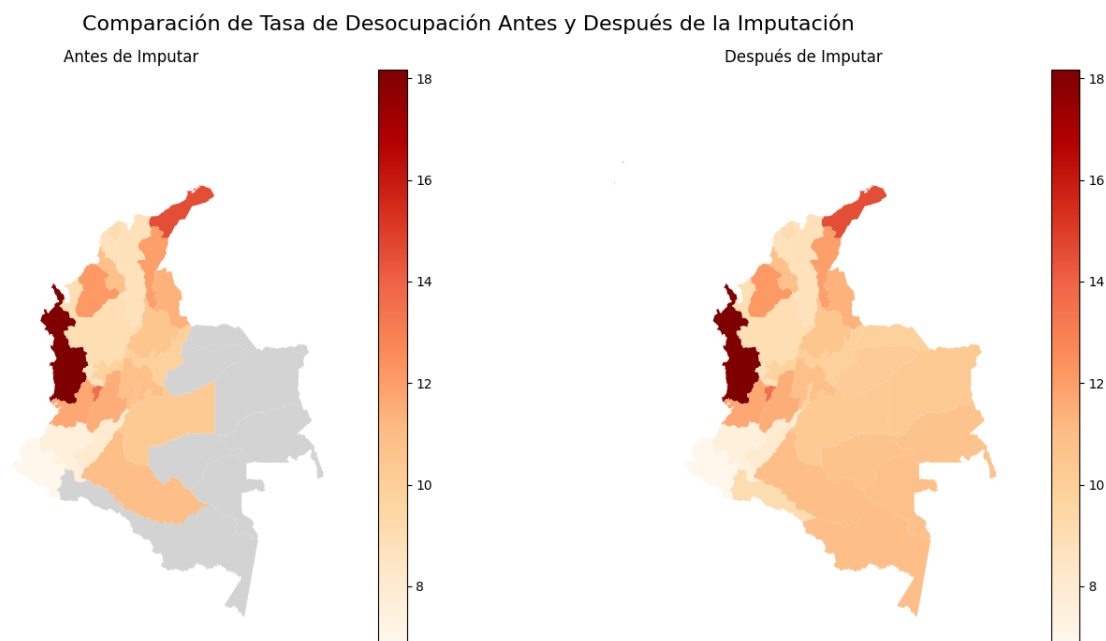


Figura 3: Imputación de la Tasa de Desocupación

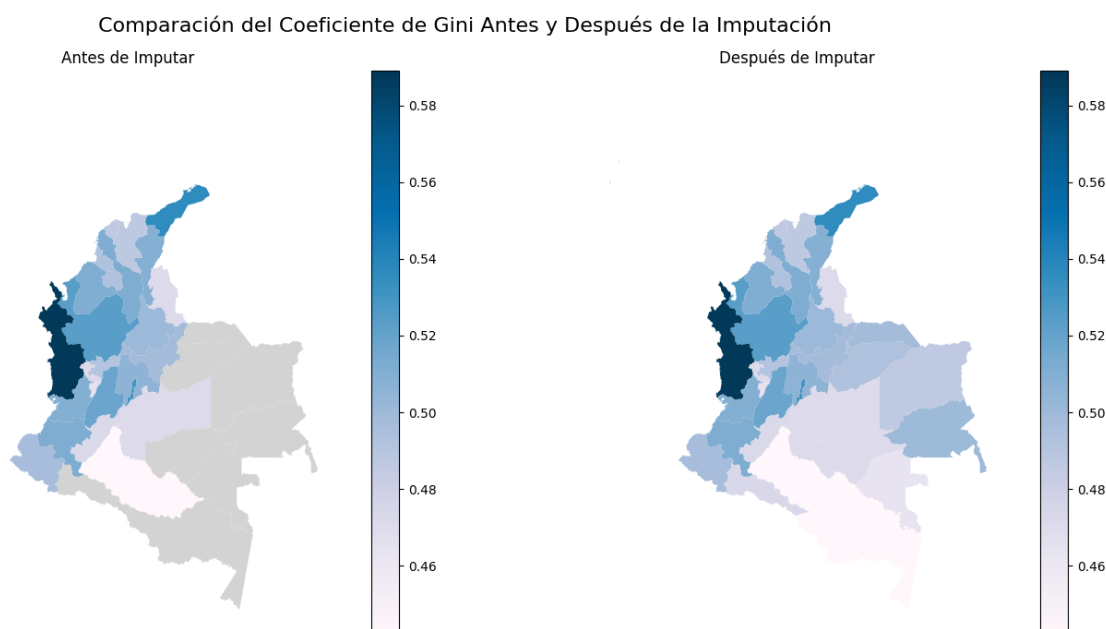


Figura 4: Imputación del Coeficiente de Gini

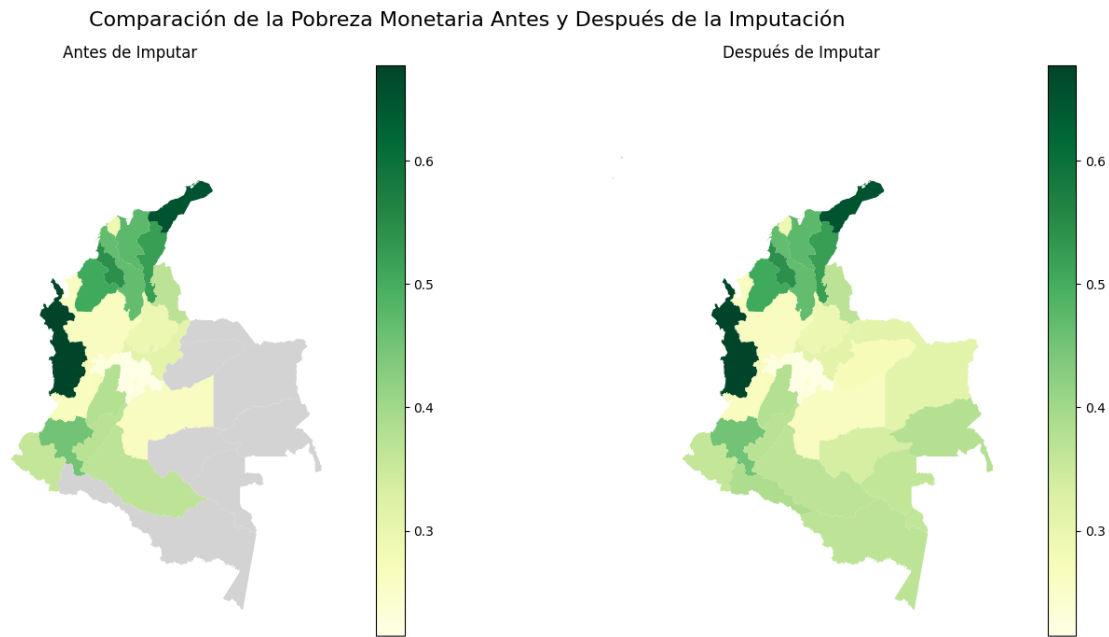


Figura 5: Imputación de la Pobreza Monetaria

Estos mapas ilustran cómo la imputación ha mejorado la cobertura de datos y preservado las relaciones espaciales en los departamentos analizados. Y su impacto en el análisis posterior es el aumento de la capacidad para identificar correlaciones entre variables de interés y se espera que permita realizar modelos más precisos, como la Regresión Poisson, al disponer de un conjunto de datos más uniforme y completo.

4.1.2. Análisis Descriptivo y Exploratorio

A continuación, se presenta un mapa de calor que permite visualizar las correlaciones entre las variables de interés. Este gráfico resalta las relaciones lineales entre las variables económicas, sociales y demográficas, y su posible influencia en los eventos de actividad criminal reportados en los diferentes departamentos. Estas correlaciones brindan una base para identificar patrones relevantes y posibles multicolinealidades entre las variables.

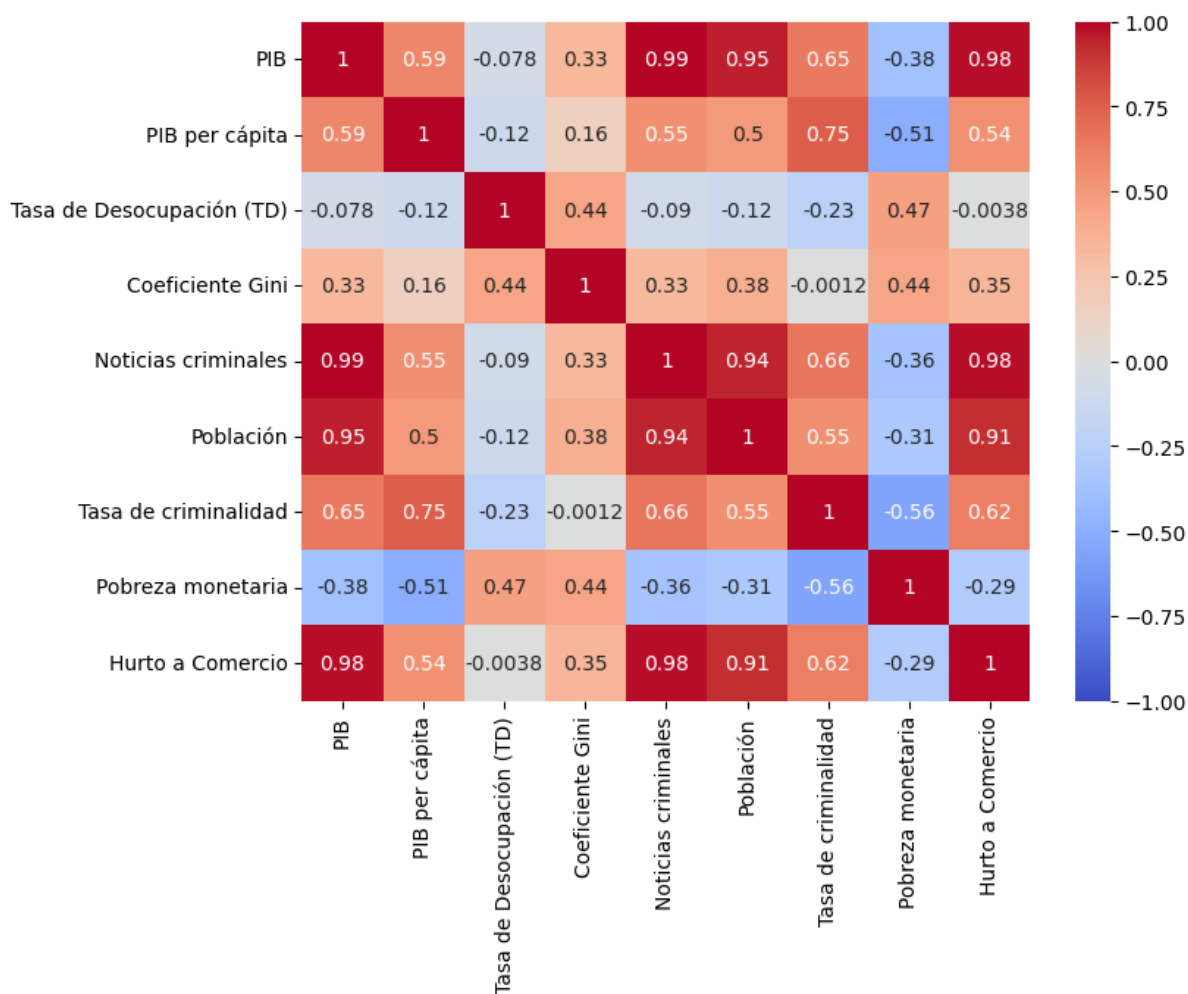


Figura 6: Mapa de Calor de las Correlaciones lineales

- **Hurto a Comercio y el PIB (0.98)**, aunque podría esperarse que mayores niveles de PIB reduzcan la actividad criminal, esta correlación sugiere lo contrario. Esto podría deberse a que las regiones con mayor PIB tienen una actividad económica más intensa, lo que las convierte en objetivos más atractivos para actividades delictivas. Este hallazgo destaca la necesidad de controlar por otras variables, como desigualdad (Coeficiente de Gini) o pobreza monetaria, para entender esta relación.
- **Hurto a Comercio y las Noticias Criminales (0.99)**, es razonable esperar una relación fuerte entre estas dos variables, ya que un mayor número de eventos criminales se espera que se realicen las denuncias respectivas, pero también esta correlación tan alta puede dar indicios de multicolinealidad.
- **El número de Eventos y la Tasa de Desocupación (-0.0038)**, la relación débil y negativa es inesperada, ya que generalmente se asocia un mayor desempleo con un aumento de la actividad criminal.
- **Coeficiente de Gini y la Tasa de Criminalidad (0.44)**, la desigualdad en la distribución del ingreso es un factor clave en la incidencia de delitos. Una correlación positiva respalda estudios previos que vinculan altos niveles de desigualdad con mayores tasas de criminalidad. Esto resalta la

importancia de incluir el Coeficiente de Gini en el modelo de análisis para capturar las dinámicas socioeconómicas.

- **Pobreza Monetaria y Tasa de Criminalidad (-0.56)**, la correlación negativa es inesperada, ya que se asume que una mayor pobreza está asociada con un aumento de la criminalidad. Esto podría estar relacionado con que los departamentos más pobres también tienen menores niveles de reporte o actividad económica, lo que reduce los eventos registrados. Este hallazgo sugiere la necesidad de usar herramientas robustas que sugieran qué variables se deben agregar en el modelo.

A continuación, se presenta un análisis de la distribución de la variable dependiente principal, Número de Eventos, mediante un histograma. Este gráfico permite visualizar la concentración de valores y posibles asimetrías en los datos, facilitando la identificación de tendencias relevantes y proporcionando una base para seleccionar las transformaciones o modelos adecuados para el análisis.

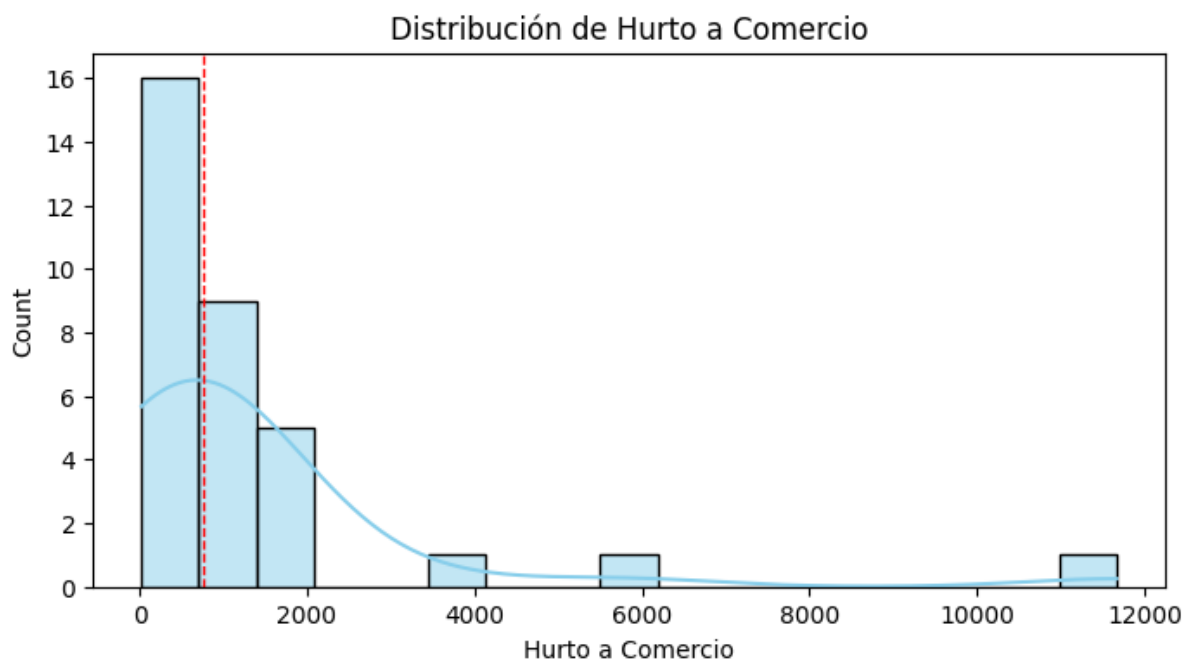


Figura 7: Distribución del número de hurtos a comercio

1. Asimetría Positiva (Sesgo a la Derecha):

- La distribución está fuertemente sesgada hacia la derecha, con una gran concentración de valores cercanos a 0 y unos pocos valores extremos elevados.
- Este comportamiento es típico en datos de conteo como los eventos de actividad criminal, donde la mayoría de los departamentos reportan un número bajo o moderado de eventos, mientras que unos pocos tienen valores excepcionalmente altos.

2. Valores Atípicos:

- Se identifican departamentos con un número de eventos significativamente superior al promedio, visibles en el extremo derecho del histograma.
- Estos valores atípicos pueden estar asociados con grandes ciudades o regiones con alta actividad económica y criminal.

3. Mediana:

- La línea punteada roja que indica la mediana, muestra que más del 50 % de los departamentos tienen un número de eventos muy por debajo del rango medio.
- Esto resalta la desigualdad en la distribución de eventos criminales entre los departamentos.

4. Cola Larga:

- La cola prolongada hacia la derecha indica que hay departamentos con valores altos pero poco frecuentes, lo que podría influir en el comportamiento de los modelos a implementar.

Implicaciones para el Modelo

- **Ajuste del Modelo:** La asimetría positiva y la presencia de valores extremos justifican el uso de un modelo como la Regresión Poisson, que es adecuado para datos de conteo con distribuciones sesgadas.
- **Transformación de Variables:** Aunque la Regresión Poisson puede manejar este tipo de distribución, podría considerarse una transformación logarítmica para estabilizar la varianza si los valores extremos tienen un impacto significativo.
- **Inclusión de Covariables:** La heterogeneidad en la distribución sugiere que las covariables (como PIB, Coeficiente de Gini y Pobreza Monetaria) pueden desempeñar un papel clave en la explicación de la variabilidad de los eventos.

El histograma de **Hurto a Comercio** confirma la necesidad de utilizar un enfoque estadístico robusto que pueda manejar la naturaleza asimétrica y dispersa de los datos. La concentración de valores bajos y la presencia de valores atípicos resaltan la importancia de incluir covariables para capturar las diferencias estructurales entre los departamentos. Este análisis preliminar proporciona una base sólida para la selección del modelo y la interpretación de los resultados.

4.2. Validación de Correlación Espacial

Para evaluar la correlación espacial de los eventos de actividad criminal, se utilizó el índice de Moran, una métrica ampliamente reconocida en el análisis espacial. El índice de Moran mide el grado de similitud espacial entre valores de una variable en ubicaciones cercanas, proporcionando una estimación cuantitativa de la autocorrelación espacial.

Teóricamente, el índice de Moran se basa en la comparación de los valores observados de una variable con un patrón aleatorio, considerando la estructura espacial definida por una matriz de pesos espaciales. Su valor oscila entre -1 y 1, donde:

- Valores positivos indican autocorrelación espacial positiva, es decir, ubicaciones cercanas tienden a tener valores similares.
- Valores negativos reflejan autocorrelación espacial negativa, es decir, ubicaciones cercanas tienden a tener valores diferentes.
- Un valor cercano a 0 indica ausencia de autocorrelación espacial.

En este estudio, el cálculo del índice de Moran se realizó utilizando la librería *esda* de Python, específicamente con el módulo *esda.moran*. Esta librería facilita el análisis estadístico espacial al proporcionar herramientas para calcular y visualizar patrones espaciales. Además, la matriz de pesos espaciales se generó considerando la vecindad entre los departamentos.

Los resultados obtenidos muestran un índice de Moran de aproximadamente 0.38 con un p-valor de 0.003. Estos valores indican una correlación espacial positiva y estadísticamente significativa, lo que sugiere que los eventos criminales tienden a agruparse en ciertos departamentos. Es importante destacar que este análisis incluye la isla de San Andrés y Providencia, de lo contrario habría reducido la correlación observada. La inclusión de este departamento permite una representación más completa de los patrones espaciales en Colombia.

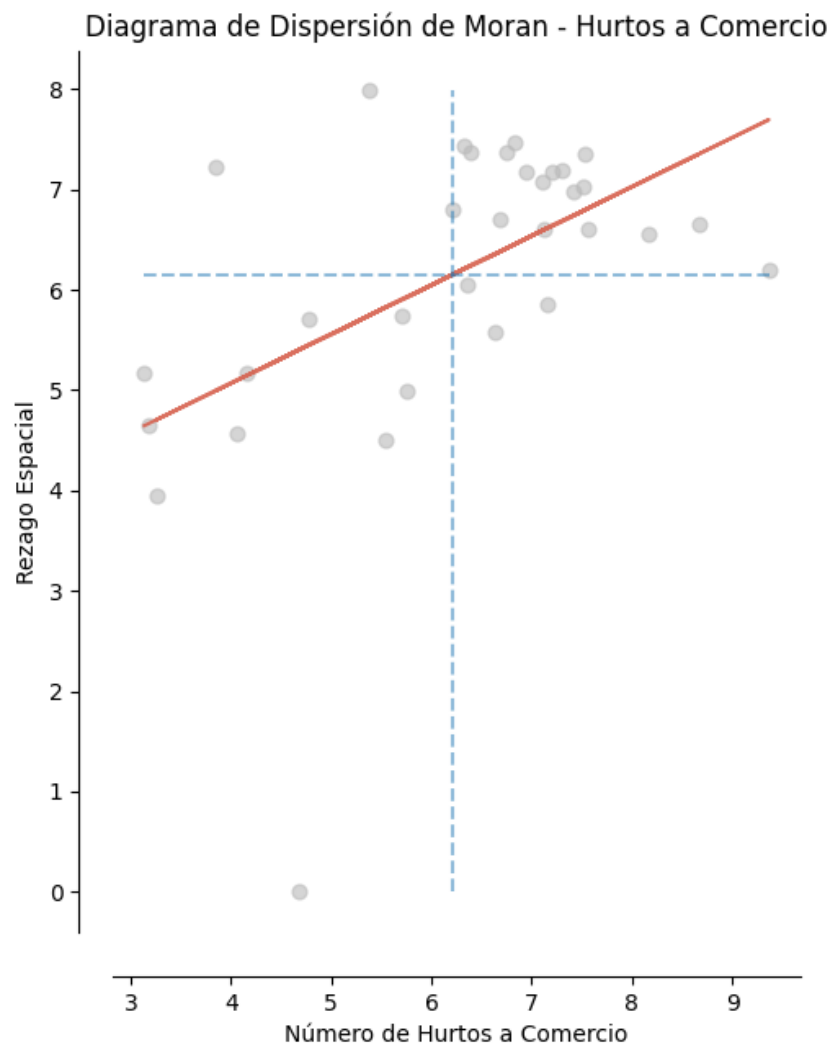


Figura 8: Diagrama de Dispersión del Índice de Moran

- **Autocorrelación Espacial Positiva:** La relación positiva sugiere que los hurtos a comercio tienden a agruparse espacialmente, es decir, las regiones con alto número de hurtos están rodeadas por regiones vecinas con valores también altos.
- **Agrupamiento Espacial:** La mayoría de los puntos están concentrados en el cuadrante I (alto-alto) y el cuadrante III (bajo-bajo), lo que refuerza la idea de que existe un agrupamiento espacial. Esto significa que los hurtos a comercio no se distribuyen de manera aleatoria, sino que tienden a formar clusters.
 - **Cluster - Cuadrante I:** Agrupa áreas donde los hurtos son altos y están rodeadas de regiones con valores altos. Estas áreas podrían ser centros urbanos o zonas comerciales importantes.

- **Cluster - Cuadrante III:** Corresponde a regiones con pocos hurtos, posiblemente áreas rurales o con menor actividad comercial.

4.3. Selección de Variables Predictoras

1. **Transformación Logarítmica a la Variable de Interés:** De acuerdo al análisis previo, se optó por realizar una transformación a la variable de interés para reducir el impacto de valores atípicos o extremos y estabilizar la varianza, lo que puede mejorar la interpretabilidad y ajuste del modelo.
2. **División de Datos y Escalado:** Los datos se dividieron en conjuntos de entrenamiento (90 %) y prueba (10 %). Posteriormente, se escaló el conjunto de variables independientes usando *StandardScaler* para normalizar las magnitudes de las variables.
3. **Selección del Mejor Valor de α :** Se utilizó *LassoCV* con validación cruzada (ésta evalúa distintos valores de penalización α) para encontrar el mejor valor de regularización.
4. **Entrenamiento del Modelo Lasso:** Se ajustó el modelo Lasso utilizando el valor óptimo de α y se evaluó el modelo con el conjunto de prueba mediante predicciones.
5. **Resultados del Modelo Lasso:** Los coeficientes obtenidos son los siguientes:

Variable	Coefficiente β	$\Delta\%$ ($1 - e^\beta$)
PIB	-1.633225	-80.47
PIB per cápita	0.583096	79.16
Noticias criminales	-0.000000	0.00
Población	2.320347	917.92
Tasa de criminalidad	0.533877	70.55
Tasa de Desocupación (TD)	0.239108	27.01
Coefficiente Gini	-0.128715	-12.07
Pobreza monetaria	0.445841	56.18

Tabla 5: Coeficientes del modelo Lasso con transformación logarítmica y cambios porcentuales.

6. **Modelo Elastic Net:** Se utilizó Elastic Net para combinar las penalizaciones L1 (Lasso) y L2 (Ridge) y seleccionar las variables más relevantes.
7. **Comparación de Modelos:**

El Modelo Lasso se destaca por su capacidad de simplificar seleccionando solo 5 variables significativas, lo que lo hace relevante a la hora de aplicar el enfoque parsimonioso. Tiene un mejor ajuste en los datos de prueba en términos de R^2 y $RMSE$, pero su capacidad de generalización es limitada (R^2 negativo en validación cruzada).

El Modelo Elastic Net mantuvo todas las variables, pero ajustó los coeficientes para reflejar mejor su relevancia. Es más estable y robusto en validación cruzada (R^2 positivo), pero con menor capacidad explicativa en los datos de prueba.

De acuerdo con lo anterior se seguirá el Modelo Lasso, porque como Box (1987) resalta que los modelos deben ser lo suficientemente simples como para ser útiles, pero no tan complejos que sean difíciles de interpretar o aplicar. Este concepto se alinea perfectamente con el principio de parsimonia, especialmente en el contexto de selección de modelos como Lasso, que busca eliminar variables irrelevantes y mantener solo aquellas que aportan significativamente al modelo.

Variable	Coefficiente β	$\Delta\%$ ($1 - e^\beta$)
PIB	-1.265363	-71.56
PIB per cápita	0.571849	77.15
Noticias criminales	-0.350658	-29.57
Población	2.306190	903.61
Tasa de criminalidad	0.585474	79.58
Tasa de Desocupación (TD)	0.251094	28.54
Coefficiente Gini	-0.174084	-15.98
Pobreza monetaria	0.502226	65.24

Tabla 6: Coeficientes del modelo Elastic Net con transformación logarítmica y cambios porcentuales.

Modelo	R ²	RMSE	MAE
Lasso	0.1940	0.9433	0.6657
Elastic Net	0.1518	0.9677	0.6583

Tabla 7: Métricas para los modelos Lasso y Elastic Net.

Modelo	R ² Promedio (CV)	Desviación Estándar (CV)
Lasso	-0.1378	0.6672
Elastic Net	0.1097	0.4674

Tabla 8: Resultados de validación cruzada para los modelos Lasso y Elastic Net.

4.4. Modelo Poisson con Efectos Espaciales

En esta sección se detalla la construcción del modelo Poisson que incorpora un componente espacial, basado en las variables seleccionadas previamente por el modelo LASSO: **PIB, PIB per cápita, Población, Tasa de criminalidad y Pobreza monetaria**. Estas variables se integraron al modelo debido a su relevancia estadística y teórica para explicar la variable dependiente, el **Hurto a Comercio**.

Adicionalmente, se incorporó un lag espacial calculado sobre la variable dependiente transformada logarítmicamente. Este lag espacial fue calculado utilizando una matriz de pesos espaciales basada en la contigüidad de los departamentos (matriz de tipo Queen), el lag espacial de una variable Y se define matemáticamente como:

$$W \cdot Y$$

donde: W es la matriz de pesos espaciales normalizada y Y es el vector de valores de la variable en las unidades espaciales.

4.4.1. Implementación

En este caso,

$$lag_hurtos = W \cdot Hurto\ a\ Comercio$$

Este término permite capturar la dependencia espacial existente entre las observaciones, reflejando la influencia de los hurtos en departamentos vecinos sobre cada unidad de análisis.

Al igual que en el modelo de Penalización Lasso, se realizará la transformación logarítmica de la variable de interés **Hurto a Comercio**, esta transformación se justifica debido a la naturaleza de los datos, que corresponden a conteos discretos y no negativos. La función de enlace logarítmica asegura que las predicciones del modelo permanezcan dentro de un rango válido (valores positivos), al tiempo que permite expresar las relaciones entre las variables en términos multiplicativos, lo cual es común en datos de conteo. Además, esta transformación facilita la interpretación de los coeficientes del modelo como cambios proporcionales en la variable de interés frente a variaciones en las variables explicativas, lo que resulta particularmente útil para identificar los factores asociados a los hurtos en locales comerciales.

Adicionalmente, la variable **Población** se utilizó como variable control, con el fin de ajustar por población (como variable de exposición), ya que permite modelar la tasa de crímenes por cada 1000 habitantes en lugar del conteo total.

La especificación matemática del Modelo Poisson con Efectos Espaciales, es:

$$\begin{aligned} \log(\mathbb{E}[\text{Hurto a Comercio}_i]) = & \beta_0 + \beta_1 \cdot \text{PIB}_i \\ & + \beta_2 \cdot \text{PIB per cápita}_i \\ & + \beta_3 \cdot \text{Tasa de criminalidad}_i \\ & + \beta_4 \cdot \text{Pobreza monetaria}_i \\ & + \beta_5 \cdot \text{Lag Hurtos}_i \\ & + \beta_6 \cdot \log(\text{Población}_i/1000) \end{aligned}$$

donde $\mathbb{E}[\text{Hurto a Comercio}_i]$ es el valor esperado de hurtos en el comercio en el departamento i por cada 1000 habitantes, y las variables independientes representan factores relevantes que influyen en esta tasa.

En la siguiente tabla, se presentan los coeficientes estimados de cada variable junto con su nivel de significancia estadística, indicado por asteriscos (***, **, *). Además, se incluyen métricas clave como el **Log-Verosimilitud**, el **Pseudo R^2** , y los **grados de libertad residuales**, las cuales permiten evaluar el ajuste y la calidad del modelo.

Este análisis progresivo permite identificar el modelo más parsimonioso y estadísticamente robusto, facilitando la interpretación de los resultados y la selección del modelo final.

Variable	Modelo I (Completo)	Modelo II (Parsimonioso)
PIB	-0.7676***	-0.7665***
PIB per cápita	0.0033	—
Tasa de criminalidad	-0.0004***	-0.0004***
Pobreza monetaria	-1.4289**	-1.4308**
Lag Hurtos	-0.5239***	-0.5238***
Log-Verosimilitud	-89.670	-89.671
Pseudo- R^2	0.3877	0.3877
GL Residuales	28	29
AIC	189.34	187.34

Nota: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Los valores — indican que la variable fue excluida del modelo.

Tabla 9: Comparación de Modelos Poisson

En un modelo Poisson con la transformación logarítmica de la variable dependiente, los coeficientes β se interpretan como el cambio en el logaritmo de la tasa relativa de hurto a comercio por cada mil habitantes asociado con un cambio unitario en la variable independiente, manteniendo constante el resto de las variables. Al aplicar la transformación exponencial e^β , se obtiene una interpretación en términos de tasas relativas:

$$1 - e^\beta$$

El resultado indica el porcentaje de cambio en la tasa de hurto a comercio por cada mil habitantes asociado con un cambio unitario en la variable independiente. A continuación, se profundiza en la interpretación de los coeficientes en los modelos presentados:

Modelo I:

1. **PIB:** Fue dividido por 10^5 , lo que implica que una unidad de PIB en el modelo corresponde a 100 mil millones de pesos en la escala original. Así pues, por cada incremento de **100 mil millones de pesos** en el PIB, se espera una disminución del 53.56 % en la tasa de hurtos a comercio por cada 1000 habitantes.
2. **PIB per cápita:** Se reescaló por 10^7 , lo que implica que una unidad de *PIB per cápita* en el modelo corresponde a **10 millones de pesos**. Entonces, por cada 10 millones de pesos de más en el PIB per cápita, se espera un aumento de aproximadamente del 0.33 % en la tasa de hurtos a comercio por cada 1000 habitantes.
3. **Tasa de Criminalidad:** Un aumento de 1 unidad en la tasa de criminalidad está asociado con una reducción muy pequeña, del 0.04 % en la tasa de hurto a comercio.
4. **Pobreza Monetaria:** Un incremento unitario en los niveles de pobreza monetaria se asocia con una reducción del 75.96 % en la tasa de hurto a comercio.
5. **Lag Espacial:** Sugiere un incremento de una unidad en el lag espacial se asocia con un impacto del -40.68 % en la tasa de hurtos a comercio por cada 1000 habitantes en el departamento que se esté analizando.

Métricas del Modelo I:

Log-Verosimilitud: -89.670, este valor refleja el ajuste del modelo a los datos. Comparado con el *Log-Verosimilitud* del modelo nulo (-146.45), el modelo ajustado mejora significativamente.

Pseudo- R^2 : 0.3877, indica que el modelo explica el 38.77 % de la variabilidad en la tasa de hurto a comercio.

AIC: 189.34, utilizado para comparar contra el modelo parsimonioso.

Modelo II:

1. **PIB:** Igual que en el Modelo I, un aumento de 100 mil millones de pesos en el PIB se asocia con una disminución del 53.54 % en la tasa de hurto a comercio.
2. **Tasa de Criminalidad:** Permanece igual que en el Modelo I.
3. **Pobreza Monetaria:** Similar al Modelo I, pero con un coeficiente levemente más fuerte, 76.04 %.
4. **Lag Espacial:** Igual que en el Modelo I.

Métricas del Modelo II:

Log-Verosimilitud: -89.671, similar al Modelo I, es decir, la omisión de la variable PIB per cápita, no interfiere en el ajuste del modelo.

Pseudo- R^2 : 0.3877, idéntico al del Modelo I.

AIC: 187.34, menor frente al del Modelo I.

Por lo tanto, el modelo final se especifica como un modelo Poisson, así:

$$\begin{aligned} \log(\mathbb{E}[\text{Hurto a Comercio}]) = & \beta_0 + \beta_1 \cdot \text{PIB}_i \\ & + \beta_2 \cdot \text{Tasa de Criminalidad}_i \\ & + \beta_3 \cdot \text{Pobreza Monetaria}_i \\ & + \beta_4 \cdot \text{Lag Hurtos}_i \\ & + \beta_6 \cdot \log(\text{Población}_i/1000) \end{aligned}$$

Índice de Moran en los Residuos: El índice de Moran calculado sobre los residuales estandarizados del modelo Poisson arrojó un valor de $I = 0.0112$ ($p = 0.405$), lo que sugiere que no existe autocorrelación espacial significativa en los residuos, el modelo ha sido eficaz capturando los patrones espaciales presentes en los datos originales, ya que los residuos no presentan dependencias espaciales adicionales. Esto valida el uso de la especificación del modelo con términos espaciales (como el lag espacial de hurtos) y ajustado por población, ya que explica adecuadamente la variabilidad espacial en los datos.

4.4.2. Validación del Supuesto de Equisdispersión

Así como se expuso en la sección teórica del Modelo de Regresión Poisson, en estos modelos se asume que la media y la varianza de la variable dependiente son iguales ($\mu = \sigma^2$). Sin embargo, en muchos casos, los datos presentan sobre-dispersión, es decir, una varianza mayor a la media, lo cual puede comprometer la validez de las inferencias del modelo. Para evaluar si el modelo Poisson ajustado presenta sobre-dispersión, se aplican los siguientes tests estadísticos:

- **Test de Cameron y Trivedi:** Evalúa la relación entre la varianza y la media. Un resultado grande sugiere que el modelo Poisson no captura adecuadamente la varianza de los datos.

Como resultado se obtuvo un valor de **2.5**, indicando una sobre-dispersión. Este resultado sugiere que la varianza de la variable dependiente excede significativamente la media, lo que contradice el supuesto fundamental del modelo *Possion* ($\mu = \sigma^2$)

- **Inspección visual:** Es una herramienta fundamental para evaluar la calidad del ajuste del modelo y diagnosticar posibles incumplimientos de los supuestos estadísticos. En el caso del modelo Poisson ajustado, se analizan los residuos de *Pearson*, que son definidos como:

$$r_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}},$$

donde y_i son los valores observados, $\hat{y}_i$ son los valores ajustados, y $\sqrt{\hat{y}_i}$ representa la desviación estándar esperada bajo el supuesto de Poisson. Los residuos de Pearson permiten detectar patrones sistemáticos que indicarían problemas como falta de ajuste, heterocedasticidad o sobre-dispersión.

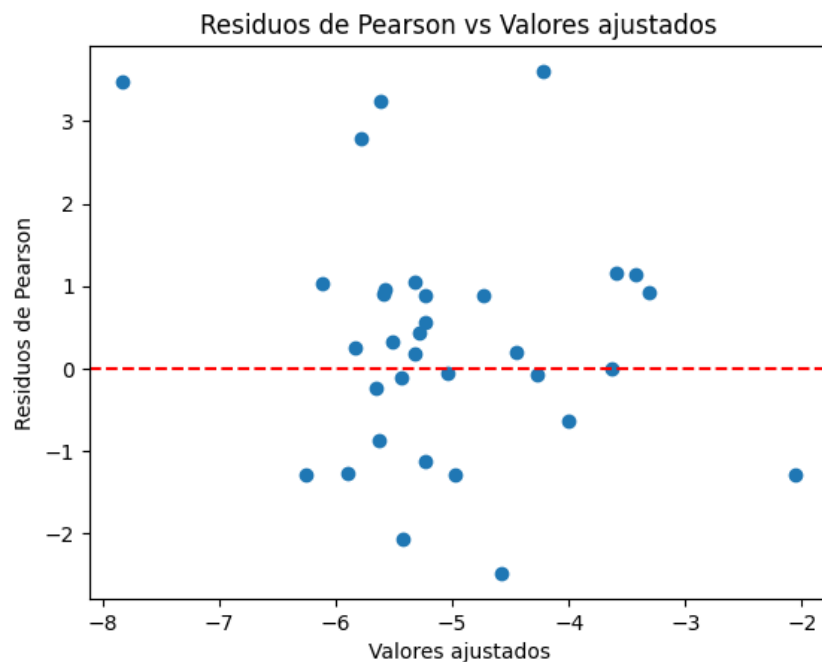


Figura 9: Residuos Pearson vs. Valores Ajustados

El gráfico de los residuos de Pearson contra los valores ajustados muestra patrones que indican un ajuste inadecuado del modelo Poisson:

- **Distribución de los Residuos:** Los residuos están distribuidos de manera aproximadamente simétrica alrededor de la línea roja, lo cual es deseable. Esto sugiere que el modelo no presenta sesgos sistemáticos en sus predicciones.
- **Homocedasticidad:** No se observa un aumento o disminución sistemática en la dispersión de los residuos a lo largo del rango de valores ajustados. La homocedasticidad (varianza constante de los residuos) parece cumplirse razonablemente bien, lo que es consistente con un modelo bien especificado.
- **Residuos Extremos:** Hay algunos puntos alejados de la mayoría de los datos, específicamente en la parte superior izquierda y en la parte inferior derecha. Estos puntos pueden corresponder a departamentos donde se ubican las principales ciudades del país.
- **Valores Ajustados:** Los valores ajustados están concentrados en un rango relativamente estrecho, lo que es típico en modelos Poisson, ya que los valores predichos están en una escala logarítmica.

En este caso, las pruebas diagnósticas realizadas previamente confirman la presencia de sobre-dispersión en los datos, lo cual invalida los supuestos del modelo Poisson para capturar adecuadamente la estructura de la variabilidad observada, así mismo, esto genera una subestimación de los errores estándar de los coeficientes en el modelo Poisson, lo que puede llevar a interpretaciones erróneas sobre la significancia estadística de las variables explicativas.

La implementación de un Modelo Binomial Negativo es esencial para manejar la sobre-dispersión presente en los datos de hurto a comercio. Este modelo no solo proporciona una especificación estadísticamente válida, sino que también mejora la precisión y robustez de las interpretaciones, permitiendo capturar adecuadamente la complejidad de los datos y los efectos espaciales subyacentes.

4.5. Modelo Binomial Negativa

Variable	β_i
PIB	-0.7290***
Tasa de Criminalidad	-0.0003***
Pobreza monetaria	-1.3604*
Lag Hurtos	-0.5241***
(α)	0.1149*
Log-Verosimilitud	-86.059
Pseudo- R^2	0.2238
GL Residuales	29
AIC	180.11
Nota: *** p<0.01, ** p<0.05, * p<0.1.	

Tabla 10: Resultados del Modelo de Regresión Negativa Binomial

A continuación, se presenta un análisis detallado de los resultados del modelo binomial negativo ajustado para la variable dependiente Hurto a Comercio, considerando las variables explicativas incluidas y los parámetros del modelo:

1. **PIB:** Por cada aumento de 100 mil millones de pesos en el PIB, la tasa de hurto a comercio disminuye en 51.97 %, manteniendo constantes las demás variables. Esta relación es estadísticamente significativa, indicando que el PIB es un factor importante en la reducción del hurto a comercio.
2. **Tasa de Criminalidad:** Por cada aumento unitario en la tasa de criminalidad, la tasa de hurto a comercio disminuye en 0.03 %. Este resultado es altamente significativo, aunque el impacto es pequeño, lo que podría reflejar efectos indirectos o saturación en la dinámica de criminalidad general.
3. **Pobreza Monetaria:** Por cada aumento de una unidad en el índice de pobreza monetaria, la tasa de hurto a comercio disminuye en 74.36 %, manteniendo constantes las demás variables. Este coeficiente no es estadísticamente significativo al nivel del 5 %, pero muestra una tendencia potencialmente relevante $p < 0.1$. Esto podría indicar que la relación entre la pobreza y los hurtos es más compleja y podría requerir variables adicionales.
4. **Lag Espacial:** Por cada aumento unitario en los hurtos en departamentos vecinos, la tasa de hurto a comercio disminuye en 40.78 %. Este resultado es altamente significativo, reforzando la importancia del efecto espacial en la dinámica de hurtos.
5. **Parámetros de Dispersión α :** El parámetro α captura la sobre-dispersión en los datos. Aunque no es altamente significativo, su valor indica que existe dispersión adicional en los datos que justifica el uso del modelo binomial negativo frente a un modelo Poisson.

Evaluación General del Modelo

- En cuanto a el Ajuste del Modelo, el R^2 de 0.2238 indica un ajuste moderado, lo cual es común en datos de conteo. Aunque no explica toda la variabilidad, captura tendencias clave.

- *PIB, Tasa de Criminalidad, y Lag Hurtos* son altamente significativos ($p < 0.05$), lo que respalda su relevancia en la explicación de la tasa de hurto a comercio.
- La inclusión del parámetro α justifica la elección del modelo binomial negativo frente al Poisson, confirmando que la varianza de los datos excede la media.

Por lo tanto, el modelo final es:

$$\begin{aligned} \log(E[\text{Hurto a Comercio}_i]) = & -0.7290 \cdot \text{PIB}_i \\ & -0.0003 \cdot \text{Tasa de Criminalidad}_i \\ & -1.3604 \cdot \text{Pobreza Monetaria}_i \\ & -0.5241 \cdot \text{Lag Hurtos}_i \\ & + \log(\text{Población}_i/1000) \end{aligned}$$

Ahora bien, el comportamiento de los Residuos del Modelo, es el siguiente:

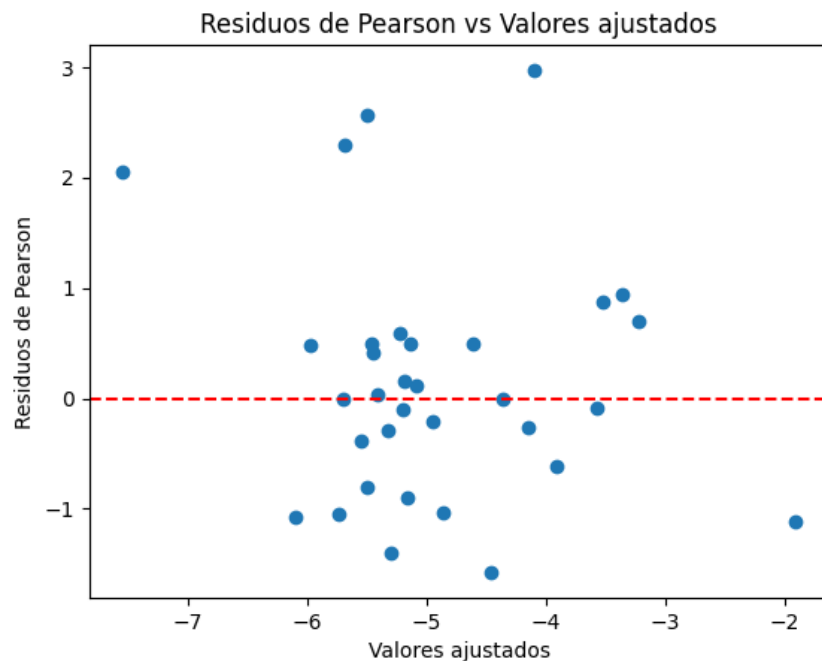


Figura 10: Residuos Pearson vs. Valores Ajustados del Modelo Binomial Negativo

El análisis de los residuos sugiere que el Modelo Binomial Negativo está bien especificado y ajusta adecuadamente los datos, capturando la sobre-dispersión presente en las observaciones. La falta de patrones evidentes en la gráfica respalda la calidad del ajuste y la validez de los supuestos del modelo.

Al igual que en el modelo Poisson, los residuos del modelo binomial negativo no presentan autocorrelación espacial significativa. Esto se evidencia por el valor del Índice de Moran = 0.0083 cercano a cero, con un p -valor de 0.345, que indica que no se rechaza la hipótesis nula de ausencia de dependencia espacial en los residuos. Estos resultados confirman que el Modelo Binomial Negativo captura adecuadamente los patrones espaciales presentes en los datos, dejando los residuos distribuidos de manera aleatoria en el espacio, lo cual valida la especificación y la calidad del ajuste del modelo.

CONCLUSIÓN Y RECOMENDACIÓN

Conclusiones

1. Validez de los Modelos Utilizados:

- La implementación de los Modelos de Regresión Poisson y Binomial Negativa demostró ser adecuada para el análisis de los hurtos a comercio en Colombia, proporcionando herramientas robustas para modelar los datos de conteo y abordar la sobre-dispersión presente en los datos.
- El Modelo Binomial Negativo, en particular, permitió capturar la variabilidad adicional de los datos al incorporar un parámetro de dispersión, logrando un ajuste superior en comparación con el Modelo Poisson.

2. Factores Asociados a los Hurtos a Comercio:

- Los resultados identificaron que variables como el Producto Interno Bruto (PIB) y los hurtos en departamentos vecinos tienen un impacto significativo en la tasa de hurtos. Estos hallazgos sugieren una relación negativa entre el desarrollo económico y la incidencia de criminalidad, así como la relevancia de los efectos espaciales en la propagación de los delitos.
- Aunque la Pobreza Monetaria mostró un efecto fuerte, su significancia estadística fue marginal, lo que resalta la necesidad de profundizar en este factor y su interacción con otras variables.

3. Ajuste por Población:

La inclusión de la población como variable de exposición (offset) permitió modelar tasas de hurto en lugar de conteos absolutos, mejorando la comparabilidad entre regiones y ofreciendo una perspectiva ajustada al contexto demográfico.

4. Validación Espacial:

La validación del Índice de Moran para los residuos confirmó la ausencia de autocorrelación espacial en los errores de los Modelo Poisson y Binomial Negativo, lo que respalda la calidad del ajuste y la correcta captura de los efectos espaciales.

Recomendaciones

1. Estrategias Basadas en Desarrollo Económico:

Los hallazgos sugieren que políticas orientadas a fomentar el crecimiento económico, especialmente en los departamentos con mayores tasas de criminalidad, podrían contribuir significativamente a la reducción de los hurtos a comercio.

2. Enfoque en la Cooperación Regional:

Dado el impacto del rezago espacial de hurtos, se recomienda implementar estrategias conjuntas entre departamentos vecinos para abordar la criminalidad desde una perspectiva regional y coordinada.

3. Refinamiento del Modelo:

Explorar modelos más complejos, como los modelos de regresión con componentes espaciales explícitos (SAR o CAR), podría proporcionar una comprensión más detallada de las dinámicas espaciales de la criminalidad.

4. Profundización en Factores Sociales:

Investigar más a fondo la relación entre la Pobreza Monetaria y la Criminalidad, posiblemente incorporando interacciones con variables económicas y de infraestructura, podría revelar patrones más claros y acciones preventivas efectivas.

APÉNDICE

Apéndice A
Datos empleados en el análisis

Departamento	PIB	PIB per cápita	TD	Coef. Gini	Noticias Criminales	Población	Tasa Criminalidad	Pobreza	Hurto a Comercio
Amazonas	711.75	13997736.23	-	-	1585	83808	1891.23	-	57
Antioquia	146798.19	33738353.41	8.99	0.524	247292	6994792	3535.37	0.264	5808
Arauca	4761.26	27403847.11	-	-	7034	308301	2281.54	-	63
Atlántico	44713.15	25039898.48	9.12	0.486	95625	2835509	3372.41	0.292	1671
Bogotá D.C.	261654.50	49730969.06	10.8	0.53	497295	7968095	6241.08	0.237	11668
Bolívar	35724.60	25197611.83	8.76	0.512	60255	2258929	2667.41	0.466	1353
Boyacá	24814.80	32177790.59	9.90	0.498	41395	1267378	3266.19	0.311	792
Caldas	15362.07	24943015.17	9.77	0.491	27815	1046418	2658.12	0.223	558
Caquetá	3804.62	14536951.43	10.98	0.442	13016	423857	3070.85	0.368	316
Casanare	12989.64	48146498.67	-	-	15283	444602	3437.46	-	300
Cauca	16898.08	18172817.15	7.54	0.512	43943	1528076	2875.71	0.453	578
Cesar	15794.62	23932464.12	12.00	0.509	35719	1359719	2626.94	0.522	1492
Chocó	3699.16	11420488.97	18.16	0.589	9353	557654	1677.20	0.677	215
Córdoba	16475.18	14614790.49	12.14	0.511	27712	1868166	1483.38	0.504	592
Cundinamarca	59702.37	28847163.49	10.81	0.506	119785	3577177	3348.59	0.215	1935
Guainía	337.44	9837191.86	-	-	1040	52627	1976.17	-	25
Guaviare	779.76	12821768.23	-	-	3581	92281	3880.54	-	254
Huila	14803.27	22376652.87	8.07	0.472	45002	1149598	3914.59	0.373	46
La Guajira	8559.60	18870183.92	14.53	0.536	15533	1015909	1528.98	0.653	1831
Magdalena	12856.67	14564363.90	8.79	0.487	35947	1476665	2434.34	0.476	849
Meta	31305.64	47525775.50	10.32	0.471	47156	1088749	4331.21	0.264	1284
Nariño	14315.39	13726930.19	6.89	0.497	46942	1631117	2877.91	0.361	763
Norte de Santander	15326.07	14714945.09	11.41	0.471	47698	1658835	2875.39	0.37	1031
Putumayo	3420.26	15128148.78	-	-	9242	374042	2470.85	-	118
Quindío	7788.60	23365782.80	13.66	0.465	20138	577543	3486.84	0.272	918
Risaralda	16074.57	26939276.74	8.75	0.471	30427	988091	3079.37	0.271	1224
San Andrés	1501.71	36188950.06	-	-	2745	65663	4180.44	-	106
Santander	61357.52	42893135.62	10.50	0.501	88983	2340657	3801.62	0.296	1871
Sucre	8029.32	12642902.65	10.84	0.492	17478	981727	1780.33	0.545	494
Tolima	19622.25	24308335.76	11.53	0.518	53892	1350060	3991.82	0.381	1231
Valle del Cauca	97380.04	32875773.41	11.72	0.51	171784	4622132	3716.55	0.266	3542
Vaupés	272.89	8984688.76	-	-	674	51133	1318.13	-	23
Vichada	620.98	8589753.94	-	-	1565	116944	1338.25	-	22

Apéndice B

Código en Python

```
#Librerías utilizadas
import zipfile
import os
import geopandas as gpd
import pandas as pd
import numpy as np
from libpysal.weights import Queen
from esda.moran import Moran
import matplotlib.pyplot as plt
from plot.esda import moran_scatterplot
from sklearn.impute import KNNImputer
import seaborn as sns
from sklearn.linear_model import Lasso
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LassoCV
import statsmodels.api as sm
from statsmodels.discrete.truncated_model import HurdleCountModel
from statsmodels.stats.outliers_influence import variance_inflation_factor
import libpysal
from sklearn.linear_model import Ridge, ElasticNet
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from sklearn.model_selection import cross_val_score

# CORRELACIÓN ESPACIAL

df = pd.read_excel('/content/Variables Auxiliares.xlsx')

#Datos Geográficos de los departamentos de Colombia
zip_folder = '/content/gadm41_COL_shp.zip'

# Extraer el contenido del archivo .zip en una carpeta
with zipfile.ZipFile(zip_folder, 'r') as zip_ref:
    zip_ref.extractall('/content/gadm41_COL_shp')

# Ver los archivos extraídos
os.listdir('/content/gadm41_COL_shp')
# Ruta al archivo .shp descomprimido
shapefile_path = '/content/gadm41_COL_shp/gadm41_COL_1.shp'

# Cargar el shapefile
gdf = gpd.read_file(shapefile_path)
#gdf

# Número de hurtos por departamento
df_agrupado = df.groupby(['DEPARTAMENTOS']).agg({'Hurto a Comercio': 'sum'}).reset_index()
df_agrupado

# Unión de los Dataframes - Hurtos con la información Geográfica
```

```
gdf_eventos = gdf.merge(df_agrupado, left_on='NAME_1', right_on='DEPARTAMENTOS', how='left')
gdf_eventos

#Cálculo del Índice de Moran
np.random.seed(999)
# Matriz de pesos espaciales utilizando el método Queen Contiguity
w = Queen.from_dataframe(gdf_eventos)

# Hurtos
eventos = np.log(gdf_eventos['Hurto a Comercio'])

# Identificar los departamentos sin vecinos
sin_vecinos = [index for index, neighbors in w.neighbors.items() if len(neighbors) == 0]

# Excluir los departamentos sin vecinos del análisis
gdf_eventos_sin_islas = gdf_eventos.drop(sin_vecinos)

# Matriz de pesos espaciales sin islas
w_sin_islas = Queen.from_dataframe(gdf_eventos_sin_islas)

# Excluir los departamentos sin vecinos del análisis
gdf_eventos_sin_islas = gdf_eventos.drop(sin_vecinos)

# Hurtos sin islas
eventos_sin_islas = gdf_eventos_sin_islas['Hurto a Comercio'].values

# Cálculo del índice de Moran
moran = Moran(eventos, w)
moran_si = Moran(eventos_sin_islas, w_sin_islas)

# Resultados
print(f'Índice de Moran: {moran.I}')
print(f'p-value: {moran.p_sim}')
moran_si.I, moran_si.p_sim

# Crear el diagrama de dispersión de Moran
fig, ax = moran_scatterplot(moran, moran.p_sim)
ax.set_title('Diagrama de Dispersión de Moran - Hurtos a Comercio')
ax.set_xlabel('Número de Hurtos a Comercio')
ax.set_ylabel('Rezago Espacial')
plt.show()

# IMPUTACIÓN DE DATOS DE LAS VARIABLES AUXILIARES
# Calcular el promedio nacional para las tres variables
mean_nacional_TD = gdf_va['Tasa de Desocupación (TD)'].mean()
mean_nacional_Gini = gdf_va['Coeficiente Gini'].mean()
mean_nacional_Pobreza = gdf_va['Pobreza monetaria'].mean()

for index, row in gdf_va.iterrows():

    if pd.isnull(row['Tasa de Desocupación (TD)']):
        neighbors = w.neighbors[index]
        neighbor_values = gdf_va.loc[neighbors, 'Tasa de Desocupación (TD)'].dropna()
```

```

    if len(neighbor_values) > 0:
        gdf_va.at[index, 'Tasa de Desocupación (TD)'] = neighbor_values.mean()
    else:
        gdf_va.at[index, 'Tasa de Desocupación (TD)'] = mean_nacional_TD

# Imputar Coeficiente Gini
if pd.isnull(row['Coeficiente Gini']):
    neighbor_values = gdf_va.loc[neighbors, 'Coeficiente Gini'].dropna()

    if len(neighbor_values) > 0:
        gdf_va.at[index, 'Coeficiente Gini'] = neighbor_values.mean()
    else:
        gdf_va.at[index, 'Coeficiente Gini'] = mean_nacional_Gini

# Imputar Pobreza monetaria
if pd.isnull(row['Pobreza monetaria']):
    neighbor_values = gdf_va.loc[neighbors, 'Pobreza monetaria'].dropna()

    if len(neighbor_values) > 0:
        gdf_va.at[index, 'Pobreza monetaria'] = neighbor_values.mean()
    else:
        gdf_va.at[index, 'Pobreza monetaria'] = mean_nacional_Pobreza

#Preparación de Datos para graficar Mapas Cloropléticos

gdf = gdf.rename(columns={"NAME_1": "Departamento"})
gdf_va = gdf_va.rename(columns={"NAME_1": "Departamento"})
df = df.rename(columns={"DEPARTAMENTOS": "Departamento"})

gdf_before = gdf.merge(df, on="Departamento")
gdf_after = gdf.merge(gdf_va, on="Departamento")

gdf_before = gdf_before.set_geometry("geometry")
gdf_after = gdf_after.set_geometry("geometry_y")

# Creación de los mapas
fig, ax = plt.subplots(1, 2, figsize=(15, 7))

# Mapa antes de la imputación
gdf_before.plot(column="Tasa de Desocupación (TD)", cmap="OrRd", legend=True, ax=ax[0],
                 missing_kwds={"color": "lightgrey", "label": "Datos faltantes"})
ax[0].set_title("Antes de Imputar")
ax[0].axis("off")

# Mapa después de la imputación
gdf_after.plot(column="Tasa de Desocupación (TD)", cmap="OrRd", legend=True, ax=ax[1])
ax[1].set_title("Después de Imputar")
ax[1].axis("off")

# Ajustar y mostrar

```

```
plt.suptitle("Comparación de Tasa de Desocupación Antes y Después de la Imputación", fontsize=16)
plt.tight_layout()
plt.show()

# Crear gráficos lado a lado para antes y después
fig, ax = plt.subplots(1, 2, figsize=(15, 7))

# Mapa antes de la imputación
gdf_before.plot(column="Coeficiente Gini", cmap="PuBu", legend=True, ax=ax[0],
                 missing_kws={"color": "lightgrey", "label": "Datos faltantes"})
ax[0].set_title("Antes de Imputar")
ax[0].axis("off")

# Mapa después de la imputación
gdf_after.plot(column="Coeficiente Gini", cmap="PuBu", legend=True, ax=ax[1])
ax[1].set_title("Después de Imputar")
ax[1].axis("off")

# Ajustar y mostrar
plt.suptitle("Comparación del Coeficiente de Gini Antes y Después de la Imputación", fontsize=16)
plt.tight_layout()
plt.show()

# Crear gráficos lado a lado para antes y después
fig, ax = plt.subplots(1, 2, figsize=(15, 7))

# Mapa antes de la imputación
gdf_before.plot(column="Pobreza monetaria", cmap="YlGn", legend=True, ax=ax[0],
                 missing_kws={"color": "lightgrey", "label": "Datos faltantes"})
ax[0].set_title("Antes de Imputar")
ax[0].axis("off")

# Mapa después de la imputación
gdf_after.plot(column="Pobreza monetaria", cmap="YlGn", legend=True, ax=ax[1])
ax[1].set_title("Después de Imputar")
ax[1].axis("off")

# Ajustar y mostrar
plt.suptitle("Comparación de la Pobreza Monetaria Antes y Después de la Imputación", fontsize=16)
plt.tight_layout()
plt.show()

# Calcular la matriz de correlación
matriz_corr = gdf_va.select_dtypes(include=['float64', 'int64']).corr()

# Mapa de calor para la matriz de correlación
plt.figure(figsize=(8, 6))
sns.heatmap(matriz_corr, annot=True, cmap='coolwarm', vmin=-1, vmax=1)

plt.show()

#Histogramas
for column in matriz_corr:
```

```

plt.figure(figsize=(8, 4))
sns.histplot(data=gdf_va, x=column, kde=True, color='skyblue')
plt.axvline(gdf_va[column].median(), color='red', linestyle='dashed', linewidth=1)
plt.title(f'Distribución de {column}')
plt.show()

# Validación de media = varianza variable y
y_mean = gdf_va[['Hurto a Comercio']].mean()
y_mean
y_var = gdf_va[['Hurto a Comercio']].var()
y_var

# REGRESIÓN LASSO
y = np.log(gdf_va['Hurto a Comercio'])
X = gdf_va[['PIB', 'PIB per cápita', 'Noticias criminales', 'Población', 'Tasa de criminalidad', 'Tasa de

# División de los datos en conjunto de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=123)

# Escalar los datos
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
print(X_train_scaled.shape) # Dimensión de X_train_scaled
print(y_train.shape)
print(np.isnan(X_train_scaled).sum())

# Rango de valores de alpha
alphas = np.logspace(-2, 2, 50)

# LassoCV para encontrar el mejor alpha
lasso_cv = LassoCV(alphas=alphas, cv=3)
lasso_cv.fit(X_train_scaled, y_train)

print(f"El mejor valor de alpha encontrado: {lasso_cv.alpha_}")

lasso = Lasso(alpha=lasso_cv.alpha_)
lasso.fit(X_train_scaled, y_train)

#Evaluar el modelo
y_pred = lasso.predict(X_test_scaled)
print(f"Coeficientes: {lasso.coef_}")
print(f"Intercepto: {lasso.intercept_}")

# REGRESIÓN ELASTIC NET
# Ridge
ridge = Ridge()
ridge_params = {'alpha': np.logspace(-2, 4, 50)}
ridge_cv = GridSearchCV(ridge, ridge_params, cv=3)
ridge_cv.fit(X_train_scaled, y_train)

print(f"Mejor alpha para Ridge: {ridge_cv.best_params_}")

```

```

# Elastic Net
elastic_net = ElasticNet()
elastic_net_params = {
    'alpha': np.logspace(-2, 4, 50),
    'l1_ratio': np.linspace(0.1, 0.9, 9)
}
elastic_net_cv = GridSearchCV(elastic_net, elastic_net_params, cv=3)
elastic_net_cv.fit(X_train_scaled, y_train)

print(f"Mejor alpha y l1_ratio para Elastic Net: {elastic_net_cv.best_params_}")

elastic_net = ElasticNet(alpha=0.01, l1_ratio=0.1)
elastic_net.fit(X_train_scaled, y_train)

# Evaluar el modelo
y_pred_en = elastic_net.predict(X_test_scaled)

# Métricas de rendimiento
def evaluar_modelo(y_true, y_pred):
    r2 = r2_score(y_true, y_pred)
    rmse = np.sqrt(mean_squared_error(y_true, y_pred))
    mae = mean_absolute_error(y_true, y_pred)
    return r2, rmse, mae

# Predicciones con Lasso
y_pred_lasso = lasso.predict(X_test_scaled)
r2_lasso, rmse_lasso, mae_lasso = evaluar_modelo(y_test, y_pred_lasso)

print(f"Lasso - R²: {r2_lasso:.4f}, RMSE: {rmse_lasso:.4f}, MAE: {mae_lasso:.4f}")

# Predicciones con Elastic Net
y_pred_elastic = elastic_net.predict(X_test_scaled)
r2_elastic, rmse_elastic, mae_elastic = evaluar_modelo(y_test, y_pred_elastic)

print(f"Elastic Net - R²: {r2_elastic:.4f}, RMSE: {rmse_elastic:.4f}, MAE: {mae_elastic:.4f}")

# Validación cruzada para Lasso
lasso_cv_scores = cross_val_score(lasso, X_train_scaled, y_train, cv=5, scoring='r2')
print(f"Lasso - Validación Cruzada R²: {lasso_cv_scores.mean():.4f} ± {lasso_cv_scores.std():.4f}")

# Validación cruzada para Elastic Net
elastic_net_cv_scores = cross_val_score(elastic_net, X_train_scaled, y_train, cv=5, scoring='r2')
print(f"Elastic Net - Validación Cruzada R²: {elastic_net_cv_scores.mean():.4f} ± {elastic_net_cv_scores.std():.4f}")

# MODELO POISSON CON COMPONENTE ESPACIAL

# Normalización de pesos
w.transform = 'r'
y = np.log(gdf_va['Hurto a Comercio'])
# X = X = gdf_va[['PIB', 'PIB per cápita', 'Población', 'Tasa de criminalidad', 'Pobreza monetaria']]
X = X = gdf_va[['PIB', 'Población', 'Tasa de criminalidad', 'Pobreza monetaria']]
X['PIB'] = X['PIB']/1e5

```

```
#X['PIB per cápita'] = X['PIB per cápita']/1e7

# Cálculo el lag espacial de la variable dependiente
gdf_va['lag_hurtos'] = w.sparse @ np.log(gdf_va['Hurto a Comercio'])

# Agregarlo como predictor al modelo Poisson
X['lag_hurtos'] = gdf_va['lag_hurtos']

# Ajustar el modelo en statsmodels
import statsmodels as sm

model = sm.discrete.discrete_model.Poisson(y, X.drop(columns='Población'), exposure=X['Población']/1

model.summary()

print("AIC:", model.aic)

# Índice de Moran en los residuos
# Valores predichos por el modelo
predicted_values = model.predict()

# Valores observados
observed_values = model.model.endog

# Estandarización de los residuales de Pearson
residuals_pearson = model.resid_pearson
standardized_residuals_pearson = residuals_pearson / np.std(residuals_pearson)

moran = Moran(standardized_residuals_pearson, w)

print(f"Moran's I: {moran.I}, p-value: {moran.p_sim}")

# Test estadístico para sobre-dispersión
# Residuos de Pearson
residuals_pearson = model.resid_pearson

# Test de Cameron y Trivedi
phi = np.sum(residuals_pearson**2) / model.df_resid
print(f"Dispersión (phi): {phi}")

if phi > 1:
    print("Posible sobre-dispersión detectada.")
elif phi < 1:
    print("Posible sub-dispersión detectada.")
else:
    print("No hay evidencia de sobre/sub-dispersión.")

# Inspección visual de los residuos

# Residuos de Pearson y predicciones
residuals_pearson = model.resid_pearson
fitted_values = model.fittedvalues
```

```
plt.scatter(fitted_values, residuals_pearson)
plt.axhline(0, color='red', linestyle='--')
plt.xlabel('Valores ajustados')
plt.ylabel('Residuos de Pearson')
plt.title('Residuos de Pearson vs Valores ajustados')
plt.show()

# MODELO BINOMIAL NEGATIVO CON COMPONENTE ESPACIAL
model_nb = sm.discrete.discrete_model.NegativeBinomial(y, X.drop(columns='Población'),
               exposure=X['Población'] / 1000).fit()

model_nb.summary()

print("AIC:", model_nb.aic)

import matplotlib.pyplot as plt

# Residuos de Pearson y predicciones
residuals_pearson = model_nb.resid_pearson
fitted_values = model_nb.fittedvalues

plt.scatter(fitted_values, residuals_pearson)
plt.axhline(0, color='red', linestyle='--')
plt.xlabel('Valores ajustados')
plt.ylabel('Residuos de Pearson')
plt.title('Residuos de Pearson vs Valores ajustados')
plt.show()

# Índice de Moran para los residuos
# Valores predichos por el modelo
predicted_values = model_nb.predict()

# Valores observados
observed_values = model_nb.model.endog

# Estandarización de los residuales de Pearson
residuals_pearson = model_nb.resid_pearson
standardized_residuals_pearson = residuals_pearson / np.std(residuals_pearson)

moran = Moran(standardized_residuals_pearson, w)

print(f"Moran's I: {moran.I}, p-value: {moran.p_sim}")
```

Referencias

- Agresti, A. (2015), *Foundations of Linear and Generalized Linear Models*, Wiley.
- Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.
*<https://ieeexplore.ieee.org/document/1100705>
- Anselin, L. (1995), 'Local indicators of spatial association—lisa', *Geographical Analysis* **27**(2), 93–115.
- Arango Sandoval, S. & Restrepo Orrego, C. (2020), 'Factores socioeconómicos y variables de disuasión que explican la criminalidad en los municipios de colombia', *Revista EAFIT* **55**(2).
*<https://repository.eafit.edu.co/server/api/core/bitstreams/dc2b2cba-bbc5-462a-a4fc-77256380c8a1/content>
- Biggeri, A., Marchi, M., Lagazio, C., Martuzzi, M. & Böning, D. (2000), 'Analysis of medical count data using regression models', *Medical Statistics Quarterly* **56**(2), 15–22.
*<https://www.jstor.org/stable/23456789>
- Box, G. E. P. & Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, Wiley, New York.
- Burnham, K. P. & Anderson, D. R. (2002), *Model selection and multimodel inference: A practical information-theoretic approach*, Springer Science & Business Media.
*<https://link.springer.com/book/10.1007/b97636>
- Cameron, A. C. & Trivedi, P. K. (1986), 'Econometric models based on count data: Comparisons and applications of estimators and tests', *Journal of Applied Econometrics* **1**(1), 29–53.
- Cameron, A. C. & Trivedi, P. K. (2013), *Regression Analysis of Count Data*, 2nd edn, Cambridge University Press.
- Claeskens, G. & Hjort, N. L. (2008), *Model selection and model averaging*, Cambridge University Press.
*<https://www.cambridge.org/core/books/model-selection-and-model-averaging/0A3D3CB63EDAA8E3179F9199A34861AB>
- Contreras Vilca, N. (2013), Análisis de votos electorales usando modelos de regresión para datos de conteo, Tesis de maestría, Pontificia Universidad Católica del Perú.
*<https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/4472>
- Cox, D. & Snell, E. (1989), *Analysis of Binary Data*, Chapman and Hall/CRC.
*<https://www.routledge.com/Analysis-of-Binary-Data-Second-Edition/Cox-Snell/p/book/9780412306204>
- De Moivre, A. (1718), *The Doctrine of Chances: Or, a Method of Calculating the Probabilities of Events in Play*, W. Pearson, London.
*<https://archive.org/details/doctrineofchance00moiv>
- Developers, P. (2024), 'esda: Exploratory spatial data analysis for python',
[urlhttps://pysal.org/esda/](https://pysal.org/esda/). Accessed: 2024-11-24.
- Dobson, A. J. & Barnett, A. (2018), *An Introduction to Generalized Linear Models*, 4th edn, CRC Press.
- Dobson, A. J. & Barnett, A. G. (2008), *An Introduction to Generalized Linear Models*, CRC Press.
*<https://www.routledge.com/An-Introduction-to-Generalized-Linear-Models/Dobson-Barnett/p/book/9781584889502>
- Elbers, C., Lanjouw, J. O. & Lanjouw, P. (2003), 'Micro-level estimation of poverty and inequality', *Econometrica* **71**(1), 355–364.

- Epidat (2014), 'Manual del programa epidat 4.0: Análisis estadístico de datos epidemiológicos'.
*<https://www.paho.org/hq/dmdocuments/2014/epidat-4-0-manual.pdf>
- García-Crespo, E. (2001), 'Count data models in economic research', *Economic Analysis Quarterly* **47**(3), 89–110.
*<https://www.jstor.org/stable/23456790>
- Gómez, J. & Morales, A. (2016), 'La percepción de inseguridad colombiana: Sus consecuencias y determinantes', *Revista de Estudios Sociales* (56), 14–28.
*<https://journal.universidadean.edu.co/index.php/plou/article/download/1574/1505/5160>
- Gómez, J., Pérez, L. & Díaz, M. (2019), 'Estrategias de seguridad en el comercio minorista: análisis y propuestas', *Revista de Seguridad Ciudadana* **12**(3), 45–62.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn, Springer Science & Business Media, New York, NY.
- Hilbe, J. M. (2014), *Modeling Count Data*, Cambridge University Press.
- King, G. (1998), *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*, University of Michigan Press, Ann Arbor, MI.
*https://www.press.umich.edu/9389/unifying_political_methodology
- Krzanowski, W. (1998), *An Introduction to Statistical Modelling*, Arnold, London, UK.
- LeSage, J. P. & Pace, R. K. (2009), *Introduction to Spatial Econometrics*, Chapman and Hall/CRC, Boca Raton, FL.
- LeSage, J. P. & Pace, R. K. (2020), *Introduction to Spatial Econometrics*, Chapman and Hall/CRC.
*<https://www.crcpress.com/Introduction-to-Spatial-Econometrics/LeSage-Pace/p/book/9780367322102>
- Lindsey, J. (1995), *Modelling Frequency and Count Data*, Clarendon Press, Oxford, UK.
- Lindsey, J. K., Jones, B. & Jarvis, P. (2001), 'Modelling count data in pharmacology', *Journal of Biopharmaceutical Statistics* **11**(3), 179–191.
*<https://www.tandfonline.com/doi/abs/10.1081/BIP-100107794>
- López Vásquez, P. A. (2016), 'Modelos de regresión poisson y binominal negativa en modelos lineales generalizados aplicados a datos correspondientes a accidentes de tránsito y lesionados entre los años 2011 y 2014 en la región de valparaíso'.
*<https://repositoriobibliotecas.uv.cl/items/ac5b6148-bfa5-43ba-a65a-1cd8ba11a90b>
- Martínez Lobo, D. S. (2018), Regresión binomial negativa: Extensión del modelo lineal generalizado poisson para el análisis de datos sobredispersos, Tesis de maestría, Universidad Nacional de Colombia.
*<https://repositorio.unal.edu.co/handle/unal/69311>
- McCullagh, P. & Nelder, J. A. (1989a), *Generalized Linear Models*, 2nd edn, Chapman and Hall/CRC.
- McCullagh, P. & Nelder, J. A. (1989b), *Generalized Linear Models*, Vol. 37, CRC Press.
*<https://www.routledge.com/Generalized-Linear-Models/McCullagh-Nelder/p/book/9780412317606>
- McFadden, D. (1974), 'Conditional logit analysis of qualitative choice behavior', pp. 105–142.
*<https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf>
- Meliciani, V. (2000), 'Innovation and count models in economics', *Journal of Economics and Statistics* **34**(1), 120–134.
*<https://www.jstor.org/stable/23456791>

- Montmort, P. R. (1714), *Essay d'analyse sur les jeux de hazard*, J. Quillau, Paris.
*<https://archive.org/details/essaydanalysesur00mont>
- Montoya, C. & Ruiz, F. (2021), 'Seguridad y desarrollo urbano: un enfoque integral para la sostenibilidad', *Revista Colombiana de Sociología* **34**(2), 101–115.
- Moran, P. A. P. (1950), 'Notes on continuous stochastic phenomena', *Biometrika* **37**(1/2), 17–23.
- Mullahy, J. (1986), 'Specification and testing of some modified count data models', *Journal of Econometrics* **33**(3), 341–365.
*[https://doi.org/10.1016/0304-4076\(86\)90002-3](https://doi.org/10.1016/0304-4076(86)90002-3)
- Nagelkerke, N. J. (1991), 'A note on a general definition of the coefficient of determination', *Biometrika* **78**(3), 691–692.
*<https://doi.org/10.1093/biomet/78.3.691>
- Navarro, V., Utzet, M., Puig, A., Caminal, J. & Martín, M. (2001), 'Epidemiology of occupational health: Count data in practice', *Occupational Medicine Reviews* **18**(3), 45–59.
*<https://academic.oup.com/occmed/article/51/3/145/1430373>
- Núñez Rivera, L. C., Tolentino Pulido, F. & Rodríguez Barrios, H. (2023), 'Factores sociodemográficos en la dinámica del comportamiento delictivo: Análisis descriptivo de criminalidad en colombia, año 2022', *Criminalidad* **65**(3), 161–280.
*https://www.scielo.org.co/scielo.php?pid=S1794-310820230003000161script=sci_arttext
- Osgoode, W. (2000), 'Criminal careers and count data models', *Criminology Journal* **38**(4), 579–606.
*<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-9125.2000.tb00905.x>
- Pascal, B. (1665), *Traité du triangle arithmétique*, Guillaume Desprez, Paris.
*<https://gallica.bnf.fr/ark:/12148/bpt6k62995>
- Paula, G. A. (2010), *Modelos de Regressão: com apoio computacional*, Universidade de São Paulo, São Paulo, Brasil.
*<https://repositorio.usp.br/item/001392542>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, (2011), 'Scikit-learn: Machine learning in python', *Journal of Machine Learning Research* **12**, 2825–2830.
- Poisson, S. D. (1837), *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*, Bachelier, Paris.
*<https://archive.org/details/recherchessurlap00pois>
- Policía Nacional de Colombia (2022), Estadísticas de criminalidad 2022, Informe oficial, Policía Nacional de Colombia.
*<https://www.policia.gov.co>
- Rincón, A. & Sánchez, J. (2020), 'Impacto económico de la inseguridad en locales comerciales: Un estudio de caso en bogotá', *Economía y Sociedad* **29**(1), 77–92.
- Romero, C., Arcos, J., Cano, M. & Sánchez, J. (2003), *Métodos Estadísticos Aplicados a las Ciencias Sociales*, Editorial Universitaria, Madrid.
*<https://www.editorialuniversitaria.com/libro/metodos-estadisticos-aplicados-a-las-ciencias-sociales12345>
- Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**(2), 461–464.
*<https://projecteuclid.org/euclid.aos/1176344136>

- Sturman, M. (1999), 'Using count data in employment research', *Journal of Applied Psychology* **84**(2), 300–315.
*<https://psycnet.apa.org/doi/10.1037/0021-9010.84.2.300>
- Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.
- Venables, W. & Ripley, B. (2002), *Modern Applied Statistics with S*, Springer.
*<https://link.springer.com/book/10.1007/978-0-387-21706-2>
- Wilson, J. Q. & Kelling, G. L. (1982), 'Broken windows: The police and neighborhood safety', *The Atlantic Monthly* **249**(3), 29–38.
- Winkelmann, R. (2000), *Econometric Analysis of Count Data*, Springer, Berlin.
*<https://link.springer.com/book/10.1007/978-3-662-10825-2>
- Zeileis, A., Kleiber, C. & Jackman, S. (2008), 'Regression models for count data in r', *Journal of Statistical Software* **27**(8), 1–25.
*<https://www.jstatsoft.org/article/view/v027i08>
- Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2), 301–320.