

# Análisis predictivo de la incidencia delictiva en Hidalgo, Mexico

Predictive analysis of the criminal incidence in Hidalgo, Mexico

*Mario Cruz Cruz* 1

*Universidad Autónoma del Estado de Hidalgo, México*  
mariocruzacruz@gmail.com

 <http://orcid.org/0000-0003-1399-9223>

Redalyc: <http://www.redalyc.org/articulo.oa?id=456057665001>

*Juan Bacilio Guerrero Escamilla* 2

*Universidad Autónoma del Estado de Hidalgo, México*  
guerreroescamilla@yahoo.com.mx

 <http://orcid.org/0000-0002-0971-7564>

*Adrián González Romo* 3

*Universidad Autónoma del Estado de Hidalgo, México*  
aromo@colpos.mx

 <http://orcid.org/0000-0003-1954-5185>.

Recepción: 05 Mayo 2018

Aprobación: 17 Junio 2018

## RESUMEN:

El objetivo es la construcción de un modelo probabilístico (con investigación de operaciones) para identificar y analizar la variable que más incide en la comisión del delito en 84 municipios del Estado de Hidalgo, México. El método de investigación fue el análisis estadístico usando una regresión binomial negativa en las variables: corrupción (ineficiencia de autoridades y poca transparencia), pobreza, desigualdad y desempleo. Los resultados indican que la variable que más incide en la comisión del delito es la corrupción. La relevancia del trabajo consiste en proponer un método comprensivo de análisis que permite estudiar todos los municipios del Estado de Hidalgo (México). Las limitaciones de la investigación están en su delimitación al espacio espacial señalado, aun cuando podría ser usado en otros estados de la República mexicana.

**PALABRAS CLAVE:** violencia, corrupción, delito, regresión binomial, diagnóstico, Hidalgo (Mexico) .

## ABSTRACT:

The objective is the construction of a probabilistic model (with operations research) to identify and analyze the variable that most affects the commission of crime in 84 municipalities of the State of Hidalgo, Mexico. The research method was statistical analysis using a negative binomial regression to the variables: corruption (inefficiency of authorities and poor transparency), poverty, inequality and unemployment. The results indicate that the variable that most affects the commission of the crime is corruption. The relevance of the work consists of proposing a comprehensive method of analysis that allows studying all the municipalities

---

## NOTAS DE AUTOR

- 1 Doctor en Ciencias Políticas y Sociales con orientación en Relaciones Internacionales por la UNAM. Profesor-Investigador de la Universidad Autónoma del Estado de Hidalgo, Instituto de Ciencias Económico Administrativas–Comercio Exterior. SNI-C. Líneas de investigación: Geopolítica del Comercio Exterior, Políticas Públicas Multisectoriales y Mercados Ilícitos. Teléfono (771) 71 72000 ext. 4171 Correo electrónico: mariocruzacruz@gmail.com. Código ORCID: <https://orcid.org/0000-0003-1399-9223>
- 2 Doctor en Ciencias Sociales por la Universidad Autónoma del Estado de Hidalgo. Profesor-investigador del Instituto de Ciencias Sociales y Humanidades de la UAEH. Líneas de investigación: Modelamiento Matemático y Estadístico en las Ciencias Sociales y en el Medio Ambiente. SNI- I. Teléfono (771) 71 72000 ext. 5206. Correo electrónico: guerreroescamilla@yahoo.com.mx. Código ORCID: <https://orcid.org/0000-0002-0971-7564>
- 3 Doctor en Ciencias, con especialidad en Estrategias para el Desarrollo Agrícola Regional, Colegio de Postgraduados-Campus Puebla. Profesor-Investigador de la Universidad Autónoma del Estado de Hidalgo, Instituto de Ciencias Económico Administrativas–Comercio Exterior. SNI I. Línea de investigación: Economía Internacional, Migración Internacional, Política Social y Pobreza. Tel: (771) 71 72000 ext. 4171. Correo electrónico: aromo@colpos.mx. Código ORCID <http://orcid.org/0000-0003-1954-5185>.

mariocruzacruz@gmail.com

of the State of Hidalgo (Mexico). The limitations of the research are in its limitation to the indicated space, although it could be used in other states of the Mexican Republic.

**KEYWORDS:** crime, violence, corruption, binomial regression, diagnosis, Hidalgo (Mexico).

## INTRODUCCIÓN

La delincuencia es uno de los principales problemas sociales que actualmente enfrenta la sociedad mexicana, debido a que dicha actividad ha venido incrementándose en los últimos años. Este aumento delictivo tiene un impacto multinivel en la productividad, inversión, trabajo y consumo. Además, este flagelo obstaculiza el desarrollo y el crecimiento del país, ocasionando que cada año sean desviados una enorme cantidad recursos económicos para tratar de prevenir y combatir los efectos negativos que genera este fenómeno social.

De acuerdo con la Encuesta Nacional de Victimización y Prevención sobre Seguridad Pública realizada por INEGI, en el año de 2016 se estimó que 23.3 millones de personas fueron víctimas de la delincuencia, esto representó una tasa de prevalencia delictiva de 28,202 víctimas por cada 100 mil habitantes, donde el robo a mano armada y la extorsión fueron los delitos de mayor frecuencia (24.2% del total de los delitos ocurridos durante el 2016), esto género que el 72.9% de la población se sintiera insegura, ya sea en su trabajo, domicilio o en la calle (INEGI, 2018).

El Estado de Hidalgo[4] no ha estado exento de dicho fenómeno, pues de acuerdo con la Encuesta Nacional de Seguridad Pública Urbana elaborada por INEGI en 2016, el 61.8% de los hidalguenses se consideraron inseguros en su trabajo, domicilio o en la calle, debido a los más de 78 mil delitos [5](27 mil por cada 100 mil habitantes) registrados en la entidad, donde los de mayor frecuencia fueron la extorsión y el robo a mano armada (23.1% del total de delitos ocurridos durante el 2016 en la entidad).

Bajo este contexto, se han generado una serie de conjeturas sobre el comportamiento de dicho fenómeno social, identificando a la corrupción (ineficiencia de autoridades y poca transparencia), pobreza, desigualdad y desempleo como los factores que más inciden en el delito. Estas aseveraciones las han realizado a partir de matrices de correlación de Pearson (Roel, 2015) explicadas en los resultados de las siguientes tablas:

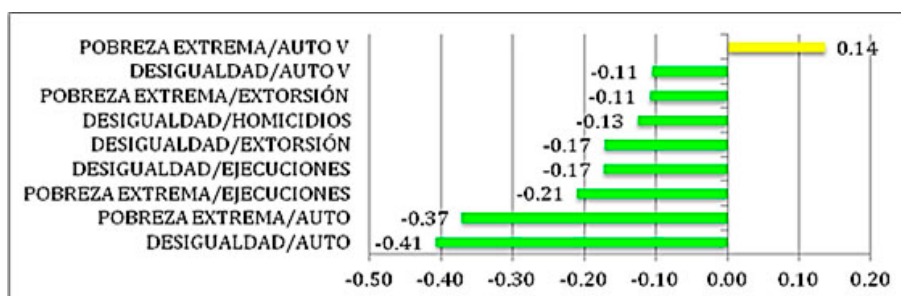
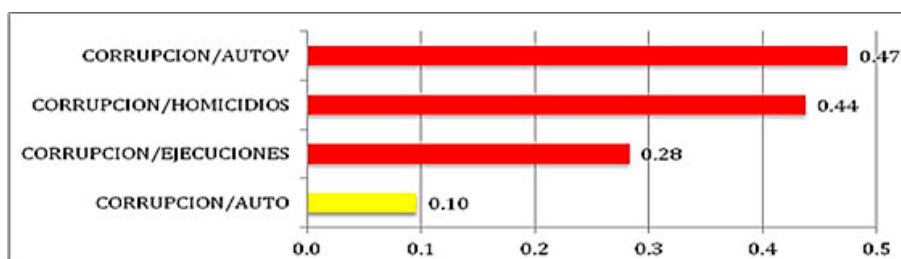
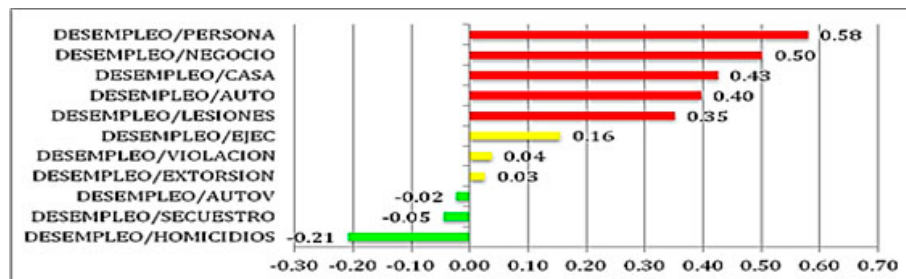


FIGURA 1  
Pobreza y desigualdad con delincuencia



Desempleo con delincuencia



Corrupción con delincuencia

(Roel, 2015)

En la figura 1 se puede identificar el grado de correlación que existe entre la pobreza, la desigualdad, el desempleo y la corrupción con la delincuencia. No obstante, esta información así mostrada no permite pronosticar la dinámica prospectiva de dicho fenómeno.

En consecuencia, para predecir el comportamiento de la delincuencia y generar política pública y modelos de intervención es fundamental que se construya un modelo matemático o estadístico, mediante el cual se pueda tener una representación simplificada del fenómeno. Así, al expresar de forma matemática los elementos y las relaciones que interviene en la incidencia delictiva, se permitiría la evaluación de distintas soluciones factibles para la toma de decisiones.

En los países desarrollados el uso de modelos matemáticos o estadísticos para la toma de decisiones se han extendido. Son utilizados en las empresas, los hospitales, las instituciones financieras, las bibliotecas, en la planeación urbana, en los sistemas de transporte y aún en la criminología (Narro, 1996).

Por ello, es de suma importancia que los países latinoamericanos difundan el uso de esta herramienta, pues en periodos de crisis económicas y sociales se podrían encontrar soluciones a dichos problemas, ya que en otros países los modelos matemáticos y estadísticos han dado soluciones a problemas complejos. En esta perspectiva, se encuentra el modelo probabilístico desarrollado por Jeff Brantingham de la Universidad de Los Ángeles, cuyo objetivo era el identificar los movimientos de los delincuentes en la ciudad de Los Ángeles, tomando como referencia las acciones delictivas, pues a través de la búsqueda de patrones se predecía la próxima acción delictiva, con este modelo se pudo reducir en un 25% la delincuencia de la ciudad de los Ángeles (Guil, 2009).

El reto en este trabajo es generar un modelo que permita identificar patrones de comportamiento del delito en Hidalgo, mediante el uso de herramientas cuantitativas que desborden aseveraciones de sentido común y nos muestren realmente el papel que juegan los actores o nodos de la red delictiva que opera en la entidad y que prediga la incidencia delictiva en el largo plazo.

Finalmente, debe de indicarse que este trabajo se desmarca de apreciaciones políticas del desempeño de la autoridad en materia de combate y contención del delito, los resultados se extraen de estadísticas oficiales, por lo que no existe sesgo metodológico, con la intención de detectar los factores que más inciden en la comisión del delito y la corrupción en los 84 municipios del Estado de Hidalgo. De hecho, el análisis y creación de este modelo fortalece la línea de investigación que los autores siguen en otros trabajos (Cruz, Guerrero, Sánchez, & Sosa, 2018), donde se muestran los costos que erogan las empresas que han sido víctimas del delito y la corrupción, que al conjuntarse estas investigaciones seguramente se enriquecerán las interpretaciones.

## OBJETIVOS DEL ESTUDIO

El presente trabajo de investigación tiene la finalidad de construir un modelo probabilístico, en el cual se pronostique el comportamiento del fenómeno delictivo en el Estado de Hidalgo, pues esta entidad se ha convertido un polo de atracción de la delincuencia organizada, ya que su ubicación y cercanía con la

Ciudad de México-Zona Metropolitana del centro del país, les permite vivir en una zona tranquila y barata pues los bajos costos que tiene la vivienda en dicha entidad son ampliamente inferiores a las demás zonas metropolitanas. En suma, el objetivo central de esta investigación es predecir la dinámica de incidencia delictiva en el Estado de Hidalgo, tomado como referencia la pobreza, el desempleo, la corrupción y desempeño de los ministerios públicos en la entidad durante el año 2015.

## MÉTODOS DE INVESTIGACIÓN

5ta. fase: Interpretación de parámetros. Al cumplir el modelo con supuestos de inferencia para su aplicación, se procede a la interpretación de sus parámetros, pues a través de ella se podrán construir escenarios alternativos del fenómeno en estudio.

Para la construcción de dicho modelo se utiliza la metodología de investigación de operaciones, la cual se conforma por cinco fases (Taha, 2012):

1era. fase: Planteamiento del problema. Consiste en definir los alcances del problema que se investigan, para lo cual se debe cumplir con tres elementos:

- Primer elemento: la delimitación de variables en el fenómeno de estudio.
- Segundo elemento: las fuentes de información para la construcción de la base de datos.
- Tercer elemento: los alcances y límites de la investigación.

2da. fase: Construcción del modelo. Radica en el planteamiento de la ecuación matemática a construir y de la selección adecuada del tipo de modelaje que se requiere para el desarrollo de dicha ecuación.

3er. fase: Solución del modelo. En esta fase se calculan los parámetros de los distintos modelos, con la finalidad de seleccionar el de mejor ajuste a la información obtenida de las distintas fuentes.

4ta. fase: Validación del modelo. Una vez seleccionado el modelo de mejor ajuste, se procede a su validación, es decir, este debe cumplir con todos los supuestos de inferencia necesarios para su ejecución.

5ta. fase: Interpretación de parámetros. Al cumplir el modelo con supuestos de inferencia para su aplicación, se procede a la interpretación de sus parámetros, pues a través de ella se podrán construir escenarios alternativos del fenómeno en estudio.

1era Fase: Planteamiento del problema

Partiendo de tales conjeturas se tiene lo siguiente:

$$Y_d = f(X_p, X_d, X_c, X_{mc}, X_{mf}) \tag{1}$$

(1)

donde:

- $Y_d$  es la tasa de delitos por cada 100 mil habitantes en cada municipio del Estado de Hidalgo.
- $X_p$  es la tasa de pobreza<sup>6</sup> en cada municipio del Estado de Hidalgo.
- $X_d$  es la tasa de desempleo<sup>7</sup> en cada municipio del Estado de Hidalgo.
- $X_c$  es el Tasa de delitos por cada 100 mil habitantes con corrupción<sup>8</sup> en cada municipio del Estado de Hidalgo.
- $X_{mc}$  es el total de ministerios públicos<sup>9</sup> del fuero común en cada municipio del Estado de Hidalgo.
- $X_{mf}$  es el total de ministerios públicos de fuero federal en cada municipio del Estado de Hidalgo.

**TABLA 1**  
Fuentes de información de cada variable a modelar en el año 2015

VARIABLES	FUENTES
Tasa de delitos por cada 100 mil habitantes (	Secretaría de Seguridad Pública
Tasa de pobreza (	CONEVAL
Tasa de desempleo (	INEGI (Anuario Estadístico 2016)
Tasa de delitos con corrupción ( )	Secretaría de Seguridad Pública
Total de ministerios públicos del fuero común (	INEGI (Anuario Estadístico 2016)
Total de ministerios públicos del fuero federal	INEGI (Anuario Estadístico 2016)

Elaboración propia

Dentro de los alcances y límites de la presente se encuentran los siguientes:

- Alcances:

- Mediante la construcción del presente modelo, se podrá predecir la dinámica de la delincuencia de la entidad, y con ello, se podrán evaluar las conjeturas que se han creado, en relación con sus causas.
- Desde el contexto del modelamiento, se podrán visualizar dos elementos: en primer lugar, el papel trascendental de las matemáticas y de la estadística sobre el pronóstico de fenómenos sociales; y en segundo lugar, la eficiencia de los modelos para la toma de decisiones.

- Límites:

- El modelo desarrollado en el presente trabajo es de corte transversal (información sólo de 2015), sería fundamental conocer su tendencia en los últimos años, a partir de las conjeturas que se tienen sobre sus posibles causas, sin embargo, no se cuenta con la suficiente información para desarrollar dicho modelo en esa magnitud.
- La tasa de delitos que se modela en el presente trabajo incluye todo tipo de delitos. Una alternativa sería el modelar alguno de estos delitos, con la finalidad de conocer sus causas y efectos sobre la población de forma diferenciada.

2da. Fase: Construcción del modelo

$$\widehat{Y}_d = \widehat{\beta}_0 + \widehat{\beta}_1 X_p + \widehat{\beta}_2 X_d + \widehat{\beta}_3 X_c + \widehat{\beta}_4 X_{mc} + \widehat{\beta}_5 X_{mf} + e_i \tag{2}$$

donde:

- $\widehat{Y}_d$  es la tasa estimada de delitos por cada 100 mil habitantes en cada municipio del Estado de Hidalgo.
- $\widehat{\beta}_i$ ;  $i = 0, 1, 2, 3, 4, 5$ . Son los parámetros por estimar.
- $e_i$  es el margen de error que no puede predecirse a través de la regresión lineal

Si la tasa de delitos es una variable aleatoria discreta (v.a.d.) (Wacklerly, Mendenhall, & Schaffer, 2010):

$$Y_d \sim \text{v. a. d.}; \text{ tal que } Y_d \in Z^+ = \{0, 1, 2, 3, \dots\} \tag{3}$$

Aunado a lo anterior, pertenece a la familia exponencial, se asume que FY (y; es una función de distribución que depende de un único parámetro, donde su función de densidad se expresa de la siguiente forma (Tusell, 2010).

$$F_Y(y; \theta) = e^{a(\theta)b(y)+c(\theta)+d(y)} \tag{4}$$

Esto debe ocurrir sobre el soporte Y, y que no depende de . En el caso de distribuciones depende de K parámetros , se tiene que (Cavazos, 2007):

$$F_Y(y; \theta) = \exp \left\{ \sum_{i=1}^K a_i(\theta)b_i(y) + c_i(\theta) + d_i(y) \right\} \tag{5}$$

Al ser una función de distribución de la familia exponencial, la media de la distribución depende de las variables independientes X, a través de la siguiente expresión (Mc Cullagh & Nelder, 1983).

$$E(Y_d) = \mu = g^{-1}(X\beta) \tag{6}$$

Bajo estos contextos, debe ser ajustado a través de los Modelos Lineales Generalizados (GLM), los cuales son una extensión de los modelos lineales que permiten utilizar distribuciones no normales de la variable respuesta.

Los GLM tienen tres componentes: el primero, la distribución de Y; el segundo, la función liga; y el tercero, la función offset (Müller, 2004). Partiendo de estos tres componentes, Y<sub>d</sub> se comportaría y se modelaría de la siguiente forma:

Comportamiento de Y<sub>d</sub>:

$$Y_d \sim P(K, \lambda) \rightarrow \text{Poisson} \tag{7}$$

$$Y_d \sim Bn(X; K, \mu) \rightarrow \text{Binomial negativo} \tag{8}$$

Función liga de Y<sub>d</sub> en cada distribución:

- Función liga de Y<sub>d</sub> en cada distribución

**TABLA 2**  
La función liga de las distribuciones poisson y binomial negativa

Distribución	Función liga
Poisson	Logarítmica, Identidad, Raíz cuadrada
Binomial negativa	Logarítmica

Modelos Lineales Generalizados (Cayuela, 2010).

La función offset sirve para ambas distribuciones, su principal función es el poder controlar el efecto de manifestación de la variable respuesta y parte de una hipótesis funcional de interés (Carrascal, 2015).

Los GLM deben cumplir determinados supuestos para su evaluación, entre ellos se encuentran los siguientes (Lindsey, 1997):

- Mayor significancia de los parámetros estimados dentro del modelo.
- Criterio de Información Akaike (AIC). Evalúa el grado de ajuste del modelo a los datos como la complejidad del modelo.

$$AIC = -2LM + 2p$$

donde:

- LM es la máxima verosimilitud del modelo actual
- p es el número de parámetros estimados

cuando más pequeño es el AIC mejor ajuste. Sirve para comparar modelos similares con distintos grados de complejidad.

- Los residuales de la devianza se deben aproximar a sus grados de libertad.
- Los residuales de la devianza deben aproximarse a una distribución normal.

$$e_i \sim N(\mu, \sigma) \tag{9}$$

(9)

No debe existir multicolinealidad entre las variables explicativas del modelo.

- El grado de variabilidad de los datos se obtiene a través de la devianza, la cual se calcula de la siguiente forma:

$$D^2 = \left[ \frac{D_N - D_R}{D_N} \right] * 100 \tag{10}$$

(10) Importar tabla donde:

donde:

- DN es la devianza nula (el modelo tiene una sola constante).
- DR es la devianza de los residuales (el modelo tiene la constante y las variables explicativas).

Considerando todo lo anterior,  $\hat{Y}_d$  es un conteo que tiene dos opciones de modelamiento (poisson ó binomial negativo), por tanto, su expresión lineal sería la siguiente:

$$E(\hat{Y}_d) = g^{-1}(F) \tag{11}$$

donde:

- $E(\hat{Y}_d)$  es el valor esperado a estimar de la tasa de delitos por cada 100 mil habitantes en cada municipio del Estado de Hidalgo.
- $F$  es el predictor lineal de  $\hat{\beta}_0 + \hat{\beta}_1 X_p + \hat{\beta}_2 X_d + \hat{\beta}_3 X_c + \hat{\beta}_4 X_{mc} + \hat{\beta}_5 X_{mf}$ .

Para la construcción de los modelos se va a utilizar el software RStudio (es un lenguaje S desarrollado por AT&T por Rick Becker, John Chambers y Allan Wilks (Sáez, 2012).

3era. Fase: solución del modelo . A través de las corridas en RStudio, ambos modelos tienen las siguientes salidas[7]:

**TABLA 3**  
Corrida en ambos modelos

```

Modelo I. Regresión Poisson
Call: glm(formula = Yd ~ log(Xc) + Xf + 1) + log(Xm + 1) + log(Xz) + Xd + offset(log(Xp)), family = poisson, data = Base2)
Deviance
Residuals:   Min       1Q   Median       3Q      Max  -16.1267  -2.3133   0.1387   2.5434  14.5276
Coefficients: Estimate
Std. Error z value Pr(>|z|) (Intercept) -7.375074  0.172644 -42.718 <2e-16 *** log(Xc)  0.477880  0.012226  39.089 <2e-
16 *** Xf  0.031270  0.004284  7.289 2.90e-13 *** Xm  -0.021051  0.003332  -6.319 2.64e-10 *** log
(Xz)  -0.102984  0.030336  -3.395 0.000687 *** Xd  0.054312  0.004978  10.910 <2e-16 *** Signif. codes: 0 '***'
0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for poisson family taken to be 1) Null deviance: 18593.8 on 83
degrees of freedom Residual deviance: 1654.7 on 78 degrees of freedom AIC: 2135.5 Number of Fisher Scoring iterations: 4

Modelo II. Regresión Binomial Negativo
Call: glm.nb(formula = Yd ~ log(Xc) + log(Xf + 1) + log(Xm + 1) + log(Xz) + Xd + offset(log(Xp)), data = Base2, init.theta =
9.4389244, link = log)
Deviance
Residuals:   Min       1Q   Median       3Q      Max  -2.13098  -0.86064  -0.07492  0.67893  2.61583
Coefficients: Estimate
Std. Error z value Pr(>|z|) (Intercept) -4.931381  0.997338  -4.945 7.63e-07 *** log(Xc)  0.497915  0.049075  10.146 <2e-16
*** log(Xf + 1) -0.131360  0.105047  -1.250 0.21112 log(Xm + 1) -0.085828  0.078883  -1.215 0.22444 log
(Xz)  -0.644386  0.209129  -3.081 0.00206 ** Xd  0.005239  0.016803  0.315 0.75235 Signif. codes: 0 '***' 0.001 '**'
0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for Negative Binomial(9.4389) family taken to be 1) Null deviance: 499.314 on
83 degrees of freedom Residual deviance: 90.067 on 78 degrees of freedom AIC: 833.24 Number of Fisher Scoring iterations: 1
    
```

Elaboración propia

Al comparar ambos modelos (poisson y binomial negativo), se puede observar que el modelo II es el más adecuado, pues tiene AIC más pequeño que el modelo I. Una vez seleccionado el modelo binomial negativo (modelo II), se puede ver que las variables de desempleo, ministerios públicos (tanto federal como del fuero común) no son significativos dentro del modelo. En consecuencia, lo adecuado, es sacarles del modelo:

Modelo III. Sacando la variable desempleo

**TABLA 4.**

Corrida del modelo binomial negativo sin la variables desempleo (Xd) y ministerios públicos (Xm)

```

Modelo III. Sacando la variable desempleo Call: glm.nb(formula = Yd ~ log(Xc) + log(Xf + 1) + log(Xz) + offset(log(Xp)),
data = Base2, init.theta = 9.391180427, link = log)
Deviance
Residuals:   Min       1Q   Median       3Q      Max  -2.13255  -0.88981  -0.07766  0.69121  2.57021
Coefficients: Estimate
Std. Error z value Pr(>|z|) (Intercept) -4.903980  0.992398  -4.938 7.87e-07 *** log(Xc)  0.49716  0.04906  10.133 <2e-16
*** log(Xf + 1) -0.13174  0.10530  -1.251 0.21092 log(Xz) + 1) -0.09676  0.07902  -1.224 0.22079 log
(Xz)  -0.64476  0.20953  -3.077 0.00205 ** Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for
Negative Binomial(9.3912) family taken to be 1) Null deviance: 437.085 on 83 degrees of freedom Residual deviance: 89.318 on
79 degrees of freedom AIC: 831.34 Number of Fisher Scoring iterations: 1

Modelo IV. Sacando la variable de ministerios públicos del fuero común Call: glm.nb(formula = Yd ~ log(Xc) + log(Xf + 1) + log(Xz) +
offset(log(Xp)), data = Base2, init.theta = 9.104439223, link = log)
Deviance
Residuals:   Min       1Q   Median       3Q      Max  -2.14109  -0.87243  -0.03707  0.69262  2.49086
Coefficients: Estimate
Std. Error z value Pr(>|z|) (Intercept) -4.75822  1.00227  -4.747 2.06e-06 *** log(Xc)  0.46414  0.04150  11.183 <2e-16
*** log(Xf + 1) -0.13186  0.09486  -2.023 0.04312 * log(Xz)  -0.65555  0.21240  -3.086 0.00203 ** Signif. codes: 0 '***'
0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for Negative Binomial(9.1044) family taken to be 1) Null deviance:
483.654 on 83 degrees of freedom Residual deviance: 89.135 on 80 degrees of freedom AIC: 830.87 Number of Fisher Scoring
iterations: 1
    
```

Elaboración propia

Al sacar las variables desempleo y ministerios públicos del fuero común, se puede observar que modelo IV es el más adecuado, pues con un P-Valor por debajo de 0.05, el intercepto, la corrupción, los ministerios públicos federales y la tasa de pobreza son significativos para predecir la incidencia delictiva. Una vez seleccionado el modelo más adecuado, se procede a su validación.

4ta. Fase: validación del modelo

En el apartado anterior se llegó a la conclusión de que el modelo IV es el más adecuado para predecir la incidencia delictiva en el Estado de Hidalgo, sin embargo, para realizar dicha acción es fundamental que el modelo cumple con los supuestos mencionados en el apartado 3:

Los residuales de la devianza del modelo IV (tabla 5) se aproximan a sus grados de libertad.

**TABLA 5**

Ajuste de los residuales de la devianza del modelo VI

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Dispersion parameter for Negative Binomial(9.1044) family taken to be 1)  
Null deviance: 483.654 on 83 degrees of freedom  
Residual deviance: 89.135 on 80 degrees of freedom

1-pchisq(89.135,80): p-value = 0.227035

ignif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for Negative Binomial(9.1044) family taken to be 1) Null deviance: 483.654 on 83 degrees of freedom  
Residual deviance: 89.135 on 80 degrees of freedom 1-pchisq(89.135,80): p-value = 0.227035

Elaboración personal

Esto se puede corroborar a través de la siguiente prueba de hipótesis (López, 2010):



$H_0$ (hipótesis nula): los residuales se aproxima a sus grados de libertad  
VS

$H_a$ (hipótesis alterna): los residuales se no se aproxima a sus grados de libertad

Si el P-valor < 0.05, se rechaza la  $H_0$ . En la tabla 5 se puede observar que el P-valor (0.227035) del modelo IV es superior a 0.05, por tanto, no se rechaza la  $H_0$ , es decir, los residuales de la devianza se aproximan a sus grados libertad.

Si el P-Valor < 0.0

Mediante la figura 2, se puede conjeturar que los residuales del modelo IV se distribuyen como una normal. Guil

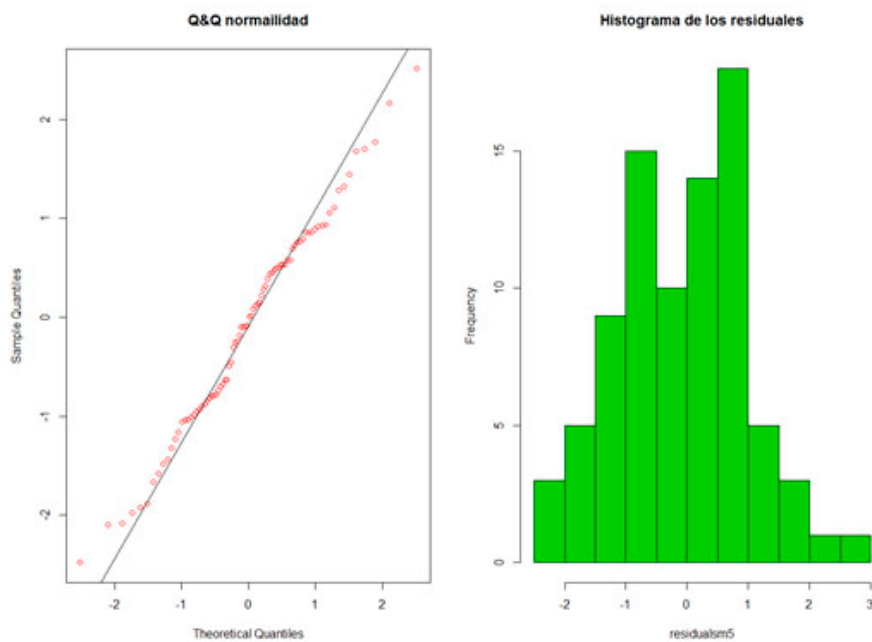


FIGURA 2.  
Histograma y Q & Q de normalidad de los residuales  
Elaboración propia

Esta conjetura se puede corroborar a través de la siguiente prueba hipótesis (Sigüeñas, 2015)

$H_0$ (hipótesis nula): los residuales se aproximan a una normal  
VS

$H_a$ (hipótesis alterna): los residuales se no se aproximan a una normal

Si el P-Valor < 0.05, se rechaza la  $H_0$ . Mediante las siguientes pruebas:

Anderson-Darling normality test: A = 0.41279, p-value = 0.3314

Cramer-von Mises normality test: W = 0.078026, p-value = 0.2168

Lilliefors (Kolmogorov-Smirnov) normality test: D = 0.074116, p-value = 0.3062

Con estas pruebas de hipótesis, se puede ver que los residuales del modelo se distribuyen como una normal, pues el P-Valor se encuentra por arriba del 0.05.

A través de la prueba VIF (Factor de inflación de la varianza: si  $VIF > 5$  va a existir multicolinealidad) se puede observar que no hay multicolinealidad, pues sus factores no inflan la varianza del modelo.

TABLA 6  
Prueba de colinealidad de las variables explicativas del modelo VI

$vif(m5)$	$\log(Xc)$	$\log(Xf + 1)$	$\log(Xz)$
	0.3327320	0.1507628	0.2404734

Elaboración personal

Como se puede ver, el modelo IV cumple con todos los supuestos (los residuales de la devianza se aproximan a sus grados de libertad, los residuales se distribuyen como normal, y las variables explicativas no presentan problemas de multicolinealidad).

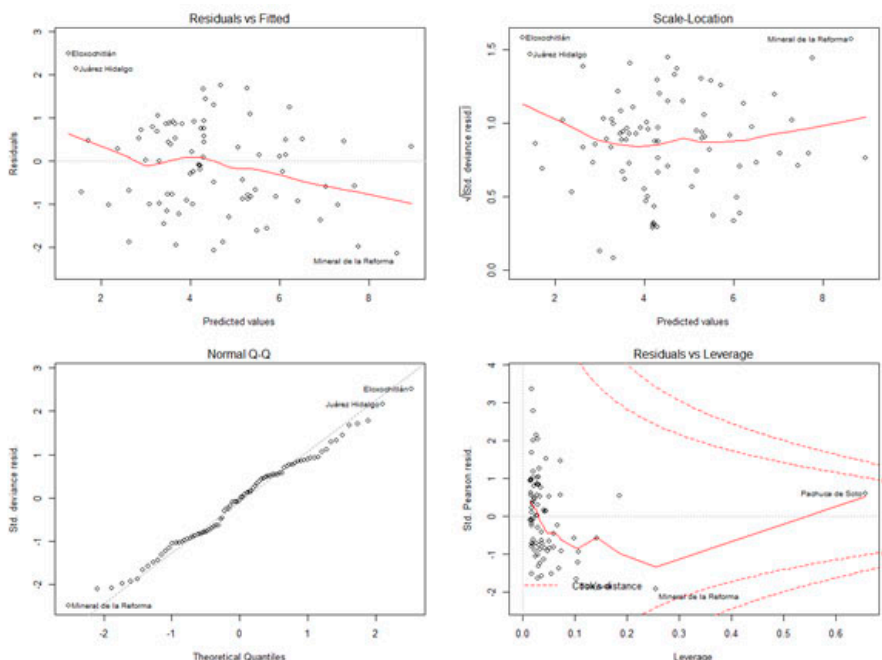


FIGURA 3.  
Observaciones influyentes sobre el modelo IV  
Elaboración propia

Sin embargo, en la figura 3 se puede ver que hay observaciones que podrían “influir” (Eloxochitlán, Mineral de la Reforma y Pachuca de Soto) sobre la dinámica del fenómeno porque se encuentran en los extremos de muy alta y muy baja incidencia delictiva por tratarse de los municipios más pequeño y más grande del Estado, no obstante, en la figura 4, se puede ver que dichas observaciones no producen sesgo en la predicción del comportamiento del fenómeno delictivo.

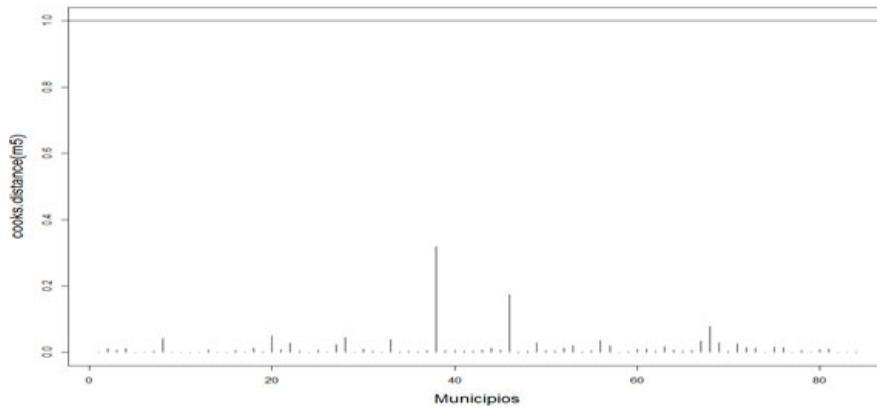
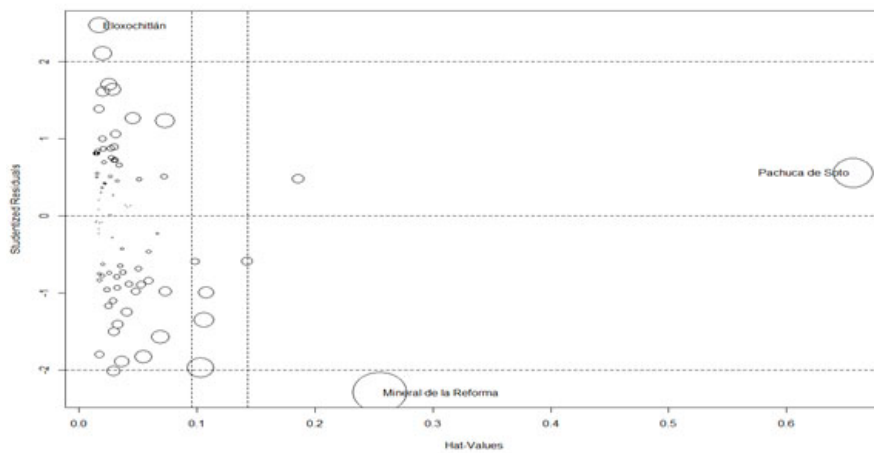


FIGURA 4

Grado de influencia de cada una de las observaciones sobre el comportamiento del fenómeno delictivo.



Elaboración propia

Aunado a lo anterior, el modelo IV presenta una devianza del 81.57%. Desde el contexto de la estadística, conserva el 81.57% de la variabilidad de los datos.

$$D^2 = \left[ \frac{D_N - D_R}{D_N} \right] * 100 = \left[ \frac{483.654 - 89.135}{483.654} \right] * 100 = 81.57$$

Desde el contexto del fenómeno, el modelo IV explica el 81.57% de la dinámica de la delincuencia en el Estado de Hidalgo.

#### ANÁLISIS Y DISCUSIÓN: INTERPRETACIÓN DE PARÁMETROS

Una vez validado el modelo IV, su estructura sería de la siguiente forma:

$$\ln E(\hat{Y}_d) = -4.76 + 0.46\ln(X_c) - 0.19\ln(X_f + 1) - 0.66\ln(X_z) + \text{offset}[\ln(P)] \quad (12)$$

Donde el offset asume que hay una relación lineal entre la población (P) y la tasa de delitos, por tanto (Seoane, Carmona, Tarjuelo, & Planillo, 2014)

$$\ln(E(\tilde{Y}_d)) - \ln(P) = -4.76 + 0.46\ln(X_c) - 0.19 \ln(X_f + 1) - 0.66\ln(X_z) \quad (13)$$

Despejan a :

Despejan a  $\tilde{Y}_d$ :

$$\ln\left(\frac{E(\tilde{Y}_d)}{P}\right) = -4.76 + 0.46\ln(X_c) - 0.19 \ln(X_f + 1) - 0.66\ln(X_z) \quad (14)$$

$$\frac{E(\tilde{Y}_d)}{P} = \exp\{-4.76 + 0.46\ln(X_c) - 0.19 \ln(X_f + 1) - 0.66\ln(X_z)\} \quad (15)$$

$$\frac{E(\tilde{Y}_d)}{P} = 0.009e^{0.46\ln(X_c) - 0.19 \ln(X_f + 1) - 0.66\ln(X_z)} \quad (16)$$

$$\frac{E(\tilde{Y}_d)}{P} = 0.009 \left( \frac{e^{0.46\ln(X_c)}}{e^{0.19 \ln(X_f + 1)} * e^{0.66\ln(X_z)}} \right) \quad (17)$$

Por tanto:

$$E(\tilde{Y}_d) = 0.009P \left( \frac{e^{0.46\ln(X_c)}}{e^{0.19 \ln(X_f + 1) + 0.66\ln(X_z)}} \right) \quad (18)$$

(18)

A través de la ecuación 18 y por medio de la figura 5, se puede observar el grado de ajuste del modelo IV con respecto a los datos reales.

Figura 5. Grado de ajuste del modelo IV con

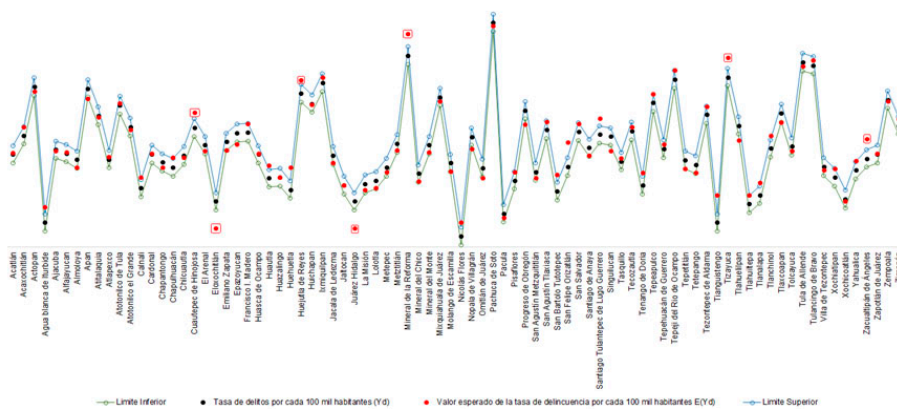


FIGURA 5  
Grado de ajuste del modelo IV con los datos reales  
Elaboración propia

Partiendo de la ecuación 18, la interpretación de sus datos sería la siguiente:

$$E(\hat{Y}_d) = 0.009P \left( \frac{e^{0.46 \ln(X_c)}}{e^{0.19 \ln(X_f+1)+0.66 \ln(X_z)}} \right) \quad (18)$$

Si la tasa de corrupción ( $X_c$ ), los ministerios públicos ( $X_f$ ) y la tasa de pobreza ( $X_z$ ) permanecen constantes, el valor esperado de delitos por cada 100 mil habitantes sería de aproximadamente 900 delitos.

$$E(\hat{Y}_d) = 0.009(100,000) \left( \frac{e^{0.46 \ln(0)}}{e^{0.19 \ln(0+1)+0.66 \ln(0)}} \right) = 900$$

Por cada dos puntos porcentuales que se incremente la tasa de corrupción en la entidad y permaneciendo constante los ministerios públicos ( $X_f$ ) y la tasa de pobreza ( $X_z$ ), el valor esperado de delitos por cada 100 mil habitantes sería de aproximadamente 1,238 delitos.

$$E(\hat{Y}_d) = 0.009(100,000) \left( \frac{e^{0.46 \ln(2)}}{e^{0.19 \ln(0+1)+0.66 \ln(0)}} \right) = 1237.98$$

Esto personifica un incremento del 37.56% sobre el valor esperado.

Por cada ministerio público federal que proporciones servicio en la entidad y permaneciendo constante la tasa de corrupción ( $X_c$ ) y la tasa de pobreza ( $X_z$ ), el valor esperado de delitos por cada 100 mil sería a aproximadamente 789 delitos.

$$E(\hat{Y}_d) = 0.009(100,000) \left( \frac{e^{0.46 \ln(0)}}{e^{0.19 \ln(1+1)+0.66 \ln(0)}} \right) = 788.94 \quad (18)$$

Esto equivaldría a un decremento del 12.33% sobre el valor esperado.

Con esto se puede decir lo siguiente: “con un nivel de confianza del 95%, con un margen de error del 5% y con el cumplimiento de los supuestos de inferencia, el modelo IV predice en un 81.57% de la dinámica de la delincuencia en el Estado de Hidalgo.

Por cada dos puntos porcentuales que se incrementa la tasa de pobreza en la entidad y permaneciendo constante la tasa de corrupción ( $X_c$ ) y los ministerios públicos ( $X_f$ ), el valor esperado de delitos por cada 100 mil habitantes sería de aproximadamente 570 delitos.

sto equivaldría a un decremento del 36.66% del valor esperado.

Con esto se puede decir lo siguiente: “con un nivel de confianza del 95%, con un margen de error del 5% y con el cumplimiento de los supuestos de inferencia, el modelo IV predice en un 81.57% de la dinámica de la delincuencia en el Estado de Hidalgo

## RESULTADOS

Con el desarrollo del presente modelo (regresión binomial negativo) se pudo evaluar tres aspectos importantes sobre el fenómeno de la delincuencia en el Estado de Hidalgo:

Primero: las conjeturas que se planearon sobre las posibles causas que incentivan al fenómeno delictivo no todas son significativas para predecir la dinámica de la delincuencia en el Estado de Hidalgo, pues mediante el presente modelo se puede afirmar que la incidencia delictiva de la entidad está más asociada a la corrupción, los ministerios públicos y la pobreza en ese orden de importancia, y que dicha relación lineal es de un 81.57%.

$$Y_d = f(X_c, X_f, X_z) \rightarrow D^2 = 81.57$$

Ixmiquilpan

En consecuencia, la elaboración de política pública y modelos de intervención para combatir y contener la incidencia delictiva deberá de estar enfocada a prevenir y sancionar los actos de corrupción en los que incurran los funcionarios públicos, particularmente en los ministerios públicos. En términos geográficos, los municipios donde más acciones anticorrupción deben de realizarse son en Pachuca, Mineral de la Reforma, Tizayuca, Tula, Tulancingo, Huejutla e Ixmiquilpan. Si no se combate este delito no habrá crecimiento ni desarrollo en estos municipios y la escalada de violencia seguirá su curso natural hacia otro tipo de delitos como el secuestro.

De acuerdo al presente modelo, dicha relación lineal dio origen a las siguientes conjeturas

- “A mayor tasa de corrupción, mayor incidencia delictiva”
- “A mayor número de ministerios públicos federales, menor incidencia delictiva”.
- “A mayor tasa de pobreza, menor incidencia delictiva”.

Bajos dichas conjeturas, las dos primeras presenta una lógica sobre el comportamiento del fenómeno delictivo, sin embargo, la última tiene una relación inversa, por lo que esta conducta puede interpretarse que la incidencia delictiva del Estado de Hidalgo no es resultado de la pobreza, pero si es producida por los actores que conocen cómo corromper el sistema de impartición de justicia, es decir, está asociado a la corrupción dentro de los ministerios públicos del fuero común. Por ello se concluye que los agentes que más saben corrompen a las instituciones son los propios funcionarios públicos, por lo que su grado de captura institucional es mucho mayor.

Segundo: a través del presente modelo se puede corroborar la importancia los métodos cuantitativos para predecir el comportamiento de los fenómenos sociales, y en particular, la incidencia delictiva, pues mediante estas herramientas se puede aproximar a la tendencia de dicho fenómeno con la finalidad identificar el patrón de comportamiento operativo de la delincuencia en el Estado de Hidalgo.

Tercero: Los municipios más afectados por el fenómeno delictivo en el largo plazo serán Pachuca de Soto, Mineral de la Reforma, Tizayuca, Tepeji del Río, Tulancingo de Bravo y Tula de Allende, por lo que es necesario tomar acciones inmediatas para evitar una expansión del delito del fuero común hacia otro tipo de delitos más rentables. De hecho, en el municipio de Tula la autoridad ha reportado un incremento del trasiego ilegal de hidrocarburo proveniente de los municipios aledaños del Valle del Mezquital.

Si no se implementan acciones para combatir la corrupción, en el corto plazo estos municipios incrementarán el grado de violencia hacia delitos del fuero federal perpetrados por el crimen organizado y, sin duda, se traducirá en la muerte de civiles y funcionarios públicos como ha ocurrido en otras entidades. Gran parte de la captura institucional, vía corrupción e inseguridad, involucra a diversos actores acostumbrados a un ingreso, por lo que al aplicar medidas de contención los delincuentes buscarán otras fuentes de financiamiento. Esto debe de saberlo la autoridad comprometida para realmente traducirse en una política pública en materia de seguridad efectiva, no esperemos a que la violencia toque la puerta; de políticos, empresarios y la nuestra propia para tomar las medidas necesarias.

Contribución de los autores: Conceptualización (Mario Cruz Cruz), Curación de dato (Juan Bacilio Guerrero Escamilla), Análisis formal (Adrián González Romo), Investigación (Adrián González Romo), Metodología (Mario Cruz Cruz), Validación (Juan Bacilio Guerrero Escamilla); Visualización (Mario Cruz Cruz); Redacción del borrador original (Juan Bacilio Guerrero Escamilla, Mario Cruz Cruz); Redacción de revisión y edición (Mario Cruz Cruz).

## REFERENCIAS

- Canavos, C. (2007). Probabilidad y Estadística. Aplicaciones y Métodos. Estados Unidos: Mc Graw Hill.
- Carrascal, L. (2015). Modelos Generalizados Lineales con SPSS. España: Depto. Biogeografía Y cambio Climático.
- Cayuela, L. (2010). Modelos Lineales Generalizados (GLM). España: Universidad de Granadas.

- Cerimedo, F. (2002). Cuadernillos de Economía No. 65. Pobreza: definición, determinantes y programas para su erradicación. Argentina: Ministerio de Economía de la Provincia de Buenos Aires República de Argentina.
- Colín, G. (1983). Derecho Mexicano de Procedimientos Penales. México: Universidad Nacional Autónoma de México.
- Cruz, M., Guerrero, J., Sánchez, y Sosa, J. (2018) Comercio, inseguridad y captura institucional. Las distorsiones en las mipymes hidalguenses. México: UAEH-Plaza y Valdés.
- Guil, J. (2009). Los policías deberían trabajar con matemáticos para resolver delitos. España: ABC.
- Hernández, R. (2009). Introducción a la economía y la hacienda pública. España: Universidad de Valencia.
- INEGI. (2016). Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE) 2016. México: Instituto Nacional de Estadística y Geografía.
- Lindsey, J. (1997). Applying Generalized Linear Models. USA. Springer.
- López, E. (2010). Análisis de datos con el Modelo Lineal Generalizado. Una Aplicación con R.
- Machado, M. (2006). Construyendo Ciudadanía Forjamos un País sin Corrupción. Perú: Fórum Solidaridad Perú.
- Machicado, J. (2010). Concepto de Delito. México: Apuntes Jurídicos.
- Mc Cullagh, P. & Nelder, J. (1983). Monographs on Statistics and Applied Probability 37. Generalized Linear Models. USA: Chapman and Hall.
- Montgomery, D. & Runger, G. (2007). Probabilidad y Estadística Aplicada a la Ingeniería: Estados Unidos. Limusa Wiley.
- Müller, M. (2004). Generalized Linear Models. Germany: Fraunhofer Institute for Industrial Mathematics (ITWM).
- Narro, A. (1996). Aplicación de algunos modelos matemáticos a la toma de decisiones. México: Política y Cultura.
- Roel, S. (2015). ¿Cuáles son las causas de la violencia en México? México: Forbes México.
- Taha, H. (2012). Investigación de Operaciones. Estado Unidos: Universidad de Arkansas. Pearson.
- Sáez, A. (2012). Métodos Estadísticos con R y R Commander. España: Departamento de Estadística e Investigación de Operaciones de la Universidad de Jaén.
- Sigüeñas, M. (2015). Pruebas de Normalidad. Perú: Universidad Nacional Agraria La Molina.
- Seoane, J., Carmona, C., Tarjuelo, R. & Planillo, A. (2014). Análisis bioestadístico con modelos de regresión en R. España: Universidad Autónoma de Madrid.
- Tusell, F. (2007). Estadística Matemática. España: Universidad del País Vasco.
- Wackerly, D., Mendenhall, W. & Schaffer, R. (2010). Estadística Matemática con Aplicaciones. Estado Unidos: Cengage Learning.

## NOTAS

- [6] La pobreza puede entenderse como la falta de capacidad para alcanzar y mantener un nivel de vida aceptable. Por tanto, un “pobre” será aquel que sufre alguna “privación” que le impida el logro de este objetivo (Cerimedo, 2002).
- [4] El Estado de Hidalgo se ubica en la región este del país y colinda con las entidades de San Luis Potosí, Ciudad de México, Veracruz, Puebla, Tlaxcala, Estado de México y Querétaro, tiene una población aproximada de 3 millones de habitantes, representa el 1.06% del territorio nacional y se divide en 84 municipios, cuya capital es Pachuca de Soto, su Producto Interno Bruto es de aproximadamente 276 mil millones pesos mexicanos y aporta el 1.7 del PIB nacional.
- [5] n delito es un comportamiento antisocial, en el cual implica una violación de las normas vigentes, lo que hace que se merezca una pena (Machicado, 2010).
- [7] Debido a que la información no tiene simetría, a todas las variables explicativas se les aplicó el logaritmo natural, aunado a esto, a las variables  $X_f$  y  $X_m$  se les sumó la unidad. A través de esto, lograr hacer simétrica la información.

## INFORMACIÓN ADICIONAL

*Clasificación JEL:* J48