

## APLICACIÓN DE REGRESIÓN CUANTIL EN LA DETERMINACIÓN Y RANKING DE TRAMOS CRÍTICOS A ACCIDENTES

Thomas Edison Guerrero Barbosa, Universidad Francisco de Paula Santander Ocaña

[teguerrerob@ufpso.edu.co](mailto:teguerrerob@ufpso.edu.co)

Shirley Yaritza Santiago Palacio, Universidad Francisco de Paula Santander Ocaña

[sysantiagoop@ufpso.edu.co](mailto:sysantiagoop@ufpso.edu.co)

### RESUMEN

La identificación de sitios peligrosos a accidentalidad permite a las entidades encargadas de la seguridad vial adecuadas inversiones sobre los tramos viales verdaderamente críticos. Por ello, esta investigación se enfoca en determinar los tramos propensos a accidentes y elaborar un ranking de peligrosidad efectuado a los tramos críticos encontrados en Ocaña (Colombia) utilizando regresión cuantil. El modelo estimado establece relaciones de causalidad entre la frecuencia de accidentes y características asociadas al entorno y operacionales. Adicionalmente se determinó un total de siete tramos críticos a los cuales se les estableció un ranking de peligrosidad.

*Palabras claves: Regresión cuantil, tramo crítico, ranking de peligrosidad.*

### ABSTRACT

The identification of dangerous areas with high accident rates allows for governmental agencies responsible of road safety to properly allocate investment in critically accident prone road sections. Thus, this research aims to determine the segments prone to accidents and development a hazard ranking of the accident prone road sections located in Ocaña (Colombia) using quantile regression. The estimated model establishes causal relationships between the frequency of accidents and environmental and operational characteristics. The results indicate a total of seven accident prone road sections, for which a hazard ranking was established.

*Key words: Quantile regression, accident prone road section, hazard ranking.*

## 1. INTRODUCCIÓN

La Organización Mundial de la Salud (OMS), confirma la creciente epidemia en la que se ha transformado la accidentalidad vial en la mayoría de regiones del mundo y ve con mucha preocupación como este fenómeno cada año cause en el planeta la muerte de 1,2 millones de personas y entre 20-50 millones más en traumatismos no mortales (WHO, 2013). Colombia no es ajena a esta problemática. Según cifras reportadas por el Instituto Nacional de Medicina Legal y Ciencias Forenses (INMLCF), en el país cada 2,5 minutos ocurre un accidente de tránsito, cada 10 minutos se presenta un herido y cada 69 minutos un muerto, lo cual significa que diariamente mueren 20 personas y 144 resultan heridos por causa de la inseguridad vial (INMLCF, 2012). Estas cifras demuestran la problemática de salud pública que enfrenta Colombia, representado a través de altos impactos económicos. De hecho, investigaciones reportan que los traumas derivados de accidentes de tránsito tienen un costo aproximado de USD \$11.370 millones anuales (Bhalla, 2013).

En vista de estas alarmantes cifras, en los últimos años los gobiernos en Colombia han declarado la seguridad vial como una política de Estado, lo que ha facilitado la implementación de una serie de medidas de índole institucional, presupuestal, normativo e investigativo, con el fin de disminuir los efectos de la accidentalidad vial sobre la población. Por otra parte, el Ministerio de Transporte tomó medidas para fortalecer la política pública de seguridad vial a través de la creación del Observatorio Nacional de Seguridad Vial, la Agencia Nacional de Seguridad Vial y una inversión de USD \$10 millones ejecutables a cuatro años a través del CONPES 3764 (Consejo Nacional de Política Económica y Social, 2013).

A pesar de las inversiones realizadas por el gobierno a través de presupuestos cuya destinación exclusiva es para seguridad vial, siguen presentándose altas cifras de accidentalidad. En esta medida, es importante tener en cuenta que en un tramo de carretera se pueden presentar falsos positivos (p.e., cuando se observan frecuencias de accidentes elevadas en un sitio relativamente seguro) y falsos negativos (p.e., cuando es posible observar un sitio crítico o peligroso que no revela frecuencias de accidentes elevadas) los cuales conducen a la pérdida de oportunidades para las inversiones de seguridad vial eficaces (Cheng y Washington, 2005). Es por esto que se hace necesario que estas inversiones se realicen en los tramos viales verdaderamente críticos en términos de accidentalidad. Por ello, esta investigación tiene por objetivo determinar los tramos más propensos a accidentes que produzcan la menor proporción de falsos negativos y falsos positivos, para posteriormente determinar un ranking de peligrosidad de los tramos críticos encontrados en el perímetro urbano de la ciudad de Ocaña (Colombia) utilizando Regresión Cuantil (RC).

La estructura del documento presenta cinco secciones. Los antecedentes referentes a RC se describen en la sección 2. La Sección 3 describe la aproximación metodológica utilizada, las características de los datos y variables contempladas. En la Sección 4, se presentan los análisis e interpretación de los resultados y en la Sección 5, se muestran las conclusiones de la investigación.

## 2. ANTECEDENTES

El modelo RC es utilizado en distintas áreas del conocimiento tales como ecología, economía, computación, epidemiología, estadística, ambiental, entre otras particularmente cuando los datos suelen tener distribuciones asimétricas y presentan una heterogeneidad significativa (Qin *et al*, 2010). RC es una regresión que intenta reproducir la mayor cantidad de variaciones sobre una población, condicionada al conjunto de covariables contempladas en el modelo, hecho presente en los eventos de conteos de accidentes. La RC controla de forma adecuada el efecto de la presencia excesiva de valores atípicos, garantizando mayor ajuste y robustez, además de la posibilidad de estimar cualquier cuantil, pudiendo valorar así lo que ocurre con valores extremos de la población.

El concepto de RC fue introducido a la literatura mundial inicialmente por Koenker y Bassett (1978), donde se plantean soluciones a una simple minimización de los pesos de la suma de los residuos absolutos. El enfoque plantea un problema de optimización a partir de la solución por programación lineal (Koenker y Hallock, 2001; Koenker, 2005).

Al contrario de otros métodos de regresión como, por ejemplo, mínimos cuadrados ordinarios, modelos lineales generalizados, entre otros, RC tiene como objetivo estimar tanto los medios condicionales u otro cuantiles de la variable de respuesta, pudiendo valorar así lo que ocurre con valores extremos de la población, ya que RC está estrechamente relacionado con la técnica estadística error mínimo absoluto (Martínez-Silva, 2011; Qin *et al*, 2010).

RC no es ampliamente utilizado en la determinación de tramos críticos (o sus diferentes versiones en inglés: *hotspots*, *blackspots*, *sites with promise*, *high risk locations*, *accident prone locations*). Qin *et al* (2010) describe este enfoque como favorable respecto a otros tradicionales de regresión, pues no implica ningún supuesto de distribución sobre el error y es menos sensible a la violación de los supuestos de distribución. Además, RC es capaz de proporcionar estimaciones a diferentes niveles de cuantiles, teniendo la capacidad de capturar la heterogeneidad en los datos y presentar una descripción más acertada de las tendencias.

Recientemente, los resultados reportados por Washington *et al* (2014) identifican los efectos de las covariables sobre diversos cuantiles y así identificar los posibles sitios críticos de segmentos de carreteras, intercepciones, cruces peatonales, cruces, rampas, entre otros, en lugar de la media poblacional como la mayoría de los métodos lo ejecutan en la práctica. Qin y Reyes (2011) exponen el método RC como un enfoque alternativo para abordar las dificultades que existen respecto a la heterogeneidad de los datos en la determinación de sitios críticos, dado que combinar datos de diferentes localizaciones y en diversos periodos de tiempo hace las estimaciones de los parámetros inestables y menos eficientes. Los hallazgos proporcionan información referente a los efectos de variables como el volumen vehicular, características geométricas y operacionales de la intersección y el control del tráfico sobre la ocurrencia de accidente, evidenciando como las estimaciones de cuantiles condicionales son más informativos que los medios condicionales, hecho característico de los cuantiles condicionales pues poseen la propiedad de ser equivariantes frente a transformaciones no decrecientes sobre la variable dependiente mientras que la media condicional no comparte tal característica (Koenker y Bassett, 1978).

En otro estudio realizado por Wu *et al* (2014) muestran la importancia de los modelos calibrados con RC con enfoque semiparamétrico. Este tipo de aproximación permite a los investigadores relajar las restricciones de la función de distribución de la frecuencia de accidentes, lo que resulta en la estimación más robusta, con mayor detalle sobre los efectos marginales a través de la distribución condicional de la variable de respuesta y proporciona predicciones más sólidas y precisas sobre el recuento de accidentes.

Existe una gran cantidad de evidencia sobre el uso de otras metodologías tal como el enfoque bayesiano (Geurts y Wets, 2003; Geurts *et al*, 2006; Miaoua y Song, 2005; Miranda-Moreno *et al*, 2007; Wood *et al*, 2015). Este método fue inicialmente introducido por Hauer (1997) a través de estudios observacionales “antes-después” los cuales pretenden evaluar el efecto de implementar tratamientos (medidas de seguridad vial o de tránsito) aplicados a sitios y/o tramos de carreteras. En los últimos años, se ha propuesto la modelación estadística compleja con técnicas *full bayesian* y *bayesian hierarchical* buscando manejar la heterogeneidad de los datos no observados y la presencia excesiva de ceros en los datos de frecuencia de accidentes arrojando excelentes resultados (Aguero-Valverde, 2013; Deublein *et al*, 2013; El-Basyouny y Sayed, 2013).

### 3. MATERIALES Y MÉTODOS

#### 3.1. Aproximación metodológica

La definición de cuantil afirma que dado un  $p \in (0,1)$  perteneciente a una variable aleatoria  $X$  con una función de distribución acumulativa de la forma  $F(x) = P(X \leq x)$ , el  $p$ -ésimo cuantil es definido por la Ecuación 1.

$$Q(p) = F^{-1}(p) = \inf \{x: F(x) \geq p\} = \int_{-\infty}^{Q(p)} f(x) dx \quad (1)$$

El cuantil de orden 0,50 es el más conocido y correspondiente a la mediana de la distribución, el cual representa el valor de la variable de posición central en un conjunto de datos ordenados (Koenker y Bassett, 1978).

Si se tiene una muestra de observaciones aleatorias e independientes de una variable  $\{x_1, x_2, \dots, x_i, \dots, x_n\}$ , es posible encontrar una estimación de la función de distribución que relaciona el número de las observaciones inferiores o iguales al valor de interés y el número total de las observaciones a través de la distribución empírica de la muestra. El cuantil  $Q(p)$  puede ser resuelto minimizando la media ponderada de la muestra cuyos valores son mayores o igual  $Q(p)$  y menores a  $Q(p)$ . Este problema de minimización sigue el modelo presentado en la Ecuación 2 (Koenker, 2005):

$$\min_{\beta \in R^k} \left[ \sum_{i \in \{i: x_i \geq Q(p)\}} p |x_i - Q(p)| + \sum_{i \in \{i: x_i < Q(p)\}} (1-p) |x_i - Q(p)| \right] \quad (2)$$

El término  $Q(p)$  corresponde al  $p$ -ésimo cuantil y se puede expresar como una función lineal de los parámetros de interés utilizando la Ecuación (3).

$$Q(p) = \beta X_i + \varepsilon_i \quad (3)$$

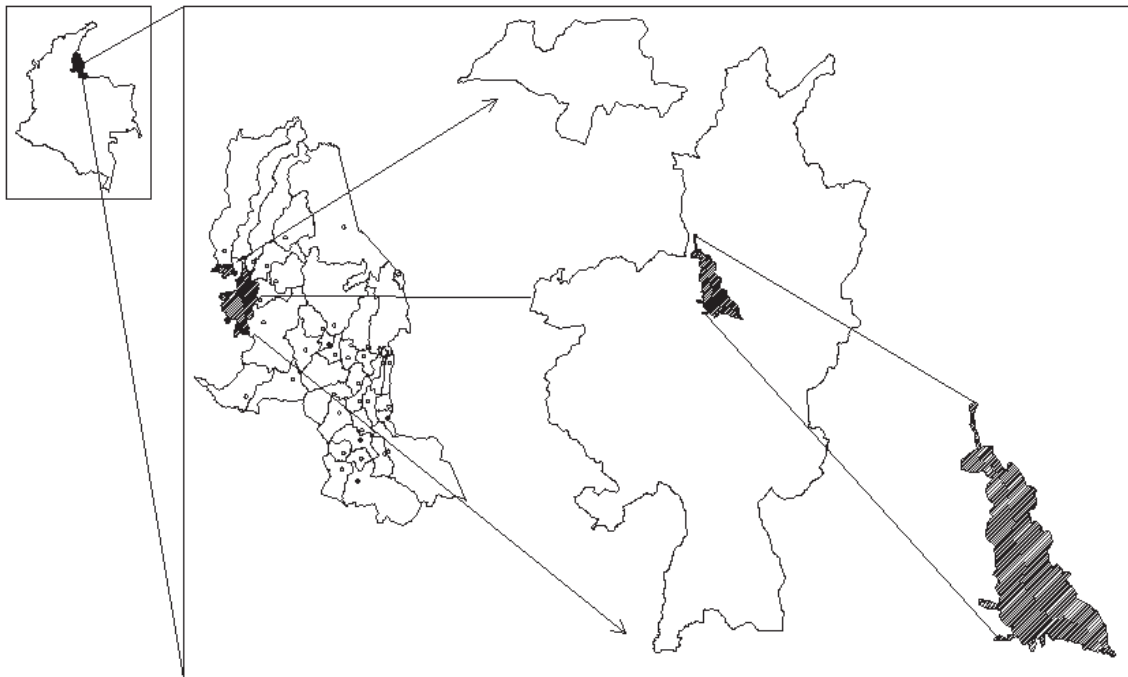
Donde  $p$  toma valores entre 0 y 1 representando el percentiles de interés;  $Q(p)$  es la variable dependiente, correspondiente al número esperado de accidentes para el percentil  $p$ ;  $\beta$  es un vector de dimensión  $k$  de parámetros desconocidos de las covariables  $X_i$  que representa las características específicas de los tramos estudiados (longitud, volumen vehicular, velocidad, entre otras); el subíndice  $i$  corresponde a la identificación de cada uno de los tramos estudiados; el término  $\varepsilon_i$  asume todos aquellos aspecto no tenidos en cuenta en la modelación y cualquier error de medición cometido. Al minimizar la suma de residuos absolutos ponderados, los parámetros  $\beta$  de regresión se estiman resolviendo el problema de optimización dado por la Ecuación (4). Esta expresión puede resolverse como problema de programación lineal a partir de diversos métodos de optimización (Cheng C., 2005).

$$\hat{\beta}(p) = \arg \min_{\beta \in R^k} \left[ \sum_{i \in \{i: x_i \geq \beta X_i\}} p |x_i - \beta X_i| + \sum_{i \in \{i: x_i < \beta X_i\}} (1-p) |x_i - \beta X_i| \right] \quad (4)$$

### 3.2. Datos y variables usadas

Se conformó una base de datos con los registros de accidentes corresponden a un total de 1,241 eventos ocurridos en las vías del perímetro urbano de la ciudad de Ocaña en Colombia (Ver Figura 1) entre el 1 de enero del 2007 y el 31 de enero del 2014. Adicionalmente, esta base de datos posee un inventario vial que contiene características geométricas y operativas para cada uno de los tramos contemplados. Para el caso particular de este estudio, se analizaron 15 tramos de carreteras.

A diferencia de otros estudios que utilizan registros de accidentes entregados por un ente gubernamental de administración de carreteras (Wu *et al.*, 2014; Ayati y Abbasi, 2011), en este estudio, los reportes de accidentes se obtienen de la Policía Nacional de Colombia y otras entidades como Cuerpo de Bomberos Voluntarios y/o Defensa Civil Seccional Ocaña, lo cual dada la naturaleza de recolección de esta información puede implicar algunas desventajas (Also y Langley, 2001; Brenac y Clabaux, 2005). La solicitud de registros de accidentes debe realizarse a través de estas entidades, dado que al momento de ocurrido un accidente la Policía Nacional de Colombia se encarga de levantar los respectivos croquis para determinar responsabilidades jurídicas y/o penales de ocurrencia de estos eventos, mientras que entidades como Cuerpo de Bomberos Voluntarios y/o Defensa Civil Seccional de Ocaña además de registrar el accidente se encargan de brindar socorro y atención médica a las víctimas las cuales son conducidas a la red de clínicas y hospitales de la ciudad.



**Figura 1. Localización perímetro urbano de Ocaña (Colombia)**

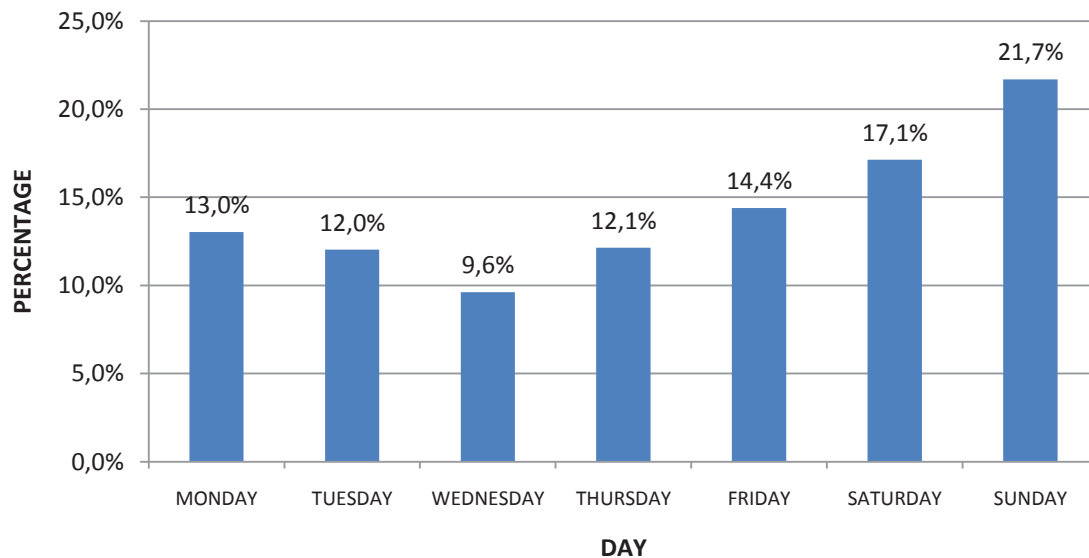
La variable dependiente corresponde al número total de accidentes ocurridos en el tramo  $i$ , mientras que las variables explicativas se dividieron en tres grupos: el primer grupo contiene aquellos factores asociados a la geometría y entorno vial, el segundo grupo corresponde a los volúmenes vehiculares y por último el tercer grupo incluye a las velocidades. Existe evidencia que soporta el uso RC con variables asociadas a la geometría, entorno vial y volúmenes vehiculares (Qin *et al*, 2010; Qin y Reyes, 2011; Liu *et al*, 2013; Washington *et al*, 2014; Wu *et al*, 2014), para el caso de velocidades remitirse a Hewson (2008). El resumen estadístico de las variables y su descripción se muestra en la Tabla 1.

**Tabla 1. Estadística descriptiva de las variables explicativas del modelo**

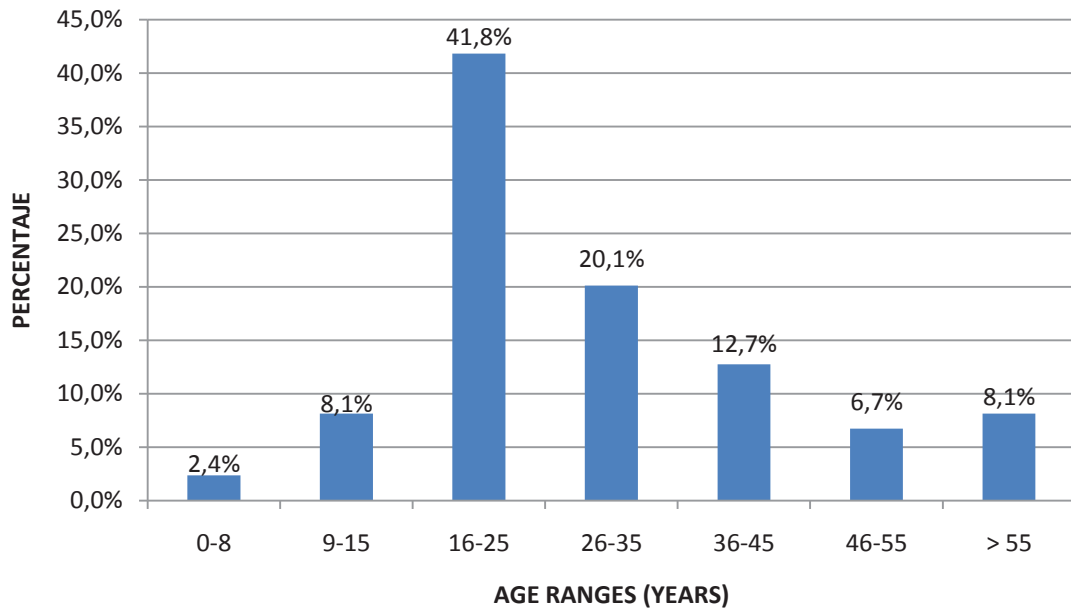
Variable	Descripción	Min	Max	Media	DS
<i>Geometría y entorno vial</i>					
LV	Longitud del tramo vial (m)	355,85	4.084,26	1.249,02	968,03
AC	Ancho de calzada (m)	5,5	7,5	6,23	0,75
SC	Variable muda que vale 1 si el tramo es de un solo sentido de circulación, 0 en otro caso	0,0	1,0	0,27	-
NC	Número de carriles	2,0	4,0	2,13	0,52
NI	Número de intersecciones	6,0	35,0	17,13	8,04
NP	Número de pasos peatonales	0,0	2,0	0,60	0,83
PR	Variable muda que vale 1 si el pavimento es rígido, 0 en otro caso	0,0	1,0	0,20	-

PF	Variable muda que vale 1 si el pavimento es flexible, 0 en otro caso	0,0	1,0	0,53	-
FV	Variable muda que toma el valor de 1 si la vía es principal, 0 en otro caso	0,0	1,0	0,067	-
<i>Volumen vehicular</i>					
TPD	Tránsito promedio diario	5.970	55.868	18.130,33	12.000
<i>Velocidades</i>					
P50	Percentil 50 (Aproximación a la velocidad media en km/h)	24,82	56,60	30,97	7,81
<i>Accidentes</i>					
NA	Cantidad total de accidentes ocurridos en el tramo	4	238	78,40	79,08

Entre los años 2007 y 2011 hubo un aumento del 78% en el número de accidentes en la ciudad y presentando una variación menor del 8% entre los años 2012 y 2013. El periodo donde se presenta la mayor cantidad de accidentes durante el día es entre 18:00 y las 21:00 horas. Los días de mayor registro de accidentes son los viernes, sábados y domingos, los cuales posiblemente puede asociarse al aumento en el consumo de bebidas alcohólicas, mientras que la menor cantidad de accidentes ocurren el día miércoles (Ver Figura 2). La población más afectada por la causa de accidentes es la comprendida entre las edades de 16 a 25 años (Ver Figura 3).



**Figura 2. Porcentajes de accidentes por día de la semana**



**Figura 3. Porcentajes de accidentes según rango de edades**

Todas las variables asociadas a la geometría y el entorno vial, los volúmenes vehiculares y las velocidades fueron medidas en campo a partir de visitas realizadas a cada tramo del estudio. La longitud de los tramos (LV) se consideró teniendo en cuenta la homogeneidad geométrica a lo largo del mismo, encontrándose valores mínimos y máximo para esta característica entre 355,85 m y 4086,26 m, respectivamente. Existe evidencia en la literatura en la cual la longitud de los tramos varía entre 25 m y 3.000 m (Washington *et al*, 2014), entre 80 m y 11.200 m (Lao *et al*, 2014), concluyendo que la longitud del tramo para este estudio se encuentra en los rangos de otras investigaciones ya reportadas. Basados en la información del Plan Básico de Ordenamiento Territorial (Alcaldía Municipal de Ocaña, 2012), se adoptó la variable funcionalidad vial (FV), la cual está comprendida por las redes viales principales y secundarias. Los volúmenes de tránsito (TPD) se obtuvieron de conteos vehiculares realizados a vehículos livianos, pesados y motocicletas; TPD varían ampliamente para el conjunto de datos analizados, donde es posible apreciar el mayor grado de dispersión de los datos con respecto al valor promedio ( $DS = 12.000$ ). Respecto a la variable P50, se estimó una media para los tramos estudiados igual a 30,97 kph. Este estudio no tuvo acceso a registros de variables asociadas al conductor, el clima, características de los vehículos involucrados y gravedad de los accidentes, lo cual constituye una limitación del mismo.

## 4. RESULTADOS

### 4.1. Estimación y Análisis del modelo

Luego de evaluar una variedad de modelos realizando combinaciones entre las variables para diferentes cuantiles en el software STATA, se presenta los coeficientes estimados en la Tabla 2 correspondiente al percentil 95. Adicionalmente, se muestra el Error Estándar (EE), el Intervalo



de Confianza (IC) para el 95% y el valor  $t$  como medida de significancia estadística para las variables que componen el modelo.

**Tabla 2. Modelo de RC para percentil 95**

Parámetro	Coficiente	EE	$t$	$p >  t $	95% IC	
Intercepto	-24,133	15,217	-1,586	0,1514	-59,224	10,958
LV	0,051	0,002	33,720	0,0000	0,047	0,054
AC	30,425	1,524	19,970	0,0000	26,912	33,939
NC	-86,148	3,53	-24,403	0,0000	-94,289	-78,007
NI	-0,405	0,119	-3,412	0,0092	-0,679	-0,131
TPD	0,007	0,0001	48,527	0,0000	0,007	0,008
P50	-2,104	0,133	-15,843	0,0000	-2,41	-1,798

Para el modelo presentado, todas las variables resultaron significativas (a excepción del Intercepto), dado que valores de  $t$  (en valor absoluto) mayores o iguales a 1,96 indican que el coeficiente es significativamente diferente de cero para un nivel de confianza del 95%.

Dentro del grupo de las variables asociadas a la geometría y entorno vial, LV presentó la mayor significancia estadística con un valor  $t$  igual a 33,72. Como se esperaba, el signo positivo que acompaña al parámetro LV indica un aumento en el número de accidentes a medida que esta variable aumenta, lo cual es acorde con los reportes de Wu *et al.*, (2014), Rangel *et al.* (2013) y Wang *et al.* (2011). Comportamiento similar evidencia la variable AC, es decir, entre más ancha sea la calzada existe mayor interacción entre los vehículos por lo que la probabilidad de colisiones aumenta, resultado congruente con Chang (2005) y Noland y Oh (2004).

La variable asociada a los volúmenes vehiculares TPD presentó la mayor significancia dentro de las variables que componen el modelo, ratificando la hipótesis que el aumento en las tasas de motorización tiene una alta incidencia en la ocurrencia de accidentes, hecho que ha sido altamente ratificado en otras investigaciones (Wu *et al.*, 2014; Qin y Reyes, 2011; Qin *et al.*, 2010). Sin embargo, su efecto es bajo dado el alto valor que presenta la media y picos máximos asociada a esta variable (media de 18.130,33 y máximo de 55.868).

Las variables NC y NI resultaron con signo negativo en su estimación, particularmente para el NC. Esto sugiere que la adición de un carril a tramos con mayor frecuencia de accidente dará lugar a mejoras en términos de seguridad vial pues los accidentes deberían disminuir. Curiosamente, la variable NC presenta el efecto de mayor peso dentro del modelo, hecho que posiblemente puede deberse a la baja variabilidad que presenta este parámetro dentro del conjunto de datos ( $DS=0,52$ ), pues son valores típicos para la red vial de carreteras urbanas (Wu *et al.*, 2014). El efecto negativo sobre la frecuencia de accidentes asociada a la variable NI puede deberse al hecho que en países como Colombia donde la circulación vehicular es por la derecha (es decir, el volante se encuentra a la izquierda), los conductores normalmente no se detienen en el carril exterior (derecho) para realizar giros a la derecha que involucren cambio de dirección en intersecciones, situación que no implica peligro de colisiones con vehículos que circulan en sentido opuesto. El efecto contrario fue reportado por Hosseinpour *et al.* (2014), pues en Malasia la circulación vehicular ocurre en el carril izquierdo provocando más conflictos entre vehículos dado que los conductores normalmente paran en el carril exterior (es decir, el carril más cercano

al tráfico en sentido contrario) cuando tienen la intención de girar a la derecha en un carril de entrada en la dirección opuesta; por lo tanto son más propensos a ser golpeado por un vehículo que se aproxima en el carril contrario.

Además, se evidencia la significancia estadística del estimador P50 para el modelo de ocurrencia de accidentes. Sin embargo, su signo negativo en el parámetro indica que a un aumento de la velocidad media en los tramos estudiados los accidentes disminuirán, el cual no es un resultado común a pesar de estar acorde con lo reportado por Navon (2003) y Rangel *et al* (2013). Uno de las posibles causas para explicar este resultado radica en que la DS de la velocidad media de los tramos es muy baja por lo que el rango de velocidades de los segmentos viales sería igualmente bajo. Para el caso en estudio, el valor de la DS para la variable P50 es de 7.81 kph, lo cual concuerda con la evidencia reportada por Rangel *et al* (2013) la cual muestra variaciones de la DS para el parámetro de velocidad de 13.83 kph, 7.00 kph y 6.42 kph para los años 2007, 2008 y 2009, respectivamente. El resto de las variables mostrada en la Tabla 1 no fueron tenidas en cuenta en el modelo presentado en la Tabla 2 dado que no presentaron significancia estadística.

**Tabla 3. Modelo con RL**

Parámetro	Coficiente	EE	<i>t</i>	<i>p</i> >   <i>t</i>	95% IC	
Intercepto	113,153	224,182	0,505	0,6274	-403,811	630,117
LV	0,067	0,035	1,935	0,0890	-0,013	0,147
AC	5,312	20,364	0,261	0,8008	-41,647	52,271
NC	-96,286	55,928	-1,722	0,1234	-225,256	32,683
NI	-0,309	2,169	-0,143	0,8901	-5,310	4,692
TPD	0,008	0,002	3,369	0,0098	0,002	0,013
P50	-2,510	2,920	-0,860	0,4149	-9,242	4,222

Igualmente, con el objetivo de mostrar las potencialidades de la estimación con RC se estimó un modelo a partir de Regresión Lineal (RL). El modelo con RL estimado (ver Tabla 3) posee los mismos parámetros mostrados en la Tabla 2. Al realizar una comparación y destacar el tipo de efectos que se dejarían de detectar al usar RL se evidencia que el EE asociado a cada una de las variables es mayor en el modelo RL con respecto a RC, lo cual es completamente reflejado en la significancia de cada uno de los parámetros, de hecho para un nivel de confianza de 95% ninguna de las variables del modelo RL son significativas (es decir, valores *t* mayores o iguales al valor absoluto de 1,95). El signo para cada una de las variables presenta es el mismo en ambos modelos a excepción del Intercepto. Indudablemente esto evidencia que el enfoque de estimación con RC es favorable respecto a otros tradicionales de regresión (como RL) ya que no asumen ningún supuesto de distribución sobre el error y es menos sensible a la violación de los mismos. Además, RC es capaz de proporcionar estimaciones a diferentes niveles de cuantiles, efecto que captura la heterogeneidad de los datos y presentar una descripción más acertada de las tendencias.

#### 4.2. Determinación de tramos críticos y ranking de peligrosidad

El criterio para determinar si un tramo es crítico en términos de accidentalidad, se basó en lo expuesto por Qin *et al* (2010), el cual consiste en la comparación entre los valores de los accidentes observados con los valores de los accidentes modelados. Un tramo se clasificó como crítico en término de accidentes si el número de los accidentes observados es mayor al número de

los accidentes modelados. El número de accidentes modelados se obtiene del modelo de frecuencias de accidentes calibrados con RC para cada tramo  $i$ .

La elaboración del ranking de peligrosidad de los tramos en estudio, se realizó solo para los tramos que fueron clasificados como críticos. Esta clasificación de los tramos se basó en la diferencia de los cuantiles, es decir, la diferencia del valor de los accidentes observados y el valor de los accidentes modelados (Qin *et al*, 2010). El ranking se realiza ordenando de mayor a menor dicha diferencia.

Los resultados de la determinación de tramos críticos y el ranking de peligrosidad se presentan en la Tabla 4, el cual arroja como tramos críticos un total de 7 de los 15 estudiados (46,67%). La distribución de los tramos declarados como críticos dentro de la red vial se puede ver en la Figura 4. Cabe destacar que una de las hipótesis de trabajo se comprueba con los resultados obtenidos. Por ejemplo, los tramos identificados con las letras D, E, F y G aunque evidencian una aparente baja frecuencia de accidentes (o por lo menos inferior a la media de 78,4 accidentes) resultaron ser críticos, mientras que el tramo L aunque aparenta tener una alta cantidad de accidentes (180 en total) no resultó ser crítico. Esto constituye una prueba adicional donde el método RC logra proporcionar una selección más específica y discriminatoria de los tramos, pues identifica tramos clasificados como falsos negativos o falsos positivos, los cuales no permiten invertir de una manera más eficiente los presupuestos destinados a seguridad vial.

**Tabla 4. Determinación de tramos críticos**

Tramo	Accidentes Modelados	Accidentes Observados	Estado	Ranking
A	187,862	188	Crítico	1
B	237,927	238	Crítico	2
C	200,948	201	Crítico	3
D	28,965	29	Crítico	4
E	52,966	53	Crítico	5
F	19,972	20	Crítico	6
G	28,981	29	Crítico	7
H	12,901	4	No Crítico	-
I	67,08	48	No Crítico	-
J	80,39	54	No Crítico	-
K	54,596	26	No Crítico	-
L	220,239	180	No Crítico	-
M	87,832	42	No Crítico	-
N	102,349	45	No Crítico	-
O	121,613	19	No Crítico	-
Promedio	100,3	78,4		

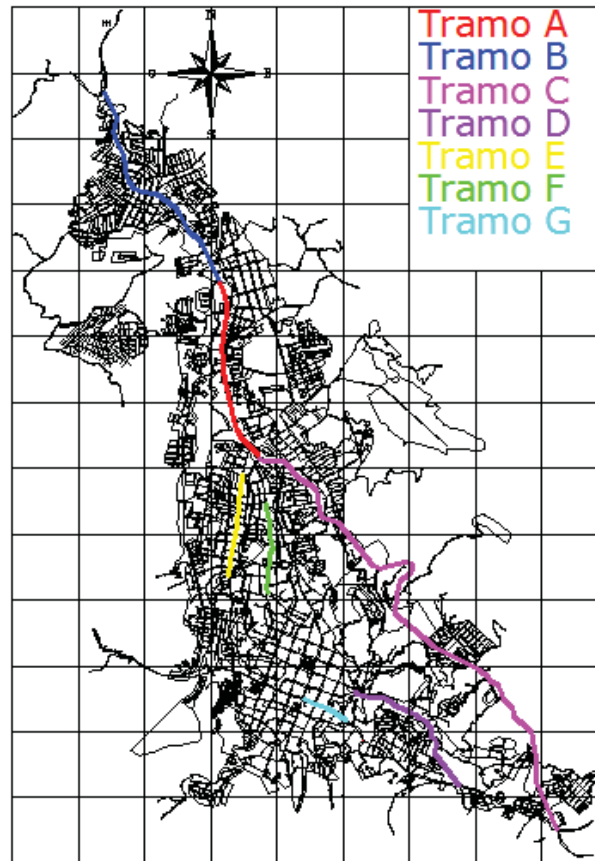


Figura 4. Distribución espacial de los tramos críticos

## 5. CONCLUSIONES

El enfoque metodológico de RC aplicado sobre la ocurrencia de accidentes sobre la red vial de Ocaña arroja resultados coherentes y acertados ratificando que esta aproximación metodológica contribuye y es pertinente para estudios del fenómeno de la accidentalidad, más exactamente en la identificación de tramos críticos de accidentes.

Fue posible encontrar relaciones de causalidad para el modelo correspondiente al percentil 95 entre características como longitud del tramo vial, ancho de calzada, número de carriles, número de intersecciones, tránsito promedio diario y velocidad media con la frecuencia de accidentes. Dichas variables resultaron tener un efecto significativo e influyen de forma directa sobre la frecuencia de accidentes, lo cual es acorde con evidencia reportada en estudios previos.

El modelo estimado muestra que la variable TPD presenta la mayor significancia estadística dentro de la frecuencia de los accidentes. Para este estudio, es claro que un aumento en el tráfico vehicular aumenta la cantidad de accidentes. El efecto producido por la variable P50 fue contrario al que inicialmente se esperaba. Sin embargo, este resultado fue acorde con la literatura explorada en los antecedentes. El efecto de variables como NC sugiere que la adición de un carril a tramos con mayor frecuencia de accidente dará lugar a mejoras en términos de seguridad vial, pues los accidentes disminuirían. Este efecto es similar para la variable NI, mientras que las características

asociadas a la geometría y el entorno vial (p.e., LV y AC) aportan una mejoría notable a medida que disminuyen las longitudes de los tramos y los anchos de calzada.

Además, se encontraron siete tramos críticos de accidentes entre un total de 15 estudiados. Este resultado corresponde al 46,67% de los tramos, valor alto comparado con otros reportes de investigación donde se encuentra alrededor del 5% (Qin *et al*, 2010) y 2,5%, 5%, 10% dependiendo del cuantil aplicado (Washington *et al*, 2014). Este resultado puede deberse a la baja cantidad de tramos estudiados con respecto a los 1710 y 1770 sitios reportados en dichos estudios, respectivamente. Igualmente a partir de la determinación de los tramos críticos se puede conocer a priori la verdad (hasta hoy desconocida) de los indiscutibles tramos críticos de accidentes en el municipio de Ocaña, minimizando de esta forma la identificación de falsos positivos o falsos negativos que desvíen los resultados de la investigación y no permitan la inversión adecuada de los recursos en tramos viales en los cuales no es necesario.

Una vez identificados los tramos críticos de accidentes, se pudo establecer un ranking de peligrosidad, el cual se constituye en una herramienta para toma de decisiones por parte de las entidades gubernamentales cuyo objetivo es mejorar la seguridad vial. De esta forma, es posible establecer prioridades de inversión a los tramos dependiendo de su peligrosidad, además de la implementación de políticas preventivas y/o correctivas que permitan maximizar los beneficios asociados a la seguridad vial.

Futuras investigaciones pueden incorporar al registro de la base de datos otras variables como las asociadas al estado físico y/o mental del conductor (p.e., uso de alcohol/drogas, edad, fatiga), condición climática, características de los vehículos involucrados y severidad de los accidentes, para desarrollar modelos que permitan una mejor comprensión de los diversos factores y sus efectos. Además, se explorará de forma alternativa la calibración de modelos multivariados como Binomial Negativa de efectos aleatorios, Poisson truncado, Binomial Negativa truncado, Poisson cero inflado, Binomial Negativa cero inflado, Empírico Bayesino, Full Bayesian, entre otros. Igualmente puede abordarse una mayor cantidad de tramos, de manera que se cubra en su totalidad la red vial urbana de la ciudad.

## REFERENCIAS

Aguero-Valverde, J. (2013). Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. **Accident Analysis and Prevention**, 50, 289-297.

Also, J., y Langley, J. (2001). Under reporting of motor vehicle traffic crash victims in New Zealand. **Accident Analysis and Prevention**, 33, 353–359.

Ayati, E., y Abbasi, E. (2011). Investigation on the role of traffic volume in accidents on urban highways. **Journal of Safety Research**, 42, 209-214.

Bhalla, K. (2013). *The Cost of road injuries in Latin America*. Inter-American Development Bank, Washington D.C.

- Brenac, T., y Clabaux, N. (2005). The indirect involvement of buses in traffic accident processes. **Safety Science**, 43, 835–843.
- Chang, L. Y. (2005). Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. **Safety Science**, 43, 541–557.
- Cheng, C. (2005). **An Introduction to Quantile Regression and the QUANTREG Procedure**. Cary, NC: SAS Institute Inc.
- Cheng, W., y Washington, S. P. (2005). Experimental evaluation of hotspot identification methods. **Accident Analysis and Prevention**, 37, 870-881.
- Consejo Nacional de Política Económica y Social. (2013). **Documento CONPES 3764**. Bogotá D.C.
- Deublein, M., Schubert, M., Adey, B. T., Köhler, J., y Faber, M. H. (2013). Prediction of road accidents: A Bayesian hierarchical approach. **Accident Analysis and Prevention**, 51, 274-291.
- El-Basyouny, K., y Sayed, T. (2013). Depth-based hotspot identification and multivariate ranking using the full Bayes approach. **Accident Analysis and Prevention**, 50, 1082-1089.
- Geurts, K., y Wets, G. (2003). **Black spot analysis methods: Literature review**. Diepenbeek (Belgium): Flemish Research Center for Traffic Safety.
- Geurts, K., Wets, G., Brijs, T., Vanhoof, K., y Karlis, D. (2006). Ranking and selecting dangerous crash locations: Correcting for the number of passengers and Bayesian ranking plots. **Journal of Safety Research**, 37, 83 – 91.
- Hauer, E. (1997). **Observational Before–after Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety**. Oxford/UK: Pergamon Press/Elsevier Science Ltd.
- Hewson, P. (2008). Quantile regression provides a fuller analysis of speed data. **Accident Analysis and Prevention**, 40, 502-510.
- Hosseinpour, M., Shukri-Yahaya, A., y Farhan-Sad, A. (2014). Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes. Case studies from Malaysian Federal Roads. **Accident Analysis and Prevention**, 62, 209-222.
- Instituto Nacional de Medicina Legal y Ciencias Forenses. (2012). **Forensis 2012 Datos para la vida**. Bogotá D.C.
- Koenker, R. (2005). **Quantile Regression**. Cambridge: Cambridge University Press.
- Koenker, R., y Bassett, G. (1978). Regression quantiles. **Econometrica**, 46, 33-50.

- Koenker, R., y Hallock, K. F. (2001). Quantile regression. **Journal of Economic Perspectives**, 15, 143-156.
- Lao, Y., Zhang, G., Wang, Y., y Milton, J. (2014). Generalized nonlinear models for rear-end crash risk analysis. **Accident Analysis and Prevention**, 62, 9-16.
- Liu, X., Saat, M. R., Qin, X., y Barkan, C. P. (2013). Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. **Accident Analysis and Prevention**, 59 (0), 87-93.
- Martinez-Silva, I. (2011). **Regresión Cuantil**. Tesis de Maestría, Universidade de Santiago de Compostela (USC), Universidad de A Coruña y Universidade de Vigo, Santiago de Compostela.
- Miaoua, S.-P., y Song, J. J. (2005). Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion AND and spatial dependence. **Accident Analysis and Prevention**, 37, 699-720.
- Miranda-Moreno, L. F., Labbe, A., y Fu, L. (2007). Bayesian multiple testing procedures for hotspot identification. **Accident Analysis and Prevention**, 39, 1192-1201.
- Navon, D. (2003). The paradox of driving speed: two adverse effects on highway accident rate. **Accident Analysis and Prevention**, 35 (3), 361-367.
- Noland, R. B., y Oh, L. (2004). The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. **Accident Analysis and Prevention**, 36 (4), 525-532.
- Qin, X., y Reyes, P. E. (2011). Conditional Quantile Analysis for Crash Count Data. **Journal of Transportation Engineering**, 137 (9), 601-607.
- Qin, X., Ng, M., y Reyes, P. E. (2010). Identifying crash-prone locations with quantile regression. **Accident Analysis and Prevention**, 42, 1531-1537.
- Rangel, T., Vassalloa, J. M., y Herraiz, I. (2013). The influence of economic incentives linked to road safety indicators on accidents: The case of toll concessions in Spain. **Accident Analysis and Prevention**, 59, 529-536.
- Wang, C., Quddus, M. A., y Ison, S. G. (2011). Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. **Accident Analysis and Prevention**, 43 (6), 1979-1990.
- Washington, S., Haqueeb, M. M., Oh, J., y Lee, D. (2014). Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots. **Accident Analysis and Prevention**, 66, 136-146.

Wood, J. S., Donnell, E. T., y Port, R. J. (2015). Comparison of safety effect estimates obtained from empirical Bayes before–after study, propensity scores-potential outcomes framework, and regression model with cross-sectional data. **Accident Analysis and Prevention**, 75, 144-154.

World Health Organization. (2013). **Global status report on road safety 2013: supporting a decade of action**. Department of Violence y Injury Prevention y Disability, Geneva.

Wu, H., Gao, L., y Zhang, Z. (2014). Analysis of Crash Data Using Quantile Regression for Counts. **Journal of Transportation Engineering**, 140 (4).