

Modelo de predicción de gravedad de accidentes de tránsito: un análisis de los siniestros en Bogotá, Colombia

Sonia E. Monroy Varela, Hernando Díaz

Profesores Titulares, Universidad Nacional de Colombia, Colombia

semonroyv@unal.edu.co hdiazmo@unal.edu.co

Edwin Urbano, Juan Sánchez, Iván Quintero

Estudiantes Facultad de Ingeniería, Universidad Nacional de Colombia, Sede Bogotá

Resumen

Las proyecciones oficiales para la ciudad de Bogotá D.C. estiman un incremento poblacional del 24% en 2020 respecto a los 10 millones actuales. Más del 75% de los habitantes se movilizan diariamente en los diversos medios de transporte o como peatones. Se trata, sin duda de un sistema complejo.

Usando datos de la Secretaría de Movilidad que incluyen registros de accidentes viales entre los años 2007-2016, se realizó un análisis descriptivo, cualitativo y cuantitativo de los siniestros viales en la ciudad con visualizaciones multidimensionales. También se realizó un estudio analítico de la gravedad de accidentes mediante técnicas de clasificación y un modelo de predicción. Para predecir la gravedad de los accidentes, cuantificada en las consecuencias sobre las personas involucradas, se diseñó un modelo de regresión logística que permite identificar las variables clave. La clasificación se realizó mediante algoritmos de estratificación (clustering) de variables mixtas, utilizando la técnica *k-prototypes*.

Los resultados muestran que cerca del 52% de las víctimas mortales son peatones. Las mujeres representan aproximadamente el 23% de los decesos y los accidentes mortales son más frecuentes entre los 20 y 30 años. En la clasificación, se observa que 4 de 7 clústeres incluyen más del 50% de víctimas peatones y en la misma medida, atropello como tipo de accidente. Los clústeres 4 y 5 están conformados en 96% y 84% por hombres, motociclistas en 67% y 44% respectivamente. La probabilidad de muerte por atropello es sólo 4% mayor que en un choque;

es mayor si el involucrado es un peatón y los hombres presentan una probabilidad de morir 1,79 veces mayor que la de las mujeres. Los resultados del análisis sugieren estrategias de prevención focalizadas principalmente en peatones, con edades entre 20 y 30 años y en motociclistas, a la hora de formular políticas públicas sobre seguridad vial.

Palabras clave

Accidentalidad, clúster, modelo de predicción, visualizaciones & víctimas.

Introducción

Cada día mueren 18 personas en las calles de Colombia (Justicia El Tiempo, 2017) a causa de algún accidente relacionado con el tránsito en las ciudades capitales. Este trabajo desarrolla un estudio descriptivo y cuantitativo de los siniestros viales en la capital colombiana, con ayuda de visualizaciones multidimensionales para analizar dichos decesos y lesiones en las vías bogotanas, utilizando datos oficiales de la Secretaria de Movilidad. Llevando el análisis a un nivel más avanzado, el estudio plantea y desarrolla la clasificación de los factores de relevancia en la gravedad de los accidentes en la ciudad y se diseña un modelo de predicción con el fin de contar con una herramienta de evaluación de políticas públicas en Bogotá, extensibles al resto del país.

Los accidentes viales no solo generan pérdidas humanas en las vías (Ruiz & Herrera, 2016), sino que también afectan la movilidad en la ciudad. Según un análisis publicado el 10 de abril de 2018 por la corporación “Bogotá cómo vamos”, en promedio una persona gasta diariamente hasta dos horas de su tiempo en llegar a su trabajo (Bogotá como vamos, 2018). Se concluyó que el tiempo que los bogotanos gastan en llegar a su lugar de trabajo es inversamente proporcional a su estrato socioeconómico. El análisis de los accidentes viales entonces no solo es favorable para el diseño de políticas de reducción de accidentes fatales, sino también de mejora de la movilidad de todos los capitalinos.

Existen ya varios estudios que evidencian la problemática de la movilidad en la ciudad, enfatizan en que un alto porcentaje de estos inconvenientes en la movilidad y el retraso en la misma son consecuencias de los accidentes de tránsito sobre las grandes arterias viales de la capital colombiana. Ruiz y Herrera por ejemplo, plantean una investigación que busca plantear políticas públicas de calidad enfocadas a los grupos sociales más vulnerables en los accidentes de tránsito, identificando fenómenos sociales en la accidentalidad vial en Colombia, (Ruiz & Herrera,

2016). Por otra parte, en un estudio planteado en el 2018 para el análisis de accidentes de tránsito en Nigeria, mediante modelos de predicción se encontró que la incorporación de factores humanos, de vehículos y algunos relacionados con el medio ambiente, produce un modelo predictivo más sólido que el uso exclusivo del conteo de bloqueos agregados (Ihuezze & Onwurah, 2018).

¿Cómo reaccionar entonces ante los problemas de movilidad causados por los accidentes viales? Ponnaluri & Santhi sugieren que este es un problema que comparten entidades públicas y privadas (Ponnaluri & Santhi) y que es necesario incluir la investigación como una herramienta fundamental para generar los recursos requeridos para el planteamiento de políticas públicas efectivas que vayan enfocadas a los actores más vulnerables en estos siniestros viales (Ponnaluri & Santhi). Por su parte (Savolainen, Manneringb, Lordc, & Quddusd, 2011) agrega la importancia de incluir a los fabricantes de automotores en la creciente preocupación por la reducción de los daños humanos en las vías de grandes urbes alrededor del mundo.

Los modelos logit y los métodos de clase latente son ampliamente utilizados para evaluación de la gravedad de los accidentes de tránsito (Cerwick, Gkritza, Shaheed, & Hans, 2014). De la evaluación realizada por (Cerwick, Gkritza, Shaheed, & Hans, 2014) se obtuvo que el método clase latente presenta una ligera superioridad sobre el Mixed Logit en términos de ajuste del modelo, aunque las probabilidades del pronóstico de este último sean más acertadas. En este trabajo, con el desarrollo de un modelo de regresión logística multinomial se identifican los factores pertenecientes a las características de los accidentes, que tienen mayor incidencia sobre la gravedad de los mismos. Adicionalmente, con un modelo basado en “*embeddings*” para las variables categóricas, se plantea un algoritmo para la predicción del número de víctimas en los siniestros viales.

Abordar completamente la correlación espacial y temporal de los datos de gravedad de las lesiones, la endogeneidad de las variables explicativas en el campo de las lesiones causadas por accidentes, el impacto potencial de la falta de los reportes de los accidentes y la disponibilidad de nuevos tipos de datos en un futuro cercano, son los cuatro desafíos metodológicos que (Savolainen, Manneringb, Lordc, & Quddusd, 2011) plantean en su investigación sobre la severidad de los accidentes de tránsito para esta década.

Finalmente se debe resaltar que a través de la caracterización de los grupos sociales, las condiciones bajo las cuales son más concurrentes los accidentes viales y los modelos de

predicción de víctimas y gravedad (Ruiz & Herrera, 2016), se proporcionan insumos suficientes para el planteamiento de políticas públicas que contemplen medidas cuya finalidad sea la reducción de los accidentes en la ciudad y en efecto la mejora en la movilidad. En concordancia desde hace un año la alcaldía mayor ha intentado adoptar una política pública de seguridad vial reconocida internacionalmente y aplicada en grandes urbes internacionales, “Visión Cero”, bajo el lema “Ninguna muerte es tolerable, todas son evitables”. Siguiendo los mismos lineamientos de esta política pública esta investigación se plantea la siguiente pregunta: ¿Hacia qué actor o actores viales se deben enfocar las políticas públicas para reducir los accidentes de tránsito de forma tal que no haya muertos en las vías y los lesionados se reduzcan al mínimo?

Metodología

Los datos fueron obtenidos a través de la participación de miembros del equipo de investigación en un concurso de análisis de datos Hackathon “Vías, Datos y Vidas”. Este fue organizado por la Secretaría de Movilidad de Bogotá. Allí, la entidad pública brindó, entre varias bases de datos, un histórico de las víctimas, lesionados y muertos, entre enero de 2007 y agosto de 2017. Los registros de las víctimas cuentan con la información del lugar del accidente (dirección, datos de georeferencia y localidad), clase de accidente (atropello, choque, caída de ocupante, incendio, volcamiento, autolesión u otro), fecha y hora de ocurrencia, condición de la víctima (peatón, pasajero, motociclista, ciclista o conductor), edad y género de la víctima. Adicionalmente, se registraron los tipos de vehículos involucrados, la causa del accidente y las características de la vía.

En primer lugar, para llegar al análisis descriptivo de las víctimas fue necesario revisar cuáles variables eran significativas. Para ello se estableció, a partir de la frecuencia, cuáles eran los niveles de la variable que no eran representativos. Estos últimos fueron agrupados en el nivel ‘otros’. En los casos en que un solo nivel de la variable era representativo, ésta no fue tomada en cuenta para el análisis.

En segundo lugar, en la medida en que fue necesario realizar un cruce de bases de datos para obtener la información de los accidentes en los que estaban involucradas las víctimas a analizar, se tuvo en cuenta la relación uno a muchos de accidentes a víctimas. Esto implicó la creación de variables binarias que representaran cada uno de los niveles de las variables propias de los accidentes. Este procedimiento fue realizado por medio de la función *createDummyFeature* del

paquete mlr de R. (Vieira, 2018). Es pertinente aclarar que se seleccionaron los registros entre enero de 2007 y diciembre de 2016 debido a que el año 2017 no había terminado y por lo tanto no estaba completo.

A partir de este tratamiento de datos, se obtuvo una base representativa de las víctimas, que para este análisis son lesionados y muertos. Se realizó una primera fase exploratoria utilizando visualizaciones multidimensionales. En seguida se realizó un análisis de clasificación por medio del algoritmo *K-prototype* para variables categóricas y numéricas por medio del paquete *clustMixType* de R (Szepannek & Mining, 2017). El algoritmo, en forma similar al conocido *k-means*, iterativamente vuelve a calcular prototipos de *cluster* y reasigna clusters. Los *clusters* se asignan usando la distancia

$$d(x, y) = d_{euclid}(x, y) + \lambda d_{simple\ matching}(x, y) \quad (1)$$

Donde la primera es la distancia euclideana medida en los atributos numéricos y el segundo término es la medida de similitud de coincidencia simple en los atributos categóricos. En este último término, el peso λ se usa para evitar favorecer cualquier tipo de atributo y la distancia de coincidencia simple se define como:

$$d_{simple\ matching}(x, y) = \sum_{j=1}^m \delta(x_j, y_j) \quad (2)$$

$$\delta(x_j, y_j) = \begin{cases} 0 & (x_j = y_j) \\ 1 & (x_j \neq y_j) \end{cases} \quad (3)$$

La medida de la diferencia entre X e Y puede ser definida por los desajustes totales de las categorías de atributos correspondientes de los dos objetos (Huang, 1998). En este procedimiento, se obtiene un número recomendado de *cluster* donde está conformado por un número de datos con un conjunto de atributos similares en cada una de las variables que se proponen. En el caso de las variables cuantitativas, el resultado es una descripción en media, mediana y cuantiles; mientras que, en el caso de las variables cualitativas, el resultado es un porcentaje de participación de cada nivel de la variable, es decir, la suma de todos los niveles debe ser 1.

Finalmente, se realizó un modelo de regresión logística multinomial para estimar la gravedad de los accidentes a partir de las variables representativas seleccionadas. Para encontrar las variables que tienen mayor influencia sobre el número de víctimas mortales causadas por un

accidente de cierto tipo, utilizamos un modelo de regresión logística para obtener los “chances” (*odds ratio*) de un resultado, comparados con los de obtener el resultado complementario. Nuestra intención es estimar cómo afectan, en un accidente vial, las diversas características de los vehículos, geográficas, temporales y de las personas involucradas, la probabilidad de que se produzcan fatalidades (una o más personas muertas como resultado del accidente). En este caso los dos resultados posibles, dado que ocurrió un accidente, se refieren a si hubo muertos o no, consecuencia del accidente.

Si π es la probabilidad de que haya víctimas fatales en un accidente, entonces $1 - \pi$ es la probabilidad de que no las haya. Definimos los “chances” (*odds ratio*) de accidente fatal comparados con los de accidente no fatal, como

$$\frac{\pi}{1 - \pi}. \quad (4)$$

Esta cantidad puede ser difícil de estimar puesto que está restringida a ser positiva. La práctica común es utilizar el logaritmo para disponer de una variable con valores positivos y negativos. Por lo tanto, definimos la función logit, en la forma

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right). \quad (5)$$

Puesto que esta cantidad puede ser positiva o negativa, podemos utilizar una función lineal de las variables independientes para calcular estos chances. Utilizando una regresión lineal estándar, se puede entonces calcular los chances η , a partir de éstos encontrar las probabilidades, invirtiendo la función, como sigue:

$$\pi = \text{logit}^{-1}(\eta) = \frac{e^\eta}{1 + e^\eta}. \quad (6)$$

El modelo usual de regresión logística supone que los resultados posibles de un experimento aleatorio (en nuestro caso, un accidente) ocurren como un proceso de Bernoulli.

Se planteó entonces una respuesta de la forma:

$$y_i = \text{número de víctimas muertas en un accidente de tipo } i$$

El tipo i de accidente hace referencia al grupo resultado de una combinación de los niveles de las variables que se tuvieron en cuenta en el modelo. Por ejemplo, las víctimas mortales presentadas en un atropello a peatón, en una vía de tres carriles, con un automóvil, etc.

El modelo usual de regresión logística supone que los resultados posibles de un experimento aleatorio (en nuestro caso, un accidente) ocurren como un proceso de Bernoulli. Por lo tanto, el número de muertos y_i en n_i ensayos (accidentes) está descrito por una distribución binomial. De esta forma, la distribución de Y_i es binomial con parámetros π_i y n_i observaciones de accidentes en el grupo tipo i . Y la probabilidad de que y_i accidentes sean mortales entre los n_i ocurridos está dada por:

$$\Pr \{Y_i = y_i\} = \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \quad (7)$$

El modelo de regresión logística utiliza, entonces, una regresión lineal sobre las variables independientes para calcular el *odd ratio* de muertos vs sólo heridos en un accidente. Por lo tanto, el modelo supone que

$$\eta_i = \text{logit}(\pi_i) = \log \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (8)$$

donde x_1, \dots, x_k , representan las k variables independientes cuyos efectos queremos observar. En nuestro caso, la regresión logística se realizó con la función *glm* del paquete *stats* en R.

A partir de los resultados de la regresión, es posible calcular las probabilidades relativas de accidentes mortales bajo diferentes condiciones. Esto es especialmente útil en el caso de variables con niveles discretos (factores) para las cuales se pueden estimar las probabilidades de accidentes mortales dado un nivel de la variable, comparativamente con otro(s) nivel(es). Por ejemplo, es posible evaluar la probabilidad de víctimas mortales en un accidente de tipo choque comparada con la probabilidad de víctimas mortales en un accidente tipo atropello.

Análisis de datos.

Para el análisis de datos se consideran las series de tiempo mensuales de lesionados y muertos. En la figura 1 se presenta la evolución en el tiempo del número de lesionados. Se observa que el mayor número de casos registrados se encuentra en el primer trimestre de 2007 con un registro superior a los 1500 lesionados. A partir de este trimestre, una tendencia decreciente inicia hasta

el menor registro encontrado en enero del 2009, presentando luego un crecimiento hasta la primera mitad del 2010, desde donde se evidencia un comportamiento uniforme en la tendencia.

La figura 2 representa los registros mensuales de muertos para los accidentes en la capital, donde los mayores registros se presentan en la segunda mitad del 2007 y en el primer semestre de 2015, superando las 55 muertes. El menor número de siniestros viales se presenta en junio de 2008 con tan solo 20 víctimas, siendo también picos bajos junio y febrero de 2011 y 2013 respectivamente, situándose por debajo de los 25 casos.

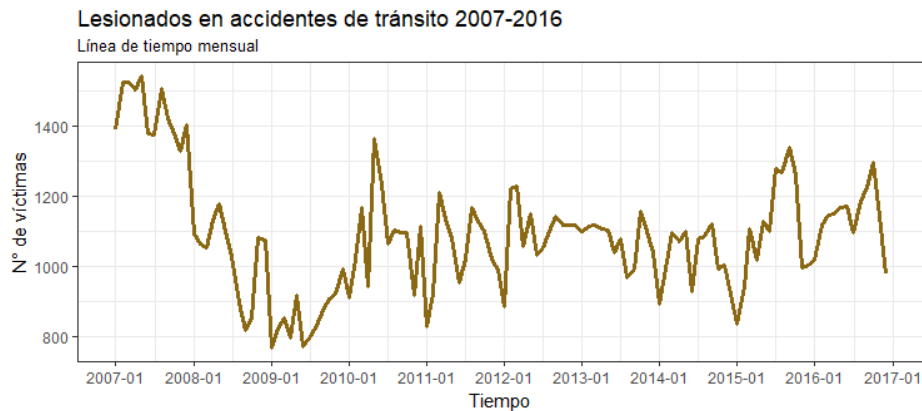


Ilustración 1. Número de registros mensuales de accidentes de tránsito con lesionado entre el 2007 y 2016 [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad y elaboración propia utilizando el software estadístico R]

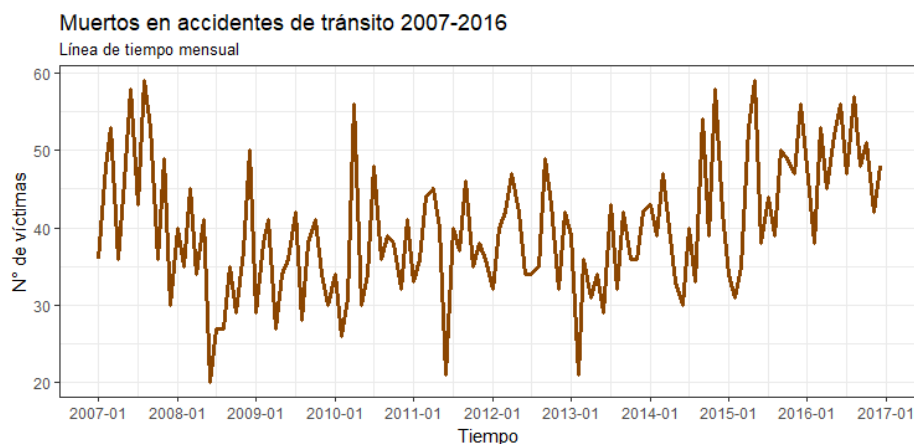


Ilustración 2. Número de registros mensuales de accidentes de tránsito con muertos entre el 2007 y 2016 [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad y elaboración propia utilizando el software estadístico R]

En las figuras 4 y 5 se visualiza el número de registros para los lesionados y los muertos en accidentes de tránsito respectivamente, discriminados por género y rangos de edad. De la gráfica de muertos se observa que en los rangos de edades de 15 a 90 años el número de hombres muertos en accidentes de tránsito supera el doble del número de mujeres. Es claro que el número total de hombres muertos supera ampliamente el de las mujeres durante el período.

También es notable el pico pronunciado de los 20 a los 30 años, aunque el de mujeres es menor y menos pronunciado. De la gráfica de los lesionados se observa que el número de registros de hombres es mayor que el de mujeres para todos los rangos de edad y que entre 20 y 40 años el número de hombres lesionados prácticamente dobla el número de mujeres. Hacia las colas en el eje de las edades la diferencia entre el número de registros discriminado por género disminuye, siendo aún los hombres los mayores afectados por los siniestros, a pesar de que, a edades avanzadas, la población de mujeres es mayor que la de hombres.

El eje de la edad presenta un mayor rango para el género masculino, incluyendo casos registrados desde los cero años y por encima de los cien. Tanto para muertos como para lesionados la “edad crítica”, o mejor, la moda está entre los 20 y los 35 años en ambos sexos.

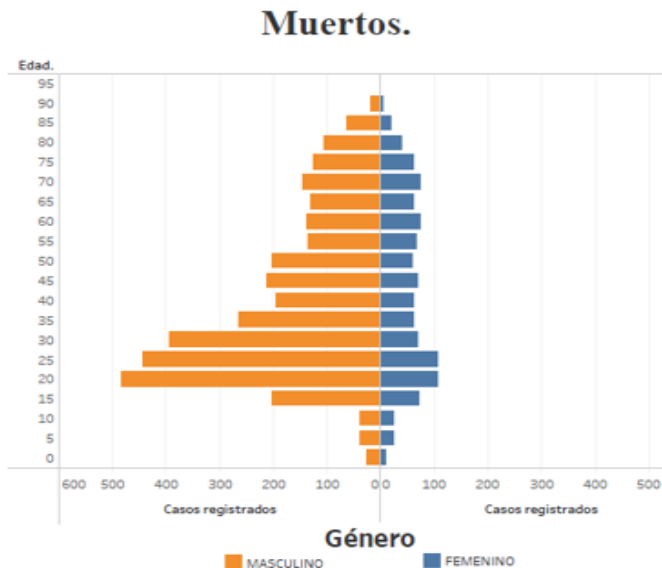


Ilustración 3. Número de registros por edades de los lesionados en accidentes de tránsito entre 2007 y 2016 [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad. Fuente: Base de datos de la Secretaría de Movilidad y elaboración propia utilizando el software Tableau]

Lesionados.

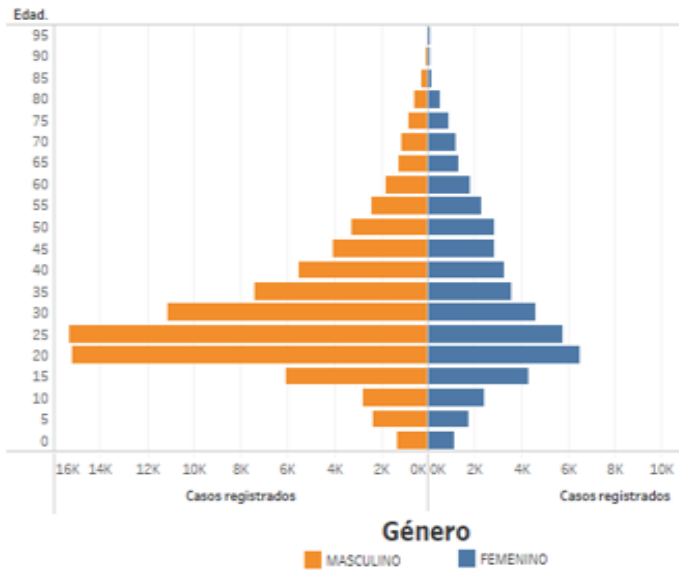


Ilustración 4. Número de registros por edades de los muertos en accidentes de tránsito entre 2007 y 2016 [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad. Fuente: Base de datos de la Secretaría de Movilidad y elaboración propia utilizando el software Tableau]

En las figuras 5 y 6 se representa, por medio de un diagrama de calor (heat map) el número de eventos para cada día de la semana y hora del día, para los registros de lesionados y muertos entre 2007 y 2016 respectivamente. La gráfica de los accidentes muestra que el mayor número de siniestros se presentan entre 7:30 de la mañana y las 7:30 de la noche, presentando el pico más alto entre las siete y nueve de la mañana hora colombiana (COT). Los mayores registros de muertos se presentan a las 7 de mañana y a las 7 de la noche, presentando además un gran número de registros en altas horas de la noche. Durante el día el mayor registro de mortalidad por siniestros viales se presenta entre las siete y las ocho de la mañana COT.

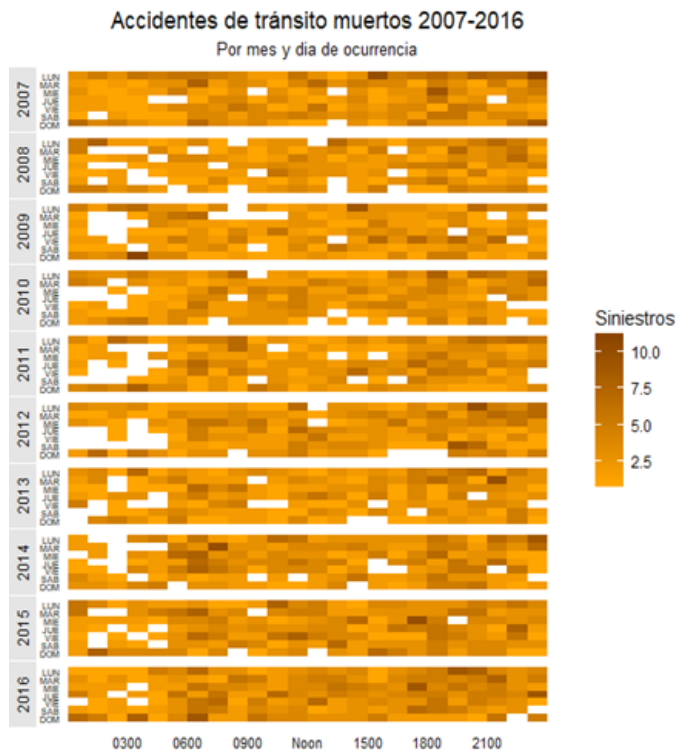


Ilustración 5. Número de registros por hora de los muertos en accidentes de tránsito entre 2007 y 2016 [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad, Fuente gráfica: Software estadístico R]

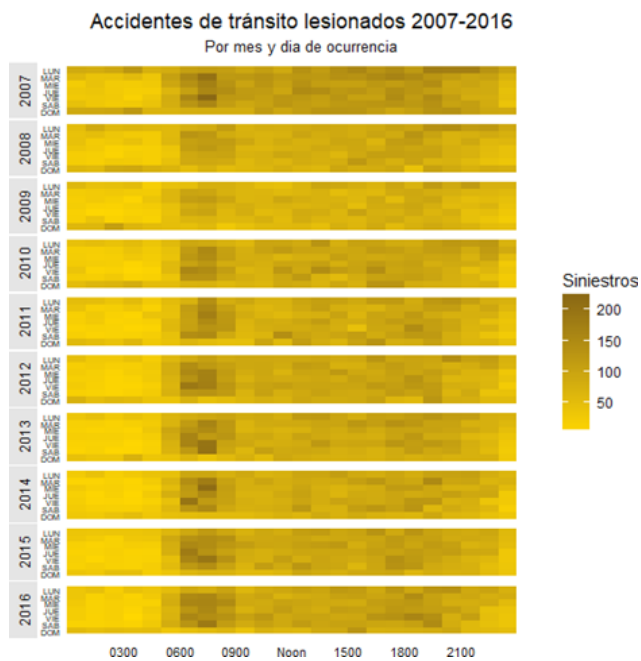


Ilustración 6. Número de registros por hora de los accidentes de tránsito entre 2007 y 2016 [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad, Fuente gráfica: Software estadístico R]

En la figura 7 se presenta un diagrama de esferas, donde el tamaño de cada una es proporcional a la fracción del número de registros ocurridos para cada tipo de vehículo y el color caracteriza el tipo de servicio que presta el vehículo. Los vehículos oficiales y diplomáticos están involucrados en pocos accidentes mientras que los particulares y públicos representan un gran porcentaje del total de los accidentes. Los vehículos que presentan un mayor número de registros son los automóviles particulares y públicos, los buses y busetas públicas y las motocicletas particulares. Los grupos nombrados representan más del 50% del total de los registros.

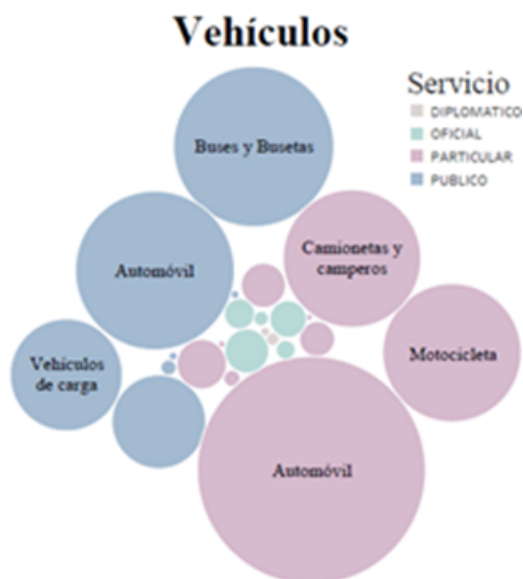


Ilustración 7. Diagrama de esferas de la cantidad por tipo de vehículos involucrados en los accidentes entre el 2007-2016 [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad, Fuente gráfica: Tableau, licencia universitaria]

En las figuras 8 y 9 se presenta una clasificación de los lesionados y muertos en los siniestros viales, respectivamente. Las víctimas se discriminan según la clase del accidente (choque, atropello, caída de ocupante, volcamiento, autolesión u otro) y la condición del afectado (conductor, pasajero, ciclista, peatón, motociclista o sin información). En los registros de lesionados, los choques representan más del 50% de los accidentes, mientras que en los de muertos los atropellos son los que presentan más de la mitad de los registros totales. En los registros de lesionados los choques se distribuyen entre pasajeros, conductores, motociclistas y ciclistas, siendo éstos los grupos más representativos de esta clase de siniestros.

Entre los lesionados por atropellos cerca del 80% son peatones. Mientras que la caída de ocupantes, el volcamiento y las autolesiones afectan principalmente a los pasajeros y los

motociclistas. Más del 90% de los muertos por atropello son peatones y más del 50% de los muertos en accidentes de tránsito se producen por atropello en la ciudad. En los choques, las principales víctimas son los motociclistas, los cuales representan más del 50% de los muertos en este tipo de accidente. Los choques también producen un número significativo, aunque un poco menor, de muertes de pasajeros y ciclistas.

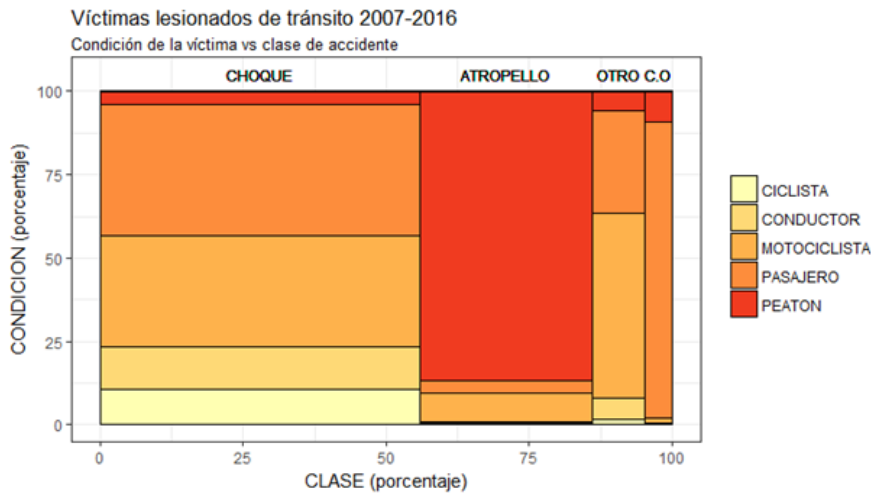


Ilustración 8. Gráfica de proporciones de condición vs Clase de los accidentes con lesionados entre 2007 y 2016 [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad, Fuente gráfica: Software estadístico R]

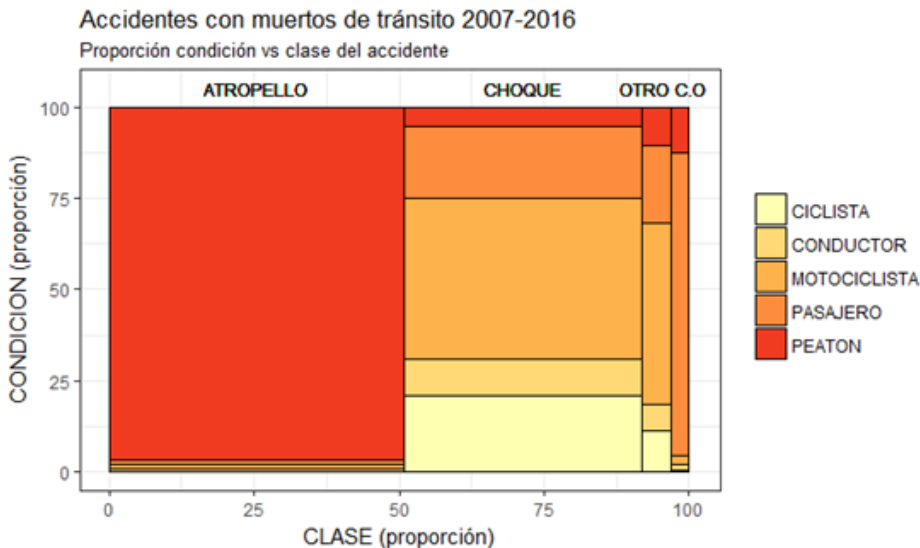


Ilustración 9. Gráfica de proporciones de condición vs1 Clase de los accidentes con muertos entre 2007 y 2016 [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad, Fuente gráfica: Software estadístico R]

Resultados.

Clasificación

Para clasificar los accidentes donde se involucran lesionados y muertos se realizó un análisis de *clusters* el cual consiste en agrupar una colección dada de patrones no etiquetados en un conjunto de grupos similares. El resultado fue el siguiente.

	HORA_OCURRENCIA	MES_OCURRENCIA	DIA_OCURRENCIA	CLASE	LOCALIDAD	CONDICION	GENERO	EDAD	GRAVEDAD
1	13.13233	AGOSTO	MARTES	ATROPELLO	KENNEDY	PEATON	FEMENINO	70.91355	0.08211792
2	13.51013	ENERO	MARTES	ATROPELLO	KENNEDY	PEATON	MASCULINO	52.59321	0.05462243
3	11.39353	AGOSTO	VIERNES	CHOQUE	KENNEDY	PASAJERO	FEMENINO	24.4087	0.01482158
4	13.19903	AGOSTO	MARTES	CHOQUE	ENGATIVA	MOTOCICLISTA	MASCULINO	25.33646	0.02610873
5	11.80432	AGOSTO	SABADO	CHOQUE	KENNEDY	MOTOCICLISTA	MASCULINO	40.5473	0.02909831
6	13.47545	MARZO	MIERCOLES	ATROPELLO	KENNEDY	PEATON	FEMENINO	30.58481	0.03389294
7	14.15411	SEPTIEMBRE	LUNES	ATROPELLO	SUBA	PEATON	MASCULINO	13.11307	0.02540065

Tabla 1. Resultados del análisis de *clusters* de lesionados y muertos. [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad, Software: R®]

Este análisis se realizó por medio del algoritmo *k-prototypes* el cual obtiene *clusters* o grupos de accidentes con la característica de que los objetos de cada *cluster* deben ser lo más similares posible entre sí, y al mismo tiempo cada grupo debe ser lo menos similar posible a los demás agrupamientos.

Se obtuvieron 7 *clusters* y cada uno de ellos tiene un porcentaje de cada una de las características que se tuvieron en cuenta: Hora de ocurrencia, Mes de ocurrencia, Día de ocurrencia, clase, localidad, condición, género y edad.

En la *Tabla 1* se puede observar que cada *cluster* está caracterizado por valores de las características de los accidentes. Las variables con valores numéricos, hora de ocurrencia y edad, están representadas por un promedio de los datos numéricos de las observaciones que hacen parte del *cluster*. Sin embargo, en el caso de la hora de ocurrencia, dicho promedio no quiere decir que la mayoría de accidentes se den a esa hora del día, sino que la mayoría de los accidentes se dan en las horas pico de la mañana y también en las horas pico de la noche, por lo cual el promedio de éstas se da al medio día. Las variables categóricas están representadas por los valores más frecuentes en el *cluster*.

La variable gravedad, para cada *cluster*, está representada por un número entre 0 y 1 donde, entre más cerca se encuentre a 1, quiere decir que hay un mayor porcentaje de muertos en el

cluster. Los datos no se acercan mucho a 1 dado que en la base de datos donde se realizó el análisis hay una mayor cantidad de lesionados que de muertos.

Modelo de gravedad de víctimas

El modelo descrito en la sección de metodología se estimó a partir de los datos de víctimas de accidentes. Se calculan los chances (odds) de que un accidente sea mortal comparado con que el accidente no lo sea. Como se describió anteriormente el logaritmo de esta función se aproxima por medio de una función lineal de las variables. A continuación, se muestran los resultados del modelo estimado:

		Estimate	Std. Error	Z value	Pr(> z)	odds _i
	GENEROFEMENINO	-4.378414	0.593828	-7.373	1.67E-13	*** 0.01254524
	GENEROMASCULINO	-3.804866	0.593034	-6.416	1.40E-10	*** 0.02226218
	HORA_OCURRENCIA	-0.015759	0.002857	-5.517	3.46E-08	*** 0.984364523
	DIA_OCURRENCIAFINDESEM	0.085324	0.0345	2.473	0.013392	* 1.089069868
Atropello	CLASECAIDA_DE_OCUPANTE	0.418496	0.108582	3.854	0.000116	*** 1.519674246
	CLASECHOQUE	-0.095284	0.065718	-1.45	0.147092	0.909114709
	CLASEOTRO	-0.764769	0.101517	-7.533	4.94E-14	*** 0.465441436
Centro	LOCALIDADNORTE	-0.06027	0.059715	-1.009	0.31283	0.941510291
	LOCALIDADOCCIDENTE	0.149219	0.054503	2.738	0.006185	** 1.160927204
	LOCALIDADSUR	0.316646	0.058308	5.431	5.62E-08	*** 1.372516615
Peatón	CONDICIONCICLISTA	-0.216014	0.08628	-2.504	0.012292	* 0.805724022
	CONDICIONCONDUCTOR	-1.079551	0.095743	-11.276	2.00E-16	*** 0.339748038
	CONDICIONMOTOCICLISTA	-0.873408	0.068864	-12.683	2.00E-16	*** 0.417526193
	CONDICIONPASAJERO	-1.404501	0.07659	-18.338	2.00E-16	*** 0.245489525
	EDAD1	1.172847	0.582933	2.012	0.044223	* 3.231178722
	CALZADAS.DOS1	0.348515	0.070709	4.929	8.27E-07	*** 1.416961797
	CALZADAS.TRES.O.MAS1	0.730953	0.075617	9.666	2.00E-16	*** 2.077059102

CALZADAS.UNA1	-0.175958	0.071504	-2.461	0.013863	*	0.838653206
CARRILES.TRES.O.MAS1	0.32052	0.037314	8.59	2.00E-16	***	1.377844057
CARRILES.UNO1	-0.244174	0.059716	-4.089	4.33E-05	***	0.783351319
ESTADO.CON.HUECOS1	0.35908	0.08054	4.458	8.26E-06	***	1.432011358
ESTADO.OTRO1	0.324865	0.124897	2.601	0.009293	**	1.383843814
CONDICIONES.SECA1	-0.132515	0.058481	-2.266	0.023455	*	0.875889796
TIPO_CAUSA.VIA	0.429628	0.053583	8.018	1.08E-15	***	1.53668577

Tabla 2. Resultados del modelo de gravedad de víctimas en accidentes de tránsito. [Fuente datos: Bases de datos de accidentalidad de la secretaría de movilidad, Software: R®]

En la tabla 1 la columna *Estimate* es el coeficiente estimado a partir de una regresión logística con el error estándar de estimación. También se incluyen los valores correspondientes de la distribución normal junto con la probabilidad de que el coeficiente sea cero. Además, se ha incluido una columna que representan los chances (odds) que representan probabilidades relativas con respecto a un valor utilizado como referencia que aparece en la primera columna. En aquellas celdas que no aparece un valor de referencia, la comparación se hace con los valores complementarios.

En el modelo final fueron consideradas las variables más significativas, cuyo valor p fuera menor que 0,05. Un resultado importante tiene que ver con el género. Tal como se aprecia en el análisis descriptivo de las víctimas, la probabilidad de una víctima mortal mujer es aproximadamente la mitad de la probabilidad de una víctima mortal hombre, 0.012 y 0.022 respectivamente.

De los odds_i se obtuvieron resultados comparativos de los niveles de la variable día de la semana, clase de accidente, localidad y condición de la víctima. Un accidente tiene 1.09 veces más probabilidad de tener víctimas mortales el fin de semana (viernes, sábado y domingo), que si ocurre en un día entre semana (lunes, martes, miércoles y jueves). En cuanto a la clase, un accidente con caída de ocupante es 1,52 veces más probable de tener víctimas mortales que un atropello. Asimismo, el choque tiene una probabilidad de tener víctimas mortales que es 0.91

veces la probabilidad de que un atropello sea mortal. En cuanto a la ubicación del accidente, es más probable la ocurrencia de un accidente grave en las localidades del sur y occidente comparado con accidentes en el centro de la ciudad. En la localidad del norte es menos probable

un accidente grave. En la condición de la víctima se encontró que los peatones tienen mayor probabilidad de morir, seguido de los ciclistas y motociclistas. Este resultado concuerda con la vulnerabilidad de estos tres tipos de actores viales a la hora de sufrir un accidente de tránsito. Los conductores y los pasajeros de los vehículos tienen aproximadamente la misma probabilidad de morir en un accidente de tránsito.

En otros resultados sobre el tipo de vía, la mayor probabilidad de ocurrencia de víctimas mortales se presenta en vías de tres calzadas o más y de tres carriles o más. El estado de la vía con presencia de huecos, así como las vías húmedas, hacen más probable un accidente grave.

Discusión

De acuerdo con la política “visión cero”, la Secretaría de Movilidad de Bogotá D.C. asume un gran reto a la hora de formular la política pública que reduzca el número de víctimas mortales a cero y el número de lesionados al mínimo posible. Los autores sugieren, a partir del estudio de los accidentes en el horizonte de tiempo 2007-2016, que es necesario priorizar las medidas a tomar dado que existen unos tipos de accidentes que son sustancialmente más peligrosos en la medida en que representan mayor incidencia en el número de víctimas mortales. En este sentido, se requiere pensar de manera primordial en unas medidas adecuadas para disminuir el número de víctimas mortales peatones, bien sea desde la víctima en sí, o desde los causantes principales, es decir, conductores. En seguida conviene adecuar las medidas a ciclistas y motociclistas quienes presentan similar vulnerabilidad en un accidente de tránsito.

Es indispensable preparar medidas de prevención de accidentes focalizadas en los jóvenes entre 20 y 30 años, fundamentalmente del género masculino. También se debe prestar especial atención a la movilidad durante las horas “pico”, tanto por la mañana como por la noche, dado que son los momentos del día donde mayor accidentalidad se presenta, lo que sugiere un momento crítico para todos los actores viales y donde se requiere la implementación de seguridad vial. De la misma manera, se deben realizar campañas dirigidas al cuidado y protección de los peatones en las vías, dado que son los más vulnerables al momento de ocurrir un accidente.

Durante este periodo de tiempo, se implementaron distintas políticas con el fin de mejorar la movilidad y reducir la accidentalidad dentro de la ciudad. Una de éstas es el pico y placa¹ durante todo el día, la cual fue implementada entre el año 2008 y 2009, tiempo durante el cual se ve una

1: Norma de tránsito que impone una restricción de circulación obligatoria en el área urbana a vehículos privados tipo automóvil y de servicio público en horarios "pico" (horarios con mayor afluencia de tráfico), dependiendo del último número de placa del automóvil.

reducción significativa del número de lesionados y muertos dentro de la ciudad. Esta disminución de accidentes también es atribuida, en el informe de accidentalidad de medicina legal para el año 2009, a las campañas de educación y concientización para todos los actores viales. Estas campañas no cuentan con información cuantitativa que pudiera ser analizada.

Conclusión.

Las técnicas de análisis descriptivo de grandes cantidades de datos a partir de la analítica, permiten encontrar tendencias y patrones. La comparación de estos patrones con los resultados del modelo de gravedad de víctimas planteado ayuda a la validación del modelo pues fundamenta la selección de las variables que se enfatizan como las más incidentes a partir de la visualización de su ocurrencia en los registros actuales.

Los actores más perjudicados por los accidentes de tránsito en Bogotá son los peatones, los cuales representan aproximadamente el 50% de los muertos en los últimos 10 años. En el caso de los lesionados se evidencia que los más afectados son los pasajeros y los motociclistas. En el caso de los motociclistas, durante los últimos años, ha aumentado significativamente su número dentro de la ciudad.

Se evidencia también que un alto porcentaje de accidentes tanto en muertos como en lesionados tienen como involucrados personas entre los 20 y los 30 años en su mayoría de género masculino, lo que sugiere que, a pesar de que la población bogotana tiene una cantidad de hombres y mujeres parecida, son los hombres los que corren más peligro de verse involucrados en este tipo de accidentes. A su vez, la mayoría de los accidentes se presentan en las horas pico, principalmente por la mañana, entre las 6:30 am y las 8:00 am.

El modelo de gravedad de víctimas planteado como un modelo de regresión logística con respuesta binaria permite encontrar fundamentalmente las variables correspondientes a las características del accidente como el tipo de vía, los vehículos involucrados en el accidente, el tiempo, la víctima, entre otros. Después de una depuración de las variables, se identificaron las más significativas. Entre éstas están todas aquellas cuyos coeficientes son significativamente diferentes de cero y cuyos *odds ratios* son comparativamente mayores.

Referencias

- Bogotá cómo vamos. (2018). *Movilidad en Bogotá*. Bogotá D.C.
- Cerwick, D. M., Gkritza, K., Shaheed, M. S., & Hans, Z. (2014). A comparison of the mixed logit and latent class methods for crash severity analysis. *Analytic Methods in Accident Research*, 11-27.
- Chen, C., & Zhang, J. (2016). Exploring background risk factors for fatigue crashes involving truck drivers on regional roadway networks: a case control study in Jiangxi and Shaanxi, China.
- Cheng, W., Gill, G., Zhang, Y., & Cao, Z. (2018). Bayesian spatiotemporal crash frequency models with mixture components for space-time interactions. *Accident Analysis & Prevention*, 84-93.
- Goel, R., Jain, P., & Tiwari, G. (2018). Correlates of fatality risk of vulnerable road users in Delhi. *Accident Analysis and Prevention*, 86-93.
- Huang, Z. (1998). Extensions to the k -Means Algorithm for Clustering Large Data Sets with Categorical Values, *304*, 283–304.
- Ihueze, C. C., & Onwurah, U. O. (2018). Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria. *Accident Analysis and Prevention*, 21-29.
- Justicia El Tiempo. (13 de Diciembre de 2017). 2017 quebró década en aumento de muertes por accidentes de tránsito. *El Tiempo*.
- Ponnaluri, R. V., & Santhi, Y. D. (s.f.). Road Crash History and Risk Groups in India. Need for New Initiatives and Safety Policies.
- Ruiz, J. I., & Herrera, A. N. (2016). Accidentes de tránsito con heridos en Colombia según fuentes de información: caracterización general y tipologías de accidentes. *CES Psicología*, 32-46.
- Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis and Prevention*, 1666– 1676.
- Szepannek, A. G., & Mining, D. (2017). Package “ clustMixType ,” 1–10. <http://doi.org/10.1023/A>
- Vieira, B. (2018). Package “ mlr .”
- Wu, Q., Zhang, G., Zhu, X., Liu, X. C., & Tarefder, R. (2016). Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways. *Accident Analysis and Prevention*, 35-45.

Yasmin, S., Eluru, N., Bhat, C. R., & Tay, R. (2014). A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. *Analytic Methods in Accident Research*, 23-38.

Ye, F., & Lord, D. (2014). Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. *Analytic Methods in Accident Research*, 72-85.