

PROCESO: Gestión de Investigación	MANUAL	Código: II-MA-009
SUBPROCESO: Gestión Editorial y Visibilidad de Publicaciones	PLANTILLA PRESENTACIÓN DE ARTÍCULOS	Versión: 1

ACCIDENTALIDAD VIAL EN CALI CON PERSPECTIVA DE GÉNERO: UN MODELO DE REGRESIÓN LOGÍSTICA

Gender-Based Analysis of Vial Accidents in Cali: A Logistic Regression Model

Santiago Caceres Medina

scaceres1025@javerianacali.edu.co
Pontificia Universidad Javeriana Cali
Colombia

Estudiante de Ingeniería Industrial y Administración de Empresas, Pontificia Universidad Javeriana Cali

Kathleen Georjajna Salazar Serna

Kathleen.salazar@javerianacali.edu.co
Pontificia Universidad Javeriana Cali
Colombia

No. ORCID – 0000-0003-3824-7044

Candidata a doctora en Ingeniería. Universidad Nacional de Colombia. Profesora asistente.
Departamento de Ingeniería Civil e Industrial.

Isabel Cristina García

Isabel.garcia@javerianacali.edu.co
Pontificia Universidad Javeriana Cali
Colombia

Profesora asistente. Departamento de Ciencias Naturales y Matemáticas

Recepción: 10/08/2024

Aceptación: 30/11/2024

DOI:

Resumen

Este trabajo analiza la accidentalidad vial en la ciudad de Cali con un enfoque de género utilizando un modelo de regresión logística. El objetivo es identificar las variables que determinan que las mujeres sean víctimas de un siniestro vial, con el fin de clasificar factores que permitan definir estrategias específicas para la prevención de accidentes. El estudio emplea estadísticas de fallecidos y lesionados proporcionados por la Agencia Nacional de Seguridad Vial para el periodo 2016-2023. Los resultados indican que el modelo tiene una efectividad del 61% en la predicción de accidentes femeninos.

Palabras clave

Accidentalidad vial, enfoque de género, Regresión logística, Movilidad.

Abstract

The paper analyzes traffic accidents in the city of Cali with a gender perspective, using a logistic regression model. The objective is to predict the variables that determine whether an accident is caused by a woman, to implement specific measures. The research is based on data from the National Road Safety Agency. The results indicate that the model has an effectiveness of 61% in predicting female accidents.

Keywords

Accidentality, Gender, Logistic regression, Mobility.

INTRODUCCIÓN

La accidentalidad vial es uno de los problemas en crecimiento a nivel mundial, en donde el 90% de los siniestros fatales son causados por accidentes en países de ingresos bajos y medianos. Según la Organización Mundial de la Salud, aproximadamente 1.19 millones de personas fallecen anualmente por accidentes de tráfico. En Colombia, fallecieron 8,405 personas, en el año 2023. Estas alarmantes cifras promueven el desarrollo, la investigación y la implementación de medidas que permitan disminuir la cantidad de accidentes fatales y no fatales a largo plazo.

En este documento se aborda la accidentalidad en la ciudad de Cali con un enfoque de género, en donde se busca predecir las variables que determinan la probabilidad de que la víctima en un accidente de tránsito es una mujer, a fin de implementar medidas en función de aquellas variables que se manifiesten en respuesta de esta. Este trabajo se realiza en el marco del proyecto de investigación: "Análisis de las dinámicas de movilidad, un reto para formular políticas públicas", el cual es financiado por el fondo para la investigación de la fundación WWB.

El documento se divide en tres secciones; en la primera sección, se presenta la metodología aplicada para la implementación de un modelo de regresión logística que permita alcanzar el objetivo presentado con anterioridad; en la segunda sección, se muestran los resultados del modelo y la capacidad de predicción de este; finalmente, se genera la discusión y conclusiones del trabajo.

MÉTODO

En esta investigación se realizó un análisis transversal de la información proporcionada por la Agencia Nacional de Seguridad Vial (ANSV), que incluyó datos sobre la siniestralidad vial en el Valle del Cauca y Valle de Aburrá. El objetivo consiste en identificar o predecir la accidentalidad vial con un enfoque de género en la ciudad de Cali, utilizando la técnica de regresión logística.

Se eligió este tipo de modelo debido a la naturaleza cualitativa de la variable de respuesta: el género, que tiene dos posibles valores (femenino y masculino). Esto se puede asociar con la ocurrencia o no de un suceso, permitiendo calcular "la probabilidad de que ocurra un hecho en cuestión como función de ciertas variables que se presumen relevantes o influyentes" (Fiuza & Rodríguez, 2000).

Las etapas de implementación del modelo incluyeron la limpieza de la base de datos, la selección de variables y el análisis estadístico exploratorio.

Selección de variables

Las variables seleccionadas buscan ser potencialmente explicativas del fenómeno en estudio, con nulo o poco error de registro, disponibilidad futura y medibles antes del evento (Redondo, 2023). La tabla 1 muestra las variables seleccionadas.

Variable	¿Cómo aporta?
Fecha	Fue utilizada para extraer el año e identificar si hay alguna tendencia con respecto a estos
Día de la semana	Se puede evaluar la incidencia de los días en la probabilidad de que el accidente tenga una víctima femenina, de modo que se implementen medidas relacionadas
Rango de horas	Identificar los rangos horarios con mayor accidentalidad permite la implementación de medidas
Usuario de la vía	Identificar la forma en la que se transportaba la víctima permite la implementación de medidas
Zona reclasificada	Identificar si hay incidencia de la zona en los accidentes con víctima femenina
Sexo	Es la variable de respuesta o la que será dependiente de los demás registros

Rango de edad	Identificar los rangos de edad con mayor accidentalidad permite la implementación de medidas
Circunstancia	A pesar de que es una variable categórica con muchos resultados, se incluye en el modelo para identificar la incidencia de las diferentes circunstancias en la accidentalidad enfocada en género

Tabla 1
Selección de variables

Limpieza de datos

La limpieza de datos es crucial antes de implementar el modelo, especialmente con grandes conjuntos de datos. “Esta etapa se compone de procesos que permiten identificar los registros incompletos, incorrectos, inexactos y/o no pertinentes” (Maté, 2019), mejorando la calidad y reduciendo sesgos en el análisis. Para este proceso, se identificaron casos de variables con respuestas similares pero con diferente escritura, las cuales fueron actualizadas al valor correspondiente. También se omitieron registros con respuestas como “sin información” u “otra”, ya que complican el análisis. Finalmente, se filtraron los datos para obtener solo los registros de la ciudad de Cali.

Análisis estadístico exploratorio

Antes de implementar el modelo de regresión, es crucial un análisis exploratorio de la variable de respuesta y las variables independientes. Dado que ambas son categóricas, las siguientes gráficas comparan el ratio de la variable estudiada respecto a las demás.

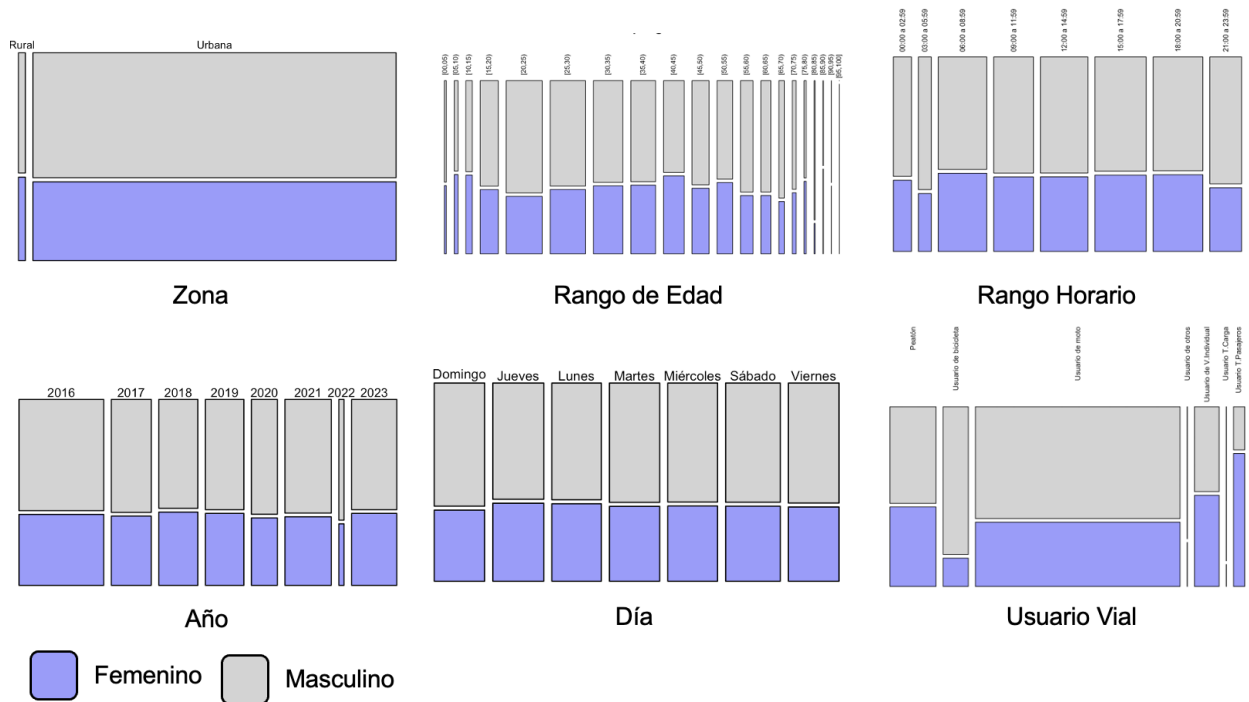


Imagen 2, 3, 4, 5, 6 y 7
Gráficas correspondientes al análisis exploratorio

as gráficas muestran que la proporción de mujeres accidentadas es menor que la de hombres en casi todas las variables, lo cual se relaciona con que las mujeres poseen solo el 30% de las licencias en Cali (ANSV, 2024). Asimismo, se ve un claro reflejo de que los accidentes rurales son mínimos en comparación con los accidentes presentados en la zona urbana. Las variables de año y día de la semana muestran uniformidad en sus categorías, sugiriendo poca relación con la incidencia de género en los accidentes. Finalmente, se observa una mayor accidentalidad en las horas pico de la ciudad y mayor proporción de accidentes en las personas jóvenes.

RESULTADOS

Una vez realizado el procedimiento metodológico, se procede a la implementación del modelo a través del Software estadístico R. En este, se introduce la variable de respuesta y sus dependencias de la siguiente manera:

$$\text{Sexo} \sim \text{as.factor(Año)} + \text{NombreDiaSemana} + \text{Rango3Hora24} + \text{UsuarioVia} + \text{ZonaReclasificada} + \text{RangoEdad} + \text{Circunstancia}$$

Esta regresión logística se realiza con una familia binomial, ya que se busca obtener una respuesta que se expresa como un 0 si el accidentado es de género masculino y un 1 si el accidentado es de género femenino.

El modelo inicial da un criterio de Akaike (AIC) de 12870, de modo que a partir de la función *step* se irán eliminando variables del modelo para determinar si este responde mejor ante dichas modificaciones. El criterio de Akaike es importante al momento de evaluar un modelo, ya que este determina la calidad de este en un análisis combinado de la bondad de ajuste y de su complejidad, por lo que cuanto menor sea su valor, mayor es la calidad global del modelo (Herrera, 2023).

Una vez realizado dicho análisis, se obtiene que la variable de accidentalidad según el género, con un criterio de Akaike de 12823, responde mejor antes las variables rango de edad, rango horario y el usuario de la vía. Variables que anteriormente se habían analizado como aquellas que mostraban una tendencia diferente de los datos con respecto al género. A continuación se presenta el modelo ajustado a las nuevas variables y los coeficientes correspondientes a este en la tabla 2.

$$\text{Sexo} \sim \text{Rango3Hora24} + \text{UsuarioVia} + \text{RangoEdad}$$

		Estimate	Std. Error	z value	Pr(> z)	Signif					
(Intercept)		-0.541059	0.249802	-2.166	0.03031	*					
Rango Horario	03:00 a 05:59	-0.368677	0.138566	-2.661	0.00780	**	[20,25)	0.211312	0.242336	0.872	0.38322
	06:00 a 08:59	0.207022	0.101091	2.048	0.04057	*	[25,30)	0.364335	0.241828	1.507	0.13192
	09:00 a 11:59	0.114935	0.104389	1.101	0.27089		[30,35)	0.419331	0.242709	1.728	0.08404
	12:00 a 14:59	0.134083	0.101317	1.323	0.18570		[35,40)	0.412830	0.243937	1.692	0.09058
	15:00 a 17:59	0.179627	0.100375	1.790	0.07353	.	[40,45)	0.639317	0.245315	2.606	0.00916
	18:00 a 20:59	0.165242	0.100617	1.642	0.10053		[45,50)	0.294918	0.247483	1.192	0.23339
	21:00 a 23:59	-0.121491	0.108730	-1.117	0.26384		[50,55)	0.388993	0.248189	1.567	0.11704
Usuario Via	Usuario de bicicleta	-1.498300	0.110743	-13.530	< 2e-16	***	[55,60)	0.040211	0.253050	0.159	0.87374
	Usuario de moto	-0.462695	0.065048	-7.113	1.13e-12	***	[60,65)	-0.039043	0.258028	-0.151	0.87973
	Usuario de otros	-1.022557	0.822264	-1.244	0.21365		[65,70)	-0.179125	0.277017	-0.647	0.51788
	Usuario de V.Individual	0.245937	0.090780	2.709	0.00675	**	[70,75)	-0.001847	0.290129	-0.006	0.99492
	Usuario T.Carga	-1.860022	1.075364	-1.730	0.08369	.	[75,80)	0.192594	0.321067	0.600	0.54860
	Usuario T.Pasajeros	1.302695	0.135363	9.624	< 2e-16	***	[80,85)	-1.010408	0.485650	-2.081	0.03748
Edad	[05,10)	0.434387	0.286843	1.514	0.12993		[85,90)	0.265234	0.578579	0.458	0.64665
	[10,15)	0.555901	0.266625	2.085	0.03707	*	[90,95)	-0.128749	0.943989	-0.136	0.89151
	[15,20)	0.307147	0.247291	1.242	0.21422		[95,100]	1.0746215	119.468.298	0.090	0.92833

Tabla 2
Coeficientes de las variables

Ahora, es importante identificar qué tan bueno es este modelo al momento de realizar predicciones, para ello, se utilizará la curva ROC (véase imagen 8). Adicionalmente, con el punto de inflexión más cercano al 100% de verdaderos positivos, se determinará la probabilidad de estimación del modelo.

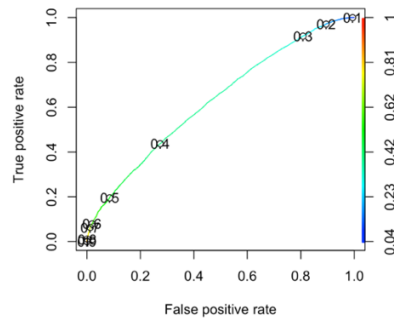


Imagen 8
Curva ROC

En la imagen 8 se observa que la curva ROC es casi una línea recta, esto indica que el modelo no predice lo suficiente como para establecer un ratio cercano al 100%, por el contrario, el modelo indica que tiene un acierto seguro del 40% en verdaderos positivos y alrededor del 30% en falsos positivos. Sin embargo, con esto, el modelo es capaz de acertar en el **61,4%** de las veces que los accidentes son sufridos por mujeres de acuerdo con las tres variables incluidas. Este cálculo se realizó mediante el uso de las predicciones y su comparación con el ratio de falsos y verdaderos positivos.

DISCUSIÓN Y CONCLUSIÓN

El modelo de regresión logística utilizado en este estudio compara la probabilidad de ocurrencia de un accidente con respecto a una categoría de referencia para cada una de las variables independientes. Si bien el modelo actual predice el 61% de los accidentes femeninos, este resultado evidencia la necesidad de mejorar su precisión. En trabajos futuros, se explorará la implementación de otros modelos de aprendizaje automático, como los árboles de decisión, para determinar si estos presentan un mejor rendimiento en la predicción de accidentes femeninos. De modo que se pueda realizar una comparación adecuada y determinar si deben introducirse cambios en la forma que se recolecta la información relacionada con la siniestralidad en el enfoque de género, ya que una implementación efectiva de las políticas es necesaria, la cual normalmente no suele ser eficaz debido a la desigualdad en el tratamiento de género en las campañas (Paricio et al., 2021).

Ahora bien, el análisis de las variables independientes reveló patrones relevantes en la siniestralidad vial femenina. En cuanto al rango horario, se observó un mayor riesgo de accidentes entre las 03:00 y las 05:59 horas, así como entre las 21:00 y las 23:59 horas. Respecto al usuario de la vía, se identificó un mayor riesgo en el caso de vehículos de transporte público y vehículos individuales. Por último, en lo que concierne al rango de edad, se evidenció un mayor riesgo en los rangos [40,45) y [95, 100).

Finalmente, la imprecisión en algunos de los registros de las variables analizadas representó un desafío para el estudio, dificultando el análisis completo de los datos y limitando el potencial de aplicación de los resultados en la creación de políticas públicas efectivas.

REFERENCIAS BIBLIOGRÁFICAS

- Agencia Nacional de Seguridad Vial (ANSV). (2024). Siniestralidad, emisión de licencias y parque automotor.
- Alegret. (2019). Análisis comparativo de la movilidad y la accidentalidad de los PTW en Barcelona en función del género (Tesis de grado, Escola de Camins).
- Cali en cifras. (2023).
- Fiuza & Pérez. (2000). La regresión Logística: una herramienta versátil. *Nefrología*, Vol. 20, No. 6, 477-565.
- Herrera. (2023). Aplicación de regresión logística al análisis de gravedad de lesiones en colisiones de vehículos para ensayo experimentales (Tesis de grado, Universidad Politécnica de Madrid).
- Maté et al. (2019). Introducción a la limpieza y análisis de los datos. FUOC, PID_00265704.
- OMS. (2023). Traumatismos causados por el tránsito.

Paricio et al. (2021). Perspectiva de género en campañas preventivas de seguridad vial y drogas. Profesional de la información, Vol. 30, No. 5, e300517. <https://doi.org/10.3145/epi.2021.sep.17>

Pérez et al. (2019). Metodología para explorar datos abiertos de accidentalidad vial usando Ciencia de Datos: Caso Medellín. Ingeniare. Revista chilena de ingeniería, 27(3), 495-509. <https://dx.doi.org/10.4067/S0718-33052019000300495>

Redondo. (2023). Modelo predictivo de accidentes de tránsito (Tesis de grado, Universidad de Antioquia).