

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

CLASIFICACIÓN Y CARACTERIZACIÓN DE INTERSECCIONES
DE RUTAS NACIONALES CON BASE EN EL USO DE DISPOSITIVOS DE
SEGURIDAD PASIVA EN OCUPANTES DE VEHÍCULOS
AUTOMOTORES PARA EL 2018

PROBLEMAS DE ESTIMACIÓN RELACIONADOS CON EL USO DE
DISPOSITIVOS DE SEGURIDAD PASIVA EN COSTA RICA A PARTIR DE
UN ESTUDIO OBSERVACIONAL

Trabajo final de investigación aplicada sometido a la consideración de la Comisión del
Programa de Estudios de Posgrado en Estadística para optar al grado y título de Maestría
Profesional en Estadística

JOSÉ JESÚS GRANADOS SALAZAR

Ciudad Universitaria Rodrigo Facio, Costa Rica

2022

Agradecimiento

A mi familia. A los profesores y amigos de la Escuela y Posgrado en Estadística Johnny Madrigal, Deiby Solano y Agustín Gómez, por su tiempo y excelente asesoramiento en el desarrollo de este trabajo.

“Este trabajo final de investigación aplicada I fue aceptado por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Profesional en Estadística.”

M.Sc. Johnny Madrigal Pana

Profesor Guía

M.Sc. Agustín Gómez Meléndez

Lector

ME. Deiby Solano Cambronerero

Lector

José Jesús Granados Salazar

Sustentante

“Este trabajo final de investigación aplicada II fue aceptado por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Profesional en Estadística”.

M.Sc. Johnny Madrigal Pana

Profesor Guía

M.Sc. Agustín Gómez Meléndez

Lector

ME. Deiby Solano Cambronerero

Lector

José Jesús Granados Salazar

Sustentante

Contenido

| | |
|--|------|
| Agradecimiento..... | ii |
| Hoja de aprobación Práctica Profesional I..... | iii |
| Hoja de aprobación Práctica Profesional II..... | iv |
| Contenido..... | v |
| Resumen Práctica Profesional I..... | viii |
| Resumen Práctica Profesional II..... | ix |
| 1. Práctica Profesional I..... | 1 |
| 1.1. Introducción..... | 1 |
| 1.2. Estado del conocimiento..... | 5 |
| 1.2.1. Aspectos relacionados con la seguridad vial y el uso de dispositivos de seguridad pasiva..... | 5 |
| 1.2.1.1. Naturaleza del accidente de tránsito..... | 5 |
| 1.2.1.2. Necesidad del uso de dispositivos de seguridad pasiva..... | 8 |
| 1.2.1.3. Antecedentes en la medición del uso de dispositivos de seguridad pasiva..... | 10 |
| 1.3. Metodología..... | 13 |
| 1.3.1. Objetivos..... | 14 |
| 1.3.2. Datos..... | 14 |
| 1.3.2.1. Población de estudio..... | 14 |
| 1.3.2.2. Marco muestral de intersecciones..... | 14 |
| 1.3.2.3. Diseño muestral..... | 15 |
| 1.3.2.4. Tamaño, distribución y selección de los puntos de muestreo..... | 16 |
| 1.3.2.5. Ponderación de intersecciones..... | 17 |
| 1.3.2.6. Trabajo de campo..... | 18 |
| 1.3.2.7. Variables de interés..... | 18 |
| 1.3.3. Técnicas estadísticas..... | 20 |
| 1.3.3.1. Análisis de conglomerados..... | 20 |
| 1.3.4. Análisis empleado..... | 27 |
| 1.4. Resultados..... | 28 |
| 1.4.1. Análisis de las variables numéricas y categóricas..... | 28 |
| 1.4.2. Agrupamiento de intersecciones..... | 30 |
| 1.4.2.1. Escogencia de indicadores y número adecuado de conglomerados..... | 30 |
| 1.4.2.2. Análisis de los grupos conformados para determinar su validez..... | 36 |
| 1.4.2.2.1. Análisis mediante variables categóricas..... | 36 |
| 1.4.2.2.2. Análisis mediante variables numéricas..... | 38 |

| | | |
|------------|--|----|
| 1.4.3. | Regresión logística multinomial | 41 |
| 1.4.3.1. | Aspectos relacionados con el ajuste del modelo | 41 |
| 1.4.3.2. | Características relevantes en el uso de dispositivos de seguridad pasiva | 43 |
| 1.4.4. | Árboles de decisión | 45 |
| 1.5. | Conclusiones | 46 |
| | Bibliografía | 49 |
| | Anexo | 51 |
| | Anexo A: Análisis de diferencias en los promedios de los grupos de uso de dispositivos de seguridad pasiva | 51 |
| | Anexo B: Bondad de ajuste y diagnósticos del modelo | 52 |
| | Anexo C Esquemas de intersecciones | 57 |
| 2. | Práctica Profesional II | 65 |
| 2.1. | Introducción | 65 |
| 2.2. | Aspectos conceptuales relacionados con el muestreo de poblaciones móviles | 67 |
| 2.3. | Metodología | 73 |
| 2.3.1. | Objetivos | 74 |
| 2.3.2. | Datos | 74 |
| 2.3.2.1. | Estudio de UPM con bajo flujo vehicular | 74 |
| 2.3.2.1.1. | UPM con bajo flujo de motocicletas | 75 |
| 2.3.2.1.2. | UPM con bajo flujo de carros | 75 |
| 2.3.2.2. | Población de estudio | 77 |
| 2.3.2.3. | Marco muestral | 77 |
| 2.3.2.4. | Diseño muestral | 78 |
| 2.3.2.5. | Trabajo de campo | 78 |
| 2.3.2.6. | Variables de interés | 79 |
| 2.3.3. | Técnicas estadísticas | 79 |
| 2.3.3.1. | Técnicas de selección | 79 |
| 2.3.3.2. | Indicadores estadísticos | 80 |
| 2.3.2.1.1. | Indicadores de desempeño en la estimación del uso de dispositivos de seguridad pasiva | 82 |
| 2.3.3.3. | Escenarios de los procesos de simulación | 83 |
| 2.3.3.3.1. | Especificación de las fracciones de muestreo en el diseño | 87 |
| 2.3.3.3.2. | Análisis comparativo de las fracciones de muestreo | 87 |
| 2.3.3.4. | Análisis de conglomerados | 89 |
| 2.4. | Resultados | 92 |

| | | |
|------------|--|-----|
| 2.4.1. | Variables e indicadores de desempeño utilizados en el análisis | 92 |
| 2.4.1.1. | Variables utilizadas en el análisis | 92 |
| 2.4.1.2. | Indicadores de desempeño utilizados en el análisis | 93 |
| 2.4.2. | Escenarios de los procesos de simulación..... | 94 |
| 2.4.2.1. | Comparación de fracciones de muestreo..... | 94 |
| 2.4.2.2. | Escenario general sin agrupación de UPM | 97 |
| 2.4.2.2.1. | Submuestreo y comparación de técnicas en la selección de vehículos | 97 |
| 2.4.2.2.2. | Rendimiento de las técnicas de selección | 99 |
| 2.4.2.3. | Posestratificación de UPM a partir del flujo vehicular | 101 |
| 2.4.2.3.1. | Escogencia del número adecuado de conglomerados | 101 |
| 2.4.2.3.2. | Submuestreo y comparación de técnicas de selección en UPM de medio y alto flujo vehicular..... | 103 |
| 2.4.2.3.3. | Rendimiento de las técnicas de selección | 106 |
| 2.4.2.4. | Efecto de los procesos de simulación en la exactitud de las estimaciones..... | 107 |
| 2.5. | Conclusiones y recomendaciones | 111 |
| | Bibliografía | 116 |
| | Anexos | 118 |

Resumen Práctica Profesional I

El objetivo de esta investigación es clasificar intersecciones entre rutas nacionales con base en el uso de dispositivos de seguridad pasiva en el año 2018 para definir perfiles que generen información de interés en el planteamiento de estrategias de intervención. Para realizarlo se utiliza como fuente de datos, el “*Estudio nacional sobre uso de dispositivos de seguridad pasiva, vestimenta y factores distractores, 2019*” realizado por el Consejo de Seguridad Vial (Cosevi) en el 2018. Previo a la clasificación de las intersecciones, se realiza una valoración de los indicadores disponibles para determinar el o los más apropiados en la formación de grupos. Posteriormente, mediante análisis de conglomerados, utilizando la técnica “K-Medoids”, se identifican diferentes escenarios para encontrar el número adecuado de grupos. Los grupos generados se validan internamente para determinar diferencias significativas, según las variables propuestas para la caracterización de las intersecciones, y una vez definidos los grupos, se procede a caracterizarlos mediante regresión logística multinomial.

Entre los resultados más importantes se encuentra que, de los tres indicadores evaluados para formar los grupos, el uso de los dos indicadores individuales (uso de casco y uso de cinturón) brinda mayor información que el compuesto (suma de indicadores individuales). Además, a partir de la razón de correlación “eta²” (η^2), análisis gráfico y criterio sustantivo, se determina que el número de agrupaciones se estabiliza en tres, denominadas “Bajo uso”, “Uso medio” y “Alto uso”. A partir de la validación interna se determinan diferencias significativas entre los grupos y las características de la intersección. La creación del perfil, usando regresión logística multinomial, sugiere que las variables relevantes corresponden a: “Zona”, “GAM”, “Flujo vehicular” y “Cantidad media de accidentes”, de las cuales, “GAM” y “Flujo vehicular” resultaron significativas al 5%, utilizando como referencia el grupo de “Bajo uso”. Lo anterior implica que, en ambas agrupaciones, las intersecciones que se encuentran dentro de la GAM tienen mayor posibilidad de ser de uso medio o alto uso de dispositivos de seguridad y que, además, aumentos en el flujo vehicular implican mayor posibilidad de pertenecer a dichos grupos de uso de dispositivos de seguridad.

A partir del análisis realizado con la clasificación y caracterización de las intersecciones, se logran identificar los focos de menor uso de dispositivos de seguridad pasiva, generando información relevante para orientar a las entidades responsables en materia de seguridad vial, creando políticas y estrategias de intervención destinadas a mejorar la situación en sitios con bajo uso de dispositivos de seguridad, mediante la aplicación de campañas de comunicación, operativos de tránsito y demás medidas que garanticen una mejoría en el porcentaje de uso de casco y cinturón de seguridad.

Resumen Práctica Profesional II

A partir del *Estudio nacional sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores, 2017*, realizado por el Consejo de Seguridad Vial en el 2016, se analizó el efecto que tendrían los cambios en el diseño muestral para disminuir la variabilidad de las estimaciones relacionadas con el uso de dispositivos de seguridad pasiva, mediante la aplicación de técnicas de selección de elementos, indicadores de desempeño y procesos de simulación aplicados al muestreo de conglomerados completos. En el trabajo, mediante procesos de simulación, se compararon técnicas de selección con variaciones en el tamaño del submuestreo y cantidad de réplicas, utilizando dos escenarios; el primero corresponde al uso de las unidades primarias de muestreo (UPM) o conglomerados completos sin agrupar (escenario general), mientras que, en el segundo, se agruparon con base en la intensidad de su flujo vehicular (posestratificación), obteniendo como resultado dos agrupaciones (medio y alto flujo vehicular).

Entre los resultados más importantes se encuentra que, en ambos escenarios, la selección aleatoria evidenció mejor rendimiento en la estimación de los indicadores sobre uso de dispositivos de seguridad pasiva (uso de cinturón de seguridad) con respecto a las técnicas sistemática y los primeros “n” casos, mostrando mayor desempeño en el 76% de los ensayos realizados en el escenario general, mientras que, en el escenario de la posestratificación, para el grupo de flujo medio y alto flujo se obtuvo un desempeño de 47% y 60% respectivamente. Asimismo, en ambos escenarios, en términos de reducción de la variabilidad y precisión, se encontró estabilidad en las estimaciones para valores del submuestreo cercanos a 20 elementos por UPM.

Con respecto a la exactitud de las estimaciones, para el caso del escenario general se encontró que la estimación de la proporción media a partir de los procesos de simulación se encuentra subestimada en cerca de 2,4% (0,76) en relación con el valor de referencia de los conglomerados completos (0,78), mostrando diferencias significativas ($\alpha = 0.05$) al realizar el cálculo de los intervalos de confianza para la proporción media. Sin embargo, dicha situación se valoró en el escenario de la posestratificación para determinar si un control de las UPM con base en el flujo vehicular contribuye a mejorar el problema de la subestimación en función de un diseño de submuestreo de tamaño fijo como el utilizado en el presente trabajo, o si, en la práctica, un diseño de submuestreo proporcional sería más apropiado.

Como parte del análisis realizado, se obtuvieron resultados de importancia práctica para el estudio del uso de dispositivos de seguridad pasiva en poblaciones móviles, ya que se logró una mejor comprensión del fenómeno abordado y pueden recomendarse cambios al diseño metodológico del estudio, en búsqueda de mejoras en el control de la variabilidad y cálculo de estimaciones.

Lista de cuadros

| | |
|---|-----|
| Cuadro 1. Resumen de las variables numéricas utilizadas en el análisis | 29 |
| Cuadro 2. Distribución porcentual de intersecciones según variables categóricas utilizadas en el análisis | 30 |
| Cuadro 3. Resultados en la conformación de grupos a partir de los indicadores propuestos | 32 |
| Cuadro 4. Distribución de intersecciones por uso de dispositivos de seguridad pasiva según características de la intersección, 2018..... | 37 |
| Cuadro 5. Correlaciones para las variables numéricas incluidas en el análisis..... | 39 |
| Cuadro 6. Comparaciones múltiples de Wilcoxon para los promedios de las covariables según grupos de uso de dispositivos de seguridad pasiva, 2018..... | 41 |
| Cuadro 7. Resultados del ajuste del modelo de regresión logística multinomial, 2018..... | 44 |
| Cuadro 8. Distribución absoluta y porcentual de UPM según flujo de motocicletas..... | 75 |
| Cuadro 9. Distribución absoluta y porcentual de UPM según flujo de carros | 77 |
| Cuadro 10. Resumen de las variables utilizadas en el análisis | 93 |
| Cuadro 11. Indicadores de desempeño definidos a partir del uso de cinturón de seguridad en los conglomerados completos..... | 94 |
| Cuadro 12. Comparación de fracciones de muestreo para estimar indicadores de desempeño según variaciones en el tamaño del submuestreo..... | 95 |
| Cuadro 13. Comparación de fracciones de muestreo según indicador de desempeño..... | 96 |
| Cuadro 14. Comparación de técnicas de selección a partir de diferencias entre indicadores de desempeño según tamaño del submuestreo | 100 |
| Cuadro 15. Consistencia en la estimación de indicadores de desempeño según técnica de selección | 101 |
| Cuadro 16. Resultados de la conformación de grupos a partir del flujo de vehículos | 102 |
| Cuadro 17. Comparación de técnicas de selección a partir de diferencias entre indicadores de desempeño por grupo de flujo vehicular según tamaños del submuestreo | 106 |
| Cuadro 18. Consistencia en la estimación de indicadores de desempeño por grupo de flujo vehicular según técnica de selección | 107 |
| Cuadro 19. Comparación de indicadores de desempeño por grupo de flujo vehicular según técnica de selección de elementos..... | 107 |
| Cuadro 20. Intervalos de confianza para la proporción media de uso de dispositivos de seguridad pasiva según escenario, grupo y técnica de selección | 108 |

Lista de tablas

| | |
|---|----|
| Tabla 1. Matriz de Haddon | 8 |
| Tabla 2. Fórmulas de cálculo de los indicadores de desempeño utilizados en los procesos de simulación para muestra y los conglomerados completos..... | 83 |
| Tabla 3. Aproximación del total de vehículos en la población de estudio..... | 88 |

Lista de figuras

| | |
|---|----|
| Figura 1. Estructura de los escenarios para la aplicación de simulaciones en la estimación de indicadores sobre uso de dispositivos de seguridad pasiva | 85 |
| Figura 2 Estructura del escenario “E1” para la aplicación de simulaciones en la estimación de indicadores de desempeño según técnica de selección..... | 86 |

Lista de gráficos

| | |
|---|-----|
| Gráfico 1 Costa Rica: Tasa de mortalidad por 100 mil habitantes en accidentes de tránsito. Periodo 2007 - 2017 | 2 |
| Gráfico 2 Costa Rica: Cantidad de vehículos inscritos por año según tipo. Periodo 2010 - 2017..... | 3 |
| Gráfico 3. Valor medio del objeto según agrupaciones generadas | 33 |
| Gráfico 4. Razón de correlación (η^2) para los indicadores sobre uso de dispositivos de seguridad pasiva y los grupos conformados | 34 |
| Gráfico 5. Distribución de intersecciones por uso de dispositivos de seguridad pasiva según características de la intersección, 2018..... | 38 |
| Gráfico 6. Promedios de las variables numéricas utilizadas en el análisis según grupos de uso de dispositivos de seguridad pasiva, 2018..... | 40 |
| Gráfico 7. Estimación de indicadores de desempeño para el uso de cinturón de seguridad con diferentes técnicas de selección y variaciones del tamaño del submuestreo | 99 |
| Gráfico 8. Inercia intraclase y valor medio del objeto según número de conglomerados | 103 |
| Gráfico 9. Estimación de indicadores de desempeño para el uso de cinturón de seguridad en el grupo de UPM de alto flujo vehicular con diferentes técnicas de selección y variaciones del tamaño del submuestreo | 104 |
| Gráfico 10. Estimación de indicadores de desempeño para el uso de cinturón de seguridad en el grupo de UPM de flujo vehicular medio con diferentes técnicas de selección y variaciones del tamaño del submuestreo | 105 |

Lista de mapas

| | |
|--|----|
| Mapa 1. Costa Rica: Distribución espacial de intersecciones entre rutas nacionales | 15 |
| Mapa 2. Costa Rica: Distribución espacial de intersecciones entre rutas nacionales seleccionadas para la medición del uso de dispositivos de seguridad pasiva..... | 16 |
| Mapa 3. Costa Rica: Distribución de intersecciones entre rutas nacionales seleccionadas según uso de dispositivos de seguridad pasiva, 2018..... | 36 |



UNIVERSIDAD DE
COSTA RICA

SEP Sistema de
Estudios de Posgrado

Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, Jose Jesús Granados Salazar, con cédula de identidad 6-0376-0809, en mi condición de autor del TFG titulado Clasificación y caracterización de intersecciones de rutas nacionales con base en el uso de dispositivos de seguridad pasiva en ocupantes de vehículos automotores para el 2018 y Problemas de estimación relacionados con el uso de dispositivos de seguridad pasiva en Costa Rica

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI NO *

*En caso de la negativa favor indicar el tiempo de restricción: año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

1. Práctica Profesional I

1.1. Introducción

Los traumatismos causados por accidentes de tránsito representan un problema social, económico y de salud pública, por lo que se requieren acciones preventivas, sostenibles y eficientes que aminoren sus efectos. De las situaciones a las que se enfrentan las personas diariamente, y que incrementan la ocurrencia de hechos fatales, los accidentes de tránsito son uno de los causantes de mayor complejidad y peligrosidad (OMS, 2004). Además, se traducen en la principal causa de muerte en jóvenes de 15 a 29 años, seguido del suicidio, VIH/sida y los homicidios (OMS, 2015), convirtiéndose en una de las principales causas de muerte a nivel mundial (FIA Foundation, 2009).

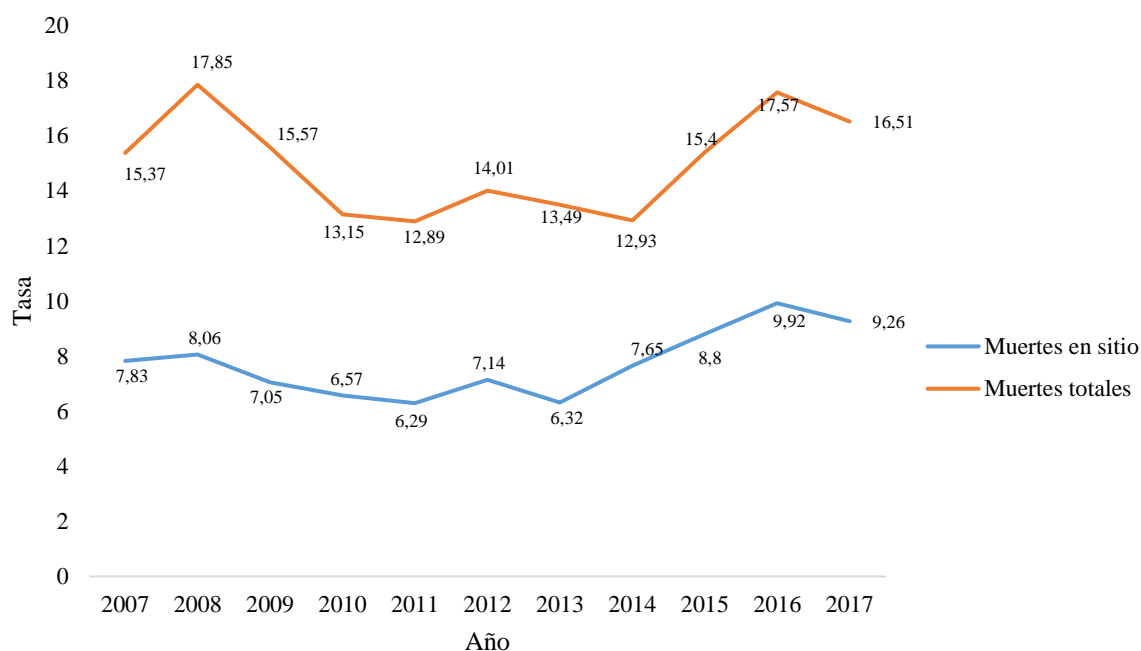
Cada año, alrededor del mundo, se estima que los accidentes de tránsito cobran la vida de aproximadamente 1.25 millones de personas y que 50 millones resultan lesionados. Esta situación es diferencial según región e ingreso de los países, como se menciona en el *“Informe sobre la situación mundial de la seguridad vial 2015”*. Desde el año 2010 se registra un aumento en el número de accidentes de tránsito, principalmente en países de ingresos bajos o medios (84%) y, según el informe, 79 países registran un descenso absoluto en la cantidad de hechos fatales, donde 56% ocurre en países de ingresos bajos o medios. Sin embargo, las tasas de mortalidad en los países de ingresos bajos representan más del doble en comparación con los de alto ingreso; por ejemplo, la tasa mundial es 17.4 muertes por cada 100 000 habitantes y mientras en Europa es de 9.3, en África llega a 26.6. La región de Las Américas presenta una tasa de 15.9 muertes por cada 100 000 habitantes, cercana a la cifra mundial (OMS, 2015).

Un dato relevante en torno a las muertes por siniestros viales, es su desproporcionada ocurrencia en relación con el nivel de motorización, dado que 90% de estas muertes suceden en países de ingresos bajos y medios, a pesar de su menor concentración de vehículos (54% del parque vehicular mundial). En el informe sobre *“La seguridad vial en la región de las Américas 2016”*, se evidencia un incremento en la tasa de motorización de la región, potenciada por el desarrollo económico que ha experimentado un grupo importante de países. Para el año 2010 se registraba una tasa de 460.08 vehículos por cada 1000 habitantes, la cual ascendió a 502.5 para el 2013. Las tasas más elevadas se presentan en países de ingresos altos, como Estados Unidos y Canadá, mientras que las menores cifras se reflejan, por ejemplo, en Guyana y Santa Lucía (OPS, 2016).

Este comportamiento, representado por un incremento en la motorización de la región y la situación socioeconómica, desempeña un factor clave en la mortalidad por accidentes de tránsito, debido a la afectación sufrida por los países en peores condiciones socioeconómicas, donde se da una mayor cantidad de muertes en usuarios vulnerables (peatones, ciclistas y motociclistas), a excepción de América del Norte, donde prevalece notablemente la defunciones en conductores de automóviles.

En el caso específico de Costa Rica, la tasa de mortalidad por siniestros viales en el 2016 y 2017, rondaba valores cercanos al promedio regional (gráfico 1) y mantiene un comportamiento creciente desde el año 2014. De igual forma, la tasa correspondiente a muertes en sitio¹ presenta una tendencia creciente desde el 2013.

Gráfico 1. Costa Rica: Tasa de mortalidad por 100 mil habitantes en accidentes de tránsito. Periodo 2007 - 2017



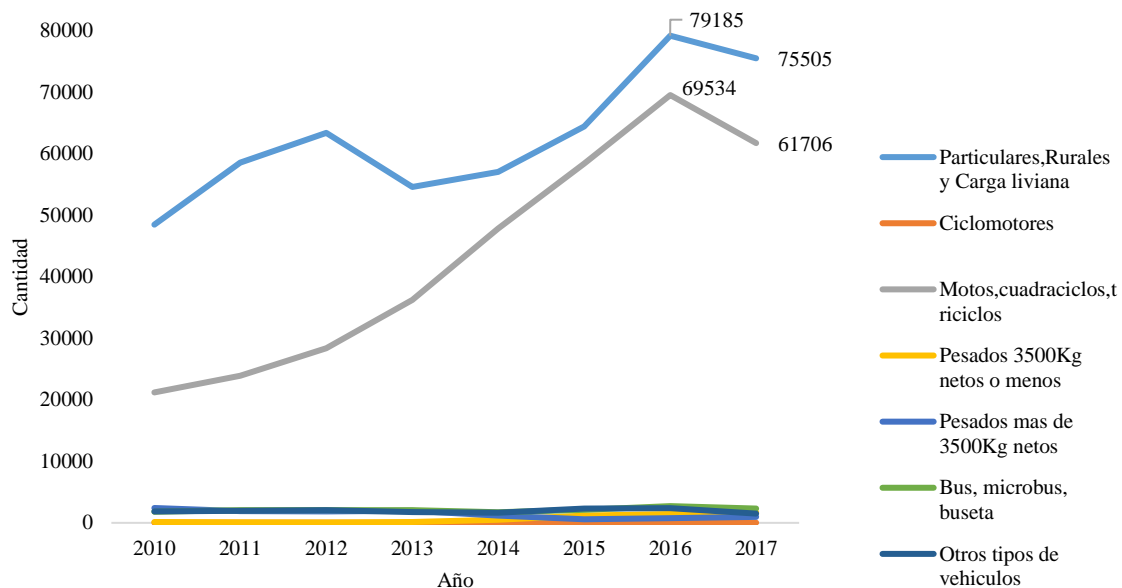
Fuente: Cosevi. Área de investigación y estadística. Muertes en sitio antes 2012 y el provisionales 2016 suministrados por Dirección General de la Policía de Tránsito, 2012-2015 del parte oficial de tránsito. Número de muertes totales del Poder Judicial. Población del CCP & INEC. Costa Rica: Estimaciones y proyecciones de población por sexo y edad 1950-2100. San José, Costa Rica: publicaciones del Instituto Nacional de Estadística y Censos (INEC). Esta información es de proyecciones existentes en las bases de datos del CCP.

Como ocurre en otras regiones de ingresos bajos y medios, en el caso de Costa Rica, la motorización presenta un incremento considerable y sostenido en los últimos años (gráfico 2). Es importante resaltar el incremento de las motocicletas como parte del parque vehicular, las cuales, como se

¹ Entiéndase muerte en sitio, como la persona que fallece al momento del accidente o durante el traslado al hospital.

mencionó anteriormente, tienen un aporte considerable en la proporción de muertes de usuarios vulnerables de la vía, en este caso, conductor y acompañante de motocicleta.

Gráfico 2. Costa Rica: Cantidad de vehículos inscritos por año según tipo. Periodo 2010 - 2017



Fuente: Valores estimados a partir de la base de datos de bienes y muebles de la Dirección de Informática, Registro Nacional, Costa Rica.

Las personas que sufren traumatismos a raíz de los siniestros viales se encuentran representados por diferentes tipos de usuarios, condiciones de vulnerabilidad, grupos etarios, etc. Se sabe que el uso de sistemas de seguridad pasiva (casco y cinturón de seguridad) contribuye a la disminución de las muertes y lesiones por causa de los siniestros viales. Como lo menciona la OPS (2016, pag. 16), en la región de América, para el caso de los motociclistas, existe una serie de factores de riesgo como el simple hecho de compartir su espacio con otros vehículos automotores, su escasa visibilidad y su exposición física al entorno vial, los cuales potencian su probabilidad de morir por causa de traumatismos craneoencefálicos. El simple hábito de utilizar casco reduce 40% el riesgo de morir y aproximadamente 70% los traumatismos graves.

Para el caso de los vehículos, el uso de cinturones reduce el movimiento interno de los ocupantes ante una eventual colisión, disminuyendo de esta forma la posibilidad de traumatismos graves o mortales. El uso del cinturón de seguridad reduce entre 45% y 50% el riesgo de traumatismo mortales para el conductor y ocupante delantero, mientras que en el caso de los ocupantes traseros, la reducción es del 25% (OPS, 2016, pag. 18). La ausencia de cinturón de seguridad en ocupantes de vehículos al momento de un siniestro vial, representa una proporción muy grande de víctimas

mortales en la vía (OMS, 2013), convirtiéndose en un factor de riesgo para los ocupantes de vehículos.

El éxito en la reducción de la gravedad de los traumatismos por accidentes de tránsito depende en gran medida de la rigurosidad con que se apliquen las leyes y el respeto a las mismas, pero esto es un aspecto que varía significativamente entre regiones o de un país a otro. Por ejemplo, en muchos países de ingresos bajos no se exige el uso adecuado de los dispositivos de seguridad, lo que es acompañado por aspectos culturales que surgen como barrera a su uso y la aplicación diferenciada de la ley según ocupante o región dentro de un país. Estos, entre otros factores, influyen de manera relevante en los índices de uso (FIA Foundation, 2009 y OPS, 2016).

En Costa Rica, la Dirección de Proyectos del Consejo de Seguridad Vial, como parte del Plan Estratégico Nacional, en el marco del decenio de acción para la seguridad Vial 2011 – 2020 (COSEVI, 2011), se conforma a partir de cinco pilares básicos: gestión de la seguridad vial, vías de tránsito y movilidad más seguras, vehículos más seguros, usuarios de las vías de tránsito más seguros y respuesta tras los accidentes. Estos pilares consideran una serie de acciones orientadas al mejoramiento de la seguridad en las vías de tránsito en el país, lo cual se traduce en la necesidad de contar con programas para sensibilizar y promover mejoras en el comportamiento de las personas, en la conducción en carretera, promoción de leyes y normas en conjunto con la educación, concientización de la ciudadanía en torno a temas relacionados con la seguridad vial y factores de riesgo en las vías de tránsito.

Entre las diferentes acciones desarrolladas se encuentra el “Estudio del uso de dispositivos de seguridad pasiva y distractores”, el que se encuentra enmarcado en el cuarto pilar (usuarios de las vías de tránsito más seguros) del plan estratégico nacional. El estudio de fenómenos de este tipo requiere sistematizar y monitorear constantemente su comportamiento como factor de riesgo, por lo que es importante realizar investigaciones de esta índole por parte del Consejo de Seguridad Vial, debido a la necesidad de contar con información oportuna y de calidad sobre el uso de dispositivos de seguridad y aspectos relacionados, como parte de las metas establecidas en el plan estratégico.

Sobre la ejecución de este estudio recae el interés del presente documento, ya que el diseño metodológico a ser utilizado en la medición de uso de dispositivos de seguridad pasiva, requiere de un marco y un diseño muestral adecuado, que garantice, en la medida de lo posible, mediciones oportunas y de calidad, permitiendo de esta forma, contar con información pertinente que pueda evidenciar la situación respecto al fenómeno analizado.

1.2. Estado del conocimiento

1.2.1. Aspectos relacionados con la seguridad vial y el uso de dispositivos de seguridad pasiva

1.2.1.1. Naturaleza del accidente de tránsito

Los accidentes de tránsito son causantes de millones de fallecidos y heridos graves a raíz de los traumatismos sufridos en carretera, lo que ha impulsado numerosas iniciativas de diversa índole a nivel mundial, orientadas a la intervención y reducción de las muertes y lesiones sufridas por causa de siniestros viales. Por tal motivo, es importante definir “accidente de tránsito” antes de ahondar en elementos afines a su prevención y mitigación.

Un accidente de tránsito se puede definir como el evento ocurrido en un camino o calle abierta al tráfico público, dando como resultado al menos una persona fallecida o herida y donde al menos un vehículo en movimiento estaba involucrado. Estos accidentes por lo tanto incluyen colisiones entre vehículos, entre vehículos y peatones, y entre vehículos y animales u obstáculos fijos. (OECD, 2018)

En torno a este fenómeno, como menciona Elvik (2009, pag. 87), han surgido interrogantes como las planteadas en relación a las posibles causas de los accidentes. ¿Existen las mismas? Si este fuera el caso, se puede dar sentido al término “causa del accidente”. Se trata de inquietudes surgidas en Noruega en la década de 1960, época en la cual se creía ampliamente que no era posible reducir los accidentes de tránsito efectivamente si no se conocían las “causas reales de los mismos”.

En relación con lo anterior, el argumento planteado consistía en la imposibilidad de implementar una planificación de la seguridad vial en forma exitosa, debido a la escasa disponibilidad de conocimiento sobre las causas reales de los accidentes y, como resultado, desconocimiento de las mejores soluciones.

Respecto al tema de la causalidad de los accidentes de tránsito, autores como Haight (citado Elvik, R. 2009), rechazan utilizar el concepto causa para explicar el fenómeno de los accidentes, pues argumentan que los accidentes son el resultado de un proceso aleatorio complejo cuyas características generales se pueden modelar estadísticamente, además, entre los factores que influyen en el proceso estocástico que implica un accidente, algunos son conocidos, mientras otros no. La necesidad de conocer las causas de un determinado evento para dar respuesta a sus posibles implicaciones, se traduce en un argumento interesante, pero necesariamente no existe una relación o dependencia entre las causas del evento y su posible solución. Por tal motivo, con base en lo mencionado, Elvik (2009,

pag. 88) comenta que, para tener un mejor entendimiento de lo argumentado es útil considerar algunos enfoques teóricos desarrollados en los últimos cien años en torno a los accidentes de tránsito. Estos se listan a continuación en forma cronológica:

- 1) Accidentes como eventos aleatorios.
- 2) Teoría de la propensión de accidentes.
- 3) Teoría de la causalidad del accidente.
- 4) Teoría de sistemas (y teoría epidemiológica).
- 5) Teoría del comportamiento.

A partir de los enfoques mencionados, el autor concluye sobre los elementos relevantes que todas las teorías aportan a la explicación de los accidentes. En cierta medida, todos los accidentes se comportan de forma aleatoria o existen casos de personas que resultan más propensas a sufrir siniestros en relación a otras. Es cierto también que los seres humanos son susceptibles a cometer errores y que esos errores pueden resultar en accidentes, así como puede alcanzarse una reducción en la cantidad de accidentes al lograr adaptar los sistemas técnicos a las limitaciones humanas. No obstante, la necesidad humana de someterse constantemente a situaciones de riesgo no puede ser mitigada en su totalidad por ningún enfoque o sistema propuesto.

En los diferentes escenarios propuestos, ningún abordaje teórico proporciona un acercamiento completo respecto a la ocurrencia de accidentes, ya que cada punto de vista presenta un enfoque parcial al fenómeno analizado, donde algún aspecto relevante en un nivel puede no serlo en otro. Hasta ahora no se ha logrado establecer una teoría general que contribuya no solo a la explicación de la problemática vial sino también que proporcione elementos de prevención.

Por otro lado, en relación con la naturaleza causal de los accidentes, al no contar con una idea sobre la frecuencia con la cual acontecen ciertos riesgos en el tránsito, es difícil obtener conclusiones contundentes sobre la importancia relativa de los factores que influyen en dichos accidentes. Por lo tanto, es necesario conocer la distribución y frecuencia de la exposición de los usuarios a los factores de riesgo para revelar la relación entre varios factores de riesgo y los accidentes, lo mismo que ocurre en el caso de los factores de riesgo asociados con la carretera o con los vehículos.

A partir de lo anterior se desprende un elemento de interés en la seguridad vial relacionado con la vulnerabilidad y el error humano, el cual es abordado dentro del enfoque sistémico o teoría de sistemas. Con base en esto, la OMS (2004, pag. 5) afirma que la atribución exclusiva otorgada tradicionalmente a los usuarios de la vía pública, en relación con la responsabilidad en la ocurrencia de accidentes, a pesar de la contribución que tienen otros factores como el mal diseño de las

carreteras, no tiene control. Asimismo, en situaciones de tránsito complejo, el comportamiento humano se torna impredecible, por lo que no es de esperar que se puedan prevenir los accidentes en su totalidad.

En el marco de la teoría de sistemas y bajo un nuevo enfoque, William Haddon describió el transporte por carretera como:

...Un sistema “hombre-máquina” mal concebido que debía ser objeto de un tratamiento sistémico integral. (OMS, 2004).

Sobre esto se cimentó lo que actualmente se conoce como la matriz de Haddon, explicando la interacción resultante en las tres fases de un choque entre los factores ser humano, vehículo y entorno (tabla 1), logrando obtener importantes avances en el estudio del comportamiento humano y su relación con el entorno y vehículo, como causantes de traumatismos debido a los siniestros viales.

En relación con lo anterior, Glizer (1993, pag. 41) comenta sobre la interrelación existente entre tres componentes que conforman la triada epidemiológica: un huésped, con un agente en un medio ambiente, cuya interrelación implica la presencia de enfermedades o lesiones. Con base en esto, el mismo autor Glizer (1993, pag. 141) menciona la importancia de analizar la ocurrencia de accidentes de tránsito como consecuencia de la interacción entre el huésped (ser humano), agente (vehículo) y el medio, donde dicha ocurrencia se encuentra relacionada con los factores de riesgo pertenecientes a cada componente, buscando con esto una mejor comprensión de la interrelación existente.

El autor evidencia la importancia que tiene el componente humano como mayor aportador en la materialización de los accidentes y la necesidad de intervenir sobre este, con el fin de prevenir la incidencia de eventos de este tipo. Respecto a esto, Feldeintein (citado en Solano, 2014, pag. 10), argumenta que el conocimiento con el que pueda contar una persona sobre elementos esenciales de conducción y legislación en torno al tránsito, no garantiza un desempeño excepcional en carretera, debido a la existencia de otros factores externos que influyen directamente en el individuo.

Tabla 1. Matriz de Haddon

| Fase | | Factores | | |
|----------------------------|---|---|---|---|
| | | Ser humano | Vehículos y equipo | Entorno |
| Previo al choque | Prevención de choques | <i>Información</i> | <i>Buen estado técnico</i> | <i>Diseño y trazado de la vía pública</i> |
| | | <i>Actitudes</i> | <i>Luces</i> | <i>Limitación de la velocidad</i> |
| | | <i>Discapacidad</i> | <i>Frenos</i> | <i>Vías peatonales</i> |
| | | <i>Aplicación de la reglamentación por la policía</i> | <i>Maniobrabilidad</i> | |
| | | | <i>Control de la velocidad</i> | |
| Durante el choque | Prevención de traumatismos durante el choque | <i>Utilización de dispositivos de retención</i> | <i>Dispositivos de retención de los ocupantes</i> | <i>Objetos protectores contra choques</i> |
| | | <i>Discapacidad</i> | <i>Otros dispositivos de seguridad</i> | |
| | | | <i>Diseño protector contra accidentes</i> | |
| Posterior al choque | Conservación de la vida | <i>Primeros auxilios</i> | <i>Facilidad de acceso</i> | <i>Servicios de socorro</i> |
| | | <i>Acceso a atención médica</i> | <i>Riesgo de incendio</i> | <i>Congestión</i> |

Fuente: Elaborado con base en la matriz de Haddon, Informe mundial sobre traumatismos causados por el tránsito (Resumen 2004, pag. 6).

1.2.1.2. Necesidad del uso de dispositivos de seguridad pasiva

En vista de la creciente problemática en torno a los traumatismos causados por accidentes de tránsito a nivel mundial, la OMS (2004) estableció en un acto histórico, el día mundial de la salud a la seguridad vial, evento que tuvo incidencia en más de 130 países donde se realizaron diferentes actividades con el objetivo de concienciar a la población en temas relacionados con los accidentes de tránsito, además de reforzar las iniciativas existentes y crear nuevos programas para la seguridad vial.

Ese mismo año, la OMS y el Banco Mundial, publicaron el “*El informe mundial sobre la prevención de traumatismo causados por el tránsito*”, donde se analizan en detalle aspectos relacionados con la prevención, impacto y causas de las heridas de tránsito, además de factores de riesgo de accidentes y la propuesta de estrategias de intervención para mejorar la seguridad vial. Como resultado, el informe concluye con seis medidas orientadas a la mejora de la situación vial. De estas seis medidas, la quinta corresponde a la “*aplicación de medidas concretas para prevenir accidentes en carretera, minimizar las heridas y consecuencias, y evaluar las repercusiones de estas medidas*”. De esta

medida se desprenden una serie de estrategias orientadas a combatir algunos de los factores causantes de traumatismos viales de mayor riesgo. Dichas estrategias, como se menciona en FIA Foundation (2009), corresponden a:

- Establecer leyes que exijan el uso de cinturones de seguridad y sistemas de retención infantil para todos los ocupantes de vehículos motorizados.
- Exigir a los motociclistas el uso de casco.
- Establecer y controlar los límites de concentración de alcohol en sangre.
- Establecer y controlar los límites de velocidad.
- Gestionar la infraestructura de carreteras existente de manera que se incremente la seguridad.
- Mejorar la seguridad de los vehículos.

En relación con lo anterior, el uso de casco y cinturones de seguridad como medidas para la reducción de la severidad de traumatismos viales, se encuentran enmarcadas dentro de lo que se conoce como dispositivos de seguridad pasiva, los cuales pueden definirse como sistemas que proporcionan mayor protección a los pasajeros relacionada con la seguridad estructural a los choques (FIT, 2017, pag. 111).

Como parte de lo mencionado, el uso de dispositivos de seguridad pasiva contribuye a reducir los efectos de algunos factores causantes de traumatismos viales, entre ellos, el exceso de velocidad como uno de los principales factores de riesgo de lesiones y sumado a ello, la ingesta de alcohol, conformación de la vía, flujo vehicular y condiciones meteorológicas que afectan el desempeño en carretera del componente humano (OMS, 2017, pag. 15).

Actualmente, a nivel mundial, las normas de seguridad en los vehículos juegan un papel relevante debido al incremento del parque vehicular y la demanda creciente de los usuarios, sin embargo, las normas de seguridad varían considerablemente en rigurosidad entre los diferentes países y regiones. Esto significa que, en los países de ingresos altos durante los últimos años, la demanda creciente impulsa la fabricación de automóviles cada vez más seguros, igual que sucede en los países de ingresos bajos y medios, debido a su motorización en incremento y su condición de vulnerabilidad en relación con la ocurrencia de accidentes con traumatismos graves. Esto está potenciando la necesidad de extender dicha normativa a nivel mundial, (OMS, 2017, pag. 28).

Sin embargo, la desventaja de los países de ingresos bajos y medios radica en que la seguridad de los vehículos está en función de la normativa existente en cada país, por lo que en lugares de escasa o nula normativa, los fabricantes pueden introducir vehículos con diseños obsoletos, los cuales no cumplen con la medidas mínimas establecidas en los países con buena reglamentación. En relación

con lo anterior, el Foro Mundial de las Naciones Unidas para la Armonización de la Reglamentación sobre vehículos, constituye el principal organismo internacionalmente capacitado para el desarrollo de normas de seguridad, que pueden ser acogidas voluntariamente por los países miembros de la ONU, de los cuales solamente 40 de los 193 aplican normas técnicas sobre seguridad (OMS y FIT, 2017).

La falta de uso de los cinturones de seguridad es un factor de riesgo que ocasiona lesiones o la muerte de ocupantes de vehículos, pues cuando el conductor o pasajero no lo llevan abrochado, en la mayoría de los casos tienen desenlaces mortales por causa de siniestros viales. Esto se debe a que el ocupante se desplaza a la misma velocidad del vehículo previo al choque, provocando una colisión contra la estructura del vehículo o la expulsión del ocupante del vehículo (FIA Foundation, 2009).

El uso de cinturones de seguridad es una de las principales medidas para reducir la mortalidad y los traumatismos por causa de accidentes de tránsito, incrementando las posibilidades de supervivencia tras un choque vehicular. En el año 1968, en su artículo 7, el Convenio de Viena sobre Circulación por Carreteras, establece:

“El uso del cinturón de seguridad es obligatorio para los conductores y pasajeros de los automóviles que ocupen asientos dotados de esos cinturones, salvo en los casos en que la legislación nacional se establezcan excepciones” (CEPE, 2007).

1.2.1.3. Antecedentes en la medición del uso de dispositivos de seguridad pasiva

En relación con la medición del uso de dispositivos de seguridad pasiva existen diversos estudios. En Estados Unidos, en la Universidad de Florida del Sur, Lee, Pino y Shultz (2014), realizaron un estudio observacional en el que analizaron ventajas y desventajas de métodos de recolección de información sobre el uso de casco en motociclistas. Igualmente, Lin, Lee y Saxena (2010), del Departamento de Transporte de la Universidad de Florida, desarrollaron un estudio para evaluar el uso de casco, donde fueron seleccionados 15 condados con base en el procedimiento de muestreo de NHTSA², para realizar observaciones de una hora, ejecutando 7305 observaciones.

En el estado de Louisiana, Elliot y Schneider (2013) realizaron un estudio observacional para medir el uso del cinturón de seguridad delantero y trasero, además del casco en motociclistas. Se

² **National Highway Traffic Safety Administration (NHTSA)**, corresponde a una dependencia del gobierno de los Estados Unidos, la cual forma parte del Departamento de Transporte y tiene como misión, la prevención y reducción de accidentes de tránsito.

seleccionaron mediante muestreo aleatorio simple 336 segmentos de carretera en función de los lugares con mayor incidencia de accidentes con heridos graves. Se trabajó con franjas horarias asignadas al azar durante el día, de 7:00 a.m. a 6:00 p.m. todos los días de la semana. Además del uso de cinturón y casco, se observó el tipo de vehículo, sexo y raza de los ocupantes. Por último, para el control de la calidad, se realizó la respectiva capacitación y posterior supervisión mediante la selección aleatoria de 19 tramos de carretera para verificar la debida ubicación y horario de los observadores.

Otro estudio sobre el uso del casco en motociclistas fue desarrollado en la Universidad de Wayne, Michigan, a cargo de Savolainen, Gates, Morden y Hacker (2013). En este se seleccionaron 40 condados con los porcentajes más altos de accidentes de tránsito con heridos, posteriormente las locaciones de observación fueron seleccionadas de estos 40 condados. Entre las variables recopiladas se encuentran el uso de casco, edad, sexo, etnia, tipo de motocicleta y vestimenta de alta visibilidad. En dicho estudio se realizaron auditorías sobre el procedimiento de recolección para garantizar su debido cumplimiento, a partir de controles aleatorios en el 5% de los puntos de muestreo.

En el Departamento de Transporte de la Universidad de Florida, Turner y Hagelin (2004), llevaron a cabo un estudio para medir el uso de casco en motociclistas, además de otros aspectos de interés a partir de fuentes externas. En el estudio se utilizó un diseño muestral estratificado multietapico, aprobado por la NHTSA, para determinar el número de sitios, días y horas requeridos en la recolección de los datos. Además de aspectos generales y el uso de casco, se midió el sexo, tipo de motocicleta y equipo de seguridad (chaquetas, pantalones, zapatos, guantes y protección para los ojos).

Continuando con estudios realizados en la región, en el caso Uruguayo, la Unidad Nacional de Seguridad Vial (UNASEV, 2013), desarrolló un estudio observacional para medir el uso de elementos de seguridad y factores de distracción en conductores. La metodología para esta investigación fue un acuerdo entre diferentes integrantes del Observatorio Iberoamericano de Seguridad Vial (OISEVI), con la finalidad de permitir comparabilidad de resultados en los países miembros.

La recolección de los datos se realizó de forma observacional y en automóviles se obtuvo información sobre el uso de cinturones de seguridad, sistemas de retención infantil, uso de luces y uso del celular. En el caso de las motocicletas, se midió el uso de casco, vestimenta de alta visibilidad, luces y uso del teléfono celular. Los puntos de observación se seleccionaron tomando en cuenta solamente zonas urbanas, obteniendo 45 puntos para las mediciones. La ubicación de dichos puntos se seleccionó

garantizando una separación suficiente para obtener mayor cobertura del casco urbano, las calles debían ser “representativas” del tránsito usual de la ciudad y no debían existir controles de tránsito cerca que pudieran afectar el comportamiento de los ocupantes del vehículo.

En Bolivia, Puppó, Centeno, Pablo y Serulnicoff (2013), como parte de un acuerdo entre los diferentes coordinadores nacionales de datos de los países miembros del OISEVI, desarrollaron un estudio observacional a nivel nacional para medir ciertas conductas viales de los conductores y ocupantes de vehículos automotores relacionadas con el uso de casco, cinturón y sistemas de retención infantil. En este estudio se obtuvo como resultado una línea basal comparable entre los países de la región miembros del OISEVI. La población de estudio se definió como los ocupantes de autos, camionetas, motos y ciclomotores en localidades con más de 5000 habitantes. El tipo de muestreo es establecido como de relevamiento observacional, probabilístico e intencionado, el cual es estratificado por ciudades en función de la cantidad de habitantes, obteniendo una muestra final de 5.503 motos y ciclomotores y 7141 unidades para el caso de autos y camionetas.

En el ámbito costarricense, los acercamientos relacionados con la medición del uso de dispositivos de seguridad pasiva son escasos, ya que su desarrollo ha estado a cargo básicamente de dos entidades: el Observatorio del Desarrollo (OdD) de la Universidad de Costa Rica y el Consejo de Seguridad Vial (COSEVI) como entidad adscrita al Ministerio de Obras Públicas y Transporte (MOPT). En el caso de la Universidad de Costa Rica, el Observatorio del Desarrollo OdD (2016) ejecutó un estudio exploratorio tipo observacional para identificar posibles riesgos en motociclistas durante la conducción en carretera. Dicho estudio consistió en una prueba a partir de 1004 observaciones, en las cuales se midió el uso de casco, presencia de factores distractores y equipo de protección. La población se definió como los motociclistas que transitaban por tres cantones del Valle Central (Curridabat, Desamparados y Alajuela) durante el 24 y 30 de julio del 2016.

Por otra parte, el Consejo de Seguridad Vial ha realizado estudios para medir el uso de dispositivos de seguridad pasiva como parte del análisis de los factores de riesgo durante la conducción, realizados en el 2014 y 2016 a nivel país en las principales rutas nacionales. Como parte del marco muestral, se definieron segmentos de un kilómetro de carretera donde se aplicaron las mediciones de lunes a viernes en franjas de cinco minutos, distribuidas cada media hora, entre 6:00 a.m. y 6:00 p.m. Para este estudio, se midió el uso de casco, cinturón de seguridad, factores distractores y características de la flota circulante. Debido a problemas en la obtención de las mediciones, producto de las limitaciones del marco muestral utilizado, para el estudio realizado en el 2018 se decidió realizar

cambios en la metodología y se construyó un nuevo marco muestral basado en intersecciones entre rutas nacionales³.

En relación con el fenómeno estudiado, las diferentes investigaciones identificadas se enfocan únicamente en la medición y estimación de aspectos relacionados con el uso de dispositivos de seguridad pasiva. Más allá de estos acercamientos, no se encontraron investigaciones dirigidas particularmente a la caracterización del entorno vial con base en el uso de dispositivos de seguridad y, en menor medida, de análisis enfocados al caso particular de intersecciones de carretera. Es importante mencionar que gran parte de los estudios desarrollados por las respectivas entidades son ejecutados con el objetivo de obtener información básica que evidencia el panorama general y que pueda ser consumida por las correspondientes instancias para el establecimiento de políticas públicas de carácter esencial.

1.3. Metodología

Este apartado contiene los aspectos metodológicos relacionados con la realización de un estudio nacional tipo observacional para medir el uso de dispositivos de seguridad pasiva: casco, cinturón, vestimenta y factores distractores. Esta información será utilizada en el documento para clasificar y caracterizar las intersecciones entre rutas nacionales. Entre los aspectos que se detallan se encuentra el abordaje metodológico empleado, descripción de los datos utilizados y características de la muestra (diseño de muestreo utilizado en la selección, puntos de muestreo seleccionados, posicionamiento en los puntos de muestreo, procedimiento de selección de los vehículos, la distribución de los puntos y tiempos de observación). De igual forma se incluye el referente teórico de las técnicas estadísticas utilizadas (análisis de conglomerados y regresión logística multinomial).

Es importante aclarar que los datos recolectados en este estudio fueron obtenidos con base en otros objetivos y no con los definidos en el presente documento de práctica profesional.

³ Para un mayor detalle sobre las limitaciones presentadas por los estudios anteriores y las mejoras realizadas, refiérase al informe sobre el “Estudio nacional para la medición del uso de dispositivos de seguridad pasiva, vestimenta y factores distractores” realizado el Consejo de Seguridad Vial (Cosevi).

1.3.1. Objetivos

Objetivo general

Clasificar intersecciones entre rutas nacionales en Costa Rica con base en el uso de dispositivos de seguridad pasiva en el año 2018 para tener insumos que contribuyan con la definición de perfiles que generen información de interés en el planteamiento de estrategias de intervención.

Objetivos específicos

- Identificar los indicadores adecuados del uso de dispositivos de seguridad pasiva para conformar agrupaciones de intersecciones entre rutas nacionales.
- Construir agrupaciones de intersecciones a partir de los indicadores del uso de dispositivos de seguridad pasiva con el uso del análisis de conglomerados.
- Verificar la consistencia de los grupos construidos a partir de técnicas estadísticas multivariadas.
- Identificar variables asociadas a las agrupaciones construidas para caracterizar las intersecciones entre las vías nacionales.

1.3.2. Datos

1.3.2.1. Población de estudio

La población que se investiga en el estudio del Área de Investigación y Estadística sobre el uso de dispositivos de seguridad pasiva, se encuentra definida por los ocupantes de vehículos automotores que circularon en las intersecciones entre rutas nacionales a lo largo del territorio nacional, en periodos de una hora, distribuidos entre 6:00 a.m. y 6:00 p.m. En este estudio se excluyó los casos de vehículos de transporte colectivo (transporte público, transporte de estudiantes, etc.) y vehículos de carga pesada (tráiler). Además, vehículos con vidrios polarizados que causaban mediciones deficientes no fueron tomados en cuenta.

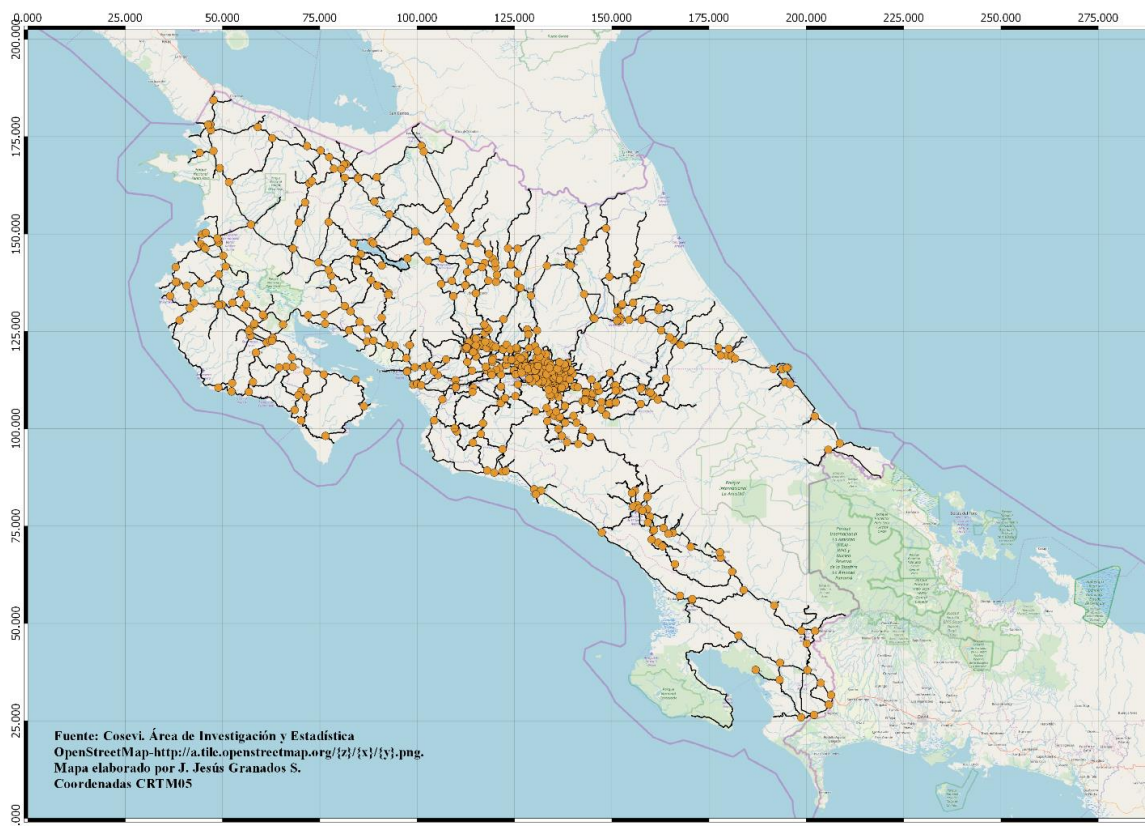
Para este estudio interesa construir una clasificación de las intersecciones entre rutas nacionales, por lo que la unidad de análisis estadístico está definida por cada intersección.

1.3.2.2. Marco muestral de intersecciones

Costa Rica cuenta con 7 759,5 kilómetros de ruta nacional y 35 085,8 kilómetros de ruta cantonal, lo cual representa 18.1% y 81.9% respectivamente de la red vial total (COSEVI, 2018). El marco

muestral está definido por las intersecciones de carretera entre rutas nacionales que abarcan la totalidad del territorio de Costa Rica, contabilizando un total de 600 puntos de muestreo. Se excluyeron los puntos formados a partir de intersecciones tipo rotonda, por dificultades de medición y porque no garantizan una reducción suficiente de la velocidad del flujo vehicular que brinde el tiempo suficiente para anotar adecuadamente las variables a medir. El siguiente mapa muestra la distribución de las intersecciones que conforman el marco muestral.

Mapa 1. Costa Rica: Distribución espacial de intersecciones entre rutas nacionales



1.3.2.3. Diseño muestral

Como se menciona en el informe sobre el “Estudio nacional para medición del uso de dispositivos de seguridad pasiva, vestimenta y factores distractores” (Cosevi, 2019):

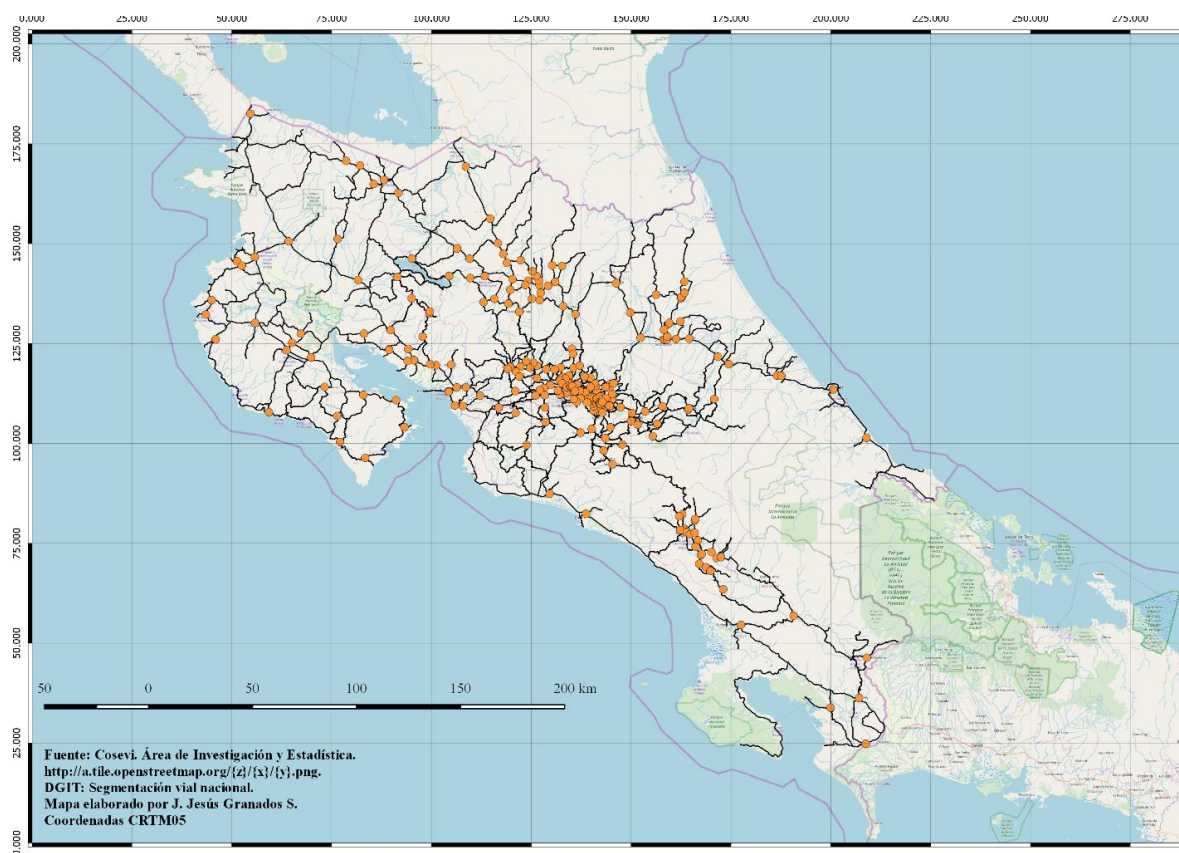
“El diseño muestral se puede definir como un diseño probabilístico de poblaciones móviles en el cual se seleccionan intersecciones entre rutas nacionales mediante muestreo sistemático y, posteriormente, en cada intersección seleccionada, para la totalidad de franjas horarias establecidas durante los siete días de la semana, se procede a realizar las mediciones sobre el uso de dispositivos de seguridad pasiva por parte de los ocupantes de vehículos automotores”.

1.3.2.4. Tamaño, distribución y selección de los puntos de muestreo

Para el estudio realizado por el Cosevi, se efectuaron estimaciones a nivel nacional y se tomaron en cuenta seis cantones de riesgo (San José, Pérez Zeledón, Alajuela, San Carlos, Puntarenas y Pococí). Estos seis cantones quedaron autorepresentados y se realizaron mediciones en la totalidad de sus intersecciones, dadas sus características en términos geográficos y ocurrencia de accidentes con personas lesionadas y fallecidas. Con base en lo anterior, el tamaño final del conjunto de puntos de muestreo, está conformado por las intersecciones de estos seis cantones (161 intersecciones) y los casos seleccionados en los 75 cantones restantes (126 intersecciones), para un total de 287 puntos de observación.

En el siguiente mapa se puede observar la distribución espacial de las intersecciones entre rutas nacionales, tanto de la muestra seleccionada como de los cantones de riesgo mencionados anteriormente.

Mapa 2. Costa Rica: Distribución espacial de intersecciones entre rutas nacionales seleccionadas para la medición del uso de dispositivos de seguridad pasiva



1.3.2.5. Ponderación de intersecciones

Como se detalla anteriormente, las intersecciones correspondientes a los seis cantones de riesgo están auto representadas, mientras que, en los 75 cantones restantes, se seleccionó una muestra aleatoria, por lo cual, el procedimiento de ponderación difiere en ambos subconjuntos. Para el caso de los cantones de riesgo, al seleccionarse la totalidad de las intersecciones, la probabilidad queda definida de la siguiente forma:

$$f = \frac{N}{N_{CR}} = \frac{161}{161} = 1$$

Donde:

N: Cantidad de intersecciones seleccionadas en los seis cantones de riesgo.

N_{CR}: Total de intersecciones en los seis cantones de riesgo.

Por lo tanto, la probabilidad de selección es “1”, implicando una ponderación definida de la siguiente forma:

$$F = \frac{1}{f} = \frac{1}{1} = 1$$

Para el caso de la muestra, la probabilidad de selección está definida de la siguiente forma:

$$f = \frac{n_{RC}}{N_{RC}} = \frac{126}{439} = 0.29$$

Donde:

n_{RC}: Cantidad de intersecciones seleccionadas en el resto de cantones.

N_{RC}: Total de intersecciones en el resto de cantones.

Por lo tanto, la probabilidad de selección es “0.29”, implicando una ponderación definida de la siguiente forma:

$$F = \frac{1}{f} = \frac{1}{0.29} = 3.48$$

1.3.2.6. Trabajo de campo

El desarrollo del trabajo de campo requirió lineamientos relacionados con aspectos como el debido posicionamiento de los observadores en los puntos de muestreo para la obtención de mediciones de calidad, un adecuado procedimiento de selección de las motocicletas y automóviles al circular por las diferentes intersecciones y la distribución de los puntos y tiempos de observación en las diferentes intersecciones de cada cantón.

En relación con los tiempos de observación, las mediciones de los 287 puntos se realizaron durante un mes de forma balanceada, cada uno de los siete días de la semana, dividiendo el día en seis franjas de una hora, de 06:00 a.m. a 06:00 p.m., para un total de 6 horas de observación por día, donde cada franja se dividió en periodos de media hora que fueron destinados a la observación de un determinado tipo de vehículo (automóvil o motocicleta)⁴, contabilizando al final del día, 3 horas de observación para motocicletas y 3 horas para automóviles.

1.3.2.7. Variables de interés

Para clasificar y agrupar las intersecciones se utilizan datos obtenidos a partir del “Estudio observacional sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores” realizado el Consejo de Seguridad Vial (Cosevi) durante el primer semestre del 2018. Como su nombre lo indica, el interés del estudio radica en la recopilación de datos relacionados con uso del casco en motociclistas, equipo de protección, cinturones de seguridad en ocupantes de vehículos y factores distractores durante la conducción, con el objetivo de contar con información adecuada que pueda evidenciar la situación respecto al fenómeno estudiado y mantener un monitoreo constante.

Además de las variables que proporciona el estudio, también se utilizan datos sobre la cantidad de accidentes en cada intersección a partir del archivo de datos de accidentes de tránsito (Periodo 2012 – 2016) a cargo del Área de Investigación y Estadística del Cosevi, e información sobre tipología de las intersecciones y señalamiento vial obtenido del estudio sobre la caracterización de intersecciones entre rutas nacionales realizado durante el 2017 por el Área de Investigación y Estadística del Cosevi. A continuación se detallan las variables utilizadas en la clasificación y caracterización de las intersecciones, tomando en cuenta la información de la encuesta como de los registros:

⁴ Los aspectos mencionados sobre características de la muestra (población de estudio, marco muestral, diseño muestral, trabajo de campo, tamaño, distribución y selección de la muestra) se pueden consultar de forma más detallada en el informe sobre el “Estudio nacional sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores” desarrollado por el Área de Investigación y Estadística del Consejo de Seguridad Vial.

Uso de casco

Variable categórica utilizada para medir la presencia ausencia de casco en los ocupantes de motocicleta.

Uso de cinturón

Variable categórica utilizada para medir la presencia o ausencia del cinturón de seguridad en los diferentes ocupantes de los automóviles.

Zona

Variable categórica en la cual se define si la región en la que se encuentra la intersección es urbana o rural. En este caso, el comportamiento que se espera, es de un menor uso de dispositivos de seguridad pasiva en las intersecciones correspondientes a la zona rural.

Gran Área Metropolitana (GAM)

Variable categórica que define si la intersección se encuentra dentro o fuera de la GAM. Para el caso de esta variable, se esperaría un mayor uso de dispositivos en las intersecciones pertenecientes a la GAM.

Flujo de vehículos

Se refiere al conteo de vehículos automotores, tanto automóviles como motocicletas. Los conteos se realizan durante cada franja horaria definida para la medición del uso de dispositivos de seguridad, lo cual implica que, en la media hora que se observan motocicletas, el conteo solo se realiza para este tipo de vehículo. De igual forma este procedimiento aplica para el caso de los automóviles.

Cantidad de accidentes

Se refiere a la cantidad de accidentes (según tipo de vehículo: automóvil o motocicleta) de tránsito con heridos o fallecidos en un rango de 50 metros de la intersección donde se realizan las mediciones. Se espera que las intersecciones con un número elevado de accidentes con víctimas presente un menor uso del dispositivo de seguridad pasiva por parte de los usuarios de la vía.

Señalamiento vial

Variable que identifica la presencia de señalamiento vial en la intersección donde se realizan las mediciones. Dicho señalamiento se clasifica en señalamiento vertical (reglamentación, prevención e información), el cual es un tipo de señalamiento que debe ser colocado al lado derecho de la vía y en casos especiales en ambos lados (SIECA, 2014). Por otra parte, el señalamiento vial horizontal, corresponde al señalamiento demarcado en el pavimento y en algunos casos sirve de complemento a otros dispositivos como el señalamiento vial vertical.

Tipo de intersección:

Se refiere a la clasificación de las intersecciones (el esquema con los tipos de intersección más frecuentes se puede consultar en el Anexo B) según su forma, de acuerdo a lo que establece el manual

de diseño geométrico de carreteras (Manual de diseño geométrico de carreteras, 2008), las intersecciones se pueden clasificar de la siguiente forma:

Intersecciones a nivel sin canalizar

- Intersección en “T” o “Y”
- Intersección en Cruz “+” o Equis “x”

Intersecciones a nivel canalizadas

- Intersección en “T” o “Y”
- Intersección en “T” o “Y” con separador y carril de giro a la izquierda
- Intersección en Cruz “+” o Equis “X”
- Intersección en “Cruz “+” o Equis “x” con separador y carril de giro a la izquierda
- Intersecciones tipo Rotonda

Intersecciones a desnivel

- Intersección a desnivel tipo “Trompeta” en carreteras no divididas
- Intersección a desnivel tipo “Trompeta” en carreteras divididas
- Intersección a desnivel tipo “Trebol” en carreteras no divididas
- Intersección a desnivel tipo “Trebol” en carreteras divididas

Sin embargo, considerando la reducida presencia de intersecciones a nivel canalizadas y a desnivel, y la mayor peligrosidad de las intersecciones tipo “+” o “x” en relación con la ocurrencia de siniestros viales, se decide recodificar la variable en dos categorías: una primer clase que abarca las tipo “+” o “x” y una segunda que resume el resto de intersecciones.

Es importante mencionar que, para las variables flujo vehicular, señalamiento vial y tipo de intersección, no se cuenta con información suficiente para establecer hipótesis formales sobre el uso de dispositivos de seguridad pasiva.

1.3.3. Técnicas estadísticas

El objetivo principal del trabajo consiste clasificar y caracterizar intersecciones entre rutas nacionales con base en la medición del uso de dispositivos de seguridad pasiva y otras variables relacionadas con la intersección. Dicho análisis se desarrollará a partir de técnicas estadísticas multivariadas con la finalidad de contar con agrupamientos que evidencien el comportamiento de los usuarios de la vía en torno al uso de dispositivos de seguridad en los diferentes puntos estudiados. Las técnicas estadísticas utilizadas se explican a continuación.

1.3.3.1. Análisis de conglomerados

El análisis de agrupamientos o de conglomerados puede definirse como un conjunto de técnicas numéricas utilizadas para la formación de grupos a partir de elementos que presentan semejanzas

según algún criterio. Para formar los grupos se parte como base de un conjunto de características correspondientes a los diferentes elementos o casos (Rodríguez, 2013).

La creación de un conglomerado está definido por la cohesión interna (homogeneidad) y aislamiento externo (separación), según comentan Cormack (1971) y Gordon (1999), citados en Everitt (2011). Lo anterior quiere decir que los grupos conformados a partir de la técnica deben ser lo más similares a lo interno de cada grupo y lo más diferente en la medida de lo posible entre los distintos grupos. En la práctica no resulta sencillo identificar una técnica única para la definición de grupos, lo que dependerá de aspectos como la medida de semejanza o desemejanza, la técnica de conglomeración como tal o el número de grupos que se requiera definir. Son aspectos que el investigador debe definir a partir de consideraciones prácticas, razones subjetivas o el contexto de los datos (Rodríguez, 2013).

En general, dentro del análisis de conglomerados, existen diferentes técnicas de agrupamiento que pueden clasificarse como métodos de partición, métodos jerárquicos, métodos basados en cuadrícula y métodos basados en densidad (Preeti, 2016). Para el presente trabajo se utilizan métodos de partición, los cuales consisten en la división del conjunto de datos en “k” grupos, donde el valor “k” requiere ser especificado por el usuario. Posteriormente, se ejecuta el algoritmo produciendo un número “k” de grupos, el cual puede ser ajustado con base en índices de calidad o de forma gráfica (Rousseeuw, 1997).

El algoritmo k-medoids consiste en un método de agrupamiento relacionado con la técnica k-medias, utilizado para la división de un conjunto de datos en “k” conglomerados distintos. Cada grupo resultante de la partición es representado por un par ordenado de puntos en el respectivo grupo de datos. A diferencia del método k-medias, en el cual cada centro del grupo es calculado como el valor medio de todos los puntos dentro de la agrupación definida, en k-medoids, determina un objeto (Medoide), donde la disimilitud media entre el objeto y los demás elementos del grupo sea mínima (Kassambara, 2018).

El algoritmo k-medoids representa un método robusto a la influencia del ruido y presencia de valores atípicos en el análisis estadístico, por lo cual se convierte en una alternativa adecuada al método de k-medias, debido al uso de objetos (medoides) como elemento central en cada agrupación formada en lugar de promedios.

Por lo tanto, como menciona Rousseeuw (1997), el método consiste en la búsqueda de “k” medoides representativos entre el conjunto de datos. Los medoides se calculan de tal manera que la disimilitud total de los objetos a su medoide más cercano sea mínima, para encontrar un subconjunto de datos $\{m_1, \dots, m_k\} \subset \{1, \dots, n\}$ el cual minimiza la función objetivo:

$$\sum_{i=1}^n \min_{t=1, \dots, k} d(i, m_t) \quad (1.1)$$

Esto implica que el objeto es asignado al grupo correspondiente al medoide más cercano o, en otras palabras, el objeto “i” es asignado al grupo “v_i” cuando el medoide “m_{v_i}” se encuentra más cercano a “i” que cualquier otro medoide “m_w” de tal forma que:

$$d(i, m_{v_i}) \leq d(i, m_w) \text{ para todo } w = 1, \dots, k. \quad (1.2)$$

1.3.3.2. Árboles de decisión

La teoría de árboles de decisión introducida por Breiman (1984) y desarrollada bajo el nombre de Árboles de Regresión y Clasificación (CART por sus siglas en inglés), consiste en clasificadores estructurados de árbol binario, los cuales se construyen a partir de divisiones repetidas de subconjuntos de “X” en dos subconjuntos descendientes, iniciando con “X”, donde “X” es el conjunto de datos inicial. Los subconjuntos que no son divididos en más subconjuntos menores son denominados subconjuntos terminales o nodos terminales, mientras que los subconjuntos que son divididos se definen como subconjuntos no terminales o nodos no terminales. En general, la construcción del árbol está definida por la selección de las divisiones, la decisión de definir un nodo como terminal o continuar con las divisiones y la asignación de cada nodo terminal a una clase.

En otras palabras, como menciona Serna (2009), el método CART es una técnica no paramétrica de segmentación binaria donde la construcción del árbol está definida por particiones repetidas del conjunto de datos en dos grupos mutuamente excluyentes. Durante el proceso de partición, cada división se selecciona de tal forma que la impureza de cada nodo sea menor que la del nodo anterior. La finalidad del procedimiento consiste en obtener grupos homogéneos a partir de la partición de la variable respuesta tratando de mantener el árbol con un tamaño razonable.

Como menciona Tomofeev (citado en Serna, 2009), la construcción y análisis de los árboles de clasificación y regresión, se puede resumir en tres pasos principales; a) *Construcción del árbol máximo*, b) *Poda del árbol* y c) *Selección del árbol óptimo mediante validación cruzada*.

Construcción del árbol máximo

La construcción del árbol máximo se realiza utilizando el método de partición binario ya mencionado, iniciando con la raíz del árbol o lo que es lo mismo, el conjunto de datos originales. La primera versión del árbol es una versión sobre ajustada del mismo, que contiene gran cantidad de nodos y niveles, los cuales no generan la mejor clasificación, por lo que la calidad del nodo debe ser

analizada mediante una medida de impureza. Entre las principales medidas de impureza o criterios de particionamiento mencionados por Breiman (1984), se encuentra:

- *El índice de información o entropía*, el cual se define de la siguiente forma:

$$i(t) = \sum_j p(j|t) \ln[p(j|t)] \quad (1.3)$$

A partir de lo anterior, se busca la partición que maximice $\Delta i(t)$ con base en la siguiente ecuación:

$$\Delta i(t) = - \sum_{j=1}^k p(j|t) \ln[p(j|t)] \quad (1.4)$$

Donde:

$j = 1, \dots, k$ representa la cantidad de clases de la variable respuesta categórica.

$p(j|t)$ corresponde a la probabilidad de clasificación correcta para la clase “j” en el nodo “t”.

$i(t)$ corresponde a la función de impureza

- *El índice de Gini*: Tiende a separar la categoría más grande en un grupo aislado, a diferencia del índice de información mencionado anteriormente, el cual forma agrupaciones con más de una categoría en las primeras particiones. El índice de Gini se define de la siguiente forma:

$$i(t) = \sum_{i \neq j} p(j|t) p(i|t) \quad (1.5)$$

A partir de lo anterior, se busca la partición que maximice $\Delta i(t)$ con base en la siguiente ecuación:

$$\Delta i(t) = - \sum_{j=1}^k [p_j(t)]^2 \quad (1.6)$$

- *El índice de Towing*: El objetivo del índice consiste en formar dos clases que representen más del 50% de los datos, formando dos categorías grandes en cada partición. Este índice genera árboles con mejor balanceo, sin embargo, su desempeño es más lento en comparación con el índice de Gini. El índice de Towing se define identificando la partición que maximice la siguiente ecuación:

$$\frac{p_L p_R}{4} \left[\sum_j |p(j|t_L) - p(j|t_R)| \right]^2 \quad (1.7)$$

Donde:

t_L y t_R representan los nodos hijos izquierdo y derecho.

p_L y p_R representan la proporción de observaciones en “t” que pasaron a t_L y t_R en cada caso.

Poda del árbol

Posterior a la construcción del árbol, este por general presenta sobreajuste, por lo que es necesario encontrar el tamaño adecuado por medio de cortes sucesivos de los nodos terminales. Para este cometido, Breiman (1984) introduce un método basado en una función de costo-complejidad “ $R_\alpha(T)$ ” por medio de la cual se comparan diferentes árboles para determinar el mejor. Para cada árbol, la función se define de la siguiente forma:

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \quad (1.7)$$

Donde:

$R(T)$ es el promedio de la suma de cuadrados entre los nodos.

$|\tilde{T}|$ es la complejidad del árbol.

α es el parámetro de complejidad.

Selección del árbol óptimo

Según De´ath y Fabricius (citado en Serna, 2009), a partir de los diferentes árboles generados es necesario realizar la selección del árbol óptimo mediante el uso de validación cruzada, con la que se estima el error de predicción. La finalidad del procedimiento consiste en identificar la proporción óptima entre la tasa de clasificación errónea y la complejidad del árbol. El procedimiento de validación cruzada puede llevarse a cabo de dos formas:

- Dividir el archivo de datos en dos grupos (entrenamiento y prueba), con un primer grupo (entrenamiento) se construyen los diferentes árboles y para cada árbol se realiza la predicción de la respuesta de los datos de entrenamiento. Luego se procede a calcular el error de predicción y se elige el árbol con el menor error de predicción. Es importante considerar que esta alternativa de validación es recomendable si se cuenta con suficientes observaciones en el archivo de datos.
- En el caso de no contar con un archivo de datos de tamaño suficiente puede recurrirse al uso de validación cruzada, la cual consiste en seleccionar una muestra de prueba del archivo de datos (archivo de aprendizaje o entrenamiento), posteriormente, del archivo de entrenamiento se calculan los estimadores, los cuales son evaluados con los datos de la muestra de prueba mediante el error de predicción, a partir del cual se obtiene el error medio absoluto.

1.3.3.3. Regresión logística multinomial

En el presente trabajo se usa la regresión logística multinomial como parte de las técnicas propuestas para el cumplimiento de los objetivos establecidos. Al tratarse de una generalización del modelo logístico binario, es importante realizar un breve repaso de dicho modelo, previo al desarrollo de su versión politómica.

1.3.3.3.1. Caso dicotómico

Los modelos de regresión logística, y en general el análisis de regresión, son utilizados fundamentalmente con tres propósitos. El primero es un acercamiento descriptivo de las variables estudiadas. En segundo lugar, la técnica proporciona la posibilidad de cuantificar la magnitud de la relación existente entre las covariables y variable dependiente y, finalmente, como último punto, mediante la relación estadística modelada, es posible realizar predicciones de un valor o valores de interés (Neter, 2004).

En las ciencias sociales, gran parte de las variables analizadas son de carácter dicotómico, por lo que se requiere un abordaje especial usando un análisis logístico como método óptimo para el estudio de variables respuesta binarias (Allison, 2003). La regresión logística binaria, como se mencionó anteriormente, es un modelo en el cual se desea conocer la relación existente entre una variable dependiente cualitativa binaria y una o más variables independientes, las cuales pueden ser cualitativas o cuantitativas.

Por lo tanto, para una variable respuesta “Y” y una variable predictora “X”, se tiene que $\pi(x) = P(Y=1|X=x) = 1 - P(Y=0|X=x)$. Con esto, el modelo de regresión logística está definido por la siguiente expresión (Agresti, 2002).

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1.8)$$

El logaritmo del odds, llamado “logit”, equipara la función de enlace logit con el predictor lineal y presenta la siguiente relación lineal:

$$\text{logit}[\pi(x)] = \log \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x \quad (1.9)$$

La transformación anterior es importante porque $\text{logit}[\pi(x)]$ cuenta con muchas de las propiedades de un modelo de regresión lineal; es lineal en sus parámetros, puede ser continuo y puede variar de $-\infty$ a $+\infty$ dependiendo de los valores de “x” (Hosmer, 2013).

El signo de “ β ” determina si $\pi(x)$ aumenta o disminuye en función de incrementos en “ x ”. Al exponenciar ambos lados de la ecuación anterior se muestra como los “odds” son una función exponencial de “ x ”, lo cual proporciona una interpretación para la magnitud de “ β ” relacionada con el aumento de los “odds” multiplicativamente por el valor e^β por cada incremento unitario en la covariable “ x ”. Esto implica que e^β es la razón de “odds”, el cociente de el “odds” de $X=x+1$ entre el “odds” de $X=x$ (Agresti, 2002).

Para el ajuste del modelo de regresión logística al conjunto de datos, se requiere estimar los valores β_0 y β_1 . Para llevar a cabo lo anterior, se cuenta con el método de estimación por máxima verosimilitud como una de las técnicas de mayor uso. El método consiste en calcular los valores para los parámetros desconocidos que maximicen la probabilidad de obtener el conjunto de datos observado, a través de una función denominada “función de verosimilitud”, la cual expresa la probabilidad de los datos observados como una función de los parámetros desconocidos. Por lo tanto, los estimadores de máxima verosimilitud de los parámetros son los valores que maximizan dicha función y están más cercanos a los datos observados (Hosmer, 2013).

1.3.3.3.2. Caso politómico

El modelo logístico es generalizable a los casos donde la variable respuesta es nominal, con más de dos niveles. Dicha modificación del caso binario se conoce, entre otras formas, como regresión logística multinomial. La técnica brinda la posibilidad de trabajar con cualquier número de niveles en la variables respuesta, no obstante, tomando en cuenta que en el presente estudio la variable respuesta cuenta con tres categorías, se presenta el desarrollo de las ecuaciones para este escenario.

A partir de lo mencionado en Hosmer (2013), en este caso, a diferencia de la versión dicotómica, el modelo multinomial cuanta con tres logitos, uno por cada categoría de respuesta, por lo que es necesario definir un grupo de referencia. Realizado lo anterior, el modelo asume que se tienen “ p ” covariables y un término constante, denotado por un vector “ \mathbf{x} ” de valores con dimensión “ $p+1$ ” con “ $x_0=1$ ”. Con base en esto, las dos funciones logito se describen a continuación:

$$g_1(\mathbf{x}) = \ln \left[\frac{P(Y = 1|\mathbf{x})}{P(Y = 0|\mathbf{x})} \right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1p}x_p = \mathbf{x}'\beta_1 \quad (1.10)$$

y

$$g_2(\mathbf{x}) = \ln \left[\frac{P(Y = 2|\mathbf{x})}{P(Y = 0|\mathbf{x})} \right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p = \mathbf{x}'\beta_2 \quad (1.11)$$

Lo anterior implica que las probabilidades condicionales para cada categoría a partir del vector de covariables están dadas por:

$$P(Y = 0|\mathbf{x}) = \frac{1}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad (1.12)$$

$$P(Y = 1|\mathbf{x}) = \frac{e^{g_1(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad (1.13)$$

$$P(Y = 2|\mathbf{x}) = \frac{e^{g_2(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad (1.14)$$

1.3.4. Análisis empleado

Como primer paso, previo a la aplicación de las técnicas estadísticas multivariadas, se procederá a realizar una valoración de los indicadores disponibles para determinar el o los más apropiados para la formación de grupos. Posteriormente, el trabajo inicia con la aplicación del análisis de conglomerados mediante el método de K-medoids para identificar diferentes escenarios con el objetivo de encontrar el número adecuado de agrupaciones. Con base en los grupos conformados, se realiza una validación interna para determinar si existen diferencias significativas entre los grupos creados a partir de un conjunto de variables propuestas para la caracterización de las intersecciones. A partir de los grupos generados mediante el análisis de conglomerados, se pretende determinar si los agrupamientos definidos, a partir del uso de los dispositivos de seguridad pasiva, presentan características que permitan la formación de tipologías mediante la aplicación de regresión logística multinomial

1.4. Resultados

En el presente apartado se detallan los principales resultados. Como etapa inicial se estudia el comportamiento de las variables numéricas y categóricas que caracterizan cada intersección. Luego se realiza la selección de indicadores para formar grupos y se define un número adecuado para el análisis. Posteriormente, se realiza una validación interna de los grupos para determinar las diferencias entre los grupos creados a partir de las variables propuestas para caracterizar las intersecciones o, dicho de otra manera, la formación del perfil. Por último, se genera la caracterización de intersecciones a partir de los grupos con base en el uso de dispositivos de seguridad pasiva.

1.4.1. Análisis de las variables numéricas y categóricas

Las variables utilizadas fueron definidas en el apartado metodológico y ahora es necesario profundizar en su interpretación y comportamiento. El cuadro 1 presenta la desagregación del flujo de vehículos (motocicletas y automóviles) y personas, además de las variables relacionadas con la ocurrencia media de accidentes y señales horizontales por intersección.

Cabe recordar que la unidad de análisis del estudio corresponde a la intersección, por lo que todas las observaciones realizadas durante las cuatro semanas, distribuidas durante los siete días en 6 franjas de una hora por día (tres horas para motocicleta y tres horas automóvil), se encuentran resumidas por intersección, implicando que las interpretaciones se encuentran en función de dicha unidad de análisis.

Por ejemplo, para el primer bloque del cuadro, correspondiente a las variables de motocicleta, el valor 80.4 corresponde a la cantidad promedio de motocicletas que circularon por intersección durante tres horas distribuidas en un día de observación. La misma lógica de interpretación aplica para las restantes tres variables:

- *Flujo_moto_obs*: Cantidad de motocicletas observadas del total que circularon por la intersección.
- *Personas_moto*: Cantidad de ocupantes del total de motocicletas observadas.
- *Uso_casco*: Cantidad de ocupantes de motocicleta con presencia de casco del total observado.

En el segundo bloque de variables correspondientes al flujo de automóviles y personas, el análisis es análogo al mencionado para el caso de las motocicletas. En el tercer bloque, las variables se

encuentran definidas por la suma de sus correspondientes en el bloque de motocicletas y automóvil. Por ejemplo, la variable *flujo_total* está definida por la suma del *flujo_moto* y *flujo_carro*, por lo que para este caso, la interpretación se realiza de la misma forma que en los casos anteriores.

Cuadro 1. Resumen de las variables numéricas utilizadas en el análisis

| Variable | Descripción | Estadísticos de resumen | | | | | | |
|-------------------|--|-------------------------|-------|-------|-------|---------|-------|--------|
| | | Min. | P25 | Med. | Prom. | D. Est. | P75 | Max. |
| flujo_moto | Cantidad de motos que circularón | 1.0 | 23.8 | 51.0 | 80.4 | 92.3 | 100.0 | 667.0 |
| flujo_moto_obs | Cantidad de motocicletas observadas | 1.0 | 18.0 | 38.0 | 42.5 | 30.3 | 59.0 | 129.0 |
| personas_moto | Cantidad de personas en motocicleta | 1.0 | 22.0 | 43.0 | 50.2 | 34.8 | 70.3 | 146.0 |
| uso_casco | Personas con uso de casco de seguridad | 1.0 | 19.8 | 41.0 | 47.0 | 34.8 | 67.0 | 139.0 |
| flujo_carro | Cantidad de carros que circularón | 2.0 | 68.0 | 219.0 | 375.7 | 440.3 | 570.0 | 2664.0 |
| flujo_carros_obs | Cantidad de carros observados | 2.0 | 54.0 | 109.0 | 102.6 | 56.1 | 145.8 | 235.0 |
| personas_carro | Cantidad de personas en carro | 2.0 | 85.0 | 151.0 | 145.4 | 76.4 | 202.0 | 355.0 |
| uso_cinturon | Personas con uso de cinturón de seguridad | 0.0 | 47.0 | 98.0 | 101.6 | 67.6 | 155.0 | 301.0 |
| flujo_total | Cantidad de carros y motocicleta que circularón | 2.0 | 95.0 | 288.0 | 455.0 | 514.2 | 658.0 | 2961.0 |
| flujo_total_obs | Cantidad de carros y motocicletas observadas | 4.0 | 78.3 | 155.0 | 146.3 | 80.0 | 206.0 | 358.0 |
| personas_vehiculo | Cantidad de personas en carro y motocicleta | 2.0 | 115.0 | 202.0 | 195.0 | 103.7 | 273.5 | 445.0 |
| uso_dsp | Personas con uso de uno o ambos disp. de seguridad | 0.0 | 69.9 | 146.0 | 148.0 | 93.2 | 214.3 | 387.0 |
| cant_acc | Cantidad promedio de accidentes con lesionados | 0.0 | 0.2 | 0.6 | 1.1 | 1.4 | 1.6 | 11.6 |
| s_horizontal_mean | Cantidad promedio de señalización horizontal | 0.0 | 1.5 | 3.8 | 3.6 | 2.4 | 5.5 | 10.0 |

Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Por último, se cuenta con dos variables correspondientes a la cantidad media de accidentes y la cantidad media de señales horizontales por intersección. La cantidad media de accidentes corresponde al promedio de accidentes con heridos o fallecidos ocurridos en un radio de 50 metros de la intersección para un periodo de 5 años. En este caso, se cuenta con un valor promedio de 1.1 accidentes con heridos o fallecidos.

La señalización horizontal se refiere a la cantidad promedio de elementos presentes en la capa asfáltica correspondientes a señalización horizontal. En este caso, para las intersecciones analizadas, se tiene un valor medio de 3.6 elementos por intersección correspondientes a señalización horizontal.

En el caso de las variables categóricas (cuadro 2), la distribución porcentual de las intersecciones por zona muestra una mayor presencia en la parte urbana (62.0%) que en la rural (38.0%), caso contrario al experimentado en la Gran Área Metropolitana, donde 62.7% aproximadamente de las intersecciones se encuentra fuera. Este comportamiento está influenciado por los seis cantones de riesgo, los cuales como se menciona anteriormente están auto representados, lo que genera un efecto en la distribución porcentual por zona y GAM.

La señalización vertical se encuentra en el 75% de las intersecciones analizadas, evidenciando mayor presencia de intersecciones con al menos un elemento vertical sobre información, reglamentación o prevención en carretera. Por último, en la valoración del tipo de intersección se genera una nueva agrupación donde se compara el tipo “X o +” respecto al resto de intersecciones, con lo que se obtiene que la configuración de mayor peligrosidad “X o +” representa el 25.8% de las intersecciones analizadas.

Cuadro 2. Distribución porcentual de intersecciones según variables categóricas utilizadas en el análisis

| Variable | Descripción | Categoría | Abs. | Rel. |
|-----------------|------------------------------------|----------------------|-------------|--------------|
| zona2 | Zona | <i>Urbano</i> | 372 | 62,0 |
| | | <i>Rural</i> | 228 | 38,0 |
| gam2 | Gran Área Metropolitana | <i>GAM</i> | 224 | 37,3 |
| | | <i>No GAM</i> | 376 | 62,7 |
| s_vertical2 | Presencia de señalización vertical | <i>Sí</i> | 465 | 77,5 |
| | | <i>No</i> | 135 | 22,5 |
| tipo_inter2 | Tipo de intersección | <i>Tipo "X o +"</i> | 155 | 25,8 |
| | | <i>Resto interse</i> | 445 | 74,2 |
| Total | | | 600 | 100,0 |

Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

1.4.2. Agrupamiento de intersecciones

Como se explicó, las mediciones sobre el uso de dispositivos de seguridad pasiva fueron realizadas en las intersecciones entre rutas nacionales, las cuales se encuentran distribuidas en los 81 cantones del país. La selección de la muestra se realizó con las intersecciones pertenecientes a 75 cantones (n=126) y en los 6 restantes se realizó un censo de sus intersecciones (N=161). Debido a esto, para realizar la agrupación, se usa la totalidad de las intersecciones en los 81 cantones.

1.4.2.1. Escogencia de indicadores y número adecuado de conglomerados

Para construir los grupos con análisis de conglomerados, se utilizan como insumos los indicadores del uso de casco en motociclistas y cinturón de seguridad en ocupantes de vehículo, ambos en forma independiente y en términos absolutos. Además, se genera un tercer indicador a partir de la suma de ambos (uso de casco y cinturón) definido como “Uso de dispositivos de seguridad pasiva (Uso DSP)”, para probar diferentes escenarios en la formación de los grupos.

La creación de un indicador compuesto responde a la necesidad de solventar la limitación presentada por el “Uso de casco de seguridad”, pues en tres intersecciones no hubo circulación de motocicletas durante las franjas horarias de medición, lo cual implicaba eliminar tres casos (intersecciones) del archivo de datos. El cómputo del nuevo indicador compuesto garantiza la permanencia de dichas intersecciones al tratarse de la suma de ambos (uso de casco y de cinturón).

Para establecer el número de grupos se realizaron diferentes pruebas y número de agrupaciones con los indicadores, haciendo uso de recursos estadísticos, como la razón de correlación “eta²” (η^2) y evaluaciones gráficas usando el método “Average silhouette”, combinando con una valoración sustantiva del fenómeno estudiado y manteniendo una relación lógica entre la cantidad de grupos, su tamaño y una interpretación aceptable desde el punto de vista teórico.

En las diferentes alternativas, la distribución del número de casos en las diferentes agrupaciones no presenta inconvenientes, ya que los conglomerados resultaron con un balance aceptable, tanto para el caso de los dos indicadores independientes como para el indicador compuesto (cuadro 3). Con base en esta información se escogió la opción de tres grupos, pues permite una mejor interpretación de la información. La opción de dos grupos limita el análisis y el uso de cuatro o más agrupaciones no ofrece información adicional que sea relevante para interpretar el uso de dispositivos de seguridad. Además, para el caso de los indicadores independientes, tampoco se observan resultados que pudieran fortalecer el análisis con cuatro o más grupos.

Por otro lado, como ya se mencionó, aunque el uso del indicador compuesto garantiza la permanencia de las tres intersecciones con ausencia de flujo de motocicletas, se prefiere el uso de los indicadores independientes porque brindan información más específica.

Adicional al abordaje anterior, como complemento para escoger el número de grupos, el método “Average silhouette” brinda la posibilidad de analizar cuáles objetos⁵ se encuentran bien ubicados dentro del grupo y cuáles no, midiendo así la calidad del agrupamiento en función de los elementos en el gráfico que presenten el mayor valor medio. Por lo tanto, el número óptimo de conglomerados “k” es aquel que maximice dicho valor medio a partir del rango de valores de “k”.

En el gráfico 3 se observa que el mayor valor medio, a partir del conjunto de agrupaciones definidas, se obtiene con dos grupos.

⁵ Par ordenado de puntos en el respectivo grupo de datos. El detalle sobre la definición de objeto se puede consultar en el apartado metodológico, pag. 14 y 15.

Cuadro 3. Resultados en la conformación de grupos a partir de los indicadores propuestos

| Agrupación | 1 Indicador (Compuesto) ^{a/} | | | | 2 Indicadores | | | | | | |
|-----------------|---------------------------------------|--------------|--------------------------------|------------------------|---------------|--------------|----------------------------------|------------------------|-------------------------------------|------------------------|-------------|
| | n | % | Uso promedio DSP ^{b/} | % de uso ^{c/} | n | % | Uso promedio casco ^{b/} | % de uso ^{c/} | Uso promedio cinturón ^{b/} | % de uso ^{c/} | |
| General | 287 | 100,0 | 153,3 | 76,9 | 284 | 100,0 | 48,3 | 92,8 | 106,3 | 71,4 | |
| 2 Grupos | C1 | 143 | 49,8 | 74,7 | 62,6 | 133 | 46,8 | 26,9 | 81,3 | 45,9 | 52,4 |
| | C2 | 144 | 50,2 | 231,4 | 83,0 | 151 | 53,2 | 67,2 | 92,8 | 159,6 | 78,6 |
| 3 Grupos | C1 | 99 | 34,5 | 51,4 | 58,3 | 67 | 23,6 | 15,9 | 68,3 | 21,1 | 41,4 |
| | C2 | 93 | 32,4 | 152,2 | 72,5 | 107 | 37,7 | 41,3 | 90,5 | 87,5 | 62,7 |
| | C3 | 95 | 33,1 | 260,6 | 85,5 | 110 | 38,7 | 74,0 | 98,8 | 176,5 | 81,1 |
| 4 Grupos | C1 | 55 | 19,2 | 26,7 | 46,1 | 57 | 20,1 | 13,3 | 63,7 | 16,9 | 37,8 |
| | C2 | 68 | 23,7 | 94,0 | 64,8 | 72 | 25,4 | 34,6 | 87,6 | 65,4 | 55,1 |
| | C3 | 75 | 26,1 | 168,2 | 76,1 | 80 | 28,2 | 57,4 | 95,7 | 125,0 | 72,8 |
| | C4 | 89 | 31,0 | 264,3 | 85,3 | 75 | 26,4 | 75,2 | 99,3 | 193,7 | 83,1 |
| 5 Grupos | C1 | 55 | 19,2 | 26,7 | 46,1 | 55 | 19,4 | 13,3 | 63,4 | 15,8 | 36,4 |
| | C2 | 63 | 22,0 | 91,3 | 63,3 | 69 | 24,3 | 33,5 | 87,7 | 64,0 | 55,8 |
| | C3 | 62 | 21,6 | 156,1 | 74,1 | 76 | 26,8 | 63,7 | 94,8 | 118,6 | 69,0 |
| | C4 | 71 | 24,7 | 223,8 | 83,2 | 53 | 18,7 | 51,2 | 99,3 | 172,0 | 83,5 |
| | C5 | 36 | 12,5 | 311,3 | 87,8 | 31 | 10,9 | 99,1 | 99,4 | 218,9 | 84,3 |
| 6 Grupos | C1 | 55 | 19,2 | 26,7 | 46,1 | 55 | 19,4 | 13,8 | 63,4 | 15,8 | 36,4 |
| | C2 | 63 | 22,0 | 91,3 | 63,3 | 65 | 22,9 | 35,0 | 87,4 | 61,7 | 53,8 |
| | C3 | 60 | 20,9 | 155,1 | 73,7 | 56 | 19,7 | 44,2 | 92,3 | 113,2 | 71,2 |
| | C4 | 37 | 12,9 | 260,6 | 84,0 | 30 | 10,6 | 101,0 | 97,9 | 135,2 | 69,0 |
| | C5 | 50 | 17,4 | 210,8 | 83,8 | 50 | 17,6 | 49,9 | 99,4 | 172,4 | 83,4 |
| | C6 | 22 | 7,7 | 331,2 | 88,4 | 28 | 9,9 | 96,4 | 99,4 | 224,9 | 85,6 |

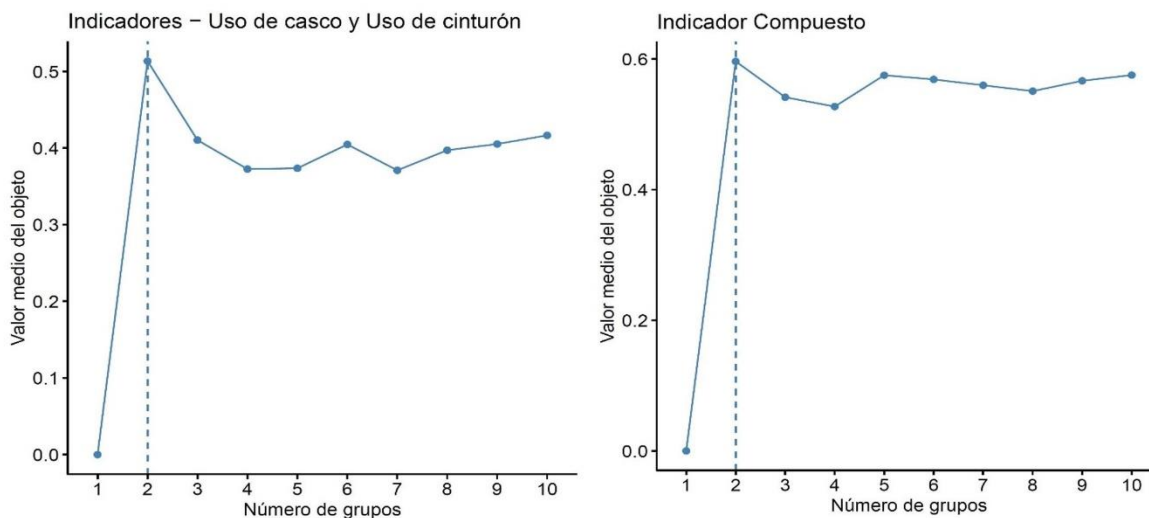
a/ Indicador compuesto se refiere a la suma de los indicadores de uso de casco y cinturón.

b/ Entiéndase “Uso promedio” como el centroide de cada grupo en la formación de los conglomerados.

c/ Porcentaje de uso del dispositivo de seguridad en cada grupo conformado.

Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

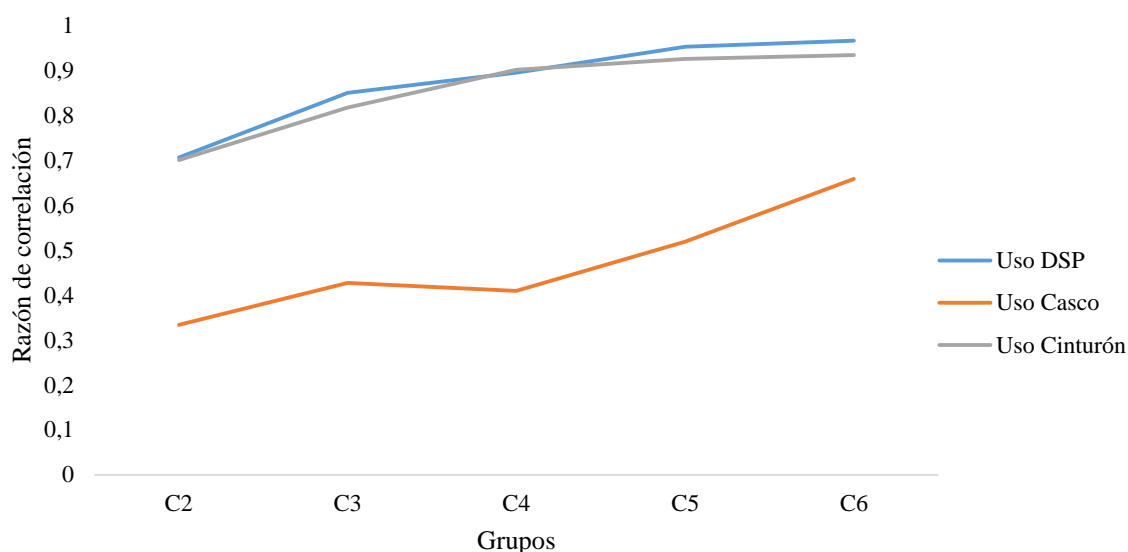
Gráfico 3. Valor medio del objeto según agrupaciones generadas



Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Además del análisis anterior puede recurrirse al cálculo de la razón de correlación “Eta²” (η^2), para verificar la relación existente entre los indicadores utilizados y las diferentes formaciones ensayadas para los conglomerados. Las agrupaciones candidatas a ser consideradas óptimas corresponden a aquellas donde exista una posible estabilización del coeficiente en relación con el número de grupos formados (gráfico 4).

Gráfico 4. Razón de correlación (η^2) para los indicadores sobre uso de dispositivos de seguridad pasiva y los grupos conformados



Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Para el indicador compuesto del uso de dispositivos de seguridad pasiva (Uso DSP) y del uso de cinturón de seguridad, el análisis gráfico de la razón de correlación evidencia estabilidad del coeficiente a partir de 4 o 5 grupos. En el caso del indicador del uso de casco de seguridad no parece estabilizarse en los diferentes escenarios planteados. Con base en la razón de correlación “eta²”, los resultados sugieren la conformación de un número elevado de grupos, mientras que el método “Average silhouette” propone 2 grupos. Pese a esto, la formación de cuatro o más agrupaciones no agrega ganancias de interpretación en términos sustantivos y la creación de dos categorías debilita la interpretación de los resultados, por lo cual se mantiene la decisión de tres grupos.

En estos tres grupos, el primero identifica el conjunto de intersecciones que presenta los promedios de uso de casco y cinturón más bajos, específicamente, de 15.9 y 21.1, respectivamente. En términos relativos, el uso de casco es 68.3% y el de cinturón es 41.1%, los que también fueron los más bajos encontrados. Por este motivo, se determinó que esta agrupación contiene intersecciones de “Bajo uso” de dispositivos de seguridad

La segunda agrupación presenta un uso promedio de casco y cinturón mayor que el grupo anterior (41.3 y 87.5 respectivamente). Además, el porcentaje de ocupantes de motocicleta que portaba casco es 90.6% y el de conductores que usaba cinturón es 62.7%. Por esta razón, esta agrupación se define como de “Uso medio” de dispositivos de seguridad pasiva.

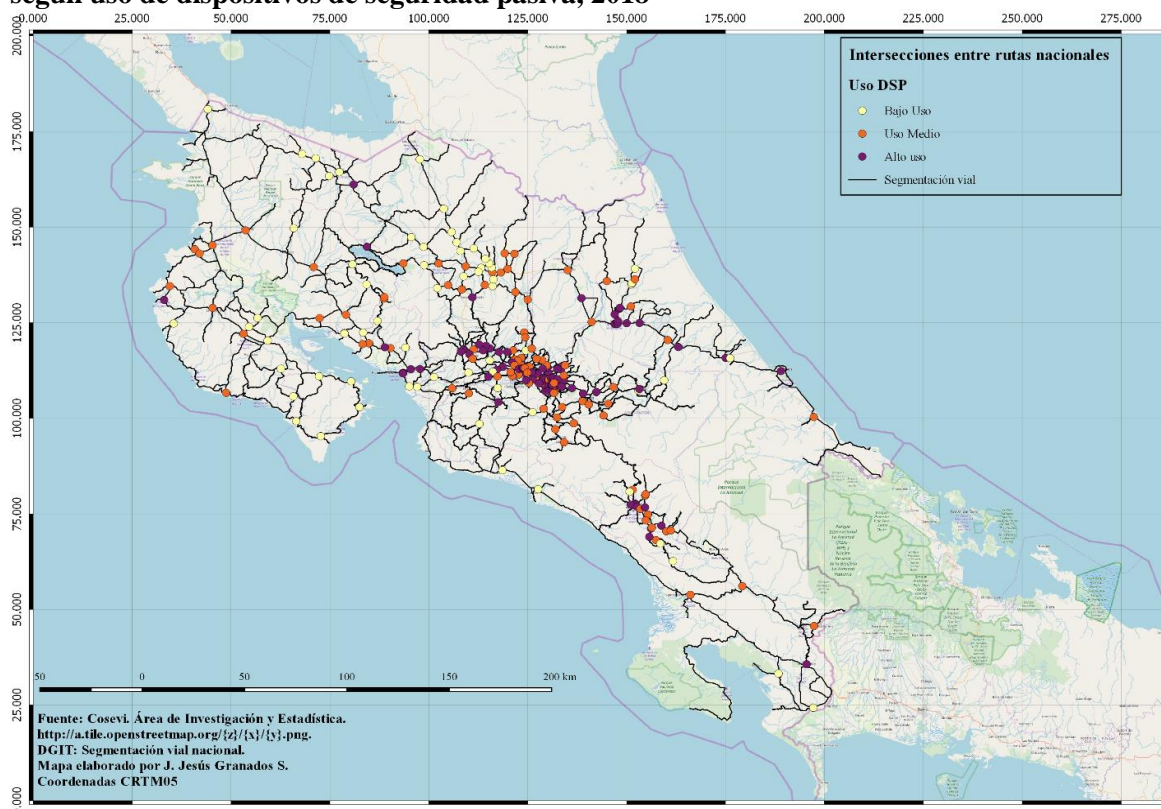
Por último, el tercer grupo se caracteriza por contar con promedios de uso de dispositivos más altas por parte de los ocupantes de motocicleta y automóvil. Además, en esta agrupación, 98.8% de los motociclistas utiliza casco de seguridad y 81.1% de los conductores usa el cinturón de seguridad, por lo que se determinó como una agrupación de “Alto uso” de dispositivos de seguridad.

El mapa 3 muestra la distribución espacial de las tres agrupaciones. Cada punto en el mapa representa una intersección del conjunto de intersecciones seleccionadas para el estudio, las cuales se encuentran distribuidas en la totalidad de la red vial nacional e identificadas según su conglomerado de pertenencia.

La representación espacial tiene como objetivo evidenciar el grado de uso de los dispositivos de seguridad pasiva a partir de las agrupaciones generadas. Como se aprecia, las zonas periféricas y rurales del país se caracterizan por un uso bajo o medio de casco y cinturón de seguridad, resultado esperable si se toma en cuenta que se trata de regiones donde la presencia policial es reducida en comparación con el Valle Central o zonas de mayor urbanización y densidad vehicular.

Al contrario de las zonas rurales, el GAM está caracterizada por un uso medio y alto de casco y cinturón de seguridad, potenciado entre otros factores, por la mayor presencia policial en relación a su contraparte rural.

Mapa 3. Costa Rica: Distribución de intersecciones entre rutas nacionales seleccionadas según uso de dispositivos de seguridad pasiva, 2018



1.4.2.2. Análisis de los grupos conformados para determinar su validez

Ahora se analizan las características de las intersecciones para determinar la existencia de asociación con los grupos establecidos. Como primer acercamiento se analiza el grado de asociación con las variables categóricas y, posteriormente, con las variables numéricas.

1.4.2.2.1. Análisis mediante variables categóricas

El análisis con las variables categóricas evidencia que este tipo de variable guarda asociación con los grupos de uso de dispositivos de seguridad pasiva (cuadro 4). El “tipo de intersección” ($C.Cont. = 0.22$ y $V Cramer = 0.22$) y la “señalización vertical” ($C.Cont. = 0.25$ y $V Cramer = 0.26$) son las que presentan valores menores, mientras que las mayores asociaciones se encontraron con las variables GAM ($C.Cont. = 0.38$ y $V Cramer = 0.41$) y zona ($C.Cont. = 0.45$ y $V Cramer = 0.50$).

Cuadro 4. Distribución de intersecciones por uso de dispositivos de seguridad pasiva según características de la intersección, 2018

| Característica | Medida de asociación | Alto Uso | | Uso Medio | | Bajo Uso | | Total | | |
|------------------------|----------------------|--------------------------|------------|-------------|------------|--------------|-----------|-------------|------------|---------------|
| | | n | % | n | % | n | % | n | % | |
| Tipo Inter. | "+ o x" | <i>C.Conting. (0,22)</i> | 41 | 52,56 | 29 | 37,18 | 8 | 10,26 | 78 | 100,00 |
| | Rest. inter. | <i>V Cramer (0,22)</i> | 69 | 33,50 | 78 | 37,86 | 59 | 28,64 | 206 | 100,00 |
| Señal. vertical | Sí | <i>C.Conting. (0,25)</i> | 93 | 43,46 | 84 | 39,25 | 37 | 17,29 | 214 | 100,00 |
| | No | <i>V Cramer (0,26)</i> | 17 | 24,29 | 23 | 32,86 | 30 | 42,86 | 70 | 100,00 |
| GAM | GAM | <i>C.Conting. (0,38)</i> | 64 | 56,14 | 46 | 40,35 | 4 | 3,51 | 114 | 100,00 |
| | No GAM | <i>V Cramer (0,41)</i> | 46 | 27,06 | 61 | 35,88 | 63 | 37,06 | 170 | 100,00 |
| Zona | Urbano | <i>C.Conting. (0,45)</i> | 100 | 53,48 | 67 | 35,83 | 20 | 10,70 | 187 | 100,00 |
| | Rural | <i>V Cramer (0,50)</i> | 10 | 10,31 | 40 | 41,24 | 47 | 48,45 | 97 | 100,00 |
| Total | | | 110 | 38,7 | 107 | 37,68 | 67 | 23,6 | 284 | 100,00 |

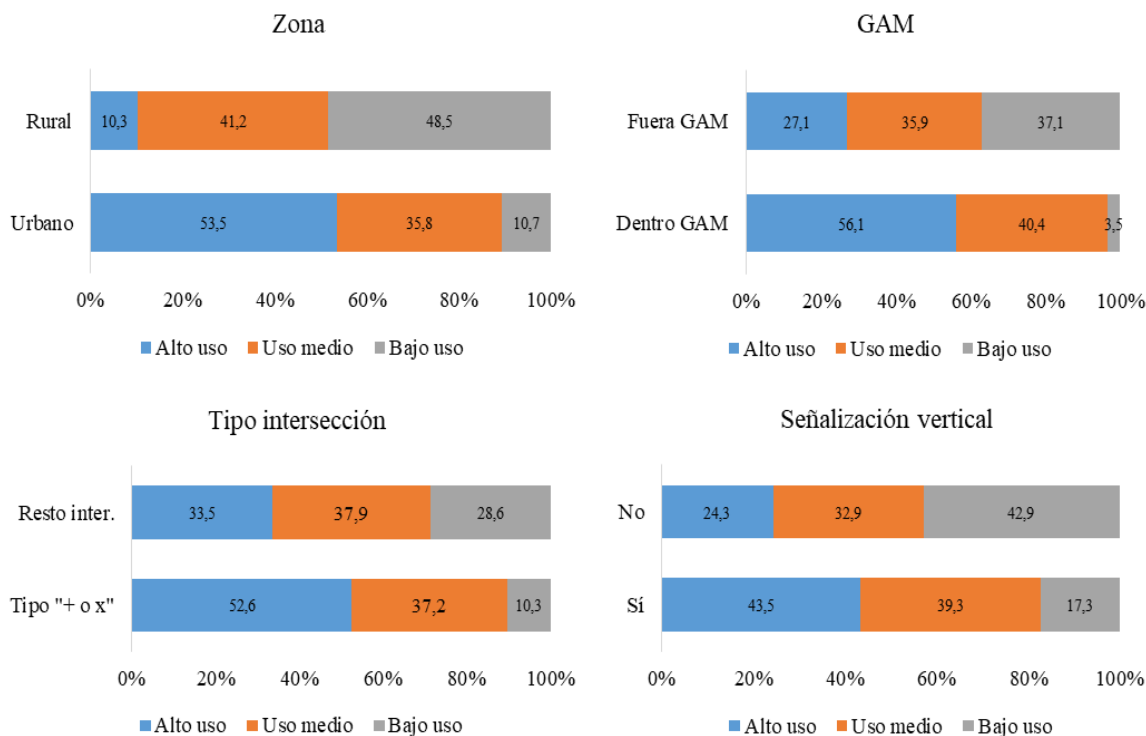
Fuente: Elaboración con base en el estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores, Área de Investigación y Estadística, Cosevi. 2018.

El gráfico 5 ilustra las relaciones encontradas. La distribución de las observaciones presenta diferencias según el grupo de uso de dispositivos de seguridad que se analice. Por ejemplo, para el caso de la zona aproximadamente el 53% de las intersecciones está caracterizado por un alto uso de dispositivos de seguridad en la parte urbana, mientras que el bajo uso es poco frecuente, con valores cercanos a 11%. El comportamiento experimentado en las intersecciones de la parte rural es contrario al observado en su contraparte urbana, predominando un bajo uso de dispositivos de seguridad con porcentajes de uso cercanos al 48%, mientras que el alto uso presenta valores apenas del 10%.

El uso de dispositivos de seguridad presenta un patrón definido dentro de la GAM, con porcentajes de alto uso que rondan 56%, seguido del uso medio con aproximadamente 40% y en menor proporción, el bajo uso (3.5%). En el caso de las intersecciones fuera de la GAM, el porcentaje de bajo uso y uso medio resulta similar, con porcentajes de 36% y 37% respectivamente, mientras que el alto uso es menos frecuente (27%).

En relación con el tipo de intersección, en las de clase “+ o x” existe mayor presencia de alto uso (52.6%) y uso medio (37.2%) en comparación con el bajo uso (10.3%). Para el resto de intersecciones, el patrón se modifica al experimentar un incremento porcentual en el bajo uso (28.6%) y una reducción en el porcentaje de intersecciones con alto uso (33.5%), mientras que los casos con uso medio permanecen prácticamente invariantes (37.9%).

Gráfico 5. Distribución de intersecciones por uso de dispositivos de seguridad pasiva según características de la intersección, 2018



Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Cuando existe presencia de señalamiento vertical en las vías de tránsito, se tiene un comportamiento similar al experimentado en el tipo de intersección, pues el porcentaje de alto uso es mayor, con valores cercanos al 44% en comparación con 17% de bajo uso. Para el grupo de intersecciones con ausencia de señalamiento vertical, el porcentaje de alto uso se reduce a 24% y la cantidad de usuarios con bajo uso del dispositivo toma relevancia, con porcentajes cercanos a 43%. El uso medio de dispositivos varía levemente en función de la ausencia o presencia de señalamiento, con una diferencia aproximada de 6%.

1.4.2.2.2. Análisis mediante variables numéricas

En el caso de las variables numéricas, existen correlaciones de magnitud importante, principalmente entre las variables flujo vehicular con uso de casco (0.74), flujo vehicular con uso de cinturón (0.55) y uso de casco con uso de cinturón (0.59) (cuadro 5).

Cuadro 5. Correlaciones para las variables numéricas incluidas en el análisis

| Covariables | S. horizontal | Cant. acciden | Flujo veh. | Uso casco | Uso cinturón |
|-------------------------|----------------------|----------------------|-------------------|------------------|---------------------|
| S. horizontal | 1 | | | | |
| Cant. accidentes | 0,34 | 1 | | | |
| Flujo veh. | 0,23 | 0,47 | 1 | | |
| Uso casco | 0,23 | 0,46 | 0,74 | 1 | |
| Uso cinturón | 0,33 | 0,45 | 0,55 | 0,59 | 1 |

Nota: Todas las correlaciones resultaron significativas al 5%.

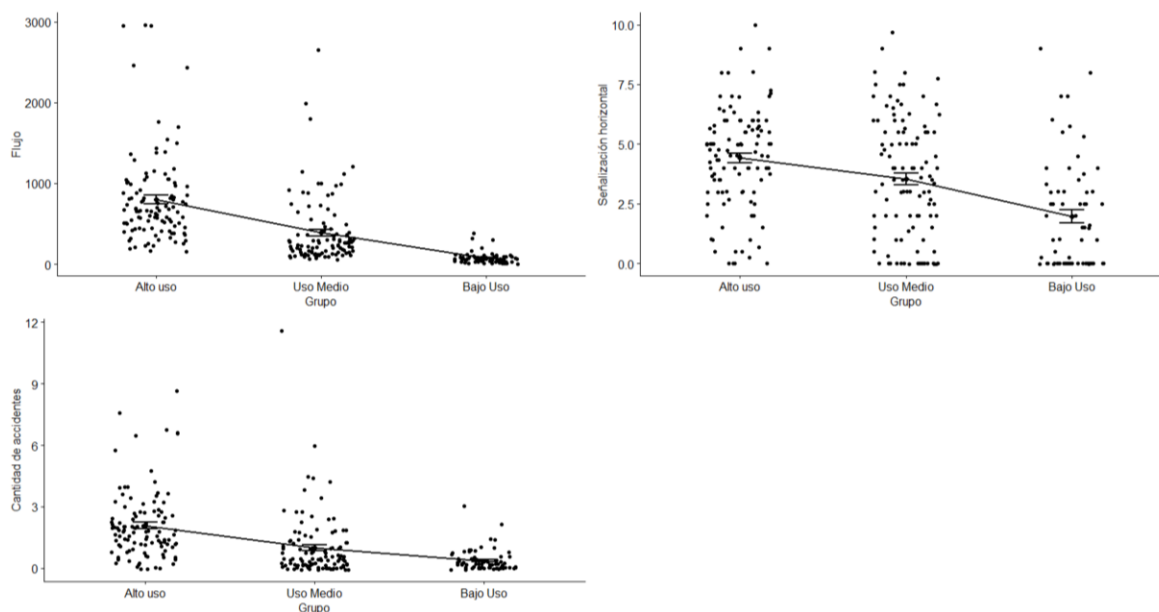
Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Los resultados a partir de las correlaciones sirven como premisa para plantear hipótesis que puedan respaldar o explicar las relaciones mostradas en el cuadro anterior, principalmente para el caso del flujo vehicular con el uso de casco y cinturón. Ambas presentan una relación directa y considerable con el flujo vehicular, lo que es consistente con el hecho de que estos indicadores son utilizados en términos absolutos y representan una consecuencia directa del flujo de automóviles y motocicletas que circularon por las intersecciones donde se realizaron las mediciones. Es decir, es razonable asumir que a mayor flujo vehicular exista un mayor uso de dispositivos de seguridad pasiva.

De la misma forma, la cantidad media de colisiones que ocurren en determinada intersección se ve influenciada por la magnitud del flujo vehicular, por lo que en aquellos puntos donde exista mayor afluencia de vehículos automotores, la propensión en la ocurrencia de siniestros viales con lesionados o fallecidos experimentaría un incremento. Por otro lado, como consecuencia del resultado anterior, en las intersecciones con mayor presencia media de colisiones con personas lesionadas o fallecidas, se esperaría que el uso de dispositivos de seguridad pasiva (uso de caso y cinturón) sea mayor, tomando en cuenta que son puntos de mayor flujo vehicular.

Como en el caso de las variables categóricas, debe verificarse la existencia de patrones relevantes entre las variables numéricas y los grupos para determinar asociación. De forma preliminar, por medio de análisis gráfico de los promedios de cada variable contra las agrupaciones, se evidencian diferencias en las medias de las covariables para los diferentes grupos al verificar que los intervalos no se traslapan en las tres variables analizadas (gráfico 6). De las tres características evaluadas, el flujo presenta la mayor diferencia en sus medias según el grupo analizado.

Gráfico 6. Promedios de las variables numéricas utilizadas en el análisis según grupos de uso de dispositivos de seguridad pasiva, 2018



Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Lo anterior proporciona indicios de posible asociación, por lo que se recurre a la aplicación de pruebas formales para determinar la existencia de diferencias entre los promedios, y si dichas diferencias se mantienen entre los grupos. Para la prueba formal, se recurre al análisis de varianza no paramétrico Kruskal-Wallis, debido al incumplimiento del supuesto de normalidad (cuadro A1, anexos).

Como resultado del análisis de varianza no paramétrico (cuadro A2, anexos), las tres variables resultan significativas, lo cual implica diferencias en al menos en una de las medias de cada variable por grupo. Esto se verifica con las comparaciones múltiples de Wilcoxon como se observa en el siguiente cuadro.

Cuadro 6. Comparaciones múltiples de Wilcoxon para los promedios de las covariables según grupos de uso de dispositivos de seguridad pasiva, 2018

| Variable | | Grupo | | |
|-------------------------|-----------|----------|-----------|----------|
| | | Bajo uso | Uso medio | Alto uso |
| Flujo | Bajo uso | *** | *** | *** |
| | Uso medio | 2E-16 | *** | *** |
| | Alto uso | 2E-16 | 1,8E-14 | *** |
| Señalamiento horizontal | Bajo uso | *** | *** | *** |
| | Uso medio | 0,000087 | *** | *** |
| | Alto uso | 8,5E-11 | 0,0073 | *** |
| Cantidad accidentes | Bajo uso | *** | *** | *** |
| | Uso medio | 0,00028 | *** | *** |
| | Alto uso | 2E-16 | 4,5E-11 | *** |

Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Estos resultados respaldan un comportamiento diferenciado en los promedios del flujo vehicular, señalamiento horizontal y cantidad de accidentes según los grupos de uso de dispositivos de seguridad pasiva, lo que representa un escenario razonable para caracterizar las intersecciones a partir de la clasificación generada y las variables propuestas para la creación de la tipología.

1.4.3. Regresión logística multinomial

Realizado el análisis preliminar de las relaciones entre variables, se procede a generar el perfil para las intersecciones a partir de las covariables propuestas y los grupos de uso de dispositivos de seguridad. Para elaborar el perfil se emplea como técnica de análisis un modelo de regresión logística multinomial, utilizando los grupos de uso de dispositivos de seguridad como variable dependiente y las características de la intersección descritas anteriormente como variables predictoras. Sin embargo, previo a la elaboración del perfil, es importante detallar elementos referentes al ajuste del modelo.

1.4.3.1. Aspectos relacionados con el ajuste del modelo

Definido el modelo preliminar, se realiza la selección de variables mediante los criterios de información de Bayes (BIC) y Akaike (AIC). A partir de este procedimiento, se elige el modelo con el menor valor del AIC y BIC, el cual resulta con las variables predictoras “Zona”, “GAM”, “Cantidad promedio de accidentes” y “Flujo vehicular” (cuadro A1, anexos).

Para el modelo ajustado se evaluaron diferentes aspectos como la bondad de ajuste, supuestos y diagnósticos (ver sección de anexos). Respecto a la multicolinealidad (cuadro A2, anexos), a partir de los valores del Factor de Inflación de la Varianza (FIV), no se encuentran indicios de correlación severa entre los predictores del modelo que pudiera ocasionar inestabilidad en la estimación de los coeficientes por causa de un incremento en su varianza.

Al evaluar la bondad de ajuste para determinar el grado de cohesión del modelo propuesto a los datos, se emplea la prueba de Hosmer & Lemeshow (cuadro A3, anexos) aplicando al modelo multinomial y de forma separada a cada uno de los dos modelos binarios⁶ resultantes, con lo cual se observa un incumplimiento en el caso multinomial. Sin embargo, su aplicación en los dos modelos binarios resulta en un adecuado ajuste para el grupo de “Alto uso”, pero no así para el caso del grupo de “Uso medio”, en el que no se puede garantizar que exista un buen ajuste con base en la prueba utilizada, a pesar de la evidencia mostrada mediante las pruebas estadísticas del apartado preliminar 1.4.2.2.2, donde se muestra la existencia de diferencias significativas en los promedios de los tres grupos a partir de las comparaciones múltiples realizadas.

Con base en el resultado anterior, se puede interpretar que el modelo logístico multinomial parece adecuado para comparar mediciones realizadas en las intersecciones de “Alto uso” en relación con el grupo de “Bajo uso”, sin embargo, resulta inadecuado para las comparaciones entre el grupo de “Uso medio” y de “Bajo uso”, al no estar captando las diferencias mostradas entre estas dos agrupaciones. Esto sugiere que es preferible ajustar un modelo logístico binario a partir de la conformación de dos grupos para el uso de dispositivos de seguridad pasiva en lugar de tres agrupaciones.

El razonamiento anterior sería válido en otras circunstancias, no obstante, dicha propuesta no procede debido a las decisiones tomadas en el apartado inicial, relacionadas con la conformación del número idóneo de grupos de intersecciones, donde se justificó la construcción de tres grupos debido a su mayor aporte interpretativo del fenómeno estudiado.

La evaluación de la capacidad predictiva del modelo a través de la curva de ROC (Receiver Operating Characteristic), arroja un elevado porcentaje correcto de clasificación de observaciones, 99% para el logito de “Alto uso” y 94% para el logito de “Uso medio”, reflejando un buen ajuste entre los datos originales y los teóricos generados por el modelo (gráfico A1 y A2, anexos).

⁶ A partir del modelo logístico multinomial, se definen dos modelos logísticos binarios para la evaluación de diferentes aspectos relacionados con la calidad del ajuste del modelo.

La identificación de valores influenciales causantes de afectación en el ajuste del modelo (gráfico A3 y A4, anexos), se realizó mediante la aplicación de diagnósticos correspondientes a la representación de los “Leverage”, el cambio en X de Pearson, el cambio en la devianza (para determinar el efecto de cada patrón de covariables en el ajuste del modelo) y por último, la medida del efecto de cada patrón de covariables en la estimación de los parámetros, cuantificado mediante el incremento en los parámetros estimados (Distancias de Cook).

Luego de aplicar los diagnósticos, se identificó de forma reiterada, un grupo de observaciones en los diferentes gráficos, por lo que se procedió a realizar el ajuste del modelo sin estas observaciones, identificando un cambio porcentual considerable en la estimación de los coeficientes. Por este motivo se decide eliminar los casos influenciales que generaban dicha afectación (cuadro A4, anexos).

1.4.3.2. Características relevantes en el uso de dispositivos de seguridad pasiva

El cuadro 7 proporciona el detalle de las variables que resultaron relevantes en el ajuste del modelo logístico multinomial y se observa cómo para ambos grupos de uso de dispositivos de seguridad, las variables con relevancia en la explicación de la variabilidad de la característica de interés (grupos de uso de dispositivos de seguridad) corresponden a: “Zona”, “GAM”, “Flujo vehicular” y “Cantidad media de accidentes”, de las cuales, “GAM” y “Flujo vehicular” resultaron significativas al 5% para ambos grupos.

Cuadro 7. Resultados del ajuste del modelo de regresión logística multinomial, 2018

| Grupo | Variable | Odds ratio | Error estandar | Prob. Asoc. | IC 95% | | |
|------------------------|------------------------|---------------|----------------|-------------------|--------|--------|-------|
| | | | | | Inf. | Sup. | |
| Alto Uso | Zona | | | | | | |
| | <i>Urbano</i> | 0,947 | 0,654 | 0,934 | -0,335 | 2,230 | |
| | <i>Rural</i> | | | Referencia | | | |
| | GAM | | | | | | |
| | <i>Dentro GAM</i> | 5,766 | 0,755 | 0,020 | 4,287 | 7,245 | |
| | <i>Fuera GAM</i> | | | Referencia | | | |
| | Flujo vehicular | 1,032 | 0,005 | 0,000 | 1,021 | 1,042 | |
| | Cantidad Acc. | 1,317 | 0,381 | 0,470 | 0,570 | 2,063 | |
| | Uso Medio | Zona | | | | | |
| | | <i>Urbano</i> | 0,329 | 0,571 | 0,052 | -0,791 | 1,449 |
| <i>Rural</i> | | | | Referencia | | | |
| GAM | | | | | | | |
| <i>Dentro GAM</i> | | 7,965 | 0,715 | 0,004 | 6,565 | 9,366 | |
| <i>Fuera GAM</i> | | | | Referencia | | | |
| Flujo vehicular | | 1,030 | 0,005 | 0,000 | 1,020 | 1,041 | |
| Cantidad Acc. | 0,991 | 0,376 | 0,980 | 0,255 | 1,727 | | |

Notas: AIC = 367.2603 y BIC = 403.4643

Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

De los resultados del modelo anterior y a partir de las variables significativas, se desprenden algunas interpretaciones relacionadas con el uso de dispositivos de seguridad pasiva en cada uno de los grupos de interés.

Al comparar el grupo de “Alto uso” de dispositivos de seguridad pasiva respecto al grupo de “Bajo uso”, se obtiene que el odds de las intersecciones dentro de la GAM es 5.8 veces mayor que las que se encuentran fuera de la GAM. Esto significa que las intersecciones que se encuentran dentro de la GAM presentan aproximadamente 6 veces más posibilidades de ser de “Alto uso”.

Para el caso del flujo, por cada vehículo o motocicleta adicional que circula por la intersección, el odds de pertenecer al grupo de “Alto uso” incrementa 3% respecto al grupo de “Bajo uso”

Respecto al grupo de “Uso medio”, al tratarse de las mismas variables significativas, se omite la interpretación de la razón de odds al ser similar al caso de “Alto uso”. Sin embargo, es importante destacar que el odds correspondiente a la categoría dentro de la GAM es aproximadamente 7.9 veces en comparación al odds fuera de la GAM y en el caso del flujo vehicular, el odds presenta un valor de 1.03, similar al grupo de “Alto uso”.

1.4.4. Árboles de decisión

La caracterización de las intersecciones entre rutas nacionales forma parte de los objetivos del presente trabajo, para lo que se planteó usar el modelo logístico multinomial (desarrollado en apartados anteriores) y los árboles de clasificación. La finalidad era generar las tipologías con ambas técnicas y compararlas para determinar la consistencia de los resultados obtenidos. Sin embargo, como se corroboró con el análisis de conglomerados, existe una fuerte relación entre la variable “flujo vehicular” y los grupos de uso de dispositivos de seguridad pasiva. Como consecuencia, la creación del perfil mediante la técnica de árboles de clasificación resultó afectada, pues la técnica consideró únicamente el “flujo vehicular” como relevante en el análisis y descartó las restantes variables de la clasificación proceso, con lo que la elaboración del perfil se vio limitada y, por ende, la comparación con los resultados del modelo multinomial.

1.5. Conclusiones

Con base en el archivo de datos del “*Estudio nacional sobre uso de dispositivos de seguridad pasiva, vestimenta y factores distractores, 2019*”, realizado por el Consejo de Seguridad Vial en el año 2018 (Cosevi, 2019), se realizaron los análisis requeridos para clasificar intersecciones entre rutas nacionales en Costa Rica con base en el uso de dispositivos de seguridad pasiva, para generar perfiles que sirvan como insumo en la planificación de estrategias de intervención para mejorar el uso de dispositivos de seguridad en el territorio nacional.

Para crear los grupos de intersecciones a partir del uso de dispositivos de seguridad pasiva, se utilizaron tres indicadores: el uso de casco de seguridad por parte de los ocupantes de motocicleta, el uso de cinturón de seguridad en ocupantes de automóvil y un tercero calculado a partir de la suma de los dos anteriores.

La decisión del número de grupos y la elección del mejor indicador se realizó mediante el uso de recursos estadísticos combinado con criterio sustantivo. A partir de lo anterior, los análisis proporcionan los siguientes resultados:

- No se detectó inconveniente en la distribución del número de intersecciones en cada uno de los grupos formados, ya que los conglomerados presentan un balance aceptable, tanto para el caso de los dos indicadores independientes como para el indicador compuesto.
- Respecto a los grupos formados con el indicador compuesto (la suma del uso de casco y uso de cinturón), aunque soluciona el problema detectado para el uso de casco, en la formación de grupos se prefiere hacerla a partir de los indicadores individuales, debido a su mayor aporte interpretativo.
- En relación con el número de grupos, las técnicas estadísticas utilizadas sugieren la creación de dos grupos mediante el método gráfico “Average silhouette” y 4 o más en el caso de la razón de correlación. Sin embargo, no se justifica la formación de cuatro o más agrupaciones, al no agregar información relevante en términos sustantivos y la creación de dos categorías resta capacidad de interpretación de los resultados, por lo que la opción de tres grupos se toma como el escenario que brinda mejor equilibrio entre el número de agrupaciones y el componente sustantivo.

Como parte de la validación de los conglomerados generados, se valoraron las agrupaciones con otras de las variables propuestas en el análisis, para identificar algún grado de relación que permita evidenciar un patrón claramente definido en los datos como justificación del número de grupos

establecido. A partir de esto, en las variables categóricas (tipo intersección, señalización vertical, zona y GAM), se corroboró asociación entre las variables, mostrando un patrón diferenciado en cada una según grupo analizado. Las diferencias que tienen mayor presencia se dan en las variables “zona” y “GAM”. Para el caso de las variables numéricas, el análisis gráfico preliminar arrojó indicios de asociación, los cuales fueron comprobados mediante el análisis en la diferencia de medias y comparaciones múltiples, verificando diferencias significativas entre los diferentes promedios para cada una de las variables (flujo vehicular, cantidad de accidentes y señalización horizontal) según grupos de uso de dispositivos de seguridad pasiva.

Para caracterizar las intersecciones entre rutas nacionales con base en los grupos creados y las variables propuestas, se usó regresión logística multinomial y árboles de decisión, con el objetivo de comparar los resultados de ambas técnicas. Sin embargo, el análisis mediante árboles de decisión se vio afectado por la fuerte relación que existe entre la variable flujo vehicular y las agrupaciones formadas a partir del uso de dispositivos de seguridad pasiva, por lo que las restantes variables no resultaron relevantes en el análisis, imposibilitando generar el perfil y la respectiva comparación con los resultados del modelo logístico multinomial.

Por el motivo expuesto, la caracterización de las intersecciones se realizó únicamente con la regresión logística multinomial, y como resultados importantes en el ajuste del modelo se obtiene que las variables “Zona”, “GAM”, “Flujo vehicular” y “Cantidad media de accidentes” resultan relevantes en la explicación de la variabilidad de la característica de interés. Además, del grupo de variables analizadas para ambos grupos de uso de dispositivos de seguridad pasiva (Alto uso y Uso medio), el “Flujo vehicular” y “GAM”, resultaron significativas al 5%, lo cual significa que en ambas agrupaciones, las intersecciones que se encuentran dentro de la GAM presentan mayores posibilidades de ser de uso medio o alto uso de dispositivos de seguridad pasiva, y que incrementos en el flujo vehicular se traducen en aumentos en la posibilidad de pertenecer a intersecciones de uso medio o alto uso.

Como aspecto de importancia en el análisis realizado, al valorar los resultados obtenidos a partir del modelo logístico en relación con las magnitudes mostradas por la razón de “odds” en las variables flujo vehicular y cantidad media de accidentes, se confirman las suposiciones o hipótesis planteadas que indican que los indicadores de uso de casco y cinturón, en términos absolutos corresponden a una consecuencia directa del flujo vehicular, por lo que es razonable asumir que a mayor flujo se experimenta mayor uso de dispositivos de seguridad pasiva. De igual forma, se verificó que en intersecciones con mayor presencia media de accidentes con personas heridas o fallecidas, el uso de dispositivos de seguridad es mayor, lo cual es provocado nuevamente por el flujo vehicular.

Como limitante del modelo logístico se encontraron problemas en el grupo de “Uso medio” para ajustarse apropiadamente a los datos, pese a la evidencia estadística obtenida en relación con las comparaciones múltiples realizadas. Esto significa que el modelo es adecuado para comparar mediciones realizadas en las intersecciones de “Alto uso” en relación con el grupo de referencia “Bajo uso”, pero no así para el grupo de “Uso medio”, lo cual implicaría una recodificación a dos categorías y el respectivo ajuste de un modelo logístico binario para captar con mayor eficiencia las diferencias mostradas entre las agrupaciones. No obstante, esta opción fue desechada debido a la línea de razonamiento que se desarrolló en el presente trabajo respecto a las decisiones tomadas en torno a la formación del número adecuado de grupos a partir del análisis de conglomerados.

A partir de los indicadores utilizados se lograron definir grupos que clasifican las intersecciones según el uso de dispositivos de seguridad, lo que permite identificar geográficamente cuáles puntos presentan bajos niveles de uso de casco y cinturón de seguridad. De igual manera, la caracterización de las intersecciones proporciona información sobre la dinámica particular en cuanto a su condición de pertenecer o no a la GAM y el flujo circulante en cada punto, lo que permite a las autoridades de tránsito optimizar en tiempo y espacio, el recurso material y humano disponible, bajo una planificación eficiente mediante operativos diferenciados según el comportamiento particular de cada intersección. Además, la intervención policial descrita funciona como un acompañamiento importante y necesario para garantizar el éxito de iniciativas que pueden desarrollar otras entidades a partir de la información generada, que consiste en la creación de políticas y estrategias de intervención dirigidas a mejorar la situación en puntos de bajo uso y uso medio, mediante la aplicación de campañas de comunicación, charlas en centros educativos, empresas y demás lugares en los que resulta necesario focalizar esfuerzos para concientizar y fomentar el uso de dichos dispositivos.

Los resultados del presente trabajo sirven como premisa al planteamiento de propuestas metodológicas para mejorar aspectos relacionados con el abordaje particular de los grupos creados en función de su dinámica, ya sea para el propio estudio en cuestión o investigaciones externas y, además, para poder evaluar si el método de selección de los vehículos debe definirse considerando variables de la intersección como flujo vehicular y pertenencia a la GAM para generar estimaciones con mayor estabilidad y validez.

Bibliografía

- Agresti, A. (2002). *Categorical data analysis*. New Jersey, EEUU: John Wiley & Sons, Inc.
- Allison, P. (2003). *Logistic regression using the SAS system: Theory and application*. EEUU.
- Arora Preeti, D. D. (2016). *Analysis of K-Means and K-Medoids algorithm for big data*. *Procedia Computer Science*, 507 - 512.
- Breiman, L. F. (1984). *Clasificación and Regression Trees*. EEUU: Chapman & Hall.
- Brian S. Everitt, S. L. (2011). *Cluster Analysis*. Londres, Inglaterra: Wiley.
- Solanoo, D. (2014). *Construcción de indicadores y clasificación de cantones por riesgo de accidentes de tránsito en Costa Rica*. San José, Costa Rica.
- COSEVI. (2011). *Plan estratégico nacional en el marco del decenio de acción para la seguridad vial 2011 - 2020 "Construyendo una cultura de paz en las carreteras"*. San José, Costa Rica.
- COSEVI. (2018, Mayo 8). *COSEVI*. <http://datosabiertos.csv.go.cr/dashboards/19737/informacion-general-de-costa-rica/>
- Cosevi. (2019). *Estudio nacional sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores, 2018*. San José, Costa Rica.
- Elliot, K., & Schneider, H. (2013). *2013 Louisiana seat belt and motorcycle helmet observation survey results*. Louisiana, Estados Unidos.
- Elvik, R., Høy, A., & Vaa, T. &. (2009). *The handbook of road safety measures*. Oslo, Noruega: Esmerald Group Publishing Limited.
- Europa, C. E. (2007). *Convención sobre la Circulación Vial de 1968*. Nueva York, Estados Unidos.
- FIT. (2017). *Cero muertes y lesiones de gravedad por accidentes de tránsito: Liderar un cambio de paradigma hacia un sistema seguro*. París, Francia.
- Foundation, F. (2009). *Cinturones de seguridad y sistemas de retención infantil: un manual de seguridad vial para decisores y profesionales*. London, United Kingdom.
- Glizer, I. M. (1993). *Prevención de accidentes y lesiones*. Washington, D.C. Estados Unidos.
- Hosmer, D. L. (2013). *Applied Logistic Regression*. New Jersey, EEUU: John Wiley & Sons, Inc. .
- Kassambara, A. (2018, Octubre 22). *Statistical tools for high-throughput data analysis*. Retrieved from <http://www.sthda.com/english/>
- Lee, C., Pino, J., & Shultz, D. (2014). *Measuring the use of motorcycle Helmets: Observational survey of motorcyclist in Florida*. Florida, Estados Unidos.
- Lin, P.-S., Lee, C., & Saxena, M. (2010). *Observational survey of motorcyclists*. Florida, Estados Unidos.
- Neter, J. K. (2004). *Applied linear statistical models*. EEUU: McGraw-Hill.

- OdD. (2016). *Estudio observacional para la determinación de los factores de riesgo en accidentes donde están involucradas motocicletas en tres distritos de Costa Rica*. San José, Costa Rica.
- OECD. (2018, Abril 21). *OECD Better policies for better lives*. Retrieved from <http://www.oecd.org/health/health-data.htm>
- OMS. (2004). *Informe mundial sobre traumatismos causados por el tránsito*. Ginebra, Suiza.
- OMS. (2013). *Informe sobre la situación mundial de la seguridad vial 2013*. Ginebra, Suiza.
- OMS. (2017). *Salve vidas - Paquete de medidas técnicas de seguridad vial*. Ginebra, Suiza.
- Puppo, C., Centeno, P., Pablo, R., & Serulnicoff, M. (2013). *Estudio de campo observacional urbano sobre el uso de cinturón de seguridad, sistemas de retención infantil en vehículos particulares y uso de cascos en motociclistas y ciclomotores en el estado Plurinacional de Bolivia*. Bolivia.
- Rodríguez, O. H. (2013). *Temas de análisis estadístico multivariante*. San José, Costa Rica: Editorial UCR.
- Rousseeuw, P. S. (1997). Clustering in an object-oriented environment. *Journal of Statistical Software*, 1-30.
- Salud, O. M. (2015). *Informe sobre la situación mundial de la seguridad vial 2015*. Ginebra, Suiza.
- Salud, O. P. (2016). *La seguridad vial en la región de las Américas*. Washington, DC. Estados Unidos.
- Savolainen, P., Gates, T., Morden, J., & Hacker, E. (2013). *2013 Direct observation survey of motorcycle helmet use*. Detroit, Estados Unidos.
- Serna, S. (2009). *Comparación de árboles de regresión y clasificación y regresión logística*. Medellín, Colombia.
- SIECA. (2014). *Manual Centroamericano de dispositivos uniformes para el control del tránsito*. Guatemala.
- Turner, P., & Hagelin, C. (2004). *Florida motorcycle helmet use observational survey and trend analysis*. Florida, Estados Unidos.
- UNASEV. (2013). *Estudio observacional de conductas viales en zonas urbanas*. Uruguay.
- Vías, I. N. (2008). *Manual de diseño geométrico de carreteras*. Colombia.

Anexo

Anexo A: Análisis de diferencias en los promedios de los grupos de uso de dispositivos de seguridad pasiva

Cuadro A1
Verificación de normalidad para las variables numéricas mediante la prueba Shapiro-Wilk

| Covariable | W | Prob. |
|-------------------------|----------|--------------|
| S. horizontal | 0,98391 | 0,002799 |
| Cant. accidentes | 0,76632 | 2,2E-16 |
| Flujo veh. | 0,74794 | 2,2E-16 |

Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Cuadro A2
Análisis de varianza no paramétrico Kruskal-Wallis

| Covariable | X2 | Prob. |
|-------------------------|-----------|--------------|
| S. horizontal | 43,349 | 3,863E-10 |
| Cant. accidentes | 91,158 | 2,2E-16 |
| Flujo veh. | 173,39 | 2,2E-16 |

Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Anexo B: Bondad de ajuste y diagnósticos del modelo

Cuadro B1

Resultados del ajuste del modelo de regresión logística multinomial, 2018

| Grupo | Variable | Odds ratio | Error estandar | Prob. Asoc. | IC 95% | | |
|------------------------|------------------------|---------------|----------------|-------------------|--------|--------|-------|
| | | | | | Inf. | Sup. | |
| Alto Uso | Zona | | | | | | |
| | <i>Urbano</i> | 0,957 | 0,633 | 0,945 | -0,284 | 2,198 | |
| | <i>Rural</i> | | | Referencia | | | |
| | GAM | | | | | | |
| | <i>Dentro GAM</i> | 6,007 | 0,727 | 0,014 | 4,582 | 7,431 | |
| | <i>Fuera GAM</i> | | | Referencia | | | |
| | Flujo vehicular | 1,027 | 0,004 | 0,000 | 1,018 | 1,035 | |
| | Cantidad Acc. | 1,503 | 0,441 | 0,356 | 0,638 | 2,368 | |
| | Uso Medio | Zona | | | | | |
| | | <i>Urbano</i> | 0,338 | 0,548 | 0,048 | -0,737 | 1,412 |
| <i>Rural</i> | | | | Referencia | | | |
| GAM | | | | | | | |
| <i>Dentro GAM</i> | | 8,042 | 0,688 | 0,002 | 6,694 | 9,391 | |
| <i>Fuera GAM</i> | | | | Referencia | | | |
| Flujo vehicular | | 1,025 | 0,004 | 0,000 | 1,017 | 1,033 | |
| Cantidad Acc. | | 1,070 | 0,435 | 0,876 | 0,219 | 1,922 | |

Notas: AIC: 385.4162 y BIC: 421.906

Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Cuadro B2

Factores de inflación de la varianza para las variables numéricas en el modelo

| Grupo | Covariables | |
|------------------|------------------|-----------------|
| | Cant. Accidentes | Flujo vehicular |
| Alto uso | 1,33 | 2,02 |
| Uso medio | 1,12 | 1,16 |

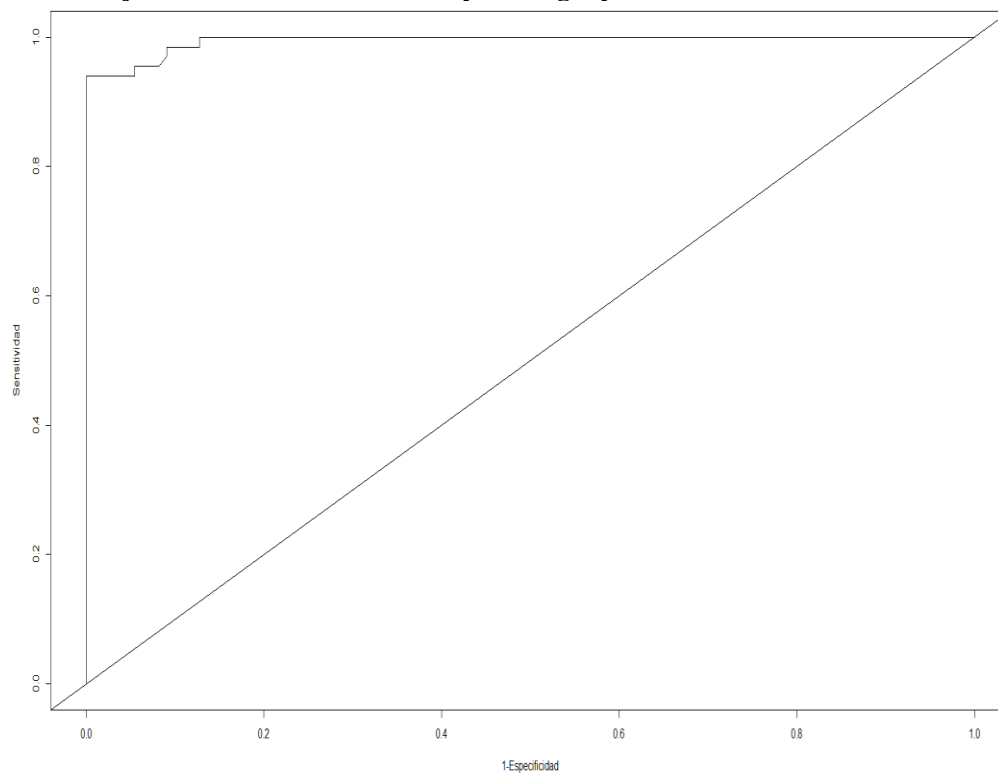
Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Cuadro B3
Resultados de la prueba (Hosmer and Lemeshow) para
la validación de bondad de ajuste del modelo

| Modelo | Estadístico X2 | Probabilidad |
|-------------|----------------|--------------|
| Multinomial | 71,9160 | 0,0000 |
| Binario | Alto uso | 2,6237 |
| | Uso medio | 78,3940 |
| | | 0,0000 |

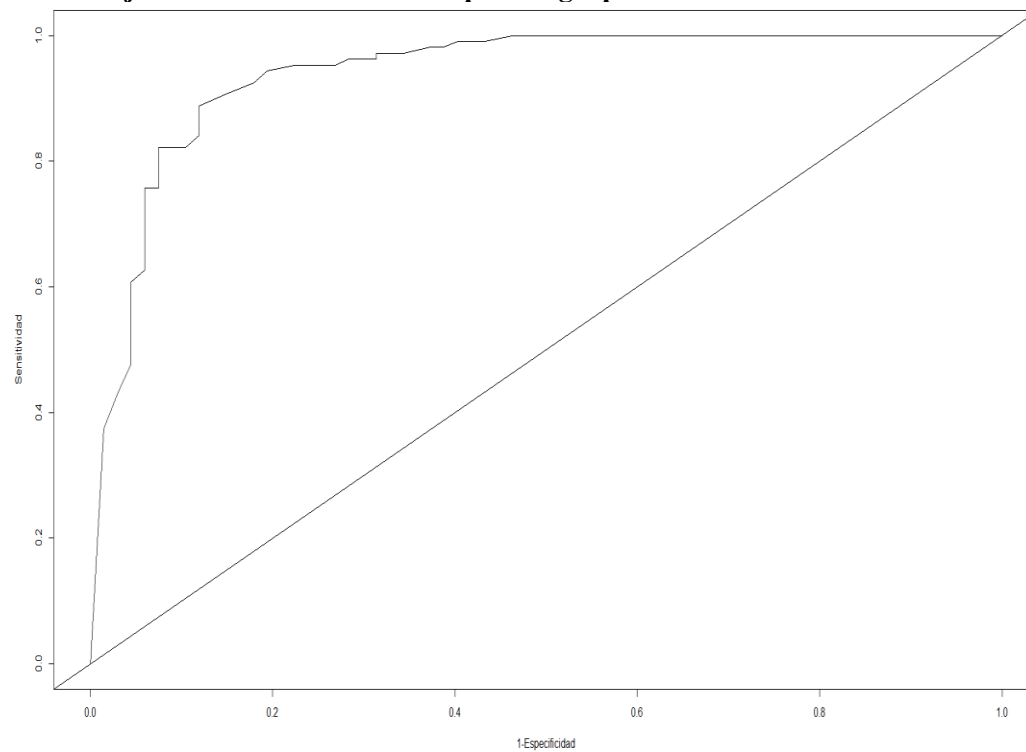
Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Gráfico B1
Porcentaje correcto de clasificación para el grupo de “Alto uso”



Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

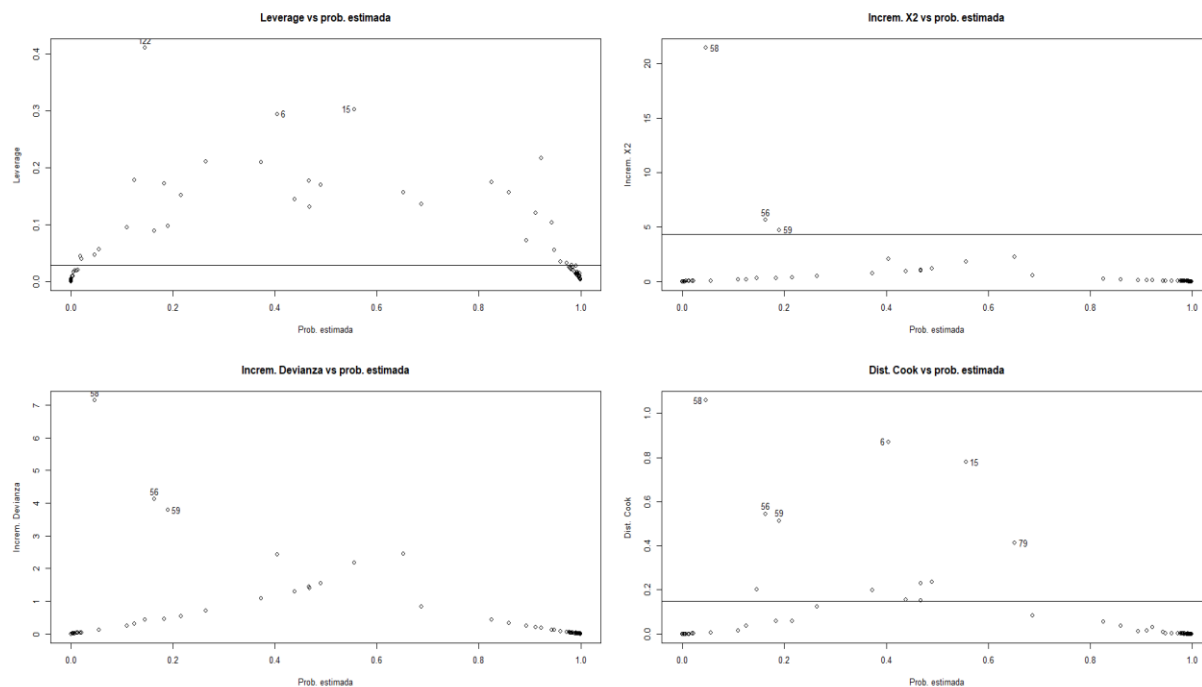
Gráfico B2
Porcentaje correcto de clasificación para el grupo de “Uso medio”



Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Gráfico B3

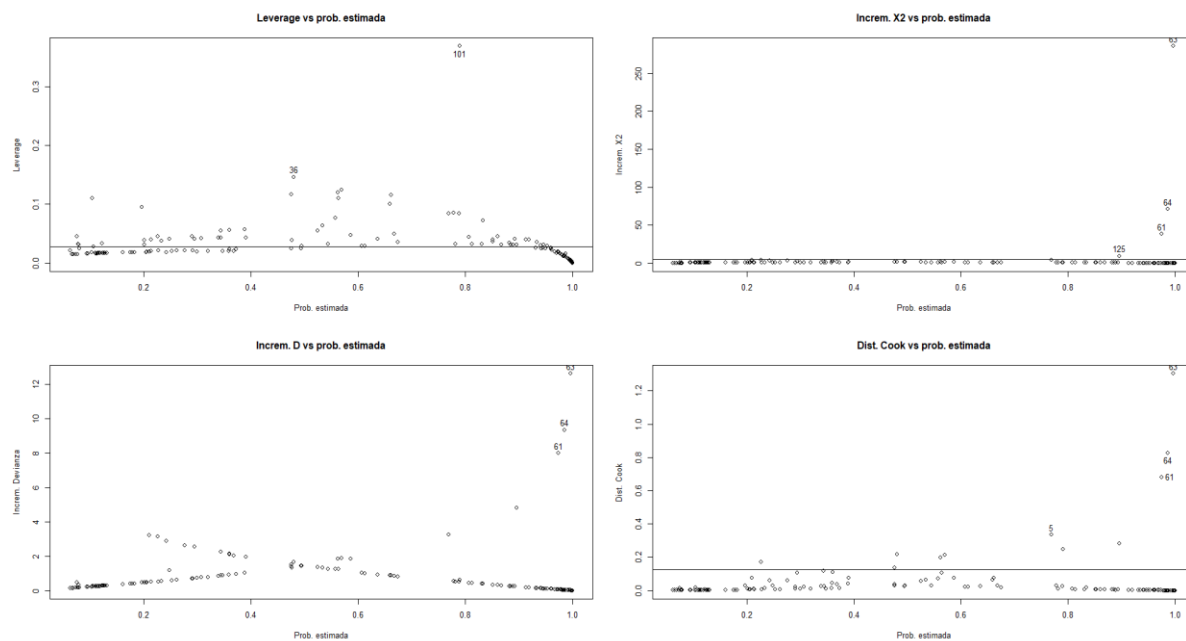
Diagnósticos del modelo para la verificación de valores influenciales en el grupo de “Alto uso”



Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Gráfico B4

Diagnósticos del modelo para la verificación de valores influenciales en el grupo de “Uso medio”



Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Cuadro B4
Cambio porcentual en los coeficientes del modelo al eliminar valores influénciales

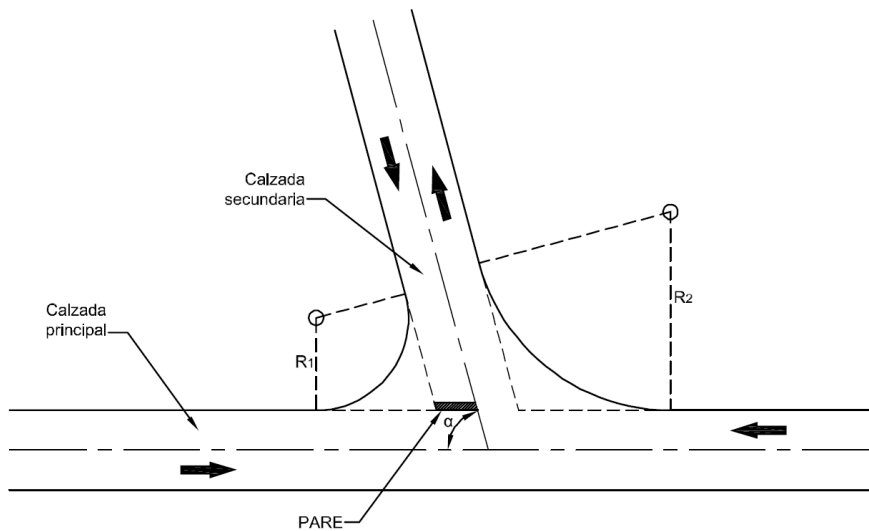
| Logito | Covariable | Coef. V. Influénciales | Coeficientes | Cambio porcentual |
|------------------|-------------------|-------------------------------|---------------------|--------------------------|
| Alto Uso | Constante | -4.8668 | -5.2572 | -7.426 |
| | Zona | -0.0449 | -0.0540 | -16.828 |
| | GAM | 1.7944 | 1.7520 | 2.424 |
| | Cant. Accidentes | 0.3382 | 0.2751 | 22.916 |
| | Flujo | 0.0263 | 0.0312 | -15.883 |
| Uso Medio | Constante | -2.9909 | -3.3800 | -11.511 |
| | Zona | -1.0861 | -1.1109 | -2.232 |
| | GAM | 2.0862 | 2.0751 | 0.536 |
| | Cant. Accidentes | 0.0554 | -0.0093 | -692.881 |
| | Flujo | 0.0248 | 0.0297 | -16.717 |

Fuente: Consejo de Seguridad Vial. Área de Investigación y Estadística. Estudio sobre el uso de dispositivos de seguridad pasiva, vestimenta y factores distractores. 2018.

Anexo C Esquemas de intersecciones

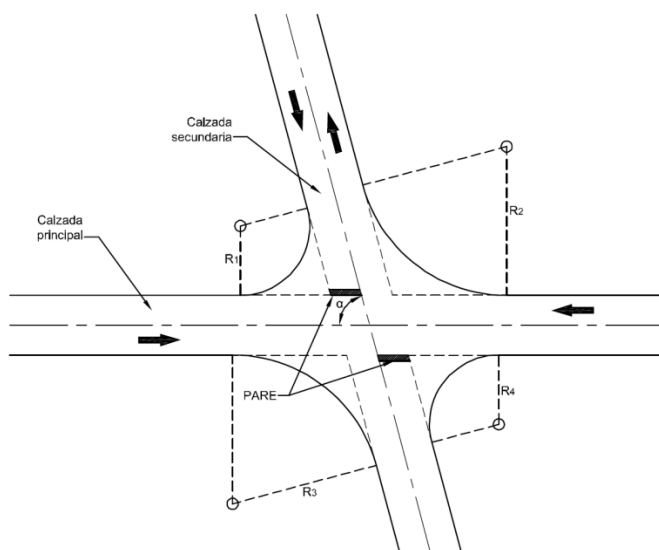
Intersecciones a nivel - Sin canalizar

Intersección en "T" o "Y"



Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

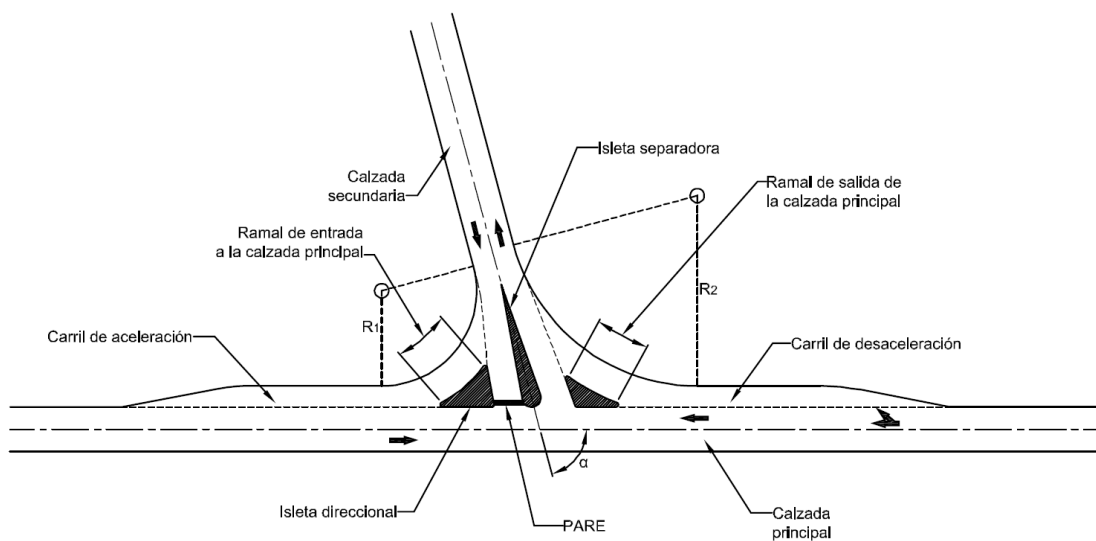
Intersección en Cruz "+" o Equis "x"



Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

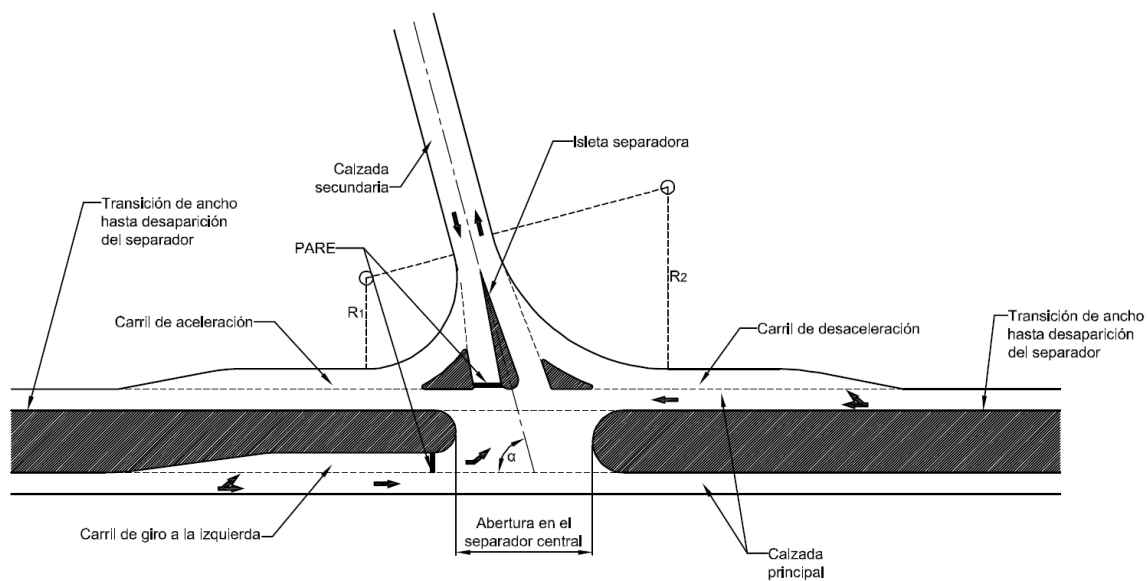
Intersecciones a nivel - Canalizadas

Intersección en "T" o "Y"



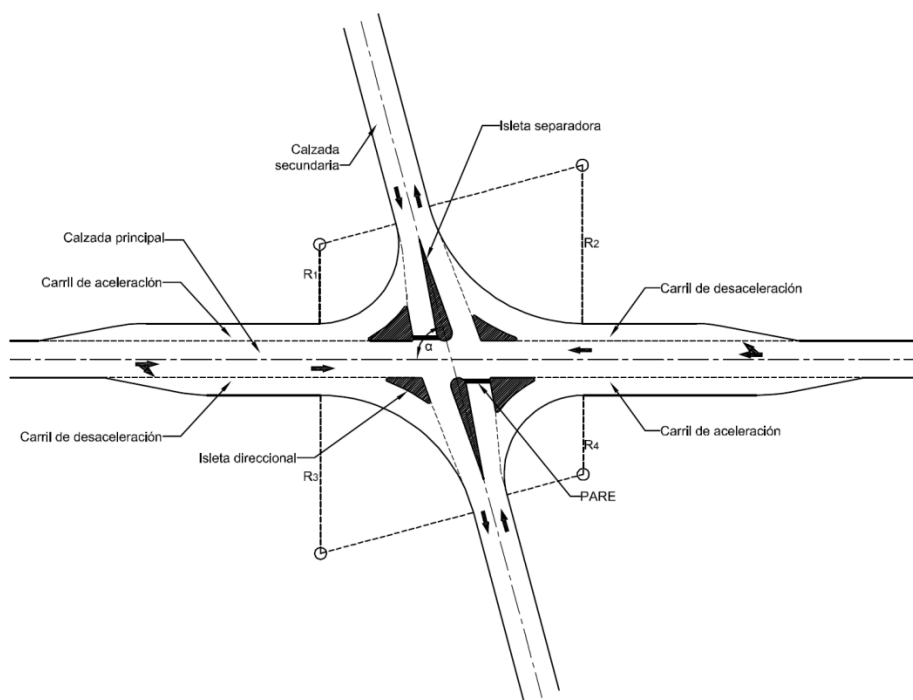
Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

Intersección en "T" o "Y" con separador y carril de giro a la izquierda



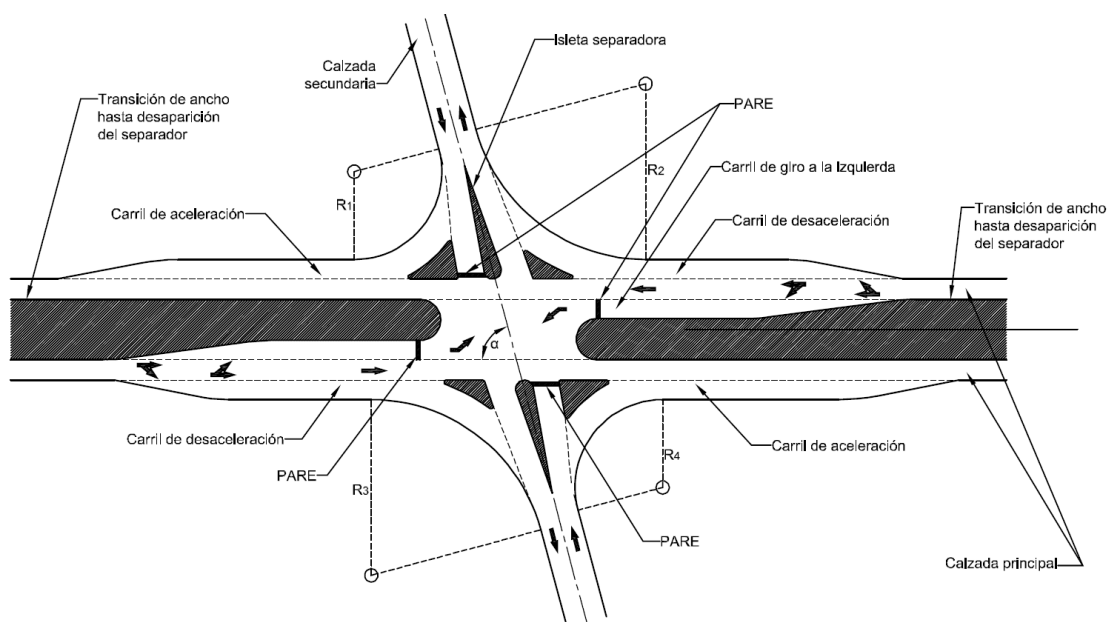
Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

Intersección en "Cruz +" o Equis "x"



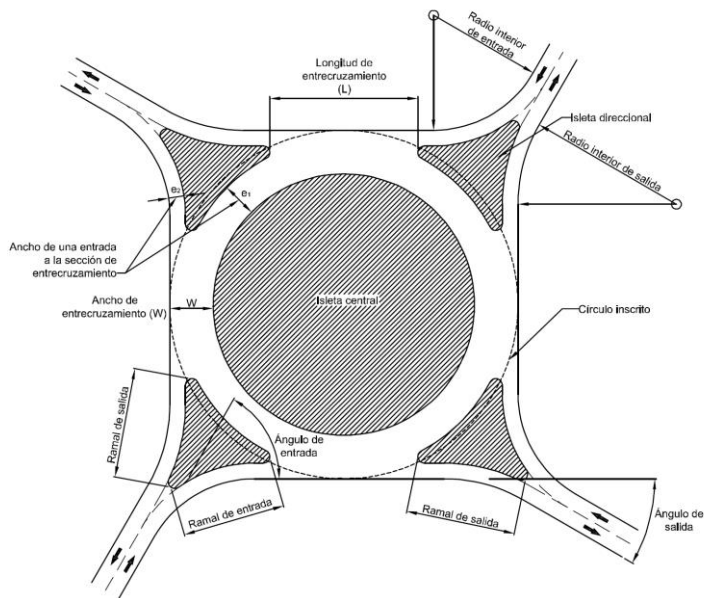
Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

Intersección en "Cruz +" o Equis "x" con separador y carril de giro a la izquierda



Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

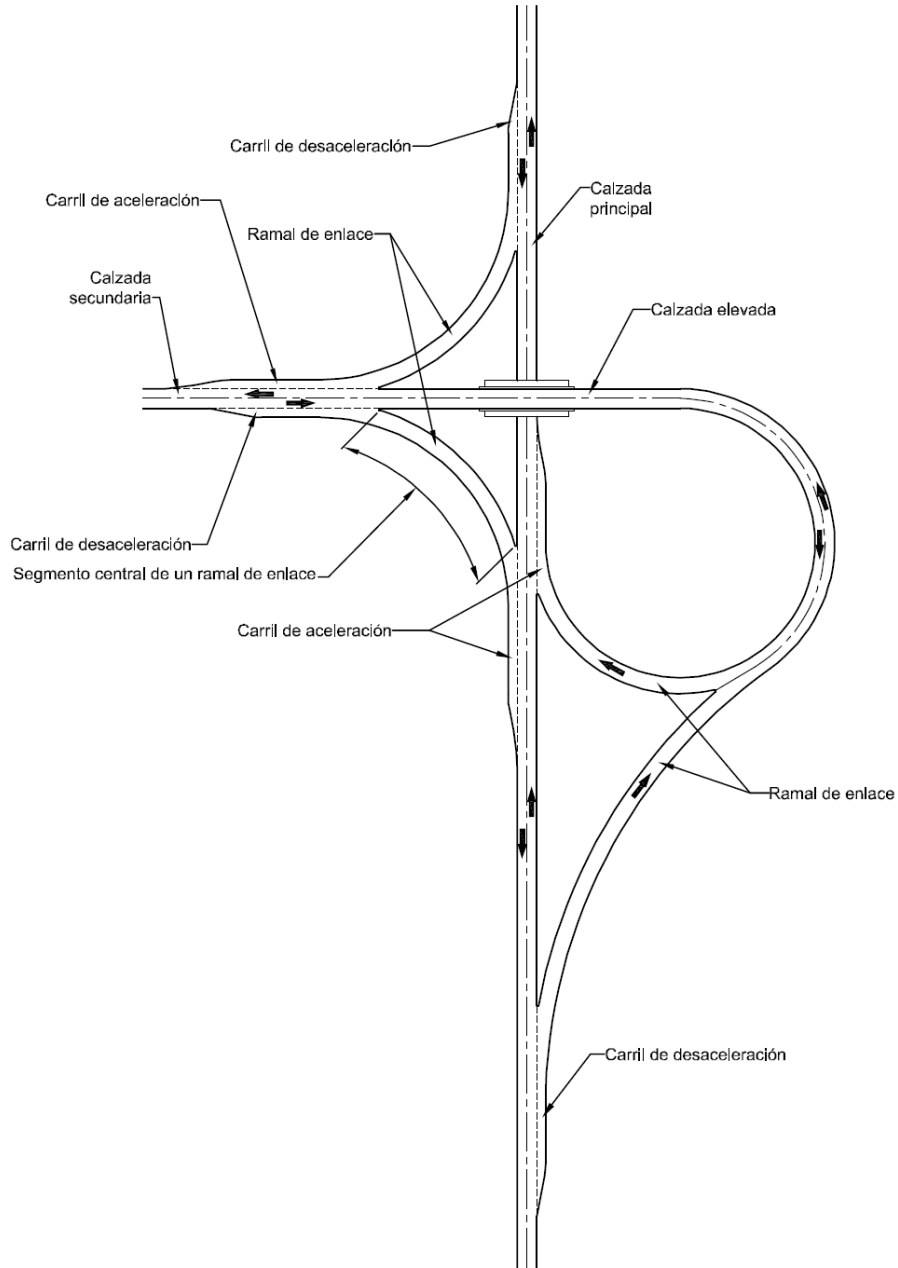
Intersección tipo rotonda



Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

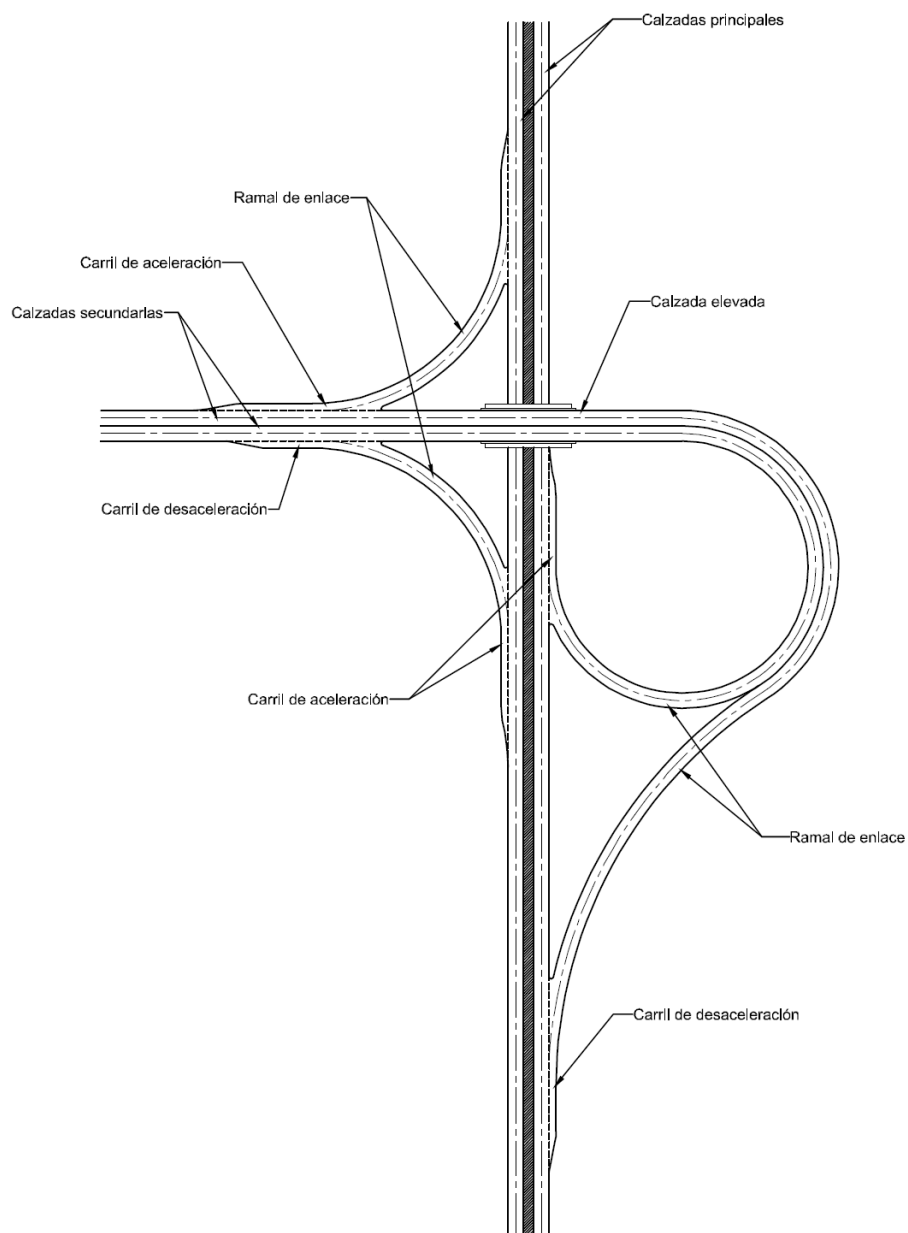
Intersecciones a desnivel

Tipo "Trompeta" en carreteras no divididas



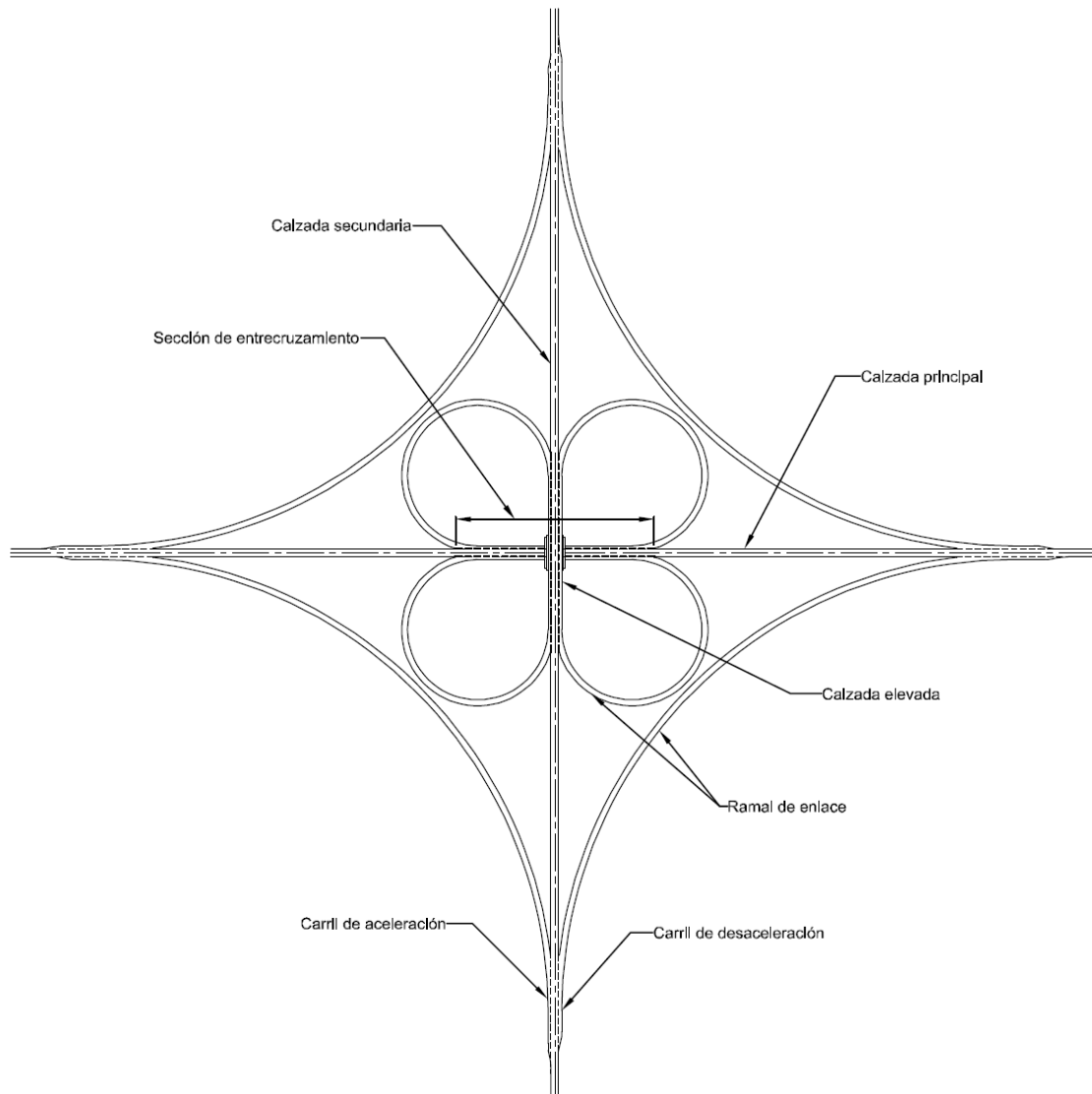
Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

Tipo "Trompeta" en carreteras divididas



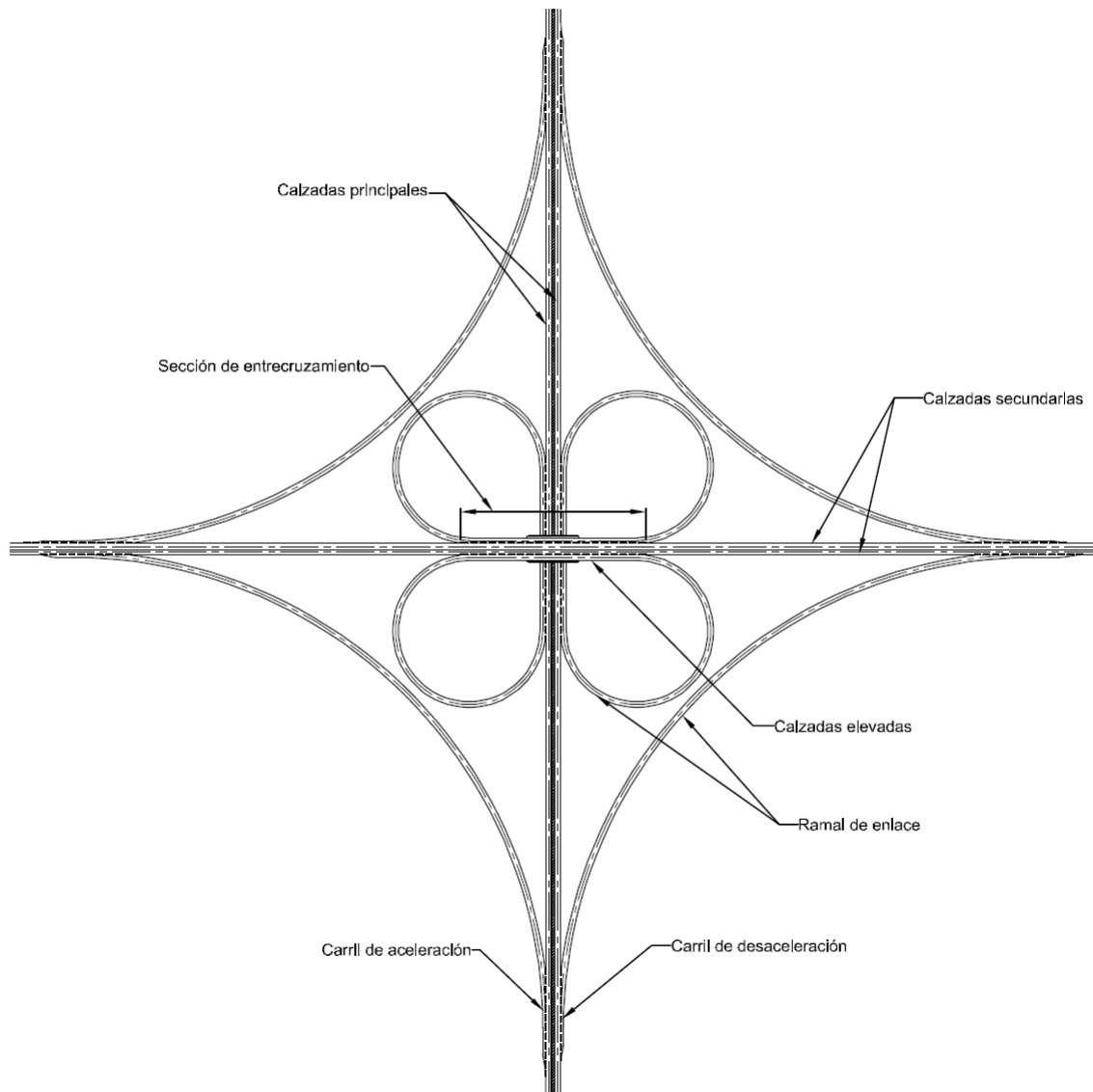
Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

Tipo "Trebol" en carreteras no divididas



Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

Tipo "Trebol" en carreteras divididas



Fuente: Manual de Diseño Geométrico de Carreteras, Instituto Nacional de Vías, 2008

2. Práctica Profesional II

2.1. Introducción

Las encuestas para estudiar temas de índole social, económico, político, demográfico y de salud, entre otras, por lo general usan diseños complejos y multietápicos con marcos muestrales en los que las unidades de análisis se encuentran establecidas en un determinado espacio. Entre estos diseños, como indican Grosh & Muñoz (1996) citados por Himelein, Eckman & Murray (2014), se encuentra el muestreo estratificado de varias etapas; sin embargo, como menciona Kalton (1991) y Kalton (2001), dichos métodos no resultan efectivos para estudiar características cuando el análisis se realiza a partir de poblaciones móviles. En humanos pueden distinguirse dos tipos de poblaciones móviles: quienes no residen regularmente en una locación fija, como las personas nómadas, y en general, quienes pertenecen a la población móvil debido a que se encuentra en tránsito, como por ejemplo visitantes de librerías, personas compradoras, viajeros y ocupantes de vehículos.

Debido a estas particularidades en el comportamiento de las poblaciones móviles, resulta de gran importancia estudiar diferentes alternativas en la formulación de diseños de muestreo para considerar la dinámica de los flujos de las personas en relación con su densidad en un determinado espacio y periodo de tiempo.

En el ámbito de la seguridad vial, como en otras áreas, la aplicación del muestreo para obtener información debe desarrollarse en función de los objetivos del estudio y de la naturaleza del fenómeno analizado. Por este motivo, el planteamiento de la presente investigación se basa en el trabajo realizado por el Consejo de Seguridad Vial durante el año 2016, en el que se midió el uso de dispositivos de seguridad pasiva. Se trata de un estudio observacional, con cobertura nacional, cuyo marco muestral fue la red de rutas nacionales (sin incluir rutas cantonales) segmentada en tramos de un kilómetro. En el muestreo, en una primera etapa se seleccionaron al azar los tramos de carretera con sus respectivas franjas horarias donde se realizaron las mediciones y, posteriormente, en una segunda etapa, en cada UPM seleccionada, con el uso de una cámara de video, un observador grabó la totalidad de los vehículos que circularon por la franja horaria, generando gran cantidad de información digital de la que posteriormente fueron extraídas las variables de interés en un archivo de datos.

Desde el punto de vista del muestreo estadístico, el procedimiento de videograbar la totalidad de vehículos que circulan en determinado tramo y franja horaria es equivalente a seleccionar conglomerados completos. Lo anterior sirve de premisa para plantear diferentes interrogantes

relacionadas con la posibilidad de usar submuestreo en la segunda etapa de selección utilizando diferentes procedimientos y analizando el efecto que tendría este cambio metodológico en el cálculo de los estadísticos de interés. Concretamente, interesa comprobar si es posible disminuir la variabilidad de las estimaciones aplicando diferentes técnicas de selección de vehículos en la segunda etapa, como lo son el muestreo simple al azar, la selección sistemática y la selección de los primeros “n” elementos que circulan por la UPM. Lo anterior brinda una segunda oportunidad de exploración, pues la posibilidad de submuestrear con diferentes técnicas lleva a determinar el número óptimo de vehículos que deben seleccionarse de las UPM para disminuir la variabilidad de las estimaciones.

Dado que el interés es analizar estrategias metodológicas para calcular estadísticos con menor variabilidad en este tipo de estudios, también se considera importante analizar el efecto que tendría la posestratificación de tramos de red vial en el proceso estimativo. Se considera que este aspecto es necesario de tomar en cuenta porque la dinámica vial del país presenta un comportamiento diferenciado según la densidad del flujo vehicular, dependiendo de la ubicación y la franja horaria. El diseño muestral utilizado en el estudio del 2016 no contempla una estratificación del marco muestral con base en el flujo vehicular, por lo que resulta de interés recurrir a la posestratificación para crear agrupaciones de UPM a partir del flujo, identificando estratos de menor y mayor circulación de vehículos, con el objetivo de analizar posibles efectos en la estimación del uso de dispositivos de seguridad, como producto de la aplicación de diferentes técnicas de selección y submuestreo en estratos con flujo vehicular diferenciado.

Tomando como parámetro de comparación los resultados obtenidos del registro de los conglomerados completos, en lo referente al uso de dispositivos de seguridad pasiva (uso de cinturón), se desarrollaron diferentes escenarios a partir de procesos de simulación para determinar si la técnica de selección de vehículos, el submuestreo y la posestratificación juegan un papel relevante en la estimación de los estadísticos de interés. Asimismo, se considera que el trabajo es indispensable para investigar con detalle el diseño metodológico utilizado para medir el uso de dispositivos de seguridad pasiva y proponer un diseño muestral que contribuya a obtener mediciones con mayor precisión.

A continuación se abordan los principales aspectos conceptuales a considerar en el diseño de muestreos complejos para el estudio de poblaciones móviles. Posteriormente, en la tercera sección, correspondiente a la metodología, se detallan elementos propios del diseño de muestreo utilizado en 2016, además de las técnicas y procedimientos estadísticos utilizados en el análisis de los datos. La cuarta sección abarca el desarrollo y análisis de los resultados del estudio y, por último, se finaliza con las principales conclusiones del trabajo.

2.2. Aspectos conceptuales relacionados con el muestreo de poblaciones móviles

El estudio de poblaciones móviles constituye un tema de interés dentro del muestreo probabilístico, ya que las unidades de análisis no se encuentran en un punto fijo del espacio, ni en un periodo de tiempo determinado, sino que mantienen un dinamismo que dificulta su estudio. La particularidad de su comportamiento representa un reto de muestreo y de análisis, lo que invita al planteamiento de nuevas estrategias y formulación de metodologías para obtener estimaciones de calidad.

Debido a lo anterior, el estudio de conceptos como *tiempo* y *espacio* adquiere especial relevancia en la planeación de marcos y diseños de muestreo para el análisis de poblaciones móviles. Usualmente, los marcos de muestreo para este tipo de poblaciones se definen a partir de listados con base en combinaciones de lugares e intervalos de tiempo como UPM, y a partir de dichas UPM pueden agregarse más etapas de selección, dependiendo de las características de la población y el procedimiento de recolección de datos. Un elemento clave corresponde a la naturaleza del flujo de la población móvil, debido a los patrones de variabilidad que pueden presentarse en las UPM y que están relacionados con comportamientos distintos del flujo entre lugares o intervalos de tiempo o, inclusive, dentro de la misma UPM, convirtiéndose en factores que influyen en el diseño de muestreo del estudio (Kalton, G., 1991).

Continuando con el razonamiento anterior, si el flujo se comporta parcialmente uniforme y los intervalos de tiempo son los mismos, la cantidad de elementos por UPM sería aproximadamente constante y, en ese caso, las UPM podrían seleccionarse con igual probabilidad y utilizar una fracción de submuestreo constante dentro de la UPM seleccionada para generar muestras con igual probabilidad (Kalton, G., 1991). Sin embargo, como mencionan Kish (1972) y Kalton (1991), en muchos casos el flujo varía a través de la UPM de forma parcialmente predecible, produciendo diferente número de elementos y, con ello, UPM de tamaño desigual. Sobre esto Kish (1972) agrega que, en la práctica, en la mayoría de ocasiones los conglomerados tanto de poblaciones humanas como no humanas contienen diferente cantidad de elementos; y aunque se diseñen muestras a partir de conglomerados de igual tamaño, estos pueden finalizar en conglomerados con diferente cantidad de elementos, causada por aspectos como la no respuesta, distribución desigual de los elementos pertenecientes a subclases en los conglomerados, o la planificación de conglomerados de igual tamaño a partir de conglomerados con tamaño real distinto (Kish, L., 1972).

La situación descrita ocasiona una serie de limitantes, como es la naturaleza aleatoria del tamaño de muestra, lo que genera problemas de costo y cálculo de varianza; además, la media muestral calculada

como una razón, a pesar de sus bondades prácticas, se convierte en un estimador sesgado de la media poblacional. En el caso de la varianza, las fórmulas utilizadas para el cálculo adquieren mayor complejidad en su desarrollo y corresponden a estimaciones sesgadas; sin embargo, bajo buenos diseños de muestreo representan aproximaciones aceptables de la varianza poblacional. Por otra parte, el control del tamaño de muestra no resulta práctico; aun así, controles razonables sobre la variabilidad de su tamaño adquieren especial importancia a partir de un manejo adecuado del coeficiente de variación, y esto implica que, para un número fijo de conglomerados, dicho coeficiente puede reducirse a partir de diferentes técnicas de selección, entre las que puede mencionarse el submuestreo con probabilidades proporcionales al tamaño de los conglomerados, división y combinación de conglomerados, estratificación con base en el tamaño y selección pareada (Kish, L., 1972).

Cochran (1976) también menciona el uso del muestreo con probabilidades proporcionales al tamaño y su extensión como parte del muestreo estratificado. Por su parte, Kalton (1991), en concordancia con lo mencionado, indica que el procedimiento usualmente utilizado en estos casos corresponde al muestreo con probabilidades proporcionales al tamaño (PPT) o al tamaño estimado (PPTE), debido al desconocimiento del tamaño real de la UPM. La técnica funciona adecuadamente siempre y cuando se logren estimaciones razonables de los tamaños; además, la aplicación de fracciones de submuestreo inversamente proporcionales a los tamaños estimados dentro de las UPM produce, en general, muestras con igual probabilidad de selección.

Como consecuencia de lo indicado, Kalton (1991) hace énfasis en un aspecto de importancia en los diseños complejos multietápicos, relacionado con la asignación de la muestra entre la primera y segunda etapa en función del número adecuado de UPM a seleccionar y cantidad de elementos a submuestrear dentro de cada una de las UPM. En el caso de personas en tránsito, aspectos como los procedimientos del trabajo de campo y la naturaleza del flujo dentro de la UPM afectan dicha decisión. En relación con este aspecto, un objetivo importante en el diseño de muestreos complejos para poblaciones móviles consiste en un uso óptimo del recurso de campo asignado a cada UPM escogida, manteniendo una muestra probabilística de elementos durante los intervalos de tiempo seleccionados, lo cual se encuentra relacionado con el procedimiento de submuestreo utilizado.

Como parte del procedimiento de submuestreo, Kalton (1991) menciona el uso de la selección sistemática de elementos como técnica probabilística y, además, sugiere procedimientos alternativos para facilitar labores de campo, como la selección sistemática de intervalos de tiempo en los cuales se escoge el primer elemento en ingresar al intervalo definido, así como también menciona la opción de seleccionar el siguiente elemento una vez finalizada la entrevista u observación. Sin embargo, se

aclara que, aunque estos procedimientos alternativos faciliten las tareas de trabajo de campo, producen muestras no probabilísticas, con el riesgo de sesgo de selección que dichos procedimientos implican.

Kish (1972) sugiere una serie de procedimientos de submuestreo que pueden escogerse y aplicarse en función de la naturaleza del estudio. Una síntesis de estos procedimientos se brinda a continuación:

- Selección sistemática de elementos a partir del intervalo $F_b = 1/f_b$ una vez realizado el correspondiente arranque aleatorio definido de 1 a F_b y tomado por separado para cada unidad primaria de muestreo.
- Calcular el tamaño esperado de la muestra $x'_\alpha = f_b x_\alpha^*$. En el caso de fracciones debe utilizarse algún esquema para un valor entero de x_α , con el objetivo de seleccionar x_α elementos al azar. El valor x_α^* debe conocerse de antemano y su aplicación en la práctica se torna más compleja respecto al caso sistemático.
- Puede dividirse cada unidad primaria de muestreo en F_b unidades secundarias y escoger una mediante selección aleatoria. Otra opción consiste en dividir la unidad primaria en $2F_b$ unidades para seleccionar dos y de esta forma permitir el cálculo de la varianza dentro de la unidad primaria de muestreo o, por último, el caso general, en el que se divide la unidad primaria en kF_b unidades para seleccionar “ k ” al azar.
- Realizar una selección de la submuestra en dos o más etapas a partir de cualquier combinación de fracciones de submuestreo donde su producto sea igual a f_b .

En relación con los elementos conceptuales abordados, existen diferentes estudios de carácter observacional en los que se emplean diseños muestrales complejos de varias etapas para el análisis de características de interés a partir del estudio de poblaciones móviles, como por ejemplo el estudio realizado por Lammers, Agent y Green (2019), del Centro de Transporte de la Universidad de Kentucky, en el que se utilizó un diseño complejo bajo los lineamientos de la NHTSA⁷ para seleccionar condados y lugares donde se aplica el estudio para medir el uso de cinturón de seguridad. Se seleccionaron aleatoriamente 15 UPM (condados) de un total de 77 casos posibles para obtener representatividad geográfica. Posteriormente se determinaron 150 puntos de muestreo con base en las millas-vehículo recorridas por UPM, los cuales se asignaron por tipo de autopista (estratos:

⁷ **National Highway Traffic Safety Administration (NHTSA)**: Corresponde a una dependencia del gobierno de los Estados Unidos, la cual forma parte del Departamento de Transporte y tiene como misión la prevención y reducción de accidentes de tránsito.

primaria, secundaria y local) en cada condado. Los accesos y carriles en cada punto de muestreo se identificaron y seleccionaron de forma aleatoria.

Los tiempos de observación para los 150 puntos de muestreo se asignaron aleatoriamente y se utilizó una hora de medición en cada sitio con uno o dos observadores dependiendo del número de direcciones del flujo vehicular. Las observaciones se realizaron entre 7:00 a.m. y 6:00 p.m., considerando la semana completa. Las horas de inicio se escalonaron para tener representatividad de los sitios por día de la semana y hora del día. En ubicaciones de alto flujo vehicular se aplicó selección aleatoria de vehículos, mientras que en lugares de bajo flujo vehicular se realizaron mediciones para totalidad de vehículos que circularon por el punto.

Turner y Hagelin (2001), de la Universidad del Sur de la Florida, utilizaron un diseño muestral multietápico como parte del estudio de observaciones sobre uso de casco en motociclistas para determinar los condados a muestrear en una primera etapa. En una segunda etapa determinaron los lugares de observación dentro de cada condado, y en la tercera definieron las ubicaciones específicas para cada lugar de observación. Por último, en la cuarta etapa, definieron los días y horas de observación. Como parte del diseño incorporan estratificación, para incrementar la precisión de las estimaciones muestrales para un tamaño de muestra dado, de acuerdo con la población, cantidad de motocicletas registradas, millas diarias recorridas y clasificaciones de las carreteras. Además, utilizaron conglomerados para lograr costo-efectividad y eficiencia mediante la agrupación de lugares dentro de las franjas horarias definidas.

Es importante mencionar que los dos diseños anteriores se rigieron bajo los lineamientos metodológicos establecidos por la NHTSA, por lo que presentan gran similitud en sus etapas y procedimientos aplicados. En Estados Unidos, esto aplica para otros estudios observacionales sobre uso de dispositivos de seguridad pasiva desarrollados en diferentes estados, como por ejemplo, los realizados por la Freeman & Schneider (2019) de la Universidad de Louisiana, Risley & Laubach (2021) de la Universidad de Oklahoma, Haas (2016) de Mountain State Criminal Justice Research Service, Vachal & Andersen (2020) de la Universidad de Dakota del Norte y Allen, Fox & Berg (2020) de la Universidad de Iowa.

En la ciudad de Buenos Aires, Argentina, se realizó un estudio observacional sobre factores de riesgo a cargo de Vergara (2016), para estudiar el uso de cinturón de seguridad, de sistemas de retención infantil y de casco, y los factores distractores, con el objetivo de analizar qué aspectos determinan dichas conductas de riesgo. El marco de muestreo se construyó a partir de datos del censo nacional, cartografías censales y viales, así como de registros del parque automotor y motocicletas. Para el

diseño muestral se seleccionaron los puntos de observación a partir de intersecciones viales con semáforo. Para ello, en una primera etapa se identificaron y clasificaron las principales vías de circulación en función de sus jerarquías y sentidos. Para facilitar el acceso al campo y reducir el error de muestreo, en una segunda etapa se procedió a estratificar la ciudad de Buenos Aires en zonas geográficas y, por último, en la tercera etapa, se seleccionaron al azar 45 puntos de muestreo (intersecciones), los cuales se definieron a partir de franjas horarias establecidas (días-horas), con lo que estimó un total de 20 observaciones por día-punto de observación.

En Chile, la Comisión Nacional de Seguridad y Tránsito (CONASET) desarrolló un estudio observacional para conocer las conductas de las personas en carretera en relación con el uso de dispositivos de seguridad pasiva, sistemas de retención infantil y factores distractores. Para el estudio observacional, se seleccionaron puntos de observación bajo criterios de seguridad del lugar, representatividad del tránsito urbano, sin controles policiales en las proximidades. En este estudio se excluyeron las carreteras de alta velocidad y rutas interurbanas, de manera que los puntos de observación se definieron a partir de zonas urbanas únicamente. Para las calles y avenidas, el método de observación en cada ciclo del semáforo consistió en seleccionar los vehículos que se detuvieron en la primera y segunda posición del carril próximo derecho, para las calles y avenidas con doble sentido de circulación se aplicó la misma técnica de selección. Por último, las observaciones se realizaron de lunes a viernes en tres horarios con los mayores flujos vehiculares: 8:00-10:00, 12:00-15:00 y 17:30-19:30 (CONASET, 2015).

En el caso de Bolivia, Puppo et al. (2013) desarrollaron un estudio observacional a nivel nacional para medir ciertas conductas viales de los conductores y ocupantes de vehículos automotores relacionadas con el uso de casco, cinturón y sistemas de retención infantil. Este estudio, al igual que el ejecutado por CANASET (2015), comparte lineamientos metodológicos en el levantamiento de observaciones, como parte de acuerdos definidos entre los países miembros del Observatorio Iberoamericano de Seguridad Vial (OISEVI). Por lo tanto, el diseño de muestreo consiste en un relevamiento observacional, probabilístico, estratificado por ciudades, según la cantidad de habitantes. Se seleccionaron puntos de observación representativos del tránsito urbano y se consideraron aspectos de seguridad para una correcta medición, contabilizando 31 puntos de muestreo en 10 localidades. Se trabajó a partir de franjas horarias, separando los autos de las motocicletas, y la selección del vehículo se realizó aleatoriamente. Por último, similar al estudio realizado por CONASET-Chile, en las observaciones en calles y avenidas con carriles de uno más sentidos de circulación se seleccionaron los vehículos que frenaron en la primera y segunda posición del carril próximo izquierdo. Adicionalmente, en calles sin semáforo, según la franja horaria, se

seleccionó cada dos minutos el primer vehículo, mientras que en zonas de accesos se seleccionó el primer vehículo cada cinco minutos.

En Costa Rica, los abordajes realizados para la medición de aspectos relacionados con el uso de dispositivos de seguridad pasiva han estado a cargo de dos entidades: el Observatorio del Desarrollo (OdD) de la Universidad de Costa Rica y el Consejo de Seguridad Vial (Cosevi), como entidad adscrita al Ministerio de Obras Públicas y Transportes (MOPT). En el caso del OdD, se implementó un estudio observacional exploratorio para identificar riesgos potenciales en los motociclistas con la finalidad de generar información relevante en la planificación de estrategias nacionales en seguridad vial. El estudio se basó en diseño exploratorio a partir de observación no participante, donde la población de estudio se definió a partir de los motociclistas no oficiales que circularon en los distritos de Curridabat, Desamparados y Alajuela, en el periodo establecido entre el 24 y el 30 de julio de 2016. La unidad de estudio correspondió a los motociclistas y la unidad de muestreo a los tramos de carretera. Para definir esta última unidad en mención se construyó el marco muestral, debido a la falta de listados para este tipo de estudios. Por causa de limitaciones de tipo financiero, tiempo y personal, la delimitación del marco se basó en los tres distritos indicados anteriormente.

Para seleccionar la muestra se utilizó un diseño polietápico, seleccionando cantones, en una primera etapa, en función de la mayor proporción de motocicletas y ocurrencia de accidentes, aunque posteriormente, debido a limitaciones financieras, se utilizaron los distritos primeros de los cantones (Curridabat, Desamparados y Alajuela. En una segunda etapa, se seleccionaron los tramos de carretera dentro de cada distrito, los cuales se asignaron a las franjas horarias de forma sistemática. Para la tercera etapa, se seleccionó al primer motociclista en llegar al punto de observación y, en los casos de llegadas simultáneas, se desarrolló un procedimiento de selección sistemático alternando la observación en función del orden de llegada de derecha a izquierda. El submuestreo por tramo-hora de observación se estableció en 16 motociclistas (OdD, 2016).

En el caso del Cosevi, se han realizado estudios para medir el uso de dispositivos de seguridad pasiva como parte del análisis de los factores de riesgo durante la conducción en carretera en los años 2014, 2016, 2018 y 2020. Para los estudios del 2014 y 2016, se definieron segmentos de un kilómetro de carretera como parte del marco muestral y se aplicaron mediciones de lunes a viernes, en franjas de cinco minutos, distribuidas cada media hora, entre 6:00 a.m. y 6:00 p.m. (Cosevi, 2017)⁸.

⁸ Para mayor detalle de la construcción del marco y diseño muestral, consultar apartado metodológico del presente documento.

Para los estudios del 2018 y 2020, se realizaron mejoras al diseño metodológico de los años anteriores, diseñando un nuevo marco de muestreo basado en intersecciones de carretera entre rutas nacionales y modificando el diseño. En este caso puede definirse el diseño como probabilístico complejo de poblaciones móviles, en el que las UPM correspondieron a la combinación intersección-franja horaria (conglomerados desiguales). Las intersecciones de la muestra se definieron a partir de selección sistemática, ordenando según la cantidad de intersecciones por cantón; asimismo, las franjas horarias se distribuyeron de 6:00 a.m. a 6:00 p.m. en seis periodos de una hora; cada franja de una hora se subdivide en dos periodos de media hora para realizar las mediciones por tipo de vehículo (carro o motocicleta). En la segunda etapa de muestreo se seleccionan los vehículos; la técnica consistió en escoger el vehículo que presentara las mejores condiciones de visibilidad al llegar al punto de muestreo para realizar el levantamiento de las características de interés y, una vez finalizada la observación de determinado vehículo, se continuó con el siguiente hasta completar la franja horaria (Cosevi, 2021).

A partir de la revisión de literatura, no se encontraron otros acercamientos relacionados con el estudio de poblaciones móviles donde se aplicaran análisis comparativos de técnicas de selección de elementos en determinada etapa de los diseños empleados; dichos estudios se enfocan en el planteamiento de diseños de muestreo de varias etapas a partir de técnicas de selección que mejor se adecuen a sus realidades, más allá de analizar alguna técnica que brinde mejor rendimiento en la reducción de variabilidades como parte de la estimación de estadísticos de interés.

2.3. Metodología

Este apartado contiene los principales aspectos metodológicos abordados como parte del análisis desarrollado en el presente trabajo. Como primer punto se explican la población de estudio, el marco de muestreo, el diseño muestral, el trabajo de campo y variables de interés, correspondientes al diseño del estudio observacional ejecutado en 2016 para la medición del uso de dispositivos de seguridad pasiva por parte del Cosevi. Además, se incluye el referente teórico de las técnicas estadísticas utilizadas y el detalle del análisis empleado.

Es importante aclarar que los datos recolectados en este estudio fueron obtenidos con base en otros objetivos y no con los definidos en el presente documento de práctica profesional.

2.3.1. Objetivos

Objetivo general

Analizar el efecto que tienen cambios al diseño muestral del estudio observacional en Costa Rica 2016 en flujos vehiculares medios y altos para disminuir la variabilidad de las estimaciones relacionadas con el uso de dispositivos de seguridad pasiva.

Objetivos específicos

- Definir variables e indicadores que permitan valorar el efecto que tienen en las estimaciones, los procesos de simulación en relación con el uso de conglomerados completos.
- Establecer procesos de simulación para analizar el efecto que tendrán, en la segunda etapa de muestreo, diferentes fracciones de muestreo, técnicas de selección, procedimientos de submuestreo y posestratificación en la estimación del uso de dispositivos de seguridad pasiva para determinar si existen ganancias en la precisión y exactitud de las estimaciones.

2.3.2. Datos

El trabajo tiene como fuente de datos los resultados del *Estudio nacional para la medición del uso de dispositivos de seguridad pasiva (casco y cinturón) y factores distractores 2017*. El estudio estuvo a cargo del Área de Investigación y Estadística del Consejo de Seguridad Vial y se ha desarrollado con una periodicidad de dos años desde el 2014. Para la presente investigación, se utilizan los datos del año 2016 (Cosevi, 2017).

Sin embargo, es importante profundizar en la explicación de ciertas limitantes presentadas en la etapa de análisis a partir del conjunto de datos indicado; esto debido a la naturaleza del fenómeno estudiado y el diseño de muestreo utilizado. En el siguiente subapartado se detallan los principales aspectos de la problemática de análisis mencionada.

2.3.2.1. Estudio de UPM con bajo flujo vehicular

Este apartado evidencia el inconveniente relacionado con las UPM de la muestra que presentaron flujo vehicular reducido. En el primer subapartado se resumen los hallazgos encontrados para el caso del flujo de motocicletas, mientras que en el segundo se estudió el flujo de carros, con la finalidad de establecer la línea de análisis a seguir en el presente trabajo.

2.3.2.1.1 UPM con bajo flujo de motocicletas

Cuadro 8. Distribución absoluta y porcentual de UPM según flujo de motocicletas

| Cantidad de motos | Tramo - franja (UPM) | | | |
|-------------------|----------------------|-------------|--------------------|-----------------|
| | n | n acumulado | % no submuestreado | % submuestreado |
| 0 | 488 | 488 | 42,5 | 57,5 |
| 1 | 115 | 603 | 52,5 | 47,5 |
| 2 | 103 | 706 | 61,5 | 38,5 |
| 3 | 87 | 793 | 69,1 | 30,9 |
| 4 | 68 | 861 | 75,0 | 25,0 |
| 5 | 74 | 935 | 81,4 | 18,6 |
| 6 | 60 | 995 | 86,7 | 13,3 |
| 7 | 41 | 1036 | 90,2 | 9,8 |
| 8 | 43 | 1079 | 94,0 | 6,0 |
| 9 | 24 | 1103 | 96,1 | 3,9 |
| 10 | 21 | 1124 | 97,9 | 2,1 |
| 11 | 16 | 1140 | 99,3 | 0,7 |
| 12 | 7 | 1147 | 99,9 | 0,1 |
| 13 | 1 | 1148 | 100,0 | 0,0 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

El cuadro 8 muestra los datos sobre el flujo de motocicletas que circularon por las UPM seleccionadas y se observa que, de los 1184 tramo-franjas de la muestra, solamente 660 presentaron flujo de motos (55,7%). Además de esta limitante, debe considerarse que en algunas UPM con presencia de flujo la cantidad de motocicletas que circuló fue reducida, dado que el aforo máximo obtenido en la totalidad de puntos estudiados corresponde a trece motocicletas, registradas en una única UPM.

Lo anterior dificulta la comparación de las técnicas de selección mediante la simulación de escenarios, por la incapacidad de poder variar el tamaño del submuestreo, debido a la baja frecuencia de motocicletas en las UPM. Por ejemplo, en escenarios con submuestreos de “b = 8”, aproximadamente 90% de las UPM no cuenta con la cantidad mínima de motocicletas requerida para realizar el estudio de las variabilidades a partir de los indicadores de desempeño. En el apartado 5 de conclusiones y recomendaciones se brindan alternativas a considerar para solucionar la limitante descrita.

2.3.2.1.2 UPM con bajo flujo de carros

Similar al caso de las motocicletas, aunque con menor afectación, se identificaron UPM con bajo flujo vehicular, lo que afecta la estimación de los indicadores de desempeño, limitando las simulaciones. Como se muestra en el cuadro 9, existen 543 UPM del total que presentaron flujos

iguales o inferiores a 20 vehículos. El primer inconveniente con este grupo de UPM está relacionado con el uso completo del aforo de vehículos dentro del conglomerado para ejecutar la simulación cuando se utiliza un determinado tamaño de submuestreo “ b_i ” que supera el tamaño del conglomerado, lo que carece de sentido práctico debido a la imposibilidad de comparar las técnicas de selección correctamente.

La segunda limitante corresponde a la incapacidad de alcanzar la convergencia de los indicadores de desempeño en relación con los valores de referencia definidos a partir de los conglomerados completos (ver cuadro 9). Como se aprecia en la columna “% uso cinturón”, se encuentra estabilidad en valores cercanos a 70% de uso, por lo que, al utilizar el grupo de UPM de bajo flujo, esto contribuye a la subestimación de los valores de referencia cuando se utiliza o submuestra la información completa de los conglomerados.

Debido a la limitante descrita, para el grupo de UPM con presencia de flujo vehicular suficiente se procedió con las simulaciones de los escenarios para comparar las técnicas de selección establecidos (apartado 4.2). En las UPM de bajo flujo, no fue posible comparar técnicas de selección, por lo que, al igual que en el caso de las UPM con bajo flujo de motocicletas, en el apartado 5 de conclusiones y recomendaciones se brindan alternativas a considerar como solución a la limitante descrita.

Cuadro 9. Distribución absoluta y porcentual de UPM según flujo de carros

| Cantidad de carros | Tramo - franja (UPM) | | | | |
|--------------------|----------------------|-------------|--------------------|---------------------|----------------|
| | n | n acumulado | % no submuestreado | % submuestreado | % uso cinturón |
| 1 | 19 | 19 | 1,7 | 98,3 | 69,2 |
| 2 | 31 | 50 | 4,4 | 95,6 | 83,8 |
| 3 | 21 | 71 | 6,2 | 93,8 | 69,6 |
| 4 | 24 | 95 | 8,3 | 91,7 | 75,5 |
| 5 | 31 | 126 | 11,0 | 89,0 | 73,6 |
| 6 | 28 | 154 | 13,4 | 86,6 | 71,9 |
| 7 | 38 | 192 | 16,7 | 83,3 | 72,4 |
| 8 | 42 | 234 | 20,4 | 79,6 | 72,5 |
| 9 | 35 | 269 | 23,4 | 76,6 | 71,8 |
| 10 | 33 | 302 | 26,3 | 73,7 | 70,6 |
| 11 | 25 | 327 | 28,5 | 71,5 | 70,7 |
| 12 | 33 | 360 | 31,4 | 68,6 | 71,2 |
| 13 | 31 | 391 | 34,1 | 65,9 | 70,6 |
| 14 | 29 | 420 | 36,6 | 63,4 | 70,1 |
| 15 | 21 | 441 | 38,4 | 61,6 | 70,4 |
| 16 | 29 | 470 | 40,9 | 59,1 | 71,1 |
| 17 | 12 | 482 | 42,0 | 58,0 | 71,3 |
| 18 | 25 | 507 | 44,2 | 55,8 | 70,5 |
| 19 | 16 | 523 | 45,6 | 54,4 | 70,1 |
| 20 | 20 | 543 | 47,3 | 52,7 | 69,9 |
| 21 o más | 605 | 1148 | 0,0 ^{a/} | 100,0 ^{a/} | 78,5 |

a/: Debe considerarse que, el porcentaje máximo submuestreado en los procesos de simulación, corresponde a 20 elementos por UPM, por lo que, para estas UPM con 21 carros o más, no existen problemas de submuestreo, lo cual explica los valores porcentuales en las columnas "% no submuestreado" y "% submuestreado"; sin embargo, se agregan para completar el cuadro con las UPM restantes de la muestra.

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

2.3.2.2. Población de estudio

La población que se investiga en el estudio sobre el uso de dispositivos de seguridad pasiva está definida por los ocupantes de vehículos automotores que circulan en tramos de carretera de las rutas nacionales de medio y alto flujo vehicular a lo largo del territorio nacional, en franjas horarias de cinco minutos, distribuidas entre 6:00 a.m. y 4:00 p.m. Este estudio excluye los casos de vehículos de transporte colectivo (transporte público, transporte de estudiantes, etc.).

2.3.2.3. Marco muestral

El estudio tiene como marco de muestreo la red vial de rutas nacionales, con una longitud de 7694 kilómetros, excluyendo las rutas cantonales. El marco se encuentra segmentado en tramos de 1 kilómetro, asumiendo que la probabilidad de cambio en el comportamiento de los usuarios en una distancia menor es baja.

Para el presente estudio, la UPM está definida por una combinación de espacio y tiempo en los que se realizaron las observaciones; en este caso se cuenta con 7694 tramos y 96 posibles franjas horarias por tramo (ver detalle en apartado 3.2.4 “Diseño muestral”), por lo que el marco final de muestreo estaría definido por 738624 tramo-franjas.

2.3.2.4. Diseño muestral

En cada tramo de carretera se realizaron las observaciones durante un día en franjas horarias distribuidas en 8 horas, de 06:00 a.m. a 04:00 p.m. (06:00 a.m.-08:00 a.m., 09:00 a.m.-01:00 p.m., y 02:00 p.m.-04:00 p.m.). Cada hora estuvo dividida en franjas de 30 minutos y, dentro de cada una de estas franjas, se generaron 6 subgrupos de 5 minutos, de los cuales se seleccionó uno al azar dentro de dicha franja, esperando que ese subgrupo fuera representativo de la media hora, con lo que se estaría seleccionado por día-tramo un total de 16 franjas de 5 minutos.

Con base en lo anterior, en una primera etapa de muestreo se seleccionaron 74 tramos de carretera de forma aleatoria, los cuales se encuentran distribuidos en los 81 cantones que conformaban al país en 2017 y, posteriormente, dentro de cada tramo se seleccionaron 16 franjas horarias, para un total de 1184 UPM. Cada combinación de tramo y franja (UPM) representó un conglomerado desigual de unidades de estudio, que fue asignado aleatoriamente al día de la semana, esperando distribución balanceada de lunes a viernes.

A partir de lo mencionado, el diseño muestral se puede definir como probabilístico bietápico de poblaciones móviles y de conglomerados desiguales. Además, el registro de los datos de interés se hizo mediante cámara de video, donde el observador grabó la totalidad de los vehículos que circularon en cada franja horaria de 5 minutos. La información digital producto de las grabaciones se extrajo, posteriormente, en un archivo de datos.

2.3.2.5. Trabajo de campo

El trabajo de campo se realizó con personal del Área de Investigación y Estadística de Cosevi, mediante lineamientos relacionados con el debido posicionamiento de los observadores en los puntos de muestreo. Para obtener datos de calidad se tomaron en cuenta, en la medida de lo posible, las condiciones del entorno vial y en otras ocasiones, por cuestiones climatológicas, el punto de muestreo debió visitarse en dos ocasiones para realizar las grabaciones.

Como parte de los lineamientos se procedió a identificar, dentro del tramo de carretera, la existencia de posibles elementos en la infraestructura vial que facilitarían la visibilidad para aplicar las grabaciones, como semáforos, curvas y reductores de velocidad, entre otros.

La duración del trabajo de campo se planificó en cuatro meses, aproximadamente, en función del recurso humano disponible, respetando, además, la asignación del día a cada punto de muestreo.

2.3.2.6. Variables de interés

En el planteamiento y desarrollo de las simulaciones para analizar las técnicas de selección, se utilizaron los datos ya mencionados. A continuación, se presenta el resumen de las variables utilizadas como parte del análisis comparativo entre técnicas de selección:

flujo_personas

Corresponde a la cantidad de ocupantes de los carros que circularon por la UPM.

uso_cinturon

Corresponde a la cantidad de ocupantes de carros que circularon por la UPM y utilizaron el cinturón de seguridad.

flujo_carros

Corresponde a la cantidad de carros que circularon por la UPM.

2.3.3. Técnicas estadísticas

El objetivo principal del trabajo consiste comparar técnicas de selección de elementos para estudiar la variabilidad en la estimación de indicadores relacionados con el uso de dispositivos de seguridad pasiva, y dicho análisis se desarrolló usando procesos de simulación y técnicas estadísticas de agrupamiento de casos. Los diferentes aspectos relacionados con las simulaciones y técnicas estadísticas se detallan en los siguientes apartados.

2.3.3.1. Técnicas de selección

Como se mencionó en el apartado introductorio, la aplicación de una determinada técnica de selección de elementos a lo interno de cada UPM no fue un aspecto metodológico considerado como parte del procedimiento de muestreo del estudio desarrollado por el Cosevi, por lo que interesa comprobar, mediante simulaciones, si existen ganancias en variabilidad para las estimaciones del uso de dispositivos de seguridad pasiva en función de la técnica de selección, o si resulta irrelevante la técnica aplicada. Las técnicas de selección aplicables en la práctica, y que son sometidas a comparación en el presente trabajo, se detallan seguidamente:

Selección aleatoria: Consiste en seleccionar al azar un número específico “b” de vehículos que circulan por determinado tramo de carretera y franja horaria de medición para realizar el levantamiento de características relacionadas con el uso de dispositivos de seguridad pasiva. En la práctica se requieren mediciones previas del flujo vehicular que circula por el punto de muestreo, para conocer su valor medio aproximado y determinar la viabilidad de aplicación de la técnica.

Selección sistemática: Para un número específico “b” de vehículos, la técnica consiste en seleccionar los elementos con un determinado espaciamiento. Dicho espaciamiento se calcula como el cociente del número “n” de vehículos requeridos y el tamaño o flujo vehicular aproximado de la UPM (tramo y franja horaria), como lo indica la siguiente fórmula:

$$k = \frac{b}{B} \quad (2.1)$$

Donde “k” corresponde al espaciamiento, “b” al número de vehículos requeridos, y “B” al tamaño de la UPM particular en la que se realizan las observaciones. Posteriormente, se selecciona de forma aleatoria un número entre 1 y k para iniciar con la selección de vehículos y dicho valor escogido sería el arranque, con lo que el siguiente valor seleccionado estaría definido por el arranque más el valor del espaciamiento “k”, y así sucesivamente hasta completar la cantidad planeada de vehículos. En la práctica se requieren mediciones previas del flujo vehicular que circula por el punto de muestreo para valorar la viabilidad de aplicación de la técnica.

Selección de los primeros “n” casos: Esta técnica es la más simple de las tres opciones analizadas y consiste en seleccionar los primeros “n” vehículos que circulan por la UPM, para realizar las observaciones relacionadas con las características de interés.

2.3.3.2. Indicadores estadísticos

Como parte de los procesos para la generación de información oportuna y relevante a partir de datos cuantitativos, los indicadores estadísticos representan una herramienta fundamental en cualquier ámbito que se utilicen para análisis y evaluación de metas y objetivos planteados a partir de evidencia objetiva. Por tal motivo, es importante conocer su significado y principales características.

Como indica Horn (1999), citado por Mondragón (2002), “Los indicadores sociales son estadísticas, serie estadística o cualquier forma de indicación que nos facilita estudiar dónde estamos y hacia dónde nos dirigimos con respecto a determinados objetivos y metas, así como evaluar programas específicos y determinar su impacto” (p. 52). Además, como menciona Mondragón (2002), existe una serie de elementos que caracteriza a un indicador estadístico:

- Estar inscrito en un marco teórico o conceptual que le permita asociarse firmemente con el evento al que el investigador pretende dar forma.
- Ser específico, es decir, estar vinculado con los fenómenos económicos, sociales, culturales o de otra naturaleza sobre los que se pretende actuar; por lo anterior, se debe contar con objetivos y metas claros, para poder evaluar qué tan cerca o lejos nos encontramos de los mismos y proceder a la toma de decisiones pertinentes.
- Ser explícito, de tal forma que su nombre sea suficiente para entender si se trata de un valor absoluto o relativo, de una tasa, una razón, un índice, etc., así como a qué grupo de población, sector económico o producto se refiere, y si la información es global o está desagregada por sexo, edad, años o región geográfica.
- Estar disponible para varios años, con el fin de que se pueda observar el comportamiento del fenómeno a través del tiempo, así como para diferentes regiones y/o unidades administrativas.
- Ser relevante y oportuno para la aplicación de políticas, describiendo la situación prevaleciente en los diferentes sectores de gobierno, permitiendo establecer metas y convertirlas en acciones.
- No ser exclusivo de una acción específica; uno puede servir para estimar el impacto de dos o más hechos o políticas, o viceversa.
- Ser claro, de fácil comprensión para los miembros de la comunidad, de forma que no haya duda o confusión acerca de su significado, y debe aceptársele, por lo general, como expresión del fenómeno a ser medido.
- Que la recolección de la información permita construir el mismo indicador de la misma manera y bajo condiciones similares, año tras año, de modo que las comparaciones sean válidas.
- Ser técnicamente sólido, es decir, válido, confiable y comparable, así como factible, en términos de que su medición tenga un costo razonable.
- Ser sensible a cambios en el fenómeno, tanto para mejorar como para empeorar.
- Si bien su selección no debe depender de la disponibilidad de información proveniente de encuestas, censos y/o registros administrativos, sino de los objetivos fijados en los programas y proyectos de gobierno (dentro de los cuales se inscribe), es importante considerar el costo-beneficio del tiempo y los recursos necesarios para su construcción, por lo que, de preferencia, debe ser medible a partir del acervo de datos recopilados.

Una vez realizado el abordaje general sobre el significado de indicadores estadísticos y sus principales características, es importante definir los conceptos correspondientes a *precisión* y *exactitud* desde el enfoque estadístico; esto debido a que, dichos elementos son la base para medir el rendimiento o desempeño de las estimaciones que se obtienen como resultado de los procesos de simulación y los indicadores que se definirán en el siguiente apartado.

Desde el enfoque del muestreo, “las muestras precisas tienen errores estándar bajos” (Kish, 1972, p. 48), lo cual está determinado por el grado de dispersión del conjunto de valores con respecto a su valor medio. Asimismo, como indica Kish (1972), “las muestras exactas tienen errores totales bajos que incluyen los sesgos y los errores no de muestreo, junto con la variabilidad del muestreo” (p. 48). En este caso, corresponde a la magnitud de la diferencia entre el estimador y el valor real.

2.3.2.1.1. Indicadores de desempeño en la estimación del uso de dispositivos de seguridad pasiva

Debido a que se requiere identificar la técnica de selección de elementos que presenta el mejor rendimiento para estimar estadísticos relacionados con el uso de dispositivos de seguridad pasiva, se hace necesario utilizar indicadores que permitan estudiar el comportamiento de las estimaciones obtenidas con cada técnica de selección en términos de su exactitud y precisión, y así determinar cuál presenta las menores diferencias con respecto a esos mismos estadísticos obtenidos a partir de los conglomerados completos. Con base en esto, los indicadores de desempeño utilizados para realizar dichas comparaciones corresponden a la proporción media de uso de dispositivos de seguridad pasiva, el error estándar de la proporción media, el efecto del diseño y el coeficiente de variación.

En la estimación de la proporción media de uso de dispositivos de seguridad pasiva, el denominador se convierte en una variable aleatoria, debido a los tamaños desiguales en cada conglomerado de elementos, por lo que, en este caso, la proporción media se define como una razón, como resultado del cociente entre dos variables aleatorias; la cantidad de casos con la característica de interés en el conglomerado i -ésimo y la cantidad de casos en el conglomerado i -ésimo. Por lo tanto, para el análisis desarrollado, se considera que $p = r$ en la muestra y $P = R$ en los conglomerados completos. La tabla 2 muestra la fórmula de cálculo de cada indicador para la muestra (a partir de los procesos de simulación) y los conglomerados completos.

Tabla 2. Fórmulas de cálculo de los indicadores de desempeño utilizados en los procesos de simulación para muestra y los conglomerados completos

| Indicador | Conglomerados completos | Muestra |
|----------------------------------|---|--|
| Proporción media | $R = \frac{Y}{X} = \frac{1}{X} \sum_{\alpha} \sum_{\beta}^{X_{\alpha}} Y_{\alpha\beta} = \frac{1}{X} \sum_{\alpha} Y_{\alpha} = \frac{\sum_{\alpha}^A Y_{\alpha}}{\sum_{\alpha}^A X_{\alpha}}$ <p>Donde:</p> <p>R: razón media poblacional correspondiente al uso de cinturón de seguridad.</p> <p>A: cantidad de conglomerados en la población.</p> <p>Y_{α}: cantidad de elementos que cumplen la característica de interés (uso de cinturón) en el conglomerado α-ésimo de la población.</p> <p>X_{α}: cantidad de elementos en el conglomerado α-ésimo de la población.</p> | $r = \frac{y}{x} = \frac{1}{x} \sum_{\alpha} \sum_{\beta}^{x_{\alpha}} Y_{\alpha\beta} = \frac{1}{x} \sum_{\alpha} y_{\alpha} = \frac{\sum_{\alpha}^a y_{\alpha}}{\sum_{\alpha}^a x_{\alpha}}$ <p>Donde:</p> <p>r: razón media muestral correspondiente al uso de cinturón de seguridad.</p> <p>a: cantidad de conglomerados seleccionados en la muestra.</p> <p>y_{α}: cantidad de elementos seleccionados que cumplen la característica de interés (uso de cinturón) en el conglomerado α-ésimo de la muestra.</p> <p>x_{α}: cantidad de elementos seleccionados en el conglomerado α-ésimo de la muestra.</p> |
| Error estándar | $ee(R) = \sqrt{var(R)} = \sqrt{\frac{1-f}{X^2} \frac{A}{A-1} (\sum Y_{\alpha}^2 + R^2 \sum X_{\alpha}^2 - 2R \sum X_{\alpha} Y_{\alpha})}$ | $ee(r) = \sqrt{var(r)} = \sqrt{\frac{1-f}{x^2} \frac{a}{a-1} (\sum y_{\alpha}^2 + r^2 \sum x_{\alpha}^2 - 2r \sum x_{\alpha} y_{\alpha})}$ |
| Coefficiente de variación | $CV(R) = \frac{ee(R)}{R}$ | $CV(r) = \frac{ee(r)}{r}$ |
| Efecto del diseño | $efd(R) = \frac{var(diseño complejo)}{var(mia)}$ $efd(R) = \frac{\frac{1-f}{X^2} \frac{A}{A-1} (\sum Y_{\alpha}^2 + R^2 \sum X_{\alpha}^2 - 2R \sum X_{\alpha} Y_{\alpha})}{\frac{p(1-p)}{n}}$ | $efd(r) = \frac{var(diseño complejo)}{var(mia)}$ $efd(r) = \frac{\frac{1-f}{x^2} \frac{a}{a-1} (\sum y_{\alpha}^2 + r^2 \sum x_{\alpha}^2 - 2r \sum x_{\alpha} y_{\alpha})}{\frac{p(1-p)}{n}}$ |

Nota: Debido a los tamaños desiguales en cada conglomerado de elementos, la proporción media se define como una razón, como resultado del cociente entre dos variables aleatorias.

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

2.3.3.3. Escenarios de los procesos de simulación

El trabajo a desarrollar en este documento consiste en simular escenarios en la segunda etapa de muestreo para estimar las variabilidades relacionadas con el uso de dispositivos de seguridad pasiva en ocupantes de vehículos motorizados, a partir de diferentes técnicas de selección, cantidad de elementos (submuestreo) y estratos (grupos de tramos según flujo vehicular). Para valorar las diferencias en las estimaciones, se estudian las variabilidades y se comparan con los valores proporcionados por los resultados originales (conglomerados completos), utilizando los indicadores de desempeño mencionados.

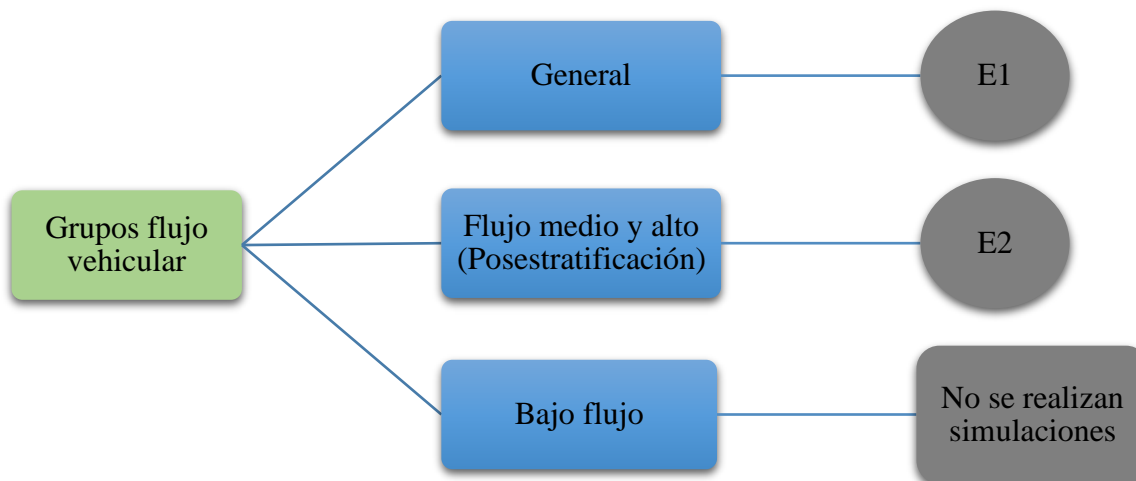
El proceso descrito anteriormente se desarrolla programando un algoritmo en R que considera los escenarios para cada variable, técnica de selección, submuestreo y réplicas. Además, el código empleado toma en cuenta la forma en la que se recolectaron los datos y, por ende, la estructura del archivo, lo cual implica que, para los diferentes escenarios, la selección de los elementos se realiza considerando la franja horaria y el correspondiente tramo de carretera en el cual se realizaron las observaciones, con el fin de obtener las estimaciones de forma correcta.

Las estimaciones dependen del tipo de diseño utilizado y requieren calcular ponderaciones para la primera etapa (definida por el inverso de la probabilidad de selección de los tramos de red vial) y segunda etapa de muestreo (definida por el inverso de la probabilidad de selección de los vehículos observados). Sin embargo, dichos cálculos se encuentran en función de la técnica de selección simulada y el submuestreo realizado en la segunda etapa para cada escenario. Además, debe considerarse que se está trabajando con conglomerados desiguales (combinación de tramo y franja), de tamaño variable, lo que implica probabilidades de selección diferenciadas en los elementos y, por lo tanto, deben realizarse los ajustes correspondientes en las fracciones de muestreo de los diferentes conglomerados para los escenarios propuestos, con el fin de obtener un procedimiento que otorgue probabilidades iguales de selección. Para realizar los ajustes mencionados, debe especificarse el diseño de muestreo de forma adecuada.

Como apoyo para complementar la explicación en los párrafos anteriores, se recurre a los siguientes esquemas para visualizar los escenarios y el detalle a lo interno del proceso de simulación. En la figura 1 se observa que el estudio de los indicadores de desempeño relacionados con el uso de dispositivos de seguridad pasiva, en su primer nivel, abarca los grupos de flujo vehicular, a partir de los cuales se desarrollan los escenarios donde se aplicarán los diferentes procesos de simulación en función de las técnicas de selección, indicadores de desempeño, tamaños de submuestreo y cantidad de réplicas.

Como se aprecia, se plantea un primer escenario general en el que se realizan las simulaciones para todas las UPM de forma conjunta, exceptuando las de bajo flujo vehicular, por las limitantes descritas en el sección 2.3.2.1. En un segundo escenario, se aplican las simulaciones en los grupos de medio y alto flujo vehicular, producto del proceso de posestratificación.

Figura 1. Estructura de los escenarios para la aplicación de simulaciones en la estimación de indicadores sobre uso de dispositivos de seguridad pasiva



Fuente: Consejo de Seguridad Vial (2106). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

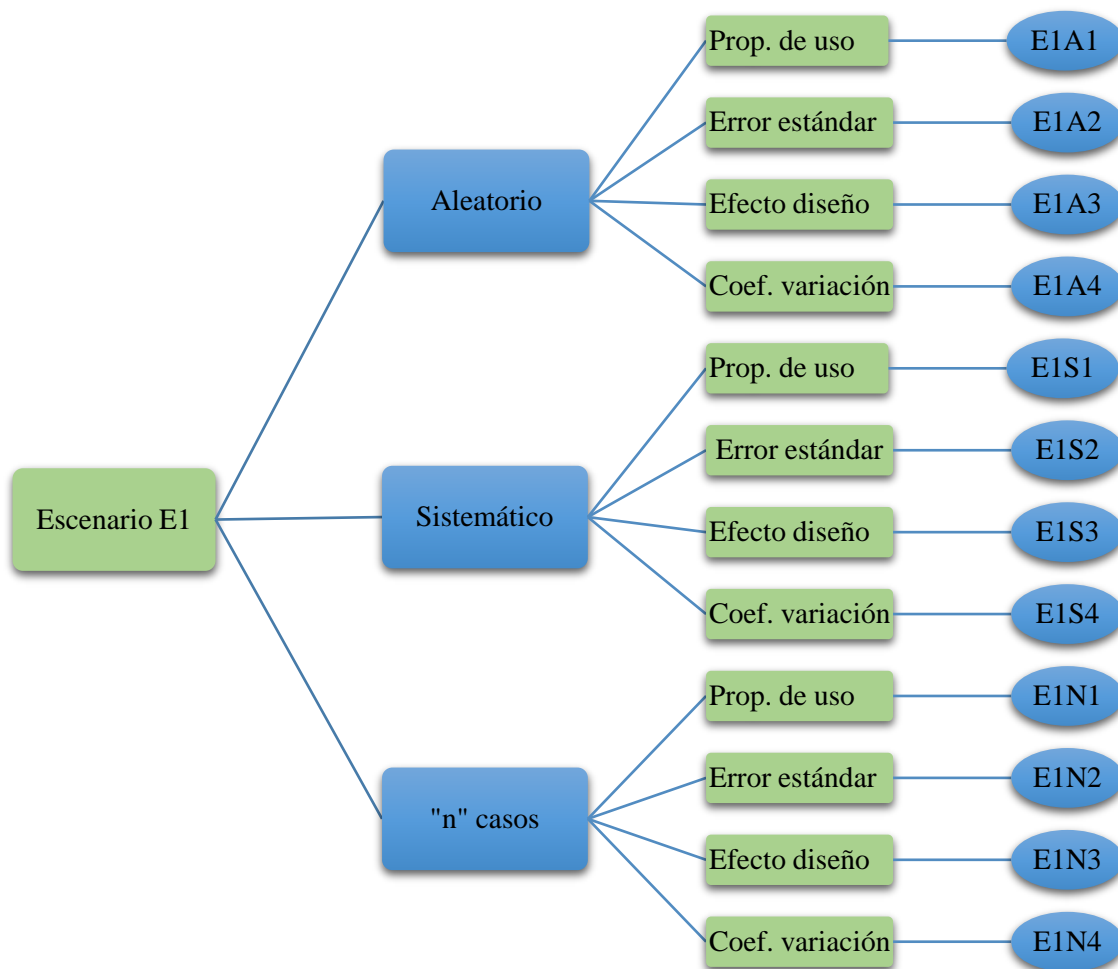
El proceso de simulación ejecutado en los dos escenarios es equivalente, por lo que en el siguiente esquema (figura 2) se aborda solamente uno de ellos para detallar los aspectos relacionados con la comparación de las técnicas de selección.

En el escenario “E1”, correspondiente al grupo general de flujo vehicular, se aplican las tres técnicas de selección de vehículos sometidas a comparación, a saber: aleatoria, sistemática y los primeros “n” casos. A partir de elementos seleccionados con cada una de las técnicas, se obtienen las estimaciones de los indicadores de desempeño con diferentes variaciones en el tamaño del submuestreo y cantidad de réplicas, representadas en la figura 2 por las elipses en cada indicador de desempeño.

Por ejemplo, la combinación “E1A1” corresponde a la cantidad de réplicas utilizadas para estimar la proporción media de uso de cinturón en cada tamaño de submuestreo obtenido con la técnica aleatoria. La mecánica descrita es equivalente para todas las combinaciones de los indicadores de desempeño y las técnicas de selección, por lo que en la siguiente fórmula se ejemplifica únicamente para el caso “E1A1”.

$$E1A1 = BxR = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} (r_1 \quad r_2 \quad \cdots \quad r_j) = \begin{pmatrix} \mathbf{b}_1 & \bar{p}b_1r_1 & \bar{p}b_1r_2 & \cdots & \bar{p}b_1r_j \\ \mathbf{b}_2 & \bar{p}b_2r_1 & \bar{p}b_2r_2 & \cdots & \bar{p}b_2r_j \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_n & \bar{p}b_nr_1 & \bar{p}b_nr_2 & \cdots & \bar{p}b_nr_j \end{pmatrix} = \begin{pmatrix} \bar{p}b_1 \\ \bar{p}b_2 \\ \vdots \\ \bar{p}b_n \end{pmatrix} \quad (2.2)$$

Figura 2 Estructura del escenario “E1” para la aplicación de simulaciones en la estimación de indicadores de desempeño según técnica de selección



Fuente: Consejo de Seguridad Vial (2106). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Con $B = 19$ variaciones de tamaño (desde $b = 2$ hasta $b = 20$) y $R = 100$ réplicas para cada tamaño de submuestreo “ b ”. A partir de las matrices descritas, el valor “ $\bar{p}_{b_n r_j}$ ” corresponde a la estimación de la proporción media de uso de cinturón de seguridad para la réplica “ j -ésima” y el tamaño de submuestreo “ n -ésimo”, mientras que “ \bar{p}_{b_n} ” representa el valor promedio a partir de las estimaciones de la proporción media de uso de cinturón de seguridad calculadas con la totalidad de las réplicas generadas para el submuestreo “ n -ésimo”. Su fórmula de cálculo se muestra a continuación:

$$\bar{p}_{b_n} = \frac{1}{R} \sum_{j=1}^R \bar{p}_{b_n r_j} = \frac{1}{R} (\bar{p}_{b_n r_1} + \bar{p}_{b_n r_2} + \dots + \bar{p}_{b_n r_j}) \quad (2.3)$$

2.3.3.3.1. Especificación de las fracciones de muestreo en el diseño

Al obtener muestras simuladas, a partir de diseños complejos con varias etapas, se genera como parte del proceso de selección de los elementos la correspondiente matriz de probabilidades conjuntas, que posteriormente se utiliza en la definición del diseño para el cálculo de las estimaciones y variabilidades de las características de interés, además del ajuste para los diseños muestrales con probabilidades desiguales.

Los datos utilizados son el resultado de la aplicación de un diseño probabilístico de conglomerados desiguales en dos etapas; sin embargo, no se cuenta con la correspondiente matriz de probabilidades conjunta, debido a que la misma no se calculó durante la etapa de selección de la muestra, por lo que debe recurrirse al uso de algún componente en la definición del diseño para el cálculo de la varianza del estimador mediante la aproximación de la matriz de probabilidades conjuntas o alguna alternativa que no requiera implementarla.

Al no contar con dicha matriz, se proponen dos escenarios para calcular y ajustar las probabilidades de selección, producto del diseño muestral de conglomerados desiguales utilizado en el estudio. Las dos modalidades para el ajuste de las fracciones de muestreo serán comparadas para estudiar su desempeño en el cálculo de las variabilidades de los indicadores de interés (sección 2.4.2.1 de resultados), escogiendo la que presente mejor rendimiento para su aplicación en la etapa de comparación de técnicas de selección de elementos (sección 2.4.2.2 y 2.4.2.3 de resultados).

2.3.3.3.2. Análisis comparativo de las fracciones de muestreo

Como primera opción, se realiza el ajuste de las probabilidades asumiendo muestreo con probabilidad proporcional al tamaño (PPT), tomando como medida de tamaño de la UPM la cantidad de vehículos que circularon al momento de realizar las observaciones. Dicho ajuste se aplica mediante la siguiente fórmula:

$$f = \frac{Mt_{\alpha}}{\frac{\sum Mt_{\alpha}}{a}} \frac{b}{Mt_{\alpha}} = \frac{ab}{\sum Mt_{\alpha}} = \frac{ab}{N} \quad (2.4)$$

Donde:

f: fracción de muestreo general.

a: cantidad de UPM seleccionadas en la primera etapa de muestreo.

b: cantidad de vehículos seleccionados de cada UPM en la segunda etapa de muestreo.

Mt_α: total de vehículos que circularon por cada UPM durante el periodo de observación.

$\sum Mt_{\alpha} = N$: cantidad de vehículos que circularon por la totalidad de UPM de la población.

De la ecuación (3.4), el valor correspondiente a “N” se desconoce en la práctica, por lo que es necesario aproximarlos a partir de la información disponible para lograr el ajuste de la fracción de muestreo. La estimación de “N” se obtiene a partir de la información de la tabla 3.

Tabla 3. Aproximación del total de vehículos en la población de estudio

| Elementos para aproximar la población de vehículos | Valor |
|---|-------------------|
| Mediana de vehículos por franja de 5 minutos de observación ^{2/} . | 22 |
| Cantidad de franjas de 5 minutos en cada hora. | 12 |
| Cantidad de horas a las que se realiza inferencia ^{1/} | 8 |
| Cantidad de tramos que conforman la red vial nacional. | 7 694 |
| Aproximación de la cantidad de vehículos ($\sum M\alpha = N$). | <u>16 249 728</u> |

1/ Horas: 06:00-08:00, 09:00-13:00,14:00-16:00

2/ Se utiliza la mediana debido a la asimetría de la variable flujo vehicular.

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

El valor de $\sum M\alpha = N$ corresponde a la aproximación de la cantidad de vehículos que circularon por la totalidad de tramos de carretera durante las franjas horarias definidas en el estudio, y se obtiene multiplicando los diferentes elementos indicados en la tabla 3 (mediana de vehículos por franja, cantidad de franjas, cantidad de horas y cantidad de tramos).

Para la opción dos se realizó el ajuste de la fracción en la segunda etapa de muestreo con los tamaños variables en la cantidad de vehículos registrados para cada UPM (conglomerados de diferente tamaño), generando un diseño no *mesip*⁹ debido a las probabilidades de selección desiguales. La fórmula con el ajuste estaría definida por la siguiente ecuación:

$$f = f_a * f_b * k = \frac{a}{A} * \frac{b}{B_{ij}} * k \quad (2.5)$$

Con

$$k = \frac{h_{ij} * n}{b}$$

⁹ Mesip: Acrónimo de ‘método de selección con igual probabilidad’.

Donde:

f : fracción de muestreo general.

f_a : fracción de muestreo en la primera etapa.

f_b : fracción de muestreo en la segunda etapa.

k : factor de ajuste.

a : cantidad de UPM seleccionadas en la primera etapa de muestreo.

b : cantidad de vehículos seleccionados de cada UPM en la segunda etapa de muestreo.

A : cantidad de UPM en la población.

B_{ij} : cantidad de vehículos que circularon por UPM (tramo “ i ” y franja horaria “ j ”).

h_{ij} : proporción de vehículos que circularon por UPM (tramo “ i ” y franja horaria “ j ”).

n : cantidad de vehículos seleccionados para realizar las observaciones en las UPM de la muestra.

En la ecuación (3.5), debido al tamaño variable de cada UPM en cuanto al flujo vehicular, y dado un submuestreo fijo “ b ”, se obtienen probabilidades de selección desiguales, por lo que se recurre al cálculo de un factor de ajuste “ k ” para lograr un diseño *mesip*. Este factor se obtiene asumiendo una asignación proporcional de la muestra en cada UPM, con lo que se obtiene la cantidad teórica de vehículos que debieron realizarse para obtener un diseño autoponderado y, a partir de este valor, se genera el factor de ajuste requerido para equilibrar las probabilidades de selección en la segunda etapa.

2.3.3.4. Análisis de conglomerados

El análisis de agrupamientos o de conglomerados puede definirse como un conjunto de técnicas numéricas utilizadas para la formación de grupos a partir de elementos que presentan semejanzas según algún criterio. Para formar los grupos se parte como base de un conjunto de características correspondientes a los diferentes elementos o casos (Rodríguez, 2013).

La creación de un conglomerado está definida por la cohesión interna (homogeneidad) y aislamiento externo (separación), según comentan Cormack (1971) y Gordon (1999), citados en Everitt, B. et al. (2011). Lo anterior quiere decir que los grupos conformados a partir de la técnica deben ser lo más homogéneos a lo interno de cada grupo y lo más diferentes, en la medida de lo posible, entre los distintos grupos. En la práctica no resulta sencillo identificar una técnica única para definir los grupos, lo que dependerá de aspectos como la medida de semejanza o desemejanza, la técnica de conglomeración como tal o el número de grupos que se requiera definir. Estos son aspectos que el investigador debe definir a partir de consideraciones prácticas, razones subjetivas o el contexto de los datos (Rodríguez, 2013).

En general, dentro del análisis de conglomerados existen diferentes técnicas de agrupamiento que pueden clasificarse como métodos de partición, métodos jerárquicos, métodos basados en cuadrícula y métodos basados en densidad (Preeti, A., 2016). Para el presente trabajo se utilizan métodos de

partición, los cuales consisten en la división del conjunto de datos en “k” grupos, donde el valor “k” requiere ser especificado por el usuario. Posteriormente, se ejecuta el algoritmo produciendo un número “k” de grupos, el cual puede ajustarse con base en índices de calidad o de forma gráfica (Rousseeuw, 1997).

Para el procedimiento de posestratificación utilizado en el presente trabajo, se recurre al método de partición conocido como k-medias. Como menciona Kassambara (2017), este procedimiento es el algoritmo de aprendizaje automático no supervisado más utilizado para dividir un conjunto de datos en determinado “k” número de grupos, donde el valor “k” se determina previamente y cada grupo es representado por su centro (centroide), el cual corresponde al promedio de los puntos asignados a determinada agrupación.

El método brinda resultados relativamente eficientes en relación con la varianza dentro del grupo, debido a las particiones generadas por la técnica (MacQueen, J., 1967, p. 281). Con lo anterior, el método busca minimizar la suma de cuadrado dentro del grupo (Hartigan, J. & Wong, M., 1979, p. 100). De tal forma que, según Kassambara (2017) y Telgarsky & Vattani (2010), los objetos dentro de un mismo grupo sean lo más parecidos posible (alta similitud intra-clase), mientras que los objetos en diferentes grupos sean lo más diferentes posible (baja similitud inter-clase).

Existen diferentes algoritmos para el método k-medias; sin embargo, como indica Kassambara (2017), el algoritmo estándar más utilizado corresponde al Hartigan-Wong (1979), el cual define la variación total dentro del grupo como la suma de cuadrados de las distancias euclídeas entre los elementos y el correspondiente centroide de la siguiente forma:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (3.6)$$

Donde:

x_i : corresponde al elemento asignado al conglomerado C_k

μ_k : corresponde al valor medio de los puntos asignados al conglomerado C_k

Cada observación “ x_i ” es asignada a determinado conglomerado de tal forma que la suma de cuadrado de la distancia de la observación al valor medio o centroide “ μ_k ” de su conglomerado de pertenencia sea mínima. La variación total dentro del conglomerado se define de la siguiente manera:

$$\sum_{k=1}^k W(C_k) = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (3.7)$$

La suma de cuadrados total dentro del conglomerado mide qué tan cercanos se encuentran los elementos dentro del grupo, y se busca que esta sea mínima. En relación con la medida de distancia utilizada, como menciona Everitt et al. (2011), existe una serie de medidas para derivar una matriz de disimilitud con base en un conjunto de observaciones multivariadas continuas: euclídea, *city block*, Minkowsky, Camberra, correlación de Pearson y separación angular. Para efectos del presente trabajo, se utiliza la distancia euclídea, la cual se define de la siguiente forma:

$$d_{ij} = \left[\sum_{k=1}^p (x_{jk} - x_{ik})^2 \right]^{1/2} \quad (3.8)$$

Donde " x_{jk} " y " x_{ik} " corresponden respectivamente al k-ésimo valor de la observación p-dimensional para los individuos "i" y "j". Con base en lo anterior y según lo indicado en Kassambara (2017), el algoritmo para el método k-medias puede resumirse como se indica seguidamente:

- 1) Como primer paso debe indicarse el número de conglomerados "k" que serán generados en la solución final.
- 2) El algoritmo inicia seleccionando aleatoriamente "k" objetos del conjunto de datos para definir los centros (centroide) iniciales de cada conglomerado.
- 3) Posteriormente, los objetos restantes se le asignan al centroide más cercano, donde la cercanía es definida utilizando la distancia euclídea entre el objeto y el valor medio del conglomerado. Esta etapa se conoce como "paso de asignación del conglomerado".
- 4) Posteriormente, el algoritmo calcula el nuevo valor promedio del conglomerado. El término "actualización del centroide" es utilizado para definir este paso. Una vez que los centros están recalculados, cada observación se revisa nuevamente para verificar si se encuentra más cercana a otro conglomerado.
- 5) Los pasos 3 y 4 se repiten de forma iterativa hasta que la asignación del conglomerado no cambie, es decir, cuando el conglomerado formado en la iteración actual sea el mismo que en la iteración previa.

2.4. Resultados

En esta sección se detallan los principales resultados. Para iniciar, se estudia el comportamiento de las variables e indicadores de desempeño utilizados en el desarrollo de las simulaciones para la comparación de técnicas de selección y variaciones del tamaño del submuestreo. Luego se realiza la comparación de escenarios, donde se estudian diferentes ajustes para la fracción de muestreo, se comparan diferentes técnicas de selección de elementos y se analiza su consistencia a partir de las simulaciones realizadas, para el caso general (conglomerados completos) y los grupos de flujo vehicular (posestratificación de UPM), estudiando el efecto de las simulaciones en la precisión exactitud de las estimaciones.

2.4.1. Variables e indicadores de desempeño utilizados en el análisis

En el apartado metodológico se indicaron las variables e indicadores de desempeño a utilizar como parte de los escenarios para la comparación de técnicas; sin embargo, es importante detallar aspectos relacionados con su significado e interpretación, como se muestra en los siguientes apartados.

2.4.1.1. Variables utilizadas en el análisis

El cuadro 10 presenta el resumen descriptivo sobre el uso de cinturón de seguridad, flujo vehicular y flujo de personas que circularon por las UPM seleccionadas. Además, como se ha indicado en el documento, la unidad de análisis corresponde a la combinación del tramo de carretera y la franja horaria en la que se aplican las mediciones, contabilizando un total de ochenta minutos distribuidos en dieciséis franjas de cinco minutos por tramo para la observación de carros, por lo que la interpretación se encuentra en función del tramo y la franja.

A partir de lo anterior, si se observa la variable “flujo_personas”, el valor 85,1 corresponde a la cantidad media de personas que circularon en sus vehículos por las tramo-franjas durante los ochenta minutos distribuidos en un día de observación.

En el caso de la variable *uso_cinturon*, el valor 66,7 corresponde a la cantidad media de personas que circularon en sus vehículos y utilizaban el cinturón de seguridad al transitar por las tramo-franjas durante los ochenta minutos distribuidos en un día de observación.

Cuadro 10. Resumen de las variables utilizadas en el análisis

| Variable | Descripción | Estadísticos de resumen | | | | | | |
|----------------|---|-------------------------|------|------|-------|---------|------|-------|
| | | Min. | P25 | Med. | Prom. | D. Est. | P75 | Max. |
| flujo_personas | Cantidad de ocupantes de vehículo que circularon por la UPM | 33,0 | 53,0 | 67,0 | 85,1 | 56,3 | 91,0 | 414,0 |
| uso_cinturon | Cantidad de ocupantes de vehículo que circularon por la UPM y utilizaron el cinturón de seguridad | 17,0 | 39,0 | 51,0 | 66,7 | 51,3 | 70,0 | 369,0 |
| flujo_carros | Cantidad de vehículos que circularon por la UPM | 30,0 | 37,0 | 46,0 | 62,0 | 42,5 | 67,0 | 302,0 |

Nota: En los estadísticos de resumen, entiéndase Min.=Mínimo, P25=Percentil 25, Med.=Mediana, Prom.=Promedio, D. Est.=Desviación Estándar, P75=Percentil 75 y Max.=Máximo.

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Por otra parte, en la variable “flujo_carros”, el valor 62,0 corresponde a la cantidad media de vehículos que circularon por la tramo-franja durante los ochenta minutos distribuidos en un día de observación.

2.4.1.2. Indicadores de desempeño utilizados en el análisis

En el apartado metodológico se presentó la definición de los indicadores de desempeño a utilizar como parte de los procesos de simulación para la comparación de técnicas de selección de elementos. Ahora interesa brindar un resumen de sus valores e interpretación en el contexto del presente trabajo. Los resultados que se muestran corresponden a los valores obtenidos de las estimaciones realizadas a partir de los conglomerados completos.

En el cuadro 11 se brindan los cuatro indicadores: el primero muestra un valor de 0,78 aproximadamente, y corresponde a la proporción media de uso de cinturón de seguridad por parte de las personas que circularon en sus vehículos por las tramo-franjas durante los ochenta minutos distribuidos en un día de observación. Asimismo, el valor 0,0070 corresponde al error estándar de la proporción media de uso de cinturón de seguridad y explica la variabilidad sufrida por este estadístico debido al diseño de muestreo utilizado.

Por su parte, el valor 10,7471 corresponde al efecto del diseño y representa el grado de inflación o deflación de la varianza del estimador de interés (en este caso, la proporción media de uso de cinturón de seguridad) debido a la implementación de un diseño de muestreo complejo o de varias etapas, en relación con la varianza que se obtendría a partir de un muestreo irrestricto aleatorio con la misma cantidad de elementos. Para efectos del presente trabajo, la variabilidad de la proporción media de uso de cinturón de seguridad se incrementa aproximadamente 11 veces, debido al uso del diseño de muestreo complejo. En el subapartado 2.4.2.2 se profundiza en el efecto del diseño y la magnitud de su valor para el caso del presente estudio.

En el caso del coeficiente de variación, como medida de dispersión relativa, contribuye a la valoración de la variabilidad del indicador de interés, en este caso, a partir del error estándar, y la proporción media se obtiene un valor próximo a 0,89 de dicho coeficiente, el cual, al tratarse de una cantidad cercana a cero, evidencia poca variabilidad, que a la vez se traduce en indicativo de buena precisión para estimar la proporción media de uso de cinturón de seguridad.

Cuadro 11. Indicadores de desempeño definidos a partir del uso de cinturón de seguridad en los conglomerados completos

| Indicador | Valor |
|--|---------|
| Proporción media de uso de cinturón de seguridad | 0,7845 |
| Error estándar de la proporción media | 0,0070 |
| Efecto del diseño | 10,7471 |
| Coeficiente de variación | 0,8865 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

2.4.2. Escenarios de los procesos de simulación

Ahora interesa definir escenarios de simulación para comparar fracciones de muestreo, diferentes técnicas de selección de elementos, procedimientos de submuestreo y posestratificación, para conocer el comportamiento de las variabilidades en las estimaciones relacionadas con el uso de dispositivos de seguridad pasiva en conductores de vehículos automotores. Para esto se propuso el uso de la proporción media, el error estándar, el efecto del diseño y el coeficiente de variación como indicadores de desempeño para el análisis de dichas variabilidades.

Los dos escenarios definidos corresponden al caso general y la posestratificación; en el primero se realizan los procesos de simulación a partir de las UPM sin agrupar, mientras que en el segundo se realiza una agrupación de UPM con base en el flujo vehicular, y en cada grupo se ejecutan los procesos de simulación ya indicados.

2.4.2.1. Comparación de fracciones de muestreo

Como se detalló en el apartado metodológico, en el presente trabajo se plantean dos escenarios para el ajuste de las fracciones de muestreo debido al tamaño desigual en la cantidad de observaciones correspondientes a cada UPM; previo a la comparación de las técnicas de selección, se valoró el rendimiento de ambas opciones de ajuste en la estimación de los estadísticos de interés o indicadores de desempeño.

Como se indicó en el cuadro 11 del subapartado 2.4.1.2, dichos indicadores corresponden a los valores obtenidos para cada estadístico a partir del total de observaciones obtenidas en el estudio (conglomerados completos), los cuales se utilizaron como valores de referencia para comparar las estimaciones obtenidas a partir de las dos propuestas de ajuste de la fracción de muestreo. El procedimiento consistió en realizar simulaciones con diferentes valores de submuestreo (b) y cantidad de réplicas, con “b” que oscilaba entre dos y veinte elementos por UPM; y para cada “b” definido, se ejecutaron 100 réplicas. A partir de las simulaciones realizadas se obtuvieron las estimaciones de los estadísticos de interés para cada “b” y, posteriormente, se calcularon las diferencias entre dichas estimaciones y el valor de referencia.

Cuadro 12. Comparación de fracciones de muestreo para estimar indicadores de desempeño según variaciones en el tamaño del submuestreo

| Submuestreo por UPM | Proporción | | Error estándar | | Efecto diseño | | C. Variación | | Total | |
|------------------------|------------|-----------|----------------|----------|---------------|-----------|--------------|----------|-----------|-----------|
| | f-PPT | f-Ajust | f-PPT | f-Ajust | f-PPT | f-Ajust | f-PPT | f-Ajust | f-PPT | f-Ajust |
| Total | 9 | 10 | 19 | 0 | 0 | 19 | 19 | 0 | 47 | 29 |
| 2 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 1 |
| 3 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |
| 4 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 1 |
| 5 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |
| 6 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 1 |
| 7 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |
| 8 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 1 |
| 9 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |
| 10 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 1 |
| 11 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 1 |
| 12 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |
| 13 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |
| 14 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |
| 15 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 1 |
| 16 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 1 |
| 17 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 3 | 1 |
| 18 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |
| 19 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |
| 20 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 | 2 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Con base en las diferencias calculadas, se identificó la fracción de muestreo que generó las menores diferencias en las estimaciones de los indicadores, según tamaño del submuestreo, como se muestra en el cuadro 12

El valor “1” en el cuadro 12 indica, para determinada combinación indicador-submuestreo, cuál fracción presentó la menor diferencia en la estimación realizada respecto al valor de referencia del cuadro 11. Como se observa, únicamente en el indicador correspondiente a la proporción se encontraron variaciones en la fracción que presentó mejor rendimiento, mientras que en el caso del error estándar y el coeficiente de variación, el ajuste de la fracción de muestreo asumiendo PPT evidenció mejor desempeño para todas las variaciones de submuestreo utilizadas.

La columna total del cuadro 12 contabiliza la cantidad de indicadores que mostraron menores diferencias en su estimación, según fracción de muestreo, lo cual puede observarse de forma general o por submuestreo, y estos resultados pueden analizarse de forma resumida en el cuadro 13.

Al analizar los datos de forma general, se observa que, del total de combinaciones realizadas, en aproximadamente 62% de los casos se obtuvo mejor rendimiento cuando se utilizó el ajuste de la fracción de muestreo asumiendo PPT (f_PPT). Lo anterior puede valorarse individualmente por indicador. Para el caso de la fracción de muestreo mediante ajuste por conglomerados desiguales (f_ajustado), se obtuvieron mayores ganancias en la estimación de la proporción y el efecto del diseño, mientras que, en el caso del error estándar y coeficiente de variación, el mejor rendimiento se obtuvo con el ajuste de la fracción asumiendo PPT.

Cuadro 13. Comparación de fracciones de muestreo según indicador de desempeño

| Indicador | Total | f_PPT | | f_Ajustado | | |
|------------------------|-----------|---------------|------------|--------------|------------|--------------|
| | | Cantidad | Porcentaje | Cantidad | Porcentaje | |
| Total | 76 | 100,00 | 47 | 61,84 | 29 | 38,16 |
| Proporción | 19 | 100,00 | 9 | 47,37 | 10 | 52,63 |
| Error estándar | 19 | 100,00 | 19 | 100,00 | 0 | 0,00 |
| Efecto diseño | 19 | 100,00 | 0 | 0,00 | 19 | 100,00 |
| Coef. Variación | 19 | 100,00 | 19 | 100,00 | 0 | 0,00 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Con base en estos resultados, se evidencia que el ajuste de la fracción asumiendo muestreo PPT presenta mejor rendimiento en la reducción de las variabilidades de los estadísticos de interés, por lo que se decide utilizar este ajuste en la etapa correspondiente a la comparación de técnicas de selección.

2.4.2.2. Escenario general sin agrupación de UPM

En este escenario, se desarrollan los procesos de simulación para el estudio de los tamaños de submuestreo, comparación de las técnicas de selección, y su consistencia en la estimación de los indicadores de desempeño a partir de las UPM sin agrupar.

2.4.2.2.1. Submuestreo y comparación de técnicas en la selección de vehículos

Una vez realizado el análisis y escogencia de la mejor metodología relacionada con la fracción de muestreo para las simulaciones, se procedió con el estudio de los tamaños de submuestreo y comparación de las técnicas de selección. Esto inicia con el desarrollo de los escenarios mediante simulaciones para diferentes tamaños del submuestreo “b” y número de réplicas para cada “b”.

El gráfico 7 muestra los resultados obtenidos luego de ejecutar los procesos de simulación para tamaños de “b” que varían entre 2 y 20 elementos por UPM. Para cada tamaño se calcularon 100 réplicas que se promediaron con el objetivo de encontrar estabilidad en las estimaciones de los cuatro indicadores. La línea negra de valor constante en el gráfico en cada indicador representa su valor cuando se utilizan los conglomerados completos, lo que sirve de referencia para realizar el análisis comparativo de las técnicas de selección.

A partir de lo anterior, para la proporción media y el efecto del diseño, se observa en las tres técnicas de selección una subestimación con respecto al valor de referencia; para la proporción, el resultado se debe al diseño de muestreo utilizado, y estos aspectos se abordan con mayor detalle en el apartado 2.4.2.4; mientras que, para el efecto del diseño, se explica por el tamaño de muestra, como se detallará en los párrafos siguientes.

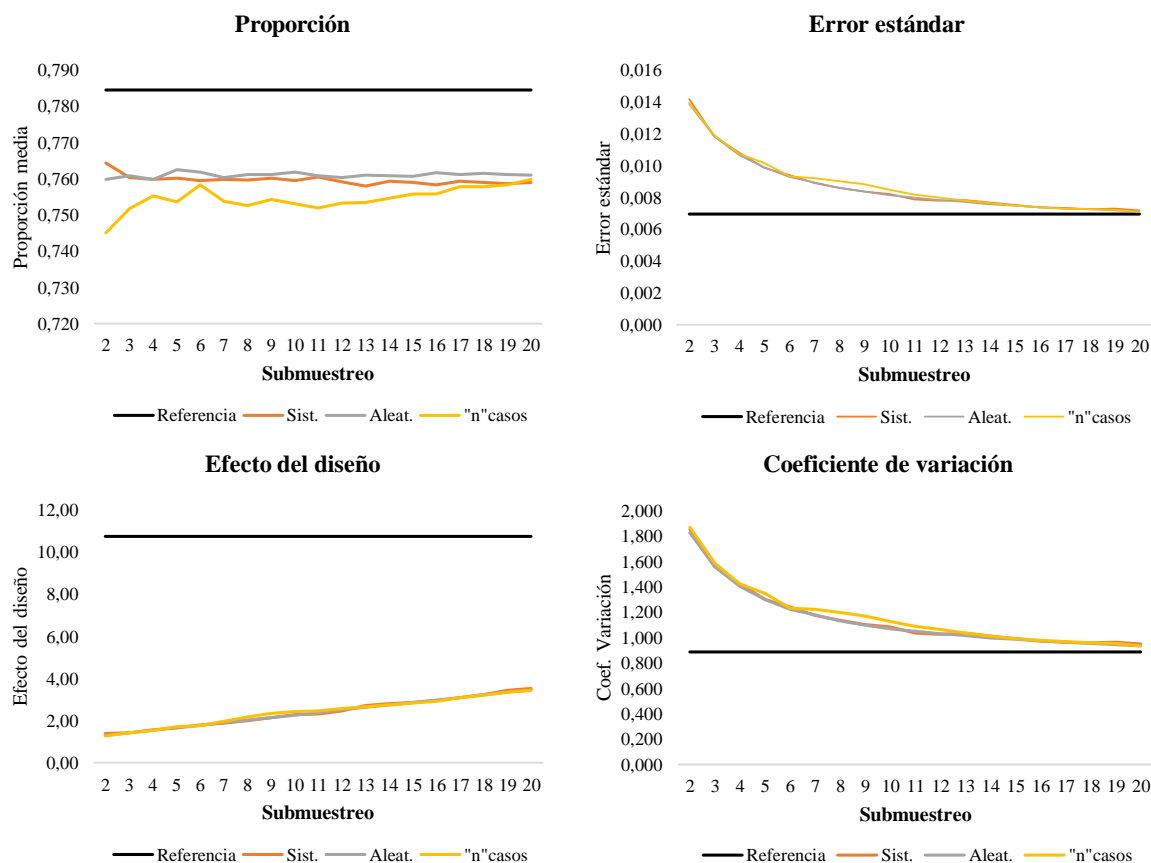
Además, como se aprecia en el gráfico 7, a excepción del efecto del diseño, las tres líneas que representan las técnicas de selección muestran un comportamiento asintótico horizontal al valor de referencia, apreciándose cierto distanciamiento, provocado en gran medida por un aspecto visual de la escala, dado que se logra estabilidad a partir de valores cercanos a 20 elementos por UPM, como se ha indicado –y dichas diferencias con respecto al valor de referencia resultan ínfimas para aspectos prácticos– (ver cuadro A2 de anexos).

En relación con la estabilidad de las estimaciones, la posibilidad de submuestrear con diferente cantidad de elementos brinda la capacidad de determinar el número óptimo de vehículos que deben seleccionarse por UPM para disminuir la variabilidad de las estimaciones. En relación con lo mencionado se observa que, tanto en la proporción media como en el error estándar y el coeficiente de variación, se logra cierta convergencia a partir de valores de “b” entre 15 y 20 elementos, lo que

sugiere que dicho umbral estaría definido por un mínimo de 20 elementos por UPM seleccionada. Sin embargo, para el caso del efecto del diseño, es importante notar que no se logró convergencia en un determinado tamaño de submuestreo como en los restantes tres indicadores; asimismo, la diferencia que se tiene entre el valor de referencia y el calculado a partir de los procesos de simulación es amplia; por ejemplo, para “ $b = 20$ ” muestra valores de 10,7 y 3,5 respectivamente.

El resultado anterior se explica por un aspecto relacionado con el tamaño de la muestra. Se sabe que el efecto del diseño surge a partir de la razón de la varianza del estimador del diseño complejo entre la varianza del estimador asumiendo un muestreo irrestricto aleatorio (‘MIA’) de igual tamaño de muestra. Con respecto a este cálculo, el MIA de elementos resulta más sensible a la cantidad de elementos en su denominador. Retomando el ejemplo anterior, para efectos del cálculo de la varianza del MIA, cuando se calcula el efecto del diseño a partir de las simulaciones con un submuestreo de “ $b = 20$ ” elementos por UPM, se obtiene un total de 12306 observaciones, mientras que, al utilizar los conglomerados completos, se obtiene un total de 37509 observaciones. Esta diferencia entre tamaños de muestra provoca los resultados que se observan entre los efectos de diseño, debido a que la varianza del MIA sufre una disminución mayor al utilizar los conglomerados completos respecto al resultado de las simulaciones.

Gráfico 7. Estimación de indicadores de desempeño para el uso de cinturón de seguridad con diferentes técnicas de selección y variaciones del tamaño del submuestreo



Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

2.4.2.2.2. Rendimiento de las técnicas de selección

A partir del análisis gráfico realizado en el subpartado anterior, se evidenció la similitud existente entre las tres técnicas de selección con respecto al valor de referencia, por lo que ahora interesa determinar de forma precisa la técnica que presenta el mejor rendimiento para el cálculo de las estimaciones de los indicadores de desempeño. Para ello, posterior al cálculo de las diferencias entre las estimaciones y el valor de referencia, se procedió a identificar y contabilizar, para cada tamaño de submuestreo, en cuántos indicadores de desempeño producían el mayor rendimiento según técnica de selección.

Como se observa en el cuadro 14, para el caso de $b = 2$, $b = 3$ y $b = 4$, de los cuatro indicadores utilizados, dos presentaron mejor desempeño en las estimaciones cuando se aplicó el sistemático para la selección de elementos y dos en el caso del aleatorio. Para la técnica definida como "n" casos, no se lograron mejores rendimientos con ninguno de los cuatro indicadores. Por lo tanto, para los

ensayos con $b = 2, 3$ y 4 elementos, no existen diferencias entre las técnicas aleatoria y sistemática. En el caso de $b = 5$, la selección aleatoria presenta mejores resultados en dos de los indicadores, mientras que en el sistemático y “n” casos se logró únicamente con un indicador, por lo que, para este ensayo, la selección aleatoria evidencia mejor rendimiento. De esta forma se analiza el rendimiento para los diferentes ensayos realizados según cantidad del submuestreo, obteniendo como resultado final el resumen presentado en el cuadro 15.

Cuadro 14. Comparación de técnicas de selección a partir de diferencias entre indicadores de desempeño según tamaño del submuestreo

| Submuestreo | Técnica | | | Mejor técnica |
|-------------|---------|--------|----------|-----------------------|
| | Sist. | Aleat. | "n"casos | |
| 2 | 2 | 2 | 0 | Sistemático/Aleatorio |
| 3 | 2 | 2 | 0 | Sistemático/Aleatorio |
| 4 | 2 | 2 | 0 | Sistemático/Aleatorio |
| 5 | 1 | 2 | 1 | Aleatorio |
| 6 | 1 | 3 | 0 | Aleatorio |
| 7 | 2 | 1 | 1 | Sistemático |
| 8 | 0 | 3 | 1 | Aleatorio |
| 9 | 0 | 3 | 1 | Aleatorio |
| 10 | 0 | 3 | 1 | Aleatorio |
| 11 | 2 | 1 | 1 | Sistemático |
| 12 | 2 | 1 | 1 | Sistemático |
| 13 | 1 | 3 | 0 | Aleatorio |
| 14 | 1 | 3 | 0 | Aleatorio |
| 15 | 1 | 2 | 1 | Aleatorio |
| 16 | 0 | 3 | 1 | Aleatorio |
| 17 | 1 | 3 | 0 | Aleatorio |
| 18 | 1 | 2 | 1 | Aleatorio |
| 19 | 1 | 3 | 0 | Aleatorio |
| 20 | 1 | 2 | 1 | Aleatorio |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del Estudio nacional sobre uso de dispositivos de seguridad pasiva.

Como puede observarse, de los 19 ensayos realizados correspondientes a los diferentes tamaños del submuestreo, en 14,5 la técnica aleatoria evidenció mejor rendimiento, lo que demuestra mayor robustez y estabilidad, según las simulaciones realizadas, por lo que se define como la mejor técnica para seleccionar vehículos, debido a que proporciona las menores diferencias en las estimaciones realizadas respecto al valor obtenido al utilizar la información de los conglomerados completos. Sin embargo, es importante indicar que dichas diferencias, particularmente entre la técnica aleatoria y sistemática, son reducidas en su magnitud (ver cuadro A3 de anexos).

Cuadro 15. Consistencia en la estimación de indicadores de desempeño según técnica de selección

| Técnica | Consistencia | |
|--------------------|--------------|---------------|
| | Cantidad | Porcentaje |
| Total | 19 | 100,00 |
| Aleatorio | 14,5 | 76,32 |
| Sistemático | 4,5 | 23,68 |
| "n" casos | 0 | 0,00 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

2.4.2.3. Posestratificación de UPM a partir del flujo vehicular

Además de analizar estrategias metodológicas basadas en procedimientos de selección de elementos y variaciones del tamaño del submuestreo para calcular estadísticos con menor variabilidad, también se consideró importante analizar el efecto que tendría la posestratificación de UPM en el proceso de estimación. De la misma forma en que se realizó para las UPM sin agrupar, en este caso se desarrollaron los procesos de simulación para el estudio de los tamaños de submuestreo, comparación de las técnicas de selección, y su rendimiento en la estimación de los indicadores de desempeño.

2.4.2.3.1. Escogencia del número adecuado de conglomerados

En la construcción de los grupos usando análisis de conglomerados, se recurrió al método de partición “k-medias”, utilizando como variable de agrupación el flujo de vehículos que circuló por la UPM. Para definir la cantidad óptima de grupos, se realizaron diferentes pruebas y cantidad de agrupaciones, utilizando recursos estadísticos, así como evaluaciones gráficas mediante el método *average silhouette* y el codo de Jambú, combinando con una valoración sustantiva del fenómeno estudiado y manteniendo una relación lógica entre la cantidad de grupos, su tamaño y una interpretación aceptable desde el punto de vista práctico.

El cuadro 16 resume los resultados de las diferentes pruebas realizadas. Como se observa, la distribución de los casos en los grupos no presentó inconvenientes y se lograron porcentajes suficientes en los tres casos ensayados. Además, se aprecia una separación clara de las agrupaciones en función del flujo medio del grupo, diferenciando los de mayor y menor flujo, resultado que puede complementarse con la información del flujo mínimo y máximo por grupo.

Como se indicó anteriormente, la finalidad de realizar posestratificación de las UPM se justifica por el comportamiento diferenciado en torno al uso del dispositivo de seguridad según el flujo vehicular. Por este motivo, en el cuadro 16 se incluyó el porcentaje de uso de cinturón de seguridad como un

indicador de consistencia de los resultados obtenidos. Como se observa en los diferentes escenarios, las agrupaciones con mayor flujo de vehículos se encuentran asociadas con un porcentaje mayor de uso de cinturón de seguridad, mientras que los grupos de menor flujo se relacionan con menor uso del dispositivo, resultado que respalda la premisa planteada en el presente apartado.

Es importante especificar que, en los tres casos analizados, el mínimo de los grupos con menor flujo vehicular está definido en 30 elementos, debido a la exclusión de las UPM con una reducida circulación de vehículos y su limitante en el desarrollo de las simulaciones, aspecto que ya fue explicado en el apartado 2.4.2.

Con base en esta información, se consideró que dos grupos es la mejor alternativa para efectos de interpretación y, además, porque proporciona una mejor separación. Si se considera que el interés en la construcción de grupos radica en la comparación de técnicas de selección y submuestreos, según el flujo vehicular, los grupos con mayores diferencias en la circulación de vehículos brindan mejores condiciones para evaluar si existen diferencias en la estimación de las variabilidades, según técnicas de selección.

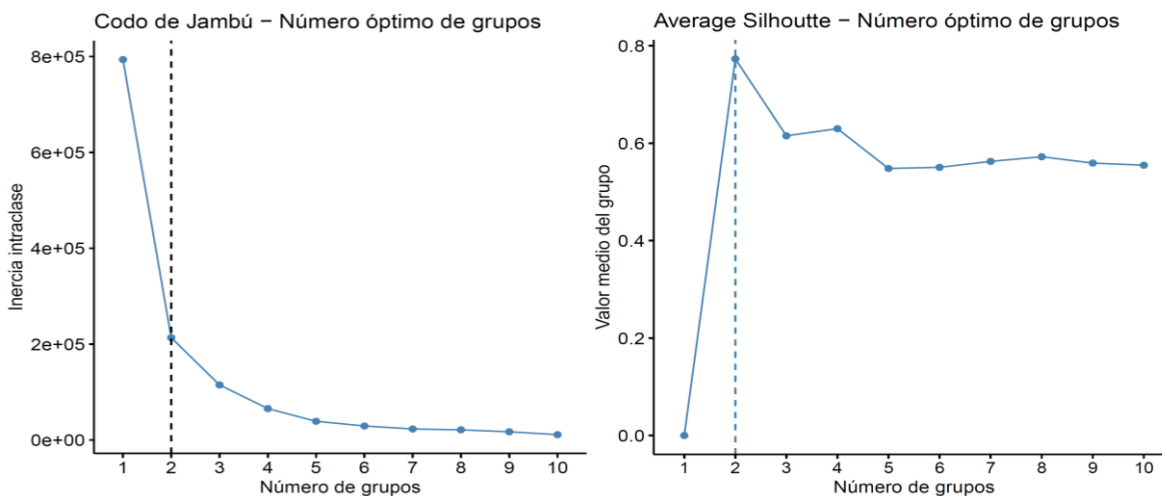
Cuadro 16. Resultados de la conformación de grupos a partir del flujo de vehículos

| Agrupación | Número de tramo-franjas por grupo (UPM's) | % de tramo-franjas por grupo | Flujo medio del grupo ^{1/} | Flujo mínimo del grupo | Flujo máximo del grupo | % de uso cinturón | |
|-----------------|---|------------------------------|-------------------------------------|------------------------|------------------------|-------------------|------|
| General | 441 | 100,0 | 89,4 | 30 | 302 | 78,5 | |
| 2 Grupos | C1 | 49 | 11,1 | <u>164,6</u> | 112 | 302 | 86,7 |
| | C2 | 392 | 88,9 | <u>49,1</u> | 30 | 106 | 75,2 |
| 3 Grupos | C1 | 32 | 7,3 | <u>187,3</u> | 140 | 302 | 88,6 |
| | C2 | 84 | 19,0 | <u>88,1</u> | 66 | 133 | 77,8 |
| | C3 | 325 | 73,7 | <u>42,9</u> | 30 | 65 | 74,7 |
| 4 Grupos | C1 | 20 | 4,5 | <u>208,3</u> | 169 | 302 | 86,8 |
| | C2 | 39 | 8,8 | <u>126,0</u> | 98 | 163 | 84,3 |
| | C3 | 127 | 28,8 | <u>66,1</u> | 53 | 95 | 76,0 |
| | C4 | 255 | 57,8 | <u>38,6</u> | 30 | 52 | 74,5 |

1/ Entiéndase “flujo medio del grupo” como el centroide de cada grupo en la formación de los conglomerados.

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Como análisis gráfico complementario al abordaje anterior, se utilizaron el codo de Jambú y el *average silhouette*. Como resultado, ambos gráficos sugieren la formación de dos grupos. En el primer gráfico, a partir de dos grupos, la ganancia en la reducción de la inercia intraclase es mínima, mientras que en el segundo, el número óptimo de conglomerados “k” que maximiza el valor medio se obtiene con dos.

Gráfico 8. Inercia intraclase y valor medio del objeto según número de conglomerados

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

2.4.2.3.2. *Submuestreo y comparación de técnicas de selección en UPM de medio y alto flujo vehicular*

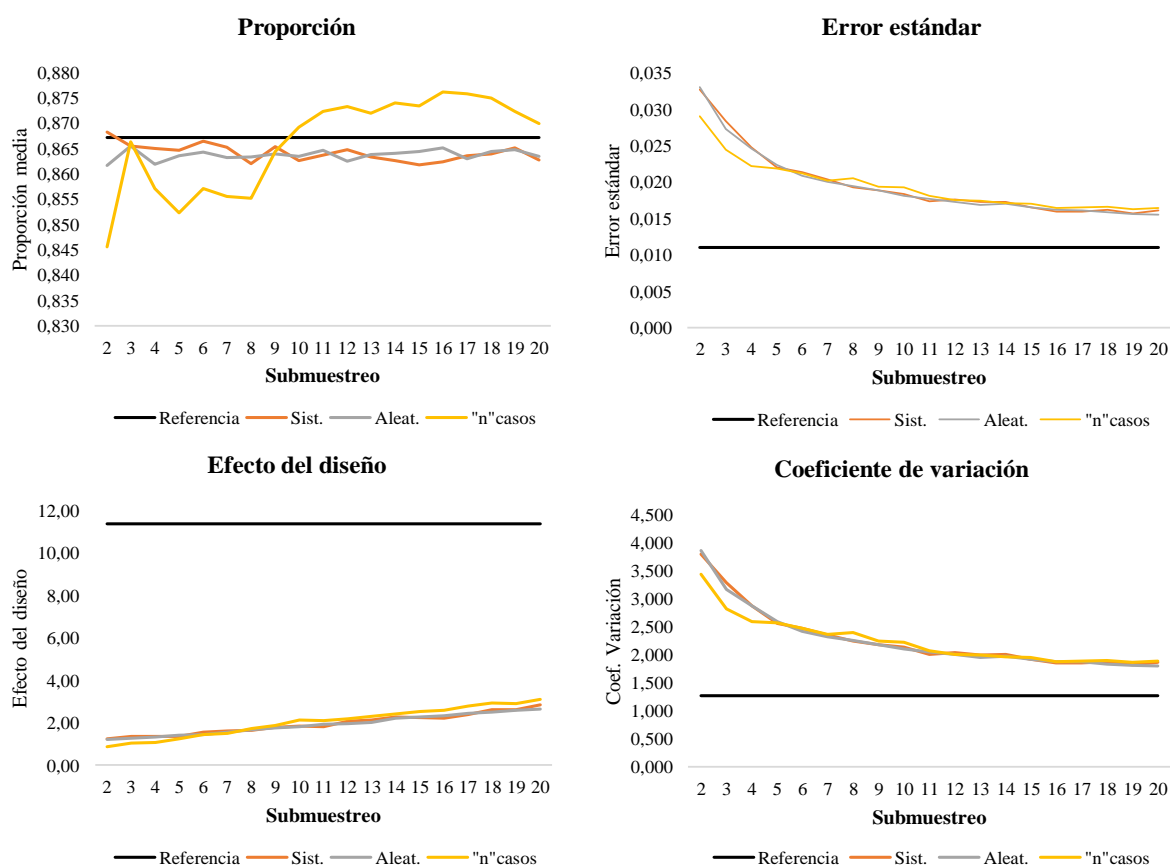
Una vez realizado el análisis de conglomerados, se procedió con las comparaciones de las técnicas de selección en los dos grupos de flujo vehicular definidos. El abordaje, en este caso, es similar al implementado en el apartado 2.4.2.2.1 para el escenario general; por lo tanto, el estudio de las técnicas de selección inicia con la ejecución de los escenarios mediante simulaciones para diferentes tamaños de submuestreo “b” y número de réplicas para cada “b” dentro del grupo de medio y alto flujo vehicular. El detalle del cálculo de las estimaciones y diferencias, con respecto al valor de referencia, puede consultarse en los anexos (cuadros A4 - A7).

El gráfico 9 muestra los resultados para el grupo de alto flujo vehicular, luego de ejecutar los procesos de simulación para tamaños de “b” que varían entre 2 y 20 elementos por UPM, con 100 réplicas que se promediaron con el objetivo de encontrar estabilidad en las estimaciones de los cuatro indicadores.

Como se observa, en el caso de la proporción media de uso de cinturón de seguridad para las tres técnicas de selección, no se aprecia el problema de subestimación del indicador como sí ocurrió en el escenario general (UPM sin agrupar). Con respecto a este resultado, se ahonda con mayor detalle en el apartado 2.4.2.4. Para el caso del efecto del diseño, las tres técnicas de selección evidenciaron una subestimación con respecto al valor de referencia, lo cual se explica por el tamaño de muestra, como se detalló anteriormente.

Con respecto a la estabilidad, en el caso de la proporción media para las técnicas de selección aleatoria y sistemática, se logra a partir de valores de “b” entre 15 y 20 elementos por UPM; no así para la técnica de los primeros “n” casos, que presenta un comportamiento con mayor variabilidad para los diferentes valores de “b” utilizados. Para el caso del error estándar y el coeficiente de variación, en las tres técnicas de selección se observa estabilidad en las estimaciones cuando el número de elementos seleccionados varía entre 15 y 20 elementos por UPM, al igual que en el caso de la proporción media, lo que sugiere que el umbral óptimo de observaciones estaría definido por un mínimo de 20 elementos por UPM. Para el efecto del diseño, similar al escenario general, se presentaron inconvenientes en la convergencia del indicador, debido a las particularidades descritas en el apartado 2.4.2.2.1 sobre el tamaño de muestra.

Gráfico 9. Estimación de indicadores de desempeño para el uso de cinturón de seguridad en el grupo de UPM de alto flujo vehicular con diferentes técnicas de selección y variaciones del tamaño del submuestreo



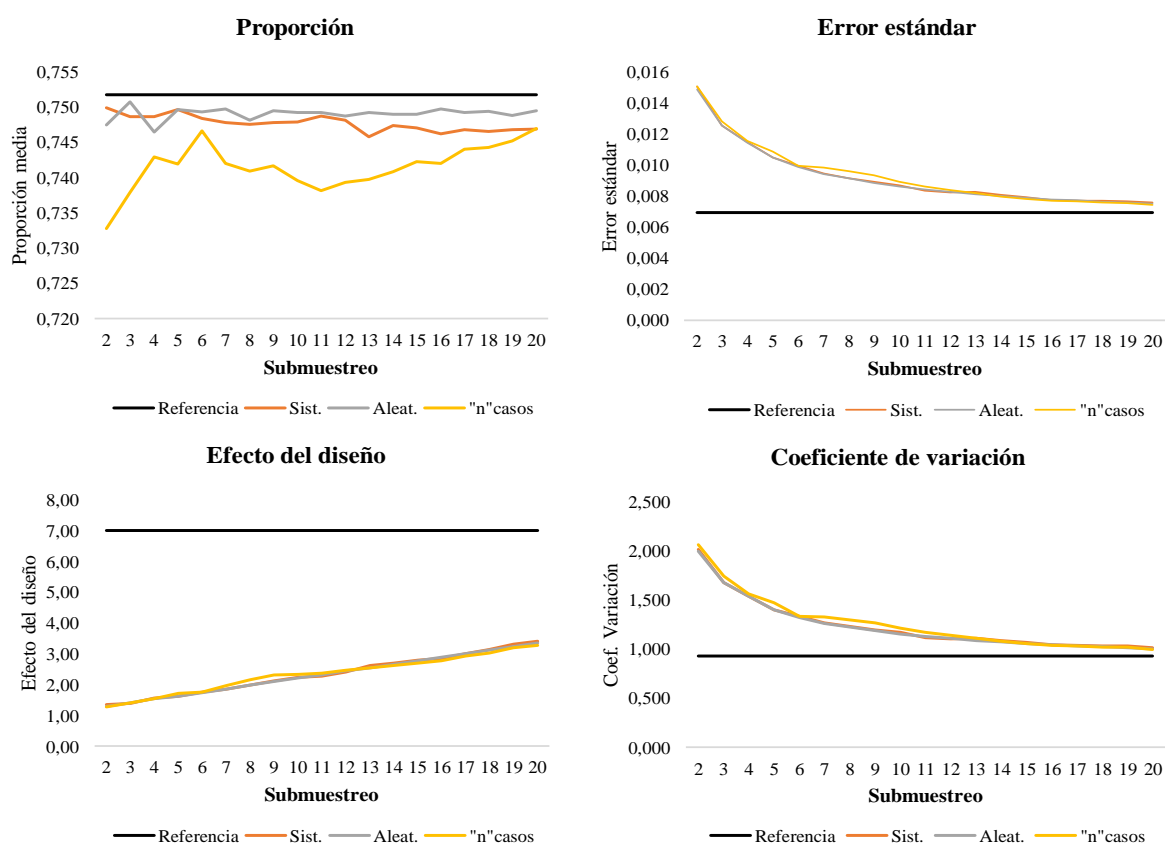
Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Similar a los resultados del grupo de alto flujo vehicular, en el caso del flujo medio, para el indicador de la proporción media, no se observaron problemas de subestimación del indicador; no así para el

efecto del diseño, en el que se mantiene la subestimación con respecto al valor de referencia para las tres técnicas de selección.

Además, con respecto a la estabilización de las estimaciones, tanto en la proporción media como en el error estándar y el coeficiente de variación se logró en valores similares al escenario general (UPM sin agrupar) y al grupo de alto flujo vehicular, con valores del submuestreo entre 15 y 20 elementos por UPM (a excepción de la técnica de los primeros “n” casos), lo que sugiere que el umbral óptimo de observaciones estaría definido por un mínimo de 20 elementos por UPM. Para el efecto del diseño, similar al escenario general y al grupo de alto flujo vehicular, se presentaron inconvenientes en la convergencia del indicador.

Gráfico 10. Estimación de indicadores de desempeño para el uso de cinturón de seguridad en el grupo de UPM de flujo vehicular medio con diferentes técnicas de selección y variaciones del tamaño del submuestreo



Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

2.4.2.3.3. Rendimiento de las técnicas de selección

Para determinar la mejor técnica, según grupo de flujo vehicular, se procedió a identificar y contabilizar, para cada submuestreo, cuántos indicadores de desempeño producían el mayor rendimiento según técnica de selección. El cuadro 17 resume los resultados obtenidos luego de identificar el desempeño de cada método de selección.

Cuadro 17. Comparación de técnicas de selección a partir de diferencias entre indicadores de desempeño por grupo de flujo vehicular según tamaños del submuestreo

| Submuestreo | UPM de alto flujo vehicular | | | UPM de flujo vehicular medio | | |
|-------------|-----------------------------|------------------|-------------------------|------------------------------|------------------|-------------------------|
| | Técnica | | Mejor técnica | Técnica | | Mejor técnica |
| | Sist. | Aleat. "n" casos | | Sist. | Aleat. "n" casos | |
| 2 | 2 | 0 | 2 Aleatorio/"n" casos | 2 | 2 | 0 Sistemático/Aleatorio |
| 3 | 1 | 0 | 3 "n" casos | 1 | 2 | 1 Aleatorio |
| 4 | 2 | 0 | 2 Sistemático/"n" casos | 3 | 1 | 0 Sistemático |
| 5 | 2 | 1 | 1 Sistemático | 2 | 1 | 1 Sistemático |
| 6 | 2 | 2 | 0 Sistemático/Aleatorio | 1 | 3 | 0 Aleatorio |
| 7 | 2 | 2 | 0 Sistemático/Aleatorio | 0 | 3 | 1 Aleatorio |
| 8 | 2 | 1 | 1 Sistemático | 0 | 3 | 1 Aleatorio |
| 9 | 2 | 1 | 1 Aleatorio | 0 | 3 | 1 Aleatorio |
| 10 | 0 | 2 | 2 Aleatorio/"n" casos | 0 | 3 | 1 Aleatorio |
| 11 | 2 | 1 | 1 Aleatorio | 2 | 1 | 1 Sistemático |
| 12 | 1 | 1 | 2 "n" casos | 2 | 1 | 1 Sistemático |
| 13 | 0 | 3 | 1 Aleatorio | 1 | 3 | 0 Aleatorio |
| 14 | 0 | 2 | 2 Aleatorio/"n" casos | 1 | 2 | 1 Aleatorio |
| 15 | 1 | 2 | 1 Aleatorio | 1 | 2 | 1 Aleatorio |
| 16 | 2 | 1 | 1 Sistemático | 0 | 2 | 2 Aleatorio/"n" casos |
| 17 | 3 | 0 | 1 Sistemático | 0 | 2 | 2 Aleatorio/"n" casos |
| 18 | 0 | 3 | 1 Aleatorio | 1 | 1 | 2 "n" casos |
| 19 | 1 | 2 | 1 Aleatorio | 1 | 2 | 1 Aleatorio |
| 20 | 0 | 2 | 2 Aleatorio/"n" casos | 1 | 1 | 2 "n" casos |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

La línea de análisis utilizada es equivalente al abordaje realizado en 2.4.2.2.2, estudiando el rendimiento para los diferentes ensayos realizados según submuestreo, obteniendo como resultado final el resumen presentado en el cuadro 18.

Este resultado muestra que, de los 19 ensayos realizados correspondientes a los diferentes tamaños del submuestreo, tanto en el grupos de alto flujo vehicular como en el de flujo medio, la técnica aleatoria evidenció mejor rendimiento, ya que presentó mayor cantidad de ensayos en los que se obtuvo mejor desempeño en la reducción de las variabilidades de los indicadores de interés: cerca de 47% y 60% respectivamente. Esto demuestra mayor robustez y estabilidad, según los diferentes escenarios simulados; por lo que se define como la mejor técnica para seleccionar vehículos.

Cuadro 18. Consistencia en la estimación de indicadores de desempeño por grupo de flujo vehicular según técnica de selección

| Técnica | UPM de alto flujo vehicular | | UPM de flujo vehicular medio | |
|--------------|-----------------------------|---------------|------------------------------|---------------|
| | Consistencia | | Consistencia | |
| | Cantidad | Porcentaje | Cantidad | Porcentaje |
| Total | 19,00 | 100,00 | 19,00 | 100,00 |
| Aleatorio | 9,00 | 47,37 | 11,50 | 60,53 |
| Sistemático | 5,50 | 28,95 | 4,50 | 23,68 |
| "n" casos | 4,50 | 23,68 | 3,00 | 15,79 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

2.4.2.4. Efecto de los procesos de simulación en la exactitud de las estimaciones

En los apartados anteriores se estudiaron las diferentes fracciones de muestreo, técnicas de selección y variaciones del submuestreo en el caso general y en los grupos formados a partir de la posestratificación de las UPM para analizar aspectos de precisión de los indicadores obtenidos a partir de los procesos de simulación, por lo que ahora interesa analizar los resultados en términos de la exactitud de las estimaciones con respecto a su valor de referencia.

El cuadro 19 resume los resultados para los indicadores de desempeño en el caso de las estimaciones generales y las obtenidas usando posestratificación. Además, se proporcionan los resultados de los indicadores para los conglomerados completos, con el objetivo de realizar las comparaciones.

Cuadro 19. Comparación de indicadores de desempeño por grupo de flujo vehicular según técnica de selección de elementos

| Indicador | C.Comp. ^{1/} | Estimación general | | | Posestratificación | | | | | | | |
|---------------------|-----------------------|--------------------|--------|-----------|-----------------------|----------------------|--------|-----------|-----------------------|-----------------------|--------|-----------|
| | | | | | C.Comp. ^{1/} | Alto flujo vehicular | | | C.Comp. ^{1/} | Flujo vehicular medio | | |
| | | Sist. | Aleat. | "n" Casos | | Sist. | Aleat. | "n" Casos | | Sist. | Aleat. | "n" Casos |
| Proporción | 0,7845 | 0,7590 | 0,7610 | 0,7598 | 0,8673 | 0,8628 | 0,8635 | 0,8700 | 0,7517 | 0,7469 | 0,7495 | 0,7470 |
| E. Estándar | 0,0070 | 0,0072 | 0,0071 | 0,0071 | 0,0111 | 0,0161 | 0,0156 | 0,0165 | 0,0070 | 0,0076 | 0,0075 | 0,0075 |
| E. Diseño | 10,7471 | 3,5124 | 3,4722 | 3,4437 | 11,3646 | 2,8361 | 2,6557 | 3,0988 | 7,0002 | 3,4013 | 3,3473 | 3,2763 |
| C. Variación | 0,8865 | 0,9459 | 0,9356 | 0,9355 | 1,2776 | 1,8721 | 1,8051 | 1,8939 | 0,9269 | 1,0170 | 1,0041 | 0,9981 |

1/: El termino "C.Comp." corresponde a los conglomerados completos, e indica el resultado general y los grupos de medio y alto flujo vehicular.

Nota: Para los valores de la estimación general y la posestratificación, se utilizaron los resultados a partir del submuestreo con b=20 elementos por UPM.

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Como puede observarse, en el caso de la estimación general y en el estrato o grupo de flujo vehicular medio, la estimación del error estándar y el coeficiente de variación evidenciaron valores similares al obtenido con los conglomerados completos, y en el caso del grupo de alto flujo vehicular, se experimentó un leve incremento en dicha estimación en comparación con el escenario general y el grupo de flujo medio.

En lo referente al efecto del diseño, se evidenció una reducción importante en sus estimaciones para el escenario general y en la posestratificación, particularmente en el grupo de alto flujo vehicular; aunque, como se detalló en apartados anteriores, este patrón se explica en gran medida por un aspecto de tamaño de muestra.

Sin embargo, en el escenario general, para el caso de la proporción media, se observa una mayor brecha entre las estimaciones obtenidas con las tres técnicas de selección y el valor de referencia, a diferencia de los resultados obtenidos en la posestratificación, donde las estimaciones presentan mayor exactitud con respecto a su valor de referencia para las tres técnicas de selección en ambos grupos. Como consecuencia de este hallazgo, se calcularon intervalos de confianza para las estimaciones de la proporción media, tanto para el caso general (conglomerados completos) como para los grupos de medio y alto flujo vehicular, con miras a determinar si existen diferencias significativas (cuadro 20).

Cuadro 20. Intervalos de confianza para la proporción media de uso de dispositivos de seguridad pasiva según escenario, grupo y técnica de selección

| Escenario | Grupo flujo vehicular | Técnica | Prop. media C.Completos | Prop. media estimada | IC 95% | |
|--------------------|-------------------------|--------------------|-------------------------|----------------------|--------|--------|
| | | | | | Inf. | Sup. |
| Estimación general | Flujo vehicular general | <i>Sistemático</i> | 0,7845 | 0,7590 | 0,7449 | 0,7731 |
| | | <i>Aleatorio</i> | | 0,7610 | 0,7471 | 0,7750 |
| | | <i>"n" Casos</i> | | 0,7598 | 0,7459 | 0,7737 |
| Posestratificación | Alto flujo vehicular | <i>Sistemático</i> | 0,8673 | 0,8628 | 0,8312 | 0,8944 |
| | | <i>Aleatorio</i> | | 0,8635 | 0,8330 | 0,8940 |
| | | <i>"n" Casos</i> | | 0,8700 | 0,8377 | 0,9023 |
| | Flujo vehicular medio | <i>Sistemático</i> | 0,7517 | 0,7469 | 0,7320 | 0,7618 |
| | | <i>Aleatorio</i> | | 0,7495 | 0,7348 | 0,7643 |
| | | <i>"n" Casos</i> | | 0,7470 | 0,7324 | 0,7616 |

Nota: Valores calculados con los resultados a partir del submuestreo con b=20 elementos por UPM.

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Como se observa, en el caso de los grupos de flujo vehicular producto de la posestratificación, los intervalos de confianza calculados en ambos grupos (medio y alto flujo vehicular) para las tres técnicas de selección contienen el valor de referencia. No obstante, en el caso del escenario general, el cálculo de la proporción media a partir de los conglomerados completos es aproximadamente 0,78, mientras que la proporción media estimada mediante los procesos de simulación para las tres técnicas de selección se encuentra subestimada en aproximadamente en 2,4% (0,76) con respecto a los conglomerados completos, por lo que, al calcular los correspondientes intervalos de confianza, estos no contienen dicho valor.

Esta situación se explica por la naturaleza de los conglomerados (UPM con tamaños variables) y el tipo de submuestreo utilizado en la segunda etapa como parte de los procesos de simulación. Además, como se mostró en la posestratificación, el grupo de flujo vehicular medio contiene la mayor cantidad de elementos, 89% aproximadamente (392 UPM), mientras que el grupo de alto flujo vehicular contiene apenas el 11% restante (49 UPM). Asimismo, debe considerarse que la proporción media de uso del cinturón de seguridad es menor en el grupo de flujo medio con respecto al grupo de alto flujo vehicular, por lo que, para el caso del escenario general, al submuestrear en la segunda etapa con un “b” de tamaño fijo en cada UPM, el valor calculado de la proporción media con base en los procesos de simulación y las tres técnicas de selección se ve subestimado debido a la mayor cantidad de UPM con flujo vehicular medio y menor porcentaje de uso de cinturón de seguridad.

El resultado anterior mejora cuando se realiza el mismo ejercicio en los grupos de flujo vehicular producto de la posestratificación. En este caso, la proporción media estimada coincide con mayor exactitud con la correspondiente proporción media calculada a partir de los conglomerados completos; además, los intervalos de confianza calculados en ambos grupos (medio y alto flujo vehicular) sí contienen el valor de dicha proporción media. Este resultado se explica por el control que proporciona la agrupación de UPM con base en el flujo vehicular, lo cual contribuye al diseño de muestreo y procedimiento de selección para una estimación con mayor exactitud de la proporción media de uso de cinturón de seguridad.

Es importante indicar que, para el escenario general, de forma complementaria se realizó el proceso de simulación para la estimación de la proporción media de uso de cinturón y demás indicadores de desempeño, pero en este caso generando un submuestreo variable por UPM definido por una fracción de muestreo “ $f_b = 1/3$ ” (ver cuadro A8 de anexos); esto para confirmar que la subestimación de la proporción media al utilizar un “b” de tamaño fijo se debe al mayor peso otorgado a las UPM de flujo vehicular medio, dado su mayor número en la población con respecto a las de alto flujo vehicular¹⁰.

Cabe destacar que el resultado descrito no invalida los objetivos establecidos en el presente trabajo para la comparación de técnicas de selección y análisis de la variabilidad; más allá de esto, invita a repensar el diseño de muestreos complejos de este tipo. Por lo tanto, en la práctica, en caso de submuestrear con “b” de tamaño fijo, un control de las UPM con base en el flujo vehicular (estratificación) es deseable, y se convierte en una herramienta a considerar para obtener ganancias

¹⁰ Se define $f_b = 1/3$ para generar un tamaño final de observaciones equivalente al que se obtiene cuando se submuestra con un “b” fijo igual a 20 elementos por UPM.

en la exactitud de las estimaciones. De lo contrario, si existen UPM (conglomerados) de tamaño variable y un control con base del flujo vehicular no resultara aplicable, es recomendable considerar en la segunda etapa un submuestreo variable (proporcional) por UPM a partir de una fracción de muestreo determinada.

2.5. Conclusiones y recomendaciones

Con respecto al primer objetivo específico del trabajo, se recurrió a variables relacionadas con la cantidad de ocupantes del vehículo que circularon por la UPM (*flujo_personas*), cantidad de ocupantes de vehículo que circularon por la UPM y utilizaron el cinturón de seguridad (*uso_cinturon*) y cantidad de vehículos que circularon por la UPM (*flujo_carros*). Asimismo, para estudiar las variabilidades se definieron como medidas de desempeño los indicadores correspondientes a la proporción media, error estándar, efecto del diseño y coeficiente de variación, con los cuales se logró valorar cuánto inciden los procesos de simulación en las estimaciones en relación con el uso de conglomerados completos.

En relación con el segundo objetivo específico, para el estudio de las variabilidades en la segunda etapa de muestreo se desarrollaron procesos de simulación, en los cuales se analizaron diferentes fracciones de muestreo, técnicas de selección, procedimientos de submuestreo y posestratificación en la estimación del uso de dispositivos de seguridad pasiva para determinar la existencia de ganancias en la precisión y exactitud de las estimaciones. Seguidamente se detallan los principales resultados obtenidos en cada uno de los elementos mencionados.

- Con respecto a la fracción de muestreo, debido al tamaño desigual en la cantidad de observaciones por UPM, se compararon dos opciones (ajuste asumiendo PPT y ajuste en la segunda etapa de muestreo). La comparación consistió en valorar su rendimiento en la estimación de los indicadores de desempeño, determinando que el ajuste asumiendo PPT mostró mejor desempeño, ya que, en 62% de los ensayos realizados, este ajuste evidenció mejor rendimiento en comparación con el ajuste en la segunda etapa de muestreo (38%), por lo que se utilizó como fracción de muestreo en la comparación de técnicas de selección y estudio del tamaño de submuestreo.
- Como parte de la comparación de técnicas de selección y estudio de tamaños del submuestreo, se establecieron dos escenarios: uno general en el que se realizaron los procesos de simulación sin agrupación de UPM, mientras que en el segundo escenario se realizó una agrupación de UPM con base en el flujo vehicular (posestratificación), y en cada grupo generado, se aplicaron dichos procesos de simulación nuevamente. Para el caso del escenario general, se detallan seguidamente los principales hallazgos:
 - Con respecto a la estabilidad y precisión de las estimaciones, se encontró que, tanto en la proporción media como en el error estándar y el coeficiente de variación, se logró convergencia a partir de valores de “b” igual a 20 elementos por UPM seleccionada, lo

que sugiere que este valor sería el umbral mínimo requerido para disminuir la variabilidad de las estimaciones. En el caso del efecto del diseño, no se logró convergencia por aspectos relacionados con el tamaño de muestra.

- En relación con la comparación de técnicas de selección, se encontró que la técnica aleatoria evidenció mejor rendimiento en aproximadamente 76% de los ensayos realizados, y mostró menores diferencias en las estimaciones realizadas respecto al valor obtenido al utilizar la información de los conglomerados completos, en comparación con la selección sistemática (24% de los ensayos) y los primeros “n” casos (0% de los ensayos).
- En relación con el escenario de la posestratificación, debido a una dinámica vial diferenciada en función de la magnitud del flujo vehicular, se agruparon con análisis de conglomerados (procedimiento k-medias) las UPM en estratos de menor a mayor circulación de vehículos, para analizar posibles efectos en la estimación del uso de dispositivos de seguridad, obteniendo como principales resultados los siguientes:
 - Como parte del proceso de escogencia del número de grupos o estratos de flujo vehicular, se determinó, mediante diferentes pruebas, evaluaciones gráficas y valoración sustantiva, la opción de dos grupos como preferible para efectos de interpretación y, además, porque proporciona una mejor separación entre grupos, los cuales se nombraron como “medio flujo vehicular” y “alto flujo vehicular”.
 - Además, de igual forma que en el escenario general, se encontró estabilidad en las estimaciones de los indicadores de desempeño a partir de valores de “b” igual a 20 elementos por UPM seleccionada, lo que sugiere que este valor sería el umbral mínimo requerido para disminuir la variabilidad de las estimaciones
 - En este caso, tanto en el grupo de flujo vehicular alto como en el de flujo medio, la técnica aleatoria evidenció mejor rendimiento, ya que presentó mayor cantidad de ensayos en los que se obtuvieron menores diferencias en las estimaciones realizadas respecto al valor obtenido al utilizar la información de los conglomerados completos, 60% y 47% respectivamente, a diferencia del 24% (alto flujo) y 29% (flujo medio) de la selección sistemática y el 16% (alto flujo) y 24% (flujo medio) de la técnica de los primeros “n” casos.

A partir de los hallazgos anteriores, se concluye que, en términos de la mejor técnica de selección y tamaño óptimo de submuestreo para lograr mayor precisión y estabilidad en las estimaciones de los indicadores de desempeño, no existen diferencias entre el escenario general y la posestratificación.

Con respecto al efecto de los procesos de simulación en la exactitud de las estimaciones, los principales hallazgos se definen en los siguientes puntos:

- En el caso del error estándar y el coeficiente de variación, tanto en la estimación general como en el grupo de flujo vehicular medio, se evidenciaron valores similares al obtenido con los conglomerados completos, y en el caso del grupo de alto flujo vehicular, se experimentó un leve incremento en dicha estimación en comparación con el escenario general y el grupo de flujo medio.
- En lo referente al efecto del diseño, se evidenció una reducción en sus estimaciones para el escenario general y en la posestratificación, particularmente en el grupo de alto flujo vehicular –aunque, como se ha indicado en los resultados, dicho comportamiento se explica por un aspecto de tamaño de muestra.
- En relación con la proporción media para el caso del escenario general, se evidenció una subestimación entre las cifras obtenidas con las tres técnicas de selección y el valor de referencia, a diferencia de los resultados obtenidos en la posestratificación, donde las estimaciones presentaron mayor exactitud con respecto a su valor de referencia para las tres técnicas de selección en ambos grupos. Como consecuencia de las diferencias indicadas, se contrastaron las proporciones medias en ambos escenarios:
 - En el escenario general, se encontraron diferencias significativas ($\alpha = 0.05$) entre la proporción media estimada a partir de las simulaciones para las tres técnicas de selección (IC95% entre 0,74-0,77) y la proporción media de los conglomerados completos (0,78).
 - Mientras que, al replicar los contrastes en el grupo de flujo medio (IC95% entre 0,73-0,76) y alto flujo (IC95% entre 0,83-0,90), no se encontraron diferencias significativas con respecto al valor de la proporción media para estos dos grupos, 0,75 y 0,87 respectivamente.
 - El hecho de no encontrar diferencias significativas en los grupos de flujo vehicular es un indicativo de mejora en el control de la variabilidad atribuible al uso de la posestratificación, ya que se obtienen ganancias en la exactitud al estimar el indicador correspondiente a la proporción media, que en el escenario general no se lograba.
- El problema de la subestimación de la proporción media se explica por la naturaleza de los conglomerados (UPM con tamaños variables) y el tipo de submuestreo utilizado en la segunda etapa como parte de los procesos de simulación (submuestreo con “b” fijo por UPM), aunque dicha situación mejora con la aplicación de la posestratificación de UPM. Puede concluirse que la agrupación con base en el flujo vehicular brindó un mejor control

de las UPM para lograr mayor exactitud en la estimación de la proporción media de uso del cinturón de seguridad.

Con respecto al objetivo general del estudio, se analizó el comportamiento de la variabilidad de las estimaciones relacionadas con el uso de dispositivos de seguridad pasiva en ocupantes de carro que circularon en UPM de flujo vehicular medio y alto en Costa Rica para el 2016. Dicha valoración se logró a partir de la aplicación de diferentes procedimientos metodológicos con base en procesos de simulación desarrollados en la segunda etapa de muestreo del estudio, para analizar el efecto en la precisión y exactitud de las estimaciones mediante el uso de indicadores de desempeño que permitieron medir el impacto de dichos procedimientos.

Recomendaciones

Con base en el análisis realizado y los hallazgos obtenidos, es viable establecer algunas recomendaciones metodológicas como parte de posibles ajustes o cambios que pueden realizarse en el actual diseño de muestreo del estudio, en aras de solventar las limitantes descritas a lo largo del documento, para mejorar aspectos relacionados con el control de la variabilidad y cálculo de las estimaciones.

- En primer lugar, como alternativa práctica para tratar de solucionar la limitante relacionada con las UPM que presentaron bajo flujo de vehicular, se propone un diseño de muestreo para contrarrestar la baja circulación, generando mayor cantidad de observaciones y así realizar un cálculo adecuado de estimaciones. La propuesta consistiría en incrementar la amplitud de la franja horaria para obtener mayor cantidad de elementos en el estrato de flujo vehicular bajo. Para definir la amplitud de la franja horaria, se requiere realizar mediciones previas del flujo vehicular en los puntos de muestreo y así estimar el rango mínimo de tiempo necesario para obtener el submuestreo deseado.
- Se propone realizar mediciones previas del flujo vehicular en función del día de la semana y el rango horario en el que se aplicarían las observaciones, pues esto permitiría calcular el flujo medio aproximado de vehículos que circulan por UPM, permitiendo identificar en qué casos se podrían aplicar las técnicas de selección en función de un determinado submuestreo “b” y en qué casos se tendrían que valorar ajustes en la amplitud de la franja de tiempo destinada a las observaciones debido al bajo flujo vehicular.
- En cuanto a la selección de elementos se recomienda la técnica aleatoria, debido a que evidenció mejor rendimiento en la estimación de los indicadores de interés. Sin embargo, en escenarios como este (poblaciones móviles), su uso debe someterse a valoración, dado su

mayor dificultad de aplicación en función de los lineamientos de trabajo de campo, además de que los elementos que conforman el marco muestral solo se conocen hasta la definición de los procedimientos de submuestreo. Como alternativa, se puede recurrir a la técnica de selección sistemática, la cual mostró el mejor rendimiento después de la técnica aleatoria, y su procedimiento de aplicación resulta más simple. Además, la magnitud de las diferencias entre la técnica aleatoria y la sistemática en las estimaciones de los indicadores de desempeño, con respecto a los valores de referencia, resultaron reducidas.

- En el caso de aplicación de la técnica aleatoria en la práctica, pueden considerarse variaciones como las descritas por Kish (1972), que también se mencionan en el apartado conceptual del presente documento. Esta alternativa consiste en dividir la UPM en " kF_b " unidades para seleccionar " k " al azar y, de esta forma, la franja horaria para un determinado tramo se subdividiría en franjas de tiempo más pequeñas, de las cuales se seleccionaría cierta cantidad y en cada una se observarían los vehículos en circulación.
- En relación con la limitante de la subestimación de la proporción media y, como se indicó en los resultados del presente trabajo, en la práctica, en caso de submuestrear con " b " de tamaño fijo, un control de las UMP con base en el flujo vehicular (estratificación) es deseable, y se convierte en una herramienta a considerar para obtener ganancias en la exactitud de las estimaciones. En el caso de que esta metodología no resulte viable, se recomienda considerar en la segunda etapa un submuestreo que sea variable a partir de una fracción determinada que proporcione tamaños de submuestreo proporcionales a la UPM.
- A partir de los hallazgos obtenidos en el presente trabajo, es relevante realizar un análisis de costos para valorar la inversión de tiempo y recursos económicos en función de las ganancias en la calidad de los datos como parte de los cambios metodológicos recomendados al estudio.

Bibliografía

Allen, M., Fox, J. & Berg, E. (2020). *Iowa Seat Belt Use Survey, 2020 Data Collection, Methodology Report*. Center for Survey Statistics & Methodology, Iowa State University. <https://dps.iowa.gov/sites/default/files/commissioners-office/governors-traffic-safety/Occupant%20Protection/Seat%20Belt%20Use%202020%20Report%20final.pdf>

Berger, Y. & Tillé, Y. (2009). Sampling with Unequal Probabilities. *Handbook of Statistics* 29, part a, 39-54. http://doc.rero.ch/record/261148/files/Berger_Y.-Sampling-20160624143303-HU.pdf

Cochran, W. (1976). *Técnicas de muestreo*. México: C.E.C.S.A.

Comisión Nacional de Seguridad de Tránsito (CONASET) (2015). *Estudio de observación de uso de sistemas de retención, casco y distractores presentes en la conducción de vehículos livianos y motocicletas*. Chile: Ministerio de Transporte y Telecomunicaciones. <https://biblioteca.digital.gob.cl/bitstream/handle/123456789/1037/N2%202015%20Informe%20Final%20Estudio%20Observacion%20Sistemas%20de%20Retencion%20y%20casco%20C%20distractores%20presentes%20en%20la%20conduccion%20de%20vehiculos%20livianos.pdf?sequence=1&isAllowed=y>

Consejo de Seguridad Vial (Cosevi) (2017). *Estudio nacional para la medición del uso de dispositivos de seguridad pasiva (casco y cinturón) y factores distractores 2016*. (Informe inédito). San José: Área de Investigación y Estadística, Cosevi.

Freeman, L. & Schneider, H. (2019). *2019 Louisiana Seat Belt Observation Survey Results*. Louisiana State University. <https://www.lahighwaysafety.org/Documents/OCCUPANTPROTECTION/2019%20LA%20State%20wide%20Belt%20Survey%20Final.pdf>

Hartigan, J. & Wong, M. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1). <https://www.jstor.org/stable/2346830>

Haas, S. (2016). *West Virginia Observational Survey of Safety Belt Use, 2016*. Charleston, WV: Mountain State Criminal Justice Research Services, LLC. Division of Motor Vehicles West Virginia Department of Transportation. https://transportation.wv.gov/DMV/DMVFormSearch/Seat_Belt_Survey_2016.pdf

Himelein, K., Eckman, S. & Murray, M. (2014). Sampling Nomads: A New Technique for Remote, Hard-to-Reach, and Mobile Populations. *Journal of Official Statistics*, 30(2), 191-213. <https://www.sciendo.com/article/10.2478/jos-2014-0013>

Kalton, G. (1991). Sampling flows of mobile human populations. *Survey Methodology*, december 1991, 17(2), 183-194. Statistics Canada. <https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X199100214503>

Kalton, G. (2001). Practical methods for sampling rare and mobile populations. *Proceedings of the Annual Meeting of the American Statistical Association*, August 5-9, 2001. <http://www.asasrms.org/Proceedings/y2001/Proceed/00454.pdf>

Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R. Unsupervised Machine Learning*. Edición 1. <http://www.sthda.com/english/>

Kish, L. (1972). *Muestreo de encuestas*. México: Trillas.

Lammers, E., Agent, K. & Green, E. (2019). *Safety Belt Usage Survey in Kentucky. Research Report KTC-19-30/KSPI-17-1F*. Kentucky Transportation Center (KTC). <https://transportation.ky.gov/HighwaySafety/Documents/2019%20Safety%20Belt%20Usage%20Survey%20in%20Kentucky.pdf>

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 281-297. <https://projecteuclid.org/proceedings/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-mathematical-Statistics-and/Chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992>

Mondragón, A. (2002). ¿Qué son los indicadores?. *Notas. Revista de información y análisis*, (19). México: INEGI. https://www.orion2020.org/archivo/sistema_mec/10_indicadores2.pdf

Observatorio de Desarrollo (OdD) (2016). *Estudio observacional para la determinación de los factores de riesgo en accidentes donde están involucradas motocicletas en tres distritos de Costa Rica*. San José: Universidad de Costa Rica. <https://odd.ucr.ac.cr/sites/default/files/comportamiento-motocicletas/informe-estudio-observacional.pdf>

Puppo, C. et al. (2013). *Estudio de campo observacional urbano sobre el uso de cinturón de seguridad, sistemas de retención infantil en vehículos particulares y uso de cascos en motociclistas y ciclomotores en el Estado Plurinacional de Bolivia*. Bolivia: OISEVI. <https://es.scribd.com/document/350094994/Bolivia-Estudio-Del-Uso-de-Casco-Ops-Policia>

Risley, L. & Laubach, L. (2021). *Seat Belt Observation Study*. Department of Adult Education and Safety Sciences, College of Education and Professional Studies-University of Central Oklahoma. https://ohso.ok.gov/sites/g/files/gmc751/f/2021_seat_belt_observation_study_-_final.pdf

Telgarsky, M. & Vattani, A. (2010). Hartigan's Method: K-Means Clustering without Voronoi. Thirteenth International Conference on Artificial Intelligence and Statistics. *PMLR-Proceedings of Machine Learning Research*, 9, 820-827. <https://proceedings.mlr.press/v9/telgarsky10a.html>

Turner, P. & Hagelin, C. (2001). *Helmet Use by Motorcyclists: Florida Observational Survey Results*, Center for Urban Transportation Research, College of Engineering-University of South Florida. *Transportation Research Record*. https://www.researchgate.net/publication/281784573_Helmet_Use_By_Motorcyclists_Florida_Observational_Survey_Results

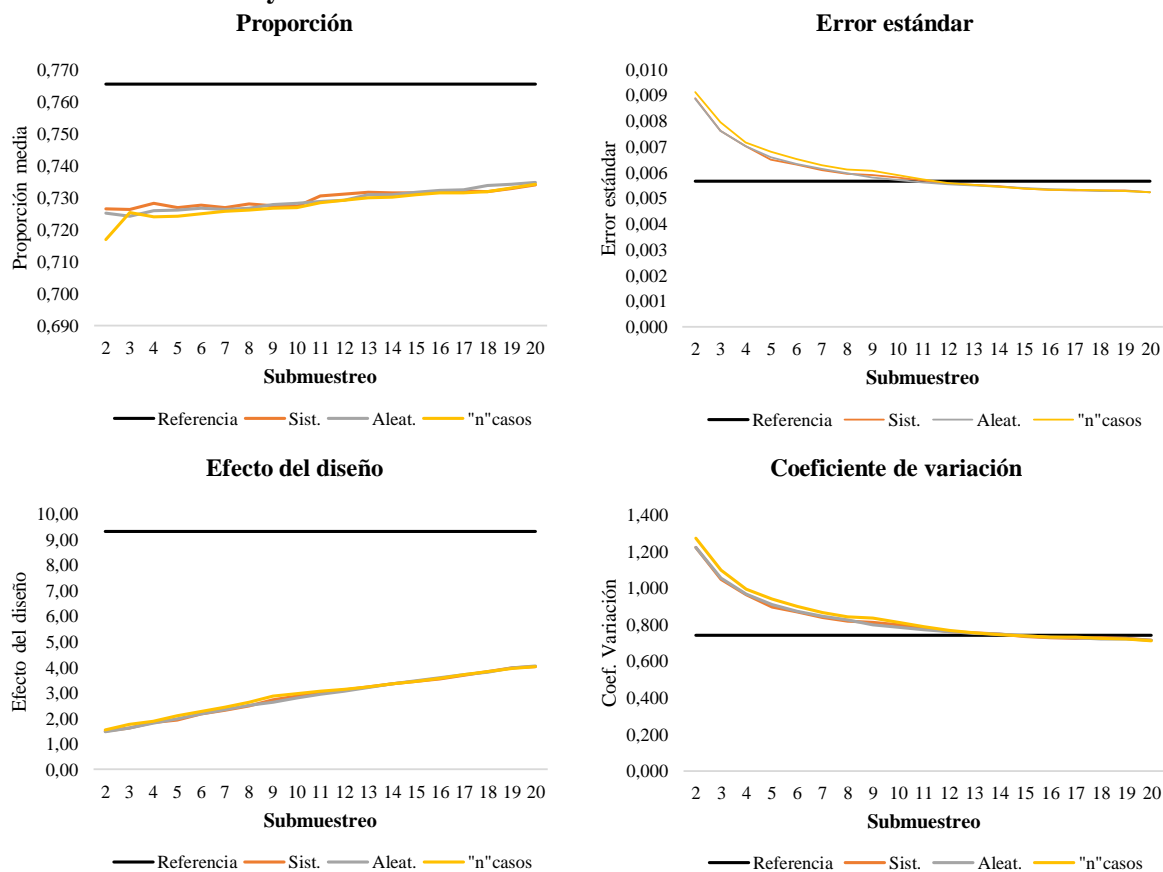
Vachal, K. & Andersen, J. (2020). *Seat Belt Use in South Dakota*. North Dakota State University. <https://dps.sd.gov/application/files/5616/1797/7908/statewide-seatbelt-use-report-sd-2020.pdf>

Vergara, A. (2016). *Primer estudio observacional sobre factores de riesgo vial de la Ciudad de Buenos Aires. Elaboración y confección de muestra representativa*. Documento-informe para el Observatorio de Seguridad Vial de la Dirección General del Cuerpo de Agentes de Tránsito y Seguridad Vial, Secretaría de Transporte de la Ciudad de Buenos Aires, Argentina.

Anexos

Gráfico A1

Estimación de indicadores de desempeño para el uso de cinturón de seguridad con diferentes técnicas de selección y variaciones del tamaño del submuestreo



Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Cuadro A1

Estimación de indicadores de desempeño para el uso de cinturón de seguridad con diferentes técnicas de selección y variaciones del tamaño del submuestreo

| Submuestreo | Proporción | | | Error estandar | | | Efecto del diseño | | | Coef. Variación | | |
|-------------|------------|--------|----------|----------------|--------|----------|-------------------|--------|----------|-----------------|--------|----------|
| | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos |
| 2 | 0,7644 | 0,7599 | 0,7452 | 0,0142 | 0,0139 | 0,0139 | 1,3843 | 1,3118 | 1,2685 | 1,8541 | 1,8286 | 1,8712 |
| 3 | 0,7603 | 0,7609 | 0,7519 | 0,0118 | 0,0118 | 0,0119 | 1,4267 | 1,4302 | 1,4153 | 1,5572 | 1,5575 | 1,5858 |
| 4 | 0,7599 | 0,7599 | 0,7553 | 0,0108 | 0,0107 | 0,0108 | 1,5845 | 1,5408 | 1,5487 | 1,4220 | 1,4023 | 1,4239 |
| 5 | 0,7601 | 0,7626 | 0,7536 | 0,0099 | 0,0099 | 0,0102 | 1,6575 | 1,6745 | 1,7121 | 1,3002 | 1,2980 | 1,3484 |
| 6 | 0,7595 | 0,7618 | 0,7583 | 0,0094 | 0,0093 | 0,0093 | 1,8018 | 1,7749 | 1,7671 | 1,2401 | 1,2220 | 1,2308 |
| 7 | 0,7598 | 0,7604 | 0,7538 | 0,0089 | 0,0090 | 0,0092 | 1,8889 | 1,9086 | 1,9882 | 1,1733 | 1,1776 | 1,2224 |
| 8 | 0,7598 | 0,7612 | 0,7527 | 0,0086 | 0,0086 | 0,0090 | 2,0151 | 2,0175 | 2,1792 | 1,1335 | 1,1303 | 1,2002 |
| 9 | 0,7603 | 0,7613 | 0,7544 | 0,0084 | 0,0084 | 0,0088 | 2,1471 | 2,1495 | 2,3595 | 1,1020 | 1,0992 | 1,1692 |
| 10 | 0,7594 | 0,7618 | 0,7531 | 0,0082 | 0,0081 | 0,0085 | 2,2965 | 2,2600 | 2,4246 | 1,0838 | 1,0683 | 1,1272 |
| 11 | 0,7605 | 0,7609 | 0,7520 | 0,0079 | 0,0080 | 0,0082 | 2,3242 | 2,3855 | 2,4734 | 1,0355 | 1,0493 | 1,0884 |
| 12 | 0,7592 | 0,7603 | 0,7533 | 0,0078 | 0,0078 | 0,0080 | 2,4658 | 2,5098 | 2,5863 | 1,0255 | 1,0321 | 1,0617 |
| 13 | 0,7580 | 0,7610 | 0,7535 | 0,0079 | 0,0077 | 0,0078 | 2,7109 | 2,6403 | 2,6705 | 1,0366 | 1,0153 | 1,0366 |
| 14 | 0,7593 | 0,7609 | 0,7547 | 0,0077 | 0,0076 | 0,0076 | 2,7995 | 2,7387 | 2,7455 | 1,0109 | 0,9964 | 1,0113 |
| 15 | 0,7591 | 0,7607 | 0,7559 | 0,0075 | 0,0075 | 0,0075 | 2,8731 | 2,8673 | 2,8346 | 0,9909 | 0,9855 | 0,9904 |
| 16 | 0,7584 | 0,7618 | 0,7559 | 0,0074 | 0,0074 | 0,0074 | 2,9563 | 2,9846 | 2,9307 | 0,9739 | 0,9710 | 0,9756 |
| 17 | 0,7594 | 0,7612 | 0,7578 | 0,0073 | 0,0073 | 0,0073 | 3,1029 | 3,0988 | 3,0868 | 0,9652 | 0,9612 | 0,9665 |
| 18 | 0,7590 | 0,7615 | 0,7579 | 0,0073 | 0,0073 | 0,0073 | 3,2402 | 3,2384 | 3,2046 | 0,9600 | 0,9539 | 0,9570 |
| 19 | 0,7587 | 0,7612 | 0,7584 | 0,0073 | 0,0072 | 0,0072 | 3,4404 | 3,3505 | 3,3645 | 0,9630 | 0,9455 | 0,9532 |
| 20 | 0,7590 | 0,7610 | 0,7598 | 0,0072 | 0,0071 | 0,0071 | 3,5244 | 3,4446 | 3,4404 | 0,9490 | 0,9348 | 0,9355 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Cuadro A2

Diferencias entre indicadores de desempeño para conglomerados completos y estimaciones según diferentes técnicas de selección y variaciones del tamaño del submuestreo

| Submuestreo | Proporción | | | Error estandar | | | Efecto del diseño | | | Coef. Variación | | |
|-------------|------------|---------|----------|----------------|--------|----------|-------------------|---------|----------|-----------------|--------|----------|
| | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos |
| 2 | -0,0201 | -0,0246 | -0,0393 | 0,0072 | 0,0069 | 0,0070 | -9,3627 | -9,4353 | -9,4785 | 0,9676 | 0,9421 | 0,9847 |
| 3 | -0,0242 | -0,0236 | -0,0326 | 0,0049 | 0,0049 | 0,0050 | -9,3204 | -9,3168 | -9,3318 | 0,6707 | 0,6711 | 0,6993 |
| 4 | -0,0246 | -0,0246 | -0,0293 | 0,0038 | 0,0037 | 0,0038 | -9,1626 | -9,2063 | -9,1984 | 0,5355 | 0,5158 | 0,5374 |
| 5 | -0,0244 | -0,0219 | -0,0309 | 0,0029 | 0,0029 | 0,0032 | -9,0896 | -9,0726 | -9,0349 | 0,4137 | 0,4115 | 0,4619 |
| 6 | -0,0250 | -0,0227 | -0,0262 | 0,0025 | 0,0024 | 0,0024 | -8,9453 | -8,9722 | -8,9800 | 0,3536 | 0,3355 | 0,3443 |
| 7 | -0,0247 | -0,0241 | -0,0307 | 0,0020 | 0,0020 | 0,0023 | -8,8582 | -8,8385 | -8,7589 | 0,2868 | 0,2911 | 0,3359 |
| 8 | -0,0247 | -0,0233 | -0,0318 | 0,0017 | 0,0016 | 0,0021 | -8,7320 | -8,7296 | -8,5679 | 0,2470 | 0,2438 | 0,3137 |
| 9 | -0,0242 | -0,0232 | -0,0301 | 0,0014 | 0,0014 | 0,0019 | -8,5999 | -8,5976 | -8,3876 | 0,2155 | 0,2127 | 0,2827 |
| 10 | -0,0251 | -0,0227 | -0,0314 | 0,0013 | 0,0012 | 0,0015 | -8,4505 | -8,4871 | -8,3224 | 0,1973 | 0,1818 | 0,2407 |
| 11 | -0,0240 | -0,0236 | -0,0325 | 0,0009 | 0,0010 | 0,0012 | -8,4228 | -8,3615 | -8,2736 | 0,1490 | 0,1628 | 0,2019 |
| 12 | -0,0254 | -0,0242 | -0,0312 | 0,0008 | 0,0009 | 0,0010 | -8,2812 | -8,2372 | -8,1608 | 0,1390 | 0,1457 | 0,1753 |
| 13 | -0,0265 | -0,0235 | -0,0310 | 0,0009 | 0,0008 | 0,0009 | -8,0361 | -8,1068 | -8,0766 | 0,1501 | 0,1288 | 0,1502 |
| 14 | -0,0252 | -0,0236 | -0,0298 | 0,0007 | 0,0006 | 0,0007 | -7,9476 | -8,0084 | -8,0015 | 0,1245 | 0,1099 | 0,1248 |
| 15 | -0,0254 | -0,0238 | -0,0286 | 0,0006 | 0,0005 | 0,0005 | -7,8740 | -7,8797 | -7,9124 | 0,1045 | 0,0990 | 0,1039 |
| 16 | -0,0261 | -0,0227 | -0,0286 | 0,0004 | 0,0004 | 0,0004 | -7,7907 | -7,7624 | -7,8164 | 0,0874 | 0,0845 | 0,0892 |
| 17 | -0,0251 | -0,0233 | -0,0267 | 0,0004 | 0,0004 | 0,0004 | -7,6442 | -7,6483 | -7,6602 | 0,0787 | 0,0747 | 0,0800 |
| 18 | -0,0255 | -0,0230 | -0,0266 | 0,0003 | 0,0003 | 0,0003 | -7,5068 | -7,5087 | -7,5424 | 0,0735 | 0,0675 | 0,0706 |
| 19 | -0,0258 | -0,0233 | -0,0261 | 0,0004 | 0,0002 | 0,0003 | -7,3067 | -7,3965 | -7,3825 | 0,0766 | 0,0590 | 0,0667 |
| 20 | -0,0255 | -0,0235 | -0,0247 | 0,0002 | 0,0002 | 0,0002 | -7,2227 | -7,3024 | -7,3066 | 0,0626 | 0,0484 | 0,0490 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Cuadro A3**Magnitud de las diferencias entre técnicas de selección en la estimación de indicadores de desempeño según variaciones del tamaño del submuestreo**

| Submuestreo | Proporción | | | Error estandar | | | Efecto del diseño | | | Coef. Variación | | |
|-----------------|---------------|--------------------|---------------------|----------------|--------------------|---------------------|-------------------|--------------------|---------------------|-----------------|--------------------|---------------------|
| | Sist./Alet. | Sist./ "n"casos | Aleat./ "n"casos | Sist./Alet. | Sist./ "n"casos | Aleat./ "n"casos | Sist./Alet. | Sist./ "n"casos | Aleat./ "n"casos | Sist./Alet. | Sist./ "n"casos | Aleat./ "n"casos |
| Promedio | 1,0604 | 0,8413 | 0,7909 | 1,0775 | 1,0004 | 0,9256 | 0,9979 | 1,0021 | 1,0042 | 1,0637 | 0,9498 | 0,8910 |
| 2 | 0,8170 | 0,5112 | 0,6256 | 1,0400 | 1,0314 | 0,9918 | 0,9923 | 0,9878 | 0,9954 | 1,0271 | 0,9827 | 0,9568 |
| 3 | 1,0273 | 0,7420 | 0,7222 | 0,9973 | 0,9821 | 0,9847 | 1,0004 | 0,9988 | 0,9984 | 0,9995 | 0,9591 | 0,9596 |
| 4 | 0,9995 | 0,8404 | 0,8408 | 1,0401 | 1,0125 | 0,9735 | 0,9953 | 0,9961 | 1,0009 | 1,0382 | 0,9964 | 0,9598 |
| 5 | 1,1112 | 0,7881 | 0,7092 | 0,9953 | 0,9127 | 0,9171 | 1,0019 | 1,0061 | 1,0042 | 1,0054 | 0,8956 | 0,8908 |
| 6 | 1,1006 | 0,9517 | 0,8646 | 1,0459 | 1,0349 | 0,9895 | 0,9970 | 0,9961 | 0,9991 | 1,0539 | 1,0268 | 0,9743 |
| 7 | 1,0258 | 0,8040 | 0,7838 | 0,9803 | 0,8669 | 0,8843 | 1,0022 | 1,0113 | 1,0091 | 0,9852 | 0,8537 | 0,8665 |
| 8 | 1,0611 | 0,7781 | 0,7334 | 1,0045 | 0,7961 | 0,7925 | 1,0003 | 1,0192 | 1,0189 | 1,0132 | 0,7873 | 0,7771 |
| 9 | 1,0436 | 0,8049 | 0,7713 | 1,0069 | 0,7623 | 0,7571 | 1,0003 | 1,0253 | 1,0250 | 1,0130 | 0,7623 | 0,7525 |
| 10 | 1,1052 | 0,7973 | 0,7215 | 1,0778 | 0,8313 | 0,7713 | 0,9957 | 1,0154 | 1,0198 | 1,0851 | 0,8197 | 0,7554 |
| 11 | 1,0156 | 0,7383 | 0,7270 | 0,8938 | 0,7476 | 0,8364 | 1,0073 | 1,0180 | 1,0106 | 0,9151 | 0,7380 | 0,8065 |
| 12 | 1,0473 | 0,8114 | 0,7747 | 0,9307 | 0,7960 | 0,8552 | 1,0053 | 1,0148 | 1,0094 | 0,9546 | 0,7934 | 0,8311 |
| 13 | 1,1288 | 0,8537 | 0,7563 | 1,1689 | 1,0536 | 0,9014 | 0,9913 | 0,9950 | 1,0037 | 1,1648 | 0,9994 | 0,8580 |
| 14 | 1,0686 | 0,8449 | 0,7907 | 1,1514 | 1,0645 | 0,9245 | 0,9924 | 0,9933 | 1,0009 | 1,1328 | 0,9973 | 0,8804 |
| 15 | 1,0672 | 0,8878 | 0,8318 | 1,0469 | 1,0663 | 1,0185 | 0,9993 | 0,9951 | 0,9959 | 1,0550 | 1,0050 | 0,9526 |
| 16 | 1,1502 | 0,9138 | 0,7945 | 0,9757 | 1,0250 | 1,0505 | 1,0036 | 0,9967 | 0,9931 | 1,0344 | 0,9804 | 0,9478 |
| 17 | 1,0786 | 0,9398 | 0,8712 | 1,0358 | 1,0143 | 0,9793 | 0,9995 | 0,9979 | 0,9984 | 1,0539 | 0,9836 | 0,9333 |
| 18 | 1,1077 | 0,9569 | 0,8639 | 1,0707 | 1,1096 | 1,0363 | 0,9998 | 0,9953 | 0,9955 | 1,0890 | 1,0415 | 0,9564 |
| 19 | 1,1048 | 0,9871 | 0,8935 | 1,4543 | 1,2822 | 0,8816 | 0,9879 | 0,9897 | 1,0019 | 1,2972 | 1,1471 | 0,8843 |
| 20 | 1,0866 | 1,0329 | 0,9506 | 1,5560 | 1,6183 | 1,0401 | 0,9891 | 0,9885 | 0,9994 | 1,2935 | 1,2761 | 0,9865 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Cuadro A4**Estimación de indicadores de desempeño para el uso de cinturón de seguridad en UPM de alto flujo vehicular con diferentes técnicas de selección y variaciones del tamaño del submuestreo**

| Submuestreo | Proporción | | | Error estandar | | | Efecto del diseño | | | Coef. Variación | | |
|-------------|------------|--------|----------|----------------|--------|----------|-------------------|--------|----------|-----------------|--------|----------|
| | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos |
| 2 | 0,8683 | 0,8618 | 0,8456 | 0,0328 | 0,0331 | 0,0291 | 1,2274 | 1,1971 | 0,8771 | 3,7954 | 3,8611 | 3,4444 |
| 3 | 0,8655 | 0,8656 | 0,8663 | 0,0284 | 0,0273 | 0,0245 | 1,3553 | 1,2567 | 1,0391 | 3,2934 | 3,1677 | 2,8242 |
| 4 | 0,8651 | 0,8619 | 0,8571 | 0,0248 | 0,0247 | 0,0222 | 1,3655 | 1,3277 | 1,0706 | 2,8764 | 2,8735 | 2,5948 |
| 5 | 0,8647 | 0,8636 | 0,8523 | 0,0221 | 0,0224 | 0,0219 | 1,3508 | 1,3972 | 1,2382 | 2,5588 | 2,6003 | 2,5734 |
| 6 | 0,8665 | 0,8643 | 0,8571 | 0,0214 | 0,0209 | 0,0212 | 1,5417 | 1,4527 | 1,4365 | 2,4722 | 2,4255 | 2,4744 |
| 7 | 0,8653 | 0,8632 | 0,8556 | 0,0204 | 0,0200 | 0,0202 | 1,6166 | 1,5469 | 1,4856 | 2,3612 | 2,3259 | 2,3634 |
| 8 | 0,8620 | 0,8634 | 0,8552 | 0,0194 | 0,0195 | 0,0206 | 1,6273 | 1,6699 | 1,7393 | 2,2480 | 2,2582 | 2,4030 |
| 9 | 0,8654 | 0,8640 | 0,8645 | 0,0189 | 0,0189 | 0,0194 | 1,7704 | 1,7587 | 1,8698 | 2,1824 | 2,1868 | 2,2439 |
| 10 | 0,8627 | 0,8635 | 0,8692 | 0,0184 | 0,0182 | 0,0193 | 1,8487 | 1,8077 | 2,1316 | 2,1392 | 2,1043 | 2,2228 |
| 11 | 0,8637 | 0,8647 | 0,8724 | 0,0174 | 0,0177 | 0,0181 | 1,8256 | 1,9215 | 2,0989 | 2,0132 | 2,0535 | 2,0766 |
| 12 | 0,8649 | 0,8626 | 0,8734 | 0,0176 | 0,0173 | 0,0175 | 2,0621 | 1,9547 | 2,1712 | 2,0408 | 2,0077 | 2,0073 |
| 13 | 0,8634 | 0,8639 | 0,8720 | 0,0173 | 0,0169 | 0,0175 | 2,1269 | 2,0259 | 2,3051 | 2,0043 | 1,9542 | 2,0030 |
| 14 | 0,8627 | 0,8641 | 0,8740 | 0,0173 | 0,0170 | 0,0172 | 2,2803 | 2,2266 | 2,4187 | 2,0062 | 1,9719 | 1,9636 |
| 15 | 0,8618 | 0,8644 | 0,8734 | 0,0166 | 0,0166 | 0,0171 | 2,2281 | 2,2767 | 2,5412 | 1,9238 | 1,9193 | 1,9552 |
| 16 | 0,8624 | 0,8652 | 0,8762 | 0,0160 | 0,0162 | 0,0165 | 2,2263 | 2,3210 | 2,5713 | 1,8559 | 1,8750 | 1,8824 |
| 17 | 0,8636 | 0,8630 | 0,8759 | 0,0160 | 0,0162 | 0,0166 | 2,3818 | 2,4286 | 2,7744 | 1,8531 | 1,8752 | 1,8944 |
| 18 | 0,8640 | 0,8645 | 0,8750 | 0,0163 | 0,0159 | 0,0166 | 2,6184 | 2,4987 | 2,9239 | 1,8822 | 1,8362 | 1,8983 |
| 19 | 0,8652 | 0,8648 | 0,8723 | 0,0157 | 0,0157 | 0,0163 | 2,6103 | 2,5740 | 2,9126 | 1,8203 | 1,8136 | 1,8682 |
| 20 | 0,8628 | 0,8635 | 0,8700 | 0,0161 | 0,0156 | 0,0165 | 2,8361 | 2,6557 | 3,0988 | 1,8721 | 1,8051 | 1,8939 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Cuadro A5

Diferencias entre indicadores de desempeño en conglomerados completos y estimaciones según diferentes técnicas de selección y variaciones del tamaño del submuestreo en el grupo de UPM de alto flujo vehicular

| Submuestreo | Proporción | | | Error estandar | | | Efecto del diseño | | | Coef. Variación | | |
|-------------|------------|---------|----------|----------------|--------|----------|-------------------|----------|----------|-----------------|--------|----------|
| | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos |
| 2 | 0,0010 | -0,0055 | -0,0217 | 0,0217 | 0,0220 | 0,0180 | -10,1373 | -10,1676 | -10,4875 | 2,5178 | 2,5835 | 2,1668 |
| 3 | -0,0017 | -0,0017 | -0,0009 | 0,0173 | 0,0162 | 0,0134 | -10,0094 | -10,1079 | -10,3255 | 2,0158 | 1,8901 | 1,5466 |
| 4 | -0,0022 | -0,0054 | -0,0101 | 0,0137 | 0,0136 | 0,0112 | -9,9991 | -10,0370 | -10,2941 | 1,5988 | 1,5959 | 1,3172 |
| 5 | -0,0026 | -0,0037 | -0,0150 | 0,0110 | 0,0113 | 0,0109 | -10,0138 | -9,9674 | -10,1264 | 1,2812 | 1,3227 | 1,2958 |
| 6 | -0,0007 | -0,0030 | -0,0101 | 0,0103 | 0,0099 | 0,0101 | -9,8229 | -9,9120 | -9,9281 | 1,1946 | 1,1479 | 1,1968 |
| 7 | -0,0026 | -0,0040 | -0,0117 | 0,0093 | 0,0090 | 0,0091 | -9,7480 | -9,8178 | -9,8790 | 1,0836 | 1,0483 | 1,0858 |
| 8 | -0,0052 | -0,0039 | -0,0121 | 0,0083 | 0,0084 | 0,0095 | -9,7373 | -9,6947 | -9,6254 | 0,9704 | 0,9806 | 1,1254 |
| 9 | -0,0019 | -0,0033 | -0,0028 | 0,0078 | 0,0078 | 0,0083 | -9,5942 | -9,6059 | -9,4949 | 0,9048 | 0,9091 | 0,9663 |
| 10 | -0,0046 | -0,0038 | 0,0020 | 0,0074 | 0,0071 | 0,0082 | -9,5160 | -9,5569 | -9,2330 | 0,8616 | 0,8267 | 0,9452 |
| 11 | -0,0035 | -0,0026 | 0,0051 | 0,0063 | 0,0067 | 0,0070 | -9,5390 | -9,4431 | -9,2657 | 0,7356 | 0,7759 | 0,7990 |
| 12 | -0,0024 | -0,0047 | 0,0061 | 0,0066 | 0,0062 | 0,0065 | -9,3025 | -9,4099 | -9,1934 | 0,7632 | 0,7301 | 0,7297 |
| 13 | -0,0039 | -0,0034 | 0,0048 | 0,0062 | 0,0058 | 0,0064 | -9,2377 | -9,3387 | -9,0595 | 0,7266 | 0,6766 | 0,7254 |
| 14 | -0,0046 | -0,0032 | 0,0068 | 0,0062 | 0,0059 | 0,0061 | -9,0843 | -9,1380 | -8,9459 | 0,7286 | 0,6943 | 0,6859 |
| 15 | -0,0055 | -0,0028 | 0,0062 | 0,0055 | 0,0055 | 0,0060 | -9,1365 | -9,0879 | -8,8234 | 0,6462 | 0,6416 | 0,6776 |
| 16 | -0,0049 | -0,0021 | 0,0089 | 0,0049 | 0,0051 | 0,0054 | -9,1384 | -9,0436 | -8,7933 | 0,5783 | 0,5974 | 0,6048 |
| 17 | -0,0037 | -0,0043 | 0,0086 | 0,0049 | 0,0051 | 0,0055 | -8,9829 | -8,9360 | -8,5902 | 0,5755 | 0,5976 | 0,6168 |
| 18 | -0,0033 | -0,0028 | 0,0077 | 0,0052 | 0,0048 | 0,0055 | -8,7462 | -8,8659 | -8,4407 | 0,6046 | 0,5586 | 0,6207 |
| 19 | -0,0021 | -0,0025 | 0,0051 | 0,0047 | 0,0046 | 0,0052 | -8,7543 | -8,7906 | -8,4520 | 0,5427 | 0,5360 | 0,5906 |
| 20 | -0,0045 | -0,0038 | 0,0027 | 0,0051 | 0,0045 | 0,0054 | -8,5286 | -8,7089 | -8,2658 | 0,5945 | 0,5275 | 0,6163 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Cuadro A6

Estimación de indicadores de desempeño en el uso de cinturón de seguridad en UPM de flujo vehicular medio con diferentes técnicas de selección y variaciones del tamaño del submuestreo

| Submuestreo | Proporción | | | Error estandar | | | Efecto del diseño | | | Coef. Variación | | |
|-------------|------------|--------|----------|----------------|--------|----------|-------------------|--------|----------|-----------------|--------|----------|
| | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos |
| 2 | 0,7499 | 0,7475 | 0,7328 | 0,0151 | 0,0149 | 0,0151 | 1,3519 | 1,3052 | 1,2862 | 2,0156 | 1,9942 | 2,0620 |
| 3 | 0,7487 | 0,7507 | 0,7379 | 0,0126 | 0,0126 | 0,0128 | 1,3956 | 1,4109 | 1,4115 | 1,6786 | 1,6770 | 1,7403 |
| 4 | 0,7487 | 0,7465 | 0,7430 | 0,0115 | 0,0115 | 0,0116 | 1,5610 | 1,5428 | 1,5557 | 1,5357 | 1,5376 | 1,5607 |
| 5 | 0,7497 | 0,7497 | 0,7419 | 0,0105 | 0,0105 | 0,0109 | 1,6311 | 1,6325 | 1,7137 | 1,4022 | 1,4022 | 1,4711 |
| 6 | 0,7484 | 0,7494 | 0,7466 | 0,0100 | 0,0099 | 0,0100 | 1,7623 | 1,7419 | 1,7490 | 1,3349 | 1,3229 | 1,3358 |
| 7 | 0,7478 | 0,7497 | 0,7420 | 0,0095 | 0,0095 | 0,0099 | 1,8519 | 1,8489 | 1,9755 | 1,2675 | 1,2613 | 1,3277 |
| 8 | 0,7476 | 0,7481 | 0,7410 | 0,0092 | 0,0092 | 0,0096 | 1,9905 | 1,9867 | 2,1526 | 1,2290 | 1,2273 | 1,2991 |
| 9 | 0,7478 | 0,7495 | 0,7417 | 0,0090 | 0,0089 | 0,0094 | 2,1239 | 2,1026 | 2,3129 | 1,1977 | 1,1860 | 1,2648 |
| 10 | 0,7479 | 0,7493 | 0,7396 | 0,0087 | 0,0086 | 0,0090 | 2,2474 | 2,2100 | 2,3410 | 1,1681 | 1,1546 | 1,2123 |
| 11 | 0,7487 | 0,7493 | 0,7381 | 0,0084 | 0,0085 | 0,0086 | 2,2810 | 2,3239 | 2,3755 | 1,1188 | 1,1288 | 1,1685 |
| 12 | 0,7482 | 0,7488 | 0,7393 | 0,0082 | 0,0083 | 0,0084 | 2,4072 | 2,4446 | 2,4779 | 1,1025 | 1,1096 | 1,1391 |
| 13 | 0,7458 | 0,7493 | 0,7397 | 0,0083 | 0,0081 | 0,0082 | 2,6262 | 2,5394 | 2,5520 | 1,1135 | 1,0853 | 1,1100 |
| 14 | 0,7474 | 0,7490 | 0,7409 | 0,0081 | 0,0080 | 0,0080 | 2,7060 | 2,6618 | 2,6154 | 1,0847 | 1,0715 | 1,0814 |
| 15 | 0,7471 | 0,7490 | 0,7423 | 0,0080 | 0,0079 | 0,0079 | 2,8041 | 2,7772 | 2,6967 | 1,0675 | 1,0570 | 1,0576 |
| 16 | 0,7462 | 0,7497 | 0,7421 | 0,0078 | 0,0078 | 0,0077 | 2,8440 | 2,8877 | 2,7769 | 1,0427 | 1,0419 | 1,0405 |
| 17 | 0,7469 | 0,7493 | 0,7441 | 0,0077 | 0,0077 | 0,0077 | 2,9992 | 3,0144 | 2,9256 | 1,0361 | 1,0340 | 1,0309 |
| 18 | 0,7466 | 0,7494 | 0,7443 | 0,0077 | 0,0077 | 0,0076 | 3,1480 | 3,1268 | 3,0343 | 1,0329 | 1,0233 | 1,0200 |
| 19 | 0,7468 | 0,7489 | 0,7452 | 0,0077 | 0,0076 | 0,0076 | 3,3173 | 3,2387 | 3,2053 | 1,0313 | 1,0151 | 1,0180 |
| 20 | 0,7469 | 0,7495 | 0,7470 | 0,0076 | 0,0075 | 0,0075 | 3,4013 | 3,3473 | 3,2763 | 1,0170 | 1,0041 | 0,9981 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Cuadro A7

Diferencias entre indicadores de desempeño en conglomerados completos y estimaciones según diferentes técnicas de selección y variaciones del tamaño del submuestreo en el grupo de flujo vehicular medio

| Submuestreo | Proporción | | | Error estandar | | | Efecto del diseño | | | Coef. Variación | | |
|-------------|------------|---------|----------|----------------|--------|----------|-------------------|---------|----------|-----------------|--------|----------|
| | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos | Sist. | Aleat. | "n"casos |
| 2 | -0,0018 | -0,0043 | -0,0189 | 0,0081 | 0,0079 | 0,0081 | -5,6483 | -5,6950 | -5,7140 | 1,0887 | 1,0673 | 1,1351 |
| 3 | -0,0031 | -0,0010 | -0,0138 | 0,0056 | 0,0056 | 0,0059 | -5,6046 | -5,5893 | -5,5887 | 0,7517 | 0,7502 | 0,8135 |
| 4 | -0,0031 | -0,0053 | -0,0087 | 0,0045 | 0,0045 | 0,0046 | -5,4392 | -5,4574 | -5,4445 | 0,6088 | 0,6108 | 0,6338 |
| 5 | -0,0021 | -0,0021 | -0,0098 | 0,0035 | 0,0035 | 0,0039 | -5,3691 | -5,3677 | -5,2865 | 0,4753 | 0,4753 | 0,5443 |
| 6 | -0,0033 | -0,0024 | -0,0051 | 0,0030 | 0,0029 | 0,0030 | -5,2379 | -5,2583 | -5,2512 | 0,4081 | 0,3960 | 0,4089 |
| 7 | -0,0039 | -0,0020 | -0,0097 | 0,0025 | 0,0025 | 0,0029 | -5,1483 | -5,1513 | -5,0247 | 0,3407 | 0,3345 | 0,4008 |
| 8 | -0,0042 | -0,0036 | -0,0108 | 0,0022 | 0,0022 | 0,0027 | -5,0097 | -5,0135 | -4,8476 | 0,3021 | 0,3005 | 0,3722 |
| 9 | -0,0039 | -0,0022 | -0,0101 | 0,0020 | 0,0019 | 0,0024 | -4,8763 | -4,8976 | -4,6873 | 0,2708 | 0,2591 | 0,3379 |
| 10 | -0,0038 | -0,0025 | -0,0121 | 0,0018 | 0,0017 | 0,0020 | -4,7528 | -4,7902 | -4,6592 | 0,2412 | 0,2277 | 0,2855 |
| 11 | -0,0030 | -0,0025 | -0,0136 | 0,0014 | 0,0015 | 0,0017 | -4,7192 | -4,6763 | -4,6247 | 0,1919 | 0,2019 | 0,2417 |
| 12 | -0,0036 | -0,0030 | -0,0124 | 0,0013 | 0,0013 | 0,0015 | -4,5930 | -4,5556 | -4,5223 | 0,1756 | 0,1827 | 0,2123 |
| 13 | -0,0060 | -0,0025 | -0,0120 | 0,0013 | 0,0012 | 0,0012 | -4,3740 | -4,4608 | -4,4482 | 0,1866 | 0,1584 | 0,1832 |
| 14 | -0,0043 | -0,0027 | -0,0108 | 0,0011 | 0,0011 | 0,0010 | -4,2942 | -4,3384 | -4,3848 | 0,1578 | 0,1446 | 0,1545 |
| 15 | -0,0047 | -0,0027 | -0,0094 | 0,0010 | 0,0009 | 0,0009 | -4,1961 | -4,2230 | -4,3035 | 0,1406 | 0,1301 | 0,1307 |
| 16 | -0,0055 | -0,0020 | -0,0097 | 0,0008 | 0,0008 | 0,0008 | -4,1562 | -4,1125 | -4,2233 | 0,1158 | 0,1151 | 0,1136 |
| 17 | -0,0049 | -0,0025 | -0,0076 | 0,0008 | 0,0008 | 0,0007 | -4,0010 | -3,9858 | -4,0746 | 0,1092 | 0,1071 | 0,1040 |
| 18 | -0,0051 | -0,0023 | -0,0074 | 0,0007 | 0,0007 | 0,0006 | -3,8522 | -3,8734 | -3,9659 | 0,1060 | 0,0965 | 0,0931 |
| 19 | -0,0049 | -0,0029 | -0,0065 | 0,0007 | 0,0006 | 0,0006 | -3,6829 | -3,7615 | -3,7949 | 0,1044 | 0,0882 | 0,0911 |
| 20 | -0,0049 | -0,0022 | -0,0047 | 0,0006 | 0,0006 | 0,0005 | -3,5989 | -3,6529 | -3,7239 | 0,0901 | 0,0772 | 0,0713 |

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.

Cuadro A8

Estimación de indicadores de desempeño para el escenario general utilizando selección aleatoria y fracción de muestreo $f_b = 1/3$ en la segunda etapa de muestreo

| Indicador | Conglomerados completos | Estimación | IC 95% para la proporción media estimada | |
|---------------------|-------------------------|------------|--|-----------|
| | | | Lim. Inf. | Lim. Sup. |
| Proporción | 0,7845 | 0,7846 | 0,7697 | 0,7996 |
| E. Estándar | 0,0070 | 0,0076 | | |
| E. Diseño | 10,7471 | 4,2493 | | |
| C. Variación | 0,8865 | 0,9734 | | |

Nota: Se define $f_b=1/3$ para generar un tamaño final de observaciones equivalente al que se obtiene cuando se submuestra con un "b" fijo igual a 20 elementos por UPM.

Fuente: Consejo de Seguridad Vial (2016). Elaborado a partir de datos del *Estudio nacional sobre uso de dispositivos de seguridad pasiva*.