

**Definición de perfiles de velocidad según el comportamiento del
usuario y sus efectos en la seguridad vial**

Por:

Cristian David Rosas López

Trabajo de investigación para optar por el título de
Magister en Ingeniería de Vías Terrestres
Modalidad investigación.

Universidad del Cauca
Facultad de Ingeniería Civil

2022

Definición de perfiles de velocidad según el comportamiento del usuario y sus efectos en la seguridad vial

Por:

Cristian David Rosas López

Informe final de investigación para optar por el título de
Magister en Ingeniería de Vías Terrestres en modalidad investigación.

Universidad del Cauca

Facultad de Ingeniería Civil

2022

Aprobado por:

Carlos Aníbal Calero Valenzuela, PhD.

Director

Fecha

Carlos Andrés Gaviria Mendoza, PhD.

Co-Director

Fecha

Sergio Andrés Córdoba Pareja, PhD.

Jurado

Fecha

Nestor Guillermo Ardila Arrieta, Mg.

Jurado Externo

Fecha

Resumen

En esta investigación se pretendió analizar el comportamiento de conductores en curvas horizontales en carreteras rurales de dos carriles debido a la importancia de conocer con rigor el comportamiento humano en estas facilidades, y aplicar con él, un análisis exhaustivo que permita una mejora considerable en la seguridad vial. A partir de datos tomados en Puerto Rico en carreteras rurales de dos carriles en curvas horizontales y en 5 puntos diferentes de la trayectoria del usuario en la curva, se aplicó el método de análisis de conglomerados, para conocer la distribución de los datos en agrupaciones, se evaluó la cantidad óptima de agrupaciones, se distribuyeron y se validó teniendo en cuenta diferentes formas de tratar los datos. Con el método se estimó la velocidad a partir de configuraciones de la geometría de la carretera, con el fin de proponer nuevas distribuciones que, al evaluar los riesgos de seguridad vial, permitan una relación con algunos tipos de comportamientos de usuarios y así plantear mejoras en las etapas del diseño de vías, tanto en prefactibilidad, factibilidad y diseños definitivos como en etapas de mejoramiento de trazados existentes. Los resultados mostraron que 3 agrupaciones es un número óptimo para el conjunto de datos, y apoyados en estudios recientes se decide distribuir en usuarios: cautelosos, moderados y agresivos. Después de validar diferentes formas de tratamiento de los datos se concluye que el método es aplicable y confiable, lo que permitió así encontrar la relación de las velocidades de cada tipo de comportamiento en relación a las diferentes variables geométricas de la carretera. Finalmente, se evaluaron segmentos de carreteras analizando la predicción de choques de estas a partir de metodologías que relacionan elementos geométricos, su presencia se corresponde a un tipo de comportamiento agresivo que pueda presentarse en las carreteras. Como conclusión, podemos decir que el comportamiento de los usuarios está directamente relacionado con la seguridad vial, si se evalúan estos comportamientos se puede ayudar a diseñar, construir y mantener carreteras más seguras. La validez del método, permite proponer estudios en contextos y tipos de carreteras distintas que generen conclusiones específicas y adaptables a las mismas. Incluir el comportamiento de usuarios en manuales de seguridad vial permitirá aterrizar la realidad para mejorarla de forma más eficaz.

Abstract

The purpose of this research was to analyze the behavior of drivers in horizontal curves on two-lane rural roads due to the importance of knowing with rigor the human behavior in these facilities, and to apply with it, an exhaustive analysis that allows a considerable improvement in road safety. Based on data taken in Puerto Rico in two-lane rural roads in horizontal curves and in 5 different points of the user's trajectory in the curve, the cluster analysis method was applied to know the distribution of the data in clusters, the optimal number of clusters was evaluated, distributed and validated taking into account different ways of treating the data. The method was used to estimate the speed from road geometry configurations, in order to propose new distributions that, when evaluating road safety risks, allow a relation with some types of user behavior and thus propose improvements in the road design stages, both in pre-feasibility, feasibility and definitive designs and in stages of improvement of existing layouts. The results showed that 3 groupings are an optimal number for the data set and, based on recent studies, it was decided to distribute them into cautious, moderate and aggressive users. After validating different ways of data treatment, it is concluded that the method is applicable and reliable, which allowed to find the connection between the speeds of each type of behavior in relation to the different geometries of the road. Finally, road segments were evaluated by analyzing the prediction of crashes of these from methodologies that relate geometric elements, their presence is related to a type of aggressive behavior that may occur on the roads. As a conclusion, we can say that the behavior of road users is directly related to road safety, if these behaviors are evaluated, it can help to design, build and maintain safer roads. The validity of the method allows us to propose studies in different contexts and types of roads that generate specific and adaptable conclusions. Including user behavior in road safety manuals will make it possible to ground reality in order to improve road safety more effectively.

Tabla de contenido

Resumen	iii
Abstract	iv
Lista de Imágenes	vii
Lista de Tablas	ix
Capítulo 1. Introducción	10
1.1. Problema de investigación	12
1.2. Justificación	15
1.3. Objetivos	17
1.3.1. Objetivo general	17
1.3.2. Objetivos específicos	17
1.4. Hipótesis	17
1.5. Base de datos seleccionada	18
Capítulo 2. Revisión de literatura	20
2.1. Diseño geométrico de carreteras	20
2.2. Consistencia de diseño	21
2.3. Seguridad vial	21
2.3.1. Método predictivo de choques	22
2.3.2. Factor humano	23
2.4. Estado infraestructura vial	24
2.5. Estadística	24
2.5.1. Estadística descriptiva	24
2.5.2. Análisis de varianza	25
2.5.3. Coeficiente de correlación de rango	25

2.5.4.	Análisis de Conglomerados	25
2.5.5.	Árbol de decisión	26
2.5.6.	Regresión lineal	27
Capítulo 3.	Metodología	29
Capítulo 4.	Resultados y conclusiones	31
4.1.	Resultados y análisis	31
4.1.1.	Estadísticas de los datos	31
4.1.2.	Numero óptimo de agrupaciones	34
4.1.3.	Análisis de conglomerados y distribuciones	35
4.1.4.	Discusión y validación	39
4.1.5.	Análisis cualitativo de árbol de decisión y cuantitativo de regresión lineal	48
4.1.6.	Análisis de estimación de accidentes y relación con perfiles de velocidad	52
4.2.	Recomendaciones	62
4.3.	Conclusiones	63
Capítulo 5.	Bibliografía	64
Capítulo 6.	Anexos	71
6.1.	Elección de cantidad óptima de conglomerados – Rstudio	71
6.2.	Distribución en conglomerados – Rstudio	74
6.3.	Gráficas de Elección de cantidad óptima de conglomerados – Matlab	79
6.4.	Árbol de decisión – Rstudio	85
6.5.	Análisis de regresión lineal – Stata	89

Lista de Imágenes

Imagen 1. Ubicación de los puntos de toma de datos. Imagen tomada (Google LLC., 2021)	18
Imagen 2. Puntos de toma de velocidad en la curva. Tomado de (Calero Valenzuela, 2015)	19
Imagen 3. Metodología específica del desarrollo del proyecto	30
Imagen 4. Resultados para el número óptimo de agrupaciones en los diferentes métodos	34
Imagen 5. Perfiles de velocidad identificados como centros de los tres conglomerados empleando diferentes estrategias de datos de entrada, de derecha a izquierda: solo velocidad, solo los parametros η y a_s^+ y velocidad con parámetros adicionales.....	37
Imagen 6. Agrupación de perfiles de velocidad usando diferentes estrategias en los datos de entrada y centro del conglomerado (línea negra). De izquierda a derecha: Agresivo (a), Moderado (b) y Cauteloso (c). De arriba a abajo: solo velocidades, solo parametros η y a_s^+ , velocidad con parametros adicionales y velocidad con parámetros adicionales ponderados.	38
Imagen 7. Histograma para el parámetro η . De izquierda a derecha: Agresivo (a), Moderado (b) y Cauteloso (c). De arriba a abajo: Solo velocidades, solo parametros η y a_s^+ y velocidad con parametros adicionales.	38
Imagen 8. Histograma para el parámetro a_s^+ . De izquierda a derecha: Agresivo (a), Moderate (b) y Cauteloso (c). De arriba a abajo: Solo velocidades, solo parametros η y a_s^+ y velocidad con parametros adicionales.	39
Imagen 9. Distribución de agrupaciones por rangos de radio en los diferentes análisis de datos	40
Imagen 10. Distribución de agrupaciones por tipo de terreno en los diferentes análisis de datos	40
Imagen 11. Distribución de agrupaciones por clasificación funcional entre ellas: local, colectora menor, colectora mayor y arterial menor.	41
Imagen 12. Distribución de agrupaciones por rangos de pendiente en los diferentes análisis de datos	41
Imagen 13. Distribución de agrupaciones por rangos de distancia de visibilidad en los diferentes análisis de datos.....	42

Imagen 14. Distribución de agrupaciones por rangos de curvatura en los diferentes análisis de datos	43
Imagen 15. Distribución de grupos de comportamiento con la misma cantidad de curvas por variable analizada.	44
Imagen 16. Distribución de grupos de comportamiento de toda la data en diferentes variables de análisis.	46
Imagen 17. Análisis de error.....	49
Imagen 18. Árbol de decisión.....	49
Imagen 19. Perfiles de velocidad tramo 110	59
Imagen 20. Perfiles de velocidad tramo 115 b	60
Imagen 21. Perfiles de velocidad tramo 102	60

Lista de Tablas

Tabla 1. Descripción estadística básica de la muestra.	31
Tabla 2. Estadísticas del conjunto de datos de 8 curvas ponderadas por clasificación funcional.	32
Tabla 3. Estadísticas del conjunto de datos de 28 curvas ponderadas por intervalos de radio.	32
Tabla 4. Estadísticas del conjunto de datos de 20 curvas ponderadas por intervalos de curvatura.	33
Tabla 5. Estadísticas del conjunto de datos de 33 curvas ponderadas por tipo de terreno.	33
Tabla 6. Estadísticas del conjunto de datos de 55 curvas ponderadas por intervalos de pendiente. Se evaluó la pendiente en ambos sentidos de circulación.	33
Tabla 7. Estadísticas del conjunto de datos de 52 curvas ponderadas por intervalos de distancia de visibilidad. Se evaluó la distancia de visibilidad en ambos sentidos de circulación.	33
Tabla 8. Perfil de velocidad tipo con variables adicionales.....	36
Tabla 9. Verificación grupo de control.....	50
Tabla 10. Resultado del modelo lineal para cada agrupación.	51
Tabla 11. Variables tramo 110, Terreno montañoso – Colector Mayor.	53
Tabla 12. Variables tramo 115b, Terreno ondulado – Arteria menor	54
Tabla 13. Variables tramo 102, Terreno plano – Colector Mayor.....	55
Tabla 14. Resumen de factores de modificación de accidentes. Tomada de SHM table 10-7 (AASHTO, 2010).....	56
Tabla 15. Tramo 110 cálculo de predicción de accidentes.	57
Tabla 16. Tramo 115-b cálculo de predicción de accidentes.....	57
Tabla 17. Tramo 102 cálculo de predicción de accidentes.	58

Capítulo 1. Introducción

Los accidentes de tránsito fueron a nivel mundial, la octava causa de muerte en el año 2016, según el informe de la situación de la seguridad vial realizado por la Organización Mundial de la Salud (World Health Organization, 2018), reportando más de 1.35 millones de muertes al año. La velocidad es uno de los factores más importantes de riesgo en la accidentalidad; dado que presenta una relación directa con la cantidad de accidentes y las víctimas fatales causadas por estos eventos (Secretary- General of the International Transport, 2020), convirtiéndose en una variable esencial para la evaluación del desempeño del diseño geométrico de carreteras y relacionándose directamente con la seguridad vial (Gibreel et al., 1999; Malaghan et al., 2020; McFadden & Elefteriadou, 2000; Misaghi & Hassan, 2005). Mediante la estrecha relación entre velocidad y seguridad vial se precisa que, al aumentar la velocidad, se incrementa la accidentalidad (Elvik et al., 2004).

El estudio de la velocidad vehicular permite analizar e identificar elementos en las vías, los cuales son susceptibles de redimensionamiento ayudando a mejorar la consistencia de diseño y la seguridad vial. El plan global para la década de Acción de la Seguridad Vial 2021-2030, continúa con los cinco pilares para lograr la meta y objetivos de la seguridad vial planteados en el plan de la década anterior: el primero, la gestión de la seguridad vial; el segundo, vías seguras y movilidad; el tercero, vehículos seguros; el cuarto, usuarios seguros de las vías y el quinto y último pilar; respuesta post-incidente. De estos pilares, los directamente relacionados con la carretera son: el segundo, el cual especifica comprometerse en el estudio de las relaciones de velocidades de diseño y de operación de las redes de carreteras, y el cuarto, que proyecta desarrollar programas integrales para mejorar el comportamiento de los usuarios en la carretera (United Nations Road Safety, 2011).

En Puerto Rico (lugar donde se obtiene la base de datos a trabajar en esta investigación, la cual se describe más adelante en este documento), los accidentes de tránsito fueron la principal causa de muerte para personas entre 11 y 27 años de edad y para el año de la toma de datos se tiene una cantidad de muertes aproximada de 4397 (Colucci Ríos & Valle Javier, 2014a). Adicional a esto se tiene un incremento de número de muertes con relación a distancia recorrida comparando a Puerto Rico con EEUU continental, por esta razón se aprueba el plan de seguridad vial estratégico de Puerto Rico, el cual analiza las principales causas que contribuyen a las muertes por accidentes

de tránsito y propone una revisión de políticas y estándares de ingeniería de seguridad vial (Colucci Ríos & Valle Javier, 2014b).

Desde este panorama, con el análisis de datos de velocidades en curvas horizontales, se pretendió identificar grupos de patrones que permitieran definir el comportamiento de los usuarios en diferentes tipos de curvas; los cuales pueden estar condicionados por ciertas variables geométricas u operacionales de estas curvas. Estudios previos (Calero Valenzuela, 2015; Castillo, 2020), han mostrado que las variables más significativas en la estimación de la velocidad operacional de curvas en vías rurales en Puerto Rico son la velocidad de diseño inferida (km/h), el radio (m) y el ángulo de deflexión (radianes). Además, los parámetros geométricos y la velocidad son en general datos sobre-dispersos como se observa en los análisis de tráfico (Rincón Numpaque et al., 2020; Rodríguez Polo & Henao Pérez, 2019). El comportamiento de los conductores se pudo identificar en diferentes puntos de las curvas, reconociendo que influyen en el incremento, descenso o constancia de la velocidad y que obedecen a obtener un recorrido eficiente y seguro.

Se logró identificar el método de aplicación de Clúster para la separación de grupos de comportamiento, definiendo la cantidad óptima de grupos de la separación. Se distribuyeron los datos en los grupos de análisis y se validó el método indagando diferentes formas de agrupación y aplicaciones de datos. Se aplicaron métodos estadísticos y matemáticos para la predicción del comportamiento de los usuarios. Finalmente, se comparó con la accidentalidad supuesta de algunos sectores, encontrada por medio de metodologías que utilizan configuraciones geométricas para su estimación, y se relacionaron con alguno de los tipos de comportamiento establecido previamente. Todo lo anterior nos permite concluir que es necesario evaluar más a fondo el comportamiento de los usuarios de las carreteras, así como definir las variables que influyen en estos comportamientos, lo cual permitirá generar acciones que pretendan carreteras más seguras.

1.1. Problema de investigación

El trazado de una carretera tiene como uno de sus fundamentos, considerar los parámetros geométricos para la elección correcta de la velocidad de diseño. La asociación estadounidense de funcionarios estatales de carreteras y transporte (AASHTO, por sus siglas en inglés), en la política sobre diseño geométrico de carretera y calles conocido como “Green book”, describe la velocidad como componente importante en el diseño geométrico y la define como un elemento determinante para la elección de las variables geométricas (AASHTO, 2018). Por otro lado, el Manual de Diseño Geométrico de Carreteras en Colombia (2008), también considera la velocidad de diseño como uno de los parámetros fundamentales para el diseño, además de la variabilidad en la topografía, en el entorno y en las dimensiones de sección transversal a lo largo de un trayecto. Establece la velocidad de tramo, que es la velocidad de diseño asignada a cada tramo (tramos divididos en longitudes de 3 km para velocidades hasta 50 km/h y de 4 km para velocidades mayores a 60 km/h, teniendo en cuenta que puede ser menor en casos especiales sin exceder una diferencia de 10 km/h con tramos adyacentes) de vía con características topográficas similares (Instituto Nacional de Vías, 2008). En ambos casos (en el Manual y en el “Green Book”) se ha considerado que, entre trayectos de diseño o tramos de estudio secuenciales, la diferencia de las velocidades de diseño no debe ser mayores de 20 Km/h, esto respondiendo al concepto de consistencia en el diseño geométrico.

La consistencia de diseño es un componente fundamental para el diseño adecuado y evaluación de las carreteras. Esta se basa, principalmente, en evaluar factores de velocidad, seguridad y desempeño de las vías terrestres (Gibreel et al., 1999). Se han identificado diversas metodologías que pueden evaluar la consistencia de diseño a partir de la velocidad operacional obtenida de modelos estadísticos, generados de muestras particulares que responden a contextos de países; estos modelos para la consistencia de diseño se realizan a partir de dos etapas: la primera etapa es la estimación de la velocidad operacional (Calero Valenzuela, 2015; Lamm et al., 1986, 1988; Llopis-Castelló et al., 2018; Morcillo et al., 2014; Polus & Mattar-Habib, 2004), estos desarrollan perfiles de velocidad de operación, que resultan de calcular el percentil 85 de una muestra de vehículos a flujo libre (Polus & Mattar-Habib, 2004); además son valorados por la segunda etapa, la cual a partir de las escalas permite comparar entre segmentos la calificación de la consistencia.

La velocidad que supone el comportamiento de los individuos es constante en cada elemento y se considera variable solo entre elementos (tangentes - curvas) (Morcillo et al., 2014). El Instituto Nacional de Vías (2008), indica que se presenta incumplimiento de la normativa por parte de los usuarios, por lo que asigna velocidades de diseño diferentes entre los elementos del alineamiento horizontal y vertical conocidas como las velocidades específicas.

El análisis de los perfiles de velocidad permite una comprensión más completa del comportamiento del conductor antes, durante y después de la curva (i.e. de la longitud de la curva). Sin embargo, algunas dificultades han sido reportadas en el proceso de establecer dichos perfiles, por ejemplo, Lareshyn et al. (2009), determinaron que generar agrupamiento de perfiles sin tener en cuenta la variación individual del mismo puede generar perfiles con sesgo que ocultan o limitan la variación en el comportamiento del usuario. Por otra parte, establecer un límite exacto para separar claramente los perfiles de velocidad no es una tarea sencilla y se han establecido indicadores adicionales como el porcentaje del tiempo en que la velocidad del conductor supera el 80% de la velocidad límite y la desviación estándar de la aceleración positiva para mejorar la definición de estos límites (Chu et al., 2017).

Adicionalmente, el Manual de Seguridad Vial estadounidense, al cual denotaremos por sus siglas en inglés HSM, evalúa el factor humano como un eje principal en la seguridad vial, dada la importancia que tiene con los efectos de la accidentalidad. Lo evalúa en cuatro grandes áreas: la atención y el procesamiento de la información, la visión, la percepción y el tiempo de reacción y la elección de velocidad (AASHTO, 2010). Esta última, de especial atención para el desarrollo de este estudio, demuestra que en vías sin divisiones, el impacto de decidir sobre qué velocidad adoptar por parte del conductor, se relaciona con la geometría de la carretera y es causante de accidentes; incluso estima que un tratamiento podría ser: dar mayor conocimiento de esta geometría para mejorar la seguridad vial. Para evaluar un método de seguridad vial, el HSM propone un cálculo predictivo de posibles accidentes, el cual a partir de eventos conocidos y analizados estadísticamente generan un valor probable de estos, el método utiliza el tránsito y la longitud del segmento para tener un valor estándar, el cual se verá modificado por los factores de modificación de accidentes, por sus siglas en inglés CMFs, los cuales son valores obtenidos a partir de elementos geométricos que presenta la vía (AASHTO, 2010). También, se han evaluado CMFs para realizar tratamientos de mejora de seguridad vial en las carreteras, los cuales disminuyen la

accidentalidad después de aplicarse en la vías y han sido comprobados por diferentes estudios (Federal Highway Administration, 2021).

Sin embargo, este método a pesar de ser altamente utilizado en pro de la seguridad vial, de evaluar sitios de accidentes antes y después de ellos y al analizar diferentes cambios de tiempos y factores tenidos en cuenta, establece que es posible que cambien las variables arrojando datos erróneos. Adicionalmente, no referencia la utilización de elementos de decisión humana de forma directa, lo que implica que no se está teniendo en cuenta el factor humano a plenitud. El HSM reduce el campo de acción a tratar con el factor humano a la educación, la legislación e incluso la penalización de dichos comportamiento, asegura que los métodos de predicción están calibrados pero no utilizan variaciones específicas en términos de contexto, distribución de edad, años de experiencia y, en general, factores de comportamiento humano (AASHTO, 2010).

1.2. Justificación

El sector del transporte cobra un significado importante en términos de economía para los países, por lo que la seguridad de las carreteras también implica una mejora sustancial en estos términos; Estados Unidos ocupa el primer puesto a nivel mundial en el valor agregado bruto por transporte, almacenamiento y comunicaciones, si se relaciona con Colombia, cuenta con cuarenta y tres (43) veces más agregado por actividad económica (NationMaster.com, 2021). Se debe reconocer que las vías terrestres son uno de los elementos más importantes de la infraestructura de un país, el plan nacional de desarrollo para Colombia en el periodo 2018-2022, en su capítulo referido al transporte y la logística para la competitividad y la integración regional, aclara entre sus objetivos se contempla la búsqueda de: *i*) la intermodalidad, *ii*) facilitación de comercio y *iii*) contar con modos de transporte competitivos, entre otras. Implementar estas estrategias y, en especial, la consolidación de concesiones viales de carreteras 4G, se podrá aumentar la productividad total del país en 0.25% a finales del 2022 (Plan-Nacional-Desarrollo, 2018).

Si el sector del transporte cobra un significado importante en términos de economía de los países, la seguridad en las carreteras también recibirá una mayor atención. La seguridad vial “se logrará mediante la implementación de varios acuerdos de infraestructura vial en el marco de las Naciones Unidas, la evaluación de la infraestructura vial y una planificación, diseño, construcción y operación de carreteras conscientes de la seguridad mejorada.” (p.13), como lo indica United Nations Road Safety (2011).

A partir de lo anterior, se puede mencionar que: Alonso Pla, (2016), identificó la relación directa que se presenta entre el factor humano y en la accidentalidad, también, Elvik et al., (2004), señalaron que la velocidad está relacionada directamente con la accidentalidad; además, se ha reconocido que la consistencia del diseño geométrico se basa en valorar la constancia de la velocidad a través de los diferentes métodos (Calero Valenzuela, 2015; Lamm et al., 1986, 1988; Llopis-Castelló et al., 2018; Morcillo et al., 2014; Polus & Mattar-Habib, 2004). Estas relaciones entre las variables: factor humano, velocidad y elementos de evaluación del diseño geométrico permiten establecer la importancia de valorar el factor humano utilizando la velocidad como variable de análisis, en especial, al considerarla con más de un perfil de comportamiento a diferencia de los modelos ya mencionados que utilizan un único perfil. Estos se desarrollan a partir

de obtener ecuaciones para estimar las velocidades, de esta manera es posible mejorar la consistencia de diseño y la seguridad en los proyectos viales existentes y proyectados.

Adicionalmente, si se logra identificar la importancia del factor humano en la predicción de choques, colocándolo como un factor de importancia como lo indica el HSM, pero adicional reconociendo de manera puntual la estrecha relación que pueden tener los comportamientos de los conductores con los elementos geométricos y la accidentalidad, se podrá abrir una brecha de estudio, que al contextualizarse y aplicarse permitirá mejorar sustancialmente la seguridad vial. Lo anterior permite evidenciar la relevancia que presenta el desarrollo del presente proyecto para el desarrollo del diseño geométrico de carreteras y ámbitos relacionados.

1.3. Objetivos

A continuación, se definen los objetivos propuestos que se pretenden alcanzar con la investigación, buscando responder al problema planteado a partir de la interpretación de los resultados obtenidos durante el proyecto.

1.3.1. Objetivo general

Incluir el factor humano en el análisis de la seguridad vial en las curvas horizontales mediante perfiles de velocidad generados por grupos de usuarios. A partir de este se plantean los objetivos específicos a continuación.

1.3.2. Objetivos específicos

- Definir la cantidad óptima de grupos para el análisis del comportamiento de los usuarios en las curvas horizontales mediante la estadística de conglomerados.
- Valorar la continuidad del comportamiento de los grupos establecidos a través de las variables geométricas de la configuración de la carretera.
- Estimar la elección de la velocidad que optan los grupos de usuarios cuando se presentan nuevas configuraciones de la geometría de la carretera.
- Proponer configuraciones de la geometría y evaluarlas a partir del Manual de Seguridad Vial (AASHTO, 2010).

1.4. Hipótesis

El análisis de la consistencia del diseño geométrico será robusto al considerar los perfiles de velocidad generados por grupos de usuarios a través de las curvas horizontales, de esta manera se mejorará la seguridad vial y los procesos de diseño geométrico.

1.5. Base de datos seleccionada

La base de datos que se utilizó para esta investigación es la recopilada en el proyecto de investigación para hallar una metodología para la evaluación de la consistencia de diseño de carreteras rurales, realizada por el PhD Carlos Aníbal Calero Valenzuela, la cual se desarrolló como requisito para obtener el grado de Doctor en Filosofía en Ingeniería Civil, en la Universidad de Puerto Rico, Recinto Universitario de Mayagüez en el año 2015. Contando con el consentimiento del autor para el uso de los datos y tomando información de su descripción de la Tesis titulada “Metodología para la Evaluación de la Consistencia de Diseño de Carreteras Rurales de Dos Carriles” (Calero Valenzuela, 2015).

Los sitios de toma de datos pueden ser vistos en la Imagen 1, estos puntos fueron seleccionados de manera que se tuviera heterogeneidad en diferentes variables como tránsito diario promedio anual, tipo de carreteras, tipos de terreno, entre otros. También cabe aclarar que la toma de puntos fue realizada en verano, donde la temperatura varía entre 24° a 32° (Consejo de Cambio Climático de Puerto Rico CCCPR, 2013).

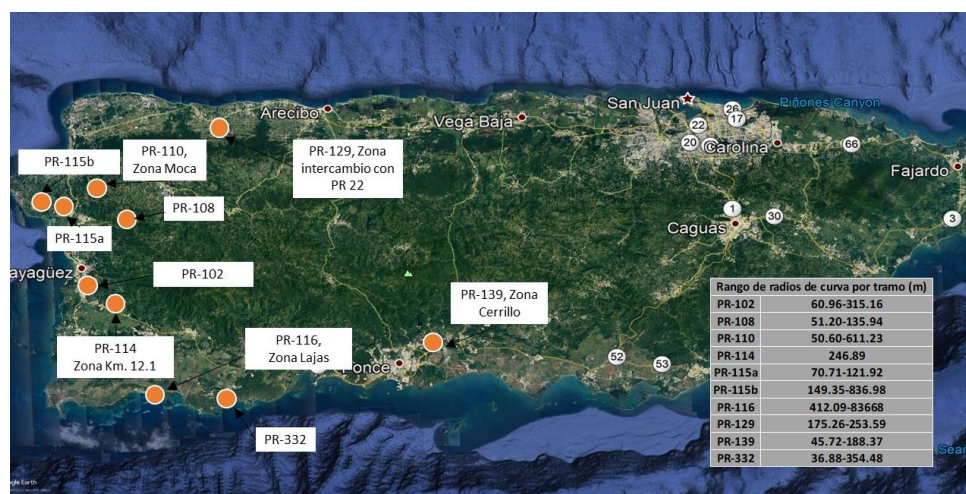


Imagen 1. Ubicación de los puntos de toma de datos. Imagen tomada (Google LLC., 2021)

Esta base de datos contiene la velocidad que presentan diferentes vehículos en un instante de tiempo, en la trayectoria que presenta el vehículo en curvas horizontales, tangentes previas y posteriores a la curva. Esta velocidad fue tomada a flujo libre, para evitar que fuera influenciada por el tráfico adicional, y como propone Poe et al., (1996), se toma a vehículos separados por lo menos por 5 segundos.

Las curvas analizadas para el estudio son escogidas de 10 diferentes carreteras rurales en Puerto Rico con un total de 38 curvas, las cuales se analizaron en los dos sentidos de avance y en 5 puntos de la trayectoria de la curva; en la mitad de la entretangencia de entrada y hasta a 200 metros del principio de curva, en el inicio de la curva (PC), en la mitad de la curva (MC), al final de la curva (PT) y en la mitad de la entretangencia de salida y a hasta 200 metros del final de la curva. La decisión de tomar la distancia de 200 metros fue debido a que a partir de una distancia superior a 400 metros se definen como curvas independientes (Instituto Nacional de Vías, 2008), por tanto, hasta 200 metros pudiera afectar la toma de decisión de la velocidad a flujo libre en los usuarios. Teniendo un total de 24250 datos de velocidad a analizar. En la Imagen 2 se puede identificar los puntos de toma de velocidades.

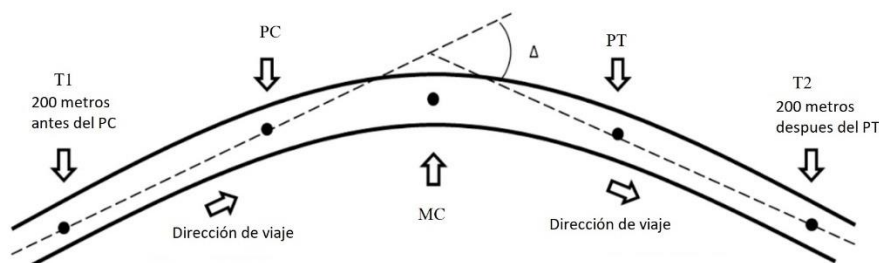


Imagen 2. Puntos de toma de velocidad en la curva. Tomado de (Calero Valenzuela, 2015)

La recolección de los datos fue realizada con el equipo MetroCount 5600, este dispositivo se instala de forma transversal a la vía y a una distancia determinada, consta de unas mangueras que capta la presión de los autos y calcula la velocidad a partir del tiempo y la distancia entre ellas. El equipo toma datos con una confiabilidad del 95% (MetroCount, 2021).

Algunos datos adicionales al respecto de cada curva, que se analizaron en el estudio y con los que se cuenta en la base de datos son: límite de velocidad, distancias de visibilidad, pendiente, ancho, derecho de vía, radio de curvatura, peralte, tránsito promedio diario, longitud de curva, entre otros. Una descripción más detallada se puede encontrar en la Tesis doctoral de Calero Valenzuela, (2015), y el análisis de estadísticos de las variables independientes se presentará en el desarrollo de la investigación.

Capítulo 2. Revisión de literatura

Los fundamentos teóricos que se presentan a continuación son el soporte en las diferentes ramas que abarcan el presente estudio. Se estudió previamente el diseño geométrico de carreteras, la consistencia de diseño, la seguridad vial, el estado de la infraestructura donde se tomaron los datos y los análisis estadísticos, en especial de conglomerados. Basándose en estos conceptos y analizando los resultados obtenidos se establecieron las conclusiones.

2.1. Diseño geométrico de carreteras

Las vías terrestres responden a la necesidad humana de comunicación, en Colombia responde al desarrollo económico y social, por lo cual cobra gran importancia en el desarrollo del país y en el bienestar de los ciudadanos, y considerando que la gran mayoría del transporte de personas y de carga es por medio de vías terrestres (Instituto Nacional de Vías, 2008).

La carretera se puede definir como la infraestructura destinada para el tránsito de vehículos motorizados, bicicletas y peatones en espacio y tiempo (Cárdenas Grisales, 2013). El diseño geométrico de la vía deberá ser funcional, que brinde comodidad y seguridad a los usuarios y que permita la movilidad, según su indicación, a una velocidad específica de operación (Cárdenas Grisales, 2013). Adicionalmente, se sugiere un enfoque sostenible al nivelar los impactos ambientales, económicos y sociales, y cumpliendo con los desempeños de la vía (AASHTO, 2018).

El proceso del correcto diseño geométrico de vías terrestres tiene como referente en Colombia un manual que el Gobierno Nacional dispuso para dicho fin, titulado “Manual de Diseño Geométrico de Carreteras” y publicado por el Instituto Nacional de Vías y el Ministerio de Transporte en el año 2008. Otras referencias en Colombia para el diseño de carreteras pueden ser el libro de Diseño Geométrico de Carreteras de Cardenas Grisales, (2013) y el libro Ingeniería de Vías de Carlos González Vergara et al., (2019). Finalmente, es de tener en cuenta la “Policy on Geometric Design of Highways and Streets” publicada en el año 2018 en su séptima edición y de autoría de la AASTHO, siendo esta una de las mayores autoridades académicas en la publicación de normas, protocolos, especificaciones y guías en relación con las carreteras y el transporte.

2.2.Consistencia de diseño

La consistencia de diseño es un término que nace como elemento importante en la evaluación de los diseños de carreteras desde el punto de vista de los conductores (Lamm et al., 1988). Se puede definir como el grado en que los elementos de una carretera están diseñados y que brindan una armonía para evitar maniobras críticas de conducción (Al-Masaeid et al., 1995). Es por esto por lo que el diseño correcto de las carreteras permite disminuir la accidentalidad y está muy relacionada con la escogencia de la velocidad de operación de las carreteras.

La consistencia de diseño se basa en el estudio de tres áreas dentro del diseño de carreteras, la consideración de la velocidad, la consideración de la seguridad y la consideración del desempeño de la vía (Gibreel et al., 1999). La consideración de velocidad está relacionada con la velocidad de operación y especialmente su estimación. La consideración seguridad se relaciona con la fricción que pueden tener los automóviles y su estabilidad. Por último, las consideraciones de rendimiento evalúan la carga de trabajo del conductores y factores de anticipación (Gibreel et al., 1999).

Para evaluación de la consistencia de diseño se han desarrollado diferentes métodos e índices que han sido estudiados por Llopis-Castelló et al., (2018); Morcillo et al., (2014); Polus & Mattar-Habib, (2004). Algunos autores que han trabajado índices de consistencia desde la velocidad en elementos verticales, horizontales o combinados son Calero Valenzuela, (2015); Fitzpatrick et al., (2000); Sánchez Ordóñez (2011). En relación con la consistencia de diseño, evaluando la fricción y seguridad se puede encontrar autores como: Lamm et al., (1991); y Sánchez Ordóñez, (2011). Finalmente, en términos de carga de trabajo podemos encontrar mediciones y referencias con Sánchez Ordóñez (2011) y Wooldridge et al., (2000).

2.3. Seguridad vial

Debido al importante número de incidentes y muertes por accidentes de tránsito evaluado en el Global Status Report on road safety (World Health Organization, 2018), la seguridad vial, su estudio y análisis ha cobrado relevancia. Algunos estudios para tener en cuenta y de importancia para evaluar la seguridad de las carreteras están cobrando significado para los países y las agencias gubernamentales. Se puede encontrar en la literatura elementos como: el riesgo de velocidad y colisión (Ms Anna Vadeby, Blair Turner, 2018) o las revisiones de capacidad de gestión de

seguridad vial y proyectos de sistemas seguros (Bliss & Breen, 2013) que permiten evaluar y mejorar los elementos que se tienen en las vías, en términos de seguridad.

Con el fin de mejorar la seguridad de las vías se han publicado planes de acción como el Global Plan For the Decade of Action for Road Safety (2011), el cual brinda pautas a considerar a nivel mundial, o el Plan Nacional de Seguridad Vial (2015) para Colombia. El Global plan for the decade of action for road safety se basa en 5 pilares los cuales procuran la mejora en varios ámbitos para la seguridad de las vías, estos pilares son: gestión de la seguridad vial, carreteras y movilidad más seguras, vehículos más seguros, usuarios viales más seguros y respuestas posteriores a los accidentes (United Nations Road Safety, 2011). En Colombia los pilares estratégicos de trabajo en seguridad vial son: la gestión institucional, el comportamiento humano, atención y rehabilitación a víctimas, infraestructura y de vehículos (Ministerio de Transporte, 2015).

El manual de seguridad vial de la AASHTO nos brinda una mirada importante en términos de seguridad vial, factores que la afectan, formas de prevenirla y estimaciones de choques para tener en cuenta su mejorarla (AASHTO, 2010). Dentro de los rasgos importantes de estudio es el factor humano muy influyente en la seguridad vial.

2.3.1. Método predictivo de choques

El manual de seguridad vial de la AASHTO propone un método predictivo para los choques según algunas características geométricas de las vías. Este método está basado en la revisión de eventos pasados y características en dichos eventos. El cálculo se basa en obtener una predicción de choques a partir del tránsito promedio diario anual (AADT) y la longitud del segmento de vía a analizar (L); en la Ecuación (1) podemos encontrar la determinación inicial de la predicción (AASHTO, 2010).

$$N_{sprfs} = AADT \times L \times 365 \times 10^{-6} \times x \quad (1)$$

Posteriormente, propone los modificadores de factores de accidentes, por sus siglas en inglés CMFs; estos según el tipo de vía, afectarán y, posiblemente, incrementarán la frecuencia de accidentes predicha. Algunos de los elementos que tienen en cuenta son: el ancho de carril, el tipo de berma, características de la curva horizontal, pendiente, grado de curvatura, tipo de separador, medidores de velocidad de usuarios, entre otros. Estos varían según el tipo de carretera a analizar. Adicionalmente, es posible encontrar soluciones propuestas para disminuir la estimación, también

por medio de CMFs, los cuales son aplicables cuando se han intervenido las vías en pro de una mejora en la seguridad vial; los estudios se pueden encontrar en la web de la Federal Highway Administration (2021), en donde están publicados los factores soportados en dichos estudios.

2.3.2. Factor humano

Se ha determinado que el factor humano y la velocidad están directamente relacionados con los accidentes (Alonso Pla, 2016). Sin embargo, al analizar los modelos y las metodologías para evaluar la consistencia del diseño geométrico, se ha identificado que: *i*) el factor humano no es considerado en la selección de la velocidad y, *ii*) no se involucra la variación de la velocidad a lo largo de los elementos geométricos ni entre ellos (i.e. se condiciona a un análisis con velocidad constante).

Una de las maneras de involucrar la variación de la velocidad a lo largo de los elementos geométricos es utilizar un perfil de velocidad analizando el factor humano; éste generaría que el estudio no se realice a partir de un dato único ponderado, esto puede influir en las variaciones de velocidad en tramos. Además, responde a los planteamientos de la falacia ecológica, la cual se refiere a posibles errores incurridos de pensar en variables que describen a los individuos en lugar de ser los mismos individuos quienes se definan, suponiendo un análisis muy globalizado sobre el particular (Piantadosi et al., 1988). Sin embargo, se han documentado varios inconvenientes en los esquemas de clasificación del perfil de velocidad que está directamente relacionado con el comportamiento del conductor; *i*) es difícil distinguir con precisión los límites de los perfiles de velocidad para diferentes estilos de conducción (Chu et al., 2017) y *ii*) una selección manual de perfiles de conglomerados de velocidad puede generar resultados más consistentes (Laureshyn et al., 2009). Por otro lado, se ha encontrado que existe una relación entre el perfil de velocidad real de la curva a flujo libre y el nivel de riesgo involucrado (Bermúdez Arbona, 2018), así como, el comportamiento de los conductores en esta relación (Dolatalizadeh et al., 2020).

El principal tema de investigación que utiliza el enfoque de simulación del conductor es el rendimiento de los usuarios de la carretera en las curvas horizontales (Bobermin et al., 2021). Las curvas horizontales son medidas de control de la velocidad para los conductores (Zolali et al., 2021). La edad y la experiencia del conductor son variables que influyen en las colisiones por deslizamiento a lo largo de una carretera rural de dos carriles (Choudhari & Maji, 2019). Generalmente, se sobreestima el tiempo y se subestima su velocidad en las curvas horizontales en

ambiente rural (Keklikoglou et al., 2018). Particularmente, la velocidad percibida puede ser menor que la velocidad real debido a un entorno monótono, sin embargo, la percepción mejora a medida que el usuario avanza más allá del punto medio de la curva (Liu & Xu, 2019). Existe una gran variabilidad en el comportamiento de la velocidad al inicio de la curva, y la desaceleración se detiene unos 135 metros después de entrar en la curva (Vos et al., 2021). Además, una desaceleración más temprana antes de entrar en la curva se convierte en una maniobra de mayor seguridad (Papadimitriou et al., 2017). En este sentido, los estudios de comportamiento agresivo personalizados deben ser preferidos sobre el análisis de datos de todo el conjunto de conductores (Jahangiri et al., 2019).

Es importante mencionar el acercamiento que existe en la conclusión empírica que realizó Rosas Palomino, (2019), donde al analizar los parámetros que influyen en la elección de velocidad para determinar una buena consistencia de diseño, pudo concluir la existencia de dichos perfiles.

2.4. Estado infraestructura vial

Debido a la importancia de conocer los elementos de estudios y así contextualizar los datos a analizar y la investigación a nivel general, es necesario reconocer el estado actual de las vías. Estudiando documentos como el reporte de infraestructura de Puerto Rico (2019), se podrá analizar elementos como capacidad de las vías, estado actual, tipos de vehículo, pavimentos, entre otros (ASCE, 2019). También es posible conocer este tipo de infraestructuras a partir de auditorías de seguridad vial en Colombia, planeadas a partir de manual titulado “La Metodología para el Desarrollo de Auditorías e Inspecciones de Seguridad Vial para Colombia” (Ministerio de Transporte, 2021).

2.5. Estadística

La estadística ha servido en diferentes estudios para analizar datos y observaciones de variados problemas, siendo una herramienta eficaz para validar conclusiones concretas. Algunos elementos para el análisis estadístico de esta investigación se presentan a continuación.

2.5.1. Estadística descriptiva

Es la estadística que permite ver de forma global y resumida los datos obtenidos, facilitando la forma de analizarlos o solo para conocerlos. Elementos como la media, la mediana, la desviación estándar, gráficas, entre otros, permiten un análisis global de los datos.

Adicionalmente, ayudan a clasificar y validar la naturaleza de los datos (Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, 2012).

2.5.2. Análisis de varianza

El análisis de varianza (ANOVA), responde a la estadística inferencial, que propone respuestas a partir de los datos obtenidos. Es utilizado cuando una variable dependiente se subdivide en componentes significativos, este procedimiento permite evaluar si las medias de dos o más variables son iguales o diferentes para concluir las hipótesis correctas (Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, 2012).

2.5.3. Coeficiente de correlación de rango

El análisis de diferentes tipos de variables permitirá relacionarlas y medir la fuerza que puedan tener entre ellas, demostrando relación nula cuando el coeficiente se acerca a cero. El coeficiente de correlación de Spearman permite obtener un equivalente no paramétrico del coeficiente de correlación convencional, permitiendo un análisis no lineal, sin hacer suposiciones de normalidad y establecer rangos sin mediciones numéricas significativas. (Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, 2012).

2.5.4. Análisis de Conglomerados

Los métodos para separar datos a partir de patrones de comportamiento, están definidos en general según la proximidad y la distancia que pueda haber entre los datos. La agrupación se definirá según el método *Kmeans*, el cual es uno de los más utilizados en formas de agrupamiento (Yang et al., 2020). El método de análisis de conglomerado, en inglés cluster analysis, pretende clasificar de forma general los elementos de un grupo en estudio, permitiendo una clasificación jerárquica (Baillo Moreno & Grané Chávez, 2008). El proceso se logra a partir de la aproximación de matrices de distancias según criterios adecuados de proximidad para llegar al objetivo de clasificar los elementos (Baillo Moreno & Grané Chávez, 2008). Finalmente, el método pretende unir elementos en similitud, proporcionando soluciones a problemas en patrones de agrupamiento, toma de decisiones, minería de datos, aprendizaje automático, entre otros. Resumiendo sus pasos de forma conceptual se puede decir: realiza representación de patrones, cálculo de similitudes, proceso de agrupación y representación de las mismas (Jain et al., 1999).

El primer análisis que se realiza a los datos, es averiguar cuál es el número óptimo de agrupación en las que se puede dividir los datos y se analiza cada curva por separado utilizando los métodos Elbow, Average Silhouette, GAP statistic y NbClust methods (DataCamp, 2021). Conocido el número de agrupaciones y con los datos normalizados (esta normalización se realiza con el fin de asegurar un peso igualitario en los diferentes elementos utilizados (Devore, 2008)), se aplica el método Kmeans para realizar la separación de los grupos, el cual consiste en agrupar según una posición espacial y las distancias a un punto aleatorio, este se va corrigiendo de manera iterativa con la Ecuación (2). A continuación, se describe el proceso paso a paso que propone el método, (i) Se asigna a cada elemento (perfil de velocidad) una posición aleatoria, (ii) Se calcula la distancia de cada elemento y se agrupan las que tengan valores similares, se muestra en la Ecuación (2) el proceso matemático, (iii) Se calcula la ubicación del centro del grupo por medio de la Ecuación (3) y (iv) se realiza un proceso iterativo entre los pasos (iii) y (iv) hasta minimizar la suma de las distancias de los elementos, este último se puede analizar en la Ecuación (4) (Fu et al., 2021).

$$\sum_{i=1}^n \left| D(X_{i+1}, C_j(k)) - D(X_i, C_j(k)) \right|^2 \quad (2)$$

$$C_j(r + 1) = \frac{1}{n} \sum_{i=1}^n X_i^{(j)} \quad (3)$$

$$E(r + 1) = \sum_{i=1}^k \sum_{w \in W} |w - C_j(r + 1)|^2, \quad (4)$$

En las ecuaciones las variables corresponden de la siguiente forma: donde, n es el número de iteraciones, k el número de conglomerados, $C_j(n)$ es el centro del conglomerado, para $j=1,2,\dots,k$, $D(X_{i+1}, C_j(k))$ es la distancia similar para $i=1,2,\dots,k$ y W_j es el centro del conglomerado. El resultado del anterior proceso le asigna un valor (1, 2, 3, ... , n) a cada perfil de velocidad analizado, lo que indica la agrupación a la cual fue asignado.

2.5.5. *Árbol de decisión*

El procedimiento de árbol de decisión, el cual se ha realizado para conocer el comportamiento de conductores (Dong & Zhou, 2020) y tomando como referencia el proceso propuesto por Yan et al., (2018). El método se basa en la partición del conjunto de datos en M regiones y se modela la respuesta como una constante para cada región. El método propuesto inicialmente por Breiman et al., (1984), propone una partición del conjunto de datos a partir de

variables independientes, pretende seleccionar de manera sucesiva grupos homogéneos. Además de ser una técnica con un resultado descriptivo importante para entender e interpretar, es apropiado para grandes conjuntos de datos y también se describe como un proceso secuencial, iterativo y descendente (Parra, 2016). Se inicia el método con la Ecuación (5), los datos un conjunto de O observaciones vectoriales, cada una de las longitudes tal que cada observación tiene una variable de respuesta y un predictor variable, el proceso completo se pueden encontrar en Hastie et al., (2011,p. 10).

$$\hat{f} = \sum_{m=1}^5 \text{cmI}\{(X_1, X_2 \in R_m)\} \quad (5)$$

2.5.6. Regresión lineal

Se conoce que un modelo de regresión lineal simple o múltiple permite relacionar variables de manera predictiva, el comportamiento de la variable dependiente estará en relación con la o las variables independientes que indican la forma en que pueda predecirse la variable dependiente. Se estudia un modelo de regresión lineal múltiple, con el objetivo de predecir el comportamiento de cada tipo de agrupación encontrada según los elementos variables de la geometría a la que corresponda. Se puede encontrar en la Ecuación (6) de forma generalizada (Hastie et al., 2011).

$$f(X) = \beta_0 + \sum_{j=1}^p X_j B_j \quad (6)$$

La estimación de los coeficientes β se realiza asumiendo que los errores aleatorios son independientes y están distribuidos de forma ideal ($\mu=0$ y $\sigma^2=1$), esto con el fin de utilizar el concepto de mínimos cuadrados y diferenciando la suma de los cuadrados de los residuos, se puede construir un sistema de ecuaciones normales para la regresión, que se soluciona por métodos de solución al sistema ya simplificado (Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, 2012, p. 449).

El coeficiente de determinación múltiple o R^2 , permite conocer cuando la variabilidad de los datos es adecuada para el modelo de regresión múltiple, valor que puede estar entre 0 y 1.0, cuando es cercano a 1.0 permite ver un modelo ajustado. La Ecuación (7) se puede inferir a partir de la suma de cuadrados de la regresión y la de la muestra.

$$R^2 = \frac{SCR}{STCC} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Finalmente, se puede pensar que si la regresión o sus componentes son significativas a partir de la suma de cuadrados de la regresión contrario de rechazar la hipótesis nula en que las constantes sean iguales a cero. Para esto con $p < 0.05$ se puede deducir que la variable contribuye al modelo, sin embargo, es tema de análisis según el investigador y la presencia de posible multicolinealidad. Para analizar a más profundidad temas como modelo ajustado, ortogonalidad, variables categóricas o modelos secuenciales se puede acudir a Probabilidad y Estadística para Ingeniería y Ciencias de Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, (2012).

Capítulo 3. Metodología

A continuación, se presenta la metodología definida para el desarrollo del proyecto, dicha estructuración fue considerada para abarcar cada uno de los objetivos específicos propuestos.

- Revisión de literatura
 - Definición de variables dependientes e independientes correspondientes al fenómeno a investigar.
 - Conocer y entender el método estadístico para conglomerados.
 - Establecer las características y estudios previos del factor humano en relación al comportamiento de los usuarios dentro de las carreteras.
 - Definir los parámetros a tener en cuenta en la seguridad vial y asociados al fenómeno que se estudiará.
- Preparación de la base de datos
 - Caracterización de base de datos de las velocidades operaciones tomadas en campo según variables independientes a trabajar.
 - Identificación de datos atípicos.
 - Organización de datos de las velocidades operaciones.
 - Normalización de los datos de velocidades operaciones.
- Análisis de datos.
 - Aplicación de métodos estadísticos para conglomerados.
 - Elección de número de grupos adecuado.
 - Elección de grupos de patrones de comportamiento de velocidades.
- Conclusiones y recomendaciones.
 - Definir a partir de los resultados encontrados los perfiles de velocidades.

En la Imagen 3, se presenta la metodología planteada, incluyendo en ella actividades, tareas y subtareas que se derivaron. Para el cumplimiento de cada objetivo específico se realiza en un diagrama. La estructura de la Imagen 3 presenta los objetivos específicos desglosados en las actividades principales que permitieron desarrollar la investigación, siempre con miras de poder llegar a cumplir el objetivo general vinculado de manera transversal en la metodología propuesta.

De manera general se observa que la revisión de literatura y la organización de datos hacen parte de las actividades para todos los objetivos planteados. La revisión de literatura parte

desde la formación en talleres de conocimiento en diferentes plataformas, búsquedas sistemáticas generales en cada aspecto a tratar, búsquedas específicas y organización de la información para el presente documento. La organización de datos, debido a que fue tomado de un estudio previo, se realizó el conocimiento de los métodos de toma de datos, depuración y organización para cada aspecto utilizado.

Las otras actividades, como la aplicación de los métodos usados, entre ellos: para número óptimo de conglomerados, del método *Kmeans* para distribución de agrupaciones, árbol de decisión, regresión lineal para perfiles de comportamiento y cálculo de frecuencia de choques, fueron soportadas en diferentes programas. Entre los softwares utilizados están: Rstudio, Matlab, Stata y Excel, con la investigación realizada y la organización de los datos, se construyeron los algoritmos de los programas y se obtuvieron los resultados que se relacionan en el Capítulo 4. Generalmente, siendo el software Excel el utilizado para visualizar de mejor manera los resultados a partir de gráficas y organización de los datos obtenidos.

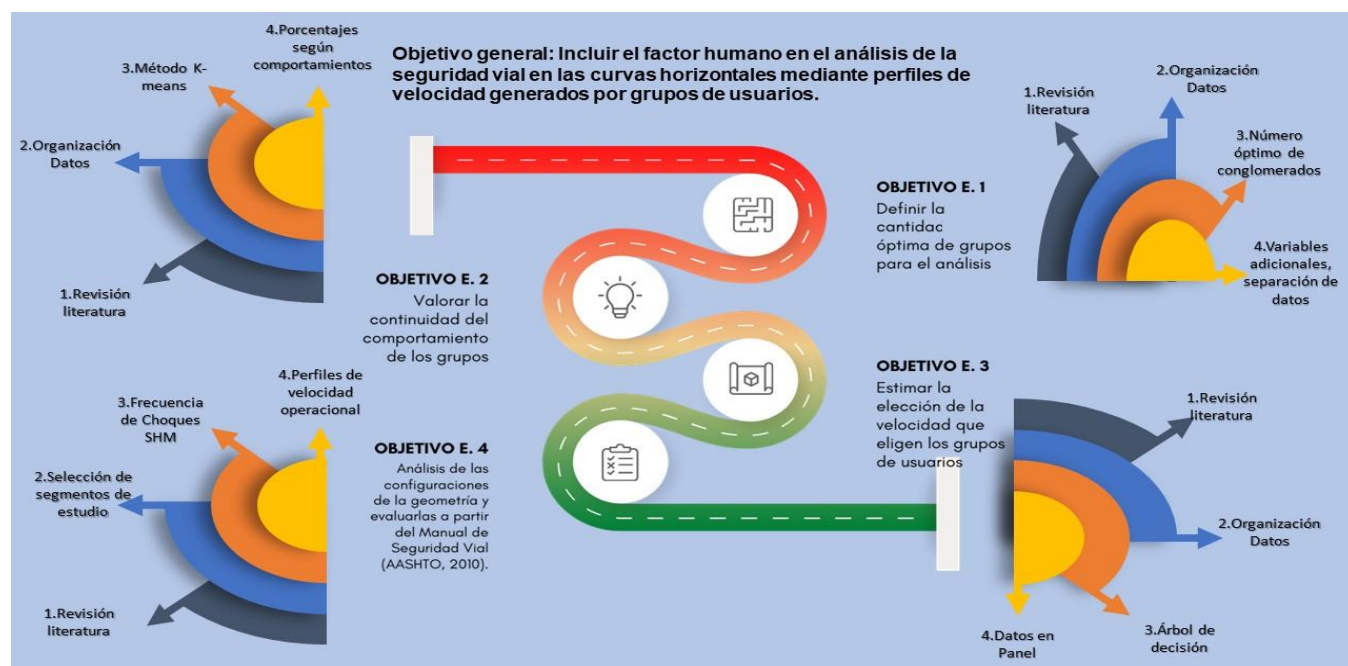


Imagen 3. Metodología específica del desarrollo del proyecto

Capítulo 4. Resultados y conclusiones

4.1. Resultados y análisis

A continuación, se presentan los resultados de los diferentes procesos metodológicos descritos anteriormente, estos se presentan de forma detallada, describiendo los resultados y conclusiones parciales que se pudieron ir obteniendo. En el Anexo 1 se puede encontrar la programación general utilizada en formato de texto en los diferentes softwares para los resultados obtenidos, esta programación podría copiarse en la interface de cada software y con los datos de la investigación o datos nuevos, se podría repetir los resultados presentados.

4.1.1. Estadísticas de los datos

En relación a los datos mencionados en el numeral 1.5 de este documento, podemos observar en la Tabla 1 la estadística descriptiva de los datos en relación a las 38 curvas estudiadas y a los 24250 perfiles de velocidad utilizados, 5 datos de velocidad puntual por perfil; en ella se observa que la varianza es mayor que el promedio para las variables asociadas a la geometría de la curva y para las velocidades en los diferentes puntos de la curva, lo cual se espera en datos reales de análisis de tráfico (ej. (Rincón Numpaque et al., 2020; Rodríguez Polo & Henao Pérez, 2019)), y demanda el uso de técnicas para análisis de datos sobre-dispersos.

Tabla 1. Descripción estadística básica de la muestra.

Variable	Min.	Max.	Promedio	Varianza
AADT (vpd)*	1491	20770	7071	26921268
r (m)*	36.9	1246.9	258.5	226064.2
L (km)*	64.5	729.3	248.1	106497.5
Pend (%) *	-10.7	10.7	± 4.76	± 12.39
V _{DIS} (km/h) *	32.2	185.1	77.1	810.2
V ₈₅ (km/h) *	42.9	97.1	68.0	102.4
V _{T1} (km/h) **	20.2	125.6	62.4	94.7
V _{PC} (km/h) **	28.8	129.6	60.8	99.1
V _{MC} (km/h) **	26.7	125.4	60.3	118.5
V _{PT} (km/h) **	23.1	125.3	60.8	115.9
V _{T2} (km/h) **	16.0	119.6	61.4	85.6

Notas: AADT: Tránsito medio diario anual r: Radio de curva; L: Longitud de curva; Pend: Pendiente; V_{DIS}: Velocidad de diseño; V_{XX}: Velocidad medida en el punto xx *La muestra es de 38 curvas; **La muestra es de 24250 Velocidades.

Debido a la dispersión, el análisis propone realizar agrupaciones utilizando criterios de separación que se apliquen a los datos, por esta razón se presentan en las Tablas 2 a la 7 las estadísticas geométricas de agrupaciones de datos por variables geométricas, teniendo en cuenta que se escogieron una misma cantidad de curvas por cada variable a estudiar y se seleccionaron 149 perfiles por cada curva, esto con el objetivo de que tuvieran un peso igualitario, en cada Tabla se menciona el total de curvas utilizadas.

Estas elecciones se realizaron de forma aleatoria y se seleccionaron: las variables geométricas; clasificación funcional, radio, curvatura, tipo de terreno, pendiente y distancia de visibilidad. El sustento de la utilización de éstas se debe a los criterios de diseño utilizados en los manuales de Colombia y Estados Unidos (Instituto Nacional de Vías, 2008; AASHTO, 2018), y a las utilizadas en investigaciones previas con los mismos datos, dichas investigaciones las consideran de importancia (Calero Valenzuela, 2015; Castillo, 2020) y, finalmente, debido al criterio de los investigadores que consideran de interés utilizar las variables mencionadas.

Tabla 2. Estadísticas del conjunto de datos de 8 curvas ponderadas por clasificación funcional.

Variable	Min	Max	Prom	Varianza
AADT (vpd)	2600	20770	6615.50	30677636.75
r (m)	36.88	496.21	221.48	74651.31
L (m)	65.53	503.27	207.36	59218.42
Pend (%)	-10.00	4.50	-2.38	24.36
V _{DIS} (km/h)	56.33	72.42	60.35	30.18

Notas: AADT: Tráfico diario promedio anual r: Radio de curva; L: Longitud de curva; Pend: Pendiente; V_{DIS}: Velocidad de diseño

Tabla 3. Estadísticas del conjunto de datos de 28 curvas ponderadas por intervalos de radio.

Variable	Min	Max	Prom	Varianza
AADT (vpd)	1491.00	20770.00	6237.82	21467791.22
r (m)	36.88	1246.94	261.99	277988.95
L (m)	64.55	729.32	242.98	122306.48
Pend (%)	-10.00	10.70	0.77	37.90
V _{DIS} (km/h)	56.33	72.42	59.20	23.61

Notas: AADT: Tráfico diario promedio anual r: Radio de curva; L: Longitud de curva; Pend: Pendiente; V_{DIS}: Velocidad de diseño

Tabla 4. Estadísticas del conjunto de datos de 20 curvas ponderadas por intervalos de curvatura.

Variable	Min	Max	Prom	Varianza
AADT (vpd)	1491.00	20770.00	5595.70	19584632.01
r (m)	36.88	628.50	167.02	82895.65
L (m)	64.55	729.32	267.56	160403.93
Pend (%)	-10.00	10.70	0.68	44.31
V _{DIS} (km/h)	56.33	72.42	59.55	25.75

Notas: AADT: Tráfico diario promedio anual r: Radio de curva; L: Longitud de curva; Pend: Pendiente; V_{DIS}: Velocidad de diseño

Tabla 5. Estadísticas del conjunto de datos de 33 curvas ponderadas por tipo de terreno.

Variable	Min	Max	Prom	Varianza
AADT (vpd)	1491.00	20770.00	7636.12	28382275.32
r (m)	36.88	1246.94	281.24	247174.02
L (m)	65.53	729.32	260.78	111569.16
Pend (%)	-10.00	10.70	1.13	30.11
V _{DIS} (km/h)	56.33	72.42	59.74	26.90

Notas: AADT: Tráfico diario promedio anual r: Radio de curva; L: Longitud de curva; Pend: Pendiente; V_{DIS}: Velocidad de diseño

Tabla 6. Estadísticas del conjunto de datos de 55 curvas ponderadas por intervalos de pendiente. Se evaluó la pendiente en ambos sentidos de circulación.

Variable	Min	Max	Prom	Varianza
AADT (vpd)	1491.00	20770.00	7385.04	30627280.44
r (m)	36.88	1246.94	247.34	210585.94
L (m)	67.17	729.32	243.52	95409.11
Pend (%)	-10.70	10.70	0.14	41.40
V _{DIS} (km/h)	56.33	72.42	59.25	23.94

Notas: AADT: Tráfico diario promedio anual r: Radio de curva; L: Longitud de curva; Pend: Pendiente; V_{DIS}: Velocidad de diseño

Tabla 7. Estadísticas del conjunto de datos de 52 curvas ponderadas por intervalos de distancia de visibilidad. Se evaluó la distancia de visibilidad en ambos sentidos de circulación.

Variable	Min	Max	Prom	Varianza
AADT (vpd)	1491.00	20770.00	6818.82	24752337.20
r (m)	36.88	1246.94	286.45	276711.78
L (m)	64.55	729.32	269.16	133110.82
Pend (%)	-10.20	10.20	0.35	30.74
V _{DIS} (km/h)	56.33	72.42	60.11	28.96

Notas: AADT: Tráfico diario promedio anual r: Radio de curva; L: Longitud de curva; Pend: Pendiente; V_{DIS}: Velocidad de diseño

En las Tablas 2 a la 7 podemos reafirmar el comportamiento sobre disperso planteado para toda la data, como se observa en la Tabla 3, y cumple el supuesto de análisis para estudios de tránsito mencionado al principio de este numeral. También podemos intuir, en primera instancia, que al utilizar toda la data el análisis se comportará de forma parecida que cuando se utiliza de forma agrupada por variables, con una igual cantidad de curvas en cada variable y mismo número de perfiles estudiados.

4.1.2. Numero óptimo de agrupaciones

El número óptimo de agrupaciones de perfiles de velocidad es mostrado en la Imagen 4 para los diferentes métodos implementados. En particular, un cambio significativo en la pendiente del método del codo (Elbow, Imagen 4 (a)) muestra que el número óptimo de clúster se ubica entre 2 y 4. Esta tendencia se presenta en la mayoría de las otras metodologías y cuyo resultado se resume en la Imagen 4 (f). Los resultados son diferentes para cada método debido a que los datos son analizados en formas alternas. Teniendo en cuenta que se pretende contar con perfiles de velocidad que puedan ser asociados a una serie de patrones de conducción de los usuarios de las vías, tres perfiles de velocidad, o número de conglomerados, son seleccionados debido a que permiten mantener una representación unificada y elemental de estilos de conducción: (a) agresivo, (b) moderado o (c) cauto que han sido usados en el pasado (Chu et al., 2017) y presentan particulares en cuanto a su velocidad inicial (i.e punto T_1) y el patrón de aceleración/desaceleración (Papadimitriou et al., 2017).

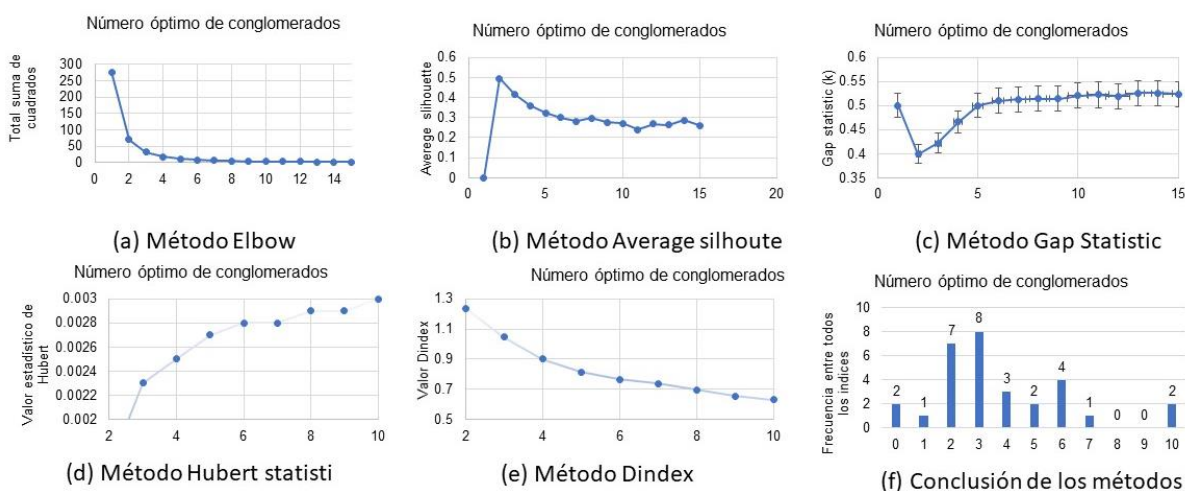


Imagen 4. Resultados para el número óptimo de agrupaciones en los diferentes métodos

4.1.3. Análisis de conglomerados y distribuciones

Una vez se define el número óptimo de conglomerados, se recurre a la técnica de clasificación Kmeans (Jahangiri et al., 2019) para generar las agrupaciones (paso 2) empleando el número de clústeres adecuado y establecido en el paso 1. En esta investigación se evaluaron inicialmente tres estrategias para el agrupamiento, con el fin de validar la capacidad de estos para definir los límites de los perfiles de velocidad para diferentes estilos de conducción (Chu et al., 2017). La primera se basa en usar solo los datos de velocidad medidos en los puntos de análisis de la curva (i.e. T_1 , P_C , M_C , P_T y T_2). La segunda estrategia se fundamenta en ampliar dos variables derivadas de los datos de velocidad únicamente: (1) la fracción del tiempo en que la velocidad del conductor es superior al 80% del límite de velocidad (η) y (2) la desviación estándar de la aceleración positiva únicamente (a_s^+). Estas variables permiten contar con indicadores de valor único por cada vehículo registrado, en lugar de una serie de valores de velocidad (como ocurre en el perfil del velocidad) y han permitido distinguir con mayor precisión los límites de los perfiles de velocidad para diferentes estilos de conducción (Chu et al., 2017). La tercera estrategia emplea los datos de velocidad y las dos variables inferidas a partir de las velocidades y las características de la curva.

Para el cálculo de la fracción del tiempo en que la velocidad del conductor es superior al 80% del límite de velocidad (η) se siguen los siguientes pasos para cada perfil de velocidad: (i) se computa la distancia a la que se encuentra cada punto de la curva de acuerdo con su geometría, (ii) con la velocidad registrada se calcula el tiempo del recorrido entre dos puntos consecutivos de medición y, (iii) se suman los tiempos en que la velocidad del conductor es superior al 80% del límite de velocidad y se divide en el tiempo total del recorrido entre el punto inicial y final de estudio. Por otra parte, la desviación estándar de la aceleración positiva únicamente (a_s^+) se estima calculando las diferencias de velocidades entre puntos consecutivos, con el tiempo de recorrido entre estos puntos, se computa la aceleración y, finalmente, se valora la desviación estándar de los valores positivos de aceleración encontrados durante el recorrido del vehículo.

En la Tabla 8 se puede ver el esquema tipo de un perfil de velocidad utilizado para el estudio con las variables adicionales. También se adjunta un ejemplo numérico de cómo se realizó el cálculo.

Tabla 8. Perfil de velocidad tipo con variables adicionales.

T ₁ (km/h)	PC (km/h)	MC (km/h)	PT (km/h)	T ₂ (km/h)	η	a _s ⁺ (m/s ²)
55.89	49.10	50.29	50.90	47.51	1.00	0.0303

Fracción del tiempo en que la velocidad del conductor es superior al 80% del límite de velocidad (η).

$$Dis_{T1PC} = 200 \text{ m (Constante por sitio de toma de datos)}$$

$$Dis_{PCMC} = 146/2 \text{ m (Longitud de curva dividida entre 2)}$$

$$Dis_{MCPT} = 146/2 \text{ m (Longitud de curva dividida entre 2)}$$

$$Dis_{PTT2} = 200 \text{ m (constante por sitio de toma de datos)}$$

$$Tiempo_{xxyy} = \frac{Dis_{xxyy}}{\frac{Vel_{xx} + Vel_{yy}}{2}}$$

$$Tiempo_{T1PC} = 14 \text{ ; } Tiempo_{PCMC} = 5 \text{ s ; } Tiempo_{MCPT} = 5 \text{ s ; } Tiempo_{PTT2} = 5 \text{ s ;}$$

$$Vel_{diseño} = 56.33 \text{ km/h}$$

$$Tiempo_{xxyy} = \text{si; } \frac{Vel_{xx} + Vel_{yy}}{2} > 80\% \text{ de } Vel_d$$

$$Tiempo_{T1PC} = 14 \text{ s ; } Tiempo_{PCMC} = 5 \text{ s ; } Tiempo_{MCPT} = 5 \text{ s ; } Tiempo_{PTT2} = 15 \text{ s}$$

$$\sum Tiempo_{xxyy} = 39 \text{ s}$$

$$\eta = \frac{\sum Tiempo_{xxyy}}{Tiempo \text{ de recorrido}} = \frac{39 \text{ s}}{39 \text{ s}} = 1.0$$

La desviación estándar de la aceleración positiva únicamente (a_s⁺).

$$DifVel_{xxyy} = Vel_{xx} - Vel_{yy}$$

$$DifVel_{T1PC} = -1.89 \text{ m/s ; } DifVel_{PCMC} = 0.33 \text{ m/s ;}$$

$$DifVel_{MCPT} = 0.17 \text{ m/s ; } DifVel_{PTT2} = -0.94 \text{ m/s ;}$$

$$Acel_{xxyy} = DifVel_{xxyy} / Tiempo_{xxyy}$$

$$Acel_{xxyy} = -0.14 \frac{m}{s^2}; Acel_{xyyy} = 0.06 \frac{m}{s^2}; Acel_{xyxy} = 0.03 \frac{m}{s^2}; Acel_{yyxy} = -0.06 \frac{m}{s^2}$$

Se considera solamente la aceleración positiva

$$a_s^+ = \text{Desviación estandar} \left(Acel_{xxyy} (\text{positivas}) \right) = 0.0303 \text{ m/s}^2$$

Los resultados de la segunda parte del esquema propuesto (es decir, agrupación de los perfiles de velocidad de los vehículos) para las tres estrategias evaluadas se presentan en las Imagen 5 y 6. La Imagen 5 muestra que hay una tendencia a separar los tres tipos fundamentales de estilos de conducción (identificados como “a”, “b” o “c”) al utilizar cualquiera de las tres estrategias de datos de entrada en el análisis. Cuando se utilizan datos de velocidad en cada punto de toma de datos, se presenta una mayor distinción en los valores de velocidad en los puntos T₁ y T₂ para cada estilo de conducción (ver Imagen 6, gráficos izquierdo y derecho).

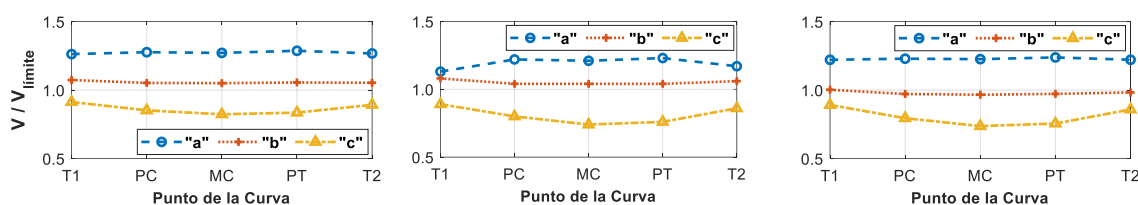


Imagen 5. Perfiles de velocidad identificados como centros de los tres conglomerados (“a”:Agresivo, “b”:Moderado “c”:Cauteloso) empleando diferentes estrategias de datos de entrada, de derecha a izquierda: solo velocidad, solo los parametros η y a_s^+ y velocidad con parámetros adicionales.

La Imagen 6 muestra que cuanto más agresivo es el conductor, se observa más dispersión en los valores de velocidad en cada uno de los puntos de medición de velocidad (gráficas del lado izquierdo). Esta figura soporta los resultados mencionados y adicionalmente considera una combinación ponderando un peso igualitario para la velocidad y para las dos variables usadas, lo que implica un peso del 50% para las velocidades y del 50% para las dos variables. También se observa que la mayoría de los conductores tienen un estilo de conducción moderado. Sin embargo, se tiene un acercamiento a la afirmación que al incorporar los datos de velocidad en los puntos de medición (i.e. estrategia 1, 3 y 4) produce la mejor delimitación de los perfiles de velocidad para reducir la dispersión en su clasificación (como se ve en los gráficos superior e inferior de la Imagen 6).

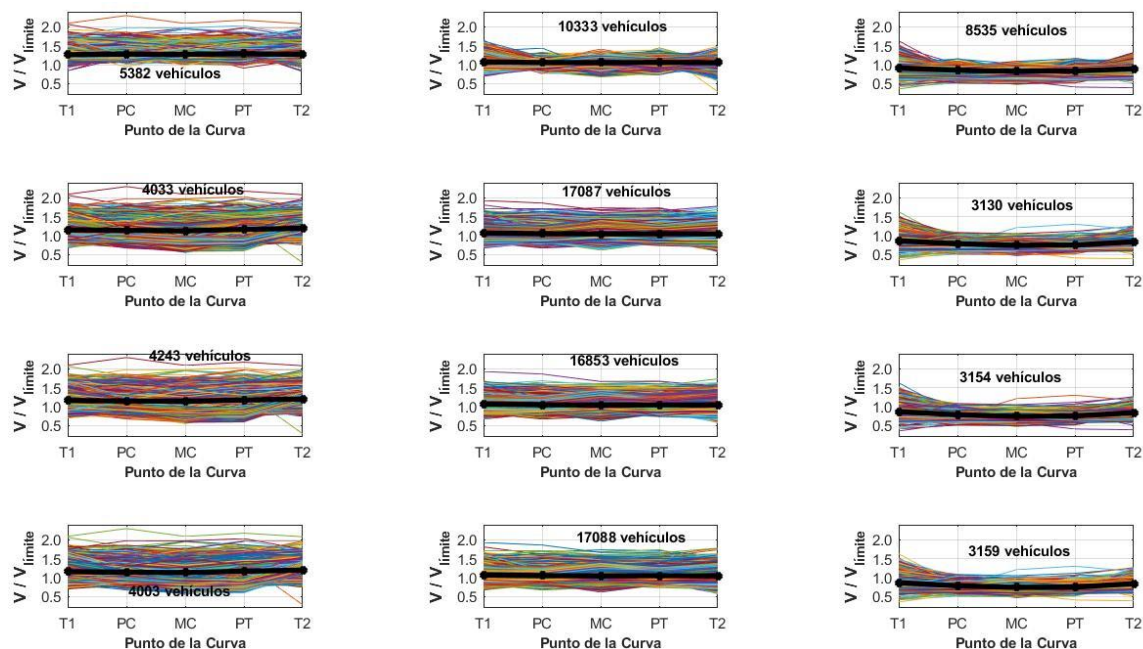


Imagen 6. Agrupación de perfiles de velocidad usando diferentes estrategias en los datos de entrada y centro del conglomerado (línea negra). De izquierda a derecha: Agresivo (a), Moderado (b) y Cauteloso (c). De arriba a abajo: solo velocidades, solo parámetros η y a_s^+ , velocidad con parámetros adicionales y velocidad con parámetros adicionales ponderados.

La Imagen 7 presenta la distribución (o frecuencia) de la variable asociada a la fracción del tiempo en que la velocidad del conductor es superior al 80% del límite permitido (η) para las tres estrategias propuestas. En estas gráficas se observa que al usar los datos de velocidad junto con las dos variables inferidas en el análisis de conglomerados (figuras inferiores) se observa que de manera más clara, el nivel de frecuencia de agresividad del conductor visto desde el tiempo durante el cual permanece cerca o por encima de la velocidad límite de la curva.

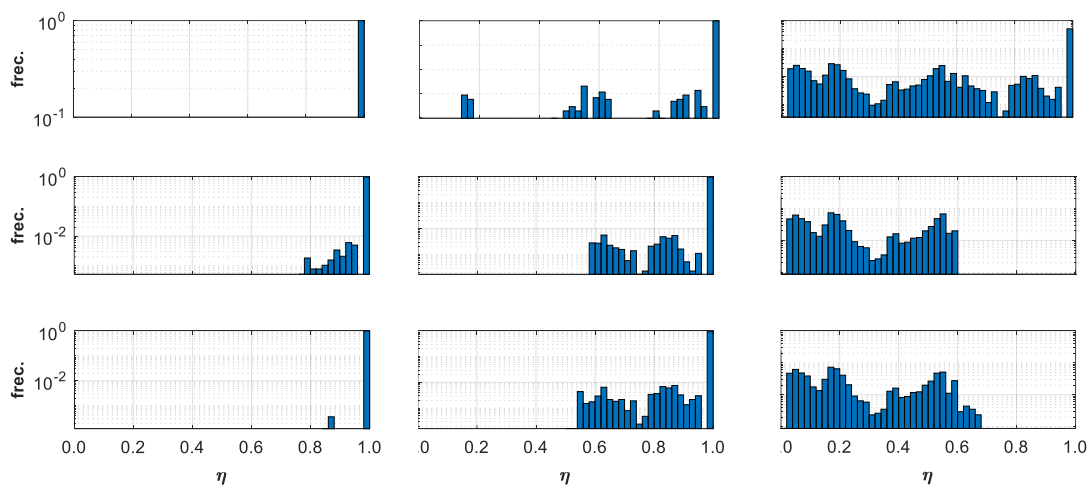


Imagen 7. Histograma para el parámetro η . De izquierda a derecha: Agresivo (a), Moderado (b) y Cauteloso (c). De arriba a abajo: Solo velocidades, solo parámetros η y a_s^+ y velocidad con parámetros adicionales.

De manera similar, la Imagen 8 presenta la distribución de la variable asociada a la desviación estándar de la aceleración positiva únicamente (a_s^+). Estas permiten observar que el valor de a_s^+ adquiere valores más altos cuando se presenta la mayor agresividad en el usuario (i.e. gráficas del lado izquierdo). A partir de las Imágenes 8 y 9 se puede inferir que el uso de las variables adicionales permite delimitar, de una manera más clara, el cambio en los tres estilos de conducción base.

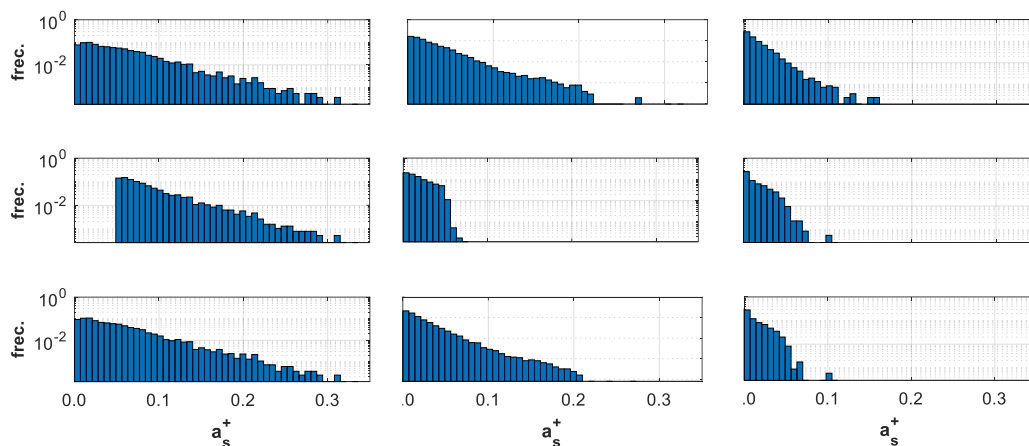


Imagen 8. Histograma para el parámetro a_s^+ . De izquierda a derecha: Agresivo (a), Moderado (b) y Cauteloso (c). De arriba a abajo: Solo velocidades, solo parámetros η y a_s^+ y velocidad con parámetros adicionales.

4.1.4. Discusión y validación

Adicionalmente, se realiza un análisis de sensibilidad gráfica que podemos encontrar en las Imagen 10 a la 14, en estas, a partir de observar la distribución y analizando diferentes variables geométricas de la curva (radio, tipo de terreno, clasificación funcional, pendiente, distancia de visibilidad y el grado de curvatura) y comparando el uso de las diferentes agrupaciones de datos propuestas, se pudo observar cómo se presenta una congruencia en los resultados cuando se tiene en cuenta las variables (η) y (a_s^+) y también se encuentra una aproximación del comportamiento de los usuarios enmarcado en las tres grupos de perfiles de velocidad definidos respondiendo a las variables analizadas.

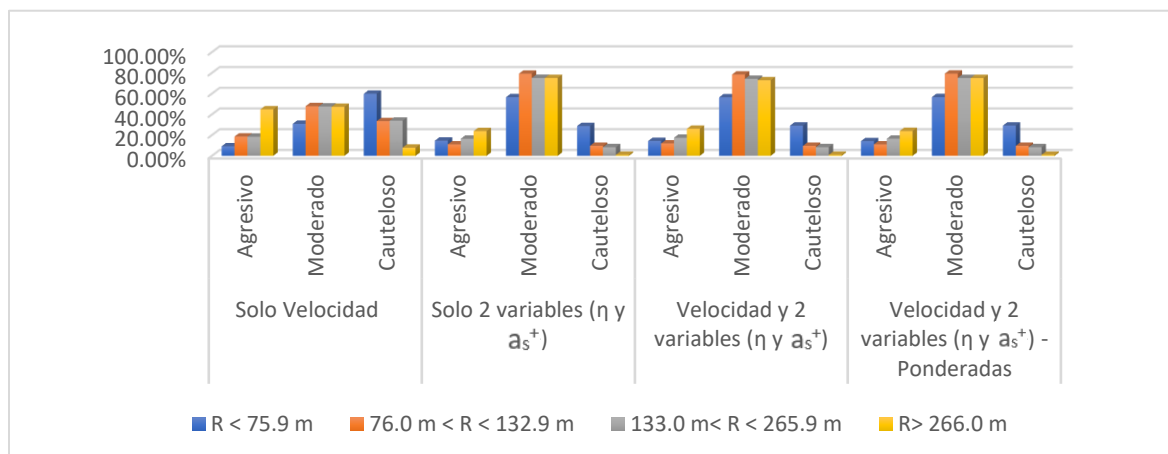


Imagen 9. Distribución de agrupaciones por rangos de radio en los diferentes análisis de datos

En la Imagen 9 podemos constatar como el comportamiento de los usuarios es predominantemente moderado cuando se utilizan las variables adicionales y la ponderación de las mismas, esto en diferentes rangos de radios en los que se distribuyó la información. También podemos observar como el comportamiento agresivo comienza a aumentar cuando el radio es mayor de 266 metros; podemos deducir que el usuario se siente cómodo con curvas más amplias y toma confianza en el proceso de conducción.

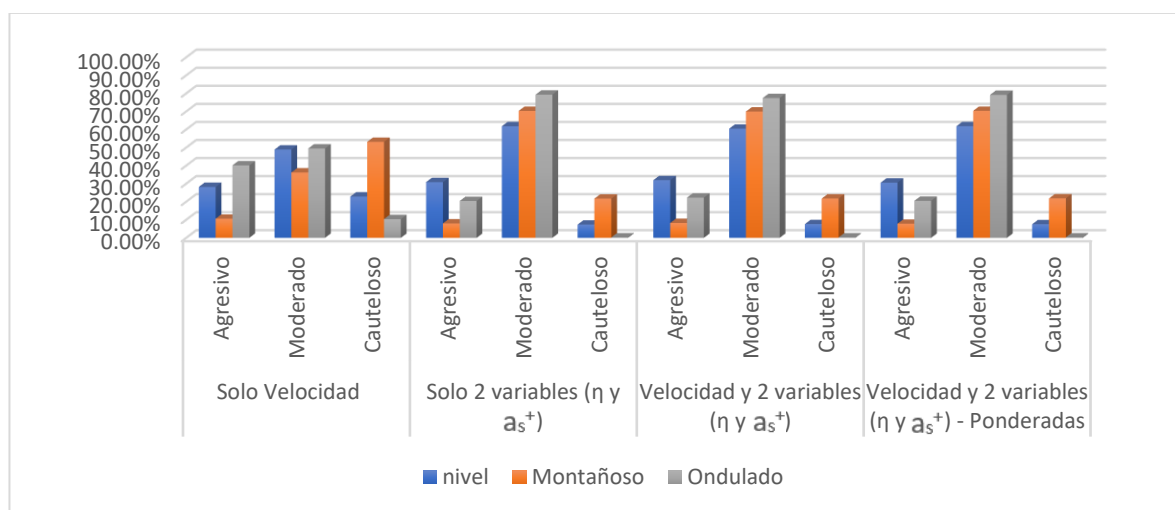


Imagen 10. Distribución de agrupaciones por tipo de terreno en los diferentes análisis de datos

En la Imagen 10 podemos comprobar que continúa un enfoque de similitud utilizando las variables adicionales y ponderadas, en este caso, cuando se divide en los tipos de terreno que pueden presentar las vías. El criterio del tipo de terreno es fundamental para el diseño geométrico de carreteras y podemos observar cómo el comportamiento agresivo en terreno plano, o a nivel

comienza a ser significativo, por lo que es una alarma interesante a tener presente, cuando se trate de una acción de mejora para la seguridad vial en diseños o mejoramientos.

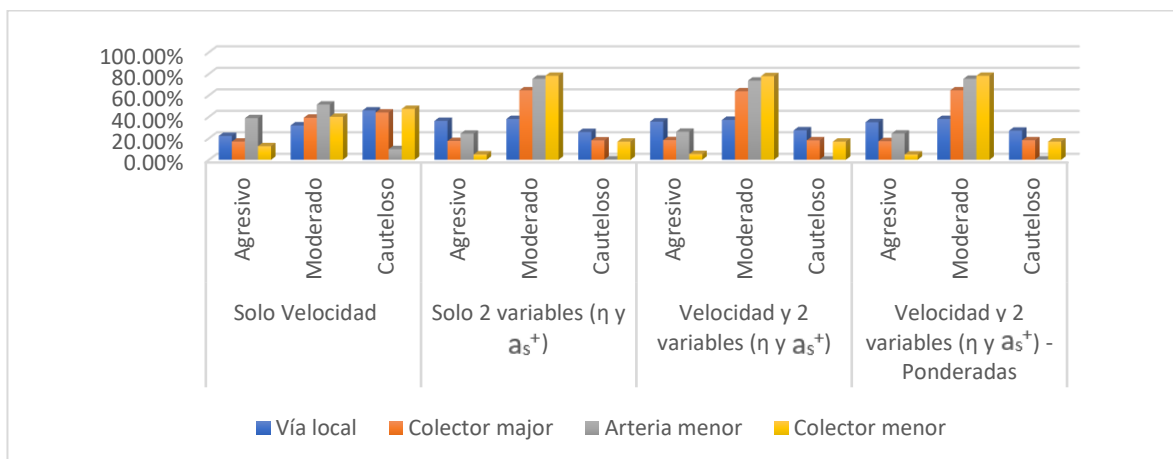


Imagen 11. Distribución de agrupaciones por clasificación funcional entre ellas: local, colectora menor, colectora mayor y arterial menor.

La Imagen 11, resultante de analizar la clasificación funcional de las carreteras nos permite continuar analizando como el comportamiento de similitud se establece utilizando las variables adicionales. De especial atención, analizar por qué los conductores pueden tomar comportamientos agresivos cuando tienen familiaridad con vías locales y posiblemente menor tránsito, lo que genera confianza y excesos en la velocidad operacional.

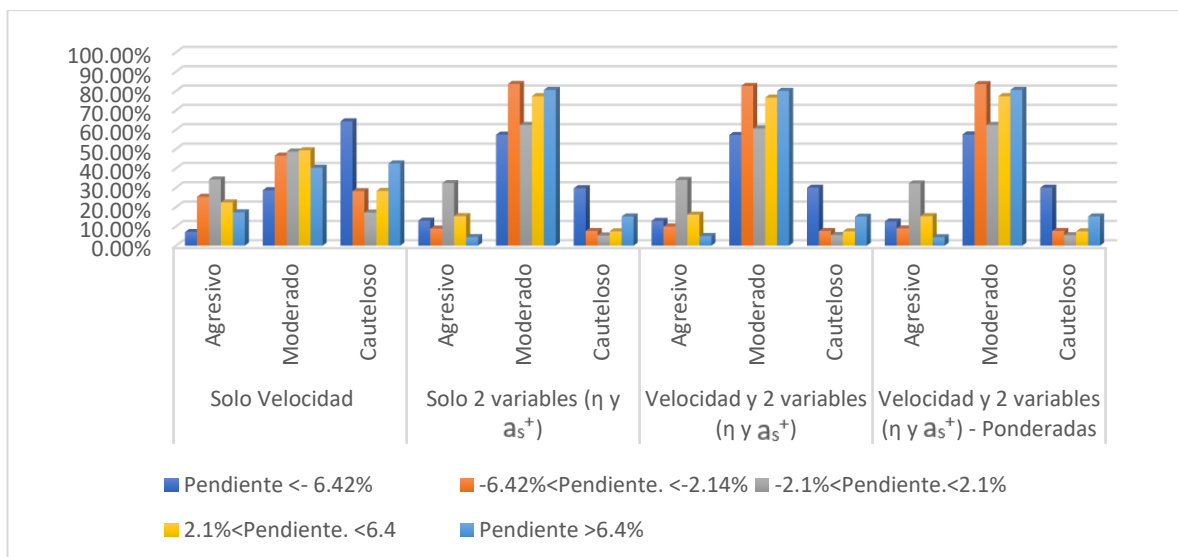


Imagen 12. Distribución de agrupaciones por rangos de pendiente en los diferentes análisis de datos

Analizar la pendiente es fundamental para el análisis del comportamiento, es por esto que para este análisis se asociaron las velocidades según el sentido de la vía, para poder así saber si el tipo de pendiente al que se enfrenta los usuarios en positiva o negativa. En la Imagen 12 el comportamiento continúa mostrando como utilizar las variables adicionales y ponderarlas es más adecuado para el conocimiento del comportamiento. También podemos analizar que el comportamiento de los usuarios, cuando la pendiente se acerca a cero, tienden a ser más agresivos y en pendientes negativas, por la sensación que genera la gravedad, tienden a ser más cautelosos.

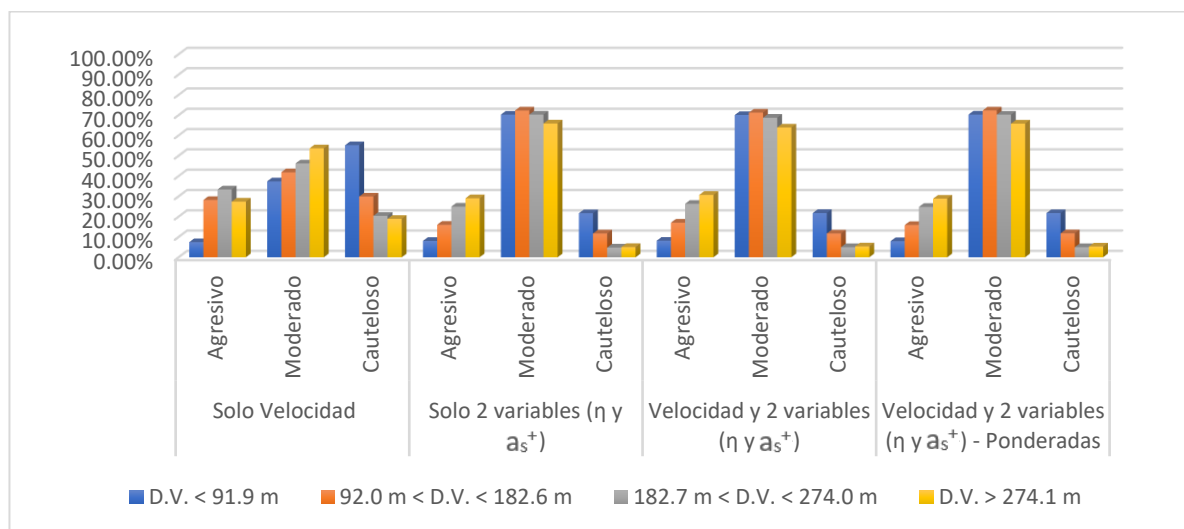


Imagen 13. Distribución de agrupaciones por rangos de distancia de visibilidad en los diferentes análisis de datos.

La distancia de visibilidad que pueda tener el usuario, cuando transita por la curva, es también de importancia para analizar el comportamiento respuesta de los conductores de las carreteras, esta distancia de visibilidad es variable, según el sentido de la curva, por lo que se evalúa de forma independiente en cada sentido. El análisis de la Imagen 13 continúa mostrando como la utilización de las variables adicionales y ponderándolas, vuelve más consistente los resultados. Como se puede esperar, a mayor distancia de visibilidad el conductor tiene mayor confianza de su entorno, por lo que puede tender a tener un comportamiento más agresivo, generado por la familiaridad que da conocer que hay a su alrededor. En contraparte, el usuario al no conocer qué pueda estar pasando a su alrededor, se siente limitado y tiende a disminuir la velocidad por su seguridad, lo que lo agrupa como cauteloso.

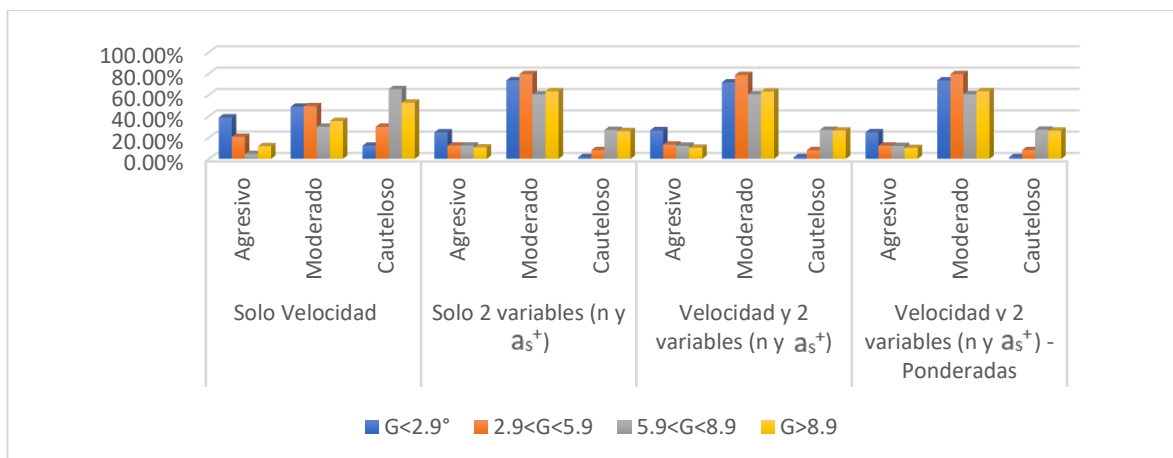


Imagen 14. Distribución de agrupaciones por rangos de curvatura en los diferentes análisis de datos

Finalmente, se estudia el grado de curvatura (G), el cual conocemos que está relacionado con el radio, en la Ecuación (8) podemos ver la relación de cálculo utilizada, la cual nos indica como la curvatura es inversamente proporcional al radio, por lo que consecuentemente con la Imagen 10, la Imagen 14 muestra que a menores grados de curvatura encontraremos conductores con comportamiento más agresivo. El estudio continúa siendo coherente con la validación, por lo que utilizar las variables adicionales y su ponderación, arroja datos más coherentes.

$$G = 2 \arcsen \frac{c}{2R}; c = 10 m \quad (8)$$

Después de analizar los resultados presentados se deciden trabajar con los datos de velocidades y de las variables (η) y (a_s^+), estos con igual peso, definiendo cada velocidad con el 10% y cada parámetro adicional con el 25%. Adicionalmente, con el fin de analizar la consistencia del método, teniendo en cuenta un peso unificado de datos a partir de cada variable, se realiza una distribución de forma aleatoria, escogiendo la misma cantidad de curvas por intervalos de variable a analizar y un número igual de perfiles en cada curva.

La distribución específica por variable analizada, en términos de organización de los datos, con una cantidad de 149 perfiles de velocidad por curva y elegidos de forma aleatoria, fue:

- Por clasificación funcional: un total de 2 curvas por categoría para 4 categorías.
- Por radio: un total de 7 curvas por intervalo, para 4 intervalos de radios.
- Por grado de curvatura: un total de 5 curvas por intervalo, para 4 intervalos de curvaturas.

- Por tipo de terreno: un total de 11 curvas por categoría, en tres categorías evaluadas.
- Por pendiente: un total de 11 curvas por intervalo, en 5 intervalos de pendientes.
- Por distancia de visibilidad: un total de 13 curvas por intervalo, en 4 intervalos.

Los intervalos utilizados son los mismos presentados en las distribuciones de las Tablas 2 a la 7 donde se presentó la distribución estadística de estos elementos. Es importante aclarar que la distribución de pendientes y distancias, de visibilidad generan duplicidad en las curvas debido a que son tomadas en los dos sentidos, por lo que el análisis incrementa el número de curvas.

El resultado de analizar los datos, evaluando la misma cantidad de curvas por intervalo y la misma cantidad de perfiles por curva, se presenta en la Imagen 15; podemos observar cómo la consistencia del método se presenta en todos los análisis, afectándose solamente en el porcentaje de usuarios agresivos cuando se igualan los elementos según la variable del radio, el cual no se considera como una inconsistencia mayor.

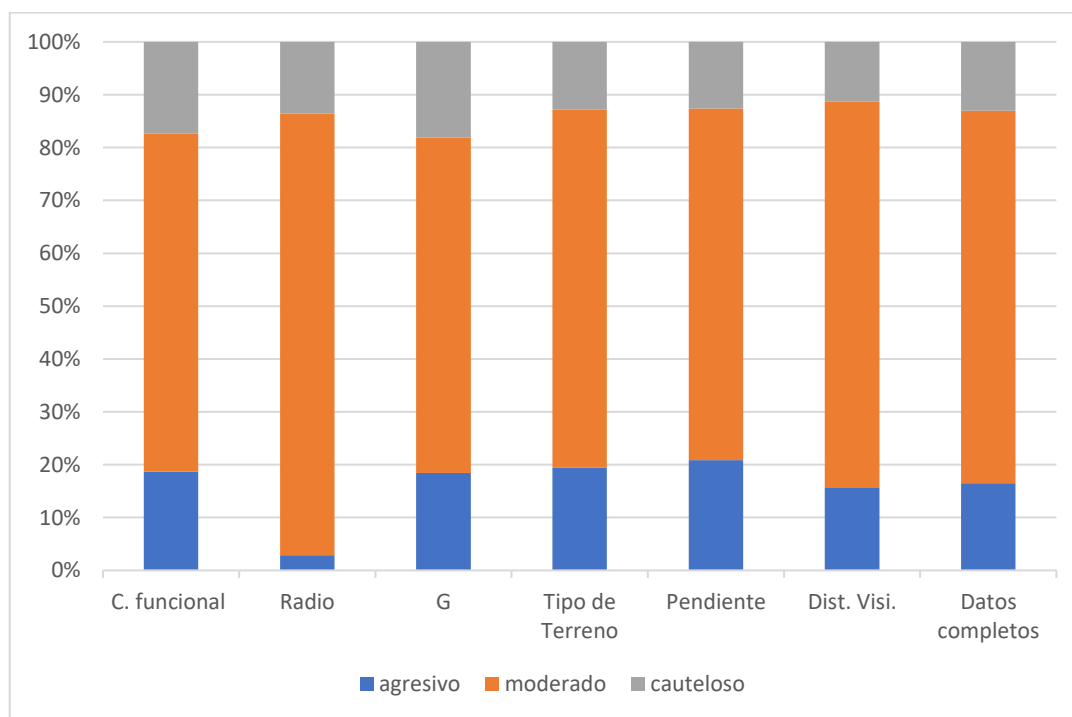


Imagen 15. Distribución de grupos de comportamiento con la misma cantidad de curvas por variable analizada.

Finalmente, después de analizar el método, la efectividad y la constancia en los resultados se valida conociendo su comportamiento con la totalidad de los datos, se evalúa el árbol de decisión por el método propuesto y se modela cada comportamiento encontrado.

En la Imagen 16 podemos observar la distribución final, según diferentes variables de las curvas en porcentaje, teniendo en cuenta las velocidades y los parámetros (η) y (a_s^+) ponderados al mismo peso. Este resultado muestra las tendencias de los conductores cuando se presentan diferentes tipos de validaciones según sus variables geométricas, lo cual puede permitir un análisis de seguridad vial de prevención.

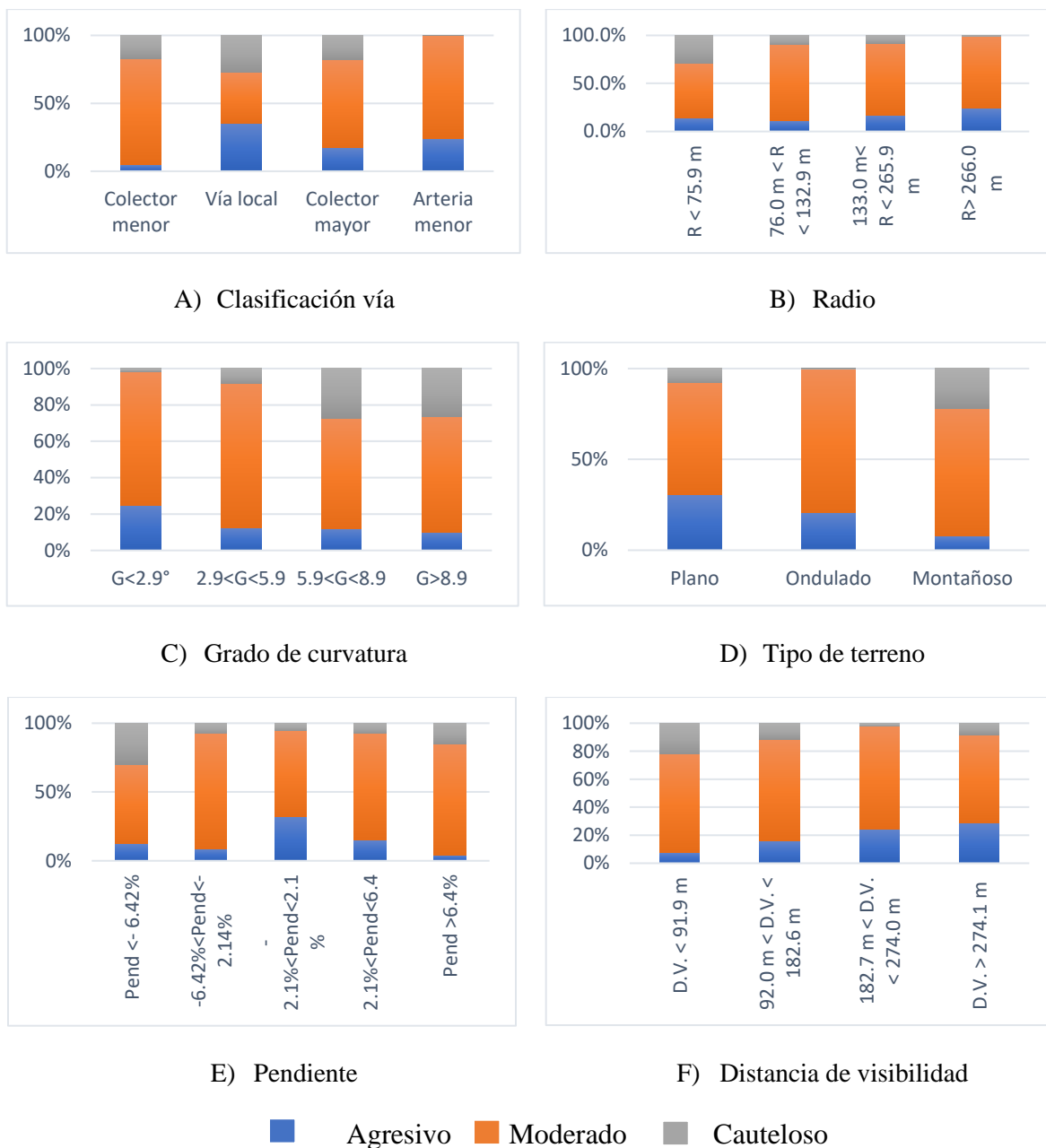


Imagen 16. Distribución de grupos de comportamiento de toda la data en diferentes variables de análisis.

Analizando el comportamiento presentado en la Imagen 16, mencionando detalles de interés, en especial, acerca de la seguridad vial. La Imagen 16(A) muestra cómo las vías locales y arterias menores presentan mayor cantidad de usuarios con comportamientos agresivos. En las vías locales puede deberse al bajo tráfico y al conocimiento y confianza que pueden tener de la vía. En las colectoras menores analizamos que se tratan de sitios con tipo de terreno ondulado, por lo que se enlaza con el análisis de la Imagen 16 (D); pero también se atribuye que las características de

la vía son diferentes, presentando carreteras amplias, anchos de zona más grandes y disposiciones de señalización mejores.

La Imagen 16 (B) muestra que si la magnitud del radio aumenta, los usuarios ganan confianza y cambian a un comportamiento agresivo (donde la variable aumenta y se espera la reducción de la aceleración). Las curvas con radios pequeños representan aproximadamente el 40% de sus usuarios más cautelosos, lo que se atribuye a una mayor percepción de peligro (tendencia a reducir la velocidad antes de iniciar la longitud de la curva). Las curvas con radios que van desde 133 metros a 260 metros tendrán usuarios predominantemente cautelosos y moderados en sus estilos de conducción, el comportamiento agresivo que representan menos del 20% del total (es decir, la condición de seguridad vial más alta).

La curvatura (relacionada con la inversa de los radios de curvatura) permite identificar el comportamiento agresivo del conductor ya que es menor a 3° (Imagen 16 (C)); se puede utilizar para cuantificar la seguridad vial y definir en el diseño geométrico de la curva. La Imagen 16 (D) muestra cómo el terreno ondulado y plano es mucho más susceptible al comportamiento agresivo del conductor, lo que afecta la seguridad e implica que este tipo de terreno debe diseñarse con especial cuidado en términos de seguridad vial de carreteras (es decir, estas son arterias menores).

Cuando la pendiente es más alta, los conductores asumieron un comportamiento más cauteloso (Imagen 16 (E)), en especial cuando es negativa. Dado que los conductores tienden a comportarse de forma más agresiva cuando la pendiente de una curva se encuentra el intervalo (-2.1% - +2.1%) debería estudiarse más en términos de seguridad vial, esto tendiendo a sitios predominantemente planos. La distancia de visibilidad define otras distancias como detenerse, pasar de vista y situaciones de maniobra complejas, estas se utilizan para diseñar los elementos de alineación vertical y horizontal y configuraciones de seguridad vial (AASHTO, 2018). La Imagen 16 (F) muestra que cuando esta distancia supera los 182 m, el comportamiento de los conductores se vuelve más agresivo (por encima del 20%), por lo que tener en cuenta este valor de distancia de visibilidad permitirá manejar elementos de mayor seguridad para estos tramos.

En general, la clasificación de los enfoques de análisis de conglomerados propuestos por tres perfiles de velocidad coincide con el comportamiento esperado por la geometría y la condición operativa de la curva. Se hace necesario el análisis de los porcentajes de usuarios agresivos

relacionados con las características evaluadas para así definir elementos en manuales de seguridad vial y criterios de diseño para no incurrir en estos porcentajes elevados.

4.1.5. Análisis cualitativo de árbol de decisión y cuantitativo de regresión lineal

Se realizó un análisis de decisión por medio de árboles de clasificación, el método propuesto inicialmente por Breiman et al., (1984), indica una partición de conjunto de datos a partir de variables independientes, pretende seleccionar de manera sucesiva grupos homogéneos. Además de ser una técnica con un resultado descriptivo importante para entender e interpretar, es apropiado para grandes conjuntos de datos y también se describe como un proceso secuencial, iterativo y descendente (Parra, 2016).

El método comienza eligiendo una variable dependiente y dividiendo el conjunto de datos de la manera más homogénea posible, y después con cada grupo elegido se continúa dividiendo en partes homogéneas, termina el proceso cuando la partición se haya seleccionado por completo. El resultado del método da un árbol conformado por nodos, ramas y hojas, los nodos se conforman como las variables de entrada, las ramas representarán los valores que puedan tomar dichas variables y las hojas serán relacionadas con los valores de la variable de salida (Parra, 2016).

Debido a la necesidad de análisis, los resultados del método y para comprobar su veracidad se realiza una partición del 70% de los datos, esto con el fin de realizar un conjunto de entrenamiento y un conjunto de prueba que permitan una posterior verificación. El árbol puede ser ajustado o podado, con el fin de obtener un modelo acorde a la necesidad y que tenga una representatividad importante; una elección de poda es no permitir que el error después de descender presente un aumento de error para la elección de la variable respuesta, se puede consultar respecto al tema a Breiman et al., (1984). En la Imagen 17 podemos ver que no fue necesario la realización de la poda del árbol en término de nivel jerárquico, esta muestra que con 7 niveles se tiene una estabilidad del error relativo. Pero en la Imagen 18 sí se puede ver como las variables fueron elegidas según su jerarquía y se utilizaron las de mayor relevancia para el modelo.

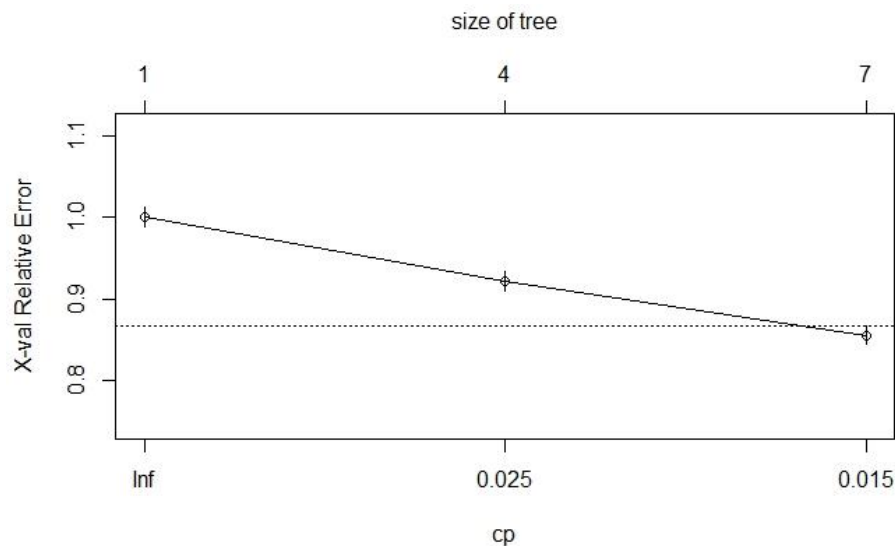


Imagen 17. Análisis de error.

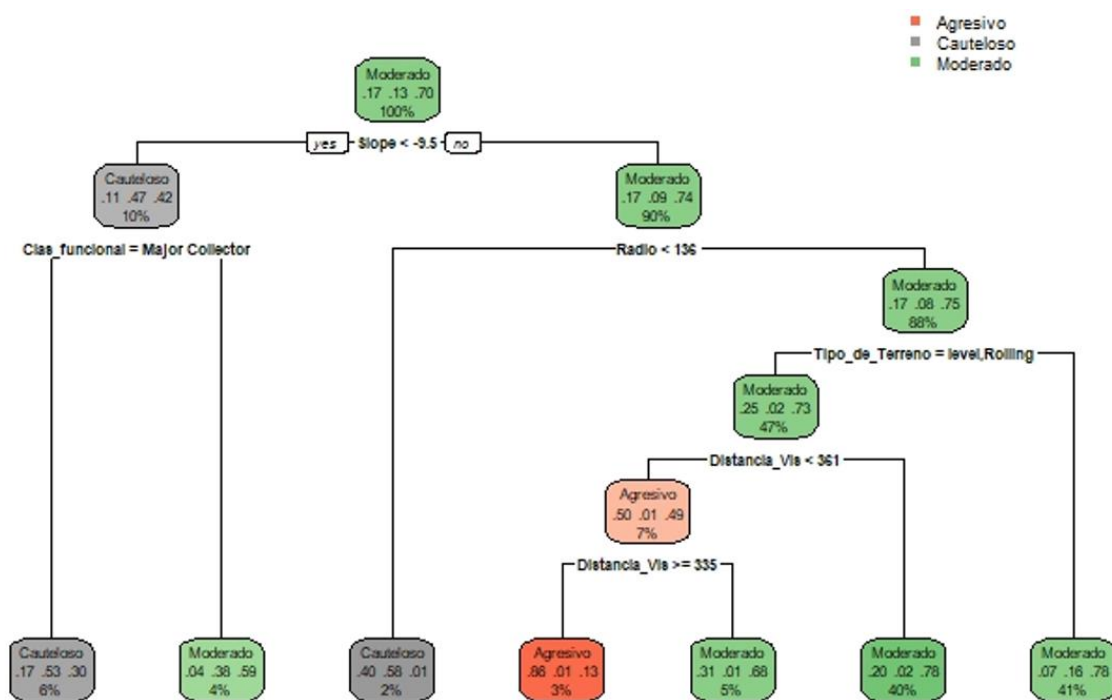


Imagen 18. Árbol de decisión.

A partir de la clasificación de los perfiles de velocidad en los tres comportamientos predominantes o estilos de conducción, la Imagen 18 presenta el resultado del análisis cualitativo para la elección del estilo de conducción según la percepción de la geometría de la curva. La Imagen 18 permite predecir el comportamiento del conductor (en cuanto a elección de la velocidad) en función de las variables geométricas que puede usarse a la hora de definir o estudiar

parámetros de diseño considerando la seguridad vial. En ella podemos observar como la variable de la pendiente tiene mayor importancia en la elección del tipo de conducción que pueda tener un usuario de una carretera; el radio y la clasificación funcional también presentan un porcentaje importante de significancia para la elección del tipo de conducción. Estos parámetros permiten observar qué variables se pueden estudiar con mayor detenimiento para que el comportamiento de los usuarios cambie y así mejorar la seguridad de la vía.

Para conocer lo acertado del modelo se utiliza el segmento de prueba separado al inicio del proceso, utilizando el set de datos restante del 30%, se aplicó el modelo obtenido para verificar los resultados. En la Tabla 9 se observa que valor que tomó el conjunto de prueba utilizando el modelo de árbol de decisión y el valor asignado al comenzar el análisis. Este proceso de comparación permite observar un ajuste del 74.5% en el modelo, utilizando la diagonal la cual muestra los casos donde hay igualdad, con el modelo de árbol de decisión y con los datos iniciales, comparado con el total de datos del conjunto de prueba.

Tabla 9. Verificación grupo de control

Grupo resultante del modelo	Grupo Cluster perteneciente		
	1	2	3
1	165	0	24
2	118	272	120
3	917	675	4982

Finalmente, se considera un modelo lineal de cada uno de los estilos de conducción, partiendo de las variables de velocidad de diseño, grado de curvatura y radio, estas elegidas por el análisis contemplado en este estudio y el realizado a los mismos datos para la obtención de la velocidad operacional por Castillo, (2020); este analiza las variables de radio, velocidad de diseño y ángulo de deflexión, de este último se puede derivar el grado de curvatura. En la Tabla 10 se muestran los resultados de la modelación de aplicar el modelo de regresión lineal multivariable, el cual arroja las Ecuaciones (9) a la (11) para cada tipo de comportamiento. Es importante mencionar que el valor p para los coeficientes individuales de las regresiones es menor a 0.01, lo que nos indica que la contribución de las mismas es significativa. También se observa que el R^2 es bajo para las regresiones cerca al 30% para el comportamiento agresivo, 25% para el moderado y menor al 1% para el cauteloso, estos valores invitan al análisis debido a que la proporción de la

incertidumbre de la variable respuesta explicada por el modelo es baja, sin embargo, se adoptaron los modelos debido a la necesidad de comparar con modelos previamente estudiados (Castillo, 2020). Así pues, ajustar estos modelos de regresión múltiple propone un análisis más riguroso a futuro para continuar conociendo el comportamiento de los usuarios a partir de perfiles de velocidad adaptados a dichos comportamientos y poder relacionarlos con perfiles de velocidad operacional encontrados en la literatura los cuales son generalizados para el tramo de estudio.

Tabla 10. Resultado del modelo lineal para cada agrupación.

VARIABLES	Modelo lineal Vel Agresivo	Modelo lineal Vel Moderado	Modelo lineal Vel Cauteloso
Radio	0.00713*** (0.000368)	-0.0102*** (0.000158)	0.000875*** (0.000255)
G	-0.441*** (0.0257)	0.666*** (0.0159)	0.0310** (0.0129)
Vdiseño	-0.0715*** (0.0150)	0.610*** (0.00762)	-0.0442*** (0.00919)
constante	41.80*** (0.541)	15.44*** (0.305)	28.80*** (0.292)
Observaciones	20,015	85,440	15,795
R-cuadrado	0.316	0.269	0.006

Nota: (Error estándar) y *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

$$V_{\text{Agresivo}} = 41.8 + 0.00713 \cdot R - 0.441 \cdot G - 0.0715 \cdot V_{\text{diseño}} \quad (9)$$

$$V_{\text{Moderado}} = 15.44 - 0.0102 \cdot R + 0.666 \cdot G + 0.610 \cdot V_{\text{diseño}} \quad (10)$$

$$V_{\text{Cauteloso}} = 28.80 + 0.000875 \cdot R + 0.310 \cdot G - 0.0442 \cdot V_{\text{diseño}} \quad (11)$$

En las Ecuaciones se observa como el comportamiento de los conductores se afecta de formas distintas por cada variable utilizada en el estudio, esto puede deberse a la cantidad de datos utilizada y la heterogeneidad que se presenta en los datos. El comportamiento agresivo es afectado de forma positiva por el radio, sin embargo, se restringe cuando el grado de curvatura y la velocidad de diseño aumentan. Por otro lado, el comportamiento cauteloso es contrario al anterior, pero la constante es menor, así que el valor que aporta la velocidad de diseño y el grado de curvatura es más limitada sin sobrepasar el agresivo. El comportamiento cauteloso muestra como la velocidad de diseño afecta negativamente la velocidad operacional de los individuos, propio de conductores cautelosos que se adaptan por debajo de la velocidad de diseño a la utilización de las carreteras.

4.1.6. Análisis de estimación de accidentes y relación con perfiles de velocidad

Se utilizó la metodología de estimación de frecuencia de choques del manual de seguridad vial (Highway Safety Manual, 2010) para tres segmentos de vía escogidos de forma aleatoria y revisando que hubiera heterogeneidad de los tramos (Tramos 110 – 115b y 102). A continuación, se presentan las variables utilizadas, el cálculo de la frecuencia de choques y los factores de modificación, estos últimos se utilizaron solamente aquellos que aplican a los elementos que tienen dichas vías.

Las Tablas 11 a la 13 muestran las características de las vías utilizadas para calcular tanto la predicción como los factores mencionados. Cada tramo tiene una cantidad de segmentos evaluados, los cuales se suman para tener la globalidad. Al ítem que no se le conoce la información o que no aplica para la vía se marca como NA.

Tabla 12. Variables tramo 115b, Terreno ondulado – Arteria menor

Variables	Segmentos			
	1	2	3	6
Longitud (millas)	0.17918535	0.27527952	0.183662503	0.23964861
Aadt (vehículos por días)	14800	14800	14800	14800
Ancho(pies)	12	12	10	10
Bermas (pies) derecha	13.45	18.07	1.18	1.18
Bermas (pies) izquierda	9.51	10.92	0.85	0.92
Tipo de berma	pav	pav	pav	pav
Longitud de curva (millas)	0.10342775	0.19952192	0.107904903	0.16389101
Radio (pies)	1628	2746	1160	490
Peralte	5.7	4.8	3.8	4.7
Curvatura (G) (°)	1.154107737	0.684195894	1.620636718	3.837090129
Densidad de conductores por milla	NA	NA	NA	NA
Tiras de sonido centrales	NA	NA	NA	NA
Líneas de sobrepaso	NA	NA	NA	NA
Giros	NA	NA	NA	NA
Clasificación de peligro	NA	NA	NA	NA
Control automático de velocidad	NA	NA	NA	NA

Tabla 13. Variables tramo 102, Terreno plano – Colector Mayor

Variables	Segmentos					
	1	2	3	4	5	6
Longitud (millas)	0.16595261	0.11647421	0.12136119	0.121156273	0.21827503	0.15047819
Aadt (vehículos por días)	6500	6500	6500	6500	6500	6500
Ancho(pies)	9	9	9	9	9	9
Bermas (pies) derecha	1.18	0.85	0.92	1.05	1.05	5.31
Bermas (pies) izquierda	0.92	1.18	0.79	1.05	0.98	0.92
Tipo de berma	pav	pav	pav	pav	pav	pav
Longitud de curva (millas)	0.09019501	0.04071661	0.04560359	0.045398673	0.14251743	0.07472059
Radio (pies)	200	522	811	830	4091	1034
Peralte	9.3	5.5	3.4	4.8	4.3	6.6
Curvatura (G) (°)	9.39	3.60	2.32	2.27	0.46	1.82
Densidad de conductores por milla	NA	NA	NA	NA	NA	NA
Tiras de sonido centrales	NA	NA	NA	NA	NA	NA
Líneas de sobrepaso	NA	NA	NA	NA	NA	NA
Giros	NA	NA	NA	NA	NA	NA
Clasificación de peligro	NA	NA	NA	NA	NA	NA
Control automático de velocidad	NA	NA	NA	NA	NA	NA

Para el cálculo de la frecuencia de choques se siguieron los pasos del capítulo 10 del Manual de Seguridad Vial de la AASHTO, el cual es un complemento de la parte C del documento y esta direccionado a carreteras rurales de dos carriles (AASHTO, 2010). La Tabla 14 muestra los factores de modificación utilizados para el cálculo de la frecuencia de choques en el HSM, debido a las condiciones de las vías utilizadas solo se utilizaron los 5 primeros factores de modificación los cuales se referencian: el ancho de carril, el tipo y ancho de berma, las curvas horizontales, la pendiente y el grado de curvatura.

Tabla 14. Resumen de factores de modificación de accidentes. Tomada de SHM table 10-7 (AASHTO, 2010).

Facility Type	CMF	CMF Description	CMF Equations and Tables
Rural Two-Lane Two-Way Roadway Segments	CMF _{lr}	Lane Width	Table 10-8, Figure 10-7, Equation 10-11
	CMF _{sr}	Shoulder Width and Type	Tables 10-9, 10-10, Figure 10-8, Equation 10-12
	CMF _{hr}	Horizontal Curves: Length, Radius, and Presence or Absence of Spiral Transitions	Equation 10-13
	CMF _{hr}	Horizontal Curves: Superelevation	Equations 10-14, 10-15, 10-16
	CMF _{gr}	Grades	Table 10-11
	CMF _{dr}	Driveway Density	Table 10-11
	CMF _{tr}	Centerline Rumble Strips	See text
	CMF _{pr}	Passing Lanes	See text
	CMF _{pr}	Two-Way Left-Turn Lanes	Equations 10-18, 10-19
	CMF _{rd}	Roadside Design	Equation 10-20
	CMF _{lr}	Lighting	Equations 10-21, Table 10-12
	CMF _{sr}	Automated Speed Enforcement	See text
	Three- and four-leg stop control intersections and four-leg signalized intersections	CMF _{si}	Intersection Skew Angle
CMF _{tl}		Intersection Left-Turn Lanes	Table 10-13
CMF _{tl}		Intersection Right-Turn Lanes	Table 10-14
CMF _{li}		Lighting	Equation 10-24, Table 10-15

A continuación, se presentan los cálculos específicos de la predicción de los choques para cada tramo de vía, sumando todos sus segmentos y modificándolos por los CMF. En las Tablas 15 a la 17 se puede encontrar las variables utilizadas y el proceso descrito en el capítulo 2.3.1 utilizado por el Manual de Seguridad Vial de Estados Unidos. Cabe aclarar que todos los CMF son de incremento a la predicción de accidentes, no se aplicó ningún CMF que decrezca este valor debido a que no hubo ningún tratamiento en la vía posterior a la toma de datos que indicara una mejora en la seguridad, con lo cual pudiera disminuir su accidentalidad.

Tabla 15. Tramo 110 cálculo de predicción de accidentes.

CMF	Segmentos								
	1	2	3	4	5	6	7	8	9
Nspfrs	0.21020 39	0.15468 978	0.24652 813	0.356873 508	0.20997 396	0.18644 3	0.15949 074	0.26412 235	0.22208 436
CMFra	1.3	1.3	1.3	1.3	1.05	1.05	1.05	1.05	1.05
CMF1r	1.1722	1.1722	1.1722	1.1722	1.0287	1.0287	1.0287	1.0287	1.0287
CMFwrs	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
CMFtra	1	1	1	1	1	1	1	1	1
CMF2	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287	1.287
CMF3	2.51558 151	3.08312 966	1.51871 598	1.586721 721	3.14111 204	4.44823 077	3.96403 595	2.35631 568	2.72183 951
CMF4	4	4	4	4	4	4	4	4	4
CMF5	1.16	1.1	1	1.1	1.16	1.16	1.16	1.16	1.16
N	7.21791 938	6.17335 097	4.40573 033	7.329639 742	6.38138 123	8.02415 312	6.11700 567	6.02149 672	5.84852 446
Sumatoria de segmentos				57.51920 16	Redondeo				58

Notas: Nspfrs: Número base de la predicción, CMFxx: Factor de modificación utilizado, N: Predicción final

Tabla 16. Tramo 115-b cálculo de predicción de accidentes.

CMF	Segmentos			
	1	2	3	6
Nspfrs	0.70852831	1.08850043	0.726231697	0.94760996
CMFra	1	1	1.3	1.3
CMF1r	1	1	1.1722	1.1722
CMFwrs	0.98	0.98	1.5	1.5
CMFtra	1	1	1	1
CMF2	0.98852	0.98852	1.287	1.287
CMF3	1.30729195	1.09443903	1.413374329	1.64430493
CMF4	4	4	4	4
CMF5	1	1	1	1.1
N	3.58923027	4.61627536	12.07834041	20.1687754
Sumatoria de segmentos		40.4526214	Redondeo	41

Notas: Nspfrs: Número base de la predicción, CMFxx: Factor de modificación utilizado, N: Predicción final

Tabla 17. Tramo 102 cálculo de predicción de accidentes.

CMF	Segmentos					
	1	2	3	4	5	6
Nspfrs	0.28819765	0.20227215	0.21075902	0.210403155	0.37906214	0.26132436
CMFra	1.5	1.5	1.5	1.5	1.5	1.5
CMF1r	1.287	1.287	1.287	1.287	1.287	1.287
CMFwrs	1.5	1.5	1.5	1.5	1.5	1.5
CMFtra	1	1	1	1	1	1
CMF2	1.287	1.287	1.287	1.287	1.287	1.287
CMF3	3.86833688	3.43444858	2.39901629	2.373160862	1.08874527	1.66970237
CMF4	4	4	4	4	4	4
CMF5	1.16	1.1	1	1	1	1
N	19.2784589	11.3916178	7.53735505	7.443531481	6.15228276	6.50456967
	Sumatoria de segmentos		58.3078157	Redondeo		59

Notas: Nspfrs: Número base de la predicción, CMFxx: Factor de modificación utilizado, N: Predicción final

Finalmente, se compararon los perfiles de velocidad. Se calculó un perfil de velocidad operacional (V_{85}) de la vía a partir de un modelo calculado con la misma base de datos, el cual trabaja con la velocidad de diseño, el radio y el ángulo de deflexión (Castillo, 2020). Este modelo, el cual se muestra en la Ecuación (12) fue analizado y se compara con los modelos de la velocidad de los perfiles de comportamientos a partir de los resultados previos del estudio.

$$V_{85} = 35.796 + 0.614V_d + 0.003R - 0.109\Delta \quad (12)$$

En las Imágenes 19 a la 21 se muestran la comparación de los perfiles de velocidad mencionados para el tramo en estudio. En ellos se observa cómo los perfiles de comportamiento se comportan de manera lógica. En relación al perfil de la velocidad operacional se puede deducir un cambio por lo mencionado por Laureshyn et al. (2009), que indica que realizar agrupamiento de perfiles sin tener en cuenta la variación individual del mismo puede generar perfiles con sesgo que ocultan o limitan la variación en el comportamiento del usuario. Atribuir los accidentes predichos a comportamientos agresivos de los conductores es una conclusión aceptable e indica el tratamiento especial que se debe dar a estos usuarios.

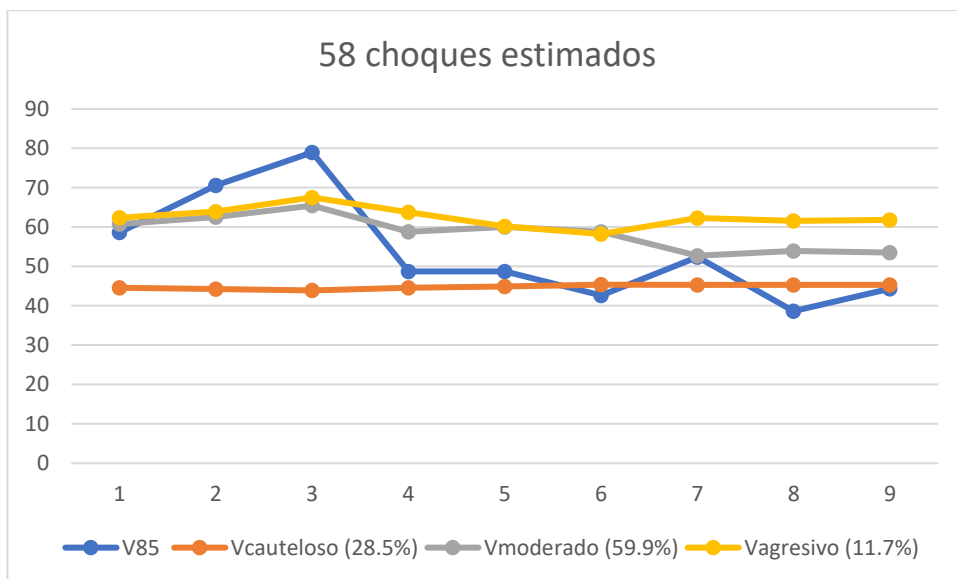


Imagen 19. Perfiles de velocidad tramo 110

La Imagen 19 nos muestra de los tres perfiles de velocidad estudiados, evaluando las curvas en el tramo 110, en este se predijeron 58 choques según la metodología del HSM. El perfil de la velocidad operacional varia a lo largo del tramo, en puntos se puede encontrar sobre el perfil agresivo y en otros debajo del perfil cauteloso. Podemos aducir que el comportamiento agresivo es el que está generando los 58 eventos de accidentalidad mencionados, a excepción de las curvas 5 y 6 en las que los perfiles tienden a combinarse en comportamiento. En este sentido, en términos de seguridad vial, podemos colocar especial cuidado en las curvas 7 a 9 debido a que el comportamiento agresivo se separa del moderado.

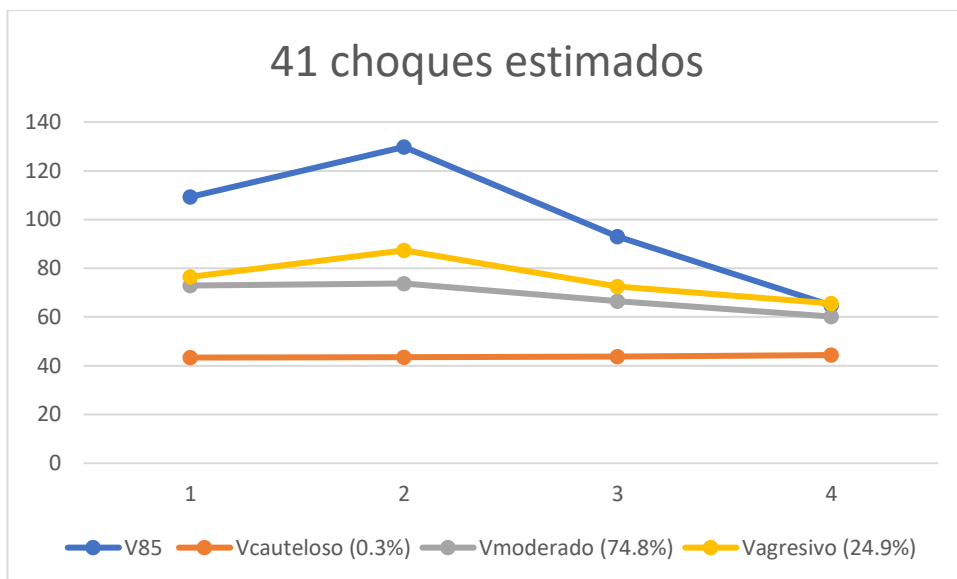


Imagen 20. Perfiles de velocidad tramo 115 b

En la Imagen 20 podemos intuir un comportamiento similar a la anterior, con la particularidad que en este caso la velocidad operacional sobrepasa en las 4 curvas a todos los perfiles estudiados. También podemos atribuir los 41 choques estimados al comportamiento agresivo, el cual puede ser una revisión más crítica debido el tramo 115b tiene menor cantidad de curvas o la longitud del tramo es menor al tramo 110.

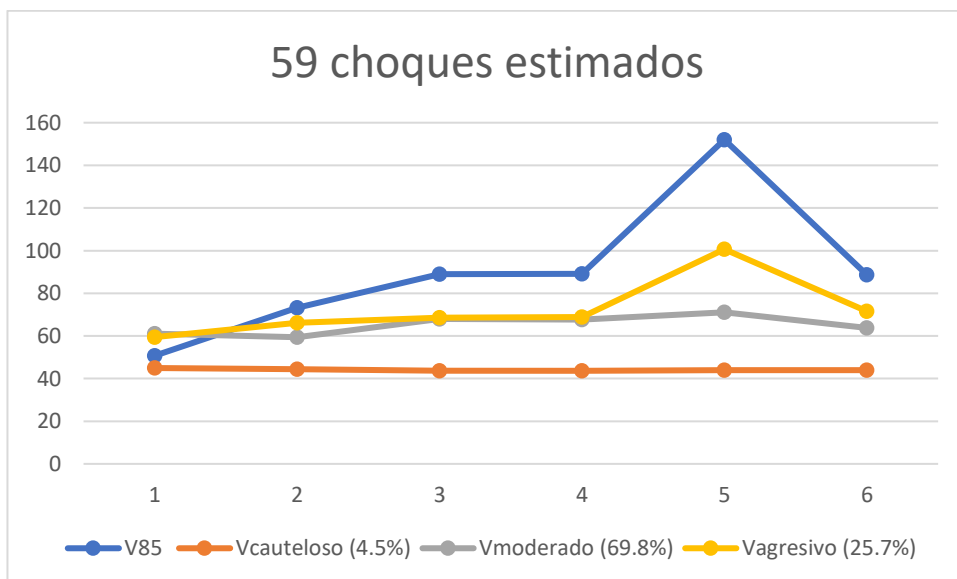


Imagen 21. Perfiles de velocidad tramo 102

En la Imagen 21 continua el comportamiento de forma parecida a los evaluados en tramos anteriores, lo cual indica una validación del método utilizado. De manera particular se estima que

en la curva número 5 la velocidad de diseño crece respecto a las otras, por lo cual el comportamiento agresivo y del modelo de velocidad operacional se dispara. En este segmento de tramo puede atribuirse de manera más clara la consecuencia de posibles accidentes.

Adicionalmente, podemos referir que la estimación de choques calculados puede responder al porcentaje de conductores agresivos que presenta la vía, al identificar este porcentaje y tratar su comportamiento, relacionándolo con elementos de la vía en este estudio podremos mejorar la seguridad vial. Analizando las variables utilizadas a lo largo del estudio se propone que la disminución del comportamiento agresivo se puede ejercer al controlar las diferentes variables, por ejemplo: algunas formas de disminuir los perfiles de velocidad agresiva serian disminuir el radio, aumentar la curvatura, pendientes más grandes y distancias de velocidad menores. Sin embargo, para no afectar otras dinámicas de seguridad vial, cuando se presenten este tipo de condiciones o terrenos muy planos o mayor categoría de las vías, se propone señalar y acudir a elementos de seguridad vial, para advertir a este tipo de conductores.

4.2. Recomendaciones

La primera recomendación que nos arroja la presente investigación es, tener en cuenta el factor humano para ser analizado a fondo en: la consistencia de diseño, la seguridad vial y la estimación de choques que pueda presentar una carretera. Al realizar este tipo de análisis se podrá re-estudiar la consistencia de diseño, agregando elementos a las metodologías existentes. Se podrá relacionar en los diferentes manuales de seguridad vial, para tener en cuenta los tramos en que el porcentaje comportamientos agresivos sean elevados. Posteriormente, se podrá mejorar la metodología de estimación de choques, relacionando el comportamiento agresivo a un factor de modificación.

El análisis de Clúster permite la distribución de grupos, se recomienda sea utilizado en diferentes estudios que impliquen velocidades de operación, en especial en el campo de las vías terrestres. La utilización de datos normalizados, variables adicionales que permita el estudio y la ponderación de los datos, arrojan una constancia en el método, por lo cual se recomienda realizar estos procesos para validar los resultados.

Los elementos estadísticos utilizados como el árbol de decisión y regresión lineal, permiten un panorama decisivo y acertado en el trabajo, por lo cual se recomienda su uso para llegar a este tipo de conclusiones.

Finalmente, se recomienda aplicar el método utilizado para contextos diferentes. En Colombia se recomienda realizar toma de datos y análisis del proceso en diferentes tipos de carreteras, con las mismas o más variables usadas. Incluso se recomienda aplicar el estudio, tomando datos de accidentes y comparándolos con los supuestos y así proponer y aplicar acciones de mejora a la seguridad vial para validar el método.

4.3. Conclusiones

Podemos concluir con el presente estudio que el análisis de la velocidad de forma variable cuando el usuario transita por la curva, teniendo en cuenta las tangencias anterior y posterior, nos permite estudiar el comportamiento de los conductores y proponer su validación para términos de diseño geométrico y de seguridad vial. De igual forma se pudo constatar que el que el método Kmeans para la distribución de grupos es consistente y aplicable para este tipo de datos a partir del análisis de diferentes formas de agrupación y muestras derivadas del conjunto inicial. El número óptimo de agrupaciones es de tres tipos de comportamiento de los usuarios.

Adicionalmente se constata que las variables propuestas (η) y (a_s^+) brindan un mejor análisis de los datos al ser utilizadas y ponderarlas con las velocidades, pudiendo encontrar resultados más consistentes. La agrupación de datos para el estilo de conducción moderado que arroja el método para la data completa es del 70.46%, esto permite entender la dinámica de los usuarios de conducción en vías en buen estado y conductores prudentes en el uso de las carreteras. Este comportamiento se validó a partir de diferentes combinaciones y ponderaciones de las variables obtenidas.

Se pudo constatar la elección de velocidad de los usuarios cuando las vías de este tipo tienen unas características o configuraciones específicas. Al analizar estos resultados con la estimación de choques, permitió conocer y relacionar con perfiles específicos de velocidad y así poder proponer elementos que permitan mejorar dicha seguridad, entre otros incluir en métodos de predicción de accidentes y evaluación de carreteras seguras Del análisis se pueden proponer configuraciones de las carreteras para evitar perfiles agresivos y así disminuir la accidentalidad.

Finalmente, se puede esperar que con este estudio se abran las puertas para continuar realizando este tipo de investigaciones en diferentes tipos de carreteras y variados contextos culturales para el estudio del comportamiento de los usuarios. Se constata que es posible incluir el factor humano en un análisis de seguridad vial de curvas horizontales mediante perfiles de velocidad mostrando resultados congruentes y aplicables a la realidad.

Capítulo 5. Bibliografía

- AASHTO. (2010). *Highway Safety Manual*. <https://trid.trb.org/view/198544>
- AASHTO. (2018). A Policy on Geometric Design of Highways and Streets, 7th Edition. In *American Association of State Highway and Transportation Officials*. <https://aashtojournal.org/2018/09/28/aashto-releases-7th-edition-of-its-highway-street-design-green-book/>
- Al-Masaeid, H. R., Hamed, M., Aboul-Ela, M., & Ghannam, A. G. (1995). Consistency of horizontal alignment for different vehicle classes. *Transportation Research Record, 1500*, 178–183.
https://www.researchgate.net/publication/294621721_Consistency_of_horizontal_alignment_for_different_vehicle_classes/citations
- Alonso Pla, M. L. (2016). *La integración del factor humano en el ámbito técnico de la gestión de las carreteras y la seguridad vial: Un enfoque investigativo*. <http://roderic.uv.es/handle/10550/51943>
- Highway Safety Manual, (2010). <http://www.highwaysafetymanual.org/Pages/default.aspx>
- ASCE. (2019). *Report Card for Puerto Rico's Infrastructure*. 1–61. <https://www.infrastructurereportcard.org/wp-content/uploads/2019/11/2019-Puerto-Rico-Report-Card-Final.pdf>
- Baillo Moreno, A., & Grané Chávez, A. (2008). *100 Problemas resueltos de Estadística Multivariante (Implementados en Matlab)*. Delta Publicaciones. Madrid, España.
- Bermúdez Arbona, A. M. (2018). Velocidad y diseño geométrico: factores para identificar curvas potencialmente peligrosas en carreteras rurales de dos carriles. In *Angewandte Chemie International Edition, 6(11)*, 951–952. <https://hdl.handle.net/20.500.11801/2106>
- Bliss, T., & Breen, J. (2013). *Road Safety Management Capacity Reviews and Safe System Projects*. 136. <https://www.worldbank.org/en/topic/transport/publication/road-safety-management-capacity-review-guidelines>
- Bobermin, M. P., Silva, M. M., & Ferreira, S. (2021). Driving simulators to evaluate road geometric design effects on driver behaviour: A systematic review. *Accident Analysis &*

- Prevention*, 150, 105923. <https://doi.org/10.1016/J.AAP.2020.105923>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* (CHAPMAN &). [https://doi.org/https://doi.org/10.1201/9781315139470](https://doi.org/10.1201/9781315139470)
- Calero Valenzuela, C. A. (2015). *Metodología para la Evaluación de la Consistencia de Diseño de Carreteras Rurales de Dos Carriles* (Vol. 1, Issue 9) [Univ. Puerto Rico]. <https://hdl.handle.net/20.500.11801/840>
- Cárdenas Grisales, J. (2013). *Diseño Geométrico de Carreteras*. Ecoe Ediciones. www.ecoediciones.com
- Castillo, V. R. (2020). *Optimización De Factores En La Estimación De Velocidad Operacional En Curvas Horizontales*. Universidad del Cauca.
- Choudhari, T., & Maji, A. (2019). Socio-demographic and experience factors affecting drivers' runoff risk along horizontal curves of two-lane rural highway. *Journal of Safety Research*, 71, 1–11. <https://doi.org/10.1016/J.JSR.2019.09.013>
- Chu, D., Deng, Z., He, Y., Wu, C., Sun, C., & Lu, Z. (2017). Curve speed model for driver assistance based on driving style classification. *IET Intelligent Transport Systems*, 11(8), 501–510. <https://doi.org/10.1049/iet-its.2016.0294>
- Plan-Nacional-Desarrollo, (2018).
- Colucci Ríos, B., & Valle Javier, D. (2014a). Development of Crash Modification Factors for Rumble Strips Treatment for Freeway Applications : Phase I Development of Safety Performance Functions. *LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2014)*.
- Colucci Ríos, B., & Valle Javier, D. (2014b). Development of Safety Performance Functions for Signalized in Puerto Rico. *1st IRF Asia Regional Congress Paper Submission Form*.
- Consejo de Cambio Climatico de Puerto Rico CCCPR. (2013). *Estado del Clima de Puerto Rico*. 27.
- DataCamp. (2021). *Run.clustering: Clustering the data*. <https://www.rdocumentation.org/packages/iCellR/versions/1.6.1/topics/run.clustering>
- Devore, J. (2008). Probabilidad y estadística para ingeniería y ciencias. In *Cengage Learning: Vol.*

7th Ed.

- Dolatalizadeh, M., Boroujerdian, A. M., & Abrishami Seyed, Ehsan, S. (2020). Analysis of Speed Profiles at an Unsignalised Intersection for Left Turning Vehicles. *IJTE*, 8(2), 149–163. <https://doi.org/https://dx.doi.org/10.22119/ijte.2020.163493.1448>.
- Dong, S., & Zhou, J. (2020). A Comparative Study on Drivers' Stop/Go Behavior at Signalized Intersections Based on Decision Tree Classification Model. *Journal of Advanced Transportation*, 1–13. <https://doi.org/10.1155/2020/1250827>
- Elvik, R., Christensen, P., & Amundsen, A. (2004). Speed and road accidents: An evaluation of the Power Model. *TOI Report 740*, 740(December), 134. <http://www.trg.dk/elvik/740-2004.pdf>
- Federal Highway Administration. (2021). *Crash Modification Factors Clearinghouse*. U.S. Department of Transportation. <http://www.cmfclearinghouse.org/>
- Fitzpatrick, K., Elefteriadou, L., Harwood, D. W., Collins, J. M., McFadden, J., Anderson, I. B., Krammes, R. A., Irizarry, N., Parma, K. D., Bauer, K. M., & Passetti, K. (2000). Speed Prediction for Two-Lane Rural Highways. *Fhwa-Rd-99-171*, 99, 1–217. http://www.nrcresearchpress.com/doi/abs/10.1139/104-103#.U_9KaEu3I6I
- Fu, H., Deng, F., Shao, Y., Liu, Y., & Zhang, J. (2021). Road Centreline Extraction of High - Resolution Remote Sensing Image with Improved Beamlet Transform and K - Means Clustering. *Arabian Journal for Science and Engineering*, 46(4), 4153–4162. <https://doi.org/10.1007/s13369-021-05412-1>
- Gibreel, G. M., Easa, S. M., Member, A., Y. Hassan, A., & El-Dimeery, I. A. (1999). State of the art of highway geometric design consistency. *J. Transp. Eng*, 125(4), 305–313. [https://doi.org/https://doi.org/10.1061/\(ASCE\)0733-947X\(1999\)125:4\(305\)](https://doi.org/https://doi.org/10.1061/(ASCE)0733-947X(1999)125:4(305)).
- González Vergara, C. J., Rincón Villalba, M. A., & Vargas Vargas, W. E. (2019). *Ingeniería De Vías Diseño, trazado y localización de Carreteras* (Vol. 1). <https://www.librosyeditores.com/ingenieria/12892-ingenieria-de-vias-diseno-trazado-y-localizacion-de-carreteras-9789587785562.html>
- Google LLC. (2021). *Google Earth Pro*. <https://www.google.com/intl/es/earth/>

- Hastie, T., Tibshirani, R., & Friedman, J. (2011). The Elements of Statistical Learning: Data mining, inference, and prediction. *Springer Series in Statistics*, 26(4), 505–516. <https://doi.org/10.1007/978-0-387-84858-7>
- Instituto Nacional de Vías, M. (2008). Manual de Diseño Geométrico de Carreteras. In *Ciencia* (Vol. 84). <http://crc.gov.co/files/ConocimientoAmbiental/POT/inza/02-SubsistemaPolitico.pdf><https://www.iadb.org/es/acerca-del-bid/politica-de-desarrollo-rural%2C6229.html>https://elpais.com/elpais/2016/08/26/ciencia/1472225923_727879.html<https://www.redalyc.html>
- Jahangiri, A., Machiani, S. G., & Balali, V. (2019). Big Data Exploration to Examine Aggressive Driving Behavior in the Era of Smart Cities. In *Data Analytics for Smart Cities*. Auerbach Publications. <https://doi.org/10.1201/9780429434983-7/BIG-DATA-EXPLORATION-EXAMINE-AGGRESSIVE-DRIVING-BEHAVIOR-ERA-SMART-CITIES-ARASH-JAHANGIRI-SAHAR-GHANIPOOR-MACHIANI-VAHID-BALALI>
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264–323. <https://doi.org/10.1145/331499.331504>
- Keklikoglou, A., Fitzpatrick, C. D., & Knodler, M. A. (2018). Investigation of Time and Speed Perception using a Driving Simulator: *Transportation Research Record*, 2672(37), 132–140. <https://doi.org/10.1177/0361198118776519>
- Lamm, R., Choueiri, E. M., Hayward, J. C., & Paluri, A. (1988). Possible design procedure to promote design consistency in highway geometric design on two-lane rural roads. *Transportation Research Record*, 1195, 111–122. <http://onlinepubs.trb.org/Onlinepubs/trr/1988/1195/1195-011.pdf>
- Lamm, R., Choueri M, E., & Mailaender, T. (1991). Side Friction Demand Versus Side Friction Assumed for Curve Design on Two-Lane Rural Highways. *Transportation Research Record*, 1303, 11–21. <http://onlinepubs.trb.org/Onlinepubs/trr/1991/1303/1303-002.pdf>
- Lamm, R., Hayward, J. C., & Cargin, J. G. (1986). Comparison of Different Procedures for Evaluating Speed Consistency. *Transportation Research Record*, 10–20. [https://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds\[\]=citjournalarticle_688320_38](https://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds[]=citjournalarticle_688320_38)

- Laureshyn, A., Åström, K., & Brundell-Freij, K. (2009). From speed profile data to analysis of behaviour. *IATSS Research*, 33(2), 88–98. [https://doi.org/10.1016/S0386-1112\(14\)60247-8](https://doi.org/10.1016/S0386-1112(14)60247-8)
- Liu, T., & Xu, J. (2019). The influence rule of highway curve radius on speed perception and operating speed. *IOP Conference Series: Earth and Environmental Science*, 371(2), 022080. <https://doi.org/10.1088/1755-1315/371/2/022080>
- Llopis-Castelló, D., Camacho-Torregrosa, F. J., & García, A. (2018). Development of a global inertial consistency model to assess road safety on Spanish two-lane rural roads. *Accident Analysis and Prevention*, 119, 138–148. <https://doi.org/10.1016/j.aap.2018.07.018>
- Malaghan, V., Pawar, D. S., & Dia, H. (2020). Modeling Operating Speed Using Continuous Speed Profiles on Two-Lane Rural Highways in India. *Journal of Transportation Engineering, Part A: Systems*, 146(11), 04020124. <https://doi.org/10.1061/jtepbs.0000447>
- McFadden, J., & Elefteriadou, L. (2000). Consistency of Two-Lane Rural Highways Development of New Procedure. *Transportation Research Record*, 1737(1), 9–17. <https://doi.org/https://doi.org/10.3141/1737-02>
- MetroCount. (2021). *MetroCount*. <https://metrocount.com/>
- Ministerio de Transporte. (2015). *Plan Nacional de seguridad vial Colombia 2011-2021*.
- Ministerio de Transporte. (2021). *Metodología para el Desarrollo de Auditorias e Inspecciones de Seguridad Vial para Colombia*.
- Misaghi, P., & Hassan, Y. (2005). Modeling operating speed and speed differential on two-lane rural roads. *Journal of Transportation Engineering*, 131(6), 408–418. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2005\)131:6\(408\)](https://doi.org/10.1061/(ASCE)0733-947X(2005)131:6(408))
- Morcillo, L. G., Poyo, F. J. C., Fernández, M. P., & de Oña, J. (2014). Measurement of Road Consistency on Two-lane Rural Highways in Granada (Spain). *Procedia - Social and Behavioral Sciences*, 162, 237–242. <https://doi.org/10.1016/j.sbspro.2014.12.204>
- Ms Anna Vadeby, Blair Turner, K. M. (2018). Speed and crash risk. *ITF (International Transport Forum)*, 82p. <https://www.itf-oecd.org/speed-crash-risk%0Ahttps://trid.trb.org/view/1508779>
- NationMaster.com. (2021). *NATIONMASTER*. <https://www.nationmaster.com/country->

info/compare/Colombia/United-States/Transport

- Papadimitriou, E., Mavromatis, S., Pavlou, D., & Yannis, G. (2017). Assessment of Speeding Profiles and Safety Margins from Tangent to Curve by means of Driving Simulation. *Proceedings of the Road Safety & Simulation International Conference*, 17–19. <https://www.nrso.ntua.gr/geyannis/wp-content/uploads/geyannis-pc279.pdf>
- Parra, F. (2016). *Curso de Estadística con R*. Instituto Cántabro de Estadística. https://www.icane.es/c/document_library/get_file?uuid=c2e9fff1-72d5-42ad-b391-bacb3ebe9dbe&groupId=10138
- Piantadosi, S., Byar, D. P., & Green, S. B. (1988). The ecological fallacy. *American Journal of Epidemiology*, 127(5), 893–904. <https://doi.org/10.1093/OXFORDJOURNALS.AJE.A114892/2/127-5-893.PDF.GIF>
- Poe, C. M., Tarris, J. P., & Mason Jr, J. M. (1996). *RELATIONSHIP OF OPERATING SPEEDS TO ROADWAY GEOMETRIC DESIGN SPEEDS*. <http://trid.trb.org/view.aspx?id=572374>
- Polus, A., & Mattar-Habib, C. (2004). New consistency model for rural highways and its relationship to safety. *Journal of Transportation Engineering*, 130(3), 286–293. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2004\)130:3\(286\)](https://doi.org/10.1061/(ASCE)0733-947X(2004)130:3(286))
- Rincón Numpaque, N., Moreno Anselmi, L. Á., Rodríguez Polo, K. A., & Gaviria Mendoza, C. A. (2020). Alternatives to improve operational traffic in roundabouts using microsimulation. *Respuestas Journal of Engineering Sciences*, 25(2), 26–36. <https://doi.org/https://doi.org/10.22463/0122820X.2372>
- Rodríguez Polo, K. A., & Henao Pérez, S. (2019). Safety performance functions in Dedicated Bus Lane of BRT on Caracas Avenue Corridor at Bogotá city. *Inge Cuc*, 15(2), 66–77. <https://doi.org/10.17981/ingecuc.15.2.2019.07>
- Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, K. Y. (2012). *Probabilidad y estadística para ingeniería y ciencias* (Novena edi). PEARSON EDUCACIÓN.
- Rosas Palomino, A. (2019). *Determinación de la velocidad específica de curvas horizontales en carreteras rurales sobre terreno montañoso*. Universidad del Cauca.
- Sánchez Ordóñez, J. F. (2011). *Metodología Para La Evaluación De La Consistencia Del Trazado*

De Carreteras Interurbanas De Dos Carriles. Universidad Politécnica de Madrid.

- Secretary- General of the International Transport. (2020). *ROAD SAFETY ANNUAL REPORT* (Vol. 2507, Issue February).
- United Nations Road Safety. (2011). Global plan for the Decade of Action for Road Safety 2011–2020. Geneva: WHO, 25. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Global+Plan+for+the+Decade+of+Action+for+Road+Safety+2011-2020#0>
- Vos, J., Farah, H., & Hagenzieker, M. (2021). Speed behaviour upon approaching freeway curves. *Accident Analysis & Prevention*, 159, 106276. <https://doi.org/10.1016/J.AAP.2021.106276>
- Wooldridge, M. D., Fitzpatrick, K., Koppa, R., & Bauer, K. (2000). Effects of horizontal curvature on driver visual demand. *Transportation Research Record*, 1737, 71–77. <https://doi.org/10.3141/1737-09>
- World Health Organization. (2018). *Global status report on road safety* (Issue 4). World Health Organization. <https://apps.who.int/iris/handle/10665/189242>.
- Yan, M., Li, M., He, H., & Peng, J. (2018). Deep learning for vehicle speed prediction. *Energy Procedia*, 152, 618–623. <https://doi.org/10.1016/j.egypro.2018.09.220>
- Yang, J., Chu, D., Wang, R., Gao, M., & Wu, C. (2020). Coupling effect modeling of driver vehicle environment factors influencing speed selections in curves. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 234(7), 2066–2078. <https://doi.org/10.1177/0954407019870349>
- Zolali, M., Mirbaha, B., Layegh, M., & Behnood, H. R. (2021). A Behavioral Model of Drivers' Mean Speed Influenced by Weather Conditions, Road Geometry, and Driver Characteristics Using a Driving Simulator Study. *Advances in Civil Engineering*, 2021, 18. <https://doi.org/10.1155/2021/5542905>

Capítulo 6. Anexos

6.1. Elección de cantidad óptima de conglomerados – Rstudio

A continuación, se presenta el proceso de programación utilizado para el desarrollo de la obtención del número óptimo de conglomerados, el cual a partir de la data utiliza diferentes métodos estadísticos para encontrar cantidad óptima de conglomerados, según lo comentado en el capítulo 2.5.4. Este número es una guía para que el investigador realice la elección adecuada.

```
#Muestreo de datos
```

```
indices <- sample( 1:nrow( practica), 149)
df<- practica[ indices, ]
```

```
#Preparación de datos
```

```
#normalizar las puntuaciones
```

```
df <- scale(df)
head(df)
```

```
#Get_dist (): para calcular una matriz de distancia entre las filas de una matriz de datos.
```

```
#Comparado con la función dist () estándar, soporta medidas de distancia basadas en la correlación incluyendo
```

```
#los métodos “pearson”, “kendall” y “spearman”.
```

```
#Fviz_dist (): para visualizar una matriz de distancia
```

```
library(factoextra)
```

```
library(ggpubr)
```

```
mat_dist <- dist(x = df,method = "euclidean")
```

```
round(as.matrix(mat_dist)[1:5, 1:5], 2)
```

```
fviz_dist(dist.obj = mat_dist, lab_size = 5) +
```

```
theme(legend.position = "none")
```

```
## Estadístico H para el set de datos cerca a 0
```

```
library(clustertend)
```

```
set.seed(321)
```

```
hopkins(data = df, n = nrow(df) - 1)
```

```
#Visual Assessment of cluster Tendency (VAT)
```

```
p1 <- fviz_dist(dist.obj = ,mat_dist, show_labels = FALSE) +
```

```
  labs(title = "Curva") + theme(legend.position = "bottom")
```

```
ggarrange(p1)
```

```
#Elbow method
```

```
fviz_nbclust(x = df, FUNcluster = kmeans, method = "wss", k.max = 15) +
```

```
  labs(title = "Número óptimo de clusters")
```

```
#Average silhouette method
```

```
fviz_nbclust(x = df, FUNcluster = kmeans, method = "silhouette", k.max = 15) +
```

```
  labs(title = "Número óptimo de clusters")
```

```
#Gap statistic method
```

```
set.seed(500)
```

```
fviz_nbclust(x = df, FUNcluster = kmeans, method = "gap_stat", nboot = 500,
```

```
  k.max = 15, verbose = FALSE, nstart = 50) +
```

```
  labs(title = "Número óptimo de clusters")
```

```
#incorpora 30 índices distintos (Hubert statistic - Dindex - Conclusión )
```

```
#https://www.rdocumentation.org/packages/NbClust/versions/3.0/topics/NbClust Debe ser uno de los
```



```
#siguientes: "kl", "ch", "hartigan", "ccc", "scott", "marriot", "trcovw", "tracew",  
"friedman", "rubin",  
#"cindex", "db", "silueta", "duda", "pseudot2", "beale", "ratkowsky", "bola", "ptbiserial",  
"gap", "frey", "mcclain",  
#"gamma", "gplus ", " tau ", " dunn ", " hubert ", " sindex ", " dindex ", " sdbw ", " all "(todos  
los índices excepto GAP, Gamma, Gplus y Tau),  
#" alllong "(todos los índices con Gap , Gamma, Gplus y Tau incluidos).  
  
library(NbClust)  
numero_clusters <- NbClust(data = df, distance = "euclidean", min.nc = 2,  
                           max.nc = 10, method = "kmeans", index = "alllong")  
  
fviz_nbclust(numero_clusters)
```

6.2. Distribución en conglomerados – Rstudio

En esta sección se puede encontrar la programación utilizada para distribuir los conglomerados en 3 grupos, según sus condiciones. El proceso matemático se puede encontrar en la sección 2.5.4 del presente documento.

- # remover todo lo corrido

```
rm(list=ls(all=TRUE))
```

```
# Se decide trabajar con la base reducida
```

```
setwd("D:/1.CRISTIAN PERSONALES/2.UNICAUCA Post-grado/Tesis/Informe  
final/Ponderaci3n y sensibilizacion/Ponderaci3n")
```

```
#ejem<-read.table("BaseDatosClusterNormSpeed.txt", header=TRUE)
```

```
#ejem<-read.table("BaseDatosClusterNormRelT1Speed.txt", header=TRUE)
```

```
#ejem<-read.table("BaseDatosClusterNormRelVpreviosSpeed.txt", header=TRUE)
```

```
# ejem<-read.table("BaseDatosClusterNorm102-1 y 116-1.txt", header=TRUE)
```

```
ejem<-read.table("D:/1.CRISTIAN PERSONALES/2.UNICAUCA Post-  
grado/Tesis/Informe final/Ponderaci3n y sensibilizacion/Ponderaci3n/Datos2Param.txt")
```

```
ejem<-Datos2Param
```

```
# ejem<-read.table("BaseDatosClusterNormCurveID.txt", header=FALSE)
```

```
#definici3n del factor Ocupacion
```

```
#ejem$OCUPACION<-
factor(ejem$OCUPACION,labels=c("Gerente","Administrador","Vendedor","Auxiliar
Contable","Servicios Generales","Produccion","Obrero"))
```

```
# Prepare Data
```

```
#ejem <- na.omit(ejem) # listwise deletion of missing
```

```
#ejem <- scale(ejem) # standardize variables
```

```
# Determine number of clusters
```

```
wss <- (nrow(ejem)-1)*sum(apply(ejem,2,var))
```

```
for (i in 2:100) wss[i] <- sum(kmeans(ejem,centers=i)$withinss)
```

```
plot(1:100, wss, type="b", xlab="Number of Clusters",
```

```
ylab="Within groups sum of squares")
```

```
#importar datos Cristian Rosas
```

```
library(readr)
```

```
ejem <- read_delim("D:/1.CRISTIAN PERSONALES/2.UNICAUCA Post-
grado/Tesis/Informe final/PonderaciÃ³n y sensibilizacion/CorrecciÃ³n dist. y
norma/2variables.txt",
```

```
"\t", escape_double = FALSE, trim_ws = TRUE)
```

```
#Se organizan los datos en un 'data.frame'
```

```
datos<-data.frame(cbind(ejem))
```

```
datos
```

```
# K-Means Cluster Analysis
```

```
fit <- kmeans(datos, 3) # 21 cluster solution
```

```
fit <- kmeans(datos, 3, iter.max = 1) # 21 cluster solution
```

```
# Cluster Plot against 1st 2 principal components
```

```
# vary parameters for most readable graph
```

```
library(cluster)
```

```
# graficar los puntos sin las elipses de los clusters
```

```
x <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
```

```
clusplot(datos, x, color=FALSE, shade=FALSE, span = TRUE,
```

```
  labels=0, lines=0, xlab = "X", ylab = "Y", col.clus = "white", main = "initial data")
```

```
# graficar un primer agrupamiento en 2 grupos iniciales (primera iteracion)
```

```
x <- c(1, 1, 1, 1, 2, 2, 2, 2, 2, 2)
```

```
clusplot(datos, x, color=FALSE, shade=FALSE, span = TRUE,
```

```
  labels=0, lines=0, xlab = "X", ylab = "Y", col.clus = "black", main = "Iteration 1")
```

```
# graficar un primer agrupamiento en 2 grupos iniciales (segunda iteracion)
```

```
x <- c(1, 1, 1, 1, 1, 2, 2, 2, 2, 2)
```

```
clusplot(datos, x, color=FALSE, shade=FALSE, span = TRUE,
```

```
  labels=0, lines=0, xlab = "X", ylab = "Y", col.clus = "black", main = "Iteration 2")
```

```

#fit$size

#install.packages("zoom")

#library(zoom) # Invoke the Library

# Call plot

#zm()

# write to csv file

write.csv(fit$cluster,"D:/1.CRISTIAN PERSONALES/2.UNICAUCA Post-
grado/Tesis/Informe final/PonderaciÃ³n y sensibilizacion/CorrecciÃ³n dist. y
norma/BaseDatosClusterNorm2variables.csv", row.names = TRUE)

write.csv(fit$centers,"D:/1.CRISTIAN PERSONALES/2.UNICAUCA Post-
grado/Tesis/Informe final/PonderaciÃ³n y sensibilizacion/CorrecciÃ³n dist. y
norma/BaseDatosClusterNorm2variablesCenters.csv", row.names = TRUE)

#write.csv(fit$cluster,"C:\\Users\\andres\\Desktop\\Ponencia
TRB\\R\\BaseDatosClusterNormRelT1Speed3clusters.csv", row.names = TRUE)

#write.csv(fit$centers,"C:\\Users\\andres\\Desktop\\Ponencia
TRB\\R\\BaseDatosClusterNormRelT1Speed3clustersCenters.csv", row.names = TRUE)

#write.csv(fit$cluster,"C:\\Users\\andres\\Desktop\\Ponencia
TRB\\R\\BaseDatosClusterNormRelVpreviosSpeed3clusters.csv", row.names = TRUE)

#write.csv(fit$centers,"C:\\Users\\andres\\Desktop\\Ponencia
TRB\\R\\BaseDatosClusterNormRelVpreviosSpeed3clustersCenters.csv", row.names = TRUE)

```

```
datoscor<-cor(datos); datoscor  
# distancia m?xima entre dos componentes de X y Y  
dd<-dist(datoscor,method="maximum")  
# conformaci?n de los conglomerados  
cl<-hclust(dd)  
plot(cl,hang = -1)
```

6.3. Gráficas de Elección de cantidad óptima de conglomerados – Matlab

La graficación obtenida para realizar el presente estudio, debido a la gran cantidad de datos y las agrupaciones que se obtuvieron, fue realizada con la siguiente programación. EN la cual se identifica el numero de datos según el numero de agrupación de cada dato. Se tiene en cuenta la normalización de los datos.

```

clc
close all
clear all
%
% % data matrix with columns: (1)Direction (2)CurveID (3)Cluster # (4)T1 (5)PC (6)MC (7)PT
(8)T2
data =0;
%
%
% % coordinates of clusters centers (1)T1 (2)PC (3)MC (4)PT (5)T2
coordclusters = 0;

% sort coordclusters to a custom cluster order
% new row assign for plot
rowcluster = [2 3 1];
coordclusters2 = zeros(size(coordclusters));
for i = 1 : numel(rowcluster)
    coordclusters2(i,:) = coordclusters(rowcluster(i,:));
end
coordclusters = coordclusters2;
% also sort the data matrix (requeried)
[numvehi,~] = size(data);
data2 = zeros(numvehi,1);
for j = 1 : numel(rowcluster)
    for i = 1: numvehi
        clusterid = data(i,3); % cluster ID
        if clusterid == j
            data2(i,1) = rowcluster(j);
        end
    end
end
end
data(:,3) = data2;

x = 1:5;
[numvehi,numvar] = size(data);
[numclusters,~] = size(coordclusters);
% coordclusters = [ones(numclusters,1),coordclusters];
% % % % Plot each cluster centers % % % %

```

```

hFig2 = figure; set(hFig2, 'Position', [50 50 400 200])
set(gca,'FontSize',12)
hold on
linetype1 = {'--o','+', '-.^'};
for j = 1 : numclusters
    disp("")
    disp('particular cluster')
    disp(['a) cluster ID: ',num2str(j)])
    k = 1;
    dataclusterj = [];
    for i = 1: numvehi
        clusterid = data(i,3); % cluster ID
        if clusterid == j
            dataclusterj(k,1:numvar) = data(i,:);
            k = k + 1;
        end
    end
    disp(['b) the number of measured vehicles on cluster are: ',num2str(k-1)])
    % plot one vehicle profile and the cluster centers coordinates
    % plot(x, coordclusters(j,:).*mean(dataclusterj(:,4)),linetype1 {j}, 'LineWidth',2)
    plot(x, coordclusters(j,:),linetype1 {j}, 'LineWidth',2)

    pause
end
% % format to plot, axes and ticks
xticks([1 2 3 4 5])
xticklabels({'T1','PC','MC','PT','T2'})
xlabel('Punto de la Curva','FontSize',12,'Fontweight','bold')
ylabel('V / V_{límite}','FontSize',14,'Fontweight','bold')
ylim([0.5 1.5])
yt=get(gca,'YTick');
for j = 1: numel(yt)
    yttemp{j} = num2str(sprintf('%3.1f',yt(j)));
end
set(gca,'YTickLabel',yttemp)
set(gca,'box','on')
% add legend
hleg2 = legend("a","b","c");
set(hleg2, 'Box', 'on','Orientation','horizontal','FontSize',12,'Fontweight','bold');
grid on



---


clc
close all

```



```

clear all

% data vector with value of the parameter
data =0;
% Orden real de los cluster 2 3 1;

numIntervals = 50;
intervalWidth = 0.35/numIntervals;
x = 0:intervalWidth:0.35;

% Find the frequency of each data range "x" in the given data set "y".
ncount = histc(data,x);

% Calculate the relative frequency of each data range by dividing the frequency by the total
number of data points
relativefreq = ncount/length(data);

% Plot the relative frequency versus the data ranges as a bar chart. On this chart,
hFig2 = figure; set(hFig2, 'Position', [0 0 333 140])
set(gca,'box','on','FontSize',11)
hold on
bar(x-intervalWidth/2, relativefreq,1)
ylabel('frec.','FontSize',11,'Fontweight','bold')
xlim([0 0.35])

xt=get(gca,'XTick');
for j = 1: numel(xt)
    xttemp{j} = num2str(sprintf('%3.1f',xt(j)));
end
set(gca,'XTickLabel',xttemp)

ylim([0 1])
% yt=get(gca,'YTick');
% for j = 1: numel(yt)
%     yttemp{j} = num2str(sprintf('%3.2f',yt(j)));
% end
% set(gca,'YTickLabel',yttemp)

set(gca,'box','on')
xlabel('a_s^+','FontSize',12,'Fontweight','bold')
set(gca, 'YScale', 'log')
grid on



---


clc
close all

```

```

clear all

% data vector with value of the parameter
data = 0;
% Orden real de los cluster 2 3 1;

% bar chart of the relative frequency of data "y" in the data range -8 to 94 at intervals of 17.
numIntervals = 50;
intervalWidth = 1/numIntervals;
x = 0:intervalWidth:1;

% numIntervals = 50;
% intervalWidth = 0.35/numIntervals;
% x = 0:intervalWidth:0.35;

% Find the frequency of each data range "x" in the given data set "y".
ncount = histc(data,x);

% Calculate the relative frequency of each data range by dividing the frequency by the total
number of data points
relativefreq = ncount/length(data);

% Plot the relative frequency versus the data ranges as a bar chart. On this chart,
hFig2 = figure; set(hFig2, 'Position', [0 0 333 140])
set(gca,'box','on','FontSize',11)
hold on
bar(x-intervalWidth/2, relativefreq,1)
    ylabel('frec.','FontSize',11,'Fontweight','bold')
%     xlabel('a_s^+','FontSize',10,'Fontweight','bold')
%     xlim([0 0.35])

xlim([0 1])
xt=get(gca,'XTick');
for j = 1: numel(xt)
    xtemp{j} = num2str(sprintf('%3.1f',xt(j)));
end
set(gca,'XTickLabel',xtemp)

ylim([0 1])
% yt=get(gca,'YTick');
% for j = 1: numel(yt)
%     ytemp{j} = num2str(sprintf('%3.2f',yt(j)));
% end
% set(gca,'YTickLabel',ytemp)

```

```

set(gca,'box','on')
xlabel('\eta','FontSize',12,'Fontweight','bold')
set(gca,'box','on')
set(gca, 'YScale', 'log')
grid on

```

```

clc
close all
clear all

```

```

% data matrix with columnas: (1)Direction (2)CurveID (3)Cluster # (4)T1 (5)PC (6)MC (7)PT
(8)T2
data =0;

```

```

% coordinates of clusters centers
coordclusters = 0;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% verifica cluster
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
disp('-- plot vehicules records at each cluster --')
x = 1:5;
[numvehi,numvar] = size(data);
[numclusters,~] = size(coordclusters);
for j = 1 : numclusters
    disp("")
    disp('particular vehicules at cluster')
    disp(['a) cluster ID: ',num2str(j)])
    k = 1;
    dataclusterj = [];
    for i = 1: numvehi
        clusterid = data(i,3); % cluster ID
        if clusterid == j
            dataclusterj(k,1:numvar) = data(i,:);
            k = k + 1;
        end
    end
    disp(['b) the number of measured vehicules on cluster are: ',num2str(k-1)])
    %%%% Plot the vehicules asociate to each cluster centers %%%%

```

```

hFig2 = figure; set(hFig2, 'Position', [0 0 333 140])
set(gca,'box','on','FontSize',11)
hold on
% plot one vehicle profile and the cluster centers coordinates
plot(x, dataclusterj(1,4:end),'color',[0 0 0])
plot(x, coordclusters(j,:),'-*','color',[0 0 0],'LineWidth',4, 'MarkerEdgeColor',[0 0 0])
% plot all vehicle profile
for i = 1: k-1
    plot (x, dataclusterj(i,4:end))
end
% plot the cluster centers coordinates again to superimpose
plot(x,coordclusters(j,:),'-*','color',[0 0 0],'LineWidth',4, 'MarkerEdgeColor',[0 0 0])
% % format to plot, axes and ticks
xticks([1 2 3 4 5])
xticklabels({'T1','PC','MC','PT','T2'})
xlabel('Punto de la Curva','FontSize',11,'Fontweight','bold')
ylabel('V / V_{límite}','FontSize',13,'Fontweight','bold')
ylim([0.2 2.4])
yt=get(gca,'YTick');
for j = 1: numel(yt)
    yttemp{j} = num2str(sprintf('%3.1f',yt(j)));
end
set(gca,'YTickLabel',yttemp)
set(gca,'box','on')
% add legend
% hleg2 = legend('individual profiles','cluster center');
% set(hleg2, 'Box', 'on','Orientation','horizontal','FontSize',10,'Fontweight','bold');
% plot notes over figure
gtext([num2str(k-1,'%5.0f'),' vehículos'],'FontSize',11,'Fontweight','bold')
grid on
hold off
pause
close all
end

```

6.4. Árbol de decisión – Rstudio

El proceso para aplicar el árbol de dedición descrito en la sección 2.5.5 se realizó con la siguiente programación, la cual a partir de los diferentes datos de entrada en variables y elementos agrupados, construye dicho según las ramificación necesarias.

```
library(caret)

library(rpart)

library(rpart.plot)

library(randomForest)

library(readxl)

library(tidyverse)
```

```
R      <-read_excel("D:/1.CRISTIAN PERSONALES/2.UNICAUCA Post-
grado/Tesis/Informe final/PonderaciÃ³n y sensibilizacion/ponderar 0.5 y
0.5/DatosPanelArbol.xlsx",
```

```
      col_types = c("skip", "skip", "skip",
                    "skip", "skip", "skip", "text",
                    "numeric", "text", "numeric", "numeric",
                    "numeric", "text", "numeric", "numeric"))
```

```
DatosF<-R%>%
```

```
  select(Cluster, Radio, `Type of Terrain`, G, Slope, `Sight distance`, `speed limit`,
```

```
         `vel diseÃ±o`, `Functional classification`)%>%
```

```
  rename(Tipo_de_Terreno='Type of Terrain',
```

```
         Distancia_Vis='Sight distance',
```

```
         Lim_vel='speed limit',
```

```
         Dis_vel='vel diseÃ±o',
```

```
Clas_funcional='Functional classification')
```

```
summary(DatosF)
```

```
filas.entrenamiento <- createDataPartition(DatosF$Cluster, p = 0.7, list = FALSE)
```

```
conjuntoEntrenamiento<-DatosF[filas.entrenamiento,]
```

```
conjuntoTest<-DatosF[-filas.entrenamiento,]
```

```
# vamos a hacer un modelo de clasificaci3n basado en 3rboles de decisi3n
```

```
modeloArbolDecision <- rpart( Cluster ~.,  
                             data = DatosF[filas.entrenamiento,],  
                             method = "class")
```

```
modeloArbolDecision
```

```
prp(modeloArbolDecision, type = 2, extra = 102)
```

```
#hacemos una predicci3n sobre el conjunto de test
```

```
prediccionArbolDecision <- predict(modeloArbolDecision,
```

```
DatosF[-filas.entrenamiento,], type="class")
```

```
#añadimos la predicción en una columna nueva
```

```
DatosF[-filas.entrenamiento, c(P_arbol_decision)]<-prediccionArbolDecision
```

```
#datos e imagen de otra forma
```

```
print(modeloArbolDecision)
```

```
rpart.plot(modeloArbolDecision)
```

```
#estadísticos y error
```

```
printcp(modeloArbolDecision)
```

```
plotcp(modeloArbolDecision)
```

```
#poda del árbol (no es necesario en este caso)
```

```
parbolRpart<-prune(modeloArbolDecision,cp=0.010041 )
```

```
printcp(parbolRpart)
```

```
rpart.plot(parbolRpart)
```

```
#capacidad de predicción
```

```
table(prediccionArbolDecision,conjuntoTest$Cluster)
```

#porcentaje de aciertos

```
sum(prediccionArbolDecision==conjuntoTest$Cluster)/length(conjuntoTest$Cluster)*1
```

00

6.5. Análisis de regresión lineal – Stata

El Proceso de regresión lineal se realizó con la siguiente formulación de programación, este se describe en la sección 2.5.6 y permite con los datos utilizados encontrar una ecuación lineal que adapte a las variables midiendo su significación.

```
import excel "D:\1.CRISTIAN PERSONALES\2.UNICAUCA Post-grado\Tesis\Informe
final\Ponderación y sensibilizacion\ponderar 0.5 y 0.5\DatosPanelArbol.xlsx",
sheet("APanelModerado (2)") firstrow case(lower)
```

```
xtset id puntos
```

```
bysort puntos: egen y_mean=mean(vel)
```

```
twoway scatter vel puntos, msymbol(circle_hollow) || connected y_mean puntos
```

```
regress vel radio g vd
```

```
estimates store modelolineal
```

```
xi: regress vel radio g vd i.puntos
```

```
estimates store modelolinealdummy
```

```
areg vel radio g vd, absorb(puntos)
```

```
estimates store modelolinealabsorb
```

```
xtreg vel radio g vd, fe
```

estimates store modelpanelijos

xtreg vel radio g vd, re

estimates store modelpanelrandom

hausman modelolinealdummy modelpanelrandom

estimates store pruebah

outreg2 [modelolineal modelolinealdummy modelolinealabsorb modelpanelijos
modelpanelrandom] using Agresivo.doc, replace see