



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

**FACTORES ASOCIADOS A SINIESTROS DE TRÁNSITO EN EL
ECUADOR, MEDIANTE LA COMPARACIÓN DE MODELO DE
REGRESIÓN Y ANÁLISIS DE CORRESPONDENCIAS EN EL
PERIODO ENERO 2018 – MAYO 2021**

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar el grado académico de:

INGENIERO ESTADÍSTICO

AUTORES:

COLOMA LLANGO LEONEL ALEXANDER

GUAYCHA IMBAQUINGO JONATHAN ROBERTO

DIRECTORA: ING. JOHANNA ENITH AGUILAR REYES, MGS.

Riobamba – Ecuador

2023

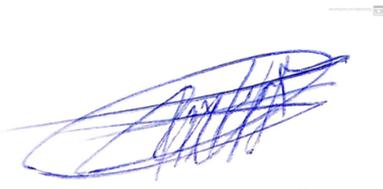
©2023, Leonel Alexander Coloma Llango y Jonathan Roberto Guaycha Imbaquingo

Autorizamos la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Nosotros, JONATHAN ROBERTO GUAYCHA IMBAQUINGO y LEONEL ALEXANDER COLOMA LLANGO, declaramos que el presente Trabajo de Integración Curricular es de nuestra autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autores asumimos la responsabilidad legal y académica de los contenidos de este Trabajo de Integración Curricular; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 14 de abril de 2023



Coloma Llango Leonel Alexander

0604213835



Guaycha Imbaquingo Jonathan Roberto

1750297317

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

El Tribunal del Trabajo de Integración Curricular certifica que: el Trabajo de Integración Curricular; Tipo: Proyecto de Investigación. **FACTORES ASOCIADOS A SINIESTROS DE TRÁNSITO EN EL ECUADOR, MEDIANTE LA COMPARACIÓN DE MODELO DE REGRESIÓN Y ANÁLISIS DE CORRESPONDENCIAS EN EL PERIODO ENERO 2018 – MAYO 2021**, realizado por el señor: **JONATHAN ROBERTO GUAYCHA IMBAQUINGO** y el señor: **LEONEL ALEXANDER COLOMA LLANGO**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Integración Curricular, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal autoriza su presentación.

	FIRMA	FECHA
Ing. Tania Paulina Morocho Barrionuevo PRESIDENTE DEL TRIBUNAL	_____	2023-04-14
Ing. Johanna Enith Aguilar Reyes DIRECTORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR	_____	2023-04-14
Ing. Paulina Fernanda Bolaños Logroño ASESORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR	_____	2023-04-14

DEDICATORIA

Este proyecto de investigación en primer lugar esta dedicado a Dios, por la fortaleza y por brindarme las herramientas para culminar esta carrera, al igual a mis padres María Imbaquingo y Robert Guaycha, hermanos, sobrinos, abuelitos, Jenny Guaycha, Micaela Guaycha, Esteban Guaycha , Emiliano Vera, Danae Vera, Daniel Cueva y demás personas que de una u otra forma fueron parte de mi vida universitaria, en especial a mis padres por la confianza brindada en todos los momentos sean buenos o malos, por la enseñanza a afrontar los problemas y los valores que son el ejemplo y referente en mi vida.

Jonathan

Quiero dedicar este trabajo de investigación a Dios por regalarme salud y fuerzas para cumplir todos mis objetivos. A mis padres Melida Llango y Marcelo Coloma, a mi hermana, sobrina Adriana Coloma, Hanna Barragán ya que con su apoyo, amor, sabiduría y ejemplo eh logrado ser la persona que soy, también a mi abuelita que me apoya desde el cielo que me da la fuerza que necesito para culminar el primer escalón de mi vida profesional.

Leonel

AGRADECIMIENTO

Agradecemos la oportunidad de ser parte de esta noble institución Escuela Superior Politécnica de Chimborazo, que la llevamos en nuestro corazón, a los directivos, personal administrativo y docente, de manera especial a la Ing. Johanna Aguilar y a la Ing. Paulina Bolaños por cada una de las enseñanzas tanto académicas como su ejemplo de mujeres valientes y por la tutoría para el desarrollo de este proyecto de investigación. Y a todas las personas que nos brindaron su apoyo en estos años de carrera, nuestro más sinceros agradecimientos, los llevamos en nuestros corazones.

Jonathan

Leonel

ÍNDICE DE CONTENIDO

ÍNDICE DE TABLAS	x
ÍNDICE DE ECUACIONES	xi
ÍNDICE DE ILUSTRACIONES	xii
ÍNDICE DE ANEXOS	xiv
RESUMEN	xv
ABSTRACT	xvi
INTRODUCCIÓN	1
CAPÍTULO I	
1. PROBLEMA DE INVESTIGACIÓN	5
1.1. Planteamiento del Problema	5
1.2. Limitaciones y delimitaciones	5
1.3. Problema General de Investigación	6
1.4. Problemas específicos de investigación	6
1.5. Objetivos	6
1.5.1. <i>Objetivo general</i>	6
1.5.2. <i>Objetivos específicos</i>	6
1.6. Justificación	7
1.6.1. <i>Justificación Teórica</i>	7
1.6.2. <i>Justificación Metodológica</i>	7
1.6.3. <i>Justificación Práctica</i>	7

CAPÍTULO II

2.	MARCO TEÓRICO	8
2.1.	Antecedentes de investigación	8
2.1.1.	<i>Antecedentes Históricos</i>	9
2.2.	Referencias Teóricas	9
2.2.1.	<i>Siniestros de tránsito</i>	9
2.2.2.	<i>¿Qué es un accidente?</i>	9
2.2.3.	<i>Principales causas de accidentes de tránsito</i>	10
2.2.4.	<i>Factores que intervienen en accidente de tránsito</i>	10
2.2.4.1.	<i>Factores humanos</i>	10
2.2.4.2.	<i>Factores Mecánicos (Vehículo)</i>	10
2.2.4.3.	<i>Factores Climáticos</i>	11
2.2.4.4.	<i>Factores viales</i>	11
2.2.5.	<i>Factor</i>	11
2.3.	Referencias Conceptuales	11
2.3.1.	<i>Modelación de accidentes</i>	11
2.3.2.	<i>Agencia nacional de tránsito</i>	11
2.3.3.	<i>Accidentes de tránsito</i>	12
2.3.4.	<i>Clases de accidentes</i>	12
2.3.5.	<i>Muestreo aleatorio estratificado proporcionado</i>	12
2.3.6.	<i>Análisis exploratorio de datos</i>	13
2.3.7.	<i>Regresión logística multinomial</i>	13
2.3.8.	<i>Análisis de correspondencia</i>	14

CAPÍTULO III

3.	MARCO METODOLÓGICO	17
3.1.	Enfoque de investigación	17
3.2.	Nivel de Investigación	17
3.3.	Diseño de investigación	17
3.3.1.	<i>Según la manipulación o no de la variable independiente</i>	17
3.3.2.	<i>Según las intervenciones en el trabajo de campo</i>	17
3.4.	Tipo de estudio	17

3.5.	Población y Planificación, selección y cálculo del tamaño de la muestra . . .	18
3.5.1.	<i>Muestreo estratificado proporcionado</i>	18
3.6.	Métodos, técnicas e instrumentos de investigación	20
3.7.	Operacionalización de las variables	20

CAPÍTULO IV

4.	MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS	25
4.1.	Análisis descriptivo	25
4.2.	Análisis de correspondencia	41
4.2.1.	<i>Análisis de correspondencia múltiple</i>	41
4.2.2.	<i>Análisis de correspondencia simple</i>	42
4.2.3.	<i>Regresión Logística Multinomial</i>	51
4.2.4.	<i>Comparación del modelo de regresión con el análisis de correspondencias</i>	53

CONCLUSIONES	54
-------------------------------	-----------

RECOMENDACIONES	56
----------------------------------	-----------

BIBLIOGRAFÍA

ANEXOS

ÍNDICE DE TABLAS

Tabla 1-3 : Muestreo estratificado proporcionado.	19
Tabla 2-3 : Operacionalización de las variables.	20
Tabla 1-4 : Frecuencias absolutas de la variable Provincia.	25
Tabla 2-4 : Frecuencias absolutas de la variable Clase.	27
Tabla 3-4 : Frecuencias absolutas de la variable Mes.	29
Tabla 4-4 : Frecuencias absolutas de la variable Feriado.	30
Tabla 5-4 : Frecuencias absolutas de la variable Días.	31
Tabla 6-4 : Frecuencias absolutas del número de fallecidos.	32
Tabla 7-4 : Frecuencias absolutas del número de lesionados.	36
Tabla 8-4 : Proporción de inercia de las variables provincia y clase.	42
Tabla 9-4 : Proporción de inercia de las variables día y clase.	44
Tabla 10-4 : Proporción de inercia de las variables mes y clase.	48
Tabla 11-4 : Área bajo la curva ROC	53

ÍNDICE DE ECUACIONES

Ecuación 1–2	: Modelo RLM	13
Ecuación 2–2	: Logit de la probabilidad	13
Ecuación 3–2	: Demostración de la probabilidad	13
Ecuación 4–2	: Logaritmo de la probabilidad	13
Ecuación 5–2	: Modelo para la probabilidad logarítmica	14
Ecuación 6–2	: Fórmula alternativa de $\pi(x)$	14
Ecuación 7–2	: Modelo de regresión logística múltiple	14
Ecuación 8–2	: Valores de π	14
Ecuación 9–2	: Fórmula de la distribución estadística de numerosidad bivalente	15
Ecuación 10–2	: Estadístico de prueba chi-cuadrado	15
Ecuación 11–2	: Valor crítico	16
Ecuación 12–2	: Región crítica	16
Ecuación 13–2	: Toma de decisión	16
Ecuación 14–2	: Coordenadas principales de X	16
Ecuación 15–2	: Coordenadas principales de Y	16
Ecuación 1–3	: Fórmula para el cálculo de Muestreo Aleatorio Simple	18
Ecuación 2–3	: Cálculo del tamaño muestral	18
Ecuación 3–3	: Muestreo estratificado proporcionado	18
Ecuación 4–3	: Cálculo de k constante positiva	18

ÍNDICE DE ILUSTRACIONES

Ilustración 1–2 : Análisis de correspondencias múltiples.....	15
Ilustración 1–4 : Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.....	25
Ilustración 2–4 : Diagrama de barras de las frecuencias absolutas de la variable Provincia.	27
Ilustración 3–4 : Diagrama de barras de las frecuencias absolutas de la variable Clase de siniestros viales.....	28
Ilustración 4–4 : Diagrama de barras de las frecuencias absolutas de la variable Mes.....	30
Ilustración 5–4 : Diagrama de barras de las frecuencias absolutas de la variable Feriado. .	31
Ilustración 6–4 : Diagrama de barras de las frecuencias absolutas de la variable Día.	32
Ilustración 7–4 : Diagrama de barras del número de fallecidos.....	34
Ilustración 8–4 : Diagrama de barras del número de fallecidos.....	35
Ilustración 9–4 : Diagrama de barras de las frecuencias absolutas de la variable número de fallecidos.....	36
Ilustración 10–4 : Diagrama de barras de las frecuencias absolutas de la variable número de lesionados.....	38
Ilustración 11–4 : Diagrama de barras de las frecuencias absolutas de la variable número de lesionados.....	39
Ilustración 12–4 : Diagrama de barras de las frecuencias absolutas de la variable número de lesionados.....	40
Ilustración 13–4 : Análisis de correspondencia múltiple.....	41
Ilustración 14–4 : Porcentaje de variabilidad explicada del análisis de correspondencia múltiple.....	42
Ilustración 15–4 : Porcentaje de variabilidad explicada de las variables provincia y clase..	43
Ilustración 16–4 : Análisis de correspondencia simple entre la variable Provincia y Clase.	44
Ilustración 17–4 : Porcentaje de variabilidad explicada de las variables Día y clase.	45
Ilustración 18–4 : Análisis de correspondencia simple entre la variable Día y Clase.....	46
Ilustración 19–4 : Porcentaje de variabilidad explicada de las variables clase y Feriado. .	47
Ilustración 20–4 : Diagrama de barras de las frecuencias de las variables Clase y Feriado.	48
Ilustración 21–4 : Porcentaje de variabilidad explicada de las variables mes y clase.....	49

Ilustración 22-4 :	Análisis de correspondencia simple entre la variable Mes y Clase.	50
Ilustración 23-4 :	Creación de variables dummy o variables ficticias.	51
Ilustración 24-4 :	Resumen del modelo de regresión logística multinomial.	52
Ilustración 25-4 :	Predicciones del modelo de regresión de los siniestros de tránsito.	52
Ilustración 26-4 :	Matriz de confusión para medir la precisión.	53

ÍNDICE DE ANEXOS

ANEXO A: CÓDIGO DE R

RESUMEN

El presente trabajo tiene como objetivo principal determinar los factores que se asocian con los siniestros de tránsito mediante la comparación de modelo de regresión y análisis de correspondencia en el periodo de enero 2018- mayo 2021. La base de datos utilizada en este trabajo investigativo se obtuvo de la página oficial de la Agencia Nacional de Tránsito (ANT), la base de datos tuvo un total de 169987 siniestros de tránsito, con la cual por fines de optimización de procesos de análisis de datos se realizó una muestra aleatoria simple con población finita, posteriormente se realizó un muestreo estratificado proporcional en el cual las provincias del Ecuador se tomaron como estratos. En el análisis estadístico descriptivo se obtuvo como resultado que las provincias con más siniestros viales son Pichincha, Guayas y Tungurahua, y las provincias con menos accidentes en las vías fueron Orellana, Sucumbíos, Pastaza y Galápagos, además que los días que no son feriados nacionales o locales se producen más accidentes de tránsito, también que el choque lateral es el más recurrente, el mes de diciembre presenta más incidentes de este tipo y por último que ocurren con más frecuencia los fines de semana, con respecto al análisis de correspondencias múltiple con las variables, mes, día, provincia, clase de siniestro vial, feriado, también se realizó entre pares de variables, además en la regresión logística multinomial y el área bajo la curva (AUC) se obtuvo el 60.32 %; por lo que, observando el valor del porcentaje de variabilidad de la gráfica, el modelo tiene un buen poder predictivo que el análisis de correspondencia. Se recomienda que las instituciones públicas o privadas interesadas en la investigación y reducción de siniestros viales, lo cual se considera en Ecuador una de las mayores causas de muerte, daños personales o materiales irreparables, difundan esta información para contribuir a la creación de estrategias que permitan reducir al máximo estos accidentes y concienticen a los conductores, peatones e instituciones gubernamentales que prestan servicios de controles viales.

Palabras clave: <ESTADÍSTICA>, <REGRESIÓN LOGÍSTICA>, <ANÁLISIS DE CORRESPONDENCIAS>, <SINIESTROS VIALES>, <ACCIDENTES DE TRÁNSITO>, <ECUADOR (PAÍS)>.



0319-DBRA-UPT-2023

ABSTRACT

The main objective of this work was to determine the factors that are associated with traffic accidents by comparing the regression model and correspondence analysis in the period from January 2018 to May 2021. The database used in this research work was obtained from the official page of the National Transit Agency (ANT), the database had a total of 169,987 traffic accidents, with which, for purposes of optimizing data analysis processes, a simple random sample with a finite population was subsequently done. A proportional stratified sampling was performed in which the provinces of Ecuador were taken as strata. In the descriptive statistical analysis, it was obtained as a result that the provinces with the most road accidents are Pichincha, Guayas and Tungurahua, and the provinces with the least accidents on the roads were Orellana, Sucumbíos, Pastaza and Galapagos, in addition more traffic accidents occur the days that are not national or local holidays, also the side collision is the most recurrent sort of accident, the month of December presents more incidents of this type and they occur more frequently on weekends, with respect to the multiple correspondence analysis with the variables, month, day, province, type of road accident, holiday, were also done between pairs of variables, in addition to multinomial logistic regression and the area under the curve (AUC) 60.32 % was obtained; Therefore, observing the value of the percentage of variability of the graph, the time model has better predictive power than the correspondence analysis. It is recommended that public or private institutions interested in the investigation and reduction of road accidents, which is considered in Ecuador one of the major causes of death, personal injury or irreparable material damage, disseminate this information and contribute to the creation of strategies that reduce these accidents to the maximum and raise awareness among drivers, pedestrians and government institutions that provide road control services.

Keywords: <STATISTICS>, <LOGISTIC REGRESSION>, <CORRESPONDENCE ANALYSIS>, <TRAFFIC ACCIDENTS>, <ECUADOR (COUNTRY)>.

Edgar Mesias Jaramillo Moyano

0603497397

INTRODUCCIÓN

La Organización Mundial de la Salud (OMS) estima que los accidentes de tránsito provocan alrededor de 1,3 millones de muertes, incluidos niños, de los cuales los peatones, ciclistas y los ancianos se encuentran entre los usuarios más vulnerables de las vías públicas (OMS, 2022a: p.1).

De acuerdo con António Guterres secretario general de la ONU menciona que “Los accidentes de tránsito son una silenciosa epidemia que esta ambulante”, en su discurso ante los representantes de 193 estados miembros representantes de la ONU que debaten el camino a seguir con el objetivo de reducir a la mitad el número de víctimas letales para el año 2030, compromiso que se verá reflejado en una declaración al final del evento político (OMS ,2022b: p.1).

Por otro lado, la Organización Internacional de Accidentología Vial (OIAV) informa que las personas, los vehículos, la infraestructura vial y el medio ambiente son factores de riesgo en la seguridad vial, el alcance de estos riesgos se reduce gradualmente con la ayuda de la gestión, un problema que implica costo y proceso (Congacha et al., 2019; citado en OIAV, 2015). Además los siniestros de tránsito son un problema de salud pública prioritario a nivel mundial, principalmente en países de Latinoamérica, por ser una de las principales causas de muerte en la región (Congacha et al., 2019; citado en Algora, Russo, Suasnavas, Merino, & Gómez, 2017).

En 2015 la Comisión Nacional de Seguridad de México informó que los principales factores de accidentes de tránsito, con el 80% son causados por los conductores, seguido de los vehículos con un 7%, los factores naturales el 9% y finalmente la carretera con el 4% (Congacha et al., 2019; citado en CNS, 2015). Al mismo tiempo (Congacha et al., 2019; citado en Román,2015), en su trabajo titulado Integración de programas de seguridad vial al modelo ecuatoriano, propone cuatro factores principales relacionados con los accidentes de tránsito, factores humanos, factores del vehículo, factores ambientales y organización del trabajo siendo el factor humano de mayor incidencia.

Un estudio de las características de la siniestralidad vial en Ecuador mostró que entre los años 1998 a 2015 fallecieron 29.148 personas en accidentes de tránsito, identificándose las principales causas de las muertes en los accidentes de tránsito en donde se atribuyen a la imprudencia del conductor con respecto a las franjas horarias para determinar el peor accidente vial que ocurre entre las 18:00 y 21:00, llegando a la conclusión que se pudiera deber a la mala iluminación (Congacha et al., 2019; citado en Gómez-García et al.,2016).

En Ecuador se produjeron 21.575 accidentes entre 2015 y 2018, con un resultado de 4.004 muertos

y 6.920 con lesiones entre leves y graves. La Dirección Nacional de Control de Tránsito y Seguridad Vial (DNT) ha definido 28 causas de los accidentes de tránsito y se dividen en las relacionadas con la imprudencia del conductor, Imprudencia de peatones, imprudencia de pasajeros, daños mecánicos, factores climáticos, defectos u obstáculos y accidentes en la vía (Congacha et al., 2019; citado en Dirección Nacional de Control de Tránsito y Seguridad Vial, 2019a) . Los siniestros de tránsito de hoy en día son problemas de salud pública, las empresas públicas y privadas les están dedicando recursos para realizar estudios de seguridad vial, de la misma manera países e instituciones y la Agencia Nacional de Tránsito ANT está desarrollando normas para reducir los problemas antes mencionados.

El artículo de investigación de (Chavez y Useche, 2021: pp.1231–1250) titulado “Perfil descriptivo de los accidentes de tránsito en Ecuador durante el período 2015-2020”, menciona que los problemas de salud pública en gran parte son causados por los siniestros de tránsito. A pesar de que existen varios estudios en Ecuador sugieren que la presencia de la pandemia del Covid-19 pudo haber cambiado la frecuencia de las causas del número de accidentes y las tendencias o pronósticos resultantes, es por ello que el objetivo es caracterizar las causas probables de los siniestros de tránsito.

En Ecuador entre los años 2015-2020 se determinó las razones del mayor número de impacto que ocurrían en los siniestros de tránsito, como resultado se describen a los accidentes de tránsito en el Ecuador, además de que hubo una disminución general en el número de accidentes debido a la pandemia de coronavirus Covid-19, dicho estudio menciona que los siniestros de tránsito ocurren más a menudo los fines de semana que los días festivos, la hora más recurrente es las 19:00, usualmente en las zonas urbanas de las ciudades, además de que las provincias donde ocurrió más accidentes de tránsito son Pichincha y Guayas, con una colisión lateral, en consecuencia se da a conocer que el sexo más recurrente en estos siniestros es el masculino, se estima que al menos 1 persona resultó herida en el 36 % al 40 % de los casos y al menos hubo un lesionado , finalmente en el 10 % de los casos termino en el fallecimiento de una persona.

Según el artículo de (Flórez et al., 2018: pp. 69–85) titulado “Análisis multicausal de ‘Accidentes’ de tránsito en dos ciudades de Colombia”, a medida que aumenta las problemáticas de salud pública, hay una necesidad de desarrollar soluciones sólidas que mejoren la seguridad de los sistemas de transporte vial. Para sacar conclusiones, se deben estudiar cuidadosamente los factores relacionados con la ocurrencia de los accidentes de tránsito, en Ibagué se constató que hasta 20 causas desencadenantes estuvieron involucrados en los accidentes de tránsito, el valor más común fue con 9 causas (19 %), seguido de siniestros relacionados con 12 causas involucradas (18 %), se obtuvo un caso con una causa asociada. En el caso de Valledupar, se encontró que incidentes similares estaban asociados con hasta 18 factores desencadenantes, que son más comunes se

encontró que los siniestros estaban relacionados con 12 causas (21 %) y 11 razones (14 %). Los incidentes con causas secundarias relacionadas tuvieron 8 y 9 causas (3 % cada una). En las dos ciudades se confirmó la multicausalidad para los accidentes de tránsito, es decir, se encontró que para las ciudades, el factor humano es el principal causante de que ocurra un siniestro de tránsito, agravado por evidentes deficiencias infraestructurales y algunos problemas organizativos.

De acuerdo con Reyes en su artículo "Siniestralidad por Accidentes de Tránsito en México", analizó las relaciones entre datos de accidentes de tránsito y factores de exposición relacionados (residentes de la ciudad, automóviles, áreas urbanas, vías urbanas y marginación). Se usó herramientas de análisis espacial y el método de estratificación de Dalenius y Hodge. La premisa es que los accidentes de tránsito son el resultado del desarrollo y crecimiento de un país. El análisis utilizado corresponde a las 32 entidades federativas de la República Mexicana, usando los datos absolutos de accidentes de tránsito en las ciudades y suburbios del país. El resultado permite relativizar, jerarquizar y visualizar distribuciones objetivas y deterministas del fenómeno de los accidentes de tránsito en las zonas urbanas del país (Reyes et al., 2012: pp. 291-309).

El objetivo del presente estudio es determinar los factores que se asocian con los siniestros de tránsito mediante la comparación de modelo de regresión y análisis de correspondencia en el periodo de enero 2018- mayo 2021, en consideración de que esta caracterización contribuirá a la toma de decisiones por parte de las instituciones públicas, privadas, para la posible reducción de los siniestros viales.

El artículo se encuentra organizado en cuatro capítulos, el capítulo I se encuentra el problema de investigación en el cual se identifica el problema, hipótesis para el proceso de investigación, en el capítulo II se presenta el marco teórico, referencias teóricas, mientras que el capítulo III marco metodológico el diseño, tipo, métodos usados en este proyecto de investigación, y el capítulo IV marco de análisis e interpretación de resultados en el cual se desarrolló el procesamiento, análisis estadístico e interpretación de resultados; comparación de modelo de regresión y análisis de correspondencias, finalmente se encuentran las conclusiones y recomendaciones en el capítulo V.

Con el transcurso del tiempo la estadística se ha incorporado poco a poco en las diferentes áreas científicas, contribuyendo a su análisis de datos y la verificación de sus resultados, especialmente en áreas de estudio como Salud, Economía, Marketing, etc. Mediante un el uso de técnicas estadísticas en el manejo de grandes volúmenes de datos y el uso de softwares estadísticos de análisis, manipulación y control de datos como: R, Spss, Excel, Minitab, etc. con el fin de solventar las interrogantes planteadas en la sección de problemas específicos de la investigación y responder

el objetivo principal, así determinar cuáles son los factores que se asocian con los siniestros de tránsito en Ecuador en el periodo de enero 2018- mayo 2021.

Finalmente, se tomaron en cuenta variables de tipo cualitativas como cuantitativas que son los factores que están asociados a siniestros de tránsito en el Ecuador; en donde se puede observar las variables: provincia, periodo de tiempo, feriado, clase, número de fallecidos, número de lesionados y mes, es necesario realizar un análisis estadístico para obtener una aproximación de la realidad de los factores asociados a siniestros de tránsito en el Ecuador, mediante la comparación de modelo de regresión y análisis de correspondencias en el periodo enero 2018 – mayo 2021.

CAPÍTULO I

1. PROBLEMA DE INVESTIGACIÓN

1.1. Planteamiento del Problema

Según (OMS, 2022a: p.1), aproximadamente 1,3 millones de personas mueren cada año como producto de los siniestros de tránsito. En América Latina debido a los accidentes de tránsito, muchas calles se tiñen de sangre y mucho de los casos por conductores que infringen las normas de tránsito, lo que lleva a la pérdida de un familiar y en la mayor parte de los casos con la mutilación de sus miembros que los deja con una discapacidad permanente, al igual que los peatones poco o nada de caso respetan las señales de tránsito.

Datos del INEC mencionan que en el Ecuador para el año 2021 se han registrado 21352 siniestros de tránsito un mayor número comparado con respecto al año 2020 que se han registrado 16972 siniestros, del total de víctimas registradas en el año 2021 estas presentaban una edad entre los 18 a 29 años destacando que la mayor cantidad de siniestros de tránsito involucraron a automóviles y motocicletas y la zona donde mayor ocurrencia se aconteció estos accidentes fue en el área urbana.

Se ha podido reflejar la importancia del análisis estadístico en el campo de la investigación científica, los trabajos de Titulación de pregrado, los cuales a su vez contienen bases de datos bibliográficas de suma relevancia. A si mismo resultados ya sea de manera descriptiva, inferencial, predictiva, etc, que evidencian las falencias de los factores que inciden para que esta problemática no se termine con el pasar del tiempo.

1.2. Limitaciones y delimitaciones

Una limitación respecto del trabajo de titulación es que existe datos limitados de los factores asociados a siniestros viales en las fuentes oficiales, la más relevante es la Agencia Nacional de Tránsito ANT. Respecto a sus delimitaciones, abarca un tiempo que corresponde al año 2018-2021, la base de datos descargada de la ANT de los factores asociados a los siniestros de tránsito en el Ecuador.

1.3. Problema General de Investigación

¿Qué factores se encuentran asociados a los siniestros de tránsito en el Ecuador aplicando una comparación de un modelo de regresión y un análisis de correspondencia durante el periodo enero 2018 - mayo 2021?

1.4. Problemas específicos de investigación

- ¿Cuáles son los posibles factores asociados a los siniestros de tránsito para determinar a través de un marco teórico?
- ¿Cómo se recolecta la información de los posibles factores, usando las bases publicadas en la Agencia Nacional de Tránsito?
- ¿Cómo se predice los siniestros de tránsito a través de un modelo de regresión y un análisis de correspondencia?
- ¿Cómo comparar la efectividad de los modelos usados con base a la precisión de la predicción?

1.5. Objetivos

1.5.1. *Objetivo general*

- Determinar los factores que se asocian con los siniestros de tránsito mediante la comparación de modelo de regresión y análisis de correspondencia que demuestra la incidencia, de un siniestro de tránsito en Ecuador en el periodo de enero 2018 – mayo 2021.

1.5.2. *Objetivos específicos*

- Determinar a través de un marco teórico pertinente los posibles factores asociados a los siniestros de tránsito que servirá como base para interpretar los resultados obtenidos a partir del estudio realizado.
- Recolectar la información de los posibles factores, usando las bases publicadas en la Agencia Nacional de Tránsito lo cual es asegurar que se obtiene información confiable.
- Predecir los siniestros de tránsito a través de un modelo de regresión y un análisis de correspondencia destinadas a reducir la posibilidad del siniestro de tránsito.
- Comparar la efectividad de los modelos usados con base a la precisión de la predicción para decidir el modelo mas adecuado en la investigación.

1.6. Justificación

1.6.1. *Justificación Teórica*

Para realizar el análisis se va a optar por; la aplicación de un modelo de regresión y análisis de correspondencia con el estudio de las variables que intervienen en los accidentes de tránsito en el Ecuador como una línea base en cuanto a estadísticas de siniestralidad, permitiendo conocer los posibles factores que originan un evento vial desafortunado, para generar propuestas que permita reducir los siniestros mostrados en el periodo de estudio.

1.6.2. *Justificación Metodológica*

El objetivo de este proyecto de investigación es determinar los factores que se asocian con los siniestros de tránsito mediante la comparación de modelo de regresión y análisis de correspondencia en el periodo de enero 2018- mayo 2021, considerando que este tipo de estudios son de gran utilidad en las líneas de investigación en accidentes de tránsito, a partir de modelos de regresión y análisis de correspondencia, con el uso del software estadístico R. La importancia de recolectar la información de los posibles factores, usando las bases publicadas en la Agencia Nacional de Tránsito radica la posibilidad de la creación de propuestas que permita reducir los accidentes de tránsito mostrados en el periodo de estudio. Además, el estudio se justifica al comparar la efectividad de los modelos de regresión para obtener predicciones lo más certeras posibles de los siniestros de tránsito en Ecuador.

1.6.3. *Justificación Práctica*

El desarrollo del presente trabajo investigativo contribuirá para determinar el comportamiento de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021 a través de un modelo de regresión y análisis de correspondencia, se podrá establecer la relación existente entre los tipos de siniestros y las clases, así como la mayor zona de peligrosidad a nivel nacional, cabe mencionar que la metodología aplicada nos dará a conocer sobre el accionar de los conductores en situaciones de riesgo al momento de estar frente a un volante, así como también dar a conocer de que falencias existen en temas relacionados al tráfico de la ciudad, a fin de tomar decisiones para reducir la siniestralidad.

CAPÍTULO II

2. MARCO TEÓRICO

2.1. Antecedentes de investigación

El artículo de investigación de (Chavez y Useche , 2021: pp.1231–1250) titulado “Perfil descriptivo de los accidentes de tránsito en Ecuador durante el período 2015-2020”, menciona que los problemas de salud pública en gran parte son causados por los siniestros de tránsito. A pesar de que existen varios estudios en Ecuador sugieren que la presencia de la pandemia del Covid-19 pudo haber cambiado la frecuencia de las causas del número de accidentes y las tendencias o pronósticos resultantes, es por ello que el objetivo es caracterizar las causas probables de los siniestros de tránsito.

En Ecuador entre los años 2015-2020 se determinó las razones del mayor número de impacto que ocurrían en los siniestros de tránsito, como resultado se describen a los accidentes de tránsito en el Ecuador, además de que hubo una disminución general en el número de accidentes debido a la pandemia de coronavirus Covid-19, dicho estudio menciona que los siniestros de tránsito ocurren más a menudo los fines de semana que los días festivos, la hora más recurrente es las 19:00, usualmente en las zonas urbanas de las ciudades, además de que las provincias donde ocurrió mas accidentes de tránsito son Pichincha y Guayas, con una colisión lateral, en consecuencia se da ha conocer que el sexo más recurrente en estos siniestros es el masculino, se estima que al menos 1 persona resultó herida en el 36 % al 40 % de los casos y al menos hubo un lesionado , finalmente en el 10 % de los casos termino en el fallecimiento de una persona.

Según el artículo de (Flórez et al., 2018: pp. 69–85) titulado “Análisis multicausal de ‘Accidentes’ de tránsito en dos ciudades de Colombia” , a medida que aumenta las problemáticas de salud pública, hay una necesidad de desarrollar soluciones solidas que mejoren la seguridad de los sistemas de transporte vial. Para sacar conclusiones, se deben estudiar cuidadosamente los factores relacionados con la ocurrencia de los accidentes de tránsito, en Ibagué se constató que hasta 20 causas desencadenantes estuvieron involucrados en los accidentes de tránsito, el valor más común fue con 9 causas (19%), seguido de siniestros relacionados con 12 causas involucradas (18%), se obtuvo un caso con una causa asociada. En el caso de Valledupar, se encontró que incidentes similares estaban asociados con hasta 18 factores desencadenantes, que son más comunes se encontró que los siniestros estaban relacionados con 12 causas (21%) y 11 razones (14%). Los incidentes con causas secundarias relacionadas tuvieron 8 y 9 causas (3% cada una). En las dos ciudades se confirmó la multicausalidad para los accidentes de tránsito, es decir, se encontró que

para las ciudades, el factor humano es el principal causante de que ocurra un siniestro de tránsito, agravado por evidentes deficiencias infraestructurales y algunos problemas organizativos.

De acuerdo con Reyes en su artículo "Siniestralidad por Accidentes de Tránsito en México", analizó las relaciones entre datos de accidentes de tránsito y factores de exposición relacionados (residentes de la ciudad, automóviles, áreas urbanas, vías urbanas y marginación). Se usó herramientas de análisis espacial y el método de estratificación de Dalenius y Hodge. La premisa es que los accidentes de tránsito son el resultado del desarrollo y crecimiento de un país. El análisis utilizado corresponde a las 32 entidades federativas de la República Mexicana, usando los datos absolutos de accidentes de tránsito en las ciudades y suburbios del país. El resultado permite relativizar, jerarquizar y visualizar distribuciones objetivas y deterministas del fenómeno de los accidentes de tránsito en las zonas urbanas del país (Reyes et al., 2012: pp. 291-309).

2.1.1. Antecedentes Históricos

Agencia Nacional de Tránsito ANT, es el organismo responsable de la planificación, regulación y control en el Transporte Terrestre, Tránsito y Gestión de la Seguridad Vial en todo el país, que brinda un transporte terrestre libre y seguro con servicios de calidad que permitan satisfacer las demandas de la población, y la protección del medio ambiente impulsando al desarrollo del país (AGENCIA NACIONAL DE TRÁNSITO, 2022a: 1A).

2.2. Referencias Teóricas

2.2.1. Siniestros de tránsito

Según datos de la Organización Mundial de la Salud (OMS), las víctimas a causa de siniestros viales están definidas como lesiones mortales y no mortales sufridas en un accidente de tránsito, por otro lado un accidente de tránsito se define como una colisión o incidente en la carretera que puede provocar lesiones en la persona e incluso su muerte (Congacha et al., 2019: pp.17-29).

2.2.2. ¿Qué es un accidente?

La palabra accidente se deriva del latín accidens, que significa algo que sucede u ocurre inesperadamente porque no es parte de lo natural (Cabrerizo, 2016, p.470).

2.2.3. Principales causas de accidentes de tránsito

Hasta el mes de febrero de 2018 hubo 4.035 accidentes de tránsito en todo el país, por otro lado según la Agencia Nacional de Tránsito; la cifra de fallecidos llegó a 347, siendo el 37 % de estos accidentes que ocurrieron al conducir sin prestar atención a las condiciones de tránsito, o ya sea por el uso del teléfono celular, pantalla de video, ingerir alimentos, maquillaje o cualquier otra distracción, mientras que el 36 % tuvo cambios inapropiados repentinos en el carril, el 26 % por exceder los límites de velocidad y el 23 % no logra mantener las distancias laterales mínimas de seguridad entre vehículos (Gómez García et al., 2018: pp.2-3).

2.2.4. Factores que intervienen en accidente de tránsito

2.2.4.1. Factores humanos

El factor humano se incluye en todos los factores de riesgo a los que nos enfrentamos, cuando se trata de vehículos y el entorno vial. De hecho, son medios al servicio del individuo. El automóvil y la vía es un medio que utiliza el hombre y es el único responsable de adaptar o no su comportamiento para un uso adecuado. En definitiva, el conductor toma la decisión de participar en el acto de manejar una "máquina" que debe controlar. En este sentido, podemos decir que el factor conductor, depende de cuando el conductor decida revisar el automóvil o cuándo quiera elegir una forma de viajar si se aproxima una tormenta, también es quien decide cumplir o no las indicaciones de una señal de tránsito, o disminuir la velocidad y tomar las precauciones en caso de un mal estado de la carretera (Montoro et al., 2000, p.384).

2.2.4.2. Factores Mecánicos (Vehículo)

Cuando se habla de factores mecánicos, se suele hacer referencia al estado del vehículo para asegurar su correcto funcionamiento, algunos autores sugieren que existe una interrelación entre el factor humano y el factor mecánico ya que están ligados entre sí, el descuido o la falta de reparación del automóvil se considera como una irresponsabilidad por parte del propietario del vehículo , además el acudir a establecimientos que no se les ha otorgado un permiso que les garantiza un funcionamiento efectivo de los trabajos de mantenimiento pone en riesgo la vida del conductor (Morales y Tierra, 2019: pp.41-44).

2.2.4.3. *Factores Climáticos*

Factores en los cuales los involucrados en un accidente de tránsito no se ven involucrados directamente. Pero cualquier evento, como los factores humanos y mecánicos pueden manejarse de forma responsable, es decir, de acuerdo con ciertos estándares y sentido común, para que los accidentes de tránsito no ocurran con mayor frecuencia. Sin embargo en general, ocurre lo contrario (Morales y Tierra, 2019: pp.41–44).

2.2.4.4. *Factores viales*

Las vías se consideran como todo espacio público o privado donde circulan toda clase de automóviles, peatones, animales, etc. De esta definición se puede decir que los factores viales que provocan accidentes están relacionados con el tipo de vía u otro espacio para la circulación de las personas, los espacios deben diseñarse y construirse a partir de normativas que regulan la correcta ejecución del trabajo de acuerdo a ciertos estándares con el fin de reducir el número de accidentes (Morales y Tierra, 2019: pp.41–44).

2.2.5. *Factor*

Un Factor es el elemento o causa que actúa junto con otros (RAE, 2007, p.307).

2.3. **Referencias Conceptuales**

2.3.1. *Modelación de accidentes*

Quiñones señaló que los modelos son expresiones matemáticas que permiten predecir eventos del futuro, observando hechos reales, a través de diferentes escenarios para la evaluación tanto del presente como del futuro. En cuanto al fenómeno de los accidentes, se destacan los siguientes puntos: modelo de análisis de correspondencias y modelo de regresión.

2.3.2. *Agencia nacional de tránsito*

La Agencia Nacional de Tránsito (ANT) realiza los trámites a través de la Secretaría General, sobre el bloqueo de documentos por la prohibición de enajenar un vehículo (por: juicio, embargo, quiebra o liquidación). Se inicia desde la recepción del pedido hasta el bloqueo en el sistema (AGENCIA NACIONAL DE TRÁNSITO, 2022b: 1A).

2.3.3. Accidentes de tránsito

Es aquello que sucede en las vías de forma inesperada, bajo la influencia de las circunstancias, y el comportamiento irresponsable potencialmente predecible en donde se involucra; factores humanos, vehículos principalmente automotores, clima, señalización y carreteras en mal estado, que causan muertes prematuras y/o lesiones, secuelas físicas y psicológicas, daños materiales y perjuicios a terceros (INEC, 2022, p.1).

2.3.4. Clases de accidentes

- **Colisión:** Una colisión es un choque que involucra a uno o más vehículos en movimiento.
- **Atropello:** Acción en la que un vehículo en circulación golpea o arrolla a uno o varios peatones.
- **Los términos colisión y atropello, atropello y vuelco, colisión y vuelco:** Se utilizan para definir una serie de incidentes interrelacionados considerados para la elaboración de estadísticas, como un solo accidente, según la secuencia en que ocurren los eventos.
- **Caída de persona o cosa del vehículo en marcha:** Se refiere a una situación en la que una persona u objeto cae de un vehículo en movimiento, causando lesiones personales o daños a la propiedad.
- **Accidentes de tránsito fatales:** Se refiere a todos los accidentes de tránsito que resultan en la muerte de una o más personas.

2.3.5. Muestreo aleatorio estratificado proporcionado

El muestreo estratificado es una técnica de muestreo estadístico que consiste en dividir la población de estudio en distintos subgrupos o estratos. El muestreo estratificado es una técnica de muestreo estadístico, con el objetivo de separar la población en segmentos diferentes, homogéneos llamados estratos, y luego extraer una muestra aleatoria de cada segmento (estrato).

Las muestras seleccionadas de diferentes estratos se combinan en una sola muestra. Este procedimiento de muestreo se denomina en ocasiones como “muestreo aleatorio por cuotas”. Una característica principal de la estratificación es que cada elemento debe pertenecer a uno y solo un único estrato, de modo que los estratos son excluyentes (no se superponen). Para lograr una estratificación adecuada es importante definir una variable que permita asignar a cada elemento un único grupo o estrato (Ochoa, 2015, p.1).

2.3.6. Análisis exploratorio de datos

Antes de aplicar técnicas de inferencia, es importante realizar una exploración de los datos que se va a procesar, para detectar errores en la codificación de las variables en estudio, eliminar inconsistencias, evaluar la magnitud y tipo de valores faltantes y conocer las características básicas de las variables en estudio (López, 2008, p.45).

2.3.7. Regresión logística multinomial

En "Predicción de la tendencia a la combustión espontánea de carbón mediante regresión logística multinomial" (Kursunoglu y Gogebakan, 2022: pp.2000-2009) mencionan que la RLM se utiliza para pronosticar el asentamiento categórico o la probabilidad de asociación de categoría en un factor dependiente de acuerdo con múltiples factores independientes. Los factores independientes pueden ser binarios o continuos (relación en escala o intervalo). El modelo RLM es una generalización bastante simple del modelo de regresión logística binomial y ambos modelos dependen principalmente del análisis logit o la regresión logística. En la ecuación 1-2 se puede utilizar para un factor de respuesta (Y) con dos categorías y un factor explicativo (X):

$$\pi(x) = P(Y = 1 | X = x) = 1 - P(Y = 0 | X = x) \quad (1-2)$$

Donde π = probabilidad de tener la característica deseada (odds probabilidad de éxito); P= probabilidad de modelos; X= factores explicativos; Y= factores de respuesta; x= valor característico deseado; y = si la característica deseada está presente; p = número de parámetros libres (factores explicativos); α = constante de regresión lineal; β = coeficiente de regresión lineal. El logit de esta probabilidad se da en la Ecuación 2-2:

$$\text{Logit} [\pi(x)] = \log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \alpha + \beta x \quad (2-2)$$

Donde la probabilidad se demuestra en la Ecuación 3-2:

$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta x) \quad (3-2)$$

El logaritmo de la probabilidad se llama logit y se puede calcular usando la Ecuación 4-2:

$$\text{Logit} [\pi(x)] = \log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \log [\exp(\alpha + \beta x)] = \alpha + \beta x \quad (4-2)$$

La regresión logística se puede extender a modelos con múltiples factores explicativos. k denota

el número de predictores para una respuesta binaria (Y) por x_1, x_2, \dots, x_n , y el modelo para la probabilidad logarítmica se puede calcular usando la Ecuación 5-2.

$$\text{Logit}[P(Y = 1)] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k. \quad (5-2)$$

La fórmula alternativa de $\pi(x)$ viene dada por la Ecuación 6-2:

$$\pi_j(x) = \frac{\exp(\sum_{i=1}^k \alpha + \beta_i x_i)}{1 + \exp(\sum_{i=1}^k \alpha + \beta_i x_i)}, \quad (6-2)$$

donde π_j = probabilidad multinomial de que una observación caiga en la categoría j. Para descubrir la relación entre esta probabilidad y los p factores explicativos X_A, X_B, \dots, X_p , el modelo de regresión logística múltiple se puede representar mediante la Ecuación 7-2:

$$\log \left[\frac{\pi_j(x_i)}{\pi_k(x_i)} \right] = \alpha_i + \beta_A j X_{1i} + \beta_B j X_{2i} + \dots + \beta_p j X_{pi}, \quad (7-2)$$

Donde $j=A,B,\dots,k-1$; $i=1,2,\dots,n$. Valores de π suman la unidad como lo de la Ecuación 8-2.

$$P_A = \frac{1}{1 + e^{\alpha\beta + \beta_{B1}X1 + \beta_{B2}X2 + \dots + \beta_{Bp}Xp + \dots + e^{\alpha k + \beta_{k1} + \beta_{k1}X1 + \beta_{k2}X2 + \dots + \beta_{kp}Xp}} \quad (8-2)$$

$$P_B = \frac{e^{\alpha\beta + \beta_{B1}X1 + \beta_{B2}X2 + \dots + \beta_{Bp}Xp + \dots + e^{\alpha k + \beta_{k1} + \beta_{k1}X1 + \beta_{k2}X2 + \dots + \beta_{kp}Xp}}}{1 + e^{\alpha\beta + \beta_{B1}X1 + \beta_{B2}X2 + \dots + \beta_{Bp}Xp + \dots + e^{\alpha k + \beta_{k1} + \beta_{k1}X1 + \beta_{k2}X2 + \dots + \beta_{kp}Xp}}$$

$$\vdots$$

$$P_K = \frac{e^{\alpha\beta + \beta_{B1}X1 + \beta_{B2}X2 + \dots + \beta_{Bp}Xp + \dots + e^{\alpha k + \beta_{k1} + \beta_{k1}X1 + \beta_{k2}X2 + \dots + \beta_{kp}Xp}}}{1 + e^{\alpha\beta + \beta_{B1}X1 + \beta_{B2}X2 + \dots + \beta_{Bp}Xp + \dots + e^{\alpha k + \beta_{k1} + \beta_{k1}X1 + \beta_{k2}X2 + \dots + \beta_{kp}Xp}}$$

2.3.8. Análisis de correspondencia

El análisis de correspondencias es una técnica descriptiva para representar tablas de contingencias, es decir, elaboramos tablas de frecuencia de dos o más variables cualitativas en conjunto de elementos (Peña, 2013, p.558).

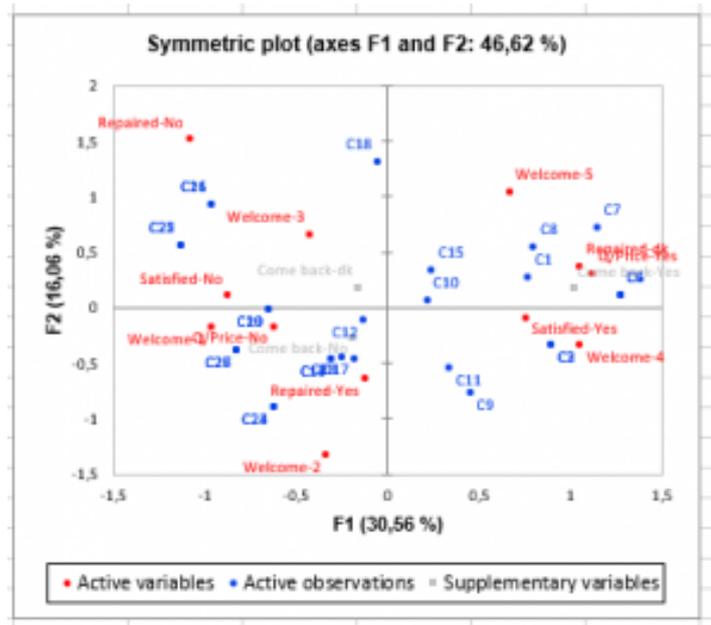


Ilustración 1–2: Análisis de correspondencias múltiples.

Fuente: (Addinsoft, 2022, p.1).

El objetivo es representar en un espacio multidimensional simplificado la relación entre las categorías de variables cualitativas.

Fórmula de la distribución estadística de numerosidad bivariante

$$N = X^t Y \tag{9-2}$$

Donde:

X es la matriz transpuesta que contiene la primera mudable estadística expresada como en la matriz de super-indicadores.

Y es la matriz que contiene la segunda mudable estadística expresada como en la matriz de super-indicadores.

Previo al análisis de correspondencia se debe aplicar el test chi-cuadrado, para determinar si las mudables estadísticas son dependientes o independientes.

Planteamiento de hipótesis:

H_0 : Las variables estadísticas categóricas cualitativas son independientes.

H_1 : Las variables categóricas cualitativas son dependientes.

Estadístico de prueba:

Estadístico de prueba chi-cuadrado

$$x^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \tag{10-2}$$

Donde:

o_{ij} es la frecuencia observada de la celda que está en la fila i , columna j .

e_{ij} es la frecuencia esperada de la celda (i,j) .

Valor crítico:

Valor crítico

$$k = x_{1-\alpha}(x^2), x^2 \sim x^2((I-1)(J-1)) \quad (11-2)$$

Región crítica:

$$(k, +\infty) \quad (12-2)$$

Toma de decisión:

$$\text{Si } x^2 \in (k, \infty) \text{ se rechaza } H_0 \quad (13-2)$$

Si no se rechaza la hipótesis nula se procede a calcular las coordenadas principales, además de los mapas perceptuales.

Coordenadas principales de X

$$A = D_r^{-1/2} U D_\lambda \quad (14-2)$$

Coordenadas principales de Y

$$B = D_c^{-1/2} V D_\lambda \quad (15-2)$$

Donde:

D_r es la matriz de las frecuencias relativas marginales de las modalidades de la variable X .

D_c es la matriz de las frecuencias relativas marginales de las modalidades de la variable Y .

D_λ es la matriz de valores singulares (Cuadras, 2018, p.305).

CAPÍTULO III

3. MARCO METODOLÓGICO

3.1. Enfoque de investigación

Por el método de investigación será mixta ya que las variables que se va a manipular son categóricas y numéricas, según el objetivo de estudio será aplicada ya que la presente investigación se centra en conocer los factores asociados a siniestros de tránsito en el Ecuador, mediante la comparación de modelo de regresión y análisis de correspondencias en el periodo enero 2018 – mayo 2021.

3.2. Nivel de Investigación

Según el nivel de profundización en el objeto de estudio será descriptiva ya que se desea conocer los factores asociados a siniestros de tránsito en el Ecuador, mediante la comparación de modelo de regresión y análisis de correspondencias en el periodo enero 2018 – mayo 2021.

3.3. Diseño de investigación

3.3.1. *Según la manipulación o no de la variable independiente*

La presente investigación utilizará un método de investigación mixto y según la manipulación de las variables es un diseño no experimental porque ya se cuenta con una base de datos descargada de la página web oficial de la Agencia Nacional de Transito ANT, es decir la información es secundaria.

3.3.2. *Según las intervenciones en el trabajo de campo*

Según el periodo temporal es transversal ya que como se mencionó anteriormente la información procede de fuentes secundarias y no hubo un seguimiento de población en el tiempo de estudio.

3.4. Tipo de estudio

El tipo de estudio es documental puesto que la base de datos, e información de la población ya la tenemos, en este proyecto de investigación se realizará un estudio de tipo documental.

3.5. Población y Planificación, selección y cálculo del tamaño de la muestra

La población de Ecuador que se vio involucrada en siniestros de tránsito en el periodo enero 2018 – mayo 2021, es parte de esta investigación, con un total de 169.987 individuos involucrados en siniestros de tránsito.

Fórmula para el cálculo de Muestreo Aleatorio Simple con tamaño de muestra finita.

$$n = \frac{Z^2 * p * q * N}{N * e^2 + Z^2 * p * q} \quad (1-3)$$

Donde:

n= tamaño muestral

N= tamaño poblacional $N = 169987$

Z= nivel de confianza $\alpha = 0,05$

p= probabilidad de éxito

q= probabilidad de fracaso

e=error de estimación máxima aceptado $e = 0,05$

$$n = \frac{(1,96)^2 * 0,5 * 0,5 * 169987}{169987 * 0,05^2 + (1,96)^2 * 0,5 * 0,5} = 383,294 \quad (2-3)$$

n= 383 individuos

A partir de una población que contiene un total de 169987 individuos en estudio, con una probabilidad de éxito y fracaso del 50%, un nivel de confianza de 95% y un error de estimación máximo del 5%, se considerará una muestra de 383 individuos involucrados en siniestros de tránsito en el periodo enero 2018 – mayo 2021.

3.5.1. Muestreo estratificado proporcionado

Para el análisis de correspondencias fue necesario realizar un muestreo estratificado proporcionado, como se detalla a continuación.

$$k = \frac{n}{N} \quad (3-3)$$

Donde:

k= constante positiva

N= 169987

$$k = \frac{383}{169987} \quad (4-3)$$

$$k = 0,00225$$

Tabla 1–3: Muestreo estratificado proporcionado.

Provincias	N_h	$n_h = kN_h$
1.-Guayas: 3 645 483	50382	113,5
2.-Pichincha: 2 576 287	53147	119,7
3.-Manabí: 1 369 780	8316	18,7
4.-Los Ríos: 778 115	6236	14,1
5.-Azuay: 712 127	8042	18,1
6.-El Oro: 600 659	4265	9,6
7.-Esmeraldas: 534 092	1549	3,5
8.-Tunguragua: 504 583	8217	18,5
9.-Chimborazo: 458 581	4011	9,0
10.-Loja: 448 996	3566	8,0
11.-Cotopaxi: 409 205	1808	4,1
12.-Imbabura: 398 244	5566	12,5
13.-Santo Domingo de las Tsáchilas: 368 013	5555	12,5
14.-Santa Elena: 308 693	3271	7,4
15.-Cañar: 225 184	960	2,2
16.-Bolivar: 183 641	873	2,0
17.-Sucumbios: 176 472	473	1,1
18.-Carchi: 164 524	753	1,7
19.-Morona Santiago: 147 940	972	2,2
20.-Orellana: 136 396	490	1,1
21.-Napo: 103 697	556	1,3
22.-Zamora Chinchipe: 91 376	533	1,2
23.-Pastaza: 83 933	405	0,9
24.-Galápagos: 25 124	41	0,1
Total:	169987	383

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

A continuación, se realizó el muestreo estratificando donde los estratos son las 24 provincias del Ecuador que se muestran en la tabla 1–3, donde se encontró que se debe tomar muestras según

dichas provincias como se observa en la tabla 1–3 dando un total de 383 individuos involucrados en siniestros de tránsito en el periodo enero 2018 – mayo 2021.

3.6. Métodos, técnicas e instrumentos de investigación

En primer lugar, se realizará un análisis exploratorio de datos de los siniestros de tránsito en el Ecuador en el periodo enero 2018 – mayo 2021, posteriormente se hará un análisis. A continuación, se efectuará una comparación de modelo de regresión y análisis de correspondencias.

3.7. Operacionalización de las variables

Tabla 2–3: Operacionalización de las variables.

Nombre de la variable	Descripción	Tipo de variable	Escala de medición	Categoría o intervalo
Día	Un día es aproximadamente el periodo durante el cual la Tierra completa una rotación alrededor de su eje, lo que lleva unas 24 horas.	Cualitativa	Nominal	Lunes Martes Miércoles Jueves Viernes Sábado Domingo
Feriado	Día no laborable por ser fiesta oficial o eclesiástica.	Cualitativa	Nominal dicotómica	Si No

Continúa en la siguiente página.

Tabla 2-3 : *Continuación de la página anterior.*

Nombre de la variable	Descripción	Tipo de variable	Escala de medición	Categoría o intervalo
Clase	Lesiones fatales y no fatales incurridas como resultado de un siniestro vial.	Cualitativa	Nominal	Arrollamiento Atropello Caída de pasajero Choque frontal Choque lateral Choque posterior Colisión Estrellamiento Otros Perdida de carril Perdida de pista Rozamiento Volcamiento

Continúa en la siguiente página.

Tabla 2-3 : *Continuación de la página anterior.*

Nombre de la variable	Descripción	Tipo de variable	Escala de medición	Categoría o intervalo
Mes	Un mes es cada uno de los doce períodos de tiempo, de entre 28 y 31 días, en que se divide el año.	Cualitativa	Nominal	Enero Febrero Marzo Abril Mayo Junio Julio Agosto Septiembre Octubre Noviembre Diciembre

Continúa en la siguiente página.

Tabla 2-3 : *Continuación de la página anterior.*

Nombre de la variable	Descripción	Tipo de variable	Escala de medición	Categoría o intervalo
Provincia	División político-administrativa conformada por un conjunto de cantones.	Cualitativa	Nominal	Azuay, Bolívar, Cañar, Carchi, Chimborazo, Cotopaxi, El Oro, Esmeraldas, Galápagos, Guayas, Imbabura, Loja, Los Ríos, Manabí, Morona Santiago, Napo, Orellana, Pastaza, Pichincha, Santa Elena, Santo Domingo de los Tsáchilas, Sucumbíos, Tungurahua, Zamora Chinchipec.

Continúa en la siguiente página.

Tabla 2-3 : *Continuación de la página anterior.*

Nombre de la variable	Descripción	Tipo de variable	Escala de medición	Categoría o intervalo
Número de fallecido	Número de individuos que fallecieron tras sufrir un accidente de tránsito.	Cuantitativa	Razón	$[0, +\infty[$
Número de lesionado	Número de individuos que resultaron lesionados tras sufrir un accidente de tránsito.	Cuantitativa	Razón	$[0, +\infty[$

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

CAPÍTULO IV

4. MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

CODIGO	ENTE_CONT	PROVINCIA	DIA	FECHA	HORA	PERIODO	FERIADO	CAUSA	CLASE	ZONA	X	Y	DIRECCION	CANTON	PARROQUIA	NUM_FALLE	NUM_LESION	MES	
1	000101MAM	MAM	TUNGURAH	JUEVES	1/1/2015	0.0053935	0	SI	CONducIR	PÉRDIDA DE URBANA	-78.612643	-1.267207	AV VICTOR I	AMBATO	HUACHI CHI	0	2	ENERO	
2	000201MAM	MAM	TUNGURAH	JUEVES	1/1/2015	0.0410648	0	SI	CONducE	B. ATROPELLO	URBANA	-78.626814	-1.245747	TRECE DE A	AMBATO	LA MATRIZ	0	1	ENERO
3	000301MAM	MAM	TUNGURAH	JUEVES	1/1/2015	0.0838657	2	SI	CONducE	B. CHOQUE LA'	URBANA	-78.621385	-1.262831	JAVIER DE A	AMBATO	CELIANO MC	0	1	ENERO
4	000401MAM	MAM	TUNGURAH	JUEVES	1/1/2015	0.0886343	2	SI	CASO FORTL	PÉRDIDA DE URBANA	-78.610359	-1.261122	AV GALO VE	AMBATO	CELIANO MC	0	1	ENERO	
5	000501MAM	MAM	TUNGURAH	JUEVES	1/1/2015	0.2258796	5	SI	TRANSITA	B. ATROPELLO	URBANA	-78.628752	-1.259393	AV LOS SHYI	AMBATO	HUACHI CHI	0	1	ENERO
6	000601MAM	MAM	TUNGURAH	JUEVES	1/1/2015	0.3779051	9	SI	MALAS CON	ESTRELLAMI	URBANA	-78.597757	-1.248922	BARRIO LA I	AMBATO	PISHILATA	0	2	ENERO
7	000701MAM	MAM	TUNGURAH	VIERNES	1/2/2015	0.6230556	14	SI	NO GUARDA	CHOQUE LA'	URBANA	-78.607122	-1.288505	AV BOLIVAR	AMBATO	CELIANO MC	0	0	ENERO
8	000801MAM	MAM	TUNGURAH	VIERNES	1/2/2015	0.5622454	13	SI	NO RESPETA	CHOQUE LA'	URBANA	-78.629346	-1.271128	AV ATAHUA	AMBATO	HUACHI CHI	0	0	ENERO
9	000901MAM	MAM	TUNGURAH	VIERNES	1/2/2015	0.7100926	17	SI	CONducE	B. CHOQUE PO	URBANA	-78.623434	-1.233366	UNIDAD NA	AMBATO	LA MERCED	0	0	ENERO
10	001001MAM	MAM	TUNGURAH	VIERNES	1/2/2015	0.7980324	19	SI	CONducIR	I VOLCAMIEN	URBANA	-78.636917	-1.307167	PANAMERIC	AMBATO	HUACHI GRI	0	0	ENERO
11	001101MAM	MAM	TUNGURAH	VIERNES	1/2/2015	0.8693981	20	SI	CONducE	B. CHOQUE PO	URBANA	-78.616928	-1.247604	CAMINO EL	AMBATO	CELIANO MC	0	0	ENERO
12	001201MAM	MAM	TUNGURAH	SABADO	1/3/2015	0.4396528	10	SI	NO RESPETA	CHOQUE LA'	URBANA	-78.601558	-1.207081	AV ATAHUA	AMBATO	HUACHI CHI	0	0	ENERO
13	001301MAM	MAM	TUNGURAH	SABADO	1/3/2015	0.4640625	11	SI	NO RESPETA	CHOQUE LA'	URBANA	-78.606982	-1.287063	AV BOLIVAR	AMBATO	TOTORAS	0	1	ENERO
14	001401MAM	MAM	TUNGURAH	SABADO	1/3/2015	0.578912	13	SI	CONducE	B. ESTRELLAMI	URBANA	-78.601111	-1.234695	AV INDOAM	AMBATO	IZAMBA	0	0	ENERO
15	001501MAM	MAM	TUNGURAH	SABADO	1/3/2015	0.8492014	20	SI	CONducE	B. VOLCAMIEN	URBANA	-78.712388	-1.273865	SAN MIGUE	AMBATO	PASA	1	5	ENERO
16	001601MAM	MAM	TUNGURAH	DOMINGO	1/4/2015	0.197037	4	SI	CONducIR	I PÉRDIDA DE URBANA	-78.616712	-1.259645	AV TRES CAI	AMBATO	CELIANO MC	0	0	ENERO	
17	001701MAM	MAM	TUNGURAH	DOMINGO	1/4/2015	0.3360417	8	SI	CONDICION	VOLCAMIEN	RURAL	-78.683417	-1.233305	CASERIO EL	AMBATO	QUISAPINCH	0	0	ENERO
18	001801MAM	MAM	TUNGURAH	DOMINGO	1/4/2015	0.3884028	9	SI	CONducE	B. CHOQUE FR	URBANA	-78.60514	-1.236672	AV INDOAM	AMBATO	IZAMBA	0	0	ENERO
19	001901MAM	MAM	TUNGURAH	DOMINGO	1/4/2015	0.490625	11	SI	NO RESPETA	ROZAMIENT	URBANA	-78.639827	-1.246662	AV QUISQU	AMBATO	HUACHI CHI	0	0	ENERO
20	002001MAM	MAM	TUNGURAH	LUNES	1/5/2015	0.7279051	17	NO	CONducE	B. ATROPELLO	RURAL	-78.611093	-1.305848	HUACHI TOT	AMBATO	TOTORAS	0	1	ENERO
21	002101MAM	MAM	TUNGURAH	LUNES	1/5/2015	0.7743056	18	NO	NO TRANSIT	ATROPELLO	URBANA	-78.61715	-1.240787	MONTES DE	AMBATO	LA MERCED	0	1	ENERO
22	002201MAM	MAM	TUNGURAH	MARTES	1/6/2015	0.407963	9	NO	NO CEDER	E CHOQUE LA'	URBANA	-78.6228	-1.266447	VICTOR HUK	AMBATO	CELIANO MC	0	1	ENERO
23	002301MAM	MAM	TUNGURAH	MIERCOLES	1/7/2015	0.3396296	8	NO	REALIZAR	CJ CHOQUE LA'	URBANA	-78.636838	-1.293351	AV ATAHUA	AMBATO	HUACHI CHI	0	1	ENERO
24	002401MAM	MAM	TUNGURAH	MIERCOLES	1/7/2015	0.4065856	9	NO	NO RESPETA	CHOQUE LA'	RURAL	-78.612128	-1.229628	AV INDOAM	AMBATO	IZAMBA	0	1	ENERO

Ilustración 1–4: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Fuente: (AGENCIA NACIONAL DE TRÁNSITO, 2022c: 1A).

4.1. Análisis descriptivo

Tabla 1–4: Frecuencias absolutas de la variable Provincia.

Nº	Provincia	Frecuencia
1	AZUAY	8042
2	BOLÍVAR	873
3	CAÑAR	960
4	CARCHI	753
5	CHIMBORAZO	4011
6	COTOPAXI	1808
7	EL ORO	4265
8	ESMERALDAS	1549
9	GALÁPAGOS	41

Continúa en la siguiente página.

Tabla 1-4 : *Continuación de la página anterior.*

N°	Provincia	Frecuencia
10	GUAYAS	50382
11	IMBABURA	5566
12	LOJA	3566
13	LOS RÍOS	6236
14	MANABÍ	8316
15	MORONA SANTIAGO	972
16	NAPO	556
17	ORELLANA	490
18	PASTAZA	405
19	PICHINCHA	53147
20	SANTA ELENA	3271
21	SANTO DOMINGO DE LOS TSÁCHILAS	5555
22	SUCUMBÍOS	473
23	TUNGURAHUA	8217
24	ZAMORA CHINCHIPE	533

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

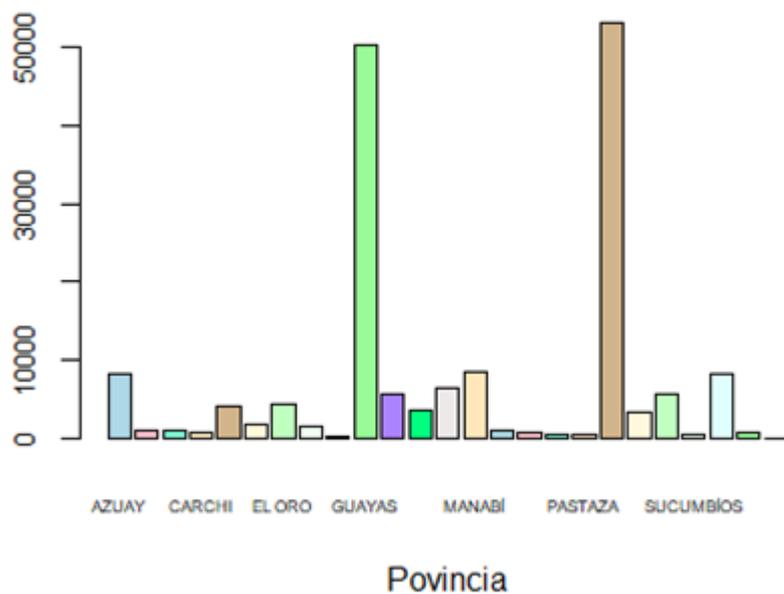


Ilustración 2-4: Diagrama de barras de las frecuencias absolutas de la variable Provincia.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 1-4 Frecuencias absolutas de la variable Provincia y en la Figura 2-4 Diagrama de barras de las frecuencias absolutas de la variable Provincia, se puede observar que las provincias del Ecuador con más presencia de siniestros de tránsito son; Pichincha, Guayas, Tungurahua seguida por Azuay, por otro lado se observa a las provincias con un menor número de accidentes viales las cuales son; Orellana, Sucumbíos, Pastaza y Galápagos con un total de 41 accidentes en el período de enero 2018 a mayo 2021.

Tabla 2-4: Frecuencias absolutas de la variable Clase.

N°	Clase	Frecuencia
1	Arrollamiento	2516
2	Atropello	25306
3	Caída de pasajero	4529
4	Choque frontal	8644
5	Choque lateral	47487

Continúa en la siguiente página.

Tabla 2-4 : Continuación de la página anterior.

N°	Clase	Frecuencia
6	Choque posterior	17546
7	Colisión	4698
8	Estrellamiento	22088
9	Otros	3261
10	Perdida de carril	3277
11	Perdida de pista	16821
12	Rozamiento	10047
13	Volcamiento	3767

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

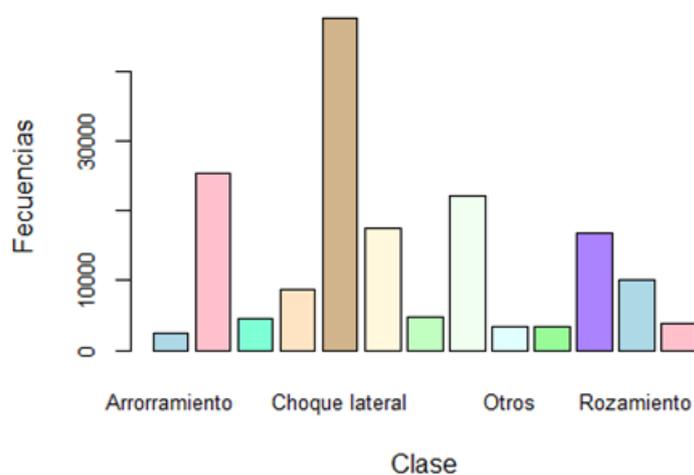


Ilustración 3-4: Diagrama de barras de las frecuencias absolutas de la variable Clase de siniestros viales.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 2-4 Frecuencias absolutas de la variable Clase y en la Figura 3-4 Diagrama de barras de las frecuencias absolutas de la variable Clase de siniestros viales, se puede observar que las clases de siniestros viales con mas presencia son choque lateral con 47487 incidentes de ese tipo, seguida por atropello con una frecuencia de 25306, por otro lado se observa que la perdida de carril y otro tipo de accidentes se registran como lo menos ocurrido en los accidentes viales, la clase de

accidente con menor ocurrencia es arrollamiento con un total de 2516 casos en el período de enero 2018 a mayo 2021.

Tabla 3-4: Frecuencias absolutas de la variable Mes.

N°	Mes	Frecuencia
1	ABRIL	13877
2	AGOSTO	13283
3	DICIEMBRE	15566
4	ENERO	15973
5	FEBRERO	15042
6	JULIO	13389
7	JUNIO	13064
8	MARZO	14836
9	MAYO	14404
10	NOVIEMBRE	13596
11	OCTUBRE	13792
12	SEPTIEMBRE	13165

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

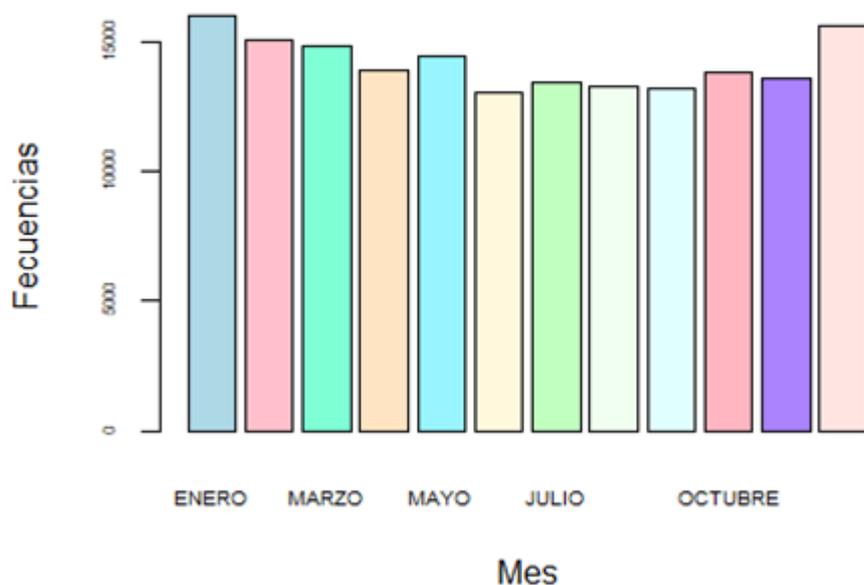


Ilustración 4-4: Diagrama de barras de las frecuencias absolutas de la variable Mes.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 3-4 Frecuencias absolutas de la variable Mes y en la Figura 4-4 Diagrama de barras de las frecuencias absolutas de la variable Mes, se observa que en los meses de; Diciembre, Enero y Febrero se registran una mayor presencia de siniestros viales con frecuencias absolutas de 15973,15973 y 15042 respectivamente, por otro lado se observa que los meses de Septiembre y Junio poseen frecuencias de 13165 y 13064 respectivamente, siendo Junio el mes con menor accidentes de tránsito registrados en Ecuador en el período de enero 2018 a mayo 2021.

Tabla 4-4: Frecuencias absolutas de la variable Feriado.

N°	Feriado	Frecuencia
1	NO	156243
2	SI	13744

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

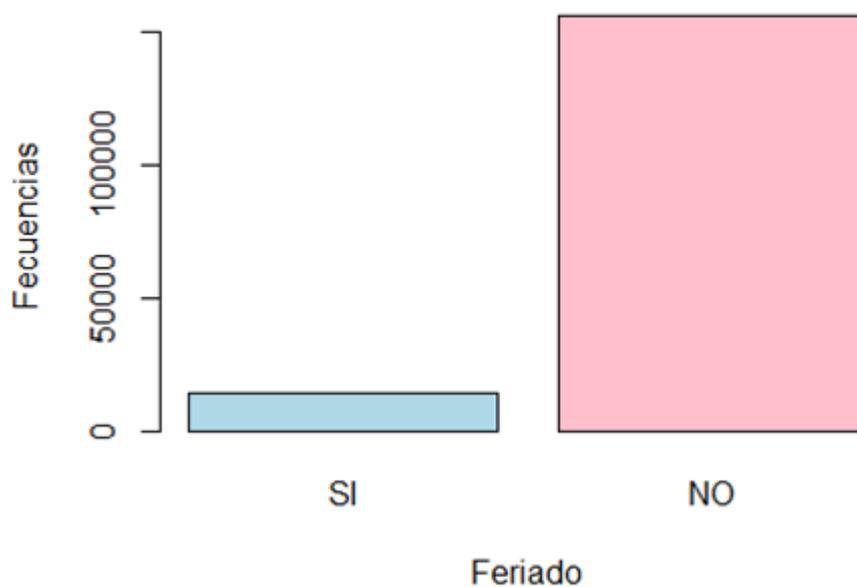


Ilustración 5-4: Diagrama de barras de las frecuencias absolutas de la variable Feriado.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 4-4 Frecuencias absolutas de la variable Feriado y en la Figura 5-4 Diagrama de barras de las frecuencias absolutas de la variable Feriado, se presenta más accidentes viales en días que no son feriados en Ecuador con un total de 156243, mientras que en días feriados se presentaron 13744 accidentes en el período de enero 2018 a mayo 2021.

Tabla 5-4: Frecuencias absolutas de la variable Días.

N°	Días	Frecuencia
1	Domingo	29237
2	Jueves	21809
3	Lunes	21481
4	Martes	20835
5	Miércoles	20595
6	Sábado	30542
7	Viernes	25488

Continúa en la siguiente página.

Tabla 5-4 : Continuación de la página anterior.

N°	Días	Frecuencia
----	------	------------

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

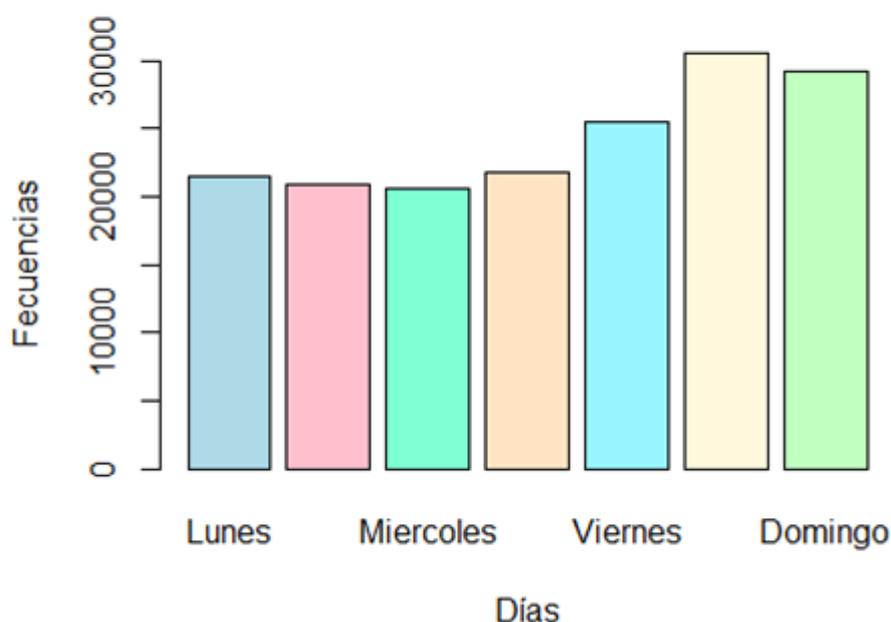


Ilustración 6-4: Diagrama de barras de las frecuencias absolutas de la variable Día.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 5-4 Frecuencias absolutas de la variable Días y en la Figura 6-4 Diagrama de barras de las frecuencias absolutas de la variable Días, se puede observar que el día domingo tiene una frecuencia absoluta de 29237, lunes con 21481, para el día martes 20835, miércoles 20835, jueves de 21809, viernes 25488 y sábado 30542, siendo el sábado como el día con más accidentes y el miércoles como el día en donde ocurren menos accidentes.

Tabla 6-4: Frecuencias absolutas del número de fallecidos.

N° Fallecidos	Frecuencia
0	158529

Continúa en la siguiente página.

Tabla 6-4 : *Continuación de la página anterior.*

Nº Fallecidos	Frecuencia
1	10439
2	807
3	116
4	44
5	21
6	10
7	2
8	3
9	1
10	1
11	4
12	5
13	2
14	1
20	1
22	1

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

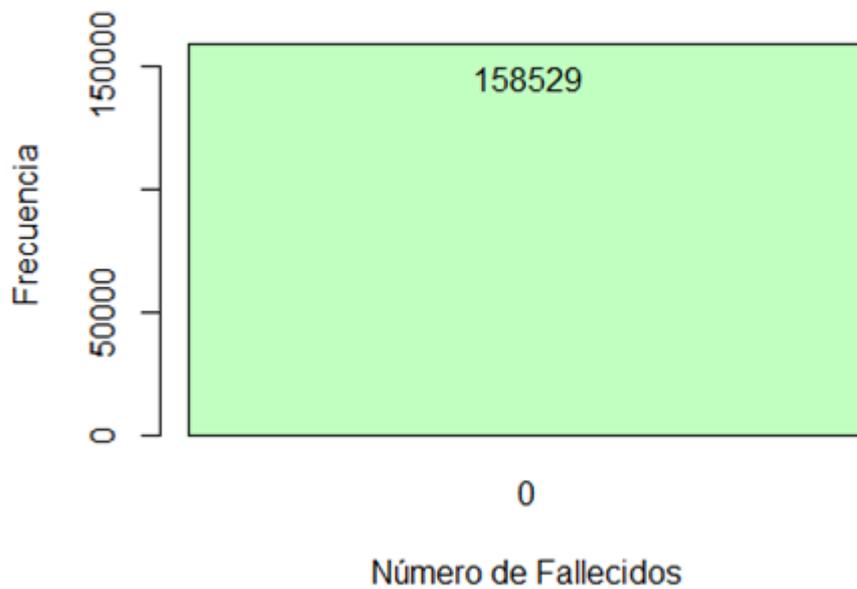


Ilustración 7-4: Diagrama de barras del número de fallecidos.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 6-4 Frecuencias absolutas del número de fallecidos y en la Figura 7-4 Diagrama de barras del número de fallecidos, se puede observar que en el período de enero 2018 a mayo 2021 existieron 0 muertos en 158529 accidentes de tránsito.



Ilustración 8–4: Diagrama de barras del número de fallecidos.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 6–4 Frecuencias absolutas del número de fallecidos y en la Figura 8–4 Diagrama de barras del número de fallecidos, se puede observar que en el período de enero 2018 a mayo 2021, existió una persona fallecida en 10439 accidentes de tránsito y dos personas fallecidas en 807 incidentes viales.

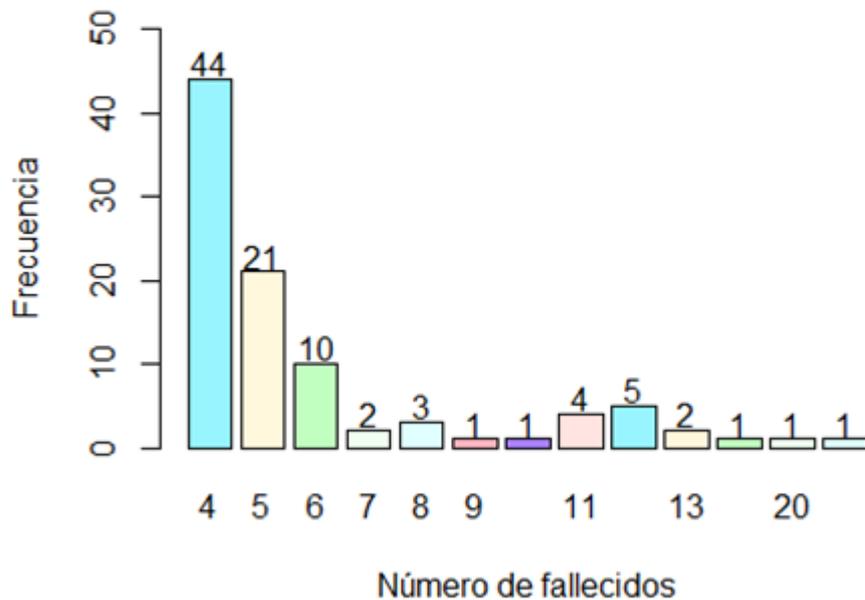


Ilustración 9-4: Diagrama de barras de las frecuencias absolutas de la variable número de fallecidos.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 6-4 Frecuencias absolutas del número de fallecidos y en la Figura 9-4 Diagrama de barras del número de fallecidos, se puede observar que en el período de enero 2018 a mayo 2021 existieron 4 muertos en 44 accidentes de tránsito, el número máximo de fallecidos fueron 22 personas en un accidente de tránsito.

Tabla 7-4: Frecuencias absolutas del número de lesionados.

N° Lesionados	Frecuencia
0	81326
1	64853
2	16548
3	4195
4	1543
5	652
6	326

Continúa en la siguiente página.

Tabla 7-4 : *Continuación de la página anterior.*

N° Lesionados	Frecuencia
7	162
8	102
9	51
10	41
11	29
12	28
13	14
14	20
15	7
16	14
17	6
18	7
19	9
20	5
21	6
22	5
23	3
24	4
25	1
26	3
27	3
28	4
29	2
30	2
31	2
32	1
33	1
34	1
35	1
36	1
43	1

Continúa en la siguiente página.

Tabla 7-4 : Continuación de la página anterior.

N° Lesionados	Frecuencia
48	1

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

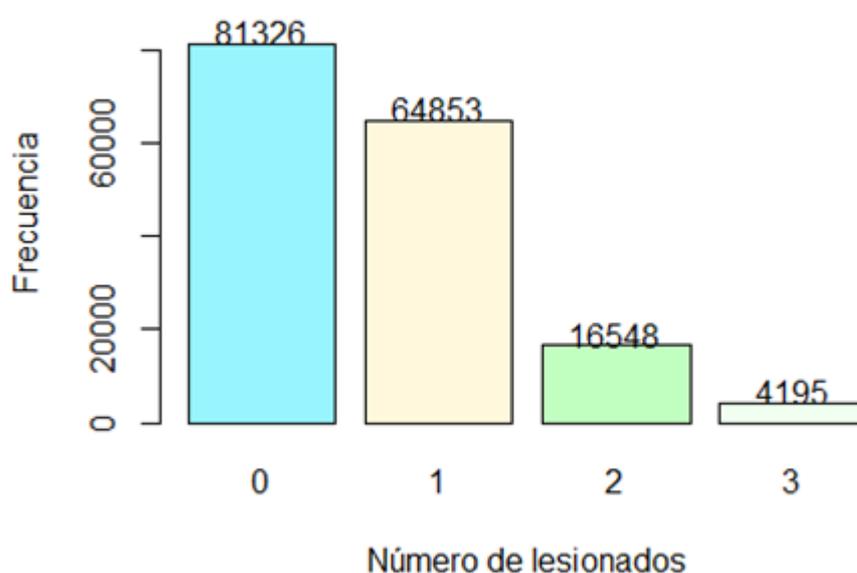


Ilustración 10-4: Diagrama de barras de las frecuencias absolutas de la variable número de lesionados.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 7-4 Frecuencias absolutas del número de lesionados y en la Figura 10-4 Diagrama de barras de las frecuencias absolutas de la variable número de lesionados, se puede observar que en el período de enero 2018 a mayo 2021 existieron 0 personas lesionadas en 81326 siniestros viales, una persona lesionada en 64853 accidentes, 16548 y 4195 con dos y tres personas lesionadas respectivamente.

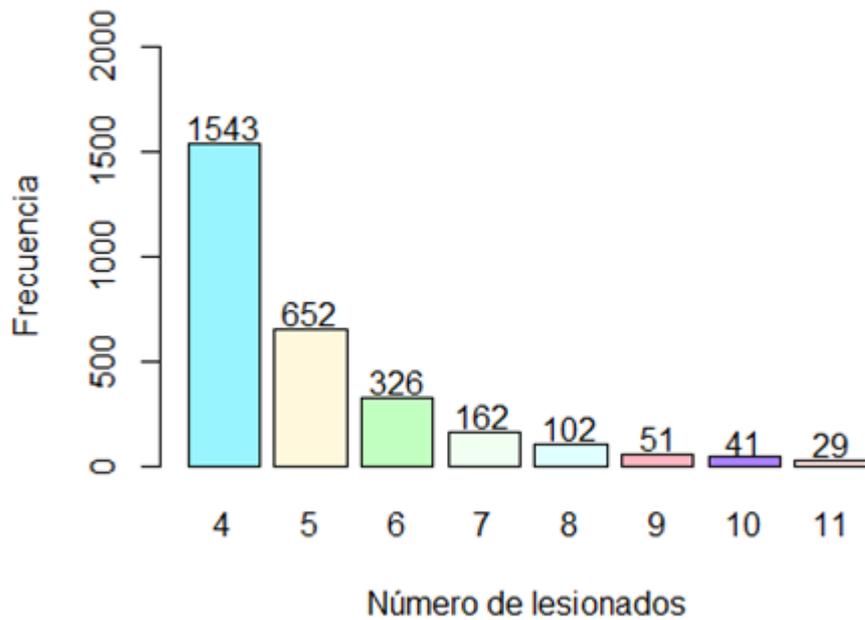


Ilustración 11–4: Diagrama de barras de las frecuencias absolutas de la variable número de lesionados.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 7–4 Frecuencias absolutas del número de lesionados y en la Figura 11–4 Diagrama de barras de las frecuencias absolutas de la variable número de lesionados, se puede observar que en el período de enero 2018 a mayo 2021 existió 4 víctimas de lesiones en 1543 accidentes, y 29 sucesos con resultado de 11 lesionados.

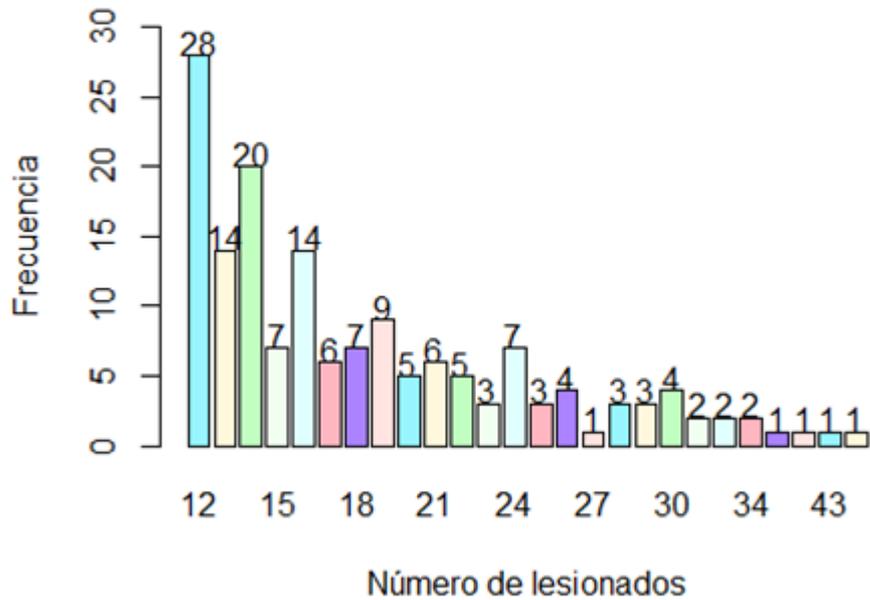


Ilustración 12–4: Diagrama de barras de las frecuencias absolutas de la variable número de lesionados.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la tabla 7–4 Frecuencias absolutas del número de lesionados y en la Figura 12–4 Diagrama de barras de las frecuencias absolutas de la variable número de lesionados, se puede observar que en el período de enero 2018 a mayo 2021 se presentaron en 7 ocasiones accidentes con 32,33,34,35,36,43 y 48 víctimas lesionadas.

4.2. Análisis de correspondencia

4.2.1. Análisis de correspondencia múltiple

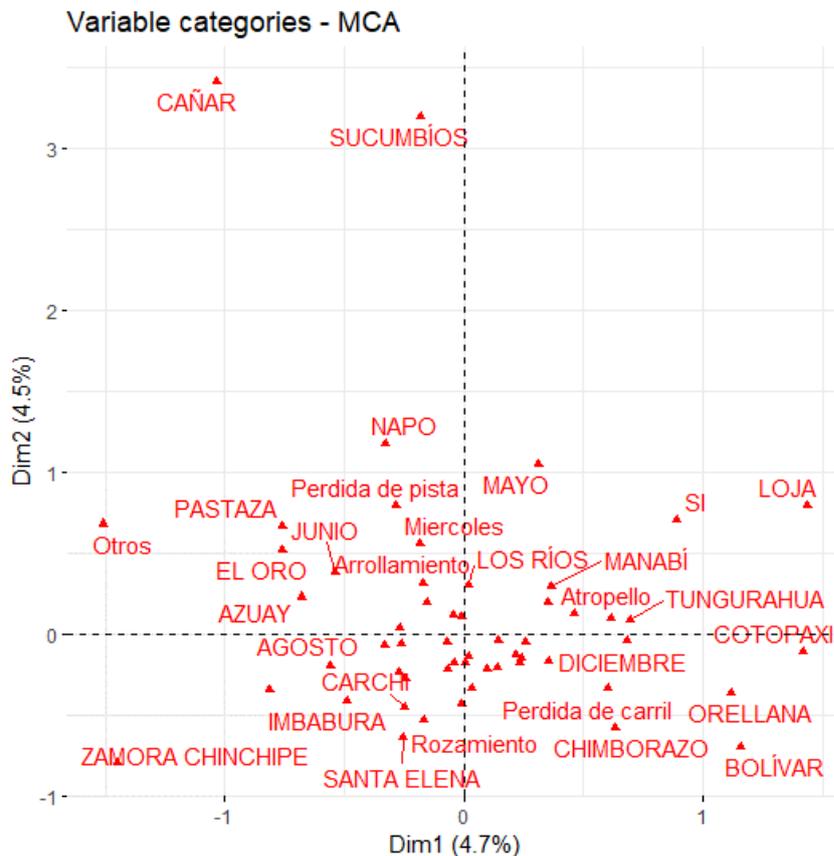


Ilustración 13-4: Análisis de correspondencia múltiple.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

Mediante la gráfica 13-4 Análisis de correspondencia múltiple, podemos observar que: en la provincia del Napo en el mes de mayo, el día en donde más común es un siniestro vial son los miércoles, en las provincias de: Pastaza, El Oro, Azuay en el mes de junio es recurrente que existan más arrollamientos, por otro lado en: Imbabura, Santa Elena, Zamora Chinchipe y el Carchi, se presentan rozamientos, mientras que en el mes de Diciembre, el siniestro vial más usual es pérdida de carril, y finalmente rozamiento en la provincia de: Chimborazo, Tungurahua, Cotopaxi y el Guayas.

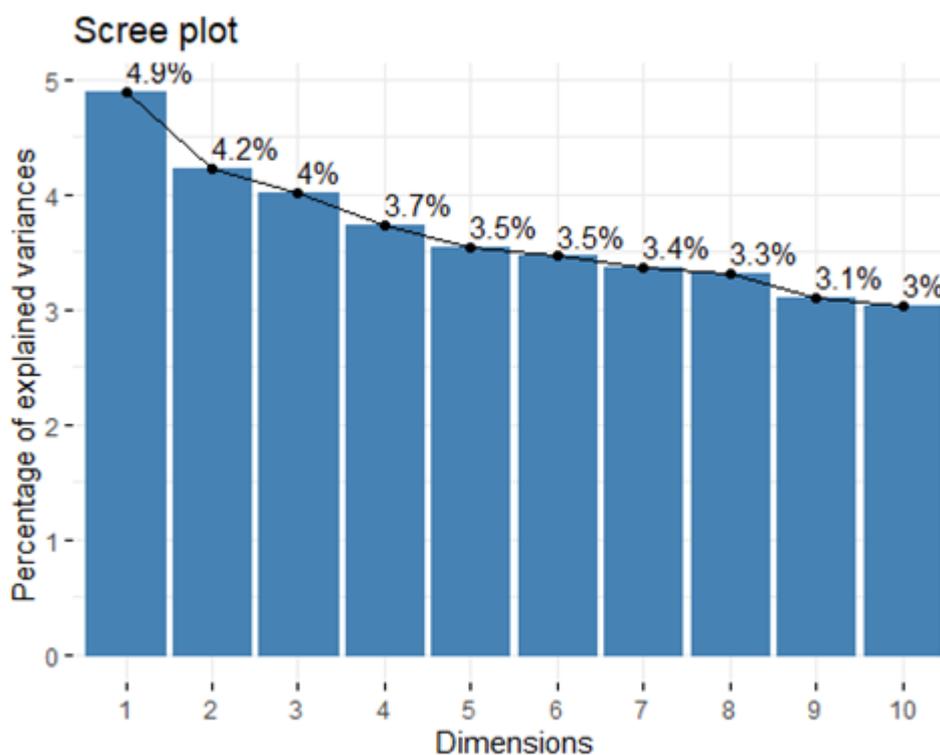


Ilustración 14–4: Porcentaje de variabilidad explicada del análisis de correspondencia múltiple.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la gráfica 14–4 del porcentaje de variabilidad explicada del análisis de correspondencia múltiple se puede observar que con una dimensión se explica el 4.9 % de la variabilidad, con dos dimensiones el 4.2%, seguido del 4% con 3 dimensiones y a partir de 4 a 10 dimensiones el porcentaje de varianza explicada fluctúa en 3 %.

4.2.2. *Análisis de correspondencia simple*

Tabla 8–4: Proporción de inercia de las variables provincia y clase.

Dimensión	Variance	% of var.	Cumulative
Dim.1	0.138	27.339	27.339
Dim.2	0.109	21.667	49.006
Dim.3	0.076	15.047	64.053
Dim.4	0.056	11.129	75.182

Continúa en la siguiente página.

Tabla 8-4 : Continuación de la página anterior.

Dimensión	Variance	% of var.	Cumulative
Dim.5	0.035	6.971	82.153
Dim.6	0.030	5.965	88.118
Dim.7	0.021	4.127	92.246
Dim.8	0.016	3.192	95.438
Dim.9	0.013	2.576	98.013
Dim.10	0.008	1.645	99.658
Dim.11	0.002	0.342	100.000

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

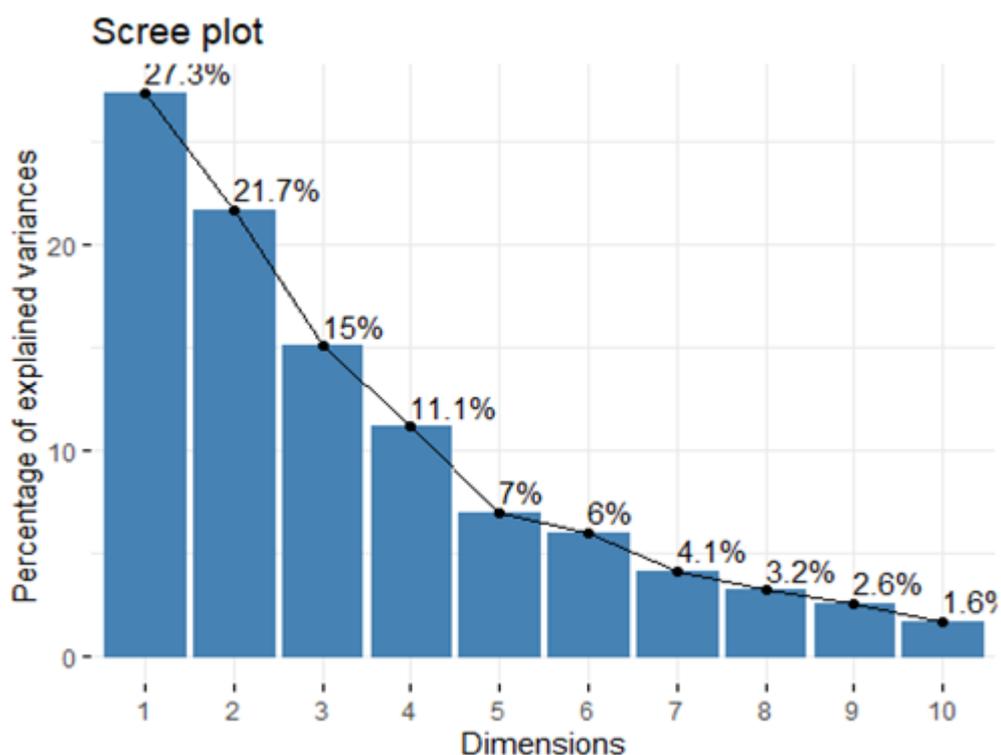


Ilustración 15-4: Porcentaje de variabilidad explicada de las variables provincia y clase.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la gráfica 15-4 del porcentaje de variabilidad explicada de las variables provincia y clase y en la tabla 8-4 proporción de inercia de las variables provincia y clase, se puede observar que en la segunda columna se encuentran los valores propios o auto valores, es decir, la correlación que existe entre las puntuaciones de las filas y las columnas. La siguiente columna pertenece a la

proporción de inercia de cada dimensión, la primera dimensión explica el 27.34% de la inercia, mientras que la segunda dimensión solamente explica el 21.67%, finalmente la ultima columna contiene la proporción de inercia acumulada de tal manera que un modelo con diez dimensiones explicaría el 99.66% de la variabilidad.

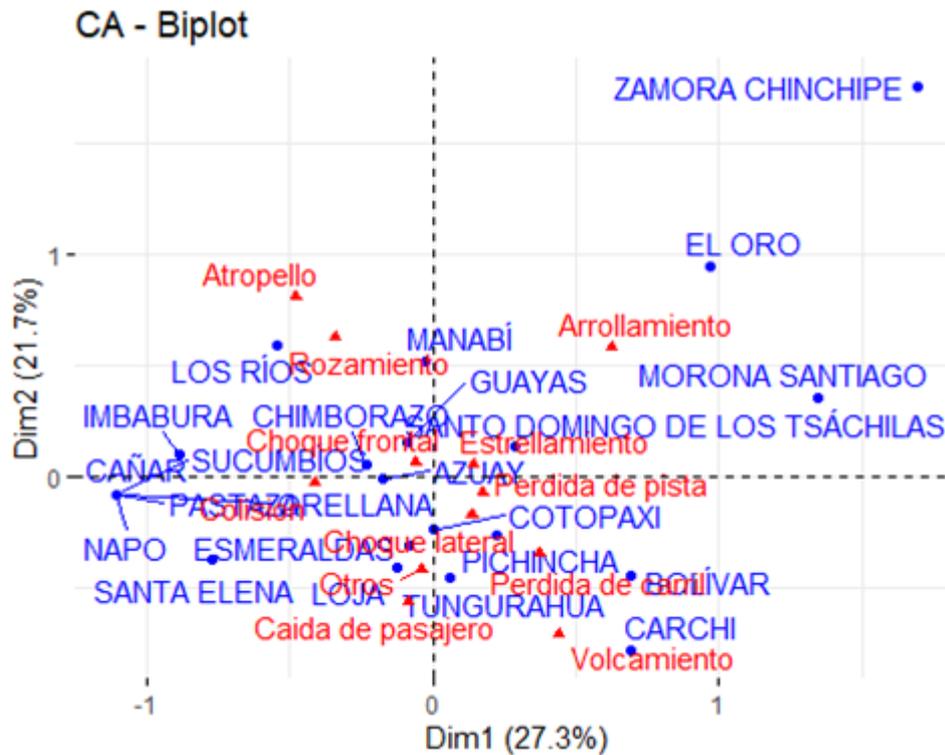


Ilustración 16-4: Análisis de correspondencia simple entre la variable Provincia y Clase.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

Mediante la gráfica 16-4 análisis de correspondencia simple entre la variable Provincia y Clase, podemos observar que: Choque frontal, Estrellamiento, Perdida de pista, ocurren con más frecuencia en las provincias de Cotopaxi, Pichincha, Azuay y Chimborazo, por otro lado, la clase del siniestro arrollamiento ocurre con más frecuencia en la provincia de El Oro, Morona Santiago, Guayas, Manabí y Santo Domingo de los Tsáchilas.

Tabla 9-4: Proporción de inercia de las variables día y clase.

Dimensión	Variance	% of var.	Cumulative
Dim.1	0.058	39.863	39.863
Dim.2	0.044	30.405	70.268

Continúa en la siguiente página.

Tabla 9–4 : Continuación de la página anterior.

Dimensión	Variance	% of var.	Cumulative
Dim.3	0.021	14.474	84.741
Dim.4	0.014	9.423	94.164
Dim.5	0.007	4.558	98.722
Dim.6	0.002	1.278	100.000

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

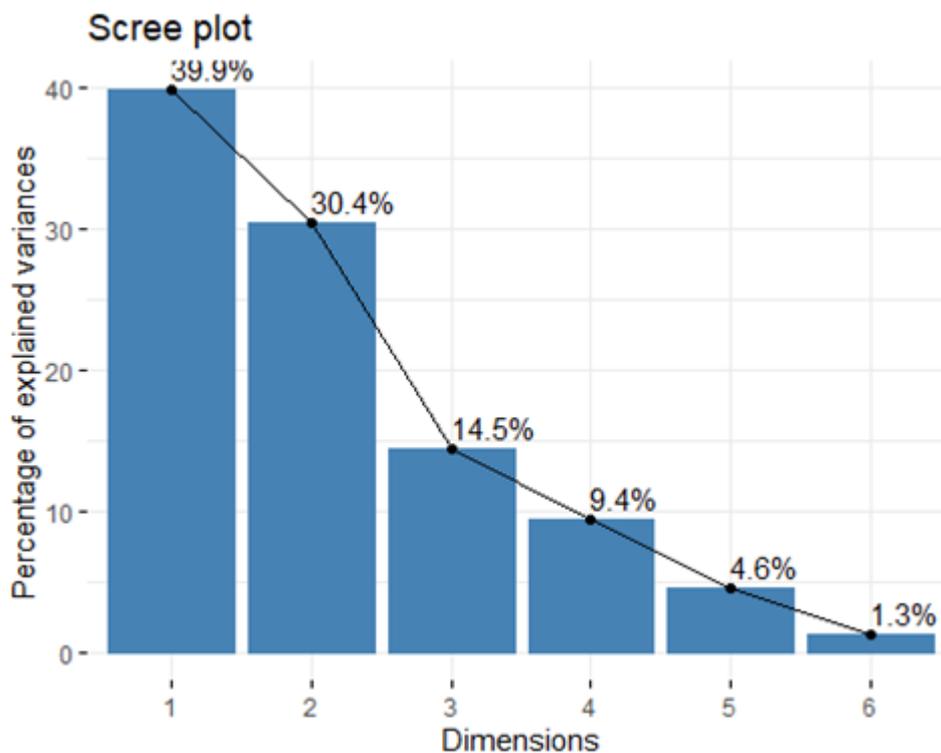


Ilustración 17–4: Porcentaje de variabilidad explicada de las variables Día y clase.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la gráfica 17–4 del porcentaje de variabilidad explicada de las variables Día y clase y en la tabla 9–4 proporción de inercia de las variables día y clase, se puede observar que en la segunda columna se encuentran los valores propios o auto valores, es decir, la correlación que existe entre las puntuaciones de las filas y las columnas. La siguiente columna pertenece a la proporción de inercia de cada dimensión, la primera dimensión explica el 39.87% de la inercia, mientras que la segunda dimensión solamente explica el 30.41%, para la tercera 14.48%, seguido de 9.42% para la cuarta dimensión, en la quinta con un 4.56%, finalmente la última columna contiene la proporción

de inercia acumulada de tal manera que un modelo con cinco dimensiones explicaría el 98.72 % de la variabilidad.

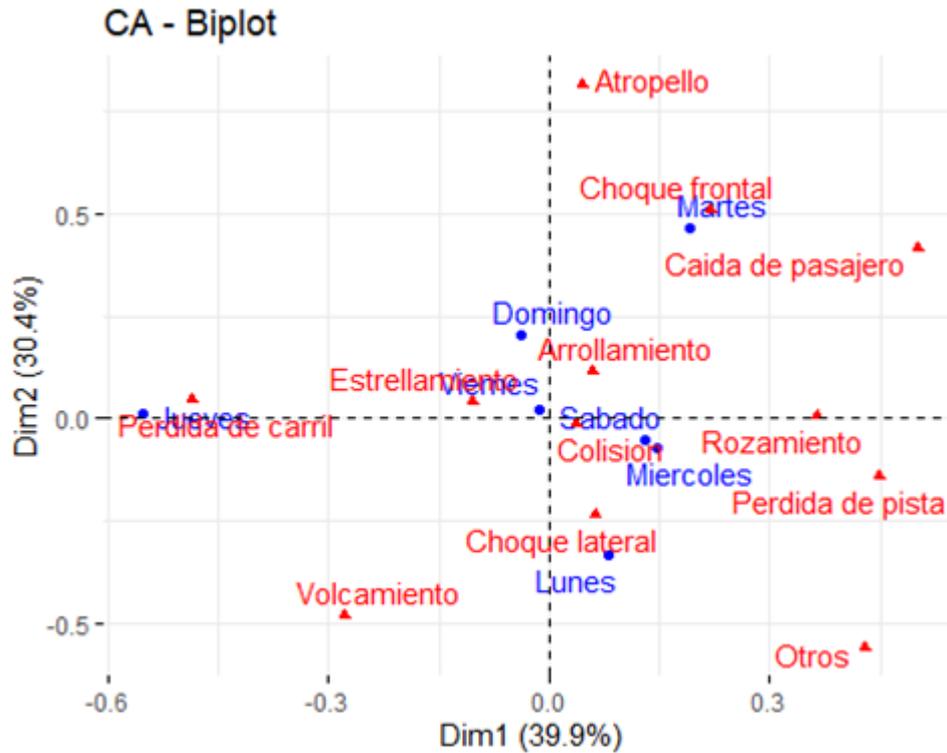


Ilustración 18–4: Análisis de correspondencia simple entre la variable Día y Clase.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

Mediante la gráfica 18–4 Análisis de correspondencia simple entre la variable Día y Clase, podemos observar que: Choque frontal, atropello, caída de pasajero, suele ocurrir los días martes, en cuanto al tipo de siniestro tenemos que: estrellamiento, arrollamiento, colisión, se presentan con frecuencia los fines de semana es decir viernes, sábado y domingo, por otro lado los lunes se presenta un siniestro de tipo choque lateral, y los días jueves de perdida de carril.

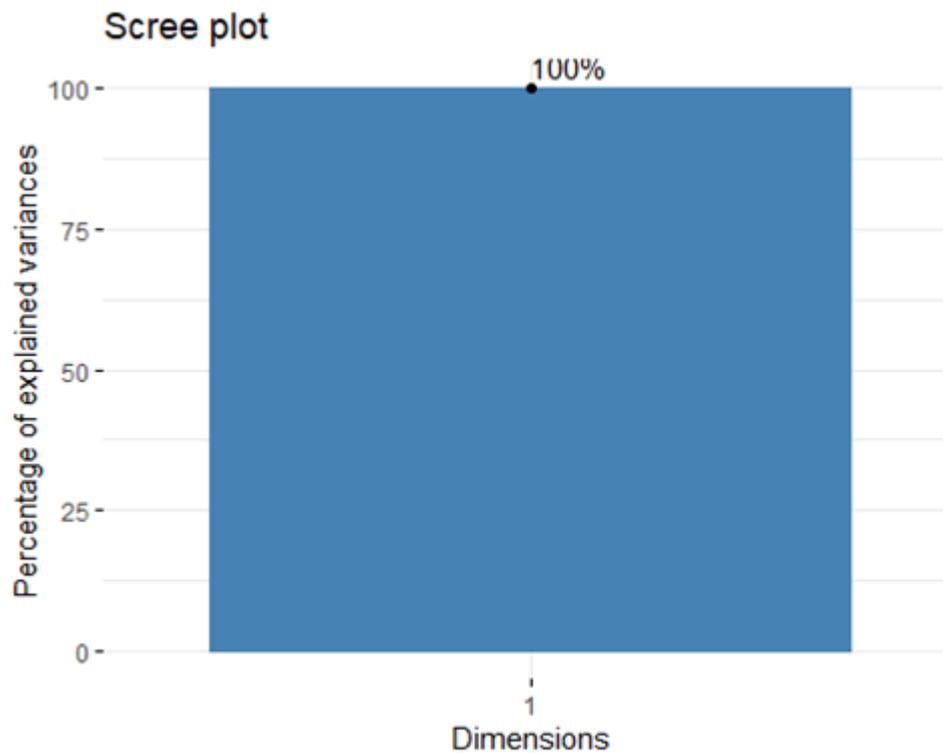


Ilustración 19–4: Porcentaje de variabilidad explicada de las variables clase y Feriado.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la gráfica 19–4 porcentaje de variabilidad explicada de las variables clase y Feriado, se observa que con una sola dimensión se explica toda la varianza, es decir para este caso puntual se implementara un diagrama de barras para su correcto análisis, que se describe a continuación:

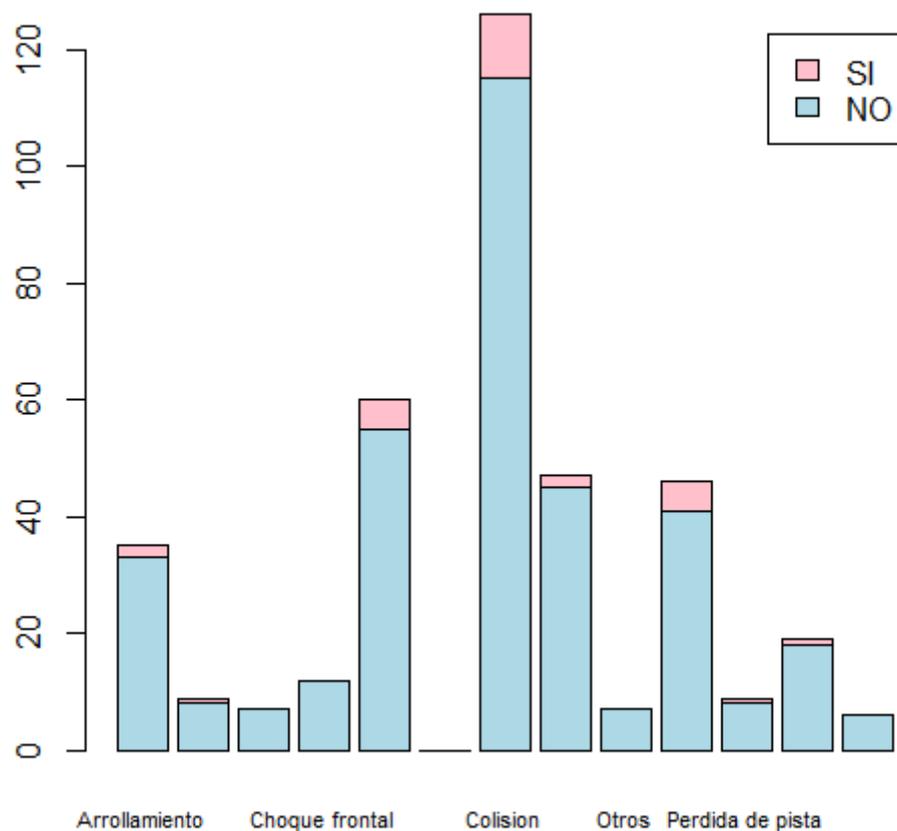


Ilustración 20-4: Diagrama de barras de las frecuencias de las variables Clase y Feriado.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

Mediante la gráfica 20-4 diagrama de barras de las frecuencias de las variables Clase y Feriado se observa que todas las clases de siniestros viales se presentan en días que no son feriados en Ecuador, esto en el período enero 2018 a mayo 2021.

Tabla 10-4: Proporción de inercia de las variables mes y clase.

Dimensión	Variance	% of var.	Cumulative
Dim.1	0.073	25.610	25.610
Dim.2	0.055	19.550	45.160
Dim.3	0.045	15.991	61.150

Continúa en la siguiente página.

Tabla 10–4 : Continuación de la página anterior.

Dimensión	Variance	% of var.	Cumulative
Dim.4	0.034	12.091	73.242
Dim.5	0.027	9.524	82.766
Dim.6	0.018	6.506	89.272
Dim.7	0.015	5.164	94.436
Dim.8	0.009	3.150	97.586
Dim.9	0.006	2.050	99.637
Dim.10	0.001	0.362	99.999
Dim.11	0.000	0.001	100.000

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

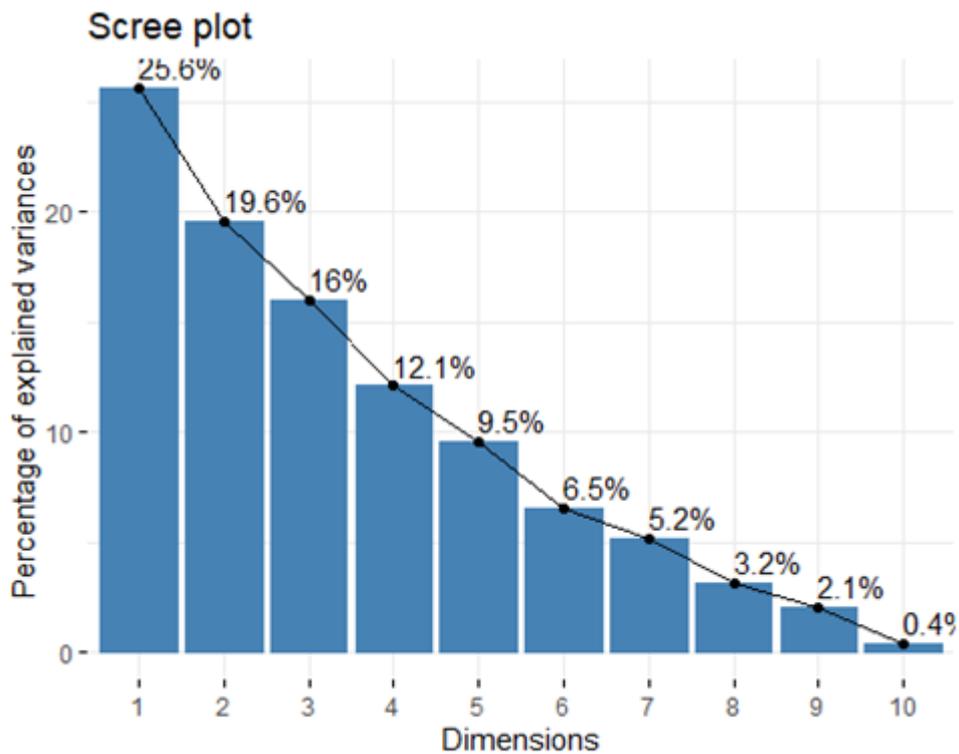


Ilustración 21–4: Porcentaje de variabilidad explicada de las variables mes y clase.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En la gráfica 21–4 del porcentaje de variabilidad explicada de las variables mes y clase. y en la tabla 10–4 Proporción de inercia de las variables mes y clase, se puede observar que en la segunda columna se encuentran los valores propios o auto valores, es decir, la correlación que existe entre las

puntuaciones de las filas y las columnas. La siguiente columna pertenece a la proporción de inercia de cada dimensión, la primera dimensión explica el 25.61 % de la inercia, mientras que para la segunda dimensión el 19.55 %, para la tercera 15.99 %, seguido de 12.10 % para la cuarta dimensión, en la quinta con un 9.52 %, para la sexta dimensión 6.51 %, seguido del 5.16 % en la séptima dimensión, en la octava, novena y décima dimensión 3.15 %, 2.05 % y 0.36 % respectivamente, finalmente la última columna contiene la proporción de inercia acumulada, de tal manera que un modelo con diez dimensiones explicaría el 99.99 %.

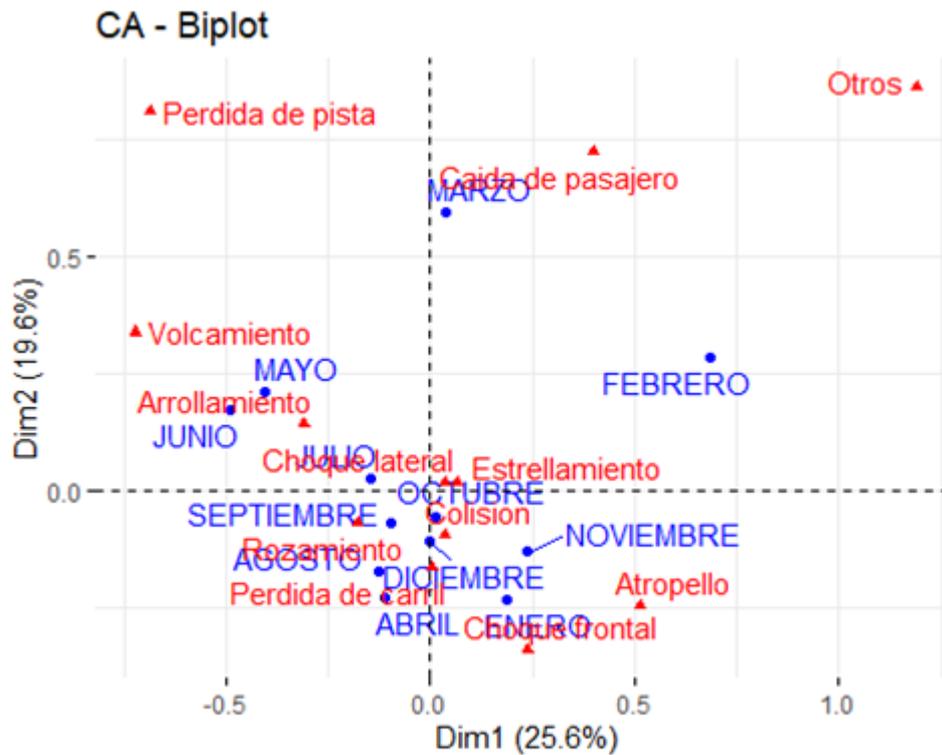


Ilustración 22–4: Análisis de correspondencia simple entre la variable Mes y Clase.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

Mediante la gráfica 22–4 Análisis de correspondencia simple entre la variable Mes y Clase, podemos observar que: noviembre, agosto y diciembre los siniestros presentados fueron pérdida de carril, colisión y rozamiento, mientras que en los meses de mayo y junio fue de tipo volcamiento y arrollamiento los cuales fueron los más comunes, y finalmente en el mes de febrero fue el mes en donde no se presentaron siniestros viales .

4.2.3. Regresión Logística Multinomial

	DIA_Domingo	DIA_Jueves	DIA_Lunes	DIA_Martes	DIA_Miercoles	DIA_Sabado	DIA_Viernes	FERIADO_NO	
1	0	0	1	0	0	0	0	0	
2	0	0	1	0	0	0	0	0	
3	0	0	1	0	0	0	0	0	
4	0	0	1	0	0	0	0	0	
5	0	0	1	0	0	0	0	0	
6	0	0	1	0	0	0	0	0	
	FERIADO_SI	MES_ABRIL	MES_AGOСТО	MES_DICIEMBRE	MES_ENERO	MES_FEBRERO	MES_JULIO	MES_JUNIO	MES_MARZO
1	1	0	0	0	1	0	0	0	0
2	1	0	0	0	1	0	0	0	0
3	1	0	0	0	1	0	0	0	0
4	1	0	0	0	1	0	0	0	0
5	1	0	0	0	1	0	0	0	0
6	1	0	0	0	1	0	0	0	0
	MES_MAYO	MES_NOVIEMBRE	MES_OCTUBRE	MES_SEPTIEMBRE	PROVINCIA_LOS RIOS	PROVINCIA_MANABI			
1	0	0	0	0	0	0			
2	0	0	0	0	0	0			
3	0	0	0	0	0	0			
4	0	0	0	0	0	0			
5	0	0	0	0	0	0			
6	0	0	0	0	0	0			
	PROVINCIA_GALÁPAGOS	PROVINCIA_SANTO DOMINGO DE LOS TSÁCHILAS	PROVINCIA_AZUAY	PROVINCIA_CARCHI					
1	0	0	0	0					
2	0	0	0	0					
3	0	0	0	0					
4	0	0	0	0					
5	0	0	0	0					
6	0	0	0	0					
	PROVINCIA_CHIMBORAZO	PROVINCIA_BOLIVAR	PROVINCIA_SUCUMBIOS	PROVINCIA_CAÑAR	PROVINCIA_COTOPAXI				
1	0	0	0	0	0				
2	0	0	0	0	0				
3	0	0	0	0	0				
4	0	0	0	0	0				
5	0	0	0	0	0				
6	0	0	0	0	0				

Ilustración 23–4: Creación de variables dummy o variables ficticias.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

El modelo de regresión logística multinomial con las variables, clase, feriado, mes, día y provincia se aplicó con la finalidad de predecir e identificar las variables que influyen que se produzcan siniestros de tránsito. Para la estimación del modelo se procedió a utilizar una muestra de train y una de test, después de seleccionar las variables más significativas para el modelo de regresión logística multinomial. Lo primero que fue hacer es crear variables dummy esto se lo puede observar en el gráfico 23–4.

	(Intercept)	DIA_Domingo	DIA_Jueves	DIA_Lunes	DIA_Martes	DIA_Miercoles	DIA_Sabado
Atropello	0.08747258	0.04874570	0.04802871	0.05883090	0.05187262	0.05559410	0.05190154
Caída de pasajero	0.13471534	0.04437080	0.03581148	0.03887728	0.04234861	0.03403963	0.03453567
Choque	0.02758447	0.03014795	0.02358713	0.02920445	0.03273552	0.02454339	0.02491468
Colisión	0.02622359	0.02749976	0.02169611	0.02745426	0.03076898	0.02299859	0.02302139
Estrellamiento	0.06011692	0.03581295	0.02803390	0.03333495	0.03816294	0.02844460	0.02873722
Pérdida de pista	0.02856462	0.03062296	0.02389663	0.03131271	0.03480370	0.02665052	0.02567131
Volcamiento	0.09938502	0.06159487	0.04538251	0.05266655	0.06292108	0.04676915	0.04628578
Atropello	DIA_Viernes	FERIADO_NO	FERIADO_SI	MES_ABRIL	MES_AGOСТО	MES_DICIEMBRE	MES_ENERO
	0.10789804	0.05081435	0.05925835	0.07271087	0.07205318	0.06844188	0.05759863
Caída de pasajero	0.03880808	0.06995307	0.07438598	0.04485407	0.04516788	0.04253691	0.04289068
Choque	0.02872200	0.02019470	0.02788171	0.03632266	0.03748776	0.03372560	0.03515845
Colisión	0.02654091	0.01875737	0.02556063	0.03398620	0.03452163	0.03106765	0.03263994
Estrellamiento	0.03357056	0.03435461	0.04059776	0.04138593	0.04230988	0.03881456	0.03942047
Pérdida de pista	0.03053369	0.02082473	0.02857654	0.03830356	0.03896154	0.03527627	0.03623302
Volcamiento	0.05788302	0.05649544	0.06615626	0.06868257	0.07159257	0.06640490	0.06210916
Atropello	MES_FEBRERO	MES_JULIO	MES_JUNIO	MES_MARZO	MES_MAYO	MES_NOVIEMBRE	MES_OCTUBRE
	0.06585970	0.06833170	0.06717129	0.07056320	0.06990931	0.07282244	0.07627011
Caída de pasajero	0.04364297	0.04520615	0.04637320	0.04541142	0.04441089	0.04409166	0.04375912
Choque	0.03563781	0.03764231	0.03894888	0.03578424	0.03645431	0.03571872	0.03451558
Colisión	0.03255603	0.03458154	0.03596629	0.03483379	0.03422214	0.03257918	0.03229660
Estrellamiento	0.03966478	0.04329087	0.04409761	0.04238703	0.04035369	0.03973010	0.04016244
Pérdida de pista	0.03652834	0.03944253	0.04083425	0.03840006	0.03763096	0.03745929	0.03710400
Volcamiento	0.06372547	0.06845953	0.06929286	0.06636167	0.06423289	0.07019290	0.06920090
Atropello	MES_SEPTIEMBRE	PROVINCIA_LOS_RIOS	PROVINCIA_MANABI	PROVINCIA_GALAPAGOS			
	0.07110583	0.20623667	0.16799487	3.3112221			
Caída de pasajero	0.04522301	0.23864215	0.23801070	5.2983091			
Choque	0.03664281	0.06609017	0.06464503	0.8944692			
Colisión	0.03400941	0.05996038	0.05640558	0.8824379			
Estrellamiento	0.04158304	0.11426213	0.11348714	2.2732958			
Pérdida de pista	0.03844751	0.07518093	0.06481211	0.9118925			
Volcamiento	0.06914179	0.19515518	0.18937191	3.8924861			

Residual Deviance: 479808.1
AIC: 480396.1

Ilustración 24–4: Resumen del modelo de regresión logística multinomial.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

El gráfico 24–4 contiene el resumen de los coeficientes y los errores estándar correspondiente al modelo de regresión logística multinomial. La primera fila representa los coeficientes para atropello, la segunda para caída de pasajeros y así sucesivamente para las demás variables.

	Arrollamiento	Atropello	Caída de pasajero	Choque	Colisión	Estrellamiento	Pérdida de pista
135991	0.03239595	0.020387520	0.04876802	0.1091466	0.4755793	0.05522290	0.21822612
135992	0.03425586	0.009269064	0.09276030	0.1844587	0.3748822	0.09520668	0.17882455
135993	0.03425586	0.009269064	0.09276030	0.1844587	0.3748822	0.09520668	0.17882455
135994	0.09434581	0.022037190	0.11878370	0.1523687	0.3873752	0.12467896	0.08732054
135995	0.09434581	0.022037190	0.11878370	0.1523687	0.3873752	0.12467896	0.08732054
135996	0.05779407	0.009325113	0.12913779	0.1988310	0.2996542	0.11835794	0.16360236
	volcamiento						
135991	0.04027353						
135992	0.03034261						
135993	0.03034261						
135994	0.01308993						
135995	0.01308993						
135996	0.02329754						

Ilustración 25–4: Predicciones del modelo de regresión de los siniestros de tránsito.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

En el gráfico 25–4 predicciones del modelo de regresión de los siniestros de tránsito. lo que más se puede destacar es que ocurra un siniestro de tránsito de clase estrellamiento esto con un 12 % de probabilidad de ocurrencia, seguido de colisión con el 47 % de que ocurra este tipo de siniestro.

4.2.4. Comparación del modelo de regresión con el análisis de correspondencias

	pre_cat						
	Arrollamiento	Atropello	Caída de pasajero	Choque	colisión	Estrellamiento	Perdida de pista
Arrollamiento	3	0	0	49	2991	0	19
Atropello	1	0	0	73	1801	0	67
Caída de pasajero	1	0	0	83	2191	0	39
Choque	4	0	0	688	5452	0	191
Colisión	9	1	0	258	11219	0	228
Estrellamiento	1	0	0	90	2932	0	61
Perdida de pista	5	0	0	112	4597	0	332
volcamiento	1	0	0	7	468	0	23

	pre_cat
	volcamiento
Arrollamiento	0
Atropello	0
Caída de pasajero	0
Choque	0
Colisión	0
Estrellamiento	0
Perdida de pista	0
volcamiento	0

Ilustración 26–4: Matriz de confusión para medir la precisión.

Fuente: Datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

La figura 26–4 matriz de confusión podemos ver reflejado que tenemos una alta precisión que frecuentemente ocurra un siniestro de tránsito de clase colisión.

Tabla 11–4: Área bajo la curva ROC

Valor Área bajo la curva	60,32 %
---------------------------------	---------

Fuente: Base de datos de los siniestros de tránsito del Ecuador en el periodo de enero 2018 a mayo 2021.

Realizado por: Coloma, Leonel y Guaycha, Jonathan, 2022.

Considerando el área bajo la curva (AUC) del modelo construido, se puede observar en la tabla 11–4 Área bajo la curva ROC que presenta un mayor porcentaje sobre el área bajo la curva. $AUC = 60.32\%$ por lo que nuestro modelo tiene un buen poder predictivo que el análisis de correspondencias observando el valor del porcentaje de variabilidad de la gráfica 14–4 del porcentaje de variabilidad explicada del análisis de correspondencia múltiple con dos componentes observamos que solo explica el 4.2% es decir un modelo de regresión logística multinomial nos muestra de mejor forma la siniestralidad de siniestros de tránsito que un análisis de correspondencias.

CONCLUSIONES

- Como resultado de aplicar un análisis de correspondencias y un modelo de regresión, se presentan la clase de tipo rozamientos, en el mes de diciembre, el siniestro vial más usual es pérdida de carril en las provincias de Pichincha, Tungurahua y el Guayas mientras que en el modelo de regresión multinomial se predice que existirá choques frontal o lateral, además que en el área bajo la curva ROC nos da un 60.32 %, por lo que nuestro modelo tiene un buen poder predictivo que el análisis de correspondencias.
- En Ecuador entre los años 2015-2020 un estudio menciona que los siniestros de tránsito ocurren más a menudo los fines de semana que los días festivos, la hora más recurrente es las 19:00, dándose lugar en las zonas urbanas de las ciudades, además de que las provincias donde ocurrió más accidentes de tránsito son Pichincha y Guayas, con una colisión lateral, en consecuencia, se da a conocer que el sexo más recurrente en estos siniestros es el masculino, y se estima que al menos una persona resultó herida en el 36 % al 40 % de los casos y al menos hubo un lesionado, finalmente en el 10 % de los casos terminó en el fallecimiento de una persona.
- De la información obtenida de la Agencia Nacional de Tránsito en el análisis descriptivo se determinó que las provincias donde surgieron mayormente los siniestros de tránsito son: Pichincha, Guayas y Tungurahua, con accidentes de tipo choque lateral seguido de atropello, que son los que ocurren con más frecuencia esto en el mes de diciembre y dándose lugar los fines de semana es decir viernes, sábado y domingo, además que los días que no son feriados nacionales o locales se producen más accidentes de tránsito, en el período de enero 2018 a mayo 2021.
- En el mapa perceptual multidimensional resultado del análisis de correspondencias múltiple se presenta que la provincia del Napo en el mes de mayo, los días miércoles ocurren mayormente los siniestros de tránsito además que en las provincias de: Pastaza, El Oro y Azuay en el mes de junio es recurrente que existan más arrollamientos, por otro lado en Imbabura, Santa Elena, Zamora Chinchipe y el Carchi, se presentan rozamientos, en el mes de diciembre, el siniestro vial más usual es pérdida de carril, rozamiento en la provincia de Chimborazo, Tungurahua, Cotopaxi y el Guayas.
- El mapa perceptual de dos dimensiones del análisis de correspondencias simple, la clase de siniestro vial y provincias, podemos observar que: choque frontal, estrellamiento, pérdida de pista, ocurren con más frecuencia en las provincias de Cotopaxi, los accidentes de tipo estrellamiento, arrollamiento, colisión, se presentan con frecuencia los fines de semana, los lunes se presenta

choque frontal, y los días jueves pérdida de carril, estos tipos de accidente se dan en su mayoría los días que no son feriados.

- En el análisis de regresión multinomial de las variables clase de siniestro de tránsito, feriado, mes, provincia y días de la semana, se predice que existirá choques de tipo frontal o lateral, además nos muestra que, en la curva de Roc, el área bajo la curva es 60.32 %, por lo que nuestro modelo tiene un buen poder predictivo.
- Considerando el área bajo la curva (AUC) del modelo construido, se puede observar que el Área bajo la curva ROC nos muestra un 60.32 % por lo que nuestro modelo tiene un buen poder predictivo que el análisis de correspondencias observando el valor del porcentaje de variabilidad explicada del análisis de correspondencia múltiple con dos componentes observamos que solo explica el 4.2% es decir un modelo de regresión logística multinomial nos muestra de mejor forma la siniestralidad de siniestros de tránsito que un análisis de correspondencias.

RECOMENDACIONES

- Efectuar un estudio investigativo más afondo de los posibles factores adicionales que están asociados a siniestros de tránsito en el Ecuador.
- Determinar los factores más influyentes en la siniestralidad vial con el fin de que el mapa perceptual presente un mayor porcentaje de variabilidad explicado en dos dimensiones, así se podrá tener una mejor representación de como los factores se encuentran relacionados y como se ven involucrados en los siniestros de tránsito.
- Aplicar nuevas técnicas estadísticas para la continuación de este estudio, aplicando la teoría estadística en base a la exploración de datos y análisis de regresión logística multinomial y análisis de correspondencias simple y múltiple se obtuvieron resultados relevantes para esta problemática, mismo que pretenden reducir la cantidad de estos siniestros en las vías.
- Concienciar a los conductores, peatones, ciclistas o pasajeros, cumplan las leyes de transito y contribuyan con la reducción de estadísticas de estos accidentes.
- Amplificar la base de datos con la información descargada de la página oficial de la Agencia Nacional de Transito u otras instituciones que posean información relevante de accidentes de transito en Ecuador, como el Ecu 911, Instituto Nacional de Estadística y Censos INEC, etc.
- Se recomienda que las instituciones públicas o privadas interesadas en la investigación y reducción de siniestros viales, la cual se considera en Ecuador una de las mayores causas de muertes, daños personales o materiales irreparables, difundan esta información para contribuir a la creación de estrategias que permitan reducir al máximo estos accidentes.

BIBLIOGRAFÍA

ADDINSOFT. Análisis de Correspondencias Múltiples (ACM). XLSTAT, Your data analysis solution [en línea]. 14 de diciembre de 2022 [consultado el 17 de noviembre de 2022]. Disponible en: <https://www.xlstat.com/es/soluciones/funciones/analisis-de-correspondencias-multiples-acm>.

AGENCIA NACIONAL DE TRÁNSITO. Ecuador - Guía Oficial de Trámites y Servicios. Inicio | Ecuador - Guía Oficial de Trámites y Servicios [en línea]. 2022a, [consultado el 21 de diciembre de 2022]. Disponible en: <https://www.gob.ec/ant>.

AGENCIA NACIONAL DE TRÁNSITO. Activación de bloqueo por prohibición de enajenar vehículo (por juicio, embargo, insolvencia o inventario) | ecuador - guía oficial de trámites y servicios. Inicio | Ecuador - Guía Oficial de Trámites y Servicios [en línea]. 2022b,[consultado el 14 de diciembre de 2022]. Disponible en: <https://www.gob.ec/ant/tramites/activacion-bloqueo-prohibicion-enajenar-vehiculo-juicio-embargo-insolvencia -inventario>

AGENCIA NACIONAL DE TRÁNSITO. Visor de Siniestralidad Nacional. www.ant.gob.ec [en línea]. 2022c [consultado el 11 de agosto de 2022]. Disponible en: <https://www.ant.gob.ec/estadisticas-siniestros-de-transito-prueba/>

CABRERIZO, Julia. “Manual para la investigación y reconstrucción de las causas de accidentes de tráfico” [en línea]. España: Wolters Kluwer España, 2016 [consultado el 22 de noviembre de 2022]. ISBN 9788490205709. Disponible en: <https://elibro.net/es/lc/epoch/titulos/55965>.

CHÁVEZ, José & USECHE Lelly. Perfil descriptivo de los siniestros de tránsito en Ecuador durante el periodo 2015-2020. Dialnet [en línea]. 2021, 7(3), 1231–1250 [consultado el 14 de diciembre de 2022]. ISSN 2477-8818. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=8229705>.

CONGACHA, Ana; et al. “Caracterización de los Siniestros Viales en el Ecuador”.

NOVASINERGIA REVISTA DIGITAL DE CIENCIA, INGENIERÍA Y TECNOLOGÍA [en línea]. 2019, 2(2), 17–29 [consultado el 28 de enero de 2023]. ISSN 2631-2654. Disponible en: doi:10.37135/unach.001.04.02.

CUADRAS, Carles. Nuevos métodos de análisis multivariante [en línea]. Barcelona: CMC Edicions Barcelona, Spain, 2018 [consultado el 5 de enero de 2023]. Disponible en: https://gc.scalahed.com/recursos/files/r161r/w24899w/Semana5/METODOS_S5.pdf

DICCIONARIO PRÁCTICO DEL ESTUDIANTE. En: REAL ACADEMIA ESPAÑOLA. Ecuador: Santillana, 2007, p. 307. ISBN 978-9978-29-668-4.

EDISON, Morales & TIERRA Mario. APLICACIÓN DE LA METODOLOGÍA DREAM 3.0, PARA DETERMINAR EL ÍNDICE DE SINIESTRALIDAD DE TRÁNSITO EN LA CIUDAD DE RIOBAMBA. En: dspace.esPOCH.edu.ec [base de datos en línea]. Proyecto de Investigación, ESPOCH, 2019.

FLÓREZ, Carlos; et al. Análisis multicausal de ‘accidentes’ de tránsito en dos ciudades de Colombia. Archivos de Medicina [en línea]. 2018, 18(1), 69–85 [consultado el 2 de noviembre de 2022]. Disponible en: doi:10.30554/archmed.18.1.2477.2018.

GÓMEZ GARCÍA, Antonio ; et al. Caracterización de 1.967 casos de fallecimientos por accidentes de tránsito en Ecuador. Espirales revista multidisciplinaria de investigación [en línea]. 2018, 2(21) [consultado el 28 de enero de 2023]. ISSN 2550-6862. Disponible en: doi:10.31876/re.v2i21.342,pp.2-3.

INEC. CONCEPTOS Y DEFINICIONES. www.inec.gob.pa [en línea]. 26 de junio de 2020 [consultado el 23 de diciembre de 2022]. Disponible en: <https://www.inec.gob.pa/archivos/P4361CONCEPTOS.pdf>

KURSUNOGLU, Nilufer & GOGEBAKAN Maruf. Prediction of spontaneous coal combustion tendency using multinomial logistic regression. Taylor & Francis [en línea]. 2022, 28(4), 2000–2009 [consultado el 11 de enero de 2023]. Disponible en:

doi:10.1080/10803548.2021.1944535

LÓPEZ, Ana. Análisis previo y exploratorio de datos. personal.us.es [en línea]. 2008 [consultado el 16 de diciembre de 2022]. Disponible en: <https://personal.us.es/analopez/aed.pdf>,p.45.

MONTORO, Luis et al. Manual de seguridad vial: El factor humano [en línea]. España: Ariel,S.A, 2000 [consultado el 16 de diciembre de 2022]. ISBN 84-344-1206-3. Disponible en: https://www.researchgate.net/publication/260789893_Manual_de_seguridad_vial_El_factor_humano

OCHOA, Carlos. Muestreo probabilístico: muestreo estratificado. Netquest - Online surveys services and Online Panel [en línea]. 16 de abril de 2015 [consultado el 28 de enero de 2023]. Disponible en: <https://www.netquest.com/blog/es/blog/es/muestreo-probabilistico-muestreo-estratificado>

OMS. Los accidentes de tráfico, la principal causa de fallecimiento de jóvenes de entre 15 y 29 años. [blog]. 2022a. [Consulta: 9 diciembre 2022]. Disponible en: <https://www.un.org/es/observances/road-traffic-victims-day>.

OMS. Los accidentes de tráfico, la principal causa de fallecimiento de jóvenes de entre 15 y 29 años. [blog].2022b. [Consulta: 1 noviembre 2022]. Disponible en: <https://news.un.org/es/story/2022/06/1511112>.

PEÑA, Daniel. Analisis de datos multivariantes [en línea]. Madrid: McGraw-Hill España, 2013 [consultado el 3 de enero de 2023]. ISBN 978-84-481-9184-9. Disponible en: <https://elibro.net/es/lc/epoch/titulos/50267>

REYES, Raquel Hinojosa; et al. Siniestralidad Por Accidentes De Tránsito En México: Una Aproximación Desde El Análisis Espacial, vol. 4, p. 291-309.

ANEXOS

ANEXO A: CÓDIGO DE R

```
1 #####
2 #####                      ANaLISIS DESRIPTIVO
3 #####
4
5 datos <- openxlsx::read.xlsx("C:/Users/Admin/Desktop
6 /Base_Completa (1).csv")
7 ## Var. PROVINCIA
8 datos4 <- datos
9 datos4$PROVINCIA <- factor(datos4$PROVINCIA, labels = c(1:24))
10 levels(datos4$PROVINCIA)
11 barplot(table(factor(datos$PROVINCIA,
12 levels = c('AZUAY', 'BOLiVAR', 'CAnAR', 'CARCHI', 'CHIMBORAZO',
13 'COTOPAXI', 'EL ORO', 'ESMERALDAS', 'GALaPAGOS', 'GUAYAS',
14 'IMBABURA', 'LOJA', 'LOS Rios', 'MANABI', 'MORONA SANTIAGO', 'NAPO',
15 'ORELLANA', 'PASTAZA', 'PICHINCHA', 'SANTA ELENA',
16 'SANTO DOMINGO DE LOS TSaCHILAS', 'SUCUMBios', 'TUNGURAHUA',
17 'ZAMORA CHINCHIPE', ''))), col = c("lightblue", "pink", "#7FFFD4",
18 "#FFE4C4", "#D2B48C", "#FFF8DC", "#C1FFC1", "honeydew1",
19 "#E0FFFF", "#98FB98", "#AB82FF", "#00FF7F", "#EEE9E9", "#FFE7BA"),
20 main = "Diagrama de barras de las frecuencias absolutas\n de
21 la variable \"PROVINCIA\"", ylab = "Frecuencias",
22 xlab = "Povincia", cex.lab = 1, cex.names = 0.5, cex.axis = 0.8)
23
24 ## Var CLASE
25 datos2 <- datos
26 datos2$CLASE <- factor(datos2$CLASE, labels = c(1:13))
27 levels(datos2$CLASE)
28 barplot(table(factor(datos$CLASE, levels = c('Arrorramiento',
29 'Atropello', 'Caida de pasajero', 'Choque frontal',
30 'Choque lateral', 'Choque posterior', 'Colision',
31 'Estrellamiento', 'Otros', 'Perdida de carril',
32 'Perdida de pista', 'Rozamiento', 'Volcamiento' )),
33 col = c("lightblue", "pink", "#7FFFD4", "#FFE4C4", "#D2B48C",
34 "#FFF8DC", "#C1FFC1", "honeydew1", "#E0FFFF", "#98FB98",
35 "#AB82FF"), main = "Diagrama de barras de las frecuencias
```

```

36 absolutas de la variable \"Clase de siniestros viales\"
37 , cex.main = 1, ylab = ("Frecuencias"), xlab = ("Clase"),
38 cex.lab =1,cex.names = 0.8,cex.axis = 0.8)
39
40 ## Var MES
41 datos6 <- datos
42 datos6$MES <- factor(datos6$MES, labels = c(1:12))
43 levels(datos6$MES)
44
45
46 barplot(table(factor(datos$MES, levels =
47 c('ENERO', 'FEBRERO', 'MARZO', 'ABRIL', 'MAYO',
48 'JUNIO', 'JULIO', 'AGOSTO', 'SEPTIEMBRE', 'OCTUBRE'
49 , 'NOVIEMBRE', 'DICIEMBRE'))), col = c("lightblue", "pink",
50 "#7FFFD4", "#FFE4C4", "#98F5FF", "#FFF8DC", "#C1FFC1"
51 , "honeydew1", "#E0FFFF", "lightpink", "#AB82FF",
52 "#FFE4E1"), main = "Diagrama de barras de las frecuencias absolutas\n
53 de la variable \"MES\"", cex.main = 1, ylab = ("Frecuencias"),
54 xlab = ("Mes"), cex.lab =1,cex.names = 0.6,cex.axis = 0.5)
55 ## Var FERIADO
56 datos3 <- datos
57 datos3$FERIADO <- factor(datos3$FERIADO, labels = c(1:2))
58 levels(datos3$FERIADO)
59 barplot(table(factor(datos$FERIADO, levels = c('SI', 'NO'))),
60 col = c("lightblue", "pink"), main = "Diagrama de barras
61 de las frecuencias absolutas\n de la variable \"Feriado\"",
62 cex.main = 1, ylab = ("Frecuencias"), xlab = ("Feriado"),
63 cex.lab =1, cex.names = 1,cex.axis = 1)
64
65 ## Var DIAS
66 datos5 <- datos
67 datos5$DIA <- factor(datos5$DIA, labels = c(1:7))
68 levels(datos5$DIA)
69
70 barplot(table(factor(datos$DIA, levels =
71 c('Lunes', 'Martes', 'Miercoles', 'Jueves',
72 'Viernes', 'Sabado', 'Domingo'))), col =
73 c("lightblue", "pink", "#7FFFD4", "#FFE4C4", "#98F5FF",
74 "#FFF8DC", "#C1FFC1"), main = "Diagrama de barras

```

```

75 de las frecuencias absolutas\n de la variable \"Dias\",
76 cex.main = 1, ylab = ("Frecuencias"), cex.lab =1,
77 xlab = ("Dias"), cex.names = 1,cex.axis = 1)
78
79 # VARIABLE NUM_FALLECIDO
80 # VARIABLES NUM_FALLECIDO y NUM_LESIONADOS
81 mDatos=data.frame(datos$NUM_FALLECIDO, datos$NUM_LESIONADO)
82 str(mDatos)
83 summary(mDatos)
84 #Diagrama de barras de la var. NUM_FALLECIDO con cero FALLECIDOS
85
86 text(barplot(table(datos$NUM_FALLECIDO)[c(1)],
87 main = "Diagrama de barras de las frecuencias absolutas de\n
88 \"0 fallecidos\",xlab = "Numero de Fallecidos" ,
89 ylab = "Frecuencia",col = c("#C1FFC1"), ylim = c(0, 170000)),
90 145000,labels = table(datos$NUM_FALLECIDO)[c(1)] )
91
92 #Diagrama de barras de la 2 y 3 FALLECIDOS
93 text(barplot(table(datos$NUM_FALLECIDO)[c(2,3)],
94 xlab = "Numero de Fallecidos" ,ylab = "Frecuencia",
95 col = c("#AB82FF","#FFE4E1"), main = "Diagrama de barras de
96 las frecuencias absolutas de\n \"1 y 2 fallecidos\",
97 ylim=c(0,12000), space = c(0)),
98 table(datos$NUM_FALLECIDO)[c(2,3)] - 350,
99 labels = table(datos$NUM_FALLECIDO)[c(2,3)])
100 #Diagrama de barras de la 4 a 22 FALLECIDOS
101 text(barplot(table(datos$NUM_FALLECIDO)
102 [-c(1,2,3,4)],xlab = "Numero de fallecidos" ,
103 main = "Diagrama de barras de las frecuencias
104 absolutas de\n \" 4 A 22 fallecidos\", ylab = "Frecuencia",
105 ylim= c(0,50),col = c("#98F5FF","#FFF8DC","#C1FFC1",
106 "honeydew1","#E0FFFF","lightpink","#AB82FF","#FFE4E1"))
107 ,table(datos$NUM_FALLECIDO)[-c(1,2,3,4)]+2,labels =
108 table(datos$NUM_FALLECIDO)[-c(1,2,3,4)])
109 #Tabla la var. NUM_FALLECIDO
110 table(datos$NUM_FALLECIDO)
111
112 # VARIABLE NUM_LESIONADOS
113 # VARIABLES NUM_FALLECIDO y NUM_LESIONADOS

```

```

114 mDatos=data.frame(datos$NUM_FALLECIDO, datos$NUM_LESIONADO)
115 str(mDatos)
116 summary(mDatos)
117 #Diagrama de barras de 0,1,2,3 LESIONADOS
118 text(barplot(table(datos$NUM_LESIONADO)[c(1,2,3,4)],
119 xlab = "Numero de lesionados" ,ylab =
120   "Frecuencia", main = "Diagrama de barras de las
121 frecuencias absolutas de\n  \"0, 1, 2 y 3 lesionados\"")
122 ,ylim= c(0,90000),col = c("#98F5FF","#FFF8DC","#C1FFC1",
123 "honeydew1","#E0FFFF","lightpink","#AB82FF","#FFE4E1")),
124 table(datos$NUM_LESIONADO)[c(1,2,3,4)]+2800,labels =
125   table(datos$NUM_LESIONADO)[c(1,2,3,4)])
126 # Diagrama de barras de 4 a 11 LESIONADOS
127 text(barplot(table(datos$NUM_LESIONADO)[c(5,6,7,8,9,10,11,12)]
128 ,xlab = "Numero de lesionados" ,ylab = "Frecuencia",
129 ylim = c(0,2000),main = "Diagrama de barras de las
130 frecuencias absolutas de\n  \"4 a 11 lesionados\"",
131 col = c("#98F5FF","#FFF8DC","#C1FFC1","honeydew1",
132 "#E0FFFF","lightpink","#AB82FF","#FFE4E1")),
133 table(datos$NUM_LESIONADO)[c(5,6,7,8,9,10,11,12)]+80,
134 labels = table(datos$NUM_LESIONADO)[c(5,6,7,8,9,10,11,12)])
135 #Diagrama de barras de 12 a 44 LESIONADOS
136 text(barplot(table(datos$NUM_LESIONADO)[-c(1:12)],
137 xlab = "Numero de lesionados" ,
138 ylab = "Frecuencia", ylim= c(0,30),
139 main = "Diagrama de barras de las frecuencias absolutas
140 de\n  \"12 a 44 lesionados\"",
141 col = c("#98F5FF","#FFF8DC","#C1FFC1",
142   "honeydew1","#E0FFFF","lightpink",
143   "#AB82FF","#FFE4E1")),table(datos$NUM_LESIONADO)
144 [-c(1:12)]+1,labels = table(datos$NUM_LESIONADO)[-c(1:12)])
145 #Tabla la var. NUM_FALLECIDO
146 table(datos$NUM_FALLECIDO)
147
148
149 #####
150 ##### ANaLISIS DE CORRESPONDENCIA #####
151 #####
152 library(tidyverse)

```

```

153 library(FactoMineR)
154 library(factoextra)
155 data <- read.csv("C:\\Users\\BELEN\\Desktop
156 \\tesis-octubre\\tesis octubre\\Base_Completa (1).csv")
157 data$PROVINCIA <- factor(data$PROVINCIA,
158   labels = c("LOS RIOS", "MANABI", "GALAPAGOS", "SANTO
159 DOMINGO DE LOS TSACHILAS", "AZUAY", "CARCHI", "CHIMBORAZO",
160 "BOLIVAR", "SUCUMBIOS", "CANAR", "COTOPAXI",
161 "EL ORO", "ESMERALDAS", "GUAYAS", "IMBABURA", "LOJA",
162 "MORONA SANTIAGO", "NAPO", "ORELLANA", "PASTAZA",
163 "PICHINCHA", "SANTA ELENA", "TUNGURAHUA", "ZAMORA CHINCHIPE"))
164
165 dt <- data %>%
166   select(., c(4, 5, 9, 11, 18, 19, 20))
167
168
169 aa <- (factor(dt$DIA, #levels = c("Lunes", "Martes",
170 # "Miercoles", "Jueves", "Viernes"),
171 labels = c("Domingo", "Domingo", "Jueves", "Jueves",
172 "Lunes", "Lunes", "Martes", "Martes", "Miercoles",
173 "Miercoles", "Miercoles", "Sabado", "Sabado", "Sabado",
174 "Viernes", "Viernes")))
175 dt$DIA <- aa
176 aaa <- (factor(dt$CLASE, labels = c("Arrollamiento",
177 "Atropello", "Caida de pasajero", "Choque frontal",
178 "Choque frontal", "Choque lateral", "Choque posterior",
179 "Colision", "Colision", "Estrellamiento", "Otros",
180 "Perdida de carril", "Perdida de pista", "Rozamiento",
181 "Volcamiento")))
182 dt$CLASE <- aaa
183
184 ##### muestra #####
185
186 muestra <- c(114, 119, 19, 14, 18, 10, 3, 19, 9, 8, 4, 13,
187           13, 7, 2, 2, 1, 2, 2, 1, 1, 1, 1)
188 prov <- c('GUAYAS', 'PICHINCHA', 'MANABI', 'LOS RIOS',
189 'AZUAY', 'EL ORO', 'ESMERALDAS', 'TUNGURAHUA', 'CHIMBORAZO',
190 'LOJA', 'COTOPAXI', 'IMBABURA', 'SANTO DOMINGO
191 DE LOS TSACHILAS', 'SANTA ELENA', 'CANAR', 'BOLIVAR',

```

```

192 'SUCUMBios', 'CARCHI', 'MORONA SANTIAGO',
193 'ORELLANA', 'NAPO', 'ZAMORA CHINCHIPE', 'PASTAZA')
194
195 muestra
196 prov
197
198
199 ### Seleccion de la muestra por provincias
200 outt <- vector("list", length(muestra))
201
202 for(i in 1:length(muestra)){
203
204 if((levels(factor(dt$PROVINCIA,
205 levels = prov))) [i] == prov[i]){
206     outt[[i]] <- dt[sample(which(dt$PROVINCIA
207 == prov[i]), muestra[i]) ,]
208
209 }
210 }
211
212 DATA_MUESTRAS <- do.call(rbind.data.frame, outt)
213 ### Analisis de correspondencia
214 # Multiple
215 dt_MCA <- MCA(DATA_MUESTRAS[, -c(5, 6)],
216 method = 'Burt', graph = F)
217
218 # Variables
219 fviz_mca_var(dt_MCA, repel = T)
220
221 # Variables e individuos
222 fviz_mca_biplot(dt_MCA, repel = T)
223
224 # Individuos
225 fviz_mca_ind(dt_MCA)
226
227 # Porcentaje de varianza explicada
228 factoextra::fviz_screplot(dt_MCA, addlabels = T)
229
230 ### Analisis de correspondencia en pares de variables

```

```

231 # PROVINICA Y CLASE
232 tabla_pro_c1 <- table(DATA_MUESTRAS[, c('PROVINCIA', 'CLASE')])
233 COP_1 <- FactoMineR::CA(tabla_pro_c1, graph = F)
234 summary(COP_1)
235
236 factoextra::fviz_screplot(COP_1, addlabels = T)
237 factoextra::fviz_ca_biplot(COP_1, repel = T)
238
239 # DIA Y CLASE
240 tabla_pro_c2 <- table(DATA_MUESTRAS[, c('DIA', 'CLASE')])
241 COP_2 <- FactoMineR::CA(tabla_pro_c2, graph = F)
242 summary(COP_2)
243
244 factoextra::fviz_screplot(COP_2, addlabels = T)
245 factoextra::fviz_ca_biplot(COP_2, repel = T)
246
247 # FERIADO Y CLASE
248 tabla_pro_c3 <- table(DATA_MUESTRAS[, c('CLASE', 'FERIADO')])
249 COP_3 <- FactoMineR::CA(tabla_pro_c3, graph = F)
250 factoextra::fviz_screplot(COP_3, addlabels = T)
251 table(DATA_MUESTRAS[, c('CLASE', 'FERIADO')])
252 par(cex=1.5)
253 barplot(table(DATA_MUESTRAS[, c('FERIADO', 'CLASE')]),
254 cex.names = 0.5, legend = rownames(table(DATA_MUESTRAS
255 [, c('FERIADO', 'CLASE')])), beside = F, col = c('lightblue', "pink"))
256
257 tabla_pro_c4 <- table(DATA_MUESTRAS[, c('MES', 'CLASE')])
258 COP_4 <- FactoMineR::CA(tabla_pro_c4, graph = F)
259 summary(COP_4)
260
261 factoextra::fviz_screplot(COP_4, addlabels = T)
262 factoextra::fviz_ca_biplot(COP_4, repel = T)
263
264 #####
265 ##### REGRESION LOGISTICA MULTINOMIAL #####
266 #####
267
268 library(fastDummies)
269 library(nnet)

```

```

270 library(VGAM)
271 library(tidyverse)
272 library(pROC)
273 file.choose()
274 ruta <- "C:/Users/Admin/Desktop/Base_Completa (1).csv"
275 datal <- read.csv(ruta)
276
277 datal <- datal %>%
278   select(., c(4, 5, 9, 11, 20))
279 View(datal)
280
281 datal$PROVINCIA <- factor(datal$PROVINCIA,
282   labels = c("LOS RIOS", "MANABi", "GALaPAGOS",
283   "SANTO DOMINGO DE LOS TSaCHILAS", "AZUAY", "CARCHI",
284   "CHIMBORAZO", "BOLiVAR", "SUCUMBios", "CAnAR", "COTOPAXI",
285   "EL ORO", "ESMERALDAS", "GUAYAS", "IMBABURA", "LOJA",
286   "MORONA SANTIAGO", "NAPO", "ORELLANA", "PASTAZA", "PICHINCHA",
287   "SANTA ELENA", "TUNGURAHUA", "ZAMORA CHINCHIPE"))
288
289
290
291 aa<- factor(datal$DIA, #levels = c("Lunes", "Martes",
292 # "Miercoles", "Jueves", "Viernes"),
293 labels = c("Domingo", "Domingo", "Jueves", "Jueves",
294   "Lunes", "Lunes", "Martes", "Martes", "Miercoles", "Miercoles",
295   "Miercoles", "Sabado", "Sabado", "Sabado", "Viernes", "Viernes"))
296 datal$DIA <- aa
297 aaa <- factor(datal$CLASE, labels = c("Arrollamiento", "Atropello",
298 "Caida de pasajero", "Choque", "Choque", "Choque", "Choque",
299 "Colision", "Colision", "Estrellamiento",
300 "Caida de pasajero", "Perdida de pista", "Perdida de pista",
301 "Caida de pasajero", "Volcamiento"))
302 datal$CLASE <- aaa
303
304 table(datal$CLASE)
305 unique(datal$CLASE)
306
307 ## imprimir esta conclusion como inicio de todo regresion
308

```

```

309 # ## El modelo de regresion logistica multinomial con
310 # las variables,## clase, feriado, mes, dia y provincia
311 # se aplico con la finalidad## de predecir e identificar
312 # las variables que influyen que se produzcan## siniestros
313 # de transito. Para la estimacion del modelo se procedio a
314 # ## utilizar una muestra de train y una de test, despues
315 # de seleccionar## las variables mas significativas para
316 # el modelo de regresion logistica multinomial.
317
318
319 # PRIMER MODELO VARIABLES DUMMY
320
321 # DIA, FERIADO, CLASE, MES, PROVINCIA
322 # creacion de variable dummy o variables ficticias
323 str(data1[, c('DIA', 'FERIADO', 'CLASE', 'MES', 'PROVINCIA')])
324 var_dummies <- dummy_cols(data1[, c('DIA',
325 'FERIADO', 'MES', 'PROVINCIA')])
326 var_dummies <- var_dummies[, -c(1:4)]
327 head(var_dummies) ###IMPMIRIR ESTA LINEA
328
329 dt <- data.frame(Y = data1$CLASE, var_dummies)
330
331 set.seed(12345)
332 train <- sample_frac(dt, 0.8)
333 test <- dt[-(as.numeric(rownames(train))), ]
334
335 # modelo
336 d_train <- multinom(Y ~., data = train)
337 sum_train <- summary(d_train)
338 sum_train ###IMPMIRIR ESTA LINEA
339
340 z_train <- sum_train$coefficients/sum_train$standard.errors
341 p_train <- round((1 - pnorm(abs(z_train), 0, 1)) * 2, 5)
342
343 round(exp(coef(d_train)), 4)
344
345 # los valores ajustados
346 head(probability_train <- fitted(d_train))#IMPMIRIR ESTA LINEA
347

```

```

348
349 # predicicones en probabilidades
350 pre_train <- predict(d_train, test, type = 'prob')
351 head(pre_train)###IMPMIRIR ESTA LINEA
352 # predicciones en categorias
353 pre_cat <- predict(d_train, test, type = 'class')
354 head(pre_cat)###IMPMIRIR ESTA LINEA
355 # matriz de confusion
356 con_table <- table(test$Y, pre_cat)###IMPMIRIR ESTA LINEA
357
358 # precision del modelo de regresion logistica con
359 # los datos de entrenamiento
360 sum(diag(con_table))/sum(con_table) * 100
361
362 roc_multi <- multiclass.roc(test$Y, pre_train) ###IMPMIRIR ESTA LINEA
363
364 auc(roc_multi)###IMPMIRIR ESTA LINEA
365 # Conclusion usando la funcion de la Libreria
366 # pROC, de la variable multivariante
367 # # predictora pre_traincon 8 niveles test$Y:
368 # # Arrollamiento, Atropello, Caida de pasajero, Choque,
369 # Colision, Estrellamiento,
370 # # Perdida de pista, Volcamiento, obteniendo asi el
371 # area bajo la curva de 0.6032
372 plot.roc(auc(roc_multi))

```



ESCUELA SUPERIOR POLITÉCNICA DE
CHIMBORAZO

DIRECCIÓN DE BIBLIOTECAS Y RECURSOS DEL
APRENDIZAJE



UNIDAD DE PROCESOS TÉCNICOS

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 25 / 04 / 2023

INFORMACIÓN DE LOS AUTORES
Nombres – Apellidos: Leonel Alexander Coloma Llango Jonathan Roberto Guaycha Imbaquingo
INFORMACIÓN INSTITUCIONAL
Facultad: Ciencias
Carrera: Estadística
Título a optar: Ingeniero Estadístico
f. Analista de Biblioteca responsable: Ing Fernanda Arévalo M.



0319-DBRA-UPT-2023