

# **CARACTERIZACIÓN DE ACCIDENTES DE TRÁNSITO EN MÉXICO APLICANDO AGRUPAMIENTO JERÁRQUICO**

## *TRAFFIC ACCIDENTS PROFILING IN MEXICO BASED ON HIERARCHY CLUSTERING*

***Luis Alberto Morales Rosales***

Universidad Michoacana de San Nicolás de Hidalgo, México  
*lamorales@conacyt.mx*

***Jaime Saavedra Rosales***

Universidad Michoacana de San Nicolás de Hidalgo, México  
*jaime\_saavedramx@yahoo.com*

***Mariana Lobato Báez***

Instituto Tecnológico Superior de Libres, México  
*mariana.lobato@upaep.edu.mx*

***Ignacio Algreto Badillo***

Instituto Nacional de Astrofísica Óptica y Electrónica, México  
*ialgreto@conacyt.mx*

***Sergio Salatiel Ramírez Bravo***

Universidad Michoacana de San Nicolás de Hidalgo, México  
*salatielramirez98@gmail.com*

**Recepción:** 22/octubre/2019

**Aceptación:** 23/noviembre/2019

### **Resumen**

Los accidentes de tránsito impactan directamente en los indicadores del bienestar social y el desarrollo socioeconómico de un país. En este artículo se plantea como premisa que los accidentes no se reparten al azar a lo largo de la red vial de un país, sino que preservan ciertos factores de comportamiento, lo que permite su caracterización. La aportación principal se centró en observar que los estados con un nivel de pobreza cercano reflejan agrupaciones con características similares de las variables que describen sus índices de accidentes de tránsito descritos por las estadísticas del INEGI (1997-2016). Se obtuvieron nueve factores que describen cómo suceden los accidentes por cada estado, permitiendo dimensionar la afectación de las variables y su interacción para la toma de decisiones por parte de las autoridades correspondientes. Estos elementos

permiten tomar decisiones respecto a medidas preventivas y de actuación por parte del gobierno con el fin de determinar políticas públicas que se ajusten a sus características propias.

**Palabras Claves:** Accidentes de tránsito, agrupamiento jerárquico, caracterización.

## **Abstract**

*Traffic accidents directly impact the indicators of social welfare and socioeconomic development of a country. This paper proposes that accidents are not distributed randomly along the road network of a country but they preserve certain behavior, which allows their characterization. The main contribution focused on observing that states with a near poverty level reflect groups with similar characteristics that describe their traffic accident rates, according by INEGI statistics (1997-2016). Nine factors were obtained that describe how accidents happen by each state, allowing to measure the affectation of the variables and their interaction in order to help in the decision making regarding what type of preventive measures and actions can be applied by the government within the public policies.*

**Keywords:** *hierarchy clustering, profiles, traffic accidents.*

## **1. Introducción**

La red carretera que conforma la infraestructura vial de un país está conformada por tres elementos: 1) el conductor, 2) el vehículo y 3) el camino. Los accidentes de tránsito son generados cuando estos tres elementos interactúan entre sí ocasionando problemas en el flujo del sistema. Por lo que se puede definir a un accidente como la consecuencia de un evento fortuito multicausal, precedido por una falla en alguno de los elementos que conforman el sistema de tránsito [Chavarría, 2001]. Otra definición propuesta por la Organización Panamericana de la Salud señala que “una colisión en las vías de tránsito es el resultado de una combinación de factores relacionados con los componentes del sistema vial que incluyen las calles, los caminos, el entorno vial, los vehículos y los usuarios de la vía pública y la manera que interactúan” [Castro, 2011].

Los accidentes de tránsito son considerados un problema global, ya que la Organización Mundial de la Salud ha afirmado que representan un problema de salud pública [Peden, 2004]. En México se tiene un registro de 36215 víctimas de accidentes carreteros a lo largo de toda su red carretera [INEGI, 2017]. Estas cifras están relacionadas con la movilidad realizada en automóviles, autobuses, camiones, y por los desplazamientos en bicicleta o a pie realizados por la población mexicana, cifra que genera una preocupación de las autoridades de la Secretaría de Salud, y en particular, de la Secretaría de Comunicaciones y Transportes.

Los actores involucrados junto con el INEGI han generado la recopilación de accidentes de tránsito con el fin de crear políticas públicas para la planificación y distribución de esfuerzos en campañas de prevención de accidentes viales. Los estudios desarrollados se concentran en una recopilación estadística de los factores que describen los accidentes, haciendo énfasis, en resumir la cantidad de daños, lesionados y pérdida de seres humanos. Tomando de base lo anteriormente descrito, se percibe la falta de estudios de accidentes viales que describan las características de agrupación de los factores y cómo se relacionan entre ellos a nivel estado de la República Mexicana. Esto permitiría generar una mejor distribución de los recursos destinados a disminuir el número de accidentes generados por año y en particular por estado.

En este artículo se propone una caracterización de los accidentes de tránsito generados dentro de los Estados Unidos Mexicanos; particularmente, se presentan los casos de estudio de ocho estados con un nivel de pobreza mayor al 50% de su población de acuerdo a lo reportado por el [CONEVAL, 2016].

México sólo cuenta con dos bases de datos públicas sobre accidentes de tránsito con cobertura nacional, una proporcionada por la Secretaría de Comunicaciones y Transportes y otra generada por el INEGI. En este trabajo se optó por la base de datos del INEGI al presentar una mayor cantidad de información estructurada para su análisis, así como una mayor temporalidad en la recopilación de datos públicos disponibles.

La disponibilidad de los datos del INEGI en internet permite analizar su registro para el período 1996-2016 y conocer los cambios y patrones espacio-temporales de la

ocurrencia de accidentes, independientemente de las discrepancias que pudieran identificarse. Esta base de datos permite conocer el impacto socioeconómico y territorial de los accidentes de tránsito en México.

[Chias, 2014] realizaron un análisis de la calidad de los datos que presenta esta base de datos. Entre los resultados del análisis observaron que se ha mejorado o ampliado la cobertura territorial del registro de datos, qué categorías de análisis y tipo de variables se consideran, así como su completitud y coherencia temporal, qué municipios han registrado datos durante todo el lapso considerado, cuáles y cuándo se han incorporado nuevos municipios, quiénes tienen o no consistencia en sus registros. A partir de este análisis de datos se podrían proponer acciones específicas para cumplir con el mandato de registrarlos y con ello contar con una base de datos más robusta y confiable que permita cumplir con las metas nacionales establecidas en el Plan Mundial para el Decenio de la Seguridad Vial 2011–2020.

De acuerdo a [Chias, 2014], México cuenta con una base de datos estadística de accidentes de Tránsito en zonas urbanas y suburbanas [INEGI, 2017] que mucha gente desconoce y en ocasiones subestima por considerar que no registra más que una pequeña parte de los accidentes que ocurren en nuestros municipios. Sin embargo, poco se ha explotado, y su potencial de uso es extenso, faltan datos ciertos, pero lo que se tiene permite tomar decisiones.

La seguridad vial al ser considerada como un problema de salud pública por la OMS representa una de las prioridades de los gobiernos. Por lo tanto, la identificación de las causas de los accidentes de tráfico se ha convertido en el objetivo principal de reducir los daños causados por los accidentes de tráfico. Por otro lado, debido al crecimiento exponencial del volumen de datos, el análisis de los factores a este respecto no es práctico.

En los modelos tradicionales de análisis estadístico, de acuerdo a [Karolien, 2003] no es posible desarrollar contramedidas efectivas para mejorar la seguridad del tráfico sin poder relacionar de manera adecuada y sistemática la frecuencia y gravedad de los accidentes con un gran número de variables como el tráfico, factores geométricos y ambientales. Además, indican que, en el pasado, los modelos estadísticos se han utilizado ampliamente para analizar los accidentes de

tránsito. Sin embargo, demuestran que pueden surgir ciertos problemas al usar el análisis estadístico clásico en conjuntos de datos con dimensiones tan grandes como un aumento exponencial en el número de parámetros a medida que aumenta el número de variables y, como consecuencia, la invalidez de las pruebas estadísticas de datos dispersos en grandes tablas de contingencia.

La minería de datos está enfocada en extracción no trivial de información implícita, y cuyos patrones no son visibles o determinados por medios estadísticos tradicionales al analizar grandes cantidades de datos [Karolien, 2003]. Cuando existen modelos y relaciones simples entre los datos a analizar los estudios estadísticos descriptivos deben ser empleados, pero cuando se requiere encontrar patrones cuyas relaciones no son triviales y resultan complejas debido a la gran cantidad de datos a analizar, las técnicas de minería de datos permitirán descubrir patrones y entender estas relaciones, como es el caso del estudio de accidentes de tráfico. Otra ventaja de la minería de datos es la identificación de las variables relevantes que contribuyen en gran medida al comportamiento de los datos, esto permite entender los factores que describen las circunstancias en que los accidentes suceden. Por lo tanto, en esta investigación se identifican factores por estado que caracterizan los accidente mediante técnicas de agrupamiento y las perfilaremos en términos de datos relacionados con causas, tipos y características de temporalidad mediante técnicas de extracción de datos para proporcionar información valiosa para las acciones del gobierno hacia la seguridad del tráfico.

[Bigham, 2010] utilizan algoritmos de minería de datos, incluidos el análisis de cestas y el control continuo para encontrar factores de accidentes de tráfico. Utilizaron un conjunto de datos denominado UKDA empleando el programa WEKA como herramienta de análisis. Obtuvieron reglas de asociación para la identificación de los factores involucrados en el accidente que ocurren juntos; ya que comprenden las circunstancias y las causas del accidente. Por lo tanto, la asociación de minería da la dirección a una investigación más profunda sobre las causas de los accidentes de tráfico. Ayuda al gobierno a adaptar las políticas de seguridad del tráfico con diferentes tipos de accidentes y situaciones. Su principal resultado fue que, aunque las características de la sociedad y el comportamiento son muy importantes en la

ocurrencia de todos los accidentes de tránsito, se pudo entender que las características espaciales y la infraestructura representan un elemento importante en el accidente. Particularmente, propusieron nuevos atributos a partir de la información recopilada, generaron modelos predictivos basados en reglas empleando el algoritmo PART de WEKA. El principal énfasis de la información fue considerar las condiciones del camino generando una descripción de diez atributos que describen tanto la separación, tipo de superficie, fallas, e iluminación que presentan, entre otras.

[Martinelli, 2018] presentan un trabajo enfocado a identificar la forma en que los conductores se desplazan en una ciudad. El enfoque utilizado para identificar la agresividad del conductor explora el uso de técnicas de aprendizaje automático sin supervisión, particularmente mediante técnicas de agrupamiento.

## **2. Métodos**

El análisis de los accidentes de tránsito propuesto en este artículo se realizó con la técnica de minería de datos denominada, agrupamiento jerárquico conglomerado. En el método de agrupamiento jerárquico en cada paso del algoritmo sólo un objeto cambia de grupo, y los grupos están anidados en los pasos anteriores. Si un objeto ha sido asignado a un grupo ya no cambia más de grupo. El método jerárquico es adecuado para determinar el número óptimo de conglomerados existente en los datos y el contenido de los mismos. Se utiliza cuando no se conoce el número de clúster a priori y cuando el número de objetos no es muy grande; además permite trabajar conjuntamente con variables de tipo mixto (cualitativas y cuantitativas). Siempre que todas las variables sean del mismo tipo, el procedimiento Análisis de Conglomerados Jerárquico podrá analizar variables de intervalo (continuas), de recuento o binarias.

El proceso de agrupación jerárquico se puede resumir gráficamente mediante una representación en dendrogamas. Los objetos similares se enlazan y su posición en el diagrama está determinada por el nivel de similitud/disimilitud entre los objetos. Un dendrograma es una representación gráfica en forma de árbol, en el que los clústeres están representados mediante trazos verticales y las etapas de fusión

mediante trazos horizontales. La separación entre las etapas de fusión es proporcional a la distancia a la que están los grupos que se funden en esa etapa. Los dendrogramas pueden emplearse para evaluar la cohesión de los conglomerados que se han formado y proporcionar información sobre el número adecuado de conglomerados que deben conservarse.

### **Descripción del Problema de Agrupamiento**

Consideremos una muestra  $X$  formada por  $n$  individuos sobre los que se miden  $p$  variables,  $X_1, \dots, X_p$  ( $p$  variables numéricas observadas en  $n$  objetos). Sea  $X_{ij}$  el valor de la variable  $X_i$  en el  $i$  –ésimo objeto  $i = 1, \dots, n$  ;  $j = 1, \dots, p$ . Este conjunto  $X$  de valores numéricos se pueden ordenar en una matriz.

La  $i$  –ésima fila de la matriz  $X$  contiene los valores de cada variable para el  $i$  –ésimo individuo, mientras que la  $j$  –ésima columna muestra los valores pertenecientes a la  $j$  –ésima variable a lo largo de todos los individuos de la muestra. Por lo que el problema puede ser planteado como: Dado un conjunto de  $n$  individuos caracterizados por la información de  $p$  variables  $X_j$  ( $j = 1, 2, \dots, p$ ), se plantea agruparlos de manera que los individuos pertenecientes a un grupo (y siempre con respecto a la información disponible de las variables) sean lo más similares posibles entre sí y los distintos grupos sean entre ellos tan disimilares como sea posible.

El proceso completo puede estructurarse de acuerdo con los siguientes elementos:

- Se inicia a partir de un conjunto de  $n$  individuos de los que se dispone de una información codificada por un conjunto de  $p$  variables (una matriz de datos de  $n$  individuos y  $p$  variables):
  - ✓ **Selección de las  $p$  variables.** En este caso se eligen y considera la codificación de los valores de cada variable a incluir que permita caracterizar los accidentes de tránsito. Para ellos se utilizó la información de accidentes del INEGI.
  - ✓ **Detección de valores atípicos.** Dado que el análisis por agrupamientos es muy sensible a la presencia de valores atípicos, es necesaria la eliminación de estos valores.

- ✓ **Transformar valores.** La técnica de agrupamiento es sensible al hecho que las variables no estén medidas en las mismas unidades y esto se ve reflejado en que la variabilidad de sus valores sea muy diferente. Es por ello que se estandarizan los datos antes de calcular las similaridades. En este caso se elige la técnica de puntuaciones  $Z$ ; es decir los datos son estandarizados a puntuaciones  $Z$ , con media 0 y desviación típica 1.
- Se establece un criterio de similaridad y se construye matriz de similaridades, que permite relacionar la semejanza de los individuos entre sí. Para medir lo similares (o disimilares) que son los individuos existe una gran cantidad de índices de similaridad y de disimilaridad o divergencia. En este caso, se elige la medida basada en coseno. Esta medida está estrechamente relacionada con el coeficiente de correlación de Pearson. Es el coseno del ángulo formado por dos vectores de puntuaciones. Tiene un máximo de 1 y un mínimo de -1, se calcula mediante la ecuación 1.

$$\text{Coseno}(X, Y) = \frac{\sum_i^n x_i y_i}{\sqrt{(\sum_i^n x_i^2)(\sum_i^n y_i^2)}} \quad (1)$$

- Se emplea un algoritmo de clasificación jerárquica para determinar la estructura de agrupación de los individuos, empleando la distancia entre los grupos como métrica de agrupamiento.
- Se especifica estructura de agrupamiento jerárquico mediante dendrograma.

### 3. Resultados

#### Selección de Información

Como punto de partida para la selección de la información se determinó que un aspecto relevante para evaluar la tasa de accidentes de tránsito es segmentar los estados de la República Mexicana, con base en un criterio que permita contrastar el comportamiento o aparición de los accidentes ante distintas condiciones. De acuerdo con lo establecido por la Ley General de Desarrollo Social (LGDS) de México, la medición de la pobreza considera los indicadores de:



- El ingreso de los hogares.
- Las carencias sociales en materia de educación.
- Acceso a los servicios de salud.
- Acceso a la seguridad social.
- Calidad y espacios de la vivienda.
- Acceso a servicios básicos en la vivienda.
- Acceso a la alimentación.
- El grado de cohesión social.

La medición de la pobreza en las entidades federativas incluye los dos grandes rubros de la pobreza nacional:

- Las carencias sociales.
- El ingreso.

El criterio de selección de los estados a analizar en esta investigación fue el índice porcentual de la pobreza de México, generado por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) en el marco de sus atribuciones, y con base en la información proporcionada por el Instituto Nacional de Estadística y Geografía (INEGI).

La estadística que se analizó fue la evolución de la pobreza 2010-2016 a nivel nacional para cada entidad federativa. En la tabla 1 se observa la evolución anual del porcentaje de la pobreza y pobreza extrema nacional de acuerdo a las entidades federativas, en los años 2010, 2012, 2014 y 2016\*.

El índice de pobreza que se seleccionó fue el calculado para el año 2016. En la figura 1 se observa la distribución por cada uno de los estados de la república mexicana.

Se consideró la hipótesis de que los estados con un índice de pobreza alto podrían presentar características similares de cómo suceden los accidentes de tránsito. Es por ello que se eligieron 8 estados de la república con un índice mayor al 50%, estos fueron: Chiapas, Oaxaca, Guerrero, Veracruz, Puebla, Michoacán, Tlaxcala y Hidalgo.

Tabla 1 Evolución del porcentaje de la pobreza y pobreza extrema nacional.

Entidad Federativa	Evolución del Porcentaje de Pobreza en México			
	2010	2012	2014	2016
Aguascalientes	38.1	37.8	34.8	28.2
Baja California	31.5	30.2	28.6	22.2
Baja California Sur	31.0	30.1	30.3	22.1
Campeche	50.5	44.7	43.6	43.8
Coahuila	27.8	27.9	30.2	24.8
Colima	34.7	34.4	34.3	33.6
Chiapas	78.5	74.7	76.2	77.1
Chihuahua	38.8	35.3	34.4	30.6
Distrito Federal	28.5	28.9	28.4	27.6
Durango	51.6	50.1	43.5	36.0
Guanajuato	48.5	44.5	46.6	42.4
Guerrero	67.6	69.7	65.2	64.4
Hidalgo	54.7	52.8	54.3	50.6
Jalisco	37.0	39.8	35.4	31.8
México	42.9	45.3	49.6	47.9
Michoacán	54.7	54.4	59.2	55.3
Morelos	43.2	45.5	52.3	49.5
Nayarit	41.4	47.6	40.5	37.5
Nuevo León	21.0	23.2	20.4	14.2
Oaxaca	67.0	61.9	66.8	70.4
Puebla	61.5	64.5	64.5	59.4
Querétaro	41.4	36.9	34.2	31.1
Quintana Roo	34.6	38.8	35.9	28.8
San Luis Potosí	52.4	50.5	49.1	45.5
Sinaloa	36.7	36.3	39.4	30.8
Sonora	33.1	29.1	29.4	27.9
Tabasco	57.1	49.7	49.6	50.9
Tamaulipas	39.0	38.4	37.9	32.2
Tlaxcala	60.3	57.9	58.9	53.9
Veracruz	57.6	52.6	58.0	62.2
Yucatán	48.3	48.9	45.9	41.9
Zacatecas	60.2	54.2	52.3	49.0

Fuente Estimaciones CONEVAL: base en el MCS-ENIGH 2010, 2012, 2014 y el MEC 2016 del MCS-ENIGH.

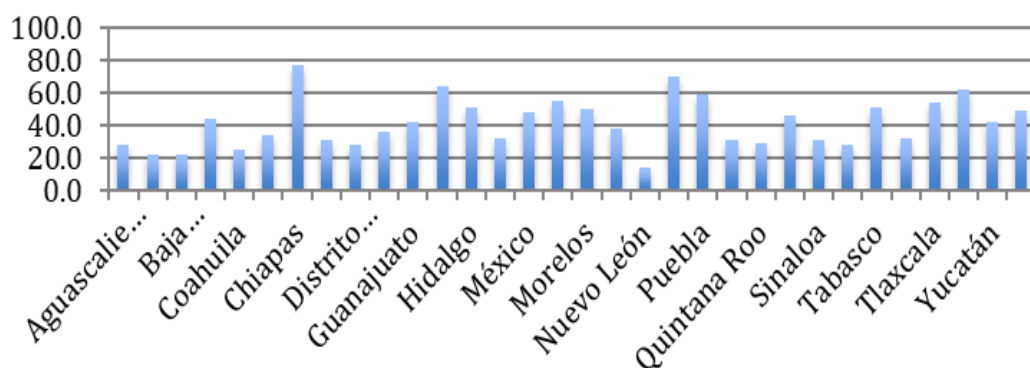


Figura 1 Índice de pobreza por entidad federativa 2016.

## Selección de Factores y Preprocesamiento de Datos (INEGI)

La base de datos seleccionada que contiene la recopilación de accidentes de tránsito fue la proporcionada públicamente por el INEGI, comprende registros de 1996 a 2016. La base de datos contiene un total de 42 campos [INEGI, 2016]. Sin embargo, para este estudio se seleccionaron solo 9 campos, tabla 2.

Tabla 2 Variables seleccionadas para análisis de accidentes.

#	Nombre del campo	Longitud	Tipo	Descripción del campo	Rango	Observaciones
1	Mes	2	Numérico	Mes de la información	1-12	Según mes de referencia
2	Hora	2	Numérico	Hora en que ocurrió el accidente	0 - 23	Clave 99.- No especificado
3	Diasemana	1	Numérico	Día del semana en que ocurrió el accidente	1-8	<ul style="list-style-type: none"> <li>• Clave 1.- Lunes</li> <li>• Clave 2.- Martes</li> <li>• Clave 3.- Miércoles</li> <li>• Clave 4.- Jueves</li> <li>• Clave 5.- Viernes</li> <li>• Clave 6.- Sábado</li> <li>• Clave 7.- Domingo</li> <li>• Clave 8.- No especificado</li> </ul>
4	Edad	2	Numérico	Edad del conductor presunto responsable	0-99	<ul style="list-style-type: none"> <li>• Clave 0 = Se ignora por que se fugó</li> <li>• Clave 99 = No especificado</li> </ul>
5	Causaacci	1	Numérico	Causa probable o presunta del accidente	1-5	<ul style="list-style-type: none"> <li>• Clave 1.- Conductor</li> <li>• Clave 2.- Peatón o pasajero</li> <li>• Clave 3.- Falla del vehículo</li> <li>• Clave 4.- Mala condición del camino</li> <li>• Clave 5.- Otra</li> </ul>
6	Caparod	1	Numérico	Capa de rodamiento	1-2	<ul style="list-style-type: none"> <li>• Clave 1.- Pavimentada</li> <li>• Clave 2.- No pavimentada</li> </ul>
7	Sexo	1	Numérico	Sexo del conductor presunto responsable	1-3	<ul style="list-style-type: none"> <li>• Clave 1.- Se fugó</li> <li>• Clave 2.- Hombre</li> <li>• Clave 3.- Mujer</li> </ul>
8	Aliento	1	Numérico	Aliento alcohólico	4-6	<ul style="list-style-type: none"> <li>• Clave 4.- Sí</li> <li>• Clave 5.- No</li> <li>• Clave 6.- Se ignora</li> </ul>
9	Cinturón	1	Numérico	Uso de cinturón de seguridad	7-9	<ul style="list-style-type: none"> <li>• Clave 7.- Sí</li> <li>• Clave 8.- No</li> <li>• Clave 9.- Se ignora</li> </ul>

En este artículo consideramos que la temporalidad en que ocurren los accidentes, así como el tipo de accidente y su causa están estrechamente relacionados con la capa de rodamiento, los cuales pertenecen al rubro de infraestructura y el vehículo.

Además, elementos asociados al conductor como la edad, el sexo, si utilizaron o no el cinturón o si presentan aliento alcohólico.

Los valores de las variables (campos) descritos en la tabla 2 indican los valores que se han normalizado y categorizado con el fin de establecer intervalos que describan cuándo, cómo, y bajo qué condiciones suceden los accidentes de tránsito. Estos valores nos ayudan a establecer un marco de referencia para realizar el análisis mediante un agrupamiento jerárquico y contrastarlo por estados.

### **Análisis de Accidentes por Agrupamiento Jerárquico**

La tabla 3 muestra los resultados del agrupamiento jerárquico para cada uno de los estados analizados con dos elementos:

- Muestra el número de grupos (clústeres) que se han formado con respecto a cada una de las variables. Esto permite entender qué tanta similaridad existe entre los grupos de valores que se obtienen para cada una de las variables.
- El dendrograma resultante muestra cuál es el agrupamiento que conservan cada una de las variables estableciendo las jerarquías por grupos formados. Con ello es posible entender la relación que existe entre los grupos y establecer un análisis de la importancia que representan estas variables para cada uno de los estados.

## **4. Discusión**

Como se observa en la tabla 3 con base en los dendrogramas, es posible distinguir cuatro casos principales donde se involucran en los elementos presentes en un accidente de tránsito: vehículo, camino y conductor:

- La primera caracterización es referente a la infraestructura vial (caminos). En este caso se observa que las variables asociadas a este elemento son:
  - ✓ Tipo de accidentes.
  - ✓ Causa del accidente.
  - ✓ La capa de rodamiento.

Por lo que estas variables establecen grupos que sobresalen en estados colindantes con características comerciales similares.

Tabla 3 Agrupamiento jerárquico de accidentes por estado.

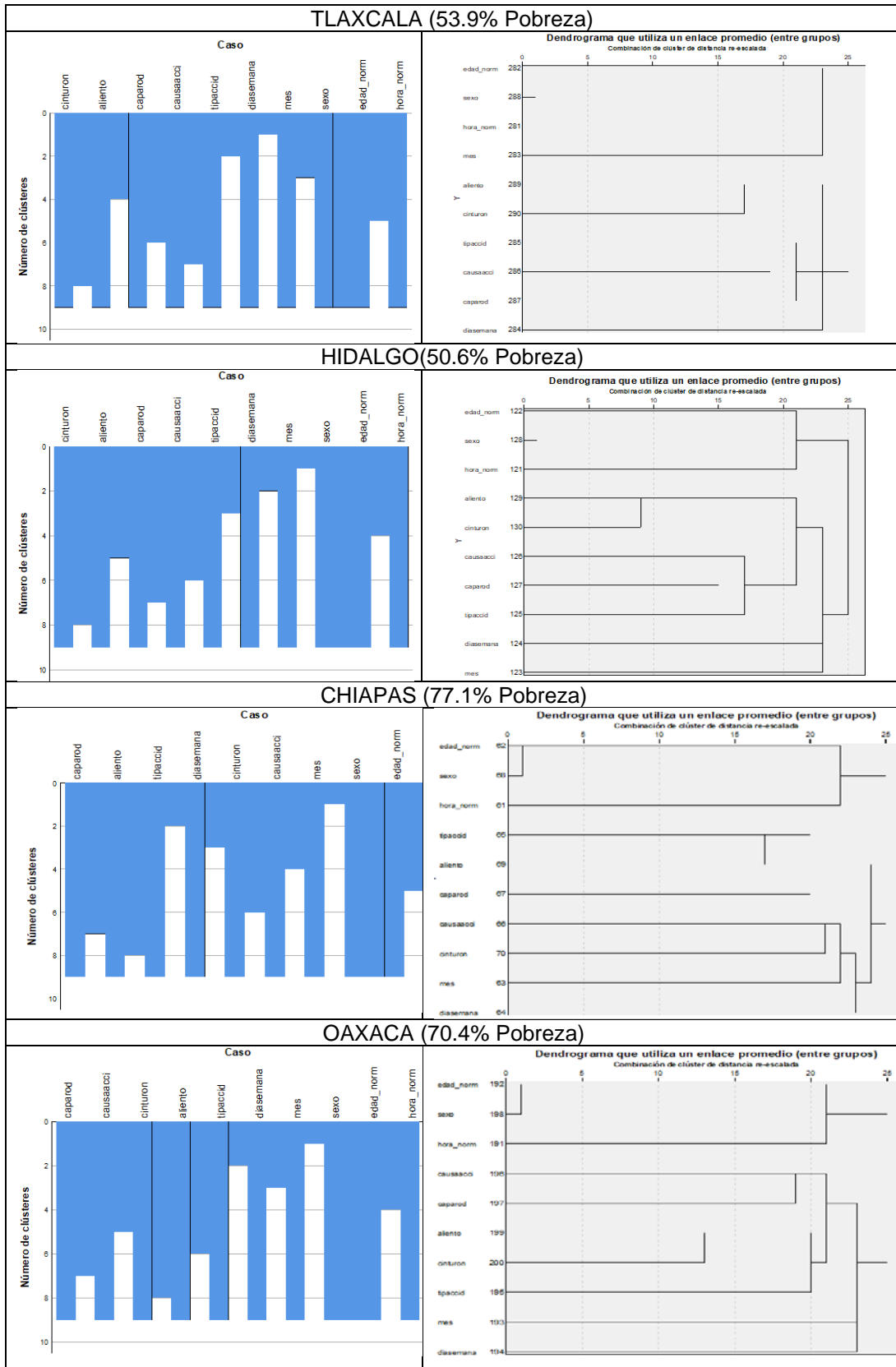
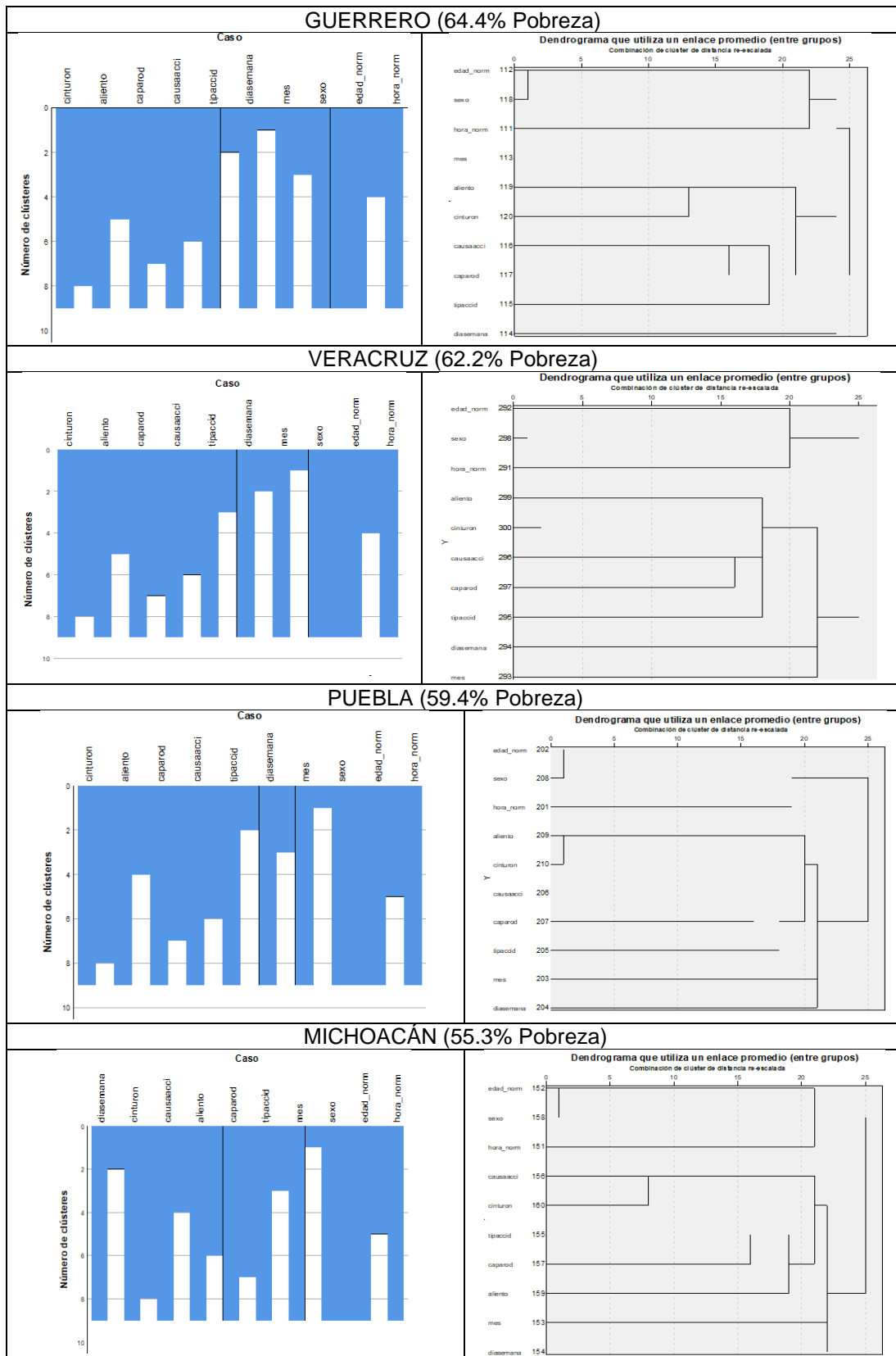


Tabla 3 Agrupamiento jerárquico de accidentes por estado (Continuación).



La agrupación refleja que existe una infraestructura de caminos deficiente en las zonas urbanas y suburbanas con que cuentan, tales como, Guerrero, Veracruz, Puebla, Tlaxcala e Hidalgo. Por otra parte, los estados de Chiapas y Michoacán generaron grupos donde el tipo de accidente, la capa de rodamiento y el aliento alcohólico representan un conglomerado que observa una relación con respecto a una orografía que presenta condiciones similares.

- En cuanto a las condiciones del vehículo y conductor se asocian las variables:
  - ✓ Aliento alcohólico.
  - ✓ Cinturón.
  - ✓ Tipo de accidente.

Con la excepción de Chiapas y Michoacán el resto de los estados presentan grupos donde el uso del cinturón y el aliento alcohólico se encuentran presentes. Además, Veracruz y Puebla presentan agrupaciones muy similares en cuanto a estas dos variables. Esto resulta interesante ya que comparten destinos turísticos y la cantidad de personas que se desplaza entre ambos estados es muy alta por el tipo de servicios comerciales y gente que se migra en busca de oportunidades laborales. Por otro lado, para el caso de Michoacán y Chiapas, el tipo de accidente y el aliento alcohólico resultaron con grupos donde se aglomeran estos accidentes, esto aunado a su orografía similar nos da un indicativo de porque se agrupan de esta forma.

- Con respecto al sexo y edad se observa que en todos los estados existen entre estas variables, grupos donde existe una similaridad muy estrecha, así como la hora en que suceden aglutina las edades y el sexo. Esto resultará muy útil cuando se requieran establecer reglas de asociación en un futuro para conocer más detalles de cómo se generan estos accidentes.
- En cuanto a la temporalidad en el caso de mes y día de la semana, se presenta una relación con el resto de las variables similar en casi todos los estados, conformando un grupo y de ahí se desprenden su unión con las distintas condiciones en cómo suceden los accidentes. Los estados de Chiapas y Guerrero presentan una variación en este grupo de temporalidad. En el caso particular de Chiapas la hora está asociada con el grupo conformado por el

sexo y la edad; y en cuanto al mes, esté se relaciona con la causa del accidente y si se tenía puesto el cinturón. Para Guerrero el mes se asocia con respecto a qué hora sucede el accidente, junto con las características del sexo y la edad de los involucrados en el accidente.

## **5. Conclusiones**

El caracterizar los accidentes mediante un agrupamiento jerárquico ha permitido dimensionar las variables y su interacción para la toma de decisiones por parte de las autoridades correspondientes. Esto se puede observar cuando se tienen grupos muy cercanos en los factores que caracterizan la infraestructura, representando a los caminos, entonces las autoridades deberán realizar una inspección de cómo se encuentran y realizar levantamientos en esos puntos, ya sea para invertir en una mejor infraestructura o evaluar si los sitios están presentando un alto riesgo debido a sus condiciones de diseño geométrico. Por otro lado, para los casos donde se observa que las variables asociadas al conductor y vehículo establecen grupos muy cercanos, la inversión debería ser en cuestiones preventivas o de señalización para concientizar a los conductores sobre su forma de conducir. El trabajo futuro de esta investigación es la aplicación de técnicas de inteligencia artificial que permitan generar reglas o asociaciones para un mejor entendimiento de cuándo y cómo están sucediendo estos accidentes en la república mexicana.

## **6. Bibliografía y Referencias**

- [1] Beshah, T., & Hill, S. (2010, March). Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia.
- [2] Castro Delgado, F. (2012). Análisis Espacial de los Accidentes de Tránsito en el Cantón de Pococí.
- [3] Chavarría Vega, J. M. (2001). Un Análisis de la relación del Factor Humano en la ocurrencia de accidentes viales. Instituto Mexicano del Transporte, Notas, (59), 1-5.
- [4] Chias L., Reséndiz D, & Martínez A, (2014). Análisis de la base de datos estadísticas de accidentes de tránsito terrestre en zonas urbanas y



suburbanas de México CISEV. IV Congreso Ibero-Americano de Seguridad Vial: "Avances de la década de acción de seguridad Vial. Juntos Podemos salvar vidas", Instituto Vial Ibero-Americano.

- [5] CONEVAL, (2017). Informa la evolución de la pobreza 2010-2016. Comunicado de prensa No. 09 Ciudad de México 30 de agosto de 2017.
- [6] K. Geurts, G. Wets, T. Brijs, K Vanhoof, "Clustering and Profiling Traffic Roads by Means of Accident Data", Proceedings of the European Transport Conference, 2003.
- [7] INEGI, (2017). Características de las Defunciones Registradas en México Durante 2017.
- [8] Martinelli, F., et al. (2018). Cluster Analysis for Driver Aggressiveness Identification. pp. 562-569.
- [9] Peden, M. (2004). Informe mundial sobre prevención de los traumatismos causados por el tránsito (No. 599). Pan American Health Org.