

Modelo de clasificación acorde a parámetros de atención primaria de adultos mayores en el municipio de Saltillo, Coahuila.

José Luis Fraga Almanza^{1*} José Refugio Reyes Valdés¹

Rina Betzabeth Ojeda Castañeda¹

Ma. del Pilar Alonso Reyes² José Antonio Flores Díaz²

¹Centro de Investigación en Matemáticas Aplicadas, UAdeC

²Facultad de Ciencias, UNAM

*jfraga@cima.udc.mx

Resumen

La magnitud e importancia del crecimiento de la población de los adultos mayores a nivel mundial, ha obligado a los gobiernos a establecer políticas públicas idóneas para este sector de la población, como es el caso de políticas para el sistema de salud y la integración plena de las personas mayores como participantes activos en el desarrollo social. En este documento se presenta un modelo estadístico de clasificación que permite identificar las variables más significativas que presenta un adulto mayor en el caso de una urgencia médica. Con la identificación el modelo clasifica cuales son los principales síntomas, a partir del registro histórico de dichas variables. La información histórica de éstas fue obtenida a partir de los registros del FRAP de Patronato de Cuerpo de Bomberos en el período del 2013 al 2014 en la atención de urgencias de adultos mayores del municipio de Saltillo Coahuila, México. Adicionalmente se presenta las características del modelo y los fundamentos teóricos de éste, así como los resultados obtenidos con su aplicación al caso señalado.

Palabras clave: Vejez; Agrupamiento; Modelo estocástico; Modelo de clasificación;

1 Introducción

En el municipio de Saltillo Coahuila al igual que en otros municipios del país los adultos mayores están siendo una base creciente de la población total. Esta situación está impulsando en su sociedad actual cambios de gran magnitud, especialmente, con relación a la plena atención de sus adultos mayores en aspectos de salud y en la integración de las personas mayores como participantes activos en el desarrollo de la sociedad.

Para lograr este objetivo es necesario contar con mecanismos que permitan obtener informa-

ción relevante sobre la salud y la atención médica en este sector de la población. Uno de estos mecanismos sería recabar la información necesaria para conocer los principales síntomas de los adultos mayores de Saltillo ante una urgencia médica. Para ello, además de analizar la estructura demográfica poblacional de dicho sector, es necesario hacer un análisis estadístico de las principales variables que representan el estado clínico de un paciente atendido en una urgencia médica, con el fin de que los paramédicos que atienden la urgencia cuenten con información de relevancia en la atención del adulto mayor. Además con esta información es posible establecer un modelo

estocástico de clasificación de las variables más significativas.

Los modelos estocásticos de clasificación en general permiten agrupar un conjunto de datos en clases, utilizando técnicas de la minería de datos. En este trabajo los datos a agrupar corresponden a variables que representan los síntomas seleccionados por el médico o paramédico que atendió la urgencia médica. Para el diseño y desarrollo de este tipo de modelos se aplica la metodología estadística relacionada con el análisis multivariado y la correlación de variables. El éxito de estos modelos de pende de que se tenga información confiable de las variables, motivo por el cual en el caso que nos ocupa fue necesario desarrollar un sistema de cómputo móvil que permitiera registrar los datos de manera confiable y segura, para tener un histórico confiable y estructurado. Este sistema no será presentado en este trabajo para mayor información ir, Almanza [1].

2 Descripción del problema

Actualmente no existe un modelo estocástico que a partir de parámetros clínicos de un paciente obtenga una clasificación y jerarquización de los síntomas que presenta a partir de la información histórica de un adulto mayor en una situación de urgencia médica. Como por ejemplo desmayo, caídas, comas diabéticos, problemas vasculares, entre otros.

En la ciudad de Saltillo la atención de las urgencias médicas le corresponde atenderlas al Patronato del Cuerpo de Bomberos, quienes registran el estado clínico del paciente para posteriormente canalizarlos a los diferentes hospitales de la ciudad. Por este motivo a los médicos o paramédicos que atienden las urgencias les serviría de ayuda contar con una clasificación de los principales síntomas.

3 Marco teórico

El *clustering* organiza los datos (un conjunto de síntomas como un vector de mediciones) extrayendo estructuras subyacentes. El agrupamiento finaliza cuando los patrones dentro de un grupo (*cluster*) son lo más similares entre sí que con otros patrones que pertenecen a otros grupos diferentes. Por lo tanto, organizar los datos usando *clustering* se emplea alguna medida de similitud entre los conjuntos de patrones. El *cluster* se define con base en los datos y propósito del análisis, Eldén [3].

El agrupamiento de datos o simplemente *clustering* es considerado uno de los problemas más importantes en el aprendizaje no supervisado, es decir, como en todos los problemas de este tipo, lo que se trata de hacer es encontrar grupos en un conjunto de datos. De manera más general puede definirse el agrupamiento de datos como *el proceso de organización de objetos que son muy similares de alguna manera*, ver figuras 1, 2.

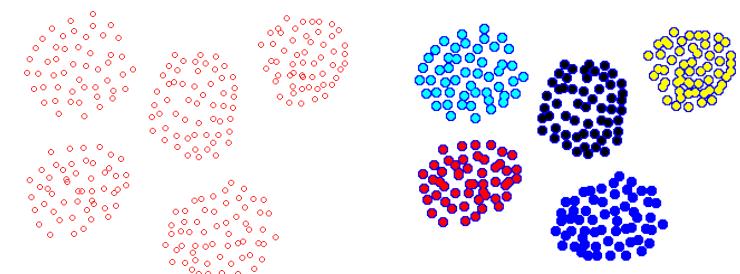


Figura 1:
Conjunto de datos.

Figura 2:
Agrupamiento.

El agrupamiento de datos a menudo suele confundirse con la clasificación de datos, en donde los objetos son asignados a clases ya establecidas. Es decir, en este tipo de problemas las clases también tienen que ser definidas, Gan et al. [4].

Este análisis trata de resolver el problema siguiente.

Definición 1 *Dado un conjunto de datos D con n objetos, una partición de D en k grupos es una familia de conjuntos $\{\pi_1, \pi_2, \dots, \pi_k\}$ que cumple con las siguientes propiedades:*

1. $\pi_i \neq \emptyset, \quad i = 1, \dots, k$

2. $\bigcup_{i=1}^k \pi_i = D$

3. $\pi_i \cap \pi_j = \emptyset \quad i \neq j$

Definición 2 *Dado un conjunto D de datos con n objetos en un espacio m -dimensional, particionar los mismos en k grupos tales que los puntos dentro de un grupo son más similares entre ellos que con otros grupos, dicha similitud se mide atendiendo a alguna función distancia.*

El problema de formar grupos es muy importante para el conocimiento del comportamiento de un conjunto de datos del cual solo se tiene una cantidad n de sus elementos. La solución a estos problemas se realiza mediante la aplicación de algoritmos de agrupamiento, o incluso la creación de nuevos algoritmos.

Definición 3 *Dado un conjunto de datos D y $\Pi = (\pi_l)_{l=1}^k$ una partición de D en k clases, decimos que π_l es un cluster si se cumplen las siguientes propiedades:*

1. *Sí x y y tienen mayor similitud $\Rightarrow x, y \in \pi_l$, donde $x, y \in D$*

2. *Sí x y y tienen menor similitud $\Rightarrow x \in \pi_l$ y $y \in \pi_m$, donde $x, y \in D$*

Se han propuesto diferentes técnicas o algoritmos para encontrar agrupamientos o clusters adecuados a diversos requerimientos. En general se pueden clasificar en jerárquicos y particionados.

Jerárquicos: Este agrupamiento es utilizado cuando no se tiene conocimiento de cuántos cluster o grupos existen. Producen una secuencia de particiones, juntando o separando clusters. En cada caso se juntan o separan dos clusters siguiendo algún criterio específico.

Particionados: Este agrupamiento realiza una división inicial de los datos en grupos y luego mueve los objetos de un grupo a otros según

se optimice alguna función objetivo. Estos algoritmos presuponen un conocimiento a priori del número de clusters en que debe ser dividido el conjunto de datos. Llegan a una división en clases que optimiza un criterio predefinido o una función objetivo.

4 Aplicación del modelo de clasificación y análisis

Ante la ausencia de un diagnóstico final en el FRAP (Formato para el Registro de la Atención Prehospitalaria de las Urgencias Médicas) Mexica [7], ya que solo se recaba la información del paciente hasta su traslado a un centro especializado, y no se da un seguimiento de las condiciones o causas determinadas de la urgencia médica, se desarrolló un modelo jerárquico del tipo no supervisado, James G. and Hastie T. [5], Lesmeister [6]. La clasificación se realizó en grupos como estrategia heurística para establecer categorías que sirvieron de base a especialistas para aportar información que sería utilizada como diagnóstico estimado.

Las principales variables seleccionadas para el análisis de clasificación por tipo se muestran en la Tabla 1.

Excluyendo las variables asociadas a la identificación del FRAP, tiempos de atención y la temperatura, ya que esta última contaba con 86 registros diferentes de cero de los cuales solamente cinco eran efectivos, fueron eliminadas del análisis; además que de acuerdo a los especialistas médicos la temperatura no es relevante, el resto de las variables fueron susceptibles de integrarse al modelo de predicción.

El total de registros en el segmento de adultos mayores fue de 855, en virtud de que en algunas variables no existían valores capturados para algún paciente específico. En este caso y dependiendo del tipo de variable se excluyeron los registros o bien se integraron como datos faltantes.

Antes de realizar la clasificación mediante un modelo jerárquico fue necesario llevar a cabo un

Tabla 1: Principales variables del análisis

Nombre	Registros
Num.FRAP	855
Estacion	855
Tiempos de atención	855
Motivo atención	788
Sexo	848
Edad	855
Origen probable	661
Nivel de conciencia	784
Via aerea	760
Ventilacion observada	774
Ventilacion auscultada	757
Circulacion: Presencia de pulsos	767
Circulacion: Calidad	691
Circulacion: Piel	757
Frecuencia respiratoria	565
Frecuencia cardiaca	640
TA Sistolica	619
TA Diastolica	619
SaO2	462
Temperatura	86
Glucosa	100
Alergias	570
Enfermedades	697
Condición del paciente	529
Prioridad	723

análisis de las correlaciones categóricas entre algunas de las principales variables. Con ello se obtuvo en primera instancia asociaciones entre ellas que permitieron establecer diagnóstico. En la primera subsección se presenta el estudio entre aquellas variables que se espera tengan fuerte interacción cuando se presentan urgencias médicas.

4.1. Correlaciones de variables principales

En esta sección se presenta el análisis de las correlaciones categóricas entre las variables: sexo, enfermedades, motivo de atención y edad, que presentaron una fuerte interacción.

4.1.1. Sexo vs Enfermedades

En adultos mayores el motivo de atención en una urgencia diferido por sexo es de relevancia,

Tabla 2: Frecuencias por sexo y enfermedad

	SI	NO
fem	58	297
mas	64	189

ya que las actividades por sexo suelen ser distintas, originando causas diversas.

Asociada a la edad del segmento de población que se estudia, la presencia de enfermedades contrastada con el sexo se consideró pertinente para su análisis. La Tabla 2 presenta las frecuencias observadas para estas dos variables.

Usando el lenguaje de programación R, la prueba de Fisher ($p\text{-value} = 0,007538$) y la de Pearson ($p\text{-value} = 0,008897$), mostraron que si existe una diferencia significativa en la presencia o no de enfermedad en relación al sexo en el caso de urgencias médicas. El resultado indica que tiende a haber más mujeres con enfermedad que hombres.

4.1.2. Sexo vs Motivo de atención (enfermedad o traumatismo)

Al contrastar la variable dicotómica sexo vs motivo de atención (enfermedad o traumatismo), se obtuvo la siguientes frecuencias presentadas en la Tabla 3.

Tabla 3: Frecuencias por sexo y motivo de atención

	enfermedad	traumatismo
fem	323	125
mas	278	56

Fuente: Elaboración propia

Bajo esta combinación la prueba de Fisher ($p\text{-value} = 0,0003037$) y χ^2 de Pearson ($p\text{-value} = 0,0003619$), arrojan una diferencia significativa. Esto quiere decir que, de acuerdo a las proporciones obtenidas, es más frecuente atender a personas del sexo femenino por traumatismo que las del sexo masculino.

Una de las posibles razones de que la emergencia asociada a mujeres es por traumatismo

radica en que a esa edad suele perder grandes cantidades de clacio y a que son más activas, esto se da particularmente en actividades asociadas al hogar .

4.1.3. Edad vs Motivo de atención (enfermedad o traumatismo)

Al contrastar las variables edad y motivo de atención se pudo identificar el rango de edades en el que el adulto mayor presenta mayor frecuencia en cuanto al motivo de la atención, ya sea por enfermedad o por traumatismo. Se obtuvieron las frecuencias que aparezcan en la tabla de frecuencias, ver Tabla 4.

Tabla 4: Frecuencias por edad y motivo de atención

	enfermedad	traumatismo
(60,65]	89	22
(65,70]	88	28
(70,75]	119	12
(75,80]	62	27
(80,85]	66	20
(85,90]	64	8
(90,95]	30	2
(95,100]	2	1
(100,105]	0	1
(105,110]	1	0

Fuente: Elaboración propia

En la prueba de χ^2 de Pearson el *p-value* indicó que el motivo de la atención si depende de la edad del adulto mayor, (*p-value*= 0,0004235). Siendo que la edad donde más frecuentemente se presenta por causa de enfermedad es la de 70 y 75 años, mientras que por traumatismo esta en las edades de 65 a 70 y 75 a 80 años de edad.

4.2. Resultados del modelo jerárquico de clasificación

Uno de los principales problemas de los registros de los FRAPs se presentó por que no contaban con un mecanismo de validación. Esta situa-

ción tuvo como resultado la anulación de datos no válidos. Después de la depuración y selección de variables que podían integrarse al modelo de clasificación, quedaron únicamente 700 registros. En la Figura 3 se representan la cantidad de datos válidos para cada variable.

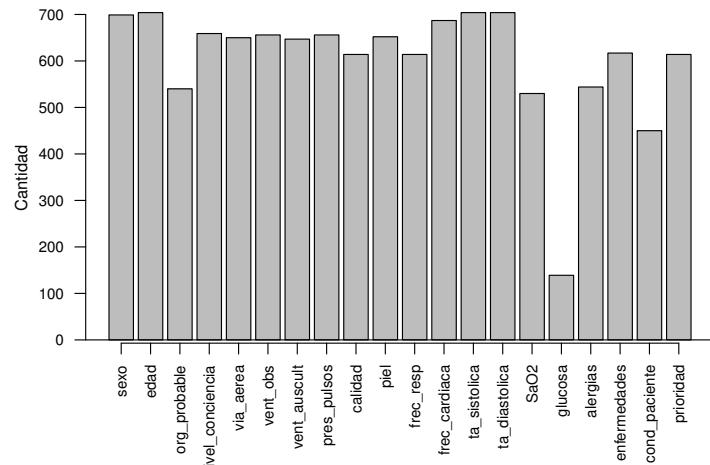


Figura 3: Datos válidos para cada una de las variables

La variable que menos datos útiles tuvo fue la glucosa. Por la importancia que tiene esta variable asociada al perfil regional, se integró inicialmente al modelo de clasificación aún y cuando con ello se redujo significativamente la cantidad total de datos válidos, ya que al eliminar registros de datos con al menos una variable con valor nulo, la combinación con otras incrementa la cantidad de datos eliminados.

En esta fase el modelo de clasificación fue de tipo no supervisado en el cual se incorporan variables cualitativas y cuantitativas. Para considerar el agrupamiento de este tipo de variables se propuso en este trabajo utilizar el método de PCAmix el cual se define a continuación, Chavane et al. [2].

Definición 4 *El PCAmix se define por la mezcla de las variables cuantitativas y cualitativas e incluye el análisis de componentes principales (PCA) y el análisis de correspondencias múltiples (ACM).*

- $V_1 = \{x_1, \dots, x_m\}$ variables cuantitativas
- $V_2 = \{z_1, \dots, z_m\}$ variables cualitativas

- Sea X y Z las matrices de datos cuantitativas y cualitativas respectivamente
- Sea $P = (C_1, \dots, C_k)$ una partición de $V = V_1 \cup V_2$
- La homogeneidad de esa partición P

$$H(P) = \sum_{k=1}^k H(C_k, y_k)$$

donde y_k es el centroide de C_k

- La función H mide el ajuste entre C_k y y_k

$$H(C_k, y_k) = \sum_{x_j \in C_k} r^2(x_j, y_k) + \sum_{z_j \in C_k} \eta^2(z_j, y_k)$$

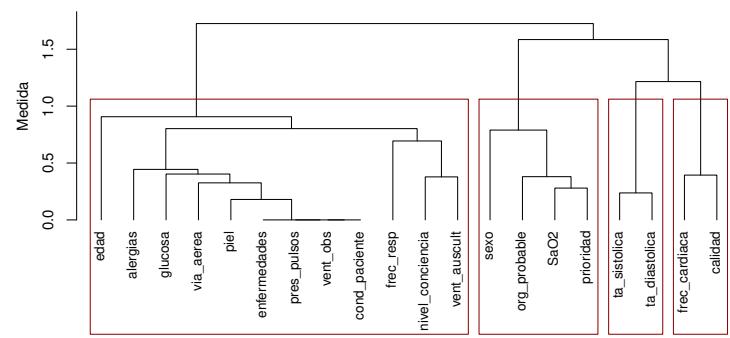
donde $r^2(x_j, y_k)$ es la correlación al cuadrado de x_j con y_k y $\eta^2(z_j, y_k)$ es la correlación ratio entre z_j y y_k

En el caso de la inclusión de la glucosa se utilizaron, en primera instancia, 20 variables para determinar los grupos correspondientes, 7 de las cuales son cuantitativas y 13 cualitativas. Las Figuras 4(a) y 4(b) muestran los grupos establecidos con resoluciones 4 y 5 respectivamente.

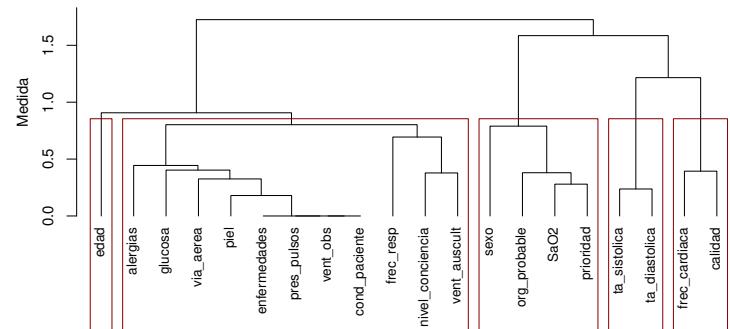
En el proceso de identificación de grupos se estableció una medida de importancia de cada una de las variables. La importancia es de naturaleza estadística ya que se asocia la medida en que cada una de las variables contribuye en la reducción del criterio de optimización, cada vez que es incorporada al árbol. El criterio establecido para la formación de grupos de las variables tipo cuantitativo fue la correlación al cuadrado del centro a cada una de las observaciones ($r^2(x_j, y_k)$) y para las variables cualitativas fue la relación de la correlación del centro a cada una de las observaciones ($\eta^2(z_j, y_k)$).

Una vez establecida la clasificación óptima (Primer componente principal de PCAmix), la importancia de las variables se muestra en forma ordenada de mayor a menor en la Figura 5.

La clasificación obtenida con el modelo propuesto no aporta información acerca de las posibles causas de los valores en los parámetros,



(a) Cluster con 4 categorías



(b) Cluster con 5 categorías

Figura 4: Clasificación jerárquica basada en parámetros principales

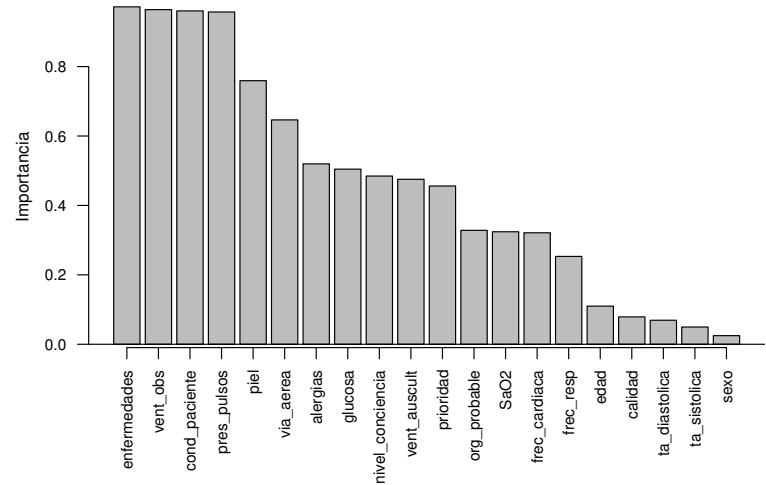


Figura 5: Orden de importancia de los parámetros principales

esto debido a que no se cuenta con una variable respuesta. Ante esta limitación, una vez seleccionada la cantidad de grupos de variables, se elabora un cuadro resumen de las características observadas en cada uno de estos grupos.

Para efecto de análisis se tomaron los cinco grupos de referencia que se presentan en la Figura 4(b). Dado que existen variables cualitativas y cuantitativas se incluyen de ambos tipos ya que esto aporta comportamiento de parámetros

en urgencias médicas que sirven de referencia a personal especializado en esa área de la medicina.

Las cuatro variables cuantitativas más importantes, de acuerdo a la clasificación jerárquica son glucosa, SaOs2, ta_sostolica y ta_diastolica. Las principales variables categóricas son enfermedades, prioridad, org_probable y sexo. Como referencia del comportamiento estadístico de cada una de estas variables en la Figura 6 se muestran los diagramas de caja para las cuantitativas y de barras para las categóricas; para estas últimas se muestran las frecuencias relativas.

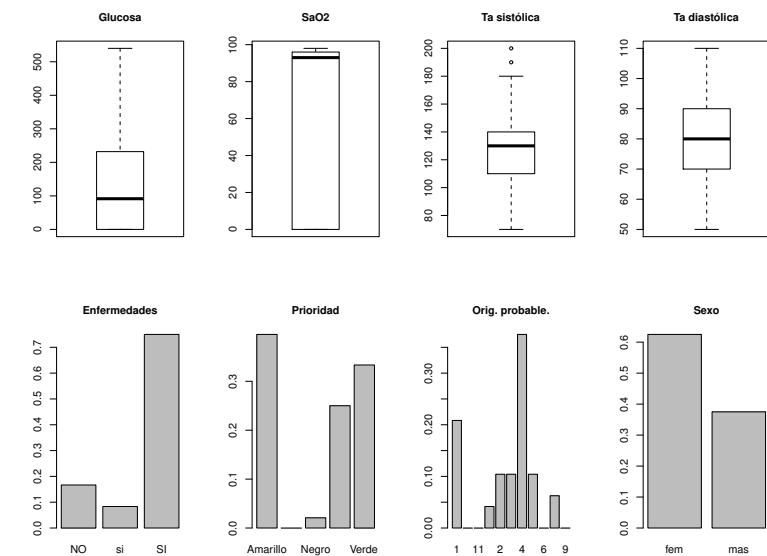
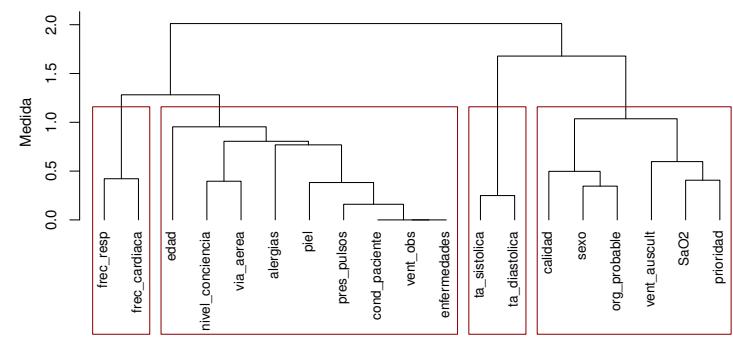


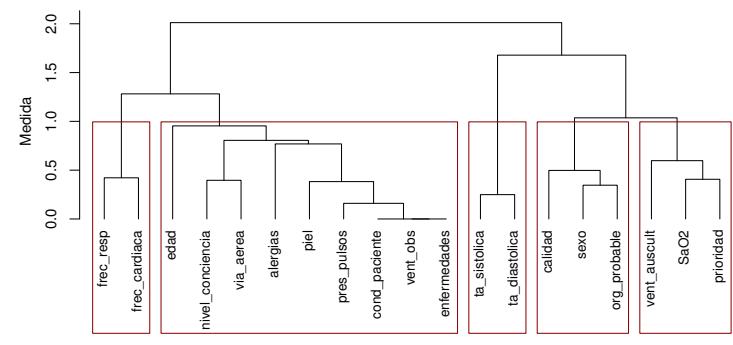
Figura 6: Variables cualitativas y cuantitativas más importantes

4.3. Clasificación sin la variable glucosa

Ante la máxima cantidad de datos anulados en la variable glucosa y con la combinación de datos nulos en otras variables, es pertinente jerarquizar sin incluir esta variable, para comparar resultados obtenidos con ella incluida, la Figura 7 muestra la clasificación jerárquica resultante. Como se observa en la Figura 7, la clasificación sin glucosa difiere sustancialmente de la obtenida al incluir esta variable. Comparando la calidad de la segmentación en grupos, sin glucosa, ver Figura 8, esta resulta mejor, ya que se incrementa sustancialmente la cantidad de datos utilizados.



(a) Cluster con 4 categorías



(b) Cluster con 5 categorías

Figura 7: Clasificación jerárquica (sin glucosa)

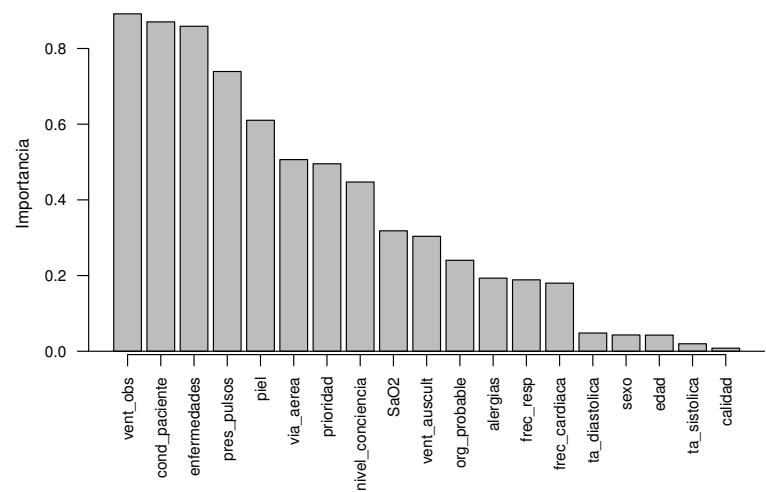


Figura 8: Importancia de parámetros principales (sin glucosa)

La deficiencia en el tamaño de la muestra en el caso de incluir glucosa y la influencia que ésta puede tener origina diferencias significativas en el orden de importancia, como se puede apreciar en las Tabla 5 y la Figura 8.

Tomando como referencia las cinco variables más importantes de cada una de las clasificaciones, las variables de origen probable y prioridad aparecen en ambos casos.

Si se consideran las cinco variables menos im-

Tabla 5: Comparación de importancia de variables con y sin glucosa

	CGLU	SGLU
1	enfermedades	vent_auscult
2	prioridad	org_probable
3	org_probable	piel
4	glucosa	vent_obs
5	SaO2	prioridad
6	sexo	calidad
7	vent_obs	nivel_conciencia
8	calidad	Sexo
9	cond_paciente	SaO2
10	nivel_conciencia	ta_diastolica
11	ta_sistolica	via_aerea
12	ta_diastolica	enfermedades
13	pres_pulsos	pres_pulsos
14	alergias	cond_paciente
15	piel	frec_cardiaca
16	frec_cardiaca	ta_sistolica
17	frec_resp	Edad
18	edad	Alergias
19	via_aerea	frec_resp
20	vent_auscult	

Fuente: Elaboración propia

portantes, se observa que las variables frecuencia cardiaca, frecuencia respiratoria y edad están presentes en ambos casos. Sin embargo existen cambios en el orden de significancia de las variables.

Dos cambios de orden significativo se aprecian, y el más drástico corresponde a la variable vent_auscult ya que al incluir la variable glucosa aparece como la menos importante y sin incluirla como la más importante. También la variable piel muestra un comportamiento similar aunque no tan drástico; con glucosa aparece en la posición 15 y sin glucosa en la posición 3.

5 Conclusiones

Con base a la información depurada de los FRAPs proporcionada por el Patronato del Cuerpo de Bomberos de Saltillo Coahuila(PCB), se pudo aplicar el modelo de clasificación con dos modalidades, incluyendo la variable glucosa presentándose diferencias significativas en ambos casos, las diferencias se presentaron dado que la variable glucosa no contaba con información suficiente para llevar un análisis confiable con dicha

variable.

Con el fin de que el modelo pueda ser aplicado en el trabajo de investigación se desarrolló un sistema de cómputo móvil que permite registrar de manera confiable y segura la información asociada al estado clínico del paciente en una situación de urgencia médica por parte de los paramédicos del PCB.

Referencias

- [1] José Luis Fraga Almanza. Sistema de monitoreo de alertas en urgencias médicas y modelo de clasificación acorde a parámetros de atención primaria de adultos mayores en el municipio de saltillo, coahuila. Master's thesis, Universidad Autónoma de Coahuila, 2016.
- [2] Marie Chavane, Vanessa Kuentz, Bnoit Liquet, and Jérôme Saracco. Clustofvar: an r package for the clustering of variables. In *The R User Conference, University of Warwick*, 2011.
- [3] Lars Eldén. *Matrix methods in data mining and pattern recognition*. SIAM, 2007.
- [4] Guojun Gan, Chaoqun Ma, and Jianhong Wu. *Data clustering: theory, algorithms, and applications*, volume 20. Society for Industrial and Applied Mathematics, 2007.
- [5] Witten D. James G. and Tibshirani R. Hastie T. *An introduction to statistical learning with applications in r*. Springer, 2013.
- [6] Lesmeister. *Mastering machine learning with r*. Packt Publishing, 2015.
- [7] Cruz Roja Mexicana. Formato para el registro de la atención prehospitalaria de las urgencias médicas (frap). <http://www.frapsedomex.org/>, 2006. Página consultada 10 julio 2014, 11:38 pm.