Universidad de Valladolid

Facultad de Ciencias

Trabajo Fin de Grado

Grado en Estadística

Mortalidad de Pacientes en Ventilación Mecánica: Influencia de la Traqueotomía

Autor:

Gonzalo Díaz Amor

Tutor:

Agustín Mayo Íscar

Cotutor:

Francisco Javier Carpena Moreno

Junio de 2020

ÍNDICE

Resumen

Muchos pacientes, sobre todo de la unidad de cuidados intensivos (UCI), suelen presentar necesidades respiratorias de carácter externo, es decir, de uso de un respirador. El respirador puede ser usado con un tubo endotraqueal introducido por la boca o bien mediante el procedimiento de la traqueotomía. El fin del estudio es analizar la asociación entre el uso de traqueotomía y la supervivencia de los pacientes.

El conjunto de datos inicial consta de 44.214 episodios de los 14 complejos asistenciales de la Comunidad de Castilla y León desde el año 2000 hasta el 2015, y disponemos de 59 variables de inicio (a veces compuestas, puesto que hay variables de identificación cuyo origen puede ser la concatenación de otras variables). En primer lugar, se realiza un filtro de los datos de los que disponemos, tanto a nivel estadístico, como son los datos anómalos, como a nivel sanitario, como puede ser edad del paciente o si presenta patologías no relevantes para el estudio. En segundo lugar, creamos variables categóricas de interés clínico y patológico, y posteriormente debemos realizar un estudio para identificar variables relacionadas con la realización de la traqueotomía. Para ello, realizaremos una regresión logística con variable respuesta traqueotomía. Por la gran cantidad de variables explicativas estudiaremos diferentes métodos de regularización de variables basados en la regresión logística, prestando atención a las tasas de acierto. Realizada dicha regresión, nos dispondremos a estimar el índice de propensión, que nos permitirá comparar a las dos poblaciones de forma directa y objetiva, sin realizar ninguna transformación para las variables de confusión. A continuación, con el índice de propensión podremos emparejar los episodios respecto al tratamiento de traqueotomía y estudiar el efecto que tiene dicho tratamiento en esos emparejamientos en la evolución del paciente. Por último, compararemos cuatro métodos de regularización de variables para el estudio de la mortalidad.

Palabras clave: Traqueotomía, regresión, métodos de regularización, matriz de confusión, índice de propensión, ventilación mecánica.

Abstract

Many patients, especially in the intensive care unit (ICU), usually have external respiratory needs, they need to use the ventilator. The ventilator can be used with an endotracheal tube or performing a tracheotomy. The aim of the study is to analyse the association between using tracheotomy and the survival of the patients.

The set of initial data consists of 44,214 episodes, which have been gathered from the 14 healthcare complexes in the Community of Castilla y León from 2000 to 2015. We also have 59 initial variables which sometimes are compound, since there are identification variables whose origin can be the concatenation of other variables. Firstly, a statistics and health related screening of the data are performed. Therefore, we shall

rule out anomalous data or the of the patient, as well as if the patients present pathologies not relevant to the study. Secondly, we create categorical variables which pose interest at a clinical and pathological level, and afterwards, we must carry out a study to identify variables related to the performing of a tracheotomy. To that aim, we will carry out a logistic regression in which the response variable is tracheotomy. Due to the great extent of explanatory variables, we will study different methods of regularization of variables based on the logistic regression. In that process we will also pay attention to success rate. Once the logistic regression is made, we can estimate the propensity score which will allows us to compare the two populations directly and objectively. We will not need to perform any transformation for the confounding variables. Next, bearing in mind the propensity score, we will be able to match the episodes regarding the tracheotomy treatment, as well as to study the effect that this treatment has on those pairings in the patient's development. Finally, we will compare four regularization models of variables in order to study the mortality.

Key words: Tracheotomy, regression, regularizations models, confussion matrix, propensity score, ventilator.

1. INTRODUCCIÓN

A lo largo del siglo XX, el avance en la medicina en occidente ha sido mayor que en ningún otro periodo de tiempo a lo largo de la historia, estos cambios han sido siempre demostrados por estudios estadísticos sin los cuales no sabríamos qué patologías son las más dañinas y qué métodos o procedimientos son más efectivos a la hora de paliar las enfermedades de una forma objetiva. Podemos mencionar a un personaje histórico como Florence Nightingale, considerada precursora de la enfermería profesional moderna, aunó la rama de la medicina y la estadística a finales del siglo XIX cuyos estudios fueron los primeros en demostrar durante la guerra de Crimea que las terribles condiciones de los heridos en los hospitales de campaña influían en la recuperación en los pacientes y que por tanto era fundamental una correcta higiene para evitar mayores patologías. Es un ejemplo de los múltiples estudios estadísticos que se han realizado en el campo de la medicina.

Es de interés conocer el valor de la traqueotomía en individuos hospitalizados que precisan ventilación mecánica. Entendemos como vía aérea respiratoria humana, la conexión que existe entre la nariz y la boca con los pulmones. Su conocimiento y manipulación fueron reflejados por primera vez de manera documentada desde el año 2000 a.C. en el libro “Rig Veda” de origen hindú y, posteriormente, también aparece en el “Papiro de Ebers” que data del 1500 a.C.

Por otro lado, la traqueotomía es definida por la Real Academia Española como la “abertura artificial de la tráquea para evitar la asfixia”. Los inicios de la intubación endotraqueal se atribuyen a Hipócrates (Siglo V a.C.) al realizar la primera intubación de introducción de un tubo metálico en la tráquea de un paciente. Sin embargo, no es hasta el siglo XVIII cuando se populariza el término de traqueotomía por el profesor alemán Lorenz Heister, y es durante el siglo XIX, debido a la epidemia de difteria (enfermedad infecciosa aguda, provocada por un bacilo, que afecta a la nariz, la garganta y la laringe, produciendo fiebre y dificultad para respirar) cuando dos cirujanos franceses describen la realización de más de 200 procedimientos consiguiendo que 50 de los pacientes tratados sobrevivan. A finales del siglo XIX, el pediatra Joseph O’Dwyer presenta una técnica de intubación cuyo éxito es notorio en casos de difteria, siendo considerado gracias a este logro uno de los padres de la intubación. Ya en el siglo XX, Chevalier Jackson describe en el año 1909 la técnica de la traqueotomía, estandarizando el instrumental necesario y estableciendo las indicaciones para su realización.

El mantenimiento prolongado de la vía aérea de forma artificial ya sea mediante intubación endotraqueal o por traqueotomía, puede acarrear una serie de complicaciones. Las dos técnicas comportan tanto ventajas como desventajas y, dependiendo de la situación de cada paciente, se requerirá una técnica u otra.

En el presente trabajo tenemos interés en evaluar la relación entre la traqueotomía y la mortalidad. Todos los pacientes de los datos a analizar están sometidos a ventilación mecánica, de los cuales a un subgrupo se les ha practicado una traqueotomía. El archivo de datos del que disponemos proviene del CMBD (Conjunto Mínimo Básico de Datos) de Castilla y León del año 2000 al año 2015 el

cuál disponemos de episodios de pacientes el estudio cuyo episodio ha necesitado ventilación mecánica durante más de 96 horas.

Por último, hay que señalar que definiremos un episodio como la atención que recibe en un momento un paciente. Esto significa que un paciente puede tener diferentes episodios por distintas razones a lo largo del periodo de estudio.

2. HIPÓTESIS Y OBJETIVO

La hipótesis de partida es que el uso de la traqueotomía en pacientes puede estar relacionado con la mortalidad, sin embargo, son muchas las variables muy sensibles que afectan a la decisión de cuándo realizar el tratamiento.

Además, como no se ha realizado un ensayo clínico, la recopilación de los datos y su posterior estudio son fundamentales por la falta de aleatorización de los estudios observacionales en el caso que nos ocupa.

Por tanto, el objetivo del estudio es determinar lo influyentes que son los diagnósticos del paciente para realizar la traqueotomía, y si este procedimiento es decisivo en la mortalidad del paciente durante el episodio clínico.

3. METODOLOGÍA

3.1. Ámbito del estudio

El estudio realizado se lleva a cabo con los datos de los 14 centros asistenciales de Castilla y León recopilados entre el año 2000 y el año 2015. Es usado el CMBD para la obtención normalizada de los datos clínicos y administrativos que se derivan de la asistencia al paciente en régimen de hospitalización.

3.2. Protocolo de actuación

Cada episodio es codificado con los datos de ingreso y guardado para su posterior recopilación e interpretación. Las variables constitutivas del CMBD serán:

Datos de identificación:
1. Tipo de actividad sanitaria
2. Identificación del centro
3. Número de historia clínica
4. Identificación del paciente
5. Fecha de nacimiento
6. Sexo
7. Municipio
8. Código Postal
9. Zona Básica de Salud

Datos no clínicos:
10. Fecha de contacto
11. Fecha de ingreso
12. Número de autorización
13. Financiación de la asistencia sanitaria
14. Circunstancias al ingreso
15. Procedencia del ingreso
16. Identificación del centro de procedencia
17. Petición del contacto
18. Identificación del centro de petición
19. Fecha de intervención quirúrgica
20. Fecha de alta
21. Identificación del Servicio
22. Identificación de la Sección
23. Identificación del médico responsable
24. Circunstancias al alta
25. Destino tras el contacto
26. Identificación del centro de traslado

Datos clínicos:
27. Diagnóstico principal
28. Diagnósticos secundarios
29. Procedimientos diagnósticos y terapéuticos
30. Identificación de las causas externas de enfermedad (códigos E)
31. Identificación de la morfología de las neoplasias (códigos M)

En los datos clínicos, tanto en el diagnóstico principal como en los diagnósticos secundarios, se utiliza como medida de homogeneidad el estándar CIE-9 (*Clasificación Internacional de Enfermedades, novena edición*) cuyo fin es catalogar las enfermedades, afecciones y causas externas de enfermedades y traumatismos con objeto de recopilar información sanitaria útil relacionada con la mortalidad y morbilidad. La mortalidad es definida por la RAE como “Tasa de muertes producidas en una población durante un tiempo dado, en general o por una causa determinada” y morbilidad como “la proporción de personas que enferman en un lugar y un periodo de tiempo determinados”.

Los procedimientos están definidos como “Una actividad dirigida o realizada en un individuo con el objetivo de mejorar la salud, tratar enfermedades o lesiones o hacer un diagnóstico" según el ‘International Dictionary of Medicine and Biology’. También utilizamos como medida clasificatoria el estándar CIE-9.

Para globalizar lo que ocurre con cada paciente, usamos el estándar GRD (*Grupo Relacional de Diagnóstico*). El sistema de clasificación de pacientes GRD es una herramienta de gestión normalizadora que utiliza el sistema CMBD para clasificar a los pacientes en grupos clínicamente similares y con parecido consumo de recursos sanitarios.

3.3. Variables analizadas

Las variables analizadas corresponden a los diagnósticos, tanto principal como los secundarios, y procedimientos que se realizaron en los episodios a los pacientes. La variable respuesta es lo que se denomina *exitus*, palabra proveniente del latín que

se emplea en medicina haciendo referencia a si la enfermedad ha desembocado en el fallecimiento del paciente durante su estancia hospitalaria. En un principio tomaremos como variable respuesta la *Traqueotomia* para poder realizar el estudio de su influencia en el desarrollo del episodio. El p-valor es respecto a la variable *Traqueotomia* y *exitus* (ver Tabla 5.2).

Se realiza la prueba de Chi-Cuadrado para ver si hay asociación entre la variable Traqueotomía con el resto de las variables de carácter clínico, obviando variables de identificación. Una variable la tomaremos como influyente cuando su p-valor sea menor que 0.2.

Las variables consideradas como 'antes del alta' las denominamos 'a priori' y son las siguientes: *edad, tiping, Traqueotomia, CARD, RESP, NEFRO, HEPAT, InmunoDef, TmDM, EnfNM, ObsVAeS, IRA, REPOC, Gripe, BrA, SDRA, TEP, Asma, BE, ICA, IM, MIO, TrTo, Coma, ACVA, TrCr, TrCC, EsEp, Shock, Quemaduras, Intox, PrDigA, AP01, AP02, AP03, AP04, AP05, AP06, AP07, AP08, NeAVM01, NeAVM02, NeAVM03 y EdadCat*. Por otro lado, agrupamos otro tipo de variables que están muy correlacionadas con las primeras, pero no son 'a priori' o 'antes del tratamiento' que denominaremos como 'a posteriori', como puede ser 'ComTraq' es decir, si ese paciente presenta complicaciones durante la traqueotomía. Este grupo de variables no las podremos usar para realizar nuestra regresión logística. Finalmente queda el grupo de variables que consideramos 'a posteriori', que son aquellas que no se pueden estudiar hasta que se da el alta al paciente, como por ejemplo la variable *estancia*. El conjunto de todas las variables se puede consultar en el anexo 1.

3.4. Depuración de los datos

Tras consultar con los facultativos, se eliminaron los episodios, donde la edad del paciente fuera menor de 16 años; la traqueotomía durase menos de 96 horas (código CIE-9 96.71 tanto en diagnóstico principal como en secundario) o fuera permanente (CIE-9 31.29), salvo que también incluyera traqueotomía temporal (CIE-9 31.1), donde el episodio acabara en traslado ya que no podríamos averiguar el tipo de alta del paciente; y estancias cuya duración fuera mayor que el cuantil 0.99 o no dispusiésemos del valor, en éste último caso lo cambiamos por 'Q' .

La siguiente acción que realizamos, por indicación de los facultativos, es transformar el tipo de alta *voluntaria* (tipalta=3) a *domicilio* (tipalta=1). Además, uniremos los traslados identificados por el *id paciente*, fechas coincidentes y alta en un hospital dentro del marco del estudio en un solo episodio, manteniendo el diagnóstico principal del primer episodio, el tipo de alta del último episodio, sumando el número total de días en estancia y seleccionando los diagnósticos secundarios usando un máximo de 10 diagnósticos. En el caso de no disponer de información de carácter relevante de un episodio (tipo de alta o si se le ha realizado la traqueotomía o no) se excluye del conjunto de datos.

La simplificación de los códigos utilizados para los diagnósticos (agrupación) respecto a los CIE-9 la podemos observar y consultar en el anexo 2.

3.5. Pacientes

De 2000 a 2015 se han registrado en total 44214 episodios con ventilación mecánica de los cuales 28212 son excluidos por la depuración de datos señalada

anteriormente. Su distribución se representa en la Tabla 1 que se detalla a continuación. De menor a mayor volumen de datos excluidos tenemos 4 cuyo valor de estancia hospitalaria es 'Q', 67 cuyo procedimiento diagnóstico es traqueotomía permanente (desde p1 a p7), 149 cuya estancia es mayor que el cuantil 0.99 muy probable traqueotomía permanente, 245 traslados a otra comunidad, 1535 episodios por estancia menor de 5 días y 22632 episodios por contener el código 96.71 en cualquiera de los apartados p1-p8, no habiendo en p8 ningún caso con traqueotomía permanente que corresponde a ventilación mecánica invasiva continua inferior a 96 horas consecutivas.

ETIQUETAS DE FILA	**NÚMERO DE CASOS EXCLUIDOS**
TRAQUEOTOMÍA PERMANENTE EN P7	3
ESTANCIA VALOR 'Q'	4
TRAQUEOTOMÍA PERMANENTE EN P5	5
TRAQUEOTOMÍA PERMANENTE EN P4	6
TRAQUEOTOMÍA PERMANENTE EN P6	7
TRAQUEOTOMÍA PERMANENTE EN P2	10
TRAQUEOTOMÍA PERMANENTE EN P3	13
TRAQUEOTOMÍA PERMANENTE EN P1	23
ESTANCIA MAYOR QUE 134 DIAS (CUANTIL .99) MUY PROBABLE TRAQUEOTOMÍA	149
TRASLADOS FUERA DE LA COMUNIDAD	245
CONTIENE 96.71 en procedimiento 8	1072
ESTANCIA < 5 DIAS	1535
CONTIENE 96.71 en procedimiento 7	1609
CONTIENE 96.71 en procedimiento 6	2408
CONTIENE 96.71 en procedimiento 1	3183
CONTIENE 96.71 en procedimiento 5	3205
CONTIENE 96.71 en procedimiento 3	3232
CONTIENE 96.71 en procedimiento 4	3412
MENORES 16	3580
CONTIENE 96.71 en procedimiento 2	4511
Total general	**28212**

Tabla 3.1: Distribución de los casos excluidos.

Finalmente, tras agrupar los traslados que pueden seguir siendo estudiados, utilizamos para nuestro estudio 13555 episodios, que podemos dividir en 10462 casos sin traqueotomía y 3093 casos con traqueotomía. De los 10462 casos sin traqueotomía 5688 tuvieron un alta diferente al *exitus* y 4774 fallecieron; y en los casos con traqueotomía 1508 tienen un alta diferente al *exitus* y 1585 fallecieron. En lo que respecta a la estancia a nivel numérico observamos en la Figura 1 cómo es mayor tanto la media como la varianza en los pacientes con traqueotomía.

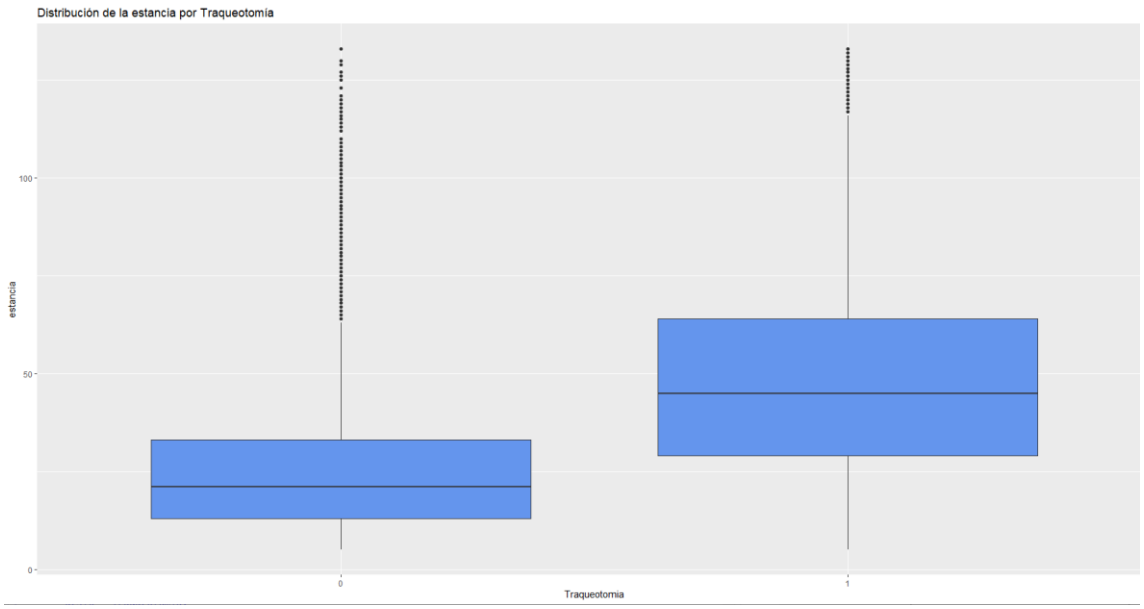

Figura 3.1: Distribución de la estancia por traqueotomía.

3.6. Análisis

3.6.1. Análisis bivariado

Realizamos un filtrado de variables que no se espera que tengan relación biológica, como puede ser el código postal o la identificación del paciente, tanto de la historia clínica como de la codificación que hemos propuesto, para hacer el conjunto de episodios de un mismo paciente que el tipo de alta ha sido traslado dentro de nuestra comunidad, para seguir el seguimiento de ese paciente. Además, hay que tener presente que las variables pueden ser recopiladas previos del acto médico y una vez realizada el alta del paciente. A continuación, vamos a trabajar con variables que se conocen al ingreso. Las variables que son conocidas al alta nos las podemos usar, las eliminamos del análisis, para poder realizar el resto del estudio de forma correcta.

A la hora de describir las variables continuas es necesario calcular su media y su desviación estándar. Por otro lado, en el caso de las variables de carácter categórico debemos describirlas expresando su frecuencia absoluta y relativa junto con su porcentaje. Para detectar asociación entre las variables del ingreso y la realización de traqueotomía, en el caso de las variables categóricas, se describirán comparando los porcentajes y también obtendremos la prueba $\chi 2$ de Pearson o Fisher para tablas 2x2 si las frecuencias esperadas son menores que 5. Por último, para identificar relaciones entre variables numéricas y la realización de traqueotomía, el test t de Student.

3.6.2 Regresión logística

La regresión logística es una técnica analítica para explicar una variable dicotómica en función de un conjunto de variables independientes. Es especialmente útil por la gran cantidad de información que nos proporciona el cálculo de los coeficientes, ya que estos tienen la interpretación de logaritmos de odds ratio. Podemos describirla como una regresión lineal con la diferencia de que la variable respuesta es binaria y, por tanto, suponemos que la variable respuesta sigue una distribución binomial.

Con este modelo de regresión logística pretendemos estimar la probabilidad de efectuar la traqueotomía bajo la premisa de buscar el más parsimonioso, es decir, aquel que con el menor número de variables posibles obtengamos la predicción más válida.

3.6.2.1. Validación cruzada

En primer lugar, tomamos una muestra de entrenamiento del 80% del tamaño total, que estará destinada a estimar los parámetros del algoritmo que vamos a utilizar que en este caso será la regresión logística. De esa muestra realizamos una validación cruzada de 10-fold (cross validation k-fold) para estudiar la regresión con selección de variables entre las que se encuentra Ridge, Lasso (Least Absolute Shrink-age and Selection Operator, -operador de mínima contracción y selección absoluta-, por sus siglas en inglés) y red elástica (elastic net), para después obtener el índice de propensión.

El método de validación cruzada 10-fold divide la muestra de entrenamiento en 10 subgrupos de igual tamaño y mutuamente excluyentes. De esos subgrupos se utilizan 9 para entrenar el método, estudiando en este caso diferentes regresiones como hemos indicado en el párrafo anterior, que nos servirán para seleccionar variables. Una vez realizada la estimación del método se prueba con el subgrupo no utilizado y se calcula el error de clasificación, que será el criterio que utilizar para comprobar la efectividad de cada método. Este paso se repite las veces que se ha dividido esta muestra de entrenamiento utilizando cada vez un subgrupo de prueba y utilizando los otros 9 subgrupos para estimar el método. Así, en este caso, obtenemos 10 estimaciones de error con las que calculamos la media de error de forma pesimista.

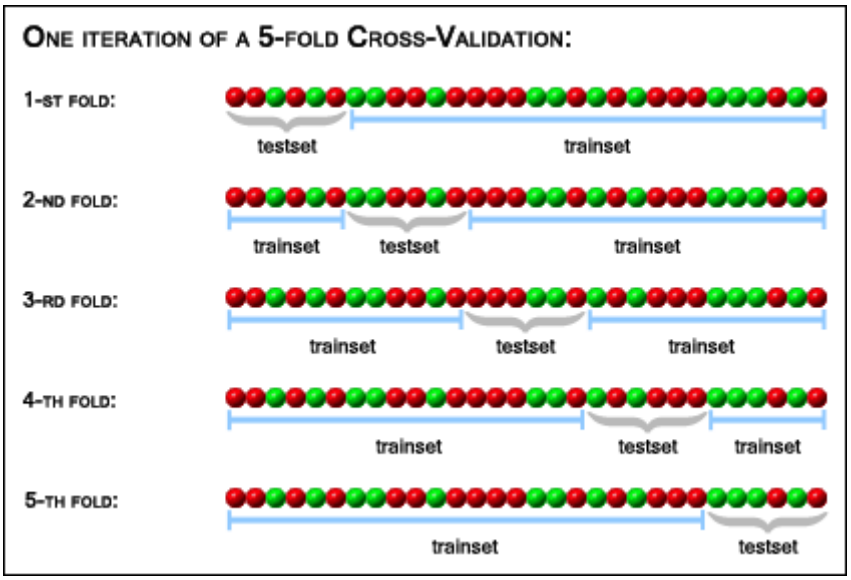

Figura 3.2: Un ejemplo de validación cruzada para k=5. FH Joanneum (2005).

3.6.2.2. Selección de variables

En el presente estudio disponemos de muchas variables explicativas las cuales pueden no ser significativas, el abuso de variables que no son significativas puede producir un modelo que no generalice el problema de forma deseada. Nuestro objetivo es minimizar la función de coste de la log-verosimilitud negativa. La regularización consiste en añadir una penalización a la función de coste. Esta penalización produce métodos más simples ya que los modelos complejos tienden al fenómeno del sobreajuste lo que significa que la solución a los datos de entrenamiento es muy precisa, no así para datos nuevos, en otras palabras, la característica de generalización que deseamos no se cumple. En el presente estudio vamos a llevar a cabo las más comunes que son Ridge, Lasso y una combinación de ambas que se denomina red elástica. Además utilizaré el método de selección de variables stepwise.

3.6.2.2.1. Método Ridge

La regresión Ridge es una técnica que es desarrollada en la década de los años 70 del siglo pasado; para añadir restricciones que hacen mermar la estimación de los parámetros. Utilizando la herramienta de validación cruzada podemos obtener el factor de penalización óptimo (lo denominamos λ) para que el error cuadrático medio sea menor a pesar de un mayor sesgo en las estimaciones. Dado β el vector de coeficientes de regresión y L la función log-verosimilitud negativa, el estimador Ridge es definido como:

$$\widehat{\beta_{Ridge}} = argmin(L(\beta) - \lambda \sum_{j=1}^{p} \beta_j^2)$$

En general, el método Ridge mejoró a los métodos clásicos de selección de variables vistos hasta entonces.

Esa penalización nos ayudará en caso de multicolinealidad en los atributos de entrada; la multicolinealidad es un fenómeno que consiste en que una variable explicativa puede ser predicha de forma lineal por otra variable explicativa con un alto grado de acierto. El método Ridge nos va a servir de ayuda haciendo que disminuya el tamaño de los coeficientes cuando sospechemos que varios de los atributos de entrada estén correlacionados entre ellos y hace que el método generalice mejor. En el método Ridge es necesario que la mayoría de los atributos sean relevantes.

3.6.2.2.2. Método Lasso

El método Lasso incluye todas las variables correlacionadas, pero en el Ridge no se lo traga (más simple) El método Lasso (Least Absolute Shrink-age and Selection Operator -operador de mínima contracción y selección absoluta-, por sus siglas en inglés) está basado en el método Ridge. A mediados de la década de los años 90 del siglo pasado esta regresión impone un límite a la suma de los valores absolutos de los coeficientes de regresión y disminuye este límite hasta que se obtiene un óptimo. El método Lasso fue introducido para mejorar la exactitud de las predicciones y la interpretabilidad de los métodos de regresión al alterar el proceso de construcción al seleccionar un subconjunto de las variables. Además, Lasso nos va a servir de ayuda cuando sospechemos que varios de los atributos de entrada sean irrelevantes. Al usar la regularización Lasso, algunos de los coeficientes pueden acabar valiendo cero. Esto puede ser útil para descubrir cuáles de los atributos de entrada son relevantes y, en general, para obtener un método que generalice mejor la respuesta. Lasso nos puede ayudar, en este sentido, a hacer la selección de entrada, funcionando mejor cuando los atributos no están muy correlacionados entre ellos.

La regresión Lasso impone una penalización en los coeficientes de regresión, minimizando la log-verosimilitud negativa. Dado β el vector de coeficientes de regresión y L la función log-verosimilitud negativa, el estimador Lasso es definido como:

$$\widehat{\beta_{Lasso}} = argmin(L(\beta) - \lambda \sum_{j=1}^{p} |\beta_j|)$$

Sujeto a la estimación de $\lambda \geq 0$ que determina la cantidad de “contracción”, es importante estimar bien λ para obtener métodos que satisfagan el principio de parsimonia.

Previo a Lasso, el método más usado para decidir qué variables incluir era el método Ridge, mientras que el método Ridge es muy eficiente con variables explicativas que pueden sufrir de multicolinealidad, el método Lasso no es tan eficiente, sin embargo, mejora la predicción al reducir en tamaño los coeficientes de regresión que sean demasiado grandes hasta llegar a cero para reducir el sobreajuste. Otra diferencia entre el modelo Ridge y el modelo Lasso es que la estimación de los parámetros en Ridge puede que nunca sean cero, es decir que pueden ser extremadamente pequeños pero nunca serán cero, esto nos puede interesar en este estudio con tantas variables para una selección más estricta.

3.6.2.2.3. Método Red Elástica

Un método más de selección de variables introducido a principios de la década de los 2000 es la red elástica que, en esencia, es una combinación de los dos métodos anteriores y es especialmente bueno en situaciones de correlación entre parámetros. La siguiente fórmula es la estimación de los parámetros, pero con un aliciente: α puede valer entre 0 y 1:

$$\widehat{\beta_{Elastic-Net}} = argmin(L(\beta) - \lambda[\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2])$$

Así, podemos comprobar en la ecuación que si α es igual 1 tenemos una regresión Lasso y si α es igual 0 tenemos una regresión Ridge. Nosotros utilizaremos diferentes valores para α entre 0 y 1 y, por tanto, comprobamos en cuál tenemos menor error. En otras palabras, estudiamos la proporción de los dos métodos anteriores idónea para nuestro estudio. Este método es especialmente útil por que combina las dos ventajas del método Ridge y del método Lasso.

3.6.2.2.4. Método Stepwise

Presentado en la década de los 60 del siglo pasado este método consiste en plantear un método predictivo completo o vacío (en nuestro caso seleccionamos el completo) e ir eliminando o añadiendo variables no significativas por “la norma de El criterio de Información de Akaike (AIC)”. El AIC es una medida de la calidad relativa de un método estadístico, para un conjunto dado de datos. Como tal, el AIC proporciona un medio para la selección del método manejando una “compensación” entre la bondad de ajuste del método y la complejidad del método. Se basa en la entropía de la información: se ofrece una estimación relativa de la información perdida cuando se utiliza un método determinado para representar el proceso que genera los datos.

AIC no proporciona una prueba de un método en el sentido de probar una hipótesis nula, es decir, AIC no puede decir nada acerca de la calidad del método en un sentido absoluto. Si todos los métodos candidatos encajan mal, AIC no dará ningún aviso de ello, por el contrario, nos es útil para comparar modelos.

Por razones computacionales, no se suele utilizar como selección de subconjuntos de variables cuando el número de variables es muy alto ya que puede sufrir problemas de estimación numérica. A mayor número de coeficientes a estimar, mayor es la posibilidad de obtener métodos que parecen funcionar muy bien en nuestro conjunto de datos de entrenamiento, pero su carácter de generalización no sea el deseado.

3.6.3. Índice de propensión

En los estudios experimentales la asignación del tratamiento es de forma aleatoria de forma que se consigue que los grupos sean comparables respecto a las covariables basales y las diferencias que pueda haber se deban al tratamiento. Nuestro estudio al no ser un ensayo clínico puede haber variables asociadas con la

variable respuesta que no estén igualmente distribuidas en cada grupo de tratamiento dando lugar a estimaciones sesgadas. Es por ello por lo que debemos “agrupar” pacientes que han sido tratados con características similares sin el tratamiento, utilizamos para dicha agrupación el índice de propensión. El índice de propensión es la probabilidad que tiene cada participante del estudio de ser asignado a cada una de las ramas del estudio en base a sus características basales.

Las estimaciones del índice de propensión se obtienen por diversas técnicas, entre otras la regresión logística, árboles de clasificación o redes neuronales. En este trabajo se utilizará la regresión logística con las variables seleccionadas anteriormente, puesto que utilizamos como variable respuesta si se realiza traqueotomía o no. Esta estimación explica la posible pertenencia de un individuo al grupo de aquellos que sí han sido intervenidos con una traqueotomía, o al grupo de aquellos que no en función de las diferencias de las variables que ambos grupos presentan. Una vez estimado el índice de propensión, se emparejarán un episodio al que se realizó traqueotomía con un número de episodios a los que no se realizó la traqueotomía. Tras ello, se comprobará la mortalidad en los pacientes entre realizarles la traqueotomía o no.

4. RESULTADOS

Realizaremos una comparación de los cuatro métodos anteriormente mencionados utilizando dos vías, la primera no seleccionando ninguna variable respecto a pruebas de asociación, y la segunda seleccionando aquellas variables que den un p-valor menor que 0.2 en el resultado obtenido del análisis bivariado. Realizada la comparación de los ocho procedimientos, se procede a estudiar la correlación que hay entre los valores obtenidos de la *logit* de la probabilidad a posteriori de recibir traqueotomía.

edad.p.value	sexo.p.value	ambito.p.value	tiping.p.value	tipo_hospital.p.value
1.503654e-04	4.239829e-01	4.143268e-01	9.952880e-03	8.645330e-05
Traqueotomia.p.value	**CARD.p.value**	**RESP.p.value**	**NEFRO.p.value**	**HEPAT.p.value**
0.000000e+00	1.204259e-07	3.074788e-02	9.448274e-03	8.122379e-05
InmunoDef.p.value	**Tm.p.value**	**DM.p.value**	**EnfNM.p.value**	**ObsVAeS.p.value**
2.902186e-01	3.091214e-02	9.150903e-04	4.057319e-23	5.364332e-02
IRA.p.value	**REPOC.p.value**	**Gripe.p.value**	**BrA.p.value**	**SDRA.p.value**
6.118191e-02	2.762995e-03	2.203280e-01	6.017383e-07	7.697146e-04
TEP.p.value	**Asma.p.value**	**BE.p.value**	**ICA.p.value**	**IM.p.value**
8.629069e-01	4.673570e-02	3.593545e-02	3.837310e-08	5.199114e-05
MIO.p.value	**TrTo.p.value**	**Coma.p.value**	**ACVA.p.value**	**TrCr.p.value**
9.942673e-01	5.909899e-08	2.353387e-01	4.445315e-07	1.132071e-03
TrCC.p.value	**EsEp.p.value**	**Shock.p.value**	**Quemaduras.p.value**	**Intox.p.value**
3.105554e-01	5.791569e-01	2.478089e-06	1.588287e-01	4.141889e-02
PrDigA.p.value	**AP01.p.value**	**AP02.p.value**	**AP03.p.value**	**AP04.p.value**
6.314237e-03	2.615041e-293	2.003915e-09	2.050651e-23	1.507056e-03
AP05.p.value	**AP06.p.value**	**AP07.p.value**	**AP08.p.value**	**NeAVM01.p.value**
1.032190e-34	4.379684e-09	5.364332e-02	1.793006e-118	4.763088e-07
NeAVM02.p.value	**NeAVM03.p.value**	**ComTraq.p.value**		
4.763088e-07	1.810829e-50	7.558481e-36		

Tabla 4.1: Matriz de p-valores de las variables respecto a traqueotomía.

Las variables que no seleccionamos al principio cuyo p-valor es mayor que 0.2 son las siguientes: *sexo, ámbito, InmunoDef, Gripe, TEP, MIO, Coma, TrCC y EsEp.* El

resto de las variables las consideramos como relacionales con la variable respuesta después del presente estudio.

4.1. Análisis bivariado.

Columna	Traqueotomía No N=10462	Traqueotomía Si N=3093	Total N=13555	p-valor traqueotomía	p-valor exitus
edad	64.2 ±16.24	65.2 ±14.73	64.48 ±15.91	0.0001504	2.2e-16
sexo	1: 6871 (65.67%) 2: 3591 (34.32%)	1: 2056 (66.5%) 2: 1037 (33.5%)	1: 8927 (65.9%) 2: 4628 (34.1%)	0.424	0.2398
ambito	1: 6025 (57.59%) 2: 4198 (40.12%) 3: 239 (2.29%)	1: 1817 (58.8%) 2: 1201 (38.8%) 3: 75 (2.4%)	1: 7842 (57.9%) 2: 5399(39.8%) 3: 314(2.3%)	0.4143	0.01108
tiping	1: 9465 (90.47%) 2: 997 (9.53%)	1: 2846 (92%) 2: 247 (8%)	1: 12311 (90.8%) 2: 1244 (9.2%)	0.009953	0.07282
exitus	0: 5688 (54.4%) 1: 4774 (45.6%)	0: 1508 (48.7%) 1: 1585 (51.3%)	0: 7196 (53.1%) 1: 6359 (46.9%)	4.378e-08	2.2e-16
tipo hospital	1: 6692 (63.9%) 2: 3713 (35.5%) 3: 57 (0.5%)	1: 2062 (66.6%) 2: 1029 (33.26%) 3: 2 (0.06%)	1: 8754 (64.6%) 2: 4742 (35%) 3: 59 (0.4%)	8.645e-05	1.657e-12
Traqueotomia	0: 10462 (100%) 1: 0 (0%)	0: 0 (0%) 1: 3093 (100%)	0: 10462 (77.2%) 1: 3093 (22.8%)	2.2e-16	4.378e-08
CARD	0: 6886 (65.9%) 1: 3576 (34.1%)	0: 2194 (71%) 1: 899 (29%)	0: 9080 (67%) 1: 4475 (33%)	1.204e-07	2.2e-16
RESP	0: 9935 (95%) 1: 527 (5%)	0: 2906 (94%) 1: 187 (6%)	0: 12841 (94.7%) 1: 714(5.3%)	0.03075	0.06967
NEFRO	0: 10066 (96%) 1: 396 (4%)	0: 3007 (97.2%) 1: 86 (2.8%)	0: 13073 (96.5%) 1: 482 (3.5%)	0.009448	4.323e-07
HEPAT	0: 10121 (96.7%) 1: 341 (3.3%)	0: 3035 (98.1%) 1: 58 (1.9%)	0: 13156 (97%) 1: 399(3%)	8.122e-05	2.4e-06
Inmuno Def	0: 10395 (99.4%) 1: 67 (0.6%)	0: 3079 (99.5%) 1: 14 (0.5%)	0: 13474 (99.4%) 1: 81(0.6%)	0.2902	0.7375

Tm	0: 10288 (98.3%) 1: 174 (1.7%)	0: 3059 (98.9%) 1: 34 (1.1%)	0: 13347 (98.5%) 1: 208 (1.5%)	0.03091	1.01e-06
DM	0: 9468 (90.5%) 1: 994 (9.5%)	0: 2860 (92.5%) 1: 233 (7.5%)	0: 12328 (91%) 1: 1227 (9%)	0.0009151	0.00492
EnfNM	0: 10334 (98.7%) 1: 128 (1.3%)	0: 2970 (96%) 1: 123 (4%)	0: 13304 (99.1%) 1: 251 (1.9%)	2.2e-16	2.766e-07
ObsVAeS	0: 10450 (99.9%) 1: 12 (0.1%)	0: 3084 (99.7%) 1: 9 (0.3%)	0: 13534 (99.85%) 1: 21 (0.15%)	0.05364	0.3034
IRA	0: 7553 (72.2%) 1: 2909 (27.8%)	0: 2179 (70.5%) 1: 914 (29.5%)	0: 9732 (71.8%) 1: 3823 (28.2%)	0.06118	0.02167
REPOC	0: 9799 (93.7%) 1: 663 (6.3%)	0: 2849 (92.1%) 1: 244 (7.9%)	0: 12648 (93.3%) 1: 907 (6.7%)	0.002763	0.2415
Gripe	0: 10368 (99.1%) 1: 94 (0.9%)	0: 3057 (98.8%) 1: 36 (1.2%)	0: 13425 (99%) 1: 130 (1%)	0.2203	0.003598
BrA	0: 10062 (96.2%) 1: 400 (3.8%)	0: 2910 (94.1%) 1: 183 (5.9%)	0: 12972 (95.7%) 1: 583 (4.3%)	6.017e-07	0.2701
SDRA	0: 9376 (89.6%) 1: 1086 (10.4%)	0: 2705 (87.5%) 1: 388 (12.5%)	0: 12081 (89.1%) 1: 1474 (10.9%)	0.0007697	2.2e-16
TEP	0: 10313 (98.5%) 1: 149 (1.5%)	0: 3047 (98.5%) 1: 46 (1.5%)	0: 13360 (98.5%) 1: 195 (1.5%)	0.8629	0.08231
Asma	0: 10226 (97.7%) 1: 236 (2.3%)	0: 3042 (98.4%) 1: 51 (1.6%)	0: 13268 (97.9%) 1: 287 (2.1%)	0.04674	0.03011
BE	0: 10397 (99.4%) 1: 65 (0.6%)	0: 3062 (99%) 1: 31 (1%)	0: 13459 (99.3%) 1: 96 (0.7%)	0.03594	0.005466
ICA	0: 9439 (90.2%) 1: 1023 (9.8%)	0: 2891 (93.5%) 1: 202 (6.5%)	0: 12330 (91%) 1: 1225 (9%)	3.837e-08	0.0008056
IM	0: 9638 (92.1%) 1: 824 (7.9%)	0: 2917 (94.3%) 1: 176 (5.7%)	0: 12555 (92.7%) 1: 1000 (7.3%)	5.199e-05	0.002813
MIO	0: 10453 (99.9%) 1: 9 (0.1%)	0: 3091 (99.9%) 1: 2 (0.1%)	0: 13544 (99.9%) 1: 11 (0.1%)	0.9943	0.3156
TrTo	0: 9476 (90.5%) 1: 986 (9.5%)	0: 2697 (87.2%) 1: 396 (12.8%)	0: 12173 (89.8%) 1: 1382 (10.2%)	5.91e-08	0.101
Coma	0: 9903 (94.6%) 1: 559 (5.4%)	0: 2910 (94.1%) 1: 183 (5.9%)	0: 12813 (94.5%) 1: 742 (5.5%)	0.2353	2.2e-16
ACVA	0: 9747 (93.1%) 1: 715 (6.9%)	0: 2797 (90.4%) 1: 296 (9.6%)	0: 12544 (92.5%) 1: 1011 (7.5%)	4.445e-07	1.439e-05

TrCr	0: 10336 (98.8%) 1: 126 (1.2%)	0: 3031 (98%) 1: 62 (2%)	0: 13367 (99.6%) 1: 188 (1.4%)	0.001132	2.896e-08
TrCC	0: 10454 (99.9%) 1: 8 (0.1%)	0: 3088 (99.8%) 1: 5 (0.2%)	0: 13542 (99.9%) 1: 13 (0.1%)	0.3106	0.05859
EsEp	0: 9770 (93.4%) 1: 692 (6.6%)	0: 2879 (93.1%) 1: 214 (6.9%)	0: 12649 (93.3%) 1: 906 (6.7%)	0.5792	3.125e-16
Shock	0: 7991 (76.4%) 1: 2471 (23.6%)	0: 2488 (80.4%) 1: 605 (19.6%)	0: 10479 (77.3%) 1: 3076 (22.7%)	2.478e-06	2.2e-16
Quemaduras	0: 10431 (99.7%) 1: 31 (0.3%)	0: 3078 (99.5%) 1: 15 (0.5%)	0: 13509 (99.7%) 1: 46 (0.3%)	0.1588	0.1327
Intox	0: 8596 (82.2%) 1: 1886 (17.8%)	0: 2591 (83.8%) 1: 502 (16.2%)	0: 11187 (82.5%) 1: 2368 (16.2%)	0.04142	0.5373
PrDiagA	0: 9521 (91%) 1: 941 (9%)	0: 2864 (92.6%) 1: 229 (7.4%)	0: 12385 (91.4%) 1: 1170 (8.6%)	0.006314	5.965e-10
AP01	0: 5181 (49.5%) 1: 5281 (50.5%)	0: 391 (12.6%) 1: 2702 (87.4%)	0: 5572 (41.1%) 1: 7983 (58.9%)	2.2e-16	2.2e-16
AP02	0: 9624 (92%) 1: 838 (8%)	0: 2737 (88.5%) 1: 356 (11.5%)	0: 12361 (91.2%) 1: 1194 (8.8%)	2.004e-09	0.09657
AP03	0: 10328 (98.7%) 1: 134 (1.3%)	0: 2966 (95.9%) 1: 127 (4.1%)	0: 13294 (98.1%) 1: 261 (1.9%)	2.2e-16	3.714e-06
AP04	0: 9267 (88.6%) 1: 1195 (11.4%)	0: 2674 (86.5%) 1: 419 (13.5%)	0: 11941 (88%) 1: 1614 (12%)	0.001507	0.196
AP05	0: 2906 (82%) 1: 1936 (18%)	0: 2204 (71.2%) 1: 889 (28.8%)	0: 10730 (79%) 1: 2825 (21%)	2.2e-16	4.525e-06
AP06	0: 8360 (80%) 1: 2102 (20%)	0: 2618 (84.6%) 1: 475 (15.4%)	0: 10978 (81%) 1: 2577 (19%)	4.38e-09	7.031e-10
AP07	0: 10450 (99.9%) 1: 12 (0.1%)	0: 3084 (99.7%) 1: 9 (0.3%)	0: 13534 (99.85%) 1: 21 (0.15%)	0.05364	0.3034
AP08	0: 2758 (26.4%) 1: 7704 (73.6%)	0: 209 (6.8%) 1: 2284 (93.2%)	0: 2967 (21.9%) 1: 10588 (78.1%)	2.2e-16	2.2e-16
NeAVM01	0: 10328 (98.7%) 1: 134 (1.3%)	0: 3013 (97.4%) 1: 80 (2.6%)	0: 13341 (98.4%) 1: 214(1.6%)	4.763e-07	0.0005881

NeAVM 02	0: 10328 (98.7%) 1: 134 (1.3%)	0: 3013 (97.4%) 1: 80 (2.6%)	0: 13341 (98.4%) 1: 214(1.6%)	4.763e-07	0.0005881
NeAVM 03	0: 9695 (92.7%) 1: 767 (7.3%)	0: 2590 (83.8%) 1: 503 (16.2%)	0: 12285 (90.6%) 1: 1270(9.4%)	2.2e-16	0.003324
ComTraq	0: 9266 (88.5%) 1: 1196 (11.5%)	0: 2469 (79.8%) 1: 624 (20.2%)	0: 11735(86.5%) 1: 1820(13.4%)	2.2e-16	0.04187
EdadCat	0: 1026 (9.8%) 1: 4459 (42.6%) 2: 4977 (47.6%)	0: 220 (7.1%) 1: 1375 (44.5%) 2: 1498 (48.4%)	0: 1246 (9.1%) 1: 5834 (43%) 2: 6475 (47.9%)	2.639e-05	2.2e-16

Tabla 4.2: Análisis bivariado respecto a traqueotomía.

4.2. Regresión logística sin selección previa.

En todos los métodos de regularización, las variables que usamos a priori y que conocemos durante el episodio del paciente sin preselección por test de correlación con la variable traqueotomía, son las siguientes: *edad, sexo, ambito, tiping, tipo_hospital, CARD, RESP, NEFRO, HEPAT, Tm, DM, EnfNM, ObsVAeS, IRA, REPOC, BrA, SDRA, Asma, BE, ICA, IM , TrTo , ACVA, TrCr , Shock, Quemaduras, Intox, PrDigA, AP01, AP02, AP03, AP04, AP05, AP06, AP07, AP08, NeAVM01, NeAVM02, NeAVM03 y EdadCat*.

4.2.1. Método Ridge

Dado que nuestro objetivo es tener la máxima precisión, vamos a utilizar el índice de Youden para obtener los valores óptimos de sensibilidad y especificidad. Este índice consiste en sumar sensibilidad con especificidad y restar uno, procedimiento que realizamos para cada punto de corte hasta conseguir el máximo. En la regresión Ridge calculada obtenemos un resultado de 0.4411 y la siguiente matriz de confusión.

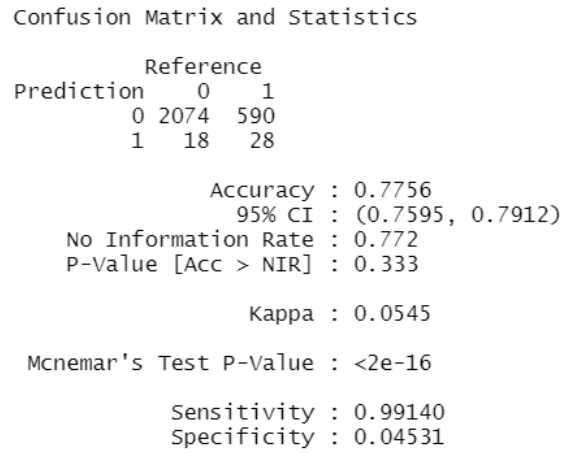

Figura 4.1: Matriz de confusión método Ridge sin selección previa.

En ella comprobamos cómo la especificidad es baja pero la sensibilidad es bastante alta. También podemos fijarnos para comparar métodos en el estadístico Kappa de Cohen o Kappa, el cual podemos considerar como una estimación de la precisión normalizada en la clasificación en caso de una descompensación entre las proporciones de cada clase (como es el caso). Consiste en la diferencia entre la proporción actual y la estimada, dividido todo ello entre uno menos la proporción estimada. Obtenemos un valor bajo debido a lo comentado anteriormente, una sensibilidad muy alta con una especificidad muy baja.

4.2.2. Método Lasso

Los resultados obtenidos, tras estimar una mejor cota de corte de clase para la predicción, que en este caso ha sido de 0.4458, están en la siguiente matriz de confusión.

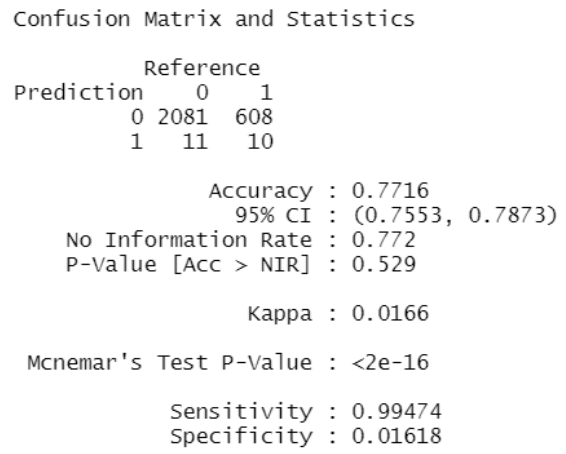

Figura 4.2: Matriz de confusión método Lasso sin selección previa.

La especificidad es menor respecto a Ridge, pero la sensibilidad aumenta y la precisión es mayor. Además, el valor de Kappa es ínfimo lo que nos indica una clasificación más descompensada.

4.2.3. Método Red Elástica

Con este método, en primer lugar, debemos especificar la proporción que se va a destinar al método Ridge y la usada para el método Lasso, para ello vamos a tomar 10 valores de Alpha entre 0 y 1 para calcular el método que obtenga una mejor tasa de acierto. Los resultados que hemos obtenido han sido con Alpha=0.2, valor que usaremos para realizar la regularización del método. Teniendo en cuenta que el punto de corte óptimo estimado es de 0.4595.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2051  561
         1   41   57

               Accuracy : 0.7779
                 95% CI : (0.7617, 0.7934)
    No Information Rate : 0.772
    P-Value [Acc > NIR] : 0.2396

                  Kappa : 0.1032

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.98040
            Specificity : 0.09223
```

Figura 4.3: Matriz de confusión método red elástica sin selección previa.

En este caso observamos en la matriz de confusión unos resultados muy próximos a la regresión Lasso con una especificidad mayor y una sensibilidad un poco menor; y debido a una mejor compensación entre sensibilidad y especificidad, el estadístico Kappa aumenta.

4.2.4. Método stepwise

Con el método de stepwise obtenemos un óptimo punto de corte de 0.5141.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2083  601
         1    9   17

               Accuracy : 0.7749
                 95% CI : (0.7587, 0.7905)
    No Information Rate : 0.772
    P-Value [Acc > NIR] : 0.367

                  Kappa : 0.035

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.99570
            Specificity : 0.02751
```

Figura 4.4: Matriz de confusión método stepwise sin selección previa.

En la matriz de confusión obtenemos unos resultados muy próximos a la regresión Lasso con una especificidad igual y una sensibilidad un poco menor. Sin

embargo, obtenemos uno de los peores resultados para Kappa respecto al resto de métodos analizados.

4.3. Regresión logística con selección previa.

En todos los métodos de regularización, las variables que usamos a priori y que conocemos durante el episodio del paciente con preselección por test de correlación con la variable traqueotomía, son las siguientes: *sexo, ambito, edad, tiping, tipo_hospital, CARD, RESP, NEFRO, HEPAT, InmunoDef, Tm, DM, EnfNM, ObsVAeS, IRA, REPOC, Gripe, BrA, SDRA, TEP, Asma, BE, ICA, IM, MIO, TrTo, Coma, ACVA, TrCr, TrCC, EsEp, Shock, Quemaduras, Intox, PrDigA, AP01, AP02, AP03, AP04, AP05, AP06, AP07, AP08, NeAVM01, NeAVM02, NeAVM03 y EdadCat*.

4.3.1. Método Ridge

Los resultados obtenidos, tras estimar una mejor cota de corte de clase para la predicción que en este caso ha sido de 0.4, están en la siguiente matriz de confusión.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2079  594
         1   13   24

               Accuracy : 0.776
                 95% CI : (0.7598, 0.7916)
    No Information Rate : 0.772
    P-Value [Acc > NIR] : 0.3165

                  Kappa : 0.0488

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.99379
            Specificity : 0.03883
```

Figura 4.5: Matriz de confusión método Ridge con selección previa.

Comprobamos cómo obtenemos una precisión por encima de 0.77 con un valor de Kappa por debajo de 0.05. Esto nos indica una descompensación que evidenciamos con la diferencia entre sensibilidad y especificidad.

4.3.2. Método Lasso

Los resultados obtenidos, tras estimar una mejor cota de corte de clase para la predicción que en este caso ha sido de 0.42, están en la siguiente matriz de confusión.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2077  603
         1   15   15

               Accuracy : 0.772
                 95% CI : (0.7557, 0.7876)
    No Information Rate : 0.772
    P-Value [Acc > NIR] : 0.5108

                  Kappa : 0.0257

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.99283
            Specificity : 0.02427
```

Figura 4.6: Matriz de confusión método Lasso con selección previa.

La especificidad es menor respecto a Ridge pero la sensibilidad aumenta y la precisión es mayor; por contra, observamos una disminución en el estadístico Kappa.

4.3.3. Método Red Elástica

Con este método, en primer lugar, debemos especificar la proporción que se va a destinar al método Ridge y la usada para el método Lasso. Para ello vamos a tomar 10 valores de Alpha entre 0 y 1 para calcular el método que obtenga una mejor tasa de acierto. Los resultados que hemos obtenido han sido con Alpha=0.3, valor que usaremos para realizar la regularización del método. Obtenemos un óptimo punto de corte de 0.4168 en la clasificación y a continuación mostramos la matriz de confusión.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2079  594
         1   13   24

               Accuracy : 0.776
                 95% CI : (0.7598, 0.7916)
    No Information Rate : 0.772
    P-Value [Acc > NIR] : 0.3165

                  Kappa : 0.0488

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.99379
            Specificity : 0.03883
```

Figura 4.7: Matriz de confusión método red elástica con selección previa.

Aquí obtenemos unos resultados muy próximos a la regresión Ridge con una especificidad igual y una sensibilidad un poco menor.

4.3.4. Método stepwise

Los resultados obtenidos, tras estimar una mejor cota de corte de clase para la predicción que en este caso ha sido de 0.4168, están en la siguiente matriz de confusión.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2079  593
         1   13   25

               Accuracy : 0.7764
                 95% CI : (0.7602, 0.792)
    No Information Rate : 0.772
    P-Value [Acc > NIR] : 0.3003

                  Kappa : 0.0512

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.99379
            Specificity : 0.04045
```

Figura 4.8: Matriz de confusión método stepwise con selección previa.

Comprobamos cómo la especificidad es baja pero la sensibilidad es bastante alta, descompensación que se comprueba en el valor bajo de Kappa.

4.4. Evaluación de los ocho métodos de selección de variables.

Tras realizar los métodos de regularización de selección de variables y estimar los logit de los métodos resultantes, hacemos la matriz de correlación para comprobar si hay similitudes entre los dos grupos. En dicha matriz de correlación disponemos de 8 columnas, de las cuales, las 4 primeras filas de las 4 primeras columnas corresponden al grupo sin selección previa de variables, mientras que las 4 últimas filas de las 4 últimas columnas corresponden al grupo con selección previa de variables. Analizando los resultados, podemos observar una alta correlación independientemente del grupo al que pertenezca cada método. Así podemos comprobar también que la menor correlación es 0.9522, un valor elevado que corresponde con el método Ridge sin preselección de variables y Lasso con preselección de variables. Además, la mayor correlación obtenida es 0.9997, la cual se da entre el método Lasso sin preselección de variables y el método Lasso con preselección de variables.

	"RidgeSin"	"LassoSin"	"ENSin"	"stepwiseSin"	"RidgeCon"	"LassoCon"	"ENCon"	"stepwiseCon"
"RidgeSin"	1.0000000	0.9537143	0.9908340	0.9814329	0.9954240	0.9522058	0.9800794	0.9806964
"LassoSin"	0.9537143	1.0000000	0.9829100	0.9548287	0.9565508	0.9997454	0.9911928	0.9607158
"ENSin"	0.9908340	0.9829100	1.0000000	0.9797669	0.9905108	0.9818879	0.9967658	0.9818025
"stepwiseSin"	0.9814329	0.9548287	0.9797669	1.0000000	0.9756480	0.9528873	0.9733749	0.9914652
"RidgeCon"	0.9954240	0.9565508	0.9905108	0.9756480	1.0000000	0.9567110	0.9846230	0.9841042
"LassoCon"	0.9522058	0.9997454	0.9818879	0.9528873	0.9567110	1.0000000	0.9914081	0.9608825
"ENCon"	0.9800794	0.9911928	0.9967658	0.9733749	0.9846230	0.9914081	1.0000000	0.9814811
"stepwiseCon"	0.9806964	0.9607158	0.9818025	0.9914652	0.9841042	0.9608825	0.9814811	1.0000000

Figura 4.9: Correlación entre los ocho métodos.

En el grupo sin preselección previa de variables, los métodos Ridge y red elástica son los más correlacionados, aunque no se da una excesiva diferencia con respecto al resto de correlaciones ya que todas son superiores a 0.95. En el grupo con selección previa de variables, comprobamos cómo el método Lasso y red elástica son los más correlacionados aunque ocurre lo mismo, todos los valores están por encima de 0.95 con respecto al resto de variables. Esto es plausible ya que la red elástica,

como hemos comentado antes, es un método que combina Ridge con Lasso. Por último, atendiendo a la matriz completa, podemos observar cómo el método stepwise arroja una correlación muy alta con el resto de los métodos pero con la desventaja de que implica un coste computacional mayor.

Tras evaluar los ocho métodos, comparándolos con la estimación del error, podemos comprobar cómo obtenemos una estimación del error menor con el método de regularización red elástica sin preselección de variables. Por tanto, aunque el resultado de la preselección de variables no tiene una diferencia muy significativa entre los ocho métodos propuestos, el uso de la red elástica con alfa igual a 0.1 tiene las ventajas del método Ridge y del método Lasso utilizando además un tiempo de ejecución menor que stepwise. Por ello, a continuación se muestran los resultados de aplicar red elástica al grupo de entrenamiento presentando los coeficientes estimados de las variables y su matriz de confusión:

	s0
(Intercept)	-2.454854512
sexo	-0.030149927
ambito	.
edad	0.004436511
tiping	-0.151459280
tipo_hospital	-0.089646633
CARD	-0.086633884
RESP	0.025642273
NEFRO	-0.029187589
HEPAT	-0.239044261
InmunoDef	.
Tm	-0.102445806
DM	-0.020440925
EnfNM	0.348371147
ObsVAeS	0.279869756
IRA	0.017051331
REPOC	0.135201397
Gripe	.
BrA	0.248539911
SDRA	0.074763985
TEP	.
Asma	-0.080877324
BE	0.237234966
ICA	-0.075063729
IM	-0.103124369
MIO	-0.012713485
TrTo	0.150883172
Coma	0.172289966
ACVA	0.035678576
TrCr	.
TrCC	.
EsEp	.
Shock	-0.129479757
Quemaduras	.
Intox	-0.035466601
PrDigA	-0.115926909
AP01	1.263961918
AP02	0.145849386
AP03	0.462053855
AP04	0.057757599
AP05	0.143180367
AP06	-0.151293569
AP07	0.280413970
AP08	0.282488003
NeAVM01	.
NeAVM02	.
NeAVM03	0.433838881
EdadCat	0.066599670

Figura 4.10: Valor de los coeficientes del método elegido.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 2051  561
         1   41   57

               Accuracy : 0.7779
                 95% CI : (0.7617, 0.7934)
    No Information Rate : 0.772
    P-Value [Acc > NIR] : 0.2396

                  Kappa : 0.1032

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.98040
            Specificity : 0.09223
```

Figura 4.11: Matriz de confusión del método elegido.

En los valores de coeficientes podemos fijarnos que los mayores valores corresponden a las variables de *AP01, AP03 y NeAVM03*. Confirmamos mediante la orden *varImp* de la librería *caret* lo influyente que son en el método. En los resultados obtenidos en la parte superior podemos observar que hay coeficientes representados por puntos, los cuales indican que esas variables no han sido seleccionadas por el método. Podemos fijarnos en variables como *ámbito, InmunoDef, Gripe, TEP, TrCr, TrCC, EsEp, Quemaduras, NeAVM01 y NeAVM02*. En la matriz de confusión, podemos observar una muy baja especificidad por lo que sería recomendable modificar el punto de corte de los logit para aumentar la especificidad en detrimento de la sensibilidad para así aumentar también el valor de Kappa. El punto de corte fue estimado para una maximización en la precisión con la orden *optimalCutoff* de la librería *InformationValue,* obteniendo así el valor 0.425. A pesar de haber indicado por qué el método red elástica ha sido seleccionado, consideramos importante señalar que presenta una desventaja contra el método Lasso: es que no desecha ninguna variable, las asigna valores muy bajos pero no las elimina.

4.5. Emparejamiento de pacientes con y sin traqueotomía basado en el índice de propensión.

Recordemos que el objetivo del estudio es comprobar el efecto del tratamiento, que en este caso es la traqueotomía, con la influencia en la evolución del paciente. Gracias a esta técnica, la estimación del efecto del tratamiento no es influenciada por covariables que, tras un primer análisis, pueden producir una confusión en forma de sesgo. La estimación del índice de propensión de los episodios la obtenemos utilizando la orden *fitted* sobre el método de regresión logística en R.

A continuación, con la librería *MatchIt* obtenemos la orden *matchit* que nos empareja a los pacientes para posteriormente poder asemejar el estudio a un ensayo clínico, proceso en el que podemos elegir diferentes métodos de emparejamiento, como por ejemplo, por máxima distancia del índice de propensión entre grupo de control y tratamiento (esta distancia la denominamos caliper); el método exacto, que es la versión más simple de emparejamiento, la cual empareja cada unidad tratada con todas las unidades de control posibles que tengan exactamente los mismos valores en todas las covariables; o, el tercer ejemplo, K-vecinos más próximos, en el que al modificar una variable para hacer el emparejamiento, que es, el número de

episodios sin tratamiento por cada episodio con tratamiento; esta variable la denominamos ratio, pudiendo trabajar en el presente estudio con una ratio 1:1, 1:2 o 1:3. La primera cifra corresponde con el número de episodios con traqueotomía, mientras que el segundo número corresponde al número de episodios sin traqueotomía. La siguiente ratio sería 1:4, pero ya no es posible realizarla salvo que usemos muestreo con reemplazamiento. Para escoger qué método de emparejamiento elegir se ha de comparar la diferencia de distancia de episodios emparejados. En nuestro estudio, tras comparar los diferentes métodos de emparejamiento, elegimos el método de k-vecinos más próximos al determinar una distancia menor. Una vez elegido el modelo debemos determinar qué ratio escoger. Finalmente, la ratio que usaremos es el 1:2.

4.6. Asociación entre traqueotomía y *exitus*

Una vez obtenido el emparejamiento, guardamos los datos añadiendo la variable *exitus* para poder comparar y obtener la odds ratio. Por otro lado, podemos comprobar, antes de calcular la odds ratio, las densidades del índice de propensión entre los grupos de traqueotomía observando al representarlo que no hay diferencias significativas como sí las obtenemos con la población original.

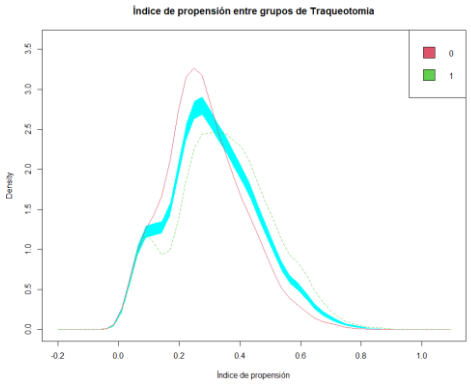

Figura 4.12: Índice de propensión entre grupos de traqueotomía tras emparejamiento.

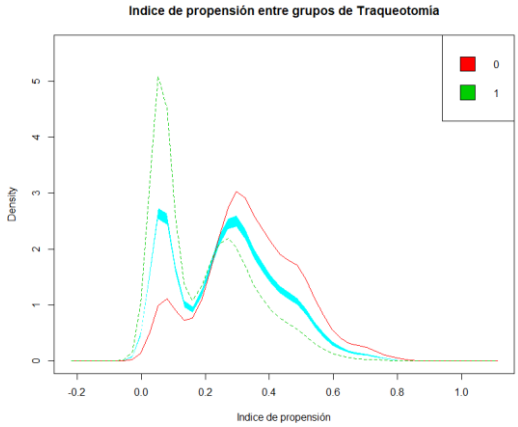

Figura 4.13: Índice de propensión entre grupos de traqueotomía sin emparejar.

Para comprobar la efectividad del índice de propensión en la agrupación, podemos comparar mediante el test de Fisher la odds ratio previa al emparejamiento con la posterior al emparejamiento, proceso que se muestra a continuación:

```
> fisher.test(table(GDA_trabajo$exitus,GDA_trabajo$Traqueotomia))

        Fisher's Exact Test for Count Data

data:  table(GDA_trabajo$exitus, GDA_trabajo$Traqueotomia)
p-value = 4.31e-08
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.154678 1.358175
sample estimates:
odds ratio
  1.252261
```

Figura 4.14: Test de Fisher de los datos sin emparejar.

```
> fisher.test(table(df.match$exitus,df.match$Traqueotomia))

        Fisher's Exact Test for Count Data

data:  table(df.match$exitus, df.match$Traqueotomia)
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.482753 1.768036
sample estimates:
odds ratio
  1.618969
```

Figura 4.15: Test de Fisher de los datos emparejados.

Tras dicha comparación vemos una diferencia significativa entre ambas figuras, se observa cómo el coeficiente de la odds ratio al haberlos emparejado aumenta de forma considerable; ya era de carácter positivo antes, sin embargo, ahora ni siquiera coincide el extremo inferior del emparejamiento con el extremo superior original. Este hecho constata cuán importante es esta variable para la evolución del paciente.

5. MÉTODO DE PREDICCIÓN DE EVOLUCIÓN DEL PACIENTE

Teniendo en cuenta que el fin del estudio es comprobar el impacto que tiene el tratamiento de traqueotomía en la mortalidad del conjunto de datos, en el apartado anterior hemos comprobado que es una variable influyente en la variable *exitus* y, por tanto, debemos de incluirla en nuestro método inicial. El resto de las variables, detalladas en el apartado 5.3 del presente estudio, las incluiremos sin una selección previa, ya que usaremos los métodos de regularización que hemos comprobado anteriormente que son efectivos para obtener un subconjunto de variables óptimo.

5.1. Método Ridge

Los resultados obtenidos tras estimar una mejor cota de corte de clase para la predicción en este caso han sido de 0.4989 y, por consiguiente, obtenemos la siguiente matriz de confusión.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1036  493
         1  403  778

               Accuracy : 0.6694
                 95% CI : (0.6513, 0.6871)
    No Information Rate : 0.531
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3334

 Mcnemar's Test P-Value : 0.002946

            Sensitivity : 0.7199
            Specificity : 0.6121
```

Figura 5.1: Matriz de confusión método Ridge.

En ella, podemos comprobar cómo la especificidad es menor que la sensibilidad, la cual se aproxima a 0.72. En el estadístico Kappa podemos observar un valor de 0.33 aproximadamente, es decir, una clasificación compensada y proporcionada.

5.2. Método Lasso

Los resultados obtenidos tras estimar una mejor cota de corte de clase para la predicción en este caso han sido de 0.526 y, por tanto, obtenemos la siguiente matriz de confusión.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1080  562
         1  359  709

               Accuracy : 0.6601
                 95% CI : (0.642, 0.678)
    No Information Rate : 0.531
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3113

 Mcnemar's Test P-Value : 2.811e-11

            Sensitivity : 0.7505
            Specificity : 0.5578
```

Figura 5.2: Matriz de confusión método Lasso.

La especificidad y la precisión son menores respecto a Ridge, pero la sensibilidad aumenta, cosa que afecta a la precisión y al estadístico Kappa que disminuye ligeramente.

5.3. Red Elástica

Con este método, en primer lugar, debemos especificar la proporción que se va a destinar al método Ridge y al método Lasso, para ello vamos a tomar 10 valores de Alpha entre 0 y 1 para calcular el método que obtenga una mejor tasa de acierto. Dicho método ha sido el obtenido con el valor Alpha=0.8, siendo por tanto el valor que usaremos para realizar la regularización del método obteniendo un óptimo punto de corte de 0.503 que da como resultado la siguiente matriz de confusión.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1050  505
         1  389  766

               Accuracy : 0.6701
                 95% CI : (0.652, 0.6878)
    No Information Rate : 0.531
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.3341

 Mcnemar's Test P-Value : 0.00012

            Sensitivity : 0.7297
            Specificity : 0.6027
```

Figura 5.3: Matriz de confusión método red elástica.

Aquí obtenemos unos resultados muy próximos a la regresión Lasso con una especificidad igual y una sensibilidad un poco menor, la precisión aumenta respecto a los dos métodos anteriores.

5.4. Stepwise

Obtenemos un punto de corte óptimo de clasificación de 0.5477, lo que resulta en la siguiente matriz de confusión.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1102  569
         1  337  702

               Accuracy : 0.6657
                 95% CI : (0.6476, 0.6834)
    No Information Rate : 0.531
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3216

 Mcnemar's Test P-Value : 1.661e-14

            Sensitivity : 0.7658
            Specificity : 0.5523
```

Figura 5.4: Matriz de confusión método stepwise.

Aquí obtenemos unos resultados muy próximos a la regresión Ridge con una especificidad y sensibilidad menor.

5.5. Método seleccionado para la variable *exitus*.

El método con menor tasa de error es la red elástica seguido muy de cerca por el método Lasso, recordemos que el alfa que obtenemos en red elástica es muy próximo a 1 y por tanto el método Ridge es poco proporcional. Obtenemos así los coeficientes de las variables junto con su matriz de confusión:

	s0
(Intercept)	-1.975185709
sexo	-0.006930258
ambito	.
edad	0.031509376
tiping	.
tipo_hospital	.
Traqueotomia	0.552405040
CARD	0.009422605
RESP	.
NEFRO	0.169884078
HEPAT	0.335012206
InmunoDef	.
Tm	0.335932917
DM	.
EnfNM	-0.181047060
ObsVAeS	.
IRA	.
REPOC	.
Gripe	.
BrA	.
SDRA	0.443425758
TEP	.
Asma	-0.102828192
BE	-0.335573833
ICA	.
IM	.
MIO	.
TrTo	.
Coma	0.584899842
ACVA	0.286686572
TrCr	.
TrCC	0.746251131
EsEp	-0.351754859
Shock	0.600409146
Quemaduras	.
Intox	-0.167418870
PrDigA	0.233837456
AP01	-1.072063108
AP02	.
AP03	.
AP04	.
AP05	0.136766844
AP06	.
AP07	.
AP08	.
NeAVM01	-0.136932686
NeAVM02	-0.010649107
NeAVM03	0.158978583
EdadCat	0.030552249

Figura 5.5: Coeficientes del método final.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1004  435
         1  444  828

               Accuracy : 0.6758
                 95% CI : (0.6578, 0.6934)
    No Information Rate : 0.5341
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.3488

 Mcnemar's Test P-Value : 0.7873

            Sensitivity : 0.6934
            Specificity : 0.6556
```

Figura 5.6: Matriz de confusión obtenida con el método final.

En los resultados obtenidos en la parte superior podemos observar que hay coeficientes representados por puntos, los cuales indican que esas variables no han sido seleccionadas por el método. Por tanto, las variables que no han sido elegidas para la predicción son: *ambito, tiping, tipo_hospital, RESP, Inmunodef, DM, ObsVAeS, IRA, REPOC, Gripe, BrA, TEP, ICA, IM, MIO, TrTo, TrCr, Quemaduras, AP02, AP03, AP04, AP06, AP07 y AP08.* Con seguridad, podemos observar cómo el coeficiente de la variable *Traqueotomia* es positivo con un valor aproximado de 0.55, eso nos indica que ante la presencia de esta variable la variable respuesta se aproxima más al valor 1. En otras palabras, ante el tratamiento de *Traqueotomia* el riesgo de *exitus* es mayor.

La curva ROC (Receiver Operating Characteristic -Característica operativa del receptor- por sus siglas en inglés), en su representación gráfica, nos sirve para visualizar la relación entre sensibilidad y especificidad para diferentes puntos de corte. Se suele representar en el eje de ordenadas la sensibilidad y en el eje de abscisas uno menos la especificidad. Es usado de forma extendida en medicina para estudios diagnósticos ya que estas curvas son útiles para comprobar cuán buen clasificador hemos podido estimar/calcular.

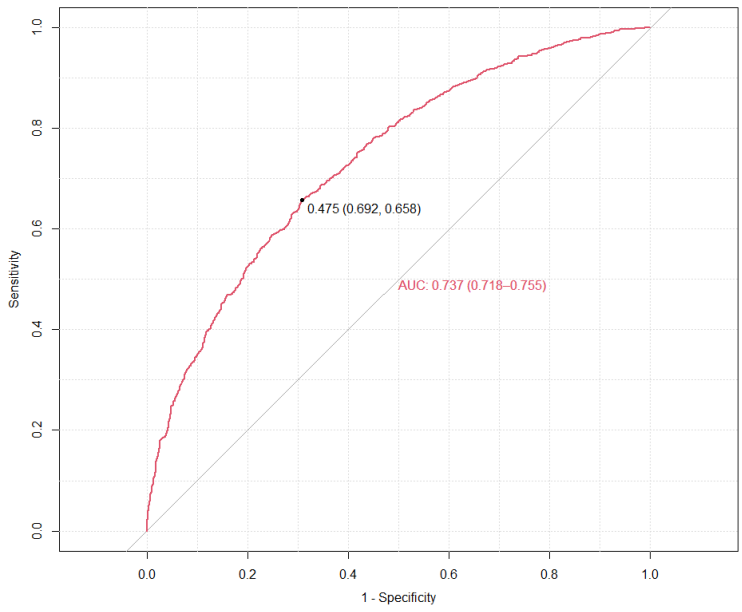

Figura 6.7: Curva ROC del método final.

Cuanto mayor sea el área por debajo de la curva roja, nuestra clasificación será mejor. El punto que vemos señalado, es el punto de concordancia entre la sensibilidad y la especificidad. El área total por debajo de la curva, en el peor de los casos de clasificación aleatoria, sería 0.5 y en el mejor de los casos, con una clasificación perfecta sería de 1. Nuestro resultado ha sido de 0.737.

6. Discusión.

En lo que respecta a la clasificación de los casos de *traqueotomía,* hemos utilizado como método final la red elástica sin preselección con Alpha igual a 0.1, siendo las variables más influyentes en este método *AP01, AP03* y *NeAVM03*. Con respecto a las variables iniciales, comprobamos que están correlacionadas y se puede escoger un subconjunto para obtener una mayor generalización en la regresión logística, gracias a los métodos de regularización. Así con el método final obtenemos una precisión del 77.45% de casos acertados.

Por otro lado, al comenzar el estudio podemos comprobar cómo la variable *Traqueotomia* es significativa en la variable respuesta *exitus*, sin embargo, no es tan alta como después de hacer el emparejamiento, donde aumenta considerablemente. Por tanto, podemos afirmar que es muy significativa al haberlo comprobado realizando el emparejamiento producido por el índice de propensión. En cuanto a la clasificación de la variable *exitus,* hemos elegido como método final la red elástica sin preselección con Alpha igual a 0.8, donde hemos obtenido un 67% de precisión. Además, tras el estudio, podemos determinar que las variables que son importantes en la evolución del paciente son *AP01, edad, Shock* y *Traqueotomia*. En el caso de *edad* el coeficiente es bajo pero como tenemos un rango muy amplio, recordemos, desde 16 hasta 96 años, hace que la diferencia en los extremos sea considerable. Esto lo podemos comprobar respecto al valor de los coeficientes y con la orden *varImp,* donde confirmamos la importancia de esta variable.

Tras evaluar los cuatro métodos, contrastando la estimación del error, podemos comprobar cómo los resultados en cada matriz de confusión son muy parejos, siendo las diferencias entre ellos tiempo de ejecución, eficacia en la selección de variables y cálculo de los coeficientes.

En el tiempo de ejecución, vemos una clara ventaja en los métodos de regularización de Ridge y de Lasso, ya que en el método de red elástica, a pesar de tener que estimar Alpha, es menor el tiempo de ejecución respecto a stepwise. Este último requería de un coste computacional notable, comparándolo con los otros tres métodos.

Respecto a la eficacia en la selección de variables, es recomendable el uso de red elástica con mayor influencia de Lasso, requiriendo estimar en un primer momento Alpha y el método Lasso para obtener el menor número de variables. El siguiente método puede ser el stepwise y por último el método Ridge, ya que no descarta ninguna variable, si bien es cierto que deja algún parámetro estimado muy próximo a cero y podemos asociar a la no selección de esa variable.

En el cálculo de los coeficientes vemos una semejanza en el signo muy próxima entre ellos, como lo es también la magnitud en los cuatro métodos. Observamos una semejanza de la red elástica con respecto a Ridge o Lasso cuando Alpha es próxima a cero en el primer caso y a uno en el último caso.

7. Conclusiones

En los datos iniciales se puede comprobar cómo hay una asociación entre el procedimiento de la traqueotomía y la mortalidad del paciente, sin embargo, para un estudio observacional tanto la elección del método de regularización como el método de emparejamiento son fundamentales para una estimación correcta de la magnitud de esa correlación.

El procedimiento de traqueotomía tiene un efecto creciente en la mortalidad en el paciente. Los episodios emparejados a partir del índice de propensión acentúan esta característica de forma abrupta.

Respecto a la mortalidad del paciente hemos comprobado cómo la variable traqueotomía es influyente junto con una de las variables en común que es *AP01*.

Por último, hay que añadir que el presente trabajo puede ser una introducción a estudios más relevantes a nivel estadístico con un análisis multivariante más exhaustivo; así como en el campo de la inteligencia artificial, más concretamente en el aprendizaje automático; todo ello para servir de apoyo clínico.

BIBLIOGRAFÍA

Agresti, Alan. (2002). “Categorical Data Analysis”. Wiley

Armitage, Peter. (1955). “Tests for Linear Trends in Proportions and Frequencies”. Biometrics.

Breiman, Leo. (1995). “Better Subset Regression Using the Nonnegative Garrote”. Technometrics 37 (4). Taylor & Francis, Ltd.: 373–84. doi:10.2307/1269730.

Cacheiro, Pilar. (2012). “Métodos de selección de variables en estudios de asociación genética. Aplicación a un estudio de genes candidatos en Enfermedad de Parkinson”. Proyecto Fin de Máster. Universidade de Santiago de Compostela. Santiago de Compostela.

Camarero, Luis *et al*. (2013). “Regresión Logística: Fundamentos y aplicación a la investigación sociológica”. UNED.

Cohen. Bernard. (1984). “Florence Nightingale”. Scientific American.

Department of Information Design. (2006). “Cross-Validation Explained”, Institute for Genomics and Bioinformatics. FH Joanneum. Recuperado el 17 de mayo de 2020 de: http://genome.tugraz.at/proclassify/help/pages/XV.html.

Doménech, Ivan. (2006). “Traqueotomía percutánea según el método de Griggs. Estudio de la técnica, como acceso instrumental de la vía aérea en pacientes UCI, sometidos a ventilación mecánica”. Tésis doctoral. Universitat de Barcelona, Barcelona.

García, Julio. (2015). “CARACTERÍSTICAS DE LAS ALTAS HOSPITALARIAS EN CASTILLA Y LEÓN. ANÁLISIS DEL PERIODO 2001-2014.” Tesis doctoral. Universidad de Valladolid. Valladolid.

Harrell, Frank. (2015). “Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression and Survival Analysis”. Springer.

Ho, Daniel *et al*. (2007) “MatchIt: Nonparametric Preprocessing for Parametric Causal Inference”.

James, Gareth *et al*. (2013). “An introduction to Statistical Learning with applications in R”. Springer.

Molina, Manuel (2015). “Índices de propensión. El deseo de parecerse al ensayo clínico”.

Pina, Koldo. (2018). “Matriz de confusión” Recuperado el 5 de marzo de 2020 de: https://koldopina.com/matriz-de-confusion/

Tibshirani, Robert. (1996). “Regression Shrinkage and Selection via the lasso”. Journal of the Royal Statistical Society. Series B (methodological) 58 (1). Wiley: 267–88. http://www.jstor.org/stable/2346178

LISTA DE FIGURAS

LISTA DE TABLAS

ANEXOS

Anexo 1. Variables de los datos originales junto con las variables compuestas.

ID	Columna	Descripción	Rango
1	Orden	ID del episodio	571-3606708
2	centro	Centro de ingreso	1:14
3	hhcc text	Historia Clínica en formato texto	6
4	Cent-hhcc	Variable conjunta de centro con hhcc	7
5	HHCC	Historia Clínica numérica	6
6	fnac	Fecha nacimiento	dd/mm/aaaa 01/01/1925 17/12/2015
7	fing	Fecha de ingreso	dd/mm/aaaa 30/04/1992* 29/12/2015
8	falta	Fecha de alta	dd/mm/aaaa 31/01/2001 31/12/2015
9	edad	Edad del paciente	16-98
10	estancia	Número de días de ingreso	5-133
11	sexo	Sexo del paciente	1-varón 2-mujer
12	provincia	Código de provincia	5-Ávila 9-Burgos 24-León 34-Palencia 37-Salamanca 39-Cantabria (Santander) 40-Segovia 42-Soria 47-Valladolid 49-Zamora 99-Desconocido
13	cp	Diferencia entre cp	Código postal del paciente
15	ambito	Rural y urbano	1-2-3
16	tiping	Tipo ingreso	1-2
17	tipalta	Tipo de alta (Solo nos interesa 1 y 4)	1,4,5
19	fintervencion	Fecha intervención en caso de que se haya realizado.	dd/mm/aaaa
20	servicio	Servicio que da el alta	32 servicios diferentes
21	c1	Diagnóstico Principal	CIE-9
22	c2	Diagnóstico Secundario (opcional)	CIE-9
23	c3	Diagnóstico Secundario (opcional)	CIE-9

24	c4	Diagnóstico Secundario (opcional)	CIE-9
25	c5	Diagnóstico Secundario (opcional)	CIE-9
26	c6	Diagnóstico Secundario (opcional)	CIE-9
27	c7	Diagnóstico Secundario (opcional)	CIE-9
28	c8	Diagnóstico Secundario (opcional)	CIE-9
29	c9	Diagnóstico Secundario (opcional)	CIE-9
30	c10	Diagnóstico Secundario (opcional)	CIE-9
31	m1	Morfología de las neoplasias	CIE-9
32	m2	Morfología de las neoplasias	CIE-9
33	p1	Código procedimiento diagnóstico	CIE-9
34	p2	Código procedimiento diagnóstico	CIE-9
35	p3	Código procedimiento diagnóstico	CIE-9
36	p4	Código procedimiento diagnóstico	CIE-9
37	p5	Código procedimiento diagnóstico	CIE-9
38	p6	Código procedimiento diagnóstico	CIE-9
39	p7	Código procedimiento diagnóstico	CIE-9
40	p8	Código procedimiento diagnóstico	CIE-9
41	grd	Sistema de clasificación de pacientes compuesto por: -Edad (la calcula con la fecha de nacimiento y la fecha de ingreso) -Sexo -Circunstancias del alta (si el paciente está vivo o fallecido, se traslada a otro hospital o ha sido alta voluntaria). -Diagnóstico Principal (el motivo del ingreso) -Intervenciones u otros procedimientos realizados durante el ingreso. -Diagnósticos secundarios que coexisten con el principal en el momento del ingreso o se desarrollan durante el mismo.	GRD
42	pesoAP27_2014	Peso GRD	GRD
43	costeAP27_2014	Coste	GRD
44	EdadEU2013	Rangos de edad EU2013	1:17
45	yearalta	Año de alta	2001:2015
46	year	Año de ingreso	2000:2015
47	mes	Mes de ingreso	1:12
48	semana	Semana de ingreso	1:53
49	dia sem ing	Día de la semana de ingreso	1:7
50	dia sem alta	Día de la semana de alta	1:7
51	diayear	Día del año	1:366
52	meserie	Mes encadenado por año	1:180
53	exitus	Paciente fallecido durante ingreso	0-1

54	CDM	Categoría Diagnóstica Mayor	0:25
55	tipo grd	Tipo episodio (quirúrgico o médico)	0-1-2
56	tipo hospital	Tipo de hospital	1:3
57	Traqueotomia	Si hubo que practicar una traqueotomía o no	0-1
58	Vent Mecánica	Si el paciente utiliza ventilación mecánica	1
59	EPOC en C1	Enfermedad Pulmonar obstructiva Crónica como diagnóstico principal (variable c1)	0-1
60	ID paciente	Concatenando fecha de nacimiento, sexo y CP	13327
61	CARD	Patología cardiaca (Diagnóstico secundario)	0-1
62	RESP	Patología respiratoria (Diagnóstico secundario)	0-1
63	NEFRO	Patología renal (Diagnóstico secundario)	0-1
64	HEPAT	Patología hepática (Diagnóstico secundario)	0-1
65	InmunoDef	Inmunodeficiencias (Diagnóstico secundario)	0-1
66	Tm	Tumores (Diagnóstico secundario)	0-1
67	DM	Diabetes	0-1
68	EnfNM	Enfermedad Neuromuscular como diagnóstico principal o secundario	0-1
69	ObsVAeS	Obstrucción via aérea superior como diagnóstico principal o secundario	0-1
70	IRA	Insuficiencia Respiratoria Aguda. Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
71	REPOC	Reagudización Enf. Pulmonar Obstructiva Crónica. Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
72	Gripe	Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
73	BrA	Broncoaspiración. Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
74	SDRA	Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
75	TEP	Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
76	Asma	Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
77	BE	BroncoEspasmo Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
78	ICA	Insuficiencia Cardiaca Aguda Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
79	IM	Infarto de Miocardio Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1

80	MIO	Miocarditis Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
81	TrTo	Traumatismo Torácico Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
82	Coma	Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
83	ACVA	Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
84	TrCr	Traumatismo Craneoencefálico Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
85	TrCC	Traumatismo Columna Cervical Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
86	EsEp	Estatus epiléptico Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
87	Shock	Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
88	Quemaduras	Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
89	Intox	Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
90	PrDiagA	Procesos digestivos agudos Patología responsable del inicio de la ventilación mecánica (Diagnóstico principal)	0-1
91	AP01	Si un episodio dura más de 21 días	0-1
92	AP02	Enf. Neurológica central como causa de la necesidad de Ventilación Mecánica.	0-1
93	AP03	Neuromuscular como causa de la necesidad de Ventilación Mecánica	0-1
94	AP04	Enf. Pulmonar Obstructiva crónica (EPOC) como causa de la Insuf. Respiratoria	0-1
95	AP05	Neumonía como causa de la Insuf. Respiratoria	0-1
96	AP06	Insuficiencia cardíaca como causa de la Insuf. Respiratoria	0-1
97	AP07	Obstrucción vía aérea superior	0-1
98	AP08	“a priori” global una o varias de las anteriores	0-1
99	NeAVM01	Neumonía Asociada a Ventilación Mecánica	0-1
100	NeAVM02	Neumonía Asociada a Ventilación Mecánica por estafilococos	0-1
101	NeAVM03	Neumonía Asociada a Ventilación Mecánica >96 horas	0-1
102	ComTraq	Complicaciones Traqueotomía	0-1
103	EdadCat	Edad categorizada <40/40:70/>70	0-1

Anexo 2. Transformación de información de diagnósticos y procedimientos a variables categóricas más generales.

Patologías en diagnósticos secundarios (Creamos nuevas variables si está presente o no (0/1)). Indicado por los facultativos.

Comorbilidades (Diagnósticos Secundarios)

Nombre de la variable	Descripción	Códigos CIE-9
CARD (Patología cardiaca):		414, 414.0, 414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.1, 414.10, 414.11, 414.12, 414.19, 414.2, 414.3, 414.4, 414.8, 414.9, 425, 425.0 ,425.1 ,425.11 ,425.18 ,425.2 ,425.3 ,425.4 ,425.5 ,425.7 ,425.8 ,425.9, 427.3, 427.31, 427.32, 428, 428.0, 428.1, 428.2, 428.20, 428.21, 428.22, 428.23, 428.3, 428.30, 428.31, 48.32, 428.33, 428.4, 428.40, 428.41, 428.42, 428.43, 428.9.
RESP (Patología respiratoria):		518.83, 491.0, 491.1, 491.20, 491.2, 491.8, 491.9, 491, 491.21, 491.22, 492.0, 496, 505, 516.2, 516.8, 516.9, 516, 516.3, 516.30, 516.31, 516.32, 516.33, 516.34, 516.35, 516.36, 516.37, 516.4, 516.5, 516.6, 516.61, 516.62, 516.63, 516.64, 516.69, 517, 517.1, 517.2, 517.3, 517.8.
NEFRO (Patología renal):		585.1 585.2 585.3 585.4 585.5 585.6 585.7 585.8 585.9
HEPAT (Patología hepática):		571 571.0 571.1 571.2 571.3 571.4 571.40 571.41 571.42 571.49 571.5 571.6 571.8 571.9.
InmunoDef (Inmunodeficiencias):		V08 042 710 710.0 710.1 710.2 710.3 710.4 710.5 710.8 710.9.
Tm (Tumores):		199 199.0 199.1 238.75 239 239.0 239.1 239.2 239.3 239.4 239.5 239.6 239.7

		239.8 239.81 239.89 239.9 200 200.0 200.00 200.01 200.02 200.03 200.04 200.05 200.06 200.07 200.08 200.1 200.10 200.11 200.12 200.13 200.14 200.15 200.16 200.17 200.18 200.2 200.20 200.21 200.22 200.23 200.24 200.25 200.26 200.27 200.28 200.3 200.30 200.31 200.32 200.33 200.34 200.35 200.36 200.37 200.38 200.4 200.40 200.41 200.42 200.43 200.44 200.45 200.46 200.47 200.48 200.5 200.50 200.51 200.52 200.53 200.54 200.55 200.56 200.57 200.58 200.6 200.60 200.61 200.62 200.63 200.64 200.65 200.66 200.67 200.68 200.7 200.70 200.71 200.72 200.73 200.74 200.75 200.76 200.77 200.78 200.8 200.80 200.81 200.82 200.83 200.84 200.85 200.86 200.87 200.88 201 201.0 201.00 201.01 201.02 201.03 201.04 201.05 201.06 201.07 201.08 201.1 201.10 201.11 201.12 201.13 201.14 201.15 201.16 201.17 201.18 201.2 201.20 201.21 201.22 201.23 201.24 201.25 201.26 201.27 201.28 201.4 201.40 201.41 201.42 201.43 201.44 201.45 201.46 201.47 201.48 201.5 201.50 201.51 201.52 201.53 201.54 201.55 201.56 201.57 201.58 201.6 201.60 201.61 201.62 201.63 201.64 201.65 201.66 201.67 201.68 201.7 201.70 201.71 201.72 201.73 201.74 201.75 201.76 201.77 201.78 201.9 201.90 201.91 201.92 201.93 201.94 201.95 201.96 201.97 201.98 202 202.0 202.00 202.01

		202.02 202.03 202.04 202.05 202.06 202.07 202.08 202.1 202.10 202.11 202.12 202.13 202.14 202.15 202.16 202.17 202.18 202.2 202.20 202.21 202.22 202.23 202.24 202.25 202.26 202.27 202.28 202.3 202.30 202.31 202.32 202.33 202.34 202.35 202.36 202.37 202.38 202.4 202.40 202.41 202.42 202.43 202.44 202.45 202.46 202.47 202.48 202.5 202.50 202.51 202.52 202.53 202.54 202.55 202.56 202.57 202.58 202.6 202.60 202.61 202.62 202.63 202.64 202.65 202.66 202.67 202.68 202.7 202.70 202.71 202.72 202.73 202.74 202.75 202.76 202.77 202.78 202.8 202.80 202.81 202.82 202.83 202.84 202.85 202.86 202.87 202.88 202.9 202.90 202.91 202.92 202.93 202.94 202.95 202.96 202.97 202.98 203 203.0 203.00 203.01 203.02 203.1 203.10 203.11 203.12 203.8 203.80 203.81 203.82 204 204.0 204.00 204.01 204.02 204.1 204.10 204.11 204.12 204.2 204.20 204.21 204.22 204.8 204.80 204.81 204.82 204.9 204.90 204.91 204.92 205 205.0 205.00 205.01 205.02 205.1 205.10 205.11 205.12 205.2 205.20 205.21 205.22 205.3 205.30 205.31 205.32 205.8 205.80 205.81 205.82 205.9 205.90 205.91 205.92 206 206.0 206.00 206.01 206.02 206.1 206.10 206.11 206.12 206.2 206.20 206.21 206.22 206.8 206.80 206.81 206.82 206.9 206.90 206.91 206.92 207 207.0 207.00 207.01 207.02 207.1 207.10

		207.11 207.12 207.2 207.20 207.21 207.22 207.8 207.80 207.81 207.82 208 208.0 208.00 208.01 208.02 208.1 208.10 208.11 208.12 208.2 208.20 208.21 208.22 208.8 208.80 208.81 208.82 208.9 208.90.
DM(Diabetes):		250.00 250.01 250.90.
Patologías tanto en diagnósticos principales como secundarios.		
EnfNM (Enfermedad Neuromuscular):		357.0 358.0 359 359.0 359.1 359.2 359.21 359.22 359.23 359.24 359.29 359.3 359.4 359.5 359.6 359.7 359.71 359.79 359.8 359.81 359.89 359.9 335.20 340 519.4 952.00 952.9.
ObsVAeS (Obstrucción Vía aérea Superior):		465 465.0 465.8 465.9 464.31 464.51.
Complicaciones en relación con traqueotomía o Inserción tubo endotraqueal (Diagnósticos secundarios)		
Neumonía Asociada a Ventilación Mecánica (NeAVM):		997.31 482.9 486.
Complicaciones Traqueotomía (ComTraq):		519.0 519.1 519.2 519.3 519.4 519.9 519 519.8 519.00 519.01 519.02 519.09 519.11 519.19.
Patología responsable del inicio de la Ventilación Mecánica (Diagnóstico Principal)		
Insuf. Respiratoria Aguda (IRA):		518.81
Reagudización Enf. Pulmonar Obstructiva Crónica (REPOC)		491.25,518.84.
Neumonía		482.0 482.2 482.30 482.31 482.32 482.39 482.4 482.81 482.82 482.83 482.89 482.9 482 482.1 482.3 482.8 482.40 482.49 482.84 482.41 482.42 483.0 483.8 483

		483.1 486.
Gripe:		488 488.0 488.1 488.01 488.02 488.09 488.11 488.12 488.19 488.8 488.81 488.82 488.89 487.0 487.1 487.8 487.
Broncoaspiración (BrA)		507.0, 997.32.
SDRA		518.82 518.52
TEP		415.0 415.1 415.19
Asma		493.00 493.01 493.10 493.1 493.20 493.21 493.2 493.90 493.9 493.12 493.22 493.92 493.11 493.91 493.02 493.8 493.81 493.82.
Broncoespasmo (BE):		519.11, 519.19
Insuf. Cardiaca Aguda (Edema Agudo de Pulmón) (ICA):		428.1 428.21 428.31 428.41 428.9 518.4
Infarto de Miocardio (IM)		410.01 410.02 410.0 410.10 410.11 410.1 410.20 410.21 410.22 410.2 410.31 410.32 410.3 410.40 410.41 410.4 410.50 410.51 410.52 410.5 410.61 410.62 410.6 410.70 410.71 410.7 410.80 410.81 410.82 410.8 410.91 410.92 410.9 410 410.00 410.12 410.30 410.42 410.60 410.72 410.90.
Miocarditis (MIO)		422.0 422.90 422.91 422.92 422.93 422.9 422 422.99

Traumatismo Torácico (TrTo):		518.5, 518.52, 959.1.
Coma		780.01 250.3 572.2.
-ACVA:		431 432.0 432.1 432.9 432 434.00 434.01 434.0 434.10 434.11 434.90 434.91 434.9 434 434.1 348.31 320.0 320.1 320.3 320.7 320.81 320.82 320.89 320.9 320 320.2 320.8 321.0 321.1 321.2 321.4 321.8 321 321.3 322.0 322.1 322.9 322 322.2 047.9 054.3 049.9 323.0 323.1 323.2 323.5 323.6 323.7 323.8 323.9 323.4 323 323.01 323.42 323.51 323.52 323.61 323.02 323.41 323.62 323.63 323.71 323.72 323.81 323.82 324.0 324.1 324.9 324.
Traumatismo Craneoencefálico (TrCr)		851.00 851.02 851.03 851.04 851.05 851.06 851.0 851.10 851.11 851.12 851.13 851.15 851.16 851.19 851.1 851.20 851.22 851.23 851.24 851.25 851.26 851.2 851.30 851.31 851.32 851.33 851.35 851.36 851.39 851.3 851.40 851.42 851.43 851.44 851.45 851.46 851.4 851.50 851.51 851.52 851.53 851.55 851.56 851.59 851.5 851.60 851.62 851.63 851.64 851.65 851.66 851.6 851.70 851.71 851.72 851.73 851.75 851.76 851.79 851.7 851.80 851.82 851.83 851.84 851.85 851.86 851.8 851.90 851.91 851.92 851.93 851.95 851.96 851.99 851.9 851 851.01 851.09 851.14 851.21 851.29 851.34

		851.41 851.49 851.54 851.61 851.69 851.74 851.81 851.89 851.94 854.00 854.01 854.02 854.03 854.05 854.06 854.09 854.0 854.10 854.12 854.13 854.14 854.15 854.16 854.1 854 854.04 854.11 854.19.
Traumatismo Columna Cervical (TrCC):		952.01 952.02 952.03 952.04 952.05 952.07 952.08 952.09 952.0 952.10 952.12 952.13 952.14 952.15 952.16 952.18 952.19 952.1 952.2 952.3 952.8 952.9 952 952.00 952.06 952.11 952.17 952.4.
Estatus Epiléptico (EsEp)		345.00 345.01 345.0 345.11 345.1 345.2 345.3 345.40 345.4 345.50 345.51 345.5 345.60 345.6 345.70 345.71 345.7 345.80 345.8 345.90 345.91 345.9 345 345.10 345.41 345.61 345.81 293.0 293.1 293.82 293.83 293.89 293.8 293.9 293.81 293 293.84 780.39.
Preeclampsia-Eclampsia (Pree):		642.41 642.42 642.43 642.44 642.4 642.51 642.52 642.53 642.54 642.5 642.61 642.62 642.63 642.64 642.6 642.71 642.72 642.73 642.74 642.7.
Shock:		785.59, 785.5 ,785.51, 785.52.
Quemaduras		941.00 941.01 941.02 941.04 941.05 941.06 941.07 941.08 941.0 941.10 941.11 941.12 941.13 941.15 941.16 941.17 941.18 941.19 941.20 941.21 941.22 941.23 941.24 941.26 941.27

		941.28 941.29 941.2 941.31 941.32 941.33 941.34 941.35 941.37 941.38 941.39 941.3 941.40 941.42 941.43 941.44 941.45 941.46 941.48 941.49 941.4 941.50 941.51 941.53 941.54 941.55 941.56 941.57 941.59 941.5 941 941.03 941.09 941.14 941.1 941.25 941.30 941.36 941.41 941.47 941.52 941.58 942.00 942.01 942.03 942.04 942.05 942.09 942.0 942.11 942.12 942.13 942.14 942.15 942.1 942.20 942.21 942.22 942.23 942.25 942.29 942.2 942.30 942.31 942.33 942.34 942.35 942.39 942.3 942.41 942.42 942.43 942.44 942.45 942.4 942.50 942.51 942.52 942.53 942.55 942.59 942.5 942 942.02 942.10 942.19 942.24 942.32 942.40 942.49 942.54 943.00 943.02 943.03 943.04 943.05 943.06 943.0 943.10 943.11 943.12 943.13 943.15 943.16 943.19 943.1 943.20 943.22 943.23 943.24 943.25 943.26 943.2 943.30 943.31 943.32 943.33 943.35 943.36 943.39 943.3 943.40 943.42 943.43 943.44 943.45 943.46 943.4 943.50 943.51 943.52 943.53 943.55 943.56 943.59 943.5 943 943.01 943.09 943.14 943.21 943.29 943.34 943.41 943.49 943.54 944.01 944.02 944.03 944.04 944.05 944.07 944.08 944.0 944.10 944.11 944.13 944.14 944.15 944.16 944.17 944.1 944.20 944.21 944.22 944.23 944.25 944.26 944.27 944.28 944.2

		944.31 944.32 944.33 944.34 944.35 944.37 944.38 944.3 944.40 944.41 944.43 944.44 944.45 944.46 944.47 944.4 944.50 944.51 944.52 944.53 944.55 944.56 944.57 944.58 944.5 944 944.00 944.06 944.12 944.18 944.24 944.30 944.36 944.42 944.48 944.54 945.00 945.01 945.02 945.03 945.04 945.06 945.09 945.0 945.10 945.11 945.13 945.14 945.15 945.16 945.19 945.20 945.21 945.22 945.23 945.24 945.26 945.29 945.2 945.30 945.31 945.33 945.34 945.35 945.36 945.39 945.40 945.41 945.42 945.43 945.44 945.46 945.49 945.4 945.50 945.51 945.53 945.54 945.55 945.56 945.59 945 945.05 945.12 945.1 945.25 945.32 945.3 945.45 945.52 945.5 946.0 946.1 946.2 946.3 946.5 946 946.4 947.0 947.1 947.2 947.4 947.8 947.9 947 947.3 948.00 948.10 948.11 948.1 948.20 948.21 948.2 948.30 948.31 948.32 948.33 948.40 948.41 948.42 948.43 948.44 948.50 948.51 948.52 948.53 948.54 948.5 948.60 948.61 948.62 948.63 948.65 948.66 948.6 948.70 948.71 948.73 948.74 948.75 948.76 948.77 948.80 948.81 948.82 948.83 948.84 948.86 948.87 948.88 948.8 948.90 948.92 948.93 948.94 948.95 948.96 948.98 948.99 948.9 948 948.0 948.22 948.3 948.4 948.55 948.64 948.72 948.7 948.85 948.91 948.97 949.0 949.2 949.3 949.4 949.5

		949 949.1.
Intoxicaciones (Intox):		969.00 969.01 969.02 969.03 969.04 969.05 969.09 969.70 969.71 969.72 969.73 969.79 960.0 960.6 961.1 961.7 962.2 962.8 963.3 964.0 964.6 965.01 965.5 966.0 967.0 967.6 968.2 968.9 969.4 969 969.0 969.1 969.2 969.3 969.5 969.6 969.7 969.8 969.9 965.61 965.69 960.1 960.2 960.3 960.4 960.5 960.7 960.8 960.9 960 961.0 961.2 961.3 961.4 961.5 961.6 961.8 961.9 961 962.0 962.1 962.3 962.4 962.5 962.6 962.7 962.9 962 963.0 963.1 963.2 963.4 963.5 963.8 963.9 963 964.1 964.2 964.3 964.4 964.5 964.7 964.8 964.9 964 965.00 965.02 965.09 965.0 965.1 965.4 965.6 965.7 965.8 965.9 965 966.1 966.2 966.3 966.4 966 967.1 967.2 967.3 967.4 967.5 967.8 967.9 967 968.0 968.1 968.3 968.4 968.5 968.6 968.7 968 970.0 970.1 970.8 970.9 970 971.1 971.2 971.3 971.9 971 972.1 972.2 972.3 972.4 972.5 972.7 972.8 972.9 972 973.0 973.2 973.3 973.4 973.5 973.6 973.9 973 974.0 974.1 974.2 974.4 974.5 974.6 974.7 974 975.1 975.2 975.3 975.4 975.5 975.7 975.8 975 976.0 976.1 976.3 976.4 976.5 976.6 976.7 976.9 976 977.0 977.1 977.2 977.4 977.8 977.9 977 978.0 978.2 978.3 978.4 978.5 978.6 978.9 978 979.0 979.1 979.2 979.4 979.5 979.6 979.7 979.9 971.0 972.0 972.6 973.1 973.8 974.3 975.0 975.6 976.2 976.8 977.3

		978.1 978.8 979.3 979 970.81 970.89 980.0 980.1 980.2 980.3 980.9 980 981 982.0 982.1 982.3 982.4 982.8 982 983.0 983.2 983.9 983 984.0 984.1 984.9 984 985.0 985.1 985.2 985.4 985.5 985.6 985.8 985.9 986 987.0 987.1 987.2 987.3 987.5 987.6 987.7 987.8 987.9 988.0 988.1 988.2 988.8 988.9 989.0 989.1 989.2 989.3 989.4 989.6 989.7 989.81 989.82 989.83 989.89 989.8 989.9 989 980.8 982.2 983.1 984.8 985.3 985 987.4 987 988 989.5 989.84 991.0 991.1 991.2 991.3 991.4 991.6 991.8 991.9 991 992.0 992.2 992.3 992.4 992.5 992.6 992.8 992.9 992 993.0 993.1 993.3 993.4 993.8 993.9 993 994.1 994.2 994.3 994.4 994.5 994.7 994.8 994.9 994 995.0 995.2 995.3 995.4 995.5 995.60 995.62 995.63 995.64 995.65 995.66 995.68 995.69 995.6 995.81 995.89 995 995.51 995.52 995.53 995.54 995.55 995.80 995.82 995.83 995.84 995.85 990 991.5 992.1 992.7 993.2 994.0 994.6 995.1 995.61 995.67 995.8 995.50 995.59 995.86 995.9 995.90 995.91 995.92 995.93 995.94 995.20 995.21 995.22 995.23 995.27 995.29 995.24 E850.0 E850.1 E850.3 E850.4 E850.5 E850.6 E850.7 E850.9 E850 E850.2 E850.8 E851 E850.0 E850.1 E850.3 E850.4 E850.5 E850.6 E850.7 E850.9 E850 E851 E852.0 E852.1 E852.3 E852.4 E852.5 E852.8 E852.9 E853.0 E853.1 E853.2 E853.8 E853.9

		E854.0E854.1E854.2 E854.3E854.8E855.2 E855.3E855.4E854 E855.0E855.6E855.8 E855.9E855 E856 E858.0E858.1E858.2 E858.3E858.4E858.6 E858.7E858.8E858.9E858 E850.2E850.8E852.2E852 E853 E855.1E855.5E857 E858.5E860.0E860.1 E860.2E860.3E860.4 E860.9E860 E861.0 E861.1E861.2E861.4 E861.5E861.6E861.9E861 E862.1E862.2E862.3 E862.4E862.9E863.0 E863.1E863.2E863.3 E863.4E863.6E863.7 E863.8E863.9E863 E864.1E864.2E864.3 E864.4E864 E865.1 E865.2E865.3E865.4 E865.5E865.9E865 E866.0E866.1E866.2 E866.4E866.5E866.6 E866.7E866.8E866 E867 E868.0E868.1E868.2 E868.8E868.9E868 E869.0E869.1E869.3 E869.4E869.8E869.9E869 E860.8E861.3E862.0E862 E863.5E864.0E865.0 E865.8E866.3E866.9 E868.3E869.2.
Sepsis		995.9 995.90 995.91 995.92 995.93 995.94.
Procesos Digestivos Agudos (PrDigA)		567 567.21 567.22 567.23 567.29 567.3 567.31 567.38 567.39 567.81 567.82 567.89 575.0 575.1 575.2 575.3 575.4 575.6 575.8 575.9 575 575.5 575.10 575.11 575.12 576.1 577.0 578.0 578.1 578.9 578.
Neumonía Asociada a Ventilación Mecánica		997.31

(Diagnóstico Secundario) (NeAVM01):		
Neumonía Asociada a Ventilación Mecánica (Diagnóstico Secundario) (NeAVM02):		997.31, 482.9, 486.
Neumonía en pacientes con Ventilación Mecánica + 96 horas con alguno de los códigos de neumonía (NeAVM03):		482, 482.0, 482.1, 482.4, 482.40, 482.41, 482.42, 482.49, 482.82, 482.83, 482.84, 482.89, 482.9.
Complicaciones Traqueotomía (ComTraq):		519.0, 519.1, 519.2, 519.3, 519.4, 519.9, 519, 519.8, 519.00, 519.01, 519.02, 519.09, 519.11, 519.19.

- ✓ 1. Duración de estancia hospitalaria ≥ 21 días (AP01).
- ✓ 2. Enf. Neurológica central como causa de la necesidad de Ventilación Mecánica (AP02)
- ✓ 3. Enf. Neuromuscular como causa de la necesidad de Ventilación Mecánica (AP03)
- ✓ 4. Enf. Pulmonar Obstructiva crónica (EPOC) como causa de la Insuf. Respiratoria (AP04)
- ✓ 5. Neumonía como causa de la Insuf. Respiratoria (AP05)
- ✓ 6. Insuficiencia cardíaca como causa de la Insuf. Respiratoria (AP06)
- ✓ 7. Obstrucción vía aérea superior. (AP07)
- ✓ 8. “a priori” global una o varias de las anteriores. (AP08).

Anexo 3. Listado de librerías usadas en R.

https://cran.r-project.org/web/packages/caret/caret.pdf

https://cran.r-project.org/web/packages/glmnet/index.html

https://cran.r-project.org/web/packages/InformationValue/InformationValue.pdf

https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf

https://cran.r-project.org/web/packages/MASS/MASS.pdf

https://cran.r-project.org/web/packages/pROC/pROC.pdf

https://cran.r-project.org/web/packages/sm/sm.pdf