

Proyecto: Calculadora de riesgos COVID-19

Alumno: Cristobal Bautista Henández, cristobal@ciencias.unam.mx

Profesor: Graciela Martínez Sánchez, graciela.mtz@ciencias.unam.mx

Ayudante: José de Jesús Ojeda González, jesusojeda@ciencias.unam.mx

Materia: Seminario de Estadística (Modelos Lineales Generalizados)



1 Introducción.

El presente proyecto tiene el objetivo de determinar la probabilidad de que una persona sea intubada a partir de sus condiciones. Así mismo, se quiere determinar la cantidad de días que en que un paciente muere a partir de una condición inicial, contando desde la fecha de inicio de los síntomas hasta su defunción.

2 Explicación de las variables.

Los datos de dicho proyecto son provenientes de la base de datos pública del gobierno con respecto al COVID-19, que se encuentran en el siguiente link:

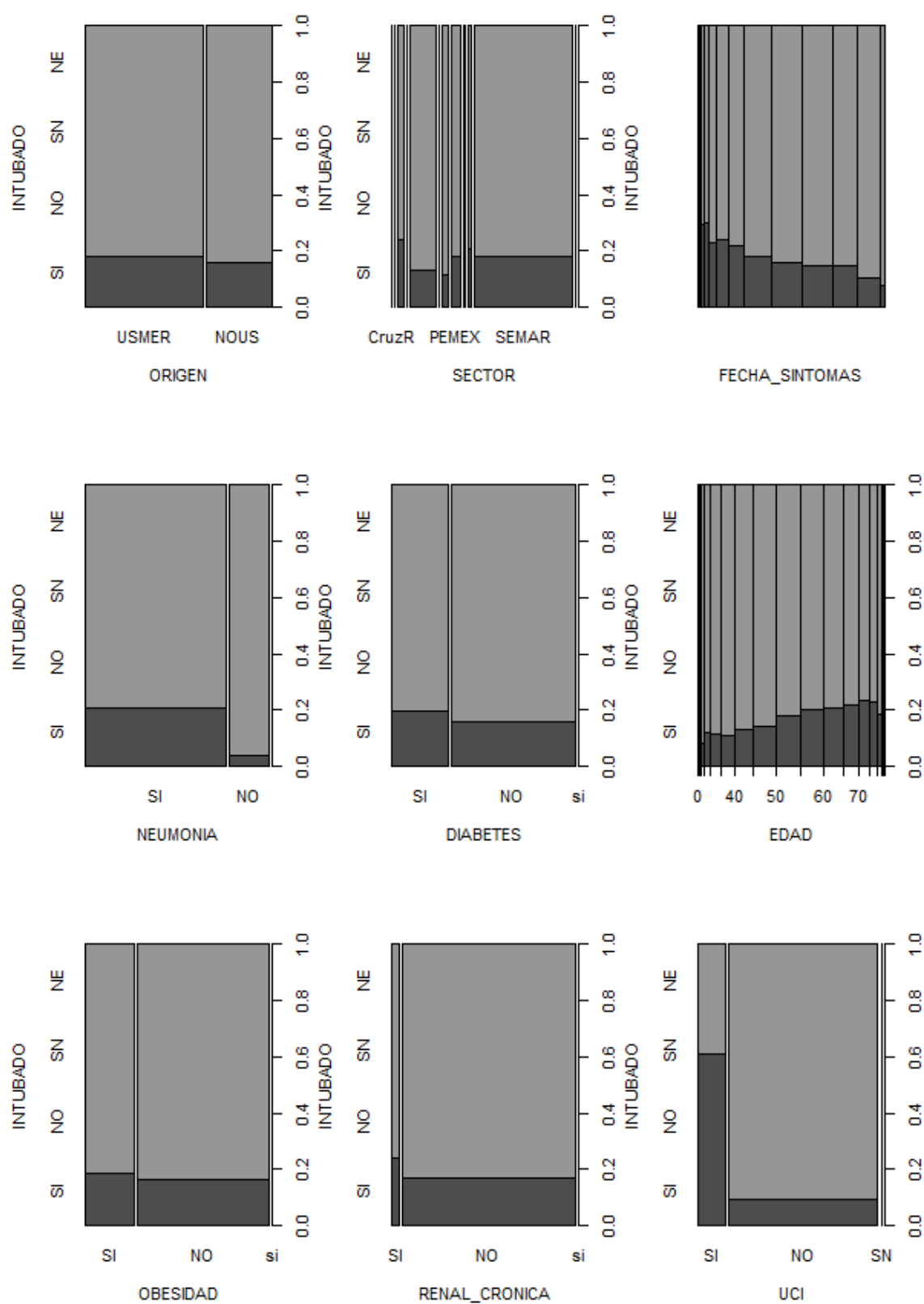
<https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico>

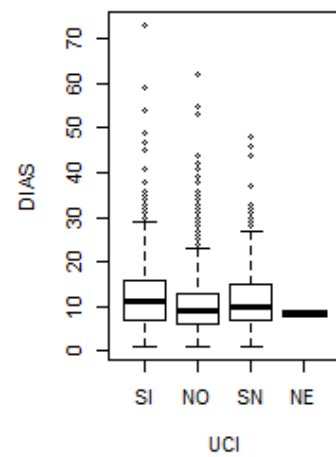
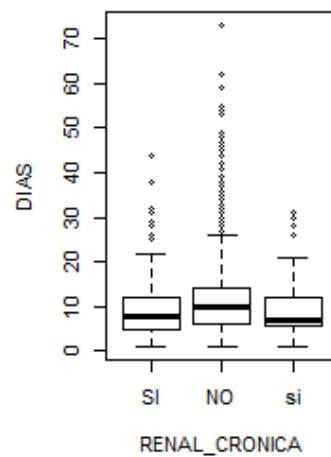
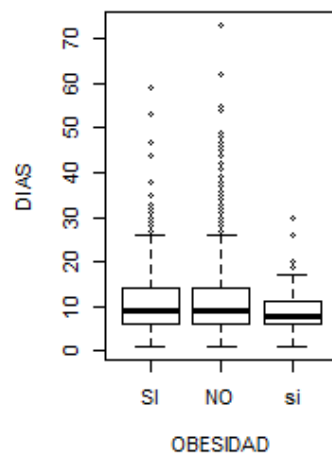
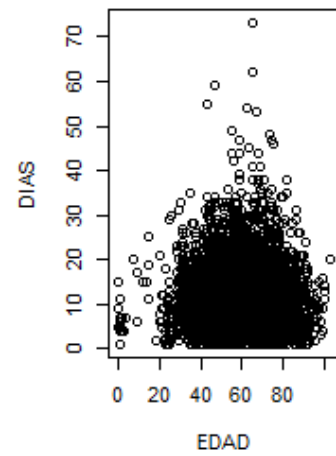
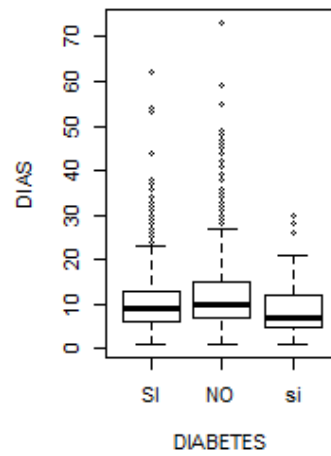
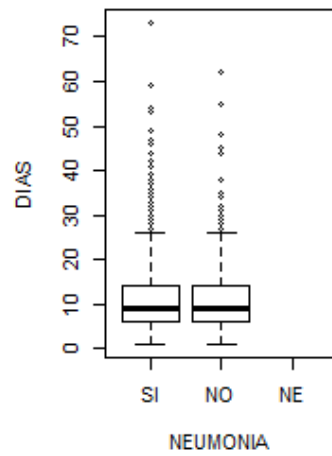
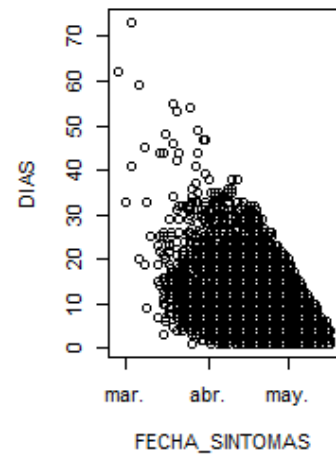
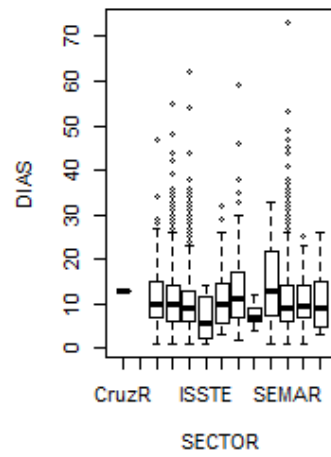
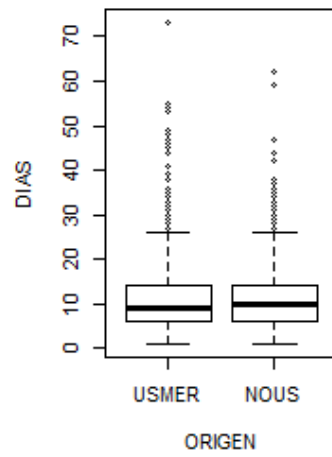
En el cual se encuentran 36 datos que recopilan de los pacientes que se les realizó o realizara una prueba al virus de COVID-19, todas estas variables están contenidas en tabla denominada variables.

3 Análisis exploratorio

Inicialmente se realiza un análisis para observar que variables podrían estar relacionadas con las variables dependientes. De lo cual para el primer caso, en el que nos interesa modelar la probabilidad de ser intubado. En la primera pagina se encuentra en el siguiente grupo de gráficos, es un análisis general de como lucen los datos en relación con las persona intubadas. De la cual, se puede notar una ligera relación, en especial con las variables Edad, Fecha (de inicio de los) síntomas, Neumonía, Diabetes y Sexo.

Por otro lado, para el segundo modelo en el cual se espera obtener la cantidad de días que una persona tarda en morir a partir de una variable se contrastan con las mismas variables del modelo anterior y corresponde a la segunda pagina de gráficos, en esta se encuentran variables como el Sector, Renal crónica o diabetes





4 Selección de modelo

Para la selección del modelo que predice si una persona está intubada a partir de variables reportadas por el sistema de salud. Lo que se usa es un modelo logístico puesto que lo que nos interesa es si la persona necesita ser intubada o no, Para ello, se toman los casos de personas intubadas o aquellas que no lo son pero son positivos.

Se hicieron modelos contemplando todas las variables, las cuales al probarlas con el modelo con función ligando *logit*, y se seleccionaron aquellas variables que se podían asumir que no son cero, para ello, primero se utilizó el método *backward* para minimizar el número de variables que explican la variables intubado y seguir siendo un modelo apropiado. Por tanto, las variables tienen relevancia para cambiar el modelo dependiendo del valor. Una vez seleccionada estas variables, se comparó el mismo modelo pero con las funciones ligando *probit* y *doble complementaria*, las cuales se fueron modelando de la misma manera y que finalmente se compararon con las estadística de Hosmer y Lemeshow, ya que nuestros datos no están ordenados

De las cuales, se puede ver que el mejor modelo que se ajusta a predecir la probabilidad de que una persona sea intubada es *probit*. La cual, al tener un valor mayor valor en la prueba de rechazo para χ^2 y menor *p-value*, resulta como mejor modelo, a comparación de los otros dos. Aunque el modelo *logit* también resultaría un buen candidato.

Modelo	p-valor de la prueba Hosmer y Lemeshow	Valor estadístico χ^2	Grados de libertad
Logit	$7.911e^{-05}$	32.396	8
Probit	$5.601e^{-05}$	33.229	8
Logaritmo complementario	0.01743	18.555	8

Por otra parte para el modelo que estima el número de días que transcurre de la fecha de inicio de los síntomas hasta la muerte de dicho paciente, se decidió estudiarlo a partir de la condición de si la persona presentaba una condición renal crónica o no era el caso. En este modelo se decidió estudiarlo como regresión Poisson, puesto que se modela el número de días a transcurrir.

5 Análisis de Resultados

De acuerdo a los análisis realizados lo que se encuentre es que una manera de obtener si una persona sera intubada depende del siguiente modelo.

$$\widehat{\pi}_{intubado_i} = \Phi(-2.04176 + \beta_1 jSector_{ij} + \beta_2 jEntidad_UM_{ij} - 0.86787 Neumonia_i + 0.00705 Edad_i + \beta_5 jEmbarazo_{ij} - 1.59720 Habla_lengua_ind_i - 0.070175 Diabetes_i - 1.49562 EPOC_i - 0.10715 Obesidad_i - 0.23497 Cronica_renal_i - 1.77718 UCI_i)$$

Neumonia, *Diabetes*, *Obesidad*, *Cronica_renal*, *Habla_lengua_ind*, *EPOC* y *UCI* son 0 si no se presenta y 1 si se presenta.

Para las variables *Entidad_UM*, *Embarazo* y *Sector* se presentan en una tabla en el anexo, donde se ven los coeficientes del modelo, y en el cual toma el valor 1, cuando se tiene la condición *j*.

De esta manera, lo que nos dice es que una persona va incrementando su probabilidad de ser intubada en algún momento si presenta algunas de las enfermedades descritas en el modelo,

sobre todo neumonia que contiene el coeficiente más grande. A su vez, se es más sensible si se es hombre y/o va incrementando su edad y cuando al parecer la cada vez se tardan más en atender personas, puesto que como avanza el tiempo, la probabilidad desciende ligeramente.

Poniendo un ejemplo, si el paciente accedió a un hospital del IMSS, en Guanajuato con un cuadro de neumonía a la edad de 60 años habla lengua indígena, presenta una enfermedad renal crónica y entro a la UCI, entonces su probabilidad de ser intubada cambia es:

$$\Phi(-2.04176+2.66213(1)-0.10032(1)+0.00705(60)+0.24767(1)-0.070175(1)-1.49562-0.10715(1)) \\ = \Phi(-2.04176+2.66213-0.10032+0.00705(60)+0.24767-0.070175-1.49562-0.10715) = 0.3148231$$

La cual quiere decir que un 31% de probabilidad que esta persona sea intubada. Comparando con una mujer embarazada de 35 que padece de todas las condiciones de una mala salud, que fue a un hospital de la SEDENA en la CDMX, ingreso a una UCI y habla una lengua indígena.

$$\Phi(-2.04176 + 2.63896(1) + 0.02022(1) + 0.00705(35)) = 0.8062527$$

Lo cual nos dice que hay una probabilidad del 80% de que la persona sea intubada.

Por otro lado, para encontrar el promedio de días que en que una persona muere dado que sufre una condición renal crónica. Lo cual, mueve a analizar el segundo modelo del segundo objetivo

Se tiene el siguiente modelo Poisson:

$$\mu_i = \exp[2.17383 + 0.20897 \text{ Cronica_renal}_j] = 8.791629 \quad \text{para } j = 1, 2 \\ \mu_i = \exp[2.17383] = 8.791629 \quad \text{para } j = 1 \text{ o } \text{Cronico_renal} = \text{si} \\ \mu_i = \exp[2.17383 + 0.20897] = 10.83487 \quad \text{para } j = 2 \text{ o } \text{Cronico_renal} = \text{no}$$

Con lo que se puede visualizar que hay un promedio de 9 días entre la fecha en que se empieza a mostrar síntomas y en la que la persona fallece. Por otra parte, en el caso de presentar una enfermedad crónica renal, este número de días incrementa hasta en un 23%

6 Conclusiones

Como se puede ver, de acuerdo a lo que se ha informado a lo largo de esta epidemia, estar afectados con enfermedades crónicas se vuelve un factor importante en la salud del mismo, puesto que, tiene mayor probabilidad de ser intubado o de morir de manera más rápida.

A su vez, se muestra que en general los servicios de salud tienen la capacidad para intubar a los pacientes que sean necesarios, a excepción de los universitarios. Mientras que la edad afecta ligeramente mientras va incrementando.

7 Material

Se usaron las bases de datos proveniente del url ya citado Todo esto se analizo en el programa de R y se utilizaron las paqueterías *MASS* y *ResourceSelection*.

8 Anexo

Variables 1: Las variables que se presentan en la base de datos. Para los casos *NE*, *SN* y *si* son *No especificado*, *No Aplica* y *Sin información*

NOMBRE DE VARIABLE	DESCRIPCIÓN DE VARIABLE	FORMATO O FUENTE
FECHA_ACTUALIZACION	La base de datos se alimenta diariamente, esta variable permite identificar la fecha de la última actualización.	AAAA-MM-DD
ID_REGISTRO	Número identificador del caso	TEXTO
ORIGEN	La vigilancia centinela se realiza a través del sistema de unidades de salud monitoras de enfermedades respiratorias (USMER). Las USMER incluyen unidades médicas del primer, segundo o tercer nivel de atención y también participan como USMER las unidades de tercer nivel que por sus características contribuyen a ampliar el panorama de información epidemiológica, entre ellas las que cuenten con especialidad de neumología, infectología o pediatría. (Categorías en Catálogo Anexo).	USMER = 0 No USMER = 1
SECTOR	Identifica el tipo de institución del Sistema Nacional de Salud que brindó la atención.	IMSS IMSS-Bienestar Universitario SEDENA ISSTE etc.
ENTIDAD_UM	Identifica la entidad donde se ubica la unidad médica que brindó la atención.	TEXTO (Estados de la República Mexicana)
SEXO	Identifica al sexo del paciente.	Mujer = 0 Hombre = 1
ENTIDAD_NAC	Identifica la entidad de nacimiento del paciente.	TEXTO (Estados de la República Mexicana)
ENTIDAD_RES	Identifica la entidad de residencia del paciente.	TEXTO (Estados de la República Mexicana)
MUNICIPIO_RES	Identifica el municipio de residencia del paciente.	TEXTO (Municipios de la República Mexicana)
TIPO_PACIENTE	Identifica el tipo de atención que recibió el paciente en la unidad. Se denomina como ambulatorio si regresó a su casa o se denomina como hospitalizado si fue ingresado a hospitalización.	Ambulatorio = 0 Hospitalizado = 1
FECHA_INGRESO	Identifica la fecha de ingreso del paciente a la unidad de atención.	AAAA-MM-DD
FECHA_SINTOMAS	Identifica la fecha en que inició la sintomatología del paciente.	AAAA-MM-DD
FECHA_DEF	Identifica la fecha en que el paciente falleció.	AAAA-MM-DD
INTUBADO	Identifica si el paciente requirió de intubación.	SI = 0 NO = 1
NEUMONIA	Identifica si al paciente se le diagnosticó con neumonía.	SI = 0 NO = 1
EDAD	Identifica la edad del paciente.	NÚMERICA EN AÑOS
NACIONALIDAD	Identifica si el paciente es mexicano o extranjero.	NACIONALIDAD
EMBARAZO	Identifica si la paciente está embarazada.	SI = 0 NO = 1
HABLA_LINGUA_INDIG	Identifica si el paciente habla lengua indígena.	SI = 0 NO = 1
DIABETES	Identifica si el paciente tiene un diagnóstico de diabetes.	SI = 0 NO = 1
EPOC	Identifica si el paciente tiene un diagnóstico de EPOC.	SI = 0 NO = 1
ASMA	Identifica si el paciente tiene un diagnóstico de asma.	SI = 0 NO = 1
INMUSUPR	Identifica si el paciente presenta inmunosupresión.	SI = 0 NO = 1
HIPERTENSION	Identifica si el paciente tiene un diagnóstico de hipertensión.	SI = 0 NO = 1
OTRAS_COM	Identifica si el paciente tiene diagnóstico de otras enfermedades.	SI = 0 NO = 1
CARDIOVASCULAR	Identifica si el paciente tiene un diagnóstico de enfermedades cardiovasculares.	SI = 0 NO = 1
OBSIDAD	Identifica si el paciente tiene diagnóstico de obesidad.	SI = 0 NO = 1
RENAL_CRONICA	Identifica si el paciente tiene diagnóstico de insuficiencia renal crónica.	SI = 0 NO = 1
TABAQUISMO	Identifica si el paciente tiene hábito de tabaquismo.	SI = 0 NO = 1
OTRO_CASO	Identifica si el paciente tuvo contacto con algún otro caso diagnosticado con SARS CoV-2	SI = 0 NO = 1
RESULTADO	Identifica el resultado del análisis de la muestra reportado por el laboratorio de la Red Nacional de Laboratorios de Vigilancia Epidemiológica (INDRE, LESP y LAVE). (Catálogo de resultados diagnósticos anexo).	+ = Positivo - = Negativo ? = Sospechoso
MIGRANTE	Identifica si el paciente es una persona migrante.	SI = 0 NO = 1
PAIS_NACIONALIDAD	Identifica la nacionalidad del paciente.	TEXTO
PAIS_ORIGEN	Identifica el país del que partió el paciente rumbo a México.	TEXTO
UCI	Identifica si el paciente requirió ingresar a una Unidad de Cuidados Intensivos.	SI = 0 NO = 1

Coeficientes 2:

	Estimado	Error Estándar	Estadístico	p-value
(Intercept)	-2.04176781962493	57.9383016313097	-0.0352403809248279	0.971888062893245
SECTOREstatal	3.35408285845027	57.9360178333538	0.0578928787977437	0.953833955506345
SECTORIMSS	2.66213983712601	57.9373976848773	0.0459485572963675	0.963351251895489
SECTORIMSSB	2.79623956758372	57.9359761133944	0.0482643040674902	0.961505602610921
SECTORISSTE	1.93182835035793	57.9371285033709	0.0333435294475379	0.973400641563462
SECTORMunicipal	2.56550618646229	57.9360369793688	0.0442816996160072	0.964679858955133
SECTORPEMEX	2.46414678883821	57.9360105853215	0.0425322137983474	0.966074432079495
SECTORPrivada	2.54104248035815	57.9366489221572	0.0438589826583214	0.965016810914624
SECTORSEDENA	2.63896116187859	57.936080714622	0.0455495285377936	0.963669297773926
SECTORSEMAR	2.90561305322399	57.9359550416757	0.0501521559648729	0.960001137548556
SECTORSSA	4.08747834142811	57.9365247211399	0.070550975590993	0.943755128938688
SECTORUniversitario	-0.184087656441433	81.9337449711897	-0.00224678679713913	0.9982073250114
ENTIDAD_UMBC	-0.256877871661681	0.364723198581978	-0.704309110745921	0.481240293765192
ENTIDAD_UMBS	-0.0618013939709723	0.48964888234641	-0.126215735804028	0.899561155607611
ENTIDAD_UMCC	-1.07489915082474	0.474638556255564	-2.26466884465656	0.0235330087256736
ENTIDAD_UMCL	-0.355613956471047	0.503792041837034	-0.705874501658129	0.480266185722412
ENTIDAD_UMCM	-0.0598391462788802	0.58533142074823	-0.102231221762173	0.918573146192687
ENTIDAD_UMCS	-0.733707608942906	0.393408897376452	-1.86500003898189	0.0621813987446246
ENTIDAD_UMCH	-0.907099653464191	0.377337820469062	-2.40394575962884	0.0162191795883564
ENTIDAD_UMDF	0.0202296826122888	0.358045116336914	0.056500372967671	0.954943198416174
ENTIDAD_UMDG	0.158730193346174	0.553626530211112	0.286709875131248	0.774334475643244
ENTIDAD_UMGT	-0.100327439196462	0.390699902759303	-0.256789004778099	0.797341658448629
ENTIDAD_UMGR	-0.627215612331501	0.378392730330359	-1.6575783889503	0.0974025946569581
ENTIDAD_UMHG	-0.100165589712046	0.367883735603272	-0.272275123954012	0.785410482126675
ENTIDAD_UMJC	-0.116078259710986	0.408356428190757	-0.284257211831528	0.77621327415976
ENTIDAD_UMMC	-0.159510574012335	0.359680287578352	-0.443478776905692	0.657419466789493
ENTIDAD_UMMN	-0.234579390968748	0.394594384389099	-0.594482334896677	0.552189553983199
ENTIDAD_UMMS	-0.0026994860404764	0.367636856708305	-0.00734280579114587	0.994141341273004
ENTIDAD_UMNT	-0.518892754716739	0.453627866808219	-1.14387318920183	0.25267623138504
ENTIDAD_UMNL	-0.663743881628938	0.428196511521891	-1.55009175406373	0.121119494943094
ENTIDAD_UMOC	0.0462217535621078	0.379139786198892	0.121912168663461	0.902968578316846
ENTIDAD_UMPL	-0.462178299727245	0.365226532688228	-1.26545652728297	0.205707722730016
ENTIDAD_UMQT	-0.010043096550624	0.388439200833017	-0.0258550026080949	0.979372990752927
ENTIDAD_UMQR	-0.482737464721087	0.372774823772825	-1.29498408673456	0.195325719105188
ENTIDAD_UMSP	-0.479216199663853	0.445688649311676	-1.07522639493727	0.282273380806685
ENTIDAD_UMSL	-0.341949202480414	0.367123802351595	-0.931427492006985	0.351632475304281
ENTIDAD_UMSR	-0.83638702150793	0.423407422266558	-1.97537165747033	0.0482259867495965
ENTIDAD_UMTC	-1.02322711977477	0.37352759247835	-2.73936153681626	0.00615586385101062
ENTIDAD_UMTS	-0.98832226049235	0.45211488178769	-2.18599807328718	0.0288157462891343
ENTIDAD_UMTL	-0.849459144022615	0.391332185813889	-2.17068560884129	0.0299549457788721
ENTIDAD_UMVZ	-0.645662470963373	0.365526777815146	-1.76638897654141	0.0773306132101995
ENTIDAD_UMYN	-0.236910002690235	0.390107956267729	-0.607293439889866	0.543656191116246
ENTIDAD_UMZS	-0.763517487287565	0.451087245452474	-1.69261599609561	0.0905285842826254
NEUMONIANO	-0.867869552434172	0.0619656663126899	-14.0056519049557	1.43952703712064e-44
EDAD	0.00705265290857143	0.00119421133564989	5.90569918240925	3.51154224993972e-09
EMBARAZONO	0.12670732801135	0.339014606207479	0.373751825706898	0.708588963633029
EMBARAZOSN	0.247672931228954	0.338117383573813	0.73250576060632	0.463859927958791
HABLA LENGUA INDIGNO	-0.178841760668712	0.111971551363686	-1.59720713422849	0.110219542969902
DIABETESNO	-0.0701747062816444	0.0370696757545904	-1.89304882908113	0.0583513762114406
EPOCNO	-0.135918736801796	0.0908776357254028	-1.49562360108586	0.134751769336759
OBESIDADNO	-0.10715853709824	0.0385340803447098	-2.78087698317033	0.00542122757489137
RENAL_CRONICANO	-0.234977740818289	0.0811645710667507	-2.89507771346491	0.00379064674686537
UCINO	-1.77718304387105	0.0433268628046359	-41.0180412065491	0