

UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

DIPLOMADO EN CIENCIA DE DATOS

Nacimientos en México

Examen Final Módulo 1

Jessica Fernanda Rodríguez Mondragón

Septiembre 2020

Índice general

Resumen	II
Introducción	I
Calidad de datos	I
0.1. Diccionario de datos	I
0.2. Exploración de Datos	I
0.3. Análisis de valores ausentes	I
0.4. Identificación de Valores Atípicos	I
0.4.1. Método IQR	I
0.5. Ingeniería de Variables	I
0.6. Reducción de dimensiones	I
0.6.1. PCA	I
0.7. Transformación entrópica	II
0.7.1. WOE	II
0.8. Conclusión y Propuesta	II

Resumen

En el siguiente trabajo se pretende hacer un análisis exploratorio sobre las causas que influyen sobre la muerte en el momento del parto en las mujeres mexicanas, que han dado a luz en el 2018.

Introducción

¿Por qué siguen muriendo mujeres durante el parto?, si bien, se sabe que la tecnología en México en el sector salud ha avanzado exponencialmente, sin embargo, hoy en día siguen muriendo mujeres al momento de dar a luz, lo cuál se debe a un sin fin de causas, como las explicadas por la Organización Mundial de la Salud (OMS), ya que menciona que la mayoría de las muertes maternas se deben a hemorragias, infecciones, abortos peligrosos y eclampsia (hipertensión grave con convulsiones), o a complicaciones de enfermedades preexistentes que empeoran durante el embarazo.

Sin embargo otro de los factores es el tema del aborto legal en México, ya que la falta de cuidados o de acceso a este servicio y su elevado costo o su escasa calidad, o bien, que en algunas partes solo es brindado a mujeres mayores de edad son elementos determinantes, que llevan a la muerte a estas mujeres, además de la falta de acceso al sector salud que se tiene por comunidades, por lo que en este documento se presenta un análisis sobre los factores que implican que una mujer muera en el parto, más allá de los aquí mencionados.

Calidad de datos

0.1. Diccionario de datos

De primer instancia, no se cuenta con un diccionario de datos proporcionado, por lo que se mencionarán las variables de acuerdo a su análisis, sin embargo para mayor información consultar el diccionario de datos del INEGI en la siguiente liga:
<https://www.inegi.org.mx/rnm/index.php/catalog/530/datafile/F1>.

Conocinedo los nombres de las variables y lo que significan se presenta la siguiente clasificación de variables:

- Continuas = 'EDADM', 'NUM_EMB', 'NUM_NACMTO', 'NUM_NACVIVO', 'HIJO_SOBV', 'ORDEN_NAC', 'TOT_CONS', 'GESTACH'
- Categóricas: 'EDADM', 'EDOCIVIL', 'CON_INDM', 'HABLA_INDM', 'ENT_RES', 'NUM_EMB', 'HIJO_SOBV', 'HIJO_ANTE', 'VIVE_AUN', 'ORDEN_NAC', 'ATEN_PREN', 'SOB_PARTO', 'NIV_ESCOL', 'OCUPHAB', 'TRAB_ACT', 'SEXOH', 'GESTACH', 'INST_NAC', 'ENT_NAC'

0.2. Exploración de Datos

Dentro de este apartado, se muestran las características de nuestro conjunto de datos original, en primer instancia se cuenta con una dimension de 1,868,214, registros con 76 variables, sin embargo, para fines de este análisis, se reduce a trabajar con las variables descritas en el diccionario de datos, por lo que se reduce a 32 variables contando la ingenería de variables.

A continuación se presenta el comportamiento de las diferentes variables continuas que se tienen, con el fin de observar los fenómenos que nos indican las variables.

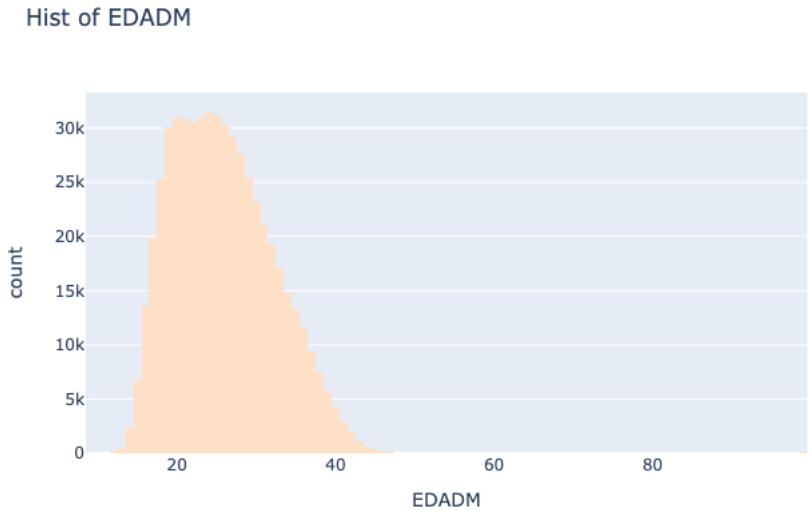


Figura 1:

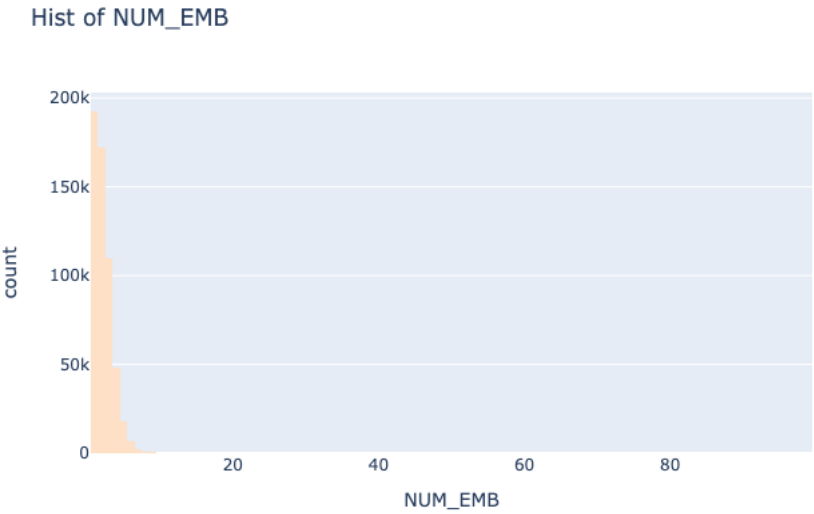


Figura 2:

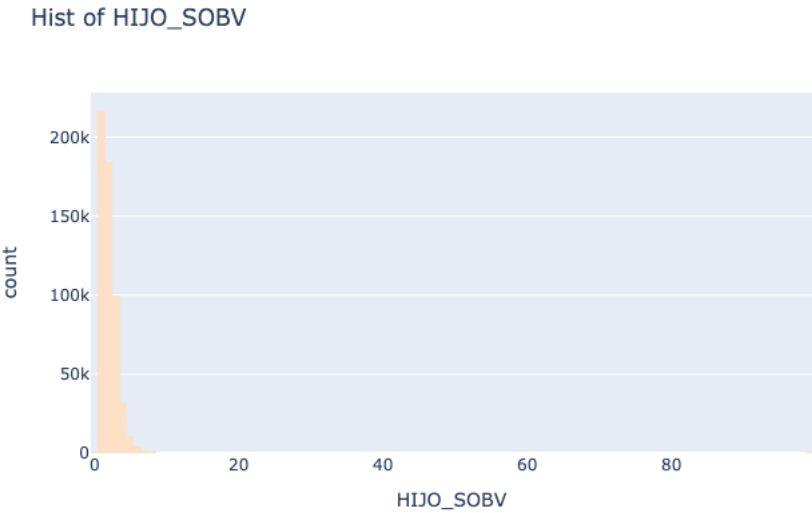


Figura 3:

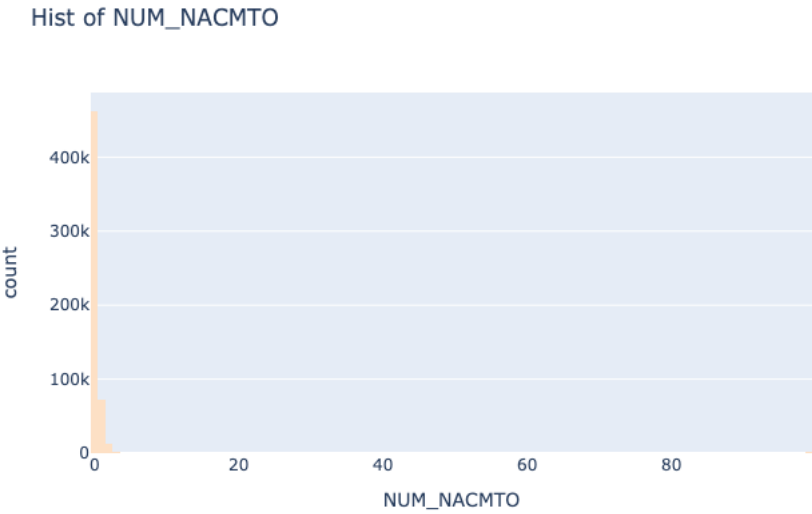


Figura 4:

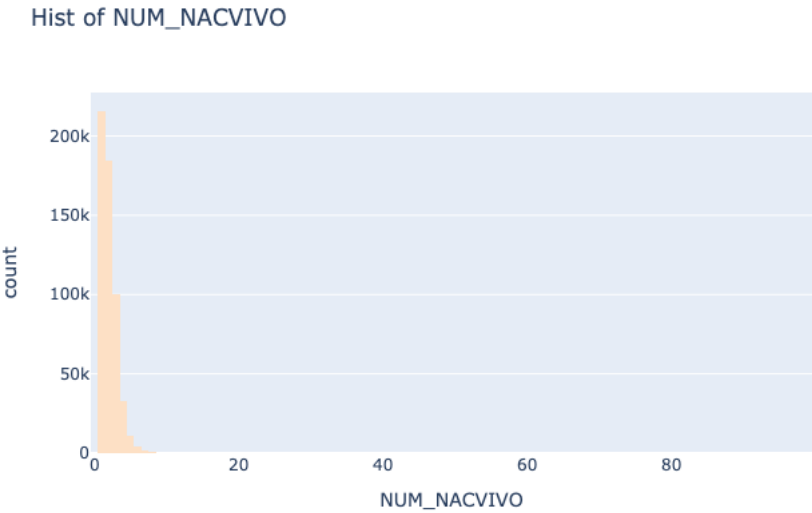


Figura 5:

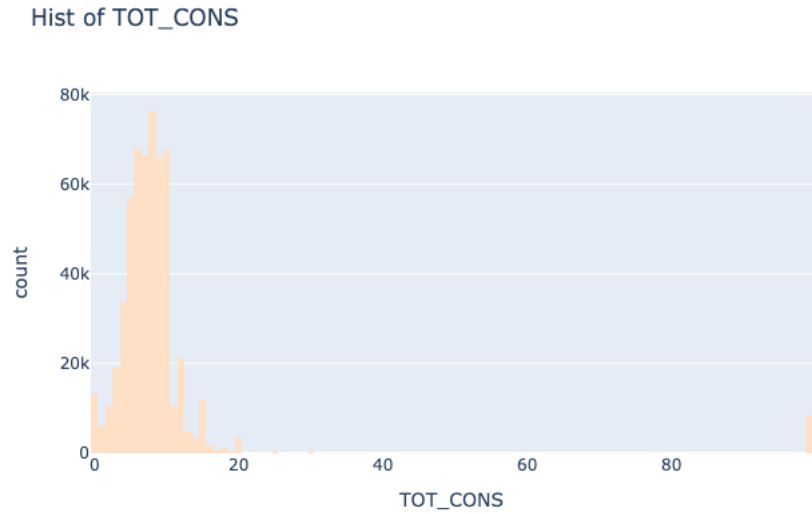


Figura 6:

Si bien, cabe mencionar que el total de bebés nacidos en 2018 fueron 1,868,214, de los cuales 1,867,085 sobrevivieron al momento del parto y 1,129 falleció o nacieron sin vida.

Tasa de evento, muerte en el parto

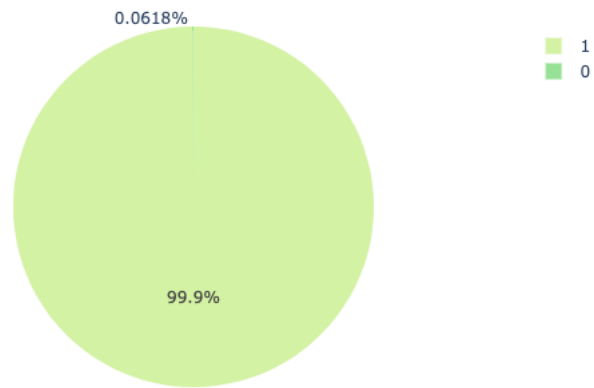


Figura 7:

0.3. Análisis de valores ausentes

Dentro de este apartado se encuentra el tratamiento realizado a la tabla de datos original sobre los valores ausentes, la siguiente imagen permite identificar de una manera visual los valores ausentes dentro de la tabla original de datos con solo las variables elegidas para el análisis:

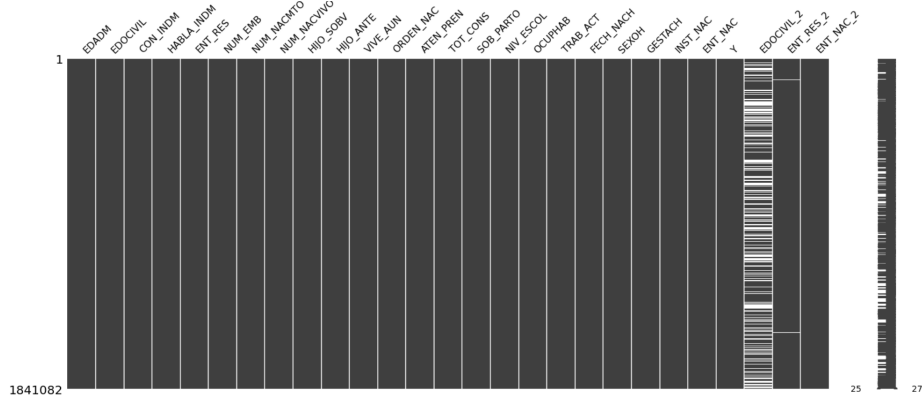


Figura 8:

Se observa que las variables de Estado Civil, Entidad Residencial, y Ocupación habitual de la madre, presentan valores ausentes, los cuales, al ser variables categóricas, se procede a distribuir uniformemente estos valores dentro de todas las categorías, y así no cambiar la distribución de estas variables.

0.4. Identificación de Valores Atípicos

Un valor atípico, o también conocido como outlier, tiene un comportamiento especial, y único a diferencia de los demás, existen técnicas estadísticas de identificación de estos valores, por lo que para poder identificarlos en los datos de los asegurados usaremos el método de IQR.

0.4.1. Método IQR

Los valores atípicos se caracterizan por estar en los extremos de las distribuciones de las variables, sin embargo, ante el problema que estamos observando, las defunciones a la hora del parto se dan justo por esos valores atípicos, es decir, si en la edad se tuviese un valor atípico ya sea mayor a 40 años, donde se considera embarazo de alto riesgo, o muy pequeño, menor a 18 años, es decir, con personas en las que el cuerpo podría no soportar el dar a luz, estaríamos eliminando la realidad del modelo al imputar o eliminar outliers, por lo que no se procede con esta técnica.

0.5. Ingeniería de Variables

Como se muestra en el apartado de diccionario de Datos, se cuenta con un gran número de variables que son útiles para realizar un análisis e incluso un modelo, sin embargo, a partir de las variables originales se pueden derivar o crear nuevas variables, a este proceso se le conoce como Ingeniería de Variables, el cual se muestra a continuación:

El siguiente listado, muestra las variables que dieron origen a otras nuevas:

- **GESTACH:**

1. prematuro: Si el recién nacido nació antes de las 34 semanas de gestación

- **'EDADM':**

1. menor_mayor_edad: Si la madre era mayor de edad o menor, al momento del parto

- **'FECHA_NACH'**

1. 'Año_nac': Año del nacimiento del bebé
2. 'mes_nac': mes del nacimiento del bebé
3. 'day_nac': día del nacimiento del bebé

0.6. Reducción de dimensiones

La Reducción de dimensiones sirve para representar de la manera más fidedigna posible la información de alta dimensionalidad en una dimensión mas baja, por lo que en esta sección se utilizará Análisis de Componentes Principales (PCA) por sus siglas en inglés para visualizar esta parte de los datos.

0.6.1. PCA

El método tiene por objeto transformar un conjunto de variables en un nuevo conjunto denominado componentes principales. Los nuevos componentes tienen la característica de ser incorrelacionados (ortogonales) y se ordenan de acuerdo a la cantidad de información (varianza) que llevan incorporada. Las componentes principales se expresan como una combinación lineal de las variables originales.

Las variables a usar son: 'EDADM', 'NUM_EMB', 'NUM_NACMTO', 'NUM_NACVIVO', 'HIJO_SOBV', 'ORDEN_NAC', 'TOT_CONS', 'GESTACH', 'Año_nac', 'mes_nac', 'day_nac'.

Las variables anteriores, explican aproximadamente el 50 % de la variable target, que en este caso es sí la madre fallece o no al momento de dar a luz, para 3 componentes principales.

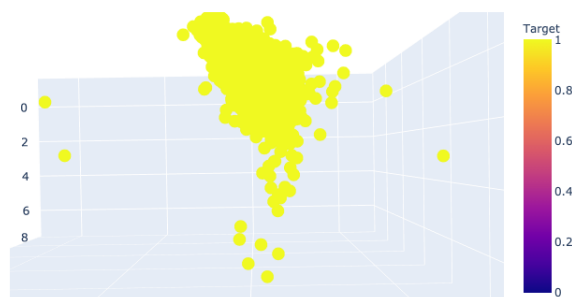


Figura 9:

0.7. Transformación entrópica

0.7.1. WOE

El WOE, indica el poder predictivo de una variable independiente en relación con la variable dependiente, por lo que, al aplicar el WOE a las variables categóricas, reduciéndolas a las variables con menos categorías, dado que nuestra tasa de incidencia es muy baja, nos queda el valor de información que aportan estas variables al momento de la muerte o no muerte en el parto:

1. 'EDOCIVIL_2': -0.06780127235429725
2. 'ENT_RES_2': -0.06008155597526918,
3. 'prematurro': 0.005915586612738008,
4. 'menor_mayor_edad': 0.03302611548276706,
5. 'mes_nac': -0.1370948062518379

0.8. Conclusión y Propuesta

No hay razón para que las mujeres tengan que morir en el parto. Hay que proporcionar a las mujeres jóvenes la información y el apoyo que necesitan para responder a sus necesidades en materia de salud reproductiva, hay que prestarles apoyo durante el embarazo, ya que dado el análisis anterior, lo que más influye al momento del nacimiento del bebé para que la mujer sobreviva o no es el hecho de que el bebé sea prematuro, así como la edad de la madre.

Lo anterior, se suma a lo mencionado por la OMS, la inmensa mayoría de las muertes maternas se evitarían si las mujeres tuvieran acceso a servicios de planificación familiar de calidad, atención de calidad y seguimiento durante el embarazo, el parto y el primer mes después de dar a luz, o atención después de un aborto , y claro, que en los lugares donde este sea legal , que sea seguro, brindado un sistema de salud de calidad.