

Análisis estadístico de los nacimientos en el instituto de salud de Bucaramanga E.S.E Isabu

Maria Paula Galvis Ojeda

Andrés Fabián Leal Archila

Estadística Descriptiva

Universidad Autónoma De Bucaramanga

Facultad de Economía Y Negocios

2025

INTRODUCCION

Una etapa esencial en la investigación o proyecto estadísticos es el análisis de datos, ya que permite explorar, comprender y depurar la información antes de aplicar modelos o extraer conclusiones. Un buen análisis puede facilitar la identificación de problemas en el conjunto de datos, como valores faltantes o atípicos, y proporcionar una visión clara de la distribución y comportamientos de las variables.

En este caso se trabajó con un conjunto de datos el cual contiene información relacionada con características demográficas y condiciones del nacimiento, información de los padres y los recién nacidos. La información encontrada es útil para estudios de población y análisis en el área de la salud pública.

El objetivo del análisis realizado es obtener una visión integral de la calidad y estructura de datos, abordando aspectos como la detección y visualización de valores faltantes, el análisis de la distribución de variables categóricas y numéricas, la identificación y tratamientos atípicos, así como la evaluación de las variables cuánticas.

Este informe detalla el procedimiento realizado, los hallazgos obtenidos y las recomendaciones sugeridas para el tratamiento de los datos, con el fin de garantizar que el conjunto de información sea confiable y adecuado para futuros análisis estadísticos.

RESULTADOS.

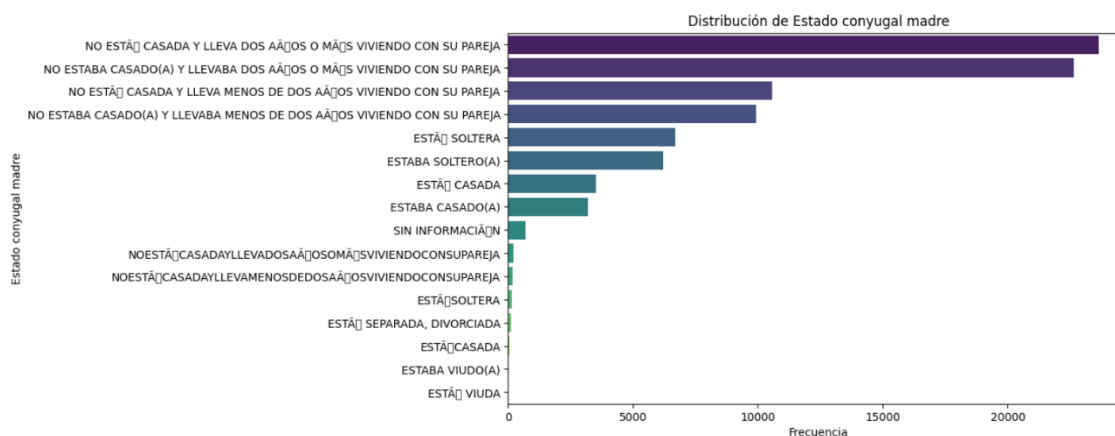
1. Análisis de las variables categóricas.

El análisis de las variables categóricas del conjunto de datos permite comprender con mayor profundidad el perfil sociodemográfico de las madres, los documentos utilizados y las características del parto. A continuación, se presentará el análisis de las variables estado conyugal de la madre, tipo de documento y tipo de parto.

- **Estado conyugal de la madre:** Esta variable proporciona información valiosa sobre la situación sentimental y la convivencia de las madres en el momento del nacimiento de los niños. Se puede observar que la mayoría de las madres no están legalmente casadas, pero conviven con su pareja. Las categorías más frecuentes son “No está casada y lleva dos años o mas viviendo con su pareja” con 23,653 casos y “No estaba casada y lleva dos años o mas viviendo con su pareja” con 22,699 casos, seguidas por “No está casada y lleva menos de dos años viviendo con su pareja” con 10,479 casos y “No estaba casado(a) y lleva menos de dos años viviendo con su pareja” con 9,886 casos.

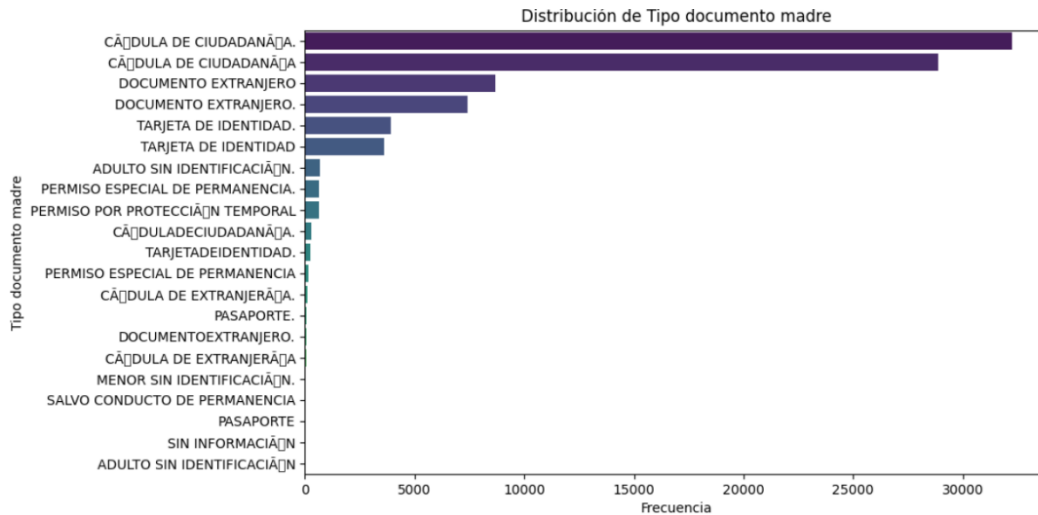
Estas categorías evidencian una tendencia generalizada a la unión libre como forma predominante de convivencia entre las madres analizadas. Por otro lado, los registros de madres solteras que no conviven con pareja también son relevantes, ya que se registraron con 6,680 casos registrados como “Está soltera” y 6,497 como “Estaba soltera”. Estas categorías que representan matrimonios formales tienen una significativamente menor con un aproximado de 3,695 y 3,633 casos respectivamente. Y finalmente otras situaciones como separación, divorcio o viudez tienen una representación mínima en los datos recolectados.

Con el análisis realizado se puede sugerir que gran parte de las madres pertenece a hogares donde la relación de pareja no esta formalizada legalmente, lo que puede estar diferentes implicaciones externas de su entorno.

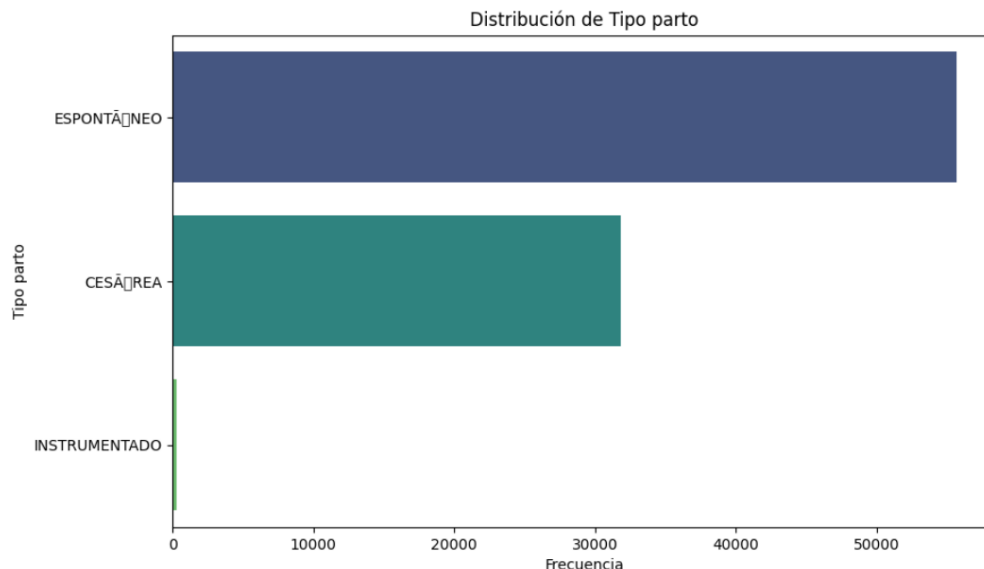


- **Tipo de Documento:** Los datos de esta variable reflejan un gran número de madres identificadas con “Cedula de Ciudadanía” entre los 28,792 a 32,554 registros encontrados en los datos recolectados. Le sigue los documentos de “Documento Extranjero” con un numero de 8,685 y 9,904 casos registrados y la “Tarjeta de Identidad” con 3,771 a 4,085 casos registrados.

14



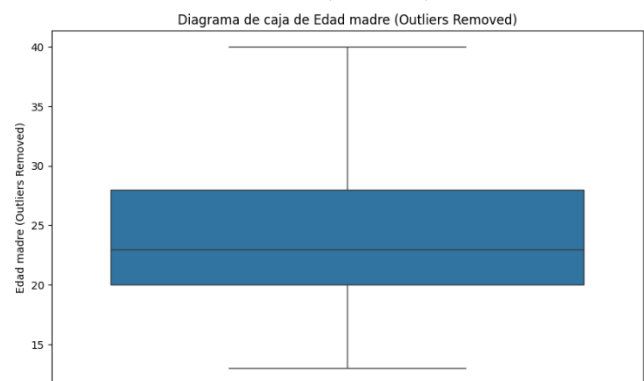
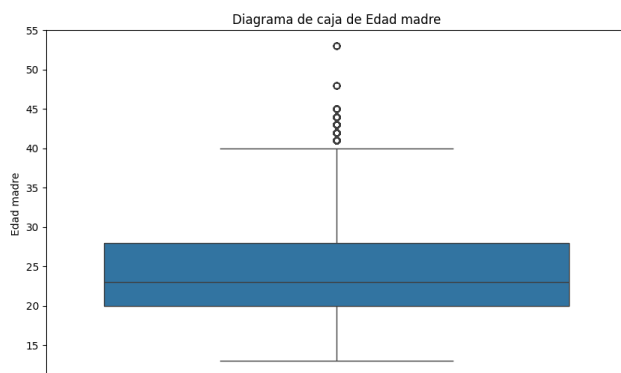
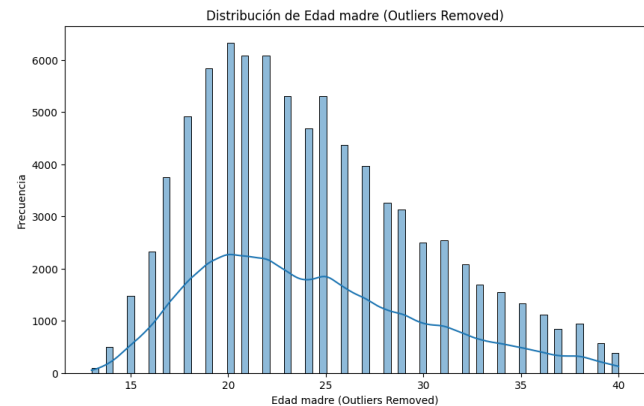
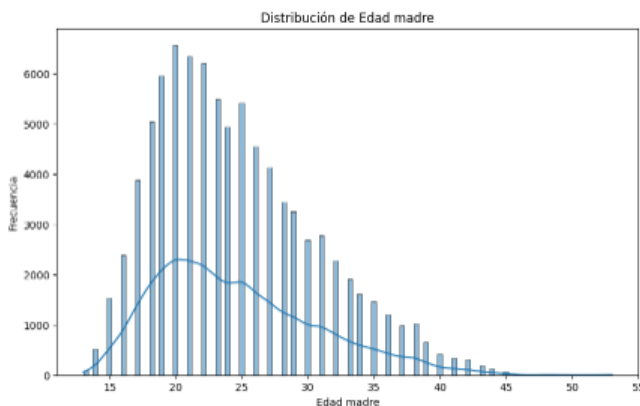
- **Tipo de parto:** La variable muestra de forma clara el tipo de parto más común en la muestra es el parto espontaneo, que se da de forma natural, tiene un total de 55,673 registros. Le sigue el parto por cesárea, con 31,893 casos, mientras que el parto instrumentado (es extremadamente poco frecuente, con tan solo 269 registros. Con esta información podemos decir que los datos recolectados nos sugieren que la mayoría de los nacimientos concurren de manera natural, sin intervenciones quirúrgicas o instrumentales, teniendo en cuenta los casos registrados.



2. Análisis de variables numéricas:

El análisis de las variables numéricas se dividió en tres: Edad de la madre, Edad del padre y el peso. Estas variables permiten describir ciertos aspectos relevantes de la población estudiada, como el perfil de los padres y el estado de salud de los recién nacidos, estos se reflejan en el peso al nacer. Se abordaron aspectos de tendencia central, dispersión, forma de distribución, presencia de valores atípicos y normalidad.

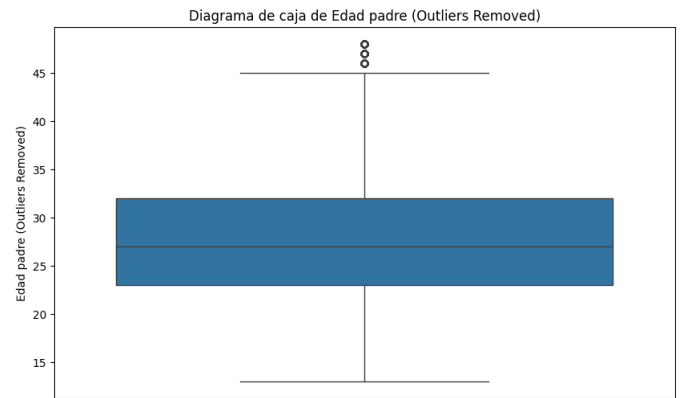
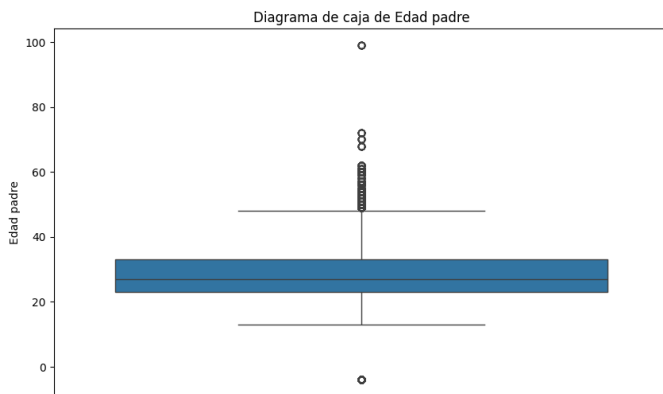
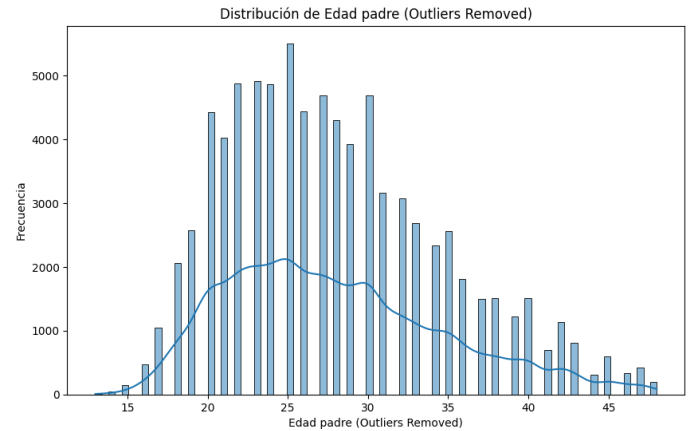
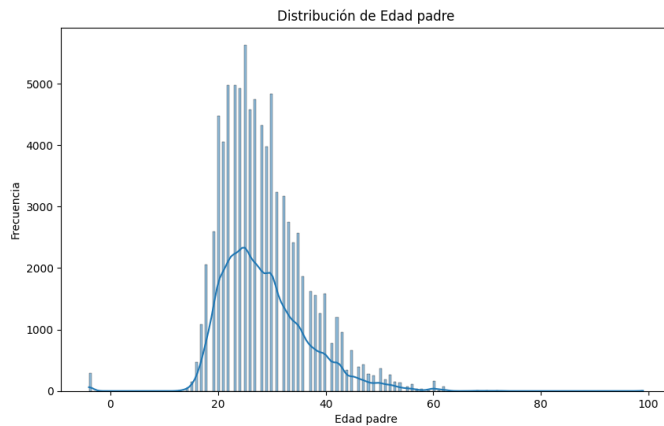
- **Edad de la madre:** La variable “Edad Madre” muestra la edad promedio de las madres que es de 24 a 25 años aproximadamente, lo cual indica que la mayoría de los nacimientos está siendo realizados por mujeres jóvenes. El rango de edad va desde los 13 a 53 años, lo cual muestra la presencia de madres adolescentes como de mujeres de avanzada edad. El histograma correspondiente revela una distribución asimétrica hacia la derecha, concentrándose principalmente en edades entre los 20 a 30 años. Los valores atípicos, detectados se sitúa en ambos extremos. Los outliers al ser tratados con el método de rango intercuartílico permitió observar una distribución más concentrada a eliminar valores.



- **Edad del Padre:** La variable presenta un promedio de edad de aproximadamente 28 a 29 años, lo que sugiere que los padres tienden a ser mayores que las madres al momento del nacimiento. Sin embargo, se identificó una amplia dispersión en los datos, con edades que oscilan entre -9 y 99 años. Cabe destacar que la presencia de un valor negativo representa un error de calidad de datos, ya que no es posible tener una edad menor a cero.

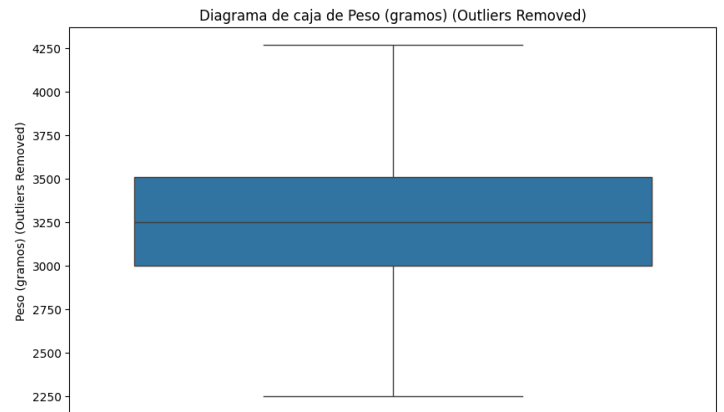
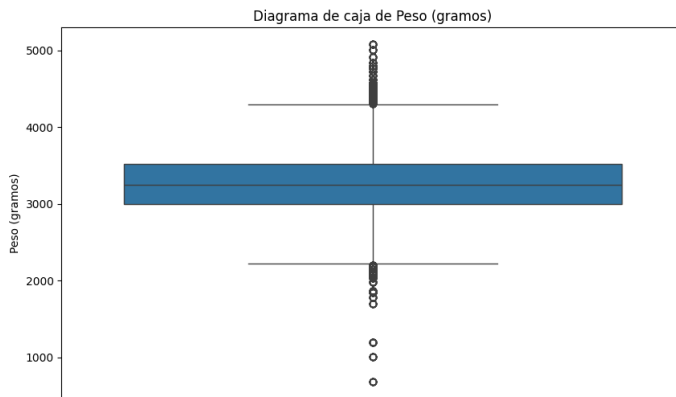
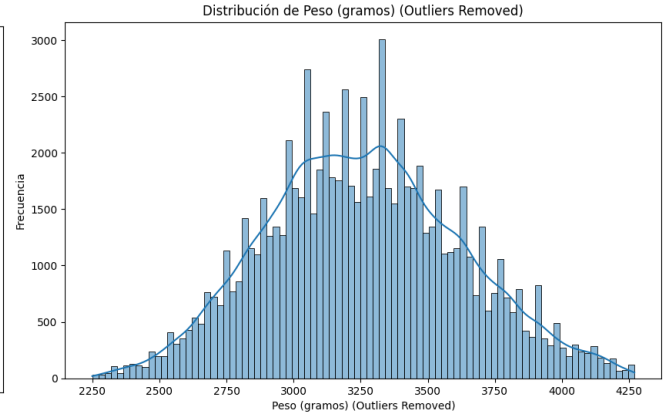
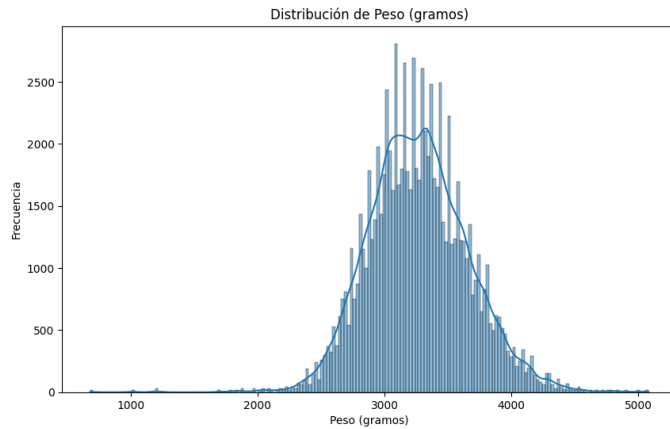
Este tipo de errores refuerza la importancia de realizar limpieza de datos antes de aplicar análisis estadístico. Al igual que en la edad materna, se observaron datos atípicos en ambos extremos. Una vez se aplican los filtros de limpieza, se obtienen un conjunto mas confiable para realizar los análisis.

La distribución, visualizada con histogramas, también presenta asimetría hacia la derecha, con una gran concentración entre los 25 y 35 años, pero con una “cola” que se extiende hasta edades mayores.



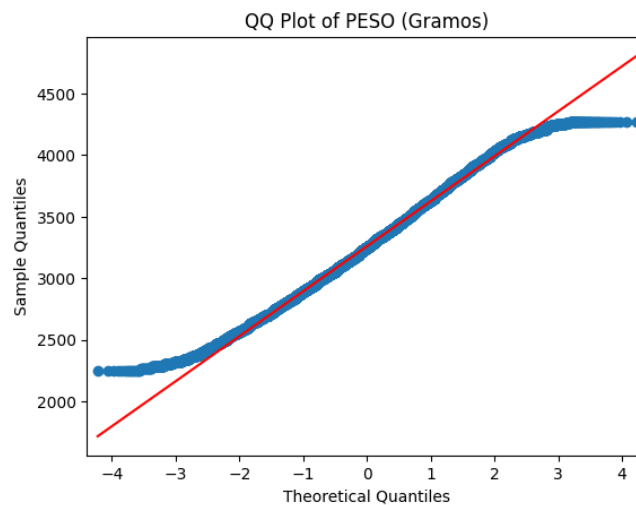
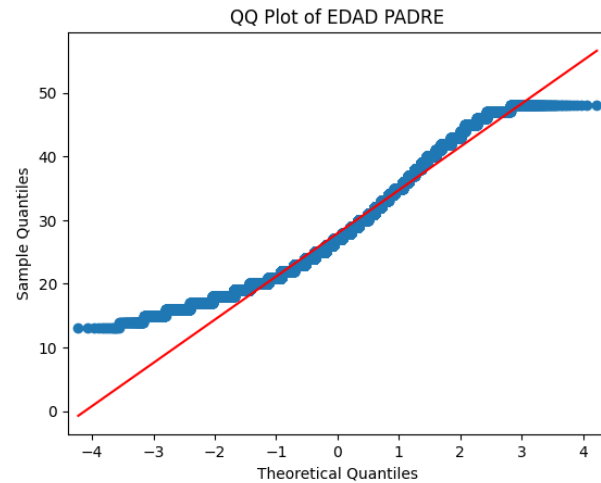
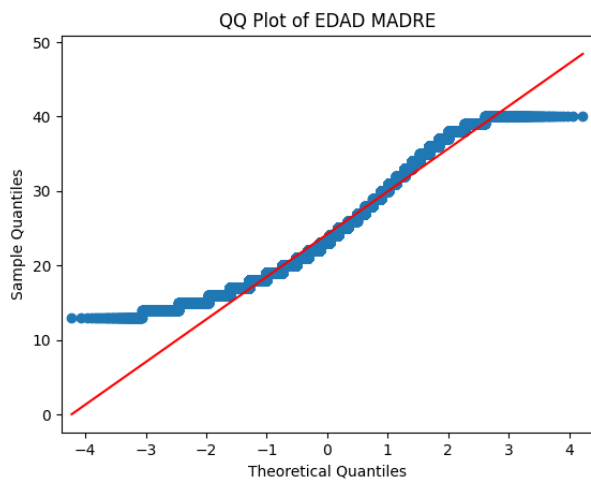
- **Peso al Nacer (Gramos):** La variable Peso (Gramos) mide el peso de los recién nacidos al momento del parto. En promedio, los bebés nacen con un peso cercano a 3,266 gramos, lo que se encuentra en un rango normal. El peso va desde 680 gramos hasta 5.080 gramos, incluyendo a nacimientos prematuros como casos de macrosomía (peso excesivo).

A diferencia de las variables de la edad, la distribución del peso es más simétrica a aproximadamente a una forma de campana o también conocida como distribución normal. Esta situación no evita que no se encuentren datos atípicos en los extremos, aunque son de menos proporción que en otras variables. Estos valores extremos fueron eliminados con el método IQR para tener un análisis más preciso.



Pruebas de normalidad:

Con el fin de determinar si estas variables seguían una distribución normal, se intentó aplicar pruebas estadísticas como Shapiro-Wilk, Kolmogórov-Smirnov, Anderson-Darling y Jarque-Bera, además de generar gráficos QQ Plot. Se logro observar que los histogramas y los boxplots permiten afirmar que las variables de edad presentan una desviación significativa de la normalidad, mientras que el peso muestra un comportamiento mas cercano a la distribución normal.



CONCLUSIÓN:

El análisis de los datos recolectados de los nacimientos de recién nacidos permitió identificar patrones importantes y detectar problemas de calidad de datos que pueden influir en la interpretación de los resultados. En términos generales, las variables numéricas muestran que la mayoría de los nacimientos se dan en madres jóvenes, con un promedio de edad cercano a los 24 años y los padres ligeramente mayores, con un promedio de 28 años. in embargo, se evidenciaron valores atípicos, como edades extremadamente altas o incluso negativas, que representan errores en la recolección o digitación de la información y que deben ser corregidos para garantizar la fiabilidad del análisis.

El peso de los recién nacidos, en parte se mantiene en un promedio de los rangos considerados normales, aunque también se registraron casos de bajo peso y macrosomía.

En cuanto las variables categóricas se observaron que la mayoría de las madres declararon unión libre o casada al momento del nacimiento, lo que refleja la tendencia hacia la convivencia en pareja en el contexto de los datos analizados. El tipo de parto más frecuente fue el natural, aunque el número de cesáreas también fue considerable lo que sugiere la necesidad de explorar factores médicos o sociales que pueden influir en esta decisión. Asimismo, en la variable tipo de documento se identificó que la gran mayoría de las madres están registradas con cédula de ciudadanía, mientras que los otros tipos de identificación presentan una proporción mucho menor.

El estudio también puso en evidencia la presencia de valores faltantes y la necesidad de depuración previa de la base de datos, ya que algunas variables presentan un porcentaje significativo de datos incompletos. Esto resalta la importancia de contar con un proceso de captura y validación de la información más riguroso para evitar inconsistencias. En conjunto, el análisis realizado proporciona una visión clara del perfil de la población estudiada y establece una base sólida para futuros estudios, siempre y cuando se tomen las medidas necesarias para mejorar la calidad de los datos.