



Modelo de regresión cuantílica para la variable peso de neonatos con bajo peso al nacer utilizando imputación múltiple

Jose Jeler Mora Mamian
Johan Javier Morales Hurtado

Universidad del Valle
Facultad de ingeniería, Escuela de estadística
Santiago de Cali, Colombia
2023

Modelo de regresión cuantílica para la variable peso de neonatos con bajo peso al nacer utilizando imputación múltiple

Jose Jeler Mora Mamian
Johan Javier Morales Hurtado

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadísticos

Director(a):
Álvaro Jose Flórez Poveda

Codirector(a):
Mercedes Andrade Bejarano

Universidad del Valle
Sede Cali
Facultad de ingeniería, Escuela de estadística
Santiago de Cali, Colombia
2023

Dedicatoria

Poniendo en primer lugar a Dios, que me ha regalado salud y fuerza de voluntad para llegar hasta aquí. Para a mi madre, mi padre y mi abuela que me han hecho la persona que soy hoy en día, y me han acompañado a lo largo de mi formación profesional.

Jose Jeler Mora

Gracias infinitas a mis padres y a mi hermano por estar presentes en cada momento, brindándome fortaleza para culminar esta etapa de mi vida.

Johan Javier Morales

Agradecimientos

Comenzamos por agradecer a la Universidad del Valle por abrirnos sus puertas, por darnos la oportunidad de crecer como personas y convertirnos en profesionales.

A nuestros profesores y a toda la escuela de estadística en general, por guiar nuestros pasos a lo largo de toda la carrera y hacer del camino una experiencia amena y espectacular.

A los profesores Álvaro Flórez y Mercedes Andrade, nuestros directores, por todo su tiempo, paciencia y dedicación.

Al Hospital Universitario del Valle y docentes del Departamento de Pediatría de la Universidad del Valle por suministrar la base de datos para que este proyecto sea una realidad.

A todos y cada uno de ustedes, Muchas gracias.

Resumen

Se aborda el problema de los datos faltantes en un contexto longitudinal utilizando registros de neonatos con bajo peso al nacer. El objetivo es introducir la incertidumbre generada por los datos faltantes en la estimación de parámetros del modelo y evitar la pérdida de información importante. Para ello, se utilizan registros tomados de historias clínicas de la unidad de recién nacidos del Hospital Universitario del Valle en Santiago de Cali, Colombia, recopilados durante el período 2002-2015. Estos registros forman parte de un programa de seguimiento de neonatos en diferentes periodos de crecimiento para evaluar su evolución en medidas antropométricas. Se ajustan modelos de regresión cuantílica para explicar el comportamiento de la variable peso en cuantiles específicos de interés, esta variable presenta un 49 % de datos faltantes a lo largo de los seguimientos. Para abordar el problema de datos faltantes, se utiliza el método de imputación múltiple, este método genera múltiples estimaciones de los datos faltantes y combina estas estimaciones para obtener resultados más precisos de los parámetros del modelo basándose en la incertidumbre generada por los registros que faltan.

Palabras clave: neonatos, bajo peso al nacer, datos longitudinales, regresión cuantílica, imputación múltiple.

The problem of missing data is addressed in a longitudinal context using records of neonates with low birth weight. The aim is to introduce the uncertainty generated by missing data in the estimation of model parameters and avoid the loss of important information. For this purpose, we use records taken from clinical histories of the newborn unit of the Hospital Universitario del Valle in Santiago de Cali, Colombia, collected during the period 2002-2015. These records are part of a follow-up program of neonates in different growth periods to evaluate their evolution in anthropometric measures. Quantile regression models are fitted to explain the behavior of the weight variable in specific quantiles of interest, this variable presents a 49 % of missing data throughout the follow-ups. To address the missing data problem, the multiple imputation method is used, this method generates multiple estimates of the missing data and combines these estimates to obtain more accurate results for the model parameters based on the uncertainty generated by the missing records.

Keywords: neonates, low birth weight, longitudinal data, quantile regression, multiple imputation.

Contenido

Agradecimientos	iv
Resumen	v
1 Introducción	2
1.1 Planteamiento del problema	2
1.2 Justificación	3
1.3 Objetivo general	4
1.4 Objetivos específicos	4
2 Antecedentes	5
3 Marco teórico	10
3.1 Marco conceptual	10
3.1.1 Neonato	10
3.1.2 Edad gestacional	10
3.1.3 Peso al nacer	11
3.1.4 Antropometría	11
3.1.5 Curvas de crecimiento	12
3.1.6 Hipertensión arterial	12
3.1.7 Control prenatal	13
3.2 Marco teórico estadístico	13
3.2.1 Datos longitudinales	13
3.2.2 Medidas de posición no central	14
3.2.3 Regresión Cuantílica	15
3.2.4 Test de Wald	16
3.2.5 Datos faltantes	17
3.2.6 Patrones de datos faltantes	17
3.2.7 Mecanismos de datos faltantes	19
3.2.8 Imputación múltiple	20
3.2.9 Métodos de imputación	21
3.2.10 Eficiencia método IM	23

4	Metodología	25
4.1	Tipo de estudio	25
4.2	Población de estudio	26
4.3	Análisis exploratorio de los datos	26
4.3.1	Patrones de datos faltantes	27
4.4	Imputación múltiple	27
4.4.1	Métodos de imputación	28
4.4.2	Modelos de imputación	29
4.5	Modelo por regresión cuantílica	32
4.5.1	Transformación del modelo de regresión	32
4.5.2	Estimación errores estándar	33
4.5.3	Test de hipótesis	34
4.6	Construcción de las curvas de crecimiento	34
5	Resultados	36
5.1	Análisis exploratorio	36
5.2	Patrón de datos faltantes	41
5.3	Imputación múltiple	43
5.4	Modelo de regresión cuantílica	44
5.5	Test de Wald	45
5.6	Curvas de crecimiento	46
6	Conclusiones y recomendaciones	49
6.1	Conclusiones	49
6.2	Recomendaciones	50
	Bibliografía	51

Lista de Figuras

3-1	Patrones típicos de datos faltantes, las áreas sombreadas representan la localización de los datos faltantes (Calafati, 2017).	18
5-1	Peso al nacer según el sexo y antecedente de hipertensión	37
5-2	Peso al nacer según el tipo de parto	37
5-3	Peso al nacer según el control prenatal y ventilación asistida	38
5-4	Diagrama de cajas del peso al nacer según el tiempo	39
5-5	Diagrama de alambres del crecimiento por individuo	39
5-6	Comportamiento del peso en el tiempo según tipo de parto, antecedente de hipertensión y control prenatal	40
5-7	Comportamiento del peso en el tiempo según sexo y ventilación asistida . . .	40
5-8	Análisis descriptivo de los datos faltantes en las covariables	42
5-9	Patrón de datos faltantes	42
5-10	Diagrama de cajas de los pesos según el seguimiento	44
5-11	Curvas de peso para hombres y mujeres	46
5-12	Curvas de peso para antecedente de hipertensión	46
5-13	Curvas de peso para ventilación asistida	47
5-14	Curvas de peso para control prenatal	47
5-15	Curvas de peso para tipo de parto	48

Lista de Tablas

4-1	Variables incluidas en el estudio	26
5-1	Distribución de la población según antecedente de hipertensión arterial, tipo de parto, control prenatal y ventilación asistida.	36
5-2	Estadísticas descriptivas según el tiempo de seguimiento	38
5-3	Datos faltantes por periodo de tiempo	41
5-4	Número de seguimientos asistidos	41
5-5	Estadísticas descriptivas de los pesos imputados	43
5-6	Distribución de las variables con observaciones imputadas	43
5-7	Parámetros por cada percentil de la regresión cuantílica	45
5-8	Valores-p del test de Wald según cada τ de la regresión cuantílica	45

1 Introducción

1.1. Planteamiento del problema

La Organización Mundial de la Salud (OMS) define el peso al nacer como el primer peso registrado del neonato inmediatamente después del parto. Este indicador es de gran importancia en el ámbito de la salud pública, ya que refleja las condiciones de salud tanto fetal como neonatal. Los bebés que nacen con un peso inferior a 2500 gramos son denominados como neonatos con bajo peso al nacer (BPN) (OMS, 2017), esta condición se asocia con la incidencia de enfermedades y la mortalidad infantil, teniendo aproximadamente hasta 20 veces más probabilidades de morir, así mismo los bebés que sobreviven tienden a padecer deficiencias en el sistema inmunológico corriendo mayor riesgo de contraer o desarrollar enfermedades que puedan poner en riesgo su vida (Unicef, 2004). El BPN está asociado con una mayor probabilidad de desnutrición, parálisis infantil y trastornos del aprendizaje, además en la etapa adulta, los niños que padecieron BPN tienen una mayor predisposición a diabetes y otras enfermedades, afectando su desarrollo a largo plazo (Velázquez Quintana et al., 2004).

Las causas asociadas con el BPN pueden ser muy diversas y variables, entre ellas algunas como la edad gestacional, parto prematuro, restricción del crecimiento intrauterino, enfermedades que pueden afectar el crecimiento del feto y el estilo de vida de la madre (fumar, alcohol, alimentación no adecuada). El nivel socioeconómico tiende a ser bastante relevante, pues madres en condiciones socioeconómicas desfavorables, con frecuencia tienen infantes con condición de BPN y propensos a padecer distintas morbilidades (Unicef, 2004).

El BPN es un problema de salud pública en todo el mundo y está asociado a una serie de consecuencias a corto y largo plazo, dado que en total, un 15 % a 20 % de los niños nacidos en todo el mundo presentan bajo peso al nacer, lo que supone más de 20 millones de neonatos cada año (OMS, 2017). Existe una variabilidad considerable en la prevalencia del bajo peso al nacer según las regiones e incluso dentro de un mismo país; sin embargo, la gran mayoría de casos de BPN se presentan en países de ingresos bajos y medios, especialmente en los grupos de población más vulnerables. En Asia Meridional, se registra una prevalencia del 31 % de bajo peso al nacer, mientras que en Asia Oriental y el Pacífico este porcentaje es del 7 %, mientras que en África subsahariana se sitúa en el 14 %. Por su parte, en Oriente Medio y África del Norte, la prevalencia es del 15 %, estos resultados se han asociado con niveles elevados de pobreza (Estrada-Restrepo et al., 2016).

En la región de América Latina, la cifra de niños que nacen con bajo peso varía entre el 5,7 % y el 12 %. Dentro de esta estadística, Perú lidera con el porcentaje más alto, alcanzando el 11 %. Le siguen Colombia, Paraguay y Venezuela, con un índice del 9 % (Estrada-Restrepo et al., 2016). Cabe destacar que los datos son limitados porque muchos partos tienen lugar en el hogar o en clínicas pequeñas y no se registran en las cifras oficiales, lo que podría llevar a subestimar la prevalencia de esta condición (OMS, 2017).

El interés de este trabajo de grado es construir curvas de crecimiento para el peso de recién nacidos con BPN a través de regresión cuantílica, teniendo en cuenta la incertidumbre generada por los datos faltantes. Se estudiarán los registros de historias clínicas de neonatos que presentaron BPN en la unidad de recién nacidos del Hospital Universitario del Valle durante el periodo 2002 a 2015 en la ciudad Santiago de Cali. Estos registros pertenecen al programa madre canguro (PMC), el cual realiza seguimiento de la evolución de distintas medidas antropométricas de cada neonato, en diferentes periodos de tiempo a lo largo de 3 años. Los registros del PMC poseen un 25 % de datos faltantes en forma general, y específicamente para la variable peso, estos llegan a ser de un 49 %. El ajuste del modelo de regresión cuantílica no requiere ningún supuesto distribucional sobre la variable respuesta, pero solo proporciona estimaciones válidas cuando la probabilidad de que la observación falte no depende de la variable respuesta (mecanismo de datos faltantes más débil). Para solucionar el problema del mecanismo de datos faltantes en los registros del PMC, se aplican técnicas de imputación múltiple a través del método de especificación totalmente condicional (FCS por sus siglas en inglés) antes de ajustar los modelos, lo que permite obtener estimaciones válidas de los parámetros de regresión cuantílica, para posteriormente construir curvas de crecimiento.

1.2. Justificación

El BPN es un problema de salud pública cada vez más frecuente, está altamente relacionado con la mortalidad infantil, además de generar una serie de complicaciones a corto, mediano y largo plazo en los neonatos que lo padecen (Unicef, 2004), por esto es necesario un seguimiento estadístico en Colombia, ya que los factores que pueden desencadenar el BPN en los neonatos varían para cada territorio. Según Castaño-Díez et al. (2020) no se recomienda generalizar las evaluaciones del desarrollo físico de los neonatos con las curvas de crecimiento existentes, ya que están realizadas con base en referentes internacionales y no clasifican adecuadamente a los bebés nacidos en esta región.

Uno de los principales inconvenientes que generan las curvas de crecimiento para estudios de tipo longitudinal, son los datos faltantes a lo largo del seguimiento, por lo tanto, se propone una alternativa para tratar este problema y mejorar las estimaciones de las curvas de

crecimiento realizadas mediante regresión cuantílica. Es así, como la regresión cuantílica al no estar basada en verosimilitud, es poco robusta en presencia de registros perdidos, ya que requiere que el mecanismo de datos faltantes sea completamente al azar (MCAR), algo que es muy poco probable en la realidad (Koenker et al., 2006), es por esta razón que se realizan estimaciones por medio de bases de datos completas generadas mediante el método de imputación múltiple antes mencionado. Además, la variable de respuesta (peso) posee una cantidad mayor de datos faltantes en contraste con las demás variables. Este es un problema muy común en las variables longitudinales (en este caso el peso) y suele hacerse notorio cuando los periodos de tiempo en los que se realiza la medición son extensos, por lo que está relacionado directamente con la recolección de las observaciones y el diseño del estudio (Calafati, 2017). En este caso específico, dado que se trata de un estudio de seguimiento de neonatos durante 3 años, los datos faltantes pueden deberse a muchas razones difíciles de controlar, como la falta de seguimiento en los controles programados para los neonatos, la deserción de participantes, problemas económicos, sociales o familiares (Calafati, 2017).

La combinación de la imputación múltiple por FCS y la regresión cuantílica es una estrategia sólida y eficiente (Verhasselt et al., 2019) para enfrentar los desafíos que plantea la alta proporción de datos faltantes en este estudio. Mediante la imputación múltiple se pueden obtener estimaciones más precisas de los parámetros globales, y al utilizar la regresión cuantílica se logra comprender mejor la relación entre las variables independientes y el peso del neonato con BPN (Koenker et al., 2006). De esta manera, se contribuye a una mejor comprensión y análisis del problema médico abordado mediante curvas de crecimiento.

1.3. Objetivo general

Modelar la variable peso por medio de regresión cuantílica, para los neonatos con bajo peso al nacer atendidos en el Hospital Universitario del Valle durante el periodo 2002 a 2015, utilizando imputación múltiple FCS para el manejo de la información faltante.

1.4. Objetivos específicos

- Realizar el proceso de imputación múltiple mediante FCS para las covariables y variable respuesta.
- Ajustar un modelo de regresión cuantílica después de realizada la imputación múltiple.
- Construir curvas de crecimiento de la variable peso para caracterizar el desarrollo de los neonatos con BPN.

2 Antecedentes

Curvas de peso al nacer

Montoya Restrepo y Correa Morales (2007) obtuvieron las curvas de crecimiento para distintos percentiles relacionados con el peso al nacer y la edad gestacional, estas fueron realizadas con los pesos de los recién nacidos de la empresa Su Salud del año 1999 al 2005, estos tuvieron entre 22 y 44 semanas de edad gestacional; en total la cantidad de registros obtenidos fueron 54,044 entre las ciudades de Cali, Medellín, Bogotá y Barranquilla. La edad gestacional se determinó siguiendo el Protocolo del Control del Proceso Reproductivo de Su Salud. Se utilizó el método de Capurro, la fórmula de Naegele y la ecografía para estimar la edad gestacional del recién nacido a través de un examen físico. El peso al nacer se obtuvo inmediatamente después del parto, y se tomó como el peso en gramos del bebé desnudo. Esta medición fue realizada por personal médico o paramédico capacitado, utilizando balanzas mecánicas o digitales calibradas. Los autores resaltan principalmente que las curvas que obtuvieron pueden ser mejoradas con la inclusión de más variables a la hora de realizar el ajuste del modelo, variables que pueden afectar directa o indirectamente el crecimiento de los neonatos; sin embargo, resaltan que puede ser una gran herramienta para realizar la clasificación de los recién nacidos en nuestro país, ya que estos datos fueron obtenidos de una población en la cual las características socio demográficas son similares.

Recién nacidos con bajo peso; causas, problemas y perspectivas a futuro

Según Velázquez Quintana et al. (2004) el BPN en los neonatos es una problemática de salud pública, no solo en países en vía de desarrollo sino también de carácter mundial, porque impacta de manera significativa en la mortalidad infantil, estos problemas se presentan más a menudo en países en vía de desarrollo debido a características económicas, socioculturales y demográficas de las madres de neonatos con bajo peso al nacer, pero esto no quiere decir que este problema del bajo peso al nacer no se presente en países desarrollados, su probabilidad de ocurrencia es menor a comparación con países en vía de desarrollo; sin embargo, también puede deberse a otras causas como *“los antecedentes obstétricos y condiciones patológicas que afectan la funcionalidad y suficiencia placentaria, así como las alteraciones propiamente fetales”* (Velázquez Quintana et al., 2004); Además las consecuencias del bajo peso al nacer

no se limitan solo al periodo neonatal inmediato sino también a mediano y largo plazo, ya que ese retardo en el crecimiento puede continuar hasta la edad adulta, acarreando consigo muchas otras complicaciones de salud. Los principales factores de riesgo para los neonatos con bajo peso al nacer pueden ser clasificados según distintas causas como socio demográfico, pre-concepcionales, concepcionales, ambientales, de comportamiento y dependientes del cuidado de la salud.

Factores de riesgo del bajo peso al nacer. Policlínico Universitario José Jacinto Milanés. 2013-2014

Según García et al. (2018) *"La identificación de factores que influyan en la morbimortalidad del recién nacido es una de las preocupaciones fundamentales de la perinatología moderna. El peso del recién nacido es el principal determinante de supervivencia durante el primer año de vida"* por lo cual las principales complicaciones de las etapas posteriores de la vida, pueden ser malas adaptaciones del infante al medio físico, así como también al periodo de escolaridad, debido a que las características neurológicas de los neonatos con bajo peso al nacer están determinadas principalmente por su edad gestacional. Para la realización del estudio recolectaron distintos datos de un policlínico, creando una base de datos en la cual estudiaron de manera descriptiva variables como edad materna, paridad, presencia de factores de riesgo, coito desprotegido, anemia, nivel socioeconómico bajo, antecedentes de parto, desnutrición, antecedentes de bajo peso fetal, entre muchas otras, esto con el fin de analizar los factores de riesgo del BPN.

Programa Madre Canguro: primeros resultados de una cohorte de niños seguidos desde la unidad neonatal hasta la semana 40 de edad postconcepcional

Torres et al. (2006) evalúan el seguimiento realizado a los niños desde la unidad neonatal del Hospital Universitario del Valle (HUV), hasta las 40 semanas de edad gestacional, en el programa Madre Canguro (PMC), para lo cual se realizó un estudio de cohorte en el cual se evaluó la eficacia del PMC para el manejo de los bebés con bajo peso al nacer. Estos bebés tenían un peso inferior a 2,001 gramos, habían superado los problemas principales de adaptación neonatal. Se excluyeron aquellos neonatos con malformaciones graves o letales tempranas, así como aquellos que estaban abandonados o en proceso de adopción. Se recopilieron datos sobre el crecimiento físico, la duración de la lactancia materna, la morbilidad, la mortalidad, los reingresos, los medicamentos utilizados, la duración de la estancia hospitalaria y la rotación de camas desde el momento del alta hasta las 40 semanas de edad gestacional. Según esto, los autores obtuvieron:

“Hasta el cumplimiento de las 40 semanas de edad gestacional, el PMC en el HUV mostró ser una alternativa segura para el manejo del recién nacido de muy bajo peso, pues se les garantizó un egreso temprano, contacto piel a piel y una alimentación inicial basada en leche materna. Los niños egresaron con 19 días de edad postconcepcional y con 1552 g de peso en promedio. La proporción de niños con lactancia materna fue excelente: 96 % al egreso y 94 % a las 40 semanas. A partir de la semana 34, la ganancia de peso promedio estuvo entre 15 y 20 g/día. La permanencia en la sala de recién nacidos del HUV durante el PMC (2002-2003) fue inferior en dos días a la estancia en el seguimiento convencional de recién nacidos de riesgo (año 2000).”

Curvas antropométricas para los recién nacidos en Chile

Según Alarcón et al. (2010) el objetivo está centrado en realizar el diseño de la curva de crecimiento antropométrico para los recién nacidos en Chile, estos poseían una edad gestacional entre las 24 y 42 semanas, la población total de registros analizados fue de 148,385 recién nacidos obtenidos del Servicio de Salud Metropolitano de Oriente, Hospital del Salvador entre 1998-2000 y Hospital Luis Tisne 2001 a 2005, esta base de datos poseía variables como fecha de la última regla, peso, talla, perímetro cefálico, sexo, género entre muchas otras, por algunos criterios de exclusión planteados en el estudio, finalmente se trabajó con 86,575 registros. Los autores clasifican la población según variables como el peso y la edad gestacional en distintas tablas de peso, perímetro cefálico y talla respectivamente, posteriormente se calcularon algunas medidas de posición y dispersión, además de la distribución de distintos percentiles ajustados con un polinomio de tercer orden. En la parte central del debate, se establece que se necesita una única curva para la variable sexo, debido a que solo se observan diferencias notables después de la semana 35 de embarazo. Además, se llevan a cabo comparaciones con otros estudios que presentan características socio-demográficas similares, encontrando algunas similitudes. La conclusión final es que las curvas obtenidas proporcionan datos precisos y actualizados en ese momento sobre las medidas antropométricas de los recién nacidos.

Quantile regression methods for reference growth charts

Según Koenker et al. (2006) ajustaron curvas de regresión para las variables peso y altura de 2500 niños en Finlandia, realizando también una depuración de la base de datos, eliminando algunos registros con información faltante, de los cuales se conservaron al rededor de 2300 registros de infantes, cabe resaltar que se realizaron varias mediciones a cada niño y niña registrado en la base de datos a lo largo del tiempo. El objetivo del estudio fue obtener

un contraste para las curvas de crecimiento ajustadas mediante LMS (Least-mean-squared algorithm) y regresión cuantílica respectivamente, obteniendo en sus resultados muy pocas diferencias en cuanto a los dos modelos. Los autores concluyen que para el caso especial de la regresión cuantílica este método es más flexible, ya que puede ser capaz de revelar distorsiones de los supuestos resultantes de modelos paramétricos, además resaltan que el modelo de regresión cuantílica permite revelar cambios y tendencias que son difíciles de observar con otros métodos paramétricos más convencionales.

Estrategias para el tratamiento de datos faltantes en estudios con datos longitudinales

Según Calafati (2017) en este estudio se analiza de manera detallada la problemática generada por los datos faltantes sobre algunas bases de datos con registros longitudinales, los cuales son medidos a través del tiempo en distintas etapas, lo que puede generar una cantidad indeterminada de información ausente en la base de datos, el autor expone los distintos métodos que se pueden utilizar para solucionar dicho problema, además de detallar las ventajas, desventajas y limitaciones de los distintos métodos de imputación de datos longitudinales. Para ello evalúa distintos métodos como imputación mediante la media, imputación por regresión y por regresión estocástica, imputación múltiple, entre otros. Según el autor, la imputación múltiple es uno de los mejores métodos para solucionar el problema de datos faltantes, ya que esta genera varios conjuntos de datos, remplazando los valores faltantes con valores factibles de acuerdo a la naturaleza de los datos estudiados, estos valores factibles se obtienen de una distribución modelada específicamente para cada dato, luego de esto se procede a estimar los parámetros de interés para cada uno de los datos imputados, y finalmente, el conjunto de estimaciones se combinan para obtener inferencias globales. El autor concluye que la imputación múltiple es uno de los métodos más efectivos para el tratamiento de datos faltantes.

The impact of incomplete data on quantile regression for longitudinal data

En este artículo Verhasselt et al. (2019) abordan el problema de los datos faltantes en estudios longitudinales, centrándose específicamente en la deserción de sujetos que abandonan el estudio y no tienen registros posteriores. Se menciona que hacer inferencias basadas únicamente en los casos completos puede llevar a estimaciones sesgadas e ineficientes. Por lo tanto, se requieren técnicas que manejen adecuadamente estos datos faltantes. Se describen tres mecanismos principales de datos faltantes: perdidos completamente al azar (MCAR), perdidos al azar (MAR) y perdidos no al azar (MNAR). Se indica que el mecanismo de da-

tos faltantes tiene un gran impacto en el rendimiento de las técnicas estadísticas utilizadas para tratarlos. Se mencionan dos métodos comunes para manejar los datos faltantes: ponderación de probabilidad inversa (IPW) e imputación múltiple (MI). La imputación múltiple utiliza los datos observados para generar valores para los datos faltantes, mientras que los métodos de IPW responderán las observaciones según el mecanismo subyacente de ausencia. El enfoque principal del artículo es el uso de técnicas para datos faltantes en el contexto de la regresión por cuantiles, que permite estimar el efecto de las covariables en diferentes cuantiles de la respuesta.

3 Marco teórico

En esta sección del documento podremos encontrar las bases teóricas, conceptuales y estructurales que se implican de manera contextual y práctica en el desarrollo del trabajo de grado, utilizando como principal herramienta toda la teoría que sustenta el estudio del BPN en los neonatos, mediante la imputación múltiple y la regresión cuantílica que tendrá como variables de respuesta el peso del neonato a lo largo de los seguimientos.

3.1. Marco conceptual

3.1.1. Neonato

El neonato, o también denominado recién nacido, es un bebé que tiene 28 días o menos contados desde su nacimiento, ya sea que haya nacido por parto natural o por cesárea. Es una etapa de la vida muy corta pero de gran importancia, ya que los cambios que se presenten durante esta etapa conllevan a consecuencias determinantes para el resto de la vida del bebé recién nacido (Saha, 2023).

3.1.2. Edad gestacional

La edad gestacional se define como el tiempo entre el primer día del último ciclo menstrual normal de la madre y el día del nacimiento del neonato, en este tiempo el feto se desarrolla y crece dentro del útero de la madre. Este representa un sistema estándar para cuantificar la progresión del embarazo y así poder describir qué tan avanzado se encuentra el feto para determinar la madurez de los órganos (Almira, 2008).

Durante el periodo anterior al nacimiento, los especialistas pueden determinar la edad gestacional por medio del ciclo menstrual, si estos ciclos no presentan irregularidades en la madre, o por medio de ecografías que permitan medir el tamaño de la cabeza, del abdomen y del fémur del bebé, permitiendo observar qué tan bien evoluciona el feto dentro del útero. Con base en esto se realizará la clasificación del neonato según el tiempo medido en semanas (Gómez-Gómez et al., 2012):

- **Término temprano o prematuros:** son los neonatos que nacen antes de la semana 37 de gestación. Estos se subdividen en 3 categorías dentro de los bebés nacidos en

término temprano, las cuales son: **premature extremo** el cual nace antes de las 28 semanas de gestación, **premature moderado** nacido entre las semanas 28 y 32 de gestación, y **moderadamente premature tardío** los cuales nacen entre la semana 33 y 37 de edad gestacional.

- **Término completo:** neonato cuyo nacimiento se produjo entre las semanas 37 y 41 de edad gestacional.
- **Post-término:** el nacimiento del bebé se produjo después de la semana 42 de edad gestacional.

Si bien la fecha de última regla es un método adecuado para el cálculo de la edad gestacional, cuando se presentan sesgos de información en la fecha de concepción o presencia de irregularidades en los ciclos menstruales los resultados no serán del todo confiables, por ello se han propuesto nuevos métodos basados en la observación post-parto de características físicas y neurológicas que presente el neonato. Una de estas es la prueba de Ballard Ballard et al. (1991) la cual se basa en evaluar la madurez física y neuromuscular del neonato, otorgándole un puntaje de acuerdo a diferentes características y pruebas a la que es sometido el recién nacido, para así calcular la edad gestacional en términos de semanas.

3.1.3. Peso al nacer

El peso al nacer hace referencia a la primera toma de peso del neonato justo después de su nacimiento, este relacionado con la edad gestacional del neonato. Un neonato con un peso inferior a los 2500 gramos se clasificará con bajo peso y dentro de esta clasificación se tienen sub-grupos, bajo peso al nacer cuando el peso es inferior a los 2500 g, muy bajo peso al nacer por debajo de los 1500 g y peso extremadamente bajo inferior a los 1000 g (Velázquez Quintana et al., 2004). Esta variable es de gran importancia, ya que se utiliza como primer indicador de la morbilidad y mortalidad con respecto a diferentes condiciones que presentó y pueda presentar a futuro el neonato, también se utiliza para entender las condiciones socio-económicas y de salud de la madre, siendo este factor de gran ayuda para la prevención del bajo peso.

3.1.4. Antropometría

La vigilancia del crecimiento en el neonato, a través de la evaluación antropométrica, reviste gran importancia en términos de detección de riesgos de morbi-mortalidad y deterioro del estado nutricional, permitiendo la toma de decisiones oportunas y convenientes (Montesinos-Correa, 2014). Para que el control de calidad en la evaluación antropométrica esté garantizado, se requiere conocer las técnicas de mediciones correctas, estar estandarizado y comparar con valores de referencia de una población similar. Las mediciones antropométricas más

utilizadas en el neonato hospitalizado incluyen: peso corporal (masa corporal total), longitud (tamaño corporal y óseo), medición de circunferencias (brazo, tórax, muslo) y pliegues cutáneos (grasa subcutánea). Además, al combinar algunas mediciones, se pueden generar índices pronósticos nutricios y de utilidad diagnóstica. La antropometría debe ser un método de rutina en las unidades de cuidado neonatal como parte de una evaluación nutricional completa (Cárdenas-López et al., 2005).

3.1.5. Curvas de crecimiento

Las curvas de crecimiento, o denominadas patrones de crecimiento, permiten por medio del peso, talla y perímetro cefálico llevar un monitoreo del crecimiento del neonato para determinados tiempos del neonato, siendo una parte vital para la evaluación personal de la salud del individuo como también definir grupos poblacionales que puedan presentar mayores riesgos de mortalidad y morbilidad neonatal, lo cual es de gran importancia para reducir los riesgos que se puedan presentar en los individuos con mayor probabilidad de afectación y a su vez llevar un control del crecimiento antropométrico de los neonatos (Fabres et al., 2010).

Uno de los principales objetivos de las curvas de crecimiento, es determinar la población en riesgo de enfermar o morir, por lo que es importante contar con un patrón de crecimiento lo más fisiológico posible y que permita el máximo de confiabilidad. La detección adecuada de este grupo de fetos y/o recién nacidos permitirá la implementación oportuna de controles o medidas para disminuir los riesgos y para poder tratar a los casos en que la situación lo amerite (Fabres et al., 2010).

Estas curvas tienen una gran utilidad cuando se ponen en uso los percentiles de las curvas, debido a que estos indican las medidas antropométricas para realizar un análisis individual más a detalle y así determinar los neonatos que necesitan de un mayor cuidado por pertenecer a percentiles alejados de los valores medios en las medidas antropométricas tomadas (Anderson y Hay, 2001).

3.1.6. Hipertensión arterial

La tensión arterial es la fuerza que ejerce la sangre contra las paredes de las arterias al ser bombeada por el corazón. Las arterias son grandes vasos por los que circula la sangre en el organismo. Se considera que la persona presenta hipertensión cuando su tensión arterial es demasiado elevada. De la tensión arterial se dan dos valores: el primero es la tensión sistólica y corresponde al momento en que el corazón se contrae o late, mientras que el segundo, la tensión diastólica, representa la presión ejercida sobre los vasos cuando el corazón se relaja entre un latido y otro (Mancia et al., 2007).

Durante el embarazo, la hipertensión es el problema médico más frecuente que pueda presentar la madre, lo que conlleva a ser la principal causa de morbilidad materna y fetal, y puede llegar a ser la principal causa de mortalidad materna si no se suministran los tratamientos adecuados. Es por ello que la importancia de su estudio puede llevar a prevenir o atenuar alteraciones hipertensivas de la gestación, como también enfermedades cardiovasculares y renales futuros (Marín-Iranzo, 2006). La presión arterial alta durante el embarazo ocasionará los siguientes problemas de salud: menor flujo sanguíneo a la placenta, desprendimiento de la placenta, restricción del crecimiento intrauterino, lesión en órganos, parto prematuro y enfermedad cardiovascular en el futuro.

3.1.7. Control prenatal

Es el conjunto de actividades de prevención, diagnóstico y tratamiento planificadas con el fin de lograr una adecuada gestación, que permita que el parto y el nacimiento ocurran en óptimas condiciones, sin secuelas físicas o síquicas para la madre o el recién nacido (Cáceres-Manrique, 2009).

Se recomienda que la primera consulta sea lo más pronto posible para confirmar el diagnóstico del embarazo. Es recomendado realizar el primer control prenatal en el primer trimestre, idealmente antes de la semana 10 de gestación. Los principales objetivos del control prenatal son identificar aquellos pacientes de mayor riesgo, con el fin de realizar intervenciones en forma oportuna que permitan prevenir dichos riesgos y así lograr un buen resultado perinatal. Esto se realiza a través de la historia médica y reproductiva de la mujer, el examen físico, la realización de algunos exámenes de laboratorio y exámenes de ultrasonido. Además, es importante promover estilos de vida saludables, la suplementación de ácido fólico, una consejería nutricional y educación al respecto (Vintzileos et al., 2002).

3.2. Marco teórico estadístico

3.2.1. Datos longitudinales

Los datos longitudinales son observaciones registradas en los mismos individuos a través del tiempo (Rodríguez y Díaz, 2004). Así, los diseños longitudinales de medidas repetidas constan de uno o más grupos de sujetos medidos en una o más variables a lo largo de puntos temporales. A diferencia de lo que ocurre en el contexto experimental, donde las medidas repetidas corresponden a los tratamientos, en las situaciones no-experimentales tales medidas representan las ocasiones temporales en las que se realizan las mediciones. Las curvas de crecimiento constituyen un ejemplo típico de este tipo de datos. En ellas, el investiga-

dor registra una serie de medidas en intervalos temporales sucesivos de amplitud constante, en una o más muestras de sujetos, con el fin de examinar el proceso de desarrollo de cada individuo y las posibles diferencias existentes en dicho proceso entre distintas muestras de individuos (Arnau y Balluerka, 2004).

Para los datos longitudinales, definiremos a Y_{ij} como la respuesta del i -ésimo individuo en el tiempo t_j , para $i=1, \dots, N$ y $j=1, \dots, n_i$. De esta forma las observaciones del mismo individuo se pueden agrupar en un vector unitario, de la forma $Y_i = (Y_{i1}, \dots, Y_{in_i})'$ (Fitzmaurice et al., 2012).

3.2.2. Medidas de posición no central

Las medidas de posición no central permiten dividir la distribución de una variable en un determinado número de segmentos (cuantiles) facilitando así la ubicación de un sujeto o caso sobre un conjunto de datos, cabe mencionar que estas particiones requieren que exista un orden en la categoría de las variables, por lo que es necesario determinarlas a partir de por lo menos una escala ordinal (Batanero, 2000). En el uso diario, los cuantiles más comunes en los distintos ámbitos de la estadística son los cuartiles, deciles y percentiles. Para este caso, se hablará principalmente de los percentiles, es una medida estadística de posición, que divide la distribución ordenada de los datos en cien partes iguales, esta medida de posición no central aporta información sobre el porcentaje de observaciones de una variable, ordenados de menor a mayor, que se sitúan por debajo del valor de este.

El τ -percentil de cualquier variable aleatoria, es la función inversa de la distribución acumulativa evaluada en τ (Davino et al., 2013). Esto es:

$$Q(\tau) = F^{-1}(\tau) = F(y) \geq \tau \text{ cuando } 0 < \tau < 1$$

El estimador de un percentil se obtiene de la siguiente forma:

$$\hat{Q}(\tau) = \underset{b \in \mathbb{R}}{\text{Min}} \left[\sum_{Y_i \geq b} \tau |y_i - b| + \sum_{Y_i \leq b} (1 - \tau) |y_i - b| \right]$$

siendo τ el cuantil, y los distintos valores que toman las observaciones de la muestra para la variable y y b el valor que minimiza la expresión (Vicèns and Sánchez, 2012).

3.2.3. Regresión Cuantílica

La regresión cuantílica, también conocida como regresión por cuantiles o regresión cuantil, es una técnica estadística utilizada para modelar la relación entre una variable respuesta y una o más variables predictoras. Siendo este tipo de regresión introducida en 1978 por Koenker y Bassed como *"los cuantiles condicionales que podrían ser estimados por una función de optimización minimizando la suma de desviaciones absolutas ponderadas, usando pesos como funciones asimétricas de cuantiles"* (Davino et al., 2013).

Esta regresión se puede observar como una extensión de la estimación clásica de medias condicionales por mínimos cuadrados, llevada así a una estimación por medio de un conjunto de funciones de cuantiles condicionales. A diferencia de los diferentes tipos de regresiones lineales ordinarias, que se centran en modelar la media condicional de la variable respuesta, la regresión cuantílica se enfoca en modelar los diferentes cuantiles condicionales de la variable respuesta con respecto a las variables predictoras. Además, la regresión cuantílica no hace ningún supuesto distribucional sobre la respuesta, lo que hace que sea más flexibles en la ejecución de esta técnica, clasificando el modelo como semiparamétrico (Hao et al., 2007).

Según Koenker y Hallock (2001) el modelo regresión cuantílica (RC) se puede especificar como:

$$y_i = x_i' \beta_\tau + u_{\tau,i}$$

donde y_i es el vector de variables respuesta, $i = 1, 2, \dots, n$ es una muestra aleatoria de n observaciones, x_i representa el vector $p \times 1$ de covariables asociadas a la i -ésima observación, β_τ es el vector $p \times 1$ de parámetros a estimar correspondiente al cuantil τ y $u_{\tau,i}$ es la perturbación aleatoria correspondiente al cuantil τ .

Los supuestos que condicionan los errores del modelo son: el τ -ésimo cuantil de los errores condicionados a las variables regresoras es cero $Q_\tau(u_{\tau,i} | x_i) = 0$, y además, que las observaciones sean independientes. Entonces el τ -ésimo cuantil de y_i con respecto a las variables regresoras (Koenker y Hallock, 2001):

$$Q_\tau(y_i | x_i) = x_i' \beta_\tau$$

Por lo tanto, según Koenker y Hallock (2001) el estimador de β_τ se obtiene como la solución al siguiente problema de minimización:

$$\hat{\beta}_\tau = \underset{\beta_\tau \in \mathbb{R}}{\text{Min}} \left[\sum_{y_i \geq x_i' \beta_\tau} \tau |y_i - x_i' \beta_\tau| + \sum_{y_i \leq x_i' \beta_\tau} (1 - \tau) |y_i - x_i' \beta_\tau| \right]$$

Para la estimación de los coeficientes en la regresión cuantílica, no hay una solución analítica disponible debido a la naturaleza no lineal de la función objetivo que se busca minimizar, ya que por su complejidad no es diferenciable para poder hallar los coeficientes óptimos. Por lo cual es necesario emplear métodos numéricos y técnicas de optimización para poder solucionar el problema, donde se utilizará para este caso la programación lineal (Hao et al., 2007).

Otra ventaja de los modelos de regresión cuantílica es que son robustos a los valores atípicos. Esto debido a que la estimación con respecto a las desviaciones en valores absolutos penalizará los errores de forma lineal, dándole pesos correspondientes a la región cuantílica en la que se encuentre la observación, contrario a las estimaciones por medio de MCO que elevan los errores al cuadrado entregando una mayor importancia o peso a estos datos extremos. Esta característica de la regresión cuantílica hace que también sea especialmente útil para el trato de datos censurados, ya que realmente solo es relevante el hecho de si el valor estimado se encuentra por encima ($y_i < x'_i\beta_\tau$) o por debajo ($y_i \geq x'_i\beta_\tau$) del real, no su magnitud (Koenker y Hallock, 2001).

3.2.4. Test de Wald

El proceso de inferencia de los regresores se lleva a cabo por medio de algunas transformaciones y con ayuda de métodos generalizados de los momentos, esto bajo determinadas condiciones, como que el tamaño de muestra sea lo suficientemente grande para la convergencia hacia la distribución asintótica de los estimadores (Barnes y Hughes, 2002), está dada por:

$$\sqrt{n}(\hat{\beta}_\tau^* - \beta_0) \rightarrow N(0, V(\hat{\beta}_\tau^*))$$

donde $\hat{\beta}_\tau^*$ representa el vector de regresores a comprobar por medio del test, y $V(\hat{\beta}_\tau^*)$ es la matriz de varianzas-covarianzas de los regresores.

Para comprobar las hipótesis planteadas se utilizará el test de Wald, cuyo estadístico de prueba está definido como:

$$W = (\hat{\beta}_\tau^* - \beta_0)' V(\hat{\beta}_\tau^*)^{-1} (\hat{\beta}_\tau^* - \beta_0)$$

donde β_0 es el valor nulo sugerido a probar dentro de la hipótesis.

El estadístico de prueba sigue una distribución χ^2 (chi-cuadrado) con grados de libertad igual a la dimensión de $\hat{\beta}_\tau^*$.

3.2.5. Datos faltantes

Los datos faltantes se definen como valores que no están disponibles y que serían significativos si se observan (Dagnino et al., 2014). Los datos faltantes se pueden deber a diferentes factores dentro del estudio o recolección de la información, desde una secuencia de falta de seguimiento, una característica incompleta, archivos faltantes, información incompleta, errores en datos de entrada, etc. Siendo este problema muy común en los diferentes conjuntos de datos recolectados en el mundo real, por lo que es necesario tener conocimiento sobre el manejo de estos, ya que puede llevar a una pérdida de información importante para la toma de decisiones dentro del estudio o en su defecto ocasionar sesgos de información (Calafati, 2017).

Teniendo en cuenta la variable de respuesta Y_{ij} , se define un indicador para determinar los datos faltantes en el conjunto de datos (Fitzmaurice et al., 2012):

$$R_{ij} = \begin{cases} 1 & \text{si } Y_{ij} \text{ es observado} \\ 0 & \text{otro caso} \end{cases}$$

Con esto se obtiene $R_i = (R_{i1}, \dots, R_{in_i})'$ el cual representa el vector para el individuo i que contiene el indicador de los valores observados y faltantes para este individuo. Luego, con el vector R_i se divide el vector Y_i en dos subvectores, tal que Y_i^o será el vector que contiene la información de los datos observados, $R_{ij} = 1$, y el vector Y_i^m contendrá el complemento del vector Y_i^o , representando el vector del conjunto de datos faltantes, con esto se puede afirmar que la variable aleatoria R_i tiene una distribución de probabilidad conjunta, donde esta describe el porcentaje de datos omitidos dentro de la base y su relación con los datos observados, esto es $f(R_i | Y_i^o, Y_i^m)$ (Fitzmaurice et al., 2012).

3.2.6. Patrones de datos faltantes

Los patrones de datos faltantes se pueden definir como la configuración de los datos observados y los valores omitidos dentro del conjunto de datos (Calafati, 2017). Interpretando la base de datos como una matriz donde las filas son los registros de los individuos y las columnas representan las variables de interés, con un análisis visual permitirá encontrar patrones de comportamiento que permitan facilitar la elección del método.

En la Figura 3-1 se observan 6 patrones típicos de datos faltantes que se pueden encontrar en los diferentes estudios, para los datos de tipo longitudinal los más comunes son el patrón monótono, no monótono y el patrón general.

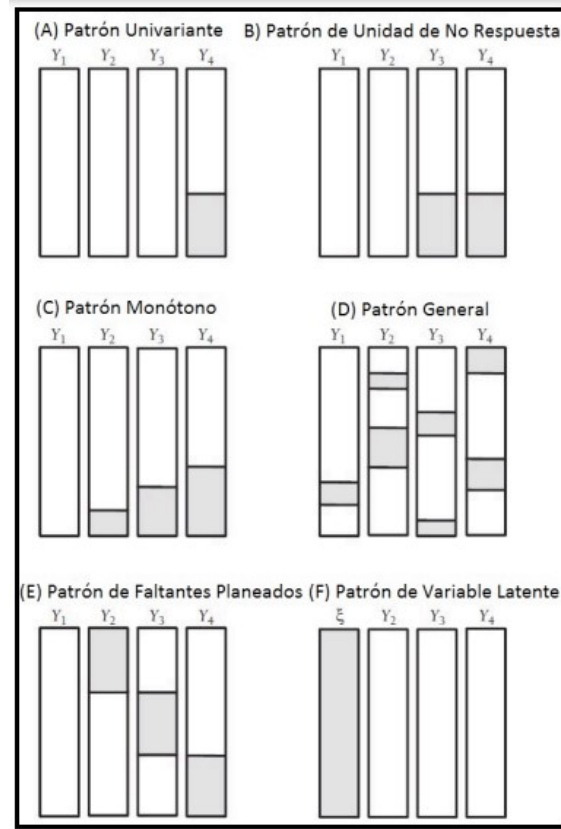


Figura 3-1: Patrones típicos de datos faltantes, las áreas sombreadas representan la localización de los datos faltantes (Calafati, 2017).

Para los datos longitudinales existen dos patrones comunes, el primero será un patrón monótono (C) de pérdida de información que indica que al faltar un seguimiento del individuo, los siguientes no serán observados debido a esta condición. Los valores faltantes en el vector de respuesta y_i toman la forma de abandono monótono si cuando falte el dato y_{ij} también lo estará y_{ik} para todo $k \geq j$. Este patrón requiere una mayor facilidad de evaluación para las funciones de probabilidad debido a que se pueden factorizar en términos de densidades condicionales (Calafati, 2017).

El caso contrario es el del patrón de pérdida no monótono, que ocurre cuando al tener un dato faltante para un individuo en un periodo del seguimiento, este puede presentar en los siguientes periodos alguna observación completa.

El patrón general (D) es el más común dentro de las bases de datos, este tiene valores faltantes dispersos en toda la matriz sin tener un patrón definido para las variables, se debe tener cuidado con este patrón, ya que puede llevar a considerar relaciones no existentes entre los datos omitidos y su comportamiento (Calafati, 2017).

3.2.7. Mecanismos de datos faltantes

Rubin (1976), y Rubin y Little (1987) argumentan que los datos faltantes se deben analizar como un fenómeno estocástico e introducen los mecanismos de datos faltantes. Para ello, primero se presenta en (3-1) la distribución conjunta de Y_i y R_i .

$$f(Y_i, R_i \mid X_i, \theta, \psi) = f(Y_i \mid X_i, \theta) f(R_i \mid X_i, Y_i^o, Y_i^m, \psi) \quad (3-1)$$

donde X_i es la matriz de covariables asociadas al i -ésimo individuo, θ y ψ son vectores de parámetros asociados a Y_i y R_i , respectivamente. El segundo término de la derecha de (3-1) define el mecanismo de datos faltantes.

Así Rubin (1976) clasificó 3 mecanismos de datos faltantes, cada uno de estos tendrá sus características especiales y causas sobre los modelos para datos longitudinales.

- **MCAR (Missing Completely At Random):** este describe un proceso completamente aleatorio donde la probabilidad de que un dato sea faltante no depende de los valores observados y no observados de la variable respuesta, sin embargo puede depender de covariables. Esto es:

$$f(R_i \mid X_i, Y_i^o, Y_i^m, \psi) = f(R_i \mid X_i, \psi)$$

- **MAR (Missing At Random):** bajo un mecanismo MAR la probabilidad de que falte una observación es condicionalmente independiente de los datos no observados, dados los valores de los datos observados. Esto es:

$$f(R_i \mid X_i, Y_i^o, Y_i^m, \psi) = f(R_i \mid X_i, Y_i^o, \psi)$$

- **NMAR (Not Missing At Random):** este mecanismo ocurre cuando la probabilidad de que una observación sea faltante depende también de los valores no observados de la respuesta. Esto es:

$$f(R_i | X_i, Y_i^o, Y_i^m, \psi)$$

Bajo el mecanismo NMAR, los métodos de estimación estándar pueden proporcionar estimaciones sesgadas.

3.2.8. Imputación múltiple

La imputación múltiple se define como un enfoque predictivo para el manejo de los datos faltantes en el análisis multivariante, combinando técnicas estadísticas clásicas o bayesianas, basándose en la introducción del error o variabilidad original de los datos observados para realizar las imputaciones. La idea de este método es el de imputar el valor faltante M veces generando M conjuntos de datos, esto por medio de métodos de simulación de Monte Carlo (Van Buuren, 2018).

Luego, estos conjuntos de datos imputados se analizan mediante procedimientos estándar que se utilizarían comúnmente para el análisis del conjunto de datos completos, para finalmente combinar estos resultados generando estimaciones robustas (Medina y Galván, 2007). El funcionamiento de este método se puede describir en tres pasos principales:

- El análisis comienza con datos observados e incompletos. La imputación múltiple crea varias versiones completas de los datos al reemplazar los valores faltantes por valores de datos plausibles. Estos valores plausibles se extraen de una distribución modelada específicamente para cada entrada faltante, a partir de los datos observados.
- El segundo paso es estimar los parámetros de interés de cada conjunto de datos imputado.
- El último paso es combinar las M estimaciones de los parámetros en una estimación y estimar su varianza. La varianza combina la varianza dentro de cada imputación y la varianza adicional causada por la varianza agregada por los datos faltantes entre la imputación. Bajo las condiciones apropiadas, las estimaciones agrupadas son insesgadas y tienen las propiedades estadísticas correctas.

Para llevar a cabo este método deberíamos conocer la distribución de $Y_i = (Y_i^o, Y_i^m)$ con vector de parámetros θ , así se podría imputar el valor de Y_i^m por medio de la siguiente distribución condicional:

$$f(Y_i^m | Y_i^o, \theta)$$

El objetivo del proceso de imputación es muestrear a partir de esta verdadera distribución predictiva. Al no conocer θ , se estimará por medio de los datos observados ($\hat{\theta}$) para así realizar la imputación de los datos a través de:

$$f(Y_i^m | Y_i^o, \hat{\theta})$$

Cabe aclarar que este método de imputación asume que el mecanismo de los datos faltantes es MAR. Este método es importante para la imputación debido a que al imputar múltiples valores para los datos faltantes se crea una distribución considerada adecuada de posibles valores completos para cada observación faltante, esta distribución tiene en cuenta la relación entre las variables observadas y los valores faltantes, utilizando los valores existentes en los datos para estimar los valores desconocidos (Van Buuren, 2018).

3.2.9. Métodos de imputación

Con base en los mecanismos y patrones de los datos faltantes se determinará el método a utilizar para la imputación de la información ausente, donde se pueden encontrar métodos de regresión paramétrica que asume normalidad multivariada o no paramétricos que utiliza puntajes de propensión o, por otro lado, implementar el método de simulación de Monte Carlo con cadenas de Markov (MCMC).

El método de imputación múltiple más usado es el algoritmo de Imputación Múltiple con Ecuaciones Encadenadas (MICE por sus siglas en inglés). Dentro de este método existe el procedimiento de regresión FCS (Full Conditional Specification), o conocido como regresión secuencial, el cual requiere especificar un modelo de regresión para cada variable a imputar, utiliza las distribuciones condicionales univariadas para cada variable incompleta dadas las demás covariables, realizando iteraciones cíclicas para cada periodo, o tiempo de registro, hasta llegar a una convergencia determinada de los valores imputados (Van Buuren et al., 2006). Es importante tener en cuenta que las primeras realizaciones estarán correlacionados al valor inicial del algoritmo de regresión secuencial, por ello se deben hacer un número considerable de realizaciones del proceso dependiendo de la cantidad de datos faltantes dentro de la base de datos.

FCS propone obtener una distribución posterior de θ muestreando iterativamente a partir de distribuciones condicionales de la forma:

$$\begin{aligned} f(Y_1 \mid Y_{-1}, \theta_1), \\ \vdots \\ f(Y_k \mid Y_{-k}, \theta_k). \end{aligned}$$

donde $Y_{-j} = (Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_k)$, $j = 1, \dots, k$, y θ_j son los parámetros asociados a Y_j para la j -ésima variable.

Los parámetros $\theta_1, \dots, \theta_k$ son específicos de las densidades condicionales respectivas y no son necesariamente el producto de una factorización de la verdadera distribución conjunta $f(Y \mid \theta)$. Específicamente, la t -ésima iteración de ecuaciones encadenadas consta del siguiente muestreador de Gibbs:

$$\begin{aligned} \theta_1^{*(t)} &\sim f(\theta_1 \mid y_1^o, y_2^{(t-1)}, \dots, y_k^{(t-1)}) \\ y_1^{*(t)} &\sim f(y_1^m \mid y_1^o, y_2^{(t-1)}, \dots, y_k^{(t-1)}, \theta_1^{*(t)}) \\ &\vdots \\ \theta_k^{*(t)} &\sim f(\theta_k \mid y_k^o, y_1^{(t)}, y_2^{(t)}, \dots, y_{k-1}^{(t)}) \\ y_k^{*(t)} &\sim f(y_k^m \mid y_k^o, y_1^{(t)}, \dots, y_{k-1}^{(t)}, \theta_k^{*(t)}) \end{aligned}$$

donde $Y_j^t = (Y_j^o, Y_j^{*(t)})$ es la j -ésima variable imputada en la iteración t .

Procedimiento de Imputación múltiple

El procedimiento para realizar la imputación múltiple propuesto por Rubin (1988) combina distintas estimaciones generadas a partir de M imputaciones. Utilizando la notación de Schafer considere un vector de estimaciones $\hat{Q}$ y U la matriz de varianzas asociada a $\hat{Q}$. Después de generar M conjuntos de datos mediante simulaciones, se tienen M vectores de estimaciones de $\hat{Q}$ y de U (Schafer, 2003). Lo cual nos lleva a combinarlos de la siguiente manera:

$$\bar{Q} = \frac{1}{M} \sum_{i=1}^M \hat{Q}_i$$

donde $\hat{Q}_i$ es el vector de estimaciones de la i -ésima base de datos imputada.

La matriz de varianza de $\bar{Q}$ se obtiene a partir de dos componentes:

- La varianza de cada imputación denotada como:

$$\bar{U} = \frac{1}{M} \sum_{i=1}^M \hat{U}_i$$

- la varianza entre las imputaciones denotada como:

$$\bar{B} = \frac{1}{M-1} \sum_{i=1}^M (\hat{Q}_i - \bar{Q})(\hat{Q}_i - \bar{Q})'$$

Por tanto, la varianza total T se obtiene sumando las expresiones de las dos componentes de la varianza, corrigiendo el número finito de imputaciones por el valor $\frac{M+1}{M}$:

$$T = \bar{U} + \frac{M+1}{M} \bar{B}$$

Según Rubin (1988) la implementación de esta metodología asume algunos supuestos:

- El patrón de datos faltantes es aleatorio (MAR).
- Se requiere que el modelo estadístico utilizado para generar los datos imputados sea apropiado, es decir que exista correlación alta entre la variable a imputar y el vector de covariables que se utilizará para modelar los datos que se utilizarán para reemplazar los datos omitidos.
- Es necesario que el modelo de análisis se relacione con el que se utilizó para efectuar el proceso de imputación de los datos.

3.2.10. Eficiencia método IM

La imputación múltiple se presenta atractiva para este problema de datos faltantes por su eficiencia en los resultados, siendo capaz de generar resultados robustos con pocas interacciones del conjunto de datos completos. De acuerdo con Rubin (1988), la eficiencia relativa de los estimadores para M imputaciones está definida por:

$$\left(1 + \frac{\lambda}{M}\right)^{-1}$$

donde λ será la tasa de registros sin información del conjunto de datos.

La eficiencia del método depende de varios factores, incluyendo la cantidad y patrón de los datos faltantes, el modelo de imputación utilizado y la calidad de los datos completos. En general, a medida que aumenta el número de simulaciones (M), se espera que la eficiencia

del método mejor. Esto se debe a que con más simulaciones, se genera un mayor número de conjuntos completos de datos imputados, lo que permite una mejor estimación de la variabilidad y la incertidumbre asociada con los valores imputados.

La elección del número óptimo de simulaciones en el método depende del equilibrio entre la eficiencia deseada y los recursos computacionales disponibles. Es comúnmente aceptado que un número de simulaciones entre 5 y 10 puede proporcionar resultados razonablemente eficientes. Sin embargo, esta elección puede variar según el contexto y los requisitos específicos del análisis de datos.

4 Metodología

En este capítulo se describirá la metodología implementada para el desarrollo de este trabajo. Tenemos como primera medida la descripción de las características de la población en estudio y el tipo de estudio implementado para la recolección de la información, seguido a esto se presentan los procesos para el análisis exploratorio tanto de la información disponible como de los datos faltantes dentro de la base de datos. Seguido se detallan los procesos de modelado para la implementación del método de imputación múltiple para el manejo de la información faltante. Por último, se describe el modelamiento por medio de regresión cuantílica para la variable respuesta peso y sus respectivas covariables.

4.1. Tipo de estudio

Se realizó un estudio de datos longitudinales para una cohorte de neonatos con bajo peso al nacer, mediante el seguimiento al neonato con respecto a diferentes variables antropométricas de interés para medir su evolución. En este trabajo la variable a analizar es el peso medido a través del tiempo con el cual se realizará el análisis de la evolución de la variable de interés en neonatos con BPN con los métodos estadísticos apropiados y así observar cómo el peso de estos se va desarrollando a medida que transcurre una cantidad determinada de tiempo.

Se tomaron medidas del peso del neonato en 6 momentos: el primero y punto de partida del estudio fue el tiempo cero, que se fijó en la semana 40 de la edad gestacional del neonato, posteriormente se realizan medidas a los 6, 12, 18, 24 y 36 meses del punto de referencia del seguimiento.

En la Tabla 4-1 se muestran las variables que se incluirán en el estudio, además de sus descripciones, las cuales son consideradas variables médicamente influyentes para el peso al nacer de los neonatos y dentro de las cuales se encuentran medidas de diferentes aspectos, tanto del neonato como de la madre y su proceso durante el embarazo, que puedan afectar el peso en el nacimiento del neonato.

Tabla 4-1: Variables incluidas en el estudio

Variable	Descripción	Niveles
Tiempo	Tiempo en meses en el que se realizan los seguimientos	0, 6, 12, 18, 24, 36
Sexo	Genero del recién nacido	Masculino, Femenino
Peso al nacer	Peso en gramos medido cuando el bebe nace	Numérica
Peso	Peso en gramos del neonato a lo largo del tiempo	Numérica
Tipo de parto	Tipo de parto mediante el cual nació el neonato	Cesárea, Espontáneo, Vaginal
Hipertensión	Si la madre presentaba hipertensión	Si, No
Edad gestacional al nacer	Edad gestacional en semanas en la que nació el neonato	Numérica
Control prenatal	El neonato requirió control prenatal	Si, No
Ventilación asistida	El neonato requirió Ventilación asistida	Si, No

4.2. Población de estudio

Para este estudio se dispuso de una muestra de 259 historias clínicas del Hospital Universitario del Valle (HUV), Cali, Colombia, tomadas entre febrero de 2002 y noviembre de 2015, estas historias pertenecientes al Programa de Seguimiento de Alto Riesgo y Canguro del HUV. Los principales criterios para la inclusión del neonato al programa definidos por la unidad de recién nacidos fueron: recién nacidos con un peso menor a los 2001 gramos, regulación térmica y del patrón respiratorio adecuado, saturación de oxígeno normal, procedentes de la ciudad de Cali y familiar con la disposición de colaborar y cumplir con el seguimiento del programa. Se excluyeron del programa de seguimiento los neonatos que fueron remitidos a otra institución, presentaron malformaciones mayores o letales, y los neonatos abandonados o que fueron dejados en adopción.

4.3. Análisis exploratorio de los datos

El objetivo de este análisis fue explorar y analizar por medio de resúmenes descriptivos y herramientas gráficas el comportamiento de los datos para tener una primera idea de los métodos estadísticos adecuados para la modelación, debido a que esto nos podría mostrar anomalías dentro de los datos como lo pueden ser datos atípicos u observaciones inusuales, también nos podría revelar patrones dentro de la variable de interés y observar el grado de relación entre las variables. Cabe mencionar que todo el análisis y el procesamiento de la base de datos se realizó con el software estadístico RStudio así como también la obtención de los modelos planteados.

En este análisis se agregaron algunas curvas de peso con los datos disponibles dentro de la base de datos para los tiempos (0, 6, 12, 18, 24, 36), con el fin de observar el comportamiento previo a la imputación de los valores faltantes y su diferencia luego de imputada la base de datos; además de esto se obtuvieron diferentes curvas de peso teniendo en cuenta las variables cualitativas como: sexo, tipo de parto, hipertensión y control prenatal, con esto se determinó

si existían diferencias notorias entre las diferentes categorías de las variables las cuales puedan influir en los resultados de los análisis estadísticos posteriores.

4.3.1. Patrones de datos faltantes

El análisis de los patrones de comportamiento de los datos faltantes se llevó a cabo por medio de análisis gráficos a través de la ubicación de estos datos dentro de la base de datos y se acompañó por medio de conteos de los registros faltantes en diferentes circunstancias, estas herramientas nos proporcionaron un primer indicio del posible método a utilizar para la imputación de la información faltante y nos dieron una idea de las posibles razones de la existencia de estos datos faltantes en los diferentes tiempos de seguimiento del estudio.

Como primera medida se realizó un marco de los patrones de datos, tomando la base de datos como una matriz, en el cual se visualizaron los valores faltantes y los datos completos para cada una de las variables, proporcionando a su vez el porcentaje de datos faltantes para cada variable y el porcentaje de ausencias para la base de datos conjunta.

Seguido a lo anterior proporcionamos algunas estadísticas descriptivas que nos permitieron entender de una manera más adecuada y con mayor claridad el comportamiento de los datos faltantes dentro de la base de datos. Como primera medida se realizó un conteo de registros faltantes en cada uno de los tiempos para identificar patrones y características de los datos faltantes en el estudio longitudinal. Luego, se contabilizó la cantidad de seguimientos realizados para cada paciente, con el objetivo de determinar la proporción de información disponible y entender el comportamiento de los datos faltantes en diferentes momentos. Por último, se propusieron estadísticas descriptivas para las variables restantes, incluyendo proporciones de clases en variables cualitativas y medidas de posición central en variables cuantitativas, junto con la cantidad de valores faltantes para cada variable.

4.4. Imputación múltiple

El principal objetivo del estudio se enfocó en el manejo de los datos faltantes dentro de las observaciones, donde se tomó la elección de la técnica del manejo de los datos faltantes con base en la distribución de la variable, los parámetros de interés para el estudio, el tamaño de la base de datos y la tasa de datos faltantes. A este problema se le dio solución por medio del método de imputación múltiple, llevándose a cabo a través de simulaciones estadísticas, para crear así diferentes conjuntos de los datos completos, donde cada conjunto tuvo estimaciones de los datos faltantes, posteriormente, se realizó un análisis estadístico combinando cada conjunto de datos completo estimado, y de esta forma se generaron estimadores robustos e insesgados.

El proceso de imputación múltiple de la base de datos se llevó a cabo a través de la función `mice()`, esta nos permitió determinar la cantidad de imputaciones M a realizar con las respectivas iteraciones de las simulaciones, además se ajustó la matriz de predictores dentro de la función dependiendo del modelo a ajustar para cada variable a imputar y con esta matriz pudimos asignar un método específico de imputación a cada variable dependiendo del tipo o características de la misma.

4.4.1. Métodos de imputación

Hicimos uso de 3 de los métodos que existen dentro de la gran variedad de métodos en la función `mice()` (Van Buuren, 2011), los cuales se seleccionaron basándonos en los tipos de variables con las que disponemos en la base, ya que se imputarán tanto variables cuantitativas como variables cualitativas, los métodos utilizados fueron:

- **Regresión lineal bayesiana (Norm):**

En el método `Norm` se utiliza una regresión lineal múltiple para imputar los valores faltantes en cada variable. En primer lugar, se ajusta un modelo de regresión lineal múltiple para cada variable con datos faltantes, utilizando las variables predictoras que tienen datos disponibles. Luego, se utiliza la distribución normal de los residuos del modelo para simular múltiples valores imputados para cada valor faltante.

Es uno de los métodos de imputación más utilizados y se basa en la suposición de que los datos faltantes tienen una distribución normal condicional en las variables predictoras disponibles. Este método es útil para imputar valores faltantes en conjuntos de datos con una estructura compleja, como los datos longitudinales o los datos de encuestas.

- **Regresión logística (Logreg):**

En este método, se ajusta un modelo de regresión logística para cada variable categórica con datos faltantes, utilizando las variables predictoras que tienen datos disponibles. Luego, se utiliza el resultado de este modelo para imputar valores faltantes de cada observación. También permite la imputación de datos faltantes en múltiples variables categóricas simultáneamente, lo que nos ayuda a mejorar la precisión de las estimaciones del método. Este tipo de método se utiliza para las covariables categóricas con dos clases o categorías.

- **Regresión logística politómica (Polyreg):**

La regresión logística politómica es un enfoque para imputar valores faltantes en variables categóricas con más de dos categorías. La regresión logística politómica se utiliza

para modelar la relación entre una variable categórica con múltiples categorías y un conjunto de variables predictoras.

Utilizamos el método **Norm** para la variable edad gestacional al ser esta de tipo continuo, el método **Polyreg** en la variable tipo parto al tener 3 clases dentro de su característica y el método **Logreg** para las variables hipertensión, control prenatal y ventilación asistida al ser estas variables de tipo dicotómicas.

4.4.2. Modelos de imputación

Se ajustaron modelos de imputación para cada una de las variables que tuvieran valores faltantes. Dentro de todas las variables de la base de datos se tienen las variables sexo y peso al nacer con la información completa, estas variables se incluyeron dentro de los modelos de imputación como covariables para poder tener un mayor rango de información y hacer de las imputaciones más robustas.

Primero se imputaron las covariables que se utilizaron en el modelo de regresión cuantílica, esto con el fin de que a la hora de imputar el peso en los tiempos de seguimiento se dispusiera de la información completa de estas variables. Para ello utilizamos el método secuencial, donde primero se imputa una variable y seguido se utiliza esta variable imputada para realizar el proceso de imputación de otra variable incompleta.

Siguiendo la secuencia de modelos para estas variables:

- Para la variable tipo parto, la probabilidad de que el neonato i haya nacido mediante tipo de parto espontáneo es:

$$\begin{aligned} \text{logit } P(\text{TipoParto}_i = 1) = & \beta_0 + \beta_1 \text{Sexo}_i + \beta_2 \text{PesoNacer}_i + \beta_3 \text{Hipertension}_i \\ & + \beta_4 \text{EgNacido}_i + \beta_5 \text{ContPrenatal}_i + \beta_6 \text{VentAsis}_i \end{aligned}$$

Y para el tipo de parto vaginal es:

$$\begin{aligned} \text{logit } P(\text{TipoParto}_i = 2) = & \beta_0 + \beta_1 \text{Sexo}_i + \beta_2 \text{PesoNacer}_i + \beta_3 \text{Hipertension}_i \\ & + \beta_4 \text{EgNacido}_i + \beta_5 \text{ContPrenatal}_i + \beta_6 \text{VentAsis}_i \end{aligned}$$

- La probabilidad de que el neonato i presentara hipertensión en el registro de la madre es:

$$\begin{aligned} \text{logit } P(\text{Hipertension}_i = 1) = & \beta_0 + \beta_1 \text{Sexo}_i + \beta_2 \text{PesoNacer}_i + \beta_3 \text{TipoParto}_i \\ & + \beta_4 \text{EgNacido}_i + \beta_5 \text{ContPrenatal}_i + \beta_6 \text{VentAsis}_i \end{aligned}$$

- Para el neonato i el valor predicho de la edad gestacional al nacer estará condicionado por el siguiente modelo:

$$\begin{aligned} \text{EgNacido}_i = & \beta_0 + \beta_1 \text{Sexo}_i + \beta_2 \text{PesoNacer}_i + \beta_3 \text{TipoParto}_i + \beta_4 \text{Hipertension}_i \\ & + \beta_5 \text{ContPrenatal}_i + \beta_6 \text{VentAsis}_i + u_i \end{aligned}$$

donde $u_i \sim N(0, \sigma^2)$

- Para la variable control prenatal, la probabilidad de que el neonato i haya tenido control prenatal es:

$$\begin{aligned} \text{logit } P(\text{ContPrenatal}_i = 1) = & \beta_0 + \beta_1 \text{Sexo}_i + \beta_2 \text{PesoNacer}_i + \beta_3 \text{TipoParto}_i \\ & + \beta_4 \text{Hipertension}_i + \beta_5 \text{EgNacido}_i + \beta_6 \text{VentAsis}_i \end{aligned}$$

- La probabilidad de que el neonato i haya necesitado ventilación asistida es:

$$\begin{aligned} \text{logit } P(\text{VentAsis}_i = 1) = & \beta_0 + \beta_1 \text{Sexo}_i + \beta_2 \text{PesoNacer}_i + \beta_3 \text{TipoParto}_i \\ & + \beta_4 \text{Hipertension}_i + \beta_5 \text{EgNacido}_i + \beta_6 \text{ContPrenatal}_i \end{aligned}$$

Para la imputación del peso en los diferentes tiempos de seguimiento se utilizó el mismo procedimiento de regresión secuencial, donde se tomó cada tiempo como variable y se realizó la imputación, contando con la información completa de las variables anteriormente imputadas y siendo agregadas estas a los modelos de imputación de los pesos en los tiempos.

Para cada uno de los pesos en los tiempos (0 a 36) se utilizó el método **Norm** debido a que son continuos y el comportamiento del peso se considera de comportamiento normal en la gran mayoría de los estudios, obteniendo así los siguientes modelos de imputación para cada peso en el tiempo:

$$\begin{aligned} \text{Peso } 0_i &= \beta_0 + \beta_1 \text{Sexo}_i + \beta_2 \text{PesoNacer}_i + \beta_3 \text{TipoParto}_i + \beta_4 \text{Hipertension}_i \\ &+ \beta_5 \text{EgNacido}_i + \beta_6 \text{ContPrenatal}_i + \beta_7 \text{VentAsis}_i + u_i \end{aligned}$$

$$\begin{aligned} \text{Peso } 6_i &= \beta_0 + \beta_1 \text{Peso } 0_i + \beta_2 \text{Sexo}_i + \beta_3 \text{PesoNacer}_i + \beta_4 \text{TipoParto}_i + \beta_5 \text{Hipertension}_i \\ &+ \beta_6 \text{EgNacido}_i + \beta_7 \text{ContPrenatal}_i + \beta_8 \text{VentAsis}_i + u_i \end{aligned}$$

$$\begin{aligned} \text{Peso } 12_i &= \beta_0 + \beta_1 \text{Peso } 6_i + \beta_2 \text{Peso } 0_i + \beta_3 \text{Sexo}_i + \beta_4 \text{PesoNacer}_i + \beta_5 \text{TipoParto}_i \\ &+ \beta_6 \text{Hipertension}_i + \beta_7 \text{EgNacido}_i + \beta_8 \text{ContPrenatal}_i + \beta_9 \text{VentAsis}_i + u_i \end{aligned}$$

$$\begin{aligned} \text{Peso } 18_i &= \beta_0 + \beta_1 \text{Peso } 12_i + \beta_2 \text{Peso } 6_i + \beta_3 \text{Peso } 0_i + \beta_4 \text{Sexo}_i + \beta_5 \text{PesoNacer}_i \\ &+ \beta_6 \text{TipoParto}_i + \beta_7 \text{Hipertension}_i + \beta_8 \text{EgNacido}_i + \beta_9 \text{ContPrenatal}_i \\ &+ \beta_{10} \text{VentAsis}_i + u_i \end{aligned}$$

$$\begin{aligned} \text{Peso } 24_i &= \beta_0 + \beta_1 \text{Peso } 18_i + \beta_2 \text{Peso } 12_i + \beta_3 \text{Peso } 6_i + \beta_4 \text{Peso } 0_i + \beta_5 \text{Sexo}_i \\ &+ \beta_6 \text{PesoNacer}_i + \beta_7 \text{TipoParto}_i + \beta_8 \text{Hipertension}_i + \beta_9 \text{EgNacido}_i \\ &+ \beta_{10} \text{ContPrenatal}_i + \beta_{11} \text{VentAsis}_i + u_i \end{aligned}$$

$$\begin{aligned} \text{Peso } 36_i &= \beta_0 + \beta_1 \text{Peso } 24_i + \beta_2 \text{Peso } 18_i + \beta_3 \text{Peso } 12_i + \beta_4 \text{Peso } 6_i \\ &+ \beta_5 \text{Peso } 0_i + \beta_6 \text{Sexo}_i + \beta_7 \text{PesoNacer}_i + \beta_8 \text{TipoParto}_i + \beta_9 \text{Hipertension}_i \\ &+ \beta_{10} \text{EgNacido}_i + \beta_{11} \text{ContPrenatal}_i + \beta_{12} \text{VentAsis}_i + u_i \end{aligned}$$

donde $u_i \sim N(0, \sigma^2)$ para cada uno de los modelos anteriores.

Cada modelo representa el peso en el tiempo para el i -ésimo neonato a imputar, donde los pesos en los tiempos 6 al tiempo 36 utilizarán los pesos anteriormente imputados.

4.5. Modelo por regresión cuantílica

Para el modelamiento del problema se utilizó regresión cuantílica debido a que este tipo de regresión nos permitió encontrar relaciones entre la variable respuesta y la variable predictora para diferentes cuantiles, dándole así una ponderación a cada cuantil en específico y permitiéndonos conocer mejor los cambios existentes en la variable respuesta con cada cambio en las variables predictoras. Además de lo anterior, la cualidad de esta técnica que nos lleva a tomar la decisión de trabajar con este tipo de regresión es que nos permitió modelar valores extremos con una mayor robustez, donde estos datos extremos representan neonatos con extremo bajo peso, siendo estos de vital importancia para el enfoque médico, y a su vez este modelo nos brindó una mayor flexibilidad para el modelamiento donde existe un alto grado de variabilidad en los registros.

Para este proceso utilizamos la función `rq()` de la librería `quantreg`, el modelo a ajustar se ejecutó para cada una de las M bases de datos completas realizadas en la imputación de la información faltante. Para cada uno de los modelos se realizó estimaciones para los cuantiles 0.1, 0.25, 0.5, 0.75 y 0.9 con el fin de observar el cambio entre las relaciones de las variables predictoras a medida que cambia el nivel de probabilidad, o cuantil, de la distribución de la variable respuesta. Siendo los cuantiles extremos de gran ayuda para explicar el comportamiento de pesos extremos dentro de la base de datos, los cuales requieren de un mayor grado de seguimiento para el estudio clínico médico.

4.5.1. Transformación del modelo de regresión

Se realizó una transformación en el modelo de regresión para mejorar su capacidad de ajuste. La transformación se aplica a la variable tiempo, ya que la relación entre el peso y el tiempo no es lineal debido a los cambios físicos variables y rápidos en los neonatos a lo largo del tiempo. La transformación del tiempo busca linealizar esta relación y reducir la influencia de valores extremos, por lo general la evolución de los neonatos para el modelado de curvas de crecimiento sigue la forma de una curva sigmoidea, esta tiene un crecimiento rápido en los primeros tiempos del estudio y a medida que transcurre el tiempo se estabiliza y desacelera su crecimiento. Se tomó como referencia a Johnson et al. (2013), los cuales utilizan modelos de efectos mixtos para generar curvas de crecimiento del peso en diferentes circunstancias. Se utilizó la función de primer orden de Berkey-Reed la cual se ajusta adecuadamente para el crecimiento de los neonatos debido a su modelamiento secuencial que establece una forma característica en la curva de crecimiento, esto puede deberse a que en esta función se considera una ecuación sigmoideal, siendo apropiada para estudios de crecimiento a lo largo del tiempo.

Con lo anterior, establecemos el modelo de regresión cuantílica con base en nuestro estudio y variables, quedando definido de la siguiente manera:

$$\begin{aligned} Q_{Peso_i}(\tau) = & \beta_{0,\tau} + \beta_{1,\tau}(Tiempo_i) + \beta_{2,\tau} \ln(Tiempo_i) + \beta_{3,\tau} \left(\frac{1}{Tiempo_i} \right) + \beta_{4,\tau}(PesoNacer_i) \\ & + \beta_{5,\tau}(Sexo_i) + \beta_{6,\tau}(TipoParto \ 1_i) + \beta_{7,\tau}(TipoParto \ 2_i) \\ & + \beta_{8,\tau}(Hipertension_i) + \beta_{9,\tau}(EgNacido_i) + \beta_{10,\tau}(ContPrenatal_i) \\ & + \beta_{11,\tau}(VentAsis_i) \end{aligned}$$

donde $TipoParto \ 1_i = 1$ si el parto del i -ésimo neonato fue espontáneo, $TipoParto \ 1_i = 0$ si lo contrario; y $TipoParto \ 2_i = 1$ si el parto del i -ésimo neonato fue vaginal, $TipoParto \ 2_i = 0$ si lo contrario. $Sexo_i = 1$ si el i -ésimo neonato es hombre y $Sexo_i = 0$ si el i -ésimo neonato es mujer. Para las covariables hipertensión, control prenatal y ventilación asistida, la categoría de referencia es la ausencia de la característica.

$\beta_{0,\tau}$ es el intercepto que está ligado con el peso en la semana gestacional 40 del neonato, $\beta_{1,\tau}$ representa el componente lineal entre el peso y el tiempo, $\beta_{2,\tau}$ se relaciona con la disminución de la tasa de crecimiento en el transcurso del tiempo del estudio y $\beta_{3,\tau}$ representara el punto de inflexión que encontraremos en los primeros tiempos de la curva de crecimiento de los neonatos. Desde $\beta_{4,\tau}$ hasta $\beta_{11,\tau}$ se ve representado el efecto debido a las covariables peso al nacer, edad gestacional del nacido, sexo, hipertensión, ventilación asistida, control prenatal, tipo de parto, respectivamente.

4.5.2. Estimación errores estándar

El proceso de estimación de los errores estándar de los coeficientes de regresión se llevó a cabo por medio de algunos ajustes a través de métodos bootstrap y de cluster para garantizar una mayor confianza en los resultados, donde se obtuvo la matriz de varianzas-covarianzas de cada modelo de regresión cuantílica ajustado para cada una de las M imputaciones realizadas en el proceso, para así poder obtener la estimación global de los errores requeridos.

Uno de los ajustes desarrollados en la estimación y cálculo de estos errores es por medio de la técnica de inferencia bootstrap, esta se llevó a cabo debido a que nos proporciona una medida de la incertidumbre de los coeficientes de regresión, permitiéndonos obtener inferencias más precisas sobre los efectos de los predictores en los diferentes cuantiles de la distribución de la variable respuesta. En este caso, se realizan múltiples remuestreos con reemplazo de cada uno de los datos completos imputados y se ajustan modelos cuantílicos a cada uno de los conjuntos de datos remuestreados. Luego, se calculan los errores estándar de los coeficientes a partir de la distribución de los coeficientes de los modelos cuantílicos ajustados en cada

uno de los conjuntos de datos remuestreados. Esta técnica permite obtener una estimación más precisa de los errores estándar de los coeficientes en modelos de regresión cuantílica, especialmente cuando el supuesto de independencia de los errores no se cumple (Efron y Tibshirani, 1994).

Adicional al proceso anterior, se realizó una corrección por cluster dado que en los datos de tipo longitudinal por lo general se asume que las observaciones dentro de cada uno de los individuos estarán correlacionados, al ignorar esta correlación se pueden producir problemas en el cálculo de los errores estándar y a su vez en la significancia estadística de los coeficientes. Para realizar la técnica se define como clusters cada uno de los neonatos referenciados por el número de historia clínica dentro del estudio, para luego obtener la matriz de varianza-covarianza por conglomerados, el cual tiene en cuenta las correlaciones de los datos dentro de cada individuo. Esta corrección se realiza mediante la reorganización de los grupos en el bootstrap y la generación de estadísticas bootstrap a nivel de grupo en lugar de a nivel de observación, lo cual mejora la precisión de las estimaciones y reduce el sesgo en la inferencia bootstrap.

4.5.3. Test de hipótesis

Se realizó la evaluación de las variables predictoras dentro del modelo de regresión cuantílica por medio de la significancia de los coeficientes de la función de cuantil requerida, para así poder determinar las covariables que tienen un peso significativo dentro del modelo y obtener las características específicas que influyen en el desarrollo del peso a lo largo del tiempo. Se comprobó lo anterior por medio de las siguientes hipótesis sobre los regresores de interés:

$$\begin{aligned} H_0 : \hat{\beta}_\tau^* &= 0 \\ H_a : \hat{\beta}_\tau^* &\neq 0 \end{aligned}$$

donde $\hat{\beta}_\tau^* = (\beta_{5,\tau}, \beta_{6,\tau}, \beta_{7,\tau}, \beta_{8,\tau}, \beta_{9,\tau}, \beta_{10,\tau}, \beta_{11,\tau})$ representa el vector de regresores comprobados por medio de la prueba de hipótesis. Estos regresores ligados a las variables sexo, tipo de parto, hipertensión, edad gestacional, control prenatal y ventilación asistida.

4.6. Construcción de las curvas de crecimiento

El proceso de construcción de las curvas de crecimiento del peso a lo largo del tiempo se basó en las estimaciones de los coeficientes del modelo de regresión cuantílica ajustado para diferentes cuantiles. Se obtuvieron diferentes funciones según las características deseadas de

las curvas y los valores cuantílicos ajustados para cada modelo. Estas funciones se componen de las estimaciones de los coeficientes ajustados en el modelo para cada uno de los 5 cuantiles evaluados, y asignando un valor específico a cada variable en la función para representar diferentes situaciones en el estudio.

Los valores asignados a las variables dependen de las comparaciones que se deseen realizar. En este caso, se compararon las variables cualitativas, para observar el comportamiento de las diferentes curvas en las clases de estas variables. Se asignaron los valores correspondientes a cada clase de estas variables en las diferentes funciones, lo que resultó en 5 curvas de crecimiento para cada clase. Para las variables peso al nacer y edad gestacional del recién nacido, que son variables continuas, se les asignó un valor fijo estimado en todas las funciones. En este caso, se utilizó la mediana de cada una de estas variables como un punto medio para los datos, considerando la posible presencia de valores extremos que pudieran afectar las curvas.

5 Resultados

Se realiza un análisis exploratorio para describir como se comportan los datos, las características iniciales de la población estudiada, para posteriormente exponer los resultados del ajuste de los modelos de regresión por cuantiles y las curvas de crecimiento.

5.1. Análisis exploratorio

Las características de la población según su sexo, 134 (51.7 %) mujeres y 125 (48.3 %) hombres, así mismo la mediana del peso (1271 gramos) es muy baja según la clasificación realizada por la OMS (2017) y oscila dentro de un rango intercuartílico de 1054 a 1450 gramos. La mediana de la edad gestacional indica que fueron bebés prematuros (30 semanas) con un rango intercuartílico que va desde 28 hasta 32 semanas. Los resultados (5-1) muestran que la gran mayoría de las madres gestantes no tenían hipertensión y sus bebés nacieron mediante cesárea. La mayoría de las madres asistieron al control prenatal; así mismo, la mayoría de los neonatos requirieron de ventilación asistida.

Tabla 5-1: Distribución de la población según antecedente de hipertensión arterial, tipo de parto, control prenatal y ventilación asistida.

Variable	N = 259
Hipertensión (HA)	
No	164 (69.8 %)
Si	71 (30.2 %)
Datos faltantes	24
Tipo de parto (TP)	
Cesárea	178 (70.6 %)
Espontáneo	14 (5.6 %)
Vaginal	60 (23.8 %)
Datos faltantes	7
Control prenatal (CP)	
No	32 (14.5 %)
Si	189 (85.5 %)
Datos faltantes	38
Ventilación asistida (VA)	
No	44 (18.9 %)
Si	189 (81.1 %)
Datos faltantes	26

En la figura 5-1 se observa que el peso al nacer se comporta de manera similar para el sexo y el antecedente de hipertensión de la madre, en ambos casos las medianas se sitúan en pesos similares y son simétricas, ahora bien, con respecto a sus rangos intercuartílicos, estos se sitúan aproximadamente entre los 1000 y los 1500 gramos, con medianas que rondan los 1250 gramos. Para finalizar se observa que la varianza esperada parece ser similar para ambos factores y sus respectivos niveles.

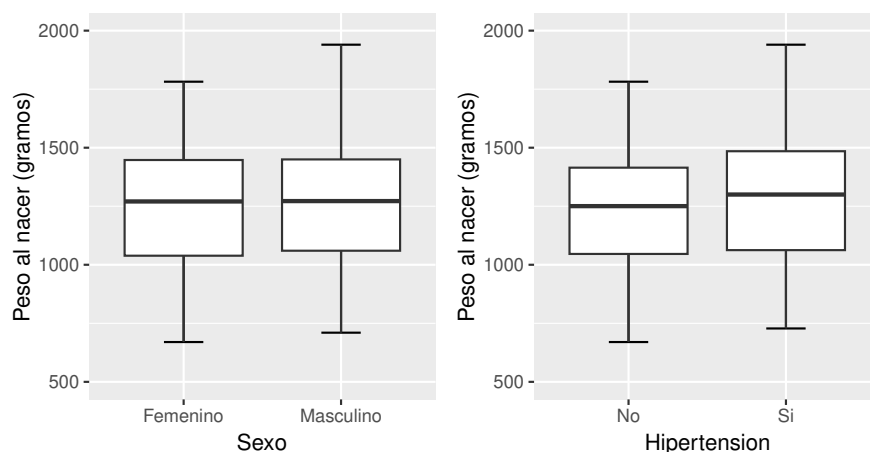


Figura 5-1: Peso al nacer según el sexo y antecedente de hipertensión

En la figura 5-2 se muestra el comportamiento de las medianas de diferentes tipos de parto con relación al peso al nacer (PN). En primer lugar, se observa que los recién nacidos por parto espontáneo tienen una mediana de PN más alta y presentan asimetría en su distribución. Mientras que los neonatos nacidos por cesárea y parto vaginal presentan una mediana de PN similar y no presentan asimetría.

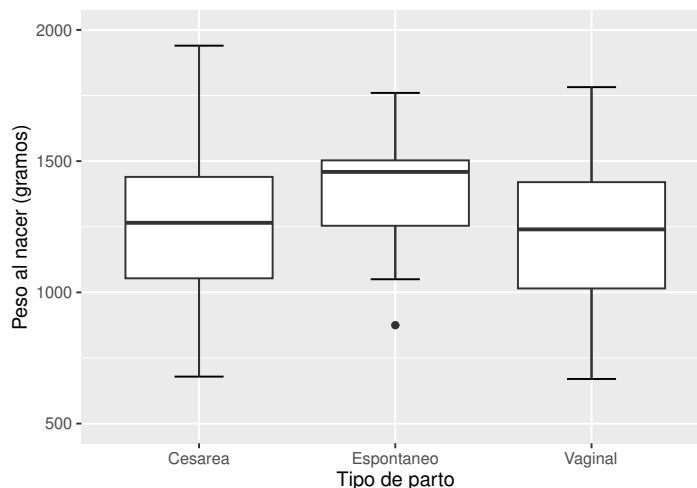


Figura 5-2: Peso al nacer según el tipo de parto

La figura 5-3 muestra que los neonatos que recibieron control prenatal tienen un mayor peso al nacer en comparación con aquellos que no recibieron este tipo de atención médica. Además, se observa que los neonatos que requirieron ventilación asistida presentan una mediana de peso menor en comparación con aquellos que no necesitaron esta intervención.

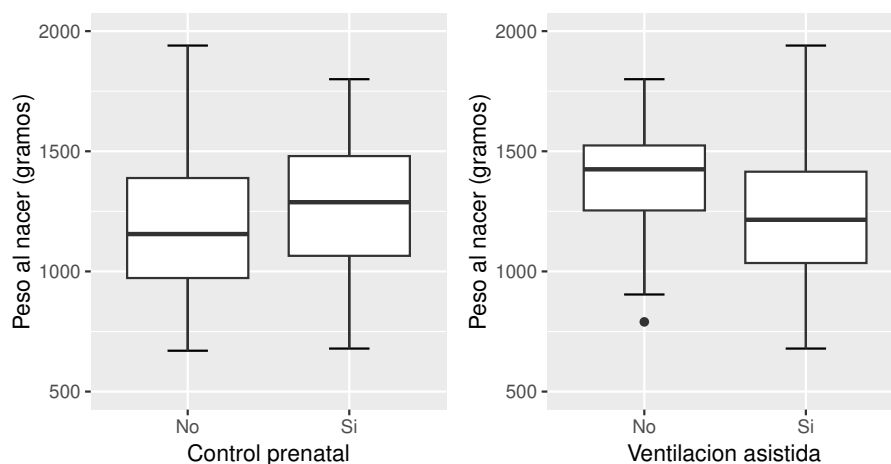


Figura 5-3: Peso al nacer según el control prenatal y ventilación asistida

En la tabla 5-2 se muestra el crecimiento de los neonatos a lo largo de los 36 meses de seguimiento según el peso, los valores de la media y la mediana van aumentando a medida que el tiempo transcurre, así mismo se puede observar que para todos los seguimientos, el CV es superior al 10 % lo que indica una dispersión moderada en los datos, llegando hasta un 21,5% en el caso específico de los 0 meses, esto sería explicado debido a que el ritmo de desarrollo en los bebés, es diferente para cada individuo. En la figura 5-4 se muestra que el crecimiento más acelerado de los neonatos ocurre de 0 a 6 meses, sin embargo, este crecimiento parece desacelerarse a medida que transcurre el tiempo. La distribución del peso en los tiempos de seguimiento es simétrico.

Tabla 5-2: Estadísticas descriptivas según el tiempo de seguimiento

Seguimiento	Media (SD)	Mediana (RIQ)	CV(%)
0 Meses	1,262.7 (271.5)	1,276.0 (1,060.0, 1,460.0)	21.5
6 Meses	6,856.0 (1,151.8)	6,870.0 (6,250.0, 7,605.0)	16.79
12 Meses	8,739.9 (1,219.8)	8,635.0 (7,762.5, 9,595.0)	13.95
18 Meses	10,129.3 (1,366.2)	9,905.0 (9,115.0, 11,130.0)	13.48
24 Meses	11,385.6 (1,483.6)	11,185.0 (10,340.0, 12,182.5)	13.03
36 Meses	13,306.8 (1,920.7)	13,050.0 (11,760.0, 14,310.0)	14.43

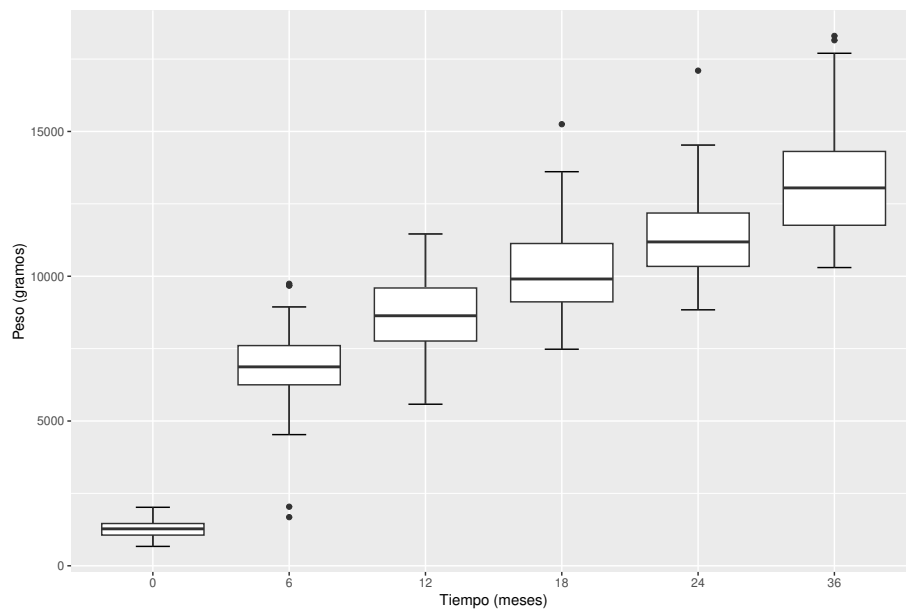


Figura 5-4: Diagrama de cajas del peso al nacer según el tiempo

La figura 5-5 muestra las curvas de peso de los neonatos a través del tiempo de seguimiento, cabe resaltar que los datos atípicos presentes evidencian formas de desarrollo inusuales en donde no parece haber un patrón de crecimiento específico, otro problema que se puede evidenciar son las curvas de desarrollo incompletas o discontinuas generadas por la gran cantidad de datos faltantes que se presentan en los seguimientos; sin embargo, se observa que en la mayoría de los casos el patrón de crecimiento se mantiene.

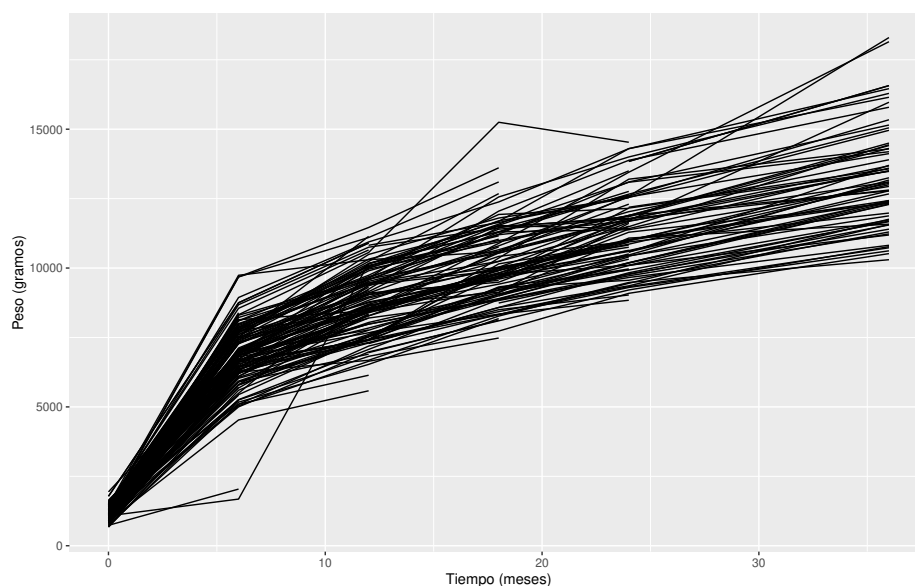


Figura 5-5: Diagrama de alambres del crecimiento por individuo

En la figura 5-6 se exponen las curvas medias del peso al nacer de los neonatos para cada uno de los niveles de los factores expuestos (Tipo de parto, Hipertensión arterial, control prenatal). Para los casos específicos de la de la hipertensión y el control prenatal, no parecen haber diferencias notorias en la curva media de crecimiento, ya que no se alteran los pesos de los neonatos con respecto a los niveles de los factores durante el tiempo de seguimiento, sin embargo, para el tipo de parto se encuentra un leve diferencia en los pesos de los 24 a los 36 meses según cada tipo de parto.

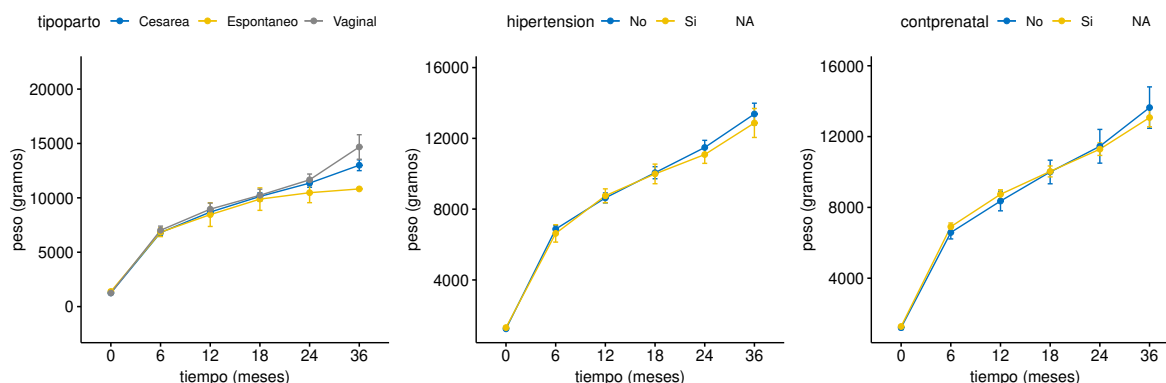


Figura 5-6: Comportamiento del peso en el tiempo según tipo de parto, antecedente de hipertensión y control prenatal

En la figura 5-7 se muestra el comportamiento del peso a través del tiempo de seguimiento por sexo y ventilación asistida, se observa entonces que para el caso de la variable sexo el peso medio a lo largo del tiempo de seguimiento de los hombres es un poco mayor al de las mujeres. Para la variable ventilación asistida, el desarrollo en el peso medio es similar en ambos niveles de la variable.

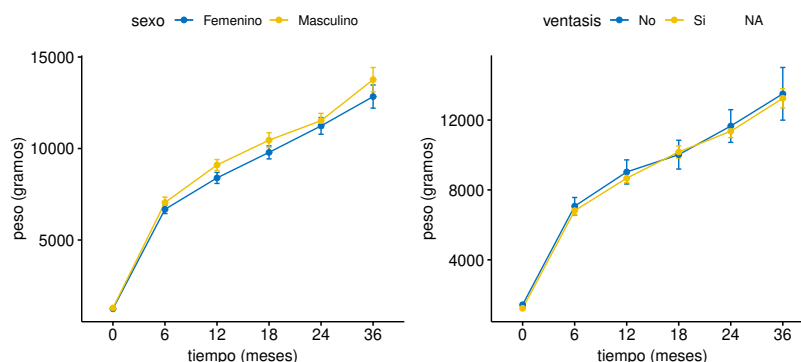


Figura 5-7: Comportamiento del peso en el tiempo según sexo y ventilación asistida

5.2. Patrón de datos faltantes

Para analizar los datos faltantes que posee la base de datos del PMC, se analizó la variable respuesta (peso) de distintas formas, en primer lugar se precisa que de manera conjunta (analizando la variable peso sin segmentarla por meses de seguimiento) no se observaron el 49 % de los pesos de los neonatos, lo que denota un problema fuerte de datos faltantes. Por consiguiente, se contabilizan cuantas observaciones para el peso se tienen en cada uno de los tiempos, los resultados se muestran en la tabla **5-3**, se puede observar que a medida que el tiempo transcurre, los datos faltantes van aumentando, a los 3 años solo se tiene el registro del 26 % de los neonatos.

Tabla 5-3: Datos faltantes por periodo de tiempo

Peso	Tiempo en meses					
	0	6	12	18	24	36
Datos faltantes n(%)	3 (1.2 %)	110 (42.5 %)	137 (52.9 %)	161 (62.2 %)	161 (62.2 %)	190 (73.4 %)

Teniendo en cuenta la tendencia al aumento de datos faltantes, a medida que el tiempo transcurre, se contabiliza en la tabla **5-4** la cantidad de seguimientos que posee cada individuo estudiado, se observa que hay una alta presencia de neonatos con información incompleta, solo el 20 % de los neonatos tienen los registros completos de los pesos. De igual forma, casi el 40 % de los neonatos solo cuentan con una medición del peso.

Tabla 5-4: Número de seguimientos asistidos

Pacientes que:	N=259	%
Asistieron a 1 seguimiento	102	39.4
Asistieron a 2 seguimientos	32	12.4
Asistieron a 3 seguimientos	20	7.7
Asistieron a 4 seguimientos	11	4.2
Asistieron a 5 seguimientos	41	15.8
Asistieron a 6 seguimientos	53	20.5

En la figura **5-8** se observa en orden descendente las covariables y sus respectivos porcentajes de datos faltantes, la gran mayoría de los registros faltantes presentes en las covariables poseen cantidades bajas y moderadas de información perdida, por lo que se puede concluir entonces que la gran mayoría de la información faltante se presenta a la hora de medir la variable peso, como se expresa en la tabla **5-3**, ya que la naturaleza longitudinal de esta variable hace que sea común encontrarse con altos porcentajes de registros faltantes.

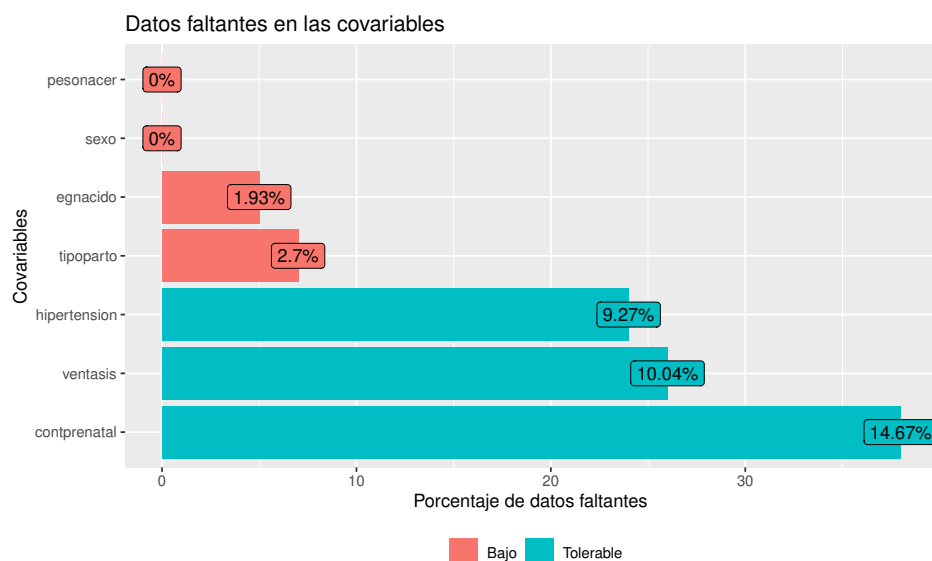


Figura 5-8: Análisis descriptivo de los datos faltantes en las covariables

En la figura 5-9 se observa el patrón de datos faltantes de la base de datos estudiada, donde analizamos que de manera conjunta el 25 % de los datos están perdidos, evidenciando un problema significativo, además se identificó que la cantidad de valores faltantes aumentan en los pesos de los neonatos a medida que el tiempo transcurre, algo que era de esperarse debido a su naturaleza longitudinal. Es importante destacar que no se han identificado mecanismos ni patrones sistemáticos en la forma en que se presenta esta falta de información en lo que a las covariables respecta.



Figura 5-9: Patrón de datos faltantes

5.3. Imputación múltiple

Mediante el modelo de imputación por FCS se obtuvieron N=25,900 valores distribuidos en 100 bases de datos, estos datos fueron agrupados en una base global. Para la tabla 5-5 se observan las estadísticas descriptivas para los seguimientos, el análisis más importante para resaltar con esta tabla es que las mediciones realizadas se están comportando de manera similar a las mediciones iniciales, por lo que mantienen el patrón de crecimiento.

Tabla 5-5: Estadísticas descriptivas de los pesos imputados

Seguimiento	Peso medio (SD) con N=25,900
0 meses	1,262.6 (270.7)
6 meses	6,949.2 (1,205.2)
12 meses	8,829.8 (1,300.4)
18 meses	10,173.5 (1,455.0)
24 meses	11,409.6 (1,648.7)
36 meses	13,393.0 (2,126.3)

Como se puede observar en la tabla 5-6, se conservan los patrones de comportamiento en las variables Tipo de parto, Control prenatal, ventilación asistida e hipertensión, además se precisa que el sexo y el peso al nacer mediano (1.271 gramos) al no contener datos faltantes siguen comportándose igual a los análisis anteriores. También en la tabla 5-6, se puede observar que la mediana de la edad gestacional se mantuvo en 30 semanas.

Tabla 5-6: Distribución de las variables con observaciones imputadas

Variable	N = 25,900
Peso al nacer (PN)	1,271 (1,053.0, 1,450.0)
Tipo de parto (TP)	
Cesárea	18,272 (70.5 %)
Espontáneo	1,461 (5.6 %)
Vaginal	6,167 (23.8 %)
Hipertensión	
No	18,186 (70.2 %)
Si	7,714 (29.8 %)
Edad gestacional del nacido (EG)	30.0 (28.0, 32.0)
Control prenatal (CP)	
No	3,921 (15.1 %)
Si	21,979 (84.9 %)
Ventilación asistida (VA)	
No	5,017 (19.4 %)
Si	20,883 (80.6 %)

Variables cualitativas: n(%)

Variables cuantitativas: Mediana (RIQ)

En la figura 5-10 se analiza de manera gráfica como se están comportando las poblaciones generadas mediante imputación, en general los registros sintéticos mantienen los patrones de desarrollo expuestos en el análisis exploratorio anterior (figura 5-4).

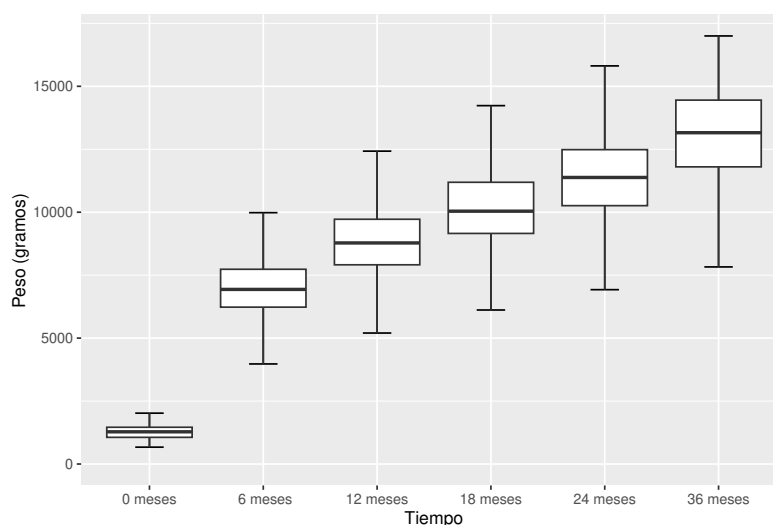


Figura 5-10: Diagrama de cajas de los pesos según el seguimiento

5.4. Modelo de regresión cuantílica

En la tabla 5-7 se muestran todos los parámetros globales de las 5 regresiones cuantílicas, agrupados por cuantil y por variables observadas, estos parámetros globales representan los gramos que aporta cada variable en el peso de los neonatos a lo largo de su desarrollo. Para todas las variables, excepto el peso al nacer, los errores estándar de las estimaciones de los parámetros son grandes.

También en la tabla 5-7 Se observa que los hombres tienen un peso mayor que las mujeres y la diferencia se hace mayor a medida que el cuantil aumenta. En cuanto al control prenatal, los pesos de los neonatos cuyas madres recibieron control prenatal son mayores. Caso contrario ocurre con la necesidad de ventilación asistida, los neonatos que la requirieron poseen pesos muy bajos. Por último, con respecto al tipo de parto, se observa que los neonatos que nacieron por parto vaginal presentan mayores pesos que los nacidos por parto espontáneo. Realizando una comparación entre cuantiles, se muestra que los neonatos, que nacieron con un muy bajo peso, tienen un crecimiento desacelerado y se ubican en los cuantiles inferiores de regresión.

Tabla 5-7: Parámetros por cada percentil de la regresión cuantílica

Variable	Regresión cuantílica									
	$\tau = 0,1$		$\tau = 0,25$		$\tau = 0,5$		$\tau = 0,75$		$\tau = 0,9$	
	β	Std.error	β	Std.error	β	Std.error	β	Std.error	β	Std.error
Intercepto	-835.19	1758.68	1156.59	1310.90	1831.33	1338.26	3057.75	1385.71	3421.09	1715.18
tiempo	40.28	35.16	84.79	27.92	108.93	27.56	146.02	28.70	174.35	36.97
log(tiempo)	2510.48	724.59	1830.54	542.39	1747.48	535.73	1387.14	558.14	1287.65	692.63
I(1/(tiempo))	726.83	1535.59	-1211.95	1120.71	-1901.90	1087.01	-3142.76	1162.75	-3712.88	1438.14
pesonacer	1.07	0.29	1.20	0.30	1.18	0.28	1.16	0.29	1.04	0.34
egnacido	-8.38	34.19	-16.30	33.22	-11.94	29.01	-8.55	24.94	6.56	35.95
sexoMasculino	123.19	184.77	151.57	146.82	153.70	146.75	241.36	188.37	294.23	234.38
hipertensionSi	-3.43	179.67	35.87	122.59	4.75	127.82	-16.59	110.35	-68.36	163.01
ventasisSi	-18.71	171.57	-30.79	145.79	-52.00	163.57	-103.08	220.32	-125.07	292.64
contprenatalSi	67.00	246.81	37.36	162.52	52.46	179.45	105.70	201.99	75.27	267.37
tipopartoEspontaneo	-372.47	15771.51	-216.82	2271.67	-208.80	270.83	-163.59	253.97	-92.85	74629.31
tipopartoVaginal	45.01	156.42	134.80	149.76	117.55	162.77	158.77	219.13	247.59	304.49

5.5. Test de Wald

Se usó el test de Wald para determinar cuáles variables predictoras de los modelos de regresión cuantílica ajustados son significativas para el modelamiento del peso con respecto al tiempo de los neonatos con bajo peso al nacer, los resultados fueron similares entre los modelos, pero distintos en cuanto los valores p. Las variables que resultaron ser significativas fueron el peso al nacer (β_4) y el tiempo ($\beta_1, \beta_2, \beta_3$), por lo tanto, estas no se evaluarán en el test de hipótesis.

Cuando se evaluó la significancia del resto de las variables predictoras incluidas en los modelos (edad gestacional del nacido, sexo, hipertensión, ventilación asistida, control prenatal, tipo de parto) los resultados obtenidos fueron los siguientes:

$$H_0 : \beta_{i,\tau} = 0 \quad \text{para todo } i \in (5, \dots, 11)$$

$$H_a : \beta_{i,\tau} \neq 0 \quad \text{para algun } i \in (5, \dots, 11)$$

Contrastando cada valor-p expuesto en la tabla 5-8 con un nivel de significancia de 0,05 No hay suficiente evidencia para rechazar la hipótesis nula y no se puede afirmar que las variables predictoras evaluadas (edad gestacional del nacido, sexo, hipertensión, ventilación asistida, control prenatal, tipo de parto) sean estadísticamente significativas para el modelo.

Tabla 5-8: Valores-p del test de Wald según cada τ de la regresión cuantílica

Valor-p	Cuantiles de la regresion				
	$\tau = 0,1$	$\tau = 0,25$	$\tau = 0,5$	$\tau = 0,75$	$\tau = 0,9$
	0,99	0,91	0,81	0,75	0,74

5.6. Curvas de crecimiento

En la figura 5-11 se muestran las curvas de peso para los neonatos con bajo peso al nacer, ajustadas por la variable sexo, se observa que se mantiene el patrón de la curva de peso a través del tiempo, además es un poco mayor en el caso de los hombres, en ambos casos el cambio más notorio ocurre entre los 0 y 6 meses.

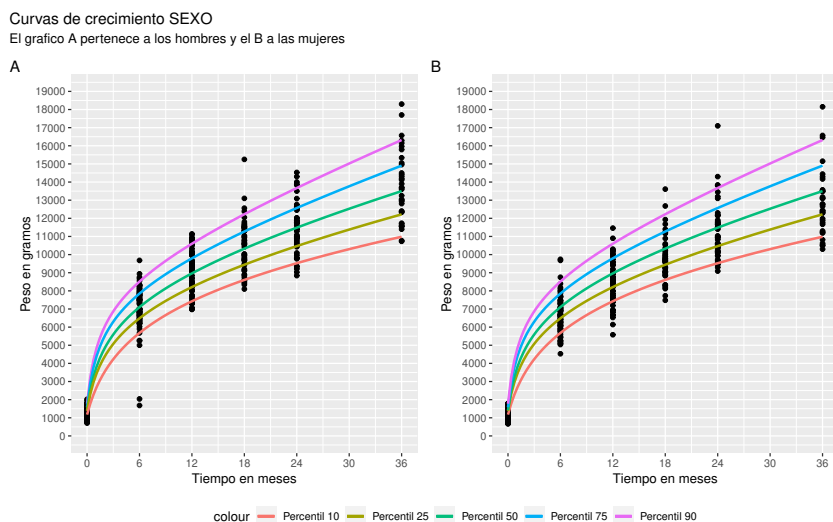


Figura 5-11: Curvas de peso para hombres y mujeres

En la figura 5-12, están expuestas las curvas de peso para los neonatos con madres que tienen o no antecedente de hipertensión, no se observan cambios notorios en cuanto a las curvas de peso a través del tiempo.

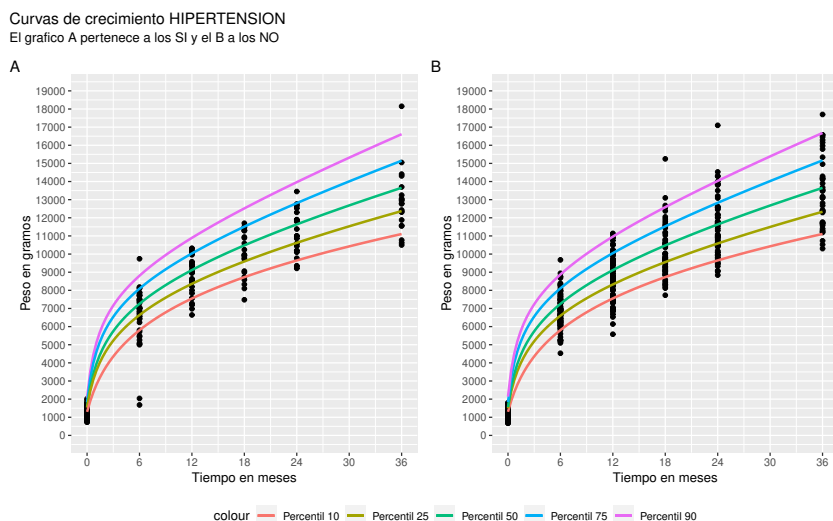


Figura 5-12: Curvas de peso para antecedente de hipertensión

En la figura 5-13 se presenta la curva de peso de los neonatos que necesitaron o no ventilación asistida respectivamente, se observa que cada una de las curvas presentadas se comportan de manera similar en ambo casos.

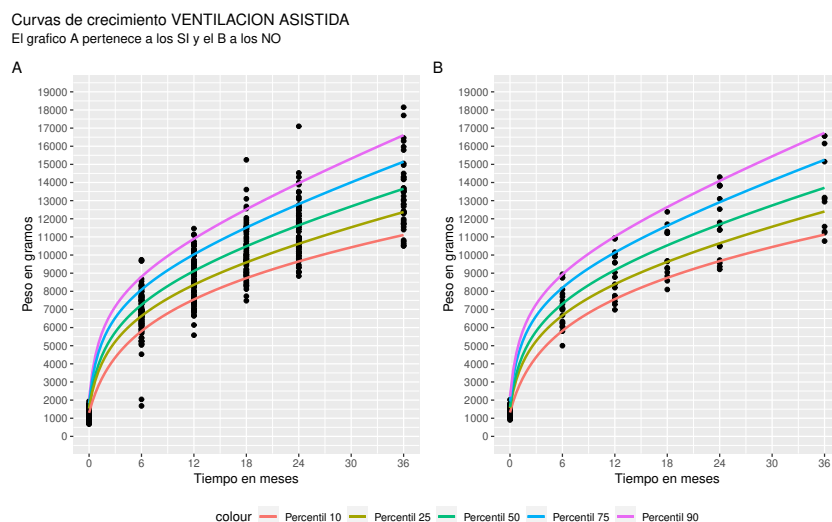


Figura 5-13: Curvas de peso para ventilación asistida

Para el caso de la variable control prenatal y sus curvas de peso a través del tiempo presentadas en la figura 5-14, se muestra que al igual que los pesos de los neonatos que fueron intervenidos con control prenatal, se comportan de manera similar a los que no fueron intervenidos con esta asistencia médica.

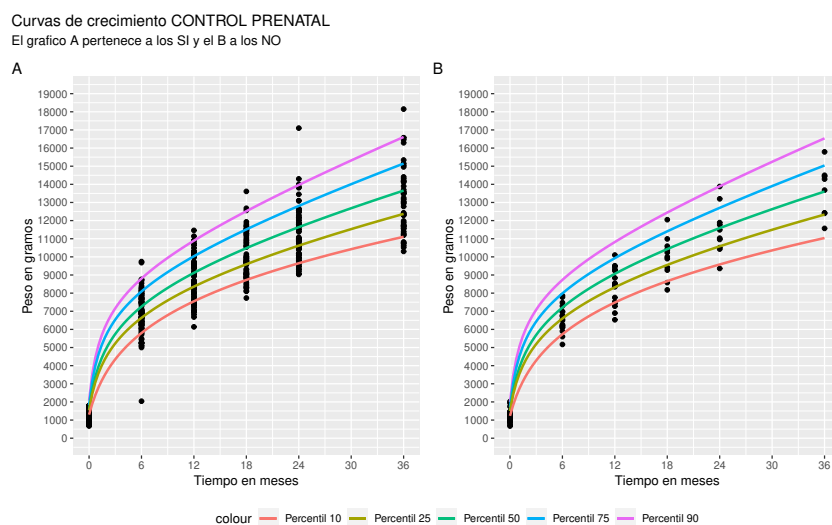


Figura 5-14: Curvas de peso para control prenatal

Con respecto a la figura 5-15 se puede observar leves variaciones en cuanto a las distintas curvas de peso, en donde los neonatos, que presentan un peso mayor, nacieron por parto vaginal. Se resalta también la poca cantidad de partos espontáneos que hubo con respecto a las otras dos categorías de la variable.

Curvas de crecimiento TIPO DE PARTO

El gráfico A pertenece a CESAREA, el B a ESPONTANEO Y el C a VAGINAL

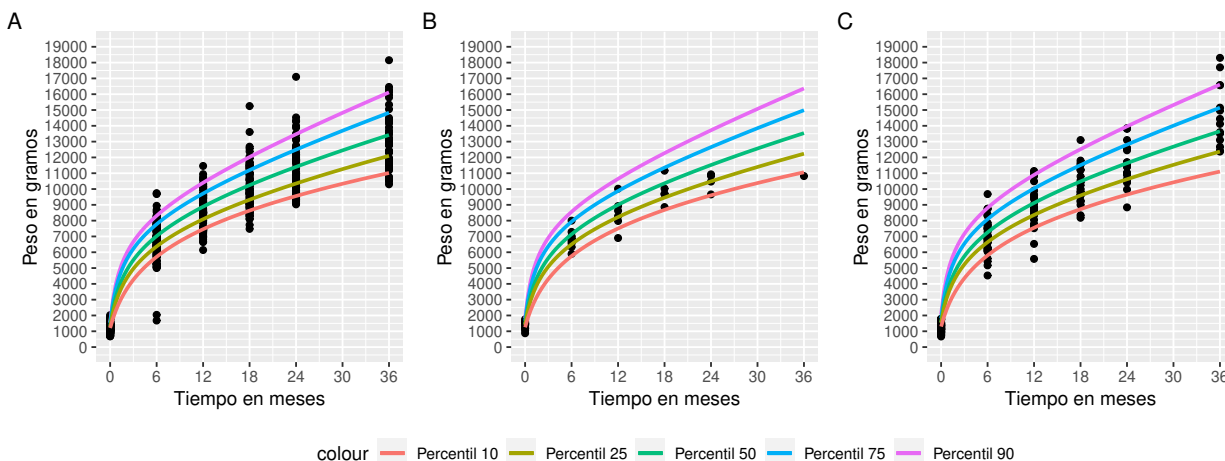


Figura 5-15: Curvas de peso para tipo de parto

6 Conclusiones y recomendaciones

6.1. Conclusiones

Los registros del PMC poseen una gran cantidad de valores faltantes para la variable longitudinal, en la que no se observaron el 49 % de los pesos a lo largo de los 3 años de seguimiento, esto genera un gran inconveniente a la hora de utilizar regresión por cuantiles, ya que este tipo de regresión requiere del mecanismo de datos faltantes más débil (MAR) para obtener estimaciones válidas (Koenker et al., 2006). Para solucionar este problema, se utilizó imputación múltiple por medio de la técnica especificación totalmente condicional (FCS), de esta forma, se tuvo en cuenta la incertidumbre debida a la información faltante, además de preservar la estructura, tendencias y relaciones existentes en los datos.

El modelo de regresión cuantílica Berkey-Reed (Johnson et al., 2013) de primer orden, describe de manera adecuada el comportamiento del crecimiento de los neonatos con BPN, donde al principio el desarrollo es rápido (0 a 6 meses) y luego desacelerado a medida que el tiempo transcurre. Con respecto a las covariables evaluadas, el peso al nacer generó un efecto significativo, en donde los bebés que nacieron con pesos no tan bajos (cuantiles superiores) ganan peso de manera más acelerada. Los hombres tienen un peso mayor a las mujeres, aumentando la diferencia a medida que el cuantil se acerca a valores altos. El control prenatal tiene un efecto positivo para todo τ , indicando que los pesos de los neonatos con BPN cuyas madres recibieron control prenatal son mayores. Caso contrario ocurre con la necesidad de ventilación asistida, los bebés que necesitaron ventilación asistida poseen pesos muy bajos. Por último, con respecto al tipo de parto, los neonatos que nacieron por parto vaginal presentan pesos mayores que los nacidos por parto espontáneo.

Dada la condición de ser neonatos prematuros y con bajo peso al nacer, estos pueden presentar un patrón de crecimiento distinto al común, por esta razón, las curvas de crecimiento construidas con regresión cuantílica, permiten realizar el seguimiento a los neonatos del PMC, ya que fueron construidas con datos de la región (condiciones socio demográficas similares) y describen de manera más precisa la posible trayectoria que puede desarrollar el peso de un neonato que padece BPN con respecto al cuantil en que se ubique.

6.2. Recomendaciones

Se recomienda incluir otras covariables que puedan ayudar a explicar y entender mejor el problema planteado, donde se tengan en cuenta otros factores que puedan explicar el peso de los neonatos nacidos con bajo peso, como la talla de los padres, estrato socioeconómico, entre otros; siendo estos factores referencias tomadas de las discusiones de los expertos en Bermudez et al. (2015). Teniendo en cuenta para la selección de estos factores influyentes, algunos antecedentes de estudios realizados en este campo de investigación y recomendaciones de expertos en el tema.

Otra recomendación será realizar evaluaciones periódicas al estudio de seguimientos de los neonatos, teniendo en cuenta que los tiempos establecidos de medición sean los correctos, asignando un grupo específico que realice visitas domiciliarias para las familias de los neonatos con el fin de observar las posibles razones de incumplimiento en los seguimientos del programa, incluyendo factores psico-sociales dentro de la familia. Esto influirá en la disminución de los registros faltantes, mejorando así los resultados del modelamiento y teniendo una caracterización mayor para esta población.

Bibliografía

- Alarcón, J., Alarcón, Y., Hëring, E., and Buccioni, R. (2010). Curvas antropométricas de recién nacidos chilenos. *Revista de la Sociedad Boliviana de Pediatría*, 49(1):58–65.
- Almira, A. G. (2008). Embarazo: diagnóstico, edad gestacional y fecha del parto. *Medisan*, 12(4).
- Anderson y Hay, W. (2001). Retardo del crecimiento intrauterino y el neonato pequeño para la edad gestacional, neonatología fisiopatología y manejo del recién nacido. *EUA: Panamericana*, pages 411–13.
- Arnau y Balluerka, N. (2004). Análisis de datos longitudinales y de curvas de crecimiento. enfoque clásico y propuestas actuales. *Psicothema*, 16(1):156–162.
- Ballard, J., Khoury, J., Wedig, K., Wang, L., Eilers-Walsman, B., and Lipp, R. (1991). New ballard score, expanded to include extremely premature infants. *The Journal of Pediatrics*, 119(3):417–423.
- Barnes y Hughes, A. T. W. (2002). A quantile regression analysis of the cross section of stock market returns.
- Batanero, C. (2000). Significado y comprensión de las medidas de posición central. *Uno. Revista de Didáctica de las Matemáticas*, 25:41–58.
- Bermudez, I. M., Bejarano, M. A., and Muñoz, J. T. (2015). Modelación de la evolución de neonatos con bajo peso al nacer, atendidos en el hospital universitario del valle, durante el período 2002 a 2010: Estudio de cohorte. *Archivos de Medicina (Manizales)*, 15(2):191–199.
- Cáceres-Manrique, F. d. M. (2009). El control prenatal: una reflexión urgente. *Revista Colombiana de Obstetricia y Ginecología*, 60(2):165–170.
- Calafati, R. O. (2017). Estrategias para el tratamiento de datos faltantes ("missing data") en estudios con datos longitudinales. *Universitat Oberta de Catalunya*.
- Cárdenas-López, C., Haua-Navarro, K., Suverza-Fernández, A., and Perichart-Perera, O. (2005). Mediciones antropométricas en el neonato. *Boletín médico del hospital infantil de México*, 62(3):214–224.

- Castaña-Díez, C., Álvarez-Castaño, L. S., Caicedo-Velásquez, B., Ruiz-Buitrago, I. C., and Valencia-Aguirre, S. (2020). Tendencia del bajo peso al nacer en recién nacidos a término y su relación con la pobreza y el desarrollo municipal en Colombia. 2000-2014. *Revista Chilena de Nutrición*, 47(1):22–30.
- Dagnino, J. et al. (2014). Datos faltantes (missing values). *Rev Chil Anest*, 43:332–4.
- Davino, C., Furno, M., and Vistocco, D. (2013). *Quantile regression: theory and applications*, volume 988. John Wiley & Sons.
- Efron y Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC Press.
- Estrada-Restrepo, A., Restrepo-Mesa, S. L., Fera, N. D. C. C., and Santander, F. M. (2016). Factores maternos relacionados con el peso al nacer de recién nacidos a término, Colombia, 2002-2011. *Cadernos de Saúde Pública*, 32.
- Fabres, J., Aspillaga, C., et al. (2010). Recomendación sobre curvas de crecimiento intrauterino. *Revista Chilena de Pediatría*, 81(3):264–274.
- Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied longitudinal analysis*. John Wiley & Sons.
- García, I. G., Conforme, G. M. G., Mesa, A. J. H., Cancino, I. I. T., García, I. G., and Mendoza, L. E. F. (2018). Factores de riesgo del bajo peso al nacer. policlínico universitario José Jacinto Milanés. 2013-2014. *Revista Médica Electrónica*, 40(1):89–98.
- Gómez-Gómez, M., Danglot-Banck, C., and Aceves-Gómez, M. (2012). Clasificación de los niños recién nacidos. *Revista Mexicana de Pediatría*, 79(1):32–39.
- Hao, L., Naiman, D. Q., and Naiman, D. Q. (2007). *Quantile regression*. Number 149. Sage.
- Johnson, W., Balakrishna, N., and Griffiths, P. L. (2013). Modeling physical growth using mixed effects models. *American Journal of Physical Anthropology*, 150(1):58–67.
- Koenker, Y., Pere, A., Koenker, R., and He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, 25(8):1369–1382.
- Koenker y Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156.
- Mancia, G., De Backer, G., Dominiczak, A., Cifkova, R., Fagard, R., Germano, G., Grassi, G., Heagerty, A. M., Kjeldsen, S. E., Laurent, S., et al. (2007). 2007 guidelines for the management of arterial hypertension: The task force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *European Heart Journal*, 28(12):1462–1536.

- Marín-Iranzo, R. (2006). Hipertensión arterial y embarazo. *Hipertensión*, 23(7):222–231.
- Medina y Galván, M. (2007). *Imputación de datos: teoría y práctica*. Cepal.
- Montesinos-Correa, H. (2014). Crecimiento y antropometría: aplicación clínica. *Acta pediátrica de México*, 35(2):159–165.
- Montoya Restrepo y Correa Morales, J. C. (2007). Curvas de peso al nacer. *Revista de Salud Pública*, 9:1–10.
- OMS, O. M. (2017). Metas mundiales de nutrición 2025: documento normativo sobre retraso del crecimiento.
- Rodríguez y Díaz, F. J. L. (2004). Estudios longitudinales: concepto y particularidades. *Revista española de salud pública*, 78(2):141–148.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.
- Rubin, D. B. (1988). An overview of multiple imputation. In *Proceedings of the survey research methods section of the American Statistical Association*, pages 79–84. Citeseer.
- Rubin y Little, R. J. (1987). *Statistical analysis with missing data*. John Wiley & Sons.
- Saha, U. (2023). Changes in the newborn at birth: Fetal-to-newborn transition. In *Clinical Anesthesia for the Newborn and the Neonate*, pages 29–47. Springer.
- Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, 57(1):19–35.
- Torres, J., Palencia, D., Sánchez, D. M., García, J., Rey, H., and Echandía, C. A. (2006). Programa madre canguro: primeros resultados de una cohorte de niños seguidos desde la unidad neonatal hasta la semana 40 de edad postconcepcional. *Colombia Médica*, 37(2):96–101.
- Unicef (2004). Low birthweight: country, regional and global estimates. *United Nations Children’s Fund and World Health Organization*, pages 1–31.
- Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC press.
- Van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., and Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12):1049–1064.
- Van Buuren, y. K. G.-O. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67.

- Velázquez Quintana, N. I., Masud Yunes Zárraga, J. L., and Ávila Reyes, R. (2004). Recién nacidos con bajo peso; causas, problemas y perspectivas a futuro. *Boletín Médico del Hospital Infantil de México*, 61(1):73–86.
- Verhasselt, A., Flórez, A. J., Van Keilegom, I., and Molenberghs, G. (2019). The impact of incomplete data on quantile regression for longitudinal data. *FEB Research Report KBL1906*.
- Vicèns, J. and Sánchez, B. (2012). Regresión cuantílica: estimación y contrastes. *Madrid, España: Universidad Autónoma de Madrid*, (21).
- Vintzileos, A. M., Ananth, C. V., Smulian, J. C., Scorza, W. E., and Knuppel, R. A. (2002). The impact of prenatal care on neonatal deaths in the presence and absence of antenatal high-risk conditions. *American Journal of Obstetrics and Gynecology*, 186(5):1011–1016.