



MODELO DE REGRESIÓN NO PARAMÉTRICA PARA EL CRECIMIENTO DE NEONATOS CON SITUACIÓN DE BAJO PESO

Juliette Barboza Candelo
Yesica Bejarano Lozano

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Cali, Colombia
2018

MODELO DE REGRESIÓN NO PARAMÉTRICA PARA EL CRECIMIENTO DE NEONATOS CON SITUACIÓN DE BAJO PESO

Juliette Barboza Candelo
Yesica Bejarano Lozano

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

Director(a):
Javier Olaya Ochoa

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2018

Dedicatoria

A Dios principalmente por darme todo lo necesario para llegar hasta aquí, desde mi familia hasta la voluntad y la determinación para lograr lo que me he propuesto. A mis padres por su amor, dedicación, entrega total, que han dado todo para hacer de este proceso el mejor y darme la fortaleza para continuar, a mis hermanos que alegran mis días y son las personitas que impulsan mis ganas de salir adelante. A mi compañera y amiga en todo este proceso de formación Yesica, eres uno de los regalos que me deja la Universidad, gracias por compartir tantos momentos y hacer parte de este camino. A mis amigas que hicieron parte de todo el proceso y no sólo académico, gracias por cada cosa, ustedes son un regalo de Dios.

Juliette Barboza Candelo

Doy muchas gracias principalmente a Dios por darme de su sabiduría y su fortaleza para culminar esta etapa de mi vida. Gracias infinitas a mis padres y a mi hermanito por estar conmigo en todo momento, por enseñarme que no debo rendirme, que debo luchar por mis sueños, por ser mi apoyo incondicional y demostrarme día tras día su amor, éste triunfo es de ustedes. Muchas gracias a mi compañera y amiga Juliette, conocerte y tener el privilegio de trabajar contigo fue y será siempre una gran bendición de Dios para mi vida. A mi grupo de amigas y compañeros de universidad por ser parte de este maravilloso proceso. “Esfuerzate y sé muy valiente”. Josué 1:9

Yesica Bejarano Lozano

Agradecimientos

Agradecemos a nuestra Alma Mater, la Universidad del Valle por abrirnos sus puertas, por darnos la oportunidad de crecer personal y profesionalmente.

A la Escuela de Estadística y a nuestros profesores por brindarnos las herramientas para desarrollar todos nuestros conocimientos.

Al profesor Javier Olaya, nuestro director, por todo el apoyo en este trabajo de grado, y durante el transcurso de nuestra carrera, por su tiempo y dedicación.

Al Hospital Universitario del Valle (HUV) y a los Doctores Javier Torres, Carlos Jiménez y Diana Palencia, docentes del Departamento de Pediatría de la Universidad del Valle por suministrar la base de datos para este proyecto.

A nuestros compañeros por hacer de este escarpado camino una gran experiencia.

Resumen

En este trabajo se proponen las curvas para el peso de los 145 neonatos nacidos en el Hospital Universitario del Valle de la ciudad de Santiago de Cali entre el periodo 2002-2010 que hicieron parte del plan madre canguro, la metodología empleada es la Regresión No Paramétrica con B-splines, también se proponen otras metodologías como la regresión cuantílica debido a la eficacia de esta regresión debido a la importancia de calcular las curvas para el percentil 5 y 95. Por otra parte, el querer tener información más a fin con la realidad se trabaja con la edad cronológica obteniendo mejores resultados con la metodología empleada.

Palabras clave: peso al nacer, neonato, regresión no paramétrica, B-splines, datos funcionales, regresión cuantílica.

Abstract

In this work we propose the curves for the weight of the 145 neonates born in the University Hospital of the Valley of the city of Santiago de Cali between the period 2002-2010 that were part of the kangaroo mother plan, the methodology used is the Non Parametric Regression with B-splines, as well as other methodologies such as quantile regression are proposed due to the effectiveness of this regression, such as the importance of calculating the curves for the 95th and 5th percentiles. On the other hand, wanting to have information more related to reality We work with the chronological age obtaining better results with the methodology used.

Keywords: birth weight, neonate, non parametric regression, B-splines, functional data, quantile regression.

Contenido

Agradecimientos	IV
Resumen	V
Lista de figuras	VII
Lista de tablas	1
1 Introducción	2
1.1 Planteamiento del problema	2
1.2 Justificación	3
1.3 Objetivos	4
1.3.1 Objetivo General	4
1.3.2 Objetivos Específicos	4
1.4 Antecedentes	5
2 Materiales y Métodos	10
2.1 Marco Conceptual	10
2.1.1 Neonato	10
2.1.2 Peso al nacer	10
2.1.3 Edad Gestacional	10
2.1.4 Edad Gestacional de Ballard	11
2.1.5 Curvas de Crecimiento	13
2.1.6 Control Prenatal	13
2.1.7 Hipertensión Arterial (HA)	14
2.2 Marco Teórico	15
2.2.1 Regresión No Paramétrica	15
2.2.2 Cuantil	17
2.2.3 Regresión cuantílica	18
2.2.4 Datos Funcionales	19
2.2.5 Mínimos cuadrados ponderados	22
2.2.6 Test de Bartlett	22
2.2.7 Test de Levene	23

2.3	Metodología	24
2.3.1	Caso de estudio	24
2.3.2	Análisis Exploratorio de Datos	24
2.3.3	Ajuste No Paramétrico (NP) del individuo promedio con B-Splines .	25
2.3.4	Estimación de los percentiles	27
2.3.5	Modificación de la variable Tiempo	27
3	Resultados	30
3.1	Análisis Exploratorio de Datos	30
3.2	Ajuste No Paramétrico (NP) del individuo promedio con B-Splines	35
3.3	Datos Funcionales	40
3.4	Estimación de los percentiles	43
3.5	Modificación de la variable Tiempo	45
3.6	Mínimos cuadrados poderados	48
3.7	Covarianza y correlación	49
4	Conclusiones y Recomendaciones	51
4.1	Conclusiones	51
4.2	Recomendaciones	52
	Bibliografía	53

Lista de Figuras

3-1. Distribución del peso al nacer.	30
3-2. Peso al nacer según sexo	31
3-3. Comportamiento del peso al nacer según la EG	32
3-4. Distribución de la Morbilidad materna respecto al PN	34
3-5. Curvas de crecimiento de los RN	35
3-6. Curvas percentílicas para el BPN	35
3-7. Suavización B-Spline	36
3-8. Residuales Vs Valores ajustados	37
3-9. Cuantiles teóricos vs cuantiles empíricos para 145 niños	38
3-10.Cuantiles teóricos vs cuantiles empíricos para 35 niños	38
3-11.Histograma de residuales para 145 niños	39
3-12.Histograma de residuales para 35 niños	39
3-13.Bases B-spline	40
3-14.Curvas Funcionales	41
3-15.Curvas suavizadas	42
3-16.Media Funcional	42
3-17.Estimación para el percentil 95	43
3-18.Estimación para el percentil 5	43
3-19.Regresión Cuantílica	44
3-20.Estimación B-spline en días para los 145 RN	45
3-21.Estimación B-spline en días para los 35 RN	45
3-22.Estimación B-spline para los percentiles con la edad en días	47
3-23.Regresión Cuantílica para edad cronológica	48
3-24.Suavización B-Spline con Mínimos Cuadrados Ponderados	49

Lista de Tablas

2-1. Variables de estudio	24
3-1. Distribución por clasificación del peso al nacer	31
3-2. Estadísticas descriptivas para el peso al nacer según el sexo	31
3-3. Estadísticas Descriptivas para el peso según la EG	32
3-4. Análisis de variables clínicas de la madre con respecto al PN	33
3-5. Matriz de covarianzas	50
3-6. Matriz de Correlación	50

1 Introducción

1.1. Planteamiento del problema

El bajo peso al nacer (BPN), definido por la Organización Mundial de la Salud (OMS 2017) y por la Clasificación internacional de enfermedades como un peso al nacer inferior a 2500 gramos, corresponde a uno de los problemas más comunes y es considerado como una condición multicausal que constituye el principal factor de morbilidad y la segunda causa de mortalidad perinatal, convirtiéndose en un problema de salud pública (López et al., 2012). Según UNICEF (2004) los infantes que pesen menos de 2500 gramos tienen aproximadamente 20 veces más probabilidades de morir que los bebés más pesados, y los que sobreviven tienen disminuidas las funciones del sistema inmunológico corriendo mayor riesgo de padecer una enfermedad posteriormente.

Son diversas las causas asociadas a este flagelo como la edad gestacional (parto prematuro o restricción del crecimiento intrauterino), el propio crecimiento del feto de la madre y su dieta antes/durante el embarazo, su estilo de vida (fumar, alcohol, entre otras) o complicaciones como enfermedades que pueden afectar el crecimiento y desarrollo del feto. Como también una causa relevante es el nivel socioeconómico, pues madres en condiciones socioeconómicas desfavorables con frecuencia tienen infantes con condición de bajo peso (UNICEF, 2004).

En el informe de Unicef et al. (2016) se menciona que alrededor de 1 millón de niños en 2015 su primer día de vida fue también el último, pues las muertes neonatales representaron el 45 % del total de muertes con un aumento proporcional del 5 % desde el año 2000. Además, los niños que viven en estados frágiles y en países afectados por conflictos enfrentan un alto riesgo de muerte, pues de los 20 países que presentan los índices de mortalidad infantil más elevados, 10 figuran en la lista del Banco Mundial relativa a los contextos frágiles.

La mejor forma de establecer el correcto desarrollo de un niño, es a través de la evaluación del crecimiento (Castaño, 2015) que puede definirse como: “movimiento de la materia viva que se desplaza en el tiempo y en el espacio”. El fenómeno del crecimiento es un proceso dinámico, que refleja el estado psicosocial, económico, nutricional, cultural, ambiental y de homeostasis orgánica en el que se desarrolla un individuo (Serrano, 2002). Por ello, es importante tener herramientas para su adecuada valoración en la práctica clínica y en la población (Briceño et al., 2012).

Siguiendo el informe de Progreso para la Infancia 2007 de la UNICEF, aproximadamente al 60 % de los recién nacidos en los países en desarrollo no se les realiza medición de peso y se estima que la proporción de BPN en estos países es del 16 % (cerca de 19 millones de niños) (Pinzón Villate et al., 2013); en Colombia, por ejemplo, según el DANE (2009), tanto el departamento de Antioquia como el de Chocó tienen una tasa de BPN del 9,2 %; sin embargo, la tasa de mortalidad neonatal temprana es de 7,1 por cada 1000 recién nacidos vivos en Antioquia y de 17,4 por cada 1000 recién nacidos vivos para el Chocó. Al mismo tiempo, la proporción de la población en condiciones de miseria corresponde a 31 % en el Chocó y a 8,4 % en Antioquia. Por lo tanto, es posible considerar la existencia del subregistro en varias de las regiones del país, especialmente en las más marginadas (Gaitán and Delgado, 2016). Por esta razón se dice que no existen datos que permitan realizar comparaciones válidas, es decir que los datos disponibles no representan la población en general. Tanto así que la UNICEF recomienda tener cuidado al comparar los datos entre países, regiones y periodos de tiempo.

Es así, como estudiar las curvas sobre BPN es de tanta importancia por el componente social que involucra, y como lo menciona Castaño 2015 las curvas existentes para evaluar el desarrollo físico de los neonatos están construidas con base en referentes internacionales que no clasifican de manera adecuada los recién nacidos en esta región.

Ahora, se quiere dar solución a este problema ajustando las curvas de crecimiento de los recién nacidos con bajo peso al nacer atendidos en el Hospital Universitario del Valle, durante el período 2002 a 2010 en la Ciudad de Cali por medio de los modelos de regresión no paramétricos combinados con datos funcionales y regresión cuantílica, ya que la naturaleza de los datos se ajusta muy bien a estas metodologías. En nuestro caso no tenemos pares únicos (x_i, y_i) , es decir, se cuenta con más de una respuesta para cada valor de X por lo que cada niño es considerado como una observación funcional, por otro lado también se tiene la necesidad de calcular las curvas para los niños más y menos pesados (percentil 5 y 95) donde se emplea la regresión cuantílica.

1.2. Justificación

Las curvas de crecimiento existentes para evaluar el desarrollo físico de los neonatos no clasifican adecuadamente los recién nacidos debido a que se encuentran realizadas con base en referentes internacionales (Castaño, 2015), es por esto que se hace necesario proponerlas a través de nuevas metodologías.

Además las curvas de crecimiento son una herramienta para observar la evolución de los neonatos con respecto a su peso o talla, que a su vez son una advertencia oportuna para evitar un futuro problema de salud. De ahí se deriva la importancia del Bajo peso al nacer

como lo menciona Faneite et al. (2011) el BPN no es sólo un indicador de la salud de la madre biológica y el estado nutricional, sino también de las posibilidades de supervivencia de un recién nacido, su crecimiento, salud a largo plazo y desarrollo psicosocial; donde con salud a largo plazo se refiere a que en algunos adultos se desarrollan enfermedades asociadas a patologías cardiovasculares, hipertensión arterial y diabetes.

Según el Observatorio Así Vamos (2016) en el país, el porcentaje de niños que nacen con bajo peso ha venido en aumento durante los últimos años. En 2000, la prevalencia de bajo peso al nacer era de 7,33 por cada 100 nacidos vivos, mientras que en 2011 fue de 9,01 %. A partir del 2012 este indicador presentó un leve descenso, con una prevalencia de 8,81 % mientras que en el 2016 fue 9,07 %. Por departamentos hay diferencia ya que Caquetá, Meta, Vichada, Putumayo y Vaupés presentaron las menores prevalencias de bajo peso al nacer, mientras que Bogotá, D.C., Cundinamarca y Chocó fueron las regiones con porcentajes más altos que el total nacional.

Siendo así como este flagelo se ha convertido en un problema de salud pública y por esta razón es importante la construcción de dichas curvas por medio de herramientas estadísticas y poder brindar una ayuda a la comunidad en una adecuada clasificación de los neonatos con bajo peso al nacer.

Por otra parte la utilización de la regresión no paramétrica se justifica por las ventajas que tiene esta regresión comparada con las de la regresión paramétrica, siendo el caso de desconocer la función de regresión f subyacente, como también que en el enfoque no paramétrico no es sensible al cumplimiento de los supuestos evitando la elección de un modelo incorrecto, adicionalmente algo que va muy acorde a nuestros objetivos es que en el enfoque no paramétrico no se cuenta con una forma formulada del modelo, por lo que sólo se basa en la estimación de las curvas gráficamente.

1.3. Objetivos

1.3.1. Objetivo General

Obtener las curvas de crecimiento para los recién nacidos con bajo peso pertenecientes al plan madre canguro del Hospital Universitario del Valle durante el período 2002 a 2010, a través de modelos de regresión no paramétrica.

1.3.2. Objetivos Específicos

- Comparar el ajuste por Regresión No Paramétrica con la estimación de la media funcional

- Estimar la media de las curvas de crecimiento de los niños con información completa, usando análisis de datos funcionales.
- Proponer la estimación del peso de los neonatos en los percentiles 5 y 95 y comparar con el ajuste por regresión cuantílica.

1.4. Antecedentes

A continuación se presentan estudios anteriores que están relacionados con la temática y teoría que se empleará en el trabajo.

Modelación del crecimiento en neonatos con situación de Bajo Peso, nacidos en el Hospital Universitario del Valle. Periodo 2002-2009

En este trabajo, Castaño (2015) construyó las curvas para los neonatos con bajo peso nacidos en el HUV de la ciudad de Cali durante el periodo 2002-2009, donde la base de datos estaba conformada por 160 neonatos.

La metodología empleada comprende el análisis del crecimiento en peso y talla de los recién nacidos con BPN, donde en la primera etapa se realizó un análisis exploratorio para el peso y la talla contra factores influyentes, con el objetivo de conocer y caracterizar el comportamiento de las variables incluidas en el estudio, como también fue útil para validar la información contenida en la base de datos, pues se eliminaron 2 registros que presentaron inconsistencias, por lo tanto finalmente quedaron 158 registros. También se ajustaron curvas de crecimiento percentílicas para el peso y la talla de manera exploratoria con la intención de obtener una visión del comportamiento del crecimiento ante diferentes percentiles, por esta razón se conformó grupos de tiempo de 14 semanas, para un total de 12 grupos y se calculó los percentiles 2,10,25,50,75,90 y 97 %. Al graficar estas curvas se encontró picos debido a la presencia de puntos atípicos por lo que se llevó a cabo un método de suavizado para las curvas con el fin de eliminar las fluctuaciones y hacerlas semejantes a las curvas de referencia propuestas por la OMS.

En la segunda etapa del trabajo se realizó el ajuste del modelo por medio de la regresión cuantílica (QR) en lo cual se centra el objetivo principal, dicho ajuste se realizó con QR paramétrica con la variable edad transformada y no transformada, pero se encontró que dicha alternativa no representa adecuadamente la dinámica del crecimiento en los individuos del estudio, por lo que se realizó el ajuste del modelo por medio de la QR no paramétrica seleccionando el modelo que mejor se ajustara a los datos bajo el criterio SIC y por último selección de variables y validación los supuestos por medio del análisis gráfico de los residua-

les, sin embargo en la QR no se tiene ningún supuesto sobre los errores del modelo por lo que puede ajustarse perfectamente a casi cualquier tipo de datos.

En la tercera etapa se realizó una comparación con estudios anteriores, donde se resalta que la regresión cuantílica resultó más efectiva, ya que utiliza mayor información de los individuos, lo que permite encontrar mayores efectos sobre el peso de los neonatos y permite modelar el crecimiento para diferentes cuantiles y no sólo para la media. Ya al comparar con las curvas de referencia de la OMS se encontró que los neonatos con BPN tardan un año en recuperar su peso “ideal”, sin embargo siempre se mantienen por debajo de los estándares de la OMS, y para la talla tardan más tiempo en recuperarse pero luego mantienen su talla semejante a las referencias internacionales. Haciéndose evidente que no es posible comparar el desarrollo físico de los neonatos con BPN con los nacidos con peso normal, de ahí la importancia y necesidad de crear curvas de referencia para la población local para los recién nacidos con bajo peso.

Quantile Regression methods for reference growth charts

En este artículo publicado por Wei et al. (2006) , se ajustó las curvas de crecimiento mediante regresión cuantílica para la altura y el peso de 2514 niños finlandeses, de estos registros se eliminaron una pequeña proporción de niños con bajo peso al nacer o sin información, gemelos u otros registros sospechosos, de los cuales quedaron 1143 niños y 1162 niñas todos a término y saludables, y para cada niño se tuvieron entre 3 y 44 mediciones, los datos se recogieron de los centros de salud y escuelas de dos cohortes 1096 niños nacidos entre 1954 y 1962, el otro de 1209 niños nacidos entre 1968 y 1972. La metodología consistió en comparar los métodos Least-Mean-Square algorithm (LMS) y QR para la estimación de las curvas de crecimiento no condicional, las cuales dependen exclusivamente de la edad y cuyos resultados en la comparación no presentaron muchas diferencias.

Para el otro caso, la estimación de las curvas de crecimiento condicional o curvas de crecimiento longitudinales, que tienen en cuenta la historia previa del crecimiento y posiblemente otras covariables, al comparar los métodos QR y LMS fueron bastante similares en casi todas las edades.

Finalmente, en el artículo Wei et al. 2006 concluye que para la estimación no condicional los dos enfoques son complementarios, pues LMS impone más estructura y puede ser más estable particularmente en regiones donde los datos son escasos, y el método QR es más flexible y por lo tanto capaz de revelar salidas de los supuestos subyacentes de modelos paramétricos. Y ya para las estimaciones condicionales según la edad, basadas en los métodos QR y LMS fueron bastante similares como se mencionó anteriormente, pero las estimaciones QR de densidades condicionales específicas de la edad sugirieron una forma claramente bimodal para 1 año de

edad. Sin embargo, se resaltó una ventaja del método de regresión QR sobre el LMS, y se refiere a su flexibilidad que permiten exponer características importantes de los datos que no serían fácilmente observables con métodos paramétricos convencionales.

Curvas de peso al nacer

En este estudio, Montoya-Restrepo and Correa-Morales 2007 planteó las curvas de los percentiles relacionando el peso al nacer con la edad gestacional, estas curvas fueron elaboradas utilizando los pesos de los recién nacidos afiliados al régimen de la empresa promotora de salud Susalud (1999-2005), en cuya base de datos se incluyeron los pesos de los neonatos producto de embarazos que finalizaron entre las 22 y 44 semanas de gestación.

Se contó con 58406 nacimientos ocurridos en la red de instituciones encargadas de atender los partos en ciudades como Cali, Medellín, Bogotá y Barranquilla, sin embargo se realizó una depuración a la base de datos donde se excluyeron registros en los cuales el peso fuera menor a 500g, aquellos en los que no se conocía la edad gestacional, fallecidos, registros de gestantes con problemas de salud, malformaciones, entre otros, quedando así un total de 54044 registros, con los cuales se construyeron las curvas porcentuales para el peso de los neonatos a los niveles 5, 10, 25, 50, 75, 90 y 95 % con sus respectivas tablas tanto a nivel general como diferenciado por género.

Así como lo recomienda la OMS, que el crecimiento intrauterino de un recién nacido perteneciente a una población específica, sea valorado por comparación con estándares obtenidos de la misma población, los autores resaltan que las curvas propuestas por ellos a pesar de que puede ser mejoradas con la inclusión de otras variables que puedan afectar el crecimiento fetal, las consideran una gran herramienta para la clasificación de los recién nacidos en nuestro país, ya que los datos provienen de una población cuyas características demográficas son similares a las de la población colombiana.

Modelación de la evolución de neonatos con BPN, atendidos en el HUV, durante el periodo 2002 a 2010: Estudio de cohorte

En la investigación realizada por Bermúdez et al. (2015) se propuso un modelo de coeficientes aleatorios que permitiera estimar la curva de evolución del peso de cada neonato que nace con bajo peso, se incluyeron solo aquellos que presentaron observaciones en los tiempos 0, 12, 18, 24, 36 meses, razón por la cual se redujo el número de neonatos inicial de 159 a 145 que cumplían con lo requerido. Los datos fueron recolectados a partir de las historias clínicas de neonatos prematuros nacidos en la unidad de recién nacidos del Hospital Universitario del

Valle (HUV) de la ciudad de Santiago de Cali, entre febrero de 2002 y noviembre de 2010, pertenecientes al Programa de Seguimiento de Alto Riesgo y Canguro.

Por medio del análisis multivariado se logró identificar algunas relaciones que posteriormente dieron pie a los investigadores para dividir la población objeto de estudio en dos grupos, los que presentaban Bajo Peso Extremo (BPE) con alguna intervención, y los de Bajo Peso (BP) y Muy Bajo Peso (MBP) que no presentan ninguna de estas, entendiéndose intervención como recibir transfusiones, alimentación parental, o que sus madres presenten morbilidad materna.

En el ajuste del modelo se evaluaron las diferentes covariables como Ventilación, Transfusiones, Requiere Oxígeno, Alimentación parenteral, Control Prenatal y Morbilidad Materna, siendo la variable Control prenatal la única que representó un aporte significativo al modelo en la explicación de la variación del peso con respecto al peso medio general en los neonatos del HUV. Así, muestran el modelo utilizado para las estimaciones de las curvas, el cual se refiere al peso estimado de un neonato prematuro que nazca con peso inferior de 2001 gramos en el HUV para un tiempo, incorporando en el modelo si la madre presentó o no Control Prenatal (CP), especificado de la siguiente manera:

$$Peso_{ti} = \gamma_{00} + \gamma_{01}CP_i + \gamma_{10}Tiempo_{ti} + \gamma_{20}Tiempo_{ti}^2 + (u_{0i} + u_{1i}Tiempo_{ti} + u_{2i}Tiempo_{ti}^2) + e_{ti} \quad (1-1)$$

Para

$t = 1, 2, 3, 4, 5$ (Tiempos de medición 0, 12, 18, 24 y 36 meses)

$i = 1, 2, \dots, 145$ (Representa al neonato)

Donde:

γ_{00} Parámetro de regresión del intercepto.

γ_{01} Parámetro de regresión de la variable control prenatal

γ_{10} Parámetro de regresión de la variable tiempo

γ_{20} Parámetro de regresión de la variable tiempo al cuadrado

u_{0i} Intercepto aleatorio relacionado con el neonato i -ésimo

u_{1i} Efecto aleatorio lineal relacionado al mes o tiempo i -ésimo

u_{2i} Efecto aleatorio relacionado al tiempo al cuadrado

e_{ti} Error en la medición tomada en el neonato i -ésimo, en el tiempo t -ésimo

Finalmente, los investigadores recomiendan modelar el BPN mediante un modelo de coeficientes aleatorios, donde se cuenten con más factores influyentes según la teoría, como la talla

de los padres, estrato socioeconómico, entre otras y con más observaciones y además, obtener modelos que permitan evaluar las variaciones tanto para la talla y el perímetro cefálico de los neonatos, haciendo uso de la metodología propuesta.

Curvas antropométricas de recién nacidos Chilenos

En esta investigación el objetivo de Alarcón et al. (2008) se centra en diseñar las curvas antropométricas locales para recién nacidos Chilenos entre 24 y 42 semanas de edad gestacional, se analizó una población de 148.395 recién nacidos del Servicio de Salud Metropolitano Oriente, de las maternidades del Hospital del Salvador entre los años 1998-2000 y Hospital Luis Tisné entre los años 2001-2005. En la base de datos se registraron antecedentes obstétricos como Fecha de la última regla, morbilidad materna; y datos neonatales como el peso, talla, perímetro cefálico, edad gestacional pediátrica, sexo, gemelaridad y patología perinatal, sin embargo en el estudio fueron excluidos todos aquellos recién nacidos hijos de madre con patología asociadas. Del total de 148.395 recién nacidos registrados 86.575 cumplieron con los criterios de selección para el análisis del peso, talla e índice ponderal.

Se clasificó la población según la edad gestacional y sexo en tablas de peso, talla perímetro cefálico y índice ponderal, se calculó promedios, desviación estándar y la distribución de los percentiles 10, 50 y 90 ajustados a un polinomio de tercer orden.

En la discusión los autores mencionan que en cuanto a sexos resulta práctica la utilización de una curva única, ya que las diferencias sólo son significativas después de las 35 semanas de edad gestacional. También realizan comparación con estudios anteriores para la misma región y encuentran similitudes, excepto con uno en especial que por ser un estudio poblacional no excluye a los recién nacidos con patologías asociadas que afectan las curvas de crecimiento intrauterino.

Concluyen que las curvas obtenidas representan una información real y actualizada de la antropometría de los recién nacidos chilenos por tener un número de casos importante, lo que las hace confiables para su uso perinatólogo y recomiendan el uso clínico de las mismas como estándar referencial.

2 Materiales y Métodos

En éste capítulo se introducen algunos conceptos que serán de gran utilidad para entender la problemática que estamos trabajando y también para tener mayor claridad al momento de interpretar de manera adecuada los resultados obtenidos.

2.1. Marco Conceptual

2.1.1. Neonato

El neonato, también conocido como recién nacido, es aquel que se encuentra en la etapa que comienza el día del nacimiento hasta los 28 días completos después de su nacimiento (de la Salud, 1995). Es un tiempo en el que los cambios son muy rápidos e importantes para el resto de la vida del recién nacido (Castaño, 2015).

2.1.2. Peso al nacer

El Peso al nacer es el primer peso del feto o del recién nacido después del parto. En el caso de los nacidos vivos, el peso al nacer debería medirse en la primera hora de vida antes de que se produzca una pérdida de peso posnatal significativa y el peso real debería registrarse según el grado de precisión con el que se mide (Organización Mundial de la Salud, 2005). El bajo peso puede subdividirse en peso bajo al nacer (menos de 2500 g), muy bajo peso al nacer (menos de 1500 g) y peso extremadamente bajo (menos de 1000 g). Un bebé de bajo peso puede indicar que es demasiado pequeño, que nació antes de tiempo (premature) o ambas cosas. Esto puede deberse a muchas causas; por ejemplo, problemas de salud de la mamá, factores genéticos, problemas con la placenta o que la madre haya abusado de sustancias durante el embarazo (Biblioteca Nacional de Medicina, 2016).

2.1.3. Edad Gestacional

La edad gestacional (EG) también conocida como edad fetal o gestación, es el periodo de tiempo entre la concepción del feto hasta su nacimiento, en este tiempo el bebé crece y se desarrolla en el útero de la madre. La edad gestacional es el tiempo medido en semanas desde el primer día del último ciclo menstrual de la mujer hasta la fecha actual o también conocido como Fecha de Última Regla (FUR) (DAM, 2016).

Antes del nacimiento, el médico empleará una ecografía para medir el tamaño de la cabeza, el abdomen y el fémur del bebé. Esto ayuda a determinar qué tan bien está creciendo el bebé dentro del útero. Después del nacimiento, la edad gestacional se puede medir observando el peso, la talla, el perímetro cefálico, el estado de la piel y del cabello, los reflejos, el tono muscular, la postura y los signos vitales del bebé (Animated Dissection of Anatomy for Medicine, 2011).

Según Vaca Romero (2013), la EG se clasifica según el tiempo medido en semanas de la siguiente manera :

- **Término temprano:** Son los nacidos antes de las 37 semanas de gestación. Los bebés nacidos a término temprano pertenecen a dos categorías: **Pretérmino (o prematuro):** que son aquellos que nacen antes de las 37 semanas de gestación y pueden o no tener un tamaño adecuado para su edad gestacional y **Pequeño para la edad gestacional:** que son los RN que pueden ser prematuros o no serlo, pero pesan menos que el 90 % de los bebés de su edad gestacional. Su tamaño pequeño generalmente es consecuencia de una inadecuada nutrición prenatal, lo cual disminuye el crecimiento fetal (PulsoMed, 2016).
- **Término completo:** Producto de la concepción de 37 semanas a 41 semanas de gestación.
- **Post-término:** El bebé nace después de las 42 semanas.

La edad gestacional puede calcularse con la FUR, a lo largo del tiempo se han propuesto distintos métodos para realizar la evaluación con la mayor precisión posible, basados en los signos físicos y neurológicos de madurez del recién nacido, entre ellos se encuentran las clasificaciones de Usher, Parkin, Dubowitz, Capurro y Ballard (Vaca Romero, 2013).

2.1.4. Edad Gestacional de Ballard

En neonatología, la prueba de Ballard (New Ballard Score, NBS) es un método clínico utilizado frecuentemente para datar la edad gestacional (EG) del recién nacido, fundamentado en los cambios intrauterinos por los que pasa el feto durante su maduración y desarrollo. Esta prueba consiste en la observación de una serie de características físicas y neurológicas a partir de las cuales se asigna una puntuación determinada, donde la suma de esta puntuación es extrapolada para inferir una EG del neonato (Gabriel et al., 2006),

De acuerdo con Nacho Rios 2012, los criterios físicos evaluados mediante el Ballard incluyen:

- **La piel:** Consiste en la maduración de la piel fetal evaluado por las estructuras intrínsecas de la piel con la pérdida gradual de la vernix caseosa.
- **Orejas / ojos:** El pabellón de la oreja cambia su configuración aumentando su contenido cartilaginosos a medida que progresa su madurez.
- **Cabello:** En la inmadurez extrema, la piel carece de las finas vellosidades que la caracterizan, las cuales aparecen aproximadamente durante las semanas 24 o 25.
- **Superficie plantar:** Relacionado con las grietas en la planta del pie.
- **Esbozo mamario:** El tejido en las mamas de los recién nacidos es notorio al ser estimulado por los estrógenos maternos dependiente del grado de nutrición fetal.
- **Genitales masculinos:** Los testículos fetales comienzan su descenso de la cavidad peritoneal al saco escrotal aproximadamente durante la semana 30 de gestación.

Ballard evalúa 6 parámetros neurológicos basados en la madurez neuromuscular e incluyen:

- **Postura:** El tono muscular total del cuerpo se refleja en la postura preferida por el neonato en reposo y la resistencia que ofrece al estirar los grupos musculares individuales.
- **Ventana cuadrada:** La flexibilidad de la muñeca y/o resistencia al estirar los extensores son los responsables del ángulo resultante de la flexión de la muñeca.
- **Rebote del brazo:** Maniobra que se enfoca en el tono del bíceps midiendo el ángulo de rebote producido luego de una breve extensión de la extremidad superior.
- **Ángulo poplíteo:** Medición de la resistencia por el tono pasivo del flexor en la articulación de la rodilla durante la extensión de la pierna.
- **Signo de la bufanda:** Estudio de los flexores a nivel del hombro, llevando uno de los codos hacia el lado opuesto sobre el cuello.
- **Talón a oreja:** Maniobra que se enfoca en el tono pasivo de flexores en la cintura al resistir la extensión de los músculos posteriores de la cadera.

Cada uno de los criterios en el Ballard, tanto físicos como neurológicos, recibe una puntuación desde 0 hasta 5. La puntuación total varía entre 5 y 50, con la correspondiente edad gestacional localizado entre 26 y 44 semanas. Un aumento de la puntuación de 5 corresponde a un incremento en la edad de 2 semanas. La nueva puntuación de Ballard permite valores de -1, por lo que es posible una puntuación negativa, creando un nuevo rango entre -10 y 50, extendiendo la edad gestacional prematura a las 20 semanas. A menudo se usa una fórmula para el cálculo de la puntuación total: $\text{Edad gestacional} = 2(\text{puntuación} + 120)/5$ (Nacho Rios, 2012).

2.1.5. Curvas de Crecimiento

Las curvas de crecimiento, también llamadas tablas de peso y talla, son esquemas de medidas que permiten valorar y comparar la estatura, el peso y el tamaño de la cabeza del niño con relación a un rango estándar, pudiendo de éste modo tener un control acerca del crecimiento del niño pues es necesario para poder evaluar problemas existentes o prevenir los que puedan estar ocurriendo sin síntomas (PulsoMed, 2016).

Existen tres tipos de curvas de crecimiento, las cuales marcan unos valores preestablecidos del niño que coinciden con los percentiles 3, 15, 55, 85 y 97, mediante los cuales se hará la comparación frente a la población de niños de la misma edad (Active, 2016), cabe aclarar que la curva de crecimiento para los niños no es la misma que para las niñas, ésta curvas son:

- Curva del perímetro craneal, para la construcción se usa la medida del tamaño de la cabeza que se toma envolviendo una cinta métrica alrededor de la parte posterior de la cabeza por encima de las cejas Biblioteca Nacional de Medicina (2016), si el resultado de ésta medición demarca que la cabeza crece de manera anormal, se puede estar tratando de una hidrocefalia, un tumor cerebral, etc., y si por el contrario se encuentra que crece muy lentamente, puede ser que el niño tenga una malformación cerebral fusión temprana de las suturas u otros problemas PulsoMed (2016).
- Curva de la altura, se trata de la medición de la estatura del niño la cual es medida en niños menores de 3 años mientras están acostados y estando de pie en niños mayores de 3.
- Curva del peso, en éste caso las mediciones son en onzas, libras o gramos y kilogramos un crecimiento lento de peso y altura puede ser un indicio de problemas en el desarrollo (Biblioteca Nacional de Medicina, 2016).

2.1.6. Control Prenatal

El control prenatal se define según Cancino et al. (2015) como el “conjunto de acciones y actividades que se realizan en la mujer embarazada con el objetivo de lograr una buena salud materna, el desarrollo normal del feto y la obtención de un recién nacido en óptimas condiciones desde el punto de vista físico, mental y emocional”.

Los componentes del control prenatal según Cancino et al. (2015) son:

- El análisis temprano y continuo del riesgo.
- La promoción de la salud, detección temprana y protección específica.

- Las intervenciones médicas y psicosociales

Con base en estos componentes Cancino et al. (2015), define una serie de objetivos generales que, a su vez, caracterizan a un buen control prenatal. Tales objetivos son:

- Identificación de factores de riesgo.
- Diagnóstico de la edad gestacional.
- Identificación de la condición y crecimiento fetal.
- Identificación de la condición materna.
- Educación materna y a su núcleo familiar en actividades de promoción y prevención.

Se recomienda realizar el primer control prenatal en el primer trimestre, idealmente antes de la semana 10 de gestación.

La Organización Mundial de la Salud (OMS) y el National Institute for Health and Care Excellence (NICE) de Reino Unido, basados en un consenso de expertos, recomiendan que la primera visita debe durar el equivalente de dos citas normales de control prenatal (entre 30 y 40 minutos), para las siguientes visitas una duración aproximada de 20 minutos, y en aquellas embarazadas que comienzan tardíamente el control prenatal (después de la semana 32) se debe asignar más tiempo que una primera visita regular (Amaya-Guío et al., 2013).

2.1.7. Hipertensión Arterial (HA)

La presión arterial es la presión de los vasos sanguíneos del cuerpo. Es la fuerza con la que la sangre se mueve por los vasos sanguíneos (of Family Physicians, 2014). Se considera hipertensión en el embarazo cuando se tiene una presión arterial sistólica mayor o igual a 140 mmHg o una presión arterial diastólica mayor o igual a 90 mmHg. Estas cifras deben ser confirmadas de manera repetida después de encontrarse la paciente en reposo, mínimo cinco minutos o durante un seguimiento horario durante 24 horas hospitalariamente. El hallazgo de dos registros iguales o superiores a estos límites con un lapso de diferencia entre cuatro y seis horas confirma el diagnóstico de hipertensión arterial (Petro et al., 2014).

La hipertensión arterial es la complicación médica más frecuente durante el embarazo, se presenta en el 8,10 % de todas las gestaciones y es la principal causa de mortalidad materna en el mundo. La importancia de su estudio reside en que, por un lado, una atención médica prenatal adecuada puede prevenir o atenuar las alteraciones hipertensivas de la gestación y, por otro, que la HA en el embarazo es un marcador precoz de HA esencial y de enfermedad cardiovascular y renal futuras Marin Iranzo et al. (2014). Según Pinheiro (2018) los procesos hipertensivos del embarazo se clasifican en:

- **Hipertensión crónica preexistente:** Individuos con valores de presión arterial a menudo por encima de 140/90 mmHg se consideran hipertensos. En el embarazo, se considera una hipertensión preexistente toda hipertensión que ya existía antes de la mujer embarazarse. También se considera hipertensión preexistente si es identificada antes de la semana 20 del embarazo.
- **Preeclampsia:** Es la aparición de hipertensión después de la semana 20 de gestación asociada a la pérdida de proteína en la orina, que se llama proteinuria. Una hipertensión que aparece después de la semana 20 del embarazo y se asocia con problemas de riñón, hígado, sistema nervioso central o disminución del número de plaquetas también puede ser preeclampsia.
- **Preeclampsia superpuesta a la hipertensión crónica:** Es la preeclampsia que ocurren en mujeres previamente hipertensas.
- **Hipertensión gestacional:** Se considera hipertensión gestacional la hipertensión que se presenta solamente después de la semana 20 de gestación y que no muestra pérdida de proteína en la orina o cualquier otra manifestación sugestiva de preeclampsia.

2.2. Marco Teórico

2.2.1. Regresión No Paramétrica

En el análisis de regresión se parte de la idea de que existe una función de regresión f subyacente, desconocida. Tal como en los modelos lineales, en el modelo de regresión no paramétrica será necesario establecer algunas suposiciones acerca de f . (Olaya, 2012)

Como menciona Faraway (2006) en el enfoque no paramétrico consiste en elegir f de una familia de funciones suaves, por lo que el rango de ajustes potenciales a los datos será mucho mayor con ciertas suposiciones, como el grado de suavidad y continuidad. Estos modelos no paramétricos no tienen una forma formulada de la relación entre los predictores y la respuesta, por lo que se hace gráficamente.

Señalando también que como en el enfoque paramétrico se puede elegir fácilmente la forma incorrecta para el modelo dando como resultado parcialidad, que por otra parte el enfoque no paramétrico asume mucho menos y por lo tanto es menos propenso a cometer errores.

B-Splines

Los splines son funciones polinómicas a trozos sobre las que se imponen restricciones en los puntos de unión llamados nodos, estos puntos dividen el rango de X en regiones. Dependiendo así de 3 elementos: (Durbán, 2006).

- Grado del polinomio
- Número de nodos
- Localización de nodos

Igualmente, es importante en el ajuste de Splines la definición de la base de funciones, donde según Bowman et al. (2010) se construyen las funciones en dos pasos, primero definiendo el sistema de funciones base y luego estableciendo un vector, matriz o arreglo de coeficientes para definir la función.

Sea $\phi_k(t)$ con $k = 1, \dots, K$ llamado funciones base las cuales son combinación lineal, de modo que (Bowman et al., 2010):

$$x(t) = \sum_{k=1}^K c_k \phi_k(t) = c' \phi(t),$$

donde los parámetros $c_1, c_2, \dots, c_k$ son llamados coeficientes de la expansión y la última matriz usa c como un vector de K coeficientes y ϕ se denota como un vector de K funciones base.

Las bases spline son construidas dividiendo el intervalo de observación en L subintervalos, separados por los puntos de corte llamados nodos tales que $\tau_l, l = 1, \dots, L - 1$ y sobre cada intervalo, se especifica un polinomio de orden m , es decir, el número de constantes requeridas para definirlo, el cual es uno más el grado ($m = n+1$). Siendo 3 el grado el comúnmente usado el que consiste en segmentos de polinomios cúbicos (Ramsey and Silverman, 2002).

Según Olaya (2012) la versión más sencilla de un objeto matemático que se comporte como un spline mecánico es llamado spline cúbico, pues supongamos que se tiene un conjunto de números reales $x_1, \dots, x_n$ en un intervalo $[a, b]$, tales que $a < x_1 < x_2 < \dots < x_n < b$. Siendo un spline cúbico si cumple con las siguientes condiciones:

1. s es una función cúbica en cada uno de los intervalos $(a, x_1), (x_1, x_2), \dots, (x_n, b)$
2. Las cúbicas se unen en los puntos x_i de tal manera que s y sus dos primeras derivadas son continuas en cada x_i y por lo tanto en todo el intervalo $[a, b]$.

Ahora, sea $L_2[0, 1]$ el espacio de las funciones cuadrado integrables, donde se asume que $x(t)$ y sus dos primeras derivadas $x'(t)$ y $x''(t)$ están en $L_2[0, 1]$ y a esta colección de funciones cuadrado integrables forman un espacio funcional llamado Espacio de Sobolev denotado como $W_2^2[0, 1]$. Entonces, al proceso de estimar una función de dicho espacio se le denomina *suavización*, por lo que una función perteneciente a $W_2^2[0, 1]$ se llama función *suave*.

Según Ramsey and Silverman (2002), una base de funciones B-spline es un conjunto de funciones $\phi_k(t)$ que cumplen las siguientes propiedades:

- Cada función de la base es en sí misma una función spline definida con un orden m y una secuencia de nodos τ .
- Los múltiplos, las sumas y restas de funciones splines son también splines, por lo tanto cualquier combinación lineal de funciones splines representa una función spline.
- Cualquier función spline definida por un orden m y una secuencia de nodos τ , se puede representar como una combinación lineal de los elementos de la base de funciones.

La notación $B_k(t, \tau)$ es usada a menudo para indicar el valor de t funciones bases definidas por la secuencia de puntos de corte τ , donde k se refiere al número más grande o el valor inmediato a la izquierda del valor t .

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau)$$

2.2.2. Cuantil

Los cuantiles son una generalización del concepto de mediana, así como la mediana divide a los datos en dos partes con el mismo número de elementos en cada una, si ésta división se hace en cuatro partes, diez partes o cien partes se llega al concepto de cuantil.

Veamos una definición más formal que se plantea en Davino et al. (2013):

Partiendo de que Y es una variable aleatoria, se puede utilizar la función de distribución acumulada (FDA), denotada como F_Y , que para cada valor de y da la proporción de la población tal que $Y \leq y$.

Los cuantiles son definidos como ubicaciones particulares de una distribución, es decir que el θ -ésimo cuantil es el valor y tal que $P(Y \leq y) = \theta$

Entonces,

$$F_Y(y) = F(y) = P(Y \leq y)$$

la función cuantil se define como la inversa de la FDA, así:

$$Q_Y(\theta) = Q(\theta) = F_Y^{-1}(\theta) = \inf y : F(y) > \theta; \text{ para } \theta \in [0, 1]$$

2.2.3. Regresión cuantílica

Se sabe que el propósito del análisis de regresión es explicar la relación promedio existente entre un conjunto variables predictoras y una variable respuesta basada en la función de la media condicional $E(y|x)$, sin embargo enfrentados a la realidad en muchas ocasiones no es posible predecir exactamente a partir de las variables regresoras, porque sólo proporciona una relación parcial, ya que podríamos estar interesados en describir la relación entre diferentes puntos de la distribución condicional de y , y la regresión cuantílica proporciona esa capacidad (Hao and Naiman, 2007).

La regresión cuantílica fue 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)

En el libro de Hao and Naiman (2007) se establece que el modelo de la media condicional tiene ciertas limitaciones entre ellas:

- Al resumir la respuesta para valores fijos de las variables predictoras, el modelo de media condicional no puede ser fácilmente extendido a las ubicaciones centrales.
- Los supuestos del modelo no siempre se cumplen en el mundo real, pues en particular con frecuencia falla el supuesto de homocedasticidad.
- El punto focal de la ubicación central (media) siempre ha dirigido la atención de los investigadores lejos de las propiedades de toda la distribución.

Los objetivos que se persiguen en el modelo QR son los mismos que en la regresión por MCO, es decir, modelizar la relación entre variables, sin embargo por lo mencionado anteriormente como problemas de heterocedasticidad, valores atípicos; el valor medio de respuesta que ofrece la estimación por MCO no siempre es el más representativo. Por esto y más la regresión cuantílica ofrece la posibilidad de crear distintas rectas de regresión para distintos cuantiles de la variable respuesta a través de un método de estimación que se ve menos perjudicado por tales inconvenientes. Es decir, aquí existen tantas rectas, como tantos vectores β_θ , como cuantiles se esté considerando.

La especificación del modelo se presenta de la siguiente forma:

$$y_i = X_i\beta_\theta + u_{\theta_i}$$

Donde y_i es la variable respuesta, X_i representa la matriz de covariables o de variables independientes, β_θ es el parámetro a estimar correspondiente al cuantil θ ; y u_{θ_i} es la perturbación aleatoria correspondiente al cuantil θ .

De forma análoga al modelo MCO, en el que $E(y_i|x_i) = X_i\hat{\beta}_{MCO}$ y por tanto $E(u_i|X_i) = 0$; aquí $Quant_\theta(y_i|X_i) = X_i\beta_\theta$ lo que implica que $Quant_\theta(u_{\theta_i}|x_i) = 0$, siendo este el único supuesto que se hace sobre los errores del modelo.

La estimación de los parámetros para el modelo QR se lleva a cabo a través de la minimización de las desviaciones absolutas ponderadas con pesos asimétricos, es decir que a cada desviación correspondiente a la observación i se le da más o menos peso según el cuantil cuya recta de regresión se esté estimando.

$$Min[\sum_{Y_i \geq b} \theta|y_i - b| + \sum_{Y_i \leq b} (1 - \theta)|y_i - b|]; \quad b \in \mathbf{R}$$

Como una primera aproximación al método de estimación de la regresión cuantílica, siendo θ 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.(Vicéns and Sánchez, 2012)

A partir de la expresión anterior donde el valor b correspondiente al cuantil θ es quien minimiza dicha función, si se considera que ese valor b es una simplificación del producto $X_i\beta_\theta$ cuando $X_i = 1$, entonces se tiene que el problema de estimación de los parámetros en QR se pueden expresar de la siguiente manera:

$$Min[\sum_{Y_i \geq X_i\beta_\theta} \theta|y_i - X_i\beta_\theta| + \sum_{Y_i \leq X_i\beta_\theta} (1 - \theta)|y_i - X_i\beta_\theta|] \quad \beta_\theta \in \mathbf{R}$$

Siguiendo a Vicéns and Sánchez, pág. 8 la principal ventaja que da el uso de las desviaciones en valor absoluto en lugar de las desviaciones al cuadrado como en el método de MCO, es que la regresión cuantílica prácticamente no se ve afectada por valores extremos, siendo realmente importante y útil a la vez para este problema donde los datos atípicos son muy comunes.

2.2.4. Datos Funcionales

El análisis de datos funcionales es una disciplina emergente en la estadística actual, la cual trata sobre el análisis de la información en las curvas o funciones transformando los datos discretos en funciones, y a estas funciones se les aplica varias técnicas con el objetivo de descubrir estructuras intrínsecas que con una simple representación no sería posible evidenciar (Voinea, 2014).

Según Ramsey and Silverman (2002) el análisis de datos funcionales puede ser clasificado como exploratorio y confirmatorio o predictivo. Y sus objetivos principales serán esencialmente los mismos que los de cualquier otra rama de la estadística. Incluyen los siguientes objetivos:

representar los datos de manera que ayuden a un análisis más profundo, mostrar los datos para resaltar varias características, estudiar importantes fuentes de patrones y variaciones entre los datos, explicar la variación de un resultado o variable dependiente mediante el uso de información de variables de entrada o independiente, comparar dos o más conjuntos de datos con respecto a ciertos tipos de variación, donde dos conjuntos de datos pueden contener diferentes conjuntos de repeticiones de las mismas funciones o funciones diferentes para un conjunto común de repeticiones.

De acuerdo con Aguilera Morillo (2009), el primer paso en el Análisis de Datos Funcionales es reconstruir la verdadera forma funcional de las curvas muestrales a partir de sus observaciones discretas. En general, para resolver este tipo de problemas se utilizan las bases de B-Splines para datos no periódicos y las bases de Fourier para datos periódicos.

La expansión de esta base más conocida es proporcionada por la serie de Fourier que sigue a continuación:

$$\hat{x}(t) = c_0 + c_1 \sin(wt) + c_2 \cos(wt) + c_3 \sin(2wt) + c_4 \cos(2wt) + \dots$$

Definido por la base $\phi_0(t) = 1$, $\phi_{2r-1}(t) = \sin(rwt)$, y $\phi_{2r}(t) = \cos(rwt)$

Cabe aclarar que esta base es periódica, y el parámetro w determina el periodo $2\pi/w$. Éste tipo de bases son especialmente útiles para funciones extremadamente estables, es decir, funciones donde no existen características locales fuertes y donde la curvatura tiende a ser del mismo orden en todas partes (Ramsey and Silverman, 2002).

Es posible utilizar la amplia teoría desarrollada para la regresión en el contexto funcional, al representar a $x(t)$ cómo una combinación lineal de funciones de una base, de la siguiente forma:

$$x(t) = \sum_{k=1}^K c_k \phi_k(t)$$

En términos de K funciones base conocidas $\phi_k(t)$ y c_k son los coeficientes desconocidos en la combinación lineal, los cuales son objetivo de estimación.

En ese orden de ideas, surge una gran pregunta y es ¿qué base de funciones elegir?, pues según Ramsey and Silverman (2002) en un sistema de bases de funciones, el parámetro K determina en cierta medida el grado en que los datos son suavizados mediante la estimación de la curva construida, así, si $K = n$, siendo n el número de observaciones en el intervalo τ se lleva a cabo una interpolación de los datos, es decir que se hace un ajuste que pasa por

todos los puntos, por otro lado un número pequeño de la base de funciones ignoraría ciertos detalles que pueden llegar a ser de gran información en el fenómeno estudiado mediante la construcción de los datos funcionales.

Por lo que Ramsey and Silverman (2002) mencionan una herramienta potente para la selección del tamaño de la base de funciones el Error Cuadrático Medio (ECM), expresado como:

$$ECM[\hat{x}(t)] = Bias^2[\hat{x}(t)] + Var[\hat{x}(t)]$$

Así, un método robusto para la elección del K y otros parámetros del modelo es el Criterio de Validación Cruzada, que tal como menciona Olaya (2012) consiste en que teniendo n observaciones en n sub-muestras de tamaño $n-1$ mediante el mecanismo de dejar por fuera una observación diferente cada vez, así:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - f_{\lambda(i)})$$

Sin embargo, la carga computacional que este procedimiento conlleva es muy grande por lo que sugieren simplificar este cálculo utilizando el cambio:

$$CV_{GS}(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - f_{\lambda i}}{1 - S_{\lambda ii}} \right)^2$$

Donde $S_{\lambda ii}$ es el elemento i de la matriz S_{λ} .

Y otra posible solución, donde sugieren utilizar otro criterio llamado validación cruzada generalizada $GCV(\lambda)$ definido como:

$$GCV_{GS}(\lambda) = n^{-1} \frac{\sum_{i=1}^n (y_i - f_{\lambda i})^2}{(1 - n^{-1} tr[S_{\lambda}])^2}$$

Estadísticas descriptivas para datos funcionales

Media y varianza funcional

Tal como se determina en Ramsey and Silverman (2002) las estadísticas descriptivas para datos univariados comunes se aplicarán de la misma manera a los datos funcionales, donde la función de la media expresa el promedio de las funciones punto a punto y se expresa:

$$\bar{x}(t) = N^{-1} \sum_{i=1}^N x_i(t)$$

Así mismo la expresión de la varianza:

$$varx(t) = (N - 1)^{-1} \sum_{i=1}^N [x_i(t) - \bar{x}(t)]^2$$

Y la desviación estándar es la raíz cuadrada de la varianza.

2.2.5. Mínimos cuadrados ponderados

Los mínimos cuadrados generalizados (MCG) es una técnica para la estimación de los parámetros desconocidos en un modelo de regresión lineal, se aplican cuando se presenta heterocedasticidad, o cuando existe un cierto grado de correlación entre las observaciones (Greene, 2003). Un caso especial de los MCG y que será utilizado en este trabajo, son los Mínimos Cuadrados Ponderados (MCP), los cuales son usados cuando las varianzas de las observaciones son desiguales.

Definición

Sea $W = V^{-1}$, donde V es una matriz diagonal conocida de $n \times n$, W también es diagonal y sus elementos diagonales, pesos, o factores de ponderación, son $w_1, w_2, \dots, w_n$ que deben ser conocidos (Montgomery et al., 2006). El estimador ponderado de mínimos cuadrados, obtenido a partir de las ecuaciones normales estará dado por:

$$\hat{\beta} = (X^t W X)^{-1} X^t W y$$

Obsérvese que se están ponderando las observaciones a la hora de construir el estimador: Las observaciones con menor varianza tienen ahora un peso mayor que las observaciones con una varianza mayor (Universidad de Alicante, 2018).

2.2.6. Test de Bartlett

El test de Bartlett sigue la línea de dividir las observaciones en grupos supuestamente homogéneos, y contrasta si las varianzas en todos los grupos considerados son iguales. Para ello, ajusta un modelo por cada grupo y compara las varianzas con un estadístico, el K2 de Bartlett, que sigue aproximadamente una distribución chi-cuadrado bajo la hipótesis nula de varianza constante. Donde el estadístico de prueba es:

$$T = \frac{(N - k) \ln(S_p^2) - \sum_{i=1}^k (n_i - 1) \ln(s_i^2)}{1 + \left(\frac{1}{3(k-1)}\right) \left(\left(\sum_{i=1}^k \frac{1}{N_i - 1} \right) - \frac{1}{N - k} \right)} \equiv \chi_{k-1}^2$$

Donde:

N = Tamaño de muestra total
 n_i = Tamaño del grupo i -ésimo
 k = Número de grupos en la población
 $S_p^2 = \sum_{i=1}^k \frac{(N_i-1)s_i^2}{N-k}$
 s_i^2 = Varianza del grupo i -ésimo.

Bajo las hipótesis:

$$H_0 : \sigma_1 = \sigma_2 = \dots = \sigma_k$$

$$H_1 : \exists i, j \text{ tal que } \sigma_i \neq \sigma_j$$

2.2.7. Test de Levene

El estadístico de prueba de Levene se define como: Correa et al. (2006)

$$W = \frac{(N - k) \sum_{i=1}^k n_i (\bar{Z}_{i.} - \bar{Z}_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{Z}_{i.} - \bar{Z}_{..})^2}$$

donde Z_{ij} puede tener una de las siguientes tres definiciones:

1. $Z_{ij} = |X_{ij} - \bar{X}_i|$ donde $\bar{X}_i$ es la media del i - *ésimo* subgrupo
2. $Z_{ij} = |X_{ij} - \tilde{X}_i|$ donde $\tilde{X}_i$ es la mediana del i - *ésimo* subgrupo.
3. $Z_{ij} = |X_{ij} - \bar{X}'_i|$ donde $\bar{X}'_i$ es la media recortada al 10 % del i - *ésimo* subgrupo

$\bar{Z}_{..}$ es la media global de Z_{ij} y $\bar{Z}_{i.}$ es la media del i - *ésimo* subgrupo de los Z_{ij}

Se caracteriza por poder comparar 2 o más poblaciones, por permitir elegir entre diferentes estadísticos de centralidad :mediana (por defecto), media, media truncada.

La prueba de Levene rechaza la hipótesis de que las varianzas son iguales con un nivel de significancia α si $W > F_{\alpha, k-1, N-k}$ donde $F_{\alpha, k-1, N-k}$ es el valor crítico superior de la distribución F con $k-1$ grados de libertad en el numerador y $N-k$ grados de libertad en el denominador a un nivel de significancia α .

La prueba de Levene ofrece una alternativa más robusta que el procedimiento de Bartlett, ya que es poco sensible a la desviación de la normalidad. Eso significa que será menos probable que rechace una verdadera hipótesis de igualdad de varianzas sólo porque las distribuciones de las poblaciones muestreadas no son normal.

2.3. Metodología

2.3.1. Caso de estudio

Los datos con los que se trabajará en este proyecto corresponden a una base de datos suministrada por el Hospital Universitario del Valle (HUV), conformada por los Doctores Javier Torres, Carlos Jiménez y Diana Palencia, docentes del Departamento de Pediatría de la Universidad del Valle, la cual cuenta con una muestra de 145 registros de neonatos correspondientes a historias clínicas de niños con BPN nacidos en la ciudad de Cali, en el HUV entre febrero de 2002 y noviembre de 2010, pertenecientes al Programa de Seguimiento de Alto Riesgo y Canguro.

A continuación se presenta en la tabla **2-1** la descripción, los niveles y la unidad de medida de las variables que se consideraron en este trabajo.

Tabla 2-1: Variables de estudio

Variable	Descripción	Niveles	Unidad de medida
Tiempo	Edad gestacional en meses		Meses
Sexo	Género del recién nacido	Femenino Masculino	
Peso al nacer	Peso al nacer		Gramos
Peso	Peso de los neonatos en el tiempo		Gramos
EGB	Edad gestacional de Ballard		Semanas
DG	Presencia de Diabetes gestacional	Si No	
HA	Presencia de Hipertensión arterial	Si No	
Morbilidad Materna	Presencia de DG y HA	Si No	
Control Prenatal	Asistencia al control prenatal	Si No	

2.3.2. Análisis Exploratorio de Datos

El objetivo del análisis exploratorio es un primer acercamiento a los datos con lo que se quiere ver la influencia de ciertas variables al peso al nacer (PN) de los recién nacidos pertenecientes a este estudio.

Es así, como primera medida se analiza el comportamiento de los pesos al nacer según su clasificación, pues, los neonatos se pueden clasificar según el peso registrado en su nacimiento de la siguiente manera Ramírez (2001):

- De bajo peso (BP), neonatos con un peso inferior a 2500 gr ($< 2500gr$)
- De muy bajo peso (MBP), neonatos con un peso inferior a 1500 gr ($< 1500gr$)
- De bajo peso extremo (BPE), neonatos con un peso inferior a 1000 gr ($< 1000gr$)

Las variables que se tendrán en cuenta en este primer análisis serán peso al nacer (PN), sexo, peso (evolución del peso a diferentes edades 0, 12, 18, 24 y 36), edad gestacional, Control prenatal (CP) y variables clínicas de la madre como diabetes gestacional (DG) e hipertensión arterial (HA). Así mismo se calculan las curvas para los percentiles 5, 10, 25, 50, 75 y 95 de manera exploratoria para la evolución del peso al nacer de los recién nacidos, pues más adelante se presentará otra forma de calcularlos.

2.3.3. Ajuste No Paramétrico (NP) del individuo promedio con B-Splines

El objetivo principal en esta sección es el ajuste de la curva promedio usando la suavización B-Spline, donde la variable respuesta (Y) es el *Peso* y la variable explicativa (X) es el *Tiempo* que corresponde a los 5 controles registrados. Así, se definió el número de nodos como 5 los cuales son los valores únicos de la variable Tiempo, pues como se menciona en Durbán (2006) habrán tantos nodos como observaciones únicas de X , definiendo también el orden 4, ya que se ajustarán B-Splines cúbicos.

El cálculo de las curvas B-Spline se lleva a cabo por medio de la función *stat_smooth* de la librería *ggplot2*, que realiza la regresión no paramétrica con B-spline calculando las funciones base, las cuales son una combinación lineal cuyo papel es ser las variables regresoras del modelo, así como se define en Bowman et al. (2010):

Sea $\phi_k(t)$ con $k = 1, \dots, K$ llamado funciones base de modo que:

$$x(t) = \sum_{k=1}^K c_k \phi_k(t) = c' \phi(t),$$

donde los parámetros $c_1, c_2, \dots, c_k$ son llamados coeficientes de la expansión y la última matriz usa c como un vector de K coeficientes y ϕ se denota como un vector de K funciones base.

Comparación de ajuste NP de datos totales Vs datos completos

Ahora bien, se presenta la comparación entre las curvas B-spline ajustadas para los datos totales, es decir, los 145 neonatos con que se cuenta en la base de datos vs el ajuste para los datos completos que corresponden a los datos del peso de los neonatos que tienen todos los 5 registros de la edad (0, 12, 18, 24 y 36), que en este caso son 35 niños.

La estimación de estas curvas se lleva a cabo de la misma manera como se mencionó anteriormente en el cálculo de la curva B-spline para el individuo promedio, usando también 5 nodos y Splines cúbicas.

Una vez realizada la comparación anterior, se lleva a cabo la construcción de los datos funcionales, donde cabe aclarar que para los 145 registros no se pudo llevar a cabo por la presencia de datos faltantes, pues la función empleada (*min.basis*) no permite NA's, por lo que sólo se construyó para las 35 observaciones completas para dar cumplimiento con el objetivo a pesar de las limitaciones.

El primer paso para la construcción de los datos funcionales es reconstruir la verdadera forma funcional de las curvas muestrales a partir de sus observaciones discretas, es decir en nuestro caso ver cada niño como una observación funcional (curva), y así se emplea la función *min.basis* de la librería *fda.usc*, con el propósito de obtener el número óptimo de bases k y el parámetro de suavización λ los cuales son calculados bajo el criterio de validación cruzada generalizada (GCV), es decir, encuentran los valores óptimos de estos parámetros que minimizan la suma de cuadrados por GCV, con la siguiente expresión:

$$GCV_{GS}(\lambda) = n^{-1} \frac{\sum_{i=1}^n (y_i - f_{\lambda i})^2}{(1 - n^{-1} \text{tr}[S_{\lambda}])^2}$$

De lo que se obtuvo que el número de bases óptimo es 4 y el parámetro de suavización fue 0.

Para la suavización B-splines de las curvas se usó la función *smooth.splines* se establecen los parámetros de orden 4 determinando también los 5 nodos que al no ser equidistantes se calcularon y se pasaron al intervalo $[0,1]$, obteniendo los siguientes nodos 0; 1/3; 1/2; 2/3; 1.

Cabe mencionar que también se calculan descriptivas funcionales como la media, cuya expresión se presenta a continuación:

$$\bar{x}(t) = N^{-1} \sum_{i=1}^N x_i(t)$$

2.3.4. Estimación de los percentiles

En esta sección se propone un método empírico para la estimación de las curvas del percentil 5 y 95, siendo estos de gran importancia ya que el interés se centra en ver el comportamiento de los niños más y menos pesados, dejando un aporte para la clasificación de los mismos.

Esta propuesta surge al no tener pares únicos (x_i, y_i) , pues como menciona Olaya (2012) en el análisis de regresión se asume que se tienen observaciones de la variable respuesta Y para n valores predeterminados de una variable independiente X , es decir, las n observaciones bivariadas disponibles, denotadas $(x_1, y_1), \dots, (x_n, y_n)$, sin embargo en algunos problemas se dispone de más de una respuesta para cada valor de X , que en este caso se cuenta con 5 valores de X (Tiempo en meses), en este orden de ideas cada niño (cada curva) contiene un máximo de 5 observaciones del peso para los 145 RN involucrados en el estudio. Por lo tanto para cada grupo de edad (0, 12, 18, 24 y 36) se hace el cálculo del 10 % de los RN con mayor y menor peso, teniendo en cuenta el número de observaciones en cada tiempo (ver tabla 3-3), una vez obtenidos los datos para cada conjunto se realiza el ajuste de la curva B-spline (5 nodos y orden 4), lo que equivale a la estimación de la curva para el percentil 5 y 95, siendo materia de comparación con las curvas percentílicas ajustadas con Regresión Cuantílica.

2.3.5. Modificación de la variable Tiempo

Ahora, contando con la información suministrada en la base de datos como la fecha de los controles prenatales, surge la idea de recalcular la variable edad, ya que en realidad las 5 edades con las que se cuenta inicialmente no están expresando el tiempo real en el que al niño se le registró el peso. Por tal razón, el objetivo se centra en encontrar la edad real de los recién nacidos para cada medición en el tiempo.

Adicionalmente, se quiere agregar la variable Peso al Nacer (PN) como parte del análisis por lo que se crea una nueva medición para las X (ahora son 6), la cual corresponde al primer registro del peso y que además todos los valores respectivos estén convertidos a días.

Sabiendo que la Edad Gestacional de Ballard (EGB) indica la edad del nacimiento en semanas, se considera esa edad como el día 1 (primer valor de X), pero se considera el cálculo de los demás registros dos formas distintas:

- Edad corregida: Corresponde a la edad del RN a partir de la semana 40 de gestación, es decir la edad en la que el niño debió haber nacido. Así, que si un niño nace en la semana 32 de gestación a las 16 semanas de nacido tendrá 8 semanas de edad corregida.
- Edad cronológica: Edad real que tiene el niño, es decir, la edad empieza a contar a partir

del nacimiento, por ejemplo, si un niño nació en la semana 30, ya para la semana 40 lleva 10 semanas de nacido, lo que equivale 70 días de nacido sumando así los 70 días a todas las edades restantes.

Dado lo anterior, con base en el objetivo de representar los datos lo más cercano posible a la realidad, se lleva a cabo el análisis del ajuste de la curva B-spline, los percentiles y demás con la variable de Edad cronológica.

Cabe mencionar que de los 145 neonatos, 5 de ellos no contaban con información de la variable EGB, lo cual causó problemas a la hora de pasar la edad a días pues no se contaba con el número de semanas de nacimiento, por lo que se optó por calcular el promedio de las edades en días para cada grupo y completar dicha información para estos 5 niños, y así contar con las 145 observaciones.

Así mismo, con el propósito de mejorar la estimación de la curva de crecimiento, debido a la presencia de heterocedasticidad en los modelos ajustados donde la propuesta inicial fue calcular la matriz de varianzas y covarianzas para ser considerada en el ajuste, sin embargo por limitaciones en el software disponible no se pudo incluir en el análisis pero se deja propuesto para un trabajo futuro, se propone entonces, la estimación de la curva asignando pesos a los grupos calculando las varianzas para cada grupo de edad, agregando en el modelo un vector de pesos que corresponde al inverso multiplicativo de la varianza de cada grupo de edad.

Entonces, para pasar los datos del tiempo en días al intervalo $[0,1]$ se tiene la siguiente propuesta:

Se parte de que la variable X se ha medido en N puntos equidistantes en el intervalo $[a, b]$. Por ejemplo, asumiendo que se mide una respuesta a lo largo de un día y que se recoge una observación cada hora, como cuando se mide la temperatura Y una vez cada hora. Aquí X es una variable continua (tiempo) y se observa una medición de la respuesta Y cada hora. Entonces, así se tendrían $N = 24$ observaciones tales que los valores de X están separados de manera uniforme (separados entre sí por una hora), en un día.

En este caso, un mecanismo sencillo de convertir las mediciones de $[1, N]$ a $(0, 1)$, es utilizar la expresión:

$$X_i = \frac{2i - 1}{2N}, \quad i = 1, \dots, N,$$

que produce N valores equidistantes entre 0 y 1, sin incluir el 0 ni el 1.

Para este caso, $\mathbf{X}$ se expresa en días y los niños no se pesan cada día, por lo que las edades

en días a las cuales se pesan los niños no son equidistantes. Es decir, la expresión genérica anterior no es aplicable. Por esta razón se debe modificar este mecanismo de conversión. La propuesta es:

Sea e_j la edad j observada y sea $E = e_N$ la edad máxima observada (en días). Sea N el número de medidas registradas. Entonces para la conversión de $[1, E]$ a $(0, 1)$ se puede utilizar la expresión:

$$X_j = \frac{2e_j - 1}{2E}, \quad j = 1, \dots, N,$$

De esta manera se convierten todos los valores observados de $\mathbf{X}$ en valores en el intervalo $(0, 1)$. Nótese que la primera observación se obtiene en la edad $e_1 = 1$.

Una vez convertida la variable edad, se procede a realizar el ajuste No Paramétrico de la curva B-Splines, como también las curvas percentílicas tanto para el método empírico antes presentado como también las curvas percentílicas con B-Splines y regresión cuantílica.

3 Resultados

En primera instancia se realiza un análisis exploratorio de datos, para analizar y conocer el comportamiento del peso de los recién nacidos con BPN y de las variables incluidas en el estudio como las mediciones antropométricas, algunas condiciones de los recién nacidos y las variables clínicas con respecto a la madre.

3.1. Análisis Exploratorio de Datos

Se incluyeron en el análisis 145 neonatos, con sus respectivas clasificaciones según el peso registrado en su nacimiento: Bajo peso (*BP*), Muy Bajo Peso (*MBP*) y Bajo Peso Extremo (*BPE*). En la figura 3-1 se presenta la distribución de los pesos al nacer de los neonatos con condición de bajo peso clasificado como se mencionó anteriormente en la metodología (línea roja); donde 21 de los 145 recién nacidos presentan BPE lo que equivale al 14.48 % de los neonatos, 76 de ellos tienen MBP equivalente al 52.41 % y 48 con BP, es decir el 33.1 % (ver tabla 3-1).

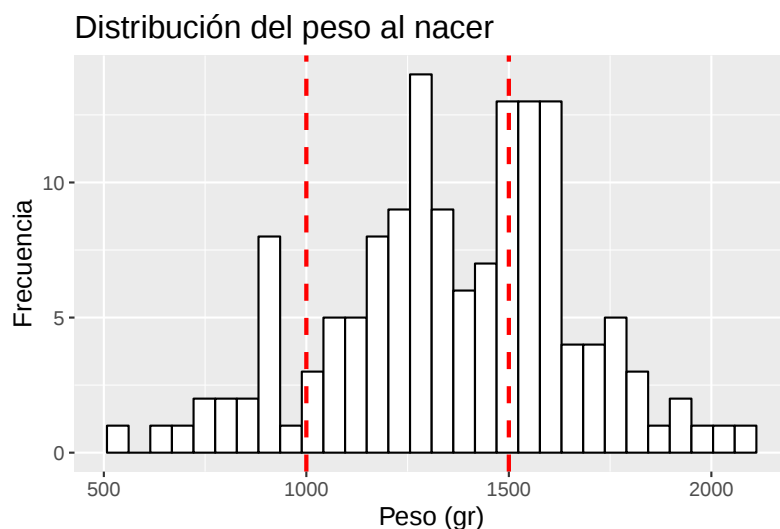


Figura 3-1: Distribución del peso al nacer.

Tabla 3-1: Distribución por clasificación del peso al nacer

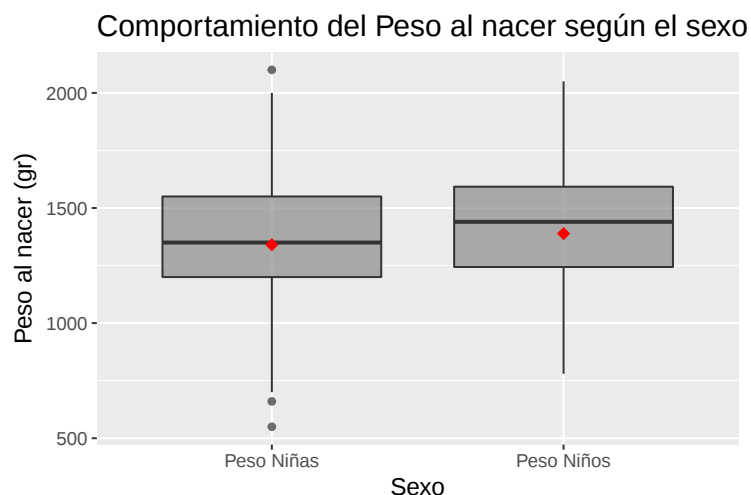
	Clasificación	Frecuencia	Porcentaje
1	BPE	21	14.48 %
2	MBP	76	52.41 %
3	BP	48	33.10 %

Ahora, con respecto al género de los neonatos se tiene que el 60.69 % de los infantes en la muestra son niñas y el 39.31 % niños; al estudiar la influencia de la variable sexo con respecto a la distribución del peso al nacer, gráficamente se encontraron muy pocas diferencias con respecto al sexo, tal como se muestra en la figura **3-2** y en la tabla **3-2**, a pesar de que la distribución de los pesos para las niñas es más variable que para los niños con un coeficiente de variación igual a 22.82 %, y el pesos de los niños es mayor que las niñas (peso promedio de 1440.56 gr).

También, se evaluó la influencia del sexo sobre el peso por medio del Test exacto de Fisher bajo un nivel de significancia del 5 % del que se obtuvo un valor-p igual a 0.1017, por lo que se concluye que no existe evidencia suficiente para afirmar que el género de los neonatos influye sobre el peso de los mismos.

Tabla 3-2: Estadísticas descriptivas para el peso al nacer según el sexo

	Medias	DS	CV(%)
Niñas	1313.20	299.67	22.82
Niños	1440.56	282.98	19.64

**Figura 3-2:** Peso al nacer según sexo

También, es importante conocer el comportamiento del peso con respecto a la edad gestacional. Como lo muestra la figura **3-3** donde se evidencia como incrementa el peso de los neonatos a medida que aumenta la edad, en la tabla **3-3** se encuentran las estadísticas descriptivas y claramente el promedio y la desviación estándar del peso de los recién nacidos incrementa con la edad, pero al observar el Coeficiente de Variación (CV), la edad 0 (semana 40 de la edad gestacional) posee el mayor grado de variabilidad con un CV de 19 %. Los pesos para las edades 12 y 18 son menos variables presentando un CV de 11 % y para las edades 18 y 36 un CV de 13 %.

Tabla 3-3: Estadísticas Descriptivas para el peso según la EG

Edad	Promedios	Frecuencias	DS	CV(%)
0	2829.71	143	539.19	19.05
12	8756.43	126	997.64	11.39
18	10067.11	111	1121.15	11.14
24	11368.22	101	1463.19	12.87
36	12730.44	45	1646.91	12.94

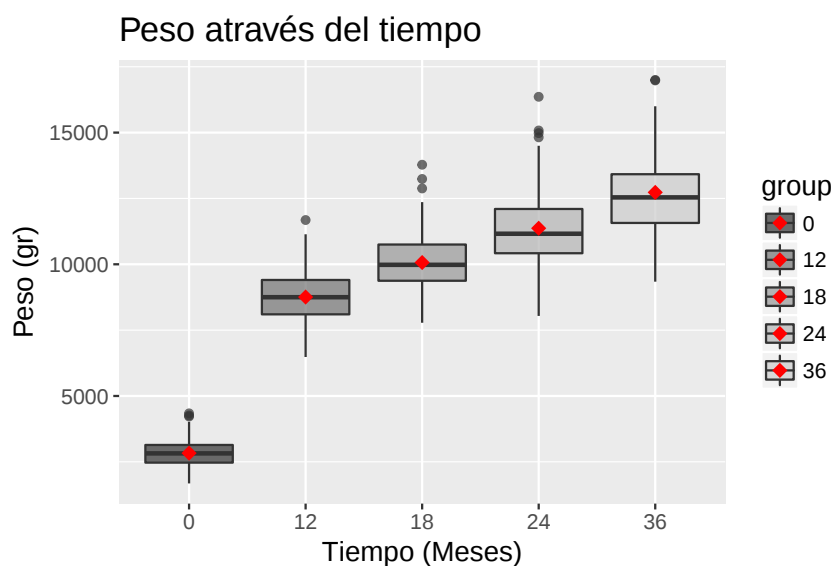


Figura 3-3: Comportamiento del peso al nacer según la EG

Por otra parte, es relevante resaltar que para cada edad no se cuenta con todos los datos, es decir que hay presencia de datos faltantes, debido a que no a todos los niños se les registró las cinco mediciones para cada edad. Ahora bien, se pretende analizar la influencia de ciertas variables clínicas de la madre en el peso al nacer, tales como la diabetes gestacional (DG), hipertensión arterial (HA) y control prenatal (CP); cabe aclarar que estas son variables cualitativas dicotómicas. Como se muestra en la Tabla **3-4** sólo 3 madres presentaron DG,

donde uno de los niños presentó BPE y los otros dos MBP; así mismo sucede con la variable hipertensión arterial, para aquellas madres que si presentaron HA la mayor frecuencia de los pesos de los RN se clasifican en BPE y MBP. Para ver de una forma más clara se analiza la variable Morbilidad materna con respecto al peso, la cual también es cualitativa dicotómica, y será “si” en caso de que la madre presente algunos de los dos casos (DG ó HA). En la Figura 3-4 se encuentra el peso al nacer para aquellos neonatos cuyas madres tiene antecedentes clínicos y se observa que el peso de estos es mucho menor comparado con el grupo de neonatos cuyas madres no presentaron antecedentes clínicos.

También, de las 145 madres 99 si realizaron el control prenatal y 46 de ellas no lo hicieron, a pesar de que la mayoría de los registros en la variable CP es “Si”, el porcentaje de RN con BPE y MBP es mayor para las que no tuvieron control prenatal comparado con la madres que si se realizaron, lo cual nos da entender que puede existir cierta influencia del CP en el peso de los neonatos.

Por lo que se contrasta con el Test exacto de Fisher para evaluar esta hipótesis bajo un nivel de significancia del 5 % obteniendo un valor-p igual a 1 por lo que se concluye no existe evidencia suficiente para afirmar que realizarse o no el control prenatal tenga influencia sobre el peso de un recién nacido. Así mismo, se realizó esta prueba de significancia para las variables DG, HA obteniendo como resultado los valores-p 0.2277 y 0.0008457 respectivamente, concluyendo así que existe evidencia para decir que el peso de los recién nacidos si se ve influido cuando la madre ha presentado Hipertensión arterial durante el embarazo.

Tabla 3-4: Análisis de variables clínicas de la madre con respecto al PN

Variable	Presencia / Ausencia	BPE (21)	MBP (76)	BP (48)	Total (porcentaje)
Diabetes Gestacional (DG)	SI	1	2	0	3 (2,0689 %)
	NO	20	74	48	142 (97,9310 %)
Hipertensión Arterial (HA)	SI	11	27	7	45 (31,0344 %)
	NO	10	49	41	100 (68,9655 %)
Control Prenatal (CP)	SI	13	53	33	99 (68,2758 %)
	NO	8	23	15	46 (31,7241 %)
TOTAL					145

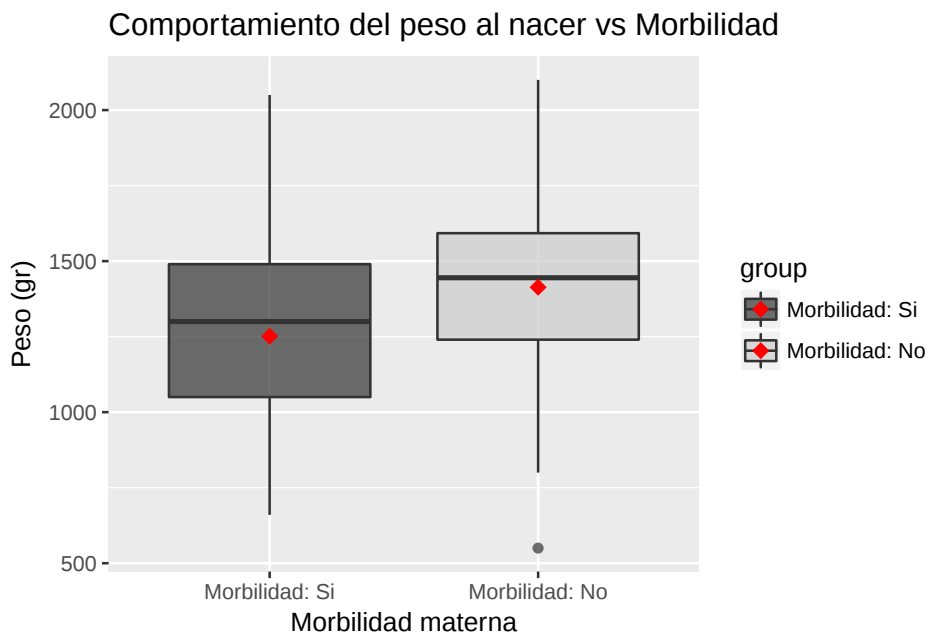


Figura 3-4: Distribución de la Morbilidad materna respecto al PN

En la Figura 3-5 se presentan las curvas de crecimiento de cada uno de los neonatos incluidos en la base de datos. Se puede observar que las curvas varían de individuo a individuo, además, no todos los neonatos tienen completos los registros para el conjunto de las 5 edades (meses), pues a medida que incrementa la edad, se tienen menos registros de los pesos de los RN, siendo esto congruente con lo que se mostró anteriormente en la Tabla 3-3, y también se observa que entre la edad 24 y 36 hay un cambio notorio, pues la cantidad de observaciones en la última medición es la menor de todas, sólo se cuenta con 45 de 145 observaciones que se tienen en total, haciendo de esta manera que sea mas evidente la amplitud entre las curvas, además hay que tener en cuenta que a medida que pasa el tiempo, el ritmo del desarrollo de cada niño es diferente (Cook et al., 2001).

También, en la figura 3-5 se presenta un acercamiento a lo que se hablará mas adelante sobre datos funcionales, donde posteriormente a cada curva se tomará como una observación funcional del peso del RN. A modo exploratorio en la figura 3-6 se presentan las curvas percentílicas (5 %, 10 %, 25 %, 50 %, 75 % y 95 %) para el peso, donde se puede observar que siguen una distribución similar a la de los datos.

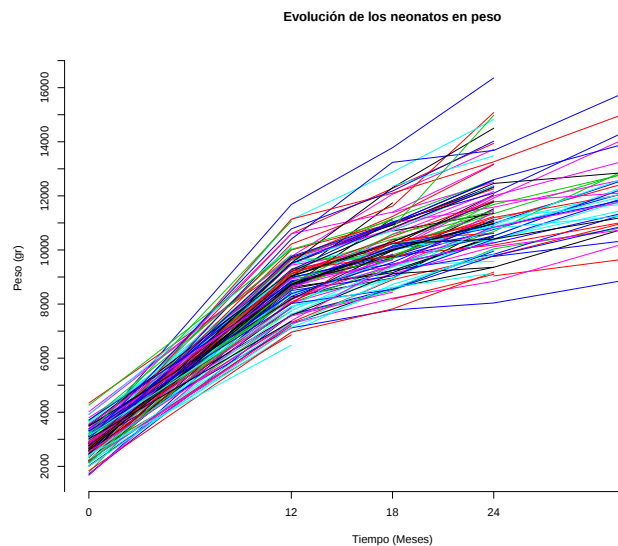


Figura 3-5: Curvas de crecimiento de los RN

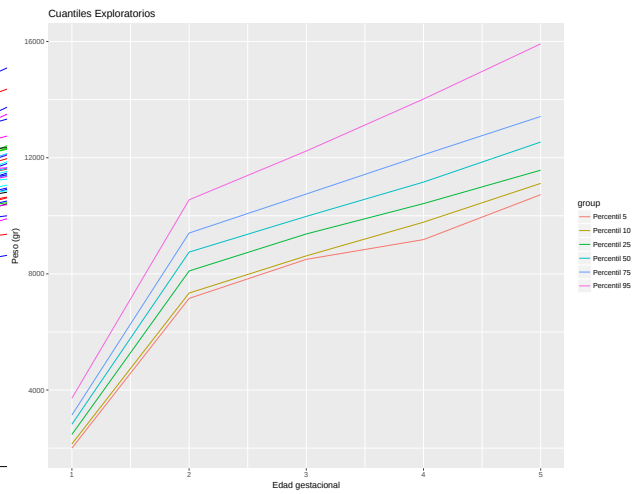
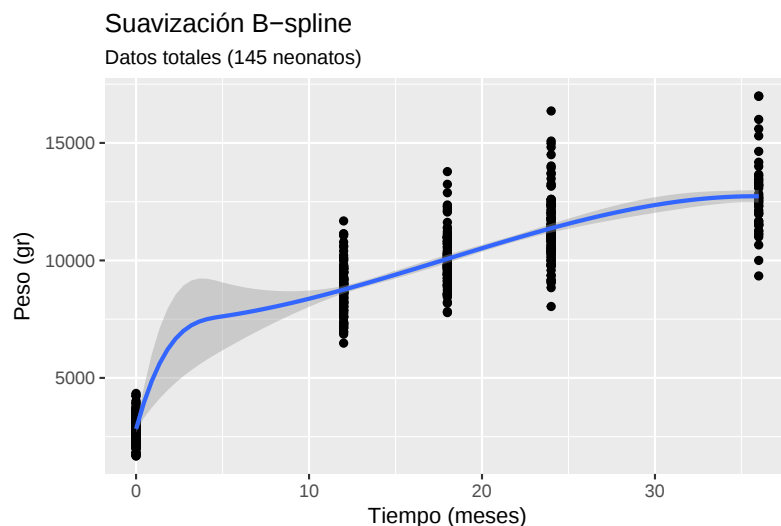


Figura 3-6: Curvas percentílicas para el BPN

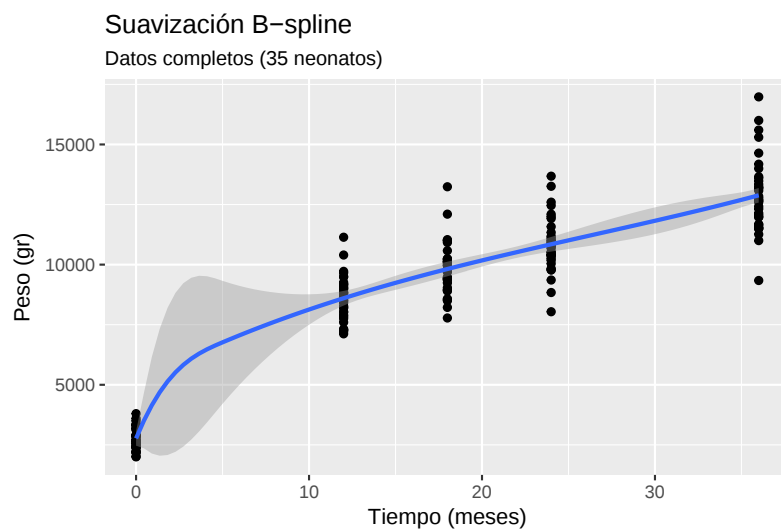
3.2. Ajuste No Paramétrico (NP) del individuo promedio con B-Splines

Se realiza la suavización B-spline para los datos totales con los parámetros definidos en la metodología, en la figura 3.7(a) se observa la curva suavizada del individuo promedio con bandas de confianza del 90%, donde se puede apreciar que hay una mayor amplitud de las bandas entre la edad de 0 y 12 meses, que puede ser provocada por la variabilidad que se presenta en ese tramo, posiblemente causada por el crecimiento acelerado de los niños en su primer año de vida, pues observando los datos del primer control (edad cero) se aprecia que el peso de los neonatos está más concentrado, es decir, tiene menor variabilidad comparados con los registros del peso para los 12 meses, donde es evidente el aumento y dispersión en los pesos de los niños para dicha edad.

Ahora bien, en la figura 3.7(b) se presenta la suavización B-spline para los 35 niños que tienen completos los 5 registros del peso (pues como se mencionó anteriormente sólo se cuenta con 35 datos completos), que comparada con la curva suavizada para los 145 niños es evidente la diferencia en la concavidad de la curva entre el tiempo 0 y 12, no obstante en la figura 3.7(a), las bandas de confianza son menos amplias que en la suavización con los 35 niños con mediciones completas. El cambio en la forma de la curva y de las bandas puede ser ocasionado por la gran disminución de los datos, pues se “pierde” la información de 110 neonatos. De igual manera, es notable la mayor amplitud de las bandas de confianza en el segmento entre 24 y 36 meses también causado por el mayor espacio entre dichas edades.



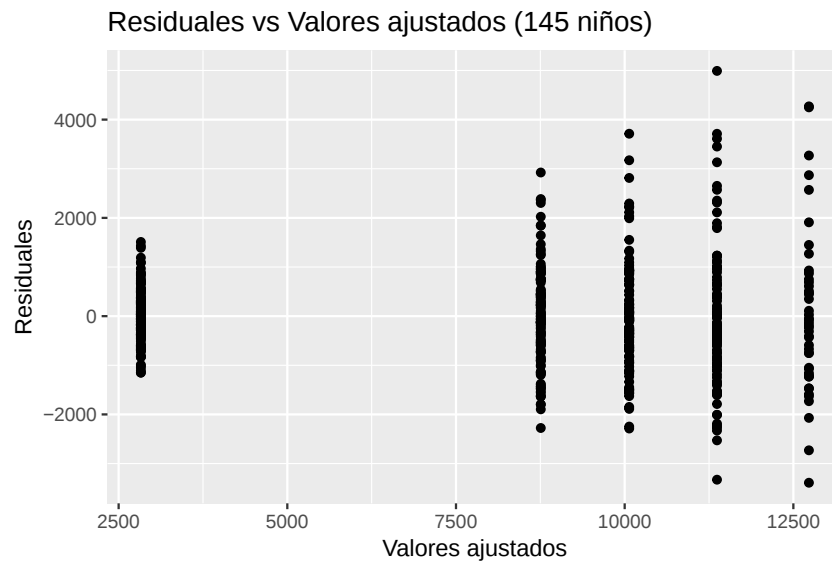
(a) Ajuste del modelo B-Spline 145 RN



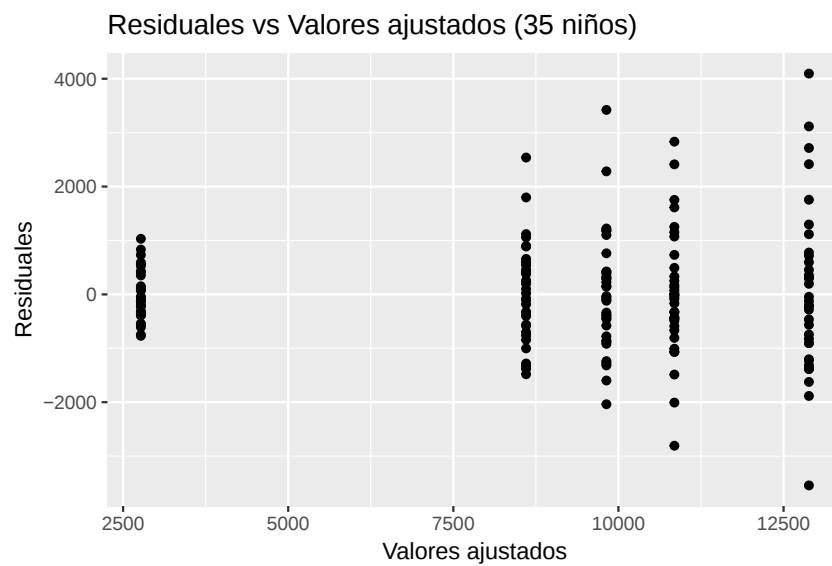
(b) Ajuste del modelo B-Spline 35 RN

Figura 3-7: Suavización B-Spline

Debido a la varianza presente en los datos que se puede observar en los gráficos de puntos, se realiza el gráfico de residuales vs valores ajustados (Figura 3-8) para evaluar el supuesto de homogeneidad de varianza, con lo que se confirma que el modelo no es homocedástico, siendo clara la tendencia creciente (o en forma de embudo como se menciona en la literatura) de los residuales del modelo para las diferentes edades, es decir, como se ya se dijo la heterocedasticidad es evidente en las observaciones del peso de los RN, cuya varianza tiende a cambiar para cada una de las edades registradas. Siendo el mismo caso para el ajuste con los datos completos (35 Neonatos).



(a) Residuales vs Valores ajustados para 145 niños



(b) Residuales vs Valores ajustados para 35 niños

Figura 3-8: Residuales Vs Valores ajustados

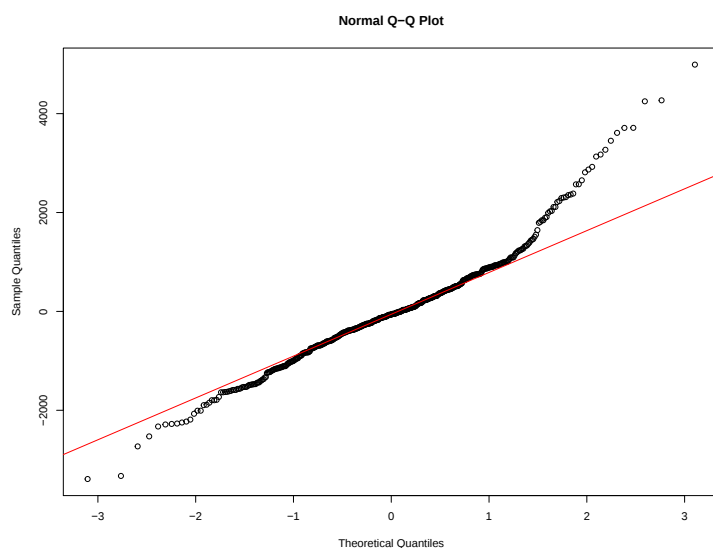


Figura 3-9: Cuantiles teóricos vs cuantiles empíricos para 145 niños

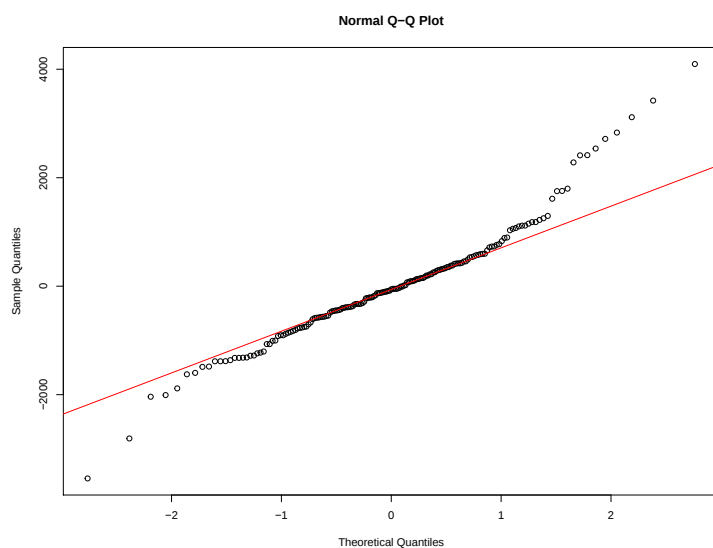


Figura 3-10: Cuantiles teóricos vs cuantiles empíricos para 35 niños

Para evaluar la normalidad de los residuales se presentan los histogramas para ambos casos (Figura 3-11 y Figura 3-12) donde se ve un comportamiento simétrico de los residuales y, en los QQ-plot (Figura 3-9 y Figura 3-10), se observa que tiene colas pesadas por lo cual no es posible asegurar el cumplimiento de la normalidad.

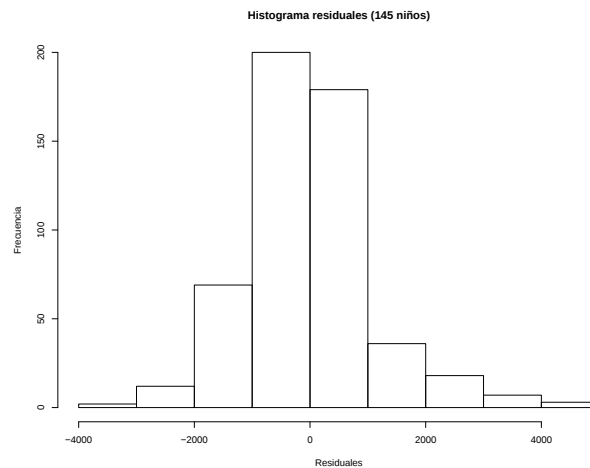


Figura 3-11: Histograma de residuales para 145 niños

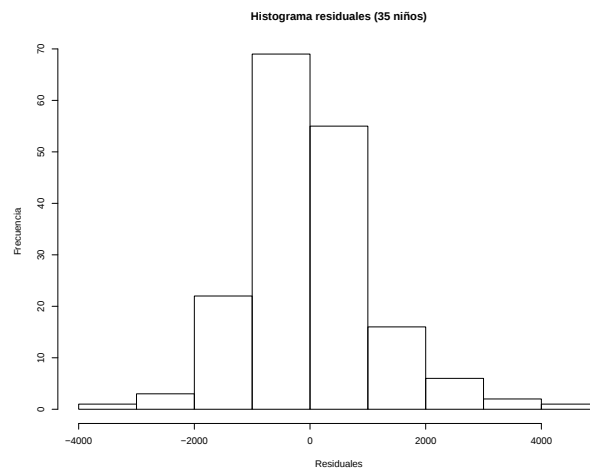


Figura 3-12: Histograma de residuales para 35 niños

Ahora bien, se aplica una prueba formal para evaluar el supuesto de varianza constante en el modelo ajustado, para ello se aplica el test de Bartlett que como se mencionó antes, consiste en dividir las observaciones en grupos supuestamente homogéneos contrastando si las varianzas en todos los grupos son iguales, pues como es evidente en este caso la variabilidad aumenta con el valor medio de la respuesta.

Siendo así, es claro que con respecto al tiempo aumenta la variabilidad en el peso, por tanto la variable Tiempo es quien clasifica o divide en grupos las observaciones y será usada para llevar a cabo el Test de Bartlett (bajo un nivel de significancia del 5 %) que ofrece el software estadístico R mediante la función `bartlett.test()`, de lo que se obtuvo un valor-p igual a 0.000,

el cual es mucho menor que $\alpha = 0.05$ por lo que se concluye que hay evidencia suficiente para rechazar la hipótesis de homogeneidad de varianza. Cabe mencionar que esta prueba es sensible a la normalidad de los errores, por lo que puede no ser tan potente en este caso.

Por lo que se opta por aplicar el test de Levene que es más robusto por ser menos sensible a la desviación de la normalidad, por medio de la función *leveneTest()*, obteniendo un valor-p de 0.000, que bajo el mismo nivel de significancia $\alpha = 0.05$ se rechaza la hipótesis de que las varianzas entre los grupos de edad son iguales.

3.3. Datos Funcionales

Para la construcción de los datos funcionales, como se detalló en la metodología, se contó sólo con 35 neonatos con sus registros de las 5 edades, ya que para poder emplear ésta metodología es necesario tener datos completos. Así, se obtuvo que el número de bases óptimo es 4 y el parámetro de suavización fue 0.

En la suavización B-splines se establecieron los parámetros de orden 4 y los 5 nodos que al no ser equidistantes se calcularon y se pasaron al intervalo $[0,1]$, obteniendo los siguientes nodos 0; $1/3$; $1/2$; $2/3$; 1, los cuales se pueden apreciar también en la Figura 3-13.

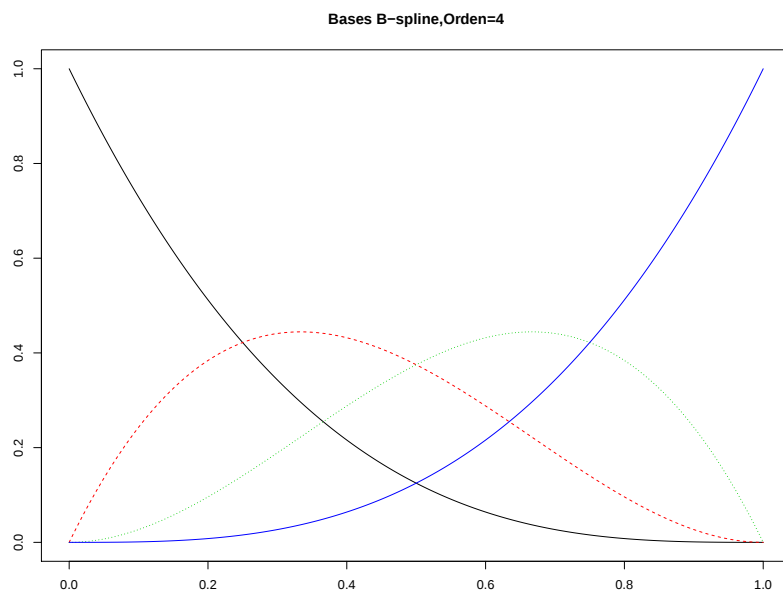


Figura 3-13: Bases B-spline

Ahora, para representar las curvas funcionales para el peso de los 35 recién nacidos para los 5 controles del peso (0, 12, 18, 24 y 36) que en este caso son los nodos especificados

anteriormente. Donde se puede observar el comportamiento del peso de los neonatos que no sea aleja de lo encontrado anteriormente, donde para las primeras mediciones se evidencia menos varianza y que a partir del segundo registro del peso comienzan siendo más dispersas las curvas a medida que transcurre el tiempo.

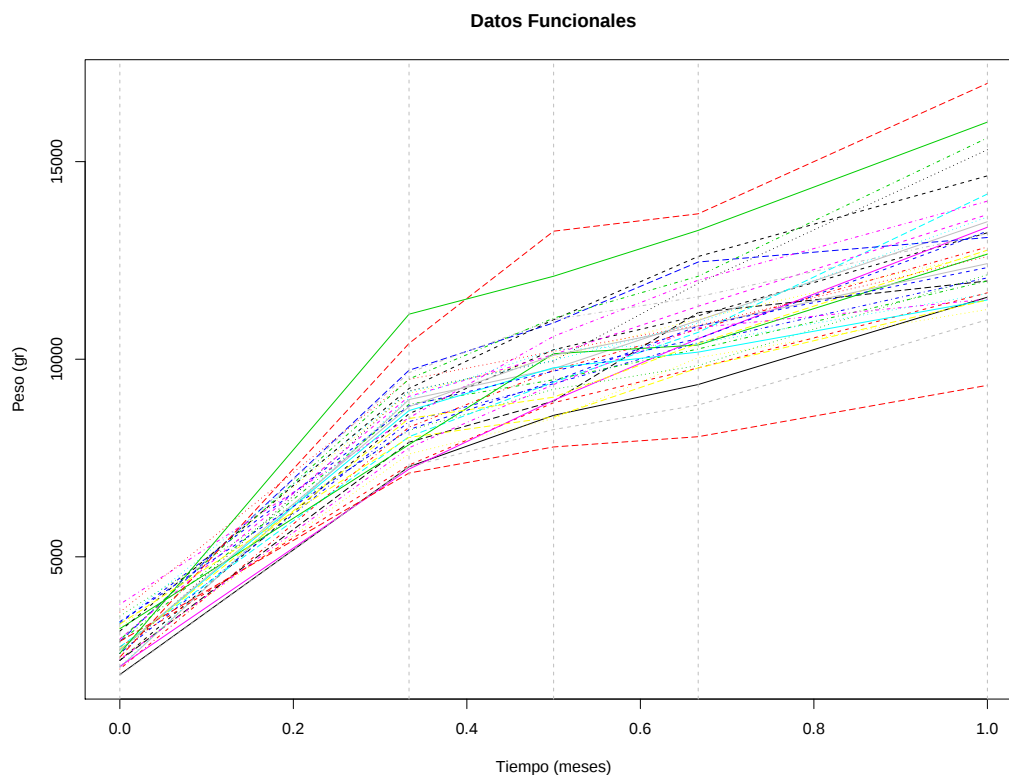


Figura 3-14: Curvas Funcionales

Ahora, se tiene la suavización de las curvas funcionales en la figura **3-15** donde se ha penalizado a primeras dos derivadas continuas por ser un B-Spline cúbico, y en la figura **3-16** que representa la curva media funcional de los 35 neonatos considerados para este análisis de datos funcionales, apreciándose una curva creciente y bastante suave para la muestra de las 35 observaciones funcionales.

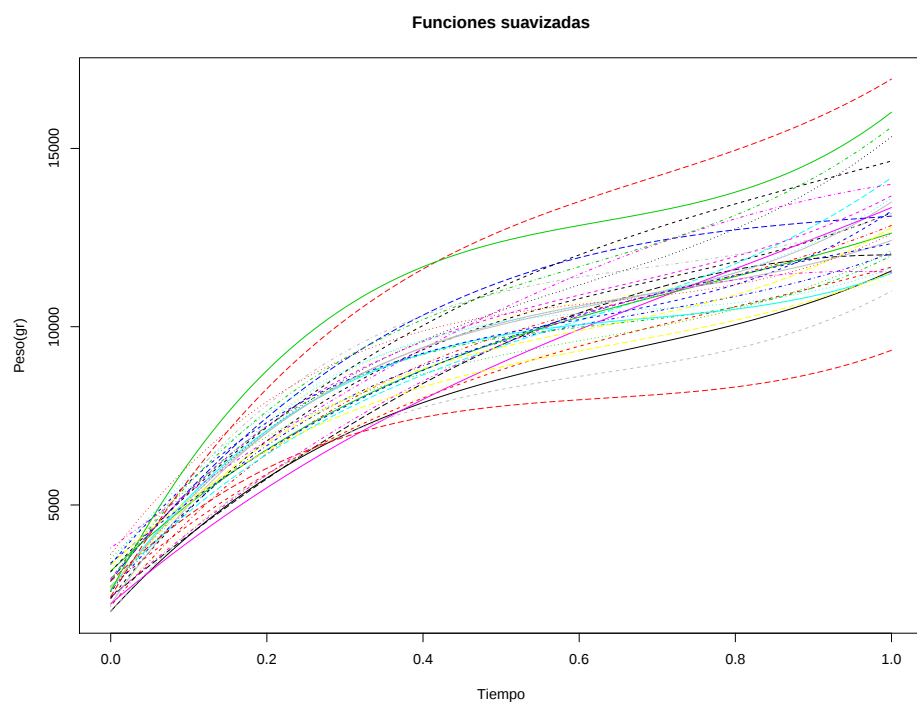


Figura 3-15: Curvas suavizadas

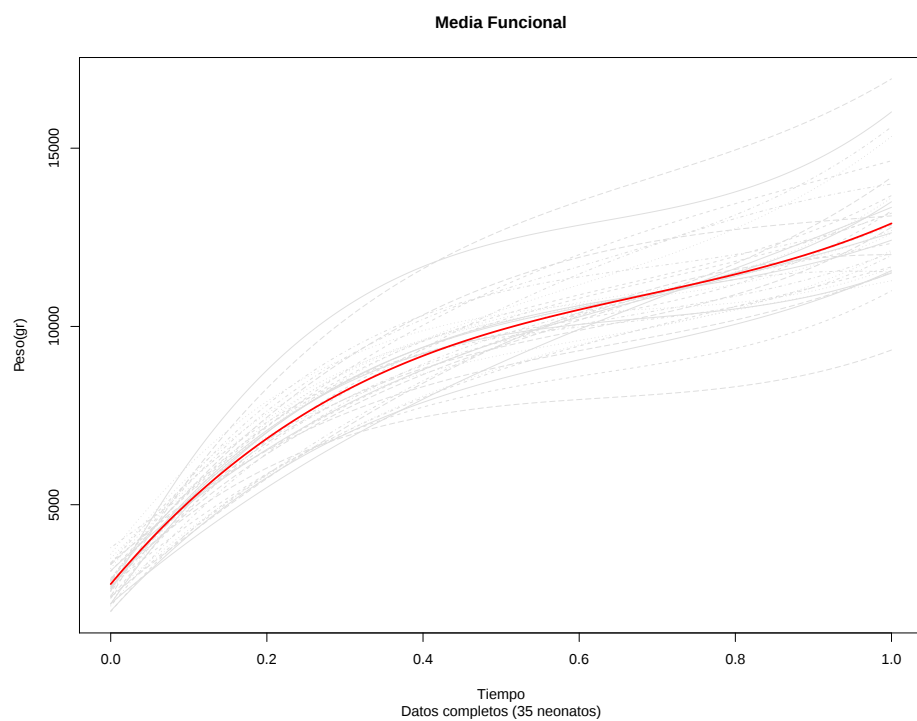


Figura 3-16: Media Funcional

3.4. Estimación de los percentiles

Tal como se mencionó en la sección 2.3.4, se estimaron los percentiles 5 y 95 sacando el 10 % de las observaciones con mayor y menor peso para cada una de las 5 edades (0, 12, 18, 24 y 36) y con ellas se realiza el ajuste de la curva B-spline, teniendo como resultado:

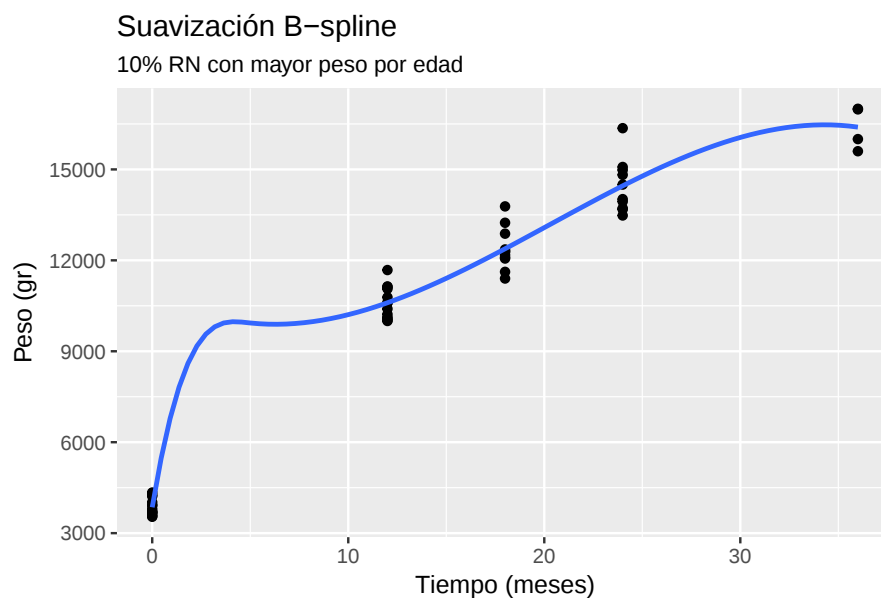


Figura 3-17: Estimación para el percentil 95

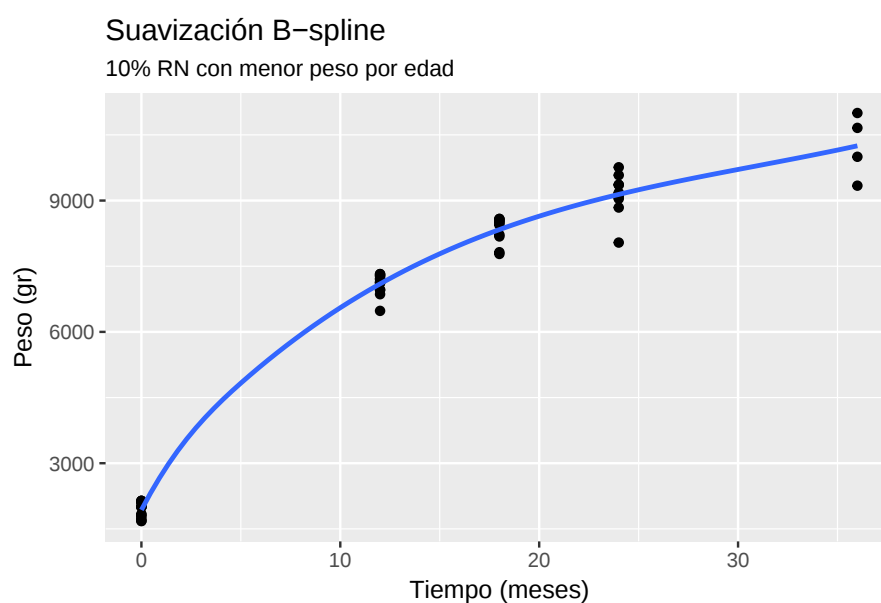


Figura 3-18: Estimación para el percentil 5

En la figura **3-17** se observa que la estimación del percentil 95 tiene un comportamiento muy similar a la estimación de la curva para el individuo promedio (Figura 3.7(a)) , ésto es congruente con dicho resultado pues se refleja la variabilidad entre las primeras dos primeras mediciones (0 a 12 meses), se aprecia también que la curva del percentil 95 es más cóncava (igualmente entre la edad 0 y 12 meses) que la del percentil 5 (Figura **3-18**), atribuible a que el comportamiento del peso de los niños no necesariamente es monótono creciente, si no que presenta cambios en el primer año de edad.

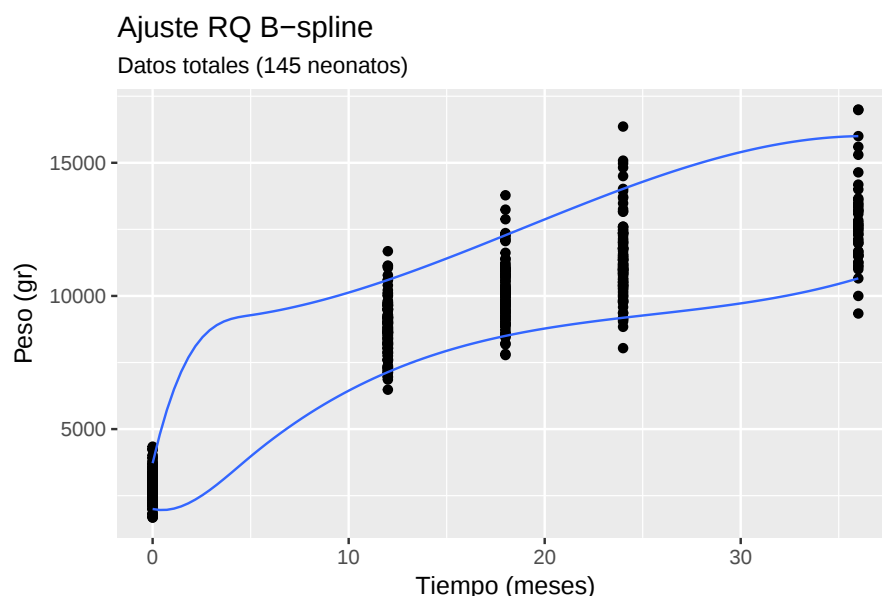


Figura 3-19: Regresión Cuantílica

Ahora, se observa la estimación de los percentiles 5 y 95 mediante la regresión cuantílica en la Figura **3-19**, donde se evidencia cierta semejanza con las curvas estimadas con la propuesta empírica ya mencionada (Figuras **3-17** y **3-18**), pues el percentil 95 sigue presentando una concavidad mayor a la del percentil 5, sin embargo las curvas percentílicas calculadas por QR parecen ser una mejor estimación.

En general el modelo se ajusta muy bien a los datos dado que se puede ver que la cantidad de observaciones por encima y por debajo de la curva de estimación son muy pocas, entonces si está representando a los percentiles 5 y 95.

3.5. Modificación de la variable Tiempo

Todos los resultados expuestos hasta el momento están dados para las cinco edades que se consideraron originalmente (0, 12, 18, 24, 36), ahora se realiza el análisis recalculando la variable edad (ver 2.3.5), teniendo así el peso en función de las edades reales en las que fueron medidos cada uno de los niños. De acuerdo a lo anterior, las figuras **3-20** y **3-21** muestra la estimación del individuo promedio con la variable edad cronológica en días para los 145 RN y para las 35 RN con las mediciones completas.

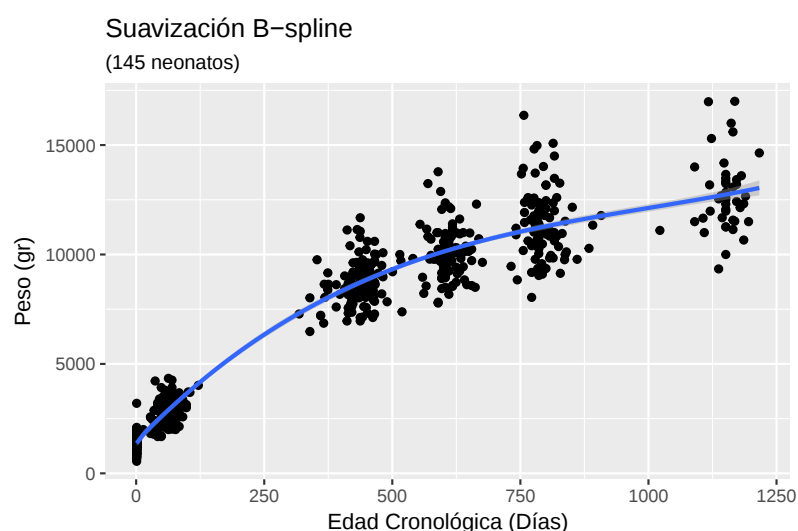


Figura 3-20: Estimación B-spline en días para los 145 RN

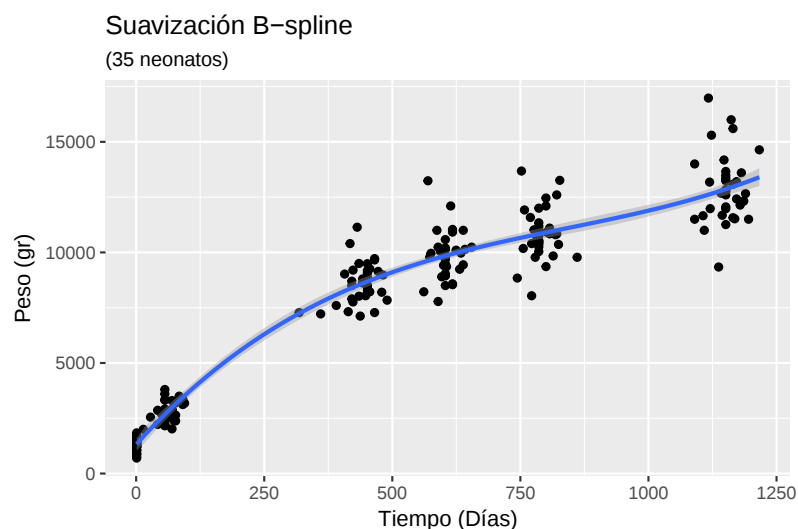
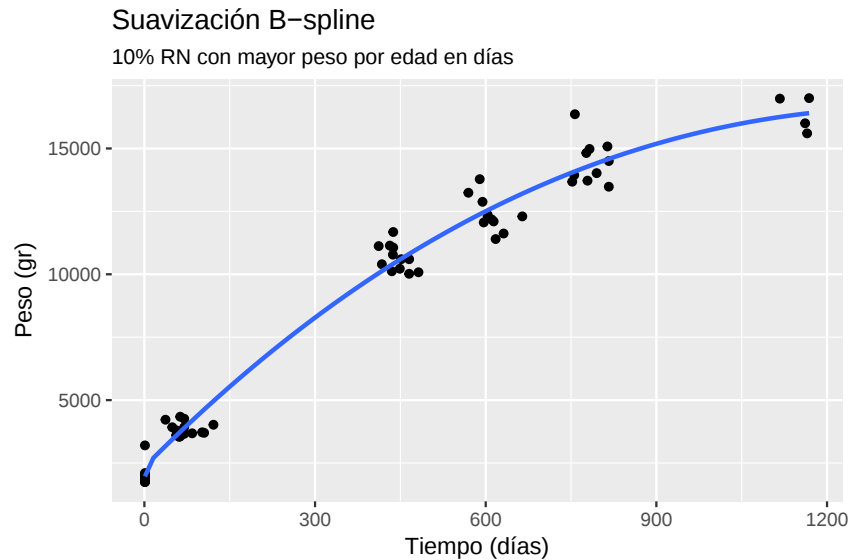


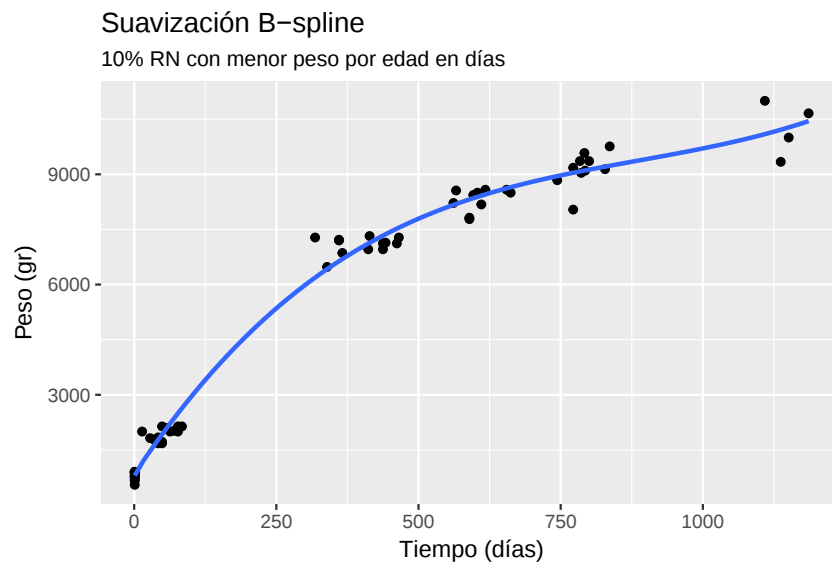
Figura 3-21: Estimación B-spline en días para los 35 RN

En general, hay un cambio notorio con respecto a la nube de puntos de la figura **3-7**, pues se aprecia que todas las observaciones no fueron realizadas en un mismo tiempo, además que mejora mucho mas la estimación B-spline disminuyendo el tamaño de las bandas de confianza. Donde se puede observar el gran cambio que tiene la curva al compararla con la curva promedio estimada para el peso en función de la edad en meses (Figura **3-7**), pues ahora con la variable edad los puntos están más dispersos formando nubes para cada grupo de edad, sin embargo esto se considera positivo, ya que hace referencia al tiempo real de medición, es decir, representa fielmente la evolución del peso de los niños a través del tiempo. Por lo tanto la curva promedio (Figuras **3-20** y **3-21**) es mucho más lisa que la anterior (**3-7**) siendo evidente que a pesar de que para el segundo grupo de edad (que está entre 14 y 122 días) y el tercer grupo hay un gran espacio de tiempo no afecta la forma de la curva, por lo que se considera como una influencia el haber agregado la medición del PN para todos los 145 RN como la primera medición del peso en el análisis que corresponde a al día 1.

Por otra parte, se realizó la estimación de los percentiles por el método aquí propuesto (Sección 2.3.4), encontrando diferencias en la figura 3.22(a) con la figura **3-17**, pues ahora la curva es mas lisa y, además, no presenta una concavidad muy pronunciada como anteriormente se evidenció, sólo al inicio de la curva se ve una pequeña elevación pero no es muy relevante. Ahora con respecto a la estimación del percentil 5 para la edad cronológica en días la figura 3.22(b) se observa que no hay diferencias muy marcadas con la estimación presentada en la figura **3-18**.



(a) Estimación para el percentil 95- Edad en días



(b) Estimación para el percentil 5- Edad en días

Figura 3-22: Estimación B-spline para los percentiles con la edad en días

Ahora bien, para la estimación de las curvas percentílicas del 5 y 95 % para los datos del peso con respecto a la edad cronológica en días por medio de Regresión cuantílica, se tienen unas curvas muy similares a las encontradas anteriormente Figura 3-22 por lo que el método propuesto no está alejado de la realidad y se puede considerar una buena estimación de estos cuantiles al tener gran similitud con la metodología de QR que es un método robusto.

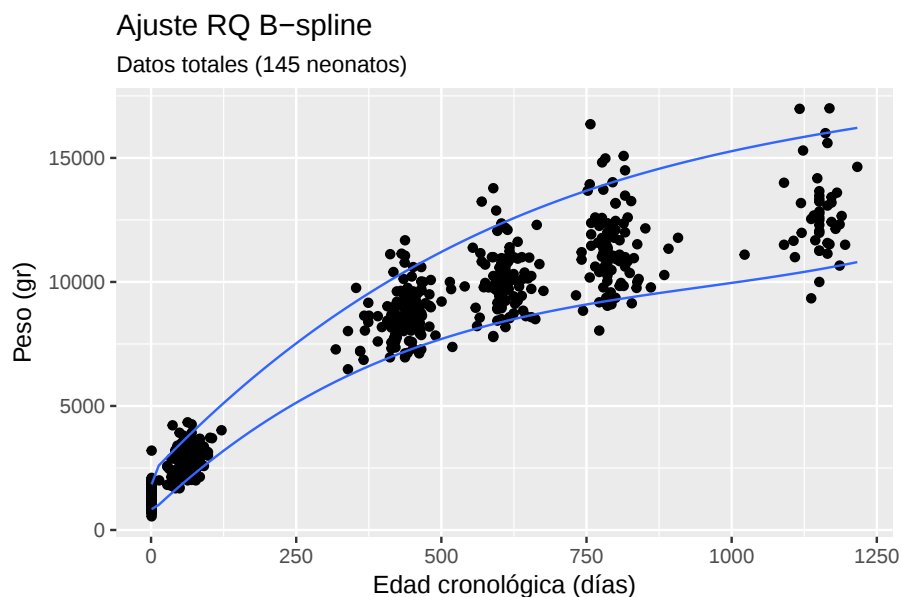


Figura 3-23: Regresión Cuantílica para edad cronológica

3.6. Mínimos cuadrados ponderados

Finalmente, se observa en la figura 3-24 la estimación de la curva con mínimos cuadrados ponderados, suavizada con B-Spline y tal como es de esperarse, por tratarse de estimación con MCP, la estimación contempla la variabilidad entre los niños de cada grupo de edad y las bandas de confianza evidencian esta variabilidad, ya que se aprecia que en las primeras mediciones hay muy poca variabilidad, por tanto las bandas son mas estrechas, por el contrario, dada la gran cantidad de información faltante en la última medición las bandas son mucho mas amplias, generando una mayor variabilidad. Ésta variabilidad tiene también mucho que ver con el comportamiento del peso de cada niño, pues en un inicio de sus vidas todos tienen comportamientos similares pero cada organismo se desarrolla diferente y luego de aproximadamente 36 a 38 meses cada niño presenta un comportamiento distinto entre ellos.

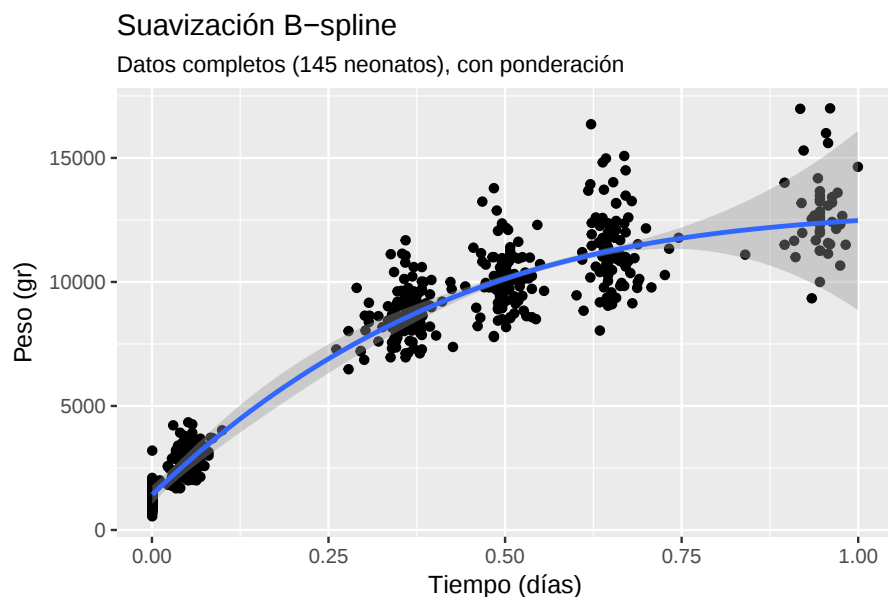


Figura 3-24: Suavización B-Spline con Mínimos Cuadrados Ponderados

3.7. Covarianza y correlación

Tal como se mencionó en la metodología, sobre la propuesta de incluir en el ajuste del modelo la matriz de varianzas y covarianzas para tratar de contrarrestar el problema de heterogeneidad en los datos, sin embargo al no ser posible se optó por agregar un vector de pesos ($1/\sigma_i$) $i = 1, \dots, 6$, el cual corresponde a un ajuste ponderado para cada grupo de edad, lo que se pudo constatar en la sección anterior.

La matriz de correlaciones **3-6** indica que los pesos al nacer y los pesos en el tiempo 0 no están correlacionados; pero los pesos de los niños de 12 meses y más sí están muy correlacionados, hasta parece que las correlaciones disminuyen a medida que las edades se diferencian más. Es decir, por ejemplo, la correlación entre los pesos a los 12 meses y a los 18 meses luce más alta que entre los 12 y los 36 meses. Por lo que estos resultados ratifican la importancia de incluir en el análisis la matriz de varianzas y covarianzas (Tabla **3-5**). Cuyo cálculo se realizó tomando como variable a los pesos de cada grupo de edad (seis variables por ser 6 grupos de edad) y por medio de las funciones *cor* y *cov* del software R se calcularon dichas matrices.

	PesoPN	Peso0	Peso12	Peso18	Peso24	Peso36
PesoPN	79845.97	5568.56	19579.65	-22952.74	-9898.45	19035.46
Peso0	5568.56	216811.85	88632.00	75800.79	64603.90	66641.34
Peso12	19579.65	88632.00	849728.24	878100.88	865637.65	1036450.00
Peso18	-22952.74	75800.79	878100.88	1165106.01	1119787.73	1372052.31
Peso24	-9898.45	64603.90	865637.65	1119787.73	1308919.66	1555449.58
Peso36	19035.46	66641.34	1036450.00	1372052.31	1555449.58	2365948.74

Tabla 3-5: Matriz de covarianzas

	PesoPN	Peso0	Peso12	Peso18	Peso24	Peso36
PesoPN	1.00	0.04	0.08	-0.08	-0.03	0.04
Peso0	0.04	1.00	0.21	0.15	0.12	0.09
Peso12	0.08	0.21	1.00	0.88	0.82	0.73
Peso18	-0.08	0.15	0.88	1.00	0.91	0.83
Peso24	-0.03	0.12	0.82	0.91	1.00	0.88
Peso36	0.04	0.09	0.73	0.83	0.88	1.00

Tabla 3-6: Matriz de Correlación

4 Conclusiones y Recomendaciones

4.1. Conclusiones

Acorde con el objetivo de proponer las curvas de crecimiento para los recién nacidos con bajo peso pertenecientes al plan madre canguro del Hospital Universitario del Valle durante el período 2002 a 2010, para ello se realizó un análisis exploratorio de los datos observando que el género no tiene influencia significativa sobre el peso al nacer de los niños, siendo el 60.69 % de los infantes en la muestra niñas y el 39.31 % niños y con respecto a las variables clínicas de la madre la única que presenta influencia sobre el BPN es la Hipertensión arterial, es decir que si la madre presenta HA durante el embarazo el niño tiene más posibilidades de nacer con BP.

Las estimaciones por medio de la regresión no paramétrica del individuo promedio, dieron mejores con la edad cronológica que con el conjunto de las 5 edades, pues la curva es mucho mas suave y las bandas de confianza presentan menor variabilidad, cabe aclarar que la estimación se realizó para los 145 RN y para los 35 RN con mediciones completas, encontrando que las estimaciones son muy parecidas a pesar de que disminuye en gran medida la cantidad de observaciones. Observando el resultado anterior, se llegó a la conclusión de que el comportamiento del peso de los niños no es monótono creciente, sobre todo en el primer año de edad. Ahora, comparando la media funcional de los 35 neonatos con la estimación del individuo promedio por medio de regresión no paramétrica, concluimos que la mejor estimación se presenta por medio de la regresión no paramétrica, pues es quien mejor representa la realidad del comportamiento del peso de los neonatos.

Tomando la edad en las 5 mediciones (0, 12, 18, 24, 36), se encontraron leves diferencias en la curva del percentil 5 entre el método de estimación propuesto en este trabajo (Sección 2.3.4), y la estimación por regresión cuantílica, se observó en esta última una leve fluctuación hacia abajo entre el tiempo 24 y 36, a diferencia de la curva con los percentiles donde la curva es mas lisa.

La realización de la transformación de la variable edad, fue un paso importante pues permitió ver y considerar mucho mejor la variabilidad en el comportamiento del peso de los RN. Al comparar el método que propusimos para estimar los percentiles 5 y 95 con la regresión cuantílica, se pudo observar que difieren muy poco, es decir, ambos métodos estiman de

manera similar los percentiles representando de esta manera la realidad del peso de los neonatos.

4.2. Recomendaciones

Se recomienda como trabajo con estos datos de crecimiento considerar la matriz de varianzas y covarianzas en el ajuste de regresión para contrarestar la heterogeneidad de varianza, y además como se comprobó en el trabajo existe correlación entre grupos de edad, siendo algo que no se puede ignorar.

Otra recomendación es aplicar más ampliamente la metodología de los Datos Funcionales a este problema, ya que por limitaciones con los datos faltantes no se pudo abordar de la mejor manera.

Bibliografía

- Active, L. (2016). Doctissimo. <http://www.doctissimo.com/es/terminos-y-privacidad/quienes-somos>. [Web; accedido el 26-11-2016].
- Aguilera Morillo, M. d. C. (2009). Estimación penalizada con datos funcionales. Universidad de Granada, Trabajo de Investigación.
- Alarcón, J., Alarcón, Y., Hering, E., and Buccioni, R. (2008). Curvas antropométricas de recién nacidos chilenos. *Revista Chilena de Pediatría*, 79(4):364–372.
- Amaya-Guío, J., Díaz-Cruz, L. A., Cardona-Ospina, A., Rodríguez-Merchán, D. M., Osorio-Sánchez, D., and Barrera-Barinas, A. (2013). Guía de práctica clínica para la prevención y detección temprana de las alteraciones del embarazo. *Revista Colombiana de Obstetricia y Ginecología*, 64(3):245–288.
- Animated Dissection of Anatomy for Medicine, A. (2011). Adam. <https://www.redbrickhealth.com/adam/Spanish{\%}20HIE{\%}20Multimedia/5/002367.htm>. [Web; accedido el 25-08-2017].
- Bermúdez, I. M., Andrade, M., and Torres, J. (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 (Col)*, 15(2):191–199.
- Biblioteca Nacional de Medicina, d. l. E. (2016). Medlineplus. <https://medlineplus.gov/spanish/>. [Web; accedido el 26-11-2016].
- Bowman, A. et al. (2010). Functional data analysis with r and matlab. *Journal of Statistical Software*, 34(1):1–2.
- Briceño, G., Durán, P., Colón, E., Line, D., Merker, A., Abad, V., Chahín, S., Del Toro, K., Matallana, A., Llano, M., et al. (2012). Protocolo del estudio para establecer estándares normativos de crecimiento de niños colombianos sanos. *Pediatría*, 45(4):235–242.
- Cancino, E., Fontibón, M., León, H., El Tunal, M., Otálora, R., Suba, M., and Ebert Pérez González, M. (2015). Secretaría Distrital de Salud de Bogotá, DC Asociación Bogotana de Obstetricia y Ginecología (ASBOG). *Guía de control prenatal y factores de riesgo*.

- Castaño, Y. S. (2015). Modelación del crecimiento en neonatos con situación de Bajo peso nacidos en el HUV (2009-2015). Universidad del Valle, Tesis de pregrado.
- Cook, D. E., Shelov, S. P., García-Fornel Florensa, S., Grimes, M. T., Hanemmann, R. E., Pérez, A., Reisberg, L. R., Suchyta, R. F., et al. (2001). *El cuidado de su hijo pequeño: desde que nace hasta los cinco años*. Number 616-053.2. Academia Americana de pedatría,.
- Correa, J. C., Iral, R., and Rojas, L. (2006). Estudio de potencia de pruebas de homogeneidad de varianza. *Revista colombiana de estadística*, 29(1):57–76.
- DAM, C. (2016). Clinica dam, edad gestacional. <http://www.sindominio.net/ash>. [Web; accedido el 26-11-2016].
- DANE (2009). Estadísticas vitales. nacimientos por peso al nacer. https://www.dane.gov.co/index.php?option=com_content&task=category§ionid=16&id=393&Itemid=889. [Web; accedido el 25-11-2016].
- Davino, C., Furno, M., and Vistocco, D. (2013). *Quantile regression: theory and applications*, volume 988. John Wiley & Sons. Italy.
- de la Salud, O. P. (1995). *Clasificación estadística internacional de enfermedades y problemas relacionados con la salud*. Number v. 1 in Publicación Científica. Organización Panamericana de la Salud.
- Durbán, M. (2006). Métodos de suavizado eficientes con p-splines. http://ciencias.bogota.unal.edu.co/fileadmin/content/eventos/simposioestadistica/documentos/memorias/MEMORIAS_2015/Cursillos/Durban_Metodos_Suavizado_P-spline.pdf.
- Faneite, P., Rivera, C., Amato, R., Faneite, J., Paradas, M., et al. (2011). ¿Tiene importancia el Bajo peso neonatal? *Revista de Obstetricia y Ginecología de Venezuela*, 71(3):151–7.
- Faraway, J. (2006). Extending the linear model with r: Generalized linear, mixed effects and nonparametric regression models.
- Gabriel, M. M., Moreiras, J. M., Fleixas, G. L., Gallego, S. D., Alonso, C. P., De la Cruz Bértolo, J., and Estévez, E. P. (2006). Valoración del test de ballard en la determinación de la edad gestacional. In *Anales de Pediatría*, volume 64, pages 140–145. Elsevier.
- Gaitán, H. and Delgado, M. (2016). El bajo peso al nacer: otro ejemplo de inequidad en Colombia. *Revista Colombiana de Obstetricia y Ginecología*, 60(2):121–123.
- Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

- Hao, L. and Naiman, D. Q. (2007). *Quantile regression*. Number 149. Sage.
- López, J. I., Lugones Botell, M., Mantecón Echevarría, S. M., González Pérez, C., and Pérez Valdés-Dapena, D. (2012). Algunos factores de riesgo relacionados con el bajo peso al nacer. *Revista Cubana de Obstetricia y Ginecología*, 38(1):45–55.
- Marin Iranzo, R., Gorostidi Pérez, M., and Álvarez-Navascués, R. (2014). Hipertensión arterial y embarazo. <http://www.revistanefrologia.com/es-publicacion-nefroplus-articulo-hipertension-arterial-embarazo-X1888970011001019>. [Web; accedido el 05-05-2018].
- Montgomery, D., Peck, E., and Vining, G. (2006). Introducción al análisis de regresión lineal. *Editorial Continental, Tercera edición, México*.
- Montoya-Restrepo, N. E. and Correa-Morales, J. C. (2007). Curvas de peso al nacer. *Revista de Salud Pública*, 9(1):1–10.
- Nacho Rios, E. G. (2012). Pediatría. <http://www.etableros.com/clinicas/clinicas/hernan/files/15.{\}-Pediatría.pdf>. [Web; accedido el 13-04-2018].
- Observatorio Así Vamos, e. S. (2016). Prevalencia de bajo peso al nacer georeferenciado. <https://www.asivamosensalud.org/indicadores/seguridad-alimentaria-y-nutricional/prevalencia-de-bajo-peso-al-nacer-georeferenciado>. [Web; accedido el 04-09-2018].
- of Family Physicians, A. A. (2014). Hipertensión inducida por el embarazo. <https://es.familydoctor.org/condicion/hipertension-inducida-por-el-embarazo/{\}?adfree{\}=true>. [Web; accedido el 05-05-2018].
- Olaya, J. (2012). *Métodos de Regresión No Paramétrica*. Universidad del Valle.
- OMS, O. M. d. I. S. (2017). Metas mundiales de nutrición 2025: documento normativo sobre bajo peso al nacer. Technical report, Organización Mundial de la Salud.
- Organización Mundial de la Salud, O. (2005). *Estadísticas sanitarias mundiales 2005*. Organización Mundial de la Salud.
- Petro, G., Bustamante, M., Martínez, E., Peña, C., Cabrera, T., et al. (2014). Guía de trastornos hipertensivos del embarazo. *Bogotá: Alcaldía Mayor de Bogotá*.
- Pinheiro, P. (2018). Hipertensión en el embarazo - causas, síntomas y tratamiento. <https://www.mdsau.de.com/es/2017/05/hipertension-embarazo.html>. [Web; accedido el 05-05-2018].

- Pinzón Villate, G., Hernández, V., Eddison, J., Márquez Beltrán, M. F. R., Quiroga Villalobos, E. F., et al. (2013). Análisis del bajo peso al nacer en Colombia 2005-2009. *Journal of Public Health*; Vol. 15, núm. 4 (2013); 577-588 *Revista de Salud Pública*; Vol. 15, núm. 4 (2013); 577-588 0124-0064.
- PulsoMed, S. (2016). Tuotromedico. http://www.tuotromedico.com/temas/tablas{_}crecimiento.htm. [Web; accedido el 26-11-2016].
- Ramírez, R. F. (2001). *Catalogación del recién nacido*. Servicio neonatología hospital clínico universidad de chile.
- Ramsey, J. and Silverman, B. W. (2002). Applied functional data analysis: methods and case studies.
- Serrano, A. d. R. T. (2002). Crecimiento y desarrollo. *Revista Mexicana de Medicina Física y Rehabilitación*, 14(2-4):54–57.
- UNICEF (2004). *Low birthweight: country, regional and global estimates*. Fondo de las Naciones Unidas para la Infancia.
- Unicef et al. (2016). Estado mundial de la infancia 2016: Una oportunidad para cada niño.
- Universidad de Alicante, R. I. (2018). Tema 2: Heterocedasticidad. https://rua.ua.es/dspace/bitstream/10045/18469/24/tema2_2010-11.pdf. [Web; accedido el 05-05-2018].
- Vaca Romero, J. E. (2013). Incidencia de patologías en neonatología del hospital teófilo dávila, cantón machala desde diciembre 2012 a julio 2013. B.S. thesis, Machala: Universidad Técnica de Machala.
- Vicéns, J. and Sánchez, B. (2012). *Regresión cuantílica: Estimación y contrastes*. Universidad Autónoma de Madrid.
- Voinea, G. C. (2014). Análisis de datos funcionales. <http://estadisticadeaaz.blogspot.com.co/2014/03/analisis-de-datos-funcionales.html>. Consultado: 28/10/2016.
- Wei, Y., Pere, A., Koenker, R., and He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in medicine*, 25(8):1369–1382.