



Universidad
Zaragoza

Trabajo Fin de Grado

Construcción de un modelo de crecimiento fetal
durante la gestación

Fetal Growth model construction during
pregnancy

Autor

Natalia Jiménez Calvo

Director

Luis Mariano Esteban Escaño

Escuela Universitaria Politécnica La Almunia

Junio 2025

Página intencionadamente en blanco.



**Escuela Universitaria
Politécnica - La Almunia**
Centro adscrito
Universidad Zaragoza

**ESCUELA UNIVERSITARIA POLITÉCNICA
DE LA ALMUNIA DE DOÑA GODINA (ZARAGOZA)**

MEMORIA

**Construcción de un modelo de crecimiento fetal
durante la gestación**

**Fetal Growth model construction during
pregnancy**

Identificador

Autor: Natalia Jiménez Calvo

Director: Luis Mariano Esteban Escaño

Fecha: 06/ 2025

Página intencionadamente en blanco.

INDICE DE CONTENIDO BREVE

1. RESUMEN	1
2. ABSTRACT	3
3. INTRODUCCIÓN	5
4. DESARROLLO	7
5. RESULTADOS	38
6. CONCLUSIONES	71
7. OBJETIVOS DE DESARROLLO SOSTENIBLE	73
8. BIBLIOGRAFÍA	75

INDICE DE CONTENIDO

1. RESUMEN	1
1.1. PALABRAS CLAVE	2
2. ABSTRACT	3
2.1. KEY WORDS	3
3. INTRODUCCIÓN	5
4. DESARROLLO	7
4.1. REVISIÓN DE LA METODOLOGÍA Y LITERATURA EN ESTUDIOS PREVIOS	7
4.2. ANÁLISIS DE LOS DATOS USADOS	10
4.2.1. Población	10
4.2.2. Base de Datos ecográfica y de partos	11
4.2.2.1. Datos de las distintas ecografías	11
4.2.2.2. Datos de los distintos partos	12
4.3. PREPROCESAMIENTO DE LOS DATOS	13
4.3.1. Limpieza	13
4.3.1.1. Limpieza base de datos ecografía	13
4.3.1.2. Limpieza base de datos embarazo	15
4.3.1.3. Dataset final	17

INDICES

4.3.1.3.1. Preparación de los datos en parto	17
4.3.1.3.2. Preparación de los datos en ecografía	18
4.3.1.3.3. Df_final	19
4.3.1.4. Depuración de datos por análisis de outliers semanales	21
4.4. METODOLOGÍA UTILIZADA EN ESTE ESTUDIO	29
4.4.1. <i>Modelo de crecimiento</i>	29
4.4.2. <i>Que son los modelos mixtos</i>	29
4.4.3. <i>Como implementamos los modelos mixtos en nuestro trabajo</i>	30
4.5. VARIANZA CONSTANTE O VARIABLE	32
4.5.1. <i>Suposición habitual de varianza constante</i>	32
4.5.2. <i>Porque la varianza cambia con la edad gestacional</i>	33
4.6. PESO POR ECOGRAFÍA.	33
4.6.1. <i>¿Por qué no es exacto?</i>	34
4.6.2. <i>Base de los cálculos</i>	34
4.6.2.1. FORMULAS DE HADLOCK	34
4.6.2.2. Intergrowth-21st	35
4.7. CONTROL DE PESO	35
4.7.1. <i>¿Qué son los percentiles fetales?</i>	35
4.8. MODELO POR GÉNERO	36
5. RESULTADOS	38
5.1. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES	38
5.1.1. <i>Variable Continuas</i>	38
5.1.2. <i>VARIABLES CATEGÓRICAS</i>	38
5.2. MODELO DE PESO HADLOCK PARA FETOS VARONES	39
5.2.1. <i>Modelo sin características de la madre</i>	39
5.2.2. <i>Modelo con característica de la madre</i>	44
5.3. MODELO DE PESO STIMERMAN PARA FETOS VARONES	47
5.3.1. <i>Modelo sin características de la madre</i>	47
5.3.2. <i>Modelo con características de la madre</i>	51
5.4. MODELO DE PESO HADLOCK PARA FETOS FEMENINOS	54
5.4.1. <i>Modelo sin características de la madre</i>	54
5.4.2. <i>Modelo con características de la madre</i>	58
5.5. MODELO DE PESO STIMERMAN PARA FETOS FEMENINOS	61
5.5.1. <i>Modelo sin características de la madre</i>	61
5.5.2. <i>Modelo con características de la madre</i>	65
5.6. CONCLUSIÓN DE LOS RESULTADOS	68

6. CONCLUSIONES	71
7. OBJETIVOS DE DESARROLLO SOSTENIBLE	73
8. BIBLIOGRAFÍA	75

INDICE DE ILUSTRACIONES

Ilustración 1, Gráfico de la dispersión de los pesos del feto a lo largo del tiempo	6
Ilustración 2, Curvas de crecimiento fetal estimadas mediante modelo mixto creado: percentiles 10, 50 y 90 del peso en función de la edad gestacional, junto con las observaciones reales.....	9
Ilustración 3, Datos iniciales de los pesos fetales estimados a lo largo de los días de la gestación	20
Ilustración 4, Datos sin outliers de los pesos fetales estimados a lo largo de los días de la gestación	23
Ilustración 5, Resultados del modelo mixto lineal sin variables maternas para fetos masculinos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock en función de los días de gestación.....	40
Ilustración 6, Porcentaje de observaciones del peso fetal estimado mediante la fórmula de Hadlock fuera del rango interpercentil [P10, P90] en los conjuntos de entrenamiento y test del modelo mixto sin variables maternas, ajustado para fetos masculinos.	42
Ilustración 7, Ajuste del modelo mixto sin variables maternas sobre el conjunto de entrenamiento: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Hadlock para fetos masculinos, junto con observaciones reales.	43
Ilustración 8, Ajuste del modelo mixto sin variables maternas sobre el conjunto de test: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Hadlock para fetos masculinos, junto con observaciones reales.	43

INDICES

Ilustración 9, Resultados del modelo mixto lineal con variables maternas para fetos masculinos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock en función de los días de gestación.	45
Ilustración 10, Curva de crecimiento fetal estimado para un feto masculino, generada para una madre específica según sus características clínicas, con percentiles 10, 50 y 90 obtenidos mediante modelo mixto y pesos estimados por la fórmula de Hadlock superpuestos.	47
Ilustración 11, Resultados del modelo mixto lineal sin variables maternas para fetos masculinos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Stimerman en función de los días de gestación.	48
Ilustración 12, Porcentaje de observaciones del peso fetal estimado mediante la fórmula de Stimerman fuera del rango interpercentil [P10, P90] en los conjuntos de entrenamiento y test del modelo mixto sin variables maternas, ajustado para fetos masculinos.	49
Ilustración 13, Ajuste del modelo mixto sin variables maternas sobre el conjunto de entrenamiento: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Stimerman para fetos masculinos, junto con observaciones reales.	50
Ilustración 14, Ajuste del modelo mixto sin variables maternas sobre el conjunto de test: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Stimerman para fetos masculinos, junto con observaciones reales.	50
Ilustración 15, Resultados del modelo mixto lineal con variables maternas para fetos masculinos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Stimerman en función de los días de gestación.	52
Ilustración 16, Curva de crecimiento fetal estimado para un feto masculino, generada para una madre específica según sus características clínicas, con percentiles 10, 50 y 90 obtenidos mediante modelo mixto y pesos estimados por la fórmula de Stimerman superpuestos.	53
Ilustración 17, Resultados del modelo mixto lineal sin variables maternas para fetos femeninos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock en función de los días de gestación.	55
Ilustración 18, Porcentaje de observaciones del peso fetal estimado mediante la fórmula de Hadlock fuera del rango interpercentil [P10, P90] en los conjuntos de	

entrenamiento y test del modelo mixto sin variables maternas, ajustado para fetos femeninos.56

Ilustración 19, Ajuste del modelo mixto sin variables maternas sobre el conjunto de entrenamiento: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Hadlock para fetos femeninos, junto con observaciones reales.57

Ilustración 20, Ajuste del modelo mixto sin variables maternas sobre el conjunto de test: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Hadlock para fetos femeninos, junto con observaciones reales.57

Ilustración 21, Resultados del modelo mixto lineal con variables maternas para fetos femeninos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock en función de los días de gestación.59

Ilustración 22, Curva de crecimiento fetal estimado para un feto femenino, generada para una madre específica según sus características clínicas, con percentiles 10, 50 y 90 obtenidos mediante modelo mixto y pesos estimados por la fórmula de Hadlock superpuestos.60

Ilustración 23, Resultados del modelo mixto lineal sin variables maternas para fetos femeninos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Stimerman en función de los días de gestación.62

Ilustración 24, Porcentaje de observaciones del peso fetal estimado mediante la fórmula de Stimerman fuera del rango interpercentil [P10, P90] en los conjuntos de entrenamiento y test del modelo mixto sin variables maternas, ajustado para fetos femeninos.63

Ilustración 25, Ajuste del modelo mixto sin variables maternas sobre el conjunto de entrenamiento: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Stimerman para fetos femeninos, junto con observaciones reales.64

Ilustración 26, Ajuste del modelo mixto sin variables maternas sobre el conjunto de test: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Stimerman para fetos femeninos, junto con observaciones reales.64

Ilustración 27, Resultados del modelo mixto lineal con variables maternas para fetos femeninos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Stimerman en función de los días de gestación.66

Ilustración 28, Curva de crecimiento fetal estimado para un feto femenino, generada para una madre específica según sus características clínicas, con percentiles

10, 50 y 90 obtenidos mediante modelo mixto y pesos estimados por la fórmula de Stimerman superpuestos..... 68

INDICE DE TABLAS

Tabla 1- Población	10
Tabla 2, Variables numéricas consideradas en el modelo y sus resultados estadísticos.....	38
Tabla 3, Número de observaciones por sexo fetal	39
Tabla 4, Distribución de observaciones según las categorías de la variable FIV (Fecundación In Vitro).....	39
Tabla 5, Distribución de observaciones según las categorías de la variable Hipertensión	39
Tabla 6, Distribución de observaciones según las categorías de la variable Diabetes	39
Tabla 7,Distribución de las mujeres Fumadoras	39
Tabla 8, Porcentaje de fetos clasificados como SGA (< P10) y LGA (> P90) según fórmula (Hadlock o Stirnemann), sexo fetal y conjunto de datos (entrenamiento o test).	69

INDICE DE ...

Incluir los diferentes índices de elementos (diferentes a ilustraciones y tablas) que se hayan utilizado en el documento.

Estos elementos se pueden crear con los rótulos predeterminados o con otros rótulos personalizados, por ejemplo: Fotografía, Plano, Esquema, etc



1. RESUMEN

En este trabajo fin de grado en Ingeniería de Datos en procesos industriales, se ha llevado a cabo un estudio del crecimiento fetal a lo largo de la edad gestacional mediante modelos lineales mixtos. Este tipo de modelización ha sido objeto de diversos enfoques, como la regresión simple, la regresión por cuantiles y los modelos mixtos, siendo estos últimos los más aceptados desde la publicación del modelo Intergrowth-21st.

El estudio se ha desarrollado en colaboración con el Hospital Universitario Miguel Servet, utilizando una base de datos real compuesta por información ecográfica longitudinal y variables clínicas maternas. El objetivo principal ha sido construir modelos predictivos del peso fetal, ajustando curvas de crecimiento personalizadas que permitan identificar desviaciones del desarrollo intrauterino.

La metodología incluye una exhaustiva limpieza y preprocessamiento de los datos, la creación de modelos mixtos diferenciados por sexo fetal y la evaluación de su rendimiento tanto en datos de entrenamiento como de prueba. Además, se ha cuestionado el supuesto clásico de **varianza constante**, demostrando que la **variabilidad del peso fetal aumenta con la edad gestacional pero no de una forma constante**, por lo que se ha optado por un enfoque que considera esta heterogeneidad, mejorando así la estimación de percentiles clínicamente relevantes.

Se han comparado modelos basados en dos fórmulas de estimación del peso fetal (Hadlock y Stirnemann), observando que el modelo basado en Hadlock ofrece un mejor ajuste general. A nivel práctico, los modelos alcanzan una cobertura de entre el **81 % y el 87 %** de los valores reales dentro del intervalo comprendido entre los percentiles 10 y 90, lo que implica que entre el **13 % y el 19 % de los casos** corresponden a fetos clasificados como **PEG** (por debajo del P10) o **GEG** (por encima del P90), cifras clínicamente consistentes.

En conclusión, este trabajo demuestra la utilidad de los modelos mixtos para el seguimiento personalizado del crecimiento fetal,

integrando datos clínicos relevantes y adaptándose a la variabilidad natural del desarrollo intrauterino.

1.1. PALABRAS CLAVE

- Peso de Stimerman
- Peso de Hadlock
- Modelos lineales mixtos
- Modelos de crecimiento fetal

2. ABSTRACT

In this Bachelor's Thesis in Data Engineering for Industrial Processes, a study of fetal growth throughout gestational age has been carried out using linear mixed models. This type of modeling has been approached through various methods, such as simple regression, quantile regression, and mixed models, with the latter being the most widely accepted since the publication of the Intergrowth-21st model.

The study was conducted in collaboration with the Miguel Servet University Hospital, using a real database composed of longitudinal ultrasound information and maternal clinical variables. The main objective was to build predictive models of fetal weight, adjusting personalized growth curves that allow the identification of deviations in intrauterine development.

The methodology includes thorough data cleaning and preprocessing, the creation of mixed models differentiated by fetal sex, and the evaluation of their performance on both training and test data. Furthermore, the classical assumption of constant variance was questioned, showing that the variability of fetal weight increases with gestational age, but not in a constant way. Therefore, an approach that considers this heterogeneity was adopted, thereby improving the estimation of clinically relevant percentiles.

Models based on two fetal weight estimation formulas (Hadlock and Stirnemann) were compared, showing that the model based on Hadlock provides a better overall fit. In practical terms, the models achieve a coverage of between 81% and 87% of actual values within the interval between the 10th and 90th percentiles, meaning that between 13% and 19% of the cases correspond to fetuses classified as SGA (below the 10th percentile) or LGA (above the 90th percentile), which are clinically consistent figures.

In conclusion, this work demonstrates the usefulness of mixed models for personalized monitoring of fetal growth, integrating relevant clinical data and adapting to the natural variability of intrauterine development.

2.1. KEY WORDS

- Stimerman weight
- Hadlock weight

- Mixed linear models
- Fetal growth models

3. INTRODUCCIÓN

Actualmente, existen numerosos estudios que analizan cómo influyen el estilo de vida y ciertos factores genéticos de la madre en el desarrollo del feto. Durante la gestación, una de las principales referencias que tienen los profesionales médicos para evaluar la salud del bebé es su peso estimado. Sin embargo, conocer el peso real del feto no es posible hasta el momento del parto, lo que ha llevado al desarrollo de métodos de estimación a partir de mediciones ecográficas.

El objetivo de este trabajo es desarrollar un modelo de crecimiento fetal empleando un modelo mixto que permita predecir el crecimiento fetal a lo largo del embarazo, teniendo en cuenta diversas características maternas. Esta herramienta podría facilitar un seguimiento más preciso de la salud del feto, ayudando a detectar posibles complicaciones de forma anticipada.

Para llevar a cabo este proyecto, se registraron datos de embarazos en el Hospital Miguel Servet durante los últimos años. Estos datos incluyen información detallada sobre ecografías, partos y variables maternas. En este estudio se ha realizado una cuidadosa selección de variables relevantes para el desarrollo fetal, tras un proceso de revisión bibliográfica y validación clínica. Además, se ha llevado a cabo un exhaustivo trabajo de limpieza del conjunto de datos para garantizar la calidad de la información, eliminando valores atípicos y registros incompletos.

Introducción

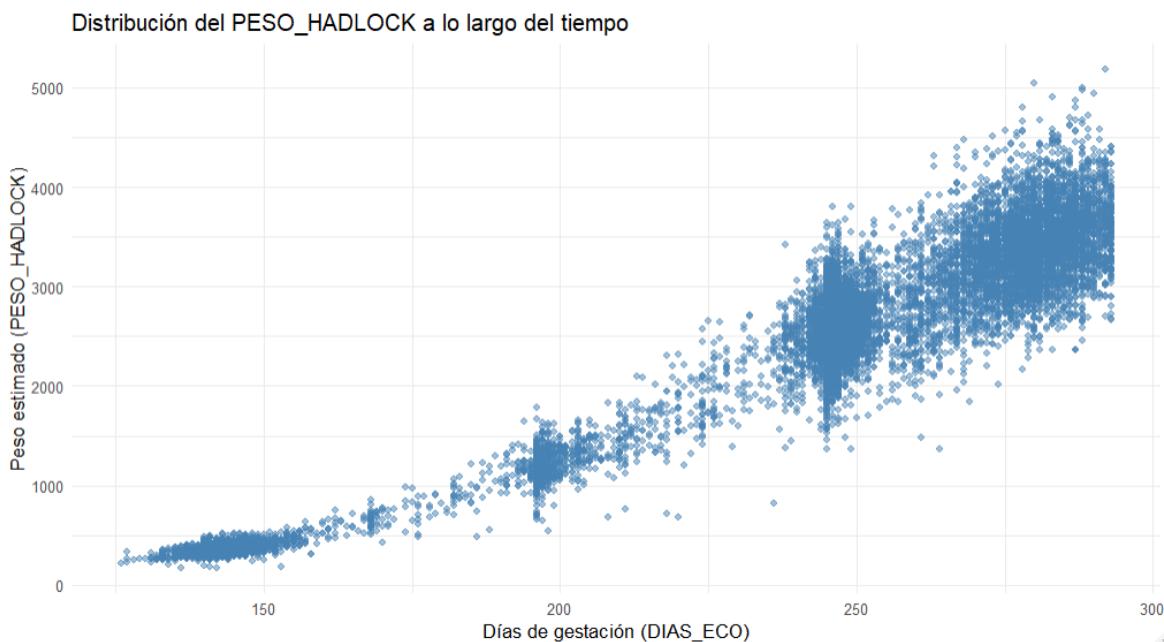


Ilustración 1, Gráfico de la dispersión de los pesos del feto a lo largo del tiempo

Este proyecto se centra exclusivamente en embarazos únicos, excluyendo gestaciones múltiples debido a las diferencias significativas en el desarrollo y el peso fetal en estos casos. La finalidad última del trabajo es que este modelo pueda aplicarse en un entorno clínico real, contribuyendo a mejorar la atención prenatal.

Existen diferentes métodos para estimar el peso fetal a partir de mediciones ecográficas como el diámetro biparietal, la circunferencia abdominal o la longitud del fémur. Estas estimaciones, combinadas con el uso de percentiles, permiten clasificar a los fetos según su adecuación al crecimiento esperado para cada edad gestacional, facilitando así un seguimiento individualizado y la toma de decisiones clínicas. En este trabajo, se emplean modelos mixtos como técnica principal de modelización, dado su potencial para manejar datos longitudinales y capturar tanto la evolución del crecimiento fetal como la variabilidad entre embarazos.

4. DESARROLLO

Dividido en epígrafes. En este apartado, el más amplio, se expondrá el cuerpo fundamental del trabajo y se argumentarán las ideas principales y secundarias del mismo.

Detalla el proceso que se ha llevado a cabo para la elaboración del TFG y la metodología adoptada dependiendo de la naturaleza del TFG escogido (contenido teórico, caso práctico, proyecto, estudio técnico, revisión bibliográfica y estado de la cuestión, justificación, resultados...), pero en todo caso debe estar dividido en capítulos homogéneos y numerados.

4.1. REVISIÓN DE LA METODOLOGÍA Y LITERATURA EN ESTUDIOS PREVIOS

El estudio del crecimiento fetal ha sido abordado desde múltiples disciplinas a lo largo de las últimas décadas, dado su papel fundamental en la evaluación de la salud prenatal y en la detección precoz de complicaciones obstétricas. Uno de los principales retos metodológicos en este campo es la construcción de modelos que permitan estimar de forma precisa el desarrollo del feto a lo largo de la gestación, considerando tanto la media como la variabilidad del crecimiento, y adaptándose a diferentes contextos clínicos y poblacionales.

A nivel internacional, uno de los referentes más importantes es el proyecto INTERGROWTH-21st(Papageorghiou et al., 2014), desarrollado por un consorcio global con el objetivo de establecer estándares prescriptivos del crecimiento fetal. A diferencia de los modelos tradicionales, basados en referencias locales, INTERGROWTH-21st se fundamenta en la idea de que, en condiciones óptimas de salud y nutrición, los fetos presentan patrones de crecimiento muy similares en todo el mundo. Para ello, se reclutó a más de 4.500 mujeres embarazadas sanas de ocho países distintos, las cuales fueron monitorizadas desde la semana 14 hasta el parto mediante ecografías seriadas estandarizadas

Una de las principales contribuciones metodológicas de INTERGROWTH-21st fue la adopción de modelos estadísticos avanzados que permiten representar de forma precisa la evolución del crecimiento fetal. Entre ellos, destacan los modelos de tipo GAM (Generalized Additive Models), que permiten un ajuste no lineal de las variables, ya

sea mediante splines o funciones polinómicas fraccionarias. Este enfoque facilita la construcción de curvas de crecimiento basadas en la media del peso fetal estimado en función de la edad gestacional, sobre las cuales se calculan bandas de confianza que definen los percentiles (por ejemplo, P10 y P90) utilizando la desviación estándar del modelo. Junto a esto, se aplicaron técnicas como el enfoque LMS (Lambda-Mu-Sigma) y sus extensiones LMST y LMSP, que permiten modelar no solo la media, sino también la variabilidad y la asimetría de la distribución del peso fetal. Estas metodologías, ampliamente documentadas por Ohuma con la colaboración de otros investigadores (Ohuma et al., 2019), son especialmente útiles en contextos con datos longitudinales o distribuciones heterogéneas, y han sido validadas mediante estrictos procedimientos de control de calidad.

Además, se ha demostrado que los modelos de referencia tradicionales —muchos de ellos diseñados con muestras locales y criterios menos exigentes— presentan importantes limitaciones metodológicas. Entre ellas destacan el uso de poblaciones no representativas, la falta de seguimiento longitudinal, y la utilización de fórmulas antiguas con escasa validez clínica actual. La mayoría de estas curvas funcionan como referencias (descriptivas) más que como estándares (prescriptivos), es decir, describen cómo han crecido los fetos en un contexto determinado, pero no necesariamente cómo deberían crecer en condiciones óptimas.

En el ámbito nacional y regional también existen propuestas interesantes. Por ejemplo, el estudio de Savirón-Cornudella (Savirón-Cornudella et al., 2018), realizado en el Hospital Universitario Miguel Servet, propone curvas específicas para embarazos gemelares, diferenciadas por corionicidad. Este trabajo, aunque centrado en gestaciones múltiples, ilustra la relevancia de adaptar los modelos al contexto poblacional concreto, utilizando datos locales y técnicas modernas como modelos mixtos con splines cúbicos y transformación logarítmica de las medidas ecográficas. En general, esta estrategia se alinea con el enfoque de los modelos personalizados o customizados, que depende de las características materno fetales (como peso, talla o paridad), en contraste con los modelos no customizados, que se basan únicamente en la edad gestacional. La elección entre uno u otro enfoque depende del objetivo clínico y del tipo de población a la que se quiere aplicar el modelo.

Todos estos estudios coinciden en la importancia de seleccionar cuidadosamente las variables que influyen en el crecimiento fetal, como el peso y altura maternos, el uso de fecundación in vitro, la edad gestacional, y otras características clínicas. Sin embargo, hay

diferencias relevantes en la metodología: mientras que INTERGROWTH-21st propone modelos internacionales aplicables a cualquier país con buenas condiciones sanitarias, otros estudios abogan por desarrollar curvas personalizadas, adaptadas a las características demográficas y clínicas locales.

En este contexto, el presente trabajo se alinea con la perspectiva de emplear **modelos mixtos** como herramienta flexible para modelar el crecimiento fetal en una población concreta (Zaragoza), utilizando datos longitudinales obtenidos en un hospital público. Estos modelos permiten incorporar tanto efectos fijos (como variables clínicas maternas) como efectos aleatorios (variación entre individuos), lo que los convierte en una opción metodológica robusta y versátil. Además, permiten evaluar si la varianza del crecimiento cambia con la edad gestacional, lo cual es fundamental para construir percentiles que reflejen fielmente la variabilidad real observada.

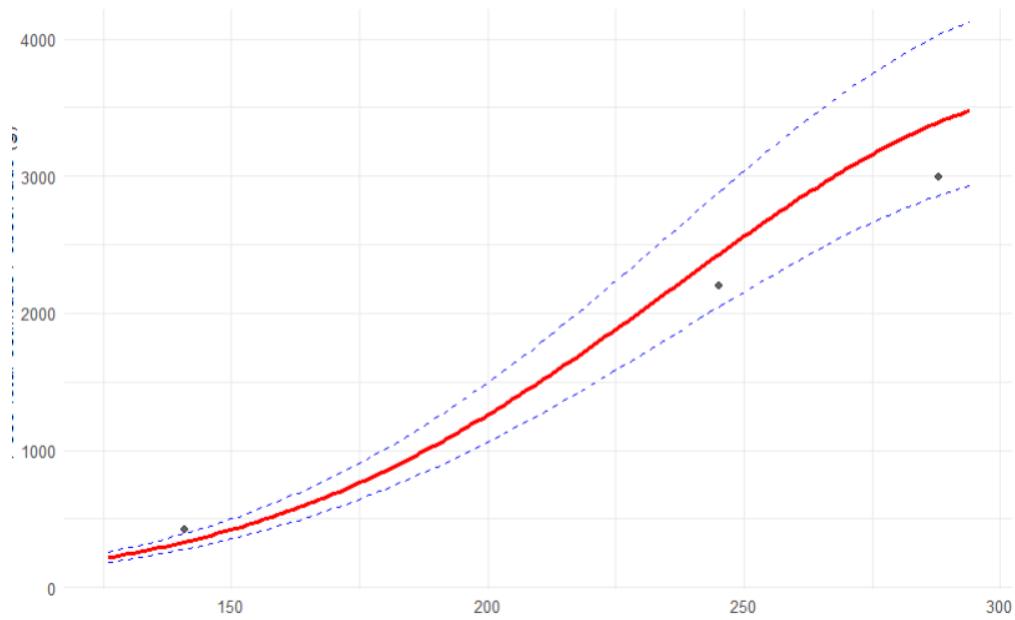


Ilustración 2, Curvas de crecimiento fetal estimadas mediante modelo mixto creado: percentiles 10, 50 y 90 del peso en función de la edad gestacional, junto con las observaciones reales.

En este sentido, se adopta una aproximación similar a la empleada por el proyecto INTERGROWTH-21st, pero adaptada a las características demográficas y clínicas específicas de la población local, asegurando así una mayor adecuación clínica y epidemiológica de las curvas generadas.

En los apartados siguientes se presentará con mayor detalle el conjunto de datos utilizado, las transformaciones aplicadas y la implementación del modelo mixto que permite integrar estas ideas en un marco metodológico coherente y clínicamente útil.

4.2. ANÁLISIS DE LOS DATOS USADOS

Siendo el estudio principal del TFG determinar qué factores afectan al peso del feto durante la gestación, la construcción de nuestro modelo mixto pasa en primer lugar por la obtención de un número suficiente de datos que nos permita crear un modelo que cuantifique la importancia de cada variable que puede participar en la predicción del peso. En la obtención de datos de este trabajo de fin de grado ha sido clave la colaboración del Hospital Universitario Miguel Servet que ha aportado 2 bases de datos con los datos de las ecografías y del parto. Por lo que al contar con datos reales podemos crear un modelo mas fiable y de mayor calidad.

4.2.1. Población

Antes de hablar de nuestros distintos datasets, considero que es importante conocer la población con la que se está trabajando y como puede afectar a nuestro estudio.

Como se ha mencionado anteriormente, la base de datos cuenta con registros de embarazos llevados en el hospital público Miguel Servet, situado en la capital aragonesa, Zaragoza.

A fin de contextualizar la evolución de los embarazos en la región, se incluye a continuación una tabla con el número de nacimientos anuales en Aragón durante los últimos años. Esta información ayuda a comprender la dinámica demográfica de la población objeto de estudio:

Tabla 1- Población

Año	2018	2019	2020	2021	2022	2023
Nacimientos en Aragón	9.506	9.245	8.708	9.061	8.575	8.268

Esta población se caracteriza por un nivel socioeconómico y de acceso sanitario relativamente homogéneo. La atención sanitaria

pública en esta región garantiza un seguimiento prenatal protocolizado y accesible para toda la población, lo que permite trabajar con datos de calidad y una cobertura representativa del entorno local.

El hecho de centrarnos en una población atendida en un hospital público en un país en el que cualquier persona tiene derecho a ser atendida reduce la variabilidad asociada a desigualdades en el acceso a la atención médica. Esto proporciona una muestra representativa de la población general de Zaragoza.

Sin embargo, al trabajar con una muestra localizada geográficamente, los resultados sobre el estudio del crecimiento fetal, especialmente en la estimación de percentiles, puede verse influenciada por factores específicos de esta población. Elementos como el origen étnico, estilo de vida, los hábitos alimenticios, características genéticas o el acceso a la sanidad pueden afectar en el desarrollo del bebé, y en consecuencia, a las medias obtenidas. En este caso al centrarnos en la población de Zaragoza, los percentiles mostrarán características propias de la región que no tienen porque ser generalizables con otros países que no sean España o incluso con otras regiones del país.

No obstante, este enfoque también permite establecer referencias locales mas ajustadas a la población aragonesa, lo que puede ser interesante para el seguimiento clínico de embarazos dentro del mismo contexto geográfico.

A nivel económico, Zaragoza es una ciudad de nivel medio-alto en comparación con otras regiones españolas, que cuenta con un sistema sanitario de calidad y amplio acceso.

Esta combinación de factores nos lleva a concluir que el estudio es representativo para una población urbana que cuenta con acceso a la sanidad y un entorno geográfico relativamente estable. No obstante este estudio podría diferir en contextos más rurales o en regiones y países con una población de diferente perfil.

4.2.2. Base de Datos ecográfica y de partos

4.2.2.1. Datos de las distintas ecografías

En esta base de datos encontramos información sobre como ha progresado el peso del feto conforme ha progresado su edad gestacional.

Como sucede en muchas bases de datos, se han podido cometer diferentes errores a la hora de añadir nueva información y no todos los casos nos interesan. Es por ello que ha sido necesario llevar una limpieza en la que nos hemos quedado con embarazos que cumplieran las siguientes características:

- Gestación de un único feto
- Se han sometido a más de 1 ecografía
- Ecografías realizadas entre las semana 18 y 42
- No faltan datos en ninguna de sus variables

En cuanto a las variables que nos interesan :

- CODIGO: identifica el embarazo
- DIAS_ECO: el día en el que se ha llevado acabo la ecografía
- PESO_HADLOCK
- PESO_STIMERMAN

4.2.2.2. Datos de los distintos partos

En esta base de datos encontramos información sobre el parto que incluye el peso del feto al final de la gestación junto con características de la madre.

Al igual que con la otra base de datos ha sido necesario realizar una limpieza para quedarnos con los partos que nos interesan. En este caso la única condición que tenían que cumplir es que todas las variables de cada embarazo estuvieran completas.

En cuanto a las variables que nos interesan:

- CODIGO: identifica el embarazo
- DIAS_ECO: el día en el que se ha llevado acabo la ecografía
- Peso_gestante1:el peso real del feto el día del parto
- SEXO: genero del feto
- EDAD_MATERNA: que edad tiene la madre a la hora de dar a luz
- FIV: el bebe a estudio ha sido gestado mediante fecundación in vitro
- DIABETES: Ha tenido la madre diabetes durante el embarazo
- HIPERTENSION: Ha tenido la madre hipertensión durante el embarazo

- Fumadora: Ha sido la madre fumadora durante el embarazo
- PESO_MATERNO: peso de la madre
- IMC_MATERNA: índice de masa corporal de la madre
- TALLA_MADRE: altura de la mujer
- TALLA_PADRE: altura del padre
- PARIDAD_PREVIA: número de partos anteriores
- DRaza: indica la etnia de la mujer
- PAPPA: Proteína Plasmática A Asociada al Embarazo
- BHCG: Hormona producida por la placenta, evalua el riesgo de ciertas anomalías crónicas

4.3. PREPROCESAMIENTO DE LOS DATOS

4.3.1. Limpieza

Como se ha mencionado en apartados anteriores, contamos con dos bases de datos que incluyen múltiples variables descriptivas sobre el crecimiento fetal a lo largo de las semanas de gestación, así como características asociadas a la madre.

No obstante, ha sido necesario llevar a cabo un proceso de limpieza en ambas bases, ya que contienen variables que no resultan relevantes para los objetivos del estudio y registros con errores o datos anómalos que podrían distorsionar los resultados, como los casos atípicos.

4.3.1.1. Limpieza base de datos ecografía

Este proceso comienza con la base de datos ecográfica, que contiene información detallada sobre el feto durante distintas semanas de gestación. Para nuestro estudio, nos centramos únicamente en embarazos **monocorial-monoamnióticos**, es decir, aquellos con un solo feto. Esta información se encuentra en las variables tipo_placenta y n_feto, por lo que se incluyeron ambas para seleccionar exclusivamente los embarazos que cumplen esta condición. Esto se debe a que hay algún estudio (Savirón-Cornudella et al., 2020) sobre cómo los

embarazos múltiples difieren significativamente en sus patrones de crecimiento.

En el análisis exploratorio detectamos que algunas ecografías no tienen registrado el día en que se realizaron. Del mismo modo, algunas observaciones carecen de información sobre las variables PESO_HADLOCK y PESO_STIMERMAN, que son fundamentales para nuestro análisis. Por tanto, todas las ecografías sin esta información han sido eliminadas.

En relación con las semanas de gestación, nos interesan únicamente las ecografías realizadas entre la semana 18 y la semana 42. Aquellas fuera de este intervalo fueron descartadas.

Finalmente, se han conservado únicamente los embarazos que cuentan con más de una ecografía. Esta decisión se basa en nuestro interés por analizar la evolución del embarazo a lo largo del tiempo. Además, algunos embarazos sólo registran una ecografía en el hospital, ya que el resto del seguimiento se ha realizado en centros privados, lo cual limita su utilidad para nuestro análisis longitudinal.

A continuación, mostraré el código con el que hemos llevado a cabo la limpieza de la base de datos parto:

```
#1. Conservamos las columnas cuyas características son importantes para nuestro estudio:  
datoseco_filtrado <- DataEcos[, c("CODIGO", "DIAS_ECO", "n_feto",  
"tipo_placenta", "PESO_HADLOCK", "PESO_STIMERMAN")]  
  
#2. Hacer un resumen de todas las variables y ver donde tenemos datos faltantes  
summary(datoseco_filtrado)  
library(psych)  
describe(datoseco_filtrado)  
#podemos ver que hay datos faltantes en los pesos y dias eco ademas de que  
#dias eco tienen dias negativos algo que no debería de suceder  
  
#3. Eliminar en la variable tipo-placenta todas aquellas que sean de los tipos:  
"BICORIAL-BIAMNIOTICA", "Bicorial-biamniótica" , "BICORIAL",  
"TRICORIAL-TRIAMNIOTICA", "MONOCORIAL-BIAMNIOTICA"  
datoseco_filtrado <- subset(datoseco_filtrado, !tipo_placenta %in% c("BICORIAL-  
BIAMNIOTICA", "Bicorial-biamniótica", "BICORIAL", "TRICORIAL-  
TRIAMNIOTICA", "MONOCORIAL-BIAMNIOTICA"))  
  
#4. Eliminar los DIAS_ECO faltantes  
datoseco_filtrado <- datoseco_filtrado[!is.na(datoseco_filtrado$DIAS_ECO), ]  
  
#5. Eliminar los DIAS_ECO que no estan entre la semana 18-42, los dias 126-294  
datoseco_filtrado <- datoseco_filtrado[datoseco_filtrado$DIAS_ECO >= 126 &  
datoseco_filtrado$DIAS_ECO <= 294, ]
```

#6. *Eliminamos las ecos que no tienen PESO_HADLOCK ni PESO_STIMERMAN*
datoseco_filtrado <- datoseco_filtrado[!is.na(datoseco_filtrado\$PESO_HADLOCK),]
datoseco_filtrado <-
datoseco_filtrado[!is.na(datoseco_filtrado\$PESO_STIMERMAN),]

#7. *Eliminar los embarazos que no tengan mas de una ecografía y por tanto su CODIGO no se repita*
codigos_repetidos <- names(table(datoseco_filtrado\$CODIGO)
[table(datoseco_filtrado\$CODIGO) > 1])
datoseco_filtrado <- subset(datoseco_filtrado, CODIGO %in% codigos_repetidos)

summary(datoseco_filtrado)

4.3.1.2. Limpieza base de datos embarazo

El primer paso consistió en cargar la base de datos y seleccionar únicamente las variables relevantes para el estudio, ya que muchas otras no aportaban información útil. Al igual que con la base de ecografías, se realizó un resumen descriptivo para identificar datos faltantes y valores atípicos.

Una de las variables con mayor cantidad de datos ausentes fue EDAD_MATERNA. Aunque normalmente estos registros se eliminarían, en este caso fue posible recuperar los valores a partir de las variables fecha_parto y fecha_nacimiento_madre. Posteriormente, ambas variables fueron eliminadas para preservar el anonimato de las participantes.

En la variable FIV, la mayoría de los valores faltantes correspondían a casos en los que no se había realizado una fecundación in vitro. Por ello, se imputó el valor “NO” en estos registros. Otro ejemplo fue la variable PARIDAD_PREVIA, que contenía un valor atípico de 316 partos anteriores, claramente erróneo, por lo que dicho embarazo fue eliminado.

Asimismo, se eliminaron los embarazos que no contaban con datos en variables clave como sexo_gestante1, peso_gestante1, PAPP-A, B-HCG e hipertensión.

En cuanto a la variable categórica Fumadora, se detectó una codificación redundante, con dos categorías diferentes tanto para el “sí” como para el “no”. Para simplificar el análisis, esta variable fue recodificada en dos clases: “SI” y “NO”.

Por último, se excluyeron los partos con el mismo CODIGO que otro embarazo registrado, lo cual indica embarazos múltiples, que no

son de interés para este estudio. También se descartaron aquellos partos que ocurrieron más allá de la semana 42 de gestación, dado que se trata de casos postérmino que podrían introducir sesgos en el análisis del crecimiento fetal típico.

A continuación, mostrare el código que corresponde con la limpieza que se acaba de explicar:

#8. Conservamos las columnas que hablan de diferentes características del parto que nos interesan:

```
dataPartos <- dataPartos[, c("CODIGO", "FECHA_PARTO", "EDAD_MATERNA",  
"FIV", "DIABETES", "HIPERTENSION", "Fumadora", "PESO_MATERNO",  
"IMC_MATERNA", "TALLA_MADRE", "TALLA_PADRE", "PARIDAD_PREVIA",  
"DIAS_PARTO", "peso_gestante1", "sexo_gestante1", "D-Raza", "PAPP-A", "B-HCG", "Fecha_parto_A")]
```

#9. Hacemos un resumen de todas las variables

```
summary(dataPartos)  
describe(dataPartos)  
#Faltan datos en edad_materna, hipertension, talla padre, paridad_previa,  
peso_gestante1, PAPP-A, B-HCG  
#El valor de paridad previa mayor es de 316 y eso no es posible  
#DIAS_PARTO mayor es 691 que lo quitaremos al no estar en el rango de  
semanas
```

#10. Calcular la EDAD_MATERNA

```
dataPartos$FECHA_PARTO <- as.Date(dataPartos$FECHA_PARTO)  
dataPartos$Fecha_parto_A <- as.Date(dataPartos$Fecha_parto_A)  
dataPartos$EDAD_MATERNA <- as.numeric(difftime(dataPartos$Fecha_parto_A,  
dataPartos$FECHA_PARTO, units = "days")) / 365.25  
summary(dataPartos)  
#ya no faltan datos de EDAD_MATERNA
```

#11. Cambiar NA de FIV por NO

```
dataPartos$FIV[is.na(dataPartos$FIV)] <- "NO"
```

#12. Quitar los 18 partos que no conocemos su PARIDAD_PREVIA

```
dataPartos <- dataPartos[!is.na(dataPartos$PARIDAD_PREVIA), ]
```

#12.1 Quitar la PARIDAD_PREVIA de 316

```
dataPartos <- dataPartos[dataPartos$PARIDAD_PREVIA != 316, ]
```

#13. Eliminar los partos en los que no conocemos peso_gestante1

```
dataPartos <- dataPartos[!is.na(dataPartos$peso_gestante1), ]
```

#14. Eliminar los partos en los que no conocemos el sexo_gestante1

```
dataPartos <- dataPartos[!is.na(dataPartos$sexo_gestante1), ]
```

#14.2 ELIMINAMOS LA HIPERTENSION QUE NO CONOCEMOS

```
dataPartos <- dataPartos[!is.na(dataPartos$HIPERTENSION), ]
```

#15. Eliminamos PAPP-A y B-HCG que no conocemos

```
dataPartos <- dataPartos[!is.na(dataPartos$`PAPP-A`), ]
```

```
dataPartos <- dataPartos[!is.na(dataPartos$ 'B-HCG'), ]  
  
#16. Convertimos en numerico Fumadora y las nivelamos bien  
library(dplyr)  
dataPartos <- dataPartos %>%  
  mutate(  
    Fumadora = recode(Fumadora,  
      "SI" = 1,  
      "Fumadora" = 1,  
      "NO" = 0,  
      "No fuma" = 0),  
    Fumadora = as.numeric(Fumadora)  
)  
  
#17. Eliminar embarazos repetidos  
duplicados <- dataPartos$CODIGO[duplicated(dataPartos$CODIGO)]  
length(duplicados)  
  
frecuencias <- table(dataPartos$CODIGO)  
codigos_unicos <- names(frecuencias[frecuencias == 1])  
data_unida <- dataPartos[dataPartos$CODIGO %in% codigos_unicos, ]  
  
#18. Eliminamos la columna FECHA_PARTO para anonimizar el dataset  
dataPartos <- dataPartos[, c("CODIGO", "EDAD_MATERNA", "FIV", "DIABETES",  
  "HIPERTENSION", "Fumadora", "PESO_MATERNO", "IMC_MATERNA",  
  "TALLA_MADRE", "TALLA_PADRE", "PARIDAD_PREVIA", "DIAS_PARTO",  
  "peso_gestante1", "sexo_gestante1", "D-Raza", "PAPP-A", "B-HCG")]  
  
#19. Queremos solo embarazos que lleguen hasta la semana 42  
dataPartos <- dataPartos[dataPartos$DIAS_PARTO >= 126 &  
  dataPartos$DIAS_PARTO <= 294,]  
summary(dataPartos)
```

4.3.1.3. Dataset final

Una vez hemos limpiado cada dataset de forma independiente queda unirlas en una misma base de datos.

Este dataset combina las mediciones ecográficas en diferentes momentos de la gestación junto con las características de la madre. Para llevar a cabo esta unión tenemos que hacer ciertos cambios en ambas bases de datos.

4.3.1.3.1. Preparación de los datos en parto

Cree un nuevo dataframe llamado fila_parto que extrae las variables relevantes del dataframe que hemos limpiado del conjunto de datos de partos. A continuación, filtramos aquellos partos que tienen ecografías asociadas en el dataset limpiado de ecografía, asegurando que solo se consideren los partos con información ecográfica disponible.

A continuación, mostrare el código que se ha empleado en la explicación anterior:

```
#20. Cambiamos el dataset de partos para que podamos unir los 2 datasets
fila_parto <- data.frame(
  CODIGO = dataPartos$CODIGO,
  DIAS_ECO = dataPartos$DIAS_PARTO,
  PESO_HADLOCK = dataPartos$peso_gestante1,
  PESO_STIMERMAN = dataPartos$peso_gestante1,
  SEXO = dataPartos$sexo_gestante1,
  EDAD_MATERNA = dataPartos$EDAD_MATERNA,
  FIV = dataPartos$FIV,
  DIABETES = dataPartos$DIABETES,
  HIPERTENSION = dataPartos$HIPERTENSION,
  Fumadora = dataPartos$Fumadora,
  PESO_MATERNO = dataPartos$PESO_MATERNO,
  IMC_MATERNA = dataPartos$IMC_MATERNA,
  TALLA_MADRE = dataPartos$TALLA_MADRE,
  TALLA_PADRE = dataPartos$TALLA_PADRE,
  PARIDAD_PREVIA = dataPartos$PARIDAD_PREVIA,
  DRaza = dataPartos`D-Raza`,
  PAPPA = dataPartos`PAPP-A`,
  BHCG = dataPartos`B-HCG`
)
```

#21. Los CODIGO de ambos datasets se tienen que unir. Para ello vamos a quedarnos en la tabla dataPartos con los embarazos que trabajamos en la tabla de datoseco. Después uniremos ambos datasets tomando en cuenta la variable CODIGO:

```
codigos_validos <- unique(datoseco_filtrado$CODIGO)
fila_parto <- fila_parto[fila_parto$CODIGO %in% codigos_validos, ]
```

4.3.1.3.2. Preparación de los datos en ecografía

En paralelo, se llevó a cabo la preparación del conjunto de datos ecográficos. Para ello, se filtró datoseco_filtrado de modo que solo se conservaran las ecografías correspondientes a los partos seleccionados previamente. A continuación, se generó una estructura denominada fila_eco, con una disposición de variables similar a la de fila_parto, pero dejando vacías (con valores NA) aquellas variables maternas que no

están disponibles en el momento de la ecografía, como pueden ser la edad materna, peso o talla.

La explicación corresponde al siguiente código:

```
#22. nos mantenemos con las ecos que corresponden a los partos que tenemos
codigos_validos2 <- unique(fila_parto $CODIGO)
datoseco_filtrado <- datoseco_filtrado [datoseco_filtrado $CODIGO %in%
codigos_validos2, ]

#23. Nos quedamos solo con las variables de eco que nos interesa
fila_eco <- data.frame(
  CODIGO = datoseco_filtrado$CODIGO,
  DIAS_ECO = datoseco_filtrado$DIAS_ECO,
  PESO_HADLOCK = datoseco_filtrado$PESO_HADLOCK,
  PESO_STIMERMAN = datoseco_filtrado$PESO_STIMERMAN,
  SEXO=NA,
  EDAD_MATERNA=NA,
  FIV=NA,
  DIABETES=NA,
  HIPERTENSION=NA,
  Fumadora=NA,
  PESO_MATERNO=NA,
  IMC_MATERNA=NA,
  TALLA_MADRE=NA,
  TALLA_PADRE=NA,
  PARIDAD_PREVIA=NA,
  DRaza=NA,
  PAPPA=NA,
  BHCG=NA
)
```

4.3.1.3.3. Df_final

Una vez preparados ambos conjuntos, se procedió a su combinación. Esta unión se realizó de manera vertical utilizando la función rbind(), de forma que por cada embarazo se obtuviera un registro por cada ecografía realizada y un registro adicional correspondiente al parto. Con el objetivo de completar la información faltante en las filas de ecografías, se imputaron los valores ausentes de las variables maternas utilizando los datos del parto correspondiente, asegurando así coherencia y completitud dentro del conjunto final.

Como resultado, el dataset df_final contiene, para cada embarazo, tanto registros ecográficos como un registro de parto. En el caso de las ecografías, identificadas por tener un valor de DIAS_ECO menor que el de DIAS_PARTO, se incluye el identificador del embarazo (CODIGO), los días de gestación en el momento de la ecografía (DIAS_ECO), las estimaciones de peso fetal (PESO_HADLOCK y/o PESO_STIMERMAN), así como las variables maternas completadas desde la información del

parto. Por otro lado, los registros de parto, que se distinguen porque DIAS_ECO es igual a DIAS_PARTO, contienen el mismo identificador del embarazo, el número de días de gestación al nacer, el peso real del recién nacido, y todas las variables maternas completamente informadas.

A continuación, se muestra un gráfico de los pesos fetales estimados en las distintas ecografías (que conforman el dataset en este punto), representados en una línea temporal según el día de gestación en que fueron registrados.

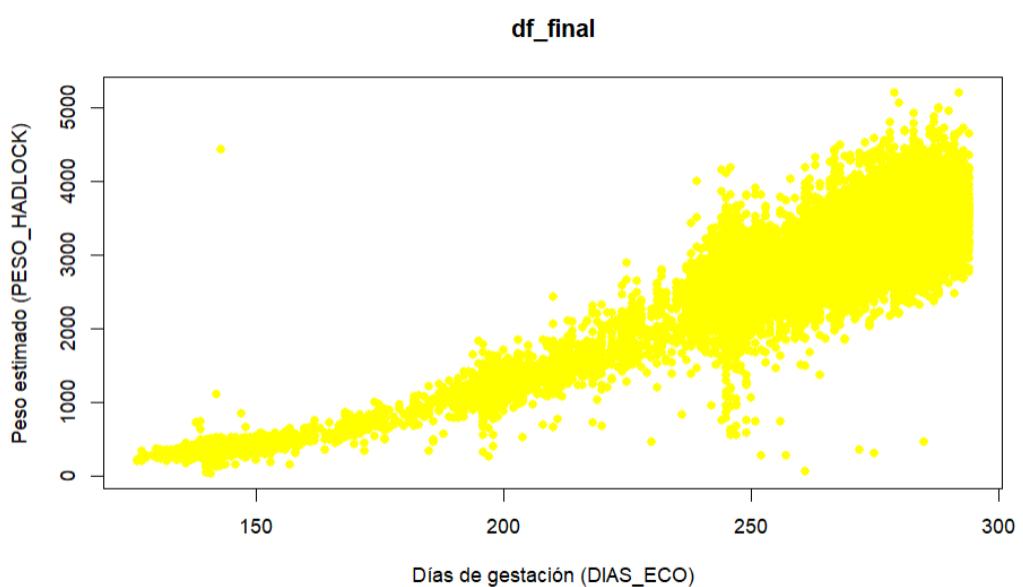


Ilustración 3, Datos iniciales de los pesos fetales estimados a lo largo de los días de la gestación

El código que se ha empleado es el siguiente:

```
#24. Vamos a juntar los datasets:  

df_final <- rbind(fila_eco, fila_parto)

# 25. Rellenar los valores de SEXO para todas las filas con el mismo CODIGO
library(dplyr)
library(tidy)

df_final <- df_final %>%
  group_by(CODIGO) %>%
  fill(
    SEXO,
    EDAD_MATERNA,
    FIV,
    DIABETES,
```

```
HIPERTENSION,  
Fumadora,  
PESO_MATERNO,  
IMC_MATERNA,  
TALLA_MADRE,  
TALLA_PADRE,  
PARIDAD_PREVIA,  
DRaza,  
PAPPA,  
BHCG,  
.direction = "downup"  
) %>%  
ungroup()  
  
summary(df_final)
```

```
#26. Imputar los datos de la variable PESO_MATERNO empleando la mediana  
sum(df_final$PESO_MATERNO == 0)  
mediana_peso <- median(df_final$PESO_MATERNO[df_final$PESO_MATERNO != 0], na.rm = TRUE)  
df_final$PESO_MATERNO[df_final$PESO_MATERNO == 0] <- mediana_peso  
sum(df_final$PESO_MATERNO == 0)  
  
#26.2 Grafico de nuestro dataset en este momento  
graficar_datos <- function(df, color = "yellow") {  
  # Filtrar valores atípicos de PESO_HADLOCK  
  df_filtrado <- df[df$PESO_HADLOCK < 6000, ]  
  
  # Graficar puntos  
  plot(  
    df_filtrado$DIAS_ECO,  
    df_filtrado$PESO_HADLOCK,  
    col = color,  
    main = deparse(substitute(df)),  
    xlab = "Días de gestación (DIAS_ECO)",  
    ylab = "Peso estimado (PESO_HADLOCK)",  
    pch = 16  
  )}  
graficar_datos(df_final)
```

4.3.1.4. Depuración de datos por análisis de outliers semanales

Para garantizar la calidad y consistencia de los datos ecográficos utilizados en el análisis, se llevó a cabo un riguroso proceso de detección y eliminación de valores atípicos (outliers) en las estimaciones de peso fetal, tanto para el método de Hadlock como el de Stímmerman. Este procedimiento se aplicó por separado según el sexo

del feto y por intervalos de semanas de gestación, definidos mediante rangos de días (DIAS_ECO). El objetivo fue preservar únicamente aquellos valores de peso fetal que se ajustaran a una distribución razonable dentro de cada grupo homogéneo.

Para ello, se desarrolló una función que filtra los datos en función de un rango temporal específico y del sexo del feto. A partir de los datos filtrados, se calcularon los límites de detección de outliers usando el criterio del rango intercuartílico (IQR), pero con un umbral más estricto de 3 veces el IQR. Concretamente, se calcularon los cuartiles primero (Q1) y tercero (Q3) del peso fetal estimado y se eliminaron aquellos valores que se encontraban por debajo de $Q1 - 3IQR$ o por encima de $Q3 + 3IQR$. Esta estrategia permitió eliminar observaciones extremas sin afectar la mayor parte de la distribución.

El procedimiento se aplicó en todos los intervalos de semanas de gestación desde la semana 18 hasta la semana 42, tanto para fetos varones como mujeres, y para ambos métodos de estimación del peso fetal. Una vez limpiados los datos por cada grupo, se unificaron de nuevo mediante la función `bind_rows()` para formar cuatro bases de datos completas: una para varones con peso estimado por Hadlock, otra para varones con Stímmerman, y lo mismo para las mujeres. Adicionalmente, se filtraron los embarazos para conservar únicamente aquellos que contenían al menos dos ecografías, lo cual es fundamental para ajustar posteriormente modelos de crecimiento fetal que incluyan variabilidad intraindividual.

Este enfoque por intervalos semanales y por sexo, junto con un criterio conservador de exclusión de outliers, asegura que los modelos de predicción se ajusten a datos fiables, evitando sesgos que podrían ser introducidos por mediciones anómalas o errores de registro.

Para finalizar este apartado de limpieza vamos a mostrar como han quedado la distribución de los pesos Hadlock con los que trabajaremos:

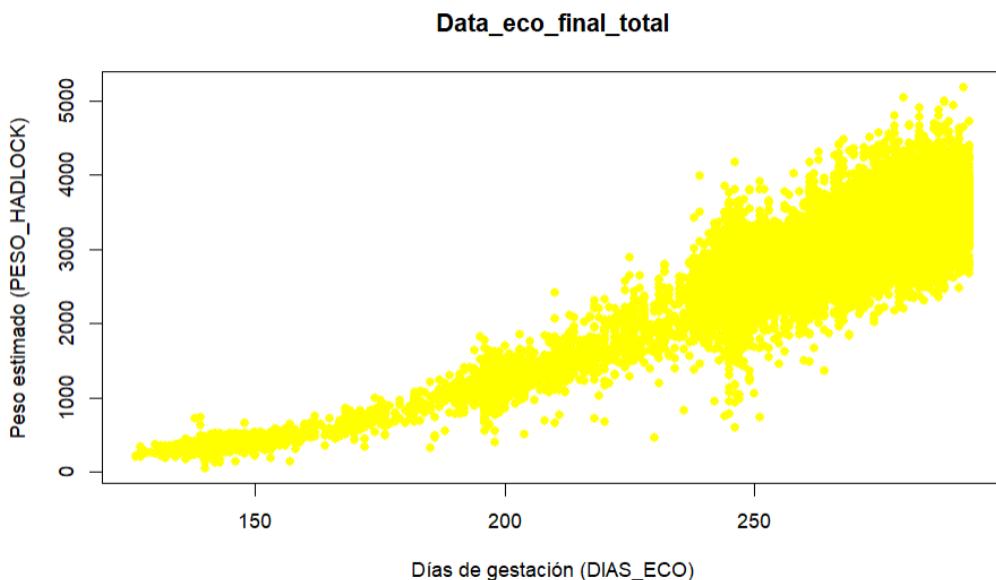


Ilustración 4, Datos sin outliers de los pesos fetales estimados a lo largo de los días de la gestación

En cuanto al código que se ha empleado es el siguiente:

```

#27. Vamos a quitar outliers por semana
# Función para filtrar datos por rango de días y eliminar outliers
#de PESO_HADLOCK y PESO_STIMERMAN segun el genero del bebe
filtrar_y_limpiar_por_sexo <- function(data, dias_min, dias_max, sexo_filtro,
columna_peso) {
  # Filtrar por rango de días y por sexo
  subset_datos <- subset(data, DIAS_ECO >= dias_min & DIAS_ECO < dias_max
  & SEXO == sexo_filtro)
  masa <- subset_datos[[columna_peso]]

  # Cálculo del IQR y límites
  q1 <- quantile(masa, 0.25, na.rm = TRUE)
  q3 <- quantile(masa, 0.75, na.rm = TRUE)
  iqr <- q3 - q1
  lower <- q1 - 3 * iqr
  upper <- q3 + 3 * iqr

  # Identificar outliers por índice
  es_outlier <- masa < lower | masa > upper
  print(paste("Outliers para sexo", sexo_filtro, "en", columna_peso, ":"))
  print(masa[es_outlier])

  # Eliminar outliers usando índice lógico
  subset_datos_clean <- subset_datos[!es_outlier, ]
  return(subset_datos_clean)
}

#27.1 SEMANA 18-20 (126-140)

```

```
Semana_18_19_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
0, dias_max = 140, sexo_filtro = "VARON", columna_peso = "PESO_HADLOCK")  
Semana_18_19_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
0, dias_max = 140, sexo_filtro = "VARON", columna_peso =  
"PESO_STIMERMAN")  
Semana_18_19_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
0, dias_max = 140, sexo_filtro = "MUJER", columna_peso = "PESO_HADLOCK")  
Semana_18_19_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final,  
dias_min = 0, dias_max = 140, sexo_filtro = "MUJER", columna_peso =  
"PESO_STIMERMAN")  
  
# Semana 20-22 (140-154)  
Semana_20_21_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
140, dias_max = 154, sexo_filtro = "VARON", columna_peso =  
"PESO_HADLOCK")  
Semana_20_21_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
140, dias_max = 154, sexo_filtro = "VARON", columna_peso =  
"PESO_STIMERMAN")  
Semana_20_21_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
140, dias_max = 154, sexo_filtro = "MUJER", columna_peso =  
"PESO_HADLOCK")  
Semana_20_21_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final,  
dias_min = 140, dias_max = 154, sexo_filtro = "MUJER", columna_peso =  
"PESO_STIMERMAN")  
  
# Semana 22-24 (154-168)  
Semana_22_24_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
154, dias_max = 168, sexo_filtro = "VARON", columna_peso =  
"PESO_HADLOCK")  
Semana_22_24_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
154, dias_max = 168, sexo_filtro = "VARON", columna_peso =  
"PESO_STIMERMAN")  
Semana_22_24_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
154, dias_max = 168, sexo_filtro = "MUJER", columna_peso =  
"PESO_HADLOCK")  
Semana_22_24_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final,  
dias_min = 154, dias_max = 168, sexo_filtro = "MUJER", columna_peso =  
"PESO_STIMERMAN")  
  
# Semana 24-26 (168-182)  
Semana_24_26_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
168, dias_max = 182, sexo_filtro = "VARON", columna_peso =  
"PESO_HADLOCK")  
Semana_24_26_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
168, dias_max = 182, sexo_filtro = "VARON", columna_peso =  
"PESO_STIMERMAN")  
Semana_24_26_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
168, dias_max = 182, sexo_filtro = "MUJER", columna_peso =  
"PESO_HADLOCK")  
Semana_24_26_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final,  
dias_min = 168, dias_max = 182, sexo_filtro = "MUJER", columna_peso =  
"PESO_STIMERMAN")
```

```
# Semana 26-28 (182-196)
Semana_26_28_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
182, dias_max = 196, sexo_filtro = "VARON", columna_peso =
"PESO_HADLOCK")
Semana_26_28_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
182, dias_max = 196, sexo_filtro = "VARON", columna_peso =
"PESO_STIMERMAN")
Semana_26_28_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
182, dias_max = 196, sexo_filtro = "MUJER", columna_peso =
"PESO_HADLOCK")
Semana_26_28_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final,
dias_min = 182, dias_max = 196, sexo_filtro = "MUJER", columna_peso =
"PESO_STIMERMAN")

# Semana 28-30 (196-210)
Semana_28_30_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
196, dias_max = 210, sexo_filtro = "VARON", columna_peso =
"PESO_HADLOCK")
Semana_28_30_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
196, dias_max = 210, sexo_filtro = "VARON", columna_peso =
"PESO_STIMERMAN")
Semana_28_30_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
196, dias_max = 210, sexo_filtro = "MUJER", columna_peso =
"PESO_HADLOCK")
Semana_28_30_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final,
dias_min = 196, dias_max = 210, sexo_filtro = "MUJER", columna_peso =
"PESO_STIMERMAN")

# Semana 30-32 (210-224)
Semana_30_32_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
210, dias_max = 224, sexo_filtro = "VARON", columna_peso =
"PESO_HADLOCK")
Semana_30_32_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
210, dias_max = 224, sexo_filtro = "VARON", columna_peso =
"PESO_STIMERMAN")
Semana_30_32_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
210, dias_max = 224, sexo_filtro = "MUJER", columna_peso =
"PESO_HADLOCK")
Semana_30_32_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final,
dias_min = 210, dias_max = 224, sexo_filtro = "MUJER", columna_peso =
"PESO_STIMERMAN")

# Semana 32-34 (224-238)
Semana_32_34_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
224, dias_max = 238, sexo_filtro = "VARON", columna_peso =
"PESO_HADLOCK")
Semana_32_34_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min =
224, dias_max = 238, sexo_filtro = "VARON", columna_peso =
"PESO_STIMERMAN")
```

```
Semana_32_34_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 224, dias_max = 238, sexo_filtro = "MUJER", columna_peso = "PESO_HADLOCK")
Semana_32_34_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 224, dias_max = 238, sexo_filtro = "MUJER", columna_peso = "PESO_STIMERMAN")

# Semana 34-36 (238-252)
Semana_34_36_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 238, dias_max = 252, sexo_filtro = "VARON", columna_peso = "PESO_HADLOCK")
Semana_34_36_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 238, dias_max = 252, sexo_filtro = "VARON", columna_peso = "PESO_STIMERMAN")
Semana_34_36_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 238, dias_max = 252, sexo_filtro = "MUJER", columna_peso = "PESO_HADLOCK")
Semana_34_36_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 238, dias_max = 252, sexo_filtro = "MUJER", columna_peso = "PESO_STIMERMAN")

# Semana 36-38 (252-266)
Semana_36_38_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 252, dias_max = 266, sexo_filtro = "VARON", columna_peso = "PESO_HADLOCK")
Semana_36_38_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 252, dias_max = 266, sexo_filtro = "VARON", columna_peso = "PESO_STIMERMAN")
Semana_36_38_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 252, dias_max = 266, sexo_filtro = "MUJER", columna_peso = "PESO_HADLOCK")
Semana_36_38_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 252, dias_max = 266, sexo_filtro = "MUJER", columna_peso = "PESO_STIMERMAN")

# Semana 38-40 (266-280)
Semana_38_40_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 266, dias_max = 280, sexo_filtro = "VARON", columna_peso = "PESO_HADLOCK")
Semana_38_40_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 266, dias_max = 280, sexo_filtro = "VARON", columna_peso = "PESO_STIMERMAN")
Semana_38_40_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 266, dias_max = 280, sexo_filtro = "MUJER", columna_peso = "PESO_HADLOCK")
Semana_38_40_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min = 266, dias_max = 280, sexo_filtro = "MUJER", columna_peso = "PESO_STIMERMAN")

# Semana 40-42 (280-294)
```

```
Semana_40_42_mas_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
280, dias_max = 294, sexo_filtro = "VARON", columna_peso =  
"PESO_HADLOCK")  
Semana_40_42_mas_stimerman <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
280, dias_max = 294, sexo_filtro = "VARON", columna_peso =  
"PESO_STIMERMAN")  
Semana_40_42_mujeres_hadlock <- filtrar_y_limpiar_por_sexo(df_final, dias_min =  
280, dias_max = 294, sexo_filtro = "MUJER", columna_peso =  
"PESO_HADLOCK")  
Semana_40_42_mujeres_stimerman <- filtrar_y_limpiar_por_sexo(df_final,  
dias_min = 280, dias_max = 294, sexo_filtro = "MUJER", columna_peso =  
"PESO_STIMERMAN")  
  
##### 28.DATASET ENTERO SIN OUTLIERS  
library(dplyr)  
Data_eco_final_hadlock_mas <- bind_rows(Semanas_18_19_mas_hadlock,  
Semanas_20_21_mas_hadlock, Semanas_22_24_mas_hadlock,  
Semanas_24_26_mas_hadlock, Semanas_26_28_mas_hadlock,  
Semanas_28_30_mas_hadlock, Semanas_30_32_mas_hadlock,  
Semanas_32_34_mas_hadlock, Semanas_34_36_mas_hadlock,  
Semanas_36_38_mas_hadlock, Semanas_38_40_mas_hadlock,  
Semanas_40_42_mas_hadlock)  
  
Data_eco_final_stimerman_mas <- bind_rows(Semanas_18_19_mas_stimerman,  
Semanas_20_21_mas_stimerman, Semanas_22_24_mas_stimerman,  
Semanas_24_26_mas_stimerman, Semanas_26_28_mas_stimerman,  
Semanas_28_30_mas_stimerman, Semanas_30_32_mas_stimerman,  
Semanas_32_34_mas_stimerman, Semanas_34_36_mas_stimerman,  
Semanas_36_38_mas_stimerman, Semanas_38_40_mas_stimerman,  
Semanas_40_42_mas_stimerman)  
  
Data_eco_final_hadlock_mujeres <- bind_rows(  
Semanas_18_19_mujeres_hadlock, Semanas_20_21_mujeres_hadlock,  
Semanas_22_24_mujeres_hadlock, Semanas_24_26_mujeres_hadlock,  
Semanas_26_28_mujeres_hadlock, Semanas_28_30_mujeres_hadlock,  
Semanas_30_32_mujeres_hadlock, Semanas_32_34_mujeres_hadlock,  
Semanas_34_36_mujeres_hadlock, Semanas_36_38_mujeres_hadlock,  
Semanas_38_40_mujeres_hadlock, Semanas_40_42_mujeres_hadlock)  
  
Data_eco_final_stimerman_mujeres <- bind_rows(  
Semanas_18_19_mujeres_stimerman, Semanas_20_21_mujeres_stimerman,  
Semanas_22_24_mujeres_stimerman, Semanas_24_26_mujeres_stimerman,  
Semanas_26_28_mujeres_stimerman, Semanas_28_30_mujeres_stimerman,  
Semanas_30_32_mujeres_stimerman, Semanas_32_34_mujeres_stimerman,  
Semanas_34_36_mujeres_stimerman, Semanas_36_38_mujeres_stimerman,  
Semanas_38_40_mujeres_stimerman, Semanas_40_42_mujeres_stimerman)  
  
#28.mirar que todos los embarazos tengan mas de una ecografía  
códigos_repetidos <- names  
(table(Data_eco_final_hadlock_mas$CODIGO)[table(Data_eco_final_hadlock_mas  
$CODIGO) > 1])
```

```
Data_eco_final_hadlock_mas <- subset(Data_eco_final_hadlock_mas, CODIGO  
%in% codigos_repetidos)  
codigos_repetidos <- names  
(table(Data_eco_final_stimerman_mas$CODIGO)[table(Data_eco_final_stimerma  
n_mas$CODIGO) > 1])  
Data_eco_final_stimerman_mas <- subset(Data_eco_final_stimerman_mas,  
CODIGO %in% codigos_repetidos)  
  
codigos_repetidos <- names  
(table(Data_eco_final_hadlock_mujeres$CODIGO)[table(Data_eco_final_hadlock_  
mujeres$CODIGO) > 1])  
Data_eco_final_hadlock_mujeres <- subset(Data_eco_final_hadlock_mujeres,  
CODIGO %in% codigos_repetidos)  
  
codigos_repetidos <- names  
(table(Data_eco_final_stimerman_mujeres$CODIGO)[table(Data_eco_final_stimer  
man_mujeres$CODIGO) > 1])  
Data_eco_final_stimerman_mujeres <-  
subset(Data_eco_final_stimerman_mujeres, CODIGO %in% codigos_repetidos)  
  
#### UNION DE DATASETS  
Data_eco_final_total <- bind_rows(Data_eco_final_hadlock_mas,  
Data_eco_final_stimerman_mas,  
Data_eco_final_hadlock_mujeres,  
Data_eco_final_stimerman_mujeres)  
  
Data_eco_final_total <- Data_eco_final_total %>%  
distinct()  
summary(Data_eco_final_total)  
#Grafico mostrando como es ahora la distribucion de nuestros datos tras la  
limpieza  
graficar_datos <- function(df, color = "yellow") {  
  # Filtrar valores atípicos de PESO_HADLOCK  
  df_filtrado <- df[df$PESO_HADLOCK < 6000, ]  
  
  # Graficar puntos  
  plot(  
    df_filtrado$DIAS_ECO,  
    df_filtrado$PESO_HADLOCK,  
    col = color,  
    main = deparse(substitute(df)),  
    xlab = "Días de gestación (DIAS_ECO)",  
    ylab = "Peso estimado (PESO_HADLOCK)",  
    pch = 16  
  )  
  graficar_datos(Data_eco_final_total)
```

4.4. METODOLOGÍA UTILIZADA EN ESTE ESTUDIO

Para el análisis del crecimiento fetal en este estudio, se ha empleado modelos lineales mixtos, útil para trabajar con datos longitudinales, es decir, que recogen múltiples mediciones por individuo a lo largo del tiempo.

4.4.1. *Modelo de crecimiento*

El modelo de crecimiento fetal es una herramienta estadística cuyo principal objetivo es estimar cuál sería el peso fetal ideal a lo largo de la edad gestacional. Este modelo se construye a partir de datos ecográficos recogidos en distintos momentos del embarazo, y permite representar la evolución esperada del peso fetal semana a semana en condiciones consideradas normales.

A través de este enfoque, se busca describir una curva de crecimiento que sirva como **patrón de referencia**, con la cual se puedan comparar otros embarazos. Esto resulta fundamental en el ámbito clínico, ya que permite identificar fetos que presentan un crecimiento anómalo, y así tomar decisiones médicas oportunas.

Para su construcción, el modelo debe incorporar la variación del peso del feto a lo largo de las semanas de gestación. No obstante, también es interesante que el modelo cuente con variables maternas relevantes como peso, talla, edad o condiciones médicas preexistentes. Esta adaptabilidad mejora la exactitud de las estimaciones y ofrece curvas de crecimiento ajustadas a las características individuales de cada embarazo.

4.4.2. *Que son los modelos mixtos*

El modelo lineal mixto es una extensión del modelo lineal general (GLM) que combina efectos fijos y aleatorios para analizar datos con estructuras jerárquicas, como medidas repetidas o agrupaciones (p. ej., fetos en embarazos múltiples).

Efecto fijo: representan influencias sistemáticas sobre la variable respuesta (como el IMC materno o la edad).

-Factores de efecto aleatorio: capturan variabilidad entre individuos o grupos (p. ej., $ui \sim N(0, \sigma^2_u)$).

La estructura del modelo se expresa como:

$$Y_{ij} = \underbrace{\beta_0 + \beta_1 X_{ij}}_{\text{Efectos fijos}} + \underbrace{u_i + \epsilon_{ij}}_{\text{Efectos aleatorios}}$$

- Y_{ij} : variable respuesta para la observación j en el grupo i
- X_{ij} : predictor fijo
- β_0, β_1 : Coeficientes fijos (intercepto y pendiente)
- u_i : Efecto aleatorio del grupo i (ej., $u_i \sim N(0, \sigma^2_u)$).
- ϵ_{ij} : Error residual ($\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$).

La principal ventaja de los modelos mixtos en este tipo de análisis es que permite tener en cuenta la estructura jerárquica de los datos, en la que las observaciones están anidadas dentro de individuos. Evita sesgos en la estimación de los cocientes y aprovechar toda la información disponible sin necesidad de reducir los datos a una sola medida por individuo.

4.4.3. Como implementamos los modelos mixtos en nuestro trabajo

En el presente estudio se ha implementado un modelo lineal mixto con el objetivo de predecir el crecimiento fetal a lo largo de la gestación, utilizando mediciones longitudinales (peso STIMERMAN y peso HADLOCK) obtenidas a partir de las ecografías.

La variable dependiente seleccionada ha sido el logaritmo del peso fetal estimado, calculado a partir de las mediciones ecográficas disponibles. Esta transformación se ha realizado con el fin de estabilizar la varianza y aproximar la normalidad de los residuos, en línea con lo recomendado en estudios previos relacionados con el crecimiento fetal.

Para la parte fija del modelo se han incluido varias variables maternas que, según la revisión clínica, podrían tener un impacto relevante en el desarrollo fetal. Entre estas variables se encuentran:

- Edad materna
- Índice de masa corporal (IMC) materno

- Talla de la madre
- Peso previo al embarazo
- Embarazo por fecundación in vitro (FIV)
- Paridad
- Fumadora
- Diabetes
- Hipertension
- PAPPA

Estas variables fijas permiten estimar el efecto medio de cada factor sobre el crecimiento fetal en toda la muestra.

En cuanto a los **efectos aleatorios**, se ha incluido un intercepto aleatorio y también un efecto aleatorio asociado a la edad gestacional, lo que permite que cada feto no solo tenga su propia línea base de crecimiento, sino también un ritmo de crecimiento diferenciado a lo largo del tiempo. Esta estructura capta la variabilidad interindividual en el patrón de desarrollo como la posibilidad de trayectorias de crecimiento no paralelas entre embarazos, reflejando así que cada embarazo puede presentar un ritmo de crecimiento ligeramente distinto, incluso si comparte características con otros embarazos de la muestra.

Desde el punto de vista técnico, los modelos mixtos en este análisis se han implementado con la función `lmer()` del paquete `lme4` en R, ajustando como variables dependientes del logaritmo del peso fetal estimado y modelando los modelos efectos fijos de diversas variables maternas, junto con un intercepto aleatorio por individuo.

Tras realizar varios modelos para el conjunto de embarazos cuyos fetos sean solo hombres o mujeres y que estén estudiando la estimación del peso STIMERMAN o HADLOCK, hemos llegado a este modelo final:

```
Modelo3_12bis<-lmer(log(PESO_STIMERMAN)~poly(DIAS_ECO,2,raw=TRUE)+  
PESO_MATERNO+EDAD_MATERNA+TALLA_MADRE+PARIDAD_PREVIA+PAPPA+Fu  
madora2 +DIABETES2+HIPERTENSION2+(DIAS_ECO|CODIGO),  
data=train_data_stimerman_mas)
```

EdadGestacional se trata como variable continua y CODIGO representa el identificador único para cada embarazo, sobre el que se aplica el efecto aleatorio.

Además, se han explorado interacciones potencialmente relevantes, como la interacción entre semana de gestación y FIV, o entre edad materna y talla, aunque solo se mantuvieron aquellas que mostraron significación estadística y coherencia clínica.

Por último, a partir del modelo ajustado es posible calcular intervalos de predicción que permiten estimar no solo el valor medio esperado del peso fetal, sino también sus límites inferiores y superiores, típicamente representados por los percentiles 10 y 90. Estos percentiles ofrecen una referencia clínica útil para detectar fetos con pesos por debajo o por encima de lo esperado según su edad gestacional, lo que puede ser indicativo de posibles alteraciones en el crecimiento intrauterino.

La estimación de estos percentiles se basa en la estimación de la varianza de las predicciones y su dependencia con la edad gestacional, que refleja la dispersión esperada de los pesos alrededor del valor medio. Esta varianza es esencial para construir los intervalos de confianza, ya que determina cuánto se espera que fluctúe el peso estimado dentro de una determinada edad gestacional. Así, los percentiles no solo ofrecen una medida de tendencia central, sino también un marco de referencia probabilístico sobre el crecimiento fetal esperado.

4.5. VARIANZA CONSTANTE O VARIABLE

Uno de los supuestos clásicos en los modelos lineales es el de **varianza constante** o *homocedasticidad*, lo que implica que el error (o la variabilidad) de las observaciones respecto a los valores predichos se mantiene uniforme a lo largo de todos los niveles del predictor. En el contexto del crecimiento fetal, esto equivaldría a suponer que la dispersión del peso fetal estimado es la misma en todas las semanas de gestación medido a través del coeficiente de variación de Pearson ($sd(x)/mean(x)$).

4.5.1. Suposición habitual de varianza constante

En muchos análisis estadísticos, asumir una varianza constante simplifica considerablemente la modelización e interpretación. Este supuesto es común en los modelos mixtos lineales, y en ocasiones se acepta sin cuestionamiento, sobre todo cuando la diferencia en la dispersión no es visualmente evidente o cuando el modelo sigue siendo suficientemente robusto. Sin embargo, esta simplificación puede llevar a una mala especificación del modelo si en realidad la variabilidad cambia significativamente con el tiempo, como ocurre en procesos de crecimiento. Hadlock realizó un primer estudio en 1991 donde analizó el coeficiente de variación de Pearson a lo largo del período de embarazo, concluyendo que este valor puede considerarse aproximadamente constante. Aunque esto es cercano a la realidad, las pequeñas variaciones en el coeficiente de variación de Pearson son muy influyentes en la estimación de los percentiles.

4.5.2. Porque la varianza cambia con la edad gestacional

En el caso del crecimiento fetal, **la variabilidad del peso estimado aumenta progresivamente con la edad gestacional**. Esta tendencia se debe, en parte, a la propia biología del proceso: a medida que el feto crece, las diferencias individuales se amplifican, lo cual genera una dispersión mayor en los pesos estimados en etapas más avanzadas del embarazo. Además, en semanas más tardías es más probable que aparezcan condiciones clínicas (como restricción del crecimiento intrauterino o macrosomía) que contribuyan a una mayor heterogeneidad.

Este fenómeno se puede visualizar fácilmente mediante gráficos de dispersión o de residuos, donde se aprecia cómo las estimaciones del peso fetal presentan menor variabilidad en etapas tempranas y una dispersión notablemente mayor en las semanas finales. Por tanto, aunque el modelo asumiera inicialmente varianza constante, es fundamental evaluar este supuesto y, de ser necesario, contemplar estructuras de varianza más complejas o transformaciones que corrijan esta heterocedasticidad.

4.6. PESO POR ECOGRAFÍA.

Es necesario explicar que el peso calculado en cada ecografía no es más que una estimación y por tanto no es una medida exacta como en el parto, pues en esa circunstancia es posible pesar al bebe con una báscula y obtener un peso exacto.

4.6.1. ¿Por qué no es exacto?

Hay varios factores por el que no es posible obtener una medida exacta del peso a través de las ecografías. Los motivos son los siguientes:

- Ecografía: la única herramienta con la que contamos es la ecografía, la cual retorna una imagen muy aproximada a la realidad aunque cuenta con error en las medidas. Estos errores dependen de la postura del feto, el grosor de la piel de la madre, la experiencia del médico encargado en tomar la radiografía y la calidad de la herramienta de ultrasonidos
- Fórmulas basadas en promedios: en la ecografía que se lleva a cabo, es necesario llevar a cabo diferentes medidas como el diámetro de la cabeza. Una vez calculadas se introducen en un programa estadístico que calcula peso tomando también en cuenta las semanas de embarazo. El error, pues que es una fórmula que da un peso promedio ya que un feto con las mismas medidas y mismas semanas gestacionales pueden no pesar lo mismo.
- Margen de error: como se ha mencionado antes, se toma en cuenta que hay un margen de error el cual suele rondar entre el 10% y 15% respecto al peso real.

4.6.2. Base de los cálculos

Como se ha explicado anteriormente el peso fetal se estima mediante ecuaciones matemáticas que relacionan medidas biométricas del feto obtenidas por ecografía. Las fórmulas que vamos a emplear provienen de los siguientes estudios

- Hadlock
- Intergrowth

A continuación hablaremos de cada uno de ellos y sus diferencias.

4.6.2.1. FORMULAS DE HADLOCK

Esta fórmula fue creada por el Dr. Frank Hadlock y su equipo en los años 80. Actualmente son las más utilizadas en la práctica clínica

Para poder desarrollar la fórmula, es necesario conocer las siguientes medidas del feto:

- Diametro Biparietal (BPD): Distancia entre los huesos parietales de la cabeza fetal
- Circunferencia Abdominal (CA): Perímetro del abdomen fetal (refleja el crecimiento nutricional)
- Longitud del Femur (LF): Mide el hueso del del muslo (longitud esquelética)
- Circunferencia cefálica (HC)

A continuación vamos a mostrar la fórmula:

$$\log_{10}(EFW) = 1.326 - 0.00326 \times AC \times FL + 0.0107 \times HC + 0.0438 \times AC + 0.158 \times FL$$

Y luego:

$$EFW = 10^{\log_{10}(EFW)}$$

4.6.2.2. Intergrowth-21st

Es un estudio multicéntrico internacional que buscaba establecer estándares de crecimiento fetal en condiciones óptimas. Para desarrollar esta fórmula empleó mujer de muchas nacionalidades y de diferentes continentes. El seguimiento se realizó desde la semana 14 de gestación hasta los 2 años de vida.

- **DBP:** Diámetro Biparietal en cm
- **CA:** Circunferencia Abdominal en cm
- **LF:** Longitud del Fémur en cm

$$\text{Peso estimado (g)} = 1.07 \times (DBP)^3 + 0.30 \times (CA)^2 \times LF$$

4.7. CONTROL DE PESO

Aunque a primeras parece que nos enfocamos más en el peso estimado del feto, el dato más importante no es su valor absoluto, sino su ubicación en los percentiles de crecimiento fetal.

Es más importante el percentil que el peso estimado, pues las tendencias en las curvas de percentiles sí es un indicador fiable de:

4.7.1. ¿Qué son los percentiles fetales?

Los percentiles son medidas estadísticas que dividen un conjunto de ordenadas en 100 partes iguales. Cada percentil indica el valor por debajo del cual se encuentra un cierto porcentaje de los datos.

Fórmula matemática para calcular el percentil P_k :

Dado un conjunto de datos ordenado de menor a mayor con n observaciones, el percentil número k se calcula mediante:

$$P_k = x_{(i)} + \left(k \cdot \frac{n+1}{100} - i \right) \cdot (x_{(i+1)} - x_{(i)})$$

- P_k : percentil deseado (por ejemplo P25 para el percentil 25)
- $x(i)$: es el valor en la posición entera i del conjunto de datos ordenado
- $k*((n+1)/100)$ es la posición teórica del percentil
- La parte decimal se interpola linealmente entre $x(i)$ y $x(i+1)$

4.8. MODELO POR GÉNERO

En este estudio se ha optado por construir modelos desagregados para fetos masculinos y femeninos, con el fin de mejorar la precisión de las estimaciones del crecimiento fetal. Esta decisión metodológica se basa en la evidencia ampliamente respaldada por la literatura médica y obstétrica que indica que, en condiciones normales, los fetos masculinos tienden a presentar un peso ligeramente superior al de los femeninos en cada etapa del embarazo.

Las diferencias de crecimiento según el sexo fetal, aunque sutiles, son sistemáticas y pueden afectar la posición de un feto respecto a los percentiles establecidos si no se tienen en cuenta. Por tanto, emplear un único modelo común para ambos sexos podría introducir un sesgo en la evaluación del desarrollo intrauterino, especialmente en los percentiles extremos (P10 o P90), que son clínicamente relevantes para detectar posibles casos de restricción del crecimiento fetal (RCIU) o macrosomía.

Al dividir la muestra en dos subconjuntos —fetos masculinos y fetos femeninos— y ajustar modelos mixtos por separado para cada grupo, se logra una mayor sensibilidad en la estimación de las curvas de crecimiento. Esta estrategia permite adaptar los modelos a las particularidades de cada grupo, proporcionando una herramienta más fiable para el seguimiento clínico y para futuras aplicaciones en la práctica obstétrica local.



5. RESULTADOS

5.1. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES

5.1.1. Variable Continuas

En la siguiente tabla se resumen las estadísticas descriptivas de las principales variables cuantitativas del estudio. Se presentan el valor mínimo, primer cuartil (Q1), mediana, media, tercer cuartil (Q3) y el valor máximo.

Tabla 2, Variables numéricas consideradas en el modelo y sus resultados estadísticos.

Variable	Min	1st Quartil	Media n	Mean	3rd Quartil	Max
DIAS_ECO	126	144	245	220	270	293
PESO_HADLOCK	165	380.6	2421	1990.5	3000	5190
PESO_STIMERMAN	287.1	471.6	2330.5	1993	3000	5190
EDAD_MATERNA	15.14	30.70	34.02	33.55	37	49.92
PESO_MATERNO	40	56	62.5	65.09	71	150
TALLA_MADRE	130	159	163	163.3	168	187
PARIDAD_PREVIA	0	0	0	0.62	1	11
PAPPA	0	0.66	0.97	1.12	1.41	9.66
BHCG	0.1277	0.65	0.9662	1.1697	1.4557	18.5659

5.1.2. VARIABLES CATEGÓRICAS

A continuación vamos a mostrar como se describen las variables categóricas recogidas:

Tabla 3, Número de observaciones por sexo fetal

Variable	Total	MUJER	VARON
SEXO	31064	15239	15825

Tabla 4, Distribución de observaciones según las categorías de la variable FIV (Fecundación In Vitro)

Variable	Total	FIV	FIV DO	FIV DO SOSPECHAS	Donante de óvulo	NO
FIV	31064	1073	162	133	22	29674

Tabla 5, Distribución de observaciones según las categorías de la variable Hipertensión

Variable	Total	No HTA	HTA crónica	Preeclampsia	HTA gestacional
Hipertensión	31064	30252	156	277	379

Tabla 6, Distribución de observaciones según las categorías de la variable Diabetes

Variable	Total	Nada	DM pregestacional	DM gestacional	Intolerancia glucosa
Diabetes	31064	27307	319	3140	298

Tabla 7, Distribución de las mujeres Fumadoras

Variable	Total	Fumadora	No Fumadora
Fumadora	31064	4630	26434

5.2. MODELO DE PESO HADLOCK PARA FETOS VARONES

5.2.1. Modelo sin características de la madre

El modelo Modelo2 fue ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock ($\log(\text{PESO_HADLOCK})$), en función de los días de gestación (DIAS_ECO). Para capturar la evolución no lineal del crecimiento fetal, se incluyó un término polinómico de segundo grado. Además, se incorporaron efectos aleatorios por individuo (identificado mediante **CODIGO**) en el intercepto y la pendiente de **DIAS_ECO**, permitiendo así una modelización individualizada del crecimiento fetal.

Composición del modelo:

Modelo2<~

```
lmer(log(PESO_HADLOCK)~poly(DIAS_ECO,2,raw=TRUE)+(DIAS_ECO|  
CODIGO),data=train_data_hadlock_mas)
```

Resultados del modelo:

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
CODIGO	(Intercept)	1.723e-02	0.1312626	
	DIAS_ECO	3.707e-07	0.0006088	-0.80
Residual		5.373e-03	0.0733020	

Number of obs: 12652, groups: CODIGO, 4853

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.216e-01	1.612e-02	38.57
poly(DIAS_ECO, 2, raw = TRUE)1	4.733e-02	1.663e-04	284.59
poly(DIAS_ECO, 2, raw = TRUE)2	-7.328e-05	4.006e-07	-182.93

Correlation of Fixed Effects:

(Intr)	p(DIAS_ECO, 2, r=TRUE)1
p(DIAS_ECO, 2, r=TRUE)1	-0.991
p(DIAS_ECO, 2, r=TRUE)2	0.978 -0.996

Ilustración 5, Resultados del modelo mixto lineal sin variables maternas para fetos masculinos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock en función de los días de gestación..

- **Efectos fijos:** los coeficientes asociados a DIAS_ECO (grado 1 y grado 2) son altamente significativos, lo que confirma la relevancia de la relación cuadrática entre los días de embarazo y el peso fetal.
 - poly(DIAS_ECO, 2)1: coeficiente positivo → tendencia creciente.

- poly(DIAS_ECO, 2)2: coeficiente negativo → desaceleración del crecimiento en fases tardías, típico del desarrollo fetal.

▪ **Efectos aleatorios**

- El modelo estima una variabilidad considerable entre individuos en el intercepto ($\sigma \approx 0.131$) y menor en la pendiente ($\sigma \approx 0.00061$), con una correlación negativa alta (-0.80), lo que sugiere que los fetos con mayor peso inicial tienden a crecer más lentamente, y viceversa.
- La varianza residual es baja ($\sigma \approx 0.073$), lo que indica un buen ajuste general del modelo a los datos observados.

Evaluación cuantitativa:

Una vez ajustado el modelo mixto lineal para predecir el peso fetal estimado mediante la fórmula de Hadlock, se procedió a evaluar su rendimiento tanto en el conjunto de entrenamiento como en el de test. Esta evaluación se llevó a cabo en dos etapas complementarias: primero, mediante el cálculo del porcentaje de observaciones reales que se encontraban por fuera de los límites definidos por los percentiles 10 (P10) y 90 (P90); y en segundo lugar, mediante la representación gráfica del ajuste del modelo y sus bandas de predicción.

El motivo de realizar primero el cálculo de los porcentajes de fetos que se encontraban por debajo del P10 y por encima del P90 fue obtener una medida cuantitativa objetiva del grado de ajuste del modelo. Esta métrica permite verificar si el modelo está cumpliendo con su propósito de cubrir, aproximadamente, el 80% de los casos dentro del rango entre P10 y P90. Además, facilita la comparación entre los conjuntos de entrenamiento y test, permitiendo detectar posibles problemas de sobreajuste o subajuste antes de recurrir al análisis visual. A continuación mostramos este resultado en la siguiente imagen:

```
> # Datos de test
> graficar_datos(test_data_hadlock_mas)
> contar_extremos(test_data_hadlock_mas)
Nº de pesos por debajo del P10: 223 (7.05%)
Nº de pesos por encima del P90: 205 (6.48%)
> #APLICAR A TRAIN Y TEST
> # Datos de entrenamiento
> graficar_datos(train_data_hadlock_mas)
> contar_extremos(train_data_hadlock_mas)
Nº de pesos por debajo del P10: 887 (7.01%)
Nº de pesos por encima del P90: 777 (6.14%)
```

Ilustración 6, Porcentaje de observaciones del peso fetal estimado mediante la fórmula de Hadlock fuera del rango interpercentil [P10, P90] en los conjuntos de entrenamiento y test del modelo mixto sin variables maternas, ajustado para fetos masculinos.

- En **datos de entrenamiento**, el 7.01% de los valores se sitúan por debajo del percentil 10 (P10) y el 6.14% por encima del percentil 90 (P90), es decir, el **86.85%** de las observaciones están dentro del rango previsto.
- En **datos de test**, el 7.05% cae por debajo del P10 y el 6.48% por encima del P90, con un **86.47%** de cobertura total. Estos resultados son coherentes y reflejan una buena capacidad de generalización del modelo.

Evaluación visual del modelo

A continuación, se visualizaran los resultados a través de gráficos que muestran el peso fetal real (PESO_HADLOCK) frente a los días de gestación (DIAS_ECO), junto con las predicciones del modelo y las

bandas de los percentiles P10 y P90.

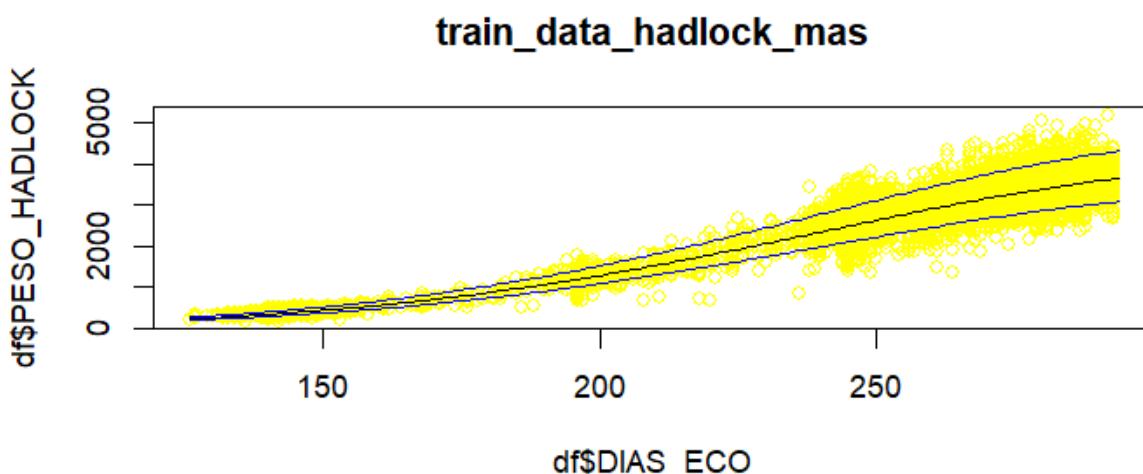


Ilustración 7, Ajuste del modelo mixto sin variables maternas sobre el conjunto de entrenamiento: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Hadlock para fetos masculinos, junto con observaciones reales.

En el conjunto de entrenamiento (train_data_hadlock_mas), los resultados muestran una adecuada capacidad del modelo para ajustarse a la tendencia general del crecimiento fetal. El 7.01% de los pesos estimados se sitúan por debajo del percentil 10 y el 6.14% por encima del percentil 90, lo que indica que el 86.85% de los valores caen dentro del rango interpercentil [P10, P90], cumpliendo así con el objetivo de abarcar el grueso de los datos dentro de este intervalo.

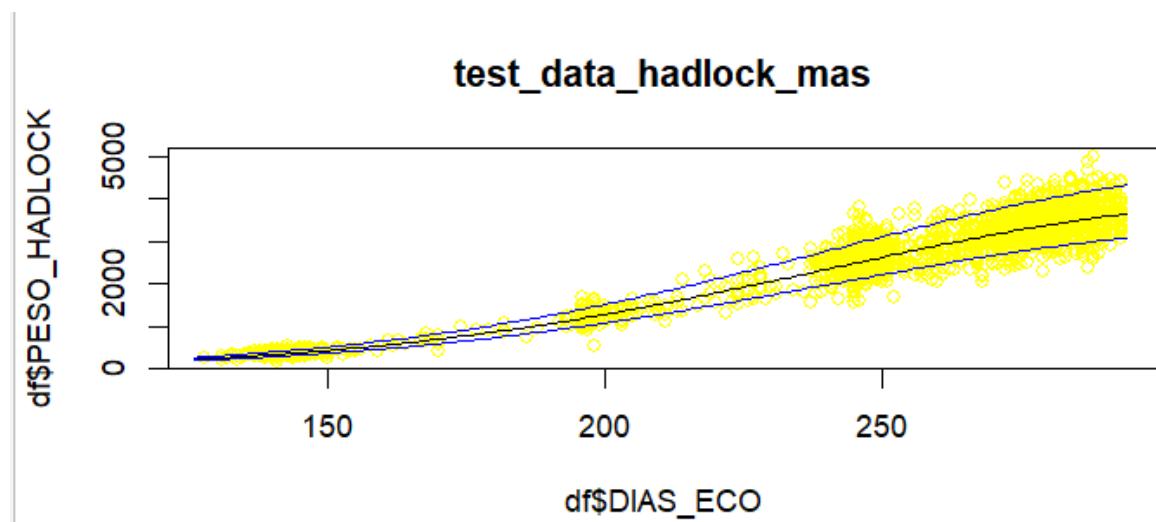


Ilustración 8, Ajuste del modelo mixto sin variables maternas sobre el conjunto de test: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Hadlock para fetos masculinos, junto con observaciones reales.

En el conjunto de test (test_data_hadlock_mas), el modelo mantiene un comportamiento similar, con un 7.05% de los pesos por debajo del P10 y un 6.48% por encima del P90. Esto se traduce en un 86.47% de los valores incluidos dentro del rango entre los percentiles 10 y 90, lo que refleja una buena capacidad de generalización del modelo a nuevos datos.

Como hemos podido observar, en ambas gráficas (entrenamiento y test), se observa que el modelo sigue correctamente la tendencia de crecimiento fetal a lo largo del embarazo, capturando la forma no lineal del proceso, especialmente en las fases más avanzadas de la gestación.

Estos análisis combinados —cuantitativo y visual— proporcionan una validación sólida del comportamiento del modelo, demostrando que es capaz de ajustarse adecuadamente a los datos y de identificar correctamente los casos extremos que podrían requerir una mayor atención clínica.

5.2.2. Modelo con característica de la madre

El modelo mixto Modelo2_12bis fue ajustado con el objetivo de predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock ($\log(PESO_HADLOCK)$), incorporando como efectos fijos variables maternas relevantes, y permitiendo efectos aleatorios por embarazo (CÓDIGO), tanto para el intercepto como para la pendiente respecto a los días de gestación (DIAS_ECO).

Composición del modelo:

```
Modelo2_12bis<-lmer(log(PESO_HADLOCK) ~
poly(DIAS_ECO,2,raw=TRUE) + PESO_MATERNO + EDAD_MATERNA +
TALLA_MADRE + PARIDAD_PREVIA + PAPPA + Fumadora2 +
DIABETES2 + HIPERTENSION2 +
(DIAS_ECO|CÓDIGO), data=train_data_hadlock_mas)
```

- **Efectos fijos:** se incluyeron los términos polinomiales de segundo grado de los días de embarazo ($\text{poly}(\text{DIAS_ECO}, 2, \text{raw}=TRUE)$), así como covariables clínicas maternas como el

peso, la edad y la talla, la paridad previa, el valor de PAPPA, si la madre es fumadora, y la presencia de diabetes (gestacional o pregestacional) e hipertensión (crónica, preeclampsia o gestacional).

- **Efectos aleatorios:** se introdujo un intercepto y una pendiente aleatoria para DIAS_ECO por cada gestante (CODIGO), lo cual permite capturar la variabilidad individual en el crecimiento fetal.

Resultado del ajuste:

```
Scaled residuals:
    Min      1Q   Median      3Q      Max
-11.3770 -0.4676  0.0173  0.5063  3.7745

Random effects:
Groups   Name        Variance Std.Dev. Corr
CODIGO   (Intercept) 7.165e-03 0.0846488
          DIAS_ECO    4.112e-07 0.0006413 -0.87
Residual            7.063e-03 0.0840393
Number of obs: 12660, groups: CODIGO, 4859

Fixed effects:
                                         Estimate Std. Error t value
(Intercept)                      2.042e+00  3.789e-02 53.892
poly(DIAS_ECO, 2, raw = TRUE)1  3.333e-02  1.877e-04 177.529
poly(DIAS_ECO, 2, raw = TRUE)2 -4.416e-05  4.530e-07 -97.467
PESO_MATERO          5.792e-04  1.078e-04  5.371
EDAD_MATERNA        6.347e-04  2.606e-04  2.435
TALLA_MADRE         9.234e-04  2.097e-04  4.403
PARIDAD_PREVIA      6.649e-06  2.627e-04  0.025
PAPPA                1.270e-02  1.958e-03  6.485
Fumadora25í        -1.003e-02  3.576e-03 -2.805
DIABETES2DM GESTACIONAL 9.040e-03  1.336e-02  0.677
DIABETES2DM PREGESTACIONAL 9.132e-03  4.283e-03  2.132
HIPERTENSION2HTA CRONICA  1.876e-02  1.898e-02  0.988
HIPERTENSION2PREECLAMPSIA -3.462e-02  1.571e-02 -2.204
HIPERTENSION2HTA GESTACIONAL -1.960e-02  1.197e-02 -1.637
```

Ilustración 9, Resultados del modelo mixto lineal con variables maternas para fetos masculinos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock en función de los días de gestación.

- **Efectos significativos:** El modelo muestra que el crecimiento fetal está influenciado significativamente por múltiples factores. Entre ellos:
 - **Peso materno** ($p < 0.001$), **edad materna** ($p < 0.001$) y **talla materna** ($p < 0.001$) presentan efectos positivos significativos sobre el peso fetal estimado.

Resultados

- **Paridad previa** también presenta un efecto positivo significativo, al igual que el valor de **PAPPA** ($p < 0.001$).
- Por el contrario, **ser fumadora** se asocia significativamente con un menor peso fetal estimado ($p < 0.01$).
- La **diabetes gestacional** muestra una asociación positiva significativa, mientras que la **hipertensión crónica** y la **preeclampsia** se asocian negativamente con el peso fetal (ambas con $p < 0.05$).
- La variable **hipertensión gestacional y diabetes pregestacional** no alcanzaron significación estadística clara.
- **Efectos aleatorios:** la variabilidad entre gestantes (CÓDIGO) se refleja en una desviación estándar del intercepto de 0.084 y una correlación negativa alta (-0.87) entre el intercepto y la pendiente de DIAS_ECO. Esto indica que las gestantes con mayor peso estimado al inicio tienden a presentar una pendiente de crecimiento más lenta, o viceversa. La varianza residual es baja, lo que sugiere un ajuste adecuado del modelo a los datos.

Evaluación visual del modelo:

En la figura correspondiente se muestra el crecimiento fetal estimado para una madre concreta (la primera fila del dataset), en función de sus características personales. La curva estimada (línea roja) se sitúa dentro de los márgenes definidos por los percentiles 10 y 90 (líneas azules discontinuas), y los pesos reales observados (puntos negros) siguen adecuadamente la trayectoria prevista por el modelo.

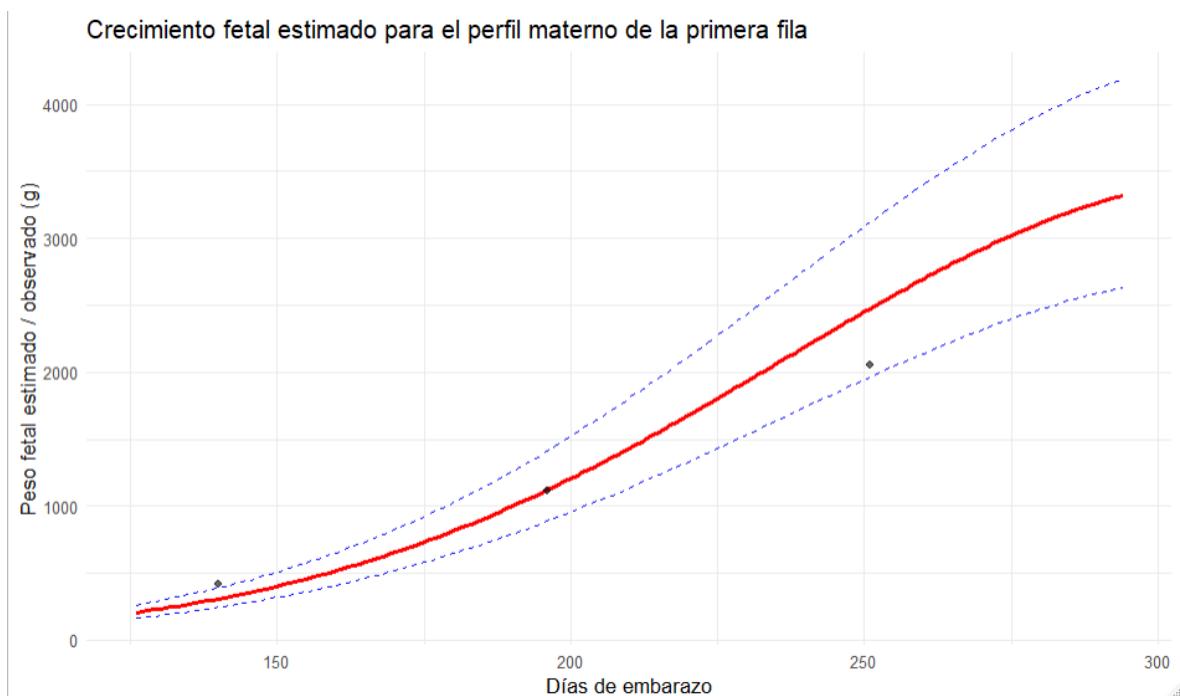


Ilustración 10, Curva de crecimiento fetal estimado para un feto masculino, generada para una madre específica según sus características clínicas, con percentiles 10, 50 y 90 obtenidos mediante modelo mixto y pesos estimados por la fórmula de Hadlock superpuestos.

Todos los puntos observados caen dentro del intervalo de predicción generado por el modelo, lo que sugiere que este no solo tiene un buen ajuste a nivel poblacional, sino también un comportamiento adecuado en la predicción individualizada, respetando la variabilidad biológica y clínica.

Como podemos ver, el modelo Modelo2_12bis logra capturar tanto la evolución temporal del crecimiento fetal como el efecto de variables clínicas maternas relevantes. Su estructura permite realizar predicciones personalizadas y generar curvas de referencia individualizadas, lo que resulta útil para la detección precoz de desviaciones en el desarrollo fetal y para una atención obstétrica más personalizada.

5.3. MODELO DE PESO STIMERMAN PARA FETOS VARONES

5.3.1. *Modelo sin características de la madre*

El modelo mixto Modelo3 fue ajustado con el objetivo de predecir el logaritmo del peso fetal estimado mediante la fórmula de Stirmann

($\log(\text{PESO_STIMERMAN})$), considerando únicamente los días de gestación (DIAS_ECO) como variable explicativa. Para capturar la naturaleza no lineal del crecimiento fetal, se utilizó un término polinómico de segundo grado. Además, se introdujeron efectos aleatorios por embarazo (CODIGO), tanto en el intercepto como en la pendiente de DIAS_ECO, permitiendo modelar la evolución individual de cada feto.

Composición del modelo:

```
Modelo3<-
lmer(log(PESO_STIMERMAN)~poly(DIAS_ECO,2,raw=TRUE)+(DIAS_ECO|CODIGO),data=train_data_stimerman_mas)
```

Scaled residuals:

Min	1Q	Median	3Q	Max
-10.6971	-0.4468	0.0066	0.4828	4.6447

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
CODIGO	(Intercept)	1.422e-02	0.1192324	
	DIAS_ECO	5.985e-07	0.0007736	-0.89
Residual		6.554e-03	0.0809565	

Number of obs: 12656, groups: CODIGO, 4839

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.269e+00	1.765e-02	128.56
poly(DIAS_ECO, 2, raw = TRUE)1	3.329e-02	1.831e-04	181.76
poly(DIAS_ECO, 2, raw = TRUE)2	-4.408e-05	4.418e-07	-99.79

Correlation of Fixed Effects:

(Intr)	p(DIAS_ECO, 2, r=TRUE)1
p(DIAS_ECO, 2, r=TRUE)1	-0.993
p(DIAS_ECO, 2, r=TRUE)2	0.980
	-0.995

Ilustración 11, Resultados del modelo mixto lineal sin variables maternas para fetos masculinos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Stimerman en función de los días de gestación.

- Efectos fijos:** los dos términos del polinomio de segundo grado para DIAS_ECO resultaron altamente significativos ($p < 0.001$). El primer término (positivo) refleja el crecimiento progresivo del feto con el tiempo, mientras que el segundo término (negativo) indica una desaceleración natural del crecimiento en las etapas finales del embarazo, fenómeno característico del desarrollo fetal.

- **Efectos aleatorios:** se identificó una variabilidad considerable entre embarazos en el intercepto (desviación estándar ≈ 0.119) y una menor variabilidad en la pendiente de DIAS_ECO (desviación estándar ≈ 0.00077). La correlación negativa elevada entre ambos componentes (-0.89) sugiere que aquellos fetos con mayor peso inicial tienden a mostrar un crecimiento más lento posteriormente, o viceversa.
- **Varianza residual:** baja ($\sigma \approx 0.081$), lo que indica un ajuste adecuado del modelo a los datos observados.

Evaluación cuantitativa:

Para evaluar el rendimiento del modelo se analizaron los porcentajes de observaciones fuera de los percentiles 10 (P10) y 90 (P90) tanto en el conjunto de entrenamiento como en el de test:

```
> # Datos de entrenamiento
> graficar_datos(train_data_stimerman_mas)
> contar_extremos(train_data_stimerman_mas)
Nº de pesos por debajo del P10: 1111 (8.78%)
Nº de pesos por encima del P90: 1047 (8.27%)
> # Datos de test
> graficar_datos(test_data_stimerman_mas)
> contar_extremos(test_data_stimerman_mas)
Nº de pesos por debajo del P10: 268 (8.47%)
Nº de pesos por encima del P90: 268 (8.47%)
```

Ilustración 12, Porcentaje de observaciones del peso fetal estimado mediante la fórmula de Stimerman fuera del rango interpercentil [P10, P90] en los conjuntos de entrenamiento y test del modelo mixto sin variables maternas, ajustado para fetos masculinos.

- **Entrenamiento:** el 8.78% de los valores se sitúan por debajo del P10 y el 8.27% por encima del P90, lo que supone que el 82.95% de las observaciones se encuentran dentro del rango previsto.
- **Test:** se obtuvo un 8.47% de casos tanto por debajo como por encima del rango, resultando en una cobertura total del 83.06%.

Estos resultados reflejan una consistencia en el comportamiento del modelo, sin grandes diferencias entre entrenamiento y test, lo cual sugiere una buena capacidad de generalización.

Evaluación visual del modelo:

A continuación, se presentan los gráficos correspondientes a los conjuntos de entrenamiento y test. En ellos se representa el peso fetal estimado (eje Y) frente a los días de embarazo (DIAS_ECO, eje X). La línea negra central corresponde a la predicción media del modelo, mientras que las bandas de color azul marcan los percentiles 10 y 90. Los puntos amarillos representan los datos observados reales.

- Entrenamiento

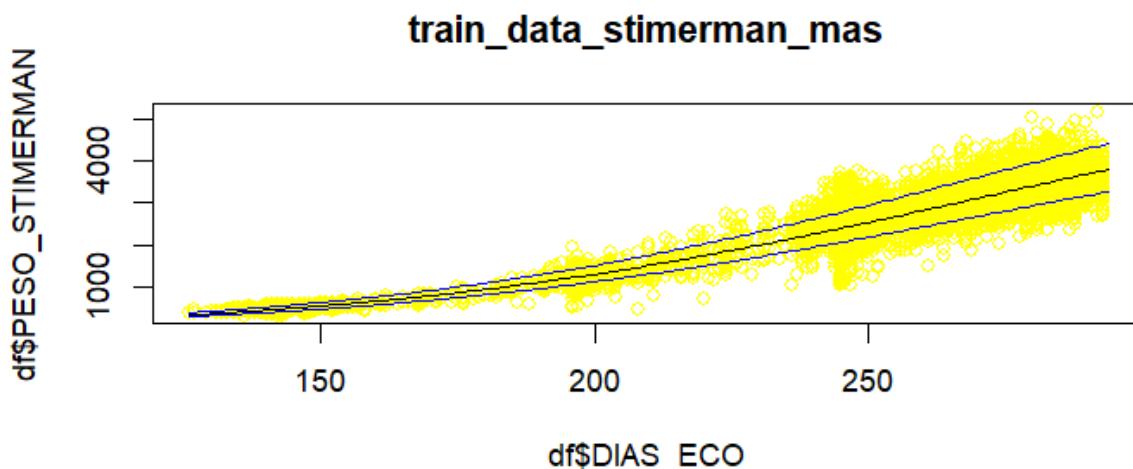


Ilustración 13, Ajuste del modelo mixto sin variables maternas sobre el conjunto de entrenamiento: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Stimerman para fetos masculinos, junto con observaciones reales.

- Test

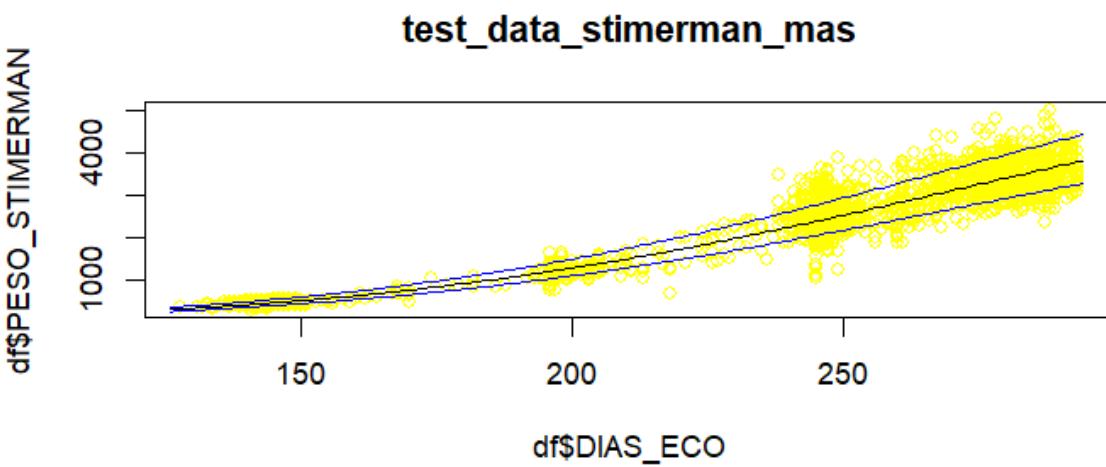


Ilustración 14, Ajuste del modelo mixto sin variables maternas sobre el conjunto de test: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Stimerman para fetos masculinos, junto con observaciones reales.

Ambas gráficas confirman visualmente los resultados obtenidos. La curva estimada sigue adecuadamente la tendencia general del crecimiento fetal, y la mayoría de los puntos se encuentran dentro de las bandas percentilares. Aunque se observa una mayor dispersión de datos en edades gestacionales avanzadas, el modelo mantiene un ajuste estable y coherente.

Podemos concluir que el modelo Modelo3, basado en la fórmula de Stirnemann y sin incluir características maternas, proporciona una estimación razonablemente precisa del crecimiento fetal masculino. Gracias a la incorporación de efectos aleatorios por embarazo, el modelo consigue reflejar la variabilidad individual en la evolución fetal. Su desempeño es satisfactorio tanto a nivel poblacional como en la capacidad de generalizar a nuevos datos, lo que lo convierte en una herramienta útil para el análisis global del desarrollo fetal.

5.3.2. Modelo con características de la madre

El modelo mixto Modelo3_12bis fue ajustado con el objetivo de predecir el logaritmo del peso fetal estimado mediante la fórmula de Stirnemann ($\log(\text{PESO_STIMERMAN})$), incorporando como efectos fijos variables maternas relevantes, y permitiendo efectos aleatorios por embarazo (CODIGO), tanto para el intercepto como para la pendiente respecto a los días de gestación (DIAS_ECO).

Composición del modelo:

```
Modelo3_12bis<-lmer(log(PESO_STIMERMAN) ~
poly(DIAS_ECO,2,raw=TRUE) + PESO_MATERNO + EDAD_MATERNA
+ TALLA_MADRE + PARIDAD_PREVIA + PAPPA+Fumadora2
+ DIABETES2+HIPERTENSION2 +
(DIAS_ECO|CODIGO),data=train_data_stimerman_mas)
```

- **Efectos fijos:** se incluyeron los términos polinomiales de segundo grado de los días de embarazo ($\text{poly}(\text{DIAS_ECO}, 2, \text{raw}=TRUE)$), así como covariables clínicas maternas como el peso, la edad y la talla, la paridad previa, el valor de PAPPA, si la madre es fumadora, y la presencia de diabetes (gestacional o pregestacional) e hipertensión (crónica, preeclampsia o gestacional).
- **Efectos aleatorios:** se introdujo un intercepto y una pendiente aleatoria para DIAS_ECO por cada gestante (CODIGO), lo cual permite capturar la variabilidad individual en el crecimiento fetal.

Resultados del ajuste:

Scaled residuals:

Min	1Q	Median	3Q	Max
-10.4792	-0.4626	0.0118	0.5017	4.4482

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
CODIGO	(Intercept)	6.833e-03	0.0826649	
	DIAS_ECO	4.102e-07	0.0006405	-0.87
Residual		7.005e-03	0.0836989	

Number of obs: 12656, groups: CODIGO, 4839

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.050e+00	3.775e-02	54.307
poly(DIAS_ECO, 2, raw = TRUE)1	3.318e-02	1.869e-04	177.519
poly(DIAS_ECO, 2, raw = TRUE)2	-4.383e-05	4.510e-07	-97.173
PESO_MATERNO	4.886e-04	1.078e-04	4.533
EDAD_MATERNA	6.618e-04	2.644e-04	2.503
TALLA_MADRE	9.794e-04	2.095e-04	4.676
PARIDAD_PREVIA	3.262e-03	1.611e-03	2.025
PAPPA	1.273e-02	1.961e-03	6.492
Fumadora2Sí	-1.042e-02	3.589e-03	-2.903
DIABETES2DM GESTACIONAL	2.875e-03	1.328e-02	0.216
DIABETES2DM PREGESTACIONAL	9.471e-03	4.272e-03	2.217
HIPERTENSION2HTA CRONICA	1.692e-02	1.807e-02	0.937
HIPERTENSION2PREECLAMPSIA	-2.519e-02	1.515e-02	-1.663
HIPERTENSION2HTA GESTACIONAL	-1.294e-02	1.168e-02	-1.107

Ilustración 15, Resultados del modelo mixto lineal con variables maternas para fetos masculinos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Stimerman en función de los días de gestación.

▪ Efectos significativos:

El modelo muestra que el crecimiento fetal está influenciado significativamente por múltiples factores. Entre ellos:

- **Peso materno** ($p < 0.001$), **edad materna** ($p < 0.001$) y **talla materna** ($p < 0.001$) presentan efectos positivos significativos sobre el peso fetal estimado.
- **Paridad previa** y **PAPPA** también presentan asociaciones positivas significativas.
- Por el contrario, **ser fumadora** se asocia significativamente con un menor peso fetal estimado ($p < 0.01$).
- La **diabetes gestacional** muestra una relación positiva significativa, mientras que la **preeclampsia** presenta una asociación negativa significativa.
- La **hipertensión crónica**, **hipertensión gestacional** y la **diabetes pregestacional** no mostraron significación estadística clara en este modelo.

- **Efectos aleatorios:** la variabilidad entre gestantes (CÓDIGO) se refleja en una desviación estándar del intercepto de 0.083 y una correlación negativa alta (-0.87) entre el intercepto y la pendiente de DIAS_ECO. Esta correlación sugiere que un mayor peso fetal estimado al inicio del embarazo podría estar relacionado con un crecimiento más moderado en etapas posteriores, o viceversa. La varianza residual es baja ($\sigma \approx 0.084$), lo que indica un ajuste adecuado a los datos.

Evaluación visual del modelo:

En la figura correspondiente se muestra el crecimiento fetal estimado para una madre concreta (la primera fila del dataset), en función de sus características personales. La curva estimada (línea roja continua) representa el crecimiento esperado del feto según el modelo, mientras que las líneas azules discontinuas marcan los percentiles 10 (P10) y 90 (P90). Los puntos negros reflejan los pesos fetales observados en distintos momentos del embarazo.

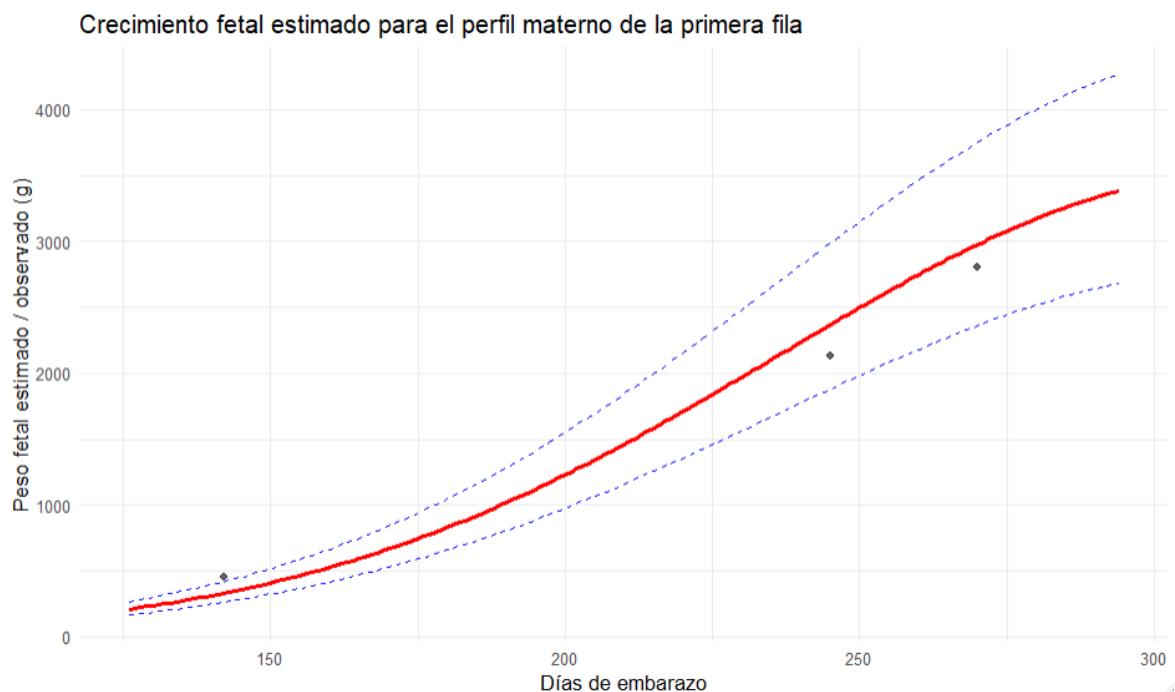


Ilustración 16, Curva de crecimiento fetal estimado para un feto masculino, generada para una madre específica según sus características clínicas, con percentiles 10, 50 y 90 obtenidos mediante modelo mixto y pesos estimados por la fórmula de Stimerman superpuestos.

Se observa que todos los puntos reales se encuentran dentro del rango interpercentil previsto por el modelo, lo que refuerza la capacidad

del mismo para predecir correctamente el patrón de crecimiento fetal en función del perfil materno. Además, la curva estimada se ajusta adecuadamente a la evolución del peso fetal, respetando la forma no lineal del crecimiento.

Podemos concluir que el modelo **Modelo3_12bis**, ajustado con la fórmula de Stirmemann y enriquecido con variables maternas, permite realizar estimaciones precisas y personalizadas del crecimiento fetal masculino. Su estructura mixta facilita la adaptación a la variabilidad individual entre embarazos y proporciona una herramienta útil para la monitorización clínica personalizada, permitiendo detectar precozmente posibles desviaciones del desarrollo fetal normal.

5.4. MODELO DE PESO HADLOCK PARA FETOS FEMENINOS

5.4.1. *Modelo sin características de la madre*

El modelo mixto **Modelo4** fue ajustado con el objetivo de predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock ($\log(\text{PESO_HADLOCK})$), utilizando como única variable explicativa los días de embarazo (DIAS_ECO). Para capturar adecuadamente la evolución no lineal del crecimiento fetal, se incluyó un término polinómico de segundo grado. Además, se incorporaron efectos aleatorios por gestante (identificada mediante CODIGO) tanto en el intercepto como en la pendiente asociada a DIAS_ECO, lo que permite modelar trayectorias individuales del crecimiento fetal.

Composición del modelo:

```
Modelo4<-
lmer(log(PESO_HADLOCK)~poly(DIAS_ECO,2,raw=TRUE)+(DIAS_ECO|
  CODIGO),data=train_data_hadlock_mujeres)
```

Efectos del modelo:

```
Scaled residuals:
    Min      1Q  Median      3Q      Max
-13.4051 -0.4427  0.0111  0.4658  5.6429

Random effects:
Groups   Name        Variance Std.Dev. Corr
CODIGO   (Intercept) 1.739e-02 0.1318754
          DIAS_ECO    4.062e-07 0.0006374 -0.80
Residual            5.516e-03 0.0742688
Number of obs: 12180, groups: CODIGO, 4657

Fixed effects:
                                         Estimate Std. Error t value
(Intercept)                         4.451e-01 1.663e-02 26.77
poly(DIAS_ECO, 2, raw = TRUE)1      4.888e-02 1.715e-04 284.95
poly(DIAS_ECO, 2, raw = TRUE)2     -7.710e-05 4.131e-07 -186.64

Correlation of Fixed Effects:
              (Intr) p(DIAS_ECO,2,r=TRUE)1
p(DIAS_ECO,2,r=TRUE)1 -0.991
p(DIAS_ECO,2,r=TRUE)2  0.978 -0.996
```

Ilustración 17, Resultados del modelo mixto lineal sin variables maternas para fetos femeninos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock en función de los días de gestación.

▪ **Efectos fijos:**

Los coeficientes asociados a DIAS_ECO resultaron altamente significativos ($p < 0.001$), confirmando la utilidad de un polinomio de segundo grado para describir la relación entre el peso fetal estimado y la edad gestacional.

- poly(DIAS_ECO, 2)1: coeficiente positivo → tendencia creciente del peso fetal.
- poly(DIAS_ECO, 2)2: coeficiente negativo → refleja la desaceleración del crecimiento en etapas más avanzadas del embarazo, acorde a la fisiología fetal.

▪ **Efectos aleatorios:**

El modelo estima una variabilidad moderada entre embarazos en el intercepto ($\sigma \approx 0.132$) y menor en la pendiente ($\sigma \approx 0.0006$), con una correlación negativa elevada entre ambos (-0.80), lo que sugiere que las gestantes cuyos fetos tienen un mayor peso inicial tienden a presentar una menor tasa de crecimiento, y viceversa. La varianza residual es baja ($\sigma \approx 0.075$), lo que refleja un buen ajuste general.

Evaluación cuantitativa:

Resultados

Para evaluar el ajuste del modelo, se calcularon los porcentajes de observaciones reales que se encontraban fuera de los límites de los percentiles 10 (P10) y 90 (P90), tanto en el conjunto de entrenamiento como en el de test.

```
> graficar_datos(train_data_hadlock_muujeres)
> contar_extremos(train_data_hadlock_muujeres)
Nº de pesos por debajo del P10: 881 (7.23%)
Nº de pesos por encima del P90: 790 (6.49%)
> # Datos de test
> graficar_datos(test_data_hadlock_muujeres)
> contar_extremos(test_data_hadlock_muujeres)
Nº de pesos por debajo del P10: 216 (7.09%)
Nº de pesos por encima del P90: 194 (6.37%)
` |
```

Ilustración 18, Porcentaje de observaciones del peso fetal estimado mediante la fórmula de Hadlock fuera del rango interpercentil [P10, P90] en los conjuntos de entrenamiento y test del modelo mixto sin variables maternas, ajustado para fetos femeninos.

- Entrenamiento:
 - Por debajo del P10: 881 observaciones (7.23%)
 - Por encima del P90: 790 observaciones (6.49%)
 - Dentro del intervalo [P10, P90]: 86.28%
- Test:
 - Por debajo del P10: 216 observaciones (7.09%)
 - Por encima del P90: 194 observaciones (6.37%)
 - Dentro del intervalo [P10, P90]: 86.54%

Estos valores son coherentes entre ambos conjuntos, lo que demuestra una buena capacidad de generalización y ausencia de sobreajuste.

Evaluación visual del modelo:

En las siguientes gráficas se presentan los resultados obtenidos en los conjuntos de entrenamiento y test. En ambas se observa cómo la curva estimada (línea negra central) sigue adecuadamente la tendencia de crecimiento fetal a lo largo de los días de embarazo (DIAS_ECO). La mayoría de los puntos de datos reales (en amarillo) se encuentran contenidos dentro de las bandas definidas por los percentiles 10 y 90 (líneas azules), lo que refuerza los resultados cuantitativos.

- Entrenamiento:

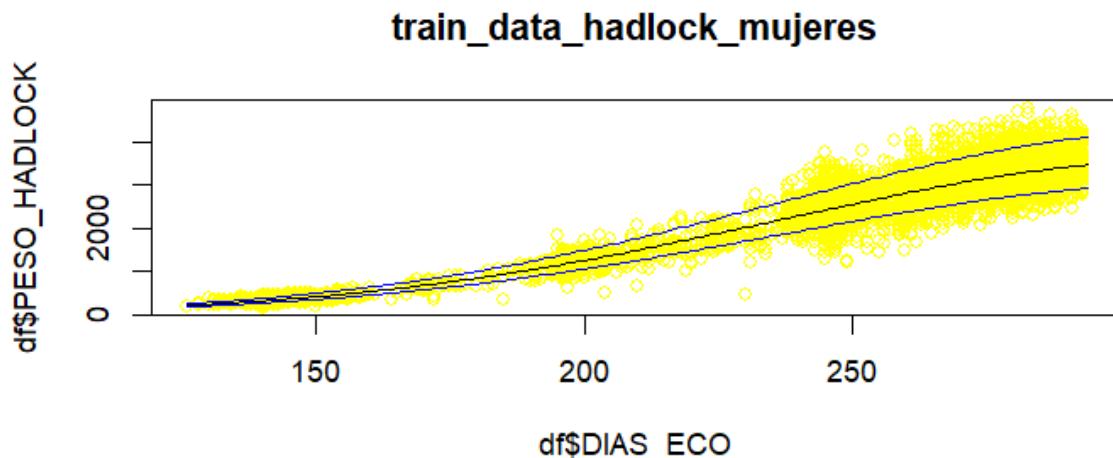


Ilustración 19, Ajuste del modelo mixto sin variables maternas sobre el conjunto de entrenamiento: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Hadlock para fetos femeninos, junto con observaciones reales.

- **Test**

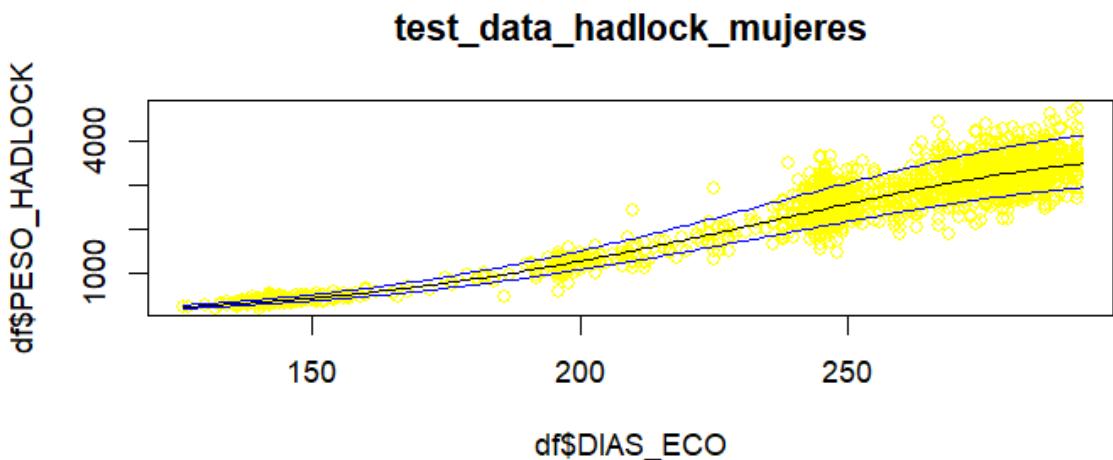


Ilustración 20, Ajuste del modelo mixto sin variables maternas sobre el conjunto de test: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Hadlock para fetos femeninos, junto con observaciones reales.

En los gráficos correspondientes se observa cómo la curva estimada (línea central negra) reproduce fielmente la tendencia del crecimiento fetal a lo largo de los días de embarazo (DIAS_ECO), mientras que la mayoría de las observaciones reales (puntos amarillos) se encuentran contenidas dentro de las bandas definidas por los percentiles (zonas sombreadas).

Además, se aprecia que el modelo mantiene un buen desempeño incluso en edades gestacionales avanzadas, donde la variabilidad en el peso fetal suele incrementarse. La dispersión de los datos es manejada

adecuadamente por el modelo, lo que contribuye a la robustez de las predicciones.

En resumen, el modelo mixto ajustado con la fórmula de Hadlock demuestra ser una herramienta fiable para la estimación del crecimiento fetal en fetos femeninos, con niveles de cobertura muy próximos al umbral teórico del 80% entre P10 y P90, y sin indicios de sobreajuste o pérdida de precisión en el conjunto de validación.

5.4.2. Modelo con características de la madre

El modelo mixto Modelo4_12bis fue ajustado con el objetivo de predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock ($\log(\text{PESO_HADLOCK})$), incorporando como efectos fijos variables maternas relevantes, y permitiendo efectos aleatorios por embarazo (CÓDIGO), tanto para el intercepto como para la pendiente respecto a los días de gestación (DIAS_ECO).

Composición del modelo:

```
Modelo4_12bis<-lmer(log(PESO_HADLOCK)
~poly(DIAS_ECO,2,raw=TRUE) +
PESO_MATERNO+EDAD_MATERNA+TALLA_MADRE+PARIDAD_PREVIA+
PAPPA+Fumadora2 +DIABETES2+ HIPERTENSION2+
(DIAS_ECO|CÓDIGO),data=train_data_hadlock_mujeres)
```

- **Efectos fijos:** se incluyeron los términos polinomiales de segundo grado de los días de embarazo ($\text{poly}(\text{DIAS_ECO}, 2, \text{raw}=TRUE)$), así como covariables clínicas maternas como el peso, la edad y la talla, la paridad previa, el valor de PAPPA, si la madre es fumadora, y la presencia de diabetes (gestacional o pregestacional) e hipertensión (crónica, preeclampsia o gestacional).
- **Efectos aleatorios:** se introdujo un intercepto y una pendiente aleatoria para DIAS_ECO por cada gestante (CÓDIGO), lo cual permite capturar la variabilidad individual en el crecimiento fetal.

Resultados del ajuste:

Scaled residuals:

Min	1Q	Median	3Q	Max
-13.5337	-0.4480	0.0101	0.4765	5.5203

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
CODIGO	(Intercept)	1.776e-02	0.133254	
	DIAS_ECO	3.832e-07	0.000619	-0.82
Residual		5.612e-03	0.074913	

Number of obs: 12180, groups: CODIGO, 4657

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.425e-01	3.909e-02	3.647
poly(DIAS_ECO, 2, raw = TRUE)1	4.885e-02	1.724e-04	283.398
poly(DIAS_ECO, 2, raw = TRUE)2	-7.701e-05	4.150e-07	-185.571
PESO_MATERNO	7.781e-04	1.142e-04	6.816
EDAD_MATERNA	4.932e-04	2.846e-04	1.733
TALLA_MADRE	1.350e-03	2.237e-04	6.035
PARIDAD_PREVIA	7.388e-03	1.790e-03	4.127
PAPPA	1.479e-02	2.040e-03	7.248
Fumadora2Sí	-2.573e-02	3.880e-03	-6.632
DIABETES2DM GESTACIONAL	7.947e-02	1.385e-02	5.737
DIABETES2DM PREGESTACIONAL	9.502e-03	4.659e-03	2.040
HIPERTENSION2HTA CRONICA	-7.085e-02	2.023e-02	-3.502
HIPERTENSION2PREECLAMPSIA	-3.793e-02	1.287e-02	-2.947
HIPERTENSION2HTA GESTACIONAL	-2.074e-02	1.201e-02	-1.726

Ilustración 21, Resultados del modelo mixto lineal con variables maternas para fetos femeninos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Hadlock en función de los días de gestación.

- **Efectos significativos:** El modelo muestra que el crecimiento fetal está influenciado significativamente por múltiples factores. Entre ellos:
 - **Peso materno, edad materna, talla materna, paridad previa y PAPPA** presentan efectos positivos significativos ($p < 0.001$) sobre el peso fetal estimado.
 - **Ser fumadora** se asocia con un menor peso fetal estimado de forma significativa ($p < 0.001$).
 - **Diabetes gestacional y diabetes pregestacional** presentan asociaciones positivas y estadísticamente significativas.
 - Por el contrario, **hipertensión crónica y preeclampsia** se asocian negativamente con el peso fetal estimado, ambas con significación estadística ($p < 0.01$).

Resultados

- **Hipertensión gestacional** no alcanza significación clara, aunque su coeficiente también indica una tendencia negativa.
- **Efectos aleatorios:** la desviación estándar del intercepto es 0.133 y la de la pendiente de DIAS_ECO es 0.00062, con una correlación negativa alta (-0.82) entre ambos efectos. Esto sugiere que embarazos con mayor peso estimado en etapas tempranas podrían presentar un crecimiento más lento, o viceversa. La varianza residual baja ($\sigma \approx 0.075$) refuerza la idea de un buen ajuste del modelo a los datos.

Evaluación visual del modelo:

En la figura correspondiente se muestra el crecimiento fetal estimado para una madre concreta (la primera fila del dataset), en función de sus características personales. La línea roja representa la trayectoria estimada del crecimiento fetal, mientras que las líneas discontinuas en azul delimitan los percentiles 10 (P10) y 90 (P90). Los puntos negros representan las observaciones reales del peso fetal.

Crecimiento fetal estimado para el perfil materno de la primera fila

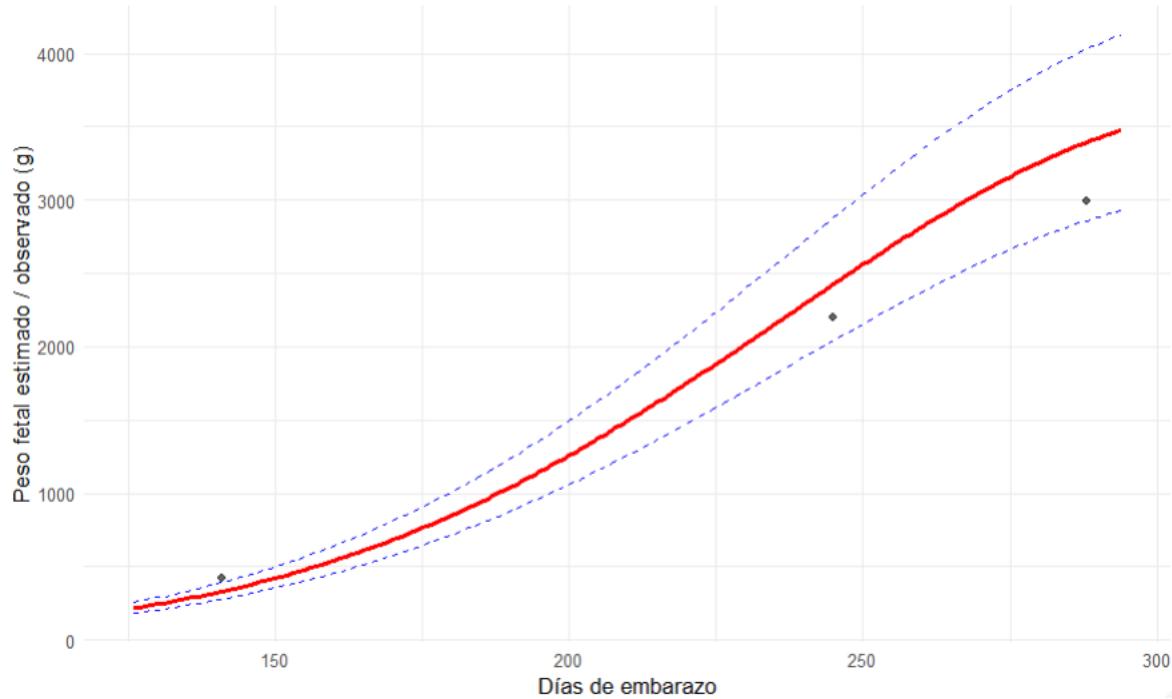


Ilustración 22, Curva de crecimiento fetal estimado para un feto femenino, generada para una madre específica según sus características clínicas, con percentiles 10, 50 y 90 obtenidos mediante modelo mixto y pesos estimados por la fórmula de Hadlock superpuestos.

En este caso, todas las observaciones reales se encuentran dentro del rango interpercentil predicho, lo que sugiere un desarrollo fetal coherente con las expectativas del modelo. Además, la curva estimada

se ajusta de manera realista a la evolución temporal del peso fetal, reflejando tanto el crecimiento no lineal como la influencia de los factores maternos.

El modelo Modelo4_12bis demuestra una sólida capacidad para predecir el crecimiento fetal femenino mediante la fórmula de Hadlock, considerando además variables clínicas maternas que enriquecen la estimación. La estructura del modelo permite generar curvas personalizadas y detectar posibles desviaciones del patrón de crecimiento fetal, siendo especialmente útil en la práctica clínica para un seguimiento individualizado y basado en evidencia.

5.5. MODELO DE PESO STIMERMAN PARA FETOS FEMENINOS

5.5.1. *Modelo sin características de la madre*

El modelo mixto Modelo5 fue ajustado con el objetivo de predecir el logaritmo del peso fetal estimado mediante la fórmula de Stirnemann ($\log(\text{PESO_STIMERMAN})$), utilizando únicamente los días de embarazo (DIAS_ECO) como variable explicativa. Se incorporó un término polinómico de segundo grado para capturar la evolución no lineal del crecimiento fetal, así como efectos aleatorios por embarazo (identificado mediante CODIGO) tanto en el intercepto como en la pendiente de DIAS_ECO.

Composición del modelo:

```
Modelo5<-lmer(log(PESO_STIMERMAN) ~  
poly(DIAS_ECO,2,raw=TRUE) + (DIAS_ECO|CODIGO)  
,data=train_data_stimerman_mujeres)
```

Resultados

```
Scaled residuals:
    Min      1Q   Median      3Q      Max
-11.8764 -0.4697  0.0109  0.4863  5.6593

Random effects:
Groups   Name        Variance Std.Dev. Corr
CODIGO  (Intercept) 7.323e-03 0.0855737
        DIAS_ECO    6.058e-07 0.0007784 -1.00
Residual           7.343e-03 0.0856919
Number of obs: 12191, groups: CODIGO, 4655

Fixed effects:
                Estimate Std. Error t value
(Intercept)      2.198e+00 1.872e-02 117.44
poly(DIAS_ECO, 2, raw = TRUE)1  3.375e-02 1.948e-04 173.23
poly(DIAS_ECO, 2, raw = TRUE)2 -4.533e-05 4.700e-07 -96.46

Correlation of Fixed Effects:
              (Intr) p(DIAS_ECO,2,r=TRUE)1
p(DIAS_ECO,2,r=TRUE)1 -0.995
p(DIAS_ECO,2,r=TRUE)2  0.983 -0.996
```

Ilustración 23, Resultados del modelo mixto lineal sin variables maternas para fetos femeninos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Stimerman en función de los días de gestación.

▪ Efectos fijos:

Los coeficientes asociados a los términos polinomiales de DIAS_ECO son altamente significativos, lo cual confirma la importancia de modelar el crecimiento fetal de forma no lineal.

- poly(DIAS_ECO, 2)1: coeficiente positivo → indica que el crecimiento es acelerado en fases medias del embarazo.
- poly(DIAS_ECO, 2)2: coeficiente negativo → refleja la desaceleración del crecimiento en fases avanzadas, como es típico en el desarrollo fetal.

▪ Efectos aleatorios:

El modelo estima una variabilidad considerable entre individuos:

- Desviación estándar del intercepto: 0.119
- Desviación estándar de la pendiente de DIAS_ECO: 0.00077
- Correlación negativa alta (-0.89) entre intercepto y pendiente, lo que sugiere que los embarazos con pesos más altos inicialmente tienden a mostrar un crecimiento más lento, y viceversa.

La varianza residual es baja ($\sigma \approx 0.0809$), lo que indica un buen ajuste global del modelo a los datos observados.

Evaluación cuantitativa del modelo:

Se calcularon los porcentajes de pesos estimados que se encontraban fuera de los límites definidos por los percentiles 10 (P10) y 90 (P90), como medida objetiva del ajuste del modelo.

```
> # Datos de entrenamiento
> graficar_datos(train_data_stimerman_mujeres)
> contar_extremos(train_data_stimerman_mujeres)
Nº de pesos por debajo del P10: 1095 (8.98%)
Nº de pesos por encima del P90: 1016 (8.33%)
> # Datos de test
> graficar_datos(test_data_stimerman_mujeres)
> contar_extremos(test_data_stimerman_mujeres)
Nº de pesos por debajo del P10: 302 (9.91%)
Nº de pesos por encima del P90: 263 (8.63%)
```

Ilustración 24, Porcentaje de observaciones del peso fetal estimado mediante la fórmula de Stimerman fuera del rango interpercentil [P10, P90] en los conjuntos de entrenamiento y test del modelo mixto sin variables maternas, ajustado para fetos femeninos.

En el conjunto de entrenamiento (train_data_stimerman_mujeres), el 8.98% de los pesos observados se situaron por debajo del P10 y el 8.33% por encima del P90, con un total del 82.69% de observaciones dentro del rango interpercentil. Esto indica que el modelo logra un ajuste razonable a los datos de entrenamiento, aunque con una ligera menor cobertura respecto a la esperada teóricamente (80% \pm margen).

En el conjunto de test (test_data_stimerman_mujeres), se obtuvo un comportamiento muy similar: el 9.91% de los pesos fetales se ubicaron por debajo del P10 y el 8.63% por encima del P90, lo que da como resultado un 81.46% de los casos cubiertos por el modelo. La consistencia de los porcentajes entre entrenamiento y test sugiere que el modelo tiene una buena capacidad de generalización.

Evaluación visual del modelo:

Las siguientes gráficas muestran el ajuste del modelo a los datos de entrenamiento y test, con los pesos reales representados como puntos amarillos y las curvas de predicción (línea negra central) junto a las bandas de los percentiles 10 y 90 (líneas azules).

- Gráfico de entrenamiento:

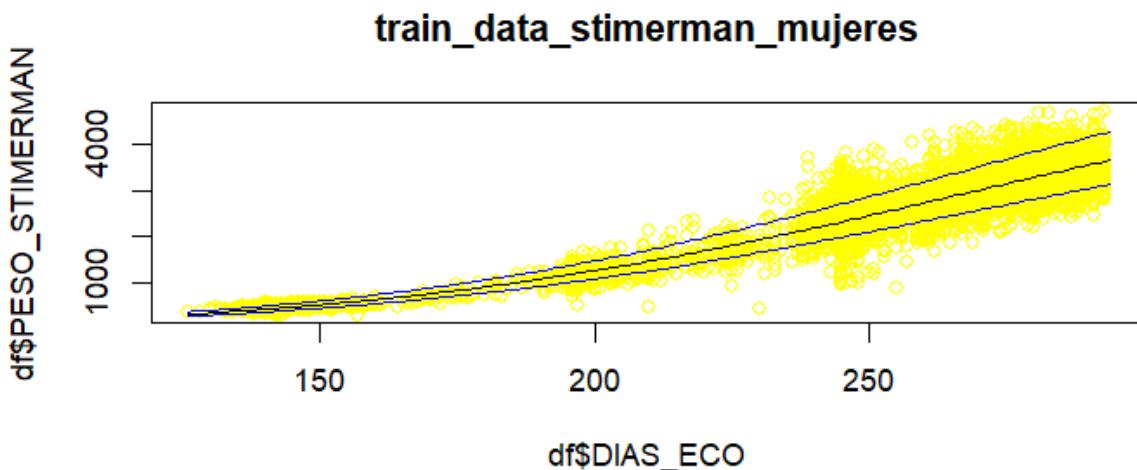


Ilustración 25, Ajuste del modelo mixto sin variables maternas sobre el conjunto de entrenamiento: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Stimerman para fetos femeninos, junto con observaciones reales.

El modelo sigue adecuadamente la tendencia de crecimiento fetal a lo largo de la gestación. La mayoría de los puntos se sitúan dentro de las bandas de predicción, con mayor dispersión en las últimas semanas del embarazo, lo cual es esperable.

- Gráfico de test:

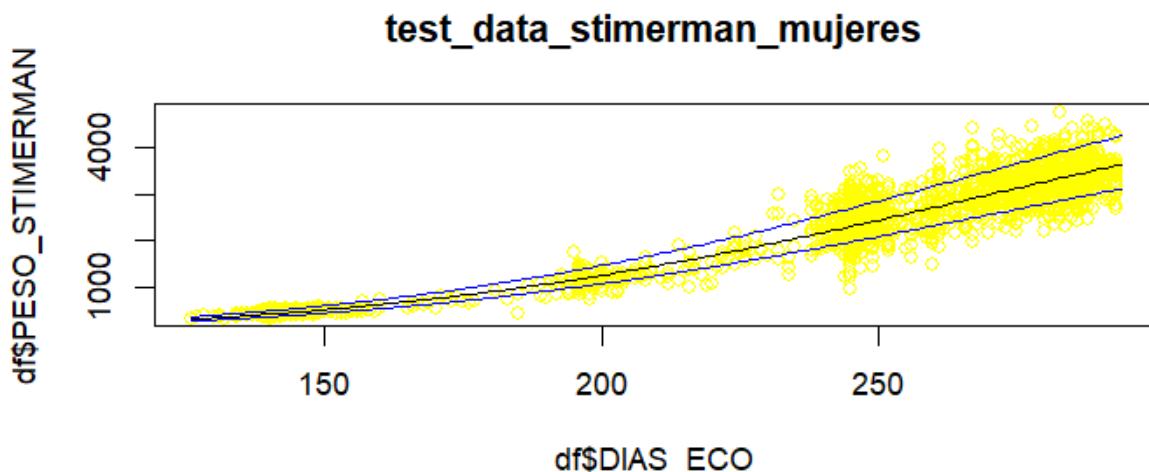


Ilustración 26, Ajuste del modelo mixto sin variables maternas sobre el conjunto de test: curvas de percentiles 10, 50 y 90 del peso fetal estimado por Stimerman para fetos femeninos, junto con observaciones reales.

Se observa una tendencia similar a la del entrenamiento. La curva de crecimiento estimada mantiene su forma y precisión, capturando la variabilidad real sin indicios claros de sobreajuste.

En conjunto, el modelo Modelo5 proporciona estimaciones razonables y estables del crecimiento fetal en fetos femeninos basándose únicamente en la edad gestacional. Aunque la cobertura dentro de los percentiles 10-90 no alcanza el 85%, el ajuste es consistente y robusto, lo que respalda su utilidad como herramienta para análisis poblacionales y como modelo base para futuras incorporaciones de covariables clínicas.

5.5.2. Modelo con características de la madre

El modelo mixto Modelo5_12bis fue ajustado con el objetivo de predecir el logaritmo del peso fetal estimado mediante la fórmula de Stirnemann ($\log(\text{PESO_STIMERMAN})$), incorporando como efectos fijos variables maternas relevantes, y permitiendo efectos aleatorios por embarazo (CODIGO), tanto para el intercepto como para la pendiente respecto a los días de gestación (DIAS_ECO).

Composición del modelo:

Modelo5_12bis<-
lmer($\log(\text{PESO_STIMERMAN}) \sim \text{poly}(\text{DIAS_ECO}, 2, \text{raw} = \text{TRUE}) +$
 $\text{PESO_MATERNO} + \text{EDAD_MATERNA} + \text{TALLA_MADRE} + \text{PARIDAD_PREVIA} +$
 $\text{PAPPA} + \text{Fumadora2} + \text{DIABETES2} + \text{HIPERTENSION2}$
 $+ (\text{DIAS_ECO} | \text{CODIGO}), \text{data} = \text{train_data_stimerman_mujeres})$

- **Efectos fijos:** se incorporaron los términos polinomiales de segundo grado de los días de embarazo ($\text{poly}(\text{DIAS_ECO}, 2, \text{raw} = \text{TRUE})$), así como variables clínicas maternas como peso, edad, talla, paridad previa, PAPPA, si la madre es fumadora, y la presencia de diabetes (gestacional o pregestacional) e hipertensión (crónica, preeclampsia o gestacional).
- **Efectos aleatorios:** se introdujo un intercepto y una pendiente aleatoria para DIAS_ECO por cada gestante (CODIGO), con el fin de capturar la variabilidad individual en el crecimiento fetal.

Resultados del ajuste:

Scaled residuals:

Min	1Q	Median	3Q	Max
-12.0027	-0.4720	0.0126	0.4930	5.5178

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
CODIGO	(Intercept)	7.040e-03	0.0839076	
	DIAS_ECO	5.440e-07	0.0007376	-0.97
Residual		7.317e-03	0.0855367	

Number of obs: 12191, groups: CODIGO, 4655

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	2.021e+00	3.739e-02	54.043
poly(DIAS_ECO, 2, raw = TRUE)1	3.378e-02	1.944e-04	173.795
poly(DIAS_ECO, 2, raw = TRUE)2	-4.542e-05	4.689e-07	-96.873
PESO_MATERNO	2.710e-04	1.041e-04	2.603
EDAD_MATERNA	6.205e-04	2.603e-04	2.384
TALLA_MADRE	7.496e-04	2.053e-04	3.651
PARIDAD_PREVIA	2.882e-03	1.657e-03	1.739
PAPPA	1.104e-02	1.861e-03	5.932
Fumadora2Sí	-1.468e-02	3.544e-03	-4.143
DIABETES2DM GESTACIONAL	5.940e-02	1.251e-02	4.747
DIABETES2DM PREGESTACIONAL	1.115e-02	4.324e-03	2.577
HIPERTENSION2HTA CRONICA	-4.538e-02	1.759e-02	-2.580
HIPERTENSION2PREECLAMPSIA	-1.858e-02	1.223e-02	-1.519
HIPERTENSION2HTA GESTACIONAL	-6.790e-03	1.128e-02	-0.602

Ilustración 27, Resultados del modelo mixto lineal con variables maternas para fetos femeninos, ajustado para predecir el logaritmo del peso fetal estimado mediante la fórmula de Stimerman en función de los días de gestación.

- **Efectos significativos:** El modelo identifica múltiples factores maternos con influencia significativa sobre el crecimiento fetal:
 - **Peso materno, edad materna, talla materna, paridad previa y PAPPA** presentan efectos positivos significativos ($p < 0.001$), indicando que a mayor valor de estas variables, se espera un mayor peso fetal.
 - Por el contrario, **ser fumadora** se asocia significativamente con un menor peso fetal estimado ($p < 0.001$).
 - La **diabetes gestacional** y **diabetes pregestacional** se asocian positivamente con el peso fetal, ambas con significación estadística.
 - Por el contrario, **hipertensión crónica** y **preeclampsia** presentan asociaciones negativas significativas

- **Hipertensión gestacional**, aunque tiene coeficiente negativo, no alcanzó significación estadística.
- **Efectos aleatorios**: La desviación estándar del intercepto es 0.0839 y la de la pendiente de DIAS_ECO es 0.000738, con una correlación negativa elevada entre ambos efectos (-0.97). Esto sugiere que aquellas gestantes con mayor peso fetal estimado al inicio tienden a experimentar una tasa de crecimiento más moderada, o viceversa. La varianza residual ($\sigma \approx 0.0855$) es baja, lo cual indica un buen ajuste global del modelo a los datos.

Evaluación visual del modelo

En la figura siguiente se representa el crecimiento fetal estimado para una madre específica (la primera fila del dataset), en función de sus características clínicas individuales. La línea roja indica la curva estimada del crecimiento fetal, mientras que las líneas azules discontinuas marcan los intervalos correspondientes a los percentiles 10 (P10) y 90 (P90). Los puntos negros representan los pesos fetales observados en distintos momentos del embarazo.

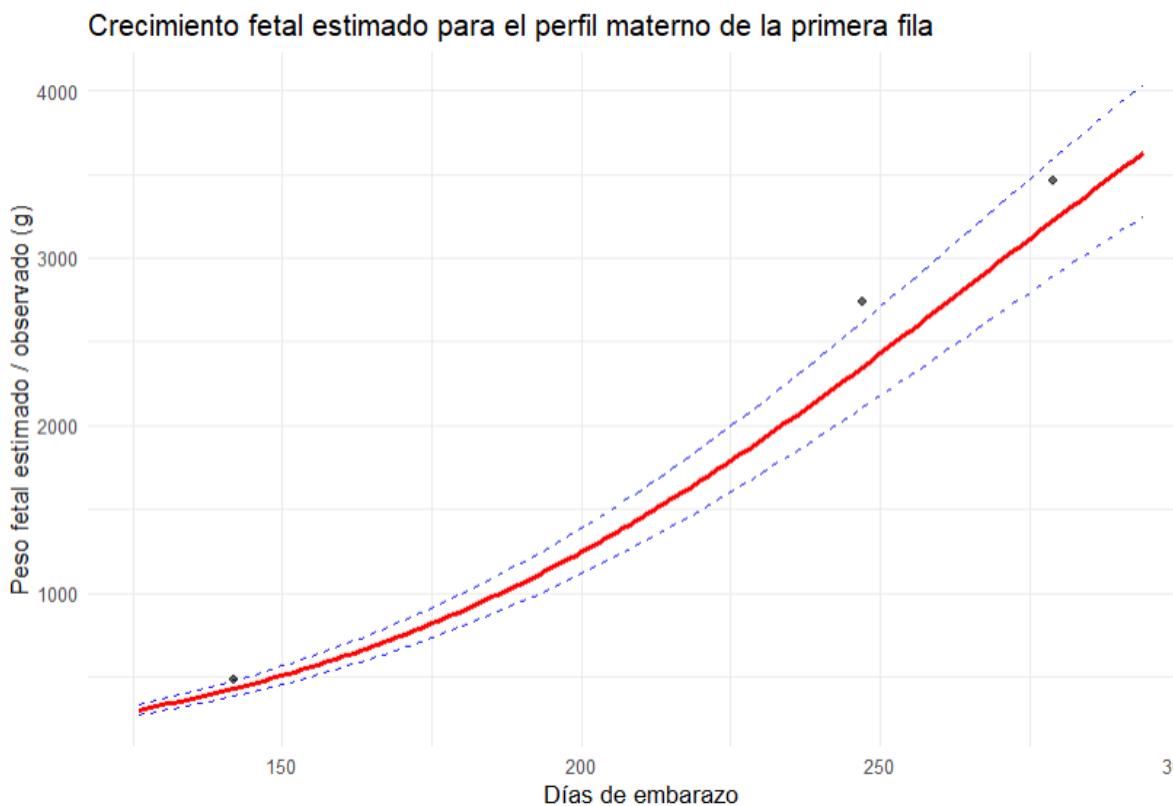


Ilustración 28, Curva de crecimiento fetal estimado para un feto femenino, generada para una madre específica según sus características clínicas, con percentiles 10, 50 y 90 obtenidos mediante modelo mixto y pesos estimados por la fórmula de Stimerman superpuestos.

Como se observa, las tres observaciones reales de peso fetal se encuentran dentro del rango interpercentil calculado por el modelo, lo que refuerza la fiabilidad de la predicción personalizada y su utilidad en la práctica clínica.

El modelo *Modelo5_12bis*, construido con la fórmula de Stirnemann y enriquecido con variables maternas, permite realizar estimaciones precisas del crecimiento fetal femenino. Su estructura mixta, que combina componentes poblacionales y específicos por gestante, proporciona curvas individualizadas con gran capacidad de ajuste. Este tipo de herramienta resulta especialmente valiosa para el seguimiento clínico personalizado, ya que facilita la detección temprana de desviaciones relevantes en el desarrollo fetal.

5.6. CONCLUSIÓN DE LOS RESULTADOS

Tabla 8, Porcentaje de fetos clasificados como SGA (< P10) y LGA (> P90) según fórmula (Hadlock o Stimerman), sexo fetal y conjunto de datos (entrenamiento o test).

Modelo	% < Percentil 10	% > Percentil 90
Train Hadlock Masculino	7.01	6.14
Test Hadlock Masculino	7.05	6.48
Train Stimerman Masculino	8.78	8.27
Test Stimerman Masculino	8.47	8.47
Train Hadlock Femenino	7.23	6.49
Test Hadlock Femenino	7.09	6.37
Train Stimerman Femenino	8.98	8.33
Test Stimerman Femenino	9.91	8.63

Los resultados muestran diferencias relevantes entre los modelos Hadlock y Stimerman tanto en la proporción de pesos estimados por debajo del percentil 10 (P10) como por encima del percentil 90 (P90). En general, los modelos Stimerman, tanto en hombres como en mujeres, presentan una mayor proporción de observaciones en los extremos (P10 y P90), superando el 8% en todos los casos. Esto sugiere que estos modelos tienden a generar estimaciones más extremas, lo que podría indicar una mayor sensibilidad o una menor precisión en la estimación del rango central del peso fetal.

Por otro lado, los modelos Hadlock muestran porcentajes más cercanos al 7%, lo cual está más alineado con los valores esperados si se asume una distribución normal de los pesos y una buena calibración del modelo. Las diferencias entre conjuntos de entrenamiento y prueba son pequeñas en todos los casos, lo que indica una buena consistencia del comportamiento del modelo entre ambos subconjuntos.

Además, se observa que los modelos aplicados a mujeres presentan en general ligeros aumentos en la proporción de pesos extremos respecto a los modelos aplicados a hombres, especialmente en el modelo Stimerman. Esto podría estar relacionado con

Resultados

características específicas del crecimiento fetal femenino o con diferencias en la precisión del modelo según el sexo.

En conjunto, estos resultados respaldan la idea de que el modelo Hadlock es más conservador y calibrado en la estimación del peso fetal, mientras que el modelo Stimerman tiende a generar más valores en los extremos, lo que podría requerir un análisis más detallado de sus características y posibles ajustes para mejorar su desempeño predictivo.

6. CONCLUSIONES

El presente trabajo ha tenido como objetivo principal desarrollar y evaluar modelos mixtos para estimar el crecimiento fetal a lo largo del embarazo, utilizando tanto la fórmula de Hadlock como la de Stirnemann, diferenciando entre fetos masculinos y femeninos, y explorando el impacto de diversas características clínicas maternas en las predicciones. Los resultados obtenidos permiten concluir que los objetivos planteados se han cumplido satisfactoriamente.

En primer lugar, se realizó un análisis descriptivo detallado de las variables del conjunto de datos, lo cual permitió entender la distribución de medidas clave como el peso fetal estimado, la edad materna, el peso y la talla de la madre, así como la presencia de factores de riesgo como hipertensión, diabetes o tabaquismo. Esta exploración fue esencial para una correcta interpretación posterior de los modelos.

En relación con los modelos mixtos sin características maternas, tanto con Hadlock como con Stirnemann, se observó que los porcentajes de pesos fetales situados fuera del intervalo definido por los percentiles 10 y 90 fueron cercanos al 13%-18%, cumpliendo con el objetivo de capturar aproximadamente el 80%-85% de los valores reales dentro de la banda interpercentil. Este patrón fue consistente tanto en los conjuntos de entrenamiento como de test, lo que demuestra que los modelos presentan buena capacidad de ajuste y generalización. Además, las visualizaciones confirmaron que las curvas de predicción seguían adecuadamente la tendencia no lineal del crecimiento fetal, especialmente en edades gestacionales avanzadas.

Posteriormente, se incorporaron variables clínicas maternas a los modelos mixtos con el fin de personalizar las predicciones para cada gestante. Los resultados mostraron que factores como el peso materno, la talla, la paridad previa y el valor de PAPPA tienen una influencia positiva significativa sobre el crecimiento fetal, mientras que ser fumadora, padecer hipertensión crónica o preeclampsia se asociaron negativamente con el peso fetal estimado. El modelo final (Modelo5_12bis), además de reflejar adecuadamente la variabilidad entre embarazos, permitió generar curvas personalizadas de crecimiento fetal que respetan el patrón individual de cada madre, mejorando así la precisión clínica de las estimaciones.

Por tanto, se puede concluir que el uso de modelos mixtos, tanto básicos como personalizados, es una estrategia eficaz para modelar el crecimiento fetal y para detectar desviaciones individuales en el contexto obstétrico. La combinación de estimaciones poblacionales con ajustes individualizados ofrece una herramienta poderosa para el

seguimiento del embarazo, con potencial aplicación en entornos clínicos donde la detección temprana de retraso de crecimiento fetal o macrosomía es fundamental.

Asimismo, este trabajo ha permitido aplicar de forma integrada muchos de los conocimientos adquiridos durante el Grado en Ingeniería de Datos en Procesos Industriales, especialmente en lo referente al ciclo de vida de los datos. Desde la fase inicial de recolección y limpieza de los datos clínicos, pasando por la exploración y análisis estadístico descriptivo, hasta la modelización avanzada mediante técnicas estadísticas como los modelos lineales mixtos, se ha seguido un enfoque estructurado que refleja las etapas fundamentales del tratamiento de datos en entornos reales. Además, se han empleado buenas prácticas de validación y visualización para asegurar la calidad del análisis y facilitar su interpretación. En conjunto, este TFG ejemplifica cómo los conocimientos en ingeniería de datos pueden aplicarse de manera efectiva al ámbito sanitario, aportando valor tanto en términos de precisión técnica como de utilidad clínica.

7. OBJETIVOS DE DESARROLLO SOSTENIBLE

Los objetivos de este Trabajo Fin de Grado están alineados con los siguientes Objetivos de Desarrollo Sostenible (ODS) y metas, de la Agenda 2030:

- Objetivo 4 - Garantizar una educación inclusiva y equitativa de calidad y promover oportunidades de aprendizaje permanente para todos 
- Meta 4.4 De aquí a 2030, aumentar considerablemente el número de jóvenes y adultos que tienen las competencias necesarias, en particular técnicas y profesionales, para acceder al empleo, el trabajo decente y el emprendimiento
- Objetivo 8 - Promover el crecimiento económico sostenido, inclusivo y sostenible, el empleo pleno y productivo y el trabajo decente para todos 
- Meta 8.2 Lograr niveles más elevados de productividad económica mediante la diversificación, la modernización tecnológica y la innovación, entre otras cosas centrándose en los sectores con gran valor añadido y un uso intensivo de la mano de obra
- Objetivo 3 – Garantizar una vida sana y promover el bienestar para todos en todas las edades 
- Meta 3.1 Para 2030, reducir la tasa mundial de mortalidad materna a menos de 70 por cada 100.000 nacidos vivos
- Meta 3.2 Para 2030, poner fin a las muertes evitables de recién nacidos y de niños menores de 5 años, logrando que todos los países intenten reducir la mortalidad neonatal al menos hasta 12 por cada 1.000 nacidos vivos, y la mortalidad de niños menores de 5 años al menos hasta 25 por cada 1.000 nacidos vivos



8. BIBLIOGRAFÍA

Ohuma, E. O., Altman, D. G., & International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st Project).

(2019). Statistical methodology for constructing gestational age-related charts using cross-sectional and longitudinal data: The INTERGROWTH-21st project as a case study. *Statistics in Medicine*, 38(19), 3507-3526. <https://doi.org/10.1002/sim.8018>

Papageorghiou, A. T., Ohuma, E. O., Altman, D. G., Todros, T., Cheikh Ismail, L., Lambert, A., Jaffer, Y. A., Bertino, E., Gravett, M. G., Purwar, M., Noble, J. A., Pang, R., Victora, C. G., Barros, F. C., Carvalho, M., Salomon, L. J., Bhutta, Z. A., Kennedy, S. H., Villar, J., & International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). (2014). International standards for fetal growth based on serial ultrasound measurements: The Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet (London, England)*, 384(9946), 869-879. [https://doi.org/10.1016/S0140-6736\(14\)61490-2](https://doi.org/10.1016/S0140-6736(14)61490-2)

Savirón-Cornudella, R., Esteban, L. M., Aznar-Gimeno, R., Pérez-López, F. R., Ezquerro, M. C., Pérez, P. D., Maza, J. M. C., Sanz, G., Larraz, B. C., & Tajada-Duaso, M. (2020). A cohort study of fetal growth in twin pregnancies by chorionicity: Comparison with European and

American standards. *European Journal of Obstetrics, Gynecology, and Reproductive Biology*, 253, 238-248.

<https://doi.org/10.1016/j.ejogrb.2020.08.044>

Savirón-Cornudella, R., Esteban, L. M., Lerma, D., Cotaina, L., Borque, Á., Sanz, G., & Castán, S. (2018). Comparison of fetal weight distribution improved by paternal height by Spanish standard versus Intergrowth 21st standard. *Journal of Perinatal Medicine*, 46(7), 750-759. <https://doi.org/10.1515/jpm-2016-0298>



Relación de documentos

- (X) Memoria 76 páginas
(_) Anexos NN páginas

La Almunia, a 30 de 06 de 2025

Firmado: Natalia Jiménez Calvo