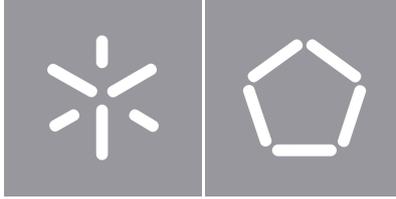


Universidade do Minho
Escola de Engenharia

Daniela Filipa Machado Lemos

**Regressão não linear em curvas de crescimento
para parâmetros placentares em R**



Universidade do Minho
Escola de Engenharia

Daniela Filipa Machado Lemos

**Regressão não linear em curvas de crescimento
para parâmetros placentares em R**

Dissertação de Mestrado
Mestrado em Bioinformática

Trabalho efetuado sob a orientação de
Professora Doutora Ana Cristina da Silva Braga
Professora Doutora Rosete Maria Amorim Novais
Nogueira Cardoso

Direitos de Autor e Condições de Utilização do Trabalho por Terceiros

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho:



Atribuição-NãoComercial-SemDerivações

CC BY-NC-ND

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Agradecimentos

Gostaria de expressar os meus sinceros agradecimentos a todos que contribuíram para a realização deste trabalho e para a conclusão bem-sucedida do meu mestrado.

Primeiramente, quero expressar a minha gratidão à minha orientadora, Professora Doutora Ana Cristina Braga, pela amizade, orientação e apoio ao longo deste projeto. A sua disponibilidade, partilha de conhecimento e conselhos foram inestimáveis na condução deste estudo. Agradeço não apenas pela confiança que depositou em mim desde o início, mas também por me incentivar a abraçar desafios, a explorar novos caminhos e a acreditar no meu potencial.

Também gostaria de agradecer à minha coorientadora, Professora Doutora Rosete Nogueira, pela sua amizade, apoio, conselhos e contribuições para este estudo. A sua orientação adicional enriqueceu profundamente esta investigação.

Agradeço ao laboratório de Patologia Embrionária Fetal do Centro de Genética Clínica no Porto, por generosamente ceder os dados essenciais para a condução deste estudo.

À minha família, principalmente ao meu pai, mãe e irmã, agradeço do fundo do coração pelo seu amor incondicional, encorajamento constante e apoio inabalável ao longo de toda a minha jornada académica. Sem vocês, nada disto teria sido possível.

Ao meu namorado, quero expressar a minha gratidão pelo incentivo, apoio e carinho durante todos os momentos deste percurso.

Aos meus amigos, quero agradecer pelo seu apoio contínuo, amizade e por todos os momentos ao longo destes anos.

Este trabalho é dedicado a todos vocês, e estou profundamente grata por fazerem parte desta jornada.

Declaração de Integridade

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

Universidade do Minho, Braga, dezembro 2023

Daniela Filipa Machado Lemos

Resumo

Nos últimos anos, tem existido um crescente interesse na avaliação dos parâmetros biométricos da placenta e a sua relação com resultados obstétricos. Evidências têm sido publicadas sugerindo que as medidas da placenta e a sua evolução são capazes de refletir alterações no desenvolvimento do feto e até mesmo doenças do recém-nascido e do adulto. Tendo em conta, que os gráficos de crescimento desempenham um papel crucial na avaliação e vigilância da população pediátrica, surgiu o tema desta dissertação. O principal objetivo passa por estudar a aplicabilidade de modelos de regressão não linear em curvas de crescimento de parâmetros como o Diâmetro 1 (D1) e 2 (D2), Espessura (EP) e Peso placentar (PP) e Peso fetal (PF). Para isso, utilizou-se um conjunto de dados de parturientes portuguesas recolhidos no CGC (Centro de Genética Clínica), Porto.

Neste estudo foi utilizada uma abordagem de regressão semiparamétrica para a construção de curvas de crescimento de referência. Esta metodologia utiliza os modelos aditivos generalizados para localização, escala e forma (GAMLSS), Lamda-Mu-Sigma (LMS), LMS com *Box-Cox t* (BCT) e LMS com *Box-Cox-powe-exponential* (BCPE), oferecendo vantagens distintas sobre os métodos tradicionais como a regressão quantílica. Uma das principais vantagens do GAMLSS é a sua flexibilidade para acomodar qualquer distribuição estatística, permitindo a modelação de vários parâmetros biométricos.

Através da aplicação da metodologia proposta, foi demonstrado que com a utilização do método GAMLSS com BCT e *P-splines* para os parâmetros D1 e D2 e do método LMST para os parâmetros EP, PP e PF, podemos alcançar curvas de crescimento representativas. Além disso, foi desenvolvida uma aplicação *web*, disponível em https://placentalgrowth.shinyapps.io/uminho_pt/, utilizando o ambiente R. Permite que profissionais de saúde e investigadores analisem e interpretem as curvas de crescimento desenvolvidas com facilidade. Os resultados deste estudo fornecem informações importantes sobre o desenvolvimento da placenta e têm implicações significativas para a prática clínica em obstetria, permitindo o seu avanço e acompanhamento da saúde materno-fetal.

Palavras-chave Placenta, Curvas de crescimento, Regressão não linear, GAMLSS, LMST

Abstract

In recent years, there has been a growing interest in assessing the biometric parameters of the placenta and their relationship with obstetric outcomes. Evidence has been published suggesting that placental measures and their evolution can reflect changes in fetal development and even impact newborn and adult health. Considering the crucial role that growth charts play in the evaluation and monitoring of the paediatric population, the topic of this dissertation emerged. The main objective was to study the applicability of nonlinear regression models in growth curves for parameters such as Diameter 1 (D1) and 2 (D2), Thickness (PT), Placental weight (PW), and Fetal weight (FW). For this purpose, a dataset of Portuguese parturients collected at CGC (Center for Clinical Genetics), Porto, was used.

This study employed a semi-parametric regression approach to construct reference growth curves. This methodology utilizes generalized additive models for location, scale, and shape (GAMLSS), Lambda-Mu-Sigma (LMS), LMS with Box-Cox t (BCT), and LMS with Box-Cox-power-exponential (BCPE), offering distinct advantages over traditional methods such as quantile regression. One of the key advantages of GAMLSS is its flexibility to accommodate any statistical distribution, allowing the modeling of various biometric parameters.

Through the application of the proposed methodology, it was demonstrated that using the GAMLSS method with BCT and P-splines to D1 and D2 parameters and the LMST method to PT, PW, and FW parameters, representative growth curves can be achieved. Furthermore, a web application, available at https://placentalgrowth.shinyapps.io/uminho_pt/, was developed using the R environment. It enables healthcare professionals and researchers to easily analyze and interpret the developed growth curves. The results of this study provide valuable insights into placental development and have significant implications for clinical practice in obstetrics, promoting its advancement and the monitoring of maternal-fetal health.

Keywords Placenta, Growth curves, Nonlinear regression, GAMLSS, LMST

Conteúdo

1	Introdução	1
1.1	Enquadramento e motivação	1
1.2	Objetivos	4
1.3	Estrutura da dissertação	5
2	Estado da arte	7
2.1	Placenta humana	7
2.1.1	Introdução histórica	7
2.1.2	Definição e características gerais	8
2.1.3	Implicações dos parâmetros placentares e fetais	10
2.2	Curvas de crescimento	12
2.2.1	Introdução histórica	12
2.2.2	Regressão não linear	19
2.2.3	Definição e características gerais das curvas de crescimento	20
2.2.4	Aplicação de curvas de crescimento em várias áreas	21
2.3	Software R	24
2.3.1	Shiny	25
2.3.2	Pacotes	26
2.3.3	Aplicação do R	26
3	Regressão	28
3.1	Conceito e definição de regressão	28
3.2	Regressão linear	29
3.3	Regressão não linear	30
3.4	Regressão paramétrica, semi e não paramétrica	32
3.4.1	Regressão paramétrica	32

3.4.2	Regressão não paramétrica	32
3.4.3	Regressão semiparamétrica	33
4	Métodos para a construção das curvas de crescimento	35
4.1	Considerações metodológicas	35
4.1.1	Aspetos distributivos	35
4.1.2	Aspetos de alisamento	36
4.2	Seleção de métodos	36
4.2.1	CrITÉrios para a seleção dos métodos	37
4.2.2	Métodos	38
4.3	GAMLSS - Modelos Aditivos Generalizados para a Localização, Escala e Forma	39
4.3.1	Modelo estatístico	39
4.3.2	Algoritmos	42
4.3.3	Estimação dos hiperparâmetros λ	44
4.3.4	Termos aditivos	44
4.3.5	Alisamento	45
4.3.6	Distribuições	49
4.4	LMS e as suas extensões LMST e LMSP	52
4.5	Estimativa de percentis	53
4.6	CrITÉrios de qualidade de ajuste	54
4.6.1	Coeficiente de Determinação	54
4.6.2	Erro Quadrático Médio	55
4.6.3	CrITÉrio de Informação de Akaike	56
4.6.4	CrITÉrio de Informação Bayesiano	57
4.6.5	Estatísticas Z	58
4.6.6	Resíduos quantílicos normalizados	59
4.7	Pacotes	62
4.8	Visão geral da metodologia implementada na construção das curvas de crescimento	64
5	Preparação dos dados para análise e modelação	66
5.1	Conjunto de dados	66
5.2	Pré-processamento	67
6	Desenvolvimento da aplicação web	70

6.1	Estratégia de desenvolvimento e ferramentas	70
6.2	Arquitetura da aplicação	72
6.2.1	Funções e eventos reativos	73
7	Resultados	76
7.1	Análise exploratória dos dados	76
7.2	Testes de significância entre géneros	86
7.3	Curvas de crescimento	87
7.3.1	Diâmetro 1	88
7.3.2	Diâmetro 2	93
7.3.3	Espessura placentar	97
7.3.4	Peso placentar	102
7.3.5	Peso fetal	106
7.4	Aplicação <i>web</i>	110
8	Discussão	116
9	Conclusões e trabalho futuro	125
A	Distribuições	141
B	Dados	144
C	Valores-p dos testes de significância de géneros	147
D	Diâmetro 1	148
E	Diâmetro 2	150
F	Espessura placentar	152
G	Peso placentar	154
H	Peso fetal	156
I	Aplicação <i>web</i>	158

Lista de Figuras

1	Passos para produzir um relatório de um exame patológico da placenta com relevância clínica. Adaptado de Turowski et al. [2018].	9
2	Imagens reais de ultrassom captadas pelos médicos: Conceição Brito, Cristina Godinho, Mónica e Márcia Marinho (DPN, CHVNG/E-Portugal). Apresentam os diâmetros da placenta (linha a vermelho), o maior diâmetro da placenta (D1) e o menor diâmetro da placenta (D2).	10
3	Gráfico de crescimento. Representa uma série de distribuições normais em idades diferentes com linhas conectando a mediana (50.º percentil) e outros percentis. Adaptado de: Wright [2002].	22
4	Representação da estrutura dos algoritmos GAMLSS CG e RS em três partes. Adaptado de Fernandes et al. [2021].	43
5	Diagrama ilustrando os diversos termos aditivos que podem ser ajustados no modelo GAMLSS . Adaptado de Stasinopoulos et al. [2017].	46
6	Diagrama ilustrando a metodologia aplicada para a construção das curvas de crescimento neste estudo.	65
7	Diagramas de caixa para as variáveis D1 , D2 , PP , EP e PF	69
8	Diagrama apresentando a interação entre os dois componentes de uma aplicação <i>shiny</i> , o UI e o servidor.	71
9	Representação gráfica da estrutura da <i>interface</i> da aplicação. Cada caixa colorida representa uma página do menu da barra lateral da aplicação. As caixas brancas representam funcionalidades dentro de cada página, estando interligadas por setas.	73
10	Gráficos de barras para as variáveis, local, paridade da mãe, género do feto e forma da placenta.	77

11	Histogramas para as variáveis idade materna e IG	78
12	Gráficos de dispersão representando a relação entre a IG e vários parâmetros (D1 , D2 , PP , EP e PF). Cada ponto nos gráficos representa uma observação individual de uma amostra de 2064 placentas.	84
13	Gráficos Q-Q para a avaliação de normalidade de D1 , D2 , PP , EP e PF	85
14	Os valores médios e os seus correspondentes intervalos de confiança de 95% para os parâmetros, D1 , D2 , PP , EP e PF , são apresentados com base no género fetal em cada IG	87
15	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo GAMLSS do D1 com apenas a função de alisamento <i>P-splines</i>	89
16	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo GAMLSS do D1 com a função de alisamento <i>P-splines</i> e a distribuição BCT	91
17	Valores observados do D1 representados em função da IG para o modelo final GAMLSS com a função de alisamento <i>P-splines</i> e a distribuição BCT , com sete curvas de percentis ajustadas (3°, 10°, 25°, 50°, 75°, 90° e 97°).	92
18	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo GAMLSS do D2 com apenas a função de alisamento <i>P-splines</i>	94
19	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo GAMLSS do D2 com a função de alisamento <i>P-splines</i> e a distribuição BCT	95
20	Valores observados do D2 representados em função da IG para o modelo final GAMLSS com a função de alisamento <i>P-splines</i> e a distribuição BCT , com sete curvas de percentis ajustadas (3°, 10°, 25°, 50°, 75°, 90° e 97°).	97
21	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo GAMLSS da EP com apenas a função de alisamento <i>P-splines</i>	98
22	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo LMST da EP com a função de alisamento <i>P-splines</i>	99
23	Valores observados da EP representados em função da IG para o modelo final LMST com a função de alisamento <i>P-splines</i> , com sete curvas de percentis ajustadas (3°, 10°, 25°, 50°, 75°, 90° e 97°).	101

24	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo GAMLSS do PP com apenas a função de alisamento <i>P-splines</i>	102
25	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo LMST do PP com a função de alisamento <i>P-splines</i>	104
26	Valores observados do PP representados em função da IG para o modelo final LMST com a função de alisamento <i>P-splines</i> , com sete curvas de percentis ajustadas (3°, 10°, 25°, 50°, 75°, 90° e 97°).	105
27	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo GAMLSS do PF com apenas a função de alisamento <i>P-splines</i>	107
28	Critérios de avaliação de qualidade de ajuste (<i>worm plot</i> , estatísticas <i>Z</i> e gráficos de resíduos) para o modelo LMST do PF com a função de alisamento <i>P-splines</i>	108
29	Valores observados do PF representados em função da IG para o modelo final LMST com a função de alisamento <i>P-splines</i> , com sete curvas de percentis ajustadas (3°, 10°, 25°, 50°, 75°, 90° e 97°).	109
30	Modal de Boas-Vindas — Introdução à aplicação.	110
31	Visão geral da Página Inicial da aplicação.	111
32	Página Gráficos — Introdução das informações do processo do paciente na aba “Informações”.	111
33	Página Gráficos — Introdução dos dados para o parâmetro D1 e visualização do gráfico com a curva de crescimento.	112
34	Página Informações — Visualização da tabela de informações para o parâmetro D1	113
35	Página Rácios — Visualização da tabela com os valores dos rácios calculados.	114
36	Página Ajuda — Guia de utilização da aplicação.	115

Lista de Tabelas

1	Métodos para a construção de curvas de crescimento. Adaptado de: Borghi et al. [2006].	17
2	Descrição da forma do <i>worm plot</i> relativamente aos resíduos e à distribuição ajustada quando um modelo está mal ajustado aos dados. Adaptado de Stasinopoulos et al. [2017].	61
3	Estatísticas descritivas do D1 em cada IG . A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.	79
4	Estatísticas descritivas do D2 em cada IG . A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.	80
5	Estatísticas descritivas da EP em cada IG . A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.	81
6	Estatísticas descritivas do PP em cada IG . A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.	82
7	Estatísticas descritivas do PF em cada IG . A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.	83
8	Resultados do teste de Shapiro-Wilk para a avaliação da normalidade de D1 , D2 , PP , EP e PF	86
9	Métricas de desempenho do modelo para o D1 em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exhibe o R² , o MSE , o AIC e o BIC	92

10	Resumo do desempenho do modelo final do D1 , método GAMLSS com a função de alisamento <i>P-splines</i> e a distribuição BCT . São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.	93
11	Métricas de desempenho do modelo para o D2 em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exibe o R² , o MSE , o AIC e o BIC	96
12	Resumo do desempenho do modelo final do D2 , método GAMLSS com a função de alisamento <i>P-splines</i> e a distribuição BCT . São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.	97
13	Métricas de desempenho do modelo para a EP em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exibe o R² , o MSE , o AIC e o BIC	100
14	Resumo do desempenho do modelo final da EP , método LMST com a função de alisamento <i>P-splines</i> . São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.	101
15	Métricas de desempenho do modelo para o PP em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exibe o R² , o MSE , o AIC e o BIC	105
16	Resumo do desempenho do modelo final do PP , método LMST com a função de alisamento <i>P-splines</i> . São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.	106
17	Métricas de desempenho do modelo para o PF em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exibe o R² , o MSE , o AIC e o BIC	109
18	Resumo do desempenho do modelo final do PF , método LMST com a função de alisamento <i>P-splines</i> . São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.	109

Acrónimos

R² Coeficiente de Determinação.

AIC *Akaike's Information Criterion*/Critério de Informação de Akaike.

AVAS *Additivity and Variance Stabilization*/Aditividade e Estabilização da Variância.

BCCG *Box-Cox Cole-Green*.

BCPE *Box-Cox Power Exponential*.

BCT *Box-Cox t*.

BIC *Bayesian Information Criterion*/Critério de Informação Bayesiano.

CCD Centro de Controlo e Prevenção de Doenças.

CG Algoritmo Cole e Green.

CGC Centro de Genética Clínica.

CV Coeficiente de Variação.

D1 Maior Diâmetro da Placenta.

D2 Menor Diâmetro da Placenta.

DP Desvio Padrão.

EMV Estimação de Máxima Verosimilhança.

EP Espessura Placentar.

FPET *Fractional Polynomials and Exponential Transformation*/Polinómios Fracionários e Transformação Exponencial.

GAIC *Generalized Akaike Information Criterion/Critério de Informação de Akaike Generalizado.*

GAM Generalized Additive Models/Modelos Generalizados Aditivos.

GAMLSS *Generalized Additive Models for Location, Scale and Shape/Modelos Aditivos Generalizados para a Localização, Escala e Forma.*

GLM Generalised Linear Model/Modelo Linear Generalizado.

HRY Método Healy-Rasbash-Yang.

IDE *(Integrated Development Environment/ Ambiente de Desenvolvimento Integrado.*

IG Idade-Gestacional.

IQR *Interquartile Range/Amplitude Interquartil.*

JSU *Johnson's SU.*

LMS Lamda-Mu-Sigma.

LMSP *Lambda-Mu-Sigma Box-Cox-power-exponential.*

LMST *Lambda-Mu-Sigma Box-Cox t.*

MADAM *Mean and Dispersion Additive Models/Modelos Aditivos de Média e Dispersão.*

MSE *Mean Square Error/Erro Quadrático Médio.*

MST Média da Soma Total dos Quadrados.

MV Máxima Verosimilhança.

PF Peso Fetal.

PN Peso ao Nascer.

PP Peso Placentar.

R-PF Rácio Peso Fetal/Peso Placentar.

R-PP Rácio Peso Placentar/Peso Fetal.

R-PPN Rácio Peso Placentar ao Nascimento.

RS Algoritmo Rigby e Stasinopoulos.

UI *User interface*/Interface do Utilizador.

VCG Validação Cruzada Generalizada.

Capítulo 1

Introdução

Este primeiro capítulo marca o início da dissertação, cujo foco recai sobre o desenvolvimento de modelos de curvas de crescimento utilizando métodos de regressão não linear, com ênfase na sua aplicação a parâmetros placentares. Este estudo é realizado no contexto da dissertação de Mestrado em Bioinformática da Universidade do Minho e tem como objetivo principal avaliar a utilidade destes modelos na detecção precoce de complicações fetais.

Inicialmente, é estabelecido o enquadramento conceitual abrangente e motivação relativamente ao tema da dissertação. Posteriormente, são delineados os objetivos a serem alcançados ao longo desta investigação e, por fim, é apresentada uma visão detalhada da estrutura deste documento. Este capítulo inicial desempenha um papel fundamental ao estabelecer as bases que sustentarão a exploração e análise ao longo desta dissertação.

1.1 Enquadramento e motivação

O desenvolvimento fetal é altamente dependente de vários fatores genéticos, nutricionais e ambientais. Durante este período, especialmente em estágios iniciais, ocorrem interações complexas entre genes e fatores ambientais durante a diferenciação das células e evolução dos tecidos [Balasundaram and Avulakunta, 2022]. Estas interações desempenham um papel importante não apenas no comportamento humano, mas também na suscetibilidade a doenças. Assim, o reconhecimento precoce de falhas no crescimento pode evitar o risco de distúrbios fetais e doenças em adulto.

Nas últimas décadas, medidas fetais e neonatais têm sido utilizadas para avaliar o desenvolvimento do feto. Recentemente, tem-se verificado um interesse crescente na avaliação dos parâmetros biométricos da placenta e a sua relação com o resultado obstétrico [Siskovicova et al., 2023, Njeze et al., 2020, Mohamed et al., 2022, Nogueira et al., 2019]. A placenta é um órgão altamente complexo e importante para manter e continuar uma gravidez saudável, pois transfere e troca oxigénio e nutrientes necessários para o feto [Burton and Jauniaux, 2015]. À medida que o feto cresce, ocorrem muitas mudanças na

forma e função da placenta que refletem as necessidades do feto em diferentes estágios de crescimento [Asgharnia et al., 2008]. Qualquer distúrbio na normal sequência do desenvolvimento e crescimento pode levar a discrepâncias dos aspetos placentares e fetais, por isso, a sua monitorização é vital [Cardoso, 2020]. Existem evidências que sugerem que a forma e o tamanho da placenta são fatores que podem estar estatisticamente associados a complicações na gravidez, como, por exemplo, restrições no crescimento fetal, redução dos movimentos fetais e saúde a longo prazo do indivíduo [Ashwal et al., 2022, Vişan et al., 2020, Almog et al., 2011, Barker et al., 2011, Eriksson et al., 2011, Cardoso, 2020].

Quando os parâmetros biométricos da placenta são adequadamente estudados podem ser muito úteis na resolução de vários casos de mortalidade perinatal e doenças neonatais. Assim, atualmente, têm existido inúmeras solicitações de estudos patológicos placentares e a sua compreensão tem sido muito importante para perceber e expandir a utilidade dos exames à placenta [Akhavan et al., 2022, Loverro et al., 2022]. Sabe-se que o diâmetro placentar (**D1** ou **D2**), espessura placentar (**EP**), peso placentar (**PP**) e peso fetal (**PF**) são parâmetros importantes na avaliação da função da placenta e do crescimento fetal [Cardoso, 2020, Hayward et al., 2016]. Portanto, avanços na compreensão destes parâmetros em função da idade-gestacional (**IG**) poderá ter um grande impacto na reconhecimento precoce de distúrbios fetais.

No entanto, o relativo desinteresse pelo estudo da placenta quando comparado ao estudo fetal é responsável pela existência de uma grande lacuna na compreensão do significado biológico das lesões placentárias relacionadas ao contexto perinatal e neonatal [Nogueira et al., 2019]. Portugal é um exemplo onde a patologia placentária continua a ser uma subespecialidade da patologia cirúrgica subvalorizada, ineficazmente manuseada e não partilhada clinicamente [Cardoso, 2020].

Nesse sentido, surgiu a necessidade de investigar e produzir ferramentas que permitam a elaboração de um diagnóstico dos parâmetros biométricos da placenta. Para tal, são utilizados modelos de curvas de crescimento que conseguem descrever mudanças, como o peso e diâmetro da placenta, ao longo do tempo (**IG**), permitindo que a informação seja combinada em alguns (geralmente 3 ou 4) parâmetros com interpretação biológica [Vitezica et al., 2010].

As curvas de percentis do peso ao nascer (**PN**) são amplamente conhecidas e utilizadas na prática clínica para monitorizar o crescimento fetal e classificar o risco de algumas condições fetais [Deter et al., 1986]. Porém, estudos referentes às curvas de percentis direcionadas para dados placentares como o **PP**, **D1**, **D2** e **EP** são muito raros. Os estudos acerca das curvas de percentis fetais e placentares maioritariamente referem-se a uma **IG** superior a 24 semanas e algumas dessas informações podem já estar desatualizadas [Nogueira et al., 2019]. Desse modo, o estudo e atualização da informação relacionada com estas curvas de crescimento é importante para evitar riscos durante a gravidez. Embora seja ne-

cessário existir evidências adicionais, as curvas de percentis são pertinentes para o acompanhamento do crescimento fetal e a sua otimização e compreensão pode permitir uma intervenção prematura em contextos adversos fetais.

Modelos de regressão não linear são amplamente utilizados no contexto da modelação das curvas de crescimento devido à sua robustez em comparação com modelos de regressão linear. Vários investigadores conduziram comparações de modelos não lineares para ajustar efetivamente uma ampla gama de curvas de crescimento [Borghi et al., 2006, Numan et al., 2021]. A vantagem destas técnicas está na capacidade de combinar medidas longitudinais em poucos parâmetros facilmente interpretáveis, simplificando a análise [Vitezica et al., 2010].

A regressão quantílica Koenker [2005] é uma abordagem popular para ajustar curvas de crescimento e tem sido aplicada em estudos sobre parâmetros placentares, como o trabalho realizado por Alves et al. [2022]. Este método não paramétrico permite ajustar diferentes percentis da distribuição, em vez de se concentrar apenas na média, como na regressão padrão [Waldmann, 2018]. No entanto, uma limitação dessa metodologia é que o processo de estimação pode resultar em cruzamento das curvas de percentis, uma vez que as estimativas para cada percentil são determinadas de forma independente [Borghi et al., 2006]. Várias tentativas foram feitas na literatura para abordar este problema, mas essas abordagens frequentemente requerem esforços adicionais e podem não ser adequadas com base no problema específico que está a ser abordado. Além disso, as curvas de percentis estimadas nas caudas da distribuição de dados podem manifestar uma instabilidade substancial, tornando a sua interpretação complexa e compromete a sua confiabilidade [Nakamura et al., 2022].

Neste estudo, foram utilizados modelos de regressão semiparamétrica dentro da estrutura dos modelos aditivos generalizados para localização, escala e forma (**GAMLSS**), conforme proposto por Rigby and Stasinopoulos [2005], para ajustar com precisão as curvas de crescimento de referência para vários parâmetros como **D1**, **D2**, **EP**, **PP**, e **PF**.

Os dados para construir e testar os modelos gerados foram recolhidos no Laboratório de Patologia Embrionária Fetal do Centro de Genética Clínica (**CGC**) no Porto. Estes envolvem a população de parturientes portuguesas de diferentes regiões.

Esta investigação oferece contribuições significativas para o estudo da análise de parâmetros placentares e modelação de curvas de crescimento. Destaca a importância dos parâmetros biométricos da placenta nos resultados obstétricos e na saúde fetal. Além disso, introduz a aplicação da estrutura **GAMLSS** e métodos **LMS** para a estimação precisa e flexível de curvas de crescimento de percentis para parâmetros placentares. Os resultados esclarecem a estabilidade dessas curvas, particularmente

nas extremidades da distribuição de dados, proporcionando informações valiosas para a avaliação clínica e aprimoramento dos resultados na gravidez.

Portanto, a motivação central deste trabalho, foi a preocupação com a melhoria dos diagnósticos de reconhecimento precoce de distúrbios fetais. Nessa sequência, na presente dissertação, serão implementadas curvas de crescimento que poderão ser úteis de diversas formas e promover avanços no acompanhamento placentar e fetal, esclarecendo a avaliação da função da placenta e cuidados perinatais. Por conseguinte, este estudo poderá fornecer de alguma forma alívio psicossocial às famílias e contribuir para um aumento do sucesso perinatal e taxa de natalidade. Assim, face às ferramentas disponíveis e aos dados recolhidos pelo **CGC**, de parturientes portuguesas de diferentes regiões, pretende-se desenvolver uma aplicação *web*, acessível aos profissionais de saúde, que permita compreender e prevenir possíveis problemas fetais.

1.2 Objetivos

O principal objetivo desta dissertação passa por estudar a aplicabilidade de modelos de regressão não linear em curvas de crescimento de parâmetros biométricos placentares. Tendo como finalidade, o desenvolvimento de uma aplicação *web* que poderá auxiliar os profissionais de saúde no diagnóstico precoce de complicações fetais.

Com base neste pressuposto, também se tem o objetivo secundário de responder à seguinte questão: “Podemos determinar a evolução do feto e prevenir complicações futuras através da evolução dos parâmetros placentares?”. Mais especificamente, este trabalho abordará os seguintes objetivos científicos e tecnológicos:

- Realizar uma revisão abrangente do estado da arte na área de curvas de crescimento e a sua aplicação em parâmetros placentares;
- Adquirir conhecimento a nível da modelação utilizando técnicas de regressão não linear paramétrica, não paramétrica e semiparamétrica;
- Realizar o tratamento e preparação dos dados;
- Explorar os pacotes/funções existentes a nível do *software* R;
- Aplicar as técnicas de regressão não linear aos dados para construir as curvas de crescimento dos parâmetros placentares;

- Desenvolver uma aplicação *web* interativa para disponibilizar os resultados aos profissionais de saúde;
- Analisar os resultados obtidos, discutindo as suas implicações clínicas e possíveis melhorias.

1.3 Estrutura da dissertação

Esta dissertação desenvolve-se ao longo de nove capítulos.

No **Capítulo 1** é apresentada uma introdução ao tema em estudo, destacando as principais razões que motivaram a sua investigação. Além disso, são apresentados os objetivos a alcançar e uma visão geral da estrutura desta dissertação.

No **Capítulo 2** é apresentado o estado da arte, onde é revista a bibliografia dos principais temas abordados ao longo desta dissertação. Começando por uma contextualização ao conceito de placenta, bem como a sua evolução histórica e estudos relacionados. Posteriormente, apresentam-se as curvas de crescimento, incluindo a sua história, as suas designações nos mais diferenciados estudos e a sua aplicação na biologia e medicina. Por último, é abordado o *software* R, desde a sua contextualização, designação e aplicação.

No **Capítulo 3** é introduzida a teoria e conceitos relacionados à regressão. Começando com uma definição abrangente de regressão e prosseguindo para a exploração da regressão linear e não linear. Além disso, é discutida as distinções entre regressão paramétrica, não paramétrica e semiparamétrica. Este capítulo cria os fundamentos necessários para a compreensão das metodologias de modelação que serão abordadas posteriormente.

Este capítulo estabelece uma base para compreender as metodologias de modelação posteriores.

No **Capítulo 4** são abordados detalhadamente os métodos utilizados no desenvolvimento das curvas de crescimento. Começando com considerações metodológicas, discutindo aspetos distributivos e de alisamento. Em seguida, é dada uma explicação de como os métodos foram selecionados e é realizada uma revisão metodológica desses métodos. Além disso, são explorados os critérios de qualidade de ajuste, que serão utilizados na seleção de modelos. Este capítulo é fundamental para entender as metodologias de construção das curvas de crescimento.

No **Capítulo 5** é descrita a preparação dos dados que serviram de base para a análise e modelação das curvas de crescimento. Isso inclui a apresentação do conjunto de dados utilizado e o processo de pré-processamento, abordando a seleção de variáveis, tratamento de valores ausentes e discrepantes.

No **Capítulo 6** é detalhado o desenvolvimento da aplicação *web* “PlacentalGrowth”. É explorada

a estratégia de desenvolvimento, as ferramentas utilizadas e a arquitetura da aplicação. São também descritas as funções e eventos reativos que tornam a aplicação interativa e dinâmica, fornecendo aos utilizadores uma plataforma para análise das curvas de crescimento.

No **Capítulo 7** são apresentados os resultados obtidos ao longo desta investigação. Inclui uma análise exploratória dos dados, testes de significância entre género para os vários parâmetros placentares, como diâmetros e pesos. As curvas de crescimento desenvolvidas e, a aplicação *web*.

No **Capítulo 8** é realizada a discussão dos resultados obtidos, destacando as principais descobertas e discutindo as implicações dos resultados. Este capítulo proporciona uma visão abrangente do estudo e o seu impacto.

No **Capítulo 9** são resumidas as conclusões deste estudo e delineados possíveis caminhos para investigações futuras. Este capítulo encerra a dissertação, destacando a contribuição do estudo para o campo da obstetria e indicando como ele pode continuar a evoluir.

Capítulo 2

Estado da arte

Neste capítulo são analisados, com fundamento na literatura, um conjunto de conceitos e definições relevantes para o projeto em causa, nomeadamente o conceito da placenta humana e os modelos de curvas de crescimento. São abordados, os métodos de regressão não linear que serão utilizados, assim como uma revisão do *software* R e do pacote *shiny*. A metodologia aplicada para o apuramento dos conceitos debatidos neste capítulo passa pela pesquisa literária, principalmente em artigos científicos e livros científicos.

2.1 Placenta humana

2.1.1 Introdução histórica

A história da biologia placentária ao longo dos anos pode ser vista de várias perspetivas. Estas incluem: embriologia, variação por espécie, forma morfológica, organização celular microscópica, circulações materna e fetal, a *interface* entre as superfícies de troca materna e fetal, gás respiratório e troca de nutrientes, funções metabólicas, endócrinas e imunológicas, entre outras [Longo and Reynolds, 2010].

Desde os filósofos gregos pré-socráticos, a placenta exerce um papel preponderante e, em simultâneo, é considerada misteriosa e até mística [Longo and Reynolds, 2010]. Em muitas culturas, mesmo presentemente, a placenta é considerada um *alter ego*, um símbolo para a preservação da saúde e da boa sorte e um talismã em caso de perigo [Burns, 2014, Longo and Reynolds, 2010].

Foram necessários vários séculos até que existissem alguns conceitos gerais acerca da placenta, como a sua natureza anatómica. Atualmente, detalhes sobre as suas diversas funções estão ainda a ser estudados e existe ainda muita investigação a ser feita para descobrir o seu potencial para diagnóstico de distúrbios obstétricos.

O exame patológico da placenta pode revelar padrões macroscópicos e microscópicos que fornecem informações valiosas para obstetras, neonatologistas e pediatras [Turowski et al., 2018]. Inicialmente, a patologia placentária apenas tinha em conta anormalidades macroscópicas, no entanto, estas não conse-

guem explicar todos os possíveis resultados adversos. A descrição macroscópica da placenta inclui uma ampla gama de parâmetros clinicamente relevantes como, por exemplo, análise das membranas, o comprimento e o diâmetro médio do cordão umbilical e o **PP** em si [Turowski et al., 2018]. Por outro lado, a descrição microscópica detalha a aparência dos tecidos, destacando as lesões e também caracterizando, pelo menos brevemente, os achados normais [Turowski et al., 2018]. Alguns autores demonstraram que exames microscópicos são um passo chave para permitir diagnósticos definitivos em vários casos [Middeldorp, 2007, Baergen, 2011, Choi et al., 2014].

A classificação das lesões placentárias tem também sofrido uma evolução ao longo dos anos. Progrediu de um exercício puramente descritivo para um estágio em que os principais processos fisiopatológicos, são descritos num sistema de classificação abrangente que inclui processos maternos e fetais, processos infecciosos e processos inflamatórios idiopáticos/imunes [Cardoso, 2020]. Os progressos nesta área têm sido alcançados através da gradual aceitação dos processos fisiopatológicos padronizados e de base biológica.

O interesse e solicitação de exames patológicos placentares têm igualmente aumentado em parte pelas exigências dos obstetras relacionadas aos avanços no diagnóstico pré-natal e às novas situações médicas desafiadoras, bem como também por ser uma etapa importante na avaliação de malformações embrionárias e fetais ou na identificação de processos patológicos previamente insuspeitos na mãe, ou no bebé que podem exigir atenção imediata [Cardoso, 2020].

Da mesma forma, tem-se verificado um interesse crescente na avaliação de parâmetros como o **D1**, **D2**, **EP**, **PP**, e **PF** apesar serem atualmente pobremente conceituados [Nogueira et al., 2019]. Estes podem estar associados com desfechos adversos e doenças fetais ou perinatais futuramente na vida.

Portugal é um exemplo onde a patologia placentária continua a ser uma subespecialidade da patologia cirúrgica subvalorizada, ineficazmente manuseada e não partilhada clinicamente [Cardoso, 2020]. No entanto, existem já vários estudos da morfologia placentária e de parâmetros biométricos adaptados à população portuguesa [Nogueira et al., 2019, 2015, 2018, Cardoso, 2020, Alves et al., 2022].

O estudo patológico da placenta tem, então, contado com um grande desenvolvimento nos últimos anos e demonstrado a sua importância no diagnóstico e prevenção de possíveis resultados adversos na gravidez.

2.1.2 Definição e características gerais

A placenta é o maior órgão fetal e o primeiro a se desenvolver. Desempenha um papel central na saúde do feto e da mãe, tendo um impacto vitalício no bem-estar futuro de ambos [Turco and Moffett, 2019]. De

facto, um desenvolvimento placentar incomum, é a principal indicação de doenças na gravidez, como pré-eclâmpsia, restrição do crescimento fetal, aborto espontâneo tardio, trabalho de parto prematuro, rutura prematura de membranas, descolamento prematuro da placenta e nado-morto [Brosens et al., 2011].

Deste modo, os exames patológicos à placenta demonstram a sua importância no diagnóstico precoce de situações adversas tanto para o feto como também para a mãe, permitindo assim, que os pais e os profissionais de saúde depreendam as causas de vários problemas que ocorrem durante ou após a gravidez e possam ter uma visão para gestações futuras. Para que estes exames possam ser reproduzíveis é importante que adotem critérios de amostragem uniformes e terminologias macroscópicas e microscópicas internacionalmente aceites, por isso, têm sido criados protocolos específicos, “Amsterdam Placental Workshop Group Consensus”, para a amostragem de placentas [Khong et al., 2016]. A standardização da amostragem macroscópica e histológica da placenta é importante porque as lesões focais não devem ser perdidas e porque a comparação entre estudos de frequências e significância das lesões depende dos métodos de amostragem [Khong et al., 2016]. Os relatórios dos exames patológicos à placenta devem, então, passar pelos seguintes passos, descritos na Figura 1: historial clínico, exame macroscópico, amostragem, exame microscópico e diagnóstico.

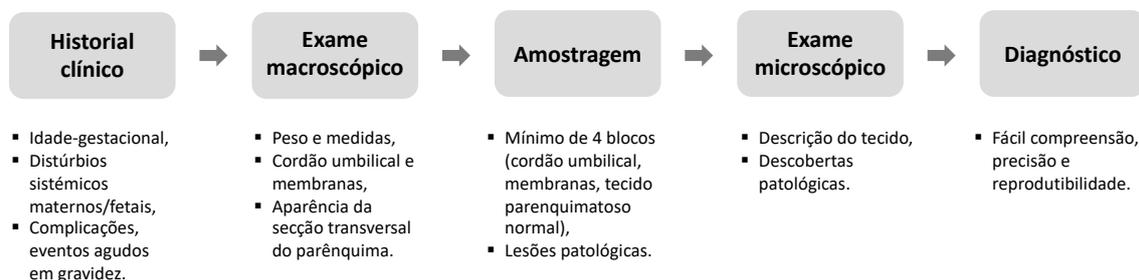


Figura 1: Passos para produzir um relatório de um exame patológico da placenta com relevância clínica. Adaptado de Turowski et al. [2018].

Até à data, existem poucos estudos publicados direcionados para o crescimento e desenvolvimento da placenta humana. Porém, as medidas da placenta são ferramentas cruciais no exame patológico e essenciais para a compreensão do resultado fetal, por exemplo, pequeno ou grande para a IG, restrição do crescimento fetal, doenças fetais e morte cerebral [Cardoso, 2020]. Estas medidas são, igualmente, essenciais na compreensão de perturbações maternas. Posto isto, o desenvolvimento de novas tecnologias capazes de monitorizar o crescimento placentar, combinadas com um exame patológico competente, poderão trazer conhecimentos adicionais e muito importantes, nomeadamente na antecipação do diagnóstico de possíveis doenças fetais ou maternas.

A descrição da placenta, na patologia de rotina, engloba o **PP** e a biometria. O **PP** é obtido retirando as membranas extraplacentárias e o cordão umbilical [Khong et al., 2016]. Já as dimensões do disco placentar devem incluir a medição da placenta em três dimensões, a dimensão linear máxima (comprimento – maior diâmetro da placenta) (**D1**), a maior dimensão do eixo perpendicular a essa medida linear (largura – menor diâmetro da placenta) (**D2**) e a espessura mínima e máxima da placenta [Khong et al., 2016]. As dimensões (**D1** e **D2**) podem ser obtidas através do ultrassom, como se pode ver na Figura 2. Para permitir a realização de estudos, como curvas de percentis, é necessária a obtenção destas medidas da placenta. Existem evidências que sugerem que a forma e o tamanho da placenta são fatores que podem estar estatisticamente associados a complicações na gravidez e saúde a longo prazo do indivíduo [Shehata et al., 2011, Eriksson et al., 2011].

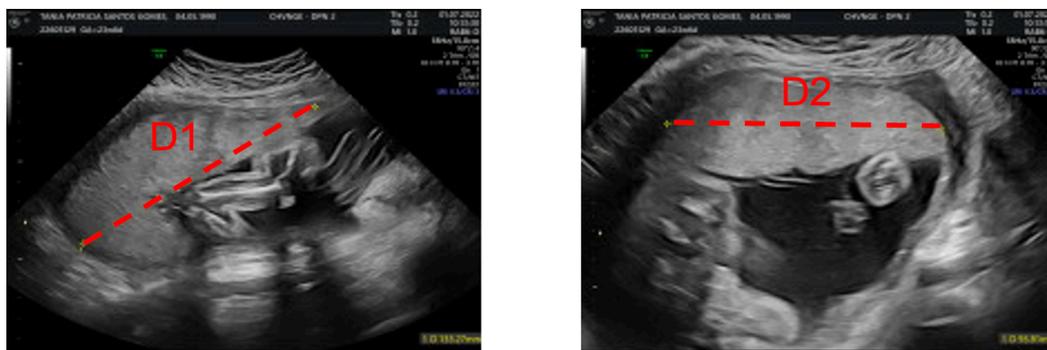


Figura 2: Imagens reais de ultrassom captadas pelos médicos: Conceição Brito, Cristina Godinho, Mónica e Márcia Marinho (DPN, CHVNG/E-Portugal). Apresentam os diâmetros da placenta (linha a vermelho), o maior diâmetro da placenta (D1) e o menor diâmetro da placenta (D2).

2.1.3 Implicações dos parâmetros placentares e fetais

O desenvolvimento fetal no útero é grandemente influenciado pela placenta. Em média, a placenta ao termo pesa cerca de 508 gramas [Panti et al., 2012]. Vale ressaltar que a proporção entre o peso da placenta e o peso do recém-nascido geralmente é de 1:6, embora medições precisas possam variar devido a diferenças na preparação da placenta e nos métodos de medição [Panti et al., 2012].

Estudos ao longo de um século demonstraram que o peso da placenta pode afetar significativamente os resultados da gravidez. Três referências reconhecidas para avaliar a saúde da placenta são o peso absoluto da placenta e os rácios **PP/PF (R-PP)** e **PF/PP (R-PF)**, pois existe uma correlação positiva entre o peso fetal e o peso da placenta [Panti et al., 2012]. Esta correlação positiva é essencial porque o crescimento fetal adequado depende da capacidade da placenta de transferir nutrientes. O rácio **R-PF**

serve como um indicador da eficiência da placenta, indicando a quantidade de gramas de feto produzida por grama de placenta [Hayward et al., 2016].

Associações clínicas com estas métricas foram documentadas, com placentas pequenas estando possivelmente relacionadas a trissomias, enquanto placentas grandes podem estar associadas a diabetes da mãe [Panti et al., 2012]. O tamanho desproporcional da placenta, indicado pelo rácio **R-PF**, pode sugerir desde lesões placentárias agudas até condições crónicas, como anemia materna ou desnutrição. A hipertensão materna pode levar a um rácio **R-PF** alto, que, por sua vez, pode causar sofrimento fetal ou baixos índices de Apgar [Panti et al., 2012]. Quando um feto não atinge o seu potencial de crescimento geneticamente determinado (restrição de crescimento fetal), tanto o peso da placenta quanto o rácio **R-PF** tendem a diminuir, sugerindo a incapacidade da placenta de adaptar a transferência de nutrientes ao seu tamanho menor [Hayward et al., 2016]. Vale ressaltar que em circunstâncias normais, o rácio **R-PF** geralmente varia entre 6,5 e 7,1 no termo (37 – 42 semanas de gestação) [Tiruneh, 2018]. No entanto, em casos de pré-eclâmpsia, essa relação tem sido encontrada significativamente menor, destacando a importância de monitorizar a saúde da placenta em gravidezes de alto risco [Tiruneh, 2018].

O rácio **R-PP** também é um fator crítico na previsão de resultados de saúde na idade adulta. Estudos sugerem que uma placenta maior em combinação com um bebé pequeno pode levar a adaptações circulatórias no feto e alterações na estrutura arterial, potencialmente resultando em hipertensão na idade adulta [Tiruneh, 2018]. Isso destaca a importância de compreender a relação peso placentar/peso fetal para prevenir problemas de saúde que podem originar-se na vida fetal.

O diâmetro e a espessura da placenta também desempenham um papel significativo no desenvolvimento fetal e no peso ao nascer. Estudos têm demonstrado uma forte correlação entre as dimensões da placenta e a probabilidade de um bebé nascer com um peso normal [Habib, 2002]. Por exemplo, um aumento no diâmetro e na espessura da placenta corresponde a uma probabilidade maior de os bebés nascerem com um peso normal. Normalmente, bebés com baixo peso ao nascer, têm um diâmetro da placenta inferior a 18 cm e uma espessura da placenta inferior a 2 cm às 36 semanas de gestação [Habib, 2002]. Além disso, alguns estudos têm explorado os efeitos dos diabetes gestacionais e da hipertensão materna na morfologia geral da placenta. Os resultados sugerem que, nos casos de diabetes, as placentas tendem a apresentar um aumento de peso, espessura central e diâmetro [Ashfaq et al., 2005]. Incorporar a medição do diâmetro e da espessura da placenta nas avaliações por ultrassom de rotina durante a gravidez, pode ser uma adição valiosa para prever com maior precisão os bebés com baixo peso ao nascer.

Apesar do papel crítico da placenta no bem-estar fetal e materno, a sua influência e relação com o peso

ao nascer não são extensamente exploradas. Frequentemente, a placenta é descartada imediatamente após o parto, sem uma análise minuciosa da sua morfologia e do seu impacto na saúde geral da mãe e do bebê. Investigações adicionais nesta área podem fornecer informações valiosas para melhorar os resultados da gravidez e a saúde materna.

2.2 Curvas de crescimento

2.2.1 Introdução histórica

O estudo do crescimento, nomeadamente o humano, tem cativado a atenção de vários investigadores ao longo dos séculos, mas somente na segunda metade do século XX foi formulada uma abordagem coerente para a recolha e análise apropriada de dados relevantes [Freeman, 1995].

A história do estudo do crescimento remonta ao Conde Philibert de Montbeillard (1720 – 1785), pois este foi o pioneiro a registar as medidas do corpo de uma criança, no caso, o seu filho, longitudinalmente desde o nascimento até ao início da idade adulta [Freeman, 1995, Cole, 2012]. O seu estudo, foi a primeira referência da ilustração do padrão de crescimento. Mais tarde, George Buffon (1707 – 1788), publicou o gráfico gerado pelos dados do Conde Montbeillard, produzindo assim a primeira curva de crescimento [Cole, 2012]. Por volta da mesma época, Christian Friedrich Jampert publicou a primeira tabela de medidas do crescimento humano em crianças durante a primavera de 1754 [Freeman, 1995]. Estes estudos foram conduzidos pela curiosidade científica motivada pela vontade de estudar e descrever os padrões do crescimento humano e, foram importantes para dar início a todo o desenvolvimento desta área. Em retrospectiva, eles tinham muito em comum com os estudos de crescimento atuais que visam produzir percentis de referência para monitorizar o crescimento da população.

Em 1997, Wright e Royston concederam uma revisão abrangente dos vários métodos existentes até à data comumente usados para a construção de curvas de crescimento de crianças [Wright and Royston, 1997]. Os métodos foram agrupados por discretização (agrupamento por faixas etárias) ou utilização da idade continuamente, e segundo o uso ou não de suposições de distribuição [Borghi et al., 2006]. Já Borghi et al. [2006], de modo a desenvolverem os padrões de crescimento infantil da Organização Mundial da Saúde, realizaram também uma revisão dos vários métodos existentes para as curvas de crescimento, incluindo os abordados por Wright e Royston. Desse modo, estas duas publicações serão usadas como base para o desenvolvimento histórico das curvas de crescimento. Apesar de estes métodos revistos serem aplicados ao crescimento infantil, estes poderão ser transpostos e utilizados para outros parâmetros como os da placenta, daí a sua revisão ser importante para fornecer um discernimento acerca da modelação

de curvas de crescimento. A Tabela 1 sumariza os vários métodos que já foram utilizados/desenvolvidos na literatura.

Métodos de discretização sem suposições de distribuição

Entre os métodos que estimam os percentis separadamente agrupando por idades e sem suposições de distribuição, o trabalho de Hamill et al. [1977] pode ser citado como as primeiras curvas de crescimento a serem testadas. Estas foram derivadas dividindo os dados em grupos de idade de um ano, calculando sete percentis do 5.º ao 95.º em cada grupo e, em seguida, usando *Splines* cúbicas para alisar cada percentil ao longo da idade. Mais tarde, Gasser et al. [1984] e Guo et al. [1990] aplicaram o alisamento de núcleo em regressão a percentis empíricos.

No ano de 2000, o Centro de Controlo e Prevenção de Doenças (CCD) [Kuczmarski et al., 2002] construiu curvas de crescimento utilizando um método de duas etapas. Na primeira, diversos métodos foram testados para modelar percentis empíricos em diferentes medidas de crescimento e faixas etárias. Modelação linear de 3 parâmetros foi aplicada para comprimento-para-idade, peso-para-idade e circunferência craniana-para-idade. Para altura-para-idade, um modelo não linear foi escolhido, enquanto peso-para-comprimento e peso-para-altura utilizaram um polinómio de quinto grau [Guo et al., 1992]. Em crianças acima de 24 meses, o alisamento dos percentis empíricos envolveu inicialmente uma regressão robusta ponderada localmente, seguida por uma regressão polinomial [Cleveland, 1979]. Na segunda etapa, foi realizada uma aproximação *a posteriori* Lambda-Mu-Sigma (LMS) [Cole and Green, 1992] desses percentis alisados para calcular os *z-scores*.

Métodos de discretização com suposições de distribuição

Dentro dos métodos que estimam os percentis separadamente usando o agrupamento por faixas etárias com suposições de distribuição, existiram diversos trabalhos que contribuíram para o seu desenvolvimento. Em 1965, Tanner et al. [1966] assumiram que a altura era normalmente distribuída ao longo da infância para construir as curvas. Os percentis foram estimados em conjunto e as variâncias foram ajustadas utilizando o método de Healy et al. [1962] para compensar o seu aumento devido ao agrupamento dos dados e, para alisar as curvas de percentis, foi utilizado o método *Eye fitting*. Já Niklasson et al. [1991], usaram o ajuste polinomial para alisamento das curvas da média e desvio padrão (DP), estimadas sob normalidade ou suposição de normalidade.

A transformação de potência *Box-Cox* [Box and Cox, 1964] foi a base para alguns métodos propostos para a construção de referências de crescimento, dada a sua capacidade de “normalizar” os dados num

grau razoável em termos de assimetria. Cole [1988] propôs a primeira versão do método **LMS** usando agrupamento de idade. No seu método, os λ da *Box-Cox* foram estimados pela máxima verosimilhança (**MV**) e a variação específica da idade da distribuição foi expressa pelo coeficiente de variação, permitindo assim caracterizar qualquer curva de percentis e calcular facilmente os *z-scores*.

Idade tratada continuamente, sem suposições de distribuição

Koenker and Bassett [1978] estimaram os percentis separadamente considerando a idade como uma variável contínua e introduzindo o conceito de quantis de regressão, onde a curva α -quantil era a função linear. O seu método minimizou uma soma ponderada de resíduos positivos com peso α , e os resíduos negativos foram tornados positivos e ponderados por $(1 - \alpha)$. Outros métodos baseados nos quantis de regressão foram propostos [Ducharme et al., 1995, Wellek and Merz, 1995], mas estimando os percentis em conjunto.

Para os percentis estimados em conjunto, existem vários trabalhos em termos de métodos não paramétricos. O método Healy-Rasbash-Yang (**HRY**) de Healy et al. [1988] estimou cada percentil como uma curva aproximada e alisou cada percentil utilizando uma função polinomial, no entanto, isto constituiu uma fraqueza do método, pois é inflexível. Pan et al. [1990] conseguiu ultrapassar esta dificuldade utilizando polinômios *grafted*. Mais tarde, Heagerty and Pepe [1999] propuseram uma versão restrita dos quantis de regressão que eliminou o problema do cruzamento de quantis (por exemplo, percentil) que surge ao modelar vários quantis simultaneamente. O cruzamento de quantis acontece quando ocorre uma falha da função quantílica condicional estimada em obedecer à restrição de monotonicidade necessária [Park et al., 2021].

Idade tratada continuamente, com suposições de distribuição

Para os percentis estimados em conjunto e para os quais a normalidade é assumida e a idade é tratada como uma variável contínua, o trabalho de Aitkin [1987] pode ser destacado. Ele estimou a média com um modelo de regressão linear e a variância residual sobre a média com um modelo log-linear, usando a estimativa de **MV**. Royston and Wright [1998] propuseram uma classe mais flexível de funções composta por polinômios fracionários de baixa ordem para cada um dos parâmetros de uma distribuição módulo-exponencial-normal (MEN).

Já Rigby and Stasinopoulos [2000] propuseram o uso da subclasse de modelos aditivos de média e dispersão (**MADAM**) para modelar a média e a variância (ou log-variância), incluindo funções paramétricas ou não paramétricas alisadas da idade para a construção de percentis relacionados à idade. Tango

[1998] propôs uma regressão não paramétrica baseada no procedimento de aditividade e estabilização da variância (**AVAS**), mas, em última análise, usando a suposição normal para construir os percentis.

Quando o uso da distribuição Normal deixou de ser tão usual, o uso de famílias de distribuições mais flexíveis para a construção de curvas de crescimento começou a ser investigado. [Cole and Green \[1992\]](#) estenderam o método **LMS** de [Cole \[1988\]](#) que se ajusta à assimetria, para permitir que os dados possam ser analisados continuamente. [Wade and Ades \[1994\]](#) desenvolveram um método semelhante para ajustar percentis com tendências de idade na transformação *Box-Cox*, e a média e as funções de **DP** foram especificadas por funções paramétricas exponenciais. [Sorribas et al. \[2000\]](#) propuseram um método paramétrico baseado na distribuição-S para construir curvas de percentis alisadas ao longo das idades. [Thompson and Theron \[1990\]](#) produziram percentis com base na família Johnson de distribuições de 4 parâmetros que especificam localização, escala e dois parâmetros de forma. Eles modelaram os parâmetros de localização e escala utilizando polinômios na idade e os parâmetros de forma como constantes e ajustaram o modelo por **MV**. [Royston and Wright \[1998\]](#) também propuseram um método completamente paramétrico, o método de polinômios fracionários e transformação exponencial (**FPET**).

[Rigby and Stasinopoulos \[2005\]](#) propuseram uma classe de modelos estatísticos univariados — modelos aditivos generalizados para a localização, escala e forma (**GAMLSS**) — que podem ser aplicados a dados que apresentam distribuições assimétricas e/ou com curtose contínua, ou discreta. Eles propuseram também o método **LMS** com distribuição *Box-Cox t* (**LMST**) de estimativa de percentis para uma variável de resposta exibindo assimetria e curtose maior do que 3, baseado na distribuição *Box-Cox t* (**BCT**). Para ultrapassar a limitação da **BCT** causada pelo fato de apenas lidar com dados com leptocurtose, [Rigby and Stasinopoulos \[2004b\]](#) desenvolveram uma distribuição mais flexível, a distribuição *Box-Cox-power-exponential* (**BCPE**), que consegue modelar qualquer tipo de curtose. Esta generalização do método **LMS** ficou conhecida como método **LMSP**.

Ambos os métodos **LMST** e **LMSP** fornecem modelos altamente flexíveis para localização, escala, assimetria e curtose da variável de resposta usando o **GAMLSS** [[Borghi et al., 2006](#)]. Cada um dos quatro parâmetros podem ser modelados usando funções paramétricas (por exemplo, polinômio fracionário) ou não paramétricas (por exemplo, *Splines* cúbicas) da variável explicativa, por exemplo, a idade. O ajuste do modelo é obtido pela **MV** (penalizada). Desse modo, é fornecida uma fórmula simples para calcular percentis e *z-scores*. Covariáveis ou fatores adicionais podem também ser incluídos no modelo, se necessário, permitindo ajustes nos parâmetros, especialmente μ , para outras variáveis explicativas (por exemplo, sexo, altura dos pais, etc.) [[Borghi et al., 2006](#)].

Métodos que incorporam correlações

Existe pouco desenvolvimento no sentido de ajustar curvas de crescimento com medidas correlacionadas, pois a maioria dos métodos existentes é baseada em gráficos transversais [Borghi et al., 2006]. Modelos de efeito misto podem ser usados para modelar as curvas de crescimento, pois permitem o particionamento da variância para a contribuição da variabilidade [Borghi et al., 2006].

Laird and Ware [1982] no seu trabalho, abordam modelos de efeitos aleatórios de dois estágios para o crescimento e as estimativas empíricas de Bayes de parâmetros num modelo específico. Este método assume que os dados são normalmente distribuídos e que a inferência baseada em dados não normais deve ser feita somente após uma transformação de normalização. Goldstein [1986] propôs o uso de modelos multiníveis para resolver problemas longitudinais, incluindo os apresentados por estudos de crescimento. A classe de modelos de efeito misto ou modelos de efeito aleatório permite estruturas complexas de covariância e também para variáveis explicativas que dependem da idade. Já Wade and Ades [1998] propuseram um método de **MV** baseado em **LMS** para as curvas de crescimento relacionadas à idade que incorporam modelação explícita de correlações.

Foram comparadas cinco estruturas de correlação diferentes combinadas com cinco modelos para a mediana, e o método de “verosimilhança de perfil” [Matthews, 1988] foi usado para a construção de intervalos de confiança com um algoritmo iterativo [Borghi et al., 2006]. Apesar da presença de uma forte estrutura de correlação, os resultados mostraram que a incorporação de uma estrutura de correlação na função de verosimilhança teve pouco efeito na escolha do modelo para a curva média ou nos percentis ajustados. Além disso, a precisão com que os percentis foram estimados não foi afetada em nenhum grau clinicamente importante quando as correlações foram incorporadas.

Visão geral dos métodos utilizados nas curvas de crescimento

Tal como foi descrito na Subsecção 2.2.1, existem vários métodos para construir os gráficos de crescimento dependentes da idade. Os percentis empíricos com idade categorizada fornecem uma aproximação discreta para os percentis populacionais [Chen and Morgan, 2005]. No entanto, as curvas de percentis contínuas são mais precisas e atraentes. Os primeiros métodos começaram por ajustar curvas alisadas em quantis de amostra de grupos de idade segmentados. Porém, estes métodos não são robustos a valores discrepantes. É preciso uma dimensão grande de amostra para estimar os percentis em cada faixa etária com precisão adequada e a segmentação pode perder informações de grupos próximos [Chen and Morgan, 2005].

Tabela 1: Métodos para a construção de curvas de crescimento. Adaptado de: [Borghetti et al. \[2006\]](#).

Método	Estimação percentis	Método de ajuste da curva	Suposições de distribuição
Métodos de discretização sem suposições de distribuição			
<i>knot splines</i> fixas [Hamill et al., 1977]	Separadamente	<i>knot splines</i> fixas *	Nenhuma
<i>Eye fitting</i> (peso/idade) [Roede, 1985]	Separadamente	<i>Eye fitting</i> *	Nenhuma
Regressão de núcleo [Gasser et al., 1984], [Guo et al., 1990]	Separadamente	Estimação de núcleo *	Nenhuma
CCD (Passo 1 e 2) [Kuczmariski et al., 2002]	Separadamente	Linear de 3 parâmetros *	Nenhuma
	Juntamente	<i>Splines</i> cúbicas *	<i>Box-Cox</i> normal
Métodos de discretização com suposições de distribuição			
<i>Eye fitting</i> (peso/idade) [Tanner et al., 1966]	Juntamente	<i>Eye fitting</i> †	Normal
Normalização <i>Box-Cox</i> [Van't Hof et al., 1985]	Juntamente	<i>Eye fitting</i> †	<i>Box-Cox</i> normal
LMS , versão 1 [Cole, 1988]	Juntamente	<i>Splines</i> cúbicas ou outros †	<i>Box-Cox</i> normal
Estabilização variância (altura/idade) [Chinn, 1992]	Juntamente	Polinômios †	Normal
Ajuste polinomial [Niklasson et al., 1991]	Juntamente	Polinômios †	Normal
Idade tratada continuamente, sem suposições de distribuição			
Regressão quantílica [Koenker and Bassett, 1978]	Separadamente	Regressão quantílica *	Nenhuma
HRV [Healy et al., 1988]	Juntamente	Polinômios *	Nenhuma
HRV adaptado [Pan et al., 1990]	Juntamente	Polinômios <i>grafted</i> *	Nenhuma
Estimativa da densidade do núcleo [Rossiter, 1991]	Juntamente	Estimativa da densidade do núcleo ‡	Nenhuma
Curvas de quantis não gaussianas [Ducharme et al., 1995]	Juntamente	Densidade do núcleo do vizinho mais próximo de cdf condicional ‡	Nenhuma

Continuação da tabela na próxima página

Tabela 1: Continuação da tabela da página anterior.

Método	Estimação percentis	Método de ajuste da curva	Suposições de distribuição
Curvas de quantis não gaussianas [Wellek and Merz, 1995]	Juntamente	Função monotônica de 4 parâmetros (média) e linear (dispersão) †	Nenhuma
Quantis de regressão [Heagerty and Pepe, 1999]	Juntamente	<i>Splines</i> naturais †	Nenhuma
Idade tratada continuamente, com suposições de distribuição			
Modelos multiníveis [Goldstein, 1986]	Juntamente	Estimativa de ML de modelos lineares e não lineares †	Normal
Aitkin [Aitkin, 1987]	Juntamente	Modelos lineares †	Normal
Thompson e Theron [Thompson and Theron, 1990]	Juntamente	Polinômios †	Sistema Johnson
LMS , versão 2 [Cole and Green, 1992]	Juntamente	<i>Splines</i> cúbicos †	<i>Box-Cox</i> normal
Wade e Ades [Wade and Ades, 1994]	Juntamente	Funções exponenciais †	<i>Box-Cox</i> normal
Wade e Ades (correlações) [Wade and Ades, 1998]	Juntamente	ML exponencial; polinomial †	<i>Box-Cox</i> normal
FPET [Royston and Wright, 1998]	Juntamente	Polinômios fracionários †	(Módulo)-exponencial-normal
AVAS [Tango, 1998]	Juntamente	Regressão não paramétrica AVAS †	Normal
MADAM [Rigby and Stasinopoulos, 2000]	Juntamente	Funções paramétricas ou não paramétricas (MADAM) †	Normal
Distribuição S [Sorribas et al., 2000]	Juntamente	Polinômios †	Distribuição S
GAMLSS [Rigby and Stasinopoulos, 2005]	Juntamente	Linear paramétrico ou aditivo não paramétrico †	Vários
LMST [Rigby and Stasinopoulos, 2004a]	Juntamente	<i>Splines</i> cúbicas ou polinômios (fracionários) †	BCT
LMSP [Rigby and Stasinopoulos, 2004b]	Juntamente	<i>Splines</i> cúbicas ou polinômios (fracionários) †	BCPE

* Aplicado aos percentis.

† Aplicado a parâmetros distributivos.

‡ Percentis calculados a partir do ajuste da densidade.

Para evitar a segmentação, [Cole and Green \[1992\]](#) desenvolveram uma abordagem semiparamétrica baseada na transformação *Box-Cox* a partir do método **LMS** que resolve equações de verosimilhança penalizadas. Devido à falta de expectativa finita para algumas das derivadas da log-verosimilhança penalizada, as soluções destas equações podem ser sensíveis a um ponto inicial [[Chen and Morgan, 2005](#)]. Existe também a questão de saber se a transformação *Box-Cox* é boa o suficiente para a suposição de distribuição especificada, por exemplo, a normalidade [[Chen and Morgan, 2005](#)].

Por isso, foi criada uma forma alternativa para contornar esta situação, a regressão quantílica, um método não paramétrico. Este método não coloca nenhuma suposição de distribuição de antemão e também é relativamente fácil acomodar outras covariáveis além da idade [[Chen and Morgan, 2005](#)]. Ajusta diferentes quantis de distribuição em vez de apenas a média como na regressão não linear padrão [[Nakamura et al., 2022](#)]. No entanto, também apresenta desvantagens, nomeadamente no processo de estimação que pode resultar num cruzamento das curvas de percentis, uma vez que as estimativas para cada percentil são determinadas de forma independente. Para superar esse desafio, foram desenvolvidos modelos de regressão semiparamétrica, como os **GAMLSS**, que permitem a estimação precisa de curvas de crescimento para parâmetros placentares e podem ser aplicados a dados que apresentam distribuições assimétricas e/ou com curtose. Como tal, este método apresenta-se como uma boa sugestão de ajustamento para as curvas de crescimento.

2.2.2 Regressão não linear

Os gráficos de crescimento podem ter problemas associados, principalmente se forem construídos com base na suposição que as medidas são normalmente distribuídas. Esta suposição pode levar a erros, pois os dados utilizados nas curvas de crescimento geralmente não são normalmente distribuídos. De forma a resolver este contratempo são utilizados modelos de regressão não linear para ajustar as curvas de crescimento, pois a regressão permite quantificar a relação entre uma variável de resposta e algumas covariáveis [[Yu et al., 2003](#)].

A análise de dados de crescimento longitudinal requer metodologias estatísticas específicas que envolvem o ajuste de modelos não lineares aos parâmetros [[Anchieta et al., 2003](#)]. Ajustar curvas de crescimento a dados longitudinais envolve a descrição e o resumo do processo de crescimento com um número limitado de parâmetros. Esses parâmetros caracterizam o padrão de crescimento e têm o mesmo significado para todos os indivíduos, permitindo comparações entre indivíduos e grupos [[Anchieta et al., 2003](#)]. Por isso, é importante obter modelos matemáticos representativos do comportamento dos dados de crescimento.

2.2.3 Definição e características gerais das curvas de crescimento

O crescimento pode ser definido como o processo pelo qual os indivíduos mudam de tamanho e forma durante um determinado período. O seu estudo é realizado a partir de medições sequenciais no mesmo indivíduo em intervalos regulares e usando dados em série para estabelecer padrões comportamentais [Anchieta et al., 2003].

Um problema atual que pode ser atenuado pelo estudo do crescimento, são as taxas de mortalidade perinatal. Estas taxas são particularmente altas nas grandes populações de países de rendimento médio e baixo, onde ocorrem 98% das mortes neonatais do mundo, razão pela qual existem esforços internacionais para disponibilizar a tecnologia de ultrassom para essas sociedades [Kiserud et al., 2018]. É de grande importância frisar também que os intervalos de referência atualmente disponíveis para a biometria ultrassonográfica fetal são derivados na maioria de populações únicas em sociedades industrializadas com aplicabilidade incerta num mundo de variação étnica [Kiserud et al., 2018]. Demonstrando assim, a necessidade da existência de gráficos de crescimento fetal e infantil para uso internacional. Se para além desses gráficos fosse possível desenvolver uma referência com parâmetros biométricos placentares, estas taxas de mortalidade perinatal poderiam ser ainda mais reduzidas e seria possível detetar complicações precoces na gravidez e evitar vários problemas futuramente.

No sentido de auxiliar a visualização do crescimento de uma forma simples e intuitiva, são utilizadas curvas de crescimento. As curvas de crescimento são um método estatístico frequentemente usado na análise de dados em que as mesmas unidades experimentais são observadas ao longo do tempo [Kshirsagar and Smith, 1995]. São uma ferramenta gráfica poderosa e no caso da descrição do desenvolvimento infantil, conseguem exibir o tamanho da criança numa série de idades e, em simultâneo, a sua taxa de crescimento ou velocidade de crescimento ao longo do tempo, com base na inclinação da curva [Cole, 2012]. Portanto, as curvas de crescimento são usadas para avaliar o bem-estar e tomar decisões sobre intervenções em indivíduos, em termos de estratégias de alimentação e práticas de cuidados com a saúde, por exemplo [Lampl, 2012].

Desde o seu primeiro surgimento no século XVIII, o uso de gráficos do crescimento expandiu-se para incluir a exibição do padrão de crescimento de grupos de crianças, bem como indivíduos, e o gráfico tornou-se uma ferramenta importante na triagem de saúde infantil e avaliação clínica pediátrica [Cole, 2012]. Estes gráficos têm sido tradicionalmente usados para fornecer um ponto de referência para responder à questão de quão bem uma criança individual está a “crescer” por inferência da sua posição no resumo estatístico [Lampl, 2012]. Como qualquer equipamento médico, o seu uso tem limitações que devem ser bem compreendidas.

O conhecimento de como os gráficos de crescimento são construídos pode ajudar na compreensão de quais perguntas eles podem ou não responder e a sua interpretação geralmente requer uma comparação entre diferentes medidas simultâneas ou sucessivas ao longo do tempo [Wright, 2002]. É importante ter em consideração que os gráficos de crescimento não podem fazer um diagnóstico, mas podem fornecer pistas importantes, bem como garantias vitais.

O gráfico de crescimento europeu, consiste em 7 linhas de percentis (3, 10, 25, 50, 75, 90, 97) [Cole, 1994]. Cada linha no percentil n marca o peso ou altura abaixo do qual $n\%$ das crianças dessa idade e gênero caem (por exemplo, 25% das crianças estão abaixo ou no 25.º percentil) [Wright, 2002]. O percentil 50 representa a mediana da população, enquanto os percentis 3 e 97 são, respectivamente, dois **DP** acima e abaixo da mediana [Wright, 2002]. A distância entre cada linha de percentil (2/3 de um **DP**) é conhecida como espaço percentil. O percentil mais baixo, o 3, foi escolhido para identificar medidas extremamente baixas.

Dessa forma, os gráficos de percentis apenas permitem dizer onde cada medida coloca uma criança relativamente a outras crianças na mesma idade. Estes são descritos como sendo “transversais”, pois são construídos usando diferentes crianças em cada idade de medição e são, na verdade, uma série de distribuições normais unidas por linhas, como podemos observar na Figura 3 [Wright, 2002]. Assim, embora muitas vezes chamadas “curvas de crescimento”, as linhas não descrevem o crescimento ao longo do tempo em indivíduos, pois a base em que se baseiam, não fornece uma compreensão do processo dinâmico pelo qual o crescimento realmente ocorre ou a biologia do crescimento [Lamp, 2012].

Apesar de existir uma grande aplicação das curvas de crescimento em várias áreas, o pressuposto da sua interpretação é igual. Ou seja, a leitura dos vários percentis nas curvas é analisada da mesma forma, quer seja do peso infantil ao longo da sua idade, como foi esclarecido anteriormente, ou do peso placentar ao longo da **IG**, ou qualquer outro parâmetro.

2.2.4 Aplicação de curvas de crescimento em várias áreas

Ao longo dos anos, uma ampla aplicação das curvas de crescimento nos mais diversos campos das ciências sociais e da vida tem ocorrido. Na literatura podem ser encontrados vários estudos que relacionam o conceito das curvas de crescimento aplicado a diversas áreas da biologia e da medicina [Thompson et al., 2007, Nogueira et al., 2019, Almog et al., 2011, Anchieta et al., 2003, Kiserud et al., 2018, Kaplan and Gürcan, 2018, Nakamura et al., 2022]. Demonstrando assim, a atualidade deste método e a existência de uma grande evidência científica.

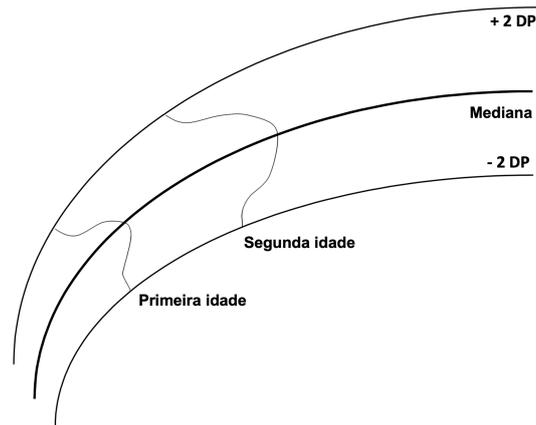


Figura 3: Gráfico de crescimento. Representa uma série de distribuições normais em idades diferentes com linhas conectando a mediana (50.º percentil) e outros percentis. Adaptado de: [Wright \[2002\]](#).

Começando pela temática abordada nesta dissertação, aplicação de curvas de crescimento para parâmetros placentares, [Thompson et al. \[2007\]](#) foram os primeiros a produzir curvas de percentis de referência da população norueguesa para o **PP** e a perceber a razão entre o **PN** e o **PP**. Estatisticamente, medianas, quartis inferiores, quartis superiores e intervalos interquartis foram calculados para cada semana gestacional completa. De forma a obter a normalidade para as faixas etárias e ser possível construir as curvas de percentis, foi utilizado o método **LMS** de [Cole and Green \[1992\]](#). Para permitir a natureza cúbica geral dos percentis placentares, os modelos foram ajustados aos pontos percentuais que incluíam um termo linear, um quadrado e um cúbico. Todos os termos do modelo foram estatisticamente significativos ($P < 0,0001$). A natureza das curvas de razão levou à inclusão de um termo linear e um quadrado e os termos do modelo ajustaram-se bem às curvas.

Em outro estudo dentro da mesma temática [Nogueira et al. \[2019\]](#), conseguiram produzir valores de referência para o **PP**, **PF**, **R-PP** e **R-PPN** em gestações unifetais em função da **IG**. Com estes valores foram produzidas tabelas e curvas de percentis dos parâmetros placentares para a população portuguesa e foi possível concluir que as curvas podem servir de referência também para outras populações. Que os diâmetros da placenta podem prever o volume placentar e podem ajudar a estimar o **R-PP** e o **R-PPN** pré-natal. E que existe uma associação significativa entre as medidas da placenta e a avaliação da função placentar (relacionada ao tamanho e volume) e a sua implicação no crescimento fetal, auxiliando os médicos na prevenção de riscos à vida fetal e melhorando a saúde materno-infantil. Quanto ao método estatístico para a criação das curvas de crescimento, foram utilizados métodos de regressão linear estatisticamente significativos ($P < 0,0001$).

Já [Almog et al. \[2011\]](#) estabeleceram curvas de percentis com parâmetros placentares utilizando

dados da população norte-americana para gestações unifetais e gemelares. No que concerne às estatísticas utilizadas, o **DP** médio, erro padrão, mediana, quartis inferiores, quartis superiores e intervalos interquartis foram calculados para cada idade gestacional e para gêneros separados. O percentil do **PP** foi calculado usando o método **LMS** de [Cole and Green \[1992\]](#). As condições do modelo foram estatisticamente significativas ($P < 0,0001$) e ajustaram-se bem às curvas.

Passando para outra utilização das curvas de crescimento, aplicadas ao crescimento infantil, [Anchieta et al. \[2003\]](#) avaliou o peso de recém-nascidos pré-termo adequados para a **IG** nas primeiras doze semanas de vida. No seu estudo testaram vários modelos não lineares como o polinômio de segunda ordem, polinômio de terceira ordem, logístico, Monomolecular (Jenss-Baykey) e Count. Estes modelos de regressão foram ajustados pelo método dos mínimos quadrados utilizando o *software* estatístico SPSS. Usando médias do peso, as curvas de crescimento foram construídas e ajustadas ao modelo de Count, pois foi o modelo que melhor se ajustou aos dados. As curvas de velocidade absoluta (g/dia) e relativa ($g\ kg^{-1}\ dia^{-1}$) foram obtidas a partir de uma derivada desse modelo. Foi concluído que a velocidade relativa descreve melhor a dinâmica do peso de prematuros, principalmente daqueles com menor peso ao nascer, que a velocidade relativa de ganho de peso foi inversamente proporcional ao peso ao nascer, e o pico de velocidade em todos os pesos ao nascer ocorreram durante a 3.^a semana de vida.

A Organização Mundial da Saúde de forma a controlar o crescimento fetal também tem apostado na utilização de gráficos de crescimento para uso internacional. Um exemplo, é o estudo de [Kiserud et al. \[2018\]](#) onde são realizados gráficos de crescimento fetal para medidas fetais comuns e peso fetal estimado baseados num estudo longitudinal de 1387 mulheres grávidas de baixo risco de 10 países. De forma a estabelecer os gráficos de crescimento, foi utilizada a regressão quantílica que possibilitou a demonstração de várias características do crescimento fetal: (1) Houve uma distribuição assimétrica do peso fetal estimado na população. (2) Para além do sexo fetal, os fatores maternos (altura, peso, idade e paridade) e o país tiveram influência significativa no peso fetal e o seu efeito foi classificado nos percentis. (3) Quando o ajuste foi feito para as covariáveis maternas, ainda houve um efeito significativo do país como covariável que indicou que a variação étnica, cultural e geográfica desempenha um papel. (4) A variação entre as populações não se restringiu ao tamanho fetal porque também houve diferenças nas trajetórias de crescimento. (5) As amplas faixas fisiológicas, conforme ilustrado pelo 5.^o - 95.^o percentil para o peso fetal estimado, significam que o crescimento fetal humano sob condições maternas otimizadas não é uniforme.

O estudo e entendimento das curvas de crescimento também desempenha um papel importante nas práticas de manuseamento, alimentação e melhoramento genético de vários animais. Cada vez

mais têm existido estudos nesta área do crescimento animal, pois é possível tirar várias ilações a partir dessa informação. Um exemplo, foi o estudo de [Kaplan and Gürçan \[2018\]](#) onde o objetivo foi comparar o crescimento de codornas machos e fêmeas com a função não linear mais adequada conforme os critérios de qualidade de ajuste (coeficiente de determinação, coeficiente de determinação ajustado, erro quadrático médio, critérios de informação de Akaike e critérios de informação de Schwarz Bayesian). Para determinar o modelo de crescimento mais consistente para as codornas, os modelos Richards, Janoschek e Levakovich, conhecidos por terem pontos de inflexão flexíveis e os modelos Gompertz, Logístico e Von Bertalanffy que apresentam comportamento fixo em termos de ponto de inflexão foram utilizados. Foi possível concluir que o peso vivo das fêmeas foi maior que o dos machos, e a primeira divergência no seu crescimento ocorreu às 4 semanas. Foi também observado que todos os modelos tiveram um bom desempenho na descrição de mudanças relacionadas à idade no peso vivo em codornas. No entanto, o modelo de Richards foi determinado como o modelo que melhor se ajustava aos dados de crescimento de ambos os sexos.

Outro estudo realizado em animais foi o de [Nakamura et al. \[2022\]](#), onde forneceram uma alternativa para criar curvas de crescimento de referência para fêmeas da raça Hereford com base na estrutura de **GAMLSS**. Ou seja, o objetivo foi desenvolver uma metodologia flexível para ajustar as curvas de crescimento de referência para que os produtores de gado possam estabelecer tomadas de decisão necessárias com base nas curvas padrão de crescimento em todas as faixas etárias, evitando possíveis processos de estimativa de inferência como problemas de cruzamento de quantis. Foi possível concluir que a idade do animal afeta diretamente as características da mediana, variabilidade e assimetria da sua curva de crescimento e que esta abordagem poderá ser aplicada em qualquer outro crescimento animal ou vegetal, e pode ser usada como uma poderosa ferramenta de tomada de decisão pelos produtores.

2.3 Software R

Atualmente, vários *softwares* são utilizados de forma a guiar e auxiliar investigadores ou até profissionais de saúde nas suas tomadas de decisão. Alguns dos *softwares* mais comumente usados são, por exemplo, o SPSS®, o MedCalc, o Stata® e o R. Cada um destes programas de análise estatística oferecem recursos diversificados e com limitações e vantagens diferentes. No entanto, entre estes, o R tem demonstrado ser uma ferramenta completa e com várias vantagens relativamente aos restantes.

O projeto GNU R, [[TEAM, 2005](#)], mais conhecido como R, é uma linguagem e um ambiente para computação estatística no formato de projeto de *software* livre de código aberto (*open source*), ou seja,

pode ser utilizado sem custos de licença. O objetivo da criação do R foi permitir o desenvolvimento de um sistema que fosse totalmente livre, que pudesse ser modificado, transformado e acrescido de funções consoante as necessidades dos utilizadores ou das organizações [Konrath et al., 2018]. Nesse contexto, mais do que um *software*, o R é uma linguagem e um ambiente de desenvolvimento integrado para computação estatística. Além de todo o trabalho que pode ser feito com as funções padrões do R, na área da Estatística, é possível utilizar pacotes desenvolvidos pela comunidade para resolver problemas em outras áreas como a Engenharia, Economia, Biologia, Ciência da Saúde e Ciências Sociais [Konrath et al., 2018].

No entanto, o R é um programa conduzido por meio de digitação de linhas de comando, o que pode dificultar a sua operacionalização por pessoas que não possuem conhecimentos de programação [Konrath et al., 2018]. Por isso, de forma a aprimorar a sua usabilidade foi desenvolvida uma *interface* gráfica, o RStudio. O RStudio [Racine, 2011] é um ambiente de desenvolvimento integrado (**IDE**) que reúne diversas funcionalidades, tornando mais amigáveis aos utilizadores a importação de dados, visualização de comandos, funções, resultados e gráficos, além da geração de documentos [Henning et al., 2016]. Este combina vários componentes do R, tal como, a consola, a edição de código-fonte, os gráficos, o *knitr*, o *rmarkdown*, entre outros, num ambiente de trabalho único [Gandrud, 2018]. Portanto, o R é demonstrado como um *software* extremamente útil na análise estatística e com imensas vantagens face aos seus competidores.

2.3.1 Shiny

De forma a viabilizar ainda mais a utilização do *software* R, e permitir que este seja mais acessível a, por exemplo, um profissional que não tenha conhecimentos de programação em R e pretenda uma análise rápida de dados, foi criado o pacote *shiny*. Com este pacote, é possível transformar análises complexas em aplicações interativas e amigáveis utilizadas via *web*, sendo assim uma alternativa promissora para contribuir com inovações na área da Estatística.

Uma ampla variedade de aplicações podem ser desenvolvidas a partir do *shiny*. Por exemplo, é possível desenvolver aplicações de valores padrão para uma classe de elementos (*default* CSS) totalmente escritas em R, páginas na *web*, por exemplo, linguagem de marcação de hipertexto (HTML), folhas de estilo em cascata (CSS), linguagem orientada por objetos (JavaScript) e bibliotecas de funções JavaScript que interage com o HTML (jQuery) e documentos interativos que contêm janelas, botões, ícones, menus, barras de rolagem e outras funcionalidades (*widgets*) *shiny* incorporados [Konrath et al., 2018].

As aplicações *shiny* são construídas utilizando dois *scripts* R que se comunicam entre si: um *script*

de interface do utilizador (ui.R), que controla o *layout* e a aparência; e um *script* de servidor (server.R), incorporando instruções para entrada do utilizador, processamento de dados e saída, utilizando a linguagem R e as funções dos pacotes instalados pelo utilizador [Wojciechowski et al., 2015].

2.3.2 Pacotes

O R tem sido a ferramenta preeminente para estatísticos e cientistas de dados, pois possui mais de 6.000 pacotes, que auxiliam os utilizadores em tarefas diversificadas como a química, biologia, física, finanças, psicologia e ciências médicas, além de desenhar gráficos estatísticos extremamente poderosos e flexíveis [Beeley, 2016].

Ao longo desta dissertação foram utilizados vários pacotes para auxiliar o seu desenvolvimento. Alguns exemplos de pacotes amplamente utilizados na estatística e que foram úteis ao longo deste trabalho são:

- **gam1ss**: Este pacote possui funções para ajustar os modelos aditivos generalizados para escala e forma de localização [gam1ss]. Os modelos usam uma abordagem de regressão distributiva onde todos os parâmetros da distribuição condicional da variável de resposta são modelados usando variáveis explicativas.
- **ggplot2**: Este pacote fornece ferramentas gráficas poderosas e elegantes, que podem ser utilizadas para traçar dados e fazer ajustes de modelo [ggplot2].
- **dplyr**: Uma ferramenta rápida e consistente para trabalhar com objetos *data frame* [dplyr].

2.3.3 Aplicação do R

O R tem sido extensivamente utilizado em várias áreas e mais recentemente, o *shiny* tem captado a atenção dos investigadores de forma a tornar o seu trabalho mais acessível a partir de uma aplicação *web*.

Já existem alguns estudos sobre as curvas de crescimento em que uma aplicação *shiny* é gerada. Um exemplo, é o estudo de Young et al. [2020], em que apresentam gráficos de crescimento derivados de medidas seriadas de peso, comprimento e circunferência da cabeça, refletindo o crescimento pós-natal de bebés prematuros. Os gráficos de percentis foram construídos utilizando o método **LMS** do pacote **gam1ss** do R. Com estes dados, os autores desenvolveram uma aplicação *shiny* para permitir que os médicos rastreiem o crescimento de bebés nos gráficos de crescimento desenvolvidos na aplicação.

Outro exemplo, da aplicação de curvas de crescimento em R é o trabalho de Alves et al. [2022]. As curvas de crescimento são construídas para parâmetros placentares como, por exemplo, o **PP**, o **D1** e o

D2, entre outros. Para a sua criação foram utilizados os pacotes `quantregGrowth` e `ggplot2`. Com os dados obtidos foi possível criar uma aplicação dinâmica `shiny` que permite ao investigador ou profissional de saúde introduzir valores de crescimento placentar e compará-los com as curvas de crescimento criadas para avaliar a evolução da placenta. A aplicação usa um ficheiro CSV com as informações recolhidas da placenta e depois coloca esses valores nas curvas de crescimento criadas.

Na área da microbiologia o R também pode ser benéfico, sendo o estudo de [Garel et al. \[2022\]](#) um exemplo de aplicação. No seu trabalho, foi proposta uma ferramenta para estimar parâmetros de crescimento usando um modelo logístico de Verhulst que considera toda a curva de crescimento para a estimativa da taxa de crescimento. Para o seu desenvolvimento, foram utilizados os pacotes `stats` e `quantreg` e foi criada uma aplicação `shiny`. A aplicação permite estimar parâmetros de crescimento, incluindo a taxa de crescimento, a população máxima e o início da fase exponencial, bem como estimar a sua variabilidade. Os resultados podem ser exibidos na forma de um gráfico de dispersão representando o modelo e a sua eficiência, sendo possível descarregar os parâmetros estimados.

Capítulo 3

Regressão

Este capítulo estabelece uma base fundamental para a compreensão das metodologias posteriormente aplicadas na construção das curvas de crescimento. Exploram-se detalhadamente as técnicas de regressão, dividindo o capítulo em quatro secções principais. Começa-se por abordar o conceito fundamental da regressão, seguido pela análise detalhada de dois tipos principais: regressão linear e não linear. Além disso, investigam-se as distinções entre regressão paramétrica, semiparamétrica e não paramétrica. Esta distinção é crucial, pois influenciará a escolha das estratégias de modelação mais adequadas para as curvas de crescimento. A compreensão destas abordagens é essencial para a seleção criteriosa das técnicas a serem aplicadas neste estudo.

Em última análise, este capítulo serve como o alicerce teórico sólido sobre o qual serão desenvolvidas as metodologias para as curvas de crescimento. A compreensão dos conceitos aqui apresentados será crucial para a análise precisa e a interpretação dos resultados obtidos nas etapas subsequentes do estudo.

3.1 Conceito e definição de regressão

A regressão é uma técnica estatística fundamental que permite elucidar as relações entre variáveis. Essencialmente, a regressão é utilizada para estabelecer relações entre variáveis preditoras (conhecidas como variáveis independentes ou covariáveis) e uma variável resposta (variável dependente). Sendo o objetivo principal explorar e compreender as variações na variável resposta que podem ser parcialmente atribuídas às variáveis preditoras [WU and QIU, 2021a]. Esta técnica pode englobar vários objetivos como a previsão, teste da presença de relações com preditores individuais (potencialmente ajustando para outros preditores) e a modelação da relação entre a variável resposta e todos os preditores envolvidos [Vittinghoff et al., 2007].

A seleção de um modelo de regressão apropriado depende de dois fatores-chave. Primeiramente, o tipo de variável resposta que está a ser modelada, visto que determina a parte aleatória do modelo. Adicionalmente, a forma como os parâmetros a serem estimados são incorporados no modelo, pois

permite determinar se o modelo é linear ou não linear. A escolha entre a regressão linear e a não linear é um ponto crucial na modelação estatística, pois irá determinar o quão bem o modelo se irá ajustar aos dados e o quão preciso este será a capturar as relações entre as variáveis. Ambas as abordagens têm as suas aplicações e limitações, e a decisão irá depender essencialmente da natureza dos dados.

3.2 Regressão linear

A regressão linear é um modelo estatístico fundamental e amplamente utilizado pela comunidade científica durante grande parte do último século devido à sua simplicidade e facilidade de interpretação [WU and QIU, 2021b]. Este modelo é uma ferramenta valiosa para compreender e quantificar a relação entre uma variável resposta e um conjunto de covariáveis, mesmo que essa relação possa não capturar exatamente a complexidade subjacente. Supondo existir $p+1$ variáveis, representadas como $(y, x_1, x_2, \dots, x_p)$, onde a variável y é selecionada como a resposta, enquanto as demais variáveis são tratadas como preditoras ou covariáveis [WU and QIU, 2021b]. Se se considerar um conjunto de dados de dimensão n , pode-se denotar como $(y_i, x_{i1}, x_{i2}, \dots, x_{ip}), i = 1, 2, \dots, n$.

Um modelo de regressão linear está expresso na Equação 3.1 [WU and QIU, 2021b].

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \dots, n$$

ou

(3.1)

$$E(y_i) = \mathbf{x}_i^T \boldsymbol{\beta},$$

Aqui, y_i é a resposta para o indivíduo i , $\boldsymbol{\beta}$ é o vetor dos parâmetros desconhecidos, x_{ij} representa o j -ésimo preditor para o indivíduo i , $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})^T$ é uma matriz de todos os preditores, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T$ é um vetor de parâmetros de regressão, e ε_i são erros aleatórios com média zero, $i = 1, 2, \dots, n$ [WU and QIU, 2021b].

No modelo linear (Equação 3.1), pressupõe-se que a relação entre a resposta e os preditores é linear. Essa linearidade permite interpretar os parâmetros de regressão β_j , como o efeito do preditor x_i na resposta y , ou seja, uma mudança de uma unidade em x_i está associada a uma mudança de β_j unidades em y [WU and QIU, 2021b]. Além disso, essa forma linear possibilita derivar a distribuição de y com base na distribuição assumida para ε e estudar as propriedades das estimativas dos parâmetros [WU and QIU, 2021b]. Na prática, a suposição de uma relação linear pode ser razoável e útil, especialmente quando alguns preditores são transformados.

3.3 Regressão não linear

Os modelos de regressão linear têm sido amplamente adotados devido à sua simplicidade, o que representou uma vantagem significativa numa época anterior à disponibilidade de computadores modernos. No entanto, é importante notar que estes modelos frequentemente limitam-se a fornecer uma descrição dos dados observados, sem necessariamente procurar uma compreensão mais profunda dos processos subjacentes que geraram esses dados. Em essência, os modelos lineares tendem a enfatizar o ajuste aos dados observados.

Em contrapartida, os modelos de regressão não linear têm como objetivo principal a compreensão dos processos subjacentes à geração dos dados. Eles oferecem várias vantagens relativamente aos modelos lineares. Em primeiro lugar, estes modelos podem fornecer previsões mais precisas quando extrapoladas para fora do intervalo dos dados observados. Em segundo lugar, os parâmetros em modelos não lineares frequentemente têm interpretações físicas naturais, tornando mais acessível a compreensão dos resultados. Terceiro, modelos não lineares podem ser mais parcimoniosos em termos de parâmetros, ou seja, requerem menos parâmetros para se ajustarem eficazmente aos dados.

Porém, é crucial destacar que em muitas situações práticas, os mecanismos subjacentes à geração dos dados permanecem desconhecidos. Nestes casos, os modelos lineares podem ser uma escolha mais apropriada, uma vez que se concentram na descrição direta dos dados observados, evitando suposições substanciais sobre a natureza do processo gerador.

É importante notar que, ao contrário dos modelos lineares, os modelos não lineares geralmente não possuem soluções analíticas diretas para a estimativa de parâmetros. Portanto, a obtenção de estimativas de parâmetros em modelos não lineares geralmente requer o uso de algoritmos iterativos [WU and QIU, 2021b]. Além disso, ao ajustar modelos não lineares, a escolha de valores iniciais adequados para os algoritmos iterativos desempenha um papel crucial, pois algumas funções de verosimilhança podem apresentar múltiplos modos, tornando o processo de otimização desafiador [WU and QIU, 2021b].

Seja y_i a resposta e $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ os preditores para o indivíduo i , onde $i = 1, 2, \dots, n$. Um modelo geral de regressão não linear pode ser expresso como na Equação 3.2 [WU and QIU, 2021b].

$$y_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3.2)$$

Nesta equação, h é uma função não linear conhecida, $\boldsymbol{\beta}$ é um vetor de parâmetros de regressão e ε_i é o erro aleatório. As suposições para um modelo de regressão não linear padrão são as mesmas de um modelo linear padrão, ou seja, (i) os erros ε_i são independentes, (ii) os erros ε_i têm média zero e

variância constante σ^2 , e (iii) os erros ε_i são normalmente distribuídos [WU and QIU, 2021b].

A inferência estatística para um modelo de regressão não linear pode ser baseada no método dos mínimos quadrados ou no método da verosimilhança. O estimador dos mínimos quadrados ordinários para o parâmetro β envolve a minimização da soma dos quadrados $\sum_{i=1}^n (y_i - g(\mathbf{x}_i, \beta))^2$. Isso pode ser alcançado resolvendo a seguinte Equação 3.3 de estimação [WU and QIU, 2021b].

$$\sum_{i=1}^n \frac{\partial g(\mathbf{x}_i, \beta)}{\partial \beta} [y_i - g(\mathbf{x}_i, \beta)] = 0 \quad (3.3)$$

Um algoritmo iterativo, como o método de Newton-Raphson, é frequentemente necessário para resolver a equação acima.

Alternativamente, sob a suposição de normalidade para os erros, ou seja, por exemplo, ε_i i.i.d. $\sim N(0, \sigma^2)$, a Estimação de Máxima Verosimilhança (**EMV**) de β pode ser obtida maximizando a função de verosimilhança como na Equação 3.4 [WU and QIU, 2021b].

$$L(\beta, \sigma^2 | \mathbf{y}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - g(\mathbf{x}_i, \beta))^2}{2\sigma^2}\right]. \quad (3.4)$$

Portanto, a **EMV** de β satisfaz a equação 3.5 de verosimilhança [WU and QIU, 2021b].

$$\frac{\partial \log L(\beta, \sigma^2 | \mathbf{y})}{\partial \beta} = 0, \quad (3.5)$$

Para modelos de regressão não linear, o estimador dos mínimos quadrados ordinários de β também é o mesmo que o **EMV** de β , e a estimação para um modelo de regressão não linear é análoga à de um modelo de regressão linear [WU and QIU, 2021b].

Para modelos de regressão não linear, não estão disponíveis expressões analíticas ou de formato fechado para estimativas de parâmetros. No entanto, a inferência estatística ainda pode ser realizada com base nos resultados assintóticos padrão dos métodos de verosimilhança sob as condições de regularidade usuais [WU and QIU, 2021b]. Ou seja, sob algumas condições de regularidade, as **EMV** dos parâmetros do modelo são consistentes, assintoticamente normais e assintoticamente mais eficientes.

Intervalos de confiança e testes de hipóteses podem ser baseados na normalidade assintótica das **EMV**. Portanto, com a disponibilidade de computadores e *software* modernos, a inferência estatística para modelos não lineares não oferece mais dificuldades do que para modelos lineares.

3.4 Regressão paramétrica, semi e não paramétrica

Para distinguir entre os três tipos de modelos de regressão, considere-se um cenário onde existe uma variável de resposta denotada como Y , com duas variáveis explicativas, x_1 e x_2 [Mahmoud, 2019]. Pode-se expressar o modelo de regressão que elucida a associação entre a variável resposta e essas duas variáveis explicativas da seguinte forma [Mahmoud, 2019]:

$$Y = f_1(x_1, \beta_1) + f_2(x_2, \beta_2) + \varepsilon \quad (3.6)$$

Aqui, β_1 e β_2 representam parâmetros que precisam de ser estimados, ε um termo de erro aleatório e $f_1(\cdot)$ e $f_2(\cdot)$ são funções que delinham as relações.

3.4.1 Regressão paramétrica

Os modelos de regressão paramétricos são caracterizados por um vetor de parâmetros (β) de dimensão finita p , uma forma funcional pré-definida ($f(\cdot)$), e uma distribuição predefinida para ε [Mahmoud, 2019]. O objetivo principal da sua utilização passa por estimar o vetor de parâmetros. Estes modelos são apropriados quando há um conhecimento prévio sobre a relação entre as variáveis, uma vez que eles impõem uma estrutura rígida e bem definida. No entanto, essa simplicidade também significa que eles podem ser limitados em termos de flexibilidade. Se a forma funcional escolhida não corresponder à verdadeira relação subjacente, as previsões do modelo podem ser imprecisas. Portanto, é fundamental realizar verificações de validação, geralmente com base na análise dos resíduos, para garantir que as suposições do modelo estão corretas.

Embora os modelos paramétricos sejam amplamente utilizados devido à sua interpretabilidade e simplicidade, é essencial ter em mente que a escolha da forma funcional deve ser fundamentada num entendimento da relação entre as variáveis, de modo a obter resultados confiáveis.

3.4.2 Regressão não paramétrica

Os modelos de regressão não paramétricos distinguem-se fundamentalmente dos modelos paramétricos pela forma como concebem a estrutura do modelo. Ao contrário dos modelos paramétricos, nos quais a estrutura do modelo é rigidamente predefinida, os modelos não paramétricos não se vinculam a uma estrutura de modelo previamente estabelecida.

Estes modelos são caracterizados por explorar um domínio onde o conjunto de parâmetros pertence

a um subconjunto de um espaço vetorial de dimensão infinita, e o objetivo principal é estimar esse vetor de parâmetros de dimensão infinita [Mahmoud, 2019]. Nestes modelos, a relação entre as variáveis explicativas e a variável resposta permanece desconhecida. Aplicando na Equação 3.6, tanto f_1 quanto f_2 são funções desconhecidas. Estas funções podem assumir qualquer forma, seja uma relação linear ou não linear, porém são desconhecidas para quem as esteja a estudar. É o conjunto de dados que ditará a estrutura do modelo de regressão e as características das funções f_1 e f_2 [Mahmoud, 2019]. Em contraste, na regressão paramétrica, a forma exata do modelo utilizado para ajustar os dados é conhecida, como, por exemplo, uma simples reta de regressão linear.

Os modelos não paramétricos oferecem a liberdade de ultrapassar os limites convencionais. São a escolha predileta quando a verdadeira natureza da relação é desconhecida ou quando se deseja evitar a imposição de restrições sobre o modelo.

3.4.3 Regressão semiparamétrica

Os modelos semiparamétricos encontram um equilíbrio entre as suposições rígidas dos modelos paramétricos e a flexibilidade dos modelos não paramétricos. Embora à primeira vista possam parecer semelhantes aos modelos não paramétricos, os modelos semiparamétricos possuem características distintas, concentrando-se principalmente na componente finita de β , enquanto os modelos não paramétricos concentram-se principalmente na estimativa do parâmetro de dimensão infinita [Mahmoud, 2019]. Esta distinção torna frequentemente o processo de estimação nos modelos semiparamétricos estatisticamente mais acessível em comparação com os seus homólogos totalmente não paramétricos. Os modelos paramétricos, embora sejam simples e de fácil uso, tendem a simplificar excessivamente e, portanto, podem não representar adequadamente a complexidade dos dados. Os modelos semiparamétricos conseguem conciliar a interpretabilidade com a capacidade de oferecer uma representação mais fiel.

Dentro do domínio dos modelos de regressão semiparamétricos, existe uma gama diversificada de estruturas. Uma é uma forma de análise de regressão em que uma parte dos preditores não assume formas predeterminadas e a outra assume formas conhecidas com a variável resposta [Mahmoud, 2019]. Por exemplo, no modelo descrito na Equação 3.6, f_1 pode ser conhecido (suponha-se linear) e f_2 desconhecido. Esta situação pode ser formulada da seguinte forma [Mahmoud, 2019]:

$$Y = \beta_0 + \beta_1 x_1 + f(x_2) + \varepsilon \quad (3.7)$$

Aqui, a relação entre x_1 e a resposta é assumida como linear, enquanto a relação entre a resposta e x_2 é desconhecida.

Em resumo, a escolha do modelo de regressão depende da complexidade da relação entre as variáveis e do nível de conhecimento prévio sobre essa relação. Os modelos paramétricos são diretos, mas dependem de suposições específicas. Os modelos não paramétricos oferecem flexibilidade, mas podem sofrer de problemas de dimensionalidade. Os modelos semiparamétricos fornecem um meio-termo, combinando as vantagens de ambas as abordagens. Logo, estes fatores devem ser cuidadosamente considerados ao selecionar um modelo de regressão para a análise de dados.

Capítulo 4

Métodos para a construção das curvas de crescimento

Este capítulo aborda a metodologia fundamental para a construção das curvas de crescimento. Começa-se com considerações metodológicas que ajudam a compreender os aspetos distributivos dos dados e as técnicas de alisamento necessárias para obter as curvas. A seleção dos métodos apropriados é abordada com critérios específicos e uma revisão abrangente dos métodos disponíveis. Destaca-se o **GAMLSS**, explorando o seu modelo estatístico, algoritmos, estimação de hiperparâmetros, técnicas de alisamento e distribuições. Além disso, apresenta-se o método **LMS** e as suas extensões, **LMST** e **LMSP**, que oferecem abordagens alternativas. Explica-se como estimar percentis e avaliar a qualidade dos ajustes, abrangendo uma gama dispersa de critérios.

Ao longo deste capítulo, são fornecidas informações detalhadas, ferramentas e pacotes essenciais para o desenvolvimento de curvas de crescimento, fornecendo um guia para explorar e compreender a metodologia que sustenta a construção das curvas de crescimento para os parâmetros biométricos da placenta.

4.1 Considerações metodológicas

O desenvolvimento das curvas de crescimento requer uma análise das propriedades de distribuição dos dados. Em casos de parâmetros amplamente estudados na literatura, como o crescimento e o peso de crianças, o conhecimento prévio pode servir como base para esse processo. No entanto, quando se trata de parâmetros relacionados à placenta, o desafio é ampliado, uma vez que esse é um campo de estudo relativamente pouco explorado.

4.1.1 Aspetos distributivos

Os métodos que se fundamentam em pressupostos distributivos têm ganho uma relevância crescente devido à sua capacidade de gerar *z-scores* e estimativas mais precisas dos percentis extremos. No entanto, é essencial assegurar uma concordância adequada entre a distribuição dos dados e os pressupostos

distributivos inerentes ao método selecionado. Tipicamente, as distribuições são caracterizadas com base em três estatísticas fundamentais: a média, o **DP** (ou o coeficiente de variação (**CV**)) e a assimetria [Borghiet al., 2006]. Contudo, reconhece-se cada vez mais a importância potencial da curtose, na obtenção de estimativas precisas dos percentis extremos.

Uma vez que as distribuições são descritas pelos quatro momentos — média, **DP**, assimetria e curtose — é de extrema importância que estas estatísticas sejam analisadas previamente nos dados antes de serem efetuadas suposições. No contexto deste estudo, que analisa parâmetros ao longo da **IG**, é essencial que estas estatísticas sejam avaliadas para cada **IG**. É de salientar que modelar a curtose por meio de extrapolação pode ser desafiador, devido à informação esparsa geralmente disponível nas caudas.

4.1.2 Aspetos de alisamento

Determinar o nível ideal de alisamento necessário para uma estimativa adequada dos percentis é uma tarefa de grande complexidade. Desconsiderar o alisamento pode resultar em curvas de crescimento excessivamente irregulares, mesmo em casos com amostras numerosas. No contexto do estudo do crescimento placentário, esse fenómeno surge devido à variabilidade da amostragem ao longo das diferentes **IG**, bem como da influência de valores discrepantes remanescentes e da variabilidade nas medições. As técnicas de alisamento são aplicadas com o objetivo de mitigar esses efeitos, sem distorcer os verdadeiros padrões de crescimento subjacentes. No entanto, uma aplicação excessiva de alisamento pode achatá-los os picos e vales característicos dos padrões de crescimento biológico, o que, em última análise, resulta em curvas de crescimento ajustadas enviesadas.

É fundamental destacar que a seleção da função de alisamento é tão crucial quanto a escolha da distribuição utilizada. A determinação de qual método de alisamento aplicar requer uma avaliação cuidadosa, considerando as características específicas dos dados e os objetivos da análise. Além disso, é importante mencionar que a adequação do alisamento deve ser ajustada para se adaptar às diferentes **IG**, uma vez que as características da amostra podem variar ao longo do período gestacional.

4.2 Seleção de métodos

A construção de curvas de crescimento requer uma seleção cuidadosa de métodos que considerem as características do conjunto de dados em análise e as suas futuras aplicações. Como tal, são necessários critérios que orientem a seleção dos métodos.

4.2.1 Critérios para a seleção dos métodos

No contexto da construção de curvas de crescimento, a seleção de métodos adequados desempenha um papel fundamental, sendo essencial considerar criteriosamente as características do conjunto de dados em questão e as futuras aplicações das curvas de crescimento de referência. Portanto, neste estudo, a escolha dos métodos para desenvolver as curvas de crescimento levou em consideração diversos critérios, com o objetivo de orientar de forma fundamentada todo o processo de seleção. Os critérios foram os seguintes:

- **Precisão na estimação de percentis extremos:** A obtenção de estimativas precisas para os percentis nas caudas da distribuição é um desafio complexo, principalmente devido à escassez de dados nessa região. Este desafio requer métodos que sejam capazes de oferecer essa precisão e por isso, que dependam de uma distribuição subjacente.
- **Evitar o cruzamento de curvas de percentis:** Garantir que os percentis não se cruzem é essencial para a integridade das curvas de crescimento. Para atingir esse objetivo, torna-se imperativo calcular simultaneamente os percentis, evitando qualquer sobreposição que comprometa a interpretação adequada das curvas.
- **Retrotransformações:** Os métodos selecionados devem permitir retrotransformações de forma a ser possível o cálculo direto de percentis e *z-scores*, tornando a interpretação dos dados mais acessível e direta.
- **Curtose:** São preferíveis métodos capazes de lidar com a curtose, além da assimetria. A curtose pode manifestar-se em certas medições, e um modelo incorreto poderá distorcer os percentis ajustados resultantes.

Além desses critérios principais, também foram considerados critérios secundários que complementam o processo de seleção de métodos. A compatibilidade com ferramentas de diagnóstico é crucial para avaliar a qualidade do ajuste dos métodos e garantir que os resultados sejam confiáveis. A eficiência computacional também desempenha um papel importante, especialmente ao comparar métodos igualmente qualificados, garantindo que o processo de construção das curvas seja eficaz em termos de recursos computacionais. Por último, a flexibilidade de aplicação é um critério que permite que o método seja aplicado a diversas medições, garantindo que as curvas de crescimento sejam abrangentes e aplicáveis a uma variedade de contextos.

Em resumo, a seleção dos métodos para construção das curvas de crescimento neste estudo é guiada por critérios rigorosos, abordando tanto os desafios fundamentais de precisão e integridade das curvas como critérios secundários que consideram a qualidade, eficiência e aplicabilidade dos métodos escolhidos. Isso assegura a utilidade das curvas de crescimento resultantes para aplicações futuras.

4.2.2 Métodos

Para a seleção dos métodos mais adequados para a construção das curvas de crescimento, foi realizada uma revisão abrangente da literatura sobre os métodos utilizados para este fim. Os detalhes dessa revisão podem ser encontrados na Secção 2.2, na qual os métodos são listados na Tabela 1. A avaliação dos métodos foi meticulosa e baseada nos critérios delineados na Subsecção 4.2.1. Nessa análise, priorizou-se a inclusão apenas de métodos que considerassem a idade como uma variável contínua, bem como aqueles que se baseassem em suposições de distribuição explícitas e conseguissem estimar percentis simultaneamente.

A lista final de métodos considerados para a construção das curvas de crescimento inclui:

- **Método GAMLSS:** Este método tem uma estrutura abrangente para ajustar modelos de regressão, nos quais a distribuição da variável resposta pode ser contínua ou discreta e apresentar alta assimetria e curtose [Stasinopoulos and Rigby, 2007]. O método **GAMLSS** oferece a flexibilidade necessária para incorporar distribuições complexas e capturar a variação em diferentes momentos da distribuição.
- **Método LMS:** Este método resume a distribuição por meio de três curvas, que descrevem a mediana, o **CV**, e a assimetria. A assimetria é expressa por meio de uma transformação de potência *Box-Cox* [Cole and Green, 1992]. Através do uso da verosimilhança penalizada, essas três curvas são ajustadas por meio de uma regressão não linear. O grau de alisamento necessário pode ser quantificado em termos de parâmetros de alisamento ou graus de liberdade equivalentes.
- **Método LMST:** Este método é uma extensão do método **LMS**, mas utilizando a distribuição **BCT** [Rigby and Stasinopoulos, 2004a] .
- **Método LMSP:** Este método é outra extensão do método **LMS**, mas utilizando a distribuição **BCPE** [Rigby and Stasinopoulos, 2004b].

Esta seleção criteriosa de métodos servirá como base para a construção das curvas de crescimento no contexto deste estudo.

4.3 GAMLSS - Modelos Aditivos Generalizados para a Localização, Escala e Forma

A modelação estatística percorreu um longo caminho desde a criação do Modelo Linear Generalizado (**GLM**) por [Nelder and Wedderburn \[1972\]](#). Embora o **GLM** tenha fornecido um enquadramento fundamental, as suas limitações tornaram-se evidentes ao longo do tempo. Os Modelos Generalizados Aditivos (**GAM**), introduzidos por [Hastie and Tibshirani \[1986\]](#), representaram uma melhoria substancial ao introduzir funções de alisamento de variáveis explicativas no processo de modelação. No entanto, foram os **GAMLSS** que marcaram um avanço significativo, abordando as restrições do **GLM** e do **GAM**.

Introduzido por [Rigby and Stasinopoulos \[2005\]](#), juntamente com contribuições de [Akantziotiou et al. \[2008\]](#), os **GAMLSS** são modelos de regressão semiparamétrica. Eles exibem características paramétricas, pois necessitam de uma suposição de distribuição paramétrica para a variável resposta. Porém, são considerados semiparamétricos porque a modelação dos parâmetros de distribuição, em relação às variáveis explicativas, pode incorporar funções de alisamento não paramétricas [[Stasinopoulos and Rigby, 2007](#)].

Este método oferece uma abordagem altamente versátil e adaptável para modelar uma variável resposta (univariada) em função de variáveis explicativas. Permite a utilização de uma ampla gama de distribuições, incluindo distribuições contínuas e discretas com grande assimetria e curtose [[Voudouris et al., 2012](#)]. O que distingue os **GAMLSS** é a sua capacidade de modelar não apenas a média (ou localização), mas também outros parâmetros críticos da distribuição, como escala, assimetria e curtose [[Stasinopoulos et al. \[2008\]](#)]. Estes parâmetros podem ser expressos como funções lineares paramétricas ou funções não paramétricas aditivas de variáveis explicativas e efeitos aleatórios [[Stasinopoulos et al., 2008](#)]. Esta flexibilidade torna os **GAMLSS** particularmente adequados para cenários de modelação onde a distribuição da variável resposta se afasta da suposição da família exponencial. Isto inclui dados de resposta com leptocurtose, platicurtose, e/ou assimetria positiva ou negativa, ou dados com heterogeneidade (por exemplo, onde a escala ou a forma da distribuição da variável resposta muda com a variável explicativa) [[Stasinopoulos and Rigby, 2007](#)].

4.3.1 Modelo estatístico

Os **GAMLSS** são uma estrutura versátil de modelação estatística usada para analisar dados com diferentes formas de distribuição. Nesta estrutura, assumem-se observações independentes, designadas por

y_i para $i = 1, 2, \dots, n$, com função de probabilidade (densidade) $f(y_i|\boldsymbol{\theta}^i)$, condicionada a um vetor de quatro parâmetros de distribuição, $\boldsymbol{\theta}^i = (\theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{4i}) = (\mu_i, \sigma_i, \nu_i, \tau_i)$. Estes parâmetros de distribuição podem, por sua vez, ser funções de variáveis explicativas [Stasinopoulos and Rigby, 2007].

Os elementos-chave do modelo **GAMLSS** são os seguintes:

1. **Parâmetros de Distribuição:** Estes parâmetros, μ_i , σ_i , ν_i , e τ_i , são os componentes essenciais do modelo. μ_i e σ_i são geralmente referidos como parâmetros de localização e escala, enquanto ν_i e τ_i são considerados parâmetros de forma, por exemplo, assimetria e curtose [Stasinopoulos and Rigby, 2007]. No entanto, o **GAMLSS** é flexível e pode ser aplicado a vários parâmetros de distribuição populacional.
2. **Funções de Ligação:** Os parâmetros de distribuição estão ligados às variáveis explicativas por funções de ligação monótonas conhecidas, designadas por $g_k(\cdot)$, para $k = 1, 2, 3, 4$. Estas funções relacionam os parâmetros às variáveis explicativas mediante combinações lineares.
3. **Formulação original do modelo:** Seja $\mathbf{y}^T = (y_1, y_2, \dots, y_n)$ o vetor de comprimento n da variável de resposta, e $g_k(\cdot)$ as funções de ligação monótonas conhecidas, para $k = 1, 2, 3, 4$. Estas funções relacionam os parâmetros de distribuição às variáveis explicativas da seguinte forma:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \gamma_{jk}, \quad (4.1)$$

Ou seja,

$$\begin{aligned} g_1(\boldsymbol{\mu}) &= \boldsymbol{\eta}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1} \gamma_{j1} \\ g_2(\boldsymbol{\sigma}) &= \boldsymbol{\eta}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2} \gamma_{j2} \\ g_3(\boldsymbol{\nu}) &= \boldsymbol{\eta}_3 = \mathbf{X}_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3} \gamma_{j3} \\ g_4(\boldsymbol{\tau}) &= \boldsymbol{\eta}_4 = \mathbf{X}_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4} \gamma_{j4} \end{aligned}$$

Aqui, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\nu}$, $\boldsymbol{\tau}$ e $\boldsymbol{\eta}_k$ são vetores de comprimento n , $\boldsymbol{\beta}_k^T$ é um vetor de parâmetros de comprimento J'_k , \mathbf{X}_k é uma matriz de delineamento fixa conhecida de ordem $n \times J'_k$, \mathbf{Z}_{jk} é uma matriz de

delineamento fixa conhecida de ordem $n \times q_{jk}$, e γ_{jk} são variáveis aleatórias que se assume serem distribuídas como $\gamma_{jk} \sim N_{q_{jk}}(\mathbf{0}, \mathbf{G}_{jk}^{-1})$, onde \mathbf{G}_{jk}^{-1} é a inversa (generalizada) de uma $q_{jk} \times q_{jk}$ matriz simétrica $\mathbf{G}_{jk} = \mathbf{G}_{jk}(\boldsymbol{\lambda}_{jk})$ [Stasinopoulos and Rigby, 2007].

O modelo descrito na Equação 4.1 oferece flexibilidade para representar cada parâmetro de distribuição como uma função linear de variáveis explicativas ou como uma combinação linear envolvendo variáveis estocásticas (efeitos aleatórios) [Stasinopoulos and Rigby, 2007]. É importante observar que é incomum que todos os parâmetros de distribuição exijam modelação usando variáveis explicativas.

4. **Formulação Semiparamétrica Aditiva:** O modelo pode também ser expresso numa formulação semiparamétrica aditiva, definindo $\mathbf{Z}_{jk} = \mathbf{I}_n$ (a matriz identidade) e $\gamma_{jk} = \mathbf{h}_{jk} = h_{jk}(x_{jk})$ para todas as combinações de j e k na Equação 4.1. Resultando na formulação simplificada:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(x_{jk}) \quad (4.2)$$

5. **Formulação Simples Paramétrica:** Se não houver termos aditivos em nenhum dos parâmetros de distribuição, na Equação 4.2 temos o modelo **GAMLSS** linear paramétrico simples:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k \quad (4.3)$$

6. **Formulação Semiparamétrica Não Linear Aditiva:** O modelo pode ser ainda estendido para incluir termos paramétricos não lineares para μ , σ , ν , e τ da seguinte forma:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = h_k(\mathbf{X}_k, \boldsymbol{\beta}_k) + \sum_{j=1}^{J_k} h_{jk}(x_{jk}) \quad (4.4)$$

Aqui, h_k representa funções não lineares, e \mathbf{X}_k é uma matriz de delineamento conhecida de ordem $n \times J''k$.

7. **Formulação Paramétrica Não Linear:** Se para $k = 1, 2, 3, 4$, $J_k = 0$, ou seja, se para todos os parâmetros de distribuição não tivermos termos aditivos, então o modelo 4.4 é reduzido a um modelo **GAMLSS** paramétrico não linear:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = h_k(\mathbf{X}_k, \boldsymbol{\beta}_k) \quad (4.5)$$

8. **Estimação do modelo:** Os vetores paramétricos β_k e os parâmetros de efeitos aleatórios γ_{jk} , para $j = 1, 2, \dots, J_k$ e $k = 1, 2, 3, 4$ são estimados dentro da estrutura **GAMLSS** (para valores fixos dos hiperparâmetros de alisamento λ_{jk}) maximizando uma função de verosimilhança penalizada ℓ_p dada por [Stasinopoulos et al., 2017]:

$$\ell_p = \ell - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{J_k} \lambda_{jk} \gamma'_{jk} \mathbf{G}_{jk} \gamma_{jk} \quad (4.6)$$

Onde $\ell = \sum_{i=1}^n \log f(y_i | \theta^i)$ é a função log de verosimilhança.

Para os modelos paramétricos **GAMLSS** 4.3 ou 4.5, ℓ_p reduz-se a ℓ , e β_k para $k = 1, 2, 3, 4$ são estimados maximizando a função de verosimilhança ℓ [Stasinopoulos et al., 2017].

4.3.2 Algoritmos

A chave para uma modelação eficaz no **GAMLSS** está na maximização da função de verosimilhança penalizada, conforme descrito na Equação 4.6. Existem dois algoritmos fundamentais para esse propósito, o algoritmo Cole e Green (**CG**) e o algoritmo Rigby e Stasinopoulos (**RS**), cada um adaptado para lidar com diferentes características da função de verosimilhança. As três partes dos algoritmos, iteração externa, interna e algoritmo de retroajustamento, estão esboçadas na Figura 4 para auxiliar na visualização do processo dos algoritmos.

O **CG**, é uma generalização do algoritmo de Cole and Green [1992]. É um algoritmo de pontuação implementado localmente e executado dentro de uma iteração externa, uma iteração interna e um algoritmo de retroajustamento modificado para ajustar cada parâmetro de distribuição. Ao contrário do **RS**, o **CG** depende das derivadas cruzadas (esperadas ou aproximadas) do logaritmo da verosimilhança relativamente a cada par de parâmetros da distribuição. Rigby and Stasinopoulos [2005] demonstraram que o algoritmo **CG**, maximiza o logaritmo da verosimilhança penalizada mantendo os λ fixos no contexto dos parâmetros β e γ .

É importante notar que o **CG** é mais adequado para casos em que os parâmetros de distribuição $\theta = (\mu, \sigma, \nu, \tau)$ não são ortogonais em termos de informação. Ou seja, este algoritmo destaca-se apenas ao lidar com famílias de distribuição onde as derivadas cruzadas da função de verosimilhança não são zero.

Em contraste, o **RS** é uma abordagem mais simplificada, derivada do algoritmo usado no ajuste de modelos **MADAM** por Rigby and Stasinopoulos [1996a,b]. Ao contrário do **CG**, o **RS** não depende de derivadas cruzadas da função de verosimilhança. Isso torna-o particularmente adequado para casos em

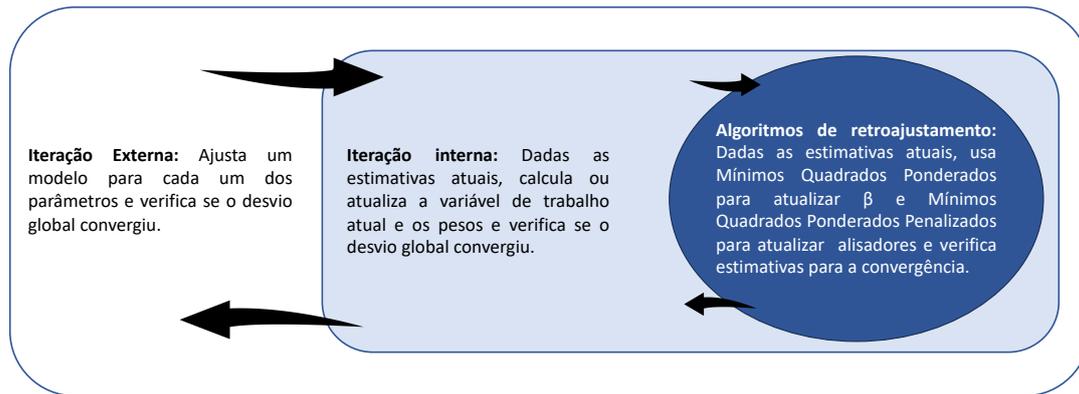


Figura 4: Representação da estrutura dos algoritmos **GAMLSS CG** e **RS** em três partes. Adaptado de [Fernandes et al. \[2021\]](#).

que os parâmetros θ exibem ortogonalidade de informação, como em modelos de localização e escala ou modelos de família de dispersão.

Embora o **RS** tenha sido eficazmente utilizado para ajustar uma ampla gama de famílias de distribuição (mesmo quando as derivadas cruzadas são zero), ocasionalmente pode apresentar uma convergência mais lenta. É importante notar que o **RS** não é um caso especial do **CG**.

O principal objetivo de ambos os algoritmos é maximizar a função de verossimilhança penalizada l_p enquanto mantêm os hiperparâmetros λ fixos. Para modelos totalmente paramétricos, conforme descrito nas Equações 4.3 e 4.5, os algoritmos concentram-se em maximizar a função de verossimilhança l .

Os algoritmos **CG** e **RS** oferecem várias vantagens:

1. **Ajuste modular:** Eles permitem um procedimento de ajuste modular, facilitando diagnósticos de modelo distintos para cada parâmetro de distribuição.
2. **Flexibilidade para expansão do modelo:** Distribuições adicionais e termos aditivos podem ser facilmente adicionados aos modelos.
3. **Simplicidade na inicialização do modelo:** Frequentemente, valores iniciais simples (por exemplo, constantes) para os parâmetros θ são suficientes, tornando os algoritmos relativamente estáveis e rápidos.

Embora haja o potencial para múltiplos máximos locais na função de verossimilhança (penalizada), esse problema geralmente é resolvido explorando diferentes valores iniciais. Na prática, não tem representado desafios significativos, principalmente devido ao uso de tamanhos de amostra amplos nos dados analisados.

No decorrer deste estudo, foi utilizado exclusivamente o **RS** para o ajuste de modelos **GAMLSS**, devido à sua eficácia no tratamento de uma ampla gama de famílias de distribuição e a sua eficiência computacional.

4.3.3 Estimação dos hiperparâmetros λ

Na análise de modelos **GAMLSS**, a seleção e ajuste dos termos de alisamento desempenham um papel crucial. Existem duas abordagens principais para esse ajuste: uma interna, também conhecida como método local, e outra externa, referida como método global.

O método interno, conforme argumentado por [Rigby and Stasinopoulos \[2014\]](#), é notavelmente mais rápido e produz frequentemente resultados comparáveis ao método externo. Nesse método, os hiperparâmetros λ são ajustados sequencialmente em ciclos no algoritmo **GAMLSS**. Essa abordagem pode ser vista como uma generalização do procedimento de [Lee et al. \[2018\]](#) e baseia-se na verosimilhança quase penalizada, sendo aplicada localmente a cada parâmetro.

Por outro lado, o método externo envolve estratégias para estimar os parâmetros de alisamento. Atualmente, três estratégias amplamente utilizadas são a **MV**, o critério de informação Akaike generalizado (**GAIC**) e a validação cruzada generalizada (**VCG**). Neste estudo foi utilizada a **MV**. Detalhes sobre esses procedimentos de estimação podem ser encontrados na secção 3.4 do trabalho de [Stasinopoulos et al. \[2017\]](#).

4.3.4 Termos aditivos

A equação 4.1 permite modelar todos os parâmetros de distribuição μ, σ, ν, τ como funções que podem ser tanto paramétricas lineares como não lineares, bem como funções não paramétricas (alisamento) das variáveis explicativas e/ou termos de efeitos aleatórios [[Stasinopoulos and Rigby, 2007](#)]. Esta abordagem permite que as variáveis explicativas tenham um impacto direto sobre os preditores dos parâmetros específicos e, por conseguinte, sobre os próprios parâmetros. Portanto, embora seja comum na modelação de regressão que as variáveis explicativas afetem a média da variável resposta, nos modelos **GAMLSS** a localização, a escala e forma distribuição da resposta podem ser afetadas pelas variáveis explicativas, que podem ser referidas como termos do modelo [[Stasinopoulos et al., 2017](#)]. Estes termos podem assumir várias formas, e as suas relações com os preditores η , podem exibir padrões lineares e não lineares. É importante destacar que as relações não lineares podem ser paramétricas, ou não paramétricas, representadas por alisadores, que permitem que os dados determinem a relação entre o preditor e as variáveis explicativas.

Além disso, o **GAMLSS** também integra o conceito de termos aditivos. Ou seja, ao avaliar a influência das variáveis explicativas no preditor para um determinado parâmetro de distribuição, somamos os seus efeitos individuais [Stasinopoulos et al., 2017]. Essa aditividade permite-nos decompor a relação complexa entre os parâmetros e as variáveis explicativas em componentes mais compreensíveis. Importante salientar que o conceito de aditividade não exclui a possibilidade de interações entre esses termos.

Na Figura 5 é possível visualizar uma classificação dos diferentes termos aditivos nos **GAMLSS**. Os termos aditivos podem ser paramétricos (esquerda do diagrama), de alisamento (centro) ou efeitos aleatórios (direita). Os termos aditivos paramétricos incluem tanto termos paramétricos lineares quanto não lineares. Os termos aditivos de alisamento são divididos em duas categorias distintas: termos aditivos penalizados e outros termos de alisamento. Os efeitos aleatórios estão conectados aos termos de alisamento penalizados, indicando que podem ser tratados como alisadores penalizados, mas também podem fornecer informações adicionais sobre os alisadores penalizados quando considerados como efeitos aleatórios.

4.3.5 Alisamento

Os alisadores são ferramentas essenciais para modelar as relações entre as variáveis explicativas e a variável preditora. Originalmente, os alisadores eram utilizados para estimar o valor esperado condicional (ou média) de y dado x sem assumir uma forma funcional paramétrica, permitindo que os dados revelassem a relação funcional [Stasinopoulos et al., 2017]. Com o tempo, esse conceito foi estendido a qualquer parâmetro de localização (por exemplo, mediana) da distribuição de y e, mais geralmente, começaram a ser utilizados para a estimativa de quantis.

Os alisadores podem ser categorizados em dois grupos, univariados e multivariados, dependendo se uma ou várias variáveis explicativas estão envolvidas [Stasinopoulos et al., 2017]. Nos casos em que apenas é utilizada uma variável explicativa, x , é empregue um alisador univariado denotado como $s(x)$. Estes alisadores capturam os principais efeitos não lineares das variáveis explicativas individuais nos parâmetros de distribuição. Ao lidar com várias variáveis explicativas, como x_1 e x_2 , entramos no domínio dos alisadores multivariados que capturam os efeitos de interações não lineares entre as variáveis.

Os alisadores podem ser ainda divididos em dois tipos principais com base nos seus princípios subjacentes e estruturas de penalização:

- **Alisadores Penalizados:** Usam penalizações quadráticas nos parâmetros do modelo para regu-

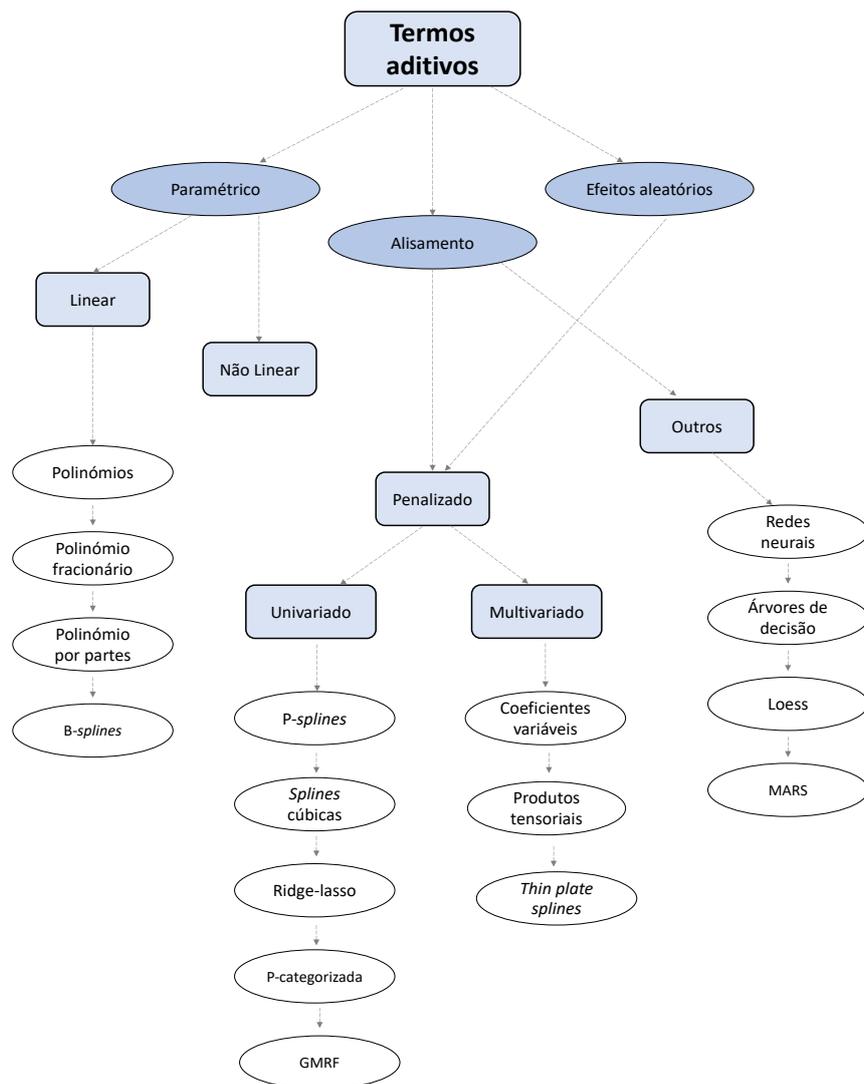


Figura 5: Diagrama ilustrando os diversos termos aditivos que podem ser ajustados no modelo **GAMLSS**. Adaptado de [Stasinopoulos et al. \[2017\]](#).

lar o grau de alisamento aplicado.

- **Alisadores Não Penalizados:** Em contraste, estes alisadores adotam estratégias diferentes, como penalizações locais ou não quadráticas, para moldar as funções de alisamento.

Um alisador univariado simples é uma generalização do modelo de regressão linear simples e pode ser expresso como um modelo estatístico [[Stasinopoulos et al., 2017](#)]:

$$E(Y_i) = \alpha + s(x_i), \quad \text{para } i = 1, \dots, n \quad (4.7)$$

Aqui, $s(x_i)$ representa uma função arbitrária que assume-se existir, enquanto α é uma constante que frequentemente pode ser absorvida na função $s(x_i)$. Essencialmente, $s(x_i)$ captura a tendência

subjacente que se pretende estimar a partir dos dados. Embora a escolha de $s(x_i)$ seja arbitrária, alisadores específicos, como *Splines* cúbicas, assumem que $s(x_i)$ possui derivadas contínuas de primeira e segunda ordem, tornando-os úteis para modelar tendências complexas.

A adequação dos alisadores pode ser abordada de diferentes maneiras, desde estimadores locais que se concentram em pontos de dados próximos até à incorporação de penalizações. Embora sejam frequentemente referidos como não paramétricos, é importante reconhecer que os alisadores de facto estimam parâmetros e apresentam um parâmetro dominante que dita o grau de alisamento dos dados - o parâmetro de alisamento.

Dentro do contexto dos modelos **GAMLSS**, foi optado por utilizar nesta investigação métodos univariados de alisamento penalizado, como *splines* penalizados (*P-splines*) e *Splines* cúbicas. Adicionalmente, foram testadas outras técnicas de alisamento, como redes neurais e *Loess*.

P-splines

As *splines* penalizadas, conhecidas como *P-splines*, são uma técnica estatística introduzida por Eilers and Marx [1996]. São utilizadas para modelar relações complexas entre variáveis, especialmente em modelos estatísticos como o **GAMLSS**.

A ideia-chave por trás das *P-splines* é garantir o alisamento na modelação, penalizando os coeficientes das funções de base *B-spline* [Stasinopoulos and Rigby, 2007]. Num contexto **GAMLSS**, onde θ representa qualquer vetor de parâmetros de distribuição, o modelo assume a forma $\theta = \mathbf{Z}(\mathbf{x})\gamma$ [Stasinopoulos and Rigby, 2007]. Aqui, $\mathbf{Z}(\mathbf{x})$ é uma matriz de delineamento de base de dimensões $n \times q$ associada à variável explicativa \mathbf{x} , geralmente definida em q nós diferentes, principalmente no intervalo de \mathbf{x} . O vetor γ , com dimensões $q \times 1$, contém coeficientes sujeitos a restrições estocásticas. Essas restrições são impostas para garantir o nível desejado de alisamento. Em particular, as restrições em γ seguem uma distribuição, como $\mathbf{D}\gamma \sim N(\mathbf{0}, \lambda^{-1}\mathbf{I})$, ou equivalente, $\gamma \sim N(\mathbf{0}, \lambda^{-1}\mathbf{K}^{-1})$, onde $\mathbf{K} = \mathbf{D}^T\mathbf{D}$. A matriz \mathbf{D} tem dimensões $(q - r) \times q$ e calcula as diferenças de ordem r do vetor γ de q dimensões.

Portanto, para definir uma *P-spline*, precisamos de considerar vários parâmetros, incluindo o número e posicionamento de nós (q) para determinar quantos nós são colocados ao longo do eixo x . O grau do polinómio por partes usado nas funções de base *B-spline*, a ordem das diferenças, que dita o tipo de penalização aplicada aos coeficientes das funções de base *B-spline*. E por fim, a quantidade desejada de alisamento, que pode ser expressa em termos de graus de liberdade equivalentes (df) ou o parâmetro de alisamento (λ).

Splines cúbicas

As *splines* cúbicas são uma técnica amplamente abordada na literatura [REINSCH, 1967, Green and Silverman, 1993, Hastie and Tibshirani, 1986]. Utilizam funções de base semelhantes às das *P-splines*, no entanto, usam polinômios por partes de grau 3. Embora estes dois alisadores sejam semelhantes, diferem em dois aspectos-chave, nas *P-splines*, os nós são equidistantes ao longo do eixo x , e o alisamento é controlado aplicando uma penalização aos parâmetros γ [Stasinopoulos et al., 2017]. Por outro lado, nas *splines* cúbicas os nós estão em valores distintos da variável x e o alisamento é obtido penalizando a segunda derivada da função [Stasinopoulos et al., 2017].

No contexto das *splines* cúbicas dentro do **GAMLSS**, é assumido no modelo 4.2 que as funções $h(t)$ são duas vezes continuamente diferenciáveis, sendo maximizada uma verosimilhança penalizada, representada por l , sujeita a termos de penalização que assumem a forma $\lambda \int_{-\infty}^{\infty} [h''(t)]^2 dt$ [Stasinopoulos and Rigby, 2007]. A solução para as funções de maximização $h(t)$ são todas *splines* cúbicas naturais, e, portanto, podem ser expressas como combinações lineares das suas funções naturais de base *spline* cúbica.

Redes neurais

As redes neurais, conhecidas pela sua notável flexibilidade, representam uma ferramenta para ajustar modelos de regressão não linear. Estes modelos, como descritos nos trabalhos de Bishop [1995], Jensen and Kendall [1993], Ripley [2007], têm uma sobreparametrização que os dota da capacidade de aproximar praticamente qualquer função contínua. No entanto, têm um grau de complexidade e desafios em termos de interpretação, diferenciando-os dos modelos de alisamento tradicionais. Outra característica peculiar desta técnica é o ajuste de modelos finais distintos quando valores iniciais diferentes são escolhidos [Stasinopoulos et al., 2017].

Para garantir a reprodutibilidade deste método, é prudente definir uma *seed* de geração aleatória no início do ajuste, permitindo assim a recriação do mesmo modelo ajustado. Além disso, para otimizar o processo de modelação e mitigar potencialmente o sobreajuste, pode-se utilizar o argumento “*decay*” — um parâmetro de alisamento. O argumento “*size*”, que especifica o número de variáveis ocultas, também desempenha um papel fundamental na modelação do comportamento da rede neural.

Em essência, as redes neurais podem ser comparadas a modelos enigmáticos de “caixa-preta”, proficientes em aplicações práticas, mas muitas vezes carentes de transparência e facilidade de interpretação [Stasinopoulos et al., 2017]. Esta opacidade decorre da sua flexibilidade, que permite capturar interações complexas de alto nível entre variáveis explicativas.

Loess

O alisador de dispersão ponderada localmente, comumente referido como *Loess*, é uma técnica utilizada para alcançar o alisamento de regressão local. Introduzido por [Cleveland and Devlin \[1988\]](#), o *Loess* é uma ferramenta valiosa para modelar curvas ou superfícies alisadas quando se lida com variáveis dependentes e uma ou mais variáveis explicativas contínuas.

Para uma única variável explicativa, o *Loess* utiliza um algoritmo padrão caracterizado pelos seguintes passos-chave [[Stasinopoulos et al., 2017](#)]:

1. Definição da largura ou extensão da janela para x , para determinar a extensão na qual as observações próximas a um determinado valor de x influenciam o processo de alisamento.
2. Os valores x são ordenados, e o foco muda para cada valor x individualmente.
3. Atribuição de pesos às observações utilizando uma função de núcleo para atribuir maior importância às observações com valores x mais próximos do valor x de interesse.
4. Ajuste de uma regressão polinomial usando mínimos quadrados ponderados, obtendo assim os valores ajustados da variável dependente no valor x de interesse.
5. Os passos 3 e 4 são repetidos para cada valor x no conjunto de dados. A culminação desse processo envolve traçar uma curva alisada conectando os valores ajustados da variável dependente em relação a x , revelando assim as relações locais nos dados.

4.3.6 Distribuições

Na estrutura do **GAMLSS**, a forma de distribuição assumida para a variável de resposta y , caracterizada por $f(y_i|\mu_i, \sigma_i, \nu_i, \tau_i)$, pode abranger uma ampla e versátil gama. Esta função de densidade de probabilidade é intencionalmente deixada não especificada, sem uma distribuição explícita fornecida. Embora geralmente englobe parâmetros associados à localização, escala, assimetria e curtose, essa generalização não é universal.

No contexto da implementação R do **GAMLSS**, a única estipulação é que a função $\log f(y_i|\mu_i, \sigma_i, \nu_i, \tau_i)$ e as suas primeiras derivadas (e, opcionalmente, segundas derivadas e derivadas cruzadas esperadas) relativamente a cada um dos parâmetros dentro de θ devem ser computacionalmente alcançáveis. Embora derivadas explícitas sejam preferidas, a utilização de derivadas numéricas permanece uma alternativa viável. Essencialmente, o algoritmo de ajuste utilizado para modelos **GAMLSS** depende apenas dessa informação, garantindo versatilidade e adaptabilidade na modelação de várias formas de distribuição.

Na tabela do Apêndice A estão representadas uma variedade de famílias de um, dois, três e quatro parâmetros de distribuições contínuas implementadas no pacote `gam1ss`. Durante este estudo foi optado por utilizar apenas as distribuições *Box-Cox Cole-Green (BCCG)*, *Box-Cox t (BCT)*, *Box-Cox Power Exponential (BCPE)* e *Johnson's SU (JSU)*. As funções de ligação para cada uma das distribuições foram as predefinidas. A decisão para a escolha destas distribuições baseou-se na revisão da literatura que pode ser consultada na Secção 2.2.

Distribuição BCCG

A distribuição **BCCG** é apropriada para lidar com dados com assimetria positiva ou negativa [Stasinopoulos et al., 2008]. Cole and Green [1992] foram uns dos primeiros a adotar esta distribuição, usando-a para modelar os três parâmetros por meio de funções de alisamento não paramétricas ligadas a uma única variável explicativa [Stasinopoulos et al., 2017]. Assumindo que Y é uma variável aleatória positiva seguindo a distribuição **BCCG** (μ, σ, ν), define-se a variável aleatória transformada Z usando a seguinte equação [Rigby et al., 2019]:

$$z = \begin{cases} \frac{1}{\sigma\nu} \left[\left(\frac{Y}{\mu} \right)^\nu - 1 \right] & \text{se } \nu \neq 0 \\ \frac{1}{\sigma} \log \left(\frac{Y}{\mu} \right) & \text{se } \nu = 0 \end{cases} \quad (4.8)$$

Aqui, $0 < Y < \infty$, $\mu > 0$, $\sigma > 0$, e $-\infty < \nu < \infty$. Presume-se que a variável aleatória Z esteja sujeita a uma distribuição normal padrão truncada [Rigby et al., 2019]. A condição $0 < Y < \infty$ garante que Y^ν seja real para todos os ν , levando ao intervalo $-1/(\sigma\nu) < Z < \infty$ se $\nu > 0$ e $-\infty < Z < -1/(\sigma\nu)$ se $\nu < 0$. Essa exigência torna necessária a distribuição normal padrão truncada para Z [Rigby et al., 2019]. Consequentemente, a função de densidade de probabilidade de Y é dada por [Stasinopoulos et al., 2008].

$$f_Y(y|\mu, \sigma, \nu) = \frac{y^{\nu-1} \exp\left(-\frac{1}{2}z^2\right)}{\mu^\nu \sigma \sqrt{2\pi} \phi\left(\frac{1}{\sigma|\nu|}\right)} \quad (4.9)$$

A variável Z é determinada pela Equação 4.8, e $\phi(\cdot)$ representa a função de distribuição acumulada de uma distribuição normal padrão [Stasinopoulos et al., 2008].

As funções de ligação dos parâmetros μ , σ e ν são log, log e identidade, respetivamente.

Distribuição BCT

Considerando Y como uma variável aleatória positiva seguindo a distribuição **BCT** (μ, σ, ν, τ) introduzida por [Rigby and Stasinopoulos \[2006\]](#). A distribuição é caracterizada pela variável aleatória transformada Z especificada na Equação 4.8, onde Z segue uma distribuição t truncada com graus de liberdade $\tau > 0$, tratada como um parâmetro contínuo [[Stasinopoulos et al., 2008](#)]. A seguinte equação fornece a função de densidade de probabilidade de Y .

$$f_Y(y|\mu, \sigma, \nu, \tau) = \frac{y^{\nu-1} f_T(Z)}{\mu^\nu \sigma F_T\left(\frac{1}{\sigma|\nu|}\right)} \quad (4.10)$$

Para valores positivos de Y , com μ e σ positivos e $-\infty < \nu < \infty$, a variável aleatória Z é determinada pela Equação 4.8. $f_T(t)$ e $F_T(t)$ são respectivamente a função de densidade de probabilidade e a função de distribuição acumulada de uma variável aleatória T caracterizada por uma distribuição t padrão com um parâmetro de graus de liberdade $\tau > 0$ [[Stasinopoulos et al., 2008](#)].

As funções de ligação dos parâmetros μ, σ, ν e τ são log, log, identidade e log, respectivamente.

Distribuição BCPE

Considerando uma variável aleatória positiva Y com uma distribuição **BCPE** (μ, σ, ν, τ), introduzida por [Rigby and Stasinopoulos \[2004b\]](#). Esta distribuição é descrita pela variável aleatória transformada Z , dada pela Equação 4.8, com Z assumindo uma distribuição exponencial de potência padrão truncada com um parâmetro de potência $\tau > 0$, considerado como um parâmetro contínuo [[Stasinopoulos et al., 2008](#)].

A função de densidade de probabilidade de Y , denominada pela Equação 4.10, está relacionada com $f_T(t)$ e $F_T(t)$, a função de densidade de probabilidade e a função de distribuição acumulada de uma variável T seguindo uma distribuição exponencial de potência padrão [[Stasinopoulos et al., 2008](#)].

As funções de ligação dos parâmetros μ, σ, ν e τ são log, log, identidade e log, respectivamente.

Distribuição JSU

A distribuição Johnson SU, introduzida por Johnson em 1949, oferece uma distribuição de probabilidade flexível adequada para várias aplicações. Na sua parametrização original, a distribuição Johnson SU é caracterizada por dois parâmetros-chave: ν e τ [[Stasinopoulos et al., 2008](#)].

O parâmetro ν desempenha um papel crucial na determinação da assimetria da distribuição. Quando $\nu > 0$, indica assimetria negativa, enquanto $\nu < 0$ representa assimetria positiva. Já o parâmetro τ

controla a curtose da distribuição. Tipicamente, τ deve assumir valores positivos, geralmente acima de 1. À medida que $\tau \rightarrow \infty$, a distribuição Johnson SU converge para a função de densidade normal padrão. Essa distribuição é particularmente adequada para lidar com dados com leptocurtose, que exibem caudas pesadas e excesso de curtose [Stasinopoulos et al., 2008].

A função de densidade de probabilidade da distribuição original de Johnson Su é denotada como **JSUo**(μ, σ, ν, τ).

$$f_Y(y | \mu, \sigma, \nu, \tau) = \frac{\tau}{\sigma} \frac{1}{(r^2 + 1)^{\frac{1}{2}}} \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} z^2 \right], \quad (4.11)$$

Para, $-\infty < y < \infty$, onde $-\infty < \mu < \infty, \sigma > 0, -\infty < \nu < \infty$, e $\tau > 0$, e onde:

$$z = \nu + \tau \sinh^{-1}(r) = \nu + \tau \log \left[r + (r^2 + 1)^{\frac{1}{2}} \right], \quad (4.12)$$

Onde $r = (y - \mu)/\sigma$, e $Z \sim \mathbf{NO}(0, 1)$. Além disso, $E(Y) = \mu - \sigma\omega^{1/2} \sinh(\nu/\tau)$ e $\text{Var}(Y) = \sigma^2 \frac{1}{2}(\omega - 1)[\omega \cosh(2\nu/\tau) + 1]$, onde $\omega = \exp(1/\tau^2)$.

As funções de ligação dos parâmetros μ, σ, ν e τ são identidade, log, identidade e log, respetivamente.

4.4 LMS e as suas extensões LMST e LMSP

O método **LMS**, desenvolvido inicialmente por Cole [1988] e subsequentemente refinado por Cole and Green [1992], é uma ferramenta para construir curvas percentilicas, particularmente no contexto de modelar uma variável de resposta Y relativamente a uma única variável explicativa x (por exemplo, idade) [Stasinopoulos et al., 2017]. O método **LMS** é especialmente vantajoso porque acomoda a suposição de uma distribuição específica para a variável Y . Isso permite a estimativa simultânea de curvas percentilicas para vários percentis (p) sem cruzamentos. Além disso, os modelos **LMS** facilitam prontamente o cálculo de percentis, $y_p(x)$, dados valores específicos de p e x , ou os *z-scores*, dando y e x [Stasinopoulos et al., 2017].

Dentro do **GAMLSS**, o método **LMS** pode ser integrado presumindo que a variável de resposta segue a distribuição **BCCG**, adaptada a dados que apresentam assimetria positiva ou negativa com $Y > 0$ como representado na Equação 4.8.

Rigby and Stasinopoulos [2004b, 2006] expandiram o método **LMS**, que aborda inerentemente a assimetria, mas não a curtose nos dados, introduzindo duas extensões importantes: as distribuições **BCT** e **BCPE** para Y . Essas extensões deram origem aos métodos de estimação percentilica **LMST** e **LMSP**, respetivamente. Mais detalhes sobre estas distribuições podem ser consultados na Secção 4.3.6.

No contexto da estimação percentilica para Y dado x , o modelo **GAMLSS** é formulado da seguinte forma:

$$\begin{aligned}
 Y &\sim \mathcal{D}(\mu, \sigma, \nu, \tau) \\
 g_1(\mu) &= s_1(u) \\
 g_2(\sigma) &= s_2(u) \\
 g_3(\nu) &= s_3(u) \\
 g_4(\tau) &= s_4(u) \\
 u &= x^\xi
 \end{aligned} \tag{4.13}$$

Aqui, \mathcal{D} geralmente representa a distribuição **BCCG**, **BCT** ou **BCPE**, para a qual μ , σ , ν e τ são a mediana aproximada, coeficiente de variação aproximado, parâmetros de assimetria e curtose da distribuição, respetivamente [Stasinopoulos et al., 2017]. As funções $g(\cdot)$ correspondem a funções de ligação apropriadas, enquanto as funções $s(\cdot)$ indicam funções de alisamento não paramétricas. A transformação de potência para x , ou seja, x^ξ , é muitas vezes necessária quando a variável de resposta apresenta um crescimento rápido nas fases iniciais ou finais, ajudando a alinhar a escala x com o ajuste da curva.

4.5 Estimativa de percentis

A estimativa de percentis desempenha um papel fundamental na compreensão e modelação do crescimento, particularmente na avaliação de como diversos fatores afetam a distribuição de variáveis relacionadas ao crescimento ao longo do tempo como, por exemplo, o peso. Nesta investigação a estimação dos percentis é feita a partir do pacote `gamlss`.

No cerne da estimativa de percentis estão duas variáveis contínuas-chave:

- **Variável Resposta:** Esta é a variável em estudo, para a qual as curvas de percentis são procuradas. Pode ser qualquer medida como o peso, o diâmetro, ou a espessura.
- **Variável Explicativa:** Normalmente, esta é a variável independente associada à variável resposta relacionada ao crescimento, frequentemente a idade no contexto do desenvolvimento humano.

Um percentil específico $100p$ para uma variável contínua aleatória Y é definido como y_p , onde $\text{Prob}(Y \leq y_p) = p$. Em termos matemáticos, $y_p = F_Y^{-1}(p)$, representando a função de distribuição cumulativa inversa de Y em p . Considerando o percentil condicional de Y dada a variável explicativa $X = x$, ou seja, $y_p(x) = F_{Y|x}^{-1}(p)$. Variando x , uma curva de percentil $100p$ de $y_p(x)$ contra x é obtida. Diferentes percentis podem ser usados para construir diversas curvas de percentis. Neste estudo,

foram utilizados os seguintes valores para os percentis: $100p = (3, 10, 25, 50, 75, 90, 97)$.

É importante ter em consideração que um *z-score*, dado os valores de y e x , é definido por $z = \Phi^{-1} [F_{Y|x}(y)]$, onde Φ^{-1} é distribuição cumulativa inversa da distribuição normal padrão. Para os valores de y e x utilizados na estimativa de um modelo, os *z-score* são os resíduos de um modelo **GAMLSS** ajustado. Enfatizando a estreita conexão entre a estimativa de percentis e técnicas de modelação estatística.

Na prática, uma abordagem eficaz para gerar curvas de crescimento de percentis de referência para populações envolve uma combinação de metodologias.

4.6 Critérios de qualidade de ajuste

A avaliação da qualidade de ajuste dos modelos é uma etapa fundamental em qualquer análise estatística, desempenhando um papel crucial na validação e interpretação dos resultados obtidos, uma vez que a qualidade do ajuste irá influenciar diretamente a confiabilidade das estimativas e a utilidade das curvas de referência geradas. Esta visa responder a perguntas fundamentais, tais como: Quão bem o modelo se ajusta aos dados observados? Em que medida as previsões do modelo se aproximam dos valores reais? Os pressupostos do modelo são cumpridos?

Para realizar uma avaliação abrangente da qualidade de ajuste, existe uma variedade de critérios, métricas e ferramentas disponíveis na literatura. No contexto deste estudo, adotou-se uma abordagem que inclui uma série de critérios quantitativos e ferramentas gráficas. Entre os critérios e métricas considerados, destaca-se o Coeficiente de Determinação (**R²**), o Erro Quadrático Médio (**MSE**), o Critério de Informação de Akaike (**AIC**), o Critério de Informação Bayesiano (**BIC**), as estatísticas Z, o *worm plot* e os gráficos de resíduos (resíduos quantílicos em relação aos valores ajustados, resíduos quantílicos em relação ao índice, estimativa de densidade e gráfico Q-Q normal).

A combinação destes critérios quantitativos e ferramentas gráficas oferece uma abordagem completa para avaliar a qualidade do ajuste dos modelos de crescimento, garantindo a confiabilidade das curvas de referência resultantes.

4.6.1 Coeficiente de Determinação

O coeficiente de determinação, frequentemente referido como R-quadrado ou R^2 , desempenha um papel significativo na análise estatística e na modelação da regressão. Apesar da ausência de uma métrica universal para avaliar o desempenho de um modelo, o **R²** destaca-se como uma escolha bem estabelecida

e utilizada pela comunidade científica.

Introduzido por Wright [1921], o R^2 na sua formulação original é uma medida de quanto a variável dependente pode ser explicada pelas variáveis independentes, quantificando essa relação em termos de proporção de variância [Chicco et al., 2021]. Em outras palavras, quantifica o quão bem o modelo se ajusta aos dados, sendo que valores mais elevados indicam um melhor ajuste. O R^2 pode assumir valores no intervalo $[0, 1]$, onde um valor de 1 indica um ajuste perfeito, implicando que o modelo prevê com precisão a variável dependente. Por outro lado, um valor de 0 corresponde a um ajuste trivial, semelhante a uma linha horizontal que representa a média dos valores-alvo [Chicco et al., 2021].

Matematicamente, o R^2 pode ser expresso como:

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2} \quad (4.14)$$

Onde X_i representa o valor previsto para a i -ésima observação, Y_i representa o valor real para a i -ésima observação, \bar{Y} é a média dos valores reais e m é o número total de observações.

É importante observar que o comportamento do R^2 não depende apenas da linearidade do modelo de regressão. Um R^2 alto não garante uma relação perfeitamente linear, e vice-versa, um R^2 baixo não implica necessariamente uma relação não linear [Chicco et al., 2021]. Logo é de extrema importância examinar os dados de forma abrangente.

Em suma, o coeficiente de determinação, é uma ferramenta versátil e fácil de interpretar, permitindo avaliar o desempenho de um modelo.

4.6.2 Erro Quadrático Médio

O erro quadrático médio, também referido como **MSE**, é uma medida frequentemente utilizada para avaliar a precisão de modelos preditivos. A sua expressão matemática é definida como:

$$MSE = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2 \quad (4.15)$$

Onde X_i representa o valor previsto para a i -ésima observação, Y_i representa o valor real para a i -ésima observação e m é o número total de observações.

No contexto do **MSE**, o objetivo é minimizar esta métrica. Um **MSE** menor indica um melhor ajuste, uma vez que significa que as previsões do modelo estão mais próximas dos dados reais e apresentam um enviesamento e variância menores [Schluchter, 2005]. Pelo contrário, um **MSE** mais elevado sugere

que as previsões do modelo afastam-se mais dos valores reais.

Uma vantagem notável desta métrica é a sua capacidade de detetar valores discrepantes nos dados. O **MSE** atribui pesos maiores a esses valores, graças à norma L_2 e, por isso, se o modelo eventualmente produzir uma única previsão exceçãoalmente má, a parte quadrada da função amplifica significativamente o erro [Chicco et al. \[2021\]](#).

É importante notar que existe uma relação matemática entre o **MSE** e o R^2 . Especificamente, R^2 pode ser expresso como $R^2 = 1 - \frac{MSE}{MST}$, onde MST é a média da soma total dos quadrados. Uma vez que o **MST** é fixo para os dados utilizados, R^2 e **MSE** partilham uma relação monótona negativa. Consequentemente, ao classificar modelos de regressão com base na qualidade do ajuste, a ordem dos modelos de regressão baseados em R^2 será idêntica (embora em ordem inversa) a uma ordenação de modelos baseados em **MSE** [\[Chicco et al., 2021\]](#).

Em conclusão, o **MSE** desempenha um papel vital na avaliação do desempenho de modelos de regressão, quantificando o grau em que as previsões do modelo se desviam dos dados reais. Permite a deteção de valores discrepantes e a sua relação com outras métricas importantes tornam-no uma ferramenta valiosa para avaliar e comparar diferentes modelos de regressão.

4.6.3 Critério de Informação de Akaike

O critério de informação de Akaike (**AIC**), introduzido por [Akaike \[1973\]](#), representa uma ferramenta fundamental no âmbito da seleção de modelos e avaliação de parcimónia na estrutura do modelo, oferecendo um enquadramento matemático para avaliar o equilíbrio entre a complexidade do modelo e a qualidade do ajuste [\[Bonakdari and Zeynoddin, 2022\]](#).

A estimativa tradicional da **MV**, fornece uma abordagem para estimar os parâmetros de um modelo com dimensão e estrutura predefinidas [\[Cavanaugh and Neath, 2019\]](#). No entanto, na modelação estatística, muitas vezes existe incerteza sobre a dimensão e a estrutura adequada do modelo. A contribuição inovadora de Akaike estendeu este paradigma e combinou num único procedimento o processo de estimativa com determinação estrutural e dimensional [\[Cavanaugh and Neath, 2019\]](#).

O **AIC** é formulado matematicamente na Equação 4.16.

$$AIC = -2 \ln(\hat{L}) + 2k \quad (4.16)$$

O \hat{L} representa a **MV** e k o número de parâmetros do modelo. Em termos práticos, o modelo selecionado como o melhor ajuste é aquele associado ao valor mínimo do **AIC**.

É importante notar que o **AIC** inclui um termo de penalização ($2k$) para evitar a seleção de modelos

com um número excessivo de parâmetros. Apesar da sua eficácia, o **AIC**, quando aplicado a conjuntos de dados com tamanhos de amostra limitados, pode tender a sobrestimar o número de parâmetros, potencialmente levando à seleção de modelos mais complexos [Bonakdari and Zeynoddin, 2022].

Em resumo, o **AIC** representa uma ferramenta versátil para a seleção de modelos, encontrando um equilíbrio entre a complexidade do modelo e a qualidade do ajuste. A sua fundamentação matemática, bem como a sua capacidade de orientar a seleção de modelos parcimoniosos, tornam-no um recurso valioso na modelação estatística e análise.

4.6.4 Critério de Informação Bayesiano

O Critério de Informação Bayesiano (**BIC**), derivado por Schwarz [1978], é reconhecido pela sua simplicidade e eficácia em várias estruturas de modelação, incluindo aplicações Bayesianas onde a definição de distribuições prévias pode ser desafiadora [Neath, 2012]. No seu cerne, o **BIC** serve como uma aproximação assintótica a uma transformação da probabilidade *a posteriori* Bayesiana de um modelo candidato [Neath, 2012].

A fórmula matemática para o **BIC** é expressa na Equação 4.17.

$$BIC = -2 \ln(\hat{L}) + k \ln(n) \quad (4.17)$$

O n representa o número de observações, \hat{L} a **MV** e k denota o número de parâmetros no modelo.

Em termos práticos, o modelo selecionado como o melhor ajuste é aquele associado ao valor mínimo do **BIC**. A seleção de modelos baseados no **BIC** é vantajosa, pois este possui a propriedade de consistência. Por exemplo, supondo que o modelo gerador $g(y)$ seja de dimensão finita e que este modelo esteja representado na coleção de candidatos. O **BIC** selecionará assintoticamente, com probabilidade um, o modelo candidato com a estrutura correta [Neath, 2012].

É importante destacar que alguns profissionais preferem o **BIC** relativamente a outros critérios de seleção de modelos, como o **AIC**, pois o **BIC** tem uma tendência a favorecer modelos mais parcimoniosos, tornando-o uma escolha valiosa em vários cenários.

Em síntese, o **BIC** é uma ferramenta poderosa para a seleção de modelos, oferecendo uma solução para encontrar o equilíbrio entre a complexidade do modelo e a qualidade do ajuste. A sua capacidade de fazer escolhas informadas dos modelos, especialmente em contextos Bayesianos onde a informação prévia pode ser escassa, sublinha a sua importância na prática estatística.

4.6.5 Estatísticas Z

As estatísticas Z são ferramentas inestimáveis, especialmente ao avaliar a normalidade dos resíduos no contexto de uma variável x independente, como a idade, frequentemente usada na estimativa de percentis, como explicado por [Royston and Wright \[2000\]](#).

Para entender as estatísticas Z, considere-se a idade como a variável explicativa. Com G como o número de grupos de idade e, dentro de cada grupo $g = 1, 2, \dots, G$, temos $r_{gi}, i = 1, 2, \dots, n_i$ como os resíduos, caracterizados por uma média \bar{r}_g e desvio padrão s_g [[Stasinopoulos et al., 2017](#)]. Dentro desse contexto, calcula-se quatro estatísticas-chave, Z_{g1}, Z_{g2}, Z_{g3} e Z_{g4} , cada uma projetada para avaliar aspetos específicos dos resíduos no grupo g . Estas estatísticas avaliam se os resíduos estão conforme as características desejadas da população, incluindo média 0, variância 1, assimetria 0 e curtose 3 [[Stasinopoulos et al., 2017](#)]. Z_{g1}, Z_{g2} , são calculadas nas Equações 4.18 e 4.19.

$$Z_{g1} = n_g^{1/2} \bar{r}_g, \quad (4.18)$$

$$Z_{g2} = \{s_g^{2/3} - [1 - 2/(9n_g - 9)]\} / \{2/(9n_g - 9)\}^{1/2} \quad (4.19)$$

Z_{g3} e Z_{g4} são estatísticas de teste para assimetria e curtose fornecidas por [D'agostino et al. \[1990\]](#). A estatística D'Agostino K^2 , é dada na Equação 4.20.

$$K^2 = Z_{g3}^2 + Z_{g4}^2 \quad (4.20)$$

As estatísticas Z_{gj}^2 , elucidam a contribuição de cada grupo de idade, ajudando a identificar quais os grupos de idade que contribuem para a significância estatística e onde são necessários ajustes no modelo [[Stasinopoulos et al., 2008](#)]. Valores de $|Z_{gj}| > 2$ podem servir como um indicador de inadequações significativas do modelo. Notavelmente, valores positivos (ou negativos) significativos $|Z_{gj}| > 2$ (ou $|Z_{gj}| < 2$) para $g = 1, 2, 3$ ou 4 indicam, respetivamente, que os resíduos têm média, variância, assimetria ou curtose maiores (ou menores) do que a distribuição normal padrão. [[Stasinopoulos et al., 2017](#)]. Nestes casos, é aconselhável considerar o aumento dos graus de liberdade no modelo para o parâmetro relevante. Por exemplo, se a média dos resíduos num grupo de idade específico exceder as expectativas, pode ser necessário aumentar os graus de liberdade para o parâmetro μ para alinhar os valores μ ajustados dentro desse grupo de idade.

Para as estatísticas Z, existe uma representação visual que ajuda a avaliar a normalidade dos resíduos. O gráfico utiliza um esquema de cores e o tamanho dos círculos para transmitir informações.

Cores claras representam valores negativos de Z , enquanto cores escuras indicam valores positivos de Z [Stasinopoulos et al., 2017]. O tamanho do círculo corresponde à magnitude do valor de Z , com círculos maiores representando desvios mais significativos e no centro de cada círculo, há um quadrado que indica que $|Z| > 2$, destacando possíveis inadequações [Stasinopoulos et al., 2017].

4.6.6 Resíduos quantílicos normalizados

Os resíduos quantílicos normalizados, frequentemente referidos como resíduos quantílicos normalizados aleatorizados quando são utilizadas distribuições de variáveis resposta discretas, oferecem um meio para avaliar a qualidade de ajuste de um modelo, particularmente no contexto de **GAMLSS** [Stasinopoulos et al., 2008].

A principal vantagem da utilização de resíduos quantílicos normalizados é a sua capacidade de fornecer um método consistente e familiar para avaliar a adequação do modelo, independentemente da distribuição da variável resposta. Quando o modelo assumido está correto, os resíduos seguem consistentemente uma distribuição Normal padrão, que está bem estabelecida na literatura estatística como um indicador fundamental de adequação do modelo [Stasinopoulos et al., 2017]. A distribuição Normal padrão possui propriedades-chave, incluindo uma média de 0, uma variância de 1, assimetria de 0 e curtose de 3 (baseada em momentos) [Stasinopoulos et al., 2017].

O processo de obtenção de resíduos quantílicos normalizados ajustados em modelos **GAMLSS** é facilitado pela função `resid()` no pacote `gamlss` no R. Este pacote oferece várias funções adicionais para auxiliar na avaliação da adequação do modelo, incluindo:

- `plot()`: Uma ferramenta geral para verificação de resíduos.
- `wp()`: Produz um *worm plot*, que ajuda a identificar se a distribuição ajustada é adequada.

Estas ferramentas de diagnóstico aproveitam as propriedades dos resíduos quantílicos normalizados para avaliar o ajuste do modelo e destacar áreas de potencial melhoria.

Em resumo, os resíduos quantílicos normalizados oferecem uma abordagem valiosa para avaliar a qualidade do ajuste nos modelos **GAMLSS**. A sua adesão a uma distribuição Normal padrão, quando os pressupostos do modelo são cumpridos, garante a sua consistência e fiabilidade como ferramenta para avaliar a adequação do modelo em diversas distribuições de variáveis resposta. Esta contribuição aumenta a robustez e interpretabilidade da modelação estatística.

Worm plot

O *worm plot*, introduzido por [Buuren and Fredriks \[2001\]](#), destaca-se como uma ferramenta fundamental, oferecendo percepções únicas sobre a qualidade de ajuste de um modelo. Estes gráficos, frequentemente utilizados no contexto de modelos aditivos generalizados e curvas de referência de crescimento condicionadas à idade, proporcionam uma representação visual das distribuições de resíduos [[Buuren and Fredriks, 2001](#)].

Essencialmente, é um gráfico Quantil-Quantil (Q-Q) pós-remoção da tendência dos resíduos. Esta técnica serve como um meio para avaliar a qualidade do ajuste, examinando os resíduos e identificando regiões (intervalos) de uma variável explicativa dentro das quais o modelo não se ajusta adequadamente aos dados (“violação do modelo”) [[Stasinopoulos et al., 2017](#)].

A disposição dos pontos de dados nestes gráficos assemelham-se a uma estrutura em forma de “minhoca”, e a sua forma revela como os dados se desviam da distribuição subjacente assumida. Na [Tabela 2](#) é possível verificar o significado para cada uma das formas do *worm plot* quando um modelo está mal ajustado aos dados.

As características-chave do *worm plot* incluem:

1. **Pontos do gráfico:** Estes pontos representam o quão longe os resíduos ordenados estão dos seus valores esperados (aproximados), representados no gráfico pela linha horizontal. A proximidade destes pontos à linha horizontal reflete a distribuição dos resíduos relativamente a uma distribuição normal padrão [[Stasinopoulos et al., 2017](#)].
2. **Intervalos de confiança de 95%:** Representados por duas curvas elípticas, estes intervalos servem como referência. Idealmente, cerca de 95% dos pontos devem estar entre estas curvas, com desvios a indicar potenciais inadequações na distribuição ajustada ou nos termos do modelo [[Stasinopoulos et al., 2017](#)]. Também podem ser identificados valores discrepantes quando estes se encontram muito afastados destas curvas [[Buuren and Fredriks, 2001](#)].
3. **Curva ajustada aos pontos da “minhoca”:** Esta curva cúbica é ajustada aos pontos do gráfico. A sua forma fornece informações valiosas sobre a natureza das inadequações no modelo, descritas na [Tabela 2](#). Por exemplo, se o nível dos pontos do gráfico estiverem consistentemente acima ou abaixo de uma linha horizontal na origem, sugere problemas relacionados com a média dos resíduos ou a localização da distribuição ajustada [[Stasinopoulos et al., 2017](#)]. Da mesma forma, tendências lineares (positivas ou negativas), quadráticas (U ou U inverso) ou cúbicas (formato em S) indicam problemas com a variância, assimetria ou curtose dos resíduos, respetivamente,

revelando áreas onde a distribuição ajustada ou os parâmetros do modelo podem necessitar de ajustes [Stasinopoulos et al., 2017].

Uma “minhoca” plana no gráfico significa um modelo bem ajustado, enquanto desvios dessa forma indicam potenciais modificações necessárias para melhorar a precisão. Esta ferramenta de diagnóstico é particularmente útil para orientar o processo de modelação, especialmente em cenários em que desvios locais do modelo podem ter implicações substanciais.

Em conclusão, o *worm plot* é uma ferramenta valiosa de diagnóstico na modelação estatística, auxiliando na avaliação da qualidade de ajuste do modelo e destacando áreas onde podem ser necessárias melhorias. A sua aplicação é particularmente relevante na modelação de curvas de referência de crescimento condicionadas à idade e fornece um meio eficaz para aumentar a fiabilidade dos modelos estatísticos.

Tabela 2: Descrição da forma do *worm plot* relativamente aos resíduos e à distribuição ajustada quando um modelo está mal ajustado aos dados. Adaptado de Stasinopoulos et al. [2017].

Forma do worm plot (ou a sua curva ajustada)	Resíduos	Distribuição ajustada
Nível: acima da origem	Média muito alta	Localização ajustada muito baixa
Nível: abaixo da origem	Média muito baixa	Localização ajustada muito alta
Linha: declive positivo	Variância muito alta	Escala ajustada muito baixa
Linha: declive negativo	Variância muito baixa	Escala ajustada muito alta
Forma 'U'	Assimetria positiva	Assimetria ajustada muito baixa
Forma 'U' invertida	Assimetria negativa	Assimetria ajustada muito alta
Forma 'S' com a esquerda curvada para baixo	Leptocurtose	Caudas da distribuição ajustada pesadas
Forma 'S' com a esquerda curvada para cima	Platicurtose	Caudas da distribuição ajustada leves

Gráficos de resíduos

A análise de resíduos é um componente crucial na modelação estatística, permitindo-nos avaliar a qualidade do ajuste de um modelo e identificar potenciais violações das suposições do modelo. Como parte do diagnóstico do modelo, é importante verificar os gráficos dos resíduos quantílicos normalizados associados ao modelo. Neste estudo foram analisados quatro gráficos-chave produzidos pela função `plot()` do pacote `gamlss`. Os gráficos são os seguintes:

- **Resíduos em relação aos valores ajustados do parâmetro μ :** Este gráfico indica os resíduos em relação aos valores ajustados do parâmetro μ , que representa a localização ou média

da variável resposta. Este gráfico ajuda a avaliar se o modelo capta adequadamente a tendência nos dados. Um modelo bem ajustado deve exibir uma dispersão aleatória de resíduos em torno da linha horizontal em zero, sem um padrão discernível [Stasinopoulos et al., 2017].

- **Resíduos em relação a um índice ou uma covariável especificada:** Este gráfico permite examinar os resíduos em relação a um índice ou a uma covariável especificada. É particularmente útil quando se avalia a influência de variáveis preditoras específicas nos resíduos. Se surgir um padrão relacionado com a covariável, isso pode sugerir que o modelo não considera totalmente os efeitos dessa covariável, o que pode levar à necessidade de um aperfeiçoamento do modelo [Stasinopoulos et al., 2017].
- **Estimativa de densidade pelo método do núcleo dos resíduos:** Este gráfico fornece informações sobre a distribuição dos resíduos do modelo. Um modelo bem ajustado deve ter resíduos que se aproximam de uma distribuição normal, resultando numa estimativa de densidade pelo método do núcleo simétrica centrada em torno de zero [Stasinopoulos et al., 2017]. Desvios da normalidade, como assimetria ou caudas pesadas, podem ser detetados através deste gráfico. Se forem observados desvios significativos, isso pode indicar que as suposições do modelo, especialmente em relação à distribuição de erros, precisam de ser reavaliadas.
- **Gráfico QQ-Normal dos resíduos:** Este gráfico permite avaliar se os resíduos seguem uma distribuição normal padrão, para tal o gráfico deve ser linear com uma interceção em 0 e um gradiente de 1 [Stasinopoulos et al., 2017]. Os quantis observados dos resíduos são comparados com os quantis esperados de uma distribuição normal. Um modelo bem ajustado produzirá pontos alinhados ao longo de uma linha diagonal. Desvios desta linha diagonal podem indicar desvios da normalidade nos resíduos.

Em suma, a utilização de gráficos informativos de resíduos auxilia na avaliação da adequação do modelo e na identificação de áreas onde podem ser necessárias melhorias, contribuindo para a interpretabilidade do modelo.

4.7 Pacotes

O pacote `gamlss` possui várias funções para ajustar os **GAMLSS** [gamlss]. Este pacote para ajustar um modelo **GAMLSS** utiliza a função principal `gamlss()` e contém métodos para lidar com os objetos `gamlss` ajustados [Stasinopoulos et al., 2017].

A função `gamlss()`, oferece uma abordagem versátil para a modelação estatística que vai além das capacidades de funções tradicionais como a função `gam()` do pacote `gam`. Enquanto `gam()` está limitado a modelar distribuições pertencentes à família exponencial, `gamlss()` é mais abrangente, acomodando uma variedade mais ampla de famílias de distribuição. Além disso, permite modelar os quatro parâmetros associados a uma distribuição: μ, σ, ν, τ , e expressá-los como funções de variáveis explicativas. Ao executar a função `gamlss()`, é retornado um objeto `gamlss` que representa um modelo **GAMLSS** ajustado. É utilizado o algoritmo **RS**, que é vantajoso devido à sua capacidade de convergência sem necessidade de valores iniciais precisos para μ, σ, ν, τ , tornando-o adequado para conjuntos de dados maiores.

Para construir um modelo com a função `gamlss()`, é necessário definir argumentos essenciais. O argumento *“formula”* representa a fórmula padrão do modelo R para o preditor do parâmetro μ , com a resposta à esquerda de um operador \sim , enquanto *“sigma.formula”*, *“nu.formula”* e *“tau.formula”* lidam com os parâmetros σ, ν, τ , respetivamente. *“family”* designa a família de distribuição da variável resposta e *“data”* refere-se ao conjunto de dados. Também é possível escolher entre os algoritmos **RS**, **CG**, ou *mixed* para a estimativa.

Outra função importante do pacote `gamlss`, que foi utilizada neste estudo, é a `lms()`, que foi criada para simplificar a seleção automática de um tipo apropriado de estimativa de percentis com o método **LMS** fornecido pelo modelo 4.13. Esta função determina a distribuição da variável de resposta, os graus de liberdade adequados para todos os parâmetros da distribuição e o cálculo do parâmetro de potência ξ . É importante observar que a função `lms()` é apenas aplicável quando há uma única variável explicativa. Ela aceita diversos argumentos, como *“y”* que representa a variável resposta, *“x”* que indica a variável explicativa, *“families”* que contém uma lista de famílias padrão (**BCCG**, **BCT** e **BCPE**), *“data”* que contém os dados de interesse, *“cent”* que define os valores dos percentis para os quais as curvas são avaliadas e *“method.pb”* que define o método usado na estimativa local dos parâmetros de alisamento.

De forma a ser possível guardar os valores dos percentis calculados nos modelos **GAMLSS** ou **LMS**, **LMST** e **LMSP** para posterior uso na aplicação *web* desenvolvida, foi utilizada a função `centiles.pred()` do pacote `gamlss`. Dando os valores de x a esta função, os valores dos percentis desejados e o modelo ajustado, ela calcula uma matriz contendo os valores dos percentis para y .

Outro pacote utilizado durante este estudo foi o `gamlss.ggpplots`, uma extensão do pacote `ggplot2`. Foi projetado para aprimorar a visualização e representação gráfica de modelos de regressão não lineares ajustados com o pacote `gamlss`. O pacote `ggplot2` também foi utilizado para a representação gráfica dos dados na análise exploratória.

O pacote `moments`, foi utilizado para calcular estatísticas dos dados como a assimetria e a curtose.

4.8 Visão geral da metodologia implementada na construção das curvas de crescimento

Ao longo deste capítulo foi abordada detalhadamente a metodologia utilizada na construção das curvas de crescimento. A metodologia é dividida em várias secções interconectadas, que delineiam o processo desde a revisão dos métodos até à avaliação da qualidade dos ajustes. Na Figura 6 é possível visualizar um resumo da metodologia implementada.

Inicialmente foi realizada uma revisão crítica dos métodos existentes para a construção das curvas de crescimento, presente na Secção 2.2. Em seguida, de acordo com essa revisão e tendo em consideração os critérios presentes na Subsecção 4.2.1, foram seleccionados alguns métodos: **GAMLSS**, **LMS**, **LMST** e **LMSP**. Cada um destes métodos oferece abordagens únicas para modelar as curvas de crescimento.

Foram seleccionadas quatro técnicas de alisamento: *P-splines*, *Splines Cúbicas*, *Redes Neurais* e *Loess*. Cada uma dessas técnicas é projetada para capturar diferentes padrões de crescimento. Dentro de cada método, foram exploradas distribuições específicas que melhor se adequam ao contexto deste estudo. Isso inclui as distribuições **BCCG**, **BCT**, **BCPE** e **JSU**. Estas distribuições desempenham um papel fundamental na representação das características dos dados.

Por fim, foram seleccionados vários critérios de qualidade de ajuste, incluindo o R^2 , o **MSE**, o **AIC**, o **BIC**, estatísticas *Z*, *Worm plots*, e gráficos de resíduos. Estes critérios fornecem informações sobre a adequação dos modelos escolhidos aos dados observados.

Esta metodologia foi, portanto, aplicada a vários modelos, com o intuito de identificar o modelo mais apropriado para cada um dos parâmetros em estudo, nomeadamente **D1**, **D2**, **EP**, **PP**, e **PF**. É relevante observar que seguiu uma ordem específica de procedimentos.

Inicialmente, foram avaliadas várias funções de alisamento num modelo **GAMLSS**, sem a imposição de uma distribuição específica para o modelo. Estas funções de alisamento são conhecidas pela sua capacidade de flexibilidade ao ajustar curvas, sem a necessidade de predefinir uma forma específica para a função subjacente, *Currie and Durban* [2002]. A justificação para esta abordagem reside na simplificação do processo de modelação. Incorporar simultaneamente as funções de alisamento e as distribuições aumentaria consideravelmente a complexidade do processo de modelação. Cada combinação adicional de função de alisamento e distribuição exigiria mais cálculos e ajustes. Ao separar estes dois aspetos, é possível simplificar o processo de seleção do modelo.

Posteriormente, foram testadas diversas distribuições em combinação com a função de alisamento escolhida, utilizando diferentes métodos. Por fim, foi avaliada a qualidade de ajuste de cada modelo.

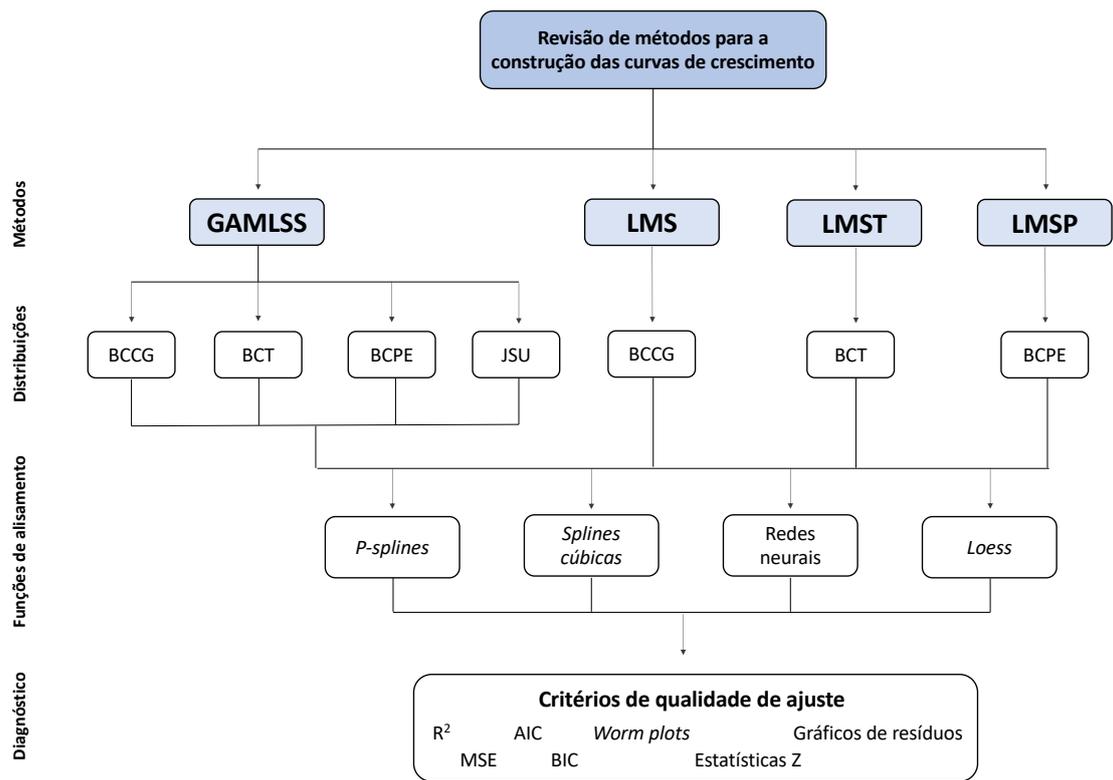


Figura 6: Diagrama ilustrando a metodologia aplicada para a construção das curvas de crescimento neste estudo.

Capítulo 5

Preparação dos dados para análise e modelação

Neste capítulo, é apresentada a metodologia utilizada para preparar e analisar os dados relativos aos parâmetros biométricos da placenta utilizados nesta investigação. Começa-se por explorar a origem do conjunto de dados, que se baseia num estudo anterior de [Nogueira et al. \[2019\]](#), abrangendo um amplo espectro de informações sobre a placenta e recém-nascidos. Em seguida, é detalhado o processo de pré-processamento dos dados, que incluiu a seleção de variáveis de interesse, tratamento de valores ausentes e discrepantes, resultando num conjunto coeso de 2064 observações com 11 variáveis relevantes. Este conjunto de dados é o alicerce fundamental para a modelação subsequente das curvas de crescimento, que serão exploradas nos capítulos seguintes.

5.1 Conjunto de dados

Neste estudo, foram modeladas e otimizadas curvas de crescimento para vários parâmetros placentares em função da **IG**, para a população portuguesa. O conjunto de dados utilizado para construir e testar os modelos gerados, foi em parte aplicado num estudo anterior de [Nogueira et al. \[2019\]](#).

O conjunto de dados utilizado no estudo de [Nogueira et al. \[2019\]](#), começou por uma análise de 7321 amostras de placenta. Essas amostras foram meticulosamente submetidas ao Laboratório de Patologia Embrionária-Fetal, localizado no **CGC**, Unilabs, Porto, Portugal. O principal objetivo foi realizar um exame histopatológico, para confirmar, ou revelar lesões potenciais que possam esclarecer resultados obstétricos. A informação recolhida dos relatórios patológicos da placenta abrangeu um período de quatro anos, de 1 de janeiro de 2014 a 31 de dezembro de 2017. Os relatórios incluíram um amplo intervalo de **IG**, variando de 12 a 41 semanas. Os parâmetros biométricos da placenta foram recolhidos de relatórios de autópsia de placentas e fetos. Além disso, também foram reunidos detalhes biométricos sobre recém-nascidos por meio das solicitações clínicas associadas a exames patológicos da placenta. Os aspetos biométricos considerados incluíram parâmetros como peso e espessura da placenta, forma e diâmetro da placenta, comprimento do cordão umbilical, diâmetros do cordão umbilical e tipo de inserção. Ademais, foram

registados meticulosamente parâmetros fetais, como o peso e o género.

Foram considerados alguns critérios na seleção das amostras, tendo sido garantido que o conjunto de dados obedecesse a condições específicas: **IG** conhecida, mulher da população portuguesa, gravidez única ≥ 12 semanas de gestação, critérios rigorosos para amostras de placenta, parâmetros de óbito fetal meticulosamente documentados e registos precisos de peso ao nascer. Foram excluídos meticulosamente casos que não se alinhavam com esses critérios, como mãe não portuguesa, casos de gestações múltiplas ou gestações únicas resultantes de tecnologias de reprodução assistida, e casos envolvendo doenças maternas crónicas conhecidas como diabetes ou hipertensão. Para mais informações acerca destes critérios, consultar [Nogueira et al. \[2019\]](#). Após exclusão de amostras seguindo estes critérios, resultou num novo conjunto de dados composto por 1951 amostras de placenta.

Para este estudo, o conjunto de dados foi estendido, tendo sido introduzidos novos casos que tiveram em consideração os critérios inicialmente delineados por [Nogueira et al. \[2019\]](#). O conjunto de dados final em estudo contou com um conjunto de $n = 2092$ amostras de placenta. Dentro dessas 2092 amostras é possível encontrar 22 variáveis que poderão ser utilizadas neste estudo, no Apêndice B é possível verificar informações acerca de todas as variáveis deste conjunto de dados, como, a sua descrição, tipo (escala de medida) e codificação quando a variável apresentada é categórica nominal. Esta codificação foi previamente realizada em outro estudo [[Nogueira et al., 2019](#)], e para as amostras adicionadas ao conjunto de dados pré-existente foi tida em consideração a codificação pré-definida. Além disso, este conjunto de dados passou por um procedimento de pré-processamento de forma a permitir uma exploração mais abrangente dos dados.

5.2 Pré-processamento

Foi realizado um processo abrangente de pré-processamento do conjunto de dados da placenta. O objetivo deste pré-processamento foi preparar os dados brutos para análise estatística e modelação posterior, eliminando erros, valores ausentes e valores discrepantes, garantindo a qualidade e a confiabilidade dos dados.

Dado o escopo deste estudo, começou-se por seleccionar um subconjunto das variáveis disponíveis que eram de interesse para a análise e posterior modelação para a construção das curvas de crescimento. Dentro das 22 variáveis disponíveis (informações no Apêndice B), foram seleccionadas 11 variáveis de interesse. Estas variáveis incluíram “local”, “GA” (idade gestacional), “MaternalAge” (idade da mãe), “Obstetriccontext” (paridade da mãe), “Fetalweight” (peso fetal), “Fetalgender” (género fetal), “Placen-

talweight” (peso placentar), “Diameter1” (diâmetro 1), “Diameter2” (diâmetro 2), “Placentalthickness” (espessura placentar) e “Placentalshape” (forma placentar).

Numa primeira análise do conjunto de dados foi possível verificar que existiam alguns registos com valores de “999”, estes casos indicam falta de informação para essa variável. Portanto, foi optado por recodificar essas entradas como NA's, ou seja, como valores omissos. Procedeu-se então à contagem do número de valores omissos, encontrando-se 22 valores (1,1% dos dados totais). Dada a quantidade mínima e ausência de um padrão destes valores, optou-se por remover todos os casos com valores omissos para pelo menos uma variável [Acuña and Rodriguez, 2004]. Além disso, foi verificado se existiam linhas duplicadas no conjunto de dados. Foram encontradas 7 linhas duplicadas, que foram removidas. Esta remoção garantiu a integridade dos dados, evitando duplicações indesejadas.

Para manter a qualidade dos dados e evitar que valores extremos afetassem as análises dos dados e posteriores modelos, realizou-se a deteção de valores discrepantes em algumas variáveis seleccionadas, incluindo **D1**, **D2**, **PP**, **EP** e **PF**. A presença de valores discrepantes pode ter implicações significativas na interpretação dos resultados e, portanto, a abordagem adotada foi baseada em boas práticas estatísticas e considerações biológicas específicas. Para identificar os valores discrepantes, foram utilizados diagramas em caixa, uma ferramenta comumente utilizada na deteção visual de valores discrepantes em conjuntos de dados. Esta técnica fornece uma representação gráfica que destaca a dispersão dos dados e a presença de pontos que se desviam significativamente do padrão geral [Krzywinski and Altman, 2014]. Os diagramas em caixa das variáveis em estudo estão presentes na Figura 7. A sua análise permitiu uma avaliação clara da distribuição dos dados e destacou a presença de pontos discrepantes em quatro variáveis cruciais para esta investigação: **D1**, **D2**, **PP** e **EP**.

Uma consideração importante neste estudo é o facto de os dados serem biológicos, mais precisamente, dados relacionados à placenta. Em estudos biológicos, é possível que os valores discrepantes possam não ser simples erros de medição, mas sim reflexões de variações naturais ou eventos biológicos excepcionais. Portanto, foi optado por uma abordagem conservadora, removendo apenas os valores discrepantes mais severos.

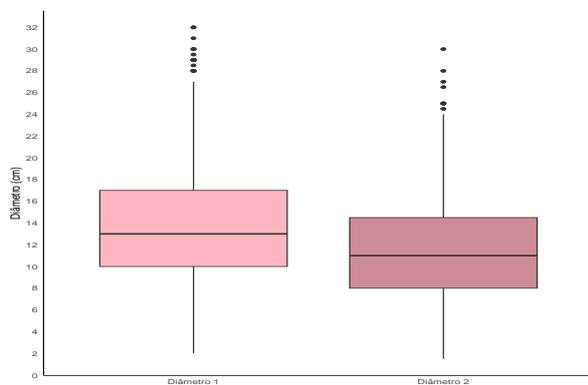
O processo de remoção dos valores discrepantes severos foi realizado conforme a seguinte metodologia:

1. Cálculo dos quartis ($Q1$ e $Q3$) e a amplitude interquartil (**IQR**) para cada variável (**D1**, **D2**, **PP**, **EP** e **PF**) utilizando a função `quantile()` do *software* R.
2. Em seguida, foi determinado os limites inferior e superior para a remoção dos valores discrepantes. Qualquer valor abaixo de $Q1 - 3 \times IQR$ ou acima de $Q3 + 3 \times IQR$ foi considerado um valor

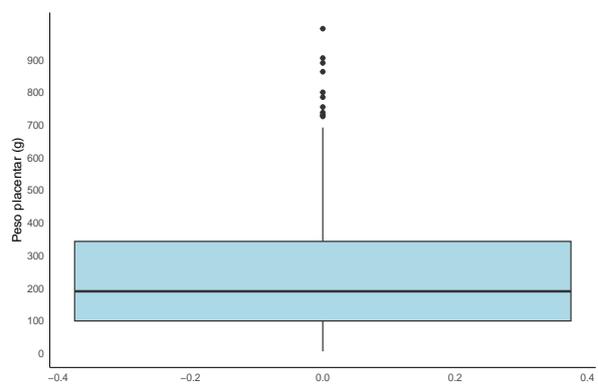
discrepante severo sendo removido do conjunto de dados.

Esta abordagem foi escolhida para preservar a integridade dos dados, ao mesmo tempo, em que reconhece a possibilidade de que os valores discrepantes possam conter informações biológicas valiosas. Com esta técnica, foram removidos apenas dois valores discrepantes severos dos dados.

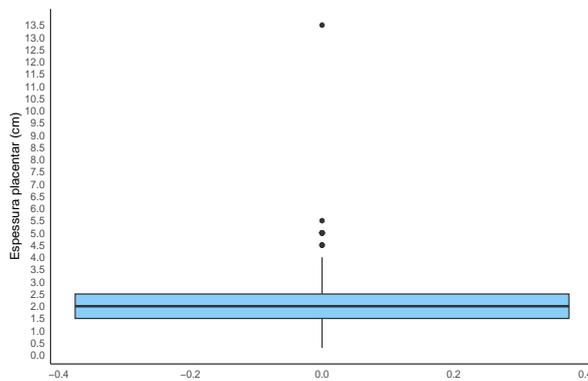
No final deste processo de pré-processamento, obteve-se um conjunto de dados limpo, consistente e dimensionado para análises subsequentes, contendo 2064 observações e 11 variáveis de interesse. Estes dados preparados serviram como base para a modelação e otimização das curvas de crescimento para os parâmetros placentares relativamente à **IG** na população portuguesa.



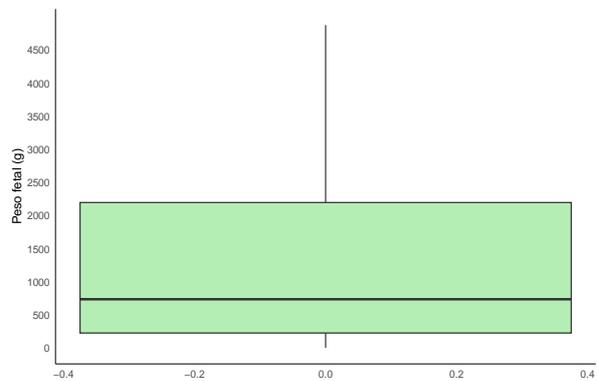
(a) Diâmetro 1 e Diâmetro 2



(b) Peso placentar



(c) Espessura placentar



(d) Peso fetal

Figura 7: Diagramas de caixa para as variáveis **D1**, **D2**, **PP**, **EP** e **PF**.

Capítulo 6

Desenvolvimento da aplicação web

Neste capítulo, é descrito detalhadamente a metodologia utilizada para o desenvolvimento da aplicação web “PlacentalGrowth” que serve como peça central desta investigação. Esta aplicação foi desenvolvida para permitir aos utilizadores analisar as curvas de crescimento de parâmetros placentares em função da **IG**. A aplicação oferece uma *interface* intuitiva que facilita a leitura, visualização e interação com os dados, além de calcular rácios relevantes para análise obstétrica. Com ênfase na usabilidade e acessibilidade, a aplicação foi construída no ambiente R, utilizando o pacote `shiny` para criar uma experiência interativa. Neste capítulo explora-se a arquitetura da aplicação, as funções reativas e os elementos da *interface* que garantem uma análise dinâmica e amigável para o utilizador.

6.1 Estratégia de desenvolvimento e ferramentas

Foi desenvolvida uma aplicação web com recursos que permitem a análise de curvas de crescimento de referência de parâmetros placentares. Existem módulos para leitura, visualização e interação dos pontos adicionados nos gráficos das curvas de crescimento, leitura e visualização de informações relativas ao paciente e aos dados do processo e cálculo de rácios importantes para a análise de possíveis resultados obstétricos. A aplicação web desenvolvida, visa, portanto, implementar vários recursos que permitam a interpretação dos parâmetros **D1**, **D2**, **PP**, **EP** e **PF** em função da **IG** para assim ser possível realizar um relatório acerca do desenvolvimento da placenta.

O desenvolvimento desta aplicação web teve como principal preocupação a facilidade de uso para o utilizador, sendo assim uma aplicação fácil de utilizar onde o utilizador não precisa de ter conhecimentos de nenhum tipo de linguagem de programação e proporcionando ao mesmo tempo, uma visualização gráfica dos dados e resultados, para que assim possam ser facilmente interpretáveis. Foi desenvolvida de forma a que todos os resultados, estejam presentes em forma de texto, tabela ou gráfico e que esses mesmos resultados possam ser transferidos para utilizações futuras.

A linguagem de programação escolhida para o desenvolvimento da aplicação foi o R [TEAM, 2005],

um ambiente de *software* integrado gratuito para manipulação de dados, computação científica e estatística e visualização gráfica. Mais detalhes acerca do R podem ser consultados na Secção 2.3. Mais especificamente, foi utilizado o pacote *shiny* [Team, 2012] que permitiu a criação da aplicação *web* interativa com o R.

Uma aplicação *shiny* tem dois componentes: a interface do utilizador (**UI**) e o servidor. A sua interação está descrita na Figura 8. O **UI** é a *frontend* que aceita os valores de entrada definidos pelo utilizador, enquanto o servidor é a *backend* que processa os valores de entrada para produzir os resultados de saída que são exibidos na aplicação *web*. O *shiny* funciona baseado em programação reativa, e a ideia principal é especificar um grafo de dependências para que, quando uma entrada for alterada, todas as saídas relacionadas sejam atualizadas automaticamente [Wickham, 2021]. Existem três tipos de objetos na programação reativa: fontes reativas, condutores reativos e terminais reativos. Mais detalhes acerca do *shiny* podem ser consultados na Subsecção 2.3.1.

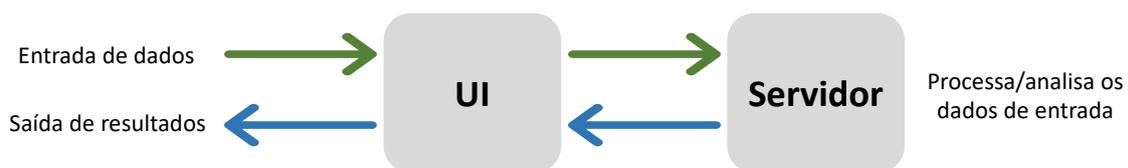


Figura 8: Diagrama apresentando a interação entre os dois componentes de uma aplicação *shiny*, o **UI** e o servidor.

Para o desenvolvimento da aplicação foram utilizados outros pacotes relevantes, que incluem:

- *ggplot2*: Fornece ferramentas gráficas poderosas e elegantes, que podem ser utilizadas para traçar dados e fazer ajustes de modelo [*ggplot2*].
- *plotly*: Cria gráficos interativos numa aplicação *web* a partir de gráficos *ggplot2* [*plotly*].
- *shinydashboard*: Permite o desenvolvimento de uma aplicação *shiny* com aparência típica de *dashboard* [*shinydashboard*].
- *prompter*: Permite armazenar informações sobre um *item* específico numa ferramenta de ajuda que podem ser consultadas pelo utilizador [*prompter*].

- `shinyjs`: Permite realizar operações em JavaScript em aplicações `shiny`, como ocultar, redefinir ou desativar elementos [[shinyjs](#)].
- `DT`: Permite que objetos de dados em R sejam renderizados como tabelas HTML [[DT](#)].
- `dplyr`: Uma ferramenta rápida e consistente para trabalhar com objetos *data frame*, tanto na memória quanto fora dela [[dplyr](#)].
- `readr`: Fornece uma maneira rápida e amigável de ler dados como “csv”, “tsv” e “fwf” [[readr](#)].

O RStudio [[TEAM, 2011](#)] foi escolhido como o **IDE** para criar os *scripts* em R e construir a plataforma. Este **IDE** é desenvolvido utilizando a linguagem de programação C++ e possui uma *interface* intuitiva complementada por uma variedade de ferramentas úteis, facilitando um fluxo de trabalho em R mais contínuo e eficiente.

Para o desenvolvimento de um relatório com os resultados das análises realizadas na aplicação foi utilizado o R Markdown [[rmarkdown](#)]. Este combina a sintaxe do Markdown com partes de código R que são executados, exibindo a saída do código no relatório final [[Baumer and Udwin, 2015](#)]. Este relatório é criado a partir do pacote `knitr`, que é utilizado para criar relatórios dinâmicos com R [[Baumer and Udwin, 2015](#)]. Os formatos de saída comuns para esses relatórios incluem documentos HTML, PDF ou Microsoft Word.

Atualmente, o código R utilizado para desenvolver a aplicação *web* pode ser acessado através do seguinte repositório usando o [GitHub](#):

```
git clone https://github.com/daniellalemos/PlacentalGrowth
```

A aplicação *web*, intitulada “PlacentalGrowth”, pode ser acessada através do seguinte [link](#):

```
https://placentalgrowth.shinyapps.io/uminho_pt/
```

6.2 Arquitetura da aplicação

A estrutura da *interface* da aplicação “PlacentaGrowth” desenvolvida neste estudo é projetada para oferecer aos utilizadores uma experiência intuitiva e completa para a análise do crescimento da placenta durante a **IG**.

Uma representação gráfica da estrutura da *interface* da aplicação é demonstrada na Figura 9. O *layout* da aplicação consiste numa *dashboard* com três componentes: *header*, *sidebar* e *body*, que foi

criado utilizando o pacote `shinydashboard`. A *sidebar* é dividida em várias páginas interconectadas, cada uma com o seu propósito específico e funcionalidades bem definidas. A “Página inicial” serve como ponto de entrada para os utilizadores e fornece informações sobre a aplicação. A página “Gráficos” é o núcleo da aplicação, onde os utilizadores podem inserir as informações do processo e parâmetros a analisar, visualizar as curvas de crescimento de referência da placenta com os pontos adicionados (via CSV ou manualmente) pelo utilizador, interagir com os mesmos e transferir um relatório em PDF contendo informações essenciais sobre o crescimento da placenta com base nos dados inseridos. A página “Informações” contém tabelas informativas que exibem os dados de entrada inseridos pelos utilizadores relativamente aos parâmetros em estudo. A página “Rácios” é dedicada ao cálculo de rácios com base nos dados de entrada seleccionados pelos utilizadores. E por fim, a página “Ajuda” fornece um guia detalhado de utilização da aplicação, orientando os utilizadores sobre como aproveitar ao máximo todas as funcionalidades disponíveis.

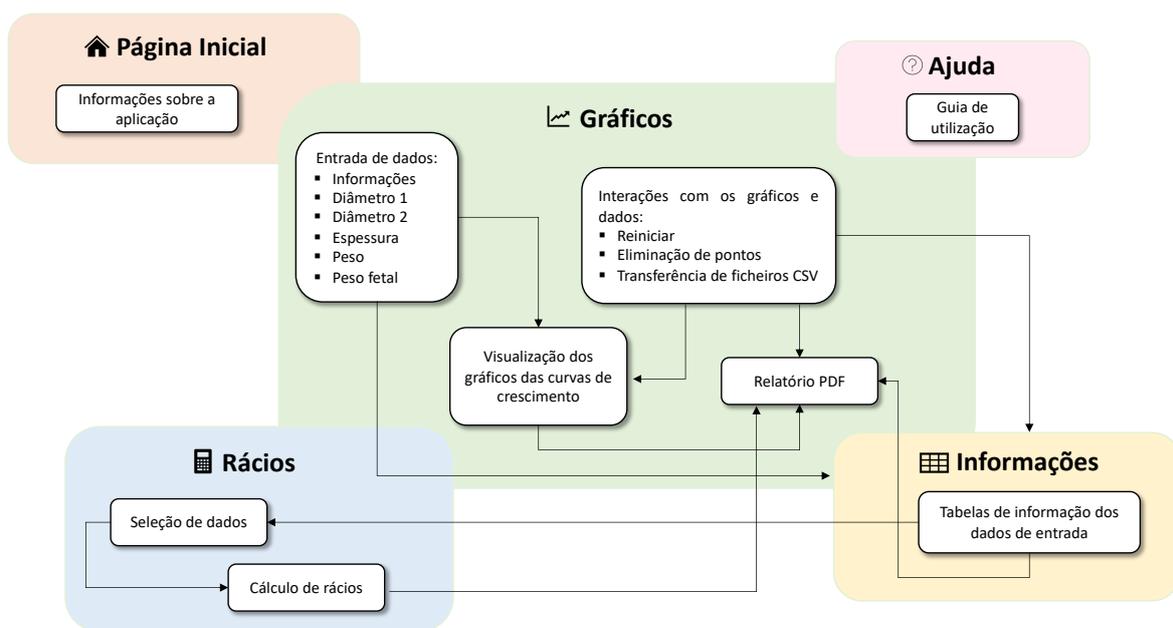


Figura 9: Representação gráfica da estrutura da *interface* da aplicação. Cada caixa colorida representa uma página do menu da barra lateral da aplicação. As caixas brancas representam funcionalidades dentro de cada página, estando interligadas por setas.

6.2.1 Funções e eventos reativos

De forma a compreender a arquitetura subjacente desta aplicação, é importante perceber as funções e eventos reativos envolvidos.

O **UI** desempenha um papel crucial na definição da *interface* do utilizador da aplicação. Nele, são especificados os elementos de entrada e saída que os utilizadores interagem diretamente. Entre esses elementos, destacam-se:

- **Número de processo, nome e idade do paciente e data:** Campos para inserção de informações relevantes.
- **Carregar CSV:** Seleção de opção para carregar pontos a partir de um ficheiro CSV.
- **Escolher ficheiro CSV:** Campo para seleccionar um ficheiro CSV, habilitado quando a opção "Carregar CSV" é seleccionada.
- **Idade gestacional e parâmetro a analisar:** Campos para inserção de dados numéricos.
- **Submeter:** Botão para submissão de dados.
- **Salvar Pontos em CSV:** *Checkbox* a seleccionar se o utilizador desejar salvar os pontos inseridos num ficheiro CSV.
- **Transferir CSV:** *Link* para transferência do ficheiro CSV.
- **Reiniciar Entradas:** Botão para redefinir as entradas.
- **Remover Ponto:** Botão para remover um ponto seleccionado no gráfico.
- **Transferir relatório:** *Link* para a transferência de um relatório PDF sobre o processo em estudo.

O servidor é responsável pela lógica da aplicação e pela manipulação de eventos reativos. Nele, encontramos:

- **Funções Personalizadas:** Incluem `addPoint()` que adiciona pontos inseridos manualmente aos gráficos, `findCentile()` que encontra o percentil para cada ponto adicionado e por fim, `update_plot()` que atualiza os gráficos sempre que mudanças são feitas.
- **Eventos Reativos:** São monitorizados por funções como `observeEvent()`, `reactive()`, `renderPlotly()` e `renderText()`, respondendo às mudanças nos dados de entrada e atualizando os dados de saída.
- **Funções de Atualização:** Permitem a modificação dinâmica dos elementos presentes na *interface* utilizando as seguintes funções: `updateNumericInput()`, `updateSelectInput()`, `updateDateInput()` e `updateActionButton()`.

- **Modalidades e transferências:** As funções `showModal()` e `removeModal()` são utilizadas para exibir mensagens modais interativas, enquanto `renderDataTable()` cria tabelas dinâmicas. O `downloadHandler()` permite a criação e transferência de relatórios em PDF e ficheiros em CSV.

Uma parte crucial implementada no servidor é um sistema de notificação eficaz para dar *feedback* em tempo real aos utilizadores. Em várias partes da aplicação, as notificações, utilizando `showNotification()`, são acionadas para orientar os utilizadores e garantir que os dados inseridos ou as ações realizadas estão conforme os requisitos estabelecidos. A seguir, são descritas as situações em que as notificações são apresentadas aos utilizadores.

Ao adicionar um novo ponto aos dados, a aplicação verifica a validade dos valores de entrada. Se algum campo estiver em branco, uma notificação de erro é exibida. Além disso, se o valor da **IG** não for um número inteiro entre 12 e 41 semanas, uma notificação de erro específica é apresentada, direcionando o utilizador a inserir um valor no intervalo aceitável. Dependendo do tipo de medida inserida (diâmetro, peso ou espessura), limites específicos são definidos para os valores inseridos, e notificações de erro correspondentes são exibidas se os valores estiverem fora desses limites.

Quando os utilizadores carregam um ficheiro CSV na aplicação, a integridade do ficheiro é verificada. Se o ficheiro não tiver exatamente duas colunas numéricas, uma notificação de erro é apresentada para indicar que o formato do ficheiro deve ser corrigido. Além disso, se os valores nas colunas da **IG** ou do parâmetro não corresponderem a valores desejáveis, uma notificação de erro também será exibida.

Durante o cálculo dos rácios entre pontos selecionados (**PP** e **PF**), as notificações são usadas para garantir a validade dos dados. Se os valores de **IG** selecionados não forem iguais, uma notificação de erro será apresentada. Da mesma forma, se mais ou menos de uma linha for selecionada em qualquer tabela, uma notificação de erro também será exibida.

Para melhorar a experiência do utilizador, as notificações também são usadas para dar *feedback* positivo. Por exemplo, quando os utilizadores criam um relatório em PDF, uma notificação de sucesso é apresentada para informar que o relatório foi gerado com êxito ou quando um ponto é adicionado ao gráfico ou um ficheiro CSV é carregado ou descarregado.

Em suma, a interação entre o **UI** e o servidor é facilitada pelos eventos reativos, que monitorizaram as mudanças nos elementos de entrada especificados no **UI**. Quando os elementos de entrada são modificados pelos utilizadores, os eventos reativos acionam as funções pertinentes no servidor, atualizando a *interface* do utilizador e fornecendo uma experiência interativa e dinâmica.

Capítulo 7

Resultados

Este capítulo marca o ponto culminante da investigação realizada nesta dissertação. Aqui, serão apresentados os resultados obtidos a partir das metodologias discutidas nos capítulos anteriores. A secção inicial aborda a análise exploratória dos dados, as secções seguintes apresentam uma análise mais aprofundada dos resultados da modelação das curvas de crescimento. Por fim, é apresentado um estudo de caso para demonstrar a utilização da aplicação *web* desenvolvida. É importante ressaltar que os resultados apresentados possuem um valor intrínseco significativo. Têm o potencial de orientar futuras investigações, contribuindo assim para o avanço contínuo da compreensão das curvas de crescimento dos parâmetros placentares e fetais.

Além disso, para facilitar o acesso à totalidade dos resultados, incluindo código elaborado e imagens produzidas, todos os recursos relevantes estão disponíveis no seguinte repositório [GitHub](https://github.com/daniellalemos/PlacentalGrowth):

```
git clone https://github.com/daniellalemos/PlacentalGrowth
```

7.1 Análise exploratória dos dados

A análise exploratória desempenha um papel fundamental na compreensão dos dados subjacentes ao desenvolvimento das curvas de crescimento para parâmetros placentares em função da **IG**. No Apêndice **B** é possível verificar uma breve descrição das variáveis exploradas neste estudo, dando assim um enquadramento para a sua subsequente análise.

O principal objetivo deste estudo passa pela exploração dos parâmetros placentares. No entanto, é imperativo ressaltar que, embora o foco principal recaia sobre as variáveis utilizadas para a construção das curvas de crescimento, a análise de outras variáveis como “local”, “Obstetriccontext” (paridade da mãe), “Fetalgender” (género do feto), “Placentalshape” (forma da placenta) e “MaternalAge” (idade materna) desempenham também um papel importante neste estudo e oferecem uma perceção do contexto e da natureza das variáveis envolvidas.

A variável local, Figura **10a**, apresenta uma ampla gama de regiões e hospitais de Portugal onde as

O histograma da idade materna, Figura 11a, indica a distribuição das idades das mães nas observações do estudo. Este histograma revela que a maioria das mães tem idades compreendidas entre 25 e 40 anos, com o intervalo de 30 a 35 anos sendo o mais frequente, com aproximadamente 630 observações. Nota-se também que a distribuição é assimétrica, com um pico significativo na faixa etária de 30 a 35 anos.

Quanto às variáveis efetivamente utilizadas neste estudo, é analisada a **IG**, a variável explicativa, e várias variáveis resposta, incluindo **D1**, **D2**, **PP**, **EP** e **PF**. A **IG** (12 a 41 semanas), representada no histograma da Figura 11b sugere uma distribuição do número de amostras por intervalo de **IG** assimétrica, tendo alguns intervalos com um número de amostras mais baixo.

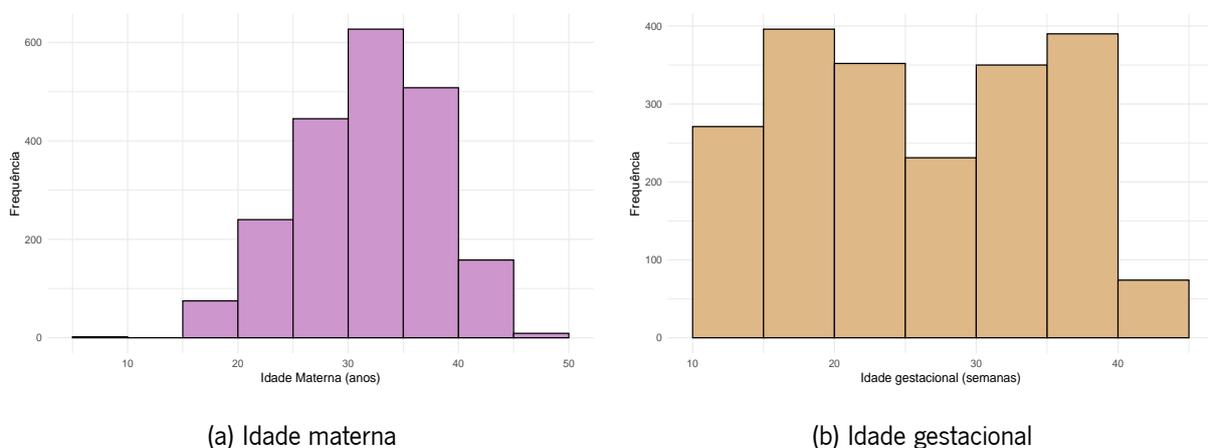


Figura 11: Histogramas para as variáveis idade materna e **IG**.

As restantes variáveis foram analisadas na Figura 12 por gráficos de dispersão, uma ferramenta eficaz para explorar a disposição das amostras e compreender o comportamento dessas variáveis, para assim ser possível encontrar o modelo que melhor as representa. Para estas variáveis também é possível consultar tabelas com as estatísticas descritivas como a dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose para cada **IG** (12–41 semanas).

Quanto ao **D1**, Figura 12a, o gráfico de dispersão revela uma associação linear positiva. À medida que a **IG** aumenta, há uma tendência geral de aumento no valor do **D1**. Isso sugere uma possível relação positiva entre estas duas variáveis, indicando que, em média, o **D1** tende a aumentar à medida que a gestação avança. Analisando a Tabela 3, é possível confirmar a relação positiva entre a **IG** e o **D1** visualizada no gráfico 12a, indicando um desenvolvimento progressivo da placenta. Por exemplo, na semana 12, a média do **D1** foi de $6,07 \pm 1,65$ cm e esse valor aumentou gradualmente para $19,07 \pm 3,92$ cm na semana 41. Esse aumento substancial no **D1** ao longo da gestação é acompanhado por um aumento leve na variabilidade dos dados, conforme indicado pelos valores do **DP**. A análise da mediana

complementa essas observações, mostrando uma tendência de aumento similar. É notável que os valores de assimetria permanecem próximos de zero na maioria das **IG**, sugerindo distribuições aproximadamente simétricas dos dados. Isso indica que o crescimento do **D1** é equilibrado. Adicionalmente, a maioria dos valores de curtose está acima de 3, refletindo caudas mais pesadas (leptocurtose) que a distribuição Normal.

Tabela 3: Estatísticas descritivas do **D1** em cada **IG**. A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.

GA	N	Média	DP	Min	Max	Mediana	Assimetria	Curtose
12	54	6,07	1,65	2,00	12,00	6,00	0,42	5,10
13	64	6,57	1,52	4,00	9,00	7,00	-0,10	1,98
14	73	7,11	1,63	4,00	12,00	7,00	0,42	3,15
15	80	7,57	1,44	4,00	12,00	8,00	0,14	3,12
16	82	8,12	1,48	4,00	11,00	8,00	-0,46	3,64
17	76	9,41	1,55	6,00	14,00	9,00	0,45	3,82
18	80	9,20	1,73	4,00	17,00	9,00	0,65	7,59
19	82	10,06	1,95	5,00	14,00	10,00	-0,16	2,42
20	76	11,10	1,81	7,00	17,00	11,00	0,19	3,94
21	78	11,76	1,91	6,00	18,00	12,00	0,25	3,84
22	75	11,71	2,24	5,50	18,00	12,00	-0,03	3,26
23	80	12,66	2,38	8,00	22,00	13,00	0,84	4,97
24	70	12,84	2,32	9,00	19,00	13,00	0,75	3,67
25	49	12,61	2,48	7,00	18,50	13,00	-0,17	2,70
26	43	13,01	2,99	8,50	24,00	12,00	1,51	6,28
27	36	14,75	2,73	9,50	23,00	14,00	1,34	5,02
28	61	13,88	2,68	9,00	25,50	14,00	1,21	7,00
29	44	14,17	2,28	10,00	19,00	14,00	0,07	2,30
30	47	14,68	2,89	8,50	22,50	14,50	0,41	3,58
31	57	15,63	3,63	8,50	29,00	16,00	0,94	5,32
32	70	16,82	4,15	11,00	31,00	16,00	1,39	4,96
33	74	16,46	3,63	11,00	29,00	16,00	1,20	4,79
34	73	17,20	3,66	11,50	29,00	17,00	1,36	5,03
35	76	18,60	4,31	11,00	32,00	17,75	0,97	3,71
36	80	17,27	2,74	12,50	27,00	17,00	0,94	4,69
37	77	18,23	3,17	12,00	29,00	18,00	1,01	4,71
38	77	18,23	3,71	13,00	30,00	17,00	1,23	4,10
39	75	18,03	3,14	13,00	26,00	18,00	0,52	2,76
40	81	18,53	2,59	10,00	25,00	18,00	0,02	3,82
41	74	19,07	3,92	12,50	32,00	18,00	1,42	4,95

No gráfico de dispersão para o **D2** em relação à **IG**, Figura 12b, observa-se uma tendência semelhante ao **D1**. À medida que a **IG** aumenta, há uma tendência geral de aumento no valor de **D2**. Isso sugere uma associação linear entre as variáveis, indicando que o **D2** tende a aumentar à medida que a gestação progride. As estatísticas descritivas para o **D2** são apresentadas na Tabela 4. Ao analisar os dados,

uma tendência clara se destaca: à medida que a **IG** avança, o **D2** aumenta progressivamente. Esse aumento progressivo é coerente com o crescimento contínuo da placenta para acomodar as crescentes necessidades do feto em desenvolvimento. Por exemplo, na semana 12, a média do **D2** foi de $4,37 \pm 1,59$ cm, e esse valor aumentou gradualmente para $16,59 \pm 3,03$ cm na semana 41. Esse aumento notável é acompanhado por um ligeiro aumento na variabilidade dos dados, indicado pelo **DP**. A análise da mediana indica uma tendência de aumento similar. Quanto à assimetria, é possível verificar que os seus valores permaneçam próximos de zero na maioria das **IG**, sugerindo distribuições aproximadamente simétricas dos dados. Adicionalmente, a maioria dos valores de curtose está entre 2 e 5, o que indica que essas distribuições são mais concentradas no centro do que a distribuição normal.

Tabela 4: Estatísticas descritivas do **D2** em cada **IG**. A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.

GA	N	Média	DP	Min	Max	Mediana	Assimetria	Curtose
12	54	4,37	1,59	2,00	8,00	4,00	0,69	2,81
13	64	5,02	1,55	1,50	9,00	5,00	0,44	3,17
14	73	5,18	1,29	2,00	8,00	5,00	-0,07	2,93
15	80	5,74	1,48	2,50	10,00	6,00	0,02	2,70
16	82	6,26	1,58	2,00	9,00	6,00	-0,32	2,74
17	76	7,41	1,70	4,00	12,00	7,00	0,13	2,47
18	80	7,27	1,72	2,00	11,00	7,00	-0,25	2,82
19	82	8,24	1,91	3,00	12,50	8,00	-0,14	3,09
20	76	8,80	1,76	4,00	12,50	9,00	-0,10	2,76
21	78	9,53	1,80	4,00	13,00	10,00	-0,47	3,16
22	75	9,57	2,09	5,00	15,00	10,00	-0,09	2,72
23	80	10,78	2,05	5,00	16,00	11,00	-0,19	3,09
24	70	10,84	2,15	7,00	18,00	10,75	0,61	3,76
25	49	10,55	2,28	6,00	17,00	11,00	0,04	2,96
26	43	10,87	2,35	6,00	16,00	11,00	0,25	2,48
27	36	12,69	2,47	8,00	19,00	12,00	0,92	3,59
28	61	11,62	2,46	7,00	18,00	12,00	0,17	2,71
29	44	12,30	2,18	8,00	16,50	12,00	0,04	1,85
30	47	12,42	2,73	6,00	21,00	12,00	0,47	4,24
31	57	13,19	3,35	6,00	25,00	13,00	1,08	5,46
32	70	14,29	3,16	8,00	25,00	13,25	1,29	5,13
33	74	13,81	2,93	8,00	25,00	14,00	0,83	4,98
34	73	15,23	3,49	10,00	27,00	14,00	1,57	5,26
35	76	15,72	2,92	8,00	24,00	15,00	0,56	3,63
36	80	15,14	2,75	10,00	28,00	15,00	1,41	7,91
37	77	15,20	2,67	11,50	30,00	15,00	2,73	14,64
38	77	15,69	2,70	10,00	24,00	15,00	0,96	4,25
39	75	15,54	2,48	12,00	23,00	15,00	0,94	3,87
40	81	16,25	2,04	9,00	21,50	16,50	-0,34	4,05
41	74	16,59	3,03	10,00	26,50	16,00	0,83	4,50

Relativamente à **EP**, o gráfico de dispersão (Figura 12c) mostra uma relação menos clara com a **IG**. Embora haja alguma variação, não é evidente uma tendência de aumento ou diminuição da **EP**. Isso pode indicar que a espessura da placenta é influenciada por outros fatores além da **IG**. No que diz respeito às estatísticas descritivas presentes na Tabela 5, é possível verificar que existe um aumento na **EP** ao longo da **IG**, no entanto, este aumento não é tão significativo como para o **D1** ou o **D2**. Uma vez que a **EP** aumentou até às 41 semanas, aproximadamente 110% do seu valor inicial médio, em contraste com o **D2** que aumentou 279%. Os valores do **DP** indicam também uma variabilidade menor comparativamente às variáveis já analisadas. Quanto à assimetria, os dados são aproximadamente simétricos e apresentam leptocurtose indicado pelos valores maiores do que 3 de curtose e por isso, apresentam caudas pesadas.

Tabela 5: Estatísticas descritivas da **EP** em cada **IG**. A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.

GA	N	Média	DP	Min	Max	Mediana	Assimetria	Curtose
12	54	1,26	0,46	0,50	2,80	1,20	0,99	4,65
13	64	1,28	0,40	0,50	2,50	1,20	0,43	3,50
14	73	1,41	0,39	0,50	2,20	1,50	0,04	2,57
15	80	1,56	0,43	0,70	3,00	1,50	0,88	4,70
16	82	1,66	0,52	0,50	3,50	1,50	0,95	5,30
17	76	1,76	0,47	0,50	3,00	1,73	0,28	3,43
18	80	1,87	0,60	1,00	4,50	1,70	1,75	7,23
19	82	1,87	0,63	1,00	4,00	1,80	0,99	4,06
20	76	1,78	0,40	0,75	3,00	1,80	0,21	3,39
21	78	1,95	0,46	0,30	3,00	2,00	-0,49	3,92
22	75	1,88	0,52	1,00	3,00	1,90	0,48	2,88
23	80	1,97	0,48	1,20	3,50	2,00	0,85	3,46
24	70	1,94	0,55	1,20	4,00	2,00	1,71	6,94
25	49	2,04	0,55	0,90	3,50	2,00	0,83	3,80
26	43	2,17	0,62	1,10	4,00	2,00	0,66	3,58
27	36	2,11	0,65	1,00	4,00	2,00	1,52	5,48
28	61	2,23	0,58	1,50	4,00	2,00	0,88	3,46
29	44	2,31	0,61	1,20	4,00	2,25	0,79	4,02
30	47	2,17	0,54	0,80	4,00	2,00	0,67	5,49
31	57	2,43	0,69	1,50	5,00	2,30	1,38	5,47
32	70	2,30	0,67	1,00	5,00	2,25	1,75	7,78
33	74	2,54	0,67	1,25	5,00	2,50	0,98	4,65
34	73	2,48	0,60	1,00	4,50	2,50	0,34	4,24
35	76	2,44	0,59	1,00	4,00	2,50	0,88	3,85
36	80	2,44	0,52	1,50	3,70	2,50	0,06	2,59
37	77	2,50	0,70	1,50	5,00	2,50	1,12	4,24
38	77	2,63	0,64	1,50	5,00	2,50	0,84	4,38
39	75	2,65	0,68	1,50	5,00	2,50	0,73	3,77
40	81	2,66	0,64	1,50	5,00	2,50	1,00	4,21
41	74	2,64	0,56	1,25	4,00	2,50	0,02	2,89

No gráfico de dispersão para o **PP** (Figura 12d), há uma tendência de aumento no **PP** à medida que a **IG** avança, sugerindo uma relação positiva. Com base nas estatísticas descritivas do **PP** (Tabela 6), é possível verificar que a média do **PP** aumenta significativamente com a **IG**, com valores oscilando entre $35,79 \pm 23,39$ g e $474,42 \pm 107,05$ g. Além disso, o **DP**, varia de forma considerável nos diferentes grupos de **IG**, refletindo níveis variados de variabilidade, especialmente nos estágios finais da gestação. Os valores mínimo e máximo do **PP** também mostram variações notáveis, como, por exemplo, nas 38 semanas pode variar entre 240g e 995g. Quanto à assimetria e curtose, esses valores também variam, indicando diferentes assimetrias e formas das distribuições dos dados. Alguns grupos apresentam distribuições muito assimétricas e um elevado grau de leptocurtose como é o caso das 17 semanas.

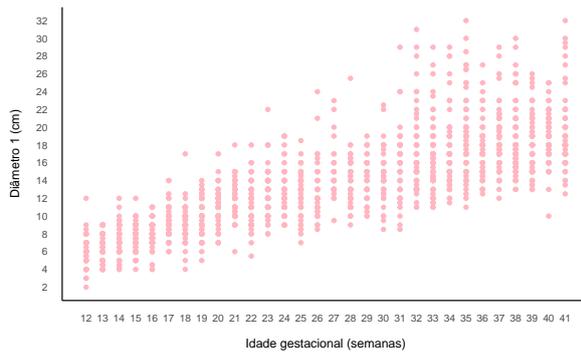
Tabela 6: Estatísticas descritivas do **PP** em cada **IG**. A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.

GA	N	Média	DP	Min	Max	Mediana	Assimetria	Curtose
12	54	35,79	23,39	6,0	143	33,0	2,19	9,99
13	64	41,64	16,92	8,0	109	42,0	0,77	5,55
14	73	51,12	17,84	12,3	90	49,0	0,28	2,48
15	80	59,80	22,32	16,0	148	57,0	0,98	5,25
16	82	71,22	24,14	12,0	145	70,0	0,26	3,86
17	76	94,13	37,66	33,0	300	88,5	2,47	13,88
18	80	95,46	31,43	32,0	250	95,5	1,46	8,76
19	82	113,24	35,99	24,0	214	113,0	0,06	3,06
20	76	126,98	44,19	36,0	360	120,5	1,93	11,88
21	78	140,67	37,14	42,0	230	143,0	-0,12	3,27
22	75	143,86	39,01	64,0	302	140,0	0,94	5,64
23	80	171,16	43,22	52,0	327	172,5	0,14	4,76
24	70	169,29	55,58	73,0	398	153,0	1,28	5,80
25	49	180,57	70,16	73,0	343	185,0	0,25	2,33
26	43	190,67	71,35	85,0	379	190,0	0,51	2,70
27	36	216,25	73,59	72,0	466	220,0	0,65	5,13
28	61	206,80	68,81	86,0	355	203,0	0,37	2,44
29	44	254,61	88,80	123,0	510	241,5	0,70	3,12
30	47	249,10	77,00	88,0	435	247,0	0,22	2,89
31	57	280,81	90,19	94,0	496	294,0	-0,00	2,78
32	70	310,73	89,20	148,0	582	305,5	0,83	3,84
33	74	334,62	88,02	153,0	524	328,0	0,25	2,55
34	73	344,96	79,02	163,0	580	345,0	0,29	3,29
35	76	373,38	123,47	177,0	890	355,0	1,42	6,12
36	80	360,07	79,82	224,0	590	349,5	0,71	3,11
37	77	378,03	89,35	222,0	682	360,0	1,00	4,27
38	77	428,03	135,38	240,0	995	404,0	1,52	6,22
39	75	445,96	126,36	255,0	905	435,0	0,99	4,31
40	81	473,92	89,09	301,0	863	458,0	1,28	6,60
41	74	474,42	107,05	290,0	738	459,5	0,42	2,62

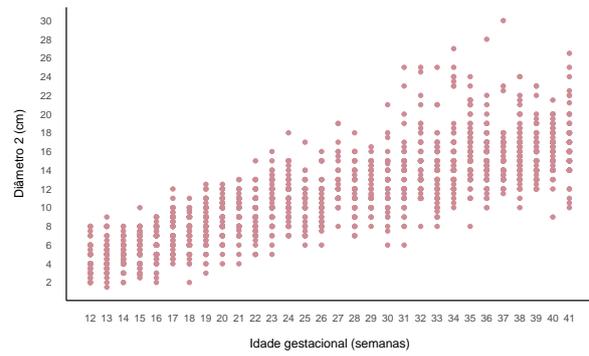
Por fim, o **PF**, representado na 12e, mostra uma tendência clara de aumento à medida que a **IG** avança. Isso é consistente com o desenvolvimento progressivo normal do feto. Observando a Tabela 7, em geral, verificou-se também um aumento notável. Na semana 12, o **PF** apresentou uma média de $23,77 \pm 38,60$ g. À medida que a gestação progride, essa média aumenta significativamente, atingindo $3301,55 \pm 503,92$ g na semana 41. Além disso, a mediana também segue essa tendência de aumento, indicando uma consistência desses resultados. A análise da assimetria mostra que, em geral, as distribuições dos dados são assimétricas devido a valores maiores que zero. Isso sugere uma maior concentração de valores menores relativamente à média, o que é coerente com o crescimento gradual do feto. A curtose, também apoia essa observação, mostrando vários valores superiores a 3.

Tabela 7: Estatísticas descritivas do **PF** em cada **IG**. A tabela inclui os valores de dimensão da amostra (N), média, desvio padrão (DP), mínimo (Min), máximo (Max), mediana, assimetria e curtose.

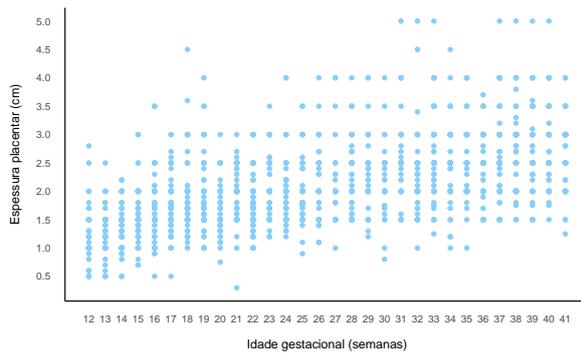
GA	N	Média	DP	Min	Max	Mediana	Assimetria	Curtose
12	54	23,77	38,60	5,70	281,00	16,25	5,84	38,42
13	64	28,72	20,61	6,00	173,00	27,00	5,44	38,84
14	73	45,15	16,87	10,50	92,00	46,00	0,13	2,81
15	80	68,09	32,09	12,00	193,40	70,50	0,57	4,53
16	82	100,71	43,87	14,00	191,00	98,35	0,11	2,39
17	76	164,23	52,08	33,00	315,00	168,00	-0,20	3,76
18	80	204,88	57,37	63,00	335,00	209,50	-0,32	3,53
19	82	259,97	84,90	29,00	449,00	269,00	-0,75	3,88
20	76	329,46	78,18	77,00	571,00	328,00	-0,34	4,77
21	78	401,82	83,96	110,00	567,00	409,50	-0,98	4,65
22	75	484,57	139,03	146,00	1305,00	489,00	2,38	17,93
23	80	558,34	104,50	313,00	833,00	562,50	-0,16	3,49
24	70	637,01	143,01	319,00	1118,00	642,50	0,05	4,11
25	49	673,89	334,22	135,00	2310,00	732,00	2,15	12,92
26	43	710,84	293,23	356,00	1880,00	640,00	2,23	9,24
27	36	910,90	189,81	470,00	1300,00	954,50	-0,46	3,02
28	61	934,85	267,50	297,00	1500,00	980,00	-0,32	2,85
29	44	1213,39	486,86	426,00	3100,00	1102,50	1,43	6,58
30	47	1267,34	349,70	540,00	2300,00	1275,00	0,34	3,33
31	57	1483,28	356,88	630,00	2380,00	1470,00	-0,09	3,02
32	70	1749,09	482,69	402,00	3765,00	1740,00	0,90	6,96
33	74	1928,49	388,52	755,00	3330,00	1960,00	0,14	4,99
34	73	2133,40	443,35	1190,00	3000,00	2100,00	-0,08	2,21
35	76	2334,57	558,02	1460,00	4470,00	2270,00	1,33	5,81
36	80	2456,01	453,46	1535,00	3420,00	2445,00	0,16	2,49
37	77	2574,39	515,95	1730,00	4780,00	2450,00	2,14	9,73
38	77	2826,90	576,36	1830,00	4800,00	2720,00	1,02	4,26
39	75	2890,00	584,18	1520,00	4535,00	2830,00	0,19	2,94
40	81	3245,64	468,57	2280,00	4745,00	3230,00	0,09	3,06
41	74	3301,55	503,92	2280,00	4880,00	3312,50	0,28	3,04



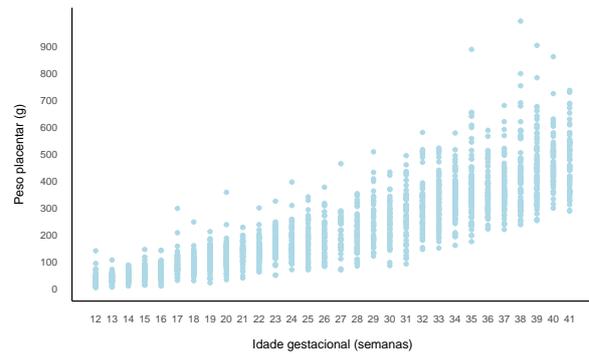
(a) Diâmetro 1



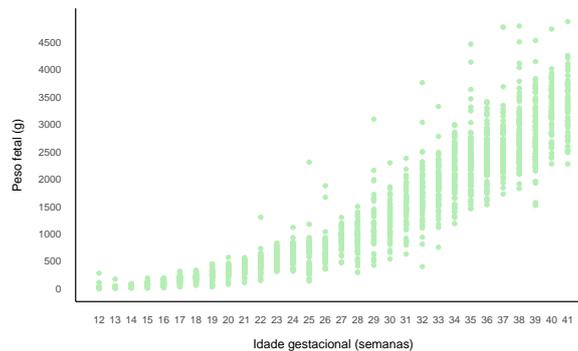
(b) Diâmetro 2



(c) Espessura placentar



(d) Peso placentar



(e) Peso fetal

Figura 12: Gráficos de dispersão representando a relação entre a **IG** e vários parâmetros (**D1**, **D2**, **PP**, **EP** e **PF**). Cada ponto nos gráficos representa uma observação individual de uma amostra de 2064 placentas.

Foi também realizada uma avaliação da normalidade das variáveis, que serão utilizadas para construir as curvas de crescimento, **D1**, **D2**, **PP**, **EP** e **PF**. A normalidade é uma suposição fundamental em muitas análises estatísticas, e a sua validade desempenha um papel crucial na interpretação dos resultados. Na Figura 13 é possível visualizar os gráficos Q-Q e na Tabela 8 estão disponíveis os resultados do teste de Shapiro-Wilk, com um nível de significância de 0,05, para a normalidade das variáveis. Os resultados sugerem que as distribuições de todas as variáveis não seguem uma distribuição normal, como indicado

por desvios visuais nos gráficos Q-Q e valores-p significativamente baixos no teste de Shapiro-Wilk, $< 2 \times 10^{-16}$ para todas as variáveis, exceto o **D2** que teve um valor-p de 6×10^{-15} . Logo, há uma rejeição clara da normalidade. É importante destacar que qualquer desvio da normalidade pode ter implicações nas análises estatísticas subsequentes, e, portanto, a sua consideração é fundamental para uma interpretação precisa dos resultados deste estudo.

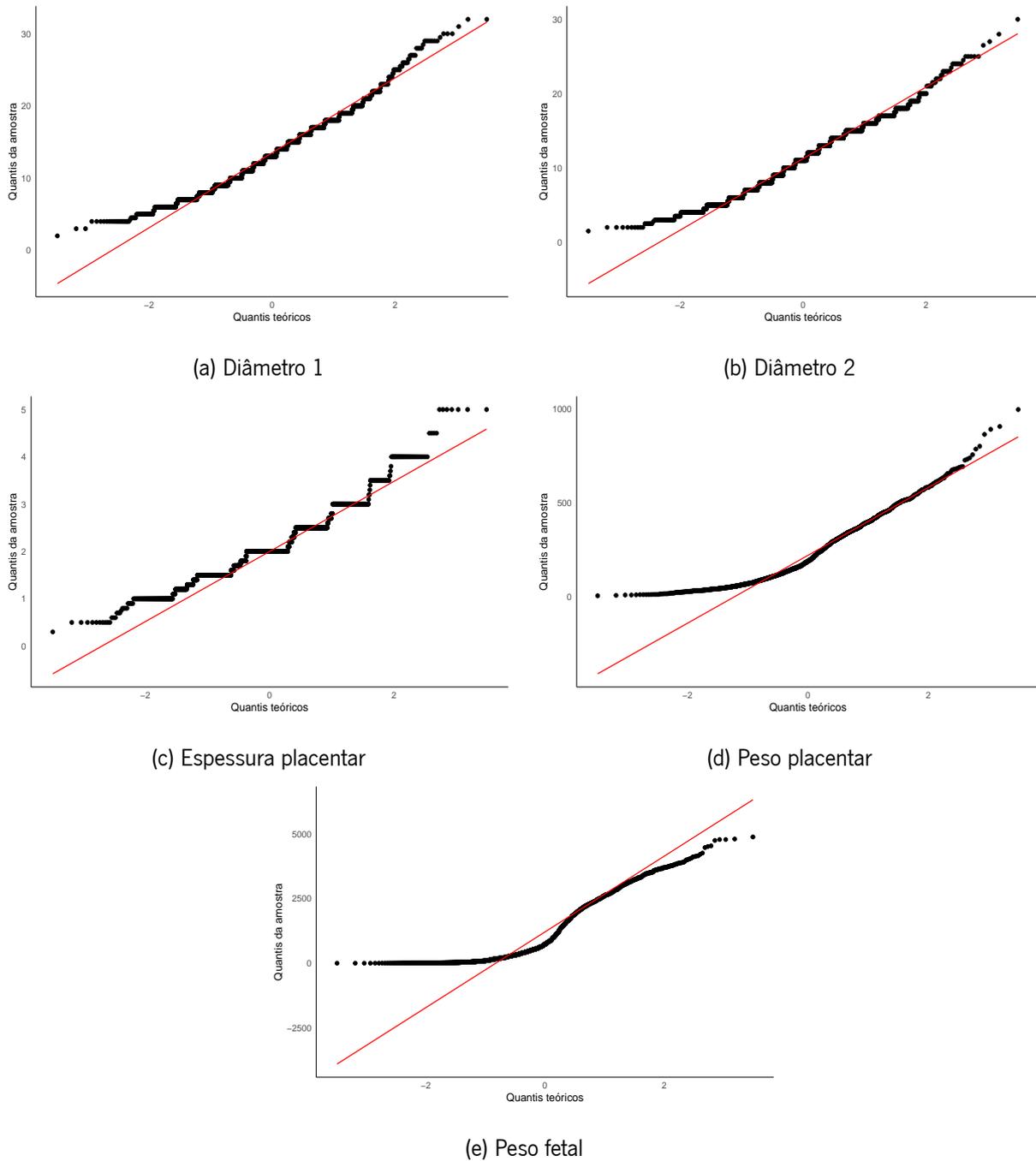


Figura 13: Gráficos Q-Q para a avaliação de normalidade de **D1**, **D2**, **PP**, **EP** e **PF**.

Tabela 8: Resultados do teste de Shapiro-Wilk para a avaliação da normalidade de **D1**, **D2**, **PP**, **EP** e **PF**.

Variável	Valor-p
Diâmetro 1	$< 2 \times 10^{-16}$
Diâmetro 2	6×10^{-15}
Espessura placentar	$< 2 \times 10^{-16}$
Peso placentar	$< 2 \times 10^{-16}$
Peso fetal	$< 2 \times 10^{-16}$

7.2 Testes de significância entre gêneros

Na análise estatística realizada visando identificar diferenças estatisticamente significativas entre as médias dos parâmetros **D1**, **D2**, **PP**, **EP** e **PF** para fetos femininos e masculinos em cada **IG**, foi aplicado o teste-t de Student, com um nível de significância de 0,05.

Os valores-p resultantes de cada teste-t podem ser encontrados no Apêndice C. Estes valores têm como propósito indicar se as divergências nas médias das variáveis entre os grupos de fetos femininos e masculinos possuem relevância estatística, ou seja, se o valor-p é inferior a 0,05. Além disso, para uma representação visual dos valores médios e os seus intervalos de confiança de 95%, é possível consultar a Figura 14.

Os resultados obtidos apresentam diferentes graus de significância. Em certas **IG** e variáveis, foram constatadas diferenças estatisticamente significativas entre os gêneros feminino e masculino. Por exemplo, para o **D1**, observou-se uma diferença estatisticamente significativa na semana 39 (valor-p = 0,04). Já para o **D2**, verificaram-se tais diferenças nas semanas 12 (valor-p = 0,009), 14 (valor-p = 0,03) e 39 (valor-p = 0,006). A **EP** apresentou diferenças estatisticamente significativas nas semanas 14 (valor-p = 0,02), 20 (valor-p = 0,04) e 40 (valor-p = 0,02). Por fim, o **PF** revelou diferenças significativas nas semanas 16 (valor-p = 0,009) e 40 (valor-p = 0,002). Importante mencionar que o **PP** não apresentou diferenças significativas em nenhuma **IG**.

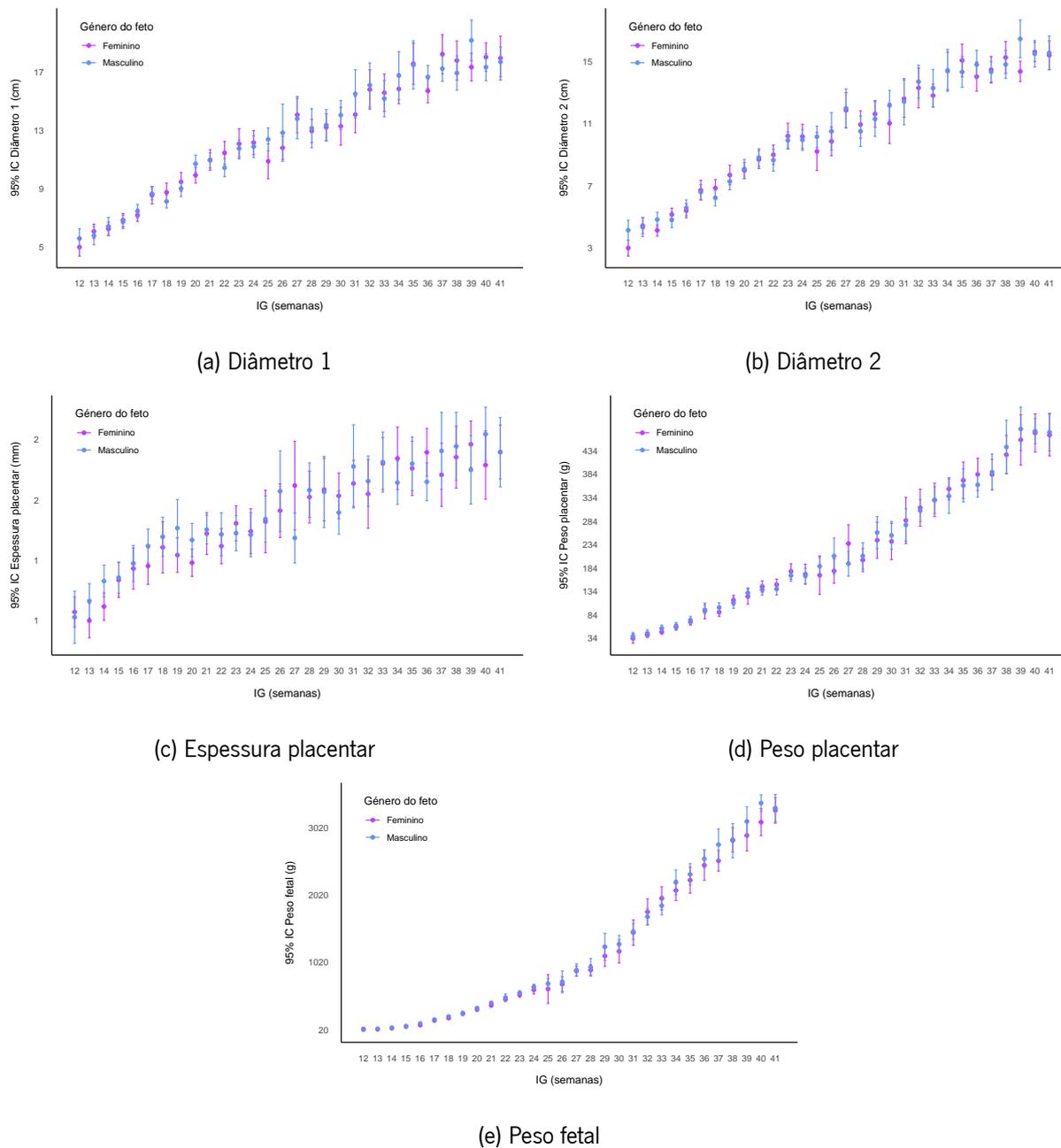


Figura 14: Os valores médios e os seus correspondentes intervalos de confiança de 95% para os parâmetros, **D1**, **D2**, **PP**, **EP** e **PF**, são apresentados com base no gênero fetal em cada **IG**.

7.3 Curvas de crescimento

Neste estudo, foi analisada a relação entre vários parâmetros como, **D1**, **D2**, **PP**, **EP** e **PF**, com a **IG** para produzir curvas de crescimento de percentis de referência ($\alpha = 3, 10, 25, 50, 75, 90, 97$). A amostra foi composta por 2064 placentas.

A análise das curvas de crescimento permitirá uma visão mais abrangente das mudanças observadas

em diferentes **IG**, contribuindo para uma compreensão mais profunda do desenvolvimento placentar e fetal na população em estudo. A representação visual dessas curvas, com os valores numéricos associados, oferece uma visão completa das tendências ao longo do tempo, enriquecendo assim a base de conhecimento para futuras investigações na área da saúde materno-fetal.

Devido à extensão da análise dos critérios de avaliação, que inclui um grande número de figuras, foi decidido incluir apenas neste documento as figuras relacionadas à função de alisamento escolhida e aos dois melhores modelos finais (combinação de função de alisamento e distribuição), sendo que o segundo melhor encontra-se nos apêndices. As demais figuras, que abrangem todos os critérios de qualidade analisados, estão disponíveis no repositório do [GitHub](#) para consulta.

7.3.1 Diâmetro 1

Para a construção das curvas de crescimento do parâmetro **D1**, foi inicialmente conduzida uma avaliação do desempenho de diferentes funções de alisamento, sem considerar nenhuma distribuição associada ao modelo. A Tabela 9 apresenta métricas que permitem classificar o desempenho dos modelos testados. As *P-splines* indicaram um desempenho notável com um valor de **R²** de 0,686, indicando um bom ajuste aos dados. O valor correspondente de **MSE** foi 4,877, revelando um desvio mínimo entre os valores previstos e observados. Os valores de **AIC** e **BIC** foram de 10044 e 10088, respectivamente, sugerindo um bom ajuste do modelo. As *splines* cúbicas também produziram resultados satisfatórios, com um **R²** de 0,684, **AIC** de 10048 e **BIC** de 10082. As redes neurais, por sua vez, mostraram desempenho comparável, alcançando um valor de **R²** de 0,686, embora acompanhado por um **MSE** ligeiramente maior de 4,901, indicando um pequeno aumento nos erros de previsão em comparação com outros métodos. O valor de **AIC** foi 10064, e o valor de **BIC** foi 10166. A função de alisamento *Loess* exibiu um valor de **R²** de 0,684, um **MSE** de 4,877 e valores de **AIC** e **BIC** de 10050 e 10082, respectivamente.

Para uma avaliação mais aprofundada da qualidade de ajuste, foram utilizados critérios de diagnóstico gráfico, como *worm plots*, estatísticas Z e gráficos de resíduos. Um exemplo dessas análises pode ser observado na Figura 15, que se refere à função de alisamento *P-splines*. O *worm plot*, Figura 15a, evidencia a presença de muitos pontos fora dos intervalos de confiança de 95%, sugerindo que o ajuste poderá ser inadequado. Além disso, a forma em S indica a existência de leptocurtose. Quanto às estatísticas Z (Figura 15b) também apontam para possíveis inadequações no modelo, uma vez que muitos valores de $|Z|$, especialmente os direcionados para a variância (Z2), assimetria (Z3) e curtose (Z4), são maiores do que 2, indicado pelos quadrados nos círculos. Observando os gráficos dos resíduos (Figura 15c), eles indicam um comportamento aceitável no gráfico em relação aos índices, mas no gráfico relativo aos

valores ajustados, é possível verificar um ajustamento em linhas verticais ao longo do gráfico, sugerindo que, à medida que os valores ajustados aumentam ou diminuem, os resíduos mudam abruptamente e formam padrões verticais distintos. Quanto à normalidade, esta é facilmente rejeitada, uma vez que há um afastamento sistemático de uma relação linear entre os resíduos ordenados observados e os seus valores esperados aproximados. Relativamente às restantes funções de alisamento estudadas, *Splines* cúbicas, redes neurais e *Loess*, o seu comportamento nas análises de qualidade de ajuste foi semelhante, não apresentando diferenças significativas.

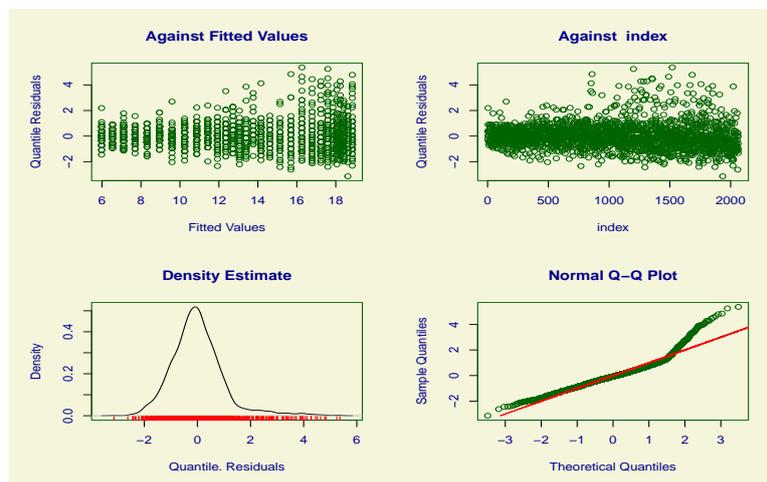
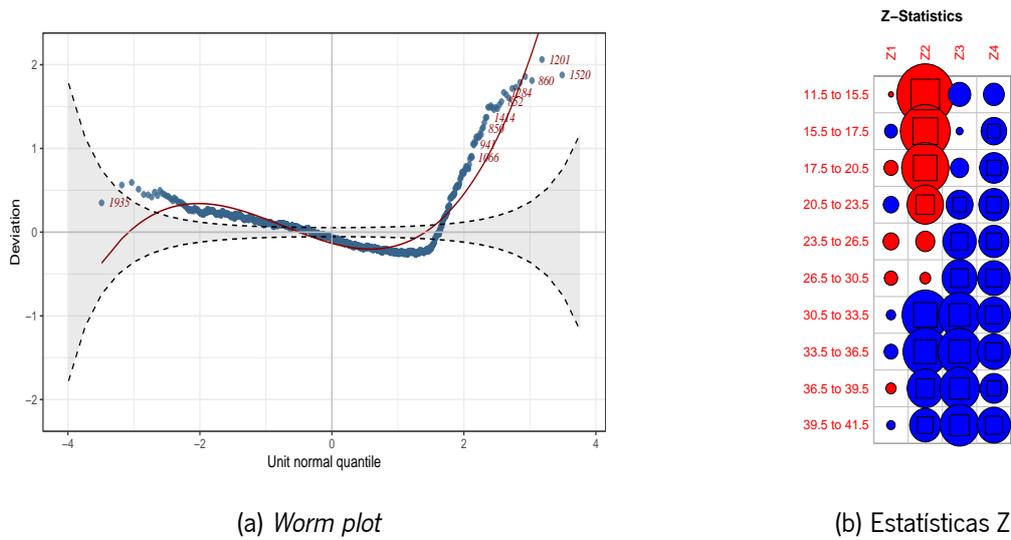


Figura 15: Critérios de avaliação de qualidade de ajuste (*worm plot*, estatísticas Z e gráficos de resíduos) para o modelo **GAMLSS** do **D1** com apenas a função de alisamento *P-splines*.

Após a seleção da função de alisamento mais apropriada, diferentes distribuições foram testadas, conforme apresentado na Tabela 9. A adaptação de um modelo **GAMLSS** com as distribuições **BCCG**, **BCT**, **BCPE** e **JSU** indicou, em geral, bom desempenho. A distribuição **BCT** teve o maior valor de R^2 , 0,755, seguida de perto pela distribuição **BCPE** com um R^2 de 0,754. Estas distribuições também apresentaram valores relativamente baixos de **MSE** de 4,593 e 4,592, valores de **AIC** e **BIC** de 9439 e 9529 para **BCT** e 9437 e 9527 para **BCPE**. Estes valores apontam que estas distribuições oferecem bom ajustes. Por outro lado, as distribuições **BCCG** e **JSU** parecem ser as menos promissoras das quatro, com valores menores de R^2 de 0,751 e 0,750, um **MSE** de 4,607 e 4,610 e valores de **AIC** e **BIC** de 9470 e 9553 para **BCCG** e 9479 e 9557 para **JSU**.

Ao analisar as técnicas de diagnóstico gráficas, é possível verificar que alguns modelos se ajustaram substancialmente melhor aos dados. O modelo **BCT**, representado na Figura 16, apresentou um ajuste particularmente promissor. Observando o *worm plot* é notável um bom ajuste aos dados, uma vez que os pontos e linha de ajuste encontram-se muito próximos da linha de referência horizontal e mais de 95% desses pontos encontram-se entre os intervalos de confiança. As estatísticas Z também apontaram para um bom ajuste global, com apenas um caso de $|Z| > 2$ no intervalo de 39,5 e 41,5 semanas para Z4. Quanto aos resíduos, eles indicam um comportamento aceitável no gráfico em relação aos índices, mas no gráfico relativo aos valores ajustados, existe um ajustamento em linhas verticais ao longo do gráfico, indicando que a variabilidade dos erros não é constante ao longo da faixa de valores previstos. É possível verificar uma suposição de normalidade nos dois gráficos inferiores dos resíduos. Em contraste, os restantes modelos não apresentaram um ajuste tão favorável, principalmente **BCCG** onde o *worm plot* apresentou uma curva em forma de S com a esquerda curvada para baixo, indicando leptocurtose. As estatísticas Z também mostraram um ajuste inadequado, com cinco situações em que $|Z| > 2$. Quanto aos resíduos apresentam-se como aproximadamente normais. **BCPE** e **JSU**, apresentaram resultados semelhantes a **BCT**, no entanto, não aparentaram um ajuste tão adequado.

Ao ajustar os restantes métodos, **LMS**, **LMST** e **LMSP** os resultados também se mostraram promissores. **LMST** teve o maior valor de R^2 , 0,755, igualando o valor de **GAMLSS** com **BCT**. Também mostrou um valor baixo de **MSE**, 4,595, **AIC**, 9441, e **BIC**, 9536. **LMSP** apresentou um bom valor de R^2 , 0,754, e valores baixos para **MSE**, **AIC** e **BIC** de 4,595, 9439 e 9534, respetivamente. **LMS** teve os valores menos favoráveis com um valor de R^2 de 0,751, **MSE**, 4,607, **AIC**, 9470, e **BIC**, 9553. Passando à análise visual, **LMST** mostrou resultados praticamente iguais aos do método **GAMLSS** com **BCT**, como pode ser verificado no Apêndice D. Indicando assim, um ajuste consistente. Quanto a **LMS** e **LMSP** apresentaram resultados semelhantes a **BCCG** e **BCPE** no **GAMLSS**.

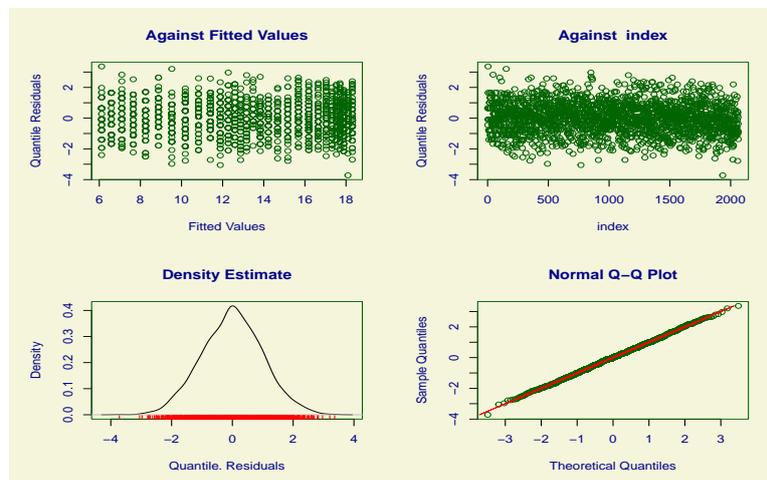
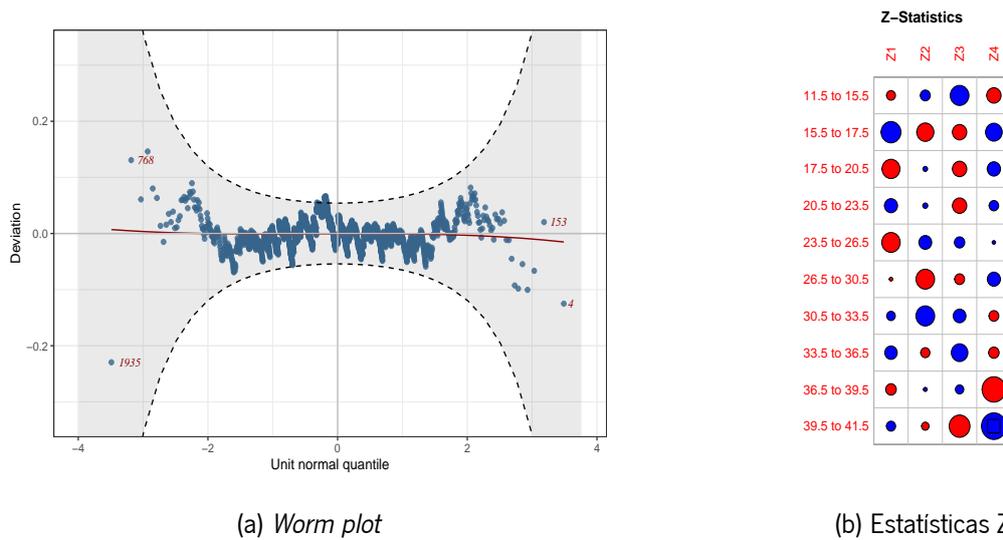


Figura 16: Critérios de avaliação de qualidade de ajuste (*worm plot*, estatísticas Z e gráficos de resíduos) para o modelo **GAMLSS** do **D1** com a função de alisamento *P-splines* e a distribuição **BCT**.

As curvas de crescimento de percentis para o modelo mais promissor, **GAMLSS** com **BCT** e *P-splines*, está representada na Figura 17. É evidente que estas curvas capturaram de forma precisa o comportamento dos dados, representados pelos pontos em cinzento no gráfico. Os valores numéricos para cada percentil em cada **IG** estão disponíveis para consulta no Apêndice D. Como diagnóstico final do modelo, as percentagens da amostra localizadas no mesmo valor ou abaixo de cada curva ajustada estão apresentadas na Tabela 10. Notavelmente, a concordância próxima entre as percentagens nominais e da amostra para o modelo sugere um bom ajuste das curvas de percentis. As curvas de crescimento de percentis para o segundo melhor modelo tambémm pode ser visualizadas no Apêndice D, sendo visualmente muito semelhantes às do **GAMLSS**.

Tabela 9: Métricas de desempenho do modelo para o **D1** em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exibe o **R²**, o **MSE**, o **AIC** e o **BIC**.

	R²	MSE	AIC	BIC
Funções de alisamento				
P-splines	0,686	4,877	10044	10088
<i>Splines</i> cúbicas	0,684	4,877	10048	10082
Redes neurais	0,686	4,901	10064	10166
<i>Loess</i>	0,684	4,877	10050	10082
Distribuições				
GAMLSS (BCCG)	0,751	4,607	9470	9553
GAMLSS (BCT)	0,755	4,593	9439	9529
GAMLSS (BCPE)	0,754	4,592	9437	9527
GAMLSS (JSU)	0,750	4,610	9479	9557
LMS	0,751	4,607	9470	9553
LMST	0,755	4,595	9441	9536
LMSP	0,754	4,595	9439	9534

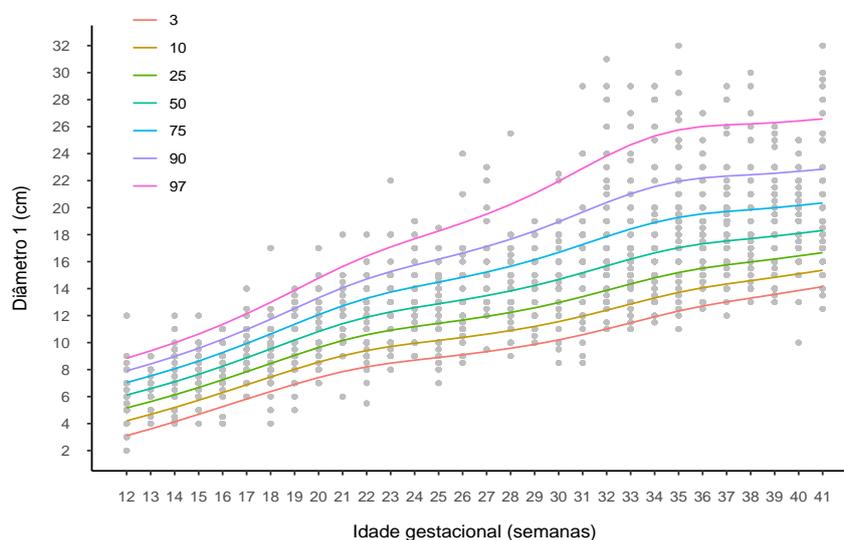


Figura 17: Valores observados do **D1** representados em função da **IG** para o modelo final **GAMLSS** com a função de alisamento *P-splines* e a distribuição **BCT**, com sete curvas de percentis ajustadas (3°, 10°, 25°, 50°, 75°, 90° e 97°).

Tabela 10: Resumo do desempenho do modelo final do **D1**, método **GAMLSS** com a função de alisamento *P-splines* e a distribuição **BCT**. São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.

Percentil nominal	3	10	25	50	75	90	97
GAMLSS (BCT)	3,05	10,37	25,05	50,29	75,63	90,02	96,75

7.3.2 Diâmetro 2

No desenvolvimento das curvas de crescimento do parâmetro **D2**, começou-se por realizar uma minuciosa avaliação do desempenho de várias funções de alisamento, sem associar nenhuma distribuição específica ao modelo. A Tabela 11 sintetiza os resultados dessa avaliação. As redes neurais sobressaíram com um **R²** de 0,723, no entanto, os outros parâmetros tiveram um desempenho pior comparativamente aos outros modelos com **MSE** de 4,588, **AIC** de 9424 e **BIC** de 9525. De modo geral, as *P-splines* apresentaram um ajuste melhor com um **R²** de 0,722, **MSE** de 4,568, **AIC** de 9407 e **BIC** de 9454. As *splines* cúbicas e *Loess* tiveram um desempenho semelhante com valores de **R²** de 0,721 e 0,720, **MSE** de 4,568 e 4,570, **AIC** de 9414 e 9417 e **BIC** de 9447 e 9449.

Além dessas métricas tradicionais, aplicaram-se critérios de diagnóstico gráficos. Na Figura 18, é possível visualizar os critérios referentes à função de alisamento *P-splines*, que se assemelham de uma forma geral aos critérios de todas as outras funções de alisamento testadas. O *worm plot* revela a presença de diversos pontos fora dos intervalos de confiança de 95%, sugerindo que o ajuste pode ser inadequado. Além disso, a forma em S indica a existência de leptocurtose. As estatísticas Z também indicam possíveis inadequações no modelo, com muitos valores de $|Z| > 2$. No que diz respeito aos gráficos de resíduos, para o gráfico relativo aos valores ajustados, os resíduos formam padrões verticais distintos e apresentam um desvio notável da normalidade.

Após determinar a função de alisamento mais adequada, foram avaliadas diferentes distribuições, conforme detalhado na Tabela 11. Na generalidade todos os modelos indicaram um bom desempenho, nomeadamente o **GAMLSS** com **BCT** e **LMST**. Estes modelos apresentaram valores de **R²** de 0,765 e 0,768, **MSE** de 4,398 e 4,394, **AIC** de 9029 e 9014 e **BIC** de 9142 e 9142. Quanto aos restantes modelos, os valores das métricas não diferiram substancialmente e apresentaram de modo geral um bom ajuste e parcimónia. À primeira vista, qualquer um destes modelos poderia representar a natureza dos dados. Ao analisar os resultados das técnicas de diagnóstico gráfico, observamos que o modelo **BCT**, ilustrado na Figura 19, se destacou com um ajuste notável. O *worm plot* apresentou pontos e uma

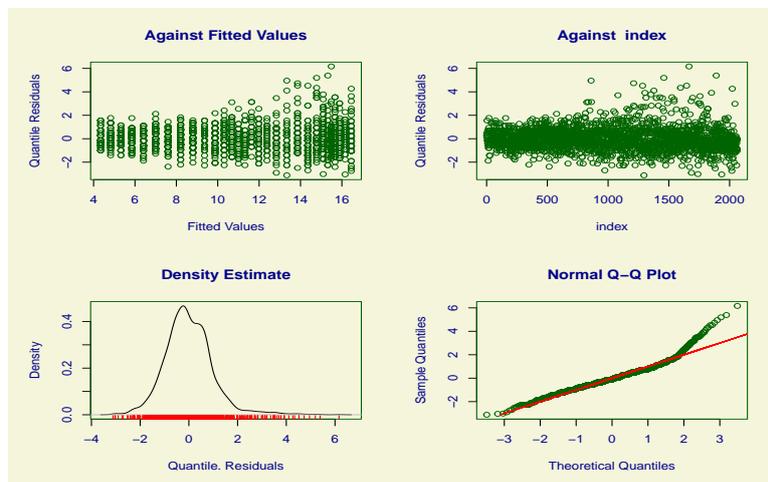
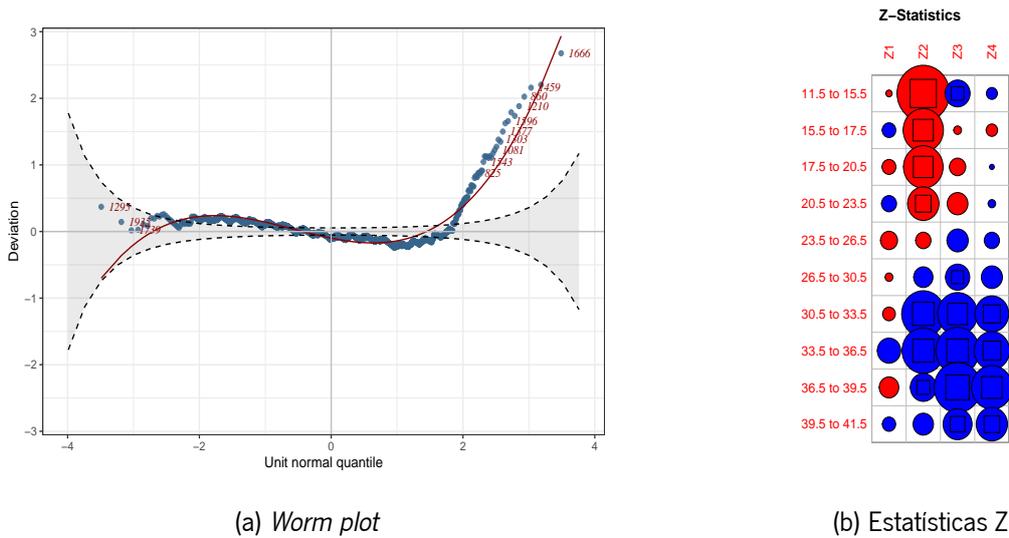


Figura 18: Critérios de avaliação de qualidade de ajuste (*worm plot*, estatísticas Z e gráficos de resíduos) para o modelo **GAMLSS** do **D2** com apenas a função de alisamento *P-splines*.

linha de ajuste próximos da linha de referência horizontal, com mais de 95% dos pontos nos intervalos de confiança. As estatísticas Z também indicaram um bom ajuste, com apenas dois valores de $|Z| > 2$ para Z4. Quanto aos resíduos, continuaram a exibir um comportamento em linhas verticais para os valores ajustados, indicando heterogeneidade na variância dos resíduos em diferentes partes da faixa de valores ajustados. Apresentaram também uma aproximação da normalidade, como pode-se constatar pelos dois gráficos inferiores em 19c.

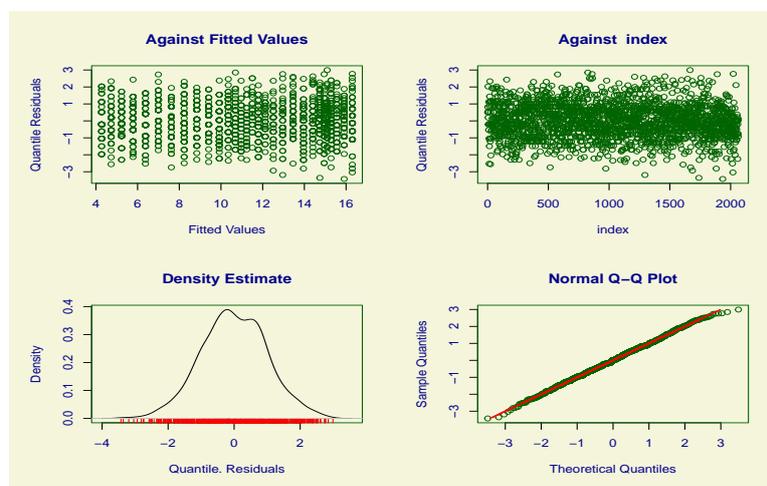
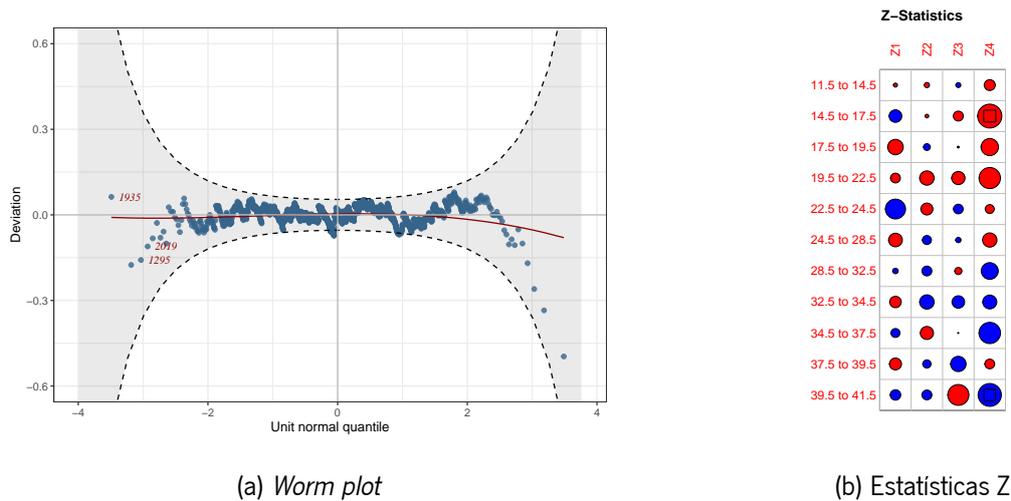


Figura 19: Critérios de avaliação de qualidade de ajuste (*worm plot*, estatísticas Z e gráficos de resíduos) para o modelo **GAMLSS** do **D2** com a função de alisamento *P-splines* e a distribuição **BCT**.

O modelo **LMST**, presente no Apêndice E, para o *worm plot* não apresentou um ajuste tão bom quanto o **GAMLSS** com **BCT**, como pode-se ver pela forma ligeiramente em S, indicando possível platicurtose, no entanto, para as estatísticas Z não apresentou nenhum valor de $|Z|$ acima de dois, indicando um bom ajuste. Os resíduos têm um comportamento como **GAMLSS BCT** e aproximação à normalidade. O modelo **GAMLSS** com **JSU** também teve um bom ajuste, identificado pelo *worm plot* semelhante a **LMST**. Para os restantes modelos, os ajustes não foram tão promissores, dado que foi possível verificar várias inadequações nos diversos diagnósticos. Os modelos **GAMLSS** com **BCCG**, **BCPE** e **LMS** e **LMSP** apresentaram para os *worm plot* formas claras em S, que indicam leptocurtose para **BCCG**, **BCPE**, **LMS** e **LMSP**. As estatísticas Z também apresentaram várias situações em que $|Z| > 2$, nome-

adamente **BCCG** e **LMS** que mostraram 6 casos. **LMSP**, não apresentou nenhuma inadequação para as estatísticas Z. No que diz respeito aos resíduos, estes mostraram o mesmo comportamento em linhas verticais no gráfico dos valores ajustados e uma aproximação à normalidade.

Tabela 11: Métricas de desempenho do modelo para o **D2** em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exibe o **R²**, o **MSE**, o **AIC** e o **BIC**.

	R²	MSE	AIC	BIC
Funções de alisamento				
P-splines	0,722	4,568	9407	9454
<i>Splines</i> cúbicas	0,721	4,568	9414	9447
Redes neurais	0,723	4,588	9424	9525
<i>Loess</i>	0,720	4,570	9417	9449
Distribuições				
GAMLSS (BCCG)	0,763	4,404	9046	9151
GAMLSS (BCT)	0,765	4,398	9029	9142
GAMLSS (BCPE)	0,761	4,399	9033	9145
GAMLSS (JSU)	0,762	4,407	9065	9140
LMS	0,763	4,404	9046	9151
LMST	0,768	4,394	9014	9142
LMSP	0,764	4,392	9014	9134

A Figura 20 exibe as curvas de crescimento de percentis geradas pelo modelo mais promissor, que combina **GAMLSS** com **BCT** e *P-splines*. Estas curvas refletem o comportamento dos dados, representados pelos pontos em cinzento no gráfico. Os valores numéricos correspondentes a cada percentil em diferentes intervalos gestacionais podem ser encontrados no Apêndice E. Como avaliação final do modelo, a Tabela 12 apresenta as percentagens da amostra que se encontram no mesmo valor ou abaixo de cada uma das curvas ajustadas. É notável a proximidade entre as percentagens nominais e as observadas na amostra, indicando um bom ajuste das curvas de percentis. As curvas de crescimento de percentis para o segundo melhor modelo podem ser consultadas no Apêndice E e são visualmente semelhantes àquelas obtidas com o **GAMLSS**.

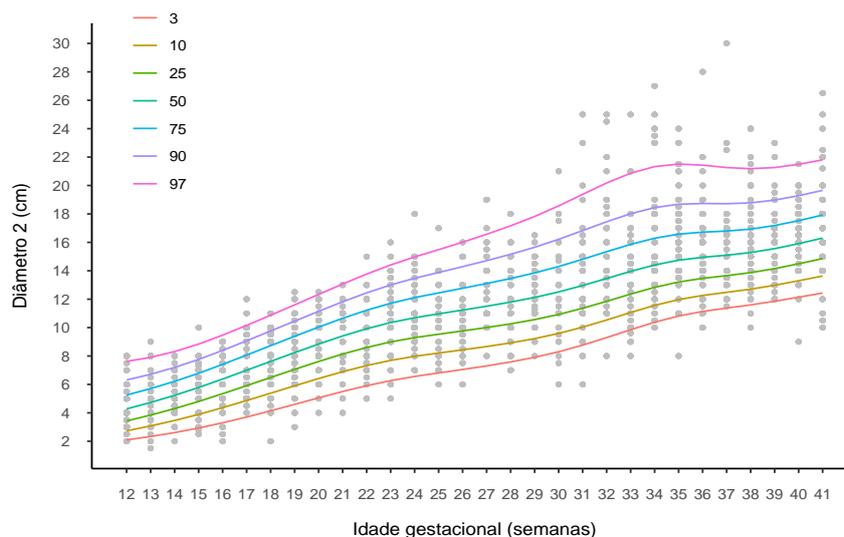


Figura 20: Valores observados do **D2** representados em função da **IG** para o modelo final **GAMLSS** com a função de alisamento *P-splines* e a distribuição **BCT**, com sete curvas de percentis ajustadas (3º, 10º, 25º, 50º, 75º, 90º e 97º).

Tabela 12: Resumo do desempenho do modelo final do **D2**, método **GAMLSS** com a função de alisamento *P-splines* e a distribuição **BCT**. São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.

Percentil nominal	3	10	25	50	75	90	97
GAMLSS (BCT)	3,00	9,21	24,37	49,08	74,61	90,07	96,56

7.3.3 Espessura placentar

No processo de desenvolvimento das curvas de crescimento para a **EP**, realizou-se uma análise abrangente do desempenho de várias funções de alisamento, sem a imposição de uma distribuição específica no modelo como pode ser visualizado na Tabela 13. Para as *P-splines*, observou-se um **R²** de 0,342, um **MSE** de 1,698, um **AIC** de 3506 e um **BIC** de 3545. As redes neurais apresentaram um **R²** de 0,343, um **MSE** de 1,700, um **AIC** de 3513 e um **BIC** de 3581. No caso das *splines* cúbicas e de *Loess*, registou-se um **R²** de 0,341 e 0,340, um **MSE** de 1,699 e 1,700, um **AIC** de 3508 e 3509 e um **BIC** de 3541 e 3541. As *P-splines* indicam um melhor ajuste geral, apesar de as redes neurais terem um **R²** ligeiramente mais elevado, as outras métricas apresentam, resultados menos favoráveis. Analisando os critérios de diagnóstico gráficos para a função de alisamento *P-splines*, na Figura 21, é possível notar que estes se assemelham aos critérios aplicados a todas as outras funções de alisamento testadas.

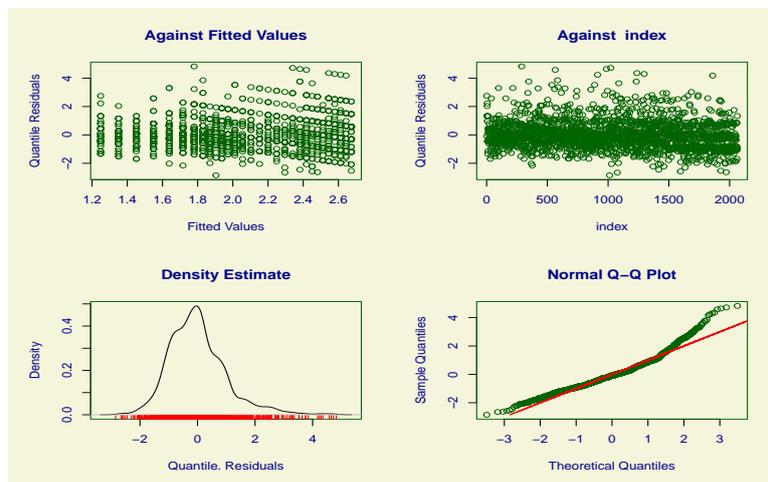
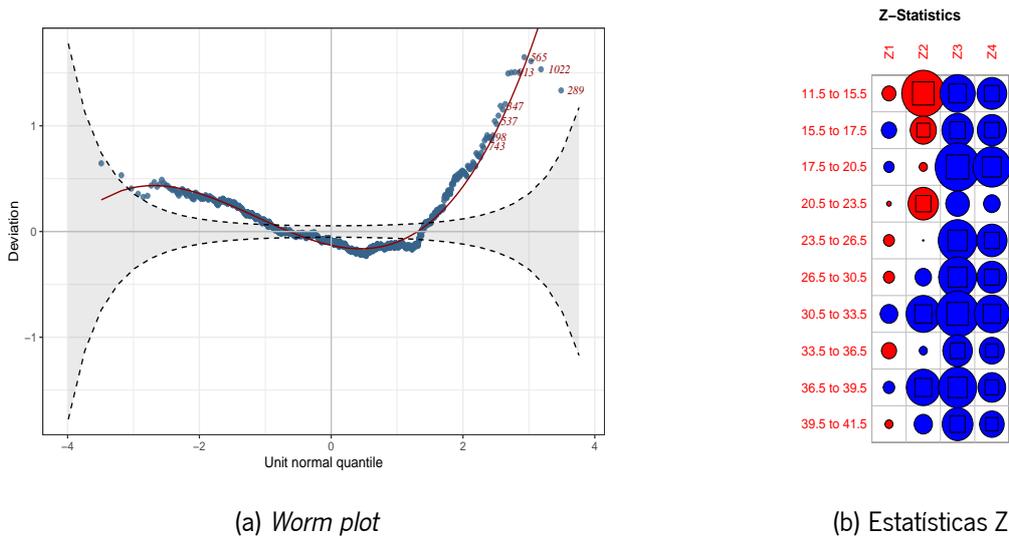


Figura 21: Critérios de avaliação de qualidade de ajuste (*worm plot*, estatísticas Z e gráficos de resíduos) para o modelo **GAMLSS** da **EP** com apenas a função de alisamento *P-splines*.

O gráfico *worm plot* chama a atenção para a presença de vários pontos que se encontram fora dos intervalos de confiança de 95%. Este cenário sugere que o ajuste do modelo pode ser inadequado. Ademais, a curvatura no gráfico sugere a possível presença de leptocurtose, uma característica que requer atenção. Outra abordagem utilizada envolveu o uso de estatísticas Z, nas quais muitos valores de $|Z|$ excederam o limiar de 2, conforme indicado pelos quadrados presentes nos círculos. Esta observação sugere que o modelo pode apresentar inadequações relativamente a vários aspetos, incluindo variância, assimetria e curtose. No que se refere aos gráficos de resíduos, estes mostraram um comportamento com linhas verticais ao longo do gráfico em relação aos valores ajustados e exibiram um desvio notável relativamente à suposição de normalidade.

Após a função de alisamento mais apropriada ser identificada, passou-se a examinar diversas distribuições, conforme descrito na Tabela 13. Os vários modelos apresentaram métricas muito semelhantes entre si, por exemplo, **GAMLSS** com **BCT** e **JSU** e **LMST** e **LMSP** com um valor de R^2 de 0,390, indicando que os modelos conseguiram capturar 39% da variação nos dados, sugerindo um ajuste razoável. Quanto aos restantes valores destes modelos, **MSE** variou entre 1,551 e 1,553, **AIC** entre 3205 e 3211 e **BIC** de 3262 a 3288. Para os restantes modelos, **GAMLSS** com **BCCG** e **BCPE** e **LMSP** apresentaram métricas ligeiramente menos favoráveis, no entanto, tiveram ajustes também razoáveis. Ao examinar-se os resultados das análises gráficas de diagnóstico, nota-se que o modelo **LMST**, que está representado na Figura 22, destacou-se com um bom desempenho.

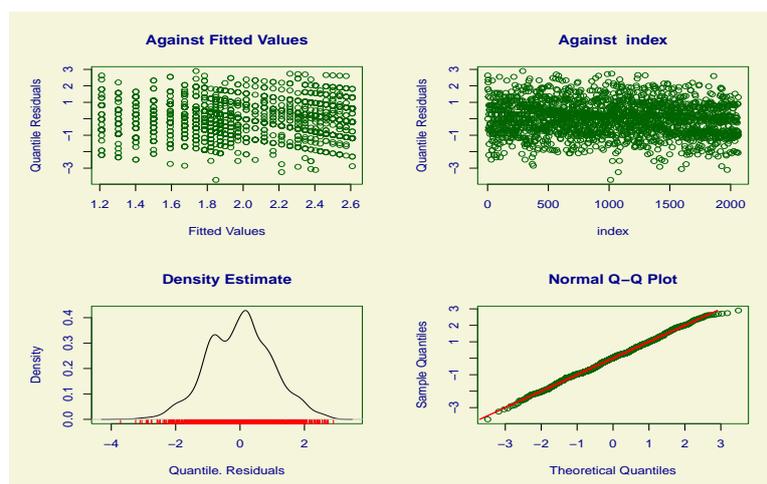
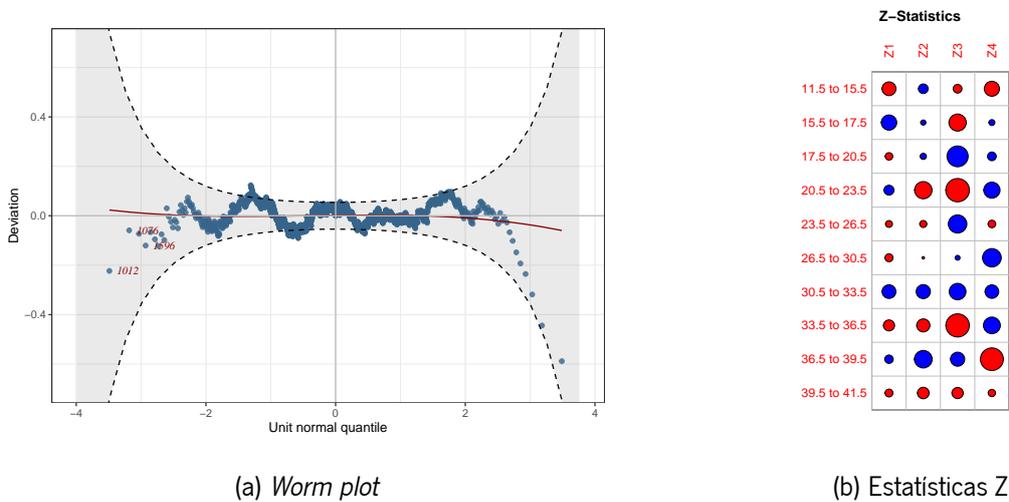


Figura 22: Critérios de avaliação de qualidade de ajuste (*worm plot*, estatísticas Z e gráficos de resíduos) para o modelo **LMST** da **EP** com a função de alisamento *P-splines*.

O *worm plot* sugeriu um bom ajuste do modelo, uma vez que a maioria dos pontos está entre as faixas de confiança elípticas aproximadas de 95%. As estatísticas Z também apontaram para um bom ajuste, sem nenhum valor de $|Z| > 2$. Relativamente aos resíduos, eles ainda mostraram uma tendência de exibir um comportamento em linhas verticais em relação aos valores ajustados. Além disso, observou-se uma aproximação à normalidade, como evidenciado pelos dois gráficos inferiores na Figura 22c.

Já o modelo **GAMLSS** com **BCT**, disponível no Apêndice F, apresentou um ajuste ligeiramente menos satisfatório quando comparado ao **LMST**, como evidenciado por dois valores de $|Z| > 2$ para as estatísticas Z , porém o *worm plot* e os resíduos apresentaram uma disposição praticamente igual a **LMST**. Os modelos **GAMLSS** com **BCPE** e **JSU** e **LMSP** também tiveram um diagnóstico favorável com *worm plots* e resíduos semelhantes aos de **LMST**, mas as estatísticas Z apresentaram mais casos de inadequações. Por outro lado, para os modelos **GAMLSS** com **BCCG** e **LMS**, os ajustes não se mostraram tão promissores. Os *worm plot* mostraram uma curvatura em forma de S indicando leptocurtose e as estatísticas Z apresentaram muitos casos em que $|Z| > 2$, por exemplo, para **BCCG** em que existem 11 casos. Já os resíduos tiveram o mesmo comportamento que os outros modelos.

Tabela 13: Métricas de desempenho do modelo para a **EP** em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exibe o **R²**, o **MSE**, o **AIC** e o **BIC**.

	R²	MSE	AIC	BIC
Funções de alisamento				
P-splines	0,342	1,698	3506	3545
<i>Splines</i> cúbicas	0,341	1,699	3508	3541
Redes neurais	0,343	1,700	3513	3581
<i>Loess</i>	0,340	1,700	3509	3541
Distribuições				
GAMLSS (BCCG)	0,379	1,574	3254	3316
GAMLSS (BCT)	0,390	1,553	3211	3283
GAMLSS (BCPE)	0,378	1,553	3211	3274
GAMLSS (JSU)	0,390	1,551	3205	3262
LMS	0,379	1,574	3254	3316
LMST	0,390	1,552	3210	3288
LMSP	0,390	1,551	3205	3262

A Figura 23 ilustra as curvas de crescimento de percentis geradas pelo modelo mais promissor, **LMST**. Estas curvas representam precisamente o comportamento dos dados, ilustrados pelos pontos em cinzento no gráfico. Os detalhes numéricos específicos em relação a cada percentil em diferentes **IG**, estão disponíveis no Apêndice F. A Tabela 14 destaca as percentagens da amostra que coincidem com ou estão abaixo de cada uma das curvas ajustadas. A notável concordância entre as percentagens nominais e as observações na amostra sugere um bom ajuste das curvas de percentis. Para visualizar as curvas de crescimento de percentis do segundo melhor modelo, consulte o Apêndice F. É importante notar que estas curvas são visualmente similares àquelas obtidas com o modelo **LMST**, reforçando a qualidade do ajuste de ambos os modelos.

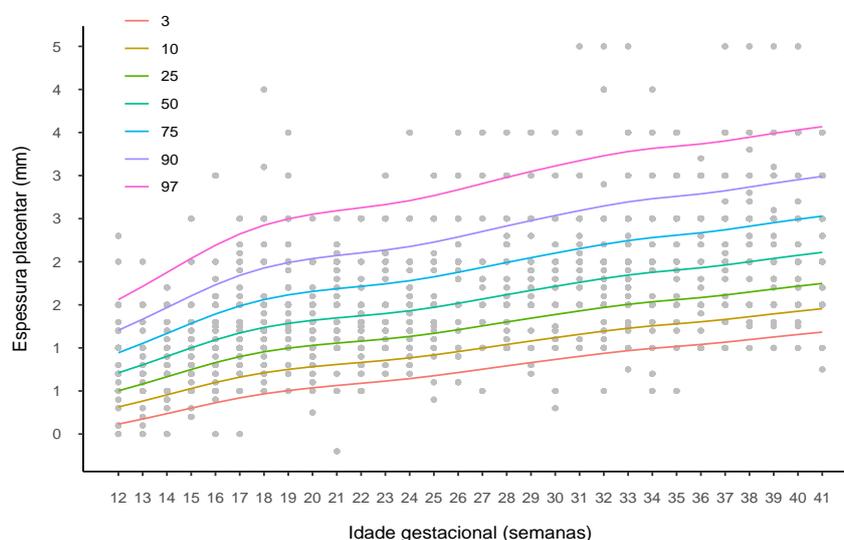


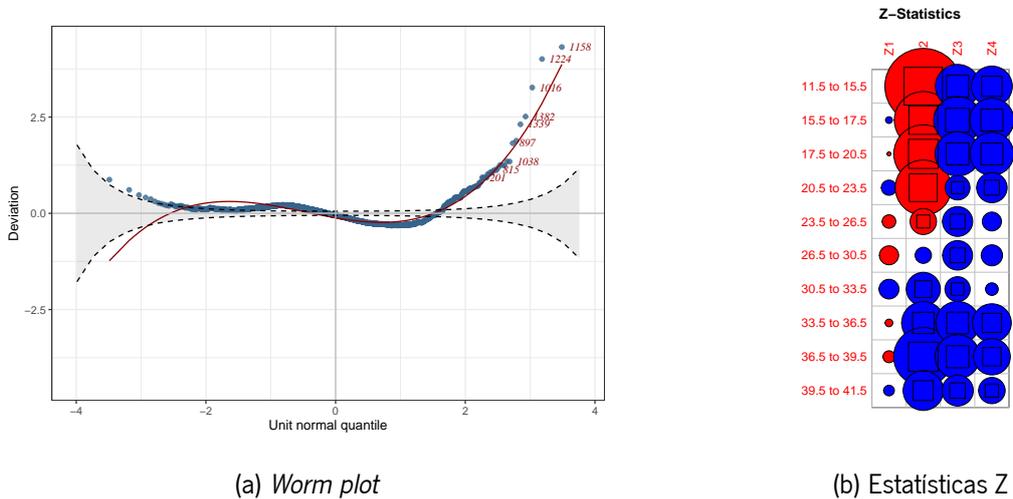
Figura 23: Valores observados da **EP** representados em função da **IG** para o modelo final **LMST** com a função de alisamento *P-splines*, com sete curvas de percentis ajustadas (3°, 10°, 25°, 50°, 75°, 90° e 97°).

Tabela 14: Resumo do desempenho do modelo final da **EP**, método **LMST** com a função de alisamento *P-splines*. São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.

Percentil nominal	3	10	25	50	75	90	97
LMST	3,59	8,67	27,33	49,32	74,37	89,83	96,27

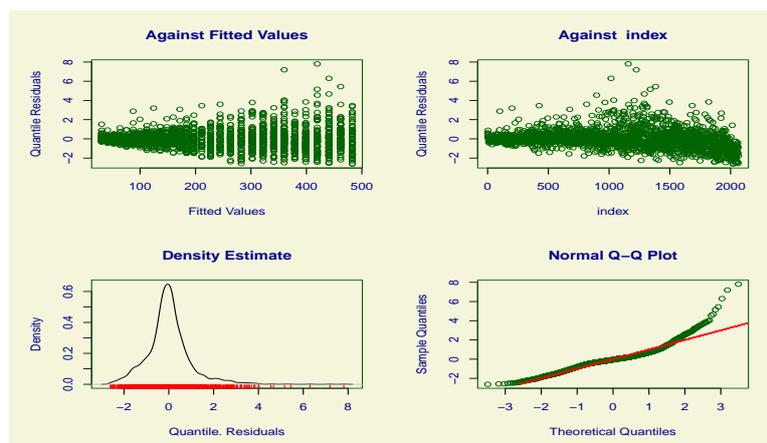
7.3.4 Peso placentar

No processo de desenvolvimento das curvas de crescimento do parâmetro **PP**, iniciou-se com uma análise de várias funções de alisamento, sem vincular uma distribuição específica ao modelo. A Tabela 15 resume os resultados dessa avaliação. As *P-splines* apresentaram o maior valor de R^2 de 0,783, e valores de **MSE**, **AIC** e **BIC** de 11,480, 23621 e 23659, respectivamente. Os restantes modelos, apresentaram valores iguais de R^2 e os modelos *splines* cúbicas e *Loess* apresentaram valores iguais para as restantes métricas, **MSE** de 11,470, **AIC** de 23622 e 23623 e **BIC** de 23655. As redes neurais, tiveram valores um pouco mais altos de **MSE**, **AIC** e **BIC** de 11,510, 23637 e 23705. Além destas métricas, aplicaram-se critérios de diagnóstico gráficos. Na Figura 24, é possível examinar os critérios relacionados à função de alisamento *P-splines*, que se assemelham aos critérios de todas as outras funções de alisamento testadas.



(a) Worm plot

(b) Estatísticas Z



(c) Resíduos

Figura 24: Critérios de avaliação de qualidade de ajuste (*worm plot*, estatísticas Z e gráficos de resíduos) para o modelo **GAMLSS** do **PP** com apenas a função de alisamento *P-splines*.

O *worm plot* revela a presença de vários pontos fora das faixas de confiança elípticas aproximadas de 95%, sugerindo um ajuste possivelmente inadequado. Além disso, a curvatura visualizada indica a presença de leptocurtose. As estatísticas Z também sugerem possíveis inadequações no modelo, com muitos valores de $|Z|$ excedendo 2, como indicado pelos quadrados nos círculos. No que se refere aos gráfico de resíduos em função dos valores ajustados, este apresentou uma condensação no início e depois um ajustamento em linhas verticais ao longo do gráfico, indicando que a variabilidade dos erros não é constante ao longo da faixa de valores previstos. Existe também um desvio notável da normalidade.

Após a identificação da função de alisamento mais apropriada, procedeu-se à análise de várias distribuições, conforme detalhado na Tabela 15. Observou-se que os vários modelos apresentaram métricas de desempenho muito similares entre si. Por exemplo, o valor de **MSE** foi de 10,890 para todos os modelos, exceto **JSU** que teve um valor de 10,870. Este modelo também apresentou o maior valor de R^2 de 0,856, para **AIC** e **BIC** teve valores de 22309 e 22394, respetivamente. **GAMLSS** com **BCT** e **LMST**, também apresentaram valores altos de R^2 de 0,851 e 0,852 e valores de **AIC** de 22287 e 22281 e **BIC** de 22404 e 22403, respetivamente. Quanto aos restantes modelos, o ajuste também foi bom tendo valores de R^2 a variar entre 0,848 e 0,849, **AIC** entre 22291 e 22320 e **BIC** de 22408 a 22423.

Ao explorar os resultados das análises gráficas de diagnóstico, destacou-se o modelo **LMST**, presente na Figura 25. O *worm plot* sugeriu um bom ajuste do modelo, uma vez que a maioria dos pontos está nas faixas de confiança elípticas aproximadas de 95%. As estatísticas Z também apontaram para um bom ajuste, com dois casos em que $|Z| > 2$, um para Z_2 e outro em Z_4 . Além disso, os resíduos quantílicos em função dos valores ajustados mostraram uma dispersão no início do gráfico e depois um ajustamento em linhas verticais, o que sugere que num determinado intervalo inicial dos valores ajustados, os resíduos quantílicos estão agrupados ou concentrados em torno de um valor central. Essa condensação pode indicar que o modelo está bem ajustado nesse intervalo e os resíduos não exibem padrões óbvios nessa faixa de valores. No entanto, os padrões observados seguidamente podem indicar heterogeneidade na variância dos resíduos em diferentes partes da faixa de valores ajustados. Foi também observada a suposição de normalidade. Por outro lado, o modelo **GAMLSS** com **JSU**, disponível no Apêndice G, apresentou um ajuste praticamente igual quando comparado ao **LMST**, a diferença está nas estatísticas Z onde apresenta cinco casos onde $|Z| > 2$. O modelo **BCT** também teve um ajuste semelhante a estes dois modelos. Os restantes modelos apresentaram algumas inadequações no modelo, principalmente evidentes a partir dos *worm plots* que apresentam uma forma curvada indicando leptocurtose para **BCCG**, **BCPE**, **LMS** e **LMSP**. As estatísticas Z apresentaram algumas inadequações, no entanto, os resíduos apresentaram o mesmo comportamento que os restantes modelos.

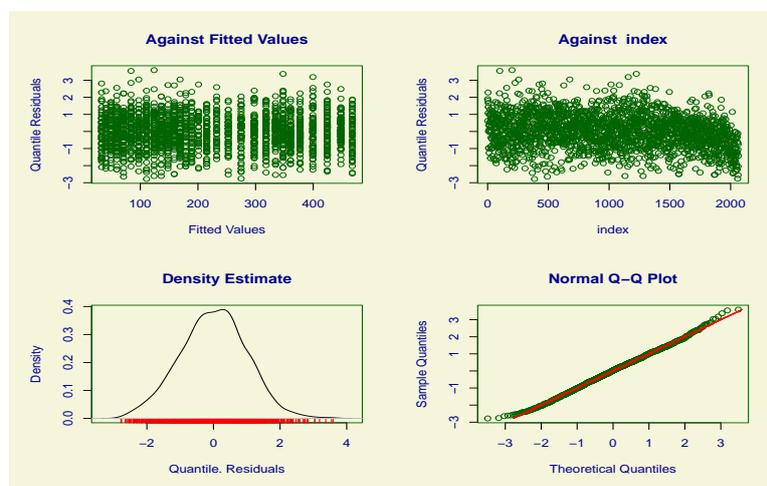
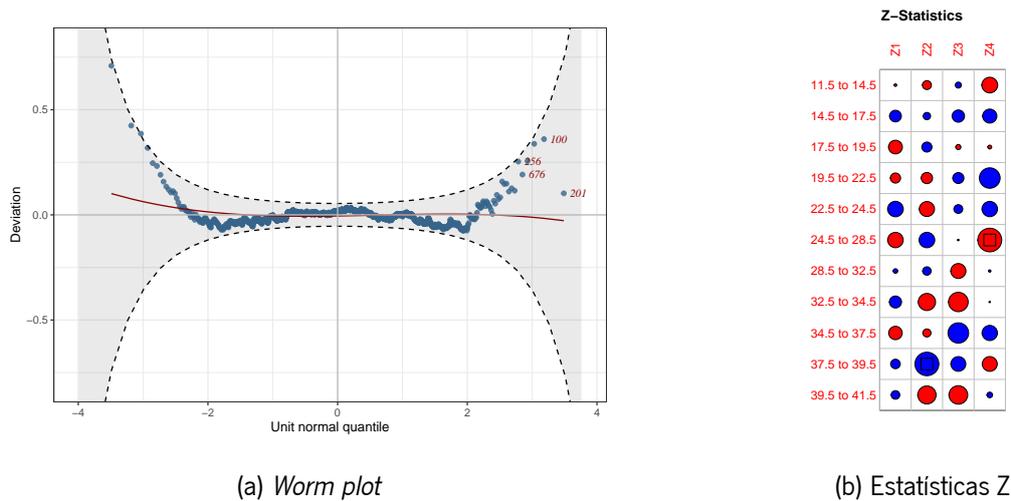


Figura 25: Critérios de avaliação de qualidade de ajuste (*worm plot*, estatísticas Z e gráficos de resíduos) para o modelo **LMST** do **PP** com a função de alisamento *P-splines*.

A representação das curvas de crescimento de percentis para o modelo mais promissor, **LMST**, pode ser observada na Figura 26. Estas curvas apontaram uma notável capacidade de capturar precisamente o comportamento dos dados, representados pelos pontos em cinzento no gráfico. Os valores numéricos para cada percentil em cada **IG** estão disponíveis para consulta no Apêndice G. Como uma análise final do modelo, a Tabela 16 apresenta as percentagens da amostra localizadas no mesmo valor ou abaixo de cada curva ajustada. É notável que a estreita concordância entre as percentagens nominais e as da amostra sugere um bom ajuste das curvas de percentis. Além disso, as curvas de crescimento de percentis para o segundo melhor modelo também estão disponíveis no Apêndice G e apresentam uma semelhança visual significativa com as curvas do modelo **LMST**.

Tabela 15: Métricas de desempenho do modelo para o **PP** em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exibe o **R²**, o **MSE**, o **AIC** e o **BIC**.

	R²	MSE	AIC	BIC
Funções de alisamento				
P-splines	0,783	11,480	23621	23659
<i>Splines</i> cúbicas	0,782	11,470	23622	23655
Redes neurais	0,782	11,510	23637	23705
<i>Loess</i>	0,782	11,470	23623	23655
Distribuições				
GAMLSS (BCCG)	0,849	10,890	22320	22423
GAMLSS (BCT)	0,851	10,890	22287	22404
GAMLSS (BCPE)	0,848	10,890	22295	22408
GAMLSS (JSU)	0,856	10,870	22309	22394
LMS	0,849	10,890	22320	22423
LMST	0,852	10,890	22281	22403
LMSP	0,849	10,890	22291	22411

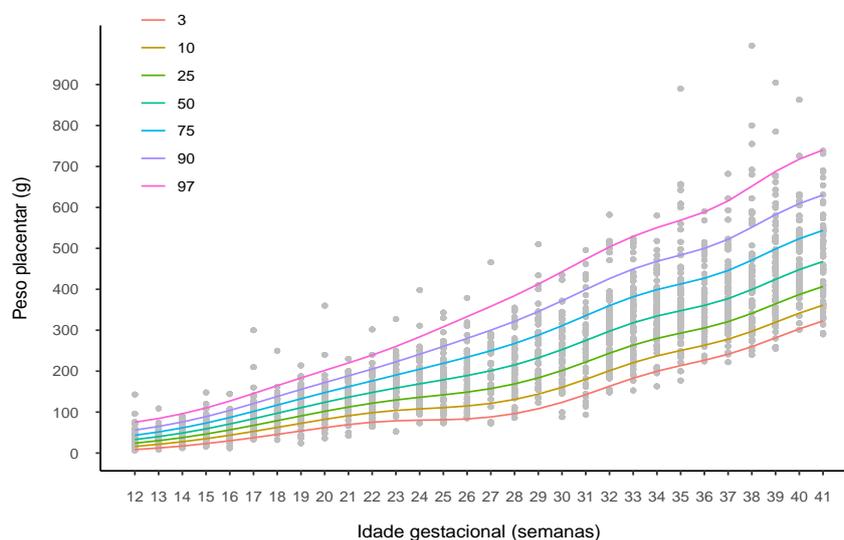


Figura 26: Valores observados do **PP** representados em função da **IG** para o modelo final **LMST** com a função de alisamento *P-splines*, com sete curvas de percentis ajustadas (3°, 10°, 25°, 50°, 75°, 90° e 97°).

Tabela 16: Resumo do desempenho do modelo final do **PP**, método **LMST** com a função de alisamento *P-splines*. São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.

Percentil nominal	3	10	25	50	75	90	97
LMST	3,44	10,37	24,32	49,32	75,39	90,55	97,43

7.3.5 Peso fetal

O desenvolvimento das curvas de crescimento para o parâmetro **PF** constituiu um desafio, uma vez que alguns modelos foram incapazes de ajustar os dados como a função de alisamento redes neurais e os modelos **GAMLSS** com **BCCG** e **BCPE** e **LMS** e **LMSP**. Portanto, os resultados destes modelos não foram discutidos.

Analisando as métricas de desempenho, presentes na Tabela 17, é possível verificar que todas as funções tiveram um bom ajuste aos dados com valores altos de R^2 de 0,914. Quanto às restantes métricas, as *P-splines* apresentaram um **MSE** de 14,540, **AIC** e **BIC** de 29897 e 29952, respetivamente. As *splines* cúbicas, um **MSE** de 14.530, **AIC** de 29909 e **BIC** de 29943. *Loess*, teve valores de 14,530, 29914 e 29945 para **MSE**, **AIC** e **BIC**, respetivamente. Quanto à análise de diagnóstico gráfica, é muito semelhante para todos os modelos. Na Figura 27 é possível verificar o diagnóstico para as *P-splines*, o *worm plot* apresenta vários pontos fora dos intervalos de confiança de 95%. Sugerindo que o ajuste do modelo pode ser inadequado. Ademais, a curvatura no gráfico sugere a possível presença de leptocurtose. Quanto às estatísticas Z, apresentam várias inadequações relativamente a vários aspetos, incluindo variância, assimetria e curtose. Já os resíduos em relação aos valores ajustados mostraram uma condensação no início do gráfico e depois um ajustamento em linhas verticais ao longo do gráfico, sugerindo que a variabilidade dos erros não é constante ao longo da faixa de valores previstos.

Após determinar a função de alisamento mais adequada, procedeu-se à análise das distribuições, como detalhado na Tabela 17. É visível que os diversos modelos apresentaram métricas de desempenho bastante semelhantes entre si. O modelo **GAMLSS** com **JSU**, apresentou o valor mais alto para R^2 de 0,968, e valores de **MSE**, **AIC** e **BIC** de 13,100, 26788 e 26918, respetivamente. Os modelos **BCT** e **LMST** apresentaram um valor de R^2 de 0,961 e 0,962, de **MSE** 13,100 e 13,110, de **AIC** 26696 e 26679 e de **BIC** 26872 e 26874.

Ao analisar os resultados das análises gráficas de diagnóstico, o modelo **LMST**, representado na Figura 28, destacou-se. O *worm plot* sugeriu um bom ajuste do modelo, visto que a maioria dos pontos encontram-se nas faixas de confiança elípticas de 95%. As estatísticas Z também apontaram para um

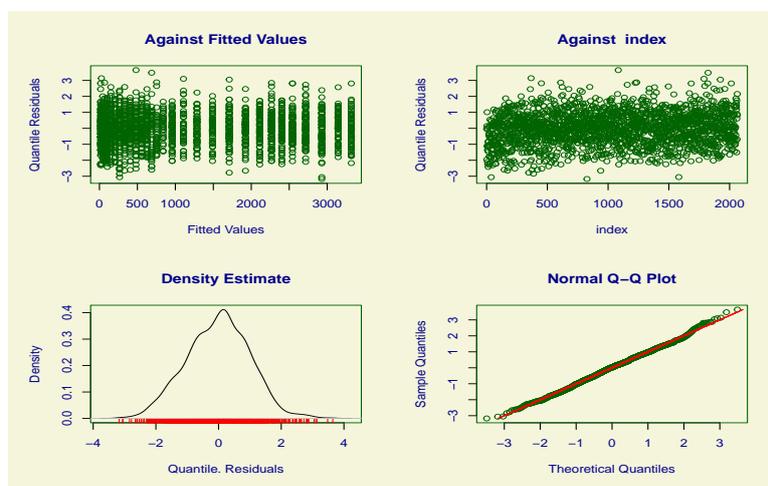
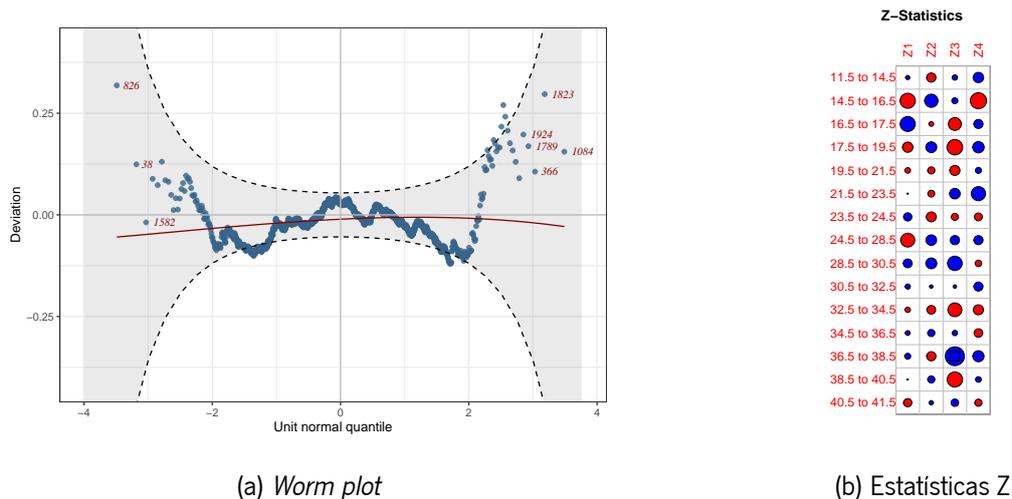


Figura 28: Critérios de avaliação de qualidade de ajuste (*worm plot*, estatísticas Z e gráficos de resíduos) para o modelo **LMST** do **PF** com a função de alisamento *P-splines*.

delo mais promissor, o **LMST**. Estas curvas conseguem capturar de forma precisa o comportamento dos dados, ilustrados pelos pontos em cinzento no gráfico. Os valores numéricos para cada percentil em cada **IG** estão disponíveis para consulta no Apêndice H. Além disso, a Tabela 18 apresenta as percentagens da amostra que coincidem com ou estão abaixo de cada uma das curvas ajustadas. A notável concordância entre as percentagens nominais e as observações na amostra sugere um bom ajuste das curvas de percentis. Para visualizar as curvas de crescimento de percentis do segundo melhor modelo, consulte o Apêndice H. É importante notar que estas curvas apresentam uma semelhança visual significativa com aquelas obtidas pelo modelo **LMST**, reforçando a qualidade do ajuste de ambos os modelos.

Tabela 17: Métricas de desempenho do modelo para o **PF** em diferentes métodos para diversas funções de alisamento e distribuições. A tabela exibe o **R²**, o **MSE**, o **AIC** e o **BIC**.

	R²	MSE	AIC	BIC
Funções de alisamento				
P-splines	0,914	14,540	29897	29952
<i>Splines</i> cúbicas	0,914	14,530	29909	29943
<i>Loess</i>	0,914	14,530	29914	29945
Distribuições				
GAMLSS (BCT)	0,961	13,100	26696	26872
GAMLSS (JSU)	0,968	13,100	26788	26918
LMST	0,962	13,110	26679	26874

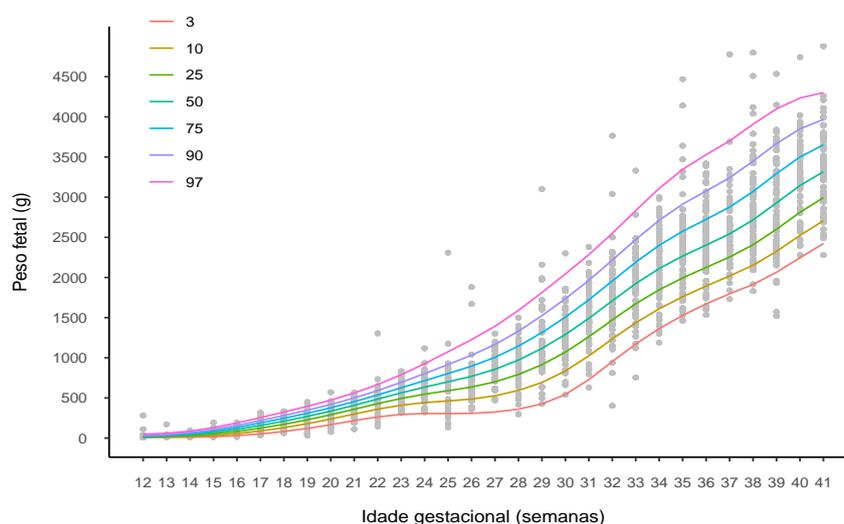


Figura 29: Valores observados do **PF** representados em função da **IG** para o modelo final **LMST** com a função de alisamento *P-splines*, com sete curvas de percentis ajustadas (3°, 10°, 25°, 50°, 75°, 90° e 97°).

Tabela 18: Resumo do desempenho do modelo final do **PF**, método **LMST** com a função de alisamento *P-splines*. São comparadas as percentagens nominais e de amostra em ou abaixo de cada percentil.

Percentil nominal	3	10	25	50	75	90	97
LMST	3,30	11,39	25,44	48,35	74,81	90,65	97,63

7.4 Aplicação web

Após a escolha das curvas de crescimento de percentis mais adequadas para os parâmetros **D1**, **D2**, **EP**, **PP** e **PF**, foi desenvolvida uma aplicação *web*, tal como descrito no Capítulo 6. Esta aplicação permite aos utilizadores explorar e analisar interativamente as curvas de referência para os vários parâmetros.

Antes de entrar-se em detalhes sobre os resultados que podem ser obtidos através desta aplicação, é importante entender a estrutura e as funcionalidades da aplicação. A aplicação “PlacentalGrowth” foi projetada com uma estrutura intuitiva e amigável para facilitar a sua utilização por profissionais de saúde e investigadores. Ao iniciar a aplicação, o utilizador é recebido com um modal de boas-vindas que fornece uma breve introdução e explica o que pode ser feito na aplicação, como pode ser visualizado na Figura 30.

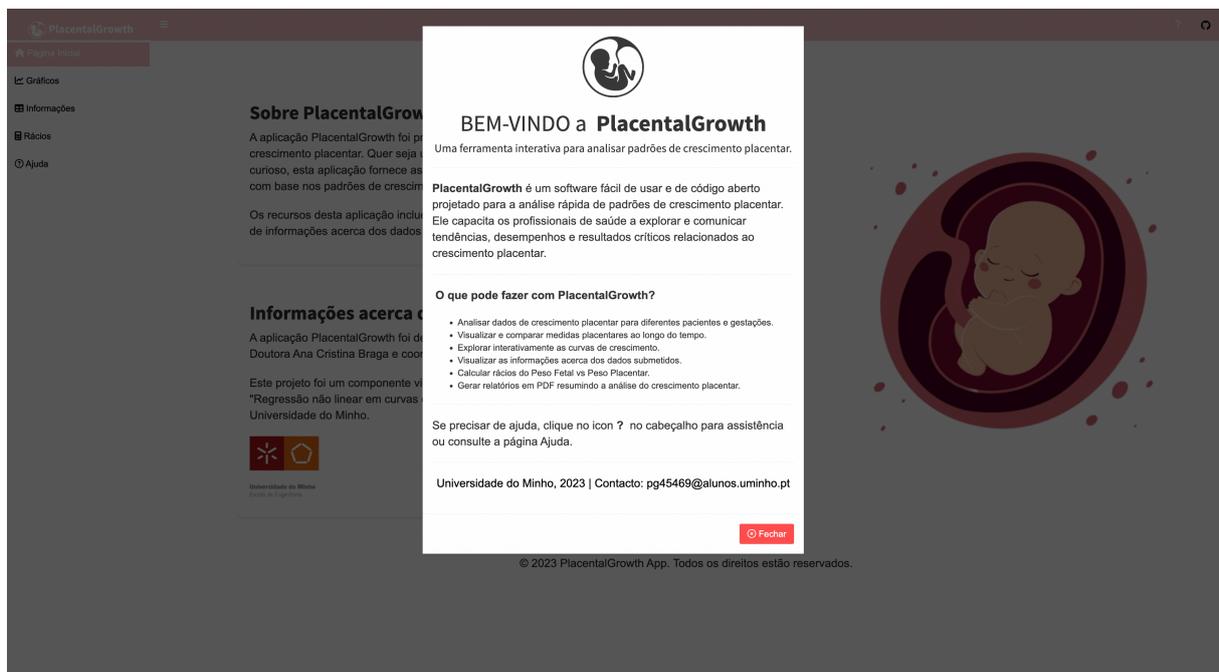


Figura 30: Modal de Boas-Vindas — Introdução à aplicação.

O modal de boas-vindas é o ponto de partida para o utilizador, orientando-o sobre como utilizar a aplicação para analisar vários parâmetros ao longo da **IG**. Ele oferece uma visão geral das funcionalidades e recursos disponíveis. Além disso, a página inicial da aplicação, ilustrada na Figura 31, serve como um recurso informativo para o utilizador, proporcionando-lhe contexto sobre a aplicação e o projeto que levou ao seu desenvolvimento. Para ilustrar a utilização da aplicação, considere-se um estudo de caso hipotético envolvendo uma paciente grávida.

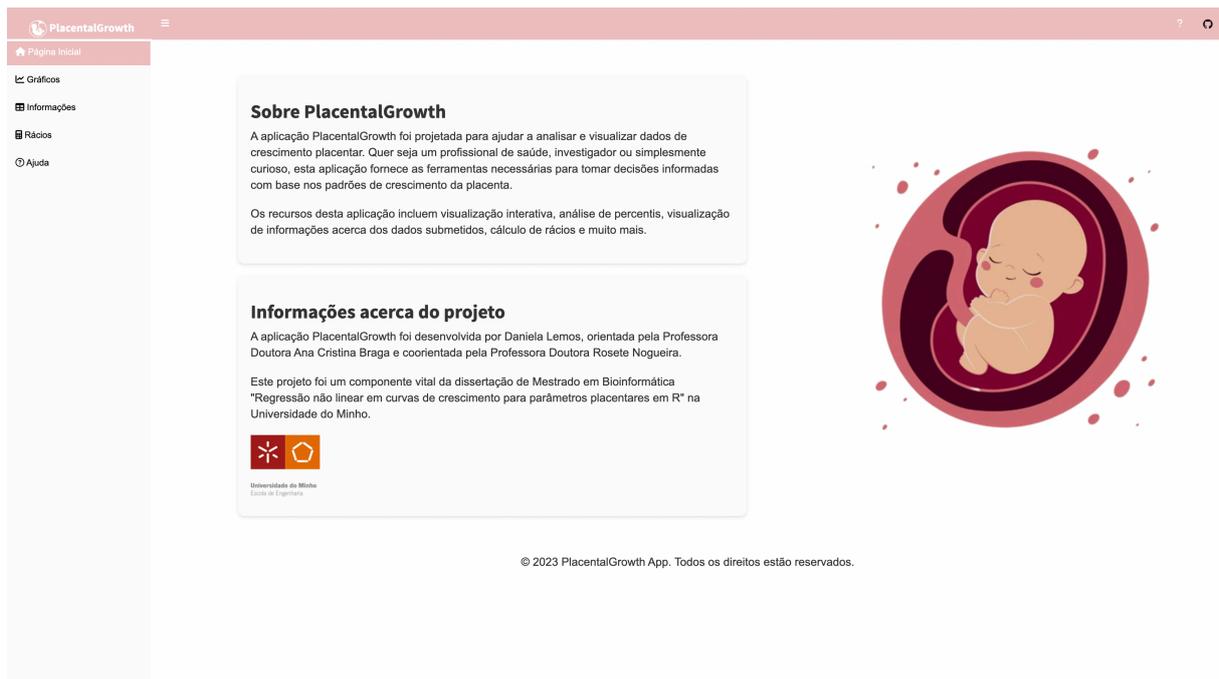


Figura 31: Visão geral da Página Inicial da aplicação.

O utilizador pode dar início à análise, através da página “Gráficos”, visível na Figura 32, onde é solicitado que insira informações do processo do paciente em estudo. Para este estudo de caso consideraram-se os seguintes dados iniciais: Número de processo – 549667, Nome – Maria Silva, Idade – 34 anos e Data de avaliação – 09/10/2023. É possível também adicionar informações extra que sejam relevantes, neste caso não foi necessário.

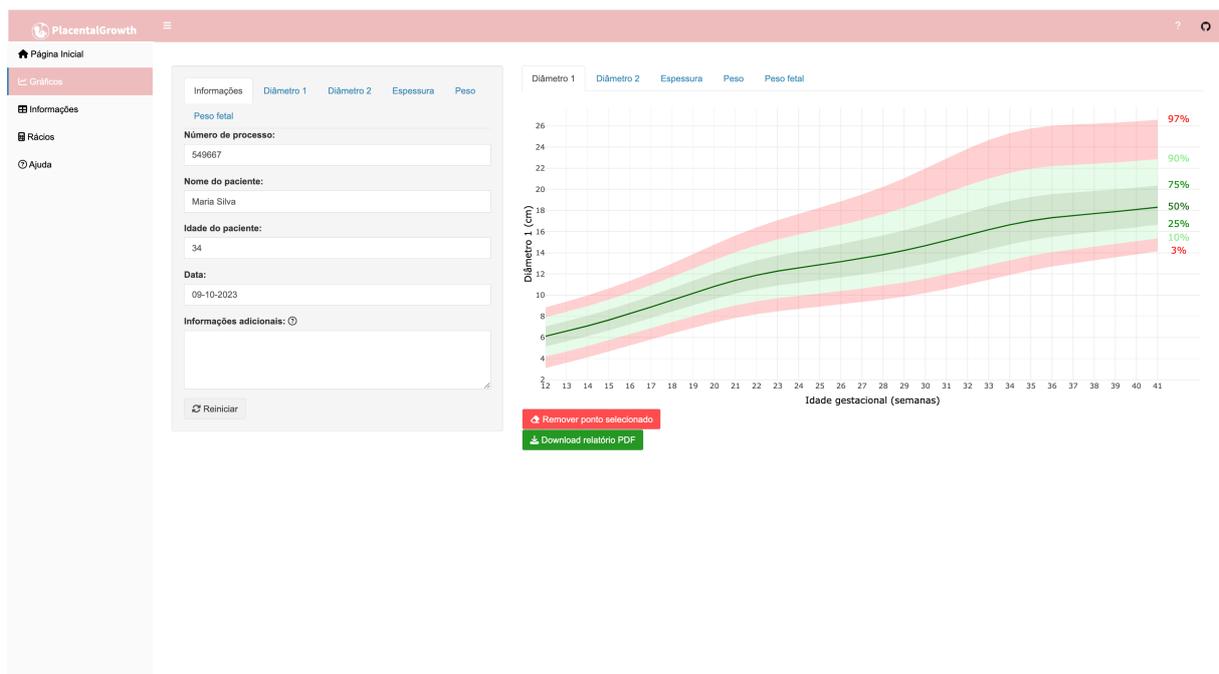


Figura 32: Página Gráficos – Introdução das informações do processo do paciente na aba “Informações”.

Posteriormente à inserção das informações do paciente, o utilizador pode prosseguir para a inclusão dos parâmetros que serão analisados. Na Figura 33, é possível visualizar a introdução das informações do processo para o parâmetro **D1** e o respetivo gráfico da curva de crescimento com os pontos adicionados pelo utilizador. Esta aplicação oferece ao utilizador a flexibilidade de adicionar esses pontos de duas maneiras: via carregamento de um ficheiro CSV ou manualmente. Neste caso, já existia informação prévia da paciente para as 14 e 25 semanas de gestação, por isso, foi carregado um ficheiro CSV. Ao selecionar a opção “Sim” e escolhendo o ficheiro, a aplicação processa os dados do CSV e exibe-os na forma de pontos adicionados aos gráficos das curvas de crescimento. Além disso, o utilizador consegue inserir pontos manualmente, permitindo uma personalização completa dos dados. Por exemplo, neste estudo de caso, acrescentaram-se manualmente pontos de dados referentes à 33.^a semana de gestação.

Na Figura 33, está ilustrado como o utilizador pode visualizar as curvas de crescimento de referência com os pontos adicionados pelo utilizador para o parâmetro **D1**. A aplicação permite a interação com os pontos, fornecendo informações detalhadas quando o utilizador adiciona um ponto. Por exemplo, quando o utilizador adicionar o ponto relativo à 33.^a semana para o **D1**, aparece a seguinte mensagem “O ponto encontra-se entre o 25.^o e o 50.^o percentil”, a informar o percentil onde o ponto caiu. É também possível eliminar pontos dos gráficos, caso o utilizador se engane, reiniciar completamente as entradas e gráficos e guardar os pontos adicionados aos gráficos num ficheiro CSV para posterior análise, selecionando a caixa “Deseja salvar os pontos num ficheiro CSV?”.



Figura 33: Página Gráficos – Introdução dos dados para o parâmetro **D1** e visualização do gráfico com a curva de crescimento.

Após todas as informações e pontos serem adicionados na página “Gráficos”, as suas informações podem ser consultadas na página “Informações”. Esta página oferece uma visão detalhada e tabular de todas as informações associadas aos pontos adicionados pelo utilizador aos gráficos de crescimento placentar. Para cada ponto inserido, a página “Informações” exibe tabelas detalhadas que incluem informações essenciais, como a **IG**, o valor do parâmetro em estudo (por exemplo, **D1**) e o percentil correspondente. Isso permite ao utilizador aceder rapidamente aos detalhes de cada ponto e compreender como eles se posicionam nas curvas de referência. A *interface* da tabela é projetada para ser intuitiva e de fácil leitura. O utilizador pode navegar pelas tabelas conforme a ordem no qual os pontos foram inseridos ou com base em qualquer uma das outras colunas da tabela, facilitando a localização e comparação de dados específicos. Na Figura 34 é possível verificar a tabela de informações do caso de estudo para o parâmetro **D1**.

Processo	Data	IG	Diâmetro1	Percentil
1 549667	2023-10-09	14	6	entre o 10º e o 25º
2 549667	2023-10-09	25	12	entre o 25º e 50º
3 549667	2023-10-09	33	16	entre o 25º e 50º

Figura 34: Página Informações — Visualização da tabela de informações para o parâmetro **D1**.

É também possível fazer uma análise mais aprofundada das métricas críticas relacionadas ao desenvolvimento da placenta durante a gravidez. Na página “Rácios” o utilizador consegue calcular e examinar dois rácios fundamentais: o rácio PF/PP (Peso Fetal/Peso Placentar) e o rácio PP/PF (Peso Placentar/-Peso Fetal). Na Figura 35, é possível visualizar o cálculo dos rácios relacionados a este estudo de caso. Para calcular estes rácios o utilizador deve escolher os valores que deseja incluir no cálculo dos rácios. Esta seleção é feita clicando nas linhas correspondentes à **IG** desejada e ao valor do parâmetro, **PP** ou **PF**. As linhas selecionadas serão destacadas em azul, indicando que estão prontas para o cálculo. Uma

vez que os valores forem escolhidos, basta clicar no botão “Calcular Rácios”. A aplicação processará instantaneamente os dados selecionados e calculará os dois rácios mencionados acima, proporcionando uma análise imediata dos resultados. Os valores dos rácios calculados são apresentados numa tabela, tornando-os facilmente acessíveis e prontos para interpretação.

Cálculo dos rácios

Para calcular os rácios (rácio1-PF/PP e rácio2-PP/PPF) é necessário:

- Selecionar os valores a calcular (PP- Peso placentar e PF- Peso fetal), clicando nas linhas correspondentes que irão ficar a azul.
- Clicar no botão “Calcular Rácios”.
- Visualizar os valores na tabela.

Processo Data IG Peso

1	2023-10-26	14	34
2	2023-10-26	25	260
3	2023-10-26	33	370

Showing 1 to 3 of 3 entries

Calcular Rácios

Processo IG rácio1 rácio2

1	14	0.971	1.03
2	25	3.3	0.303
3	33	6.027	0.166

Showing 1 to 3 of 3 entries

Valores dos rácios adicionados com sucesso!

Figura 35: Página Rácios — Visualização da tabela com os valores dos rácios calculados.

A aplicação “PlacentalGrowth” oferece uma funcionalidade que permite aos utilizadores criar relatórios em PDF abrangentes contendo todas as informações e análises realizadas durante a sua sessão na aplicação. Este relatório é uma ferramenta valiosa para documentar e partilhar os resultados da análise de parâmetros placentares ao longo da gestação. O Relatório em PDF Inclui:

- **Informações:** Informações acerca da aplicação e dados do processo do paciente.
- **Gráficos das curvas de crescimento de referência:** Todos os gráficos exibidos na página “Gráficos”, incluindo os pontos adicionados pelo utilizador, são automaticamente incluídos no relatório em PDF. Isso permite ao utilizador capturar visualmente o crescimento da placenta e do feto ao longo da gestação.
- **Tabelas de Informações:** As tabelas da página “Informações” que apresentam dados detalhados dos pontos adicionados aos gráficos também fazem parte do relatório. Isso inclui alguns dados da aba “Informações”, os valores de **IG** e parâmetros correspondentes e o percentil onde cada ponto caiu.

- **Tabela de Rácios Calculados:** Se o utilizador realizar cálculos de rácios na página “Rácios”, a tabela dos rácios calculados também será incluída no relatório em PDF.

Após concluir a análise e interação com a aplicação, o utilizador pode clicar no botão “Transferir relatório PDF” na página “Gráficos”. O relatório em PDF será automaticamente gerado com base em todas as informações inseridas e ações realizadas durante a sessão do utilizador. O relatório é salvo no formato PDF, tornando-o facilmente acessível e legível numa variedade de dispositivos. O relatório gerado neste estudo de caso está disponível para consulta no Apêndice H.

Por fim, o utilizador pode consultar a página de “Ajuda” da aplicação, ilustrada na Figura 36. Nesta página é possível consultar um guia de utilização que fornece uma explicação detalhada de como utilizar a aplicação “PlacentalGrowth”. Este guia está disponível para consulta na aplicação ou no repositório do [GitHub](#).

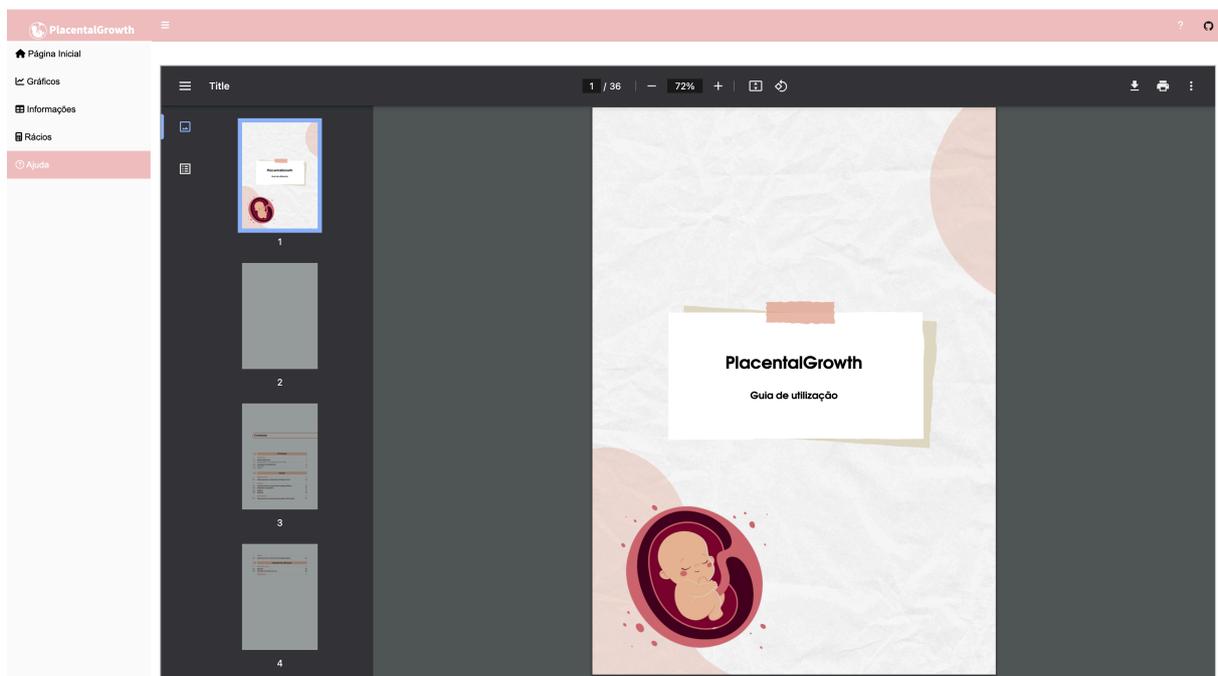


Figura 36: Página Ajuda – Guia de utilização da aplicação.

Capítulo 8

Discussão

Ao longo do tempo, o desenvolvimento de curvas de percentis para o **PF** em relação à **IG** tem sido realizado para fornecer orientações valiosas aos profissionais de saúde e pais relativamente ao crescimento e desenvolvimento de fetos e recém-nascidos [Grantz, 2021, Aybuke et al., 2023]. Os resultados do presente estudo mostram que também é possível produzir curvas de percentis para parâmetros placentares como **D1**, **D2**, **EP** e **PP**, para tirar conclusões adicionais sobre o bem-estar do feto, uma vez que estudos anteriores demonstram uma associação estatisticamente significativa entre a forma e o tamanho da placenta e complicações durante a gravidez, bem como os resultados de saúde a longo prazo dos indivíduos [Shehata et al., 2011, Mitsuda et al., 2020].

A análise exploratória dos dados, apresentada na Secção 7.1, forneceu uma base para o desenvolvimento dessas curvas de percentis. O objetivo principal deste estudo é a exploração dos parâmetros placentares e construção das curvas de crescimento. Contudo, é importante destacar que, a análise de outras variáveis, como o local de recolha das amostras, paridade da mãe, género do feto, forma da placenta e idade materna, também desempenham um papel relevante, fornecendo uma visão mais abrangente do contexto e da natureza das variáveis envolvidas. Há estudos que demonstraram que o peso da placenta pode ser influenciado por estas variáveis [Wallace et al., 2013, Haavaldsen et al., 2011, Asghar-nia et al., 2008]. Os dados geográficos mostraram uma ampla gama de regiões e hospitais em Portugal, onde as amostras foram recolhidas, evidenciando diversidade geográfica. No que diz respeito à paridade da mãe, a distribuição nos dados amostrais mostra variação no número de filhos das mães representadas neste estudo, mas a maioria tem apenas um filho. No que se refere ao género do feto, os dados revelam que a maioria das amostras corresponde a fetos do sexo masculino, no entanto, só existem 85 amostras de diferença entre os dois géneros. A análise da forma da placenta revela que a maioria das placentas apresenta uma forma normal. Quanto à idade materna, a maioria das mães tem idades compreendidas entre 25 e 40 anos.

Também foi possível observar tendências claras na análise exploratória das variáveis de interesse,

como **D1**, **D2**, **EP**, **PP**, e **PF**, em relação à **IG**. À medida que a **IG** aumenta, observou-se um crescimento progressivo em **D1**, **D2**, **PP**, e **PF**, sugerindo uma correlação positiva entre essas variáveis e a **IG**. Observou-se também um aumento da variabilidade com o aumento da **IG**, indicado pelos valores crescentes de **DP**, o que é algo expectável, pois à medida que se aproxima do fim da gravidez existe mais diferenças entre os fetos e placentas. O aumento nos valores médios ao longo das semanas é apoiado pelas estatísticas descritivas, incluindo média, mediana, assimetria e curtose. Vale destacar que a assimetria próxima de zero sugere uma distribuição aproximadamente simétrica dos dados, enquanto a curtose com valores superiores a 3 indica distribuições com concentração em torno da média e apresentam caudas mais pesadas em comparação com a distribuição normal. As estatísticas desejadas são uma média de 0, variância de 1, assimetria de 0 e curtose de 3.

A **EP** apresentou um aumento menos pronunciado, indicando uma relação mais complexa com a **IG**. A assimetria e a curtose das distribuições da **EP** indicaram uma assimetria positiva e leptocurtose na maioria dos casos. As observações para todos os parâmetros estão alinhadas com estudos anteriores que destacam o desenvolvimento progressivo da placenta e do feto ao longo da gestação [Turco and Moffett, 2019, DiPietro et al., 2015]. A presença de curtose positiva nos dados indica a necessidade de utilização de métodos que possam abordá-la, como o método **GAMLSS** implementado, para evitar percentis ajustados distorcidos [Borghi et al., 2006]. Além disso, é importante destacar que a normalidade das variáveis de resposta foi avaliada, e os resultados indicaram que as distribuições não seguem estritamente uma distribuição normal. Isso tem implicações nas análises estatísticas subsequentes, e é crucial considerar essa não normalidade ao interpretar os resultados. Estas tendências dos parâmetros são de grande importância, uma vez que podem servir como referência para profissionais de saúde e futuros estudos na área da obstetria.

Considerando a associação conhecida entre o gênero do feto e o peso ao nascer, pode ser importante categorizar as curvas de percentis em curvas específicas para o gênero feminino e masculino. Embora alguns estudos concluam não haver diferenças significativas no peso placentar entre neonatos do gênero masculino e feminino [Tamayev et al., 2020, Asgharnia et al., 2008, Nogueira et al., 2019], foi encontrado alguns casos onde existiu diferenças significativas. Para o **D1**, observou-se uma diferença estatisticamente significativa na semana 39, para o **D2** nas semanas 12, 14 e 39 e para o **PF** nas semanas 16 e 40. O **PP** não apresentou diferenças significativas em nenhuma amostra. Apesar de ter sido identificado algumas diferenças de gênero estatisticamente significativas em relação aos parâmetros em certas **IG**, é importante reconhecer que essas variações foram relativamente limitadas, representando apenas 6,2% dos casos totais. Este cenário sugere que, em geral, as curvas de percentis para parâmetros placentares

podem ser consideradas um conjunto único, independentemente do género fetal e, por isso, foi optado por não categorizar as curvas por género.

Para o desenvolvimento das curvas de crescimento foi importante selecionar o melhor método para ajustar os dados, pois curvas de percentis imprecisas podem levar a interpretações incorretas sobre o desenvolvimento da placenta/feto, resultando em cuidados clínicos subótimos. Conforme os critérios para a seleção de métodos, presentes na Secção 4.2, os métodos **GAMLSS**, **LMS**, **LMST** e **LMSP** foram testados nos dados. Inicialmente, diferentes funções de alisamento, *P-splines*, *Splines* cúbicas, Redes neurais e *Loess* foram avaliadas num modelo **GAMLSS**, sem nenhuma suposição de distribuição, uma vez que são conhecidas pela sua flexibilidade para ajustar curvas sem escolher antecipadamente uma forma rígida para a função subjacente [Currie and Durban, 2002]. A justificação para esta abordagem reside na simplificação do processo de modelação, pois incorporar simultaneamente as funções de alisamento e as distribuições aumentaria consideravelmente a complexidade do processo de modelação.

Analisando os resultados para todos os parâmetros, é possível verificar que as *P-splines* destacaram-se como a função de alisamento preferencial para ajustar os dados. No entanto, as diferenças com as restantes funções de alisamento foram mínimas na maioria dos casos, as *P-splines* apresentaram um desempenho notável em todos os critérios avaliados e, por isso, foram selecionadas para inclusão nos modelos testados para todos os métodos e parâmetros em estudo. As *P-splines* oferecem vantagens sobre outros alisadores, como as *splines* cúbicas, devido à sua eficiência computacional, permitindo que os modelos sejam ajustados a conjuntos de dados maiores [Currie and Durban, 2002].

Após a identificação da função de alisamento mais apropriada, procedeu-se ao teste de várias distribuições em diferentes métodos. Diferentemente da função de alisamento escolhida, que permaneceu constante para ajustar os diversos parâmetros em análise, a seleção dos métodos e distribuições específicas para o modelo global variou para cada parâmetro.

No contexto do parâmetro **D1**, a procura por um modelo eficaz e uma distribuição apropriada revelou dois modelos que se destacaram notavelmente: o **GAMLSS** com a distribuição **BCT** e o modelo **LMST** que também utiliza a distribuição **BCT**. Este destaque é fundamentado pelos valores elevados de **R²** e na redução dos erros, refletidos nos valores relativamente baixos de **MSE**, **AIC** e **BIC**. Além disso, a análise dos gráficos de diagnóstico confirmou a qualidade desses modelos. Dessa forma, optou-se pelo modelo **GAMLSS** com a distribuição **BCT** para representar a curva de crescimento de percentis de referência para o parâmetro **D1**, devido ao seu bom desempenho. Esta curva de percentis conseguiu capturar com precisão o comportamento dos dados, e a concordância entre as percentagens nominais e da amostra sugerem um bom ajuste do modelo.

No que diz respeito ao parâmetro **D2**, o modelo **GAMLSS** com a distribuição **BCT** e o modelo **LMST**, indicaram um desempenho notável. A escolha destes modelos foi aliciada por métricas de desempenho superiores relativamente aos outros modelos. Ambos os modelos exibiram altos valores de R^2 , indicando um bom ajuste aos dados. Além disso, os valores de **MSE**, **AIC** e **BIC** foram notavelmente baixos, sugerindo uma minimização dos erros entre os valores observados e previstos, bem como a eficiência do modelo com parcimónia. Os gráficos de diagnóstico adicionalmente reafirmaram a qualidade destes modelos, o que fortaleceu a escolha do modelo **GAMLSS** com a distribuição **BCT** como o representante da curva de crescimento de percentis de referência para o parâmetro **D2**. Essa decisão baseou-se principalmente na consistência de desempenho em relação aos outros métodos testados e apesar das métricas serem inferiores ao modelo **LMST**, os gráficos de diagnóstico revelaram um ajuste melhor para o **GAMLSS** com **BCT**, reforçando a necessidade de várias técnicas de diagnóstico para se fazer escolhas corretas. As curvas de crescimento de percentis geradas pelo modelo escolhido conseguiram retratar com precisão o comportamento dos dados do parâmetro **D2**. A concordância entre as percentagens nominais e as da amostra sugeriram uma adequada calibração do modelo.

Para o parâmetro **EP**, dois modelos destacaram-se de maneira proeminente na tentativa de modelar o seu crescimento: o modelo **LMST** e o modelo **GAMLSS** com a distribuição **BCT**. A seleção destes modelos foi motivada por métricas de desempenho e gráficos de diagnóstico preferíveis relativamente a outras alternativas consideradas. Estes modelos mostraram valores muito mais baixos de R^2 , comparativamente a **D1** ou **D2**, mostrando uma maior dificuldade, que já tinha sido previamente destacada na análise exploratória, para ajustar estes dados. Dentro das possibilidades de ajuste testadas, ambos os modelos selecionados tiveram ajustes razoáveis. No entanto, o modelo **LMST** foi escolhido como o modelo final para representar a curva de crescimento de percentis de referência para o parâmetro **EP**. Existiu uma concordância entre as percentagens nominais e aquelas derivadas da amostra, indicando que a curva selecionada conseguiu representar de maneira precisa o comportamento dos dados.

Em relação ao parâmetro **PP**, também foram selecionados dois modelos que se destacaram entre os vários testados: o modelo **LMST** e o modelo **GAMLSS** com a distribuição **JSU**. Estas escolhas foram fundamentadas num desempenho notável de ambos os modelos, conforme evidenciado por diversas métricas de avaliação e gráficos de diagnóstico. Os dois modelos exibiram valores de R^2 notavelmente elevados, indicando uma boa adaptação aos dados do parâmetro **PP**. Além disso, os valores do **MSE**, bem como o **AIC** e **BIC**, revelaram níveis de erro reduzidos e a eficácia do modelo. As análises dos gráficos de diagnóstico aprofundaram a validação desses modelos, consolidando a decisão de adotar o modelo **LMST** como a representação das curvas de crescimento de percentis de referência para o parâmetro **PP**.

Esta escolha baseou-se no desempenho consistente relativamente às alternativas consideradas. As curvas de crescimento de percentis desenvolvidas com base nesse modelo proporcionaram uma representação precisa da dinâmica do parâmetro **PP**. A concordância entre as percentagens nominais e aquelas obtidas na amostra realça a boa qualidade do ajuste.

No que se refere ao parâmetro **PF**, a sua modelação foi a mais desafiadora entre todos os parâmetros devido à natureza dos seus dados. A tentativa de utilizar a função de alisamento baseada em redes neurais, bem como os modelos **GAMLSS** com distribuições **BCCG**, **BCPE**, e os modelos **LMS** e **LMSP**, resultou em dificuldades significativas. Estes métodos não foram capazes de se ajustar adequadamente aos dados, tornando-se impossível traçar curvas de crescimento coerentes. No entanto, entre os modelos avaliados, estes destacaram-se de maneira notável, apesar das complexidades associadas ao parâmetro **PF**. Por exemplo, estes modelos alcançaram o maior valor de R^2 entre todos os parâmetros, indicando uma forte capacidade de ajuste aos dados. No entanto, é relevante mencionar que esse valor de R^2 foi acompanhado por valores mais elevados de **MSE**, **AIC**, e **BIC**. Estes resultados ressaltam a complexidade inerente à modelação do parâmetro **PF** e a necessidade de um equilíbrio delicado entre ajuste aos dados e parcimónia no modelo. O modelo escolhido para representar este parâmetro foi o **LMST**, que teve um ajuste razoável conforme os gráficos de diagnóstico e métricas indicaram. O segundo melhor modelo, **GAMLSS** com a distribuição **BCT** também apresentou um ajuste aceitável. As curvas de crescimento de percentis geradas pelo modelo escolhido para o parâmetro **PF** conseguiram retratar com precisão o comportamento dos dados observados. A concordância entre as percentagens nominais e as da amostra sugeriram uma adequada calibração do modelo, indicando que as curvas de crescimento são representativas do crescimento fetal ao longo da gestação para o parâmetro **PF**.

Como resultado, a análise abrangente de todos os parâmetros estudados aponta para um padrão notável. Independentemente do parâmetro em questão, a utilização quer de um modelo **GAMLSS** com a distribuição **BCT**, quer o modelo **LMST** juntamente com a função de alisamento *P-splines* oferecem um ajuste confiável aos dados. Este sucesso abrange desde a modelação do **D1** até ao **PF**, comprovando a versatilidade destas abordagens no contexto dos parâmetros placentares e fetais e também a precisão dos critérios de seleção dos modelos.

Estas escolhas metodológicas foram feitas com base em critérios rigorosos de desempenho, que incluem métricas como o R^2 , o **MSE**, o **AIC** e o **BIC**. Bem como, gráficos de diagnóstico como *worm plot*, estatísticas Z e gráficos de resíduos. Ao optar por esta metodologia esta-se a assegurar que os modelos oferecem um bom ajuste, possibilitando a criação de curvas de crescimento de referência altamente precisas para cada parâmetro. Estas curvas de referência têm uma aplicação direta e valiosa para os

profissionais de saúde, permitindo-lhes avaliar com eficácia o desenvolvimento placentar num contexto clínico. Ao comparar, por exemplo, o **D1** com a curva de percentil de 3% para uma determinada **IG**, é possível identificar um subdesenvolvimento placentar esperado durante a gravidez. Estas informações são cruciais para a deteção precoce de potenciais complicações e fornecem orientações importantes para o acompanhamento do bem-estar materno e fetal. Como foi descrito na Subsecção 2.1.3, a análise destes parâmetros e as suas curvas é muito importante, dado que existe uma associação clínica bem documentada na literatura.

É fundamental destacar que esta metodologia tem um potencial de aplicação mais amplo e pode ser utilizada numa variedade de estudos que procuram desenvolver curvas de crescimento padrão em diferentes contextos. Para obter informações mais detalhadas sobre esta abordagem, recomenda-se a consulta de referências pertinentes como [Stasinopoulos et al. \[2017\]](#), [Rigby et al. \[2019\]](#), que fornecem uma base para a implementação bem-sucedida desta metodologia em estudos futuros.

Durante o curso desta investigação, deparou-se com um desafio significativo e constante: a escassez de dados e estudos que permitam comparar e contextualizar os resultados desta investigação no âmbito dos parâmetros placentares. Ao longo dos anos produziram-se curvas de referência de percentis, principalmente para o **PF** em função da **IG**, mas estes gráficos são geralmente restritos ao terceiro trimestre da gravidez e tendo em consideração casos específicos como o género do feto, paridade e etnia [[Flatley et al., 2022](#), [Panti et al., 2012](#), [Wallace et al., 2013](#)]. Estas condições não se alinham com as deste estudo, uma vez que se explorou diversas **IG** incluindo fases mais iniciais como as 12 semanas, não se categorizou por géneros ou paridade e apenas foram usados dados da população portuguesa. Tornando assim, mais complicada e limitada a disponibilidade de investigações prévias nesta área e restringindo, assim, a capacidade de alinhar as descobertas deste estudo com trabalhos anteriores.

No entanto, apesar dessa carência de dados comparativos, conseguiu-se identificar alguns estudos notáveis que conseguiram explorar eficazmente os parâmetros placentares. Um exemplo é o estudo de [Alves et al. \[2022\]](#), onde os autores usaram uma parte do conjunto de dados utilizado nesta investigação para construir as curvas de crescimento para os parâmetros placentares **D1**, **D2**, **EP** e **PP**, aplicando a regressão quantílica. Os autores, apenas apresentam os gráficos das curvas de crescimento, por isso foi feita uma comparação visual. As curvas produzidas neste estudo têm diferenças com as de [Alves et al. \[2022\]](#), tendo um perfil mais alisado e visualmente captam melhor o comportamento dos dados, mostrando um melhor ajuste utilizando a metodologia imposta nesta investigação. Outro estudo que utiliza uma parte do conjunto de dados desta investigação é o de [Nogueira et al. \[2019\]](#). São apresentadas curvas de crescimento e valores dos percentis para os parâmetros **PP** e **PF**. Comparando os valores dos

percentis e gráficos, existem diferenças nos valores estimados para cada percentil em cada **IG** e na forma das curvas de percentis. Como não são apresentadas métricas ou gráficos de diagnóstico, não é possível fazer uma comparação de qualidade de ajuste com as curvas deste estudo, mas aparentemente ambos os estudos têm curvas que podem representar com confiança os parâmetros **PP** e **PF**.

Mais um estudo de curvas de percentis é o de [Flatley et al. \[2022\]](#). Este estudo recente demonstra o uso bem-sucedido do método **GAMLSS** para criar curvas de referência relacionadas ao **PP**, uma vez que obtiveram uma boa qualidade de ajuste. Os resultados deste estudo ilustram claramente a eficácia da abordagem **GAMLSS** na análise de parâmetros placentares, fornecendo um precedente valioso para este trabalho. No entanto, como o estudo é feito categorizando por gênero e paridade para **IG** entre 28 e 42 semanas, não podem ser realizadas comparações com as curvas desenvolvidas nesta investigação.

Além disso, o estudo de [Borghi et al. \[2006\]](#) também é pertinente. Este estudo desempenhou um papel fundamental na construção dos padrões de crescimento infantil da Organização Mundial da Saúde. Os autores destacaram a viabilidade do uso de modelos **GAMLSS** com distribuições semelhantes às exploradas neste trabalho e também dos modelos **LMS**, **LMST** e **LMSP**, devido ao seu potencial para produzir resultados promissores quando a idade é tratada como uma variável contínua, semelhante à **IG**.

Portanto, embora haja uma falta de referências diretas, estes estudos citados demonstram que a análise de parâmetros placentares, especialmente com a utilização de métodos como o **GAMLSS**, tem potencial e capacidade para produzir resultados significativos.

De forma a conciliar todo o trabalho feito com as curvas de crescimento e torná-las acessíveis para análise fácil e eficaz, foi desenvolvida a aplicação “PlacentalGrowth”. Esta ferramenta responde a uma necessidade crucial na obstetrícia, fornecendo uma solução prática para a interpretação de dados placentares e fetais ao longo da gravidez. A aplicação foi projetada com foco na usabilidade, fornecendo uma interface intuitiva e amigável para os utilizadores. A facilidade de inclusão de informações do paciente e de pontos de dados para análise é um ponto forte da aplicação, tornando-a acessível a uma ampla gama de utilizadores, incluindo profissionais de saúde com diferentes níveis de experiência em análise de dados. Outro ponto notável é a capacidade de os utilizadores personalizarem a análise ao adicionar manualmente ou carregar pontos de dados de pacientes específicos através de ficheiros CSV. Isso permite que a aplicação seja aplicada a uma ampla variedade de cenários clínicos e contextos de investigação. A flexibilidade na entrada de dados é uma característica fundamental que a torna relevante para uma audiência diversificada.

A aplicação permite que os utilizadores calculem rácios clínicos essenciais, como o rácio **PF/PP** e **PP/PF**. Estes rácios são de grande importância na avaliação do desenvolvimento placentar durante a

gravidez e podem ser usados para tomar decisões clínicas bem fundamentadas. Sabe-se que o **R-PP** e **R-PF** são parâmetros importantes para o equilíbrio entre o crescimento fetal e placentar e considerando a capacidade de reserva funcional da placenta, estes podem ser os maiores preditores de restrição de crescimento intrauterino e doenças mais tarde na vida do que o **PP** e o **PF** isoladamente [Nogueira et al., 2019]. Na Subsecção 2.1.3, é possível verificar mais informações acerca destes rácios. O facto de a aplicação automatizar o seu cálculo economiza tempo e minimiza erros potenciais.

A capacidade de gerar relatórios em PDF abrangentes é um dos recursos mais significativos da aplicação. Isso permite que os utilizadores documentem e compartilhem facilmente os resultados das análises, tornando-a uma ferramenta valiosa para relatórios médicos, investigação científica e colaboração interdisciplinar. A capacidade de transmitir informações detalhadas de maneira clara e concisa é fundamental em ambientes clínicos e de investigação.

Durante o estudo de caso apresentado na Secção 7.4, foi possível analisar parâmetros da placenta e fetais ao longo da **IG** inserindo os dados na aplicação via ficheiros CSV e manualmente. O resultado de todas as análises foi guardado num relatório PDF, presente no Apêndice H, e revela informações valiosas sobre o desenvolvimento placentar e o crescimento fetal ao longo da gravidez. Analisando o relatório, nota-se que o **D1** mostra um aumento constante ao longo das avaliações, com os dados mantendo-se consistentemente entre o 10.º e o 50.º percentil. Isso sugere um desenvolvimento placentar saudável nesse aspeto, apesar de nas 14 semanas apresentar um percentil mais baixo. Para o **D2** os seus percentis variam de 25.º a 75.º, indicando um desenvolvimento moderado. Ambos os parâmetros **D1** e **D2** refletem uma evolução positiva ao longo das avaliações. Para a **EP** observou-se novamente um aumento na espessura com o progresso da gravidez, com os percentis oscilando entre o 10.º e o 75.º. Isso sinaliza um desenvolvimento placentar consistente, com valores na faixa esperada. Para o **PP**, os resultados mostram um crescimento moderado ao longo da gravidez, com a paciente mantendo-se entre o 10.º e o 90.º percentil, refletindo variações normais. Por fim, o **PF** revelou que o feto também se está a desenvolver conforme o esperado. Os percentis variam de 25.º a 90.º, com valores aumentando à medida que a gravidez avança, o que é uma indicação positiva. A Tabela 6 destaca a importância dos rácios **R-PP** e **R-PF** na avaliação do desenvolvimento da placenta. Os resultados indicam que esses rácios variam ao longo da gravidez, sugerindo diferentes fases de crescimento e interação entre o feto e a placenta. Analisando os valores obtidos com os esperados para cada **IG** em cada rácio no estudo de Nogueira et al. [2019], verifica-se que estes se encontram em intervalos normais, indicando um bom desenvolvimento do feto e da placenta. Globalmente, a análise destes dados no relatório reforça a utilidade da aplicação “PlacentalGrowth” na monitorização e avaliação do desenvolvimento placentar e fetal.

A aplicação “PlacentalGrowth” tem, portanto, o potencial de ter um impacto significativo na prática clínica. Os resultados obtidos com a aplicação podem ser usados para monitorizar o desenvolvimento da placenta durante a gravidez e identificar anormalidades precoces. Isso pode levar a intervenções mais oportunas e melhores resultados de saúde materna e fetal. No entanto, também é importante reconhecer as limitações da aplicação. A qualidade dos resultados depende da precisão dos dados de entrada e da seleção adequada dos pontos. Além disso, a aplicação atualmente concentra-se em cinco parâmetros específicos. Para abordar a diversidade dos parâmetros placentares e a complexidade dos dados, considerações futuras podem envolver a expansão da aplicação para incluir um leque mais amplo de parâmetros.

Capítulo 9

Conclusões e trabalho futuro

Ao longo deste estudo, foi realizada uma exploração detalhada dos parâmetros placentares e o seu impacto na saúde materna e fetal. Foram feitas análises abrangentes, desenvolvidas abordagens metodológicas inovadoras e esses conhecimentos foram aplicados para a construção das curvas de crescimento de referência no intervalo das 12 às 41 semanas de gestação. Os resultados obtidos revelaram informações importantes sobre o desenvolvimento placentar e o **PF** ao longo da gravidez, representando uma contribuição significativa para a comunidade médica, especialmente para obstetras e clínicos gerais. Estas curvas de percentis fornecem informações valiosas sobre o desenvolvimento embrionário e placentar, e têm o potencial de serem úteis na identificação precoce de problemas de saúde fetal.

À medida que se enfrenta desafios contemporâneos relacionados à baixa taxa de natalidade e ao aumento da idade materna, a promoção da saúde placentar, fetal e infantil torna-se uma prioridade fundamental. As curvas de percentis desenvolvidas oferecem uma compreensão detalhada de como os parâmetros **D1**, **D2**, **EP**, **PP** e **PF** evoluem ao longo da gravidez. Além disso, estas curvas representam padrões populacionais que podem ser particularmente relevantes no contexto do resultado fetal.

No desenvolvimento das curvas de crescimento, foi constatado que a aplicação de modelos **GAMLSS** e **LMST**, aliados à função de alisamento *P-splines* e distribuição **BCT**, oferecem uma boa abordagem para modelar os parâmetros placentares e fetais, evidenciado por critérios promissores de qualidade de ajuste e permitindo assim, a criação de curvas de percentis de referência precisas. Mais precisamente, a utilização do método **GAMLSS** com **BCT** e *P-splines* para modelar os parâmetros **D1** e **D2** e do método **LMST** com *P-splines* para modelar os parâmetros **EP**, **PP** e **PF**.

Além disso, a aplicação *web* “PlacentalGrowth” desenvolvida representa um marco na acessibilidade e usabilidade desse conhecimento, tornando-o disponível para profissionais de saúde e investigadores de maneira simples e interativa. A possibilidade de gerar relatórios personalizados, calcular rácios importantes e fornecer informações detalhadas sobre cada ponto de dados adicionado à aplicação aprimora a capacidade de tomar decisões informadas.

Um dos principais desafios encontrados nesta investigação foi a relativa escassez de estudos existentes na área de parâmetros biométricos placentares. Antes deste estudo, as curvas de crescimento placentar para **IG** compreendidas entre 12 e 41 semanas, para a população portuguesa e para os parâmetros **D1**, **D2**, **EP**, **PP** e **PF**, utilizando metodologias como o **GAMLSS** ou o **LMST** eram amplamente inexploradas. Portanto, a literatura limitada e as metodologias estabelecidas neste campo específico colocaram desafios na elaboração de comparações e na validação dos resultados. No entanto, a abordagem pioneira deste estudo preenche uma lacuna crucial na literatura, oferecendo informações valiosas sobre os padrões de crescimento placentar e o seu significado clínico.

Entretanto, reconhece-se que, apesar das conquistas, este trabalho enfrenta limitações, e há caminhos para investigações futuras. É fundamental continuar a refinar as metodologias e expandir as amostras de estudo para aprimorar a precisão e aplicabilidade das curvas de crescimento de referência placentares e fetais.

Numa perspetiva de trabalho futuro, poderia ser realizada uma investigação mais ampla e inclusiva para aprimorar a precisão das curvas de crescimento de referência. A inclusão de amostras mais diversificadas de gestantes, abrangendo diferentes grupos étnicos, idades gestacionais e condições médicas, pode permitir uma compreensão mais completa das variações no desenvolvimento placentar e fetal. A integração de mais parâmetros placentares, além dos parâmetros explorados nesta dissertação, dado que a inclusão de outros parâmetros placentares relevantes pode permitir a elaboração de diagnósticos mais precisos. A expansão da aplicação “PlacentalGrowth” para fornecer suporte adicional aos profissionais de saúde, como a integração de sistemas de registo médicos eletrónicos que guardem os dados dos processos estudados na aplicação, para assim aprimorar a acessibilidade e utilidade desta ferramenta. Explorar e analisar de uma maneira mais aprofundada os rácios **R-PP** e **R-PF** dada a sua importância biológica documentada na literatura. E por fim, estudar a relação de outras variáveis com o desenvolvimento placentar como, por exemplo, a paridade ou a idade da mãe, para tentar perceber se existem outros fatores que influenciem os parâmetros placentares.

Em última análise, esta dissertação reforça a importância da investigação dos parâmetros placentares, enfatizando como a combinação de estatísticas, modelação matemática e desenvolvimento de aplicações pode resultar em avanços significativos na medicina e na saúde materna e fetal. Apesar dos desafios enfrentados, os resultados obtidos demonstraram a viabilidade e eficácia da abordagem desenvolvida. Espera-se que este trabalho possa inspirar futuras investigações e, ao mesmo tempo, beneficiar diretamente as vidas das gestantes e os seus bebés, promovendo uma gravidez saudável e segura. É importante reconhecer que este estudo não é o ponto final, mas sim um ponto de partida.

Bibliografia

- Edgar Acuña and Caroline Rodriguez. The treatment of missing values and its effect on classifier accuracy. In *Classification, Clustering, and Data Mining Applications*, pages 639–647. Springer Berlin Heidelberg, 2004.
- Murray Aitkin. Modelling variance heterogeneity in normal regression using glim. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 36:332–339, 1987.
- Hirotsugu Akaike. Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*, pages 267–281, 1973.
- C. Akantziliotou, R.A. Rigby, and D.M. Stasinopoulos. A framework for modelling overdispersed count data, including the poisson-shifted generalized inverse gaussian distribution. *Computational Statistics & Data Analysis*, 53:381–393, 2008.
- Setareh Akhavan, Sedigheh Borna, Alireza Abdollahi, Mamak Shariat, and Narges Zamani. Pathologic examination of the placenta and its benefits in treatment plan or follow-up of patients: a cross-sectional study. *European Journal of Medical Research*, 27:1–5, 2022.
- B. Almog, F. Shehata, S. Aljabri, I. Levin, E. Shalom-Paz, and A. Shrim. Placenta weight percentile curves for singleton and twins deliveries. *Placenta*, 32:58–62, 2011.
- Samuel Alves, Ana Cristina Braga, and Rosete Nogueira. Percentile growth curves for placenta measures: A dynamic shiny application. *Computational Science and Its Applications – ICCSA 2022 Workshops*, pages 543–554, 2022.
- LM Anchieta, CC Xavier, EA Colosimo, and MF Souza. Weight of preterm newborns during the first twelve weeks of life. *Braz J Med Biol Res*, 36:761–770, 2003.
- M. Asgharnia, N. Esmailpour, M. Poorghorban, and Z. Atrkar-Roshan. Placenta weight and its association with maternal and neonatal characteristics. *Acta Medica Iranica*, 46:467–472, 2008.

- Muhammad Ashfaq, Muhammad Zahoor Janjua, and Muhammad Aslam Channa. Effect of gestational diabetes and maternal hypertension on gross morphology of placenta. *Journal of Ayub Medical College Abbottabad*, 17, 2005.
- Eran Ashwal, Jasmine Ali-Gami, Amir Aviram, Stefania Ronzoni, Elad Mei-Dan, John Kingdom, and Nir Melamed. Contribution of second trimester sonographic placental morphology to uterine artery doppler in the prediction of placenta-mediated pregnancy complications. *J. Clin. Med.*, 11:6759, 2022.
- Yazici Aybuke, Buyuktiryaki Mehmet, Sari Fatma Nur, Akin Mustafa Senol, Ertekin Omer, and Alyamac Dizar Evrim. Comparison of different growth curves in the assessment of extrauterine growth restriction in very low birth weight preterm infants. *Archives de Pédiatrie*, 30:31–35, 2023.
- Rebecca N. Baergen. *Manual of pathology of the human placenta*. Springer New York, NY, 2nd edition, 2011.
- Palanikumar Balasundaram and Indirapriya Darshini. Avulakunta. *Human Growth and Development*. StatPearls Publishing, Treasure Island (FL), 2022.
- DJP Barker, C. Osmond, KL Thornburg, Eero Kajantie, and Johan G. Eriksson. The lifespan of men and the shape of their placental surface at birth. *Placenta*, 32:783–787, 2011.
- Benjamin Baumer and Dana Udwin. R markdown. *WIREs Computational Statistics*, 7:167–177, 2015.
- Chris Beeley. *Web application development with R using Shiny*. Packt Publishing Ltd, 2016.
- Christopher M. Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.
- Hossein Bonakdari and Mohammad Zeynoddin. Chapter 5 - goodness-of-fit & precision criteria. In *Stochastic Modeling*, pages 187–264. Elsevier, 2022.
- Elaine Borghi, Mercedes de Onis, Cutberto Garza, Jan Van den Broeck, Edward A. Frongillo, Laurence Grummer-Strawn, S. Van Buuren, H. Pan, L. Molinari, Reynaldo Martorell, et al. Construction of the world health organization child growth standards: Selection of methods for attained growth curves. *Statistics in Medicine*, 25:247–265, 2006.
- George EP Box and David R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26:211–243, 1964.

- Ivo Brosens, Robert Pijnenborg, Lisbeth Vercruyssen, and Roberto Romero. The "great obstetrical syndromes" are associated with disorders of deep placentation. *American Journal of Obstetrics and Gynecology*, 204:193–201, 2011.
- Emily Burns. More than clinical waste? placenta rituals among Australian home-birthing women. *The Journal of Perinatal Education*, 23:41–49, 2014.
- Graham J. Burton and Eric Jauniaux. What is the placenta? *American Journal of Obstetrics and Gynecology*, 213:S6.e1, S6–8, 2015.
- Stef van Buuren and Miranda Fredriks. Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, 20:1259–1277, 2001.
- Rosete Maria Amorim Novais Nogueira Cardoso. *A Multicenter Study of Singleton Placentas Biometric Parameters and Fetal Weight In Function of Gestational-Age*. PhD thesis, Universidade do Minho, 2020.
- Joseph E. Cavanaugh and Andrew A. Neath. The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*, 11, 2019.
- Colin Chen and J. P. Morgan. Growth charts of body mass index (bmi) with quantile regression. *Algorithmic Mathematics and Computer Science*, 5:114–120, 2005.
- Davide Chicco, Matthijs J Warrens, and Giuseppe Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ computer science*, 7, 2021.
- S. Chinn. A new method for calculation of height centiles for preadolescent children. *Annals of Human Biology*, 19:221–232, 1992.
- Tae Yeong Choi, Hye Min Lee, Won Kyoung Park, So Yeong Jeong, and Hwa Sook Moon. Spontaneous abortion and recurrent miscarriage: A comparison of cytogenetic diagnosis in 250 cases. *Obstetrics & Gynecology Science*, 57:518, 2014.
- William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836, 1979.
- William S. Cleveland and Susan J. Devlin. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, 83:596–610, 1988.

- T. J. Cole. Do growth chart centiles need a face lift? *BMJ*, 308:641–642, 1994.
- T. J. Cole. The development of growth references and growth charts. *Annals of Human Biology*, 39: 382–394, 2012.
- T. J. Cole and P. J. Green. Smoothing reference centile curves: The lms method and penalized likelihood. *Statistics in Medicine*, 11:1305–1319, 1992.
- Timothy J. Cole. Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 151:385–406, 1988.
- Iain D Currie and Maria Durban. Flexible smoothing with p-splines: a unified approach. *Statistical Modelling*, 2:333–349, 2002.
- Ralph B. D’agostino, Albert Belanger, and Ralph B. D’Agostino Jr. A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44:316–321, 1990.
- Russel L. Deter, War K. Rossavik, Ronald B. Harrist, HAalock, and Frank P. Mathematic modeling of fetal growth: development of individual growth curve standards. *Obstet. Gynecol.*, 68:156–161, 1986.
- Janet A. DiPietro, Kathleen A. Costigan, and Kristin M. Voegtline. Studies in fetal behavior: Revisited, renewed, and reimagined. *Monographs of the Society for Research in Child Development*, 80:vii, 2015.
- dplyr. Cran - package dplyr, 2023. URL <https://CRAN.R-project.org/package=dplyr>. Acedido em: 23-07-2023.
- DT. Cran - package dt, 2023. URL <https://CRAN.R-project.org/package=DT>. Acedido em: 20-08-2023.
- Gilles R Ducharme, Ali Gannoun, Marie-Claude Guertin, and Jean-Claude Jequier. Reference values obtained by kernel-based estimation of quantile regressions. *Biometrics*, 51:1105–1116, 1995.
- Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11:89–121, 1996.
- Johan G. Eriksson, Eero Kajantie, Kent L. Thornburg, Clive Osmond, and David JP Barker. Mothers body size and placental size predict coronary heart disease in men. *European Heart Journal*, 32:2297–2303, 2011.

Adolfo Fernandes, Altemir Braga, Djair Frade, Jhennifer Nascimento, Joana Oliveira, and Dandara Medeiros. An overview of the development of a new probability distribution: Odd log-logistic skew t-student. *Brazilian Journal of Development*, 7:30536–30555, 2021.

Christopher Flatley, Pol Sole-Navais, Marc Vaudel, Øyvind Helgeland, Dominika Modzelewska, Stefan Johansson, Bo Jacobsson, and Pål Njølstad. Placental weight centiles adjusted for age, parity and fetal sex. *Placenta*, 117:87–94, 2022.

Jennifer Vanessa Freeman. *The construction of growth reference curves for British children, 1990*. PhD thesis, University of London, 1995.

gamlss. Cran - package gamlss, 2023. URL <https://cran.r-project.org/web/packages/gamlss/index.html>. Acedido em: 28-01-2023.

Christopher Gandrud. *Reproducible research with R and R studio*. Chapman and Hall/CRC, 2018.

Marc Garel, Lloyd Izard, Marthe Vienne, David Nerini, Badr Al Ali, Christian Tamburini, and Séverine Martini. A ready-to-use logistic verhulst model implemented in r shiny to estimate growth parameters of microorganisms. *bioRxiv*, 2022.

Th Gasser, W. Köhler, HG Müller, A. Kneip, R. Largo, L. Molinari, and A. Prader. Velocity and acceleration of height growth using kernel estimation. *Annals of Human Biology*, 11:397–411, 1984.

ggplot2. Cran - package ggplot2, 2022. URL <https://cran.r-project.org/web/packages/ggplot2/index.html>. Acedido em: 28-01-2023.

Harvey Goldstein. Efficient statistical modelling of longitudinal data. *Annals of human biology*, 13:129–141, 1986.

Katherine L. Grantz. Fetal growth curves: Is there a universal reference? *Obstetrics and Gynecology Clinics of North America*, 48:281–296, 2021.

P. J. Green and Bernard W. Silverman. *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall/CRC, 1993.

Shumei Guo, Alex F. Roche, Richard N. Baumgartner, W. Cameron Chumlea, and Alan S. Ryan. Kernel regression for smoothing percentile curves: reference data for calf and subscapular skinfold thicknesses in Mexican Americans. *The American Journal of Clinical Nutrition*, 51:908S–916S, 1990.

- Shumei Guo, Roger M. Siervogel, Alex F. Roche, and Wm Cameron Chumlea. Mathematical modelling of human growth: A comparative study. *American Journal of Human Biology*, 4:93–104, 1992.
- C Haavaldsen, SO Samuelsen, and A Eskild. The association of maternal age with placental weight: a population-based study of 536 954 pregnancies. *BJOG: An International Journal of Obstetrics & Gynaecology*, 118:1470–1476, 2011.
- Fawzia A. Habib. Prediction of low birth weight infants from ultrasound measurement of placental diameter and placental thickness. *Annals of Saudi medicine*, 22:312–314, 2002.
- Peter V. Hamill, Terence A Drizd, Clifford L. Johnson, Robert B. Reed, and Alex F. Roche. Nchs growth curves for children birth-18 years. *Vital Health Stat 11*, pages 1–74, 1977.
- Trevor Hastie and Robert Tibshirani. Generalized additive models. *Statistical Science*, 1:297–310, 1986.
- Christina E. Hayward, Samantha Lean, Colin P. Sibley, Rebecca L. Jones, Mark Wareing, Susan L. Greenwood, and Mark R. Dilworth. Placental adaptation: What can we learn from birthweight:placental weight ratio? *Frontiers in Physiology*, 7:28, 2016.
- Patrick J. Heagerty and Margaret S. Pepe. Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in us children. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48:533–551, 1999.
- MJR Healy, J. Rasbash, and Min Yang. Distribution-free estimation of age-related centiles. *Annals of human biology*, 15:17–22, 1988.
- MJR Healy et al. The effect of age grouping on the distribution of a measurement affected by growth. *American Journal of Physical Anthropology*, 20:49–50, 1962.
- Elisa Henning, Marcelo Savio Ramos, Rogério de Aguiar, Ivanete Zuchi Siple, and Luciane Mulazani dos Santos. Para além da computação estatística: o uso do ambiente r para o ensino de métodos numéricos. *RENOTE*, 14, 2016.
- JL Jensen and Wilfrid S Kendall. *Networks and chaos-statistical and probabilistic aspects*, volume 50. CRC Press, 1993.
- Selçuk Kaplan and Eser Kemal Gürcan. Comparison of growth curves using non-linear regression function in japanese quail. *Journal of Applied Animal Research*, 46:112–117, 2018.

- T. Yee Khong, Eoghan E. Mooney, Ilana Ariel, Nathalie CM Balmus, Theonia K. Boyd, Marie-Anne Brundler, Hayley Derricott, Margaret J. Evans, Ona M. Faye-Petersen, John E. Gillan, et al. Sampling and definitions of placental lesions amsterdam placental workshop group consensus statement. *Archives of Pathology and Laboratory Medicine*, 140:698–713, 2016.
- Torvid Kiserud, Alexandra Benachi, Kurt Hecher, Rogelio González Perez, José Carvalho, Gilda Piaggio, and Lawrence D. Platt. The world health organization fetal growth charts: concept, findings, interpretation, and application. *American Journal of Obstetrics and Gynecology*, 218:S619–S629, 2018.
- Roger Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005.
- Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46:33–50, 1978.
- Andréa Cristina Konrath, Silvio Aparecido da Silva, Elisa Henning, Luciane Mulazani dos Santos, Rodrigo Gabriel de Miranda, and Robert Wayne Samohyl. Desenvolvimento de aplicativos web com r e shiny: inovações no ensino de estatística. *Abakós*, 6:55–71, 2018.
- Martin Krzywinski and Naomi Altman. Visualizing samples with box plots. *Nat. Methods*, 11:119–120, 2014.
- Anant Kshirsagar and William Smith. *Growth Curves*, volume 145. CRC Press, 1st edition, 1995.
- Robert J. Kuczmarski, Cynthia L. Ogden, Shumei S. Guo, Laurence M. Grummer-Strawn, Katherine M. Flegal, Zuguo Mei, Rong Wei, Lester R. Curtin, Alex F. Roche, and Clifford L. Johnson. 2000 CDC growth charts for the united states: methods and development. *Vital Health Stat. 11*, (246):1–190, 2002.
- Nan M. Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.
- Michelle Lampl. *Limitation of growth chart curves in terms of individual growth biology*. Springer, New York, NY, 2012.
- Youngjo Lee, John A Nelder, and Yudi Pawitan. *Generalized linear models with random effects: unified analysis via H-likelihood*, volume 153. CRC Press, 2018.
- Lawrence D. Longo and Lawrence P. Reynolds. Some historical aspects of understanding placental development, structure and function. *International Journal of Developmental Biology*, 54:237–255, 2010.

- Maria Teresa Loverro, Edoardo Di Naro, Vittorio Nicolardi, Leonardo Resta, Salvatore Andrea Mastrolia, Federico Schettini, Manuela Capozza, Matteo Loverro, Giuseppe Loverro, and Nicola Laforgia. Pregnancy complications, correlation with placental pathology and neonatal outcomes. *Frontiers in Clinical Diabetes and Healthcare*, 2, 2022.
- Hamdy FF Mahmoud. Parametric versus semi and nonparametric regression models, 2019.
- David E. Matthews. Likelihood-based confidence intervals for functions of many parameters. *Biometrika*, 75:139–144, 1988.
- Saskia Middeldorp. Thrombophilia and pregnancy complications: Cause or association? *Journal of Thrombosis and Haemostasis*, 5:276–282, 2007.
- Naomi Mitsuda, Naw Awn JP, Masamitsu Eitoku, Nagamasa Maeda, Mikiya Fujieda, Narufumi Sukanuma, Michihiro Kamijima, Shin Yamazaki, Reiko Kishi, Nobuo Yaegashi, et al. Association between maternal hemoglobin concentration and placental weight to birthweight ratio: The japan environment and children's study (jecs). *Placenta*, 101:132–138, 2020.
- Mariam L. Mohamed, Magda M. Elbeily, Maisara M. Shalaby, Yara H. Khattab, and Omima T. Taha. Umbilical cord diameter in the prediction of foetal growth restriction: a cross sectional study. *J. Obstet. Gynaecol.*, 42:1117–1121, 2022.
- Luiz R. Nakamura, Thiago G. Ramires, Ana J. Righetto, Rodrigo R. Pescim, Fernanda V. Roquim, Taciana V. Savian, and Dimitrios M. Stasinopoulos. Cattle reference growth curves based on centile estimation: A gamlss approach. *Computers and Electronics in Agriculture*, 192:106572, 2022.
- Joseph E. Neath, Andrew A. e Cavanaugh. The bayesian information criterion: background, derivation, and applications. *WIREs Computational Statistics*, 4:199–203, 2012.
- J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135:370–384, 1972.
- A. Niklasson, A. Ericson, JG Fryer, J. Karlberg, C. Lawrence, and P. Karlberg. An update of the swedish reference standards for weight, length and head circumference at birth for given gestational age (1977-1981). *Acta Paediatrica*, 80:756–762, 1991.
- Ngozi R. Njeze, Joseph O. Ogbochukwu, and Josephat M. Chinawa. Correlation of ultrasound placental diameter & thickness with gestational age. *Pakistan Journal of Medical Science Q.*, 36:1058–1062, 2020.

- R. Nogueira, Filipa Pinto-Ribeiro, SM Pereira, and F. Valente. Macroscopic and histopathological study of the placenta - an essential resource in litigation processes. *Journal of Clinical Research & Bioethics*, 06, 2015.
- R. Nogueira, P. Cardoso, AC Braga, C. Almeida, C. Nogueira-Silva, and JC Pinto. Morphological placental portraits in congenital diaphragmatic defects: A pathological study approach. *Clin Surg*, 3, 2018.
- Rosete Nogueira, Pedro Luis Cardoso, Ana Azevedo, Marcos Gomes, Cláudia Almeida, Catarina Varela, Ana Cristina Braga, and Jorge Correia Pinto. Placental biometric parameters: The usefulness of placental weight ratio and birth/placental weight ratio percentile curves for singleton gestations as a function of gestational age journal of clinical and anatomic pathology. *J Clin Anat Pathol*, 4:104 (1–15), 2019.
- Sharif Numan, Ali Asad, Mohsin Imran, and Ahmad Nisar. Evaluation of nonlinear models to define growth curve in lohi sheep. *Small Ruminant Research*, 205:106564, 2021.
- Hui Qi Pan, Harvey Goldstein, and Qi Yang. Non-parametric estimation of age-related centiles over wide age ranges. *Annals of Human Biology*, 17:475–481, 1990.
- Abubakar A. Panti, Bissala A. Ekele, Emmanuel I. Nwobodo, and Ahmed Yakubu. The relationship between the weight of the placenta and birth weight of the neonate in a nigerian hospital. *Nigerian medical journal: journal of the Nigeria Medical Association*, 53:80, 2012.
- Youngsuk Park, Danielle Maddix, François-Xavier Aubet, Kelvin Kan, Jan Gasthaus, and Yuyang Wang. Learning quantile functions without quantile crossing for distribution-free time series forecasting. *PMLR*, 151:8127–8150, 2021.
- plotly. Cran - package plotly, 2023. URL <https://CRAN.R-project.org/package=plotly>. Acedido em: 5-08-2023.
- prompter. Cran - package prompter, 2021. URL <https://CRAN.R-project.org/package=prompter>. Acedido em: 18-08-2023.
- Jeffrey S. Racine. Rstudio: A platform-independent ide for r and sweave. *Journal of Applied Econometrics*, 27:167–172, 2011.
- readr. Cran - package readr, 2023. URL <https://CRAN.R-project.org/package=readr>. Acedido em: 18-07-2023.
- CH.H. REINSCH. Smoothing by spline functions. *Numerische Mathematik*, 10:177–183, 1967.

- R A Rigby and D M Stasinopoulos. A semi-parametric additive model for variance heterogeneity. *Stat. Comput.*, 6:57–65, 1996a.
- R. A. Rigby and D. M. Stasinopoulos. *Statistical Theory and Computational Aspects of Smoothing*. 1996b.
- R. A. Rigby and D. M. Stasinopoulos. Construction of reference centiles using mean and dispersion additive models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49:41–50, 2000.
- R. A. Rigby and D. M. Stasinopoulos. Box-cox t distribution for modelling skew and leptokurtotic. 2004a.
- Robert A. Rigby and D. Mikis Stasinopoulos. Smooth centile curves for skew and kurtotic data modelled using the box–cox power exponential distribution. *Statistics in Medicine*, 23:3053–3076, 2004b.
- Robert A. Rigby and D. Mikis Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54:507–554, 2005.
- Robert A Rigby and D Mikis Stasinopoulos. Using the box-cox t distribution in gamlss to model skewness and kurtosis. *Statistical Modelling*, 6:209–229, 2006.
- Robert A. Rigby and Dimitrios M. Stasinopoulos. Automatic smoothing parameter selection in gamlss with an application to centile estimation. *Statistical Methods in Medical Research*, 23:318–332, 2014.
- Robert A. Rigby, Mikis D. Stasinopoulos, Gillian Z. Heller, and Fernanda De Bastiani. *Distributions for modeling location, scale, and shape: Using GAMLSS in R*. CRC press, 2019.
- Brian D Ripley. *Pattern recognition and neural networks*. Cambridge university press, 2007.
- rmarkdown. R markdown, 2023. URL <https://rmarkdown.rstudio.com/>. Acedido em: 1-09-2023.
- Machteld J. Roede. Growth diagrams 1980. netherlands third national survey. *Tijdsch Soc Gezondh*, 63:1–34, 1985.
- J. E. Rossiter. Calculating centile curves using kernel density estimation methods with application to infant kidney lengths. *Statistics in Medicine*, 10:1693–1701, 1991.
- Patrick Royston and Eileen M Wright. A method for estimating age-specific reference intervals (‘normal ranges’) based on fractional polynomials and exponential transformation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161:79–101, 1998.

- Patrick Royston and Eileen M. Wright. Goodness-of-fit statistics for age-specific reference intervals. *Statistics in medicine*, 19:2943–2962, 2000.
- Mark D. Schluchter. *Mean Square Error*. John Wiley & Sons, Ltd, 2005.
- Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.
- F. Shehata, I. Levin, A. Shrim, B. Ata, B. Weisz, R. Gamzu, and B. Almog. Placenta/birthweight ratio and perinatal outcome: a retrospective cohort analysis. *BJOG: An International Journal of Obstetrics & Gynaecology*, 118:741–747, 2011.
- shinydashboard. Cran - package shinydashboard, 2021. URL <https://CRAN.R-project.org/package=shinydashboard>. Acedido em: 23-07-2023.
- shinyjs. Cran - package shinyjs, 2021. URL <https://CRAN.R-project.org/package=shinyjs>. Acedido em: 18-08-2023.
- Andrea Siskovicova, Vladimir Ferianec, Marian Krizko, Martin Alfoldi, Ivana Kunochova, Jozef Záhumensky, and Martin Gabor. Analysis of factors influencing ultrasound-based fetal weight estimation. *Bratisl. Lek. Listy*, 124:25–28, 2023.
- A. Sorribas, J. March, and E. O. Voit. Estimating age-related trends in cross-sectional studies using s -distributions. *Statistics in Medicine*, 19:697–713, 2000.
- D. Mikis Stasinopoulos and Robert A. Rigby. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, 23:1–46, 2007.
- Mikis Stasinopoulos, Bob Rigby, and Calliope Akantziliotou. Instructions on how to use the gamlss package in r second edition. 2008.
- Mikis D. Stasinopoulos, Robert A. Rigby, Gillian Z. Heller, Vlasios Voudouris, and Fernanda De Bastiani. *Flexible regression and smoothing*. Chapman & Hall/CRC, 2017.
- Liliya Tamayev, Letizia Schreiber, Adi Marciano, Jacob Bar, and Michal Kovo. Are there gender-specific differences in pregnancy outcome and placental abnormalities of pregnancies complicated with small for gestational age? *Arch. Gynecol. Obstet.*, 301:1147–1151, 2020.
- Toshiro Tango. Estimation of age-specific reference ranges via smoother *avas*. *Statistics in medicine*, 17: 1231–1243, 1998.

- James M. Tanner, RH Whitehouse, and MT Takaishi. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. i. *Archives of disease in childhood*, 41:454, 1966.
- R CORE TEAM. R: A language and environment for statistical computing, 2005. URL <https://www.r-project.org/>. Acedido em: 2023-1-27.
- R CORE TEAM. Rstudio, 2011. URL <https://www.rstudio.com/>. Acedido em: 2023-1-27.
- RStudio Team. Shiny package, 2012. URL <https://shiny.posit.co/>. Acedido em: 2023-1-27.
- JMD Thompson, LM Irgens, R. Skjaerven, and S. Rasmussen. Placenta weight percentile curves for singleton deliveries. *BJOG: An International Journal of Obstetrics and Gynaecology*, 114:715–720, 2007.
- ML Thompson and GB Theron. Maximum likelihood estimation of reference centiles. *Statistics in Medicine*, 9:539–548, 1990.
- Shibabaw Tedla Tiruneh. Correlation between gross morphology of the human placenta and birth weight in normotensive and pre-eclamptic pregnancies in northwest ethiopia. *Anatomy*, 12:27–32, 2018.
- Margherita Y. Turco and Ashley Moffett. Development of the human placenta. *Development (Cambridge)*, 146, 2019.
- Gitta Turowski, W Tony Parks, Susan Arbuckle, Anne F Jacobsen, and Alexander Heazell. The structure and utility of the placental pathology report. *APMIS*, 126:638–646, 2018.
- MA Van't Hof, JM Wit, and MJ Roede. A method to construct age references for skewed skinfold data, using box-cox transformations to normality. *Human Biology*, 57:131–139, 1985.
- Valeria Vişan, Raluca Anca Balan, Claudia Florida Costea, Alexandru Căărăuleanu, Raluca Maria Haba, Mihai Ştefan Cristian Haba, Gabriela Socolov, Raluca Anamaria Mogoş, Camelia Margareta Bogdănici, Dragoş Nemescu, et al. Morphological and histopathological changes in placentas of pregnancies with intrauterine growth restriction. *Rom. J. Morphol. Embryol.*, 61:477–483, 2020.
- ZG Vitezica, C. Marie-Etancelin, MD Bernadet, Xavier Fernandez, and C. Robert-Granie. Comparison of nonlinear and spline regression models for describing mule duck growth curves. *Poultry Science*, 89: 1778–1784, 2010.

- Eric Vittinghoff, Charles E. McCulloch, David V. Glidden, and Stephen C. Shiboski. Linear and non-linear regression methods in epidemiology and biostatistics. In *Epidemiology and Medical Statistics*, volume 27, pages 148–186. Elsevier, 2007.
- Vlasios Voudouris, Robert Gilchrist, Robert Rigby, John Sedgwick, and Dimitrios Stasinopoulos. Modelling skewness and kurtosis with the bcpe density in gamlss. *Journal of Applied Statistics*, 39:1279–1293, 2012.
- A. M. Wade and A. E. Ades. Age-related reference ranges: Significance tests for models and confidence intervals for centiles. *Statistics in medicine*, 13:2359–2367, 1994.
- AM Wade and AE Ades. Incorporating correlations between measurements into the estimation of age-related reference ranges. *Statistics in medicine*, 17:1989–2002, 1998.
- Elisabeth Waldmann. Quantile regression: A short story on how and why. *Statistical Modelling*, 18:203–218, 2018.
- JM Wallace, Sohinee Bhattacharya, and GW Horgan. Gestational age, gender and parity specific centile charts for placental weight for singleton deliveries in aberdeen, uk. *Placenta*, 34:269–274, 2013.
- S. Wellek and E. Merz. Age-related reference ranges for growth parameters. *Methods of information in medicine*, 34:523–528, 1995.
- Hadley Wickham. *Mastering shiny*. "O'Reilly Media, Inc.", 2021.
- Jessica Wojciechowski, Ashley M Hopkins, and Richard Neil Upton. Interactive pharmacometric applications using r and the shiny package. *CPT: pharmacometrics & systems pharmacology*, 4:146–159, 2015.
- Charlotte M. Wright. The use and interpretation of growth charts. *Current Paediatrics*, 12:279–282, 2002.
- Eileen M Wright and Patrick Royston. A comparison of statistical methods for age-related reference intervals. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160:47–69, 1997.
- Sewall Wright. Correlation and causation. 1921.
- Lang WU and Jin QIU. *Chapter 8 Linear and Nonlinear Regression Models*, pages 111–130. EDP Sciences, 2021a.
- Lang WU and Jin QIU. *Chapter 8 Linear and Nonlinear Regression Models*, pages 111–130. 2021b.

Aneurin Young, Edward T Andrews, James John Ashton, Freya Pearson, R Mark Beattie, and Mark John Johnson. Generating longitudinal growth charts from preterm infants fed to current recommendations. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 105:646–651, 2020.

Keming Yu, Zudi Lu, and Julian Stander. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52:331–350, 2003.

Apêndice A

Distribuições

Na seguinte tabela é possível verificar as distribuições contínuas **GAMLSS** e funções de ligação correspondentes. Adaptado de: [Stasinopoulos et al. \[2017\]](#).

Distribuições	Intervalos	Funções de Ligação de Parâmetros			
		μ	σ	ν	τ
beta	(0,1)	logit	logit	-	-
Box-Cox Cole-Green	(0, ∞)	ident.	log	ident.	-
Box-Cox Cole Green orig.	(0, ∞)	log	log	ident.	-
Box-Cox power exponential	(0, ∞)	ident.	log	ident.	log
Box-Cox power expon. orig.	(0, ∞)	log	log	ident.	log
Box-Cox t	(0, ∞)	ident.	log	ident.	log
Box-Cox t orig.	(0, ∞)	log	log	ident.	log
exponential	(0, ∞)	log	-	-	-
exponential Gaussian	($-\infty$, ∞)	ident.	log	log	-
exponential gen. beta 2	($-\infty$, ∞)	ident.	log	log	log
gamma	(0, ∞)	log	log	-	-
generalized beta type 1	(0,1)	logit	logit	log	log
generalized beta type 2	(0, ∞)	log	log	log	log
generalized gamma	(0, ∞)	log	log	ident.	-
generalized inv. Gaussian	(0, ∞)	log	log	ident.	-
generalized t	($-\infty$, ∞)	ident.	log	log	log
Gumbel	($-\infty$, ∞)	ident.	log	-	-
inverse Gamma	(0, ∞)	log	log	-	-
inverse Gaussian	(0, ∞)	log	log		

Continuação da tabela na próxima página

Continuação da tabela da página anterior.

Distribuições	Intervalos	Funções de Ligação de Parâmetros			
		μ	σ	ν	τ
Johnson's SU repar.	$(-\infty, \infty)$	ident.	log	ident.	log
Johnson's SU original	$(-\infty, \infty)$	ident.	log	ident.	log
logistic	$(-\infty, \infty)$	ident.	log	-	-
logit normal	(0,1)	ident.	log	-	-
log normal	$(0, \infty)$	ident.	log	-	-
log normal 2	$(0, \infty)$	log	log	-	-
log normal (Box-Cox)	$(0, \infty)$	ident.	log	fixed	-
NET	$(-\infty, \infty)$	ident.	log	fixed	fixed
normal	$(-\infty, \infty)$	ident.	log	-	-
normal family	$(-\infty, \infty)$	ident.	log	-	-
Pareto 2	$(0, \infty)$	log	log	-	-
Pareto 2 original	$(0, \infty)$	log	log	-	-
Pareto 2 repar	$(0, \infty)$	log	log	-	-
power exponential	$(-\infty, \infty)$	ident.	log	log	-
reverse gen. extreme	$y > \mu - (\sigma/\nu)$	ident.	log	log	-
reverse Gumbel	$(-\infty, \infty)$	ident.	log	-	-
sinh - arcsinh	$(-\infty, \infty)$	ident.	log	log	log
sinh - arcsinh original	$(-\infty, \infty)$	ident.	log	ident.	log
sinh - arcsinh original 2	$(-\infty, \infty)$	ident.	log	ident.	log
skew normal type 1	$(-\infty, \infty)$	ident.	log	ident.	-
skew normal type 2	$(-\infty, \infty)$	ident.	log	log	-
skew power exp. type 1	$(-\infty, \infty)$	ident.	log	ident.	log
skew power exp. type 2	$(-\infty, \infty)$	ident.	log	ident.	log
skew power exp. type 3	$(-\infty, \infty)$	ident.	log	log	log
skew power exp. type 4	$(-\infty, \infty)$	ident.	log	log	log
skew t type 1	$(-\infty, \infty)$	ident.	log	ident.	log
skew t type 2	$(-\infty, \infty)$	ident.	log	ident.	log
skew t type 3	$(-\infty, \infty)$	ident.	log	log	log

Continuação da tabela na próxima página

Continuação da tabela da página anterior.

Distribuições	Intervalos	Funções de Ligação de Parâmetros			
		μ	σ	ν	τ
skew t type 3 repar	$(-\infty, \infty)$	ident.	log	log	log-2
skew t type 4	$(-\infty, \infty)$	ident.	log	log	log
skew t type 5	$(-\infty, \infty)$	ident.	log	ident.	log
t Family	$(-\infty, \infty)$	ident.	log	log	-
t Family repar	$(-\infty, \infty)$	ident.	log	log - 2	-
Weibull	$(0, \infty)$	log	log	-	-
Weibull (PH)	$(0, \infty)$	log	log	-	-
Weibull (μ the mean)	$(0, \infty)$	log	log	-	-

Apêndice B

Dados

Variável	Descrição	Tipo	Codificação
Local	Local de recolha da placenta	Qualitativa nominal	-----
NumeroProcesso	Número associado ao paciente	Qualitativa nominal	-----
MaternalAge	Idade materna	Numérica contínua	-----
GA	Idade gestacional	Numérica discreta	-----
Fetalweight	Peso do feto	Numérica contínua	-----
Placentalweight	Peso da placenta	Numérica contínua	-----
Diameter1	Maior valor do diâmetro da placenta	Numérica contínua	-----
Diameter2	Menor valor do diâmetro da placenta	Numérica contínua	-----
Placentalthickness	Espessura da placenta	Numérica contínua	-----
UCvesselsnumber	Número de vasos no cordão	Numérica discreta	-----
UClenght	Comprimento do cordão	Numérica discreta	-----

Continua na próxima página

Continuação da página anterior

Variável	Descrição	Tipo	Codificação
UCdiameter	Diâmetro do cordão	Numérica contínua	-----
ratio	Rácio Peso do feto/Peso placentar	Numérica contínua	-----
ratio2	Rácio Peso placentar/Peso do feto	Numérica contínua	-----
Placental_vol	Volume da placenta	Numérica contínua	-----
Obstetriccontext	Paridade da mãe	Numérica discreta	-----
Fetalgender	Gênero do feto	Catégorica nominal	0 - Feminino 1 - Masculino 2 - Ambíguo
Pregnancyoutcome	Resultado da gravidez	Catégorica nominal	0 - Morte Fetal 1 - Morte Neonatal 2 - Recém-Nascido 3 - IMG
Placentalshape	Forma da placenta	Catégorica nominal	0 - Normal (simples) 1 - Bilobada/Polilobada 2 - Circunmarginada/ Circunvalada 3 - Membranácea
UCinsertiontype	Inserção do cordão	Catégorica nominal	0 - Normal 1 - Marginal 2 - Velamentosa 3 - Bifurcada

Continua na próxima página

Continuação da página anterior

Variável	Descrição	Tipo	Codificação
Fetalpathology	Patologia fetal	Categórica nominal	0- Ausente / Normal
			1 - RCF
			2 - Síndrome polimalformativo
			3 - DTN
			4 - Defeito parede abdominal
			5 - Defeito cardiovascular
			6 - Defeito toraco-respiratório
			7 - Defeito gastro-intestinal
			8 - Defeito Geniturinário
			9 - Infecções
			10 - BSLC
			11 - Macrosomia
			12 - Aneuploidias
			13 - Doenças monogénicas
Maternalpathology	Patologia materna	Categórica nominal	0 - Ausente / Normal
			1 - Diabetes
			2 - HTA
			3 - SAF / Trombofilias
			4 - Doenças sistémicas
			5 - Infecções
			6 - Neoplasias
			7 - Anomalias cromossómica
			8 - Defeitos anatómicos uterinas
			9 - Alergias / Asma / Rinite
			10 - Patologia psiquiátrica
			11 - Defeito génico
			12 - Álcool / Fumadora
13 - Consanguinidade			
			14 - IRC

Apêndice C

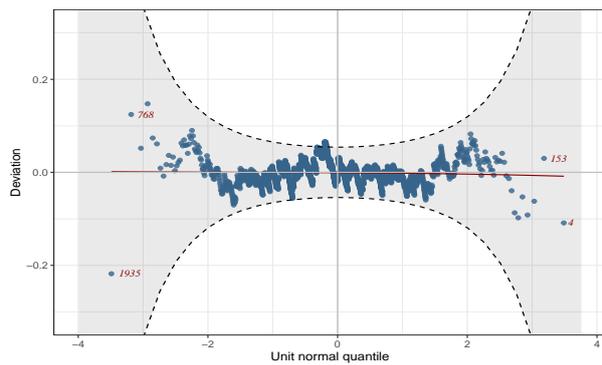
Valores-p dos testes de significância de géneros

IG	Diâmetro 1	Diâmetro 2	Peso	Espessura	Peso fetal
12	0,2	0,009	0,5	0,7	0,5
13	0,5	0,8	0,4	0,1	0,2
14	0,7	0,03	0,07	0,02	0,1
15	0,7	0,3	0,7	0,8	0,4
16	0,4	0,6	0,5	0,7	0,009
17	0,8	0,8	0,8	0,1	0,3
18	0,1	0,1	0,2	0,5	0,07
19	0,3	0,3	0,5	0,1	0,8
20	0,06	0,8	0,4	0,04	0,3
21	1	0,8	0,4	0,8	0,08
22	0,05	0,5	0,3	0,4	0,4
23	0,6	0,6	0,4	0,4	0,2
24	0,6	0,7	0,8	0,8	0,2
25	0,05	0,2	0,4	0,9	0,5
26	0,3	0,4	0,2	0,4	0,7
27	0,8	0,9	0,09	0,05	0,8
28	0,8	0,5	0,7	0,7	0,6
29	0,8	0,7	0,5	0,9	0,3
30	0,4	0,2	0,6	0,3	0,3
31	0,2	0,8	0,8	0,5	0,9
32	0,8	0,6	0,8	0,6	0,5
33	0,4	0,5	1	0,9	0,3
34	0,4	1	0,5	0,2	0,3
35	0,9	0,3	0,7	0,8	0,5
36	0,1	0,3	0,3	0,06	0,5
37	0,2	0,8	0,8	0,3	0,1
38	0,3	0,5	0,7	0,6	0,9
39	0,04	0,006	0,5	0,2	0,2
40	0,3	0,8	0,9	0,02	0,002
41	0,8	0,8	0,8	1	0,8

Apêndice D

Diâmetro 1

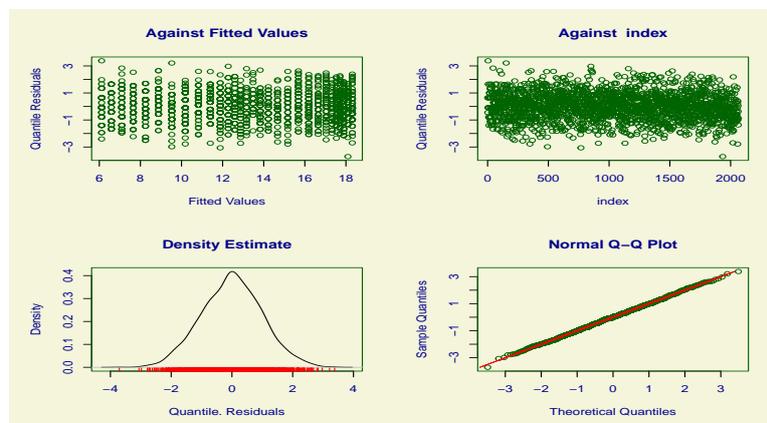
Crerios de avaliaço de qualidade de ajuste para o modelo **LMST** do **D1** e respectiva curva de crescimento.



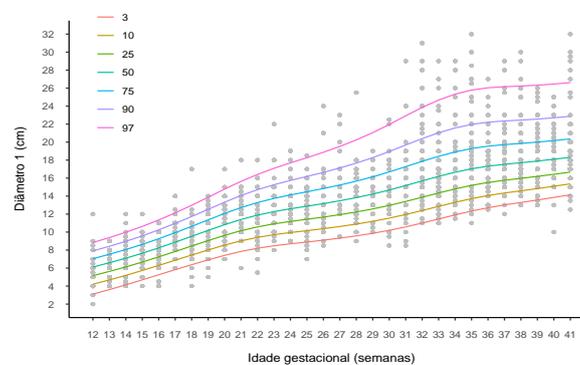
(a) Worm plot



(b) Estatisticas Z



(c) Residuos



(d) Curva de crescimento

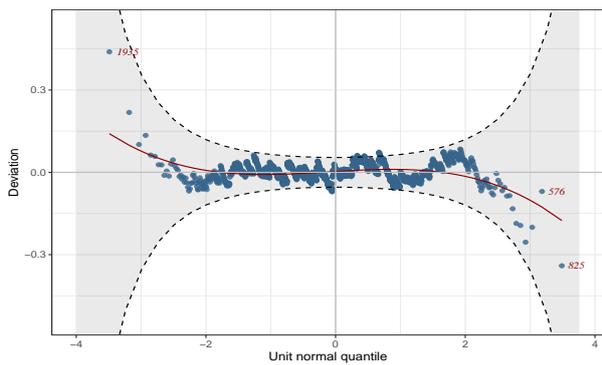
Representação numérica dos percentis traçados no gráfico da curva de crescimento para o **D1**, Figura 17, para o modelo final **GAMLSS** com *P-splines* e **BCT**.

IG	3º	10º	25º	50º	75º	90º	97º
12	3,11	4,21	5,16	6,12	7,04	7,91	8,84
13	3,60	4,68	5,63	6,59	7,53	8,42	9,39
14	4,13	5,19	6,13	7,09	8,05	8,97	9,98
15	4,68	5,73	6,67	7,65	8,63	9,57	10,63
16	5,25	6,29	7,24	8,25	9,26	10,24	11,35
17	5,82	6,87	7,84	8,88	9,93	10,97	12,13
18	6,37	7,45	8,45	9,53	10,64	11,73	12,98
19	6,91	8,02	9,05	10,18	11,36	12,53	13,88
20	7,41	8,56	9,64	10,83	12,07	13,32	14,78
21	7,85	9,03	10,16	11,40	12,72	14,06	15,64
22	8,20	9,41	10,58	11,88	13,28	14,71	16,40
23	8,48	9,71	10,92	12,27	13,74	15,26	17,08
24	8,70	9,95	11,19	12,59	14,13	15,73	17,69
25	8,89	10,16	11,42	12,88	14,48	16,17	18,27
26	9,10	10,38	11,67	13,16	14,84	16,63	18,87
27	9,33	10,62	11,94	13,48	15,23	17,12	19,52
28	9,58	10,89	12,24	13,83	15,65	17,66	20,24
29	9,87	11,20	12,58	14,23	16,14	18,27	21,05
30	10,19	11,55	12,97	14,68	16,68	18,95	21,95
31	10,58	11,96	13,40	15,17	17,27	19,67	22,91
32	11,00	12,40	13,87	15,69	17,86	20,38	23,83
33	11,46	12,86	14,35	16,19	18,42	21,03	24,65
34	11,91	13,31	14,80	16,65	18,90	21,56	25,31
35	12,34	13,72	15,19	17,03	19,28	21,95	25,76
36	12,71	14,06	15,51	17,32	19,54	22,20	26,01
37	13,02	14,34	15,76	17,53	19,71	22,34	26,13
38	13,30	14,59	15,97	17,70	19,84	22,43	26,20
39	13,57	14,84	16,19	17,89	19,99	22,54	26,29
40	13,86	15,10	16,42	18,09	20,16	22,69	26,42
41	14,15	15,36	16,66	18,31	20,35	22,85	26,57

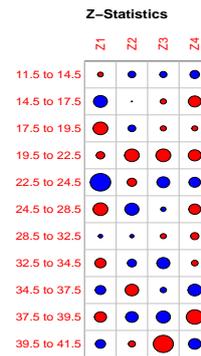
Apêndice E

Diâmetro 2

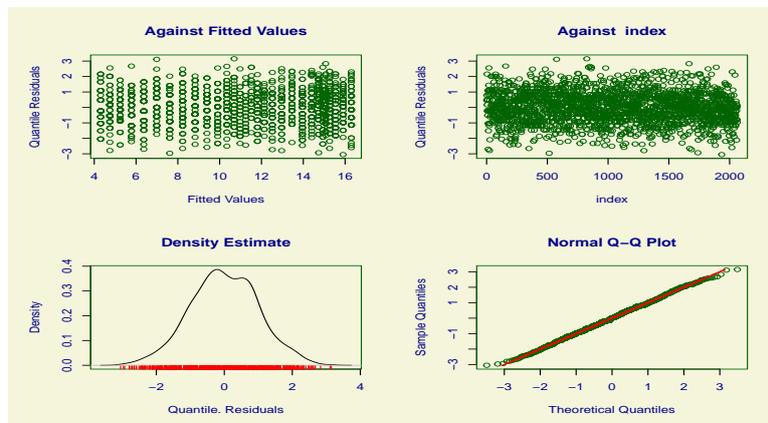
Crerios de avaliaço de qualidade de ajuste para o modelo **LMST** do **D2** e respetiva curva de crescimento.



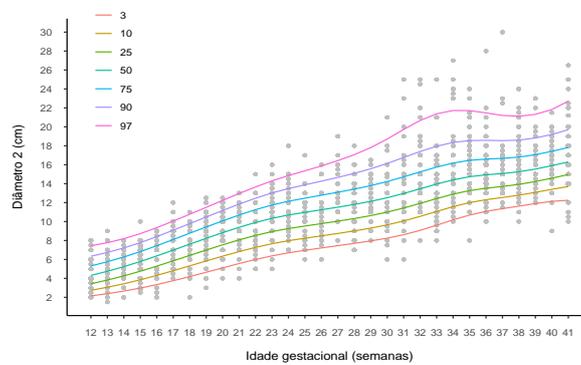
(a) Worm plot



(b) Estatisticas Z



(c) Residuos



(d) Curva de crescimento

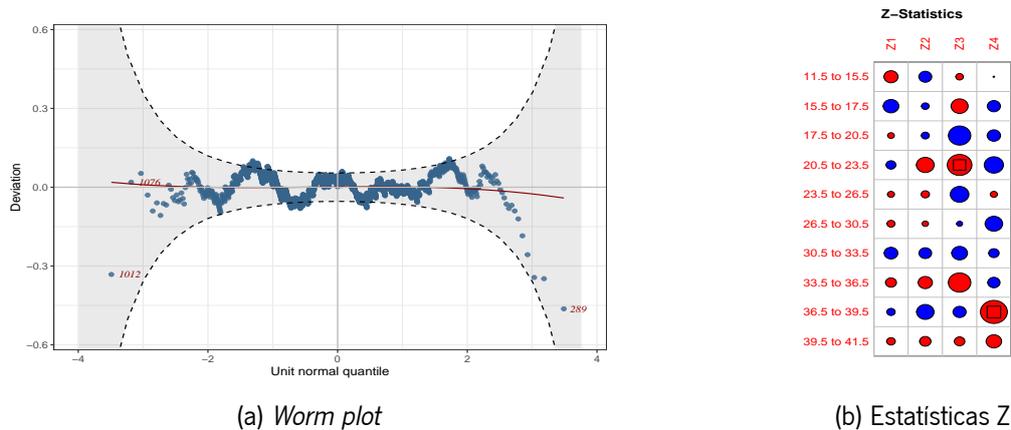
Representação numérica dos percentis traçados no gráfico da curva de crescimento para o **D2**, Figura 20, para o modelo final **GAMLSS** com *P-splines* e **BCT**.

IG	3º	10º	25º	50º	75º	90º	97º
12	2,09	2,74	3,43	4,28	5,25	6,31	7,62
13	2,33	3,08	3,84	4,73	5,71	6,72	7,91
14	2,61	3,46	4,30	5,23	6,22	7,20	8,31
15	2,93	3,89	4,80	5,78	6,79	7,76	8,84
16	3,30	4,36	5,35	6,38	7,41	8,39	9,47
17	3,70	4,86	5,91	6,99	8,07	9,07	10,16
18	4,14	5,38	6,49	7,62	8,73	9,76	10,87
19	4,60	5,90	7,06	8,24	9,39	10,45	11,60
20	5,05	6,41	7,62	8,84	10,03	11,15	12,33
21	5,50	6,89	8,13	9,41	10,65	11,82	13,07
22	5,90	7,33	8,60	9,92	11,22	12,44	13,77
23	6,27	7,69	8,99	10,35	11,71	13,00	14,41
24	6,56	7,98	9,29	10,69	12,11	13,47	14,97
25	6,80	8,21	9,53	10,97	12,44	13,88	15,49
26	7,05	8,43	9,77	11,23	12,77	14,29	16,01
27	7,30	8,67	10,00	11,51	13,10	14,71	16,57
28	7,58	8,92	10,26	11,79	13,46	15,16	17,16
29	7,91	9,22	10,56	12,12	13,85	15,65	17,82
30	8,30	9,59	10,93	12,51	14,30	16,21	18,57
31	8,77	10,03	11,36	12,96	14,81	16,83	19,38
32	9,29	10,53	11,85	13,46	15,34	17,45	20,18
33	9,85	11,06	12,36	13,96	15,86	18,01	20,86
34	10,38	11,56	12,83	14,41	16,29	18,45	21,33
35	10,81	11,97	13,21	14,74	16,58	18,68	21,51
36	11,13	12,26	13,47	14,95	16,71	18,74	21,43
37	11,37	12,48	13,65	15,09	16,80	18,72	21,26
38	11,60	12,71	13,87	15,28	16,93	18,78	21,18
39	11,85	12,98	14,15	15,56	17,18	18,97	21,26
40	12,14	13,30	14,49	15,91	17,53	19,28	21,49
41	12,43	13,63	14,86	16,30	17,92	19,66	21,81

Apêndice F

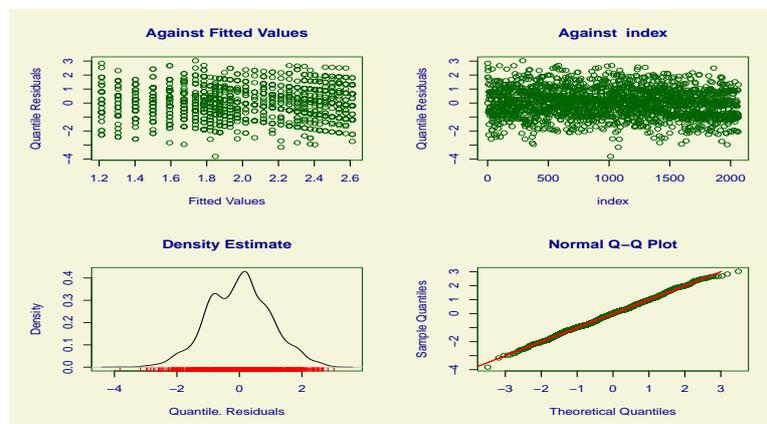
Espessura placentar

Crêterios de avaliaç o de qualidade de ajuste para o modelo **GAMLSS** com **BCT** da **EP** e respetiva curva de crescimento.

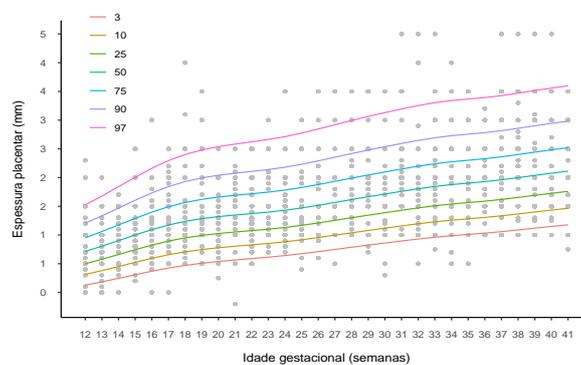


(a) Worm plot

(b) Estatisticas Z



(c) Residuos



(d) Curva de crescimento

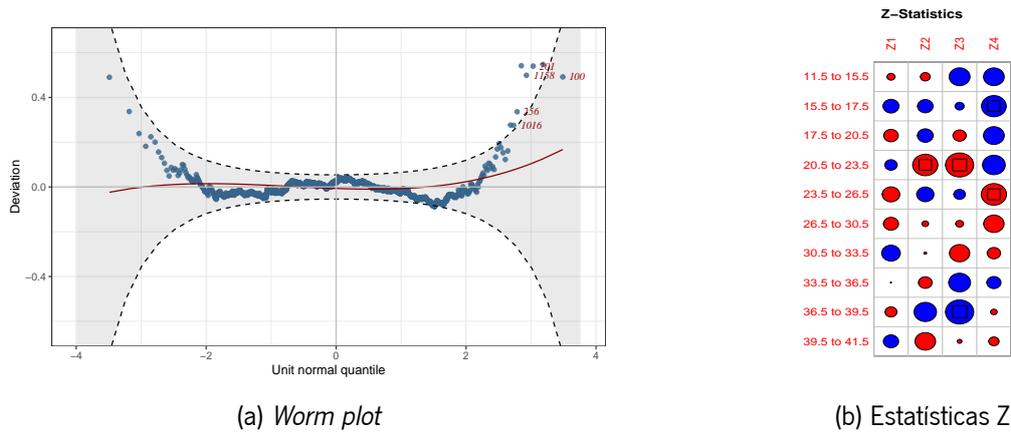
Representação numérica dos percentis traçados no gráfico da curva de crescimento para a **EP**, Figura 23, para o modelo final **LMST** com *P-splines*.

IG	3º	10º	25º	50º	75º	90º	97º
12	0,62	0,81	1,00	1,21	1,44	1,70	2,06
13	0,67	0,88	1,08	1,30	1,55	1,83	2,21
14	0,74	0,95	1,16	1,40	1,67	1,97	2,38
15	0,80	1,03	1,25	1,50	1,78	2,10	2,54
16	0,86	1,10	1,33	1,59	1,89	2,23	2,69
17	0,92	1,16	1,40	1,67	1,98	2,34	2,82
18	0,97	1,21	1,45	1,74	2,06	2,43	2,92
19	1,00	1,25	1,49	1,78	2,12	2,49	3,00
20	1,04	1,28	1,53	1,82	2,16	2,54	3,05
21	1,06	1,31	1,56	1,85	2,19	2,57	3,09
22	1,08	1,33	1,58	1,87	2,21	2,60	3,13
23	1,11	1,36	1,60	1,90	2,24	2,63	3,16
24	1,14	1,38	1,63	1,93	2,28	2,68	3,21
25	1,18	1,42	1,67	1,97	2,32	2,73	3,27
26	1,21	1,46	1,71	2,02	2,38	2,79	3,34
27	1,25	1,50	1,76	2,07	2,43	2,85	3,41
28	1,29	1,54	1,80	2,11	2,49	2,92	3,48
29	1,33	1,58	1,84	2,16	2,55	2,98	3,55
30	1,37	1,62	1,89	2,21	2,60	3,04	3,61
31	1,40	1,66	1,93	2,26	2,65	3,09	3,67
32	1,43	1,69	1,97	2,31	2,70	3,15	3,73
33	1,47	1,73	2,01	2,35	2,75	3,19	3,78
34	1,49	1,76	2,03	2,38	2,78	3,23	3,81
35	1,52	1,78	2,05	2,40	2,81	3,26	3,84
36	1,54	1,80	2,08	2,43	2,84	3,29	3,87
37	1,57	1,83	2,11	2,46	2,87	3,32	3,90
38	1,60	1,86	2,14	2,49	2,91	3,36	3,94
39	1,63	1,89	2,18	2,54	2,95	3,41	3,99
40	1,66	1,93	2,22	2,57	3,00	3,45	4,03
41	1,68	1,95	2,24	2,61	3,03	3,49	4,07

Apêndice G

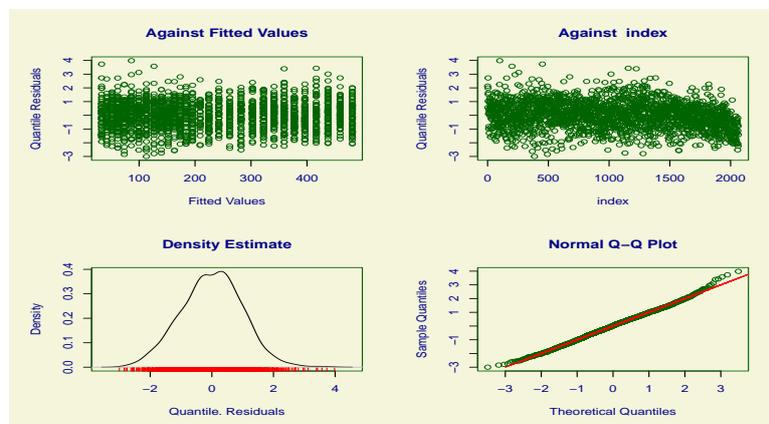
Peso placentar

Crêrios de avaliação de qualidade de ajuste para o modelo **GAMLSS** com **JSU** do **PP** e respetiva curva de crescimento.

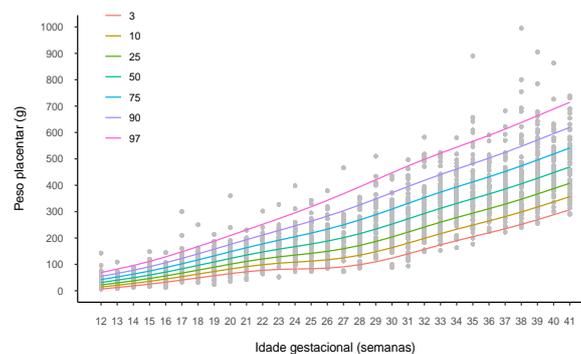


(a) Worm plot

(b) Estatísticas Z



(c) Resíduos



(d) Curva de crescimento

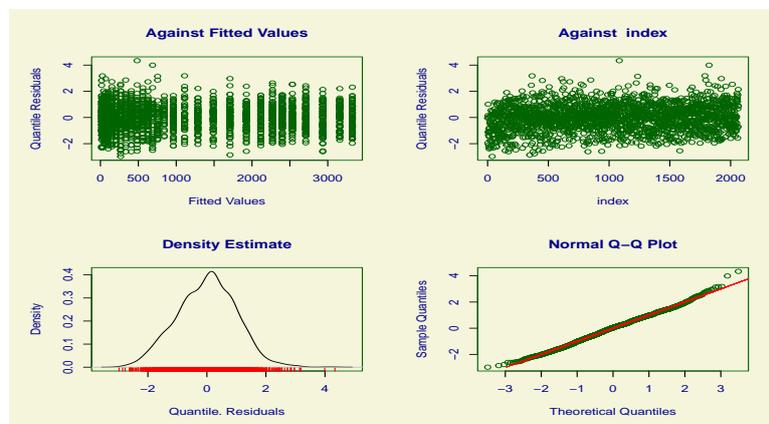
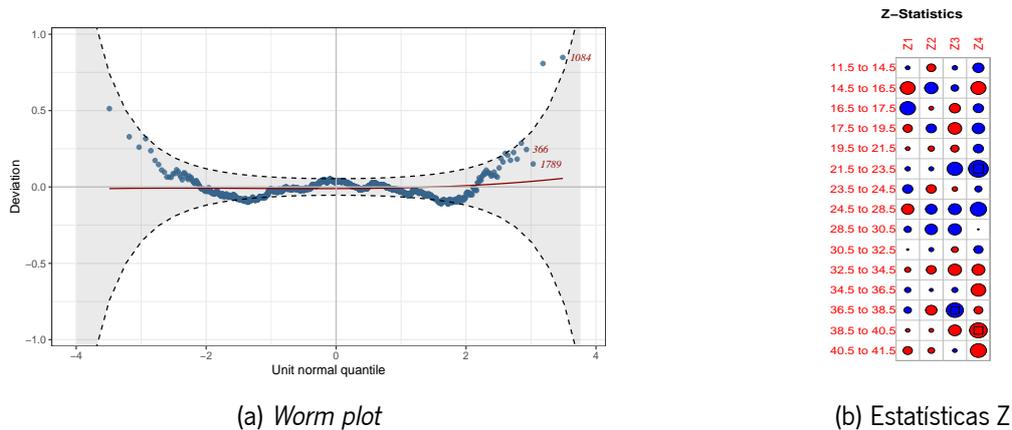
Representação numérica dos percentis traçados no gráfico da curva de crescimento para o **PP**, Figura 26, para o modelo final **LMST** com *P-splines*.

IG	3º	10º	25º	50º	75º	90º	97º
12	8,44	16,14	23,99	32,94	43,37	56,02	75,32
13	12,28	21,29	30,17	40,17	51,57	64,97	84,47
14	17,17	27,57	37,63	48,85	61,45	75,84	95,97
15	23,07	34,99	46,42	59,12	73,21	88,93	110,18
16	29,89	43,48	56,49	70,93	86,81	104,18	126,99
17	37,43	52,77	67,48	83,82	101,66	120,86	145,43
18	45,49	62,54	78,98	97,26	117,11	138,16	164,49
19	53,82	72,49	90,58	110,75	132,55	155,37	183,37
20	62,07	82,22	101,88	123,89	147,58	172,13	201,74
21	69,48	91,11	112,39	136,32	162,06	188,52	219,96
22	75,24	98,52	121,68	147,93	176,20	205,09	239,05
23	78,79	104,06	129,60	158,85	190,47	222,73	260,36
24	80,33	107,84	136,16	169,01	204,78	241,29	283,69
25	81,12	110,85	142,06	178,79	219,14	260,44	308,28
26	83,28	114,93	148,78	189,18	233,98	279,98	333,21
27	88,07	121,30	157,40	201,01	249,80	300,09	358,26
28	96,23	130,79	168,77	215,15	267,45	321,57	384,22
29	108,14	143,89	183,54	232,41	287,90	345,59	412,45
30	123,76	160,55	201,62	252,64	310,97	371,89	442,67
31	142,55	180,10	222,26	274,98	335,66	399,35	473,62
32	163,11	201,05	243,84	297,71	360,15	426,07	503,31
33	182,83	220,71	263,68	318,15	381,79	449,48	529,31
34	199,73	237,18	279,95	334,62	399,07	468,26	550,55
35	213,59	250,46	292,89	347,62	412,84	483,58	568,63
36	226,60	263,11	305,45	360,59	427,03	499,94	588,66
37	241,40	278,06	320,90	377,21	445,81	521,98	615,86
38	259,84	297,20	341,15	399,38	471,01	551,41	651,72
39	281,35	319,45	364,44	424,34	498,50	582,36	687,90
40	303,17	341,68	387,23	448,03	523,56	609,34	717,88
41	322,64	361,18	406,81	467,82	543,79	630,34	740,31

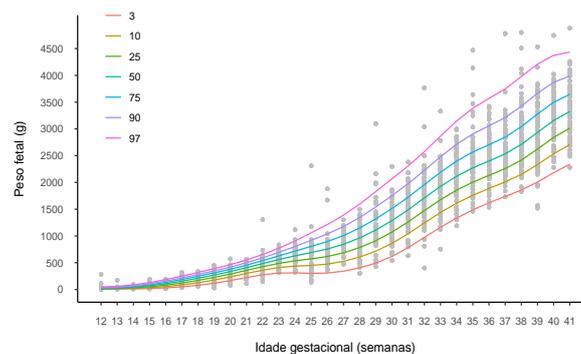
Apêndice H

Peso fetal

Crêterios de avaliaç o de qualidade de ajuste para o modelo **GAMLSS** com **BCT** do **PF** e respetiva curva de crescimento.



(c) Residuos



(d) Curva de crescimento

Representação numérica dos percentis traçados no gráfico da curva de crescimento para o **PF**, Figura 29, para o modelo final **LMST** com *P-splines*.

IG	3º	10º	25º	50º	75º	90º	97º
12	7,27	9,62	12,28	15,95	21,30	29,79	48,79
13	9,40	14,51	19,83	26,29	34,25	44,34	60,37
14	10,74	21,19	31,69	43,05	55,30	68,82	87,18
15	16,98	33,82	51,48	69,73	87,97	106,62	130,10
16	30,53	55,79	81,88	108,03	133,10	157,84	188,00
17	52,11	87,82	122,56	155,95	187,08	217,35	254,05
18	81,87	129,09	171,56	210,47	246,07	280,76	323,37
19	120,07	178,86	227,44	270,02	308,64	346,83	394,95
20	166,98	236,94	290,52	335,88	376,98	418,47	472,43
21	217,78	299,56	359,12	408,43	453,24	499,47	561,52
22	263,50	359,98	429,06	485,68	537,37	591,71	666,58
23	294,26	409,49	493,52	562,68	626,18	693,85	788,82
24	305,61	442,10	546,54	634,21	715,43	802,77	926,70
25	304,60	462,14	589,89	700,28	803,84	915,85	1075,54
26	307,71	485,76	635,98	768,96	895,05	1031,57	1225,74
27	324,33	526,76	700,21	856,12	1005,08	1165,77	1392,31
28	360,29	593,61	792,09	972,14	1156,57	1331,22	1589,73
29	424,74	692,10	913,59	1116,50	1313,74	1524,23	1812,51
30	542,58	835,51	1070,69	1289,91	1506,09	1735,57	2043,84
31	727,63	1024,11	1262,05	1491,11	1721,32	1964,52	2284,20
32	952,88	1235,10	1471,92	1710,54	1956,27	2215,59	2550,10
33	1171,99	1436,08	1672,77	1923,88	2190,22	2472,92	2833,63
34	1362,53	1610,28	1846,97	2111,15	2400,79	2712,41	3109,80
35	1527,66	1760,39	1994,46	2267,28	2575,85	2913,27	3345,89
36	1671,21	1893,49	2124,98	2402,75	2723,20	3076,47	3528,80
37	1796,00	2018,72	2255,30	2543,05	2876,41	3241,36	3700,60
38	1913,85	2152,32	2407,13	2716,56	3070,61	3449,09	3908,67
39	2064,55	2325,24	2600,78	2929,06	3293,53	3667,84	4100,06
40	2245,41	2523,76	2811,47	3144,47	3500,50	3850,45	4234,77
41	2422,58	2707,31	2994,58	3317,78	3651,82	3968,23	4302,08

Relatório de crescimento placentar e fetal

Este relatório apresenta uma análise detalhada do crescimento placentar e fetal ao longo das diferentes fases da gestação. Acompanhar o desenvolvimento destes parâmetros é fundamental para avaliar a saúde e o bem-estar do feto, bem como a função placentar. O relatório inclui informações essenciais do paciente. Além disso, são fornecidas informações detalhadas sobre os diversos parâmetros, juntamente com gráficos e tabelas que destacam os percentis de crescimento e os rácios.

Os gráficos das curvas de crescimento são acompanhados por áreas sombreadas que indicam diferentes percentis de crescimento, variando de 3% a 97%. Os pontos adicionais marcados a vermelho foram inseridos pelo utilizador para representar dados específicos do paciente para um certo parâmetro. Cada secção do relatório é acompanhada por tabelas que fornecem detalhes específicos sobre os pontos marcados nos gráficos, incluindo o número do processo, a data, a idade gestacional, o valor do parâmetro medido e o percentil correspondente. Também é apresentada uma tabela com os valores dos rácios calculados.

Este relatório foi gerado a partir da aplicação web **PlacentalGrowth**.

Informações acerca do Paciente

- **Número de processo:** 549667
- **Nome do paciente:** Maria Silva
- **Idade:** 34
- **Data:** 2023-10-09
- **Informações adicionais:**

Diâmetro 1 placentar

Gráfico da curva de crescimento

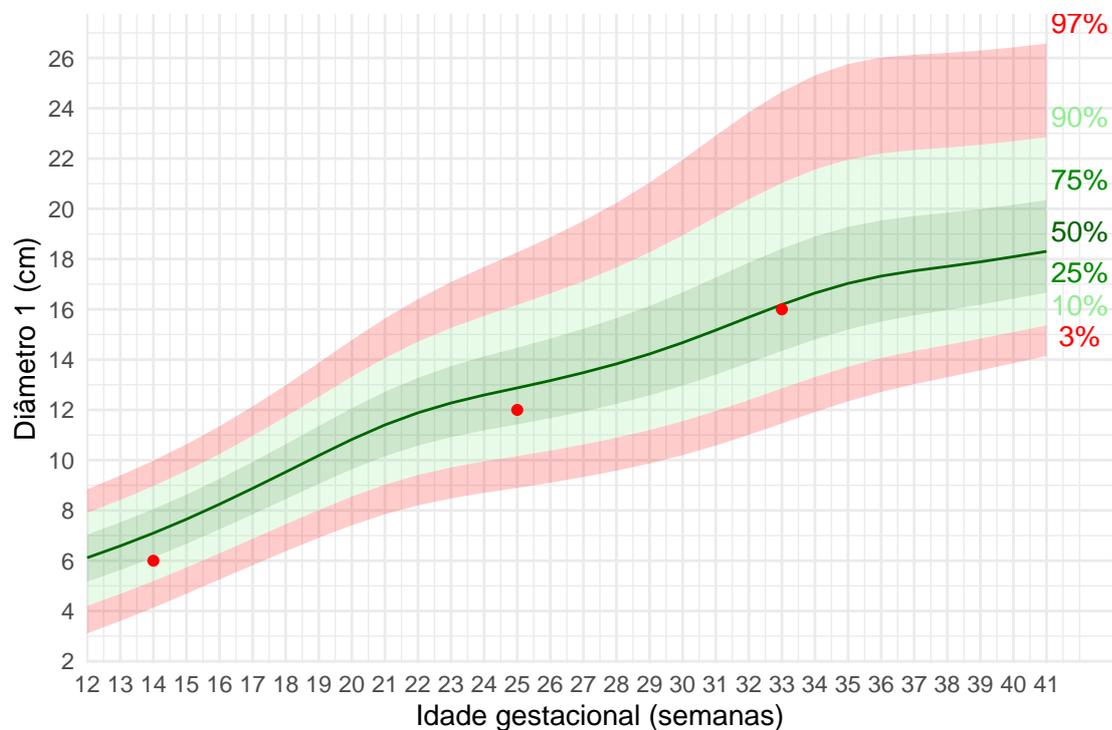


Tabela com informações acerca dos pontos adicionados no gráfico

Table 1: Informações acerca dos pontos para o Diâmetro 1.

Processo	Data	IG	Diâmetro1	Percentil
549667	2023-10-09	14	6	entre o 10 ^o e o 25
549667	2023-10-09	25	12	entre o 25 ^o e 50 ^o
549667	2023-10-09	33	16	entre o 25 ^o e 50 ^o

Diâmetro 2 placental

Gráfico da curva de crescimento

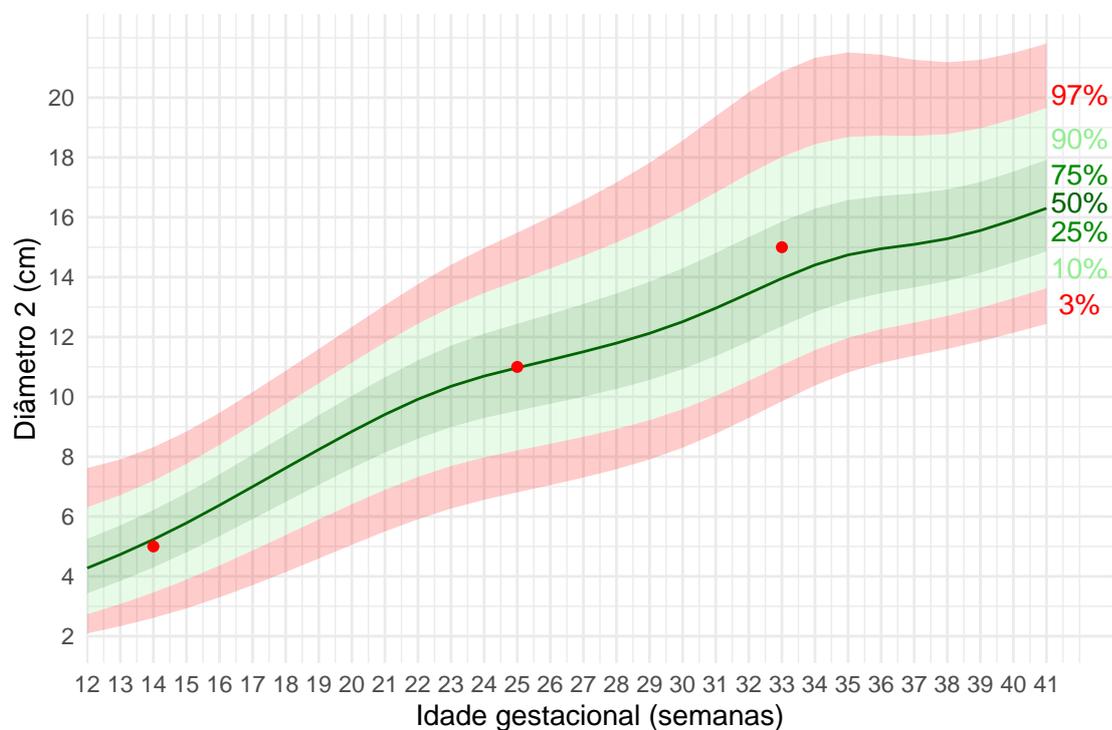


Tabela com informações acerca dos pontos adicionados no gráfico

Table 2: Informações acerca dos pontos para o Diâmetro 2.

Processo	Data	IG	Diâmetro2	Percentil
549667	2023-10-09	14	5	entre o 25º e 50º
549667	2023-10-09	25	11	entre o 50º e o 75
549667	2023-10-09	33	15	entre o 50º e o 75

Espessura placentar

Gráfico da curva de crescimento

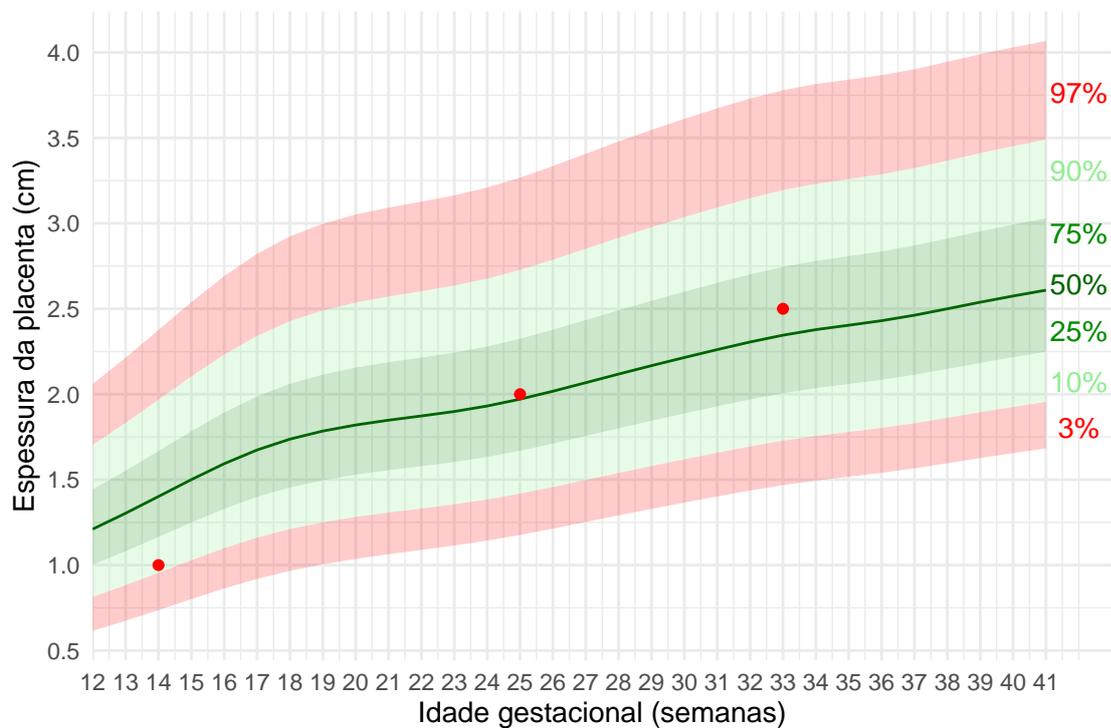


Tabela com informações acerca dos pontos adicionados no gráfico

Table 3: Informações acerca dos pontos para a Espessura.

Processo	Data	IG	Espessura	Percentil
549667	2023-10-09	14	1.0	entre o 10º e o 25
549667	2023-10-09	25	2.0	entre o 50º e o 75
549667	2023-10-09	33	2.5	entre o 50º e o 75

Peso placentar

Gráfico da curva de crescimento

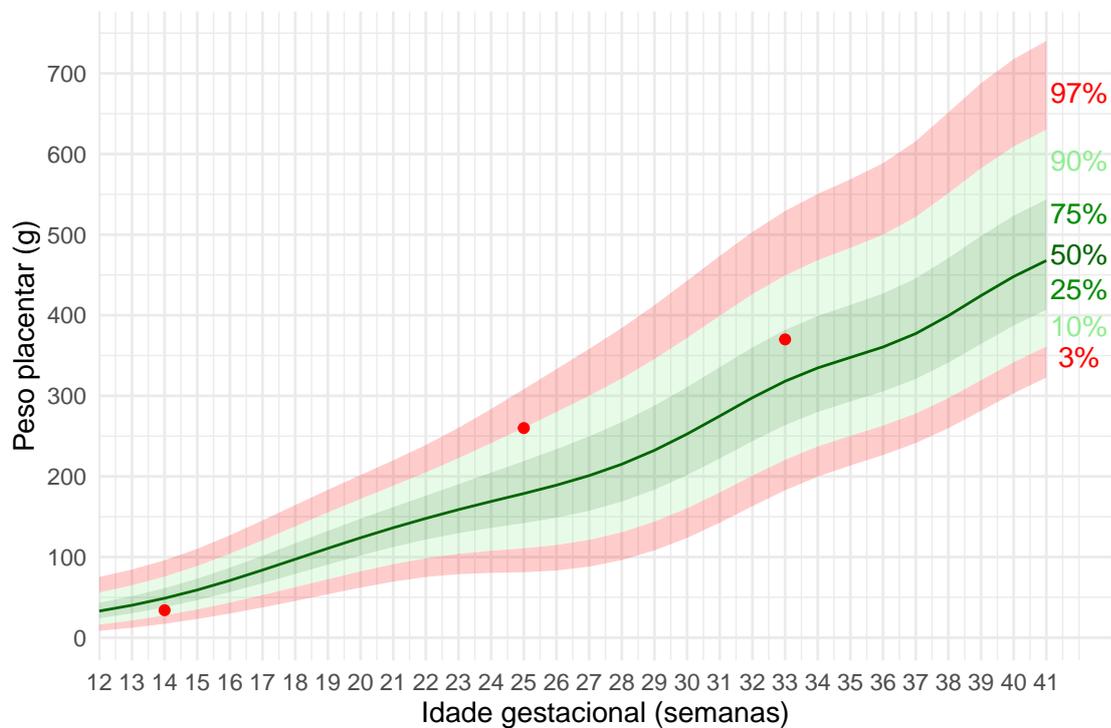


Tabela com informações acerca dos pontos adicionados no gráfico

Table 4: Informações acerca dos pontos para o Peso.

Processo	Data	IG	Peso	Percentil
549667	2023-10-09	14	34	entre o 10 ^o e o 25
549667	2023-10-09	25	260	entre o 75 ^o e o 90
549667	2023-10-09	33	370	entre o 50 ^o e o 75

Peso fetal

Gráfico da curva de crescimento

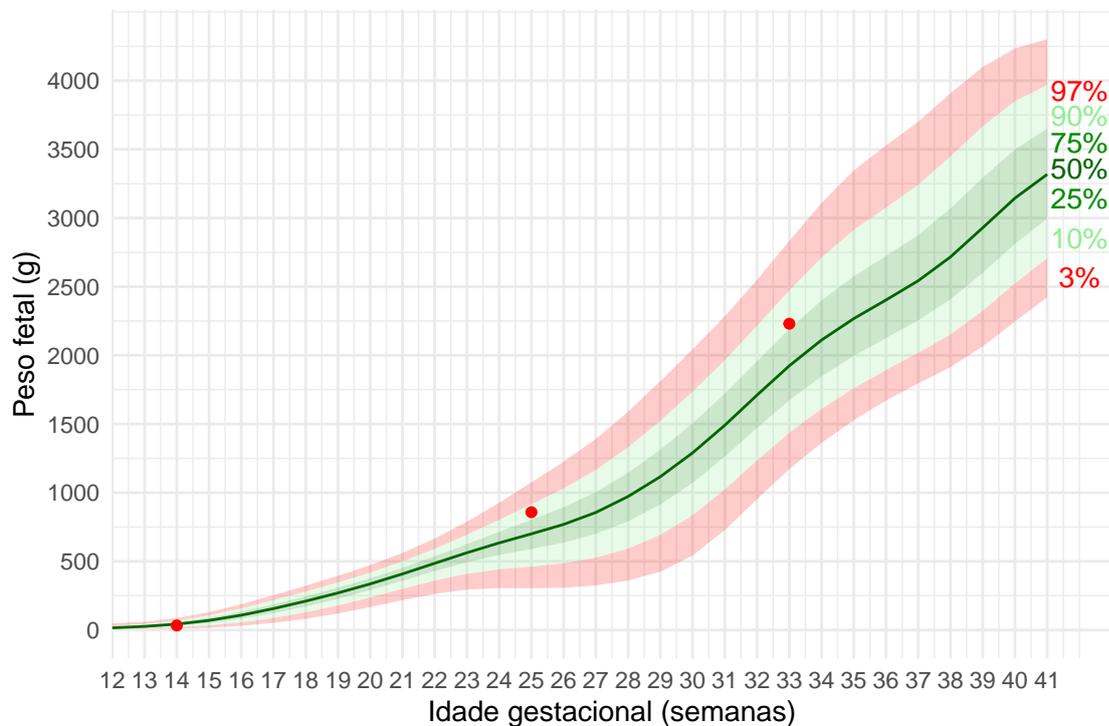


Tabela com informações acerca dos pontos adicionados no gráfico

Table 5: Informações acerca dos pontos para o Peso Fetal.

Processo	Data	IG	Peso_fetal	Percentil
549667	2023-10-09	14	33	entre o 25º e 50º
549667	2023-10-09	25	858	entre o 75º e o 90
549667	2023-10-09	33	2230	entre o 75º e o 90

Rácios

- O valor do rácio1 corresponde ao Peso Fetal/ Peso Placentar.
- O valor do rácio2 corresponde ao Pesseo Placentar/ Peso fetal.

Table 6: Informações acerca dos rácios.

Processo	IG	rácio1	rácio2
549667	14	0.971	1.030
549667	25	3.300	0.303
549667	33	6.027	0.166

