



DENIZE PALMITO DOS SANTOS

**MODELAGEM GAMLSS DA PRESSÃO ARTERIAL
SANGUÍNEA EM UM ESTUDO DE HIPOTENSÃO
PÓS-EXERCÍCIO**

LAVRAS - MG

2018

DENIZE PALMITO DOS SANTOS

**MODELAGEM GAMLSS DA PRESSÃO ARTERIAL SANGUÍNEA EM UM ESTUDO
DE HIPOTENSÃO PÓS-EXERCÍCIO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Dra. Izabela Regina Cardoso de Oliveira
Orientadora

LAVRAS - MG

2018

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Santos, Denize Palmito dos.

Modelagem GAMLSS da pressão arterial sanguínea em um estudo de hipotensão pós-exercício / Denize Palmito dos Santos. – Lavras : UFLA, 2018.

60 p. : il.

Dissertação (mestrado acadêmico)–Universidade Federal de Lavras, 2018.

Orientadora: Dra. Izabela Regina Cardoso de Oliveira.

Bibliografia.

1. Dados longitudinais. 2. Modelos mistos. 3. Modelos semiparamétricos. I. Oliveira, Izabela Regina Cardoso de. II. Título.

DENIZE PALMITO DOS SANTOS

**MODELAGEM GAMLSS DA PRESSÃO ARTERIAL SANGUÍNEA EM UM ESTUDO
DE HIPOTENSÃO PÓS-EXERCÍCIO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 17 de Dezembro de 2018.

Prof. Dr. Marcelo Ângelo Cirillo UFLA
Prof. Dr. Cristian Marcelo Villegas Lobos ESALQ/USP

Dra. Izabela Regina Cardoso de Oliveira
Orientadora

**LAVRAS - MG
2018**

*À minha família,
meus amigos,
pelo amor e carinho.
DEDICO*

AGRADECIMENTOS

Agradeço à Deus, por sempre me abençoar e me iluminar, guiando-me pelos melhores caminhos que eu deveria percorrer. À Universidade Federal de Lavras (UFLA), especialmente ao Departamento de Ciência Exatas, por me dar a oportunidade de conseguir conquistar o título de Mestra em Ciências. Ao CNPq, pela concessão da bolsa de mestrado. À Professora Dra. Izabela Regina Cardoso de Oliveira, por ter me dado a oportunidade de fazer o mestrado sob sua orientação. Sendo sempre um exemplo de profissional e honrando sempre os compromissos. Agradeço também pela confiança e por todo o ensinamento não apenas nesse trabalho, mas na vida. Aos meus pais Antenor e Dina pelo imenso amor, proteção, força e carinho. Ao meu querido irmão Denisar pelo amor, incentivo e exemplo de pessoa. Ao meu namorado Julio Cezar pelo amor, compreensão, cuidado e exemplo de pessoa. Sou muito grata aos professores e funcionários do Departamento de Estatística da UFLA, aos professores componentes da minha banca, Dr. Marcelo Angelo Cirillo, Dr. Cristian Marcelo Villegas Lobos, Dr. Renato Ribeiro de Lima. À professora Dra. Giancarla Aparecida Botelho pela contribuição e por ter fornecido os dados que foram analisados nessa dissertação. À todos os amigos que conheci durante essa etapa, muito obrigada pela troca de conhecimento, pelos momentos de descontração e pelo incentivo durante esses dois anos de mestrado.

RESUMO

Em Ciências da Saúde, os estudos controlados desempenham um papel importante no desenvolvimento de técnicas que ajudam na prevenção ou na cura de doenças. Frequentemente, nessa área, utilizam-se análises estatísticas de medidas repetidas no tempo, pois essas aplicadas a dados dessa natureza permitem, vantajosamente, observar as possíveis mudanças na variável resposta ao longo do tempo. A metodologia mais flexível, para a análise de dados longitudinais, envolve o uso de modelos mistos, em que o método da máxima verossimilhança (ML) ou máxima verossimilhança restrita (REML) podem ser usados na estimação dos parâmetros. A pressão arterial (PA) é aferida em termos da pressão arterial sistólica (PAS) e diastólica (PAD) e seus altos valores levam à hipertensão. Essa doença é um problema mundial que, se não tratada corretamente, pode levar o hipertenso a sofrer consequências graves. Atualmente, buscam-se métodos não farmacológicos, como a prática de atividades físicas, para o controle da PA. Um estudo longitudinal foi desenvolvido em Lavras, MG, para avaliar a PA de 26 indivíduos adultos normotensos submetidos a dois programas de treinamento (resistência e hipertrofia). Aos dados de PAS e PAD obtidos, ao longo de 18 sessões, foram ajustados modelos semiparamétricos com efeito aleatório. Para isso, utilizou-se a classe de modelos aditivos generalizados para localização, escala e forma (GAMLSS). Em todos os modelos (para PAS e PAD), a *baseline* foi incluída como parte do vetor de respostas, assumindo homogeneidade entre grupos antes do início dos treinamentos. De maneira geral, os resultados indicaram que ambos os programas de treinamento provocam adaptações da PA que são capazes de diminuir os níveis tensionais pré-treinamento. Observou-se que o treinamento de resistência ocasiona maior variabilidade nos dados da PAS do que o treinamento de hipertrofia e esse último resultou em um efeito hipotensor crônico mais acentuado nos dados da PAD. Todas as análises foram realizadas no programa R.

Palavras-chave: Dados longitudinais. Efeito aleatório. Modelos Mistos. Modelos semiparamétricos.

ABSTRACT

In Health Sciences, controlled studies play an important role in the development of techniques that promote prevention or cure of diseases. Frequently, statistical analyzes of repeated measures in time are used, since they allow to investigate the possible changes in the response variable over time. The more flexible methodology for the analysis of longitudinal data involves the use of mixed models, where the maximum likelihood (ML) or restricted maximum likelihood (REML) methods can be used in parameters estimation. Blood pressure (BP) is measured in terms of systolic (SBP) and diastolic (DBP) pressure and their high values lead to hypertension. This disease is a worldwide problem that, if not treated correctly, can lead the patient to suffer serious consequences. Currently, non-pharmacological methods, as the practice of physical activities, are prescribed for the control of BP. A longitudinal study was carried out in Lavras, MG, to evaluate the BP of 26 normotensive adults submitted to two training programs (resistance and hypertrophy). Semiparametric models with random effect were fitted to data from SBP and DBP obtained over 18 sessions. For this, we use the class of generalized additive models for location, scale and shape (GAMLSS). In all models (for SBP and DBP) the baseline was included as part of the response vector, assuming homogeneity between groups before training. Overall, the results indicated that both training programs lead to BP adaptations that are able to decrease pre-training blood pressure levels. Resistance training has been found to cause greater variability in SBP data than hypertrophy training and the latter showed a higher hypotensive effect on DBP. All analyzes were performed in R.

Keywords: Longitudinal data. Random effects. Mixed model. Semiparametric model.

LISTA DE FIGURAS

Figura 3.1 – Gráficos boxplot para os Grupos 1 (Treinamento de Resistência) e 2 (Treinamento de Hipertrofia) para as variáveis respostas PAS na <i>baseline</i> (a), PAS considerando todas as sessões (b), PAD na <i>baseline</i> (c) e PAD considerando todas as sessões (d).	43
Figura 3.2 – Gráficos de perfis médios dos Grupos 1 (Treinamento de Resistência) e 2 (Treinamento de Hipertrofia), para as variáveis respostas PAS (a) e PAD (b), ao longo das 18 sessões de treinamento.	44
Figura 3.3 – Gráficos do termo aditivo para os dados da PAS (a) e PAD (b) ao longo das sessões de treinamento.	46
Figura 3.4 – Gráfico de valores preditos de PAD, para o Grupo 1 (Treinamento de Resistência) e para o Grupo 2 (Treinamento de Hipertrofia) ao longo das sessões de treinamento.	47
Figura 3.5 – Gráficos de índices dos resíduos quantílicos dos modelos ajustados aos dados de PAS (a) e PAD (b), gráficos <i>worms</i> dos modelos para PAS (c) e PAD (d) e gráficos de probabilidade normal com envelope simulado dos modelos para PAS (e) e PAD (f).	48

LISTA DE TABELAS

Tabela 2.1 – Funções de ligação canônicas de algumas distribuições estatísticas.	15
Tabela 3.1 – Protocolos de realização dos treinamentos físicos.	39
Tabela 3.2 – Medidas descritivas das variáveis respostas PAS e PAD considerando todas as sessões.	42
Tabela 3.3 – Média e erro padrão da média dos valores de diferença nas medidas antropométricas antes e após o treinamento. Os valores-p reportados referem-se ao teste de Wilcoxon-Mann-Whitney para comparar as diferenças entre os dois grupos de treinamento.	45
Tabela 3.4 – Estimativas dos efeitos fixos dos modelos ajustados aos dados de PAS e PAD.	46

SUMÁRIO

	PRIMEIRA PARTE	10
1	INTRODUÇÃO	10
1.1	Justificativa e contribuições	12
2	REFERENCIAL TEÓRICO	13
2.1	Modelo linear	13
2.2	Modelo linear generalizado	14
2.3	Modelos aditivos generalizados	15
2.4	Modelagem de dados longitudinais	16
2.4.1	Modelos lineares mistos	18
2.5	Principais estudos clínico-epidemiológicos e o ajuste para medidas de referência (<i>baseline</i>)	20
2.6	Modelos aditivos generalizados para locação, escala e forma	22
2.6.1	O modelo estatístico	22
2.6.2	Resíduos quantílicos	24
2.6.3	Seleção de covariáveis	26
2.6.4	Suavização por <i>P-spline</i>	26
3	CONCLUSÃO	29
	REFERÊNCIAS	30
	SEGUNDA PARTE- ARTIGO	34
	ARTIGO - Modelos mistos semiparamétricos para avaliar o efeito hipotensor de diferentes treinamentos físicos prescritos a adultos normotensos	34
1	INTRODUÇÃO	35
2	MATERIAL E MÉTODOS	37
2.1	Descrição dos grupos de estudo	37
2.2	Aferição da pressão arterial	39
2.3	Coleta de dados antropométricos	40
2.4	Análise estatística	40
3	RESULTADOS E DISCUSSÃO	42
3.1	Análise exploratória	42
3.2	Avaliação das medidas antropométricas nos grupos de treinamento	44
3.3	Os efeitos dos treinamentos físicos na pressão arterial	45

4	CONCLUSÃO	50
	REFERÊNCIAS	51
	APÊNDICES	56

PRIMEIRA PARTE

1 INTRODUÇÃO

A Organização Mundial de Saúde (OMS) define qualidade de vida como uma percepção intrínseca que cada indivíduo possui sobre si em seu contexto cultural, indicando, assim, um grau de condições básicas que envolvem o bem-estar físico e psicológico do ser humano. Nessa vertente, qualidade de vida e saúde são termos intimamente relacionados, pois se entendem à saúde como um conjunto de fatores (econômicos, sociais, culturais e políticos) que estão diretamente associados com a qualidade de vida.

A prática de atividade física realizada regularmente favorece a manutenção de uma vida saudável (LANDAL et al., 2014; SATO et al., 2018). Indivíduos não sedentários apresentam maior resistência contra problemas psicológicos e físicos, como ansiedade e depressão, estabilidade emocional e diminuição nos níveis de estresse. Além disso, é considerado um método não farmacológico para a prevenção da hipertensão arterial (RADOVANOVIC et al., 2014; SOARES-MIRANDA et al., 2016).

A Hipertensão Arterial Sistêmica (HAS) ou usualmente chamada de pressão alta é um problema grave de saúde pública no Brasil e no mundo. Ela provoca graves consequências na saúde, desencadeando inúmeras doenças, entre as quais insuficiência cardíaca, infarto do miocárdio, arritmias cardíacas, morte súbita e outras.

Uma análise conjunta, envolvendo a OMS e centenas de cientistas, revelou, em novembro de 2016, que a hipertensão atingiu cerca de 1,13 bilhão de pessoas no mundo, quase o dobro do número de casos registrados em 1975. Segundo a Sociedade de Cardiologia do Estado de São Paulo (SOCESP), acredita-se que a hipertensão atinja mais de 30 milhões de pessoas no Brasil e, de acordo com o Ministério da Saúde, apenas 10% fazem o controle adequado. Diversos fatores, como a obesidade, sedentarismo, estresse e herança familiar, contribuem para a elevação da pressão arterial.

Os estudos desenvolvidos na área da saúde são responsáveis pelo desenvolvimento de técnicas que ajudam na prevenção de doenças e proporcionam melhor qualidade de vida aos indivíduos de uma sociedade. Independente do delineamento utilizado, em estudos desse tipo, sempre haverá heterogeneidade entre as unidades analisadas (seres humanos, animais, material biológico), que são provenientes da própria estrutura física, gênero, fatores ambientais, genéticos, sociais e até comportamentais (FITZMAURICE et al., 2004).

Além disso, os estudos nessa área utilizam com muita frequência dados com medidas repetidas ao longo do tempo, denominados estudos longitudinais. Esses estudos, por meio de uma modelagem estatística, permitem detectar características individuais, partindo de um modelo misto e globais, dado por modelos marginais. Em ambas as abordagens, a matriz de covariâncias referentes à estrutura que explica a evolução temporal das respostas, em função das covariáveis, é contemplada (SINGER; NOBRE; ROCHA, 2008).

Em estudos relacionados à pressão arterial, muitos pesquisadores têm utilizado métodos estatísticos flexíveis, como os modelos aditivos generalizados (GAM) na análise de dados (RATHNAYAKE; CHOUDHARY, 2017; YU et al., 2018). Liu e Tu (2012), por exemplo, propõem o uso de modelos de regressão semiparamétricos para avaliar a influência da altura e peso de crianças na pressão sistólica e diastólica medidas em um estudo longitudinal. Dada a correlação entre os dois tipos de pressão aferidos, os autores propõem a sua modelagem simultânea, tratando a resposta como um vetor bivariado.

Uma classe de modelos ainda mais flexível é a dos modelos aditivos generalizados para a posição, escala e forma (GAMLSS) proposta por Rigby e Stasinopoulos (2005). O GAMLSS foi introduzido com o intuito de superar algumas limitações associadas aos modelos lineares generalizados (MLG) e aos modelos aditivos generalizados (GAM). Além desse, podem-se ajustar vários modelos de regressão. Essa classe apresenta uma modelagem flexível na regressão paramétrica e/ou semiparamétrica e a distribuição da variável resposta não necessariamente precisa pertencer à família exponencial, sendo substituída por uma família de distribuições mais geral, permitindo avaliar inferências não apenas no parâmetro de posição. Dessa forma, é possível relacionar as variáveis explicativas e efeitos aleatórios a cada um dos parâmetros da distribuição e não apenas ao de localização (RIGBY; STASINOPOULOS, 2005).

O principal objetivo desta dissertação é divulgar o uso dos modelos GAMLSS aplicando-os em um estudo sobre hipotensão pós exercício. No artigo resultante, são analisadas medidas de pressão arterial (PA) de dois grupos de indivíduos que foram submetidos a diferentes tipos de treinamento físico, resistência e hipertrofia. Nesse ensaio, os valores de pressão arterial sistólica (PAS) e pressão arterial diastólica (PAD) foram obtidos, antes do início dos treinamentos (medida de referência ou *baseline*) e, ao longo de 18 sessões, caracterizando um estudo longitudinal. O objetivo é comparar os dois tipos de treinamento em relação à redução na PA. Na análise, são usados modelos semiparamétricos com efeitos fixos e aleatórios no do contexto dos GAMLSS.

1.1 Justificativa e contribuições

Mesmo com a disponibilidade de métodos estatísticos mais poderosos e flexíveis, muitos pesquisadores de áreas aplicadas preferem utilizar modelos lineares ou técnicas ainda mais simples como testes de médias (paramétricos ou não paramétricos), sem se atentar para as pressuposições e adequabilidade desses procedimentos. Então, uma das contribuições deste trabalho é divulgar o uso apropriado de uma modelagem estatística flexível e disponibilizar os códigos computacionais utilizados. Assim, esperamos encorajar pesquisadores de áreas aplicadas a conhecer e aplicar esses métodos de análise a seus problemas de pesquisa.

Os modelos GAMLSS têm sido aplicados a problemas práticos de diversas áreas do conhecimento, como: Ciências Agrárias, Ciências Sociais, Ciência Ambiental, Biologia, Medicina (em estimativas de curvas de crescimento, por exemplo) e outras (STASINOPOULOS et al., 2017). Apesar dessa ampla difusão, a utilização desses modelos no contexto prático, apresentado neste trabalho, é inédita.

Finalmente, apesar da forte tradição e atuação da Universidade Federal de Lavras (UFLA) em Ciências Agrárias, com os cursos recentemente criados em Ciências da Saúde, há novas demandas por análise estatística nessas áreas do conhecimento. Nesse aspecto, contribuiu-se ao elucidar alguns termos frequentemente utilizados em estudos dessa natureza e ao desenvolver um trabalho aplicado na área.

2 REFERENCIAL TEÓRICO

Nesta seção, algumas definições introdutórias serão apresentadas e expandidas com o intuito de fundamentar os principais conceitos relacionados aos modelos GAMLSS.

2.1 Modelo linear

Os modelos lineares desempenham um papel central nos métodos estatísticos. São aplicados a dados (x_i, y_i) , para $i = 1, 2, 3, \dots, n$, em que, convencionalmente, y é a variável resposta e x é chamada de variável explicativa ou independente. Em muitas situações, a variável resposta é influenciada por mais de uma variável explicativa que, na maioria das vezes, são controladas pelo pesquisador. No entanto existem influências externas, não controláveis.

Dessa forma, o modelo linear clássico é definido, matricialmente, por Searle (1971), como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

em que:

\mathbf{y} representa o vetor da variável resposta de dimensão $n \times 1$ dos dados observados;

\mathbf{X} é a matriz das variáveis explicativas, de dimensões $n \times p$;

$\boldsymbol{\beta}$ é o vetor de parâmetros desconhecidos de efeitos fixos, de dimensão $p \times 1$;

$\boldsymbol{\varepsilon}$ é o vetor de erros aleatórios, de dimensão $n \times 1$.

O objetivo, ao se aplicar uma modelagem linear a um conjunto de dados, é modelar a média de \mathbf{y} usando o vetor de parâmetros de efeitos fixos $\boldsymbol{\beta}$. Admitindo que $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I})$ é um vetor de componentes aleatórios independentes e identicamente distribuídos, com vetor de média $\mathbf{0}$ e matriz de variância e covariância $\boldsymbol{\sigma}^2\mathbf{I}$, tem-se que $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\sigma}^2\mathbf{I})$.

As estimativas dos parâmetros podem ser obtidas utilizando o método da máxima verossimilhança ou o método dos mínimos quadrados. No método de mínimos quadrados, as estimativas são obtidas, resolvendo o Sistema de Equações Normais (S.E.N.) que é dado por:

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{y}.$$

Se a matriz \mathbf{X} for de posto coluna completo, então, $\mathbf{X}^T\mathbf{X}$ é não singular, admite inversa simples e a solução do S.E.N. é:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

Se \mathbf{X} é de posto coluna incompleto, pode-se utilizar uma inversa generalizada e a solução do sistema passa a ser:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}.$$

2.2 Modelo linear generalizado

Propostos por Nelder e Wedderburn (1972), os modelos lineares generalizados (MLG) são uma extensão dos modelos lineares normais e diferem da modelagem clássica em alguns aspectos. Nesse caso, não são necessárias as pressuposições de normalidade e homocedasticidade, para a variável resposta, a qual deve pertencer à família de distribuição exponencial, e os dados são trabalhados na sua forma original sem necessidade de serem transformados (NELDER; WEDDERBURN, 1972).

Cordeiro e Demétrio (2008) mencionam que a classe dos modelos lineares generalizados são caracterizados por três aspectos:

i) **Componente aleatório:** o qual é constituído por um conjunto de variáveis aleatórias e independentes Y_i , com $i= 1,2,\dots,n$, provenientes de uma mesma distribuição que pertence à família exponencial escrita da seguinte forma:

$$f(y_i; \theta_i, \phi) = \exp[\phi^{-1} \{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)],$$

em que as funções $b(\cdot)$ e $c(\cdot)$ são conhecidas, θ_i é um parâmetro canônico e $\phi^{-1} > 0$ é um parâmetro de dispersão.

ii) **Componente sistemático:** é dado pelo pesquisador ao atribuir as covariáveis no preditor e essas variáveis explicativas entram na forma de uma soma linear.

$$\eta_i = x_i^T \boldsymbol{\beta} \quad \text{ou} \quad \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta},$$

sendo $\mathbf{X} = (x_1, \dots, x_n)^T$ a matriz do modelo, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ o vetor de parâmetros e o preditor linear é $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$.

iii) **função de ligação**: é uma função que relaciona o componente aleatório ao componente sistemático, ou seja, relaciona a média ao preditor linear, isto é,

$$\eta_i = g(\mu_i),$$

sendo $g(\cdot)$ uma função monótona e diferenciável.

Algumas distribuições pertencentes à família exponencial são a Poisson, Binomial e a Binomial Negativa, a Normal, a Gama e Normal Inversa. Para essas distribuições, na Tabela 2.1 encontram-se as funções de ligação comumente utilizadas.

Tabela 2.1 – Funções de ligação canônicas de algumas distribuições estatísticas.

Distribuição	Função de ligação canônica
Normal	Identidade: $\eta = \mu$
Poisson	Logarítmica: $\eta = \log(\mu)$
Binomial	Logística: $\eta = \log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{\mu}{m-\mu}\right)$
Gama	Recíproca: $\eta = \frac{1}{\mu}$
Normal Inversa	Recíproca do quadrado: $\eta = \frac{1}{\mu^2}$

Fonte: CORDEIRO; DEMÉTRIO (2008, p. 43)

2.3 Modelos aditivos generalizados

Os modelos aditivos generalizados (GAM), introduzidos por Hastie e Tibshirani (1990), proporcionam uma relação menos rígida do que os (MLG), de modo que o termo linear correspondente a cada variável preditora é substituído por uma função não especificada dessa variável. Em outras palavras, um GAM consiste em um modelo linear generalizado com o preditor linear envolvendo uma soma de funções suavizadoras. De forma mais explícita, os preditores em MLG e em GAM são dados, respectivamente, por:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p,$$

$$\eta = \beta_0 + \beta_1 f_1(x_1) + \beta_2 f_2(x_2) + \dots + \beta_p f_p(x_p).$$

Os procedimentos referentes a testes de hipóteses e métodos numéricos de ajuste foram desenvolvidos de maneira análoga aos MLG. Contudo incorpora-se a forma em que as funções f_1, f_2, \dots sejam estimadas por meio de suavizadores.

Em síntese, na maioria das técnicas de suavização, o valor predito suavizado da resposta é obtido com base em uma “média” de r observações na vizinhança de um dado valor x_i . A questão é que diferentes formas de cálculo dessa média, em uma vizinhança de x_i , definem diferentes métodos de suavização.

A escolha do tamanho da vizinhança está associada a um parâmetro (λ) denominado parâmetro de suavização, que deve ser fixado antes do processo de suavização. A escolha de valores, para esse parâmetro, depende da relação entre o viés e a variância da curva estimada.

Dentre as funções de suavização que podem ser usadas, citam-se LOESS (trata-se de uma suavização de regressão local) e *splines* cúbico. O ajuste do modelo GAM no R (R Core Team, 2018) pode ser feito, por exemplo, utilizando os pacotes *gam* e *mgcv*. No pacote *gam* as estimativas das funções suavizadoras são feitas por um método denominado retroajuste, uma maneira simples, porém robusta, de estimar um suavizador por vez. Já o pacote *mgcv* usa *splines* e os quais requerem um pouco mais de compreensão matemática do que os métodos utilizados na *gam*.

A principal vantagem do GAM no pacote *gam* é sua simplicidade, sendo fácil de entender e explicar. A vantagem do GAM no *mgcv* é que ele permite validação cruzada e uma modelagem mista aditiva generalizada (GAMM), incluindo correlações espaciais e temporais, bem como dados aninhados e vários padrões de heterogeneidade. O leitor mais interessado no tema pode consultar Hastie e Tibshirani (1990), Cleveland (1979) e Wood (2006).

2.4 Modelagem de dados longitudinais

O principal objetivo, ao se realizar um estudo longitudinal, é avaliar as mudanças individuais na variável resposta ao longo do tempo. Dessa forma, estudos dessa natureza consistem em estimar as mudanças individuais e examinar os fatores que influenciam a heterogeneidade entre os indivíduos observando como eles mudam com o tempo.

Estudos longitudinais são muito comuns na área da saúde e seus principais aspectos serão apresentados a seguir. Ao leitor interessado em aprofundar o tema, sugere-se consultar Fitzmaurice et al. (2008), Fitzmaurice, Laird, Ware (2004), Molenberghs e Verberke (2005) e Verbeke e Molenberghs (2000).

Dados longitudinais são um caso especial de medidas repetidas, multivariados por natureza e possuem uma estrutura complexa nos erros que deve ser considerada nas análises. Segundo Fitzmaurice et al. (2008), um dos métodos mais antigos propostos, para a análise

de dados longitudinais, foi a Análise de Variância (ANOVA) com um efeito aleatório para a unidade experimental, que induz à correlação entre medidas repetidas no mesmo indivíduo. Como denominado pelos autores, o modelo de ANOVA univariado para medidas repetidas, considerando N observações, é dado por

$$Y_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + b_i + \varepsilon_{ij},$$

$$i = 1, 2, \dots, N \quad \text{e} \quad j = 1, 2, \dots, n_i,$$

em que Y_{ij} é variável resposta de interesse, \mathbf{X}_{ij} é uma matriz de delineamento conhecida, $\boldsymbol{\beta}$ é um vetor de parâmetros de regressão, b_i é um vetor de parâmetros de efeitos aleatórios e $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ são efeitos aleatórios. Historicamente o modelo ANOVA apresenta notáveis semelhanças com os dados coletados, em um experimento casualizado em blocos, para os quais o efeito de blocos é aleatório. Por esse motivo, assume-se a estrutura de simetria composta para a matriz de variâncias e covariâncias, isto é:

$$\text{Var}(Y_{ij}) = \sigma_b^2 + \sigma_\varepsilon^2 \quad \text{e} \quad \text{Cov}(Y_{ij}, Y_{ik}) = \sigma_b^2 \quad (j \neq k).$$

De acordo com Fitzmaurice et al. (2008), uma abordagem relacionada é a ANOVA multivariada (MANOVA), para medidas repetidas que, como o nome sugere, trata as medidas repetidas no tempo como um vetor de respostas multivariado. Um caso especial desses modelos é a abordagem conhecida como análise de perfis. Segundo os autores, esse procedimento consiste na obtenção de conjuntos de variáveis derivadas, baseadas em combinações lineares das medidas repetidas, que são analisadas pela MANOVA. O método fornece informação sobre a média da resposta, ao longo de todas as medidas, bem como sobre a mudança da resposta com o tempo.

Apesar de ser computacionalmente mais trabalhosa que a ANOVA univariada, a MANOVA permite mais flexibilidade para a estrutura de covariâncias entre medidas repetidas do mesmo indivíduo. Ainda assim, ambos os métodos baseados no procedimento de ANOVA possuem diversas limitações. A ANOVA univariada impõe uma pressuposição muito restritiva na estrutura de covariâncias que, muitas vezes, não é apropriada a estudos longitudinais, porque, em geral, espera-se uma redução na correlação com o aumento da distância entre os pontos no tempo. Também a variabilidade na resposta, no início do estudo, pode ser diferente da variabi-

lidade ao longo do tempo, especialmente, quando a primeira medida é uma *baseline* (medida de referência).

Dentre as desvantagens da MANOVA, cita-se a impossibilidade de usá-la para ensaios não balanceados, ao longo do tempo e de modelar padrões gerais de dados faltantes (*missing*). Se um indivíduo possui um valor ausente, em qualquer ocasião, todo o vetor com informações do indivíduo é excluído da análise. Para contornar essas limitações, técnicas mais versáteis foram propostas.

2.4.1 Modelos lineares mistos

Os modelos lineares mistos (MLM) são uma das técnicas utilizadas para a análise de dados longitudinais. Esses modelos contêm efeitos fixos, além da média geral, e efeitos aleatórios, além do erro e incluem a ANOVA para medidas repetidas como caso especial. Sua aplicação, nesse contexto, ocorreu com o clássico trabalho de Laird e Ware (1982). O modelo linear de efeitos mistos pode ser definido por:

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}) \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i) \end{aligned} \tag{2.1}$$

sendo Y_i a variável resposta de interesse, $\boldsymbol{\beta}$ é um vetor de parâmetros ($p \times 1$) dos efeitos fixos, \mathbf{X}_i é uma matriz de delineamento ($n_i \times p$) dos efeitos fixos, \mathbf{b}_i é um vetor de parâmetros ($q \times 1$) de efeitos aleatórios, \mathbf{Z}_i é uma matriz de delineamento conhecida de efeitos aleatórios de ordem ($n_i \times q$) e $\boldsymbol{\varepsilon}_i$ é um vetor ($n_i \times 1$) de erros aleatórios. Os MLM fornecem uma estrutura flexível para as covariâncias \mathbf{D} e \mathbf{R}_i .

Quando a matriz de variâncias e covariâncias do erro aleatório for $\mathbf{R}_i = \sigma^2\mathbf{I}$, o modelo é chamado de modelo de independência condicional homocedástico, indicando que as observações são condicionalmente independentes. Com frequência, essa estrutura é mais usada, no entanto, outras estruturas de covariância mais gerais podem ser acomodadas.

As matrizes \mathbf{X}_i e \mathbf{Z}_i podem se diferenciar variando entre unidades experimentais e permitem a modelagem de dados desbalanceados e incompletos. Para estimação e inferência sobre os parâmetros dos modelos, são comumente utilizados os métodos da máxima verossimilhança (ML) e da máxima verossimilhança restrita (REML) (VERBEKE; MOLEMBERGHS, 2000;

FITZMAURICE; LAIRD; WARE, 2004), sendo esse último recomendado para diminuir o viés do ML para pequenas amostras.

Laird e Ware (1982) mostram que o modelo (2.1) pode ser interpretado como um modelo linear em dois estágios. No primeiro, são considerados fixos os efeitos aleatórios \mathbf{b}_i , de forma que o modelo pode ser reescrito como:

$$\mathbf{Y}_i | \mathbf{b}_i \sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \mathbf{R}_i).$$

No segundo estágio, os vetores \mathbf{b}_i são considerados independentes com distribuição $N \sim (\mathbf{0}, \mathbf{D})$ e marginalmente o modelo é:

$$\mathbf{Y}_i \sim N(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T + \mathbf{R}_i).$$

O processo de estimação também está relacionando com a escolha da estrutura da matriz de variância e covariância, visto que essa escolha afeta os erros padrões de efeitos fixos, diagnósticos e inferências. De acordo com Littell, Pendergast e Natarajan (2000), busca-se escolher uma estrutura mais parcimoniosa, envolvendo o menor número de parâmetros a serem estimados, sugerindo a utilização dos critérios de AIC e BIC para a escolha das estruturas de covariâncias.

Os MLMs são adequados para respostas contínuas que são, ao menos aproximadamente, normalmente distribuídas. Para respostas discretas, esses modelos já não são válidos e a classe mais geral dos modelos lineares generalizados mistos (MLGM) deve ser considerada. Esses modelos são uma extensão dos MLG (Seção 2.2) e a não linearidade entre o preditor e a resposta média exige alguns cuidados adicionais na interpretação dos coeficientes do modelo.

Segundo Fitzmaurice et al. (2008), os MLG foram estendidos, para acomodar dados longitudinais, por meio de diversas maneiras. Em termos gerais, as três classes de modelos são: (i) modelos marginais, (ii) modelos de efeitos aleatórios e (iii) modelos condicionais às respostas. Essas estratégias de modelagem diferem tanto na forma como a correlação entre medidas é considerada quanto na forma de interpretação dos parâmetros da regressão.

O ajuste de modelos mistos no programa R ganhou popularidade com o pacote *lme4* (BATES et al., 2015). Os MLM são ajustados, por meio da função *lmer*, enquanto, para o ajuste de MLGM, é utilizada a função *glmer* do pacote *lme4*. Nesse último caso, sete distribuições

podem ser consideradas para a variável resposta. Modelos não-lineares mistos também foram implementados por Pinheiro et al. (2018) e são ajustados pela função *nlme* do pacote *nlme*.

2.5 Principais estudos clínico-epidemiológicos e o ajuste para medidas de referência (*baseline*)

Nesta seção, serão apresentados os conceitos associados aos principais estudos clínico-epidemiológicos, abordando os ensaios clínicos e os estudos observacionais. Em geral, os estudos observacionais são divididos em estudos de corte transversal, estudos de coorte e caso-controle.

Em ensaios clínicos (estudos experimentais), o pesquisador tem o objetivo de avaliar o impacto de um determinado tratamento ou realizar uma investigação, ao longo da duração do estudo, que é feito sob condições controladas. Quase sempre, os indivíduos são incluídos no experimento de maneira elegível e a alocação aos tratamentos é feita de forma aleatória.

Os estudos observacionais não provêm de um experimento, por isso, não há influência do pesquisador. Segundo Giolo (2017), o que caracteriza os estudos transversais é o método utilizado para a coleta de dados, os quais são obtidos em um ponto específico no tempo (corte transversal), e cujo objetivo é obter informações instantâneas sobre a variável de interesse. Por outro lado, um estudo de coorte é constituído por um grupo de indivíduos (coorte) que apresentam algumas características semelhantes e são observados em um período de tempo a fim de se avaliar as possíveis mudanças entre eles. São classificados em retrospectivo (ocorrência de desfecho antes do início da pesquisa) e prospectivo (não ocorrência do desfecho).

O estudo caso-controle é essencialmente semelhante ao estudo de coorte. O que o diferencia é a forma de seleção e a coleta das informações. Nesses estudos, que são sempre retrospectivos, dois grupos são separados e comparados. O primeiro grupo é denominado de caso e é constituído de indivíduos que apresentam uma determinada característica de interesse. No segundo grupo, chamado controle, os indivíduos são livres dessa característica (FONTELLES, 2012; GIOLO, 2017).

Em ensaios clínicos e em alguns estudos observacionais (coorte), em regra, as medidas são tomadas ao longo do tempo. Fitzmaurice, Laird e Ware (2004) relatam que, em estudos longitudinais, normalmente, os sujeitos possuem uma medida inicial ou frequentemente denominada medida *baseline*. Ela é utilizada para diversos fins: como critério para a inclusão ou não

de indivíduo no experimento ou para medir os efeitos dos tratamentos em termos da mudança em relação às condições iniciais.

Os mesmos autores apresentam uma discussão sobre o ajuste para a resposta *baseline* dentro do contexto de análise de perfis considerando os tipos de estudos. Nos ensaios clínicos, essa medida é obtida antes da intervenção do pesquisador. Em consequência, a resposta média, na primeira ocasião (*baseline*), é independente do tratamento e as médias podem ser consideradas iguais na *baseline*.

Assim, o principal interesse está na interação tratamento \times tempo. Por outro lado, em estudos observacionais, em muitas situações, não é possível assumir que as médias são as mesmas na *baseline*. Nesses estudos, tanto o efeito da interação quanto os efeitos principais são relevantes.

Logo, uma questão que surge, quando há a presença de medidas *baseline*, é como incluir essa informação no processo de modelagem. Segundo os autores, há quatro estratégias para lidar com essa medida:

1. considerá-la como parte do vetor resposta sem fazer pressuposições sobre as diferenças entre grupos na *baseline*.
2. considerá-la como parte do vetor resposta e assumir que as médias nos grupos são iguais na *baseline*.
3. subtrair a resposta na *baseline* das medidas nos outros tempos e analisar as diferenças em relação à *baseline*, ou
4. usar a *baseline* como covariável.

A primeira estratégia corresponde a um modelo com efeitos principais de tratamento e tempo e com interação, enquanto, na segunda, o efeito principal de tratamento é suprimido. Ambas as estratégias são válidas, em ensaios clínicos e em estudos observacionais, em que é possível assumir que as médias dos grupos são iguais na *baseline*.

A terceira e a quarta estratégias focam na mudança em relação à medida de referência e o vetor resposta consiste de dados obtidos a partir do início do estudo. A interpretação dos testes de hipóteses é, então, diferente. Os autores mostram que as estratégias 1 e 2 são equivalentes às suas correspondentes em 3 e 4. No entanto, por questões práticas, as estratégias 1 e 2 são preferidas e a escolha entre elas deve ser direcionada pelo delineamento do estudo.

2.6 Modelos aditivos generalizados para locação, escala e forma

Os modelos aditivos generalizados para locação, escala e forma (GAMLSS) foram introduzidos inicialmente por Rigby e Stasinopoulos (2005) com o objetivo de superar algumas limitações dos MLG e GAM. O GAMLSS apresenta algumas peculiaridades. Inicialmente, pode-se destacar que não há mais restrição da família exponencial, para a variável resposta y , sendo substituída por uma família de distribuições \mathcal{D} .

A família de distribuições \mathcal{D} é mais geral e implica que a distribuição deva estar na forma paramétrica, podendo conter distribuições de até quatro parâmetros, incluindo termos ligados à assimetria e/ ou à curtose. Dessa forma, define-se que a variável resposta y segue distribuição $D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})$, de modo que $D \in \mathcal{D}$. Vale ressaltar que $\boldsymbol{\mu}$ e $\boldsymbol{\sigma}$ são parâmetros de locação e escala, respectivamente e $\boldsymbol{\nu}$ e $\boldsymbol{\tau}$ parâmetros de forma, que estão frequentemente relacionados com a assimetria e curtose.

Os modelos GAMLSS permitem funções paramétricas lineares, não lineares ou funções de suavização não paramétricas de variáveis explicativas. A parte sistemática do modelo é, então, expandida e todos os parâmetros da distribuição podem ser modelados como funções de variáveis explicativas, não apenas a locação (posição). Uma variedade de termos aditivos pode ser incluída no modelo, o que fornece uma modelagem bastante flexível para superdispersão, excesso de zeros, assimetria e curtose nos dados.

2.6.1 O modelo estatístico

Stasinopoulos et al. (2017) definem que, no modelo GAMLSS, as observações y_i são independentes, $i = 1, 2, 3, \dots, n$, com função densidade de probabilidade $f(y_i | \boldsymbol{\theta})$ condicionais a $\boldsymbol{\theta}$, sendo:

$$\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau})^T.$$

De acordo com a formulação original do GAMLSS, \mathbf{y}^T é o vetor da variável resposta de tamanho n , de modo que $\mathbf{y}^T = (y_1, y_2, \dots, y_n)$. Para $k = 1, 2, 3, 4$, considere uma função monótona $g_k(\cdot)$, denominada função de ligação, que relaciona as variáveis explicativas a cada um dos parâmetros da distribuição, ou seja,

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}, \quad (2.2)$$

isto é,

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1} \boldsymbol{\gamma}_{j1},$$

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2} \boldsymbol{\gamma}_{j2},$$

$$g_3(\mathbf{v}) = \boldsymbol{\eta}_3 = \mathbf{X}_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3} \boldsymbol{\gamma}_{j3},$$

$$g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \mathbf{X}_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4} \boldsymbol{\gamma}_{j4},$$

sendo $\boldsymbol{\theta}_k^T$ um vetor de quatro parâmetros da distribuição;

$\boldsymbol{\eta}_k^T$ é o preditor linear;

$\boldsymbol{\beta}_k^T = (\beta_{1k}, \beta_{2k}, \dots, \beta_{J'_k})$ é um vetor de parâmetros de tamanho J'_k ;

\mathbf{X}_k é uma matriz de delineamento conhecida de efeitos fixos de tamanho $n \times J'_k$;

\mathbf{Z}_{jk} é uma matriz de delineamento conhecida de efeitos aleatórios e de ordem $n \times q_{jk}$;

$\boldsymbol{\gamma}_{jk}$ é uma variável aleatória q_{jk} -dimensional com distribuição $\boldsymbol{\gamma}_{jk} \sim N_{q_{jk}}(\mathbf{0}, \mathbf{G}_{jk}^-)$.

Stasinopoulos e Rigby (2007) mencionam que os termos de efeitos aleatórios $\boldsymbol{\gamma}_{jk}$ são independentes com distribuição $\boldsymbol{\gamma}_{jk} \sim N_{q_{jk}}(\mathbf{0}, \mathbf{G}_{jk}^-)$, em que \mathbf{G}_{jk}^- é uma inversa generalizada de ordem $q_{jk} \times q_{jk}$ da matriz simétrica $\mathbf{G}_{jk} = \mathbf{G}_{jk}(\boldsymbol{\lambda}_{jk})$ que pode depender de um vetor hiperparamétrico $\boldsymbol{\lambda}_{jk}$, sendo \mathbf{G}_{jk} singular. Então, $\boldsymbol{\gamma}_{jk}$ admite uma função de densidade imprópria proporcional a $\exp(-\frac{1}{2} \boldsymbol{\gamma}_{jk}^T \mathbf{G}_{jk} \boldsymbol{\gamma}_{jk})$.

No GAMLSS, existem vários submodelos importantes. Stasinopoulos e Rigby (2007) mostram que, quando $\mathbf{Z}_{jk} = \mathbf{I}_n$, em que \mathbf{I}_n é uma matriz identidade de ordem $n \times n$ e considerando $\boldsymbol{\gamma}_{jk} = h_{jk} = h_{jk}(x_{jk})$, para todas as combinações de j e k em (2.2), obtém-se a formulação aditiva semiparamétrica dada por:

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

para $k = 1, 2, 3, 4$, que representa os vetores de parâmetros $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, \mathbf{v} e $\boldsymbol{\tau}$ em que x_{jk} para $j = 1, 2, \dots, J_k$, são vetores de tamanho n . A função h_{jk} é desconhecida da variável explicativa e $h_{jk} = h_{jk}(x_{jk})$ é um vetor que avalia a função h_{jk} em x_{jk} .

Quando não ocorre termo aditivo no modelo, obtém-se o modelo linear paramétrico, definido por:

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

Quando são atribuídos, no modelo semiparamétrico, funções não lineares nos k parâmetros da distribuição, é obtido um modelo GAMLSS semiparamétrico não-linear, para $i = 1, \dots, n$ e $k = 1, 2, 3, 4$, definido por:

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

em que h_k são funções não lineares para $k = 1, 2, 3, 4$ e X_k é uma matriz conhecida de ordem $n \times J'_k$. Quando a esse modelo não é atribuída a parte aditiva, tem-se um modelo GAMLSS paramétrico não-linear, definido por:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = h(\mathbf{X}_k, \boldsymbol{\beta}_k).$$

No R (R CORE TEAM, 2018), os modelos GAMLSS são implementados por meio do pacote de mesmo nome (*gamlss*). Atualmente existem cerca de 100 distribuições implementadas em *gamlss.family*, além de pacotes adicionais, para termos espaciais (*gamlss.spatial*), dados com censura (*gamlss.cens*), implementação de distribuições de misturas (*gamlss.mx*) e outros.

2.6.2 Resíduos quantílicos

Nos diagnósticos dos modelos GAMLSS, são utilizados os resíduos quantílicos (aleatorizados) normalizados, que foram introduzidos por Dunn e Smyth (1996) e fundamentados na ideia de inverter a função de distribuição estimada para cada observação. Esse procedimento fornece os resíduos normais padrão e são utilizados para a avaliação da adequação de ajustes de modelos de regressão. Considerando uma distribuição $f(x|\boldsymbol{\theta})$, ajustada a um conjunto de observações y_i com $i = 1, 2, \dots, n$, os resíduos quantílicos (aleatorizados) normalizados ajustados são dados por:

$$\hat{r}_i = \Phi^{-1}(\hat{u}_i),$$

em que $\Phi^{-1}(\cdot)$ é o inversa da função de distribuição acumulada da normal padrão. Os \hat{u}_i são resíduos quantílicos, que são definidos de forma diferente para variáveis respostas contínuas e discretas.

Supondo que y seja uma observação de uma variável contínua, então, $u = F(y|\boldsymbol{\theta})$ e $\hat{u} = F(y|\hat{\boldsymbol{\theta}})$ são os valores da função na distribuição acumulada do modelo e do ajuste, respectivamente. Se o modelo estiver correto, então, u tem distribuição uniforme entre zero e um. Esse processo é denominado de transformação de probabilidade integral (STASINOPOULOS et al., 2017).

Caso y seja uma observação de uma variável discreta, então, $F(y|\boldsymbol{\theta})$ é uma função descontínua nos inteiros $y \in R_Y$. Dessa forma, a distribuição de $u = F(y|\boldsymbol{\theta})$ tem amplitude de zero a um, mas é discreta com probabilidade positiva nos pontos $F(y|\boldsymbol{\theta})$, $y \in R_Y$.

Os autores sugerem uma alternativa para contornar esse fato que é definindo u como um valor aleatório da distribuição uniforme no intervalo $[u_1, u_2] = [F(y-1|\boldsymbol{\theta}), F(y|\boldsymbol{\theta})]$ e, da mesma forma, \hat{u} é um valor aleatório de uma distribuição uniforme em $[\hat{u}_1, \hat{u}_2] = [F(y-1|\hat{\boldsymbol{\theta}}), F(y|\hat{\boldsymbol{\theta}})]$.

Stasinopoulos et al. (2017) consideram que a principal vantagem dos resíduos quantílicos (aleatorizados) normalizados é que, se o modelo estiver correto, qualquer que seja a distribuição da variável resposta, os resíduos verdadeiros sempre têm uma distribuição normal padrão. No pacote *gamlss*, esses resíduos podem ser obtidos, por meio da função *resid()* e existem outras funções que utilizam os resíduos quantílicos para obter informações sobre o modelo ajustado. Algumas delas são:

- *plot()*, para verificação geral de resíduos.
- A função *wp()* que fornece o *worm plot*. Este gráfico pode ser usado para identificar se a distribuição ajustada é adequada. Nesse gráfico, espera-se que, pelo menos, 95% dos pontos (resíduos) estejam contidos entre as duas curvas elípticas.

O formato da curva adaptada aos pontos do *worm* revela possíveis inadequações do modelo. Por exemplo, se o nível de pontos de plotagem no gráfico de *worms* estiver acima de uma linha horizontal na origem indica que a média residual é muito alta, o que implica que a locação da distribuição ajustada é muito baixa.

Uma tendência linear (positiva ou negativa), forma quadrática (**U** ou **U** invertido) ou forma cúbica (**S** ou **S** invertido) indicam um problema com a variância, assimetria ou curtose

dos resíduos, respectivamente. Esse fato, por sua vez, destaca um problema com a distribuição ajustada.

2.6.3 Seleção de covariáveis

Um dos critérios disponíveis no *gamlss*, para a seleção de covariáveis, é a função *stepGAIC()*, que é utilizada na construção do modelo para qualquer um dos parâmetros da distribuição. Os modelos são selecionados por etapas usando o Critério de Informação de Akaike Generalizada (GAIC).

Stasinopoulos et al. (2017) relatam que essa função foi baseada na função *stepAIC()* da biblioteca MASS (VENABLES; RIPLEY, 2002). Basicamente a seleção de covariáveis é feita por um procedimento para frente, para trás ou passo a passo, em que todas as variáveis consideradas no modelo são, individualmente analisadas, para serem descartadas em cada etapa, enquanto todas as variáveis, que não estão presentes no modelo, podem ser adicionadas.

2.6.4 Suavização por *P-spline*

No GAMLSS existem várias funções de suavização como *splines* cúbicos, *LOESS*, redes neurais e *P-splines*. Essa última será descrita com mais detalhes por ter sido a função de suavização utilizada na análise dos dados deste trabalho.

Os *P-splines* ganharam popularidade com o trabalho de Eilers e Marx (1996) e são polinômios definidos por funções de base *B-spline*, na variável explicativa, em que os coeficientes das funções de base são penalizados para garantir suavidade suficiente (STASINOPOULOS; RIGBY, 2007).

Lee e Durbán (2011) apresentam de maneira objetiva a definição da função *P-spline*, inicialmente, considerando um modelo não-paramétrico para uma resposta \mathbf{y} e uma covariável \mathbf{x} , representado por:

$$\mathbf{y} = f(\mathbf{x}) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}),$$

em que $f(\cdot)$ é uma função de suavização e $\boldsymbol{\varepsilon}$ é o erro aleatório com variância $\sigma^2 \mathbf{I}$. O método proposto por Eilers e Marx (1996) considera $f(\cdot)$ como uma soma de funções de base local, ou seja, $\mathbf{B}\boldsymbol{\theta}$, em que $\mathbf{B} = (\mathbf{B}_1(\mathbf{x}), \mathbf{B}_2(\mathbf{x}), \dots, \mathbf{B}_c(\mathbf{x}))$ é uma matriz $n \times c$ de *B-splines* (c depende do grau e número de nós da *B-spline*) construída a partir da covariável \mathbf{x} e $\boldsymbol{\theta}$ é o vetor de coeficientes de regressão (LEE; DURBÁN, 2011).

O P -spline minimiza a soma de quadrados penalizada

$$\mathbf{S}(\boldsymbol{\theta}; \mathbf{y}; \lambda) = (\mathbf{y} - \mathbf{B}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{B}\boldsymbol{\theta}) + \boldsymbol{\theta}^T \lambda \mathbf{P}\boldsymbol{\theta}, \quad (2.3)$$

em que a matriz \mathbf{P} é de penalidade discreta que depende de um parâmetro de suavização λ . Este termo de penalidade controla a suavidade do ajuste, aplicando penalidades sobre os coeficientes adjacentes. Definimos a matriz \mathbf{P} de ordem $c \times c$ como:

$$\mathbf{P} = \mathbf{D}^T \mathbf{D},$$

em que \mathbf{D} é uma matriz de diferença de ordem “ d ”, aplicada diretamente aos coeficientes de regressão. Em geral, é considerado uma diferença de segunda ordem a qual é definida como

$$(\boldsymbol{\theta}_1 - 2\boldsymbol{\theta}_2 + \boldsymbol{\theta}_3)^2 + \dots + (\boldsymbol{\theta}_{c-2} - 2\boldsymbol{\theta}_{c-1} + \boldsymbol{\theta}_c)^2 = \boldsymbol{\theta}^T \mathbf{D}^T \mathbf{D} \boldsymbol{\theta}.$$

Nesse caso, $\mathbf{D}^T \mathbf{D}$ define essa penalidade de segunda ordem, porém penalidades de primeira ou terceira ordem podem ser consideradas.

Para um dado valor de λ , a minimização de (2.3) produz

$$\hat{\boldsymbol{\theta}} = (\mathbf{B}^T \mathbf{B} + \mathbf{P})^{-1} \mathbf{B}^T \mathbf{y}.$$

A escolha de λ está, em geral, sujeita a um determinado critério. Três possibilidades são o critério de informação de Akaike (AIC) (Akaike, 1973), o critério de informação bayesiano (Schwarz, 1978) e validação cruzada generalizada (Craven e Wahba, 1979). Esses critérios são diretos para calcular P -splines, mas o cálculo da dimensão efetiva do modelo é necessário. Hastie e Tibshirani (1990) sugerem aproximar a dimensão efetiva ou graus de liberdade (gl) com a matriz *hat* do suavizador. A matriz *hat* da P -splines é definida como (LEE; DURBÁN, 2011).

$$\mathbf{H} = \mathbf{B}(\mathbf{B}^T \mathbf{B} + \mathbf{P})^{-1} \mathbf{B}^T.$$

As propriedades do traço de uma matriz permitem usar uma expressão computacionalmente mais eficiente para o grau de liberdade (gl), ou seja,

$$gl = \text{traço}(\mathbf{H}) = \text{traço}(\mathbf{B}(\mathbf{B}^T \mathbf{B} + \mathbf{P})^{-1} \mathbf{B}^T \mathbf{B}),$$

que envolve apenas matrizes $c \times c$, em vez do cálculo de H de ordem $n \times n$.

Nos *P-splines*, a quantidade de funções bases é menor que a quantidade de observações, diferentemente do que acontece no caso de suavização de *splines*, em que há tantos nós quanto dados e torna-se necessário trabalhar com matrizes de alta dimensão. O número de nós no *P-splines* não excede 40, o que os torna computacionalmente eficientes, em especial, quando se trabalha com uma grande quantidade de dados (LEE; DURBÁN, 2011).

3 CONCLUSÃO

As informações contidas no contexto apresentado incluem desde os modelos mais simples até a classe GAMLSS, que fornece uma estrutura eficiente e flexível na análise de dados longitudinais. Entre as principais vantagens desses modelos, podem-se destacar a grande variabilidade na escolha da distribuição para a variável resposta e a facilidade de modelar outros parâmetros da distribuição, o que os tornam modelos mais parcimoniosos.

O artigo resultante desta dissertação será apresentado na sequência. Os GAMLSS foram utilizados, para modelar dados de pressão arterial, que é uma variável de natureza biológica e apresenta grande variabilidade nas medições ao longo do tempo. Dessa forma, considerar modelos que sejam capazes de investigar e/ou avaliar as influências dessa variabilidade pode ser uma contribuição valiosa para os estudos epidemiológicos.

REFERÊNCIAS

AKAIKE, H. Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**, London, v. 60, n. 2, p. 255-265, Aug. 1973.

BATES, D et al. Fitting Linear Mixed-Effects Models Using lme4. **Journal of Statistical Software**, Los Angeles, v. 67, n. 1, p. 1-48, Oct. 2015.

CLEVELAND, W. S. Robust locally weighted regression and smoothing scatterplots. **Journal of the American Statistical Association**, New York, v. 74, n. 368, p. 829-836, Apr. 1979.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B. **Modelos Lineares Generalizados e extensões**. Piracicaba: ESALQ, 2008. 186 p.

CRAVEN, P.; WAHBA, G. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of cross-validation. **Numerical Mathematics**, Berlin, v. 31, n. 4, p. 377-403, Dec. 1979.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Alexandria, v. 5, n. 3, p. 236-244, Sept. 1996.

EILERS, P. H. C.; MARX, B. D. Flexible smoothing with B-splines and penalties (with comments and rejoinder). **Statistical Science**, Hayward, v. 11, p. 89-121, 1996.

FITZMAURICE, G. M. et al. (Ed.). **Longitudinal Data Analysis: A handbook of modern statistical methods**. Boca Raton: Chapman & Hall, 2008. 632 p.

FITZMAURICE, G. M.; LAIRD, N. M.; WARE, J. H. **Applied longitudinal analysis**. 2 ed. Canada: John Wiley & Son, 2004. 740 p.

FONTELLES, M. J. **Bioestatística aplicada à pesquisa experimental**. São Paulo: Livraria da Física, 2012. 420 p.

GIOLO, S. R. **Introdução à análise de dados categóricos com aplicações**. São Paulo: Edgard Blücher, 2017. 256 p.

HASTIE, T.; TIBSHIRANI, R. **Generalized Additive Models**. Boca Raton: Chapman & Hall, 1990. 352 p.

- LAIRD, N. M.; WARE, J. H. Random-Effects Models for longitudinal Data. **Biometrics**, Washington, v. 38, n. 4, p. 963-974, Dec. 1982.
- LANDAL, A. C. et al. Fatores associados à melhora da composição corporal em indivíduos com DPOC após treinamento físico. **Fisioterapia em Movimento**, Curitiba, v. 27, n. 4, p. 633-641, out./dez. 2014.
- LEE, D. J.; DURBÁN, M. P-spline ANOVA-type interaction models for spatio-temporal smoothing. **Statistical Modelling**, London, v. 11, n. 1, p. 49-69, Jan. 2011.
- LITTELL, R. C.; PENDERGAST, J.; NATARAJAN, R. Modelling covariance structure in the analysis of repeated measures data. **Statistics in Medicine**, New York, v. 19, n. 13, p. 1793-1819, July 2000.
- LIU, H.; TU, W. A. Semiparametric regression model for paired longitudinal outcomes with application in childhood blood pressure development. **The Annals of Applied Statistics**, Cleveland, v. 6, n. 4, p. 1861-1882, 2012.
- MIRANDA, L. S. Physical Activity and Risk of Coronary Heart Disease and Stroke in Older Adults: The Cardiovascular Health Study. **The Cardiovascular Health Study**, v. 133, n. 2, p. 147-155, 2016.
- MOLENBERGHS, G.; VERBEKE, G. **Models for Discrete Longitudinal Data**. New York: Springer, 2005. 683 p.
- NELDER, J. A.; WEDDERBURN, R. W.M. Generalized linear models. **Journal of Royal Statistical Society Series A, General**, London, v. 135, n. 3, p. 370-384, 1972.
- ORGANIZAÇÃO MUNDIAL DE SAÚDE- OMS. **Publicações da OMS**. Genebra: OMS, 2018. Disponível em: <<http://www.who.int/eportuguese/publications/pt/>>. Acesso em 10 mar. 2018
- ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE – OPAS. Doenças cardiovasculares. **OPAS**, Brasília, maio 2017. Disponível em: <https://www.paho.org/bra/index.php?option=com_content&view=article&id=5253:doencas-cardiovasculares&Itemid=839 />. Acesso em 10 mar. 2018.
- PINHEIRO, J. et al. 2014. **Linear and Nonlinear Mixed Effects Models** [S.l.: s.n.], 2018. 336 p.

R Core Team **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2018.

RADOVANOVIC, C. A. T. et al. Hipertensão arterial e outros fatores de risco associados às doenças cardiovasculares em adultos. **Revista Latino-Americana de Enfermagem**, Ribeirão Preto, v. 22, n. 4, p. 547-553, jul./ago. 2014.

RATHNAYAKE, L. N.; CHOUDHARY, P. K. Semiparametric modeling and analysis of longitudinal method comparison data. **Statistics in Medicine**, New York, v. 36, n. 13, p. 2003-2015, June 2017.

RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. **Journal of Statistical Software**, Los Angeles, v. 54, n. 3, p. 507-554, 2005.

SATO, K. et al. Association of physical activity with a systolic blood pressure difference between arms in older people. **Geriatrics Society**, Baltimore, v. 18, n. 1, p. 95-100, Jan. 2018.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, Hayward, v. 6, n. 2, p. 461-64, Mar. 1978.

SEARLE, S. R. **Linear Models**. New York: John Wiley & Sons, 1971. 532 p.

SINGER, J. M.; NOBRE, J. S.; ROCHA, F. M. M. **Análise de dados longitudinais**. São Paulo: Ed. USP, 2008. 203 p.

SOARES-MIRANDA, L. S. et al. Physical activity and risk of coronary heart disease and stroke in older adults: the cardiovascular health study. **Circulation**, Dallas, v. 133, n. 2, p. 147-155, Jan. 2016.

SOCIEDADE DE CARDIOLOGIA DO ESTADO DE SÃO PAULO - SOCESP. Hipertensão atinge mais de 30 milhões de pessoas no Brasil. **Blog do Coração**, São Paulo, abr. 2018. Disponível em: <<http://www.socesp.org.br/blogdocoracao/2018/04/23/hipertensao-atinge-mais-de-30-milhoes-de-pessoas-no-brasil/>>. Acesso em: 18 nov. 2018.

STASINOPOULOS, M. D. et al. **Flexible Regression and Smoothing: Using GAMLSS in R**. 1. ed. Boca Raton: Chapman & Hall, 2017. 549 p.

STASINOPOULOS, D. M.; RIGBY, R. A. Generalized additive models for location scale and shape (GAMLSS) in R. **Journal of Statistical Software**, Los Angeles, v. 20, n. 2, p. 1-64, 2007.

VENABLES, W. N.; RIPLEY, B. D. **Modern Applied Statistics with S**. 4. ed. New York: Springer, 2002. 495 p.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. New York: Springer-Verlag, 2000. 570

WOOD, S. N. **Generalized Additive Models: an introduction with R**. 2. ed. Boca Raton: Chapman & Hall, 2006. 476 p.

YU, Z. et al. Association between inflammation and systolic blood pressure in RA compared to patients without RA. **Arthritis Research & Therapy**, London, v. 20, n. 1, p. 107, June 2018.

SEGUNDA PARTE - ARTIGO**ARTIGO**

Modelos mistos semiparamétricos para avaliar o efeito hipotensor de diferentes treinamentos físicos prescritos a adultos normotensos

Modelos mistos semiparamétricos para avaliar o efeito hipotensor de diferentes treinamentos físicos prescritos a adultos normotensos

Denize Palmito dos Santos ¹, Carlos Roberto Barbosa Junior ², Izabela Regina Cardoso de Oliveira ¹, Giancarla Aparecida Botelho Santos ³

¹Departamento de Estatística, Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil

²Departamento de Educação Física, Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil

³Departamento de Ciências da Saúde, Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil

Resumo:

A hipertensão é uma doença crônica decorrente da elevação nas medidas de pressão arterial (PA) e que pode levar o hipertenso a sofrer graves consequências. A prática de exercício físico tem sido recomendada como um método não farmacológico para controle da PA. O objetivo deste trabalho é comparar o efeito hipotensor de dois treinamentos físicos (hipertrofia e resistência) em adultos normotensos não praticantes de atividades físicas. Este estudo foi desenvolvido em Lavras, MG, em que se avaliou a PA de 26 indivíduos, ao longo de 18 sessões. Medidas antropométricas também foram coletadas antes dos indivíduos serem submetidos aos treinamentos e ao final do estudo. Para avaliar o efeito (fixo) do tipo de exercício na redução da PA, foram utilizados modelos semiparamétricos dentro do contexto dos modelos aditivos generalizados para locação, escala e forma (GAMLSS). De maneira geral, os resultados indicaram que ambos os programas de treinamento provocam adaptações da PA e são capazes de diminuir os níveis tensionais pré-treinamento. Observou-se que o treinamento de resistência ocasiona maior variabilidade na pressão arterial sistólica se comparado ao treinamento de hipertrofia e esse último resultou em um efeito hipotensor crônico mais acentuado na pressão arterial diastólica.

Palavras-chave: Dados longitudinais, pressão arterial, treinamento físico, GAMLSS, hipertensão.

1. INTRODUÇÃO

A hipertensão arterial (HA) é uma doença crônica decorrente da elevação nas medidas da pressão arterial sistólica (PAS) e/ou pressão arterial diastólica (PAD). É considerada um problema de saúde pública global e está associada a doenças cardiovasculares, acidente vascular cerebral (AVC), insuficiência, morte súbita e incapacidade (WHO, 2013).

Alguns estudos recomendam a prática de atividades físicas para o controle e redução de pressão arterial (PA). Essas atividades contribuem também para a redução de peso corporal,

melhora da função imunológica, maior percepção da qualidade de vida, aumento dos níveis de força, resistência e flexibilidade e auxílio na manutenção do peso corporal em longo prazo. (LANDAL et al., 2014; SATO et al., 2018; MIRANDA et al., 2016; RADOVANOVIC et al., 2014; CUNHA et al., 2012; DIMEO, et al. 2012).

É sabido que a influência do exercício físico, sendo ele aeróbio ou resistido, na resposta da PA depende de diversos fatores, como o tipo de exercício, a intensidade, a demanda bioenergética determinante e a duração. Tais fatores determinam o grau e o efeito na PA, podendo estes efeitos ser de aumento ou diminuição da PA de forma aguda. Mas as ocorrências de respostas hipotensivas do exercício físico crônico em diversos indivíduos vêm sendo estudadas e estas, em geral, provocam diminuição significativa na PA (NEGRÃO, 2001; DAMORIM et al., 2016; PRATA et al., 2018).

Meta-análises têm apontado que a prática de atividade física realizada regularmente é capaz de reduzir a PAS em 3 a 5 mmHg e a PAD em 2 a 3 mmHg em indivíduos normotensos. Uma redução de 5/2 mmHg na PA diminui a primeira incidência de acidentes cardiovasculares fatais e não fatais em torno de 30% dos casos (DIMEO, et al. 2012; CORNELISSEN et al. 2011; WHELTON et al, 2002). Já em indivíduos hipertensos o treinamento aeróbico reduz a PAS em cerca de 7 mmHg e a PAD em 5mmHg (MEDINA et al, 2010). O treinamento resistido também promove queda pressórica em torno de 2 a 3 mmHg em adultos hipertensos (CORNELISSEN et al. 2011).

Com o avanço de técnicas estatísticas e computacionais, a modelagem de dados de pressão arterial tem sido feita usando abordagens mais flexíveis. O uso dos modelos aditivos generalizados (GAM) e dos modelos aditivos generalizados para localização, escala e forma (GAMLSS) é proeminente (MERKER, 2018; XI et al, 2016; FINUCANE et al, 2014; BARBA et al, 2014; YAN et al. 2013; LIU; TU, 2012). Esses últimos referem-se a uma ampla classe

de modelos que foi introduzida por Rigby e Stasinopoulos (2005). Entre as suas principais vantagens destacam-se a grande variedade na escolha da distribuição para a variável resposta e a facilidade de modelar outros parâmetros da distribuição. Apesar de as diversas aplicações dos modelos GAMLSS em várias áreas, até o momento não se encontraram na literatura trabalhos que avaliassem o efeito hipotensor de diferentes treinamentos físicos utilizando essa modelagem.

Este trabalho tem como objetivo comparar o efeito hipotensor de diferentes treinamentos físicos (hipertrofia e resistência), usando dados de um estudo realizado em adultos normotensos, de ambos os sexos, avaliados ao longo de 18 sessões. Para avaliar o efeito do tipo de exercício e de outras variáveis, propôs-se o uso de modelos mistos semiparamétricos para as respostas PAS e PAD. Para isso, utilizou-se o pacote *gamlss* do programa R (R Core Team, 2018). Os códigos utilizados nas análises estão disponíveis no Anexo A.

2. MATERIAL E MÉTODOS

2.1 Descrição dos grupos de estudo

Os dados utilizados neste trabalho foram obtidos por Barbosa (2013) em 2012. Os participantes da pesquisa foram policiais e não policiais (dependentes) voluntários do Programa de Controle e Combate à Obesidade, desenvolvido pelo 8º Batalhão da Polícia Militar de Minas Gerais, localizado na cidade de Lavras.

Para cada participante, foi entregue um termo de compromisso livre e esclarecido (TCLE) e explicados os objetivos e exigências do projeto. Posteriormente, foram aplicados dois questionários: questionário sobre prontidão para a atividade física - Par-q e questionário internacional de nível de atividade física - IPAQ versão curta. A segunda fase de avaliações consistiu

na realização de um teste progressivo em ciclo ergômetro para predição das cargas de treinamento.

O Par-q é um instrumento para identificar os indivíduos assintomáticos que não precisam de um exame clínico mais específico antes de se engajarem em programas de atividades físicas. O IPAQ versão curta foi utilizado para avaliar o nível de atividade física do indivíduo. Somente participaram da pesquisa os indivíduos que não apresentaram nenhum risco cardiovascular pelo questionário Par-q e nível sedentário de atividade física pelo IPAQ versão curta.

Foram avaliados 26 indivíduos normotensos de ambos os sexos, não praticantes de atividades físicas. Os participantes foram aleatoriamente divididos em dois grupos. Ao primeiro grupo (Grupo 1), composto de 10 mulheres e 5 homens, com idades entre 40 a 50 anos, foi prescrito um Treinamento de Resistência. Ao segundo grupo (Grupo 2), composto de 4 mulheres e 7 homens, com idades de 30 a 50 anos, foi prescrito um Treinamento de Hipertrofia.

O Treinamento de Resistência constituiu em um programa de exercício aeróbico com séries contínuas de 30 a 45 minutos associadas a um treinamento resistido de mesma duração, voltado para o desenvolvimento de resistência e força composto por 3 séries com 15 repetições e pausas de 30 segundos entre cada série.

O segundo grupo foi submetido a um programa de exercícios físicos aeróbicos intermitentes de 30 a 45 minutos e, ao longo de cada série, 2 minutos foram destinados para a fase ativa e 3 minutos para a fase de recuperação. Essa atividade também foi associada a um treinamento resistido de força objetivando hipertrofia com 3 séries compostas por 12 repetições e pausas de 45 a 90 segundos.

Em termos gerais, um exercício aeróbico consiste em atividades com intensidade moderada e com longa duração cujo objetivo principal é estimular toda a estrutura muscular. Por outro lado, um exercício resistido oferece uma determinada resistência muscular, proporcionada

por algum método proveniente da utilização de pesos, equipamentos ou até mesmo o peso do próprio corpo.

A Percepção Subjetiva do Esforço (PSE) foi o método utilizado, para fazer o monitoramento do treinamento, utilizando a Escala de Borg (6-20) que se baseia na frequência cardíaca. O PSE entre 12 e 14 desta escala sugere que a atividade física seja realizada em um nível de intensidade moderada e, acima dessa escala, em um nível intenso a exaustivo (BARBOSA, 2013). O protocolo de realização de ambos os treinamentos é resumido na Tabela 3.1.

Tabela 3.1 – Protocolos de realização dos treinamentos físicos.

	Treinamento de Resistência			Treinamento de Hipertrofia		
	Duração	Carga	Séries	Duração	Carga	Séries
Aeróbico	30-45 min	12-14 PSE	Contínua	30-45 min	14-17 PSE na fase ativa; 8-10 PSE na fase de recuperação	2 min na fase ativa; 3 min na fase de recuperação
Resistido	30-45min	60 a 70% carga estimada	3x15, com pausa 30 seg entre séries	30-45min	70 a 85 % carga estimada	3x12, com pausa de 45 a 90 seg

Fonte: BARBOSA (2013, p. 26)

2.2 Aferição da pressão arterial

A duração do período de treinamento consistiu de 6 semanas com frequência semanal de 3 vezes, totalizando 18 sessões de treinamento. Em ambos os grupos, cada sessão durava, em média, 150 minutos de treinamento concorrente. A aferição da PA foi obtida na *baseline*, ou seja, antes dos indivíduos serem submetidos aos treinamentos físicos e, durante as sessões, após 5 minutos de repouso e imediatamente ao término do programa de exercício físico. Em todas as ocasiões, a aferição da PA foi realizada com os indivíduos sentados e com o braço à altura do coração.

A PA dos participantes foi aferida de acordo com as VI Diretrizes Brasileiras de Hipertensão. Como a PA é um mecanismo que sofre influência direta e indireta de fatores extrínsecos, alguns cuidados foram tomados para que fosse minimizada esta influência. O aparelho utilizado para a verificação da PA foi o esfigmomanômetro aneroide da marca *WelchAllyn* devidamente calibrado.

2.3 Coleta de dados antropométricos

As medidas antropométricas foram tomadas no início e ao final do estudo. Os indivíduos foram pesados em uma balança eletrônica microdigital com capacidade de 150 kg e precisão de 100 g, em um mesmo horário pré-definido. A altura foi determinada, utilizando-se um estadiômetro (Personal Caprice, Brasil) vertical milimetrado, com escala de 0,5 cm. Em ambas as situações, os indivíduos estavam em pé, em posição firme, com os braços relaxados e cabeça no plano horizontal. A partir dos valores obtidos, foi calculado o índice de massa corporal (IMC).

O somatório de pregas cutâneas (bicipital, tricipital, subescapular e suprailíaca) foi utilizado para a aferição do percentual de gordura corporal total. Para tanto, foi utilizado um adipômetro (CESCORF Científico, Brasil) e obtida a média das 3 medidas. O lado direito do corpo foi padronizado para a aferição de todas as pregas cutâneas.

2.4 Análise estatística

A modelagem de pressão arterial foi feita usando GAMLSS, assumindo a distribuição normal para os dados de PAS e PAD. Inicialmente consideraram-se os efeitos das covariáveis sexo, idade, grupo, sessão de treinamento, interação grupo×sessão para a modelagem dos parâmetros de locação (μ) e escala (σ). Essas covariáveis foram, então, selecionadas pelo critério de informação de Akaike generalizado (GAIC). Os modelos foram ajustados e sele-

cionados usando o pacote *gamlss* (Rigby e Stasinopoulos, 2005) disponível do programa R (R Core Team, 2018).

Para os dados de PAS, a estrutura do modelo ajustado foi:

$$Y_{ijk} \sim N(\mu_{ij}, \sigma_k)$$

$$g_1(\mu_{ij}) = \mu_{ij} = \beta_{10} + s_{11}(\text{sessão}_j) + \gamma_{11}(id_i)$$

$$g_2(\sigma_k) = \log(\sigma_k) = \beta_{20} + \beta_{21} \text{grupo}_k,$$

em que Y_{ijk} é a resposta de PAS com $i = 1, 2, \dots, 26$, $j = 1, 2, \dots, 18$, e $k = 1, 2$. Para o efeito aleatório, assumiu-se que $id \sim N(0, \sigma_{id})$. A função suavizadora *P-spline* foi atribuída à covariável sessão por não apresentar efeito linear na resposta.

No modelo referente aos dados de PAD, foram detectados efeitos de covariáveis apenas no parâmetro de posição. Assim, foi selecionado como melhor modelo:

$$Y_{ijk} \sim N(\mu_{ijk}, \sigma)$$

$$g_1(\mu_{ijk}) = \mu_{ijk} = \beta_{10} + s_{11}(\text{sessão}_j) + \beta_{11} \text{grupo}_k + \beta_{12} \text{sessão} \times \text{grupo}_{jk} + \gamma_{11}(id_i)$$

$$g_2(\sigma) = \log(\sigma) = \beta_{20},$$

em que Y_{ijk} é a resposta de PAD com $i = 1, 2, \dots, 26$, $j = 1, 2, \dots, 18$, e $k = 1, 2$. A interação sessão \times grupo apresentou comportamento linear, portanto não foi necessário utilizar função suavizadora nesse efeito.

As funções disponíveis no pacote *gamlss* e utilizadas nas análises foram a *stepGAIC*, para a seleção das covariáveis, que foram inseridas na regressão de cada um dos parâmetros, a

getSmo() utilizada para a obtenção da estimativa do efeito aleatório que é dada em termos do desvio padrão e a função *pvc()* para estudar o comportamento da interação sessão \times grupo.

Os dados das diferenças (antes - após o período de treinamento), nas variáveis antropométricas, foram comparados entre os dois grupos, utilizando o teste de Wilcoxon-Mann-Whitney. Em todos os casos, os testes foram aplicados separadamente para cada sexo.

3. RESULTADOS E DISCUSSÃO

3.1 Análise exploratória

Algumas medidas descritivas, para os dados de PAS e PAD, estão apresentadas na Tabela 3.2, considerando todo o conjunto de dados e separados por grupos. Em ambas as respostas e ao considerar os dados de todos os indivíduos, nota-se que as distribuições apresentaram uma leve assimetria à direita (as médias são maiores do que as medianas).

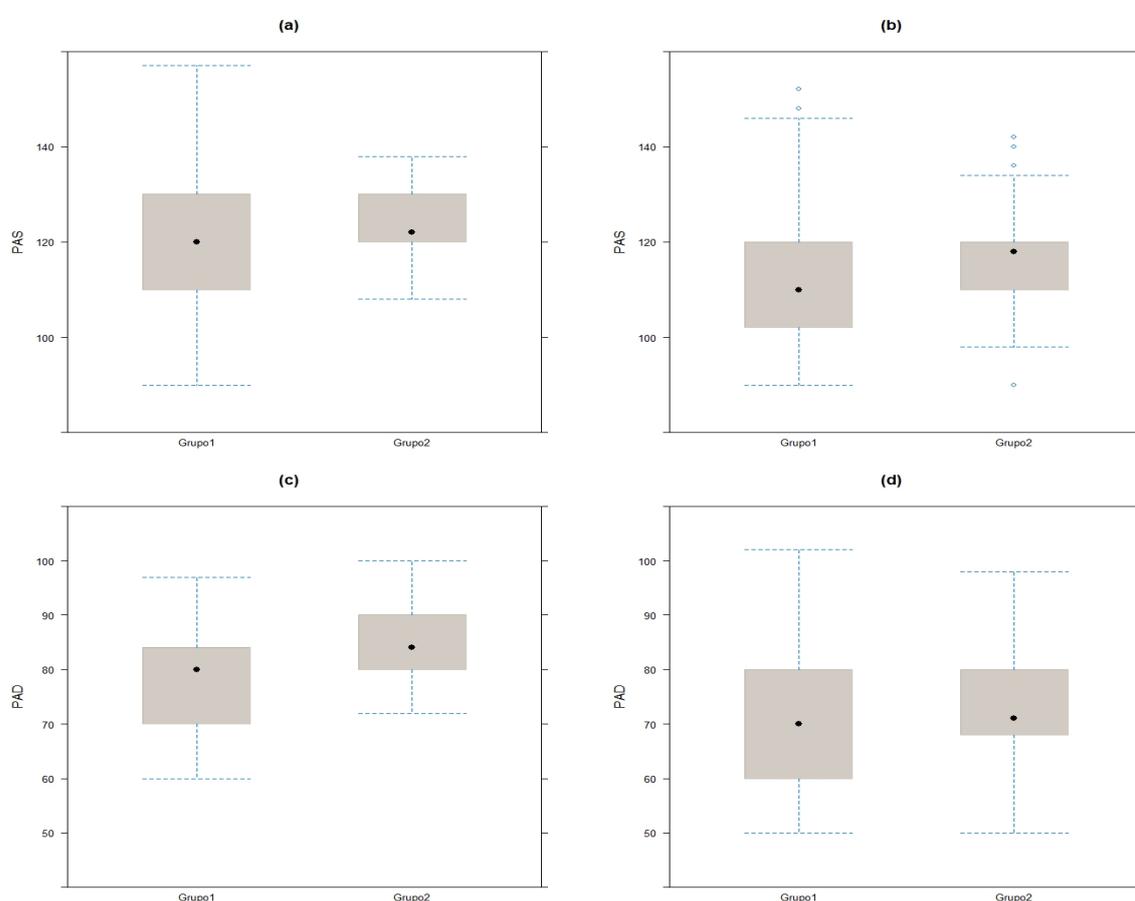
No entanto, quando essas mesmas medidas são obtidas, considerando os diferentes grupos de treinamento, observa-se que o Grupo 2 (treinamento de hipertrofia) apresentou assimetria à esquerda, para os dados da PAS e simetria para a os dados da PAD. Possivelmente essa mudança é decorrente da presença de valores extremos (Figura 3.1 (b)), que é mais frequente nesse grupo. No Grupo 1 (treinamento de resistência), a assimetria se manteve levemente à direita em ambas as respostas. Pelo coeficiente de variação (CV), observa-se maior variabilidade no Grupo 1.

Tabela 3.2 – Medidas descritivas das variáveis respostas PAS e PAD considerando todas as sessões.

Resposta	Categorias	Média	Mediana	DP	CV
PAS	Geral	114,49	110,00	11,74	10,25
	Grupo 1	113,31	110,00	12,76	11,26
	Grupo2	116,11	118,00	9,99	8,60
PAD	Geral	73,27	70,00	10,62	14,49
	Grupo 1	72,70	70,00	10,96	15,07
	Grupo2	74,05	74,00	10,11	13,65

Os boxplots da Figura 3.1 (a) e 3.1 (c) são referentes às medidas iniciais de PAS e PAD, respectivamente, que são obtidas antes dos indivíduos serem submetidos aos treinamentos físicos, ou seja, são as medidas na *baseline*. Os gráficos da Figura 3.1 (b) e 3.1 (d) apresentam as medidas dos participantes considerando todas as sessões. Observa-se uma aparente redução nos valores das medianas em ambos os gráficos, dando indícios de que ambos os treinamentos foram eficientes na redução das medidas da PAS e PAD.

Figura 3.1 – Gráficos boxplot para os Grupos 1 (Treinamento de Resistência) e 2 (Treinamento de Hipertrofia) para as variáveis respostas PAS na *baseline* (a), PAS considerando todas as sessões (b), PAD na *baseline* (c) e PAD considerando todas as sessões (d).

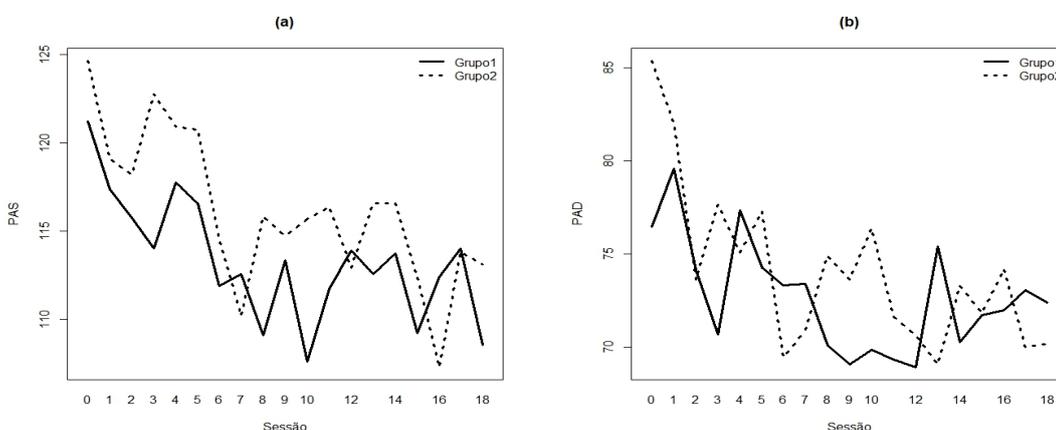


As médias iniciais e, considerando todas as sessões, também, foram calculadas. No início do estudo, as médias da PAS dos Grupos 1 e 2 foram 120,0 (mmHg) e 122,0 (mmHg), respectivamente. Ao longo das sessões, essas médias foram de 110,0 (mmHg) para o Grupo 1 e 118,0 (mmHg) para o Grupo 2. As médias de PAD dos respectivos grupos são 80,0 (mmHg)

e 84,0 (mmHg) na *baseline* e, considerando todas as sessões, corresponderam a 70,0 (mmHg) e 71,0 (mmHg) para os grupos 1 e 2, respectivamente.

Na Figura 3.2, têm-se o comportamento médio da PAS e PAD para os dois grupos, considerando as sessões treinamento. Pela Figura 3.2 (a), são observadas quedas mais acentuadas nas medidas da PAS, em ambos os grupos, já a partir da quarta sessão. Em média, a menor medida da PAS do Grupo 1 foi identificada na 10^a sessão e a do Grupo 2 ocorreu na 16^a sessão. Na Figura 3.2 (b), observa-se que entre a 8^a e 12^a sessão a PAD se manteve nos níveis mais baixos para o Grupo 1, enquanto, no Grupo 2, isso é identificado na 6^a e na 13^a sessão. De maneira geral, pode-se observar que houve oscilações da PAS e PAD, ao longo das sessões, cujo efeito se dá de forma não linear. Em ambos os gráficos, há evidências da redução dessas medidas ao longo do tempo.

Figura 3.2 – Gráficos de perfis médios dos Grupos 1 (Treinamento de Resistência) e 2 (Treinamento de Hipertrofia), para as variáveis respostas PAS (a) e PAD (b), ao longo das 18 sessões de treinamento.



3.2 Avaliação das medidas antropométricas nos grupos de treinamento

Medidas antropométricas dos participantes foram tomadas antes e após o período de treinamento. Na Tabela 3.3, são apresentadas as médias e erros padrão para as diferenças entre

as medidas iniciais e finais do índice de massa corporal (IMC), circunferência de cintura (CC) e porcentagem de gordura (PG) dos participantes, em ambos os treinamentos.

Tabela 3.3 – Média e erro padrão da média dos valores de diferença nas medidas antropométricas antes e após o treinamento. Os valores-p reportados referem-se ao teste de Wilcoxon-Mann-Whitney para comparar as diferenças entre os dois grupos de treinamento.

Sexo	IMC			CC			PG		
	\bar{X}_Δ	EP_Δ	Valor-p	\bar{X}_Δ	EP_Δ	Valor-p	\bar{X}_Δ	EP_Δ	Valor-p
F	2,382	0,850	0,198	9,095	2,728	0,278	3,016	0,553	0,075
M	1,921	0,201	0,073	4,192	0,663	0,007	3,336	0,274	0,518

Os valores positivos das médias indicam que as medidas antropométricas finais foram menores do que as iniciais em ambos os treinamentos físicos. O teste de Wilcoxon-Mann-Whitney foi aplicado aos dados das diferenças, para verificar se elas foram as mesmas, em ambos os grupos de treinamento. Pelo valor-p dos testes, houve diferença significativa entre os grupos em relação à redução da circunferência da cintura em homens. Aos demais, os exercícios físicos não diferiram significativamente.

3.3 Os efeitos dos treinamentos físicos na pressão arterial

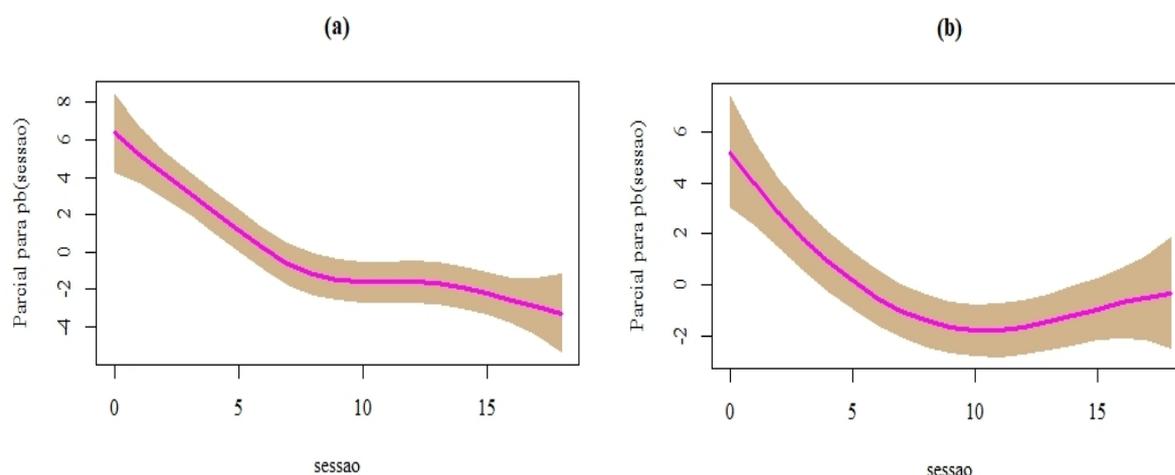
Na Tabela 3.4, apresentam-se as estimativas dos parâmetros, erros padrão e valor-p dos efeitos fixos dos modelos ajustados aos dados de PAS e PAD. Em relação à PAS e independente do grupo, observa-se sua redução ao longo das sessões de treinamento. Como a essa covariável foi atribuída uma função suavizadora, sua estimativa não é passível de interpretação direta e seu efeito é avaliado de forma gráfica na Figura 3.3 (a). Observa-se que, ao longo das sessões, a PAS decresce gradativamente, tornando-se, aproximadamente, estável entre a 10^a e 15^a sessão. A partir da 15^a sessão, seus valores médios decrescem novamente.

Ainda em relação à PAS, a covariável grupo de treinamento apresentou efeito significativo no parâmetro de escala. A estimativa para o Grupo 2 indica que sua variabilidade (em termos de desvio padrão) foi menor que a do Grupo 1 em 7,31 unidades ($e^{2,15-0,16}$). Essa vari-

Tabela 3.4 – Estimativas dos efeitos fixos dos modelos ajustados aos dados de PAS e PAD.

Resposta	Parâmetros	Efeito	Estimativa	Erro Padrão	Valor-p
PAS	μ	Intercepto	118,79	0,69	<0,01
		pb(sessão)	-	-	-
	σ	Intercepto	2,15	0,04	<0,01
		Grupo 2	-0,16	0,06	0,01
PAD	μ	Intercepto	75,07	0,88	<0,01
		Grupo 2	3,70	1,33	<0,01
		pb(sessão)	-	-	-
		Grupo2×sessão	-0,26	0,12	0,04
	σ	Intercepto	2,03	0,03	<0,01

Figura 3.3 – Gráficos do termo aditivo para os dados da PAS (a) e PAD (b) ao longo das sessões de treinamento.



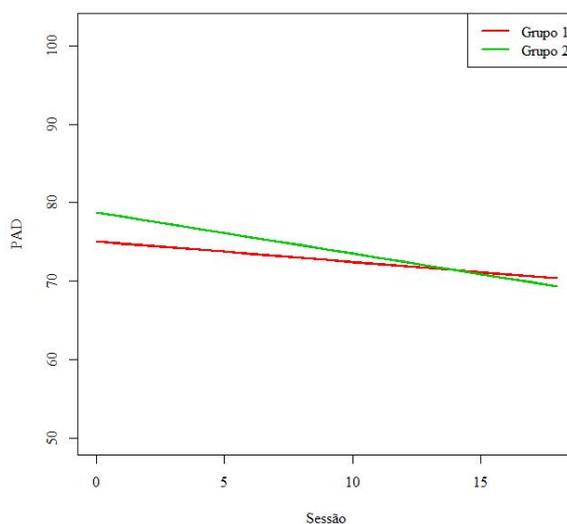
abilidade nos dados pode ser proveniente de outras características individuais não consideradas neste estudo. Por fim, o valor do desvio padrão do efeito aleatório é 7,79 mmHg para o modelo ajustado aos dados da PAS.

Para o modelo ajustado aos dados da PAD, a variável grupo e a interação sessão×grupo apresentaram efeitos significativos. A estimativa para o Grupo 2 indica que sua média é maior que a do Grupo 1 em 3,13 mmHg. No entanto, ao avaliar a interação, nota-se que a cada aumento de uma unidade da sessão há um decréscimo de 0,26 mmHg nas medidas do Grupo 2 em relação ao Grupo 1. Na Figura 3.4, representa-se a interação grupo×sessão e observa-se

que, ao longo das sessões de treinamento, o Grupo 2 obteve um efeito hipotensor maior em relação à PAD do que o Grupo 1.

Na Figura 3.3 (b), nota-se o efeito das sessões de treinamento, independente do grupo, no comportamento da PAD. Assim como para a PAS, há uma redução nos valores médios, que ocorre de maneira não linear, apresentando-se estável entre 10^a e 15^a sessão. O valor do desvio padrão do efeito aleatório é 6,61 mmHg para o modelo ajustado aos dados da PAD.

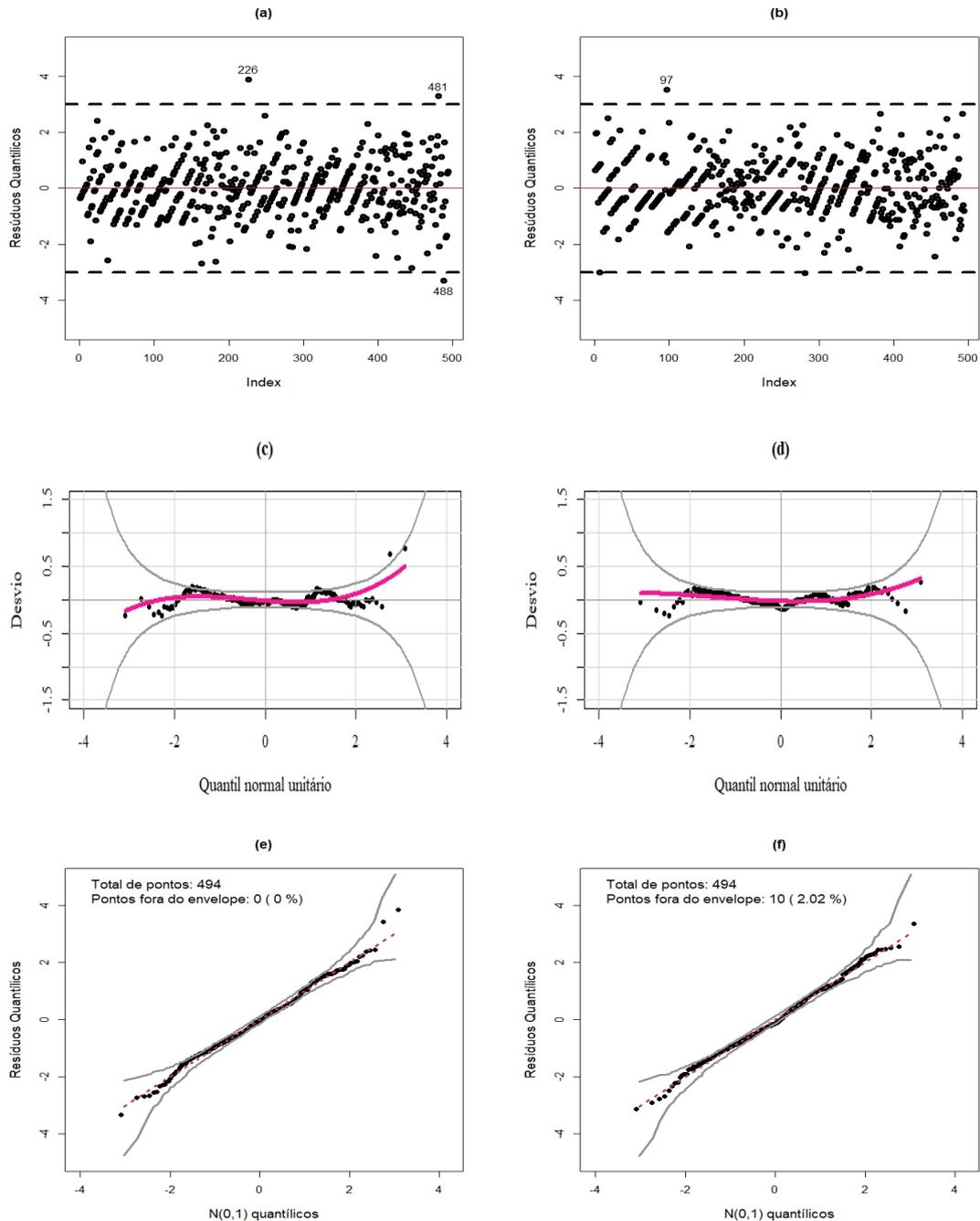
Figura 3.4 – Gráfico de valores preditos de PAD, para o Grupo 1 (Treinamento de Resistência) e para o Grupo 2 (Treinamento de Hipertrofia) ao longo das sessões de treinamento.



Uma análise dos resíduos dos modelos ajustados, para os dados de PAS e PAD, é apresentada na Figura 3.5. Ao avaliar os resíduos quantílicos (Figura 3.5 (a) e (b)), pode-se notar que, para ambos os modelos (PAS e PAD), eles se apresentaram de forma aleatória. Além disso, foram verificadas três observações fora do intervalo $[-3, 3]$ para os dados da PAS e apenas uma para os dados da PAD.

A Figura 3.5 (c) e 3.5 (d) refere-se aos gráficos *worm*, propostos no contexto dos modelos GAMLSS e indicam que a distribuição normal foi uma boa escolha para esses dados, já que nenhuma tendência (deslocamento vertical, inclinação) foi observada. Por fim, a qualidade de

Figura 3.5 – Gráficos de índices dos resíduos quantílicos dos modelos ajustados aos dados de PAS (a) e PAD (b), gráficos *worms* dos modelos para PAS (c) e PAD (d) e gráficos de probabilidade normal com envelope simulado dos modelos para PAS (e) e PAD (f).



ajuste dos modelos foi verificada pelos envelopes simulados (ATKINSON, 1982) apresentados nas Figuras 3.5 (e) e (f). Observa-se que a maioria dos resíduos está contida entre as faixas de confiança simuladas, indicando que os modelos são plausíveis e podem ser considerados para explicar os dados de PAS e PAD.

No ensaio longitudinal apresentado, buscou-se comparar dois tipos de treinamento quanto ao efeito hipotensor pós-exercício físico em indivíduos normotensos. Para isso, apresentou-se um procedimento semiparamétrico, para modelar dados da PA, utilizando a classe GAMLSS, que oferece uma alternativa prática e flexível a outras técnicas de modelagem estatística.

Neste trabalho, foi importante modelar o parâmetro de escala nos dados de PAS, pois detectou-se efeito significativo de grupo que não seria captado com o uso de uma modelagem mais simples. Silva (2017) relata que a pressão arterial é uma variável biológica que não é constante no decorrer do tempo, pois é influenciada por fatores emocionais e atividades cotidianas. Sendo assim, considera-se que seja importante ajustar modelos capazes de investigar e/ou captar as influências dessa variabilidade na PA.

A modelagem do parâmetro de escala começou na década 70 com Harvey (1976) e Aitkin (1987) que foram os primeiros a modelar a variância da distribuição normal. O parâmetro de dispersão, nos modelos lineares generalizados, foi modelado por Nelder e Pregibon (1987), Smyth (1989) e Verbyla (1993).

Rigby e Stasinopoulos (1996) introduziram as funções suavizadoras, para modelar os parâmetros μ e σ , metodologia chamada de modelos aditivos de média e dispersão (MADAM). Na formulação original da MADAM, a distribuição da resposta tinha que pertencer à família exponencial e o método de estimação usado era o da quase-verossimilhança. O GAMLSS é visto como uma classe mais geral desses modelos, em que é usado o método da máxima verossimilhança e que possibilita a modelagem de todos os parâmetros da distribuição, não só de locação e escala.

Assumiu-se a distribuição normal para ambas as respostas e o ajuste foi satisfatório. Em trabalhos dessa natureza, muitas outras distribuições podem ser consideradas e outros parâme-

tros podem ser modelados. Para isso, podem ser usados termos paramétricos e/ou funções de suavização.

Os pontos extremos identificados nas análises são esperados em estudos desse tipo. Parte dessa variabilidade também pode ser provocada por fatores genéticos ou ambientais que não foram medidos, sendo assim, uma alta heterogeneidade entre e dentro os indivíduos é natural. De acordo com a V Diretrizes Brasileiras de Monitorização Ambulatorial da Pressão Arterial (MAPA), a alta variabilidade da pressão arterial de um mesmo indivíduo pode ter correlação direta com eventos cardiovasculares e o desenvolvimento de lesão em órgãos-alvo. Porém, dispõe que é necessário investigar essa variabilidade de maneira apropriada, o que deve ser feito por meio do registro contínuo da pressão arterial (batimento a batimento), o que não é obtido pelo método usual de registro da PA.

Em ambos os modelos, foi possível verificar o efeito hipotensor dos exercícios físicos ao longo das sessões. Em relação à PAD, destaca-se que sua redução foi maior no grupo que recebeu o treinamento de hipertrofia. Este estudo concorda com outras pesquisas que mostram os benefícios desses exercícios combinados (aeróbico + resistido) na redução da pressão arterial (PIRES et al, 2018; SILVA, 2017; AZEVEDO et al, 2017; SANTOS et al, 2016).

4. CONCLUSÃO

Modelos semiparamétricos pertencentes à classe GAMLSS foram ajustados a dados de PAS e PAD e se mostraram eficientes no estudo do efeito hipotensor de diferentes treinamentos físicos. O efeito de treinamento, no parâmetro de escala da distribuição, foi detectado graças à flexibilidade dessa metodologia e, possivelmente, não seria captado com o uso de modelos mais simples. Além disso, por se tratar de um estudo com medidas repetidas no tempo, o efeito aleatório de indivíduos pôde ser devidamente inserido nos modelos.

Do ponto de vista prático, pelas análises apresentadas, constatou-se que o exercício aeróbio intermitente, juntamente com o exercício resistido com cargas mais elevadas (treinamento de hipertrofia), resultou em um efeito hipotensor mais acentuado do que os exercícios aeróbios contínuos associados com treinamento resistido, em um volume maior de repetições (treinamento de resistência).

No entanto, o treinamento de resistência também pode ser usado como uma ferramenta não farmacológica por provocar adaptações da PA, diminuindo os níveis tensionais pré-treinamento. De maneira geral, do ponto de vista da Medicina e da Saúde Pública, a prática de atividade física é um método preventivo de doença e proporciona melhor qualidade de vida a uma população.

REFERÊNCIAS

AITKIN, M. Modelling variance heterogeneity in normal regression using GLIM. **Applied Statistics**, Abingdon, v. 36, p. 332-339, 1987.

ATKINSON, A. C. The simulation of generalized inverse Gaussian and hyperbolic random variables. **Journal on Scientific and Statistical Computing**, Philadelphia, v. 3, p. 502-515, 1982.

AZEVEDO, L. M. et al. Fractionated concurrent exercise throughout the day does not promote acute blood pressure benefits in hypertensive middle-aged women. **Frontiers in Cardiovascular Medicine**, Lausanne, v. 4, n. 6, p. 1-7, 2017.

BARBA, G. et al. Blood pressure reference values for European non-overweight school children: the IDEFICS study. **International Journal of Obesity**, London, v. 38, p. 548-556, 2014.

BARBOSA, C. R. J. **Influências do treinamento aeróbico e resistido na pressão arterial de indivíduos normotensos com sobrepeso e obesidade**. 2013. Trabalho de Conclusão de Curso (Graduação em Educação Física)-Universidade Federal de Lavras, Lavras, 2013.

CORNELISSEN, V. A. et al. Impact of resistance training on blood pressure and other cardiovascular risk factors: a meta-analysis of randomized, controlled trials. **Hypertension**, Dallas, v. 58, n. 5, p. 950-958, 2011.

CUNHA, E. S. et al. Intensidades de treinamento resistido e pressão arterial de idosas hipertensas: um estudo piloto. **Revista Brasileira de Medicina do Esporte**, São Paulo, v. 18, n. 6, p. 373-376, 2012.

DAMORIM, I. R. et al. Cinética hipotensiva durante 50 sessões de treinamento de força e aeróbio em hipertensos: ensaio clínico randomizado. **Sociedade Brasileira de Cardiologia**, São Paulo, v. 108, n. 4, p. 323-330, 2017.

DIMEO, F. et al. Aerobic exercise reduces blood pressure in resistant hypertension. **Hypertension**, Dallas, v. 60, n. 3, p. 653-658, 2012.

FINUCANE, C. et al. Age-related normative changes in phasic orthostatic blood pressure in a large population study findings from the irish longitudinal study on ageing (TILDA). **Hypertension**, Dallas, v. 130, n. 20, p. 1780-1789, 2014.

HARVEY, A. C. Estimating regression models with multiplicative heteroscedasticity. **Econometrica**, Chicago, v. 44, n. 3, p. 461-465, 1976.

LANDAL, A. C. et al. Fatores associados à melhora da composição corporal em indivíduos com DPOC após treinamento físico. **Fisioterapia em Movimento**, Curitiba, v. 27, n. 4, p. 633-641, out./dez. 2014.

LIU, H.; TU, W. A. Semiparametric regression model for paired longitudinal outcomes with application in childhood blood pressure development. **The Annals of Applied Statistics**, Cleveland, v. 6, n. 4, p. 1861-1882, 2012.

MEDINA, F. L. Atividade física: impacto sobre a pressão arterial. **Revista Brasileira de Hipertensão**, São Paulo, v. 17, n. 2, p. 103-106, 2010.

MERKER, A. et al. Growth in achondroplasia: Development of height, weight, head circumference, and body mass index in a European cohort. **American Journal of Medical Genetics. Part A**, Hoboken, v. 176, n. 8, p. 1723-1734, 2018.

MIRANDA, L. S. et al. Physical activity and risk of coronary heart disease and stroke in older adults: the cardiovascular health study. **Circulation**, Baltimore, v. 133, n. 2, p. 147-155, 2016.

NEGRÃO, C. E.; RONDON, M. U. P. B. Exercício físico, hipertensão e controle barorreflexo da pressão arterial. . **Revista Brasileira de Hipertensão**, São Paulo, v. 8, n. 1, p. 89-95, 2001.

NELDER, J. A.; PREGIBON D. An extended quasi-likelihood function. **Biometrika**, London, v. 74, n. 2, p. 221-232, 1987.

PRATA, N. S. et al. Associação entre nível de exercício físico e alteração dos valores de pressão arterial em idosos da cidade de Valença/RJ. **Saber Digital**, Valença, v. 10, n. 2, p. 1-10, 2018.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2018.

RADOVANOVIC, C. A. T. et al. Hipertensão arterial e outros fatores de risco associados às doenças cardiovasculares em adultos. **Revista Latino-Americana de Enfermagem**, Ribeirão Preto, v. 22, n. 4, p. 547-553, jul./ago. 2014.

RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. **Journal of Statistical Software**, Los Angeles, v. 54, n. 3, p. 507-554, 2005.

RIGBY, R. A.; STASINOPOULOS, D. M. A semi-parametric additive model for variance heterogeneity. **Statistics and Computing**, London, v. 6, p. 57-65, 1996.

SATO, K. et al. Association of physical activity with a systolic blood pressure difference between arms in older people. **Geriatrics Society**, Baltimore, v. 18, n. 1, p. 95-100, Jan. 2018.

SANTOS, L. P. et al. Effects of aerobic exercise intensity on ambulatory blood pressure and vascular responses in resistant hypertension: a crossover trial. **Journal of Hypertension**, London, v. 34, n. 7, p. 1317-1324, 2016.

SILVA, G. V. Blood pressure variability: another puzzle piece. **Jornal Brasileiro de Nefrologia**, São Paulo, v. 39, n. 2, p. 102-103, 2017.

SMYTH, G. K. Generalized linear models with varying dispersion. **Journal of the Royal Statistical Society**. Series B, London, v. 51, n. 1, p. 47-60, 1989.

SOCIEDADE BRASILEIRA DE CARDIOLOGIA.V Diretrizes Brasileiras de Monitorização Ambulatorial da Pressão Arterial (MAPA). **Arquivos Brasileiros de Cardiologia**, São Paulo, v. 97, n. 3, p. 9-10, 2011.

SOCIEDADE BRASILEIRA DE CARDIOLOGIA; SOCIEDADE BRASILEIRA DE HIPERTENSÃO; SOCIEDADE BRASILEIRA DE NEFROLOGIA. VI Diretrizes Brasileiras de Hipertensão. **Arquivos Brasileiros de Cardiologia**, São Paulo, v. 95, p. 1-51, 2010.

VERBYLA, A. P. Modelling variance heterogeneity: residual maximum likelihood and diagnostics. **Journal of the Royal Statistical Society**. Series B, London, v. 55, n. 2, p. 49-508, 1993.

WHELTON, S. P. et al. Effect of aerobic exercise on blood pressure: a meta-analysis of randomized, controlled trials. **Annals of Internal Medicine**, Philadelphia, v. 136, n. 7, p. 493-503, 2002.

WORLD HEALTH ORGANIZATION. **A global brief on hypertension**. Geneva, 2013.

XI, B. et al. Establishing international blood pressure references among nonoverweight children and adolescents aged 6 to 17 years. **Hypertension**, Dallas, v. 133, n. 4, p. 398-408, 2016.

YAN, W. et al. Blood pressure percentiles by age and height for non-overweight Chinese children and adolescents: analysis of the china health and nutrition surveys. **BMC Pediatrics**, London, v. 13, n. 1, p. 1991-2009, 2013.

APÊNDICES

APÊNDICE A - Código em R: Análise da pressão arterial

```
1 rm(list=ls())
2 dados <- read.table("pressao.txt", h=T)
3 attach(dados)
4 str(dados)
5 grup <- factor(dados$grupo)
6 sex <- factor(dados$sexo)
7 dados <- transform(dados, id=factor(id))
8 require(gamlss)
9 require(lattice)
10 require(fBasics)
11 # Analise descritiva dos dados
12 basicStats(dados$PAS)
13 basicStats(dados$PAD)
14 basicStats(dados$PAS{[]grup==1{}})
15 basicStats(dados$PAS{[]grup==2{}})
16 basicStats(dados$PAD{[]grup==1{}})
17 basicStats(dados$PAD{[]grup==2{}})
18
19 #Grafico boxplot das medidas na baseline e ao longo das sessoes
20
21 par(family = "serif")
22 bwplot(dados$base0S ~ grup, data=dados, main="(a)", ylab="PAS",
23 col="red", par.settings=list(box.rectangle=list (col="salmon",
```

```
24 fill="salmon", alpha=0.4)))
25 bwplot(dados\[extract_itex]PAS ~ grup, data=dados, main="(b)", ylab="PAS",
26 col="red", par.settings=list(box.rectangle=list(col="salmon",
27 fill="salmon", alpha=0.4)))
28
29
30 #Grafico de perfis medios
31
32 # Perfis medios
33 par(mar=c(4, 4, 0, 0)+.1)
34 interaction.plot(sessao, grup, dados1[/extract_itex]PAS, col=1, lty=c(1, 3),
35 lwd=3, xlab="Sessao", ylab="PAS", legend =FALSE)
36 legend("topright", c("Grupo1", "Grupo2"), xpd=TRUE, horiz=FALSE,
37 inset=c(0, 0), bty="n", cex=1.0, lty=c(1, 3), col=1, lwd=3)
38 par(mar=c(4, 4, 0, 0)+.1)
39 interaction.plot(sessao, grup, dados1[/extract_itex]PAD, col=1, lty=c(1, 3),
40 lwd=3, xlab="Sessao", ylab="PAD", legend =FALSE)
41 legend("topright", c("Grupo1", "Grupo2"), xpd=TRUE, horiz=FALSE,
42 inset=c(0, 0), bty="n", cex=1.0, lty=c(1, 3), col=1, lwd=3)
43
44 # Medidas antropometricas
45
46 dados<- read.table("diferenca.txt", h=T)
47 attach(dados)
48 str(dados)
49 grup <- factor(dados[/extract_itex]grupo)
```

```
50 sex <- factor(dados$sexo)
51 datos<- transform(dados, id=factor(id))
52 require(gamlss)
53 dadosf<-subset(dados, sexo=="F")
54 View(dadosf)
55 wilcox.test(IMC~grupo, data=dadosf)
56 wilcox.test(CC~grupo, data=dadosf)
57 wilcox.test(PG~grupo, data=dadosf)
58
59 dadosm<-subset(dados, sexo=="M")
60 wilcox.test(IMC~grupo, data=dadosm)
61 wilcox.test(CC~grupo, data=dadosm)
62 wilcox.test(PG~grupo, data=dadosm)
63
64 mean(dadosf$IMC)
65 mean(dadosf$CC)
66 mean(dadosf$PG)
67 mean(dadosm$IMC)
68 mean(dadosm$CC)
69 mean(dadosm$PG)
70
71 se <- sqrt(var(dadosf$IMC)/length(dadosf)); se
72 se <- sqrt(var(dadosf$CC)/length(dadosf)); se
73 se <- sqrt(var(dadosf$PG)/length(dadosf)); se
74 se <- sqrt(var(dadosm$IMC)/length(dadosm)); se
75 se <- sqrt(var(dadosm$CC)/length(dadosm)); se
```

```
76 se <- sqrt(var(dadosm$PG)/length(dadosm));se
77
78 # Modelagem Estatística
79 #Modelo final para os dados da PAS.
80 m1 <- gamlss(PAS ~ {re(random=} 1|id) + pb(sessao),
81 sigma.fo=~ grup, family= NO ,data=dados1, trace=FALSE)
82 summary(m1)
83 wp(m1)
84
85 # Modelo completo (considerando todas as covariáveis)
86 m.1<-gamlss(PAD ~ pb(sessao)+grup*sessao+sex+idade+re(random=~ 1|id),
87 data=dados1, trace=FALSE)
88 mod1<-stepGAIC(m.1, trace=FALSE)
89 mod1.1 <- stepGAIC(m.1, parameter="sigma",
90 scope=~ pb(sessao) + grup*sessao + sexo + idade +
91 re(random= ~ 1|id), k=log(41), trace=FALSE)
92
93 # Modelo final para os dados da PAD
94
95 m2<- gamlss(PAD~ pb(sessao)+grup*sessao+re(random= ~ 1 | id),
96 data = dados1, trace=FALSE)
97 #Funcao pvc()
98 interacao <- gamlss(PAD~ grup+pb(sessao)+re(random= ~ 1 | id)
99 +pvc(sessao,by=grup),data = dados1, trace=FALSE)
100 term.plot(interacao, pages=1)
101 summary(m2)
```

```
102 wp (m2)
103
104 #Grafico de interacao
105
106 PAD1 <- gamlss(PAD~(sessao), data=subset(dados,
107                                     grupo=="1"))
108 PAD2 <- gamlss(PAD~(sessao), data=subset(dados,
109                                     grupo=="2"))
110
111 x11()
112 par(family = "serif")
113 with(dados, plot(sessao, PAD, pch=(20) [unclass(grupo)], xlab="Sessao",
114                col=c("white", "black") [unclass(grupo)]), legend =
115                FALSE)
116 with(dados, lines(sessao[grupo=="1"], fitted(gasB), col="red",
117                lwd=2))
118 with(dados, lines(sessao[grupo=="2"], fitted(gasA), col="green3",
119                lwd=2))
120 legend("topright", lwd=2, col=c("red", "green3"), c("Grupo 1", "Grupo 2"))
```