



UFSC



## Statistics

# Análise de dados longitudinais sobre desenvolvimento infantil por meio de modelos lineares mistos trivariados derivados de cópulas

Analysis of longitudinal data on child development using trivariate linear mixed models derived from copulas

Alexandro Oliveira <sup>1</sup>, Ana Claudia Batista <sup>1</sup>, Marisleane Oliveira <sup>1</sup>,  
Paulo Henrique Ferreira <sup>2</sup>, Rosemeire Fiaccone <sup>2</sup>

<sup>1</sup>Programa de Pós-Graduação em Matemática, Universidade Federal da Bahia, BA, Brazil

<sup>2</sup>Departamento de Estatística, Universidade Federal da Bahia, BA, Brazil

## RESUMO

Estudos longitudinais são bastante comuns na área de saúde pública e, consequentemente, métodos estatísticos adequados são requeridos para analisar a evolução temporal de uma ou mais variáveis respostas, separada ou simultaneamente. Especificar a função de densidade conjunta de todas as variáveis respostas, assim como sua estrutura de correlação, são os principais obstáculos dos procedimentos de modelagem multivariada. Vale destacar também as dificuldades numéricas frequentemente encontradas na inferência estatística quando a dimensão do problema aumenta. Como alternativa, neste trabalho são apresentadas duas propostas para lidar com dados longitudinais multivariados: (i) uma abordagem univariada, com modelos lineares mistos ajustados a cada uma das variáveis respostas separadamente; e (ii) uma modelagem conjunta dessas variáveis, por meio do uso de funções cópula. Ambas as metodologias são aplicadas a um conjunto de dados reais trivariados referentes ao desenvolvimento nutricional de crianças de um município brasileiro do estado da Bahia.

**Palavras-chave:** Cópulas; Dados longitudinais; Índices antropométricos; Modelos lineares mistos

## ABSTRACT

Longitudinal studies are quite common in the area of public health and, consequently, adequate statistical methods are required to analyze the temporal evolution of one or more response variables, separately or simultaneously. Specifying the joint density function of all response variables, as well as their correlation structure, are the main obstacles of multivariate modeling procedures. It is also important to highlight the numerical difficulties often encountered in statistical inference when the



response dimension increases. As an alternative, this work presents two proposals to deal with multivariate longitudinal data: (i) a univariate approach, with linear mixed models fitted to each of the response variables separately; and (ii) a joint modeling of these variables, through the use of copula functions. Both methodologies are applied to a set of real trivariate data referring to the nutritional development of children in a Brazilian municipality in the state of Bahia.

**Keywords:** Copulas; Longitudinal data; Children's anthropometric indices; Linear mixed models

## 1 INTRODUÇÃO

Estudos longitudinais são aqueles que envolvem medições repetidas da mesma unidade de análise em alguma escala ordenada de tempo em que não é possível aleatorizar os instantes. O interesse especial nos referidos estudos é avaliar mudanças globais ou individuais ao longo do tempo. Esse tipo de delineamento tem sido utilizado em diversas áreas, como bioestatística, sociologia, economia e administração, entre outras (Zhou and Chen, 2021). Um dos desafios dos estudos longitudinais é a escolha da estrutura de dependência entre as observações repetidas realizadas na mesma unidade de investigação.

Os modelos lineares mistos (MLMs) são extensões dos modelos lineares e bastante utilizados na análise de dados longitudinais (Liu, 2016). Tal classe de modelos apresenta tanto fatores de efeitos fixos como aleatórios. Condicionados aos efeitos aleatórios, tais modelos assumem que as medidas repetidas para cada unidade de análise são independentes e a correlação intraunidades surge do compartilhamento dos fatores não mensurados. Ou seja, a correlação é modelada explicitamente na equação do modelo, através da heterogeneidade dos coeficientes de regressão entre as unidades de análise por meio da inclusão dos efeitos aleatórios. Desta forma, uma estrutura de dependência apropriada é essencial para que as inferências sobre as médias sejam válidas. Dentre os vários trabalhos que surgiram na literatura voltados para a análise de dados longitudinais, vale ressaltar o modelo de regressão linear em dois estágios proposto por Laird and Ware (1982), em que as distribuições de probabilidade para os vetores de respostas das diferentes unidades de investigação pertencem a uma única família, mas alguns parâmetros de efeitos aleatórios variam através de indivíduos com uma distribuição especificada para o segundo estágio. Para mais detalhes, veja Laird and Ware (1982), Pinheiro (2000), Brunner et al. (2002), Molenberghs (2005), Demidenko (2013), Brown (2015), entre outros.

Nos últimos anos, tem-se observado uma notável evolução na metodologia estatística para a análise de dados longitudinais, possibilitando aos estatísticos e investigadores empíricos o acesso a um conjunto de ferramentas cada vez mais sofisticadas. Como alternativa, as funções cópula vêm sendo empregadas na modelagem multivariada da estrutura de dependência no contexto longitudinal. Por serem flexíveis, as cópulas podem ser utilizadas mesmo quando há violação da suposição de normalidade multivariada dos dados (ver, por exemplo, Lambert, 2002; Sun et al., 2008), além de capturar a dependência não linear e de cauda entre as variáveis. Para mais detalhes sobre cópulas, veja Nelsen (2006), Joe (2014), Hofert et al. (2018), entre outros.

Objetivando caracterizar a evolução temporal de uma ou mais variáveis respostas, dois procedimentos são empregados neste trabalho: uma abordagem univariada, na qual são considerados MLMs para cada variável resposta separadamente; e, em seguida, uma abordagem multivariada baseada em cópulas para caracterizar as relações entre as variáveis respostas conjuntamente. A abordagem longitudinal trivariada considerada neste trabalho (dados reais sobre nutrição infantil), por meio de técnicas estatísticas apropriadas, como as citadas anteriormente e descritas em detalhes nas seções subsequentes, reveste-se como uma importante estratégia para avaliar o perfil antropométrico infantil nos primeiros meses de vida por meio de indicadores antropométricos.

Vale ressaltar que este estudo é uma extensão do trabalho desenvolvido em Ferreira et al. (2019). Porém, diferentemente deste último, que considera somente duas variáveis respostas (peso e estatura das crianças), no presente trabalho são utilizadas três variáveis dependentes (indicadores antropométricos 1, 2 e 3 do estado nutricional das crianças, definidos na subseção 4.1). São também consideradas cópulas elípticas assimétricas (Skew-normal e Skew-t), ainda pouco exploradas na literatura da área. O artigo está organizado da seguinte forma. Na seção 2 são apresentadas as abordagens univariada (subseção 2.1) e multivariada (subseção 2.2) utilizadas neste estudo. A seção 3 diz respeito à inferência estatística das duas abordagens metodológicas. A seção 4 fornece uma aplicação para um conjunto de dados reais (estudo antropométrico). Finalmente, a seção 5 conclui o artigo com algumas considerações finais e direções para trabalhos futuros.

## 2 MÉTODOS

Nesta seção são descritas algumas técnicas estatísticas que podem ser utilizadas na análise de dados longitudinais multivariados (trivariados, aqui). Primeiramente, são apresentados os MLMs (finais) assumidos para cada uma das três variáveis respostas (índices antropométricos 1, 2 e 3) separadamente, isto é, sem levar em consideração uma possível associação (correlação) entre elas (subseção 2.1). Depois, são utilizados os resultados de tais ajustes de modo que, mediante o uso de funções cópula, possa-se relacionar conjuntamente as distribuições marginais em cada tempo, além de descrever a estrutura de dependência entre elas (subseção 2.2).

### 2.1 Abordagem Univariada - MLMs

Modelos que possuem parâmetros (ou coeficientes) de efeitos fixos e aleatórios são denominados modelos mistos, ou ainda, como também são chamados: modelos de efeitos mistos, modelos de efeitos aleatórios de dois estágios, modelos multiníveis e modelos hierárquicos. Todo modelo estatístico linear que possui mais de um efeito aleatório (que não somente o termo representando o erro), além dos efeitos fixos, é chamado de modelo linear de efeitos mistos, ou simplesmente, modelo linear misto (MLM). Pode-se expressar o MLM, para cada variável resposta separadamente, da seguinte forma:

$$Y_{jit} = \mathbf{X}_{jit}^\top \boldsymbol{\beta}_j + \mathbf{Z}_{jit}^\top \boldsymbol{\omega}_{ji} + \epsilon_{jit}, \quad \text{para } j = 1, 2, \dots, d, \quad i = 1, 2, \dots, m \quad \text{e} \quad t = 1, 2, \dots, n_{ji} \quad (1)$$

em que  $Y_{jit}$  é a  $j$ -ésima variável resposta para o  $i$ -ésimo indivíduo no tempo  $t$ ,  $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jp_j-1})^\top$  é o correspondente vetor de parâmetros (desconhecidos) de efeitos fixos,  $\mathbf{X}_{jit} = (X_{jit}^{(0)}, X_{jit}^{(1)}, \dots, X_{jit}^{(p_j-1)})^\top$  é o vetor de  $p_j$  covariáveis (conhecidas),  $\boldsymbol{\omega}_{ji} = (\omega_{ji0}, \omega_{ji1}, \dots, \omega_{jiq_j-1})^\top$  é o vetor de parâmetros (desconhecidos) de efeitos aleatórios,  $\mathbf{Z}_{jit} = (Z_{jit}^{(0)}, Z_{jit}^{(1)}, \dots, Z_{jit}^{(q_j-1)})^\top$  é o vetor de  $q_j$  covariáveis (conhecidas) que caracterizam a variação aleatória na  $j$ -ésima resposta de cada unidade e, finalmente,  $\epsilon_{ji} = (\epsilon_{ji1}, \epsilon_{ji2}, \dots, \epsilon_{jin_{ji}})^\top$  é o vetor de erros aleatórios (não observáveis) que descrevem a variação devido a outras fontes, como por exemplo, variações dentro da própria unidade e erros de medição. Neste trabalho,  $d = 3$  indicadores antropométricos (variáveis dependentes),  $m = 76$  crianças e  $n_{ji} = n_i = n = 6$  instantes de tempo diferentes.

Geralmente, assume-se que, além de independentes entre si,  $\omega_{ji}$  e  $\epsilon_{ji}$  são normalmente distribuídos, isto é,  $\omega_{ji} \sim N_{q_j}(\mathbf{0}_j, \Sigma_j)$  e  $\epsilon_{ji} \sim N_{n_{ji}}(\mathbf{0}_{ji}, \mathbf{R}_{ji})$ , em que  $\mathbf{0}_j$  ( $\mathbf{0}_{ji}$ ) é o vetor nulo de dimensão  $q_j$  ( $n_{ji}$ ),  $\Sigma_j$  é a matriz de variância-covariância dos efeitos aleatórios e  $\mathbf{R}_{ji}$  é a matriz de variância-covariância residual. Contudo, apesar de bastante comum, a condição/suposição de normalidade não é necessária. Inclusive, diversos trabalhos da literatura recente atribuíram distribuições assimétricas aos efeitos e/ou erros aleatórios em MLMs. Por exemplo, Arellano-Valle et al. (2005) definiram um MLM Skew-normal que assume que os efeitos aleatórios e os erros do modelo seguem distribuição Skew-normal, que inclui a normalidade como um caso especial e proporciona flexibilidade na captura de uma ampla gama de comportamentos não normais. Kahrari et al. (2019) introduziram uma classe flexível de MLMs, assumindo que os efeitos aleatórios e os erros do modelo seguem distribuição Skew-normal-Cauchy multivariada. Ferreira et al. (2022) desenvolveram uma nova família de MLMs, em que o efeito aleatório segue uma distribuição mistura de escala Skew-normal e o termo de erro segue uma distribuição mistura de escala normal multivariada, com ênfase nas distribuições Skew t-Student normal, Skew-slash e Skew-normal contaminada.

Os casos mais simples de MLMs são os modelos com intercepto aleatório, em que existe um único efeito aleatório (além do erro experimental ou de medida) que é normalmente distribuído. Assim, no contexto deste estudo, isto é, para os dados longitudinais trivariados sobre crescimento infantil, os modelos de intercepto aleatório para as variáveis respostas *índice 1* (peso/idade), *índice 2* (peso/estatura) e *índice 3* (estatura/idade), podem ser expressos como segue ( $p_j = p = 5$  e  $q_j = q = 1$ ):

$$Y_{jit} = \beta_{j0} + \beta_{j1}t + \beta_{j2}\text{Sexo}_i + \beta_{j3}\text{Peso}_i + \beta_{j4}\text{Hemoglobina}_{it} + \beta_{j5}\text{Comprimento}_i + \omega_{ji} + \epsilon_{jit} \quad (2)$$

em que  $Y_{jit}$  é o índice  $j$  da criança  $i$  no instante  $t$ ,  $\beta_{j0}$  é o intercepto fixo,  $\beta_{j1}$  é o efeito do tempo (em meses) de medição das crianças,  $\beta_{j2}$  é o efeito do sexo (1 - masculino, 2 - feminino),  $\beta_{j3}$  é o efeito do peso (em gramas, g) ao nascer,  $\beta_{j4}$  é o efeito da hemoglobina (em gramas por decilitro, g/dL),  $\beta_{j5}$  é o efeito do comprimento (em centímetros, cm) ao nascer,  $\omega_{ji} \sim N(0, \sigma_{\omega_j}^2)$  é o intercepto aleatório e  $\epsilon_{jit}$  é o erro aleatório não observável.

Para a escolha da melhor estrutura de covariância, levou-se em conta a construção do semivariograma (Diggle et al., 2002) e os melhores ajustes dos modelos (segundo o Critério de Informação de Akaike, ou AIC; Akaike, 1977). Sendo assim, neste trabalho foi assumida a estrutura autoregressiva de ordem 1, em que a  $(k, l)$ -ésima entrada da matriz  $\mathbf{R}_{ji}$  é definida como  $\sigma_j^2 \rho_j^{|k-l|}$ , para algum  $\rho_j \in [-1, 1]$ . Segundo Arnau et al. (2012), o AIC é um indicador mais eficaz da qualidade do ajuste do que o critério de informação Bayesiano (BIC; Schwarz, 1978). Entretanto, os autores observaram, por meio de um estudo de simulação, que o AIC nem sempre selecionava a estrutura verdadeira, visto que outras estruturas poderiam fornecer aproximações mais adequadas. Nesse sentido, Vaida (2005) propuseram o AIC condicional, ou CAIC, como uma alternativa eficaz e útil para a seleção de modelos de efeitos mistos, argumentando que fornecia uma boa aproximação que era adequada à quantidade de dados e à quantidade de informações que eles continham.

## 2.2 Abordagem Multivariada - Cópulas

É comum em pesquisas biomédicas e de saúde pública a observação repetida de múltiplas variáveis respostas na mesma unidade de análise. Esses experimentos resultam em dados longitudinais multivariados, os quais vêm ganhando bastante popularidade, pois permitem a evolução conjunta desses resultados ao longo do tempo (ver Bandyopadhyay et al., 2011; Verbeke et al., 2014; Cho, 2016). A utilização de métodos especiais (isto é, adequados) faz-se necessária para analisar dados multivariados no contexto longitudinal, pois tanto as observações repetidas no mesmo indivíduo quanto as diferentes variáveis respostas possivelmente exibem associação entre si. De acordo com Cho (2016), a correlação natural existente em dados longitudinais geralmente dificulta a especificação da função de verossimilhança completa.

Como alternativa, as *cópulas* são funções de distribuição conjunta que vêm sendo utilizadas na modelagem multivariada da estrutura de dependência no contexto longitudinal. Conforme outrora mencionado, por serem bastante flexíveis, as cópulas podem ser empregadas mesmo quando há violação da suposição de normalidade multivariada dos dados (ver Lambert, 2002), além de capturar dependência não linear e de cauda entre as variáveis. Assim, sua aplicabilidade na

análise de dados longitudinais multivariados dá-se em áreas como: biomédica (Lambert, 2002; Joe, 2014), saúde pública (Ferreira et al., 2019), atuária (Frees, 2006; Frees et al., 2021), hidrologia (Genest and Favre, 2007), dentre outras.

Neste trabalho, uma abordagem multivariada (no caso, trivariada) por meio de algumas funções cópula foi desenvolvida com o intuito de considerar a dependência entre os três índices antropométricos de 76 crianças ao longo do período estudado, uma vez que um dos principais desafios da análise de dados longitudinais multivariados é a natureza da complexa estrutura de correlação.

Uma cópula  $C$  é uma função de distribuição  $d$ -dimensional  $C: [0,1]^d \rightarrow [0,1]$ , de modo que suas marginais univariadas tenham distribuição uniforme no intervalo  $[0, 1]$ , isto é,  $U_j \sim U(0, 1)$ , para  $j = 1, 2, \dots, d$ . Assim, em um contexto trivariado ( $d = 3$ ), pode-se escrever uma cópula como sendo  $C(u_1, u_2, u_3) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2, U_3 \leq u_3)$ . Essas variáveis  $U_1, U_2, U_3 \sim U(0, 1)$  são originadas por uma transformação especial, chamada de transformação integral de probabilidade, segundo a qual  $U_j = F_j(Y_j)$ , em que  $F_j(Y_j)$  representa a função de distribuição acumulada (FDA) da variável aleatória  $Y_j$ , com  $j = 1, 2, 3$ . Neste trabalho,  $Y_1$  representa o índice 1, enquanto  $Y_2$  e  $Y_3$  representam os índices 2 e 3, respectivamente. Nesta subseção serão omitidos os índices/subscritos  $i$  e  $t$  de  $Y_{jit}$ , bem como de  $U_{jit}$  (ou  $u_{jit}$ ), por simplicidade de notação.

Selecionar uma cópula adequada para modelar a estrutura de dependência dos dados é extremamente importante. De acordo com Frahm et al. (2003) e Nelsen (2006), as cópulas mais comumente utilizadas na literatura são as *elípticas* e as *arquimedianas*. As cópulas elípticas são originárias de distribuições elípticas multivariadas, sendo a Gaussiana e a t-Student (ou simplesmente, cópula t) as mais famosas. As cópulas arquimedianas, por sua vez, possuem boas propriedades matemáticas e a sua construção dá-se por meio de funções convexas, sendo Clayton, Frank e Gumbel as mais utilizadas. De acordo com Sun et al. (2008), em um contexto de dados longitudinais multivariados, as cópulas elípticas são tipicamente preferidas às cópulas arquimedianas.

Ainda pouco exploradas na literatura (ver, por exemplo, Zhang et al., 2018, 2019), as cópulas Skew-normal (SN) e Skew-t (ST) pertencem a uma das poucas classes de cópulas assimétricas existentes, conhecida como *elípticas assimétricas* (Genest and Favre, 2007; Kollo et al., 2013). Sua principal vantagem em relação às cópulas com

estrutura de dependência simétrica (como é o caso das cópulas Gaussiana e t) é, de acordo com Kim and Kim (2016), a flexibilidade em detectar dependência direcional e não linear entre as variáveis. Ademais, a cópula ST apresenta um parâmetro adicional ( $\nu$ ) que regula a curtose dos dados, o que faz dela uma alternativa útil nas aplicações a dados com caudas mais pesadas.

Assim, para o estudo da dependência entre as variáveis respostas de interesse (índices 1, 2 e 3), são consideradas neste trabalho as cópulas elípticas Gaussiana e t, além das cópulas elípticas assimétricas SN e ST, com diferentes estruturas de dispersão. Essas cópulas, em suas versões trivariadas, são obtidas a partir da seguinte expressão:

$$C(u_1, u_2, u_3; \Psi) = H_3(H^{-1}(u_1), H^{-1}(u_2), H^{-1}(u_3); \Psi) \quad (3)$$

em que  $H^{-1}(.)$  representa a função quantil da distribuição univariada subjacente e  $H_3(.; \Psi)$  é a FDA conjunta da distribuição trivariada subjacente, a qual geralmente possui vetor de parâmetros de posição/localização  $\mathbf{0} = (0, 0, 0)^\top$  e matriz de correlação  $\Psi$  que, por sua vez, pode apresentar as seguintes estruturas (que são as mais conhecidas):

$$\begin{bmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & \rho_1 & \rho_1^2 \\ \rho_1 & 1 & \rho_1 \\ \rho_1^2 & \rho_1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{bmatrix} \quad (4)$$

Essas estruturas são, respectivamente, chamadas de *permutável* (ex), *autoregressiva de ordem 1* (ar1), *Toeplitz* (toep) e *não estruturada* (un), com  $\rho_j \in [-1, 1]$ , para  $j = 1, 2, 3$ . Pode-se, então, observar associação positiva ( $\rho_j > 0$ ), negativa ( $\rho_j < 0$ ) ou nula ( $\rho_j = 0$ ) entre as marginais.

No caso da cópula Gaussiana,  $H^{-1}(.) = \Phi^{-1}(.)$  e  $H_3(.; \Psi) = \Phi_3(.; \Psi)$ , em que  $\Phi^{-1}(.)$  é a função quantil da distribuição  $N(0, 1)$  e  $\Phi_3(.; \Psi)$  representa a FDA conjunta da distribuição normal trivariada com vetor de médias  $\mathbf{0}$  e matriz de correlação  $\Psi$ . Tal cópula é empregada quando, mesmo diante de não normalidade das marginais, há interesse em preservar uma dependência Gaussiana em sua modelagem conjunta. A cópula t, por sua vez, é obtida de (3) tomando  $H^{-1}(.) = T_\nu^{-1}(.)$  e  $H_3(.; \Psi) = T_{3,\nu}(.; \Psi)$ , em que  $T_\nu^{-1}(.)$  é a função quantil da distribuição t-Student com  $\nu$  graus de liberdade e

$T_{3,\nu}(\cdot; \Psi)$  representa a FDA conjunta da distribuição t-Student trivariada com vetor de parâmetros de posição/localização 0 e matriz de correlação  $\Psi$ . Geralmente, assume-se  $\nu$  fixo e pequeno, sendo  $\nu = 3$  ou 4 (adotado neste trabalho) uma escolha comum para esse parâmetro (EBA, 2014; Glogger, 2015). No caso da cópula SN,  $H^{-1}(\cdot) = G^{-1}(\cdot; \alpha_j)$  e  $H_3(\cdot; \Psi) = G_3(\cdot; \alpha, \Psi)$ , em que  $G^{-1}(\cdot; \alpha_j)$  é a função quantil da distribuição SN com parâmetros de posição/localização 0, escala 1 e de forma/assimetria  $\alpha_j$ , e  $G_3(\cdot; \alpha; \Psi)$  representa a FDA conjunta da distribuição SN trivariada com vetor de parâmetros de posição/localização 0, vetor de parâmetros de assimetria  $\alpha = (\alpha_1, \alpha_2, \alpha_3)^\top$  e matriz de correlação  $\Psi$ . Se  $\alpha_1 = \alpha_2 = \alpha_3$ , a cópula SN exibe estrutura de dependência permutável, sendo radialmente simétrica nos casos em que  $\alpha_1 = \alpha_2 = \alpha_3 = 0$  (reduzindo-se, então, à cópula Gaussiana). Finalmente, para a cópula ST, tem-se que:  $H^{-1}(\cdot) = Q_\nu^{-1}(\cdot; \alpha_j)$  e  $H_3(\cdot; \Psi) = Q_{3,\nu}(\cdot; \alpha, \Psi)$ , em que  $Q_\nu^{-1}(\cdot; \alpha_j)$  é a função quantil da distribuição ST com parâmetros graus de liberdade  $\nu$ , posição/localização 0, escala 1 e forma/assimetria  $\alpha_j$ , e  $Q_{3,\nu}(\cdot; \alpha; \Psi)$  representa a FDA conjunta da distribuição ST trivariada com  $\nu$  graus de liberdade, vetor de parâmetros de posição/localização 0, vetor de parâmetros de assimetria  $\alpha$  e matriz de correlação  $\Psi$ . Aqui, também será assumido  $\nu$  fixo e igual a 4. Ademais, para todos os quatro modelos de cópulas descritos anteriormente, serão consideradas as mesmas estruturas de dispersão apresentadas em (4).

É importante mencionar que a cópula captura toda informação de dependência entre as variáveis. As medidas de dependência baseadas nas cópulas foram estudadas por décadas. Por exemplo, Schweizer (1981) provaram muitas propriedades matemáticas para algumas dessas medidas. O coeficiente de correlação de Pearson é, possivelmente, a medida mais utilizada quando o interesse é estudar o grau de associação (ou dependência) entre as variáveis. Porém, a distribuição conjunta não pode ser completamente determinada levando-se em consideração sua medida, uma vez que não é possível construir uma distribuição conjunta das marginais. De acordo com Ida et al. (2014), as versões populacionais do  $\tau$  de Kendall e  $\rho$  de Spearman, por sua vez, podem ser representadas em termos das cópulas. Neste trabalho, considera-se a medida de dependência  $\tau$  de Kendall, que apresenta a vantagem de ser invariante a transformações monótonas das variáveis aleatórias, e pode ser calculada de duas formas: (i) *pareada*, isto é, obtida para as variáveis respostas ( $Y_j$ , para  $j = 1, 2, 3$ ) tomadas duas a duas, por meio da expressão

$\tau_{jk} = \tau(Y_j, Y_k) = 4 \int \int_{[0,1]^2} C(u_j, u_k) dC(u_j, u_k) - 1$ , para  $j, k = 1, 2, 3$ , com  $j \neq k$ ; e (ii) *média*, ou seja, mediante o cálculo<sup>1</sup>  $\tau_s = (\tau_{12} + \tau_{13} + \tau_{23})/3$ . A tabela 1 apresenta a fórmula utilizada para o cálculo de  $\tau_{jk}$ , segundo cada um dos modelos de cópulas considerados.

Tabela 1 – Os modelos de cópulas considerados e suas medidas de dependência

Cópula	$\tau_{jk}$	$\lambda_{jk}^L$	$\lambda_{jk}^U$
Gaussiana	$\frac{\arcsin(\rho_{jk})}{\pi/2}$	0	0
t	$\frac{\arcsin(\rho_{jk})}{\pi/2}$	$2T_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho_{jk})}{1+\rho_{jk}}} \right) \dagger$	$2T_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho_{jk})}{1+\rho_{jk}}} \right)$
SN	Sem forma fechada*	0	0
ST	Sem forma fechada*	$\geq \lambda_{jk}^{L,t} \frac{T_{\nu+2} \left( -(\alpha_j + \alpha_k) \sqrt{\frac{(\nu+2)(1+\rho_{jk})}{2}} \right)}{T_{\nu+1} \left( -(\alpha_k + \alpha_j \rho_{jk}) \sqrt{\frac{\nu+1}{1+\alpha_j^2(1-\rho_{jk}^2)}} \right)} \ddagger$	$\geq \lambda_{jk}^{U,t} \frac{T_{\nu+2} \left( (\alpha_j + \alpha_k) \sqrt{\frac{(\nu+2)(1+\rho_{jk})}{2}} \right)}{T_{\nu+1} \left( (\alpha_k + \alpha_j \rho_{jk}) \sqrt{\frac{\nu+1}{1+\alpha_j^2(1-\rho_{jk}^2)}} \right)} \ddagger$

\* Cálculo numéricico (aproximado) mediante o uso da expressão geral de  $\tau_{jk}$  e a aplicação da função *adaptIntegrate()* do pacote *cubature* do software estatístico R (Team, 2018). A exemplo do que fora feito em Wei et al. (2016).

†  $T_{\nu+1}(\cdot)$  é a FDA da distribuição t-Student com  $(\nu + 1)$  graus de liberdade.

§  $\lambda_{jk}^{L,t}$  e  $\lambda_{jk}^{U,t}$  representam, respectivamente, o coeficiente de dependência caudal inferior e superior da cópula t com  $\nu$  graus de liberdade. Esses resultados acerca das cópulas SN e ST foram extraídos de Bortot (2010).

Outro conceito importante no contexto de cópulas é o de *dependência caudal* que, segundo Embrechts et al. (2002), se refere ao grau de associação (ou dependência) no quadrante superior direito ou inferior esquerdo de uma distribuição bivariada, sendo de grande relevância para o estudo da dependência entre valores extremos. Os coeficientes de dependência caudal inferior e superior (pareados), considerando as variáveis de interesse ( $Y_j$ ,  $j = 1, 2, 3$ ), podem ser calculados por

$$\lambda_{jk}^L = \lambda_L(Y_j, Y_k) = \lim_{t^* \rightarrow 0^+} \mathbb{P}(Y_k \leq F_k^{-1}(t^*) \mid Y_j \leq F_j^{-1}(t^*)) = \lim_{t^* \rightarrow 0^+} \frac{C(t^*, t^*)}{t^*} \text{ e}$$

$$\lambda_{jk}^U = \lambda_U(Y_j, Y_k) = \lim_{t^* \rightarrow 1^-} \mathbb{P}(Y_k > F_k^{-1}(t^*) \mid Y_j > F_j^{-1}(t^*)) = \lim_{t^* \rightarrow 1^-} \frac{1 - 2t^* + C(t^*, t^*)}{1 - t^*},$$

respectivamente, em que  $C$  representa a função de distribuição da cópula bivariada correspondente. As expressões para o cálculo dos coeficientes de dependência caudal dos modelos de cópulas aqui considerados são apresentadas na tabela 1. Observe, por exemplo, que, ao contrário das cópulas t e ST, as cópulas Gaussiana e SN não exibem dependência caudal (isto é,  $\lambda_{jk}^L = \lambda_{jk}^U = 0$ , para  $j, k = 1, 2, 3$ ).

<sup>1</sup>Conforme observado por Heilpern (2014), a medida  $\tau_s$  é a mesma obtida tomando-se  $d = 3$  na expressão da versão multivariada do  $\tau$  de Kendall apresentada por Nelsen (1996):  $\tau_2 = \frac{1}{2^{d-1}-1} \left( 2^d \int \cdots \int_{[0,1]^d} C(u_1, \dots, u_d) dC(u_1, \dots, u_d) - 1 \right)$ .

Por fim, a obtenção da *densidade* da cópula é também de grande importância, como ficará evidenciado na seção 3. Tal densidade, denotada por  $c$ , pode ser calculada da seguinte maneira (caso trivariado):

$$c(u_1, u_2, u_3) = \frac{\partial^3 C(u_1, u_2, u_3)}{\partial u_1 \partial u_2 \partial u_3}$$

Deste modo, a partir da expressão (3), obtém-se:

$$c(u_1, u_2, u_3; \Psi) = \frac{h_3(H^{-1}(u_1), H^{-1}(u_2), H^{-1}(u_3); \Psi)}{\prod_{j=1}^3 h(H^{-1}(u_j))}$$

em que  $h(\cdot)$  representa a função densidade de probabilidade (FDP) da distribuição univariada subjacente e  $h_3(\cdot; \Psi)$  é a FDP conjunta da distribuição trivariada subjacente.

### 3 INFERÊNCIA

Nesta seção é feita uma discussão acerca dos métodos inferenciais clássicos para os modelos de dados longitudinais propostos nas formas univariadas e trivariadas.

A função de verossimilhança do MLM (1) é dada por:

$$\mathcal{L}_j(\delta_j | \mathbf{y}_j) = \prod_{i=1}^m \int \cdots \int f_j(\mathbf{y}_{ji} | \boldsymbol{\omega}_{ji}, \boldsymbol{\beta}_j, \mathbf{R}_{ji}, \boldsymbol{\Sigma}_j) f_j^*(\boldsymbol{\omega}_{ji} | \boldsymbol{\Sigma}_j) d\boldsymbol{\omega}_{ji} \quad (5)$$

em que  $\delta_j = \{\boldsymbol{\beta}_j, \mathbf{R}_j, \boldsymbol{\Sigma}_j\}$  representa o conjunto de parâmetros a serem estimados, sendo  $\mathbf{R}_j$  uma matriz quadrada diagonal de ordem  $\sum_{i=1}^m n_{ji}$ , cujos elementos são  $\mathbf{R}_{ji}$ , para  $j = 1, 2, \dots, d$  e  $i = 1, 2, \dots, m$ ;  $\mathbf{y}_j = (y_{j11}, \dots, y_{jmn_{jm}})^\top$  é o vetor que contém todas as respostas observadas, sendo  $\mathbf{y}_{ji} = (y_{ji1}, \dots, y_{jin_{ji}})^\top$  o vetor de tamanho  $n_{ji}$  com as respostas observadas na  $j$ -ésima variável dependente para a  $i$ -ésima unidade;  $f_j(\mathbf{y}_{ji} | \boldsymbol{\omega}_{ji}, \boldsymbol{\beta}_j, \mathbf{R}_{ji}, \boldsymbol{\Sigma}_j)$  representa a densidade condicional do vetor de respostas  $\mathbf{Y}_{ji}$  no ponto  $\mathbf{y}_{ji}$  dado  $\boldsymbol{\omega}_{ji}$ ; e  $f_j^*(\boldsymbol{\omega}_{ji} | \boldsymbol{\Sigma}_j)$  é a densidade conjunta de  $\boldsymbol{\omega}_{ji}$ .

Da subseção 2.1, tem-se que:  $\mathbf{Y}_{ji} | \boldsymbol{\omega}_{ji} \sim \mathcal{N}_{n_{ji}}(\mathbf{X}_{ji}\boldsymbol{\beta}_j + \mathbf{Z}_{ji}\boldsymbol{\omega}_{ji}, \mathbf{R}_{ji})$  e  $\boldsymbol{\omega}_{ji} \sim \mathcal{N}_{q_j}(\mathbf{0}_j, \boldsymbol{\Sigma}_j)$ , em que  $\mathbf{X}_{ji}$  é uma matriz de delineamento  $n_{ji} \times p_j$  dada por  $\mathbf{X}_{ji} = [\mathbf{X}_{ji}^{(0)} \mathbf{X}_{ji}^{(1)} \cdots \mathbf{X}_{ji}^{(p_j-1)}]$  e  $\mathbf{Z}_{ji}$  é uma matriz  $n_{ji} \times q_j$  definida como  $\mathbf{Z}_{ji} = [\mathbf{Z}_{ji}^{(0)} \mathbf{Z}_{ji}^{(1)} \cdots \mathbf{Z}_{ji}^{(q_j-1)}]$ , ambas de posto completo. Além disso, tem-se que:  $n_{ji} = n_i = n = 6$ ,  $p_j = p = 5$  e  $q_j = q = 1$ . Observe que (5) é a função densidade marginal

de  $\mathbf{y}_j$ , que é obtida integrando-se a densidade conjunta de  $\mathbf{y}_j$  e  $\omega_j$  sobre  $\omega_j$ , sendo  $\omega_j = (\omega_{j1}^\top, \dots, \omega_{jm}^\top)^\top$ .

Para obter os estimadores de máxima verossimilhança para os parâmetros do modelo (1), a função (vetor) escore, que é definida como a derivada parcial de primeira ordem do logaritmo da função de verossimilhança (5) com relação a  $\delta_j$ , deve ser igualada ao vetor zero. No entanto, nem sempre é possível encontrar expressões com forma fechada para esses estimadores. Logo, é frequentemente necessário o uso de métodos numéricos para resolver tais equações. Ademais, os procedimentos usuais de estimação por máxima verossimilhança não levam em consideração a presença de efeitos aleatórios; uma alternativa a esses procedimentos é aplicar o *método da máxima verossimilhança restrita* (MVR). Conforme Marcelino (2000), no método da MVR, a função de verossimilhança dada em (5) é dividida em duas partes independentes: uma referente aos efeitos fixos e a outra referente aos efeitos aleatórios. Neste trabalho é empregado o método da MVR para os MLMs propostos, mediante o uso da função *lme()* do pacote *nlme* do *software R*. Vale comentar que o método da MVR gera estimadores menos viesados (e, em alguns casos, não viesados) para as componentes de variância que o método de máxima verossimilhança usual.

Com relação aos modelos de cópulas, foi utilizado um método paramétrico de estimação clássica conhecido como *método da inferência para as marginais* (ou IFM, sigla do inglês “*Inference Function for Margins*”), proposto por Joe and Xu (1996). O método IFM tem por finalidade obter as estimativas dos parâmetros dos modelos derivados de cópulas em duas etapas. Na primeira etapa, estima-se (via método da MVR, por exemplo) os parâmetros  $\delta_j$ , para  $j = 1, 2, 3$  (caso trivariado), por meio da seguinte expressão:

$$\hat{\delta}_j = \arg \max_{\delta_j} \ell_j(\delta_j \mid \mathbf{y}_j)$$

em que  $\ell_j(\delta_j \mid \mathbf{y}_j)$  denota o logaritmo da função de verossimilhança (5) para o MLM do índice  $j$ , ou seja,  $\ell_j(\delta_j \mid \mathbf{y}_j) = \log(\mathcal{L}_j(\delta_j \mid \mathbf{y}_j))$ . Na segunda etapa, uma pseudo log-verossimilhança é utilizada para estimar o parâmetro da cópula  $\theta$  (o qual pode ser um vetor, isto é,  $\boldsymbol{\theta}$ ) da seguinte maneira:

$$\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^m \sum_{t=1}^n \log(c(\hat{u}_{1it}, \hat{u}_{2it}, \hat{u}_{3it} \mid \theta))$$

em que  $\hat{u}_{jit} = F_j(y_{jit} | \hat{\delta}_j)$ , para  $j = 1, 2, 3$ . O emprego do termo *pseudo*, neste caso, decorre do uso de  $\hat{u}_{1it}$ ,  $\hat{u}_{2it}$  e  $\hat{u}_{3it}$  ao invés dos verdadeiros  $u_{1it}$ ,  $u_{2it}$  e  $u_{3it}$ . A grande vantagem do método IFM é o custo computacional reduzido, uma vez que um número menor de parâmetros é estimado em cada etapa; ao contrário do método da máxima verossimilhança completa, em que todos os parâmetros são estimados simultaneamente (isto é, em uma única etapa).

Embora o uso de cópulas seja uma alternativa viável para a modelagem da estrutura de dependência dos dados sobre crescimento infantil, tratar essa dependência como constante, considerando MLMs derivados de cópulas trivariadas, pode não ser uma abordagem adequada, visto que o(s) parâmetro(s) das cópulas pode(m) variar com o tempo. Ou seja, a dependência (isto é, o grau de associação) entre as variáveis de interesse (índices 1, 2 e 3) pode mudar com o tempo, bem como o efeito das covariáveis sobre elas e as componentes de variância. Para mais detalhes sobre as *cópulas variantes no tempo* (do inglês *"time-varying copulas"*), veja os trabalhos de Patton (2009) e Manner (2012).

A função de log-verossimilhança para os modelos lineares (neste caso, modelos de regressão linear múltipla) derivados de cópulas trivariadas variantes no tempo, pode ser escrita da seguinte forma:

$$\ell(\boldsymbol{\xi}) = \sum_{i=1}^m \sum_{t=1}^n \log(c(u_{1it}, u_{2it}, u_{3it} | \boldsymbol{\theta}_t)) + \sum_{i=1}^m \sum_{t=1}^n \sum_{j=1}^3 \log(f_{jt}(y_{jit} | \boldsymbol{\delta}_{jt}))$$

em que  $\boldsymbol{\xi} = (\boldsymbol{\delta}_1^\top, \boldsymbol{\delta}_2^\top, \boldsymbol{\delta}_3^\top, \boldsymbol{\theta}^\top)^\top$  é o conjunto formado por todos os parâmetros do modelo,  $\boldsymbol{\delta}_{jt} = (\boldsymbol{\beta}_{jt}^\top, \sigma_{jt}^2)^\top$  é o vetor de parâmetros da  $j$ -ésima distribuição marginal univariada, sendo  $\boldsymbol{\beta}_{jt}$  e  $\sigma_{jt}^2$  o vetor de coeficientes da regressão e a variância, respectivamente, no tempo  $t$ ;  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^\top$  é o vetor de parâmetros da cópula variante no tempo;  $f_{jt}(y_{jit} | \boldsymbol{\delta}_{jt})$  denota a densidade da variável resposta  $Y_{jit}$  no ponto  $y_{jit}$ ; e  $u_{1it}$ ,  $u_{2it}$  e  $u_{3it}$  representam as marginais uniformes (FDA's) para os modelos lineares dos índices 1, 2 e 3, respectivamente.

Com o objetivo de avaliar o efeito das covariáveis sobre a força da dependência entre as distribuições marginais (isto é, os MLMs para os índices 1, 2 e 3), pode-se conectar o parâmetro de associação da cópula trivariada a uma combinação linear das covariáveis através de uma função de ligação apropriada, isto é,  $\theta_i = \theta(\mathbf{X}_i) = g(\mathbf{X}_i^\top \boldsymbol{\gamma})$ ,

para  $i = 1, 2, \dots, m$ , em que  $g^{-1}(\cdot)$  denota a função de ligação empregada e  $\gamma$  é o vetor de coeficientes dessa regressão. No caso das cópulas Gaussiana, t, SN e ST, cujo parâmetro de associação é  $\rho_j \in [-1, 1]$ , com  $j = 1, 2$  ou  $3$  (a depender da estrutura de dispersão considerada), pode-se utilizar a função de ligação *rholit* (tal como em Oliveira, 2018), definida como:

$$\log\left(\frac{1 + \rho_{ji}}{1 - \rho_{ji}}\right) = \mathbf{X}_{ji}^\top \gamma_j \in \mathbb{R} \implies \rho_{ji} = \frac{\exp\{\mathbf{X}_{ji}^\top \gamma_j\} - 1}{\exp\{\mathbf{X}_{ji}^\top \gamma_j\} + 1} \in [-1, 1]$$

Se o vetor de covariáveis utilizadas nas diferentes regressões acima for o mesmo (como é o caso do presente estudo), então tem-se que:  $\mathbf{X}_{ji} = \mathbf{X}_i$ . Dentre outros trabalhos relevantes, que consideram/propõem a inclusão de covariáveis no(s) parâmetro(s) da cópula, pode-se citar: Li (2018) e Li (2019).

Neste trabalho, para a estimação dos parâmetros, emprega-se o método IFM tanto para os MLMs derivados de cópulas trivariadas (com parâmetro(s) de associação constante(s) ou dependente(s) de covariável(eis)), quanto para os modelos lineares derivados de cópulas trivariadas variantes no tempo. Para tal, utiliza-se a função *optim()* (método “L-BFGS-B”) do R.

Finalmente, para a obtenção dos erros-padrões dos estimadores IFM para os parâmetros dos modelos de cópulas propostos, recorrer-se-á a um procedimento de reamostragem *bootstrap*, conforme sugerido por Joe and Xu (1996). Para mais detalhes sobre a técnica de reamostragem *bootstrap*, veja Tibshirani (1993). A grande vantagem do *bootstrap* é que não é necessário calcular derivadas analíticas a fim de obter a matriz de covariâncias assintóticas para os estimadores dos parâmetros do modelo em questão. No caso, aplica-se um procedimento *bootstrap* paramétrico, tal como em Ferreira et al. (2019), em que 500 novas amostras (também chamadas de amostras *bootstrap*) foram geradas dos modelos estimados, isto é, cujos parâmetros foram igualados às estimativas pontuais IFM. Em seguida, para cada uma das amostras geradas, foram obtidas as estimativas IFM dos parâmetros e, por último, calculou-se os desvios-padrões amostrais dessas estimativas (que, neste caso, correspondem às estimativas *bootstrap* dos erros-padrões).

## 4 APLICAÇÃO AOS DADOS LONGITUDINAIS DE CRESCIMENTO INFANTIL

Nesta seção são apresentados alguns resultados iniciais, obtidos por meio de uma análise descritiva (ou exploratória) dos dados longitudinais sobre crescimento infantil (subseção 4.1); os principais resultados obtidos a partir da abordagem univariada, ou seja, mediante o ajuste dos MLMs marginais (subseção 4.2); e, por último, são mostrados os resultados da abordagem multivariada, isto é, da modelagem conjunta baseada em algumas funções cópula (subseção 4.3). Todas as análises foram realizadas no *software* R, versão 3.5.1 (Team, 2018), e os códigos desenvolvidos e utilizados estão disponíveis mediante solicitação aos autores.

Os dados utilizados neste trabalho fazem parte de um estudo de coorte coordenado pelo Departamento de Ciências da Nutrição da Escola de Nutrição da Universidade Federal da Bahia (ENUFBA), cujo objetivo geral era contribuir para o desenvolvimento do Sistema Municipal de Segurança Alimentar e Nutricional da cidade de Mutuípe (SANMUTUÍPE), localizada no Vale do Jiquiriçá, que é uma importante região do estado brasileiro da Bahia, caracterizada pela atividade pecuária e agricultura familiar ou de subsistência, com destaque para a produção de cacau (Panelli-Martins et al., 2008). Em particular, o interesse deste artigo é avaliar a magnitude da dependência temporal dos indicadores antropométricos das crianças nascidas na maternidade de Mutuípe no período de junho de 2005 a maio de 2006. Esses recém-nascidos foram acompanhados por um período de seis meses, com um mínimo de 2 e máximo de 6 observações por criança/índice. Como os indicadores antropométricos são calculados utilizando informações de peso, estatura, idade e gênero de cada criança, somente as informações de 76 recém-nascidos continham todas essas variáveis em cada mensuração. Vale ressaltar a importância do tema, pois um déficit de crescimento linear nos primeiros anos de vida pode repercutir, a curto e longo prazo, na suscetibilidade a infecções, nas habilidades cognitivas e em menor altura na vida adulta, por exemplo. Desta forma, uma avaliação mais rica e complexa do estado nutricional de crianças, por meio de indicadores antropométricos, é uma estratégia de baixo custo bastante empregada na epidemiologia nutricional (Fausto et al., 2008).

Conforme comentado anteriormente na seção 1, esse mesmo conjunto de dados foi utilizado/analizado por Ferreira et al. (2019), porém com a diferença de que

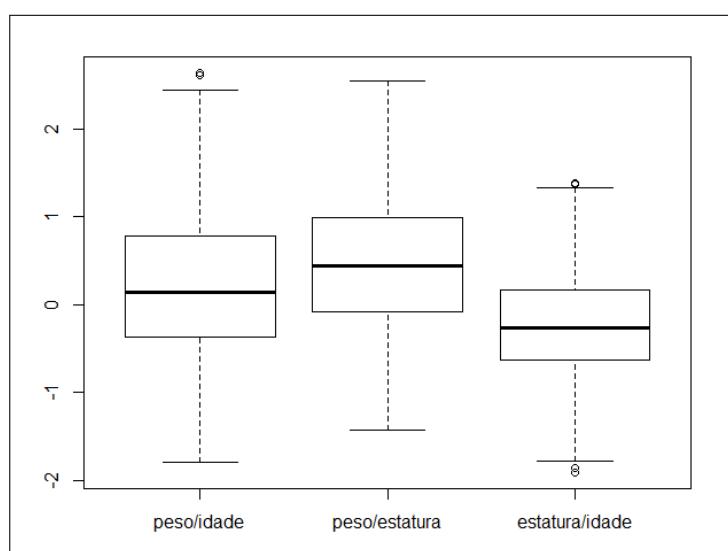
os autores consideraram um número maior de observações ( $m = 150$  crianças) e somente duas variáveis respostas (peso e estatura das crianças).

#### 4.1 Análise Descritiva

Inicialmente, uma análise descritiva (ou exploratória) de dados foi realizada a fim de verificar como os índices antropométricos 1 (peso/idade), 2 (peso/estatura) e 3 (estatura/idade) das crianças se comportam ao longo do período estudado. Como os índices são mensurados em escala normal padronizada (escores z), geralmente assumem valores entre -3 e 3. As próximas análises são das variáveis de ajuste: a média da altura dos recém-nascidos foi de 48,59 cm, variando de 44,85 cm a 52,50 cm, enquanto que o peso médio ao nascer foi de 3.214,43 g, variando de 2.405,00 g a 4.500,00 g. Os níveis de hemoglobina variaram de 7,10 g/dL a 18,60 g/dL, com média 10,78 g/dL.

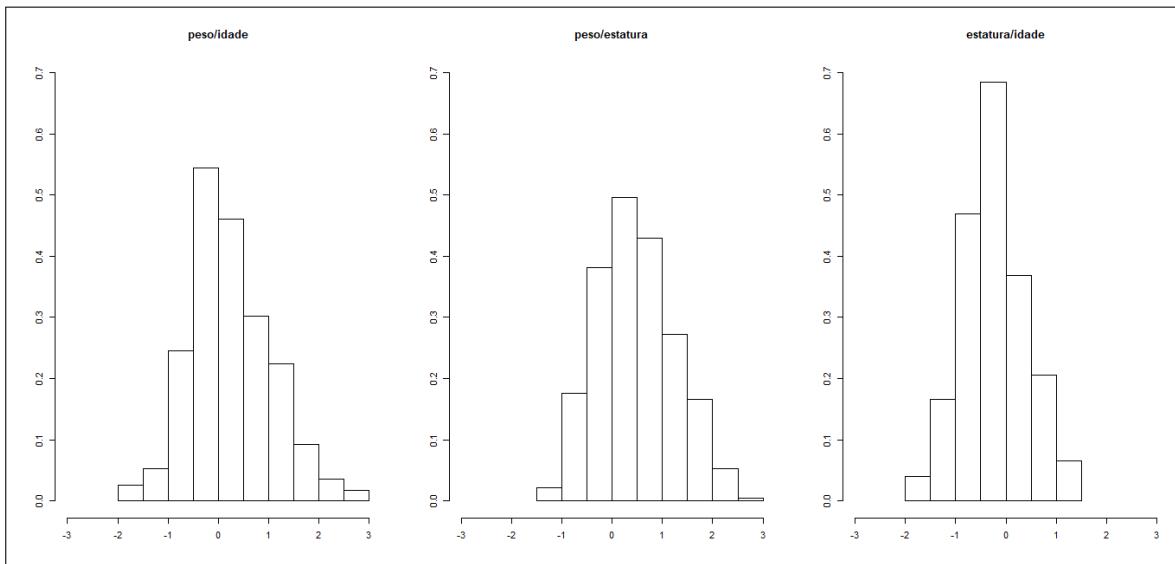
As figuras 1 e 2 mostram, respectivamente, os *boxplots* e histogramas das medidas antropométricas das crianças em estudo. Observa-se que os índices 1 e 2 apresentam similaridade em termos de variabilidade. Os índices de peso e estatura relacionados à idade (índices 1 e 3, respectivamente) possuem valores discrepantes (*outliers*), visto que existem crianças com peso muito acima do esperado em relação à sua idade, bem como crianças com estatura muito superior e também muito inferior ao esperado para a sua idade (em comparação ao conjunto de dados em análise).

Figura 1 – *Boxplots* dos índices antropométricos 1 (peso/idade), 2 (peso/estatura) e 3 (estatura/idade) das crianças de Mutuípe, Bahia, 2005-2006



Fonte: autores (2022)

Figura 2 – Histogramas dos índices antropométricos 1 (peso/idade), 2 (peso/estatura) e 3 (estatura/idade) das crianças de Mutuípe, Bahia, 2005-2006



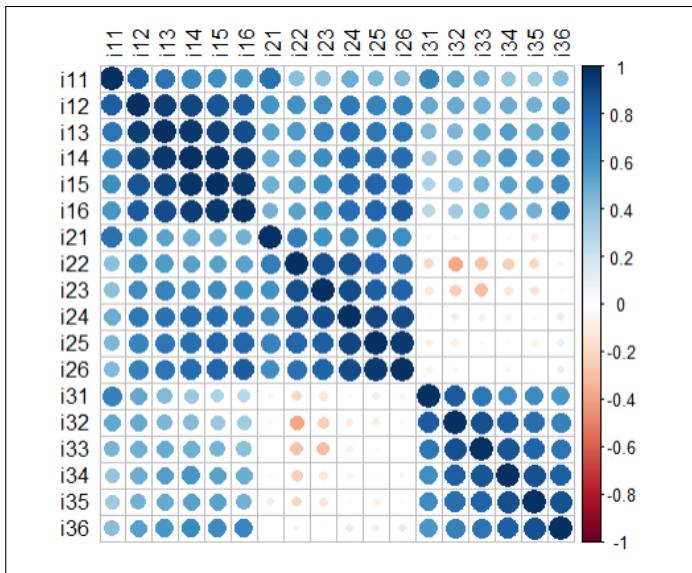
Fonte: autores (2022)

Estudar descritivamente o grau de dependência entre as variáveis respostas, assim como suas relações ao longo do tempo, auxilia na identificação de uma estrutura de variância/correlação para os modelos de regressão, como por exemplo, modelos de efeitos mistos. Assim, primeiramente foi estimada a correlação entre os indicadores antropométricos, cujos resultados revelaram que o índice peso/idade apresenta correlação positiva moderada com o índice peso/estatura (0,749) e com o índice estatura/idade (0,653), enquanto que o índice peso/estatura exibe uma correlação negativa fraca com o índice estatura/idade (-0,004). Com o objetivo de caracterizar a magnitude da dependência dos três indicadores antropométricos ao longo do tempo, um *correlograma* é exibido na figura 3.

Como mencionado na subseção 2.1, a escolha das matrizes de covariâncias dos modelos foi feita com base nos valores de AIC e pela construção de semivariogramas (dados desbalanceados). A figura 4 mostra o semivariograma para cada um dos três índices antropométricos, em que a linha horizontal tracejada representa a estimativa da variância do processo. O objetivo é descrever a associação entre esses indicadores ao longo do tempo. Observa-se que a correlação entre eles diminui à medida que a idade aumenta. Segundo Diggle et al. (2002), na maioria dos fenômenos que são mensurados ao longo do tempo, a correlação serial diminui à medida que a separação entre os tempos aumenta. Além disso, nota-se uma variabilidade mais acentuada

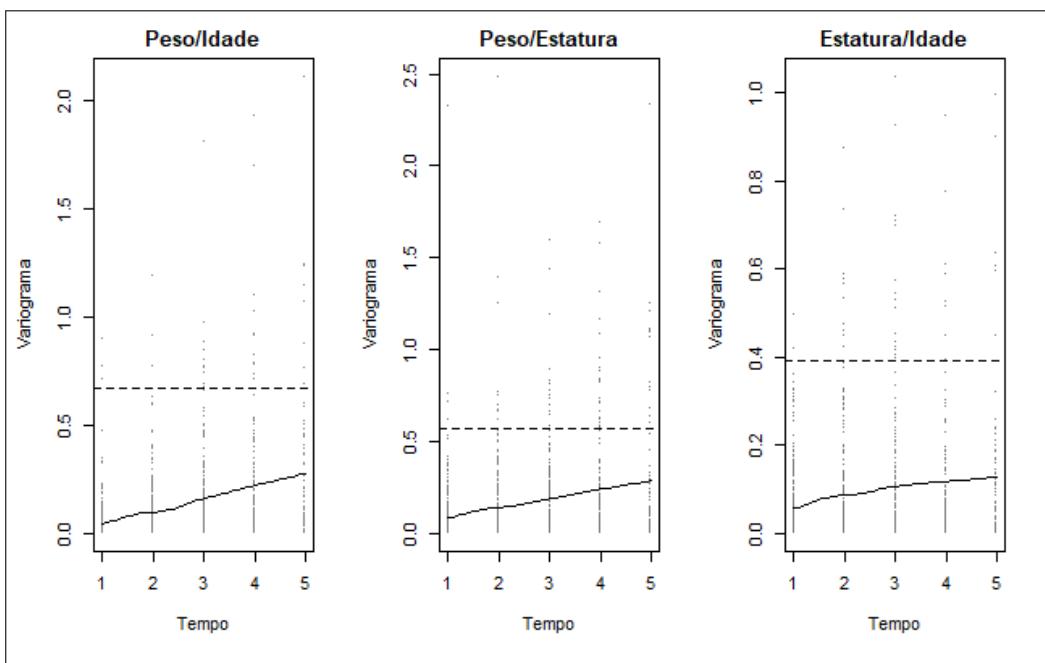
entre as crianças, sugerindo a utilização dos modelos mistos com intercepto aleatório como alternativa na abordagem univariada, isto é, no ajuste de cada um dos três índices em estudo.

Figura 3 – Correlação entre os índices antropométricos 1 (peso/idade), 2 (peso/estatura) e 3 (estatura/idade) ao longo do tempo, em que  $i[j, t]$  denota o índice  $j$  ( $j = 1, 2, 3$ ) no tempo  $t$  ( $t = 1, 2, 3, 4, 5, 6$ ), das crianças de Mutuípe, Bahia, 2005-2006



Fonte: autores (2022)

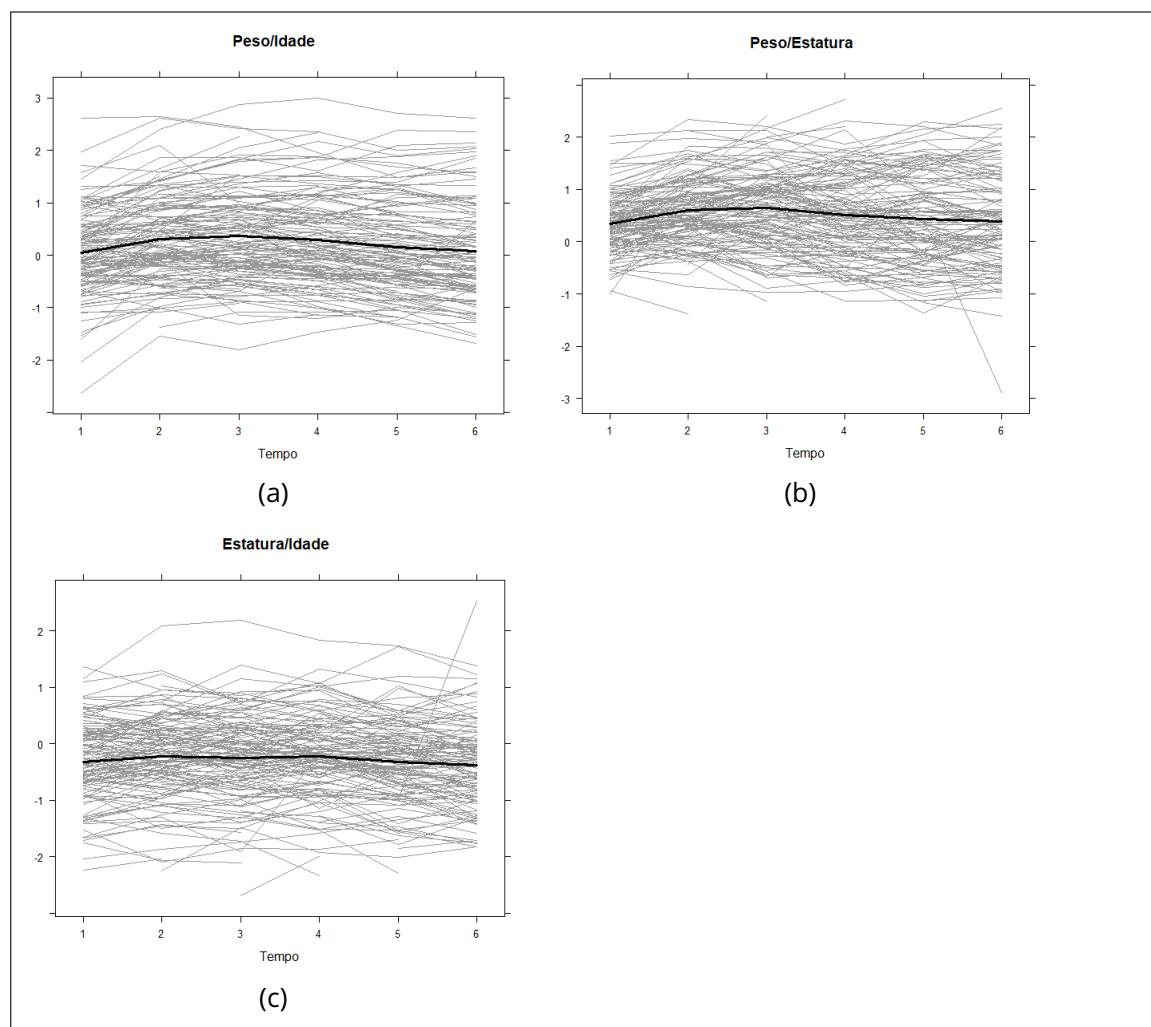
Figura 4 – Semivariograma para os índices antropométricos 1 (peso/idade), 2 (peso/estatura) e 3 (estatura/idade) das crianças de Mutuípe, Bahia, 2005-2006



Fonte: autores (2022)

Outra ferramenta descritiva importante na análise de dados longitudinais é a construção do *gráfico de perfis*, cujo objetivo é verificar padrões individuais da variável resposta ao longo do tempo, bem como comparar a variabilidade inter e intra-crianças. A figura 5 mostra os perfis individuais para os índices 1, 2 e 3. A linha sólida preta representa o perfil médio da variável resposta ao longo do tempo. Nota-se que, para cada um dos indicadores antropométricos, a variabilidade no início é menor do que ao final do estudo, com exceção do índice 3, que parece manter a mesma variabilidade. Contudo, é notória uma variabilidade acentuada entre as crianças.

Figura 5 – Distribuição longitudinal dos índices antropométricos 1 (peso/idade), 2 (peso/estatura) e 3 (estatura/idade), perfis individual e médio das crianças de Mutuípe, Bahia, 2005-2006



Fonte: autores (2022)

## 4.2 MLMs Univariados

Os principais resultados dos ajustes dos modelos lineares de efeitos mistos, separadamente para cada índice antropométrico, encontram-se na tabela 2. Observa-se, dentre outros, que as crianças do sexo feminino possuem indicadores antropométricos 1, 2 e 3, em média, superiores aos das crianças do sexo masculino (0,5862, 0,3381 e 0,3289, respectivamente). O índice 2 reduz 0,2149 pontos, enquanto que o índice 3 aumenta em 0,2936 pontos a cada aumento de 1 cm no comprimento da criança ao nascer. A cada aumento de 100 g no peso ao nascer, os índices 1 e 2 acrescem, em média, 0,09 e 0,12 pontos, respectivamente. Esses mesmos índices decrescem 0,0432 e 0,0388 pontos, respectivamente, a cada aumento de 1 g/dL no nível da hemoglobina.

Tabela 2 – Principais resultados da estimação, via método da MVR, para os MLMs univariados aplicados aos dados longitudinais sobre desenvolvimento infantil. EP = erro-padrão

Parâmetro de Efeito Fixo †	Peso/Idade			Peso/Estatura			Estatura/Idade		
	Estimativa	EP	p-valor*	Estimativa	EP	p-valor*	Estimativa	EP	p-valor*
Intercepto	-0,1072	0,0994	0,2815	0,2886	0,0834	0,0006	-0,4243	0,0594	<0,0001
Tempo	-0,0163	0,0129	0,2070	-0,0067	0,0157	0,6688	-0,0139	0,0110	0,2085
Sexo (Feminino)	0,5862	0,1460	0,0001	0,3381	0,1225	0,0070	0,3289	0,0871	0,0003
Peso ao nascer	0,0009	0,0002	0,0001	0,0012	0,0002	<0,0001	0,0001	0,0001	0,6558
Hemoglobina	-0,0432	0,0103	<0,0001	-0,0388	0,0135	0,0045	-0,0199	0,0108	0,0664
Comprimento ao nascer	0,0793	0,0621	0,2045	-0,2149	0,0519	0,0001	0,2936	0,0369	<0,0001
Log-Verossimilhança	-219,6845	-	-	-301,0674	-	-	-183,1495	-	-
AIC	457,3689	-	-	620,1349	-	-	384,2989	-	-
BIC	494,3521	-	-	657,1181	-	-	421,2822	-	-

† Parâmetro de Efeito Aleatório - Índice 1:  $\hat{\sigma}_{\omega_1} = 0,0001$ ;  $\hat{\rho}_1 = 0,8971$ ;  $\hat{\sigma}_1 = 0,7251$

- Índice 2:  $\hat{\sigma}_{\omega_2} = 0,0002$ ;  $\hat{\rho}_2 = 0,7888$ ;  $\hat{\sigma}_2 = 0,6656$

- Índice 3:  $\hat{\sigma}_{\omega_3} = 0,2744$ ;  $\hat{\rho}_3 = 0,6061$ ;  $\hat{\sigma}_3 = 0,3889$

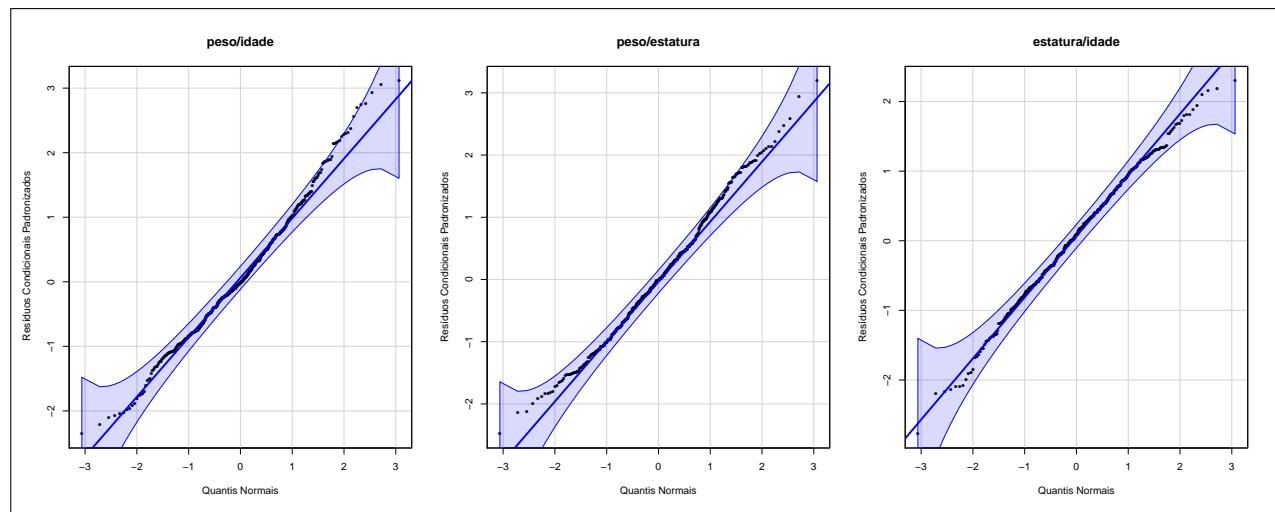
\* Do teste de Wald

Conforme apresentado na subseção 2.1, os três modelos finais ajustados consideram um efeito aleatório do intercepto. A matriz de covariâncias utilizada foi a autoregressiva de ordem 1, na qual a correlação diminui ao longo do tempo. Os valores do AIC e do Critério de Informação Bayesiano (BIC; Schwarz, 1978), para cada modelo ajustado, são exibidos na tabela 2, juntamente com o valor do logaritmo da função de verossimilhança (Log-Verossimilhança). Com o objetivo de comparar os resultados dos modelos ajustados nessa abordagem univariada com os resultados obtidos na abordagem conjunta/trivariada (ver subseção 4.3), calculou-se as somas

simples dos valores individuais dos critérios para compor um valor único total. Assim,  $AIC_{Total} = 457,3689 + 620,1349 + 384,2989 = 1.461,8027$  e  $BIC_{Total} = 494,3521 + 657,1181 + 421,2822 = 1.572,7524$ .

Foram usados os resíduos condicionais de cada modelo estimado para verificar a adequabilidade do ajuste. Tais resíduos são dados, a partir de (1), pelas diferenças entre os valores observados e os valores preditos das respostas, isto é,  $e_{jit} = Y_{jit} - \mathbf{X}_{jit}^\top \hat{\beta}_j - \mathbf{Z}_{jit}^\top \hat{\omega}_{ji}$ , para  $j = 1, 2, \dots, d$ ,  $i = 1, 2, \dots, m$  e  $t = 1, 2, \dots, n_{ji}$ , em que  $\hat{\beta}_j$  e  $\hat{\omega}_{ji}$  são, respectivamente, as estimativas MVR de  $\beta_j$  e  $\omega_{ji}$ .

Figura 6 – Gráficos de quantis com envelopes simulados para os resíduos condicionais padronizados obtidos a partir dos modelos ajustados aos índices antropométricos 1 (peso/idade), 2 (peso/estatura) e 3 (estatura/idade) das crianças de Mutuípe, Bahia, 2005-2006

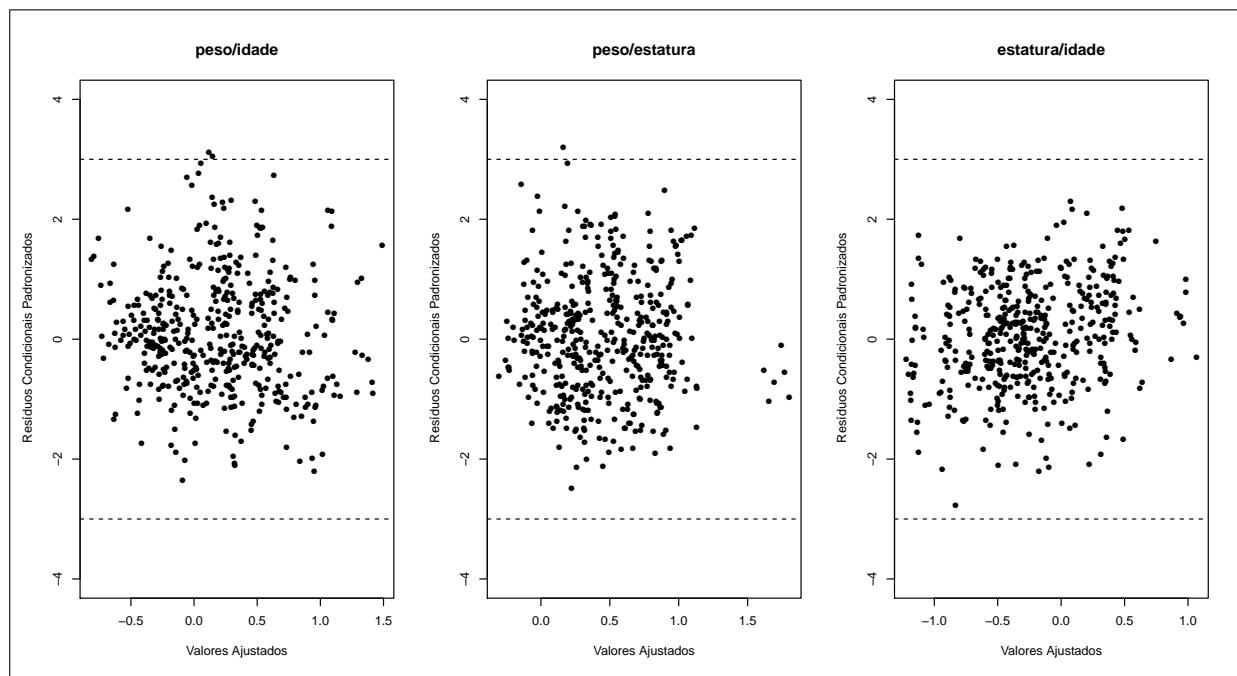


Fonte: autores (2022)

A figura 6 mostra a técnica gráfica para diagnóstico, conhecida como *gráfico de quantis normais* com envelopes simulados dos resíduos condicionais padronizados para os modelos ajustados aos índices 1, 2 e 3. O *resíduo condicional padronizado* é definido como  $e_{jit}^* = e_{jit} / \sqrt{\hat{r}_{jit}}$ , sendo  $\hat{r}_{jit}$  a  $(t, t)$ -ésima entrada de  $\hat{\mathbf{R}}_{ji}$ , que é a estimativa MVR de  $\mathbf{R}_{ji}$ . Tal figura, assim como os *gráficos de resíduos condicionais padronizados versus valores preditos* (figura 7), indicam a adequabilidade dos modelos de efeitos mistos propostos para os índices antropométricos 1, 2 e 3, uma vez que o pressuposto de normalidade dos resíduos condicionais padronizados foi validado e também não foram observados, em geral, valores grandes desses resíduos. A fim de confirmar a suposição de normalidade dos resíduos, foi realizado o teste de hipóteses

(aderência) Qui-quadrado de Pearson. Assim, verificou-se que não existem evidências suficientes para rejeitar a hipótese de normalidade dos resíduos condicionais padronizados dos modelos mistos ajustados aos índices antropométricos peso/idade, peso/estatura e estatura/idade, com os respectivos p-valores: 0,1103, 0,3995 e 0,9551. Cabe destacar que, apesar de importante, não é do interesse deste estudo realizar uma análise de diagnóstico (análise de resíduos e de influência) mais refinada e aprofundada, tal como proposto, por exemplo, em Nobre (2007, 2011) e Singer et al. (2017), visto que os resultados aqui obtidos indicaram um ajuste satisfatório dos três MLMs marginais, que serão considerados nas propostas de abordagem multivariada via cópulas desenvolvidas na subseção seguinte.

Figura 7 – Gráficos de resíduos condicionais padronizados *versus* valores preditos, obtidos a partir dos modelos ajustados aos índices antropométricos 1 (peso/idade), 2 (peso/estatura) e 3 (estatura/idade) das crianças de Mutuípe, Bahia, 2005-2006



Fonte: autores (2022)

### 4.3 MLMs Trivariados Derivados de Cópulas

Para a modelagem conjunta das três distribuições marginais, foram utilizadas as cópulas elípticas Gaussiana e t com  $\nu = 4$  graus de liberdade, além das cópulas elípticas assimétricas SN e ST, esta última também com  $\nu = 4$ . A tabela 3 exibe as estimativas IFM para os MLMs trivariados derivados de cada uma dessas cópulas, quando aplicadas aos

dados longitudinais sobre crescimento infantil. Os resultados apresentados incluem ainda os valores dos critérios de seleção AIC e BIC, a Log-Verossimilhança, além do erro-padrão *bootstrap* dos estimadores dos parâmetros das cópulas ( $\rho_j$  e  $\alpha_j$ , com  $j \in \{1, 2, 3\}$ ), de acordo com as diferentes estruturas de dispersão.

Tabela 3 – Principais resultados da estimação, via método IFM, para os parâmetros dos modelos de cópulas. Erro-padrão *bootstrap* entre parênteses

Cópula	$\rho_1$	$\rho_2$	$\rho_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	Log-Veross.	AIC	BIC
Gaussiana (ex)	0,7196 (0,0893)	-	-	-	-	-	-495,6793	1.047,3585	1.162,7883
Gaussiana (ar1)	0,7244 (0,0891)	-	-	-	-	-	-501,4665	1.058,9330	1.174,3628
Gaussiana (toep)	0,7264 (0,0915)	0,7002 (0,1398)	-	-	-	-	-495,4934	1.048,9868	1.168,5391
Gaussiana (un)	0,8330 (0,1102)	0,7127 (0,1349)	0,6498 (0,0946)	-	-	-	-484,5074	1.029,0148	1.152,6895
t ( $\nu = 4$ ; ex)	0,6798 (0,0841)	-	-	-	-	-	-503,8282	1.063,6564	1.179,0862
t ( $\nu = 4$ ; ar1)	0,6808 (0,0882)	-	-	-	-	-	-509,6813	1.075,3625	1.190,7923
t ( $\nu = 4$ ; toep)	0,6829 (0,0870)	0,6698 (0,1311)	-	-	-	-	-503,8011	1.065,6022	1.185,1545
t ( $\nu = 4$ ; un)	0,7849 (0,0982)	0,6672 (0,1365)	0,6090 (0,0992)	-	-	-	-497,4688	1.054,9375	1.178,6123
SN (ex)	0,7524 (0,1034)	-	-	-1,6500 (139,8895)	0,1303 (2,3516)	0,1531 (1,7734)	-460,7587	983,5174	1.111,3146
SN (ar1)	0,7038 (0,1159)	-	-	-1,9862 (91,0228)	0,1670 (16,6147)	0,4404 (2,6659)	-479,7865	1.021,5730	1.149,3703
SN (toep)	0,6771 (0,1289)	0,8669 (0,1340)	-	-2,0713 (20,0327)	0,2632 (4,1920)	0,0142 (0,1645)	-450,3966	964,7932	1.096,7129
SN (un)	0,8252 (0,1323)	0,8640 (0,1380)	0,5189 (0,1420)	-1,1933 (22,5031)	0,0719 (4,0310)	0,0056 (8,2313)	-400,9779	867,9558	1.003,9981
ST ( $\nu = 4$ ; ex)	0,7243 (0,1084)	-	-	-1,8014 (27,9925)	0,1179 (0,9473)	0,1355 (1,4039)	-461,3932	984,7864	1.112,5837
ST ( $\nu = 4$ ; ar1)	0,6718 (0,1213)	-	-	-1,9814 (21,3558)	0,1411 (10,0404)	0,3861 (2,7416)	-480,7035	1.023,4069	1.151,2042
ST ( $\nu = 4$ ; toep)	0,6452 (0,1337)	0,8496 (0,1547)	-	-2,2138 (27,1607)	0,2158 (2,8863)	0,0150 (1,9561)	-451,6232	967,2464	1.099,1662
ST ( $\nu = 4$ ; un)	0,7545 (0,1385)	0,8331 (0,1443)	0,3981 (0,1608)	-1,4187 (25,2313)	0,1316 (2,5577)	0,0317 (1,3210)	-415,1994	896,3988	1.032,4410

Fonte: autores (2022)

Uma vez que na tabela 2 encontram-se as estimativas MVR dos parâmetros marginais, esta subseção concentra-se nas estimativas dos parâmetros das cópulas, obtidas usando o método de estimação IFM descrito na seção 3. Nota-se claramente, observando-se as tabelas 2 e 3, que a abordagem conjunta por meio das cópulas trivariadas apresenta-se como uma melhor alternativa de modelagem quando comparada com a abordagem univariada (repare que todos os modelos baseados em cópulas têm valores de AIC e BIC inferiores ao  $AIC_{Total}$  e  $BIC_{Total}$  dos modelos univariados), justificando a utilização das cópulas como uma importante ferramenta

na estimação conjunta dos MLMs para os indicadores antropométricos 1, 2 e 3, melhorando, assim, a eficiência estatística. Observa-se também, na tabela 3, que o modelo linear de efeitos mistos trivariado derivado da cópula SN com matriz de dispersão não estruturada, foi selecionado como o melhor por apresentar os menores valores dos dois critérios utilizados: AIC = 867,9558 e BIC = 1.003,9981. As medidas de dependência estimadas para esse melhor modelo, assim como para todos os demais modelos de cópulas ajustados, estão disponíveis na tabela 4. Considerando o melhor ajuste, tais medidas revelam um relacionamento positivo e moderado entre as marginais (isto é, entre os MLMs para os índices 1, 2 e 3), com um  $\tau$  de Kendall médio de 0,4528. Ou seja, esse resultado indica a existência da associação entre os três indicadores antropométricos, bem como a importância da modelagem trivariada.

Tabela 4 – Medidas de dependência  $\tau$  de Kendall pareado ( $\hat{\tau}_{12}$ ,  $\hat{\tau}_{13}$  e  $\hat{\tau}_{23}$ ),  $\tau$  de Kendall médio ( $\hat{\tau}_s$ ) e coeficientes de dependência caudal (pareados) inferior ( $\hat{\lambda}_{12}^L$ ,  $\hat{\lambda}_{13}^L$  e  $\hat{\lambda}_{23}^L$ ) e superior ( $\hat{\lambda}_{12}^U$ ,  $\hat{\lambda}_{13}^U$  e  $\hat{\lambda}_{23}^U$ ), para os modelos de cópulas ajustados

Cópula	$\hat{\tau}_{12}$	$\hat{\tau}_{13}$	$\hat{\tau}_{23}$	$\hat{\tau}_s$	$\hat{\lambda}_{12}^L$	$\hat{\lambda}_{13}^L$	$\hat{\lambda}_{23}^L$	$\hat{\lambda}_{12}^U$	$\hat{\lambda}_{13}^U$	$\hat{\lambda}_{23}^U$
Gaussiana (ex)	0,5113	0,5113	0,5113	0,5113	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Gaussiana (ar1)	0,5158	0,3517	0,5158	0,4611	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Gaussiana (toep)	0,5176	0,4938	0,5176	0,5097	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Gaussiana (un)	0,6268	0,5051	0,4503	0,5274	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
t ( $\nu = 4$ ; ex)	0,4759	0,4759	0,4759	0,4759	0,3738	0,3738	0,3738	0,3738	0,3738	0,3738
t ( $\nu = 4$ ; ar1)	0,4767	0,3068	0,4767	0,4201	0,3746	0,2338	0,3746	0,3746	0,2338	0,3746
t ( $\nu = 4$ ; toep)	0,4786	0,4672	0,4786	0,4748	0,3763	0,3657	0,3763	0,3763	0,3657	0,3763
t ( $\nu = 4$ ; un)	0,5746	0,4650	0,4169	0,4855	0,4727	0,3636	0,3205	0,4727	0,3636	0,3205
SN (ex)	0,4580	0,4614	0,5276	0,4823	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
SN (ar1)	-0,0437	-0,1408	0,4597	0,0917	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
SN (toep)	-0,0498	0,0996	0,4656	0,1718	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
SN (un)	0,5532	0,4584	0,3467	0,4528	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
ST ( $\nu = 4$ ; ex)	0,4351	0,4375	0,5090	0,4605	0,4398	0,4428	0,3004	0,0002	0,0003	0,4110
ST ( $\nu = 4$ ; ar1)	0,3902	0,2983	0,4431	0,3772	0,4077	0,5233	0,2317	0,0001	4,75e-05	0,3536
ST ( $\nu = 4$ ; toep)	0,3752	0,5392	0,4416	0,4520	0,3949	0,5614	0,1982	3,68e-05	0,0005	0,3727
ST ( $\nu = 4$ ; un)	0,4816	0,5525	0,2581	0,4307	0,4297	0,5735	0,1300	0,0014	0,0060	0,2441

Fonte: autores(2022)

As tabelas 5 e 6 mostram os resultados das estimativas, obtidas pelo método IFM, para os vetores de parâmetros de associação ( $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6$ ) dos modelos lineares (isto é, modelos de regressão linear múltipla) trivariados derivados de cópulas variantes no tempo, bem como os valores de AIC, BIC, Log-Verossimilhança e os erros-padrões *bootstrap*. A partir dessas tabelas, conclui-se que o melhor modelo trivariado variante no tempo, segundo o critério BIC, é aquele derivado da cópula Gaussiana com matriz de dispersão não estruturada (BIC = 453,6909), mas de acordo com o critério AIC, seria aquele derivado da cópula SN com também matriz de

dispersão não estruturada (AIC = 152,3854). As medidas de dependência estimadas para esses melhores modelos estão apresentadas na tabela 7. Percebe-se, portanto, que tais valores de AIC e BIC são inferiores aos observados na tabela 3, em que o MLM trivariado derivado da cópula SN (un) revelou-se pior do que eles, de acordo com os dois critérios considerados. Assim, segundo esses critérios, os modelos lineares trivariados derivados de cópulas variantes no tempo superam os modelos lineares de efeitos mistos trivariados derivados de cópulas com parâmetro(s) constante(s) (ver tabelas 3, 5 e 6). Tal achado difere dos resultados obtidos por Ferreira et al. (2019), que realizaram um estudo similar ao proposto neste trabalho (porém com somente duas variáveis respostas: peso e estatura das crianças) e não encontraram vantagem (em termos de AIC e BIC) no uso das cópulas variantes no tempo. A saber, os modelos de cópulas (bivariadas) considerados pelos autores foram: Gaussiana, t com  $\nu = 2$ , Clayton, Frank e Gumbel.

Tabela 5 – Principais resultados da estimação, via método IFM, dos parâmetros dos modelos de cópulas Gaussiana e t ( $\nu = 4$ ) variantes no tempo. Erro-padrão *bootstrap* entre parênteses

Cópula	Parâm.	Mês						Log-Veross.	AIC	BIC
		1	2	3	4	5	6			
Gaussiana (ex)	$\rho_1$	0,5351 (0,0634)	0,4763 (0,0703)	0,4750 (0,0760)	0,5578 (0,0673)	0,5270 (0,0745)	0,6194 (0,0530)	-1.004,9370	2.237,8740	2.503,0436
Gaussiana (ar1)	$\rho_1$	0,4713 (0,0619)	0,3889 (0,0757)	0,3994 (0,0759)	0,4889 (0,0707)	0,4720 (0,0767)	0,5643 (0,0568)	-1.069,5408	2.367,0815	2.632,2512
Gaussiana (toep)	$\rho_1$	0,4735 (0,0809)	0,3914 (0,0849)	0,4023 (0,0848)	0,4913 (0,0868)	0,4771 (0,0878)	0,5657 (0,0608)	-986,3003	2.212,6006	2.491,7265
	$\rho_2$	0,6556 (0,0661)	0,6409 (0,0688)	0,6162 (0,0735)	0,6874 (0,0697)	0,6239 (0,0792)	0,7250 (0,0505)			
Gaussiana (un)	$\rho_1$	0,8102 (0,0380)	0,7566 (0,0530)	0,7884 (0,0470)	0,7858 (0,0489)	0,8526 (0,0383)	0,8294 (0,0379)	45,6956	160,6087	453,6909
	$\rho_2$	0,6389 (0,0711)	0,6159 (0,0705)	0,5727 (0,0786)	0,6314 (0,0746)	0,5865 (0,0808)	0,6701 (0,0596)			
	$\rho_3$	0,0945 (0,1097)	-0,0432 (0,1129)	-0,0498 (0,1141)	0,0203 (0,1232)	0,0790 (0,1214)	0,1440 (0,1035)			
t ( $\nu = 4$ ; ex)	$\rho_1$	0,5770 (0,0700)	0,5212 (0,0747)	0,5245 (0,0766)	0,5964 (0,0722)	0,5419 (0,0800)	0,6126 (0,0655)	-960,3476	2.148,6951	2.413,8647
t ( $\nu = 4$ ; ar1)	$\rho_1$	0,4764 (0,0705)	0,3866 (0,0830)	0,4400 (0,0853)	0,4953 (0,0800)	0,4915 (0,0811)	0,5356 (0,0744)	-1.030,1072	2.288,2145	2.553,3841
t ( $\nu = 4$ ; toep)	$\rho_1$	0,4937 (0,0959)	0,4171 (0,0995)	0,4451 (0,0977)	0,5204 (0,0961)	0,4983 (0,0967)	0,5505 (0,0826)	-935,5364	2.111,0727	2.390,1986
	$\rho_2$	0,7275 (0,0617)	0,7007 (0,0668)	0,6444 (0,0805)	0,7372 (0,0658)	0,6205 (0,0954)	0,7285 (0,0602)			
t ( $\nu = 4$ ; un)	$\rho_1$	0,8160 (0,0492)	0,7419 (0,0601)	0,7792 (0,0584)	0,8076 (0,0528)	0,8244 (0,0505)	0,8354 (0,0385)	-163,9420	579,8841	872,9663
	$\rho_2$	0,6696 (0,0720)	0,6346 (0,0857)	0,5817 (0,0938)	0,6666 (0,0789)	0,5770 (0,0968)	0,7044 (0,0683)			
	$\rho_3$	0,1442 (0,1315)	-0,0301 (0,1406)	-0,0439 (0,1392)	0,1060 (0,1450)	0,0206 (0,1433)	0,2049 (0,1267)			

Fonte: autores (2024)

**Tabela 6 – Principais resultados da estimação, via método IFM, dos parâmetros dos modelos de cônjuges SN e ST variantes no tempo. Erro-padrão *bootstrap* entre parênteses**

Côpula	Parâm.	Mês						Log-Veross.	AIC	BIC
		1	2	3	4	5	6			
(ex)	$\rho_1$	0,5300 (0,0666)	0,4547 (0,0727)	0,4515 (0,0716)	0,5578 (0,0695)	0,5196 (0,0726)	0,6194 (0,0541)	-1.002,5007	2.269,0014	2.576,0399
	$\alpha_1$	0,0605 (0,0002)	0,1527 (0,0238)	0,1600 (0,0169)	0,0004 (0,0003)	0,0750 (0,0004)	0,0001 (0,0003)			
	$\alpha_2$	-0,2211 (0,0002)	-0,4578 (0,0178)	-0,4303 (0,0177)	-0,0012 (0,0002)	-0,2636 (0,0003)	1,45e-05 (0,0004)			
	$\alpha_3$	-0,2282 (0,0003)	-0,4942 (0,0144)	-0,4311 (0,0179)	-0,0013 (0,0003)	-0,2744 (0,0003)	0,0003 (0,0004)			
	$\rho_1$	0,4753 (0,0729)	0,4162 (0,0837)	0,5810 (0,0764)	0,4823 (0,0843)	0,4787 (0,0785)	0,5505 (0,0616)			
	$\alpha_1$	0,3308 (0,2089)	0,3377 (0,2505)	0,1865 (0,1910)	0,3556 (0,2119)	0,3277 (0,4597)	-0,3141 (0,1062)			
(ar1)	$\alpha_2$	-0,3588 (0,2196)	-0,4112 (0,2840)	0,6603 (0,3711)	-0,3208 (0,2901)	-0,3511 (0,2055)	0,2815 (0,0953)	-1.033,9611	2.331,9222	2.638,9607
	$\alpha_3$	-1,0605 (1,6624)	-1,1927 (2,6322)	7,1785 (3,0358)	-0,9946 (2,6677)	-1,0422 (2,3105)	0,9050 (1,4133)			
	$\rho_1$	0,4735 (0,0813)	0,3665 (0,1152)	0,3795 (0,1063)	0,4913 (0,0816)	0,4696 (0,0948)	0,5657 (0,0676)			
	$\rho_2$	0,6556 (0,0646)	0,6325 (0,0667)	0,6015 (0,0755)	0,6874 (0,0634)	0,6214 (0,0767)	0,7250 (0,0473)			
	$\alpha_1$	0,0001 (0,0197)	0,1311 (0,0652)	-0,1352 (0,0429)	-0,0001 (0,0292)	-0,0065 (0,0807)	-0,0002 (0,0143)			
	$\alpha_2$	0,0001 (0,0890)	-0,3361 (0,2531)	0,3641 (0,1751)	-0,0003 (0,0668)	-0,0334 (0,0729)	-0,0004 (0,0209)			
(toep)	$\alpha_3$	-1,81e-05 (0,0234)	-0,0496 (0,0623)	0,0998 (0,0424)	-3,03e-05 (0,0256)	-0,0667 (0,1036)	-0,0003 (0,0170)	-986,2583	2.248,5167	2.569,5115
	$\rho_1$	0,7759 (0,0745)	0,7387 (0,0513)	0,7771 (0,0565)	0,7895 (0,0496)	0,8119 (0,0488)	0,8195 (0,0374)			
	$\rho_2$	0,3783 (0,1172)	0,5925 (0,0747)	0,5495 (0,0776)	0,5923 (0,0765)	0,5508 (0,0806)	0,6978 (0,0617)			
	$\rho_3$	-0,2769 (0,1703)	-0,1000 (0,1115)	-0,0962 (0,1153)	-0,244 (0,1168)	-0,0369 (0,1181)	0,1636 (0,1008)			
	$\alpha_1$	0,4183 (0,1206)	-0,0117 (0,0892)	-0,1811 (0,1313)	0,1722 (0,1101)	0,0236 (0,1618)	0,2369 (0,1157)			
	$\alpha_2$	1,5032 (0,4257)	-0,3443 (0,2919)	-0,4635 (0,3027)	0,4389 (0,3128)	0,1329 (0,2375)	-0,1658 (0,2330)			
(un)	$\alpha_3$	-0,8246 (0,3422)	0,4003 (0,3519)	0,2865 (0,3674)	-0,2550 (0,3870)	-0,1508 (0,4369)	0,6644 (0,3673)	67,8073	152,3854	487,3365
	$\rho_1$	0,5948 (0,0827)	0,5266 (0,0750)	0,5205 (0,0873)	0,6002 (0,0744)	0,5478 (0,0859)	0,6128 (0,0705)			
	$\alpha_1$	-0,0224 (6,8200)	-0,0243 (0,1630)	-0,0051 (0,2230)	-0,0389 (0,2071)	-0,0202 (0,1938)	-0,0292 (4,4366)			
	$\alpha_2$	-0,0635 (2,9522)	-0,2081 (0,2653)	0,2299 (0,1868)	-0,1106 (0,4877)	-0,1128 (0,2412)	-0,0063 (0,3118)			
	$\alpha_3$	-0,3426 (3,4347)	-0,1225 (0,1977)	0,3154 (5,2296)	-0,0696 (0,2214)	-0,2698 (0,4185)	0,0070 (0,3205)			
	$\rho_1$	0,4711 (0,0814)	0,3572 (0,0989)	0,4233 (0,0987)	0,4685 (0,0901)	0,4893 (0,0990)	0,4975 (0,0752)			
(ST; $\nu = 4$ ; ex)	$\alpha_1$	0,2319 (0,7852)	0,3100 (6,7186)	-0,2196 (1,9432)	0,2040 (0,2777)	0,1896 (1,7454)	-0,2421 (0,1919)	-956,9801	2.177,9602	2.484,9988
	$\alpha_2$	-0,1529 (0,5381)	-0,3032 (3,4493)	0,3013 (0,5251)	-0,2262 (2,5318)	-0,2040 (2,0762)	0,1544 (0,1288)			
	$\alpha_3$	-0,9794 (6,3070)	-0,6579 (5,2815)	0,9688 (10,0672)	-0,6170 (3,4525)	-0,9596 (8,2273)	0,6393 (2,1818)			
	$\rho_1$	0,5116 (0,0978)	0,4263 (0,1153)	0,4502 (0,1121)	0,5241 (0,1027)	0,5070 (0,0953)	0,5507 (0,0832)			
	$\rho_2$	0,7441 (0,0606)	0,7042 (0,0623)	0,6436 (0,0768)	0,7385 (0,0653)	0,6241 (0,0947)	0,7290 (0,0591)			
	$\alpha_1$	-0,0574 (0,0761)	-0,0585 (0,1079)	0,0012 (0,1338)	-0,0420 (0,0737)	-0,0416 (0,1523)	-0,0340 (0,0636)			
(ST; $\nu = 4$ ; ar1)	$\alpha_2$	0,0411 (0,1517)	-0,1194 (0,2576)	0,1232 (0,2123)	-0,0727 (0,1400)	-0,0550 (0,1352)	-0,0110 (0,1030)	-1.011,0487	2.286,0973	2.593,1358
	$\alpha_3$	-0,2682 (0,0769)	-0,0743 (0,0986)	0,1540 (0,1235)	-0,0436 (0,0736)	-0,2006 (0,1578)	-0,0056 (0,0644)			
	$\rho_1$	0,8300 (0,0535)	0,7661 (0,0780)	0,7633 (0,0709)	0,8162 (0,0535)	0,8092 (0,0590)	0,8306 (0,0436)	-144,0619	576,1238	911,0749
	$\rho_2$	0,6472 (0,0944)	0,6621 (0,1037)	0,5273 (0,1025)	0,6650 (0,0777)	0,5760 (0,1007)	0,6803 (0,0740)			
	$\rho_3$	0,1299 (0,1484)	0,0385 (0,1573)	-0,1362 (0,1478)	0,1182 (0,1426)	-0,0068 (0,1452)	0,1629 (0,1291)			
	$\alpha_1$	-0,2530 (0,2772)	-0,5014 (0,3548)	-0,1855 (0,2170)	-0,1027 (0,1360)	-0,1486 (0,1978)	-0,1099 (0,1082)			
(ST; $\nu = 4$ ; un)	$\alpha_2$	-0,3475 (0,2963)	-0,3894 (0,3343)	-0,5433 (0,2833)	-0,1245 (0,2477)	0,0704 (0,2482)	-0,3162 (0,2148)	-144,0619	576,1238	911,0749
	$\alpha_3$	-0,1068 (0,3205)	-0,2975 (0,3637)	0,3885 (0,2639)	-0,0180 (0,2995)	-0,3697 (0,3436)	0,2169 (0,2462)			

Fonte: autores (2024)

Tabela 7 – Medidas de dependência  $\tau$  de Kendall, pareado e médio, e coeficientes de dependência caudal (pareados) inferior e superior, para os melhores modelos de cópula variante no tempo ajustados - Gaussiana (un) e SN (un)

Cópula	Medida	Mês					
		1	2	3	4	5	6
Gaussiana (un)	$\hat{\tau}_{12}$	0,6013	0,5463	0,5782	0,5755	0,6500	0,6226
	$\hat{\tau}_{13}$	0,4412	0,4224	0,3882	0,4350	0,3990	0,4675
	$\hat{\tau}_{23}$	0,0603	-0,0275	-0,0317	0,0129	0,0503	0,0920
	$\hat{\tau}_s$	0,3676	0,3137	0,3116	0,3412	0,3664	0,3940
	$\hat{\lambda}_{12}^L$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{13}^L$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{23}^L$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{12}^U$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{13}^U$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{23}^U$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
SN (un)	$\hat{\tau}_{12}$	0,3312	0,5169	0,5282	0,4623	0,5954	0,6076
	$\hat{\tau}_{13}$	0,2965	0,3911	0,3770	0,4086	0,3704	0,3262
	$\hat{\tau}_{23}$	-0,0185	-0,0127	-0,0123	0,0242	-0,0153	0,1275
	$\hat{\tau}_s$	0,2031	0,2984	0,2976	0,2984	0,3168	0,3538
	$\hat{\lambda}_{12}^L$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{13}^L$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{23}^L$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{12}^U$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{13}^U$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	$\hat{\lambda}_{23}^U$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Fonte: autores (2022)

Por fim, nas tabelas 8 e 9, são apresentados os resultados dos ajustes dos modelos de cópulas com covariáveis no(s) parâmetro(s) de associação. Observa-se que o melhor modelo encontrado é aquele derivado da cópula ST com  $\nu = 4$  e matriz de dispersão não estruturada (AIC = 933,5663 e BIC = 1.131,4459). Contudo, tal modelo não supera (em termos desses dois critérios de qualidade de ajuste) aqueles selecionados anteriormente (isto é, segundo as outras duas propostas de abordagem multivariada via cópulas apresentadas neste trabalho).

Tabela 8 – Principais resultados da estimação, via método IFM, para os modelos de cônjuges Gaussiana e t ( $\nu = 4$ ) com covariáveis no(s) parâmetro(s) de associação

	Parâmetro	Gaussiana (ex)	Gaussiana (ar1)	Gaussiana (toep)	Gaussiana (un)	t ( $\nu = 4$ ; ex)	t ( $\nu = 4$ ; ar1)	t ( $\nu = 4$ ; toep)	t ( $\nu = 4$ ; un)
$\rho_1$	Intercepto	0,5571	0,5680	0,1251	0,1358	0,6393	0,2358	0,0451	0,0292
	Tempo	-0,4757	-0,4687	-0,6295	-0,6357	-0,3932	-0,5265	-0,4211	-0,8026
	Sexo (Fem.)	0,1130	0,1417	0,1774	-0,1367	0,0645	0,2574	0,0579	-0,4562
	Peso ao nascer	0,0004	0,0005	0,0002	0,0012	0,0002	0,0003	0,0004	-0,0006
	Hemoglobina	0,0615	0,0764	0,1059	0,1660	0,0475	0,0470	0,1596	0,0972
	Comprimento ao nascer	-0,0270	-0,0326	0,0683	-0,2617	0,0301	0,0352	0,0084	0,3103
$\rho_2$	Intercepto	-	-	0,2652	0,2482	-	-	0,0897	0,3076
	Tempo	-	-	-0,6278	-0,1904	-	-	-0,0331	-0,2957
	Sexo (Fem.)	-	-	-0,2048	0,1077	-	-	-0,0069	0,1458
	Peso ao nascer	-	-	0,0003	-0,0002	-	-	-0,0003	0,0010
	Hemoglobina	-	-	0,0798	0,0249	-	-	0,1244	-0,2761
	Comprimento ao nascer	-	-	-0,0327	0,0067	-	-	0,0692	-0,1036
$\rho_3$	Intercepto	-	-	-	0,2480	-	-	-	0,3236
	Tempo	-	-	-	-0,2062	-	-	-	-0,1209
	Sexo (Fem.)	-	-	-	0,4363	-	-	-	0,5548
	Peso ao nascer	-	-	-	-4,66e-05	-	-	-	0,0001
	Hemoglobina	-	-	-	0,1245	-	-	-	-0,1331
	Comprimento ao nascer	-	-	-	-0,0236	-	-	-	0,2397
Log-Veross.		-494,4520	-499,6120	-494,1813	-488,6339	-502,8726	-508,1552	-506,2319	-497,7849
AIC		1.054,9039	1.065,2241	1.066,3627	1.067,2678	1.071,7453	1.082,3105	1.090,4639	1.085,5698
BIC		1.190,9462	1.201,2663	1.227,1399	1.252,7800	1.207,7875	1.218,3528	1.251,2411	1.271,0820

Fonte: autores (2022)

Tabela 9 – Principais resultados da estimação, via método IFM, para os modelos de cônjuges SN e ST com covariáveis no(s) parâmetro(s) de associação

	Parâmetro	SN (ex)	SN (ar1)	SN (toep)	SN (un)	ST ( $\nu = 4$ ; ex)	ST ( $\nu = 4$ ; ar1)	ST ( $\nu = 4$ ; toep)	ST ( $\nu = 4$ ; un)
$\rho_1$	Intercepto	-0,2056	0,2855	0,2098	-1,4992	0,7521	0,5975	-0,1313	-0,3290
	Tempo	-0,9626	-0,4976	-0,4896	-1,6023	-0,6460	-0,3113	-0,4606	-0,8682
	Sexo (Fem.)	-0,3493	0,3720	-0,1232	0,1149	-1,1448	0,3477	0,2795	0,2498
	Peso ao nascer	0,0002	0,0003	0,0005	0,0006	-0,0004	0,0002	0,0008	0,0001
	Hemoglobina	0,1159	0,1018	0,0849	0,0483	0,0922	0,0997	0,0100	0,3028
	Comprimento ao nascer	-0,0337	0,0581	-0,0238	-0,3054	0,0304	0,0950	-0,0417	0,1408
$\rho_2$	Intercepto	-	-	-0,0471	-0,1531	-	-	0,2787	0,2762
	Tempo	-	-	-0,1573	-0,2437	-	-	0,0838	-0,4078
	Sexo (Fem.)	-	-	0,1414	0,4417	-	-	-0,1591	0,7232
	Peso ao nascer	-	-	0,0015	-0,0008	-	-	-0,0001	-0,0004
	Hemoglobina	-	-	0,0573	0,4103	-	-	0,1936	0,1950
	Comprimento ao nascer	-	-	0,0896	0,4903	-	-	0,1130	0,2158
$\rho_3$	Intercepto	-	-	-	0,3069	-	-	-	0,4698
	Tempo	-	-	-	0,1124	-	-	-	-0,2201
	Sexo (Fem.)	-	-	-	0,9487	-	-	-	-0,3911
	Peso ao nascer	-	-	-	-0,0012	-	-	-	-0,0007
	Hemoglobina	-	-	-	0,2957	-	-	-	0,3746
	Comprimento ao nascer	-	-	-	0,4470	-	-	-	0,3335
$\alpha_1$		-0,7963	-2,2037	0,0644	-0,5410	-0,8075	-1,9568	-0,0146	-0,6391
$\alpha_2$		-0,0794	0,1499	-0,0693	-0,0651	0,1007	0,1291	0,0469	0,0368
$\alpha_3$		0,0697	0,4138	0,3261	0,1828	0,1329	0,3878	0,0468	0,2964
Log-Veross.		-466,0851	-475,6855	-504,4435	-453,0006	-474,6855	-478,1814	-503,2456	-418,7831
AIC		1.004,1702	1.023,3709	1.092,8869	1.002,0012	1.021,3709	1.028,3629	1.090,4912	933,5663
BIC		1.152,5799	1.171,7807	1.266,0316	1.199,8809	1.169,7807	1.176,7726	1.263,6359	1.131,4459

Fonte: autores (2022)

## 5 CONSIDERAÇÕES FINAIS

No presente trabalho foram abordados dois procedimentos para lidar com dados longitudinais multivariados. O modelo linear misto (ou MLM) foi a primeira técnica longitudinal apresentada e a mais tradicional delas. Nesta, a dependência entre observações do mesmo indivíduo é introduzida por efeitos aleatórios. Os modelos resultantes dessa abordagem mostraram-se adequados para descrever o comportamento longitudinal individual dos índices antropométricos 1 (peso/idade), 2 (peso/estatura) e 3 (estatura/idade) das crianças nos primeiros seis meses de vida. No entanto, um dos principais interesses foi a modelagem conjunta dos três indicadores em estudo, dada a correlação natural entre os mesmos, além de sua trajetória temporal. Para alcançar tais objetivos, explorou-se outro procedimento baseado em modelagem via funções de cópulas multivariadas (trivariadas, no caso).

A metodologia de cópulas foi introduzida para melhorar a modelagem de dados multivariados por oferecer formas mais flexíveis de se estimar a dependência. Além disso, as cópulas permitem a construção de modelos multivariados com uma variedade de diferentes distribuições marginais. Os parâmetros das distribuições marginais também podem ser estimados separadamente do parâmetro (ou parâmetros, se houver mais de um) da cópula, com pequenos ajustes.

Os MLMs são muito populares na prática, uma vez que são fáceis de manipular e interpretar. A principal limitação deles é a impossibilidade de modelar conjuntamente algumas variáveis ao longo do tempo, tendo também em consideração a correlação entre as variáveis respostas. Embora a estrutura de covariância dos MLMs possa ser adaptada de forma flexível, a dependência sempre permanece Gaussiana por definição. Existem também, por outro lado, algumas limitações quanto ao uso de cópulas. A complexidade adicional pode tornar o procedimento de estimação mais exigente do ponto de vista computacional, resultando numa maior dificuldade em obter estimativas. Além disso, embora existam usos específicos e diretrizes para a escolha entre diferentes tipos de modelos de cópulas, um passo extra de seleção entre os modelos candidatos é necessário quando se utiliza a abordagem proposta. Foi observado que aumentar a complexidade do modelo, analisando cada mês separadamente, se justifica por acarretar ganho adicional, visto que alguns modelos de cópula variante no tempo apresentaram valores de AIC e BIC inferiores

aos do melhor modelo de cópula com parâmetro constante. Ademais, o presente estudo demonstrou um ganho de desempenho quando considerou-se as cópulas elípticas assimétricas (SN, principalmente), ainda pouco exploradas na literatura, no lugar das comumente utilizadas cópulas elípticas simétricas (Gaussiana e t). Finalmente, também foram obtidos resultados promissores com a inclusão de covariáveis no(s) parâmetro(s) de associação da cópula. Visto que um modelo de regressão no(s) parâmetro(s) da cópula pode fornecer discernimento sobre o relacionamento dentro dos dados, dando informações sobre como a associação/dependência entre as variáveis é influenciada pelas covariáveis, pretende-se explorar isto mais a fundo em trabalhos posteriores.

Assim, pesquisas futuras podem ser desenvolvidas para os modelos de cópulas. Como, por exemplo, estratégias que incorporam dependências cruzadas podem ser criadas, numa tentativa de modelar a dependência de todos os valores anteriores de uma variável na outra variável. Além disso, outro desafio para modelar a estrutura de dependência seria como lidar com a falta de valores (*missing data*) sem ser forçado a descartar dados.

## AGREDECIMENTOS

Os autores agradecem aos pareceristas anônimos e aos editores da Revista Ciência e Natura pelos valiosos comentários e sugestões que melhoraram significativamente este artigo. Alexandre Oliveira, Ana Claudia Batista e Marisleane Oliveira agradecem ao colega de mestrado Caio Batalha (*in memoriam*), pelas trocas e companheirismo. Paulo Henrique Ferreira agradece o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, processo nº 307221/2022-9). A pesquisa de Rosemeire L. Fiaccone recebeu apoio da Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB, n. app 0071/2016).

## REFERÊNCIAS

- Akaike, H. (1977). On entropy maximization principle. In Krishnaiah, P.R., Ed. *Application of Statistics*. (pp. 27-41). Amsterdam: Springer-Verlag.
- Arellano-Valle, R., Bolfarine, H., & Lachos, V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science*, 3(4), 415-438.
- Arnau, J., Bono, R., Blanca, M. J., & Bendayan, R. (2012). Using the linear mixed model to analyze nonnormal data distributions in longitudinal designs. *Behavior Research Methods*, 44, 1224-1238.
- Bandyopadhyay, S., Ganguli, B., & Chatterjee, A. (2011). A review of multivariate longitudinal data analysis. *Statistical Methods in Medical Research*, 20(4), 299-330.
- Bortot, P. (2010). *Tail dependence in bivariate skew-normal and skew-t distributions*. Recovered from: [www2.stat.unibo.it/bortot/ricerca/paper-sn-2.pdf](http://www2.stat.unibo.it/bortot/ricerca/paper-sn-2.pdf).
- Brown, H. & Prescott, R. (2015). *Applied Mixed Models in Medicine*. John Wiley & Sons.
- Brunner, E., Domhof, S., & Langer, F. (2002). *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*. John Wiley & Sons.
- Cho, H. (2016). The analysis of multivariate longitudinal data using multivariate marginal models. *Journal of Multivariate Analysis*, 143, 481-491.
- Demidenko, E. (2013). *Mixed Models: Theory and Applications with R*. John Wiley & Sons.
- Diggle, P. J., Heagerty, P., Liang, K.-Y., & Zeger, S. L. (2002). *Analysis of Longitudinal Data*. (2nd ed). Oxford University Press.
- EBA (2014). *Consultation Paper, Draft Regulatory Technical Standards on assessment methodologies for the Advanced Measurement Approaches for operational risk under Article 312 of Regulation (EU) No 575/2013*. Technical report, European Banking Authority.
- Embrechts, P., McNeil, A., & Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls *Risk Management: Value at Risk and Beyond*, 1, 176-223.

- Fausto, M. A., Carneiro, M., Antunes, C. M. d. F., Pinto, J. A., & Colosimo, E. A. (2008). Mixed linear regression model for longitudinal data: application to an unbalanced anthropometric data set. *Cadernos de Saúde Pública*, 24, 513-524.
- Ferreira, C. S., Bolfarine, H., & Lachos, V. H. (2022). Linear mixed models based on skew scale mixtures of normal distributions. *Communications in Statistics - Simulation and Computation*, 51(12), 7194-7214.
- Ferreira, P. H., Fiaccone, R. L., Lordelo, J. S., Sena, S. O., & Duran, V. R. (2019). Bivariate copula-based linear mixed-effects models: An application to longitudinal child growth data. *TEMA (São Carlos)*, 20, 37-59.
- Frahm, G., Junker, M., & Szymayer, A. (2003). Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3), 275-286.
- Frees, E. W., Bolancé, C., Guillen, M., & Valdez, E. A. (2021). Dependence modeling of multivariate longitudinal hybrid insurance data with dropout. *Expert Systems with Applications*, 185, 115552.
- Frees, Edward W & Wang, P. (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics*, 38(2), 360-373.
- Genest, C. & Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4), 347-368.
- Glogger, S. (2015). *Visualization of Trivariate Vine Copulae* (Master's Thesis). Fakultät für Mathematik, Technische Universität München, München, Deutschland.
- Heilpern, S. (2014). Multivariate measures of dependence based on copulas. *Mathematical Economics*, 10(17), 17-32.
- Hofert, M., Kojadinovic, I., Mächler, M., & Yan, J. (2018). *Elements of Copula Modeling with R*. Springer.
- Ida, A., Ishimura, N., & Nakamura, M. (2014). Note on the measures of dependence in terms of copulas. *Procedia Economics and Finance*, 14, 273-279.
- Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall.

- Joe, H. & Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models. *Technical Report 166, Department of Statistics*.
- Kahrari, F., Ferreira, C. S., & Arellano-Valle, R. B. (2019). Skew-normal-cauchy linear mixed models. *Sankhya B*, 81(2), 185-202.
- Kim, S. & Kim, D. (2016). Directional Dependence Analysis Using Skew-Normal Copula-Based Regression. *Statistics and Causality: Methods for Applied Empirical Research*, 131-152.
- Kollo, T., Selart, A., & Visk, H. (2013). From multivariate skewed distributions to copulas. In Bapat, Ravindra B., Kirkland, Steve J., Prasad, K. M. & Puntanen, Simo. (Eds.). *Combinatorial Matrix Theory and Generalized Inverses of Matrices*. (pp. 63-72). Springer.
- Laird, N. M. & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4), 963-974.
- Lambert, Philippe & Vandenhende, F. (2002). A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. *Statistics in Medicine*, 21(21), 3197-3217.
- Li, Feng & Kang, Y. (2018). Improving forecasting performance using covariate-dependent copula models. *International Journal of Forecasting*, 34(3), 456-476.
- Li, Feng & He, Z. (2019). Credit risk clustering in a business group: Which matters more, systematic or idiosyncratic risk?. *Cogent Economics & Finance*, 7(1), 1632528.
- Liu, X. (2016). *Methods and Applications of Longitudinal Data Analysis*. Elsevier.
- Manner, Hans & Reznikova, O. (2012). A survey on time-varying copulas: specification, simulations, and application. *Econometric Reviews*, 31(6), 654-687.
- Marcelino, Sandra Denisen do Rocio & lemma, A. F. (2000). Métodos de estimação de componentes de variância em modelos mistos desbalanceados. *Scientia Agricola*, 57(4), 643-652.
- Molenberghs, Geert & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer.

- Nelsen, R. B. (1996). Nonparametric measures of multivariate association. *Lecture Notes - Monograph Series*, 223-232.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer Science & Business Media.
- Nobre, Juvêncio S & Singer, J. M. (2007). Residual analysis for linear mixed models. *Biometrical Journal*, 49(6), 863-875.
- Nobre, Juvêncio S & Singer, J. M. (2011). Leverage analysis for linear mixed models. *Journal of Applied Statistics*, 38(5), 1063-1072.
- Oliveira, A. T. (2018). *Modelagem Estatística para Dados Financeiros Bivariados Utilizando Técnicas de Análise de Sobrevida e Funções de Cúpula* (Dissertação de mestrado). Departamento de Estatística, Universidade Federal da Bahia, Salvador, BA, Brasil.
- Panelli-Martins, B. E., Santos, S. M. C. d., & Assis, A. M. O. (2008). Segurança alimentar e nutricional: desenvolvimento de indicadores e experimentação em um município da Bahia, Brasil. *Revista de Nutrição*, 31, 65s-81s.
- Patton, A. J. (2009). Copula-based models for financial time series. In Andersen, T. G., Davis, R. A., Kreiß, J. & Mikosch, T. V. *Handbook of Financial Time Series*. (pp. 767-785). Springer.
- Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461-464.
- Schweizer, B. & Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4), 879-885.
- Singer, J. M., Rocha, F. M. M., & Nobre, J. S. (2017). Graphical tools for detecting departures from linear mixed model assumptions and some remedial measures. *International Statistical Review*, 85(2), 290-324.
- Sun, J., Frees, E. W., & Rosenberg, M. A. (2008). Heavy-tailed longitudinal data modeling using copulas. *Insurance: Mathematics and Economics*, 42(2), 817-830.

Team, R. D. C. (2018). *Version 3.5.1*. Vienna, Austria: R Foundation for Statistical Computing. Recovered from: <https://www.r-project.org>.

Tibshirani, Robert J & Efron, B. (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

Vaida, F. & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, 351–370.

Verbeke, G., Fieuws, S., Molenberghs, G., & Davidian, M. (2014). The analysis of multivariate longitudinal data: a review. *Statistical Methods in Medical Research*, 23(1), 42-59.

Wei, Z., Kim, S., & Kim, D. (2016). Multivariate skew normal copula for non-exchangeable dependence. *Procedia Computer Science*, 91, 141-150.

Zhang, Y., Gomes, A. T., Beer, M., Neumann, I., & Nackenhorst, Udo & Kim, C.-W. (2019). Modeling asymmetric dependences among multivariate soil data for the geotechnical analysis - The asymmetric copula approach. *Soils and Foundations*, 59(6), 1960–1979.

Zhang, Y., Kim, C.-W., Beer, M., Dai, H., & Soares, C. G. (2018). Modeling multivariate ocean data using asymmetric copulas. *Coastal Engineering*, 135, 91-111.

Zhou, S. & Chen, Y. (2021). *Industrial Data Analytics for Diagnosis and Prognosis: A Random Effects Modelling Approach*. John Wiley & Sons.

## Contribuições dos autores

### 1 – Alexandre Oliveira

Estatístico, MSc

<https://orcid.org/0000-0001-5708-8152> • [atotelles@hotmail.com](mailto:atotelles@hotmail.com)

Contribuição: Primeira Redação, Investigação, Análise Formal, Visualização de Dados

### 2 – Ana Claudia Batista

Estatística, MSc

<https://orcid.org/0000-0002-8634-5477> • [anacsbatista87@gmail.com](mailto:anacsbatista87@gmail.com)

Contribuição: Primeira Redação, Investigação, Análise Formal, Visualização de Dados

### 3 – Marisleane Oliveira

Estatística, MSc

<https://orcid.org/0000-0002-2543-3545> • [marisleane@ufba.br](mailto:marisleane@ufba.br)

Contribuição: Primeira Redação, Investigação, Análise Formal, Visualização de Dados

**4 – Paulo Henrique Ferreira (Corresponding Author)**

Estatístico, PhD

<https://orcid.org/0000-0001-6312-6098> • paulohenri@ufba.br

Contribuição: Conceituação, Metodologia, Revisão e Edição, Supervisão

**5 – Rosemeire Fiaccone**

Estatística, PhD

<https://orcid.org/0000-0001-5439-1551> • fiaccone@ufba.br

Contribuição: Conceituação, Metodologia, Revisão e Edição, Supervisão

**Como citar este artigo**

Oliveira, A., Batista, A. C., Oliveira, M., Ferreira, P. H., & Fiaccone, R. (2024). Análise de dados longitudinais sobre desenvolvimento infantil por meio de modelos lineares mistos trivariados derivados de cópulas. *Ciência e Natura*, Santa Maria, v.46, e70316. DOI: <https://doi.org/10.5902/2179460X70316>.