

ANÁLISE DA COMPOSIÇÃO CORPORAL VIA MODELO DE REGRESSÃO BETA

Ricardo Rasmussen PETTERLE¹
Thaís Hoffmann JONASSON¹
Adriana Regina NASCIMENTO¹
César Luiz BOGUSZEWSKI¹
Victória Zeghibi Cochenski BORBA²

- RESUMO: O processo de envelhecimento populacional é uma realidade social. Com o avançar da idade surgem mudanças na composição corporal, tais como aumento da massa gorda e redução das massas magra e óssea. Dessa forma, a composição corporal é definida como a proporção entre os diferentes componentes do corpo, sendo expressa pelas percentagens de massas gorda, magra e óssea. Sua análise permite determinar os componentes do corpo, auxiliando em programas de emagrecimento e condicionamento físico. Além disso, sua avaliação constitui um mecanismo importante para detecção e prevenção de algumas doenças crônicas. Nesse contexto, o objetivo deste estudo é analisar e investigar a relação da composição corporal com a idade, gênero, nível de atividade física (IPAQ) e índice de massa corporal (IMC) de indivíduos hígidos avaliados no Serviço de Endocrinologia e Metabologia do Hospital de Clínicas da Universidade Federal do Paraná. Como os percentuais de massas gorda, magra e óssea pertencem ao intervalo unitário (0, 1) assumiu-se distribuição Beta para cada uma delas, fazendo-se uso do modelo de regressão Beta disponível no pacote `betareg` do *software R*. Os resultados mostraram que a composição corporal de homens e mulheres foi diferente e com o avançar da idade ocorreu aumento no percentual de massa gorda e redução nos percentuais de massas magra e óssea, tanto para homens quanto para mulheres. Além disso, o nível de atividade física, estimado pelo IPAQ, mostrou ser um fator determinante na composição corporal, de tal forma que os indivíduos ativos apresentaram maior percentual de massa magra e menor percentual de massa gorda, porém, para o percentual de massa óssea, o IPAQ não foi relevante.

¹Universidade Federal do Paraná - UFPR, Setor de Ciências da Saúde, Curso de Medicina, CEP: 80060-240, Curitiba, PR, Brasil. E-mail: estatisticoufpr@gmail.com; thaisajonasson@gmail.com; adriana.reginanascimento@hotmail.com.br; clbogus@uol.com.br

²Universidade Federal do Paraná - UFPR, Hospital de Clínicas, CEP: 80060-900, Curitiba, PR, Brasil. E-mail: vzborba@gmail.com

■ PALAVRAS-CHAVE: Modelos de regressão; intervalo unitário; IPAQ.

1 Introdução

O envelhecimento populacional está cada vez mais presente na sociedade, o censo de 2010 mostrou um aumento no número de idosos, correspondendo a 10,8% da população. Sendo o aumento mais expressivo na população acima de 80 anos, correspondendo a 1,54% da população (IBGE, 2010). Atualmente, existem no Brasil cerca de 23,5 milhões de idosos e estima-se que até 2050 existam mais idosos do que crianças com até 15 de anos de idade (SECRETARIA DE DIREITOS HUMANOS, 2016). O aumento do número de idosos e da expectativa de vida, levará a maior número de comorbidades, bem como mudanças na composição corporal, que ocorrem principalmente na massa magra, massa gorda e água (CRUZ-JENTOFT *et al.*, 2010). Diante disso, é necessário o acompanhamento progressivo dos indivíduos para um envelhecimento mais saudável e ativo, livre de doenças.

Dessa forma, composição corporal é definida como a proporção entre os diferentes componentes do corpo, sendo expressa pelas percentagens de massas gorda, magra e óssea. A avaliação da composição corporal permite determinar os componentes do corpo humano de forma quantitativa, ou seja, massa magra, massa gorda e óssea. Além disso, mudanças na composição corporal acompanham o envelhecimento com alteração na quantidade e na proporção entre os tecidos (GONÇALVES e MOURÃO, 2008).

A massa muscular reduz aproximadamente 50% entre os 20 e 90 anos de idade (ARNOLD *et al.*, 2010; RECH *et al.*, 2010; SANTOS *et al.*, 2013) e esta redução é a principal responsável pela piora da mobilidade e da capacidade funcional percebidas nos idosos (BUFFA *et al.*, 2011; ORSATTI *et al.*, 2011; PENHA *et al.*, 2012). Grande quantidade de gordura corporal indica obesidade, que é fator de risco para diabetes, doença cardiovascular, câncer e apnéia do sono (AHIMA e LAZAR, 2013). Com o envelhecimento, a distribuição da gordura corporal muda, diminui a gordura apendicular e aumenta a gordura central (SCHAAP *et al.*, 2013).

Ocorre, também, redução da massa óssea, com maior impacto no sexo feminino. Uma mulher, saudável, de 70 anos de idade, chega a apresentar 20% de redução na densidade mineral óssea vertebral e 25 a 40% de redução no colo do fêmur, enquanto nos homens da mesma idade, a redução é de 3% na densidade mineral óssea vertebral e redução de 20 a 30% na densidade do fêmur (GOING, *et al.*, 1995; MATSUDO *et al.*, 2000; MONACO *et al.*, 2011).

O principal objetivo deste estudo é analisar e investigar a relação da composição corporal (percentuais de massa gorda, magra e óssea) com a idade, gênero, nível de atividade física (IPAQ) e índice de massa corporal (IMC) de indivíduos saudáveis. Tais informações são importantes para se estabelecer parâmetros no trato de doenças relacionadas a obesidade e perda progressiva das massas muscular e óssea, principalmente, na população idosa.

Em diversos tipos de estudos é comum investigar o relacionamento entre uma variável resposta e um conjunto de variáveis explicativas, sejam elas numéricas ou categorizadas. Quando a variável resposta é contínua e restrita a um intervalo, geralmente o intervalo unitário $(0, 1)$, é necessário um modelo de regressão adequado para análise dos dados. Assim, é comum o estudo de proporções, taxas, percentuais, dentre outras variáveis com tais características. Diante disso, é inviável o uso do clássico modelo de regressão linear, uma vez que esse modelo não considera dados assimétricos e heterocedásticos. Consequentemente, diversos modelos de regressão foram propostos supondo diferentes distribuições de probabilidades para a variável resposta.

O modelo de regressão Beta foi sugerido por Kieschnick e McCullough (2003), sendo posteriormente apresentado na sua forma reparametrizada por Ferrari e Cribari-Neto (2004). A aplicação do modelo é ampla (SMITHSON e VERKUILEN, 2006; SOUZA e CRIBARI-NETO, 2013; SILVA e SOUZA, 2014 ; OLIVEIRA e SOUZA, 2016) sendo usado em diversas áreas de pesquisa como em agronomia, ciências sociais, medicina dentre outras.

Como os percentuais de massas gorda, magra e óssea pertencem ao intervalo unitário $(0, 1)$ assumiu-se a distribuição Beta para cada uma delas, fazendo-se uso do modelo de regressão Beta disponível no pacote `betareg` (CRIBARI-NETO e ZEILEIS, 2010) do *software* R (R CORE TEAM, 2017).

Assim, o artigo está organizado da forma que segue. A Seção 2 descreve a metodologia de coleta dos dados, bem como a descrição do modelo de regressão Beta, sua especificação, método de estimação dos parâmetros e apresentação das funções de ligação usadas nos modelos. A Seção 3 apresenta os principais resultados obtidos com a análise dos dados. Nesta Seção, primeiro apresentamos a análise exploratória por meio de histogramas, diagramas de dispersão, gráficos box-plot e uma tabela com medidas descritivas e correlações. Na sequência, apresentamos os resultados obtidos com o ajuste dos modelos de regressão Beta para cada variável resposta, seguida pela comparação dos ajustes por meio do critério informação de *Akaike* (AIC), critério de informação Bayesiano (BIC) e valor maximizado do logaritmo da função de verossimilhança (LL). Seguindo a análise, apresentamos os parâmetros estimados pelos modelos, gráficos de resíduos usados para análise de diagnóstico do modelo e curvas de predição com bandas de confiança. Por fim, a Seção 4 apresenta as principais conclusões assim como sugestões para futuros trabalhos.

2 Material e métodos

2.1 Material

Trata-se de um estudo observacional, transversal com voluntários saudáveis recrutados após divulgação da pesquisa no Hospital de Clínicas da UFPR, consultórios, salas de aula e entre familiares.

O estudo foi realizado no Serviço de Endocrinologia e Metabologia do Hospital de Clínicas da Universidade Federal do Paraná (SEMPR). Após informações sobre a pesquisa, os participantes assinaram o termo de consentimento livre e esclarecido

(TCLE), aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do Paraná. Foram incluídos homens e mulheres saudáveis, entre 18 e 90 anos, sem uso de drogas ou derivados hormonais, seja para reposição ou suplementação, com IMC entre 18,5 e 29,9 Kg/m², sem qualquer incapacidade física e que andassem sem ajuda de órteses ou próteses.

Foram excluídos os indivíduos portadores de doenças crônicas e medicamentos ou drogas lícitas ou ilícitas que, sabidamente, afetem a composição corporal, como diabetes insulino-dependente, corticoesteróides, hormônio tireoidiano em doses supressivas e aqueles com baixo peso, condizente com IMC menor que 18,5 Kg/m² ou obesos, com IMC de no mínimo 30 Kg/m². Logo, a amostra utilizada na análise dos dados foi de 298 indivíduos saudáveis.

Todos os participantes realizaram no mesmo dia medidas antropométricas (peso e altura) e responderam ao questionário sobre dados sociodemográficos, seguida do exame de densitometria de corpo total (Aparelho *Lunar Prodigy Advance PA+302284*) para análise das massas gorda, magra e óssea do corpo total. O exame foi avaliado segundo a recomendação da *International Society for Clinical Densitometry* (ISCD) (PETAK *et al.*, 2013; KENDLER *et al.*, 2013).

Todos os participantes responderam ao IPAQ, validado em português (MATSUDO *et al.*, 2001), que é um instrumento para estimar o nível de atividade física praticada habitualmente. Foi utilizado o IPAQ-curto, composto por oito questões sobre realização, frequência e duração de atividades físicas moderadas, vigorosas ou de caminhada. Levando-se em conta, para as respostas, a última semana ou uma semana habitual de exercícios físicos. O IPAQ foi respondido na forma de autoadministração para a maioria dos voluntários ou como entrevista individual, aplicada pelo investigador ou avaliador treinado, nos casos em que houve dificuldade de entendimento. Os voluntários foram, então, divididos em três grupos, conforme o nível de atividade física realizada (NAHAS, 2001; SONATI, 2012): sedentários são aqueles que não realizam nenhuma atividade física por pelo menos 10 minutos contínuos durante a semana; insuficientemente ativos, realizam no mínimo 10 minutos contínuos de atividade física, pelo menos 5 dias na semana ou 150 minutos por semana, porém de maneira insuficiente para serem classificados como ativos. Os ativos são os indivíduos que realizam no mínimo 20 minutos de atividade física vigorosa por sessão, pelo menos 3 vezes na semana ou atividades moderadas, ou caminhada de 30 minutos por sessão, pelo menos 5 vezes na semana ou qualquer atividade somada por 5 dias da semana ou mais, com duração total de 150 minutos por semana (SILVA *et al.*, 2007).

Toda análise estatística dos dados foi feita no *software R* (R CORE TEAM, 2017) com auxílio do pacote *betareg* (CRIBARI-NETO e ZEILEIS, 2010).

2.2 Métodos

2.2.1 Modelo de regressão Beta

O modelo de regressão Beta foi introduzido por Ferrari e Cribari-Neto (2004) com objetivo de modelar variáveis respostas contínuas pertencentes ao intervalo

(0, 1). Esse modelo foi estruturado a partir da reparametrização da densidade Beta que permite modelar diretamente sua resposta média, bem como seu parâmetro de dispersão em função de covariáveis. Assim, a função de densidade Beta na sua forma reparametrizada é expressa por

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad (1)$$

em que μ ($0 < \mu < 1$) é o parâmetro associado a média da variável resposta e $\phi > 0$ o parâmetro de precisão (ou dispersão). Sua esperança e variância são dadas, respectivamente, por $E(Y) = \mu$ e $V(Y) = \frac{\mu(1-\mu)}{1+\phi}$.

Então, seja $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ um vetor de variáveis aleatórias independentes, em que cada variável y_i ($i = 1, 2, \dots, n$) segue a função de densidade Beta (1), isto é, $y_i \sim \mathcal{B}(\mu_i, \phi_i)$, de tal forma que o modelo de regressão Beta é definido por $f(y; \mu, \phi)$ e pelas seguintes relações funcionais

$$g_1(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta},$$

$$g_2(\phi_i) = \xi_i = \mathbf{z}_i^\top \boldsymbol{\gamma},$$

onde $(x_{i1}, x_{i2}, \dots, x_{ik})$ e $(z_{i1}, z_{i2}, \dots, z_{ih})$ são vetores de k e h covariáveis, com $(k + h < n)$ assumidas fixas e conhecidas, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^\top$ e $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_h)^\top$ são vetores de coeficientes associados aos parâmetros de média ($\boldsymbol{\beta} \in \mathbb{R}^k$) e de dispersão ($\boldsymbol{\gamma} \in \mathbb{R}^h$), respectivamente, sendo η_i e ξ_i seus respectivos preditores lineares que são ligados através de uma função monótona e duplamente diferenciável chamada de função de ligação $g_l(\cdot)$ com $l = 1, 2$.

Os estimadores de máxima verossimilhança de μ e ϕ são obtidos pela maximização do logaritmo da função de verossimilhança, a partir de uma amostra com n observações independentes. Logo, sua expressão é dada por

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^n \ell_i(\mu_i, \phi_i),$$

em que

$$\begin{aligned} \ell_i(\mu_i, \phi_i) &= \log \Gamma(\phi_i) - \log \Gamma(\mu_i \phi_i) - \log \Gamma((1 - \mu_i) \phi_i) + (\mu_i \phi_i - 1) \log y_i \\ &\quad + \{(1 - \mu_i) \phi_i - 1\} \log(1 - y_i). \end{aligned}$$

É importante observar que estimadores de máxima verossimilhança de μ e ϕ não possuem forma fechada, eles precisam ser obtidos através da maximização numérica do logaritmo da função de verossimilhança utilizando-se de algum algoritmo numérico de maximização não-linear, como os algoritmos de Newton ou de quasi-Newton; para maiores detalhes, ver Nocedal e Wright (1999).

Para relacionar a média de y ao preditor linear ($\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$) foram consideradas seis relações funcionais (funções de ligação) para $g(\mu_i)$, com as seguintes expressões:

- Logito: $g(\mu_i) = \ln\left(\frac{\mu_i}{1 - \mu_i}\right)$;
- Probit: $g(\mu_i) = \Phi^{-1}(\mu_i)$, onde $\Phi(\cdot)$ denota a função de distribuição da Normal padrão;
- Log: $g(\mu_i) = \ln(\mu_i)$;
- Log-log: $g(\mu_i) = -\ln[-\ln(\mu_i)]$;
- Complemento log-log: $g(\mu_i) = \ln[-\ln(1 - \mu_i)]$;
- Cauchy: $g(\mu_i) = \tan\left(\pi\left(\mu_i - \frac{1}{2}\right)\right)$.

Cada função de ligação, assume o compromisso de linearizar a relação entre os componentes aleatório e sistemático do modelo. Desse modo, a função $g(\mu_i)$ tem como objetivo mapear todo intervalo unitário retornando valores de μ_i para qualquer conjunto de valores das covariáveis, $(0, 1) \mapsto \mathbb{R}$. Cabe ressaltar, que essas seis relações funcionais foram propostas no texto com o intuito de obter o melhor ajuste aos dados, além de ilustrar outras formas de ajuste.

Para relacionar o parâmetro de dispersão (ϕ_i) ao preditor linear ($\xi_i = \mathbf{z}_i^\top \boldsymbol{\gamma}$) foi usada a função de ligação log representada por

$$g(\phi_i) = \ln(\phi_i),$$

em que $\phi_i = \exp(\mathbf{z}_i^\top \boldsymbol{\gamma})$ representa sua respectiva função inversa. A comparação entre os ajustes foi feita através do valor maximizado do logaritmo da função de verossimilhança (LL), uma vez que o melhor ajuste apresenta maior LL. Adicionalmente, considerou-se o critério de informação de *Akaike* (AIC; Akaike, 1974), que os penaliza com covariáveis desnecessárias e o critério de informação Bayesiano (BIC; Schwarz, 1978). Esses critérios são definidos, respectivamente, por:

$$\text{AIC} = -2\ell + 2p$$

$$\text{BIC} = -2\ell + \log(n)p$$

em que ℓ é o logaritmo da função de verossimilhança maximizada e p é o número de parâmetros estimados pelo modelo. Quanto menor forem os valores dos critérios de informação de *Akaike* (AIC) e Bayesiano (BIC) mais parcimoniosos são os modelos.

Para verificar a correta especificação dos modelos selecionados, usamos o teste de especificação RESET (RAMSEY, 1969; PEREIRA e CRIBARI-NETO, 2013). O teste considera as hipóteses H_0 : o modelo está bem especificado *versus* H_1 : o modelo está mal especificado.

3 Resultados

3.1 Análise exploratória

A análise exploratória foi apresentada por meio de histogramas, diagramas de dispersão, gráficos box-plot e uma tabela com medidas descritivas e correlações.

A maior parte da amostra é constituída por mulheres ($n = 150$). Os indivíduos apresentaram média de idade de 46 anos com desvio-padrão de 19,88. A idade dos indivíduos variou entre 18 e 87 anos. O IMC médio foi estimado em 24,72 Kg/m² com desvio padrão de 3,15 Kg/m². Quanto ao IPAQ, 76 indivíduos foram classificados em insuficientemente ativos e 60 indivíduos em sedentários. Por outro lado, 162 indivíduos tiveram uma boa avaliação, foram classificados segundo o questionário IPAQ, em indivíduos ativos.

A Figura 1 apresenta três histogramas relacionados com a distribuição das três variáveis respostas (percentuais de massas gorda, magra e óssea). A partir dessa Figura, observa-se que o gráfico (B) apresenta assimetria à direita, enquanto que os gráficos (A e C) apresentam distribuições um pouco mais simétricas.

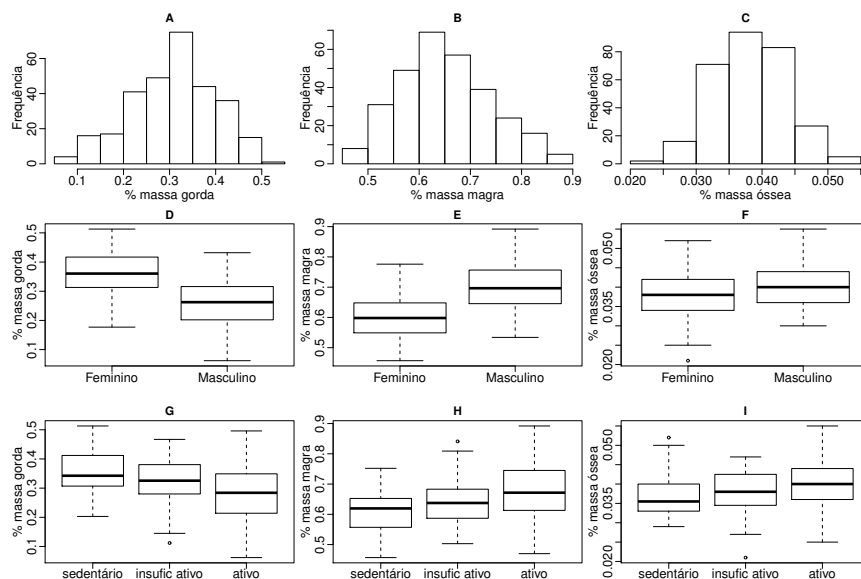


Figura 1 - Histogramas (A até C). Boxplots para sexo (D até F) e IPAQ (G até I) para o percentual de massa gorda, magra e óssea, respectivamente.

Na Figura 1, os gráficos (D até F) mostram as três variáveis respostas em função do sexo (feminino ou masculino). Assim, esses gráficos sugerem possíveis diferenças na composição corporal entre homens e mulheres, de tal forma que as mulheres tendem a apresentar um maior percentual de massa gorda (D) e menor percentual de massas magra (E) e óssea (F), quando comparadas aos homens.

Apesar da diferença no percentual de massa óssea, entre homens e mulheres, ser pequena, visualmente, há indícios de que isso pode ser relevante. Possíveis diferenças também são percebidas no nível de atividade física estimado pelo IPAQ (sedentário, insuficientemente ativo ou ativo), sugerindo, portanto, que os indivíduos ativos são aqueles que apresentam menor percentual de massa gorda (G) e maior percentual de massas magra (H) e óssea (I). A partir disso, tem-se um indicativo de que o nível de atividade física ajuda na redução da gordura corporal, bem como ajuda no aumento das massas magra e óssea.

A Figura 2 apresenta seis gráficos de dispersão (A até F). Esta figura foi construída a partir da combinação das três variáveis respostas com as duas covariáveis contínuas presentes no estudo. A partir disso, sugere-se que os indivíduos mais velhos são aqueles que apresentam maior percentual de massa gorda (A), menor percentual de massa magra (C) e menor percentual de massa óssea (E). Os gráficos (B, D e F) sugerem que o IMC exerce influência nos percentuais de massa gorda, magra e óssea, de tal forma que os indivíduos com um IMC mais elevado são aqueles que apresentam maior percentual de massa gorda (B), menor percentual massa magra (D) e menor percentual massa óssea (F).

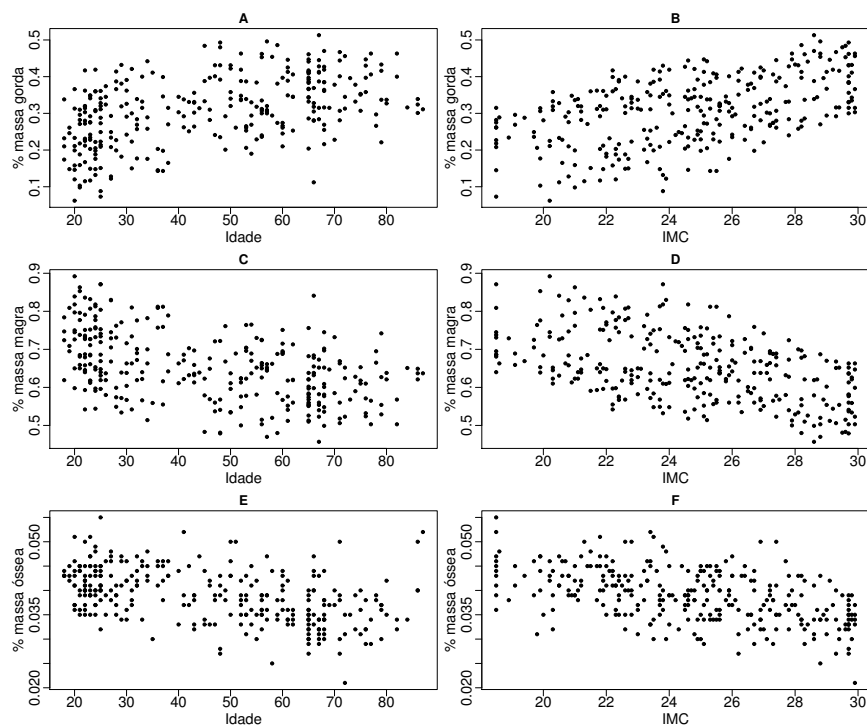


Figura 2 - Gráficos de dispersão (A, C e E) para a idade, IMC (B, D e F) para % de massa gorda, magra e óssea, respectivamente.

Conforme análise exploratória apresentada nas Figuras 1 e 2 tem-se uma idéia das principais e possíveis covariáveis associadas aos percentuais de massas gorda, magra e óssea.

A Tabela 1 apresenta medidas descritivas (mínimo, máximo, média e desvio padrão) para cada variável resposta conforme os níveis das covariáveis sexo e IPAQ, além das correlações entre as variáveis respostas com idade e IMC.

Tabela 1 - Análise descritiva e correlação para cada variável resposta em função da idade e IMC de acordo com os níveis das covariáveis sexo e IPAQ

	% de massa gorda			
	(min-max)	$\bar{x} \pm s$	ρ_{idade}	ρ_{IMC}
Sexo				
Feminino	(17,7-51,3)	36,1 \pm 7,2	0,56	0,81
Masculino	(6,2-43,2)	25,6 \pm 8,1	0,53	0,77
IPAQ				
Sedentário	(20,3-51,3)	35,2 \pm 6,9	0,39	0,46
Insufic. ativo	(11,2-46,7)	32,2 \pm 7,6	0,28	0,50
Ativo	(6,2-49,6)	28,6 \pm 10,0	0,48	0,53
	% de massa magra			
	(min-max)	$\bar{x} \pm s$	ρ_{idade}	ρ_{IMC}
Sexo				
Feminino	(45,7-77,6)	60,2 \pm 6,9	-0,53	-0,79
Masculino	(53,4-89,2)	70,3 \pm 7,8	-0,53	-0,76
IPAQ				
Sedentário	(45,7-75,2)	61,1 \pm 6,6	-0,39	-0,44
Insufic. ativo	(50,3-84,1)	63,9 \pm 7,3	-0,25	-0,47
Ativo	(47,0-89,2)	67,4 \pm 9,7	-0,47	-0,52
	% de massa óssea			
	(min-max)	$\bar{x} \pm s$	ρ_{idade}	ρ_{IMC}
Sexo				
Feminino	(2,1-5,2)	3,8 \pm 0,5	-0,61	-0,61
Masculino	(3,0-5,5)	4,0 \pm 0,4	-0,30	-0,52
IPAQ				
Sedentário	(2,9-5,2)	3,7 \pm 0,5	-0,24*	-0,51
Insufic. ativo	(2,1-4,7)	3,8 \pm 0,5	-0,55	-0,57
Ativo	(2,5-5,5)	4,0 \pm 0,5	-0,39	-0,46

Nota: *Indica p -valor $> 0,05$; ρ : coeficiente de correlação linear de Pearson.

De acordo com os resultados apresentados na Tabela 1, a média do percentual de massa gorda é maior para as mulheres (36,1%) e para os indivíduos classificados em sedentários (35,2%). Para o percentual de massa magra, a média é maior para os homens (70,3%) e para os indivíduos ativos (67,4%). Já a média do percentual de massa óssea não apresenta muita variação entre homens e mulheres e entre as classificações do IPAQ. As correlações entre o percentual de massa gorda com idade e IMC são positivas, enquanto as correlações dessas mesmas covariáveis com os percentuais de massas magra e óssea são negativas. Cabe ressaltar que todas as correlações apresentadas na Tabela 1 são estatisticamente significativas, com

exceção da correlação entre o percentual de massa óssea e idade para indivíduos sedentários.

Nas próximas subseções são apresentados os resultados dos modelos de regressão beta, ajustados para cada variável resposta em função das covariáveis em estudo.

3.2 Resultados do modelo de regressão Beta

Nesta subseção são apresentados os resultados obtidos através dos modelos de regressão Beta. Para cada variável resposta, apresentamos a comparação entre os modelos, parâmetros estimados, análise de resíduos e curvas de predição com bandas de confiança. Primeiro, mostramos os resultados para o percentual de massa gorda, na sequência para o percentual de massa magra e, por fim, os resultados para o percentual de massa óssea.

A seleção das covariáveis, para modelagem da média e dispersão, foi feita separadamente para cada variável resposta por meio de modelos encaixados, considerando-se num primeiro momento a função de ligação logito (*default* do pacote **betareg**). Os resultados dessas análises se encontram no Apêndice. Assim, para o percentual de massa gorda foi selecionado o modelo 9, enquanto para os percentuais de massas magra e óssea foram selecionados os modelos 9 e 7, respectivamente.

Com o intuito de obter o melhor ajuste aos dados, os modelos selecionados foram ajustados com diferentes funções de ligação e comparados via AIC, BIC e LL.

A Tabela 2 apresenta a comparação entre os modelos de regressão Beta na análise do percentual de massa gorda.

Tabela 2 - Critérios de informação de *Akaike* (AIC), Bayesiano (BIC), valor maximizado do logaritmo da função de verossimilhança (LL) e número de parâmetros (NP) de acordo com diferentes funções de ligação para o modelo de regressão Beta na análise do percentual de massa gorda

Função de ligação	AIC	BIC	LL	NP
Logito	-977,07	-943,79	497,53	9
Probit	-981,59	-948,32	499,79	9
Log	-958,44	-925,17	488,22	9
Log-log	-992,55	-959,28	505,28	9
Complementar log-log	-968,50	-935,23	493,25	9
Cauchy	-949,04	-915,76	483,52	9

De acordo com os resultados apresentados na Tabela 2, a função de ligação log-log é a que apresenta os menores valores dos critérios de informação de *Akaike* (AIC = -992,55) e Bayesiano (BIC = -959,28), bem como o maior valor maximizado do logaritmo da função de verossimilhança (LL = 505,28). Dessa forma, o modelo selecionado para análise do percentual de massa gorda tem função de ligação log-log.

Na sequência, a Tabela 3 apresenta as estimativas dos parâmetros, erros-padrão e p -valores referente a modelagem da média e da dispersão do modelo selecionado. Para a modelagem da média, o modelo final conta com o efeito de duas covariáveis contínuas (idade e IMC) e duas covariáveis categóricas (IPAQ e sexo) cujas categorias de referência são: IPAQ-sedentário e sexo feminino. Já a estrutura da dispersão têm o efeito de duas covariáveis: idade e sexo.

Tabela 3 - Estimativas dos parâmetros, erros-padrão e p -valores do modelo de regressão Beta com função de ligação log-log para o percentual de massa gorda

Covariável	Coefficiente	Erro-Padrão	p -valor
β_0 : constante	-1,1376	0,0604	<0,0001
β_1 : idade	0,0015	0,0004	0,0010
β_2 : IMC	0,0458	0,0026	<0,0001
β_{31} : IPAQ (insufic. ativo)	-0,0565	0,0202	0,0051
β_{32} : IPAQ (ativo)	-0,0882	0,0198	<0,0001
β_4 : sexo (masculino)	-0,3365	0,0150	<0,0001
γ_0 : constante	4,0814	0,2251	<0,0001
γ_1 : idade	0,0163	0,0041	<0,0001
γ_2 : sexo (masculino)	-0,4493	0,1632	0,0059

A partir dos coeficientes estimados (Tabela 3) é possível verificar que a idade e o IMC dos indivíduos influenciam positivamente no percentual de massa gorda. Assim, o sinal positivo do coeficiente da covariável idade indica que um acréscimo na idade dos indivíduos corresponde a um aumento no percentual de massa gorda, quando as demais covariáveis são mantidas constantes no modelo. Da mesma forma, o sinal positivo do coeficiente relacionado ao IMC aponta que indivíduos com IMC mais elevado são aqueles que apresentam maior percentual de gordura no corpo, o que já era esperado.

Por outro lado, as covariáveis IPAQ (insufic. ativo e ativo) e sexo (masculino) exercem efeito negativo no percentual de massa gorda. Logo, o percentual de massa gorda dos homens é menor do que o das mulheres. Os indivíduos classificados pelo IPAQ em insuficientemente ativos (insufic. ativo) e ativos (ativo) apresentam menos percentagem de gordura corporal quando comparados aos indivíduos classificados em sedentários. Isto indica que, quanto mais exercício físico o indivíduo pratica, menor seu percentual de massa gorda. Ademais, o sinal positivo do coeficiente γ_1 mostra que os indivíduos mais velhos tendem a apresentar respostas mais dispersas. Por outro lado, quando se considera o sexo dos indivíduos, a dispersão é menor para os homens ($\gamma_2 = -0,4493$).

O pseudo- R^2 do modelo foi de 0,7586. Isso demonstra que 75,86% da variabilidade da variável resposta é explicada pelas covariáveis pertencentes ao modelo final.

Com o intuito de verificar possíveis afastamentos da variável resposta para a

distribuição Beta, a Figura 3 apresenta o gráfico de quantis com envelope simulado (A), gráfico de resíduos ponderados *versus* o número de observações (B) e o gráfico da distância de *Cook versus* o número de observações (C).

Desse modo, o modelo de regressão Beta, com ligação log-log selecionado para explicar o percentual de massa gorda, parece estar bem ajustado aos dados, uma vez que no gráfico (A) os resíduos permanecem dentro das bandas de confiança, no gráfico (B) os resíduos permanecem dentro do intervalo (-2, 2) e no gráfico (C) os valores da distância de *Cook* são pequenos confirmando, portanto, que não há indícios de desvios dos pressupostos feitos para o modelo. Além disso, os teste REST mostrou que o modelo selecionado para análise dos dados está bem especificado (p -valor= 0,0001).

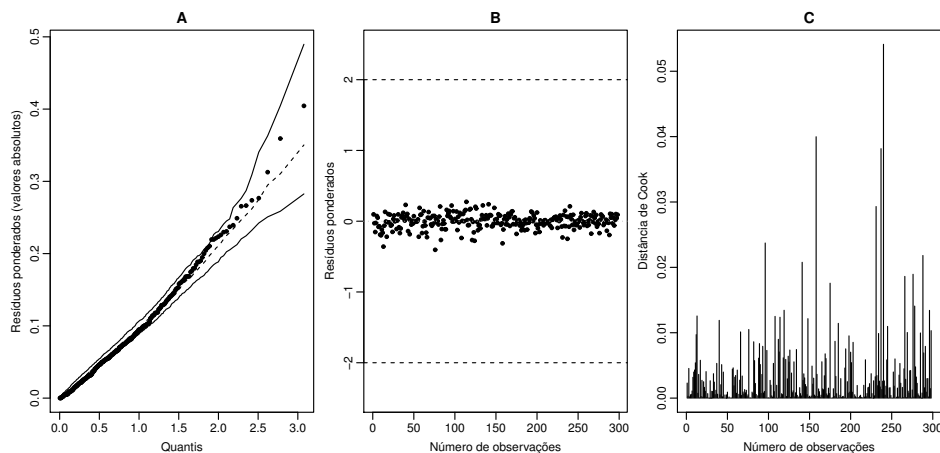


Figura 3 - Gráfico de quantis com envelope simulado (A), gráfico de resíduos ponderados *versus* o número de observações (B) e gráfico da distância de *Cook versus* o número de observações (C) - percentual de massa gorda.

A Figura 4 apresenta curvas de predição com bandas de confiança obtidas pelo modelo selecionado para análise do percentual de massa gorda.

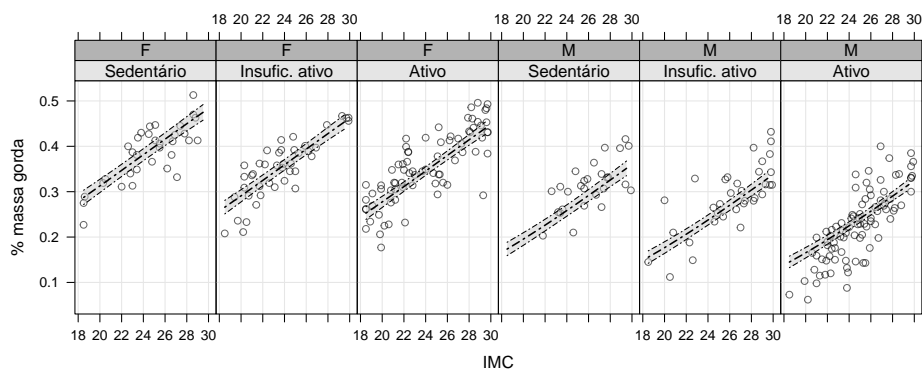


Figura 4 - Gráfico do percentual de massa gorda e intervalos com 95% de confiança obtidos pelo modelo de regressão Beta com ligação log-log em função do IMC, sexo (F-feminino ou M-masculino) e IPAQ (Sedentário, Insuficientemente ativo ou Ativo).

Por meio da Figura 4, observamos que os indivíduos classificados pelo IPAQ em sedentários são aqueles que apresentam mais porcentagem de gordura corporal quando comparados aos indivíduos classificados em insuficientemente ativos e ativos. Além disso, independente da classificação do IPAQ, fica claro pelo gráfico que o percentual de gordura corporal dos homens é menor do que o das mulheres.

A comparação entre os modelos de regressão Beta na análise do percentual de massa magra encontra-se na Tabela 4. A partir disso, pode-se ver que o modelo de regressão Beta com função de ligação log apresenta os menores valores dos critérios de informação de Akaike (AIC = -1014,18) e Bayesiano (BIC = -980,90), assim como o maior valor maximizado do logaritmo da função de verossimilhança (LL = 516,09) sendo, portanto, o modelo selecionado para análise do percentual de massa magra.

Tabela 4 - Critérios de informação de Akaike (AIC), Bayesiano (BIC), valor maximizado do logaritmo da função de verossimilhança (LL) e número de parâmetros (NP) de acordo com diferentes funções de ligação para o modelo de regressão Beta na análise do percentual de massa magra

Função de ligação	AIC	BIC	LL	NP
Logito	-988,54	-955,27	503,27	9
Probit	-991,66	-958,39	504,83	9
Log	-1014,18	-980,90	516,09	9
Log-log	-980,34	-947,07	499,17	9
Complementar log-log	-1001,12	-967,84	509,56	9
Cauchy	-968,30	-935,03	493,15	9

Na Tabela 5 são apresentadas as estimativas dos parâmetros, erros-padrão

e p -valores referente a modelagem de média e dispersão do modelo selecionado. Portanto, o modelo final para modelagem da resposta média conta com o efeito de duas covariáveis contínuas (idade e IMC) e duas covariáveis categóricas (IPAQ e sexo) cujas categorias de referência são: IPAQ-sedentário e sexo feminino. A estrutura de dispersão têm o efeito de uma covariável contínua (idade) e outra categórica (sexo).

Os coeficientes estimados pelo modelo (Tabela 5) indicam que a idade e o IMC dos indivíduos exercem efeito negativo no percentual de massa magra. Desse modo, o sinal negativo do coeficiente β_1 indica que um acréscimo na idade dos indivíduos corresponde a uma redução no percentual de massa magra, para as demais covariáveis fixas no modelo. Ainda, o sinal negativo referente ao coeficiente β_2 aponta que indivíduos com maior IMC são aqueles que apresentam menor percentual de massa magra.

Em contrapartida, as covariáveis IPAQ (insufic. ativo e ativo) e sexo (masculino) exercem efeito positivo no percentual de massa magra, uma vez que os homens apresentam um percentual de massa magra maior do que o das mulheres e os indivíduos classificados pelo IPAQ nas categorias insuficientemente ativos (β_{31}) e ativos (β_{32}) apresentam um percentual de massa magra maior, quando comparados a categoria de referência (IPAQ- sedentário). Isto mostra que a atividade física ajuda na manutenção e/ou no aumento da massa magra de indivíduos saudáveis. Com relação às covariáveis pertencentes à estrutura de dispersão, podemos observar que à medida que a covariável idade aumenta, a dispersão também aumenta. Quando comparamos os homens com as mulheres, estes apresentam respostas menos dispersas.

Tabela 5 - Estimativas dos parâmetros, erros-padrão e p -valores do modelo de regressão Beta com função de ligação log para o percentual de massa magra

Covariável	Coeficiente	Erro-Padrão	p -valor
β_0 : constante	0,0889	0,0318	0,0051
β_1 : idade	-0,0007	0,0002	0,0015
β_2 : IMC	-0,0248	0,0014	<0,0000
β_{31} : IPAQ (insufic. ativo)	0,0311	0,0115	0,0072
β_{32} : IPAQ (ativo)	0,0483	0,0112	<0,0000
β_4 : sexo (masculino)	0,1845	0,0078	<0,0000
γ_0 : constante	4,3787	0,2253	<0,0000
γ_1 : idade	0,0117	0,0041	0,0043
γ_2 : sexo (masculino)	-0,3878	0,1633	0,0017

A Figura 5 apresenta o gráfico de quantis com envelope simulado, gráfico de resíduos ponderados *versus* o número de observações e o gráfico com a distância de Cook *versus* o número de observações.

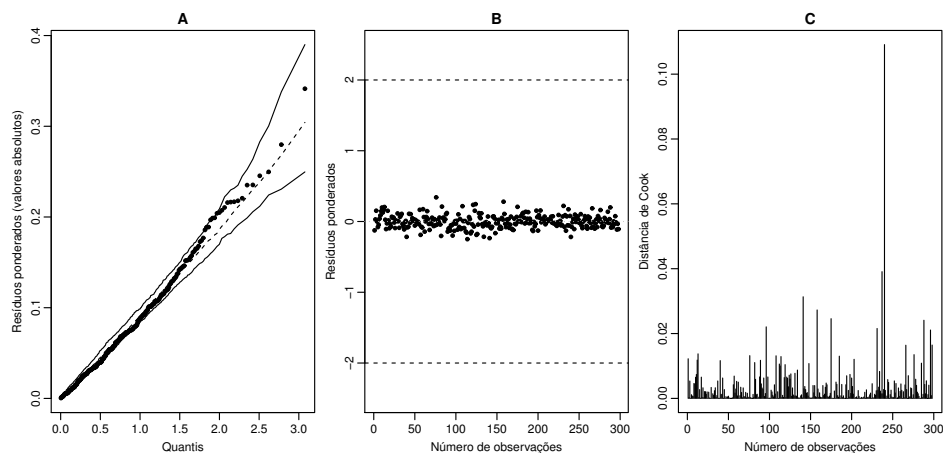


Figura 5 - Gráfico de quantis com envelope simulado (A), gráfico de resíduos ponderados *versus* o número de observações (B) e gráfico da distância de Cook *versus* o número de observações (C) - percentual de massa magra.

A partir da Figura 5, pode-se ver que o modelo de regressão Beta com função de ligação log selecionado para modelagem do percentual de massa magra parece estar bem ajustado, pois os resíduos permanecem dentro das bandas de confiança do envelope simulado e dentro do intervalo (-2, 2). O gráfico (C) mostra apenas uma observação diferente das demais, o que não interfere na qualidade do ajuste. Segundo o teste RESET, o modelo selecionado para análise dos dados está bem especificado (p -valor= 0,024).

Portanto, a variabilidade da variável resposta explicada pelas covariáveis pertencentes ao modelo final, ajustado com função de ligação log, foi de 76,05% (pseudo- R^2).

A Figura 6 apresenta as curvas de predição com intervalos com 95% de confiança para percentual de massa magra.

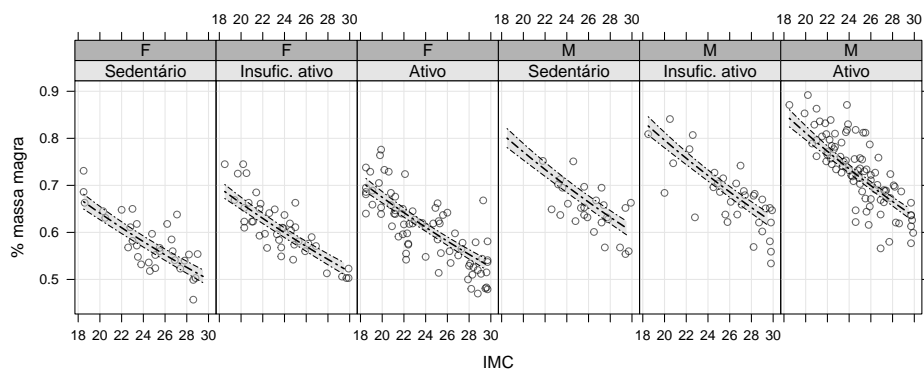


Figura 6 - Gráfico do percentual de massa magra e intervalos com 95% de confiança obtidos pelo modelo de regressão Beta com ligação log em função do IMC, sexo (F-feminino ou M-masculino) e IPAQ (Sedentário, Insuficientemente ativo ou Ativo).

Conforme os resultados apresentados na Figura 6, podemos observar que os indivíduos ativos são aqueles que apresentam o maior percentual de massa magra quando comparados aos indivíduos dos demais grupos. Isso vale tanto para homens quanto para mulheres.

A Tabela 6 apresenta a comparação entre os modelos de regressão Beta usados na análise do percentual de massa óssea. A partir desta tabela, selecionamos o modelo com função de ligação log-log, pois esse modelo apresentou o maior valor maximizado do logaritmo da função de verossimilhança ($LL = 1209,24$), bem como os maiores valores dos critérios de informação de Akaike ($AIC = -2406,48$) e Bayesiano ($BIC = -2384,30$).

Tabela 6 - Critérios de informação de Akaike (AIC), Bayesiano (BIC), valor maximizado do logaritmo da função de verossimilhança (LL) e número de parâmetros (NP) de acordo com diferentes funções de ligação para o modelo de regressão Beta na análise do percentual de massa óssea

Função de ligação	AIC	BIC	LL	NP
Logito	-2405,64	-2383,46	1208,82	6
Probit	-2406,10	-2383,92	1209,05	6
Log	-2405,51	-2383,33	1208,75	6
Log-log	-2406,48	-2384,30	1209,24	6
Complementar log-log	-2405,58	-2383,39	1208,79	6
Cauchy	-2402,07	-2379,89	1207,03	6

Para o percentual de massa óssea, o modelo final para a resposta média, conta com o efeito de duas covariáveis contínuas (idade e IMC) e uma covariável categórica

(sexo) cuja categoria de referência é o sexo feminino. Já a estrutura da dispersão conta com o efeito de apenas uma covariável contínua (idade). Desse modo, as estimativas dos parâmetros, erros-padrão e p -valores referente a modelagem da média e da dispersão encontram-se na Tabela 7.

Tabela 7 - Estimativas dos parâmetros, erros-padrão e valores- p do modelo de regressão Beta com função de ligação log-log para o percentual de massa óssea

Covariável	Coefficiente	Erro- Padrão	p -valor
β_0 : constante	-1,0153	0,0153	<0,0000
β_1 : idade	-0,0005	0,0001	<0,0000
β_2 : IMC	-0,0061	0,0007	<0,0000
β_3 : sexo (masculino)	0,0250	0,0040	<0,0000
γ_0 : constante	8,0707	0,2069	<0,0000
γ_1 : idade	-0,0090	0,0040	0,0287

De acordo com os resultados apresentados na Tabela 7, é possível verificar que a idade e o IMC dos indivíduos exercem efeito negativo no percentual de massa óssea. Logo, o sinal negativo do coeficiente β_1 mostra que um acréscimo na idade dos indivíduos corresponde a uma redução no percentual de massa óssea, isto é, os indivíduos mais velhos apresentam menos massa óssea, considerando as demais covariáveis mantidas fixas no modelo. O sinal negativo do coeficiente β_2 aponta que indivíduos com IMC mais elevado tendem a apresentar menos percentagem de massa óssea. Já os homens, quando comparados às mulheres, tendem a apresentar maior percentual de massa óssea. O sinal negativo do coeficiente γ_1 (Tabela 7), associado ao parâmetro de dispersão, indica que quando aumenta-se a idade dos indivíduos, a dispersão diminui, ou seja, os indivíduos mais velhos apresentam respostas menos dispersas.

A Figura 7 apresenta o gráfico de quantis com envelope simulado (A), gráfico de resíduos ponderados *versus* o número de observações (B) e gráfico com a distância de *Cook versus* o número de observações (C).

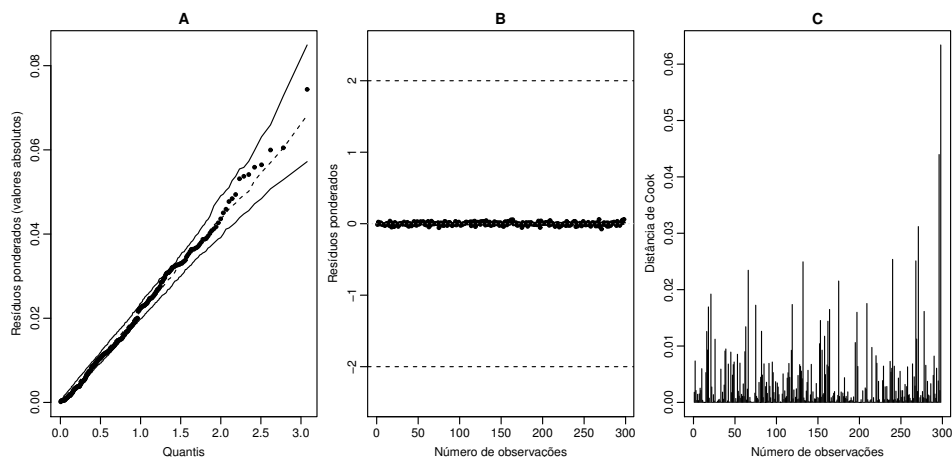


Figura 7 - Gráfico de quantis com envelope simulado (A), gráfico de resíduos ponderados *versus* o número de observações (B) e gráfico da distância de Cook *versus* o número de observações (C) - percentual de massa óssea.

De acordo com a Figura 7, observa-se que o modelo de regressão Beta com função de ligação log-log selecionado para modelagem do percentual de massa óssea parece estar bem ajustado, uma vez que os resíduos permanecem dentro das bandas de confiança do envelope simulado e dentro do intervalo $(-2, 2)$. Porém, o gráfico (C) indica que há pelo menos duas observações distante das demais. Como são poucas observações, acreditamos que isso não afete o ajuste. Os resultados do teste RESET mostraram que o modelo selecionado para análise do percentual de massa óssea não está mal especificado (p -valor= 0,0175).

Logo, as covariáveis pertencentes ao modelo final explicam 40,95% (pseudo- R^2) da variação da variável resposta, indicando que aproximadamente 59% da variação faltante, pode ser explicada por covariáveis não observadas neste estudo.

A Figura 4 apresenta as curvas de predição com bandas de confiança obtidas pelo modelo selecionado para análise do percentual de massa óssea. Conforme a Figura 4, na medida que aumenta o IMC dos indivíduos o percentual de massa óssea diminui gradativamente, tanto para homens quanto para mulheres. Ainda, pela Figura 4 podemos ver que os pontos estão muito dispersos, o que indica uma possível falta de covariáveis para explicar a redução da massa óssea nos indivíduos.

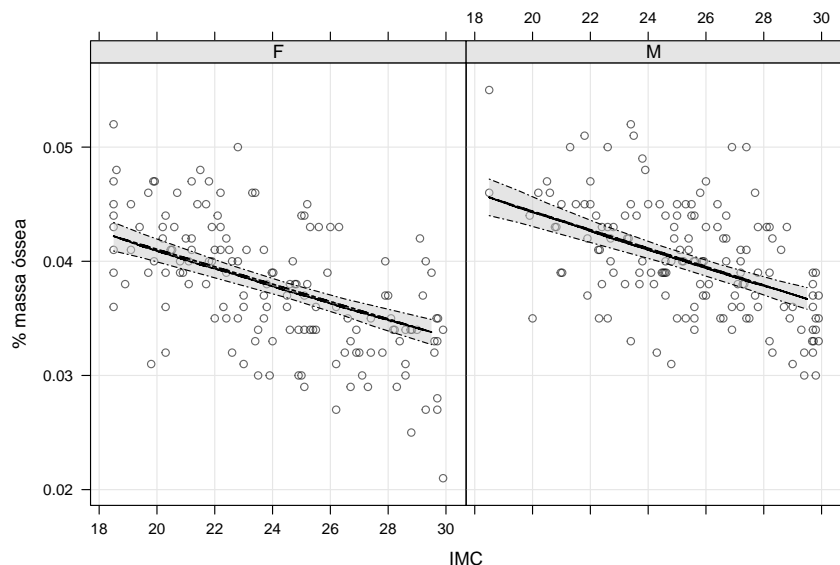


Figura 8 - Gráfico do percentual de massa óssea e intervalos com 95% de confiança obtidos pelo modelo de regressão Beta com ligação log-log em função do IMC e sexo (F-feminino ou M-masculino).

Conclusões

O presente estudo teve como objetivo analisar e investigar o relacionamento da composição corporal com a idade, gênero, nível de atividade física (IPAQ) e IMC de indivíduos saudáveis. A modelagem dos dados foi feita via modelo de regressão Beta, testando-se seis funções de ligação para cada variável resposta. Segundo o critério de informação de *Akaike*, critério de informação Bayesiano e valor maximizado do logaritmo da função de verossimilhança, a função de ligação log-log se mostrou mais adequada para análise dos percentuais de massas gorda e óssea, porém, para análise do percentual de massa magra foi selecionada a função de ligação log. Além disso, o modelo de regressão Beta possibilitou a modelagem dos parâmetros de média e de dispersão em função das covariáveis em estudo.

Dessa forma, a análise dos dados mostrou que a composição corporal de homens e mulheres foi diferente. O avançar da idade indicou aumento no percentual de massa gorda e redução nos percentuais de massas magra e óssea, tanto para homens quanto para mulheres. O nível de atividade física, estimado pelo IPAQ, se mostrou um fator determinante na composição corporal, de tal forma que os indivíduos ativos apresentaram maior percentual de massa magra e menor percentual de massa gorda. Porém, para o percentual de massa óssea, o IPAQ não foi relevante.

De maneira geral, pode-se concluir que o modelo escolhido para cada variável resposta apresentou um bom ajuste, nenhum desvio dos pressupostos foi observado

evidenciando que o modelo de regressão Beta foi adequado para análise dos dados.

Sugerimos como futuros trabalhos realizar um estudo longitudinal para melhor avaliar a composição corporal dos indivíduos e suas possíveis mudanças ao longo do tempo.

Agradecimentos

Os autores agradecem aos revisores e editores pelos comentários e sugestões.

PETTERLE, R. R.; JONASSON, T. H.; NASCIMENTO, A. R.; BOGUSZEWSKI, C. L.; BORBA, V. Z. C. Body composition analysis using Beta regression model. *Rev. Bras. Biom.*, Lavras, v.36, n.2, p.336-359, 2018.

■ **ABSTRACT:** *Population aging is a social reality. With advancing age come changes in body composition, such as increased weight, fat mass and reduced bone mass and lean. Body composition is defined as the ratio between the different components of the body, being expressed by percentages of fat, lean and bone mass. His analysis to determine the components of the body aiding in weight loss programs and fitness. In addition, review is an important mechanism for detection and prevention of some chronic diseases. The main objective of this paper is to analyze and investigate the relationship of body composition with age, gender, level of physical activity (IPAQ) and body mass index (BMI) of healthy individuals evaluated in Endocrinology and Metabolism Department of Clinical Hospital of Federal University of Paraná. As the percentage of fat, lean and bone masses belong to the unit interval (0,1) assumed Beta distribution for each of them, making use of the Beta regression model available in betareg software package R. The results showed that body composition of men and women was different. With advancing age there was an increase in body fat percentage and reduction in the percentage of lean and bone mass, both men and women. The level of physical activity, estimated by the IPAQ, proved to be a determining factor in body composition, such that active individuals had higher percentage of lean mass and a lower percentage of body fat, however, for the bone mass percentage, the IPAQ was not relevant.*

■ **KEYWORDS:** *Regression models; unit interval; IPAQ.*

Referências

AHIMA, R. S.; LAZAR, M. A. The Health Risk of Obesity—Better Metrics Imperative. *Science*. v.341, n.6148, p.856-858, 2013.

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v.19, n.6, p.716-723, 1974.

ARNOLD, A. M.; NEWMAN, A. B.; CUSHMAN, A.; DING, J.; KRITCHEVSKY, S. Body Weight Dynamics And Their Association With Physical Function, And

- Mortality In Older Adults: The Cardiovascular Health Study. *The Journals Of Gerontology. Series A, Biological Sciences And Medical Sciences*, v.65, n.1, p.63-70, 2010.
- BUFFA, R.; FLORIS, G. U.; PUTZU, P. F.; MARINI, E. Body Composition variations in ageing. *Collegium Antropologicum*, v.35, n.1, p.259-265, 2011.
- CRIBARI-NETO, F.; ZEILEIS, A. Beta regression in R. *Journal of Statistical Software*, v.34, n.2, p.1-24, 2010.
- CRUZ-JENTOFT, A. J. *et al.* Sarcopenia: European Consensus On Definition And Diagnosis: Report Of The European Working Group On Sarcopenia In Older People. *Age Ageing*, v.39, n.4, p.412-423, 2010.
- FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modeling rates and proportions. *Journal of Applied Statistics*, v.31, n.7, p.799-815, 2004.
- GOING, S.; WILLIAMS, D.; LOHMAN, T. Aging and Body Composition: Biological Changes and Methodological Issues. In: HOLLOZY, J. O. (Ed.) *Exercise and Sport Sciences Reviews*. Baltimore: Williams & Wilkins, v.23, p.411-449, 1995.
- GONÇALVES, F.; MOURÃO, P. A Avaliação da Composição Corporal - A Medição de Pregas Adiposas como Técnica para a Avaliação da Composição Corporal. *Revista do Desporto e Saúde da Fundação Técnica e Científica do Desporto*, v.4, n.4, p.14-22, 2008.
- IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo Demográfico e Contagem da População: População Residente por Sexo, Situação e Grupos de Idade, 2010*. Disponível Em: <http://censo2010.ibge.gov.br/>. Acesso em: 12 Ago. 2016.
- KENDLER, D. L. *et al.* The Official Positions of the International Society for Clinical Densitometry: Indications of Use and Reporting of DXA for Body Composition. *Journal of Clinical Densitometry*, v.16, n.4, p.496-507, 2013.
- KIESCHNICK, R.; MCCULLOUGH, B. D. Regression analysis of variates observed on (0,1): percentages, proportions and fractions. *Statistical Modelling*, v.3, n.3, p.193-213, 2003.
- MATSUDO, S. M. *et al.* Questionário Internacional de Atividade Física (IPAQ): Estudo de Validade e Reprodutibilidade no Brasil. *Revista Brasileira de Atividade Física & Saude*, v.6, n.2, p.5-18, 2001.
- MATSUDO, S. M.; MATSUDO, V. K. R.; BARROS, T. L. N. Impacto de Envelhecimento nas Variáveis Antropométricas, Neuromotoras e Metabólicas da Aptidão Física. *Revista Brasileira de Ciência e Movimento*, v.8, n.4, p.21-32, 2000.
- MONACO, D. M.; VALLERO, F.; MONACO, D. R.; TAPPERO, R. Prevalence of Sarcopenia and its Association with Osteoporosis in 313 Older Women Following a Hip Fracture. *Archives of Gerontology and Geriatrics*, v.52, n.1, p.71-74, 2011.
- NAHAS, M. V. *Atividade Física, Saúde e Qualidade de Vida: Conceitos e Sugestões para um Estilo de Vida Ativo*. Londrina: Midiograf, 2001.

- NOCEDAL, J.; WRIGHT, S. J. *Numerical Optimization*. New York: Springer-Verlag, 1999.
- OLIVEIRA, A. A.; SOUZA, T. C. Avaliação da Proporção de Crianças Obesas Beneficiadas pelo Programa Bolsa Família nas Regiões do Brasil. *Revista de Ciências Exatas e Naturais*, v.18, n.1, p.55-80, 2016.
- ORSATTI, F. L.; DALANESI, R. C.; MAESTA, N.; NAHAS, E. A. P.; BURINI, R. C. Muscle strength reduction is related to muscle loss in women over the age of 40. *Revista Brasileira de Cineantropometria & Desempenho Humano*, v.13, n.1, p.36-42, 2011.
- PENHA, J. C. L.; PIÇARRO, I. C.; BARROS NETO T. L. Evolution of physical fitness and functional capacity in active elderly women over 50 years of age according to chronological age in Santos city. *Ciência & Saúde Coletiva*, v.17, n.1, p.245-253, 2012.
- PEREIRA, T. L.; CRIBARI-NETO, F. Detecting model misspecification in inflated Beta regressions. *Communications in Statistics - Simulation and Computation*, v.43, n.3, p.631-656, 2013.
- PETAK, S. *et al.* The Official Positions of the International Society for Clinical Densitometry: body composition analysis reporting. *Journal of Clinical Densitometry*, v.16, n.4, p.508-519, 2013.
- RAMSEY, J. B. Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society. Series B*, v.31, n.2, p.350-371, 1969.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.
- RECH, C. R.; CRUZ J. L. S.; ARAÚJO, E. D. S.; KALINOWSKI, F. G.; DELLAGRANA, R. A. Associação Entre Capacidade Funcional e Excesso de Peso em Mulheres Idosas. *Motricidade*, v.6, n.2, p.47-53, 2010.
- SANTOS, V. R. *et al.* Associação entre composição corporal, equilíbrio e mobilidade de idosos com 80 anos ou mais. *Medicina (Ribeirão Preto. Online)*, v.46, n.2, p.135-140, 2013.
- SECRETARIA DE DIREITOS HUMANOS. *Dados sobre o envelhecimento no Brasil*. Disponível em: <http://www.sdh.gov.br/assuntos/pessoa-idosa/dados-estatisticos/DadossobreoenvelhecimentoonoBrasil.pdf>. Acesso em: 10 de fev. 2016.
- SCHAAP, L. A.; KOSTER, A.; VISSER, M. Adiposity, muscle mass, and muscle strength in relation to functional decline in older persons. *Epidemiologic Reviews*, v.35, n.1, p.51-65, 2013.
- SCHWARZ, G. Estimating the dimension of a model. *The annals of statistics*, v.6, n.2, p.461-464, 1978.
- SILVA, G. S. F.; BERGAMASCHINE, R.; ROSA, M.; MELO, C.; MIRANDA, R.; BARA, M. F. Avaliação do Nível de Atividade Física de Estudantes de Graduação

das Áreas Saúde/Biológica. *Revista Brasileira de Medicina do Esporte*, v.13, n.1, p.39-42, 2007.

SILVA, C. R.; SOUZA, T. C. Modelagem da taxa de analfabetismo no estado da Paraíba via modelo de regressão Beta. *Revista Brasileira de Biometria*, v.32, n.3, p.345-359, 2014.

SMITHSON, M. J. ; VERKUILEN, J. A better lemon squeezer? Maximum likelihood regression with Beta-distributed dependent variables. *Psychological Methods*, v.11, n.1, p.54-71, 2006.

SOUZA, T. C.; CRIBARI-NETO, F. Uma estimativa do impacto eleitoral do Programa Bolsa-Família. *Revista Brasileira de Biometria*, v.31, n.1, p.79-103, 2013.

SONATI, J. G. *Qualidade de Vida e Composição Corporal: Características do Envelhecimento Bem Sucedido*. 84 p. Tese (Doutorado em Educação Física) – Universidade Estadual de Campinas, Campinas. 2012.

Recebido em 29.09.2016.

Aprovado após revisão em 30.08.2017.

APÊNDICE

Tabela 8 - Comparação entre modelos encaixados para as porcentagens de massas gorda, magra e óssea ajustados via modelo de regressão Beta com ligação logito

% de massa gorda				
Estrutura da média	Estrutura da dispersão	LL	2(dif. LL)	
M1 : $g(\mu_1) = \beta_0$	$g(\phi_1) = \gamma_0$	280,09		
M2 : $g(\mu_2) = \beta_0 + \beta_1 \text{idade}$	$g(\phi_2) = \gamma_0$	319,74	79,30*	
M3 : $g(\mu_3) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC}$	$g(\phi_3) = \gamma_0$	341,22	42,97*	
M4 : $g(\mu_4) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ}$	$g(\phi_4) = \gamma_0$	345,65	8,85*	
M5 : $g(\mu_5) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_5) = \gamma_0$	484,57	277,84*	
M6 : $g(\mu_6) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_6) = \gamma_0 + \gamma_1 \text{idade}$	493,30	17,46*	
M7 : $g(\mu_7) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_7) = \gamma_0 + \gamma_1 \text{idade} + \gamma_2 \text{IMC}$	493,44	0,28	
M8 : $g(\mu_8) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_8) = \gamma_0 + \gamma_1 \text{idade} + \gamma_{2j} \text{IPAQ}$	493,92	1,24	
†M9 : $g(\mu_9) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_9) = \gamma_0 + \gamma_1 \text{idade} + \gamma_2 \text{sexo}$	497,54	8,47*	
% de massa magra				
M1 : $g(\mu_1) = \beta_0$	$g(\phi_1) = \gamma_0$	293,26		
M2 : $g(\mu_2) = \beta_0 + \beta_1 \text{idade}$	$g(\phi_2) = \gamma_0$	330,60	74,67*	
M3 : $g(\mu_3) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC}$	$g(\phi_3) = \gamma_0$	350,50	39,79*	
M4 : $g(\mu_4) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ}$	$g(\phi_4) = \gamma_0$	354,64	8,29*	
M5 : $g(\mu_5) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_5) = \gamma_0$	494,95	280,62*	
M6 : $g(\mu_6) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_6) = \gamma_0 + \gamma_1 \text{idade}$	500,85	11,79*	
M7 : $g(\mu_7) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_7) = \gamma_0 + \gamma_1 \text{idade} + \gamma_2 \text{IMC}$	500,86	0,006	
M8 : $g(\mu_8) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_8) = \gamma_0 + \gamma_1 \text{idade} + \gamma_{2j} \text{IPAQ}$	501,34	0,96	
†M9 : $g(\mu_9) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_9) = \gamma_0 + \gamma_1 \text{idade} + \gamma_2 \text{sexo}$	503,27	4,84*	
% de massa óssea				
M1 : $g(\mu_1) = \beta_0$	$g(\phi_1) = \gamma_0$	1128,00		
M2 : $g(\mu_2) = \beta_0 + \beta_1 \text{idade}$	$g(\phi_2) = \gamma_0$	1165,50	74,80*	
M3 : $g(\mu_3) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC}$	$g(\phi_3) = \gamma_0$	1185,50	40,21*	
M4 : $g(\mu_4) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ}$	$g(\phi_4) = \gamma_0$	1187,64	2,87	
M5 : $g(\mu_5) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_3 \text{sexo}$	$g(\phi_5) = \gamma_0$	1206,50	41,86*	
M6 : $g(\mu_6) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_{3j} \text{IPAQ} + \beta_4 \text{sexo}$	$g(\phi_6) = \gamma_0$	1207,70	2,47	
†M7 : $g(\mu_7) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_3 \text{sexo}$	$g(\phi_7) = \gamma_0 + \gamma_1 \text{idade}$	1208,80	4,67*	
M8 : $g(\mu_8) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_3 \text{sexo}$	$g(\phi_8) = \gamma_0 + \gamma_1 \text{idade} + \gamma_2 \text{IMC}$	1209,00	0,32	
M9 : $g(\mu_9) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_3 \text{sexo}$	$g(\phi_9) = \gamma_0 + \gamma_1 \text{idade} + \gamma_{2j} \text{IPAQ}$	1210,00	2,37	
M10 : $g(\mu_{10}) = \beta_0 + \beta_1 \text{idade} + \beta_2 \text{IMC} + \beta_3 \text{sexo}$	$g(\phi_{10}) = \gamma_0 + \gamma_1 \text{idade} + \gamma_2 \text{sexo}$	1208,80	0,001	

Nota: *Indica p -valor $< 0,05$; M_i , $i = 1, 2, \dots, 10$ modelos; LL = logaritmo da função de verossimilhança; dif. LL = diferença em LL; † : modelo selecionado.