

# MODELOS DE REGRESSÃO BETA E SIMPLEX NA ANÁLISE DO ÍNDICE DE DESENVOLVIMENTO HUMANO MUNICIPAL DE 2010

André Felipe Berdusco MENEZES<sup>1</sup>  
Wesley Oliveira FURRIEL<sup>1</sup>

- RESUMO: O objetivo deste artigo foi investigar o IDHM de 2010 a partir de alguns indicadores de aspecto socioeconômico e espacial, com intuito de averiguar a existência de relações entre eles. Foram aplicadas técnicas de estatística descritiva para visualização do comportamento dos dados e os modelos de regressão Beta e Simplex, adequados para a modelagem de taxas e proporções, foram aplicados a fim de mensurar e avaliar o impacto das covariáveis no índice de desenvolvimento humano dos municípios brasileiros. Os resultados encontrados permitiram constatar que os modelos apresentaram estimativas e conclusões bastante similares, quando considerado a modelagem da média. Além disso, verificou-se uma considerável conformidade na relação entre os indicadores sociais selecionados e a variável resposta.
- PALAVRAS-CHAVE: Desenvolvimento humano; IDHM; regressão Beta; regressão Simplex.

## 1 Introdução

O primeiro Relatório de Desenvolvimento Humano (RDH) foi publicado em 1990 pelo Programa das Nações Unidas para o Desenvolvimento (PNUD). Idealizado pelo economista Mahbub ul Haq o objetivo era desviar o foco do desenvolvimento da economia para políticas voltadas as pessoas (HAQ, 1995). Nesse âmbito foi lançado também o Índice de Desenvolvimento Humano (IDH) uma medida do grau de desenvolvimento humano de um país. O IDH engloba três requisitos importantes para ampliação da liberdade das pessoas: a oportunidade de se levar uma vida longa e saudável (saúde), ter acesso ao conhecimento (educação) e poder desfrutar de um padrão de vida digno (renda). A criação e popularização do índice contribuiu

---

<sup>1</sup>Universidade Estadual de Maringá - UEM, Departamento de Estatística, CEP: 87020-900, Maringá, PR, Brasil. E-mail: [andrefelipemaringa@gmail.com](mailto:andrefelipemaringa@gmail.com); [wesleyofurriel@gmail.com](mailto:wesleyofurriel@gmail.com)

fortemente para superar a limitada visão que identificava o PIB per capita como medida de desenvolvimento.

O IDH fornece auxílio as autoridades para avaliar o progresso e a realidade social bem como definir prioridades de políticas públicas e comparações de diferentes países. Nesse sentido, em 2012 o PNUD Brasil, o Ipea e a Fundação João Pinheiro criaram o Índice de Desenvolvimento Humano Municipal (IDHM) uma adaptação do IDH que torna possível a identificação, sumarização e hierarquização do desenvolvimento humano entre diferentes grupos da população dentro de um mesmo país. O IDHM foi calculado a partir das informações dos três últimos censos – 1991, 2000, 2010 – e conforme a malha municipal existente em 2010. Portanto, o IDHM ajusta o IDH para a realidade dos municípios e reflete as especificidades e desafios regionais no alcance do desenvolvimento humano no Brasil (BRASIL, 2013). As dimensões são as mesmas do IDH Global – saúde, educação e renda – porém alguns dos indicadores usados são diferentes. A Figura 1 apresenta as variáveis utilizadas para construção do IDHM.

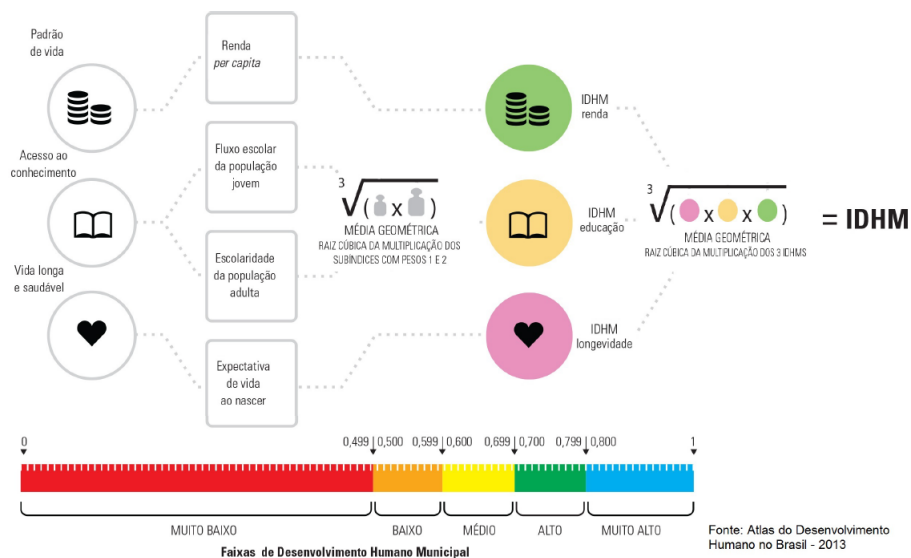


Figura 1 - Representação da construção do Índice de Desenvolvimento Humano Municipal.

O trabalho de Kieling (2014) fornece um panorama sobre o alcance do IDHM, evidenciando seus usos e analisando sua apropriação pela administração pública brasileira. Ao todo são estudadas 14 políticas públicas, em âmbito federal e governamental, realizadas entre 1998 e 2013, para avaliar como o IDHM é utilizado, seja como critério de elegibilidade de municípios para o recebimento de programas, recursos, políticas públicas, financiamentos públicos ou até mesmo parte de elaboração de projetos de lei.

Tendo em vista o panorama apresentado, o presente artigo teve como objetivo investigar a relação do IDHM de 2010 com variáveis de caráter socioeconômico e espacial. No que tange a seleção das variáveis explicativas consideramos indicadores que não são empregados diretamente na construção do IDHM, mas que nos permitiram averiguar sua consonância com demais medidas que buscam expressar aspectos da realidade social, tais como: desigualdade, vulnerabilidade infantil, habitação e região. Para atingir o objetivo desejado foi utilizado o banco de dados disponibilizado pelo *Atlas do Desenvolvimento Humano*, que conta, também, com mais de 200 indicadores socioeconômicos, permitindo qualificar e ampliar a análise do desenvolvimento humano nos municípios brasileiros.

Os modelos de regressão são utilizados para investigar a relação entre a variável resposta e as covariáveis, permitindo também realizar predições a partir da relação estabelecida. Sabendo que o domínio da variável resposta (IDHM) é restrita ao intervalo unitário  $(0, 1)$ , os modelos usuais de regressão, baseados na suposição de erros normais, são considerados inapropriados. Uma solução bastante utilizada consiste em realizar alguma transformação na variável resposta de forma que ela assuma valores no conjunto dos  $\mathbb{R}$ . Entretanto, esta transformação compromete a interpretação dos parâmetros do modelo em termos da resposta original e os valores preditos pelo modelo podem extrapolar os limites do intervalo unitário.

Dessa forma, consideramos os modelo de regressão Beta introduzido por Cepeda-Cuervo (2001) e Ferrari e Cribari-Neto (2004) assim como o modelo de regressão Simplex proposto por Barndorff-Nielsen e Jørgensen (1991). Tais modelos tem sido comumente utilizados para modelar variáveis que assumem valores no intervalo unitário. Por exemplo, Bonat, Ribeiro e Zeviani (2013) compararam os ajustes dos modelos em quatro situações práticas. Por outro lado, Cribari-Neto e Souza (2013) avaliaram a relação entre crença religiosa e inteligência de diferentes países utilizando o modelo de regressão Beta. Outras aplicações dos modelos Beta e Simplex podem ser encontradas em Song e Tan (2000), Kieschnick e McCullough (2003), Hunger, Baumert e Holle (2011), Verkuilen e Smithson (2012), Souza e Cribari-Neto (2013), López (2013), Cepeda-Cuervo e Garrido (2015), Souza e Cribari-Neto (2015), Almeida e Souza (2015), Bonat et al. (2018) e Souza e Cribari-Neto (2018).

Para melhor sistematização e organização este artigo foi dividido nas seguintes seções. A Seção 2 é dedicada a parte metodológica dos modelos de regressão Beta e Simplex e as técnicas de discriminação e diagnóstico utilizadas. Uma análise exploratória dos dados é realizada na Seção 3. Já a Seção 4, apresenta a seleção do modelo e as discussões acerca dos resultados encontrados. Por fim, têm-se algumas considerações finais.

Ressaltamos que embora não apresente nenhuma inovação metodológica a contribuição deste estudo relaciona-se ao fato que, em princípio, nenhum trabalho anterior verificou a relação do IDHM de 2010 com algumas variáveis de caráter socioeconômico e espacial, bem como comparou o ajuste dos modelos de regressão Beta e Simplex.

## 2 Modelos de regressão

### 2.1 Modelo de regressão Beta

O modelo de regressão Beta inicialmente proposto por Cepeda-Cuervo (2001) e mais tarde introduzido de forma independente por Ferrari e Cribari-Neto (2004) consiste em uma nova parametrização da distribuição Beta indexada por sua média e um parâmetro de precisão. A função densidade de probabilidade da distribuição Beta em sua forma reparametrizada é definida por

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

em que  $y, \mu \in (0, 1)$  denota a esperança e  $\phi > 0$  o parâmetro de precisão.

Considerando a reparametrização da distribuição Beta em função de  $\mu$  e  $\phi$  podemos definir o modelo de regressão Beta com modelagem conjunta dos parâmetros média e precisão. De forma geral, assume-se uma amostra aleatória  $Y_i \sim \text{Beta}(\mu_i, \phi_i), i = 1, \dots, n$ , onde ambos os parâmetros de média e precisão seguem a estrutura de regressão definida por

$$g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad (2)$$

$$h(\phi_i) = \mathbf{z}_i^\top \boldsymbol{\gamma}, \quad (3)$$

em que  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^\top$  e  $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_k)^\top$  são, respectivamente, os vetores de parâmetros associados a média e precisão do modelo de regressão,  $\mathbf{x}_i$  e  $\mathbf{z}_i$  são os vetores de covariáveis associados a média e precisão do modelo na  $i$ -ésima observação, respectivamente e  $g(\cdot)$  e  $h(\cdot)$  são funções de ligações apropriadas, duas vezes diferenciáveis em relação aos parâmetros de regressão. Possíveis escolhas para a função de ligação  $g(\cdot)$  são *logit*, *probit* ou *cloglog*. Enquanto que para  $h(\cdot)$  pode-se utilizar as funções logarítmica ou raiz quadrada.

Segue de (1) que o logaritmo da função de verossimilhança é dado por

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^n \{ \log \Gamma(\phi_i) - \log \Gamma(\mu_i \phi_i) - \log \Gamma[(1 - \mu_i) \phi_i] + (\mu_i \phi_i - 1) \log y_i + [(1 - \mu_i) \phi_i - 1] \log(1 - y_i) \}. \quad (4)$$

As estimativas de máxima verossimilhança  $\hat{\boldsymbol{\beta}}$  e  $\hat{\boldsymbol{\gamma}}$  dos parâmetros  $\boldsymbol{\beta}$  e  $\boldsymbol{\gamma}$  do modelo de regressão Beta são obtidas pela maximizando a função definida (4).

### 2.2 Modelo de regressão Simplex

Proposta por Barndorff-Nielsen e Jørgensen (1991) a distribuição Simplex também é parametrizada por sua esperança e um parâmetro extra de dispersão.

Sua função densidade de probabilidade é dada por

$$f(y | \mu, \phi) = [2\pi\phi^2\{y(1-y)\}^3]^{-1/2} \exp\left\{-\frac{1}{2\phi^2} \left[\frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}\right]\right\} \quad (5)$$

em que  $y, \mu \in (0, 1)$  e  $\phi > 0$ .

Na realidade, a distribuição Simplex faz parte dos modelos de dispersão propostos por Jørgensen (1997) e que estendem os modelos lineares generalizados. Assim como no modelo Beta os parâmetros de média e dispersão da regressão Simplex tem estrutura definida pelas equações (2) e (3), respectivamente.

Logo, a função de log-verossimilhança é dada por

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^n \left\{ -\frac{n}{2} \log 2\pi - \frac{1}{2} \{ \log \phi_i^2 + 3 \log [y_i(1-y_i)] \} - \frac{1}{2\phi_i^2} \left[ \frac{(y_i - \mu_i)^2}{y_i(1-y_i)\mu_i^2(1-\mu_i)^2} \right] \right\}. \quad (6)$$

A estimação dos parâmetros,  $\boldsymbol{\beta}$  e  $\boldsymbol{\gamma}$ , do modelo de regressão Simplex é realizada maximizando a função de log-verossimilhança definida em (6).

Uma importante propriedade do modelo de regressão Simplex é que os parâmetros  $\boldsymbol{\beta}$  e  $\boldsymbol{\gamma}$  são ortogonais. A ortogonalidade dos parâmetros fornece muitas vantagens nos resultados inferências, por exemplo, assintoticamente as estimativas de máxima verossimilhança são independentes (COX; REID, 1987).

### 2.3 Medidas para comparação e diagnóstico de modelos

Para discriminar e escolher o melhor entre os modelos propostos, os critérios de informação de Akaike (AIC) (AKAIKE, 1974), Bayesiano (BIC) (SCHWARZ, 1978) e o teste de Vuong foram considerados.

Seja  $\widehat{L}$  o valor máximo da função de verossimilhança do modelo na amostra observada,  $k$  o número de parâmetros do modelo e  $n$  o tamanho da amostra. Os critérios de discriminação de modelo AIC e BIC são definidos, respectivamente, por

$$\text{AIC} = 2k - 2 \log \widehat{L} \quad \text{e} \quad \text{BIC} = \log(n)k - 2 \log \widehat{L}. \quad (7)$$

De acordo com Held e Bové (2014) a regra de decisão, em ambos os critérios, é favorável ao modelo com o menor valor das estatísticas.

O teste introduzido por Vuong (1989), baseado no critério de informação proposto por Kullback e Leibler (1951) é uma generalização do teste da razão de verossimilhanças quando os modelos não são encaixados. A estatística do teste é dada por

$$T_{LR,NN} = \frac{1}{\widehat{\omega}^2 \sqrt{n}} \sum_{i=1}^n \log \frac{f(y_i | \mathbf{x}_i, \widehat{\boldsymbol{\theta}})}{g(y_i | \mathbf{x}_i, \widehat{\boldsymbol{\gamma}})}$$

em que

$$\hat{\omega}^2 = \frac{1}{n} \sum_{i=1}^n \left( \log \frac{f(y_i | \mathbf{x}_i, \hat{\boldsymbol{\theta}})}{g(y_i | \mathbf{x}_i, \hat{\boldsymbol{\gamma}})} \right)^2 - \left( \frac{1}{n} \sum_{i=1}^n \left( \log \frac{f(y_i | \mathbf{x}_i, \hat{\boldsymbol{\theta}})}{g(y_i | \mathbf{x}_i, \hat{\boldsymbol{\gamma}})} \right) \right)^2$$

é um estimador para a variância de  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \log \frac{f(y_i | \mathbf{x}_i, \hat{\boldsymbol{\theta}})}{g(y_i | \mathbf{x}_i, \hat{\boldsymbol{\gamma}})}$ ,  $f(y_i | \mathbf{x}_i, \hat{\boldsymbol{\theta}})$  e  $g(y_i | \mathbf{x}_i, \hat{\boldsymbol{\gamma}})$  são as correspondentes densidades dos modelos de regressão avaliadas nas estimativas de máxima verossimilhança.

Assim, sob a hipótese nula e para  $n \rightarrow \infty$ ,  $T_{LR,NN} \xrightarrow{D} N(0, 1)$ . Logo, rejeita-se  $\mathcal{H}_0$  se  $|T_{LR,NN}| < z_{\alpha/2}$ . Note que,  $f(y_i | \mathbf{x}_i, \hat{\boldsymbol{\theta}})$  será “melhor” (“pior”) do que  $g(y_i | \mathbf{x}_i, \hat{\boldsymbol{\gamma}})$  se  $T_{LR,NN} > z_{\alpha}$  (ou  $T_{LR,NN} < -z_{\alpha}$ ).

No que diz respeito a análise de diagnóstico dos modelos em estudo, os resíduos padronizados ponderados 2 foram empregados. Este tipo de resíduo é baseado no algoritmo iterativo score de Fisher para estimar  $\boldsymbol{\beta}$  quando  $\phi$  é fixo. Para o modelo de regressão Beta este resíduo foi proposto por Espinheira, Ferrari e Cribari-Neto (2008), já sob a suposição de distribuição Simplex foi Miyashiro (2008) que propôs.

## 2.4 Aspectos computacionais

Importante enfatizar que neste trabalho os modelos de regressão Beta e Simplex foram implementados no software SAS. As estimativas e erro padrão dos parâmetros foram obtidas numericamente pelo método **quasi-Newton** disponível na *procedure NLIMIXED* (SAS, 2010). Além disso, vale mencionar que os gráficos foram produzidos no software R utilizando a biblioteca *ggplot2* (WICKHAM, 2016). Os códigos utilizados podem ser obtidos mediante pedido aos autores.

## 3 Análise exploratória dos dados

Os dados analisados sobre o Índice de Desenvolvimento Humano Municipal dos 5565 municípios brasileiros de 2010 tem como fonte o *Atlas do Desenvolvimento Humano* no Brasil <sup>1</sup>. Na Tabela 1 apresentamos a descrição da variável resposta e das covariáveis utilizadas. Consideramos 7 variáveis explicativas que não foram empregados diretamente na construção do IDHM, no entanto, nos permitiram verificar sua consonância com medidas que buscam expressar circunstâncias da realidade social, como desigualdade, vulnerabilidade infantil, habitação e região. As 6 primeiras variáveis apresentadas são de natureza contínua e seus valores de mínimo e máximo são listados na última coluna da Tabela 1. Já a variável região é categórica, representando as cinco regiões do Brasil.

<sup>1</sup><http://atlasbrasil.org.br/2013/pt/>

Tabela 1 - Descrição e característica da variável resposta e das covariáveis utilizadas

Codificação	Descrição	Característica
IDHM (Y)	Índice de Desenvolvimento Humano Municipal	[0.418; 0.8620]
PINDCRI	Proporção de crianças em extrema pobreza	[0; 0.7243]
AGUAESGOTO	Proporção de pessoas em domicílios com abastecimento de água e esgoto inadequados	[0; 0.8536]
TFLSUPER	Taxa de frequência líquida ao ensino superior	[0; 0.4389]
MORT1	Mortalidade até um ano de idade	[0.0849; 0.4680]
DENSIDADE	Densidade demográfica	[0.1310; 13103.71]
GINI	Coefficiente de Gini	[0.2800; 0.8000]
REGIAO	Grandes regiões do Brasil	Centro-Oeste, Nordeste, Norte, Sul e Sudeste

A Figura 2 exibe a distribuição empírica da variável resposta de acordo com as regiões do Brasil. Fica evidente pelo mapa que os municípios com alto IDHM estão localizados nas regiões Sul, Sudeste e Centro-Oeste. Tal fato é reforçado analisando os Box-Plot do IDHM segundo as grandes regiões, em que estas mesmas regiões mostram medianas superiores ao Norte e Nordeste.

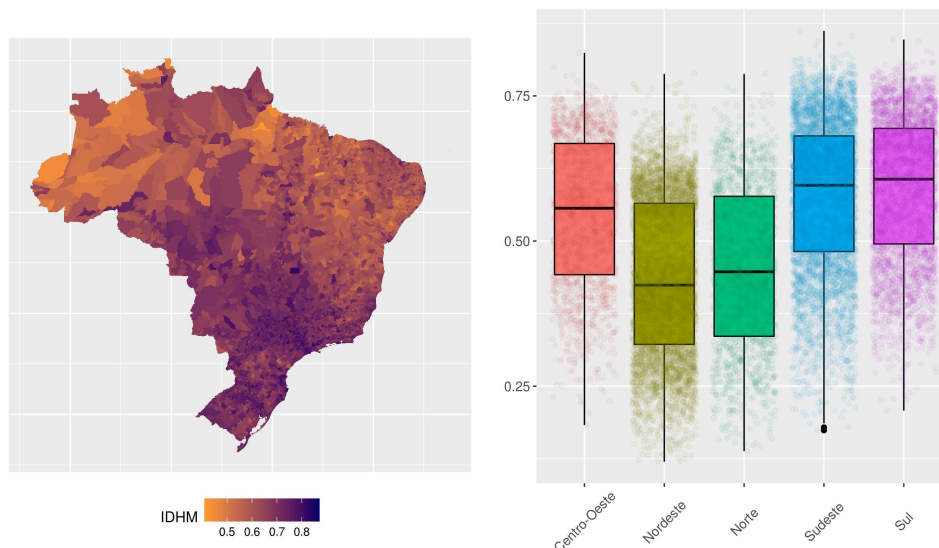


Figura 2 - Distribuição do IDHM de 2010 de acordo com as grandes regiões do Brasil.

Por fim, na Figura 3 observamos a dispersão entre o IDHM e as covariáveis contínuas, segundo as regiões do Brasil. Nota-se que para as covariáveis PINDCRI, AGUAESGOTO, TFLSUPER e MORT1 existem visíveis tendências em sua relação ao IDHM. Por exemplo, verificamos que municípios com maior proporção de crianças

extremamente pobres apresentam menor Índice de Desenvolvimento Humano. Contudo, nas variáveis densidade demográfica e índice de Gini não é possível identificar tendências claras. É perceptível também que a relação entre o IDHM e as covariáveis difere conforme a região.

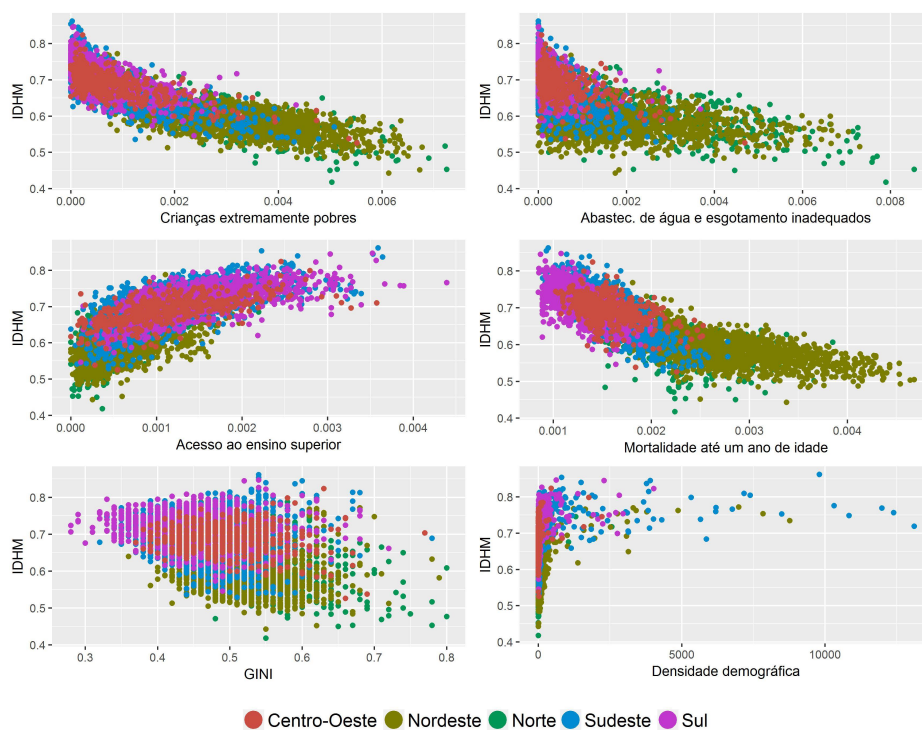


Figura 3 - Comportamento do IDHM de 2010 de acordo com as covariáveis e as regiões do Brasil.

#### 4 Avaliação dos resultados

Na presente seção, apresentamos a modelagem empírica relacionada ao IDHM de 2010 dos municípios brasileiros. Como será visto adiante, os resultados apresentados conduzem a conclusões importantes em termos teórico e prático.

Para ambos os modelos, Beta e Simplex, assumimos a seguinte estrutura de



regressão para os parâmetros  $\mu$  e  $\phi$ :

$$\begin{aligned} \text{logit}(\mu_i) &= \beta_0 + \beta_1\text{PINDCRI}_i + \beta_2\text{AGUAESGOTO}_i + \beta_3\text{TFLSUPER}_i \\ &+ \beta_4\text{MORT1}_i + \beta_5\text{GINI}_i + \beta_6\text{DENSIDADE}_i + \beta_7\text{CENTROESTE}_i \\ &+ \beta_8\text{NORDESTE}_i + \beta_9\text{NORTE}_i + \beta_{10}\text{SUDESTE}_i \end{aligned}$$

$$\begin{aligned} \log(\phi_i) &= \gamma_0 + \gamma_1\text{PINDCRI}_i + \gamma_2\text{AGUAESGOTO}_i + \gamma_3\text{TFLSUPER}_i \\ &+ \gamma_4\text{MORT1}_i + \gamma_5\text{GINI}_i + \gamma_6\text{DENSIDADE}_i + \gamma_7\text{CENTROESTE}_i \\ &+ \gamma_8\text{NORDESTE}_i + \gamma_9\text{NORTE}_i + \gamma_{10}\text{SUDESTE}_i, \end{aligned}$$

$i = 1, \dots, 5565$ .

É importante destacar que em relação as grandes regiões, foi selecionada a região Sul como referência, na construção da matriz de variáveis *dummy* empregada no modelo.

Após a seleção das estruturas de regressão para os parâmetros do modelo realizamos a discriminação entre os modelos utilizando as medidas de AIC, BIC e o teste de Vuong. Na sequência, procedemos com a análise de diagnóstico a fim de avaliar a qualidade do ajuste dos modelos Beta e Simplex.

Dos resultados apresentados na Tabela 2 verificou-se que o modelo de regressão Beta apresentou menores valores para o AIC e BIC. Corroborando com os critérios de informação, a hipótese nula do teste de Vuong, de que não existe diferenças significativa do ajuste entre o modelo Simplex e Beta foi rejeitada, favorecendo a escolha do modelo de regressão Beta.

Tabela 2 - Medidas de discriminação entre modelos Beta e Simplex

Modelo	$-2 \times \log L(\hat{\theta}   y_i)$	AIC	BIC	Vuong
Beta	-25366.0944	-25322.0944	-25176.3648	3.9156 (<0.0001)
Simplex	-25345.4891	-25301.4891	-25155.7595	

Na Figura 4 pode-se observar que ambos os modelos não apresentaram afastamentos das especificações impostas, com exceção das caudas, em geral nota-se que os percentis da Normal e os resíduos se aproximam da reta com inclinação um e intercepto zero, indicando assim que os modelos empregados podem ser utilizados para descrever os dados.

Após a comparação entre os modelos e análise de diagnósticos, serão apresentadas as principais interpretações dos parâmetros estimados pelos modelos. A Tabela 3 apresenta as estimativas, os erros padrão assintóticos e o valor- $p$  do teste de significância para os coeficientes dos modelos. É preciso ressaltar que, o parâmetro  $\gamma$ , foi estimado em escalas distintas, tendo em vista a parametrização dos modelos. Desse modo, no caso do modelo Beta temos as estimativas para o parâmetro de precisão e na Simplex do parâmetro de dispersão.

Conforme os resultados expostos na Tabela 3, no que tange a modelagem das médias, é possível constatar que ambos os modelos levam as mesmas conclusões,

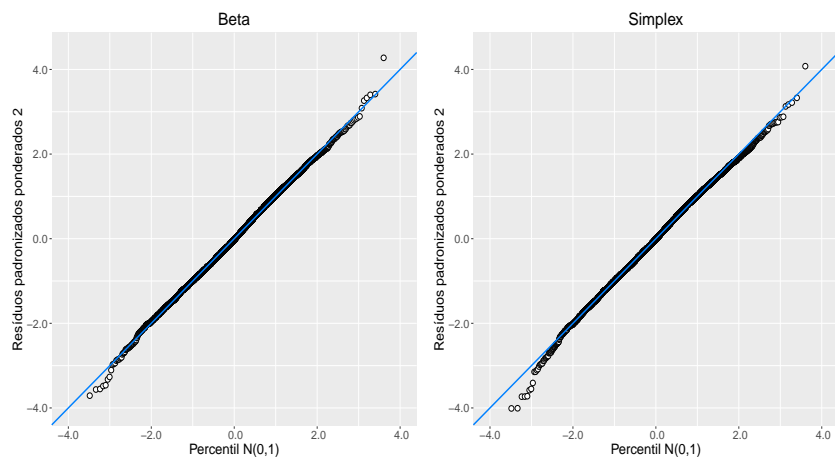


Figura 4 - Gráficos dos resíduos padronizados ponderados 2 versus percentil da Normal padrão.

uma vez que as estimativas dos coeficientes são próximas. Das variáveis explicativas selecionadas apenas a estimativa para a região Nordeste ( $\beta_8$ ) não foi estatisticamente significativa, considerando os usuais níveis de significância. Além disso, verifica-se que municípios com alta proporção de crianças em extrema pobreza tendem a ser menos desenvolvidos. Essa mesma relação também foi identificada para municípios com alta proporção de pessoas em domicílios com abastecimento de água e esgoto inadequados e elevada mortalidade infantil (até um ano de idade). Os municípios que apresentam alto índice de Gini, portanto possuem maior desigualdade na distribuição de renda, tendem a apresentar maiores Índices de Desenvolvimento Humano. Verificamos também que as covariáveis TFLSUPER e DENSIDADE influenciam positivamente o desenvolvimento dos municípios, ou seja, podemos concluir que as cidades com maiores densidades demográficas e com alta taxa de frequência líquida ao ensino superior apresentam maiores Índices de Desenvolvimento Humano. Em relação as regiões brasileira, constatamos que quando comparadas ao Sul, com exceção da região Sudeste todas as outras apresentam IDHM inferiores.

No que diz respeito aos parâmetros de precisão e dispersão, não é possível compará-los diretamente, uma vez que estão em escalas distintas, sendo eles inversos. Todavia, é possível comparar as conclusões acerca da significância dos parâmetros, ou seja, a rejeição da hipótese  $\mathcal{H}_0 : \gamma_k = 0$ . Desse modo, verificamos que ocorreram diferenças na significância dos parâmetros  $\gamma_1$  e  $\gamma_3$  entre os modelos Beta e Simplex. Nestes casos, nota-se que o primeiro foi estatisticamente significativo para o modelo Beta, ao passo que, o segundo se apresentou significativo para o modelo Simplex. Destaca-se que os demais parâmetros demonstraram conclusões similares em termos de significância. Ressaltamos que as estimativas positivas no

Tabela 3 - Estimativas, erro padrão e valor- $p$  referentes aos parâmetros dos modelos Beta e Simplex

Parâmetro	Beta			Simplex		
	Estimativas	E.P.	Valor- $p$	Estimativas	E.P.	Valor- $p$
$\beta_0$	0.6521	0.0166	< 0.0001	0.6529	0.0167	< 0.0001
$\beta_1$ (PINDCRI)	-0.9652	0.0204	< 0.0001	-0.9595	0.0204	< 0.0001
$\beta_2$ (AGUAESGOTO)	-0.0440	0.0179	0.0139	-0.0452	0.0177	0.0108
$\beta_3$ (TFLSUPER)	1.5580	0.0348	< 0.0001	1.5598	0.0349	< 0.0001
$\beta_4$ (MORT1)	-1.2393	0.0382	< 0.0001	-1.2300	0.0381	< 0.0001
$\beta_5$ (GINI)	0.5221	0.0316	< 0.0001	0.5204	0.0319	< 0.0001
$\beta_6$ (DENSIDADE)	0.0002	0.0001	< 0.0001	0.0002	0.0001	< 0.0001
$\beta_7$ (CENTROESTE)	-0.0150	0.0060	0.0116	-0.0168	0.0060	0.0048
$\beta_8$ (NORDESTE)	-0.0031	0.0076	0.6870	-0.0063	0.0076	0.4055
$\beta_9$ (NORTE)	-0.0389	0.0084	< 0.0001	-0.0417	0.0084	< 0.0001
$\beta_{10}$ (SUDESTE)	0.0236	0.0051	< 0.0001	0.0232	0.0051	< 0.0001
$\gamma_0$	5.6114	0.2020	< 0.0001	-2.8961	0.2008	< 0.0001
$\gamma_1$ (PINDCRI)	-0.9179	0.2665	0.0006	0.3018	0.2665	0.2575
$\gamma_2$ (AGUAESGOTO)	-1.2337	0.2235	< 0.0001	1.2646	0.2232	< 0.0001
$\gamma_3$ (TFLSUPER)	-0.2795	0.3944	0.4786	1.6881	0.4016	< 0.0001
$\gamma_4$ (MORT1)	1.2107	0.5371	0.0242	-1.4696	0.5369	0.0062
$\gamma_5$ (GINI)	0.4149	0.3891	0.2863	0.3197	0.3902	0.4126
$\gamma_6$ (DENSIDADE)	-0.0003	0.0001	< 0.0001	0.0004	0.0001	< 0.0001
$\gamma_7$ (CENTROESTE)	0.4466	0.0821	< 0.0001	-0.4943	0.0826	< 0.0001
$\gamma_8$ (NORDESTE)	0.5009	0.0960	< 0.0001	-0.5685	0.0970	< 0.0001
$\gamma_9$ (NORTE)	0.3197	0.1123	0.0044	-0.4046	0.1127	0.0003
$\gamma_{10}$ (SUDESTE)	-0.2746	0.0571	< 0.0001	0.2797	0.0573	< 0.0001

caso do modelo Beta, sugerem que elevações nos valores da covariável aumentam a precisão do IDHM, por outro lado, no modelo Simplex a interpretação é inversa, devido sua parametrização. Para os casos em que as estimativas foram negativas a interpretação ocorre de modo análogo.

Por exemplo, considerando o modelo Beta verificamos que conforme a covariável MORT1 aumenta, a precisão também aumenta, ou seja, os municípios com maior mortalidade infantil apresentam respostas mais precisas. Em contrapartida, os municípios com alta proporção de crianças em extrema pobreza e com alta proporção de pessoas em domicílios com abastecimento de água e esgoto

inadequados, tendem a apresentar respostas menos precisas.

É importante salientar que a relação entre os indicadores socioeconômicos e o IDHM se mostrou consonante, ou seja, indicadores que apontaram para melhores condições sociais e econômicas, se mostraram positivamente relacionados ao IDHM.

## Considerações finais

Considerando a importância do IDHM como indicador social para diagnósticos rápidos acerca do desenvolvimento dos municípios brasileiros, o que ocorre em virtude da sua natureza, isto é, restrito ao intervalo  $(0, 1)$ , os modelos de regressão Beta e Simplex, adequados para determinados fenômenos, foram utilizados a fim de compreender a relação do IDHM e alguns indicadores socioeconômicos e espacial.

No que refere-se a modelagem das médias constatamos que ambos os modelos fornecem as mesmas conclusões. Salientamos que os municípios com maior proporção de crianças em extrema pobreza influenciam negativamente o desenvolvimento. Tal relação também foi observada no que tange as variáveis abastecimento de água e esgotamento sanitário inadequados e mortalidade infantil. Por outro lado, a taxa de frequência líquida ao ensino superior, o coeficiente de Gini e a densidade demográfica do município influenciam positivamente o desenvolvimento humano. Em contrapartida, para os parâmetros de precisão e dispersão as conclusões acerca da significância estatística não foram equivalentes, especificamente para as covariáveis proporção de crianças extremamente pobres e taxa de frequência líquida ao ensino superior.

Vale destacar que o ajuste de ambos os modelos foi satisfatório, entretanto o modelo de regressão Beta obteve vantagem, considerando os critérios de informação AIC e BIC e o teste de Vuong. Além disso, verificamos que as conclusões supracitadas na análise descritiva previamente conduzida foram reforçadas pelos resultados encontrados na modelagem via regressão.

A dependência espacial foi incorporada seguindo as abordagens disponíveis na literatura e apresentada em Cepeda-Cuervo et al. (2012) e Souza and Cribari-Neto (2013). Contudo, não observamos ganhos significativos nos modelos. Por esta razão não incluímos no artigo.

## Agradecimentos

Aos revisores e editores pelas valiosas sugestões que melhoraram substancialmente a qualidade do artigo.

MENEZES, A. F. B.; FURRIEL, W. O. *Beta and simplex regression models in the analysis of the municipal human development index 2010*, Lavras, v.37, n.3, p.394-408, 2019.

■ **ABSTRACT:** *The goal of this paper was to investigate the MHDI of 2010 from some indicators of socioeconomic and spacial, in order to find out the existence of relationships between them. To this end, descriptive statistics techniques were applied to visualize the data behavior and the Beta and Simplex regression models, which are appropriate to modeling rates and proportions, were employed in order to measure and evaluate the impact of covariates on the human development index of Brazilian cities. The results showed that the models presented estimates and conclusions quite similar, in respect of the mean regression structure. Furthermore, it was observed a considerable accordance among the social indicators selected and the response variable.*

■ **KEYWORDS:** *Human development; MHDI; Beta regression; Simplex regression.*

## Referências

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, 1974. ISSN 0018-9286.

ALMEIDA, P. M.; SOUZA, T. C. Estimates of votes for Dilma Rouseff in 2010 elections under the scope of the Bolsa Família Program. *Ciencia & Natura*, v. 37, n. 1, p. 12–22, 2015.

BARNDORFF-NIELSEN, O.; JØRGENSEN, B. Some parametric models on the Simplex. *Journal of Multivariate Analysis*, v. 39, n. 1, p. 106–116, 1991. ISSN 0047-259X.

BONAT, W. H. et al. Likelihood analysis for a class of Simplex mixed models. *Chilean Journal of Statistics*, v. 9, n. 2, p. 3–17, 2018.

BONAT, W. H.; RIBEIRO, P. J.; ZEVIANI, W. M. Regression models for responses in the unit interval: Specication, estimation and comparison. *Rev. Bras. Biom. (São Paulo)*, v. 20, n. 1, p. 1–10, 2013.

BRASIL. *Índice de Desenvolvimento Humano Municipal Brasileiro*. [S.l.], 2013.

CEPEDA-CUERVO, E. *Variability modeling in generalized linear models*. Tese (Doutorado) — Mathematics Institute, Universidade Federal do Rio de Janeiro, 2001.

CEPEDA-CUERVO, E.; GARRIDO, L. Bayesian Beta regression models with joint mean and dispersion modeling. *Monte Carlo Methods and Applications*, v. 21, n. 1, p. 49–58, 2015.

COX, D. R.; REID, N. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 49, n. 1, p. 1–39, 1987.

CRIBARI-NETO, F.; SOUZA, T. C. Religious belief and intelligence: Worldwide evidence. *Intelligence*, v. 41, n. 5, p. 482–489, 2013. ISSN 0160-2896.

ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. On Beta regression residuals. *Journal of Applied Statistics*, Taylor & Francis, v. 35, n. 4, p. 407–419, 2008.

FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, v. 31, n. 7, p. 799–815, 2004.

HAQ, M. *Reflections on Human Development*. [S.l.]: Oxford University Press, 1995.

HELD, L.; BOVÉ, D. S. *Applied Statistical Inference Likelihood and Bayes*. [S.l.]: Springer, 2014.

HUNGER, M.; BAUMERT, J.; HOLLE, R. Analysis of SF-6D index data: Is Beta regression appropriate? *Value in Health*, v. 14, n. 5, p. 759–767, 2011.

JØRGENSEN, B. *The Theory of Dispersion Models*. [S.l.]: Chapman & Hall/CRC, 1997.

KIELING, L. M. *O Índice de Desenvolvimento Humano: Adaptações metodológicas e práticas no Brasil*. 2014. Monografia, Universidade Federal de Santa Catarina. Disponível em: <https://repositorio.ufsc.br/xmlui/handle/123456789/124693>.

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.

KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 22, n. 1, p. 79–86, 1951. ISSN 00034851.

LÓPEZ, F. O. A bayesian approach to parameter estimation in Simplex regression model: A comparison with Beta regression. *Revista Colombiana de Estadística*, v. 36, n. 1, p. 1–21, 2013.

MIYASHIRO, E. S. *Modelos de regressão Beta e Simplex para a análise de proporções*. Dissertação (Mestrado) — Universidade de São Paulo - USP, 2008.

SAS. *The NLMIXED Procedure, SAS/STAT® User's Guide, Version 9.4*. Cary, NC: SAS Institute Inc.: [s.n.], 2010. 4967–5062 p.

SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, The Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.

SONG, P. X.-K.; TAN, M. Marginal models for longitudinal continuous proportional data. *Biometrics*, v. 56, n. 2, p. 496–502, 2000.

SOUZA, T. C.; CRIBARI-NETO, F. An estimate of the impact of assistance programs on the 2006 presidential election in Brazil. *Rev. Bras. Biom., São Paulo*, v. 31, n. 1, p. 79–103, 2013.

SOUZA, T. C.; CRIBARI-NETO, F. Intelligence, religiosity and homosexuality non-acceptance: Empirical evidence. *Intelligence*, v. 52, p. 63–70, 2015. ISSN 0160-2896.

SOUZA, T. C.; CRIBARI-NETO, F. Intelligence and religious disbelief in the United States. *Intelligence*, v. 68, p. 48–57, 2018. ISSN 0160-2896.

VERKUILEN, J.; SMITHSON, M. Mixed and mixture regression models for continuous bounded responses using the Beta distribution. *Journal of Educational and Behavioral Statistics*, v. 37, n. 1, p. 82–113, 2012.

VUONG, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, [Wiley, Econometric Society], v. 57, n. 2, p. 307–333, 1989.

WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <http://ggplot2.org>.

Recebido em 13.10.2018.

Aprovado após revisão em 20.03.2019.