

Estimativas de votos da presidente Dilma Rousseff nas eleições presidenciais de 2010 sob o âmbito do Bolsa Família

Estimates of votes for Dilma Rousseff in 2010 elections under the scope of the Bolsa Família Program

Pedro M. de Almeida Junior*¹ e Tatiene C. Souza²

¹Departamento de Estatística, Universidade Federal de Pernambuco, Cidade Universitária, Recife/PE, 50740–540, Brasil

²Departamento de Estatística, Universidade Federal da Paraíba, Cidade Universitária, João Pessoa/PB, 58089–900, Brasil

Resumo

O objetivo deste artigo é avaliar o impacto dos gastos com o programa Bolsa Família nas eleições presidenciais de 2010. Ajustou-se o modelo de regressão beta para explicar o percentual de votos válidos da presidente Dilma Rousseff no segundo turno nas eleições de 2010 no Nordeste. Alguns fatores como o percentual de pobres, o PIB municipal, o percentual de votos de Lula em 2006 e o gasto per capita com o programa Bolsa Família exerceram efeito positivo em relação ao percentual de votos da candidata Dilma nas eleições de 2010. Com o impacto do Bolsa Família nas eleições de 2010, verificou-se que se não houvesse gasto com o Bolsa Família, a votação da candidata Dilma teria uma redução de aproximadamente 2.125 milhões de votos na região Nordeste nas eleições de 2010.

Palavras-chave: Bolsa Família, percentual, região Nordeste, modelo de regressão beta.

Abstract

The purpose of this paper is to evaluate the impact of the Bolsa Família program's expenses during the presidential elections of 2010. The beta regression model was adjusted in order to explain the percentage of valid votes from the Northeast region in Dilma Rousseff during the second round of the 2010 elections. Factors such as the poverty ratio, the municipal GDP, the percentage of votes Lula got in 2006 as well as the Bolsa Família program's per capita spending all had a positive impact on the percentage of votes in Dilma in the 2010 elections. We've established the impact of the Bolsa Família program in the 2010 elections: had the program being given no budget during the 2010 elections, President Dilma would have lost approximately 2,125 million votes from the Northeast region.

Keywords: Bolsa Família program, percentage, region Northeast, beta regression model .

*e-mail: pedroallmeida@gmail.com

1 Introdução

Os programas assistenciais ou também conhecidos como programas de transferência de renda, tem sido apontado como possibilidade de solução para a crise do emprego e enfrentamento da pobreza, sendo defendidos por políticos, organizações sociais e estudiosos das questões sociais (Silva et al., 2012). Assim, estes programas tem como objetivos principais a redistribuição de renda para tornar a sociedade mais igualitária no aspecto econômico e social, além de permitir a inserção social e profissional dos cidadãos, numa conjuntura de pobreza e desemprego.

No Brasil, a primeira proposta para a criação de um programa nacional de assistencialismo, foi introduzida pelo senador Eduardo Suplicy, do Partido dos Trabalhadores de São Paulo. O Programa de Garantia de Renda Mínima (PGRM), foi aprovado no Senado Federal, em 1991 (Silva et al., 2012). Este foi o marco inicial dos programas de assistencialismo no Brasil. Apesar das críticas e algumas limitações, foi de suma importância para o início da discussão em relação as políticas de redistribuição de renda. A partir daí, surgiram novas propostas de programas com o objetivo de erradicar a pobreza extrema e diminuir a desigualdade social.

As cidades de Campinas (SP), Ribeirão Preto (SP), Santos (SP) e Brasília (DF), foram as pioneiras na implantação de programas assistenciais no Brasil em 1995. Destes, o Programa de Garantia de Renda Familiar Mínima (PGRF) de Campinas e o Programa Bolsa Escola de Brasília destacaram-se como modelos de programas de transferência de renda (Silva et al., 2012). No governo do ex-presidente Fernando Henrique Cardoso, no penúltimo ano de seu segundo mandato (1999-2002), tem-se em destaque a transformação do Programa de Garantia de Renda Mínima, em Programa Nacional de Renda Mínima vinculado à educação chamado Bolsa Escola, a criação do Programa Bolsa Alimentação e Auxílio Gás. No Governo do ex-presidente Lula estes programas foram unificados em um só programa intitulado como Bolsa Família (Presidência da República, 2004).

O Cadastro Único (CadÚnico) é um instrumento que possibilita conhecer a realidade socioeconômica das famílias, possibilitando identificar as famílias de baixa renda existente no país, trazendo informações de todo o núcleo familiar, das características do domicílio e das formas de acesso a serviços públicos essenciais (Ministério de Desenvolvimento Social e Combate à Fome, 2010a). O Bolsa Família é executado pelos municípios e cabe a eles realizarem o cadastramento das famílias, através do CadÚnico. A seleção das famílias que vão ingressar no programa é feita pelo Ministério do Desenvolvimento Social e Combate a Fome (MDS), ver Almeida (2011).

O Bolsa Família (BF) é um programa de transferência

de renda com condicionalidades¹ que destina benefícios monetário para famílias em situações de vulnerabilidade social. O programa Bolsa Família oferece às famílias três tipos de benefícios, são eles: básico, variável e variável para jovem. O benefício básico é permitido às famílias em situação de extrema pobreza. Os benefícios variável e variável para jovem são destinados às famílias pobres e extremamente pobres, contudo, famílias que apresentam em sua composição gestante, nutrizes, crianças ou adolescentes com até 15 anos de idade recebem o benefício variável, enquanto que as famílias que tenham adolescentes entre 16 e 17 anos de idade recebem o benefício chamado variável para jovem.

Nas eleições presidenciais em 2006, o ex-presidente Luiz Inácio Lula da Silva, filiado ao PT, destacou-se pelo seu desempenho eleitoral, principalmente nas regiões mais pobres. O eleitorado de Lula havia sofrido alteração, em 2006 os votos recebidos migraram para lugares menos desenvolvidos e urbanizados (Canêdo-Pinheiro, 2009). Existem evidências que esta migração de votos foi causada pelo programa de assistencialismo Bolsa Família, ver Zucco (2008), Carraro et al. (2007) e Souza e Cribari-Neto (2013).

A região do Nordeste é o segundo maior sítio eleitoral no Brasil (Tribunal Superior Eleitoral, 2010b). No ano de 2010, 22% do eleitorado nordestino era beneficiado pelo Bolsa Família e mais da metade das famílias beneficiadas pelo programa são da região Nordeste totalizando 6.385.327 milhões de famílias beneficiadas (Marques, 2013), importante citar que o Nordeste é a região com maior gasto per capita com o programa Bolsa Família. Estes fatores contribuíram para a escolha da região Nordeste como unidade de observação do nosso estudo.

A candidata Dilma Rousseff filiada ao Partido dos Trabalhadores (PT) foi ministra de minas e energia em 2002, e ministra chefe da casa civil em 2005. Foi escolhida pelo PT para ser a candidata à presidência nas eleições de 2010. Uma das principais propostas eleitorais feita por Dilma na campanha das eleições de 2010 foi a ampliação de programas sociais, dentre eles se destacam, o Bolsa Família, Minha Casa Minha Vida e ProUni.

O resultado da eleição de 2010 na região Nordeste, em favor a candidata Dilma foi muito bom, dado que venceu em todos os Estados. Assim, seria crucial analisar o desempenho da candidata Dilma nessa região do Brasil. Como em eleições passadas, também existem

¹As condicionalidades são em relação à saúde e à educação. No tocante à saúde, as famílias devem participar de ações de acompanhamento, cuidar da frequência escolar das crianças e adolescentes e participar de ações de educação alimentar, saúde e do estado nutricional dos filhos; com relação à educação, os responsáveis legais devem matricular e acompanhar. Em caso de descumprimento destas condicionalidades o benefício de toda família pode ser bloqueado ou até cancelado, ver Ministério do Desenvolvimento Social e Combate à Fome (2004).

evidências que nas eleições de 2010, os programas assistências contribuíram para a vitória da presidente Dilma Rousseff (Zucco, 2013).

Esta pesquisa tem por objetivo ajustar um modelo de regressão beta proposto por Ferrari e Cribari-Neto (2004), com a finalidade de identificar os fatores que influenciaram o percentual de votos válidos recebidos pela presidente Dilma Rousseff nas eleições presidenciais de 2010 na região Nordeste e adicionalmente quantificar o impacto dos gastos com o programa de assistencialismo Bolsa Família nas eleições presidenciais de 2010 na região Nordeste.

O presente artigo encontra-se dividido em cinco seções. A Seção 2 apresenta o modelo de regressão beta, bem como algumas medidas de avaliação da qualidade do ajuste. Na Seção 3, foi descrito os dados utilizados. Na Seção 4, foi especificado o modelo de regressão beta considerado. Por fim, as conclusões estão apresentadas na Seção 5.

2 Modelo de regressão beta

O modelo de regressão beta, introduzido por Ferrari e Cribari-Neto (2004) é útil para a modelagem de variável resposta Y , que assume valores no intervalo $(0,1)$, ver Cribari-Neto e Souza (2012), Simas et al. (2010), Smithson e Verkuilen (2006), Espinheira et al. (2008b) e Espinheira et al. (2008a).

Uma variável aleatória Y segue uma distribuição beta com parâmetros $\alpha, \beta > 0$ denotado por $Y \sim B(\alpha, \beta)$, se a distribuição de Y tem função de densidade:

$$f(y; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1}, \quad 0 < y < 1,$$

sendo $\Gamma(\cdot)$ a função gama. A média e a variância de Y são dadas por:

$$E(Y) = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Ferrari e Cribari-Neto (2004) propuseram uma parametrização alternativa para a densidade beta que permite a modelagem da média da variável resposta utilizando uma estrutura de regressão e que envolve também um parâmetro de precisão.

Seja $\mu = \frac{\alpha}{\alpha + \beta}$ e $\phi = \alpha + \beta$, i.e, $\alpha = \mu\phi$ e $\beta = (1 - \mu)\phi$, a função de densidade beta de acordo com esta reparametrização tem a forma

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

em que $0 < \mu < 1$, $0 < \phi < 1$ e $\phi > 0$. Aqui, $E(Y) = \mu$ e $\text{Var}(Y) = \frac{\mu(1-\mu)}{1+\phi}$, sendo $V(\mu) = \mu(1-\mu)$ a “função de variância”, em que μ é a média da variável resposta e ϕ pode ser interpretado como um parâmetro de precisão.

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes, em que cada Y_t , $t = 1, \dots, n$, possui densidade (1), a média μ_t e o parâmetro de precisão ϕ_t sendo desconhecidos. O modelo de regressão beta (Ferrari e Cribari-Neto, 2004) é obtido assumindo que a média de Y_t pode ser escrita como

$$g(\mu_t) = \sum_{i=1}^k x_{ti}\beta_i = \eta_t,$$

em que $\beta = (\beta_1, \dots, \beta_k)^\top$ é um vetor de parâmetros desconhecidos ($\beta \in \mathbb{R}^k$), x_{t1}, \dots, x_{tk} são observações de k covariáveis e η_t é um preditor linear. Por fim, $g(\cdot)$, a função de ligação $g : (0,1) \rightarrow \mathbb{R}$ é estritamente monótona e duas vezes diferenciável. Assim, temos que $\mu_t = g^{-1}(\eta_t)$ e $\text{Var}(Y_t) = \mu_t \frac{(1-\mu_t)}{(1+\phi)}$. Admitimos que o parâmetro de precisão ϕ_t é dado por

$$h(\phi_t) = \sum_{j=1}^q z_{tj}\gamma_j = \nu_t,$$

em que $\gamma = (\gamma_1, \dots, \gamma_q)^\top$ é um vetor de parâmetros desconhecidos, ν_t é um preditor linear e z_{t1}, \dots, z_{tq} são observações de q covariáveis ($q + k < n$), assumidas fixas e conhecidas, e $h : (0,\infty) \rightarrow \mathbb{R}$ é uma função estritamente monótona e duas vezes diferenciável. Há várias possíveis escolhas para as funções de ligação $g(\cdot)$ e $h(\cdot)$, por exemplo, pode-se utilizar as funções logito e log, entre outras. Para maiores detalhes sobre estas funções de ligação, ver McCullagh e Nelder (1989).

No modelo de regressão beta, pode-se considerar o método de máxima verossimilhança, em que os parâmetros são estimados através da maximização da função de verossimilhança. O método da máxima verossimilhança (MV) proposto por Fisher (1922), que consiste em determinar os valores dos parâmetros que maximizam a função de verossimilhança dada por:

$$L(\beta) = \prod_{t=1}^n f(y_t|\beta),$$

em que β é um vetor de parâmetros a ser estimado, y_t ($t = 1, 2, \dots, n$) são valores observados da variável resposta Y e $f(y_t|\theta)$ é a função densidade (probabilidade) proposta para a variável aleatória Y .

O logaritmo da função de verossimilhança para a distribuição beta, a partir de uma amostra de n observações

independentes é

$$\ell(\beta, \gamma) = \sum_{t=1}^n \ell_t(\mu_t, \phi_t),$$

em que

$$\begin{aligned} \ell_t(\mu_t, \phi_t) &= \log \Gamma(\phi_t) - \log \Gamma(\mu_t \phi_t) \\ &\quad - \log \Gamma((1 - \mu_t)\phi_t) + (\mu_t \phi_t - 1) \log y_t \\ &\quad + \{(1 - \mu_t)\phi_t - 1\} \log(1 - y_t). \end{aligned}$$

É possível utilizar testes assintóticos para fazer inferência dos parâmetros desconhecidos. Considere, a hipótese nula, $\mathcal{H}_0 : \beta = \beta^{(0)}$ versus a hipótese alternativa $\mathcal{H}_1 : \beta \neq \beta^{(0)}$, em que, $\beta = (\beta_1, \dots, \beta_m)^\top$ e $\beta^{(0)} = (\beta_1^{(0)}, \dots, \beta_m^{(0)})^\top$ para $m < k$. A estatística da razão de verossimilhanças é:

$$\bar{w}_1 = 2\{l(\hat{\beta}, \hat{\phi}) - l(\tilde{\beta}, \tilde{\phi})\},$$

sendo $l(\hat{\beta}, \hat{\phi})$ o logaritmo natural da função de verossimilhança e $l(\tilde{\beta}, \tilde{\phi})$ o estimador de máxima verossimilhança restrito de (β, ϕ) obtido pela imposição da hipótese nula. Sob condições gerais e de regularidade e sob \mathcal{H}_0 $\bar{w}_1 \xrightarrow{D} \chi_m^2$. Assim, o teste pode ser avaliado usando os valores críticos aproximados da distribuição χ_m^2 , em que m é a dimensão do espaço paramétrico sob a hipótese nula.

2.1 Medidas de diagnóstico e adequabilidade para modelos de regressão beta

Uma das etapas importante de qualquer ajuste no modelo de regressão são as análises de diagnóstico e resíduo, com o objetivo de detectar pontos mal ajustados ou aberrantes, bem como a existência de observações extremas (*outliers*) que podem causar inconsistência no resultado final do ajuste, e verificar possíveis afastamentos das suposições sobre o modelo. As análises de resíduos pode basear-se nos resíduos ordinários ou em suas padronizações. Para o modelo de regressão beta, Ferrari e Cribari-Neto (2004) propuseram uma medida global de qualidade de ajuste, o resíduo componente do desvio. Espinheira et al. (2008b) propuseram dois novos tipos de resíduos para regressão beta, são elas: os resíduos ponderados e os resíduos padronizados ponderados. Vale salientar que esses dois tipos de resíduos possuem um desempenho superior ao resíduo proposto por Ferrari e Cribari-Neto (2004), especialmente no sentido de identificar observações influentes para as estimativas das médias, ver Rocha e Simas (2011).

Para avaliar a qualidade global do ajuste de um modelo de regressão geralmente utiliza-se o coeficiente de determinação, R^2 , bastante utilizado nos modelos de regressão linear. Ferrari e Cribari-Neto (2004) propuse-

ram uma medida global chamada de pseudo R^2 , que é definida como o quadrado do coeficiente de correlação entre $\hat{\eta}$ e $g(y)$. Na evolução dos métodos de diagnóstico uma etapa que se mostrou relevante foi a detecção de observações que exercem efeito desproporcional no ajuste. Neste contexto, encontram-se a distância de Cook (Cook, 1977) e a alavancagem generalizada (Wei e Fung, 1998).

Na identificação de pontos que exercem influência em modelos de regressão, uma ferramenta chave é a alavancagem. É medida pelos elementos h_{tt} da matriz H que é conhecida como matriz de projeção ou matriz chapéu e é usada para avaliar a importância individual de cada observação no próprio valor ajustado. Nos modelos lineares de regressão, é comum utilizar o elemento h_{tt} como medida da influência da t -ésima observação sobre os valores ajustados. Uma opção é verificar as observações tais que $h_{tt} \geq \frac{3k}{n}$, em que k é o número de parâmetros e n é o número de observações, estes pontos são conhecidos como pontos de alavancagem. A alavancagem generalizada proposta por Wei e Fung (1998) para modelos de regressão beta com dispersão variável pode ser visto em Ferrari et al. (2011).

Um outro fator bastante importante para uma análise de resíduo é a detecção de pontos influentes, isto é, observações que exercem um peso desproporcional nas estimativas dos parâmetros do modelo. Uma ferramenta a ser usada para quantificar o impacto dessas observações nas estimativas dos coeficientes de regressão é a distância de Cook, ver Cook (1977). Este tipo de diagnóstico pode ser feito utilizando o método de influência local, ver Cook (1968). A distância de Cook para o modelo de regressão beta definido em (1) apresentada em Espinheira et al. (2008a) é dada por

$$LD_t = \frac{h_{tt}^*}{1 - h_{tt}^*} (r_t^{pp})^2,$$

em que r_t^{pp} é o resíduo padronizado ponderado.

A distância de Cook será grande quando o t -ésimo ponto fornecer um resíduo padronizado ponderado grande ou quando h_{tt} for próximo de 1. Usualmente na prática utiliza-se de ferramentas gráficas para detectar pontos de influência. Uma forma comum de detectar estes pontos é construindo um gráfico de LD_t versus os índices das observações, para detectar possíveis pontos influentes.

2.2 Autocorrelação espacial: Índice de Moran

A autocorrelação espacial mede o quanto o valor observado de uma variável em uma região está relacionado aos valores desta mesma variável nas localidades vizinhas. A autocorrelação espacial está associada à ideia de que valores observados em áreas geográficas adjacentes

se mostram mais parecidos ou diferentes do que o esperado sob a hipótese de que a distribuição das variáveis é invariante por permutação dos índices que localizam as áreas no espaço, ver Assunção (2001). Neste contexto, uma estatística padrão que costuma ser utilizada para medir a associação espacial entre unidades de área é o Índice de Moran (Moran, 1950). Considere que z_t, \dots, z_n são as variáveis aleatórias medidas nas n áreas, o Índice de Moran é obtido por:

$$I = \frac{n \sum_{t \neq j} w_{tj} (z_t - \bar{z})(z_j - \bar{z})}{S_0 \sum_{t=1}^n (z_t - \bar{z})^2},$$

em que \bar{z} é o valor médio do atributo na região de estudo e S_0 é $\sum_{i=1}^n w_{ij}$ e w_{ij} são os elementos da matriz quadrada, $n \times n$, W . Essa matriz, denominada matriz de proximidade espacial (matriz de vizinhanças), é uma ferramenta essencial para descrever o arranjo espacial dos dados. Os elementos w_{ij} ($w_{ij} \geq 0$ e $w_{tt} = 0$) da matriz W refletem a intensidade da interdependência existente entre as regiões A_t e A_j , ou seja, representam o peso ou o grau de proximidade espacial entre as áreas A_t e A_j . A escolha dos elementos w_{ij} é arbitrária e pode ser feita considerando-se vários critérios, neste artigo foi utilizado o critério do vizinho mais próximo.

O Índice de Moran foi utilizado para verificar a correlação entre os municípios da região Nordeste. O teste de Moran realizado apresentou p -valor $< 2,2 \times 10^{-16}$, então rejeita-se a hipótese nula de que os municípios não são correlacionados espacialmente aos níveis de significância usuais, logo identifica-se que existe correlação espacial entre os municípios do Nordeste. A correlação espacial existente nos dados é incorporada ao modelo de regressão, através da adição de uma variável explicativa que considera esta dependência espacial (Cribari-Neto e Pereira, 2013). Assim, foi incorporado aos modelos como covariáveis, a latitude e longitude dos municípios da região Nordeste.

2.3 Aspectos computacionais

Para realizar as análises estatísticas, referentes a modelagem empírica via modelos de regressão beta e análises descritivas, foi utilizado o *software* estatístico R, disponível no site <http://www.r-project.org>. O pacote utilizado para modelagem de regressão beta no R foi o *betareg*, veja Cribari-Neto e Zeleis (2010).

3 Descrição dos dados

Os dados referentes as covariáveis e a variável resposta utilizadas nesta pesquisa, foram obtidos a partir do Tribunal Superior Eleitoral (TSE), Instituto Brasileiro

de Geografia e Estatística (IBGE), Atlas do Desenvolvimento Humano do Brasil (ADHB) e o Instituto de Pesquisa Econômica Aplicada (IPEA). Na Tabela 1, são apresentadas as descrições da variável resposta (Y) e das covariáveis que foram utilizadas.

Na Tabela 2 estão apresentadas algumas estatísticas descritivas das variáveis contínuas usadas: mínimo, máximo, primeiro quartil ($Q_{1/4}$), mediana, média e terceiro quartil ($Q_{3/4}$). Estas estatísticas foram calculadas com base nos municípios da região Nordeste do Brasil, com exceção a Fernando de Noronha, distrito estadual de Pernambuco, pelo motivo de não participarem das eleições e a cidade de Nazária (PI), por não fornecer informações necessárias sobre as variáveis utilizadas. No total foram 1792 municípios avaliados.

Observa-se na Tabela 2 em relação a variável *VOTOS* que o percentual mínimo de votos válidos que a candidata Dilma obteve nas eleições de 2010 na região Nordeste foi de 0,31 e o percentual máximo, 0,97. O percentual médio de votos da presidente Dilma foi de 0,73 nos municípios do Nordeste. Também pode ser visto que os gastos per capita no ano de 2010 com o programa assistencial Bolsa Família na região Nordeste variam entre R\$ 50,54 e R\$ 482,80 por família. Apresentando um benefício com valor médio de aproximadamente 183 reais por família. O terceiro quartil da variável *GASTO2010* é 201,4. Isto significa que em 75% dos municípios o valor do benefício por família é igual ou menor a R\$ 201,4.

Em 50% dos municípios do Nordeste o percentual de analfabetismo é de até 35,52% (Tabela 2). O menor percentual de analfabetos observado foi de aproximadamente 5% e o maior foi de 57%. Em relação ao percentual de pobres na região Nordeste varia entre 9,84% e 78,23%. O primeiro quartil da variável *POBRES* é 34,63, ou seja, em 25% dos municípios apresentam um percentual de pobres menor ou igual a 35%.

O percentual da população residente em áreas rurais no Nordeste é em média de 45% (Tabela 2). O maior percentual de pessoas que residem na zona rural foi de 92%. Também podemos notar que em 50% dos municípios é de até 45% das pessoas vivem em áreas rurais.

Em relação a variável *RENDA* o valor médio da renda per capita da população residente na região Nordeste é de aproximadamente R\$ 276. Também foi observado que, a menor renda per capita registrada foi de 96,25 reais por habitante e o maior foi de 1.144 mil reais por habitante.

O percentual mínimo de votos válidos que a candidata Dilma obteve nas eleições de 2010 na região Nordeste foi no município de Pilar, no Estado de Alagoas, e a proporção máxima observada foi em Calumbi, no Estado de Pernambuco. Vale salientar, que nas eleições de 2010, a candidata Dilma venceu em todos os Estados do Nordeste (ver Figura 1).

O menor gasto per capita em 2010 com Bolsa Família

Tabela 1: Descrição da variável resposta (Y) e das covariáveis utilizadas.

Variável	Definição
$VOTOS$ (Y)	Percentual de votos válidos de Dilma Rousseff no segundo turno da eleição de 2010.
$VOTOSLULA$	Percentual de votos válidos de Lula no segundo turno da eleição de 2006.
$POBRES$	Percentual de pessoas com renda per capita igual ou inferior a R\$ 140,00 em 2010.
$RENDA$	Renda per capita em 2010.
$GASTO2010$	Gastos, em milhões de reais, nos programas assistenciais em 2010.
$PIB2010$	PIB per capita municipal em 2010.
$PREF$	Variável <i>dummy</i> : 1 se prefeito do PT em 2010, 0 caso contrário.
LAT	Latitude do município.
$LONG$	Longitude do município.
$ANALF$	Taxa de analfabetismo da população de 25 anos ou mais em 2010.
$RURAL$	Percentual da população vivendo em áreas rurais em 2010.

Tabela 2: Estatísticas descritivas das variáveis contínuas.

Variável	Mínimo	$Q_{1/4}$	Mediana	Média	$Q_{3/4}$	Máximo
$VOTOS$	0,31	0,65	0,75	0,73	0,82	0,97
$GASTO2010$	50,54	162,60	182,90	180,50	201,40	482,80
$POBRES$	9,84	34,63	41,95	41,53	48,72	78,23
$PIB2010$	1,073	1,791	2,080	2,691	2,550	132,40
$RURAL$	0	30,44	45,42	44,79	59,5	91,68
$ANALF$	4,69	29,84	35,52	34,98	40,65	57,18
$RENDA$	96,25	222,00	259,30	276,50	301,80	1144,00
$VOTOSLULA$	42,35	71,89	79,10	77,82	85,28	97,20

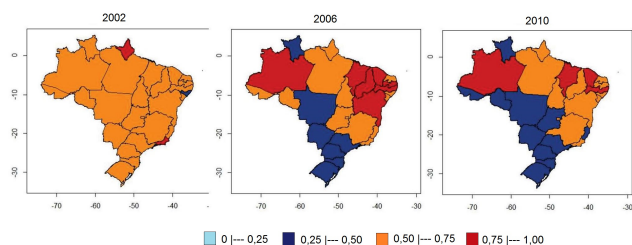


Figura 1: Percentual de votos válidos no segundo turno das eleições de 2002 e 2006 de Lula e Dilma em 2010, nos Estados brasileiros.

foi registrado no município Mucuri, no Estado da Bahia e o maior, em Sebastião Barros, localizado no Estado do Piauí.

O menor percentual de analfabetos encontra-se na cidade de Salvador (BA) e o maior, na cidade de Caxingo (PI). Em que, mais da metade (57%) da população de 25 anos ou mais são consideradas analfabetas.

O menor percentual de pobres foi observado na cidade de Toritama, seguido do município de Santa Cruz, ambas no Estado de Pernambuco. E a maior, no município de Marajá do Sena, no Estado do Maranhão. O maior Produto Interno Bruto (PIB) per capita observado foi no município de São Francisco do Conde, na Bahia, é o

município com o maior PIB per capita no Brasil, devido a produção e refino de petróleo pela refinaria RLAM, da Petrobrás. O menor PIB per capita observado, foi no município de São Vicente Ferrer (MA). O município com maior percentual de pessoas que residem em áreas rurais foi encontrado em Limoeiro de Anadia, no Estado de Alagoas.

Em 2010, dos 1792 municípios do Nordeste, apenas 142 dos municípios são governados por prefeitos do Partido dos Trabalhadores (PT), que corresponde a 7% das cidades do Nordeste.

O município Marajá do Sena (MA) possui a população com menor renda per capita e a maior é vista na capital de Pernambuco, Recife. Em que, a renda per capita chega a aproximadamente 1.144 mil reais por habitante.

4 Especificação do modelo de regressão beta

Nesta seção apresenta-se uma modelagem empírica relacionada à eleição presidencial nos municípios da região Nordeste. Foi utilizado o modelo de regressão beta (Ferrari e Cribari-Neto, 2004). Pretende-se investigar a relação do percentual de votos com variáveis políticas, sociais e econômicas, além dos gastos com o programa

Bolsa Família, que é uma das principais fontes de discussão sobre o resultado das eleições presidências em 2010.

Na seleção das covariáveis utilizadas para explicar a variabilidade da percentual de votos válidos da atual presidente Dilma Rousseff no segundo turno das eleições presidências de 2010, foi utilizado o critério de seleção de modelo AIC (*Akaike's information criterion*), proposto por Akaike (1974).

O modelo de regressão beta, pode admitir estrutura de regressão para a parâmetro de precisão. Assim inicialmente, é necessário verificar se a precisão é fixa, isto é, se há ou não estrutura de regressão, para o parâmetro de precisão ϕ . Neste contexto adicionalmente, foi realizado o teste da razão de verossimilhanças sob um modelo com precisão variável *versus* o modelo sem precisão variável, isto é, testamos a hipótese de que a precisão é fixa, $H_0 : \phi_1 = \dots = \phi_q$. A partir do teste da razão de verossimilhanças, a hipótese nula não foi aceita ao níveis usuais de significância (Espinheira, 2007). Sendo assim, além de modelar a média, existe a necessidade de modelar um parâmetro de precisão.

A variável resposta Y , é o percentual de votos válidos da candidata Dilma Rousseff no segundo turno das eleições de 2010. As covariáveis selecionadas para o modelo final foram *POBRES*, *GASTO2010*, *PREF*, *LAT*, *LONG*, *VOTOSLULA*, *PIB2010*, *RURAL*, *RENDA* e *ANALF*. A Figura 2 apresenta o histograma e o *box-plot* da variável resposta *VOTOS* e identifica-se que a distribuição dessa variável é assimétrica. Neste contexto, foi aplicado o modelo de regressão beta com precisão variável.

O modelo selecionado é dado por:

$$\begin{aligned} \text{cloglog}(\mu_t) &= \beta_1 + \beta_2 \text{GASTO2010}_t + \beta_3 \text{PREF}_t \\ &+ \beta_4 \text{POBRES}_t + \beta_5 \text{LAT}_t \\ &+ \beta_6 \text{LONG}_t + \beta_7 \text{VOTOSLULA}_t \\ &+ \beta_8 \text{PIB2010}_t + \beta_9 \text{ANALF}_t + \beta_{10} \text{RURAL}_t, \\ \log(\phi_t) &= \gamma_1 + \gamma_2 \text{RENDA}_t, \end{aligned}$$

para $t = 1, \dots, 1792$.

Os coeficientes estimados para o modelo selecionado acima, está apresentado na Tabela 3. Das covariáveis selecionadas para o modelo apenas a covariável *LAT* não foi significativa, em que apresentou um p -valor de 0,40. Entretanto como a covariável *LONG* foi significativa, então justifica-se o uso da covariável *LAT* no modelo, pois não faz sentido utilizar apenas a longitude do município.

É possível verificar que as covariáveis *POBRES*, *GASTO2010*, *PIB2010*, e *VOTOSLULA* tiveram uma influência positiva em relação ao percentual de votos da presidente Dilma Rousseff nas eleições de 2010. Isto é, quanto maior for o gasto com o programa assistencial Bolsa Fa-

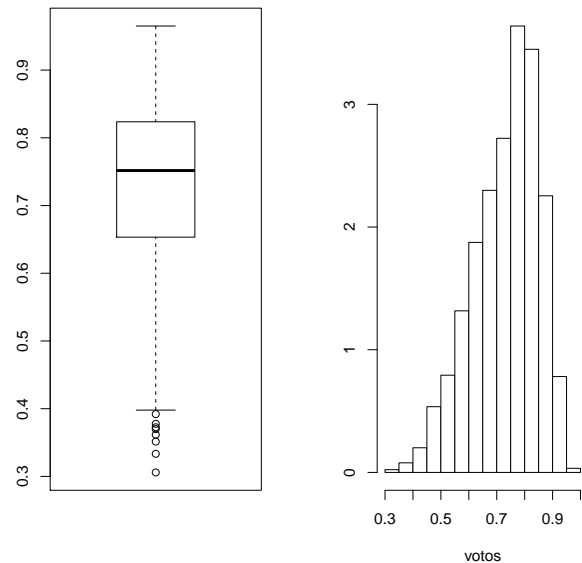


Figura 2: Histograma e box-plot do percentual de votos válidos no segundo turno de Dilma Rousseff na eleição de 2010.

mília, percentual de pobres, valor do PIB per capita, percentual de votos de Lula em 2006 e a proporção de pessoas que residem na zona rural, estes municípios tendem a apresentar, em média, um maior percentual de votos a favor de Dilma. Em contrapartida, as covariáveis *PREF*, *LAT*, *LONG* e *ANALF* exercem efeito negativo. Em relação a estrutura de regressão para o parâmetro de precisão (ϕ), é visto que à medida que a covariável *RENDA* aumenta, a precisão também aumenta, isto é, os municípios que apresentam maior renda per capita tendem a apresentar respostas mais precisas. Importante citar que, o pseudo- R^2 obtido no modelo selecionado foi de 0,695.

Com o interesse de realizar uma análise de resíduo e diagnóstico, existem algumas ferramentas gráficas que podem auxiliar na detecção de possíveis afastamentos das suposições feitas para o modelo de regressão. Entre elas, na Figura 3 é apresentado o gráfico da probabilidade normal com envelopes simulados. Foi utilizado o resíduo de *pearson* para o modelo selecionado, pode-se ver na Figura 3 que os resíduos em geral, permanecem dentro dos limites de confiança dos envelopes simulados, então designa-se que o modelo selecionado aparenta fornecer uma boa representação dos dados. Adicionalmente foi realizado o teste RESET proposto por Ramsey (1969). A hipótese nula de que o modelo estimado está bem especificado não foi rejeitada, em que o p -valor do teste obtido foi de 0,70, acima dos níveis usuais de significância. Portanto, o modelo estimado está bem especificado.

Tabela 3: Coeficientes estimados para as estruturas de regressão para μ e ϕ , e seus respectivos p -valores.

Modelo para estrutura μ			
Variáveis	$\hat{\beta}$	erro padrão	p -valor
Intercepto	-2,3151	0,0734	$< 2 \times 10^{-16}$
POBRES	0,0016	0,0007	0,0163
GASTO2010	0,0009	0,0002	$< 2,65 \times 10^{-07}$
PREF	-0,0964	0,0165	$< 5,05 \times 10^{-09}$
LAT	-0,0012	0,0014	0,4044
LONG	-0,0107	0,0016	$< 2,61 \times 10^{-11}$
VOTOSLULA	0,0248	0,0005	$< 2 \times 10^{-16}$
PIB2010	0,0023	0,0011	0,0346
ANALF	-0,0028	0,0008	0,0003
RURAL	0,0021	0,0003	$< 9,76 \times 10^{-13}$
Modelo para estrutura ϕ			
Variável	$\hat{\gamma}$	erro padrão	p -valor
Intercepto	3,6597	0,1009	$< 2 \times 10^{-16}$
RENDA	0,0006	0,0003	0,0858

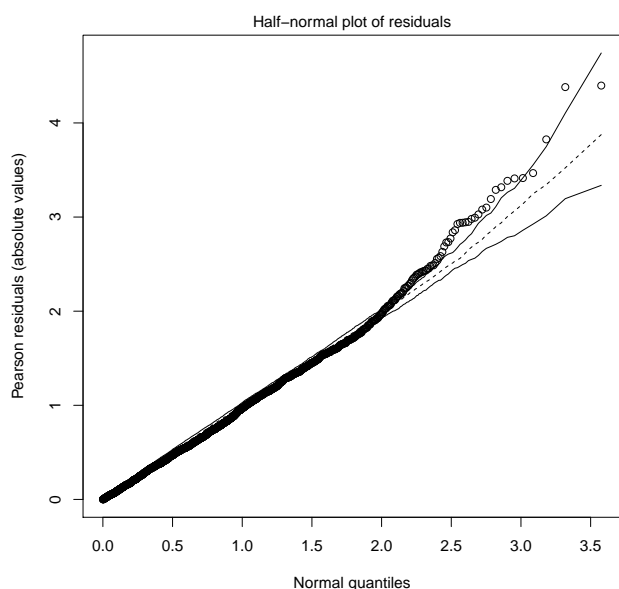


Figura 3: Gráfico da probabilidade normal com envelopes simulados (resíduo pearson).

Ainda neste contexto, temos na Figura 4 o gráfico dos resíduos *pearson* versus os índices das observações. Pode-se ver, que o modelo de regressão beta selecionado aparenta estar bem ajustado, dado que apenas 4 resíduos se encontram fora do intervalo de $(-3, 3)$.

Na detecção de pontos influentes e de alavancagem, foi considerado a distância de Cook e a alavancagem generalizada. A Figura 5 apresenta os gráficos da distância de Cook e da alavancagem generalizada, respectivamente. Nota-se que no gráfico das distâncias de Cook e da alavancagem generalizada temos uma obser-

vação que destaca-se entre as demais, correspondente ao município de São Francisco do Conde, no estado da Bahia. É importante frisar que esse município é o que possui o maior Produto Interno Bruto per capita entre os municípios brasileiros. Segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE) a arrecadação municipal de impostos ligados à produção e refino de petróleo pela refinaria RLAM, da Petrobras, é de cerca de 200.000.000 de reais por ano. Em 2006 o PIB per capita era de 128,73, que em 2010 passou para 132,42. Assim, foi ajustado o modelo sem essa observação e identificou-se que a covariável *PIB2010*, deixou de ser significativa para o modelo. Entretanto, os resultados apresentados neste artigo são baseados nos 1792 municípios conforme citado anteriormente (Seção 3).

Em 2010, foram gastos aproximadamente mais de 14 bilhões de reais com o programa Bolsa Família, no Brasil. Vale salientar, que mais da metade desses gastos foram aplicados na região Nordeste, em que foram gastos 7 de bilhões de reais, ver Ministério de Desenvolvimento Social e Combate à Fome (2010b). No ano de 2006 o gasto per capita com Bolsa Família era R\$ 78,77 por família, em 2010 esse valor era de R\$ 142,84, ou seja, a valor do gasto per capita com o programa na região Nordeste em 2010 aumentou 64,07 reais, quando se é comparado com o ano de 2006.

Para obter o impacto dos gastos com programas assistenciais nas eleições presidenciais de 2010, na região Nordeste, foi calculado os valores ajustados para cada $t = 1, \dots, 1792$, foi definido a covariável $GASTO2010 = 0$, $PREF = 1$ para o cenário em que todos os municípios são governados por prefeitos filiados ao PT, e $PREF = 0$, caso contrário. As demais covariáveis, iguais aos seus respectivos valores em cada observação. Assim obtemos μ_t , em que $\mu_t = 1 - \exp(-\exp(\eta_t))$, depois

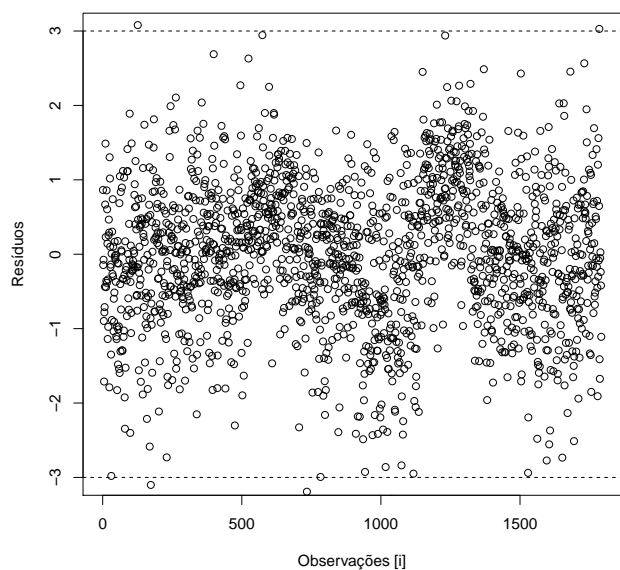


Figura 4: Gráfico dos resíduos *pearson* versus os índices das observações.

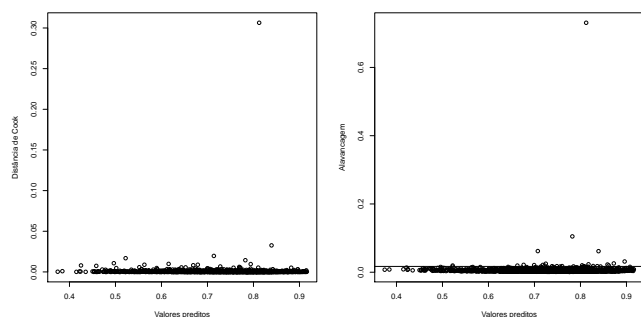


Figura 5: Gráficos das distâncias de Cook e da alavancagem generalizada.

de obter μ_t , foi calculado o somatório dos produtos entre μ_t e o número de votos válidos nos municípios. O resultado dessa soma, fornece a quantidade de votos da candidata Dilma Rousseff nas eleições presidenciais de 2010, estimada pelo modelo.

Os resultados obtidos revelam que se não houvessem gastos com o programa Bolsa Família, o quantitativo de votos da presidente Dilma reduziriam em aproximadamente 1.369.568 milhões. Adicionalmente, foi considerado dois cenários para estimar a quantidade de votos da candidata Dilma na região Nordeste, no primeiro cenário, foi identificado que se não houvessem gastos com o programa Bolsa Família e se em todos os municípios os prefeitos fossem filiados ao PT, a candidata Dilma teria

aproximadamente 16.261.767 milhões de votos, ou seja, o quantitativo de votos da presidente Dilma teria uma redução de aproximadamente 2.125.827 milhões na região Nordeste, mantendo as demais covariáveis fixadas na mediana. No segundo cenário, se não houvessem gastos com o Bolsa Família e se em todos os municípios da região Nordeste os prefeitos não fossem filiados ao PT, a presidente Dilma teria aproximadamente 17.154.058 de votos, ou seja, ela teria uma redução de 1.233.536 milhões de votos nessa região nas eleições presidenciais de 2010, mantendo as demais covariáveis fixadas na mediana. Assim, mesmo que não houvessem gastos com Bolsa Família, a presidente Dilma venceria na região Nordeste e seria eleita a Presidente da República. Vale salientar que, nas eleições presidenciais de 2010 a candidata Dilma obteve 18.387.594 milhões de votos na região Nordeste, enquanto que o candidato José Serra teve 7.633.467 milhões de votos, ver Tribunal Superior Eleitoral (2010a).

5 Conclusões

Ajustando o modelo de regressão beta para explicar o percentual de votos da presidente Dilma em 2010, verifica-se que o gasto per capita com o programa Bolsa Família do Governo Federal, teve uma influência positiva em relação ao percentual de votos da presidente Dilma Rousseff no Nordeste nas eleições de 2010. Isto é, os municípios que tinham maiores gastos com o programa assistencial Bolsa Família, apresentaram maior percentual de votos a favor da presidente Dilma. Assim como Zucco (2013), foi constatado que os gastos com programas assistenciais influenciam positivamente no percentual de votos do candidato ligado ao governo. Adicionalmente, constata-se os municípios que apresentaram, maior percentual de pobres, maior foi o percentual de votos a favor da candidata Dilma Rousseff. Da mesma forma que Zucco (2008), verificou-se que o candidato ligado ao governo tem melhor desempenho nas regiões menos desenvolvidas, quanto maior o percentual de pessoas pobres, menor o desenvolvimento do município. Quanto maior o percentual de votos de Lula nas eleições presidenciais de 2006 na região Nordeste, maior seria o percentual de votos da candidata Dilma. O ex-presidente Lula encerrou seu mandato na Presidência da República no auge de sua popularidade. Segundo a pesquisa realizada pela Datafolha em dezembro de 2010, 83% dos brasileiros adultos avaliam sua gestão como ótima ou boa, a mais alta já alcançada por um presidente na série histórica do Datafolha, ver Instituto de pesquisa Datafolha (2010). Verifica-se que, a alta popularidade do ex-presidente Lula teve influência positiva para a candidata Dilma nas eleições presidenciais de 2010. Adicionalmente, foi identificado que quanto maior número

de pessoas residentes na zona rural e maior o Produto Interno Bruto per capita do município, maior seria o percentual de votos da presidente Dilma Rousseff em 2010.

Os municípios que têm um número mais elevado de analfabetos com mais de 25 anos de idade e que são governados por prefeitos do PT apresentaram uma redução no percentual de votos da candidata Dilma em 2010. No que diz respeito a estrutura de regressão para o parâmetro de precisão, foi visto que os municípios que apresentam maior renda per capita tentem a apresentar respostas mais precisas.

Por fim, foi calculado o impacto do gasto per capita do programa Bolsa Família nas eleições presidenciais de 2010, considerando o cenário em que não houvessem gastos com o Bolsa Família, se em todos os municípios os prefeitos fossem filiados ao PT e mantendo as demais covariáveis constantes, a votação da presidente Dilma teria uma redução de aproximadamente 2.125 milhões de votos na região Nordeste, que equivale a 11,6% de votos a menos. Em contrapartida, se não houvesse gastos com o Bolsa Família, se todos os prefeitos da região Nordeste não fossem filiados ao PT e mantendo as demais covariáveis constantes, a votação da presidente Dilma teria uma redução de aproximadamente 1.233 milhões de votos nessa região nas eleições presidenciais de 2010, que equivale a 6,7% de votos a menos. Em vista disso, o fato do município ser administrado por um prefeito filiado ao PT, não favoreceu a presidente Dilma nas eleições pelo contrário, de certa forma prejudicou seu desempenho na votação. Portanto, mesmo que não houvesse gasto com Bolsa Família, a candidata Dilma venceria na região Nordeste e seria eleita a Presidente da República, dado que a diferença de votos válidos entre os candidatos Dilma e José Serra foi de 10.754.127 votos.

Referências

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6), 716–723.
- Almeida, L., 2011. A percepção das mulheres beneficiárias acerca do programa bolsa família da cidade de João Pessoa/pb. In: (Monografia em Serviço Social), Universidade Federal da Paraíba, João Pessoa. p. 69.
- Assunção, R. M., 2001. Estatística espacial com aplicações em epidemiologia, economia e sociologia. In: São Carlos: 7a Escola de Modelos de Regressão.
- Canêdo-Pinheiro, M., 2009. Bolsa família ou desempenho da economia? determinantes da reeleição de lula em 2006. In: Anais do XXXVII Encontro Nacional de Economia.
- Carraro, A., Shikida, C., Monasterio, L., Araújo, A., Damião, O., 2007. "it is the economy, companheiro!": an empirical analysis of lula's re-election based on municipal data. *Economics Bulletin* 29 (2), 976–991.
- Cook, R., 1968. Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B* 48, 133–169.
- Cook, R., 1977. Detection of influential observation in linear regression. *Technometrics* 19 (1), 15–18.
- Cribari-Neto, F., Souza, T. C., 2012. Testing inference in variable dispersion beta regressions. *Journal of Statistical Computation and Simulation* 82, 1827–1843.
- Cribari-Neto, F., Zeleis, A., 2010. Beta regression in r. *Journal of Statistical Software* 34 (2).
- Cribari-Neto, F., F., Pereira, T., T. L., 2013. Avaliação da eficiência de administrações municipais no estado de São Paulo: Uma nova abordagem via modelos de regressão beta. *Revista Brasileira de Biometria* 31, 270–294.
- Espinheira, P., 2007. Regressão beta. Tese de doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo (USP).
- Espinheira, P., Ferrari, S., Cribari-Neto, F., 2008a. Influence diagnostics in beta regression. *Computational Statistics & Data Analysis* 52 (9), 4417–4431.
- Espinheira, P. L., Ferrari, S., Cribari-Neto, F., 2008b. On beta regression residuals. *Journal of Applied Statistics* 35, 407–419.
- Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31 (7), 799–815.
- Ferrari, S., Espinheira, P. L., Cribari-Neto, F., 2011. Diagnostic tools in beta regression with varying dispersion. *Statistica Neerlandica* 65, 337–351.
- Fisher, R., 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A* 222, 309–68.
- Instituto de pesquisa Datafolha, 2010. Avaliação do governo do presidente Lula. Datafolha. URL <http://datafolha.folha.uol.com.br/opiniaopublica/avaliacaodegoverno/presidente/lula/indice-1.shtml>
- Marques, R., 2013. Políticas de transferência de renda no Brasil e na Argentina. *Revista de Economia Política* 33 (2 (131)), 298–314.
- McCullagh, P., Nelder, J., 1989. *Generalized Linear Models*, 1989. Chapman and Hall.

- Ministério de Desenvolvimento Social e Combate à Fome, 2010a. Bolsa Família. Cadastro Único. Governo Federal do Brasil, Brasil.
URL <http://www.mds.gov.br/bolsafamilia/cadastrounico>
- Ministério de Desenvolvimento Social e Combate à Fome, 2010b. Gastos com Programas Assistenciais. Governo Federal.
URL <http://aplicacoes.mds.gov.br/sagi/miv/miv.php>
- Ministério do Desenvolvimento Social e Combate à Fome, Janeiro 2004. Bolsa família. condicionalidades.
URL <http://www.mds.gov.br/bolsafamilia/condicionalidades>
- Moran, P. A. P., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37, 17–23.
- Presidência da República, janeiro 2004. Lei nº 10.836. cria o programa bolsa família e da outras providências.
URL http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Lei/L10.836.htm
- Ramsey, J. B., 1969. Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society, B* 31, 350–371.
- Rocha, A., Simas, A., 2011. Influence diagnostics in a general class of beta regression models. *Test* 20, 95–119.
- Silva, M., Yasbek, M., Geraldo, G., 2012. A política Social brasileira no século XXI (A prevalência dos programas de transferência de renda), 6th Edition. São Paulo: Cortez.
- Simas, A., Barreto-Souza, W., Rocha, A., 2010. Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis* 54, 348–366.
- Smithson, M., Verkuilen, J., 2006. A better lemon-squeezer? maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11, 54–71.
- Souza, T. C., Cribari-Neto, F., 2013. Uma estimativa do impacto eleitoral do programa bolsa-família. *Revista Brasileira de Biometria* 31, 79–103.
- Tribunal Superior Eleitoral, 2010a. Eleições presidenciais. Justiça Eleitoral, Brasil.
URL <http://www.tse.jus.br/eleicoes/eleicoes-antiores>
- Tribunal Superior Eleitoral, 2010b. Estatística do eleitorado. Quantitativo.
URL <http://www.tse.jus.br/eleicoes/eleicoes-antiores/eleicoes-2010/estatisticas-de-eleitorado>
- Wei, B.C. and Hu, Y., Fung, W., 1998. Generalized leverage and its applications. *Scandinavian Journal of Statistics* 25, 25 – 37.
- Zucco, C., 2008. The president's 'new' constituency: Lula and the pragmatic vote in brazil's 2006 presidential elections. *Journal of Latin American Studies* 40, 29–39.
- Zucco, C., 2013. When payouts pay off: Conditional cash transfers and voting behavior in brazil 2002–10. *American Journal of Political Science* 57 (0), 810–822.