

Universidade Estadual de Maringá
Programa de Iniciação Científica – PIC
Departamento de Estatística
Orientador: Prof. Dr. Diogo Francisco Rossoni
Acadêmico: Wesley Oliveira Furriel

**Determinantes do voto à presidência:
*análise espacial das eleições gerais no
Brasil no período de 1994 a 2014.***

Maringá 25 de Agosto de 2017

Universidade Estadual de Maringá
Programa de Iniciação Científica – PIC
Departamento de Estatística
Orientador: Prof. Dr. Diogo Francisco Rossoni
Acadêmico: Wesley Oliveira Furriel

**Determinantes do voto à presidência:
*análise espacial das eleições gerais no
Brasil no período de 1994 a 2014.***

Relatório contendo os resultados finais do projeto de iniciação científica vinculado ao Programa PIC-UEM.

Maringá 25 de Agosto de 2017

Resumo

O voto é um dos principais recursos para eleições de representantes, e tomada de decisões políticas dentro de um sistema democrático. Assim, buscar compreender os fatores que o determinam tem sido tema de pesquisa na agenda de estudiosos de diferentes áreas, que detém em comum, o uso de recursos estatísticos para explicar as características deste fenômeno. Desse modo, o presente trabalho visou apresentar a espacialização da proporção dos votos recebidos pelos candidatos à presidência da República do PT e do PSDB, por município brasileiro, no primeiro turno nas eleições gerais de 1998, 2002, 2006, 2010 e 2014. Para tal, foram utilizadas técnicas de exploração de dados espaciais (ESDA – Exploratory Spatial Data Analysis), que nos permitiram verificar a existência de dependência espacial em nossas variáveis de interesse. Além disso, para o ano de 2010 foi realizado um modelo de regressão Beta com o intuito de verificar alguns dos determinantes do voto, no Partido dos Trabalhadores. Os dados desta pesquisa foram obtidos nos diretórios oficiais do TSE e do IBGE, Censo 2010.

Palavras-chave: Voto, PT, Estatística espacial, Regressão Beta.

Conteúdo

1	Introdução	2
2	Objetivos	4
2.1	Objetivo Geral	4
2.2	Objetivos Específicos	4
3	Desenvolvimento	4
3.1	Materiais e Métodos	4
3.1.1	Detalhamento do banco de dados	4
3.1.2	Estatística de Moran	5
3.1.3	Padronização dos votos	6
3.1.4	Modelo de regressão Beta	6
3.2	Resultados e Discussão	8
3.2.1	Análise exploratória	8
3.2.2	Modelo de regressão Beta	15
3.2.3	Análise de resíduos	18
4	Conclusão	21

1 Introdução

Os fatores que determinam o voto do eleitor em uma figura política tem sido uma questão amplamente discutida dentro da academia, principalmente, em áreas como a Ciência Política, Sociologia e Economia. Diante das diferentes abordagens realizadas por essas áreas, há em comum, na maioria dos trabalhos empíricos, a utilização do conhecimento estatístico para auxiliar no entendimento desse fenômeno, sistematizando e aprofundando as descobertas dessas pesquisas, e também realizando “previsões” sobre futuras eleições.

De acordo com Downs (1957), um dos expoentes da Teoria da Escolha Racional. O eleitor observa o comportamento dos partidos e dos candidatos e avalia se as ações deles estão em conformidade com os seus interesses, guardando percepções positivas e negativas sobre estes. Assim, em uma próxima eleição, ele pode utilizar este conhecimento adquirido previamente para tomar suas decisões, a fim de diminuir os custos informacionais, economizando tempo e recursos cognitivos, em uma espécie de cálculo de interesse, próximo ao comportamento de consumidores. Além disso, independentemente de seu grau de conhecimento acerca do cenário político, os eleitores são sensíveis ao impacto das medidas governamentais. Desse modo, eles realizam um monitoramento das propostas e do desempenho dos partidos no decorrer do tempo, criando a partir disso, um julgamento sumário utilizado para nortear suas decisões políticas, inclusive seu voto (FIORINA, 1987).

No que se refere ao Brasil Peixoto e Rennó (2011) afirmam que, as disputas presidenciais têm sido alvo de diversas pesquisas acerca dos fatores que influenciam nas escolhas eleitorais, desde as eleições diretas de 1989. No entanto, após a eleição de Luiz Inácio Lula da Silva no ano de 2002, e a sua reeleição, em 2006, os trabalhos científicos que visam explicar o comportamento dos eleitores aumentaram e avançaram em seus resultados, consideravelmente. Segundo os autores, hoje sabemos que o eleitor brasileiro leva em consideração o desempenho do governante em seu mandato quando decide se votará novamente nele em uma próxima eleição, que a condição da economia em nível nacional é fundamental para explicar o voto e que a identificação e rejeição aos partidos são pontos importantes do cálculo eleitoral dos indivíduos. Além disso, alguns trabalhos apontam para a importância de se investigar a espacialização do voto, já que, as regiões parecem diferenciar os padrões de decisão entre os brasileiros.

Nas eleições presidenciais ocorridas na última década, mais especificamente, as de 2006, 2010 e 2014, em que o Partido dos Trabalhadores elegeu seus candidatos, sendo Lula em 2006 e Dilma Rousseff em 2010 e 2014. Bastante discussão tem sido direcionada a elucidar o que manteve o partido no poder desde 2002, mais especificamente o que garantiu a reeleição de Lula e posteriormente de Dilma e manteve o PT no poder por quatro eleições seguidas.

Diante dessa discussão, algumas das variáveis mais exploradas enquanto tentativa de evidenciar o direcionamento do voto, são os programas sociais de transferência de renda, que seguem como uma das principais bandeiras nas campanhas eleitorais do PT. Acerca da relação entre essas variáveis e o desempenho de Lula nas eleições de 2006, seguido de Dilma em 2010 e 2014. Existe um amplo debate sobre os possíveis dividendos eleitorais gerados por programas como o Bolsa Família. Dentro das produções acadêmicas a respeito do tema, os resultados são bastante destoantes, chegando a conclusões de que o programa não foi fator determinante para a sua reeleição, bem como, corolários

que sugerem que o programa teve papel determinante na vitória dos petistas. (PINHEIRO, 2009; RENNÓ e PEIXOTO, 2010).

Tendo em vista tais apontamentos essa pesquisa buscou investigar a espacialização da proporção dos votos recebidos pelos candidatos à presidência da República do PT e PSDB, tomando como unidade de análise os municípios brasileiros, nos anos de 1998, 2002, 2006, 2010 e 2014. Além disso, investigamos os determinantes do voto no Partido dos Trabalhadores para o ano de 2010, uma vez que neste ano, ocorreu o Censo Demográfico realizado pelo IBGE, que nos forneceu os dados necessários para a investigação. Dessa forma, foi possível selecionar algumas covariáveis socioeconômicas para a explicação do voto, como IDHM, GINI, população ocupada, população urbana e proporção de indivíduos no ensino superior. Para a modelagem dos dados foi utilizada o modelo de regressão Beta, dados as características da variável resposta que é continuamente medida no intervalo de 0 a 1, os dados foram extraídos das bases do TSE e do IBGE.

2 Objetivos

2.1 Objetivo Geral

O objetivo geral deste trabalho foi verificar se há existência de dependência espacial na proporção dos votos dos dois principais partidos que concorreram ao cargo de presidência, nas eleições de 1998 à 2014. Bem como, realizar um modelo de regressão, para investigar os fatores que influenciaram o voto no PT em 2010.

2.2 Objetivos Específicos

- Investigar a existência de dependência espacial na proporção dos votos dos dois principais partidos brasileiros, PT e PSDB, no primeiro turno das eleições presidenciais de 1994 à 2014, por município brasileiro;
- Investigar os fatores sociais e econômicos que influenciam o voto no PT em 2010, a partir do modelo de regressão Beta;
- Avaliar a qualidade do modelo de regressão Beta para os dados empregados.

3 Desenvolvimento

3.1 Materiais e Métodos

Nesta seção foram apresentadas as variáveis utilizadas na investigação do voto no PT e PSDB, bem como, os bancos de dados dos quais elas foram retiradas. Além disso, foram apresentadas as metodologias utilizadas para a verificação da dependência espacial e da modelagem dos dados, via regressão Beta.

3.1.1 Detalhamento do banco de dados

Os dados utilizados neste trabalho foram obtidos a partir dos repositórios de dados oficiais do TSE e do IBGE 2010. Para a realização das investigações foram utilizadas as variáveis expostas na tabela 1.

Tabela 1: variáveis utilizadas

Código	Descrição
VPT	Proporção de votos no PT, 1º turno para 1998, 2002, 2006, 2010 e 2014
VPSDB	Proporção de votos no PSDB, 1º turno para 1998, 2002, 2006, 2010 e 2014
IDHMR	IDHM Renda
IDHML	IDHM Longevidade
IDHME	IDHM Educação
TFBSUPER	Taxa de frequência bruta ao superior
GINI	Grau de desigualdade na distribuição de renda
PESOURB	População urbana
TRABCC	Empregados com carteira
regiao	Grandes regiões brasileiras
LAT	Latitude
LONG	Longitude

Ressalta-se que a unidade investigada neste trabalho foram os municípios que compõe o Brasil. Oficialmente existem 5570 municípios no Brasil, entretanto, não foi possível obter todas as informações para este montante, neste caso trabalhamos com n=5464, que foram obtidos pelas fontes oficiais citadas.

3.1.2 Estatística de Moran

O índice de Moran foi utilizado para verificação da existência de dependência espacial da variável de interesse, neste caso a proporção de voto no PT e PSDB, por município, nos anos de 1998, 2002, 2006, 2010 e 2014. Seu uso exige variáveis quantitativas e distribuídas espacialmente, para que seja possível calcular sua estatística e testar a hipótese de dependência espacial, sendo ela dada por

$$H_0 : \text{Não há dependência espacial}$$

$$H_1 : \text{Há dependência espacial}$$

partindo disso, a estatística do Moran global pode ser obtida da seguinte forma:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

em que w : A matriz de vizinhança

y_i : A média da partição

y_j : A média da vizinhança

\bar{y} : A média geral

n : O número de partições

Os resultados obtidos a partir do índice de Moran estão entre -1 e 1, sendo que quanto mais próximo de -1 temos uma dependência espacial negativa e quanto mais próximo de 1, dependência espacial positiva.

No caso do índice de Moran local temos a mesma estrutura e interpretação, entretanto, nele local calcula-se a estatística para cada partição no espaço. Além disso, o Moran local nos permite empregar o gráfico Local Indicator Spatial

Association (LISA) que nos auxilia a verificar visualmente na forma de clusters à força da dependência segundo as regiões.

3.1.3 Padronização dos votos

Para a comparação espacial gráfica, entre PT e PSDB realizamos uma transformação dos votos por ano, fazendo a diferença entre os votos destes dois partidos, isto é, foi subtraída a proporção de votos obtida pelo PSDB pela proporção obtida pelo PT, partindo disto, foi feita uma padronização desta diferença de acordo com a seguinte expressão:

$$D = \frac{(X_i - Y_i)}{\max|X_i - Y_i|}, \text{em que } -1 \leq D \leq 1 \quad (2)$$

X_i : Porcentagem de votos obtidos pelo PT

Y_i : Porcentagem de votos obtidos pelo PSDB

Tal padronização nos permitiu criar um indicador em que de propensão ao voto por município, no qual, quanto mais próximo seus valores chegam de -1 temos uma proporção maior de votos no PSDB e quanto mais próximo de 1, uma proporção maior de votos para o PT.

3.1.4 Modelo de regressão Beta

O modelo de regressão Beta é indicado para situações em que a variável resposta Y é medida continuamente num intervalo $0 < Y < 1$, ou em intervalos em que seja possível identificar valores limitados por um máximo e um mínimo. Ferrari e Cribari-Neto (2004) propõe uma reparametrização da distribuição Beta para sua utilização no modelo de regressão. Já que, na análise de regressão, geralmente é mais útil modelar resposta média, então, a fim de obter uma estrutura de regressão para tal, juntamente com um parâmetro de dispersão, é interessante trabalhar com uma parametrização diferente da densidade beta original. O modelo segue as propriedades da Beta, sendo então, adequado para casos em que a variável resposta Y é medida continuamente no intervalo $0 < Y < 1$. Seguindo a nova parametrização, $\mu = \alpha / (\alpha + \beta)$ e $\phi = \alpha + \beta$, ou seja, $\alpha = \mu\phi$ e $\beta = (1 - \mu)\phi$. Isto posto, a média e a variância de Y são dadas pelas seguintes expressões,

$$\mathbb{E}(Y) = \mu \text{ e } \text{Var}(Y) = \frac{V(\mu)}{(1 + \phi)} \quad (3)$$

em que $V(\mu) = \mu(1 - \mu)$, de modo que μ é a média da variável resposta e ϕ pode ser interpretado como um parâmetro de dispersão ou precisão. Tendo em vista essa nova parametrização em termos de μ e ϕ , Y tem-se uma função de densidade Beta dada por:

$$f(y | \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1 - \mu)\phi)} y^{\mu\phi-1} (1 - y)^{(1-\mu)\phi-1}, \quad 0 < y < 1 \quad (4)$$

com $0 < \mu < 1$ e $\phi > 0$.

Assim como na parametrização original é possível obter diferentes formas no comportamento da distribuição de acordo com os valores de seus parâmetros μ e ϕ . Ferrari e Cribari-Neto (2004) apontam que a distribuição pode ser simétrica quando $\mu = 1/2$ e assimétrica quando $\mu \neq 1/2$. Além disso, a dispersão da

distribuição, para um μ fixado diminui quando os valores de ϕ aumentam.

No que diz respeito ao componente sistemático para a média e a precisão consideramos Y_1, \dots, Y_n como uma a.a. da resposta, que segue uma densidade de probabilidade Beta com média μ e parâmetro de precisão ϕ desconhecidos, o modelo de regressão Beta para a média tem à seguinte forma:

$$g(\mu_t) = \sum_{i=1}^k x_{ti}\beta_i = x'_t\beta = \eta_t \quad (5)$$

em que $\beta = (\beta_1, \dots, \beta_k)'$ é um vetor de parâmetros desconhecido e x'_t é um vetor de observações fixas e conhecidas de k covariáveis. No que tange a função de ligação ela é contínua e duas vezes diferenciável, pertencendo ao intervalo $(0, 1)$ e η_t preditor linear. Além disso, o parâmetro de precisão ϕ_t temos

$$h(\phi_t) = \sum_{j=1}^q z_{tj}\gamma_j = z'_t\gamma = \nu_t \quad (6)$$

em que ν_t é o preditor linear.

É preciso ressaltar que o modelos Beta faz parte da família exponencial biparamétrica, dessa forma, utiliza-se o método de máxima verossimilhança para obter os estimadores, a partir da maximização da função de verossimilhança dada por:

$$L(\beta, \gamma) = \prod_{t=1}^n f(y_t | \beta, \gamma) \quad (7)$$

na qual β é um vetor de parâmetros que será estimado, y_t , $t = 1, 2, \dots, n$ são valores observados correspondentes a variável resposta e $f(y_t | \theta)$ é a função densidade de probabilidade Beta para a variável aleatória Y . Partindo disso, aplica-se o log na função $L(\beta, \gamma | y_t)$, obtendo $\ell(\beta, \gamma | y_t)$.

Assim sendo, considere $y = (y_1, y_2, \dots, y_n)$ uma amostra aleatória de n observações, têm-se que o logaritmo natural da função verossimilhança pode ser escrito na forma:

$$\begin{aligned} \ell(\beta, \gamma | y) &= n \log(\Gamma(\gamma)) - n \log(\Gamma(\beta\gamma)) - n \log(\Gamma((1-\beta)\gamma)) \\ &\quad + (\beta\gamma - 1) \sum_{i=1}^n y_i + [(1-\beta)\gamma - 1] \sum_{i=1}^n (1-y_i) \end{aligned} \quad (8)$$

Sabe-se que as estimativas de máxima verossimilhança são os valores de $\theta = (\beta, \gamma)$ que maximizam a função log-verossimilhança da amostra. Fazendo com que os dados observados sejam tão prováveis quanto possíveis. Partindo disso, derivá-se a função de log verossimilhança em relação aos parâmetros e igualá-se a zero, resolvendo o sistema não linear decorrente deste processo, via métodos iterativos computacionais é possível obter as estimativas do modelo.

Para atingir os objetivos desejados contamos com auxílio dos ambientes estatísticos Rstudio 1.0.136, com os pacotes betareg, rgdal, spdep, sp, maptools e tidyverse e SAS 9.4 com as procs nlmixed e proc sgplot, os resultados são apresentados nas seções seguintes.

3.2 Resultados e Discussão

Nesta seção foram apresentados os resultados decorrentes da análise exploratória dos dados, bem como os referentes ao modelo de regressão Beta para o ano de 2010.

3.2.1 Análise exploratória

Antes de iniciar a investigação espacial realizamos algumas análises gráficas com o objetivo de expor o comportamento do voto nos partidos PT e PSDB, no decorrer das eleições presidenciais, e segundo as grandes regiões.

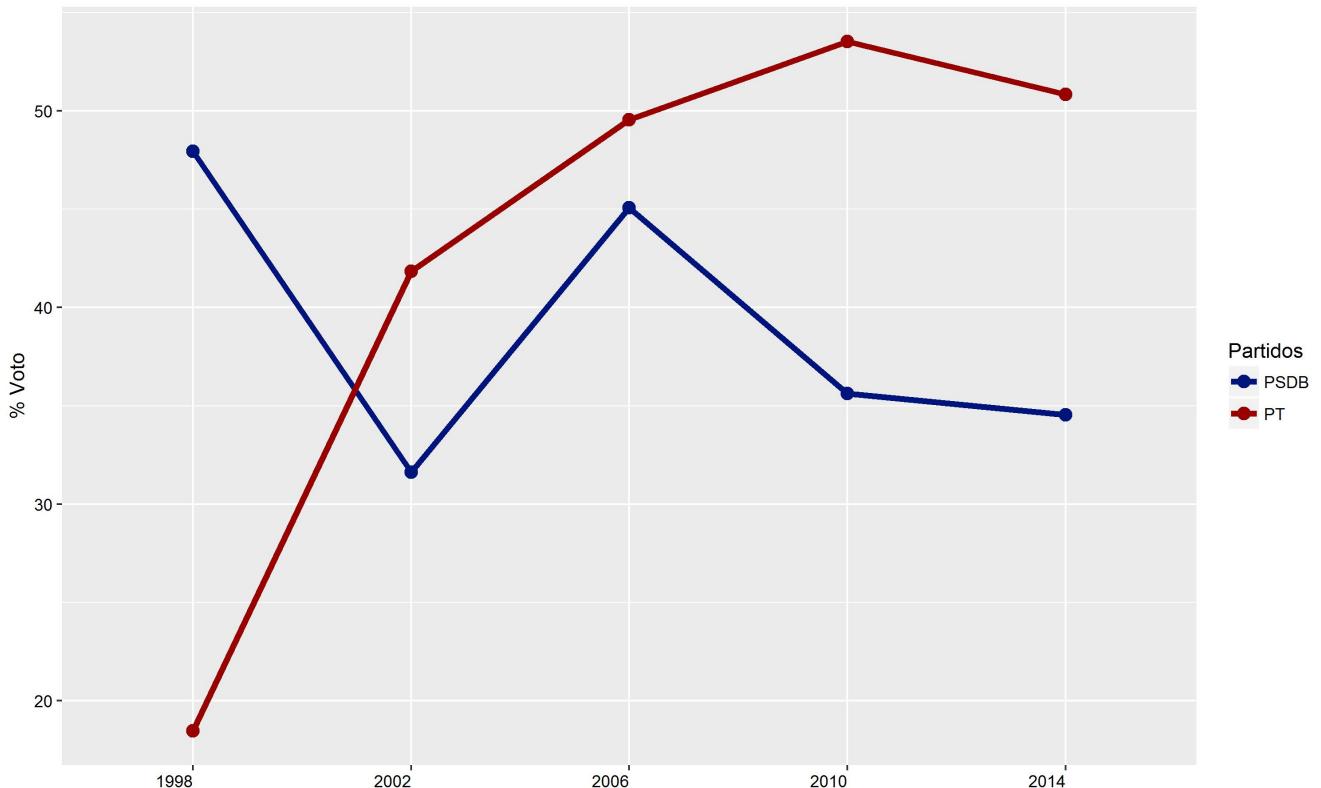


Figura 1: Votos do PT e PSDB segundo as eleições presidenciais

Pelo gráfico exposto na figura 1 observamos o percentual de votos obtidos pelo PT e PSDB no primeiro turno, no decorrer das eleições presidenciais, como é possível observar o PT teve maior porcentagens de votos no primeiro turno para todas as eleições à partir de 1998, o PSDB ganhou as eleições de 1998 e teve uma retomada na porcentagem de voto em 2006.

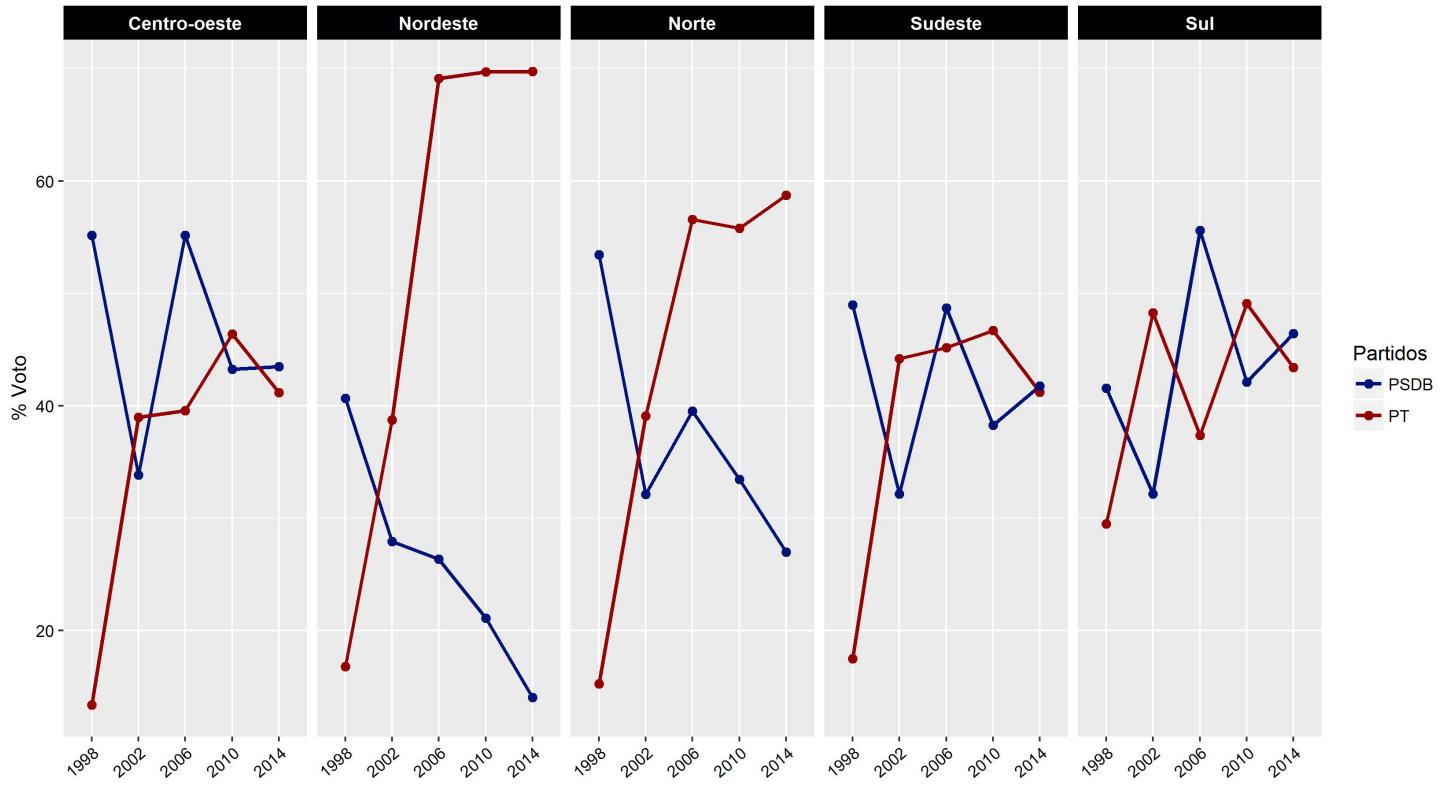


Figura 2: Votos do PT e PSDB segundo as eleições presidenciais, por região

No que tange a porcentagem de votos segundo as regiões, constatamos que Nordeste e Norte nas últimas três eleições, apresentaram maior porcentagem de votos no o PT, tal fenômeno se destaca principalmente no Nordeste em que o voto no PSDB cai acentuadamente no decorrer das eleições, enquanto o do PT aumenta. Nas demais regiões é possível observar oscilações entre os partidos no decorrer dos anos eleitorais.

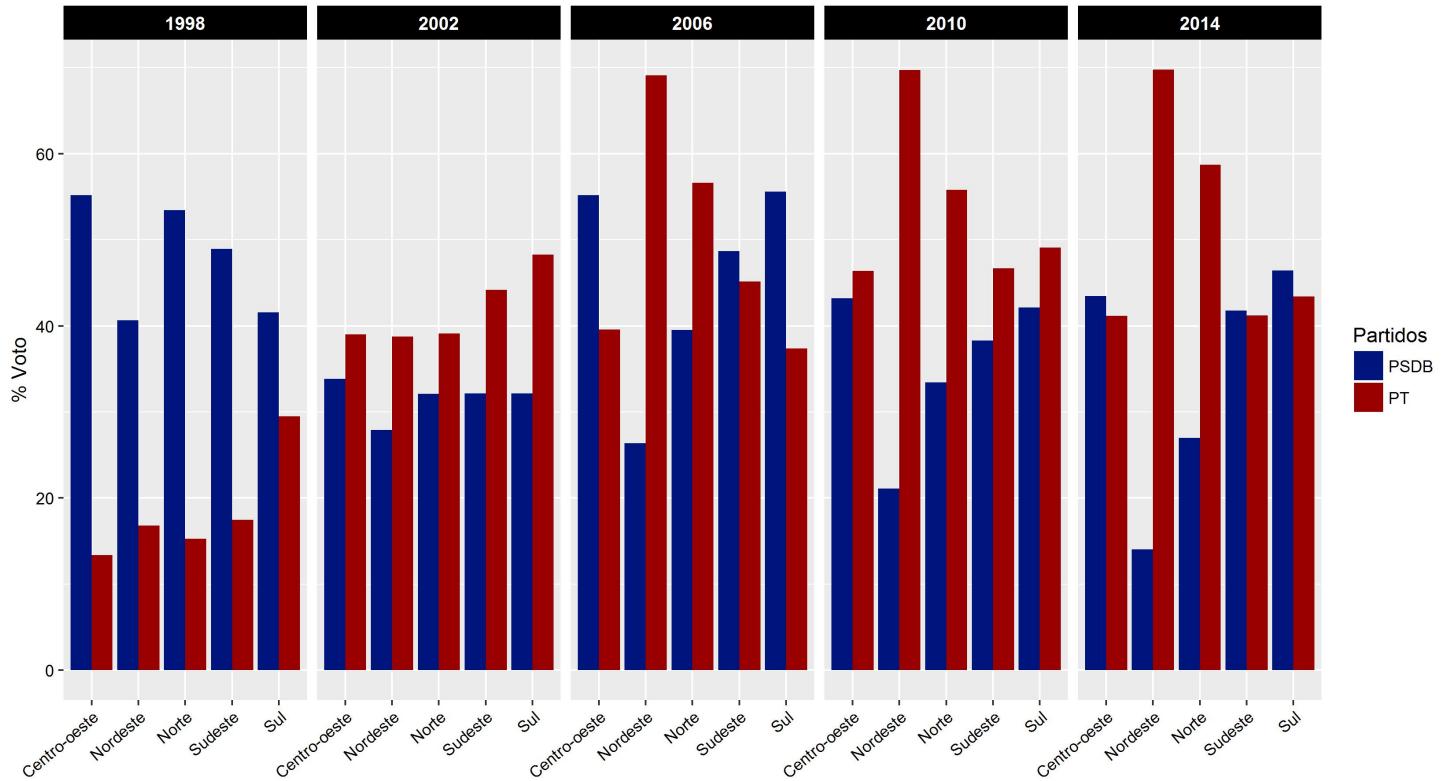
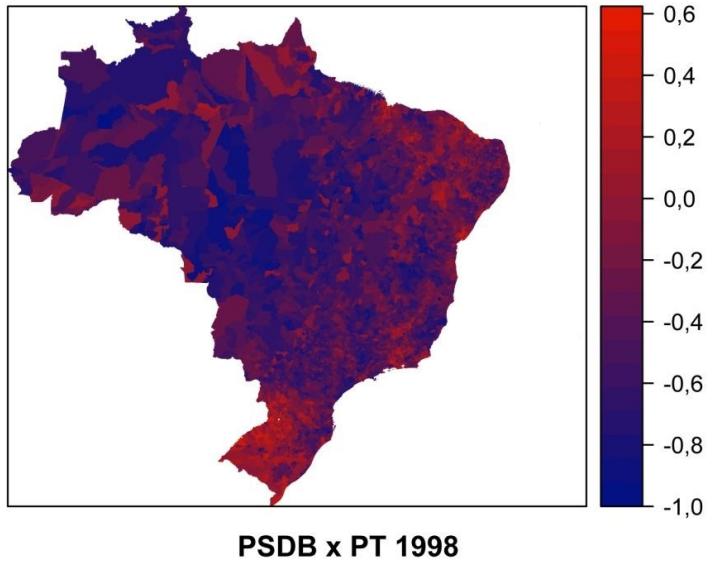
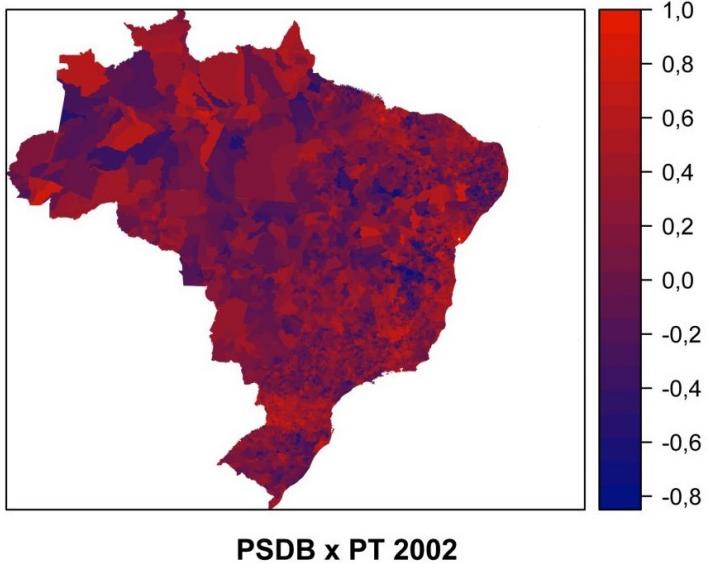


Figura 3: Votos do PT e PSDB segundo as regiões, por ano de eleição presidencial

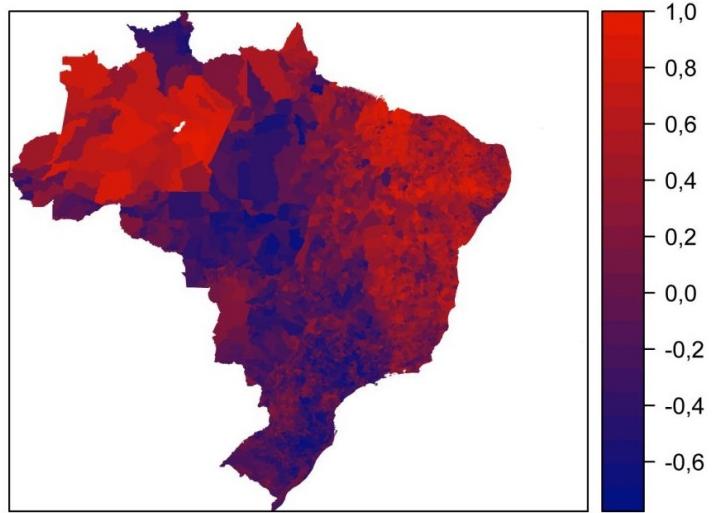
Para melhor visualizar os votos nos dois partidos por região foi feita uma modificação na estrutura gráfica dos votos por eleição e região, neste gráfico notamos que o PT recebeu a maior porcentagem de votos em todas as regiões nos anos de 2002 e 2010, já no caso do PSDB o mesmo ocorre em 1998. É preciso ressaltar também as grandes diferenças na porcentagem de votos entre os partidos nas regiões Norte e Nordeste, como já destacado no gráfico anterior.



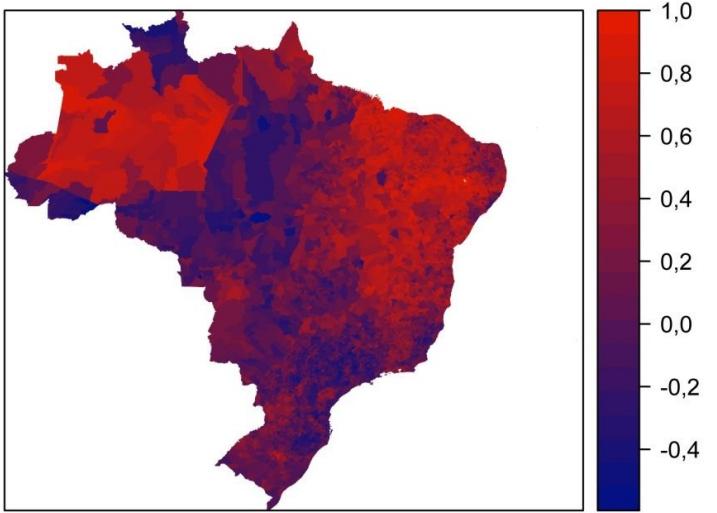
PSDB x PT 1998



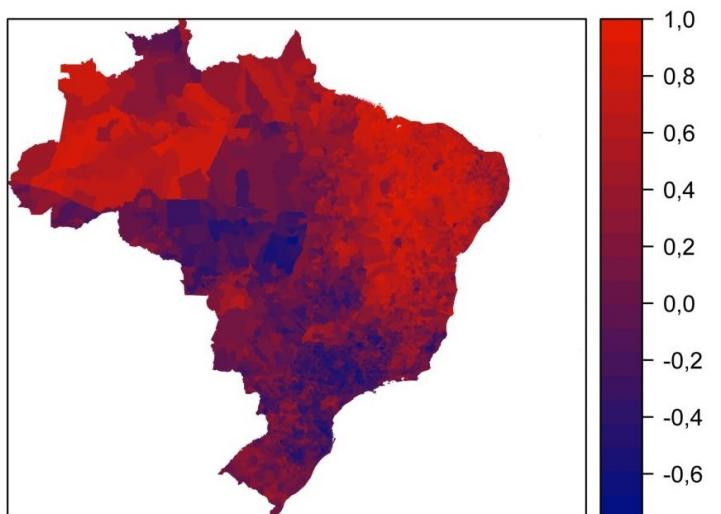
PSDB x PT 2002



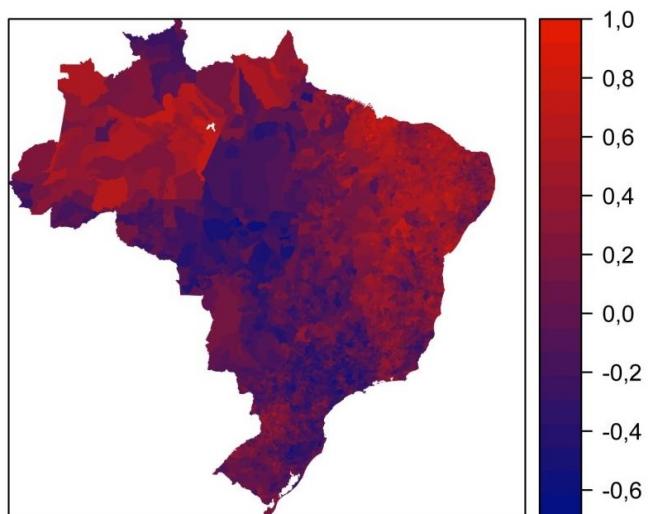
PSDB x PT 2006



PSDB x PT 2010



PSDB x PT 2014



PSDB x PT Média

Figura 4: Diferença de votos entre PT e PSDB

Na figura 4 apresentamos alguns mapas para as eleições presidenciais destacando a diferença de votos entre PT e PSDB a partir da padronização exposta em 2. É válido ressaltar que quanto maior próximo de -1, maior a porcentagem de votos no PSDB no município e quando mais próximo de 1 maior a porcentagem de votos no PT. Dessa forma, é possível constatar nas eleições de 2006, 2010 e 2014 uma clara vantagem na porcentagem de votos no PT, no Norte e Nordeste, quando comparado aos demais anos. É preciso também, destacar que em 1998 o sul apresentava mais disposição ao voto no PT, fato que mudou principalmente em 2006 em que Sul e Centro-Oeste, mostram claras vantagens para o PSDB.

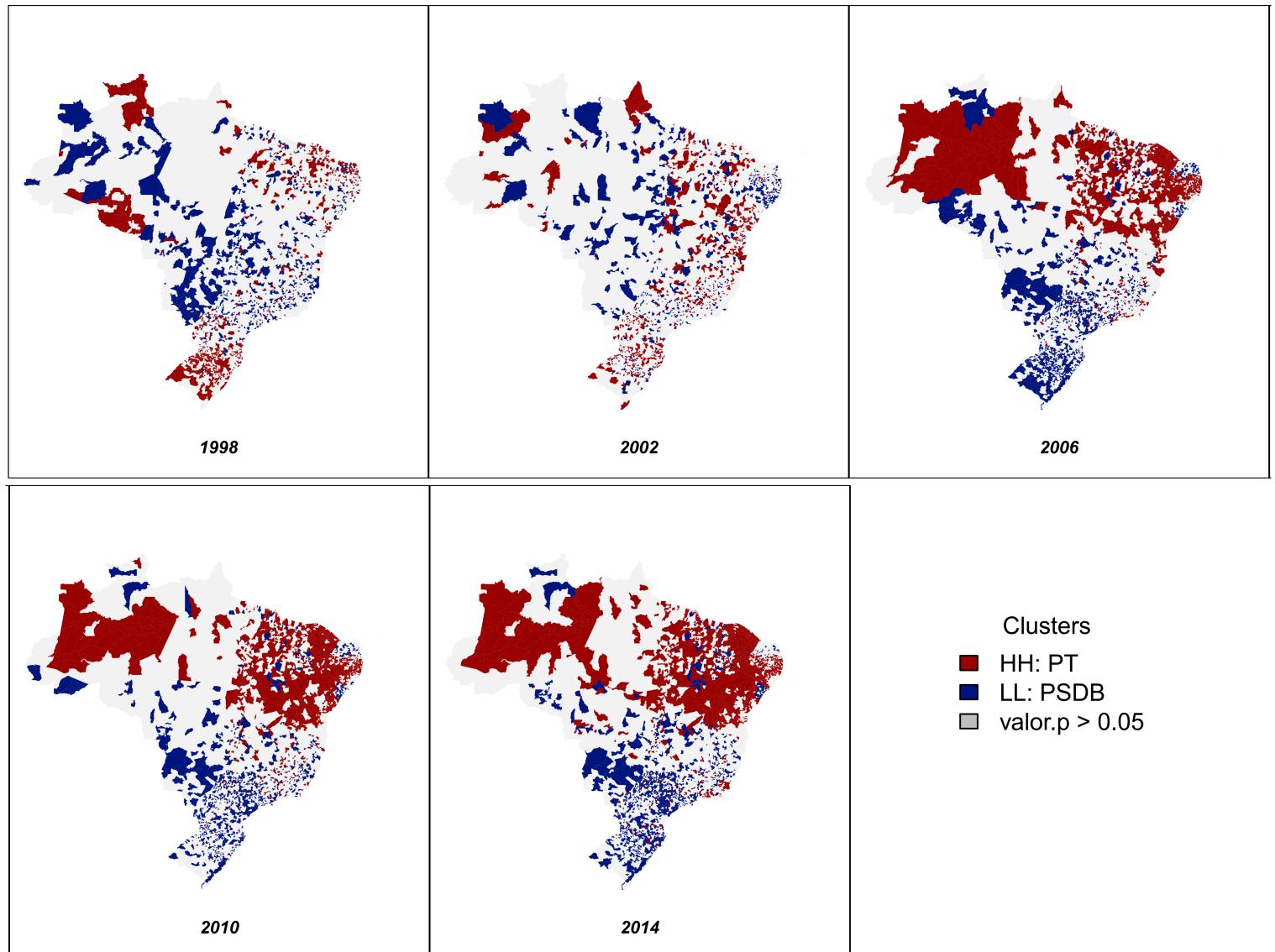


Figura 5: Índice de Moran local

No que diz respeito a dependência espacial, utilizamos o índice de Moran local e apresentamos seus resultados a partir do gráfico LISA, como é possível verificar em 5. Para melhor visualizar a dependência espacial consideramos apenas os relacionamentos que se mostraram fortes. Como após a padronização das diferenças as porcentagens mais altas de votos no PSDB são dadas por valores próximos a -1 e

no PT próximo a 1, temos que o HH representa a relação alta-alta(agrupamento de valores altos e próximos espacialmente) para o caso do PT e LL ou baixa-baixa (agrupamento de valores baixos e próximos espacialmente), representando dependência espacial quanto aos votos no PSDB. Dessa forma, verifica-se que o indicador de dependência espacial confirma os resultados expostos nos mapas anteriores.

Por fim, iremos apresentar a análise descritiva referente as covariáveis empregadas no modelo de regressão beta, para investigar o voto no PT no ano de 2010, a ideia desse gráficos foi demonstrar o comportamento dessas covariáveis diante da resposta e das grandes regiões.

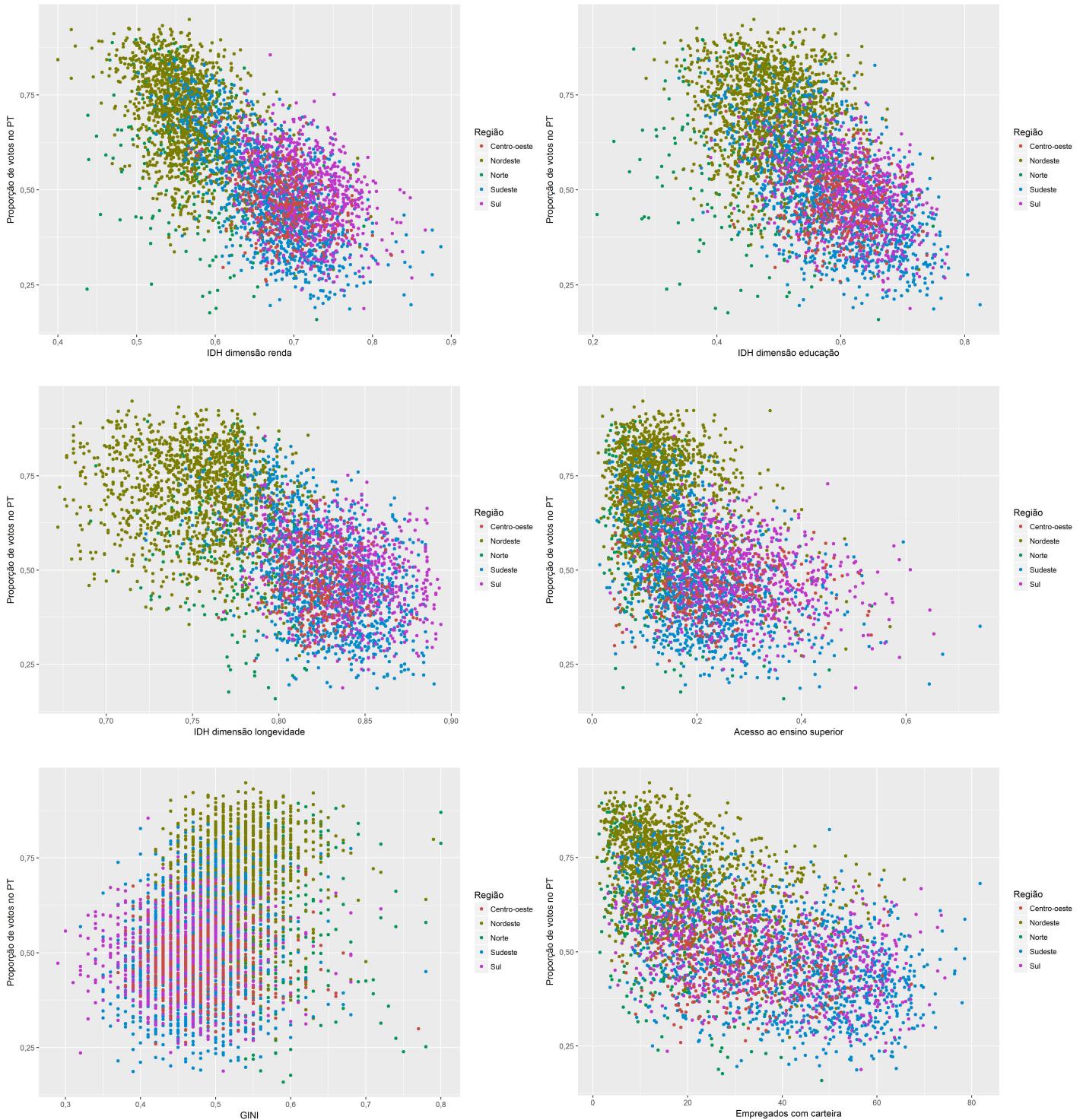


Figura 6: Resposta vs covariáveis

Nos gráficos expostos em 6 identificamos relações bastante semelhantes, no que tange a proporção de votos no PT e as covariáveis, por região. De modo geral, observamos que conforme os valores dos indicadores aumentam, a proporção de votos no PT diminui, além disso, verificamos que Norte e Nordeste estão ligados a maiores votos no PT e menores valores quanto aos indicadores. Sendo que Sul e Sudeste apresentam comportamentos bastante distintos destas duas regiões, o Centro-Oeste

aparece mais localizado nas regiões centrais dos gráficos. O único indicador que não apresenta um comportamento muito definido é o GINI, neste não foi possível verificar uma tendência crescente ou decrescente com a proporção de votos no PT.

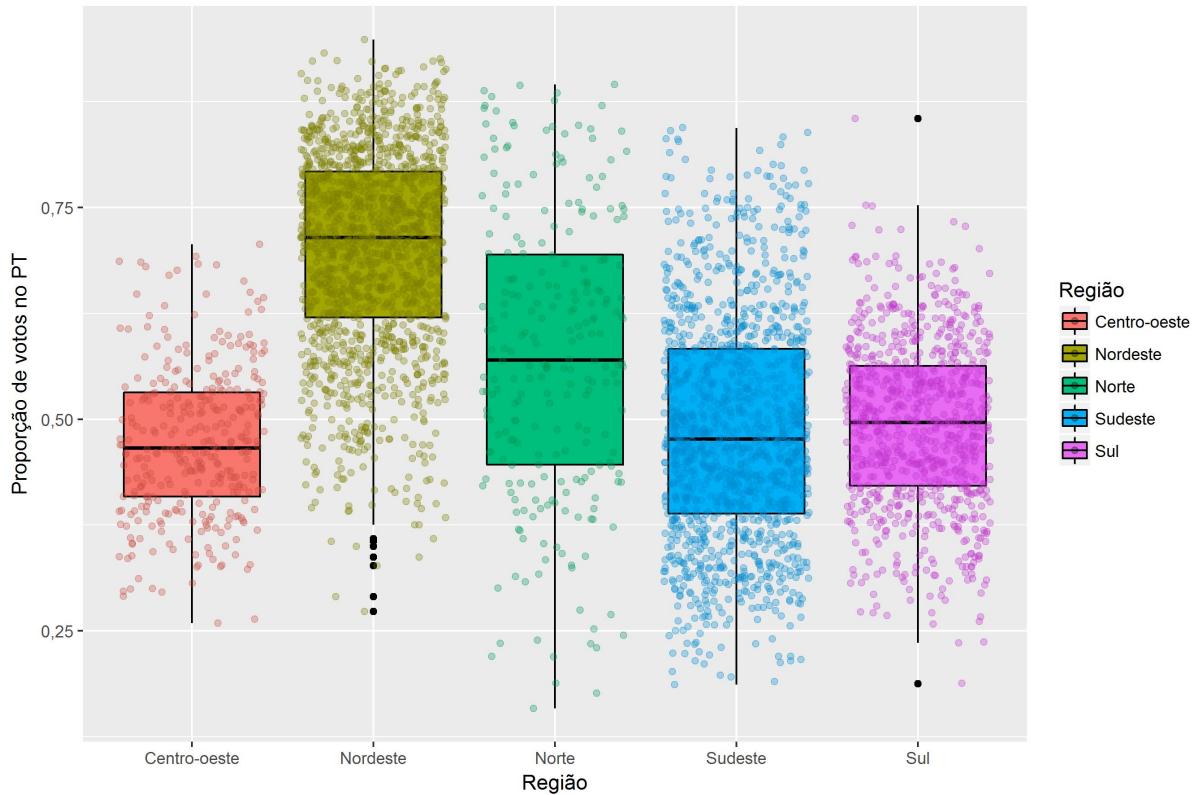


Figura 7: Boxplot da proporção de votos no PT segundo as grandes regiões

Para finalizar a parte descritiva, apresentamos um gráfico de boxplot da proporção de votos no PT segundo as grandes regiões visto em 7. Pelos resultados, constatamos de modo mais direto o que já havia sido levantado anteriormente, as médias de voto para o PT não maiores para o Norte e principalmente para o Nordeste, além do mais, neste último é possível verificar uma menor variabilidade da proporção que quando comparado ao Norte, as demais regiões apresentaram resultados bastante semelhantes.

3.2.2 Modelo de regressão Beta

Nesta seção nos ocupamos em apresentar os resultados do modelo de regressão Beta, antes de seguir para as estimativas do modelos, foi testada a necessidade de empregar a estrutura de precisão, assim, como expõe Almeida e Souza(2015) o modelo de regressão beta, pode admitir estrutura de regressão para o parâmetro de precisão. Dessa forma, é preciso identificar se a precisão é fixa, ou seja, se existe estrutura de regressão para o parâmetro de precisão ϕ . Tendo isto em vista, foi realizado TRV sobre um modelo com precisão variável versus em que a precisão é fixa, $H_0 : \phi_1 = \dots = \phi_q$ o teste é dado por $TRV = -2\ln(\ell_1) + 2\ln(\ell_2)$, em que ℓ_1 representa o modelo com precisão variável adicionada e ℓ_2 o modelo reduzido ou com precisão fixa.

Pelos resultados, foi possível rejeitar a hipótese nula ao nível de significância nominal de 0,05. Dessa forma, além de modelar a média, foi preciso modelar o

Tabela 2: TRV para o parâmetro de precisão

	#gl	LogLik	gl	Chisq	Valor-p
1	15	3503.46			
2	19	3625.29	4	243.67	0.0000

parâmetro de precisão, considerando as regiões. Assim, nosso modelo assume a seguinte forma:

$$\begin{aligned} Logit(\mu_t) = & \beta_0 + \beta_1(IDHMR) + \beta_2(IDHML) + \beta_3(IDHME) + \beta_4(TFBSUPER) + \\ & \beta_5(POPURB) + \beta_6(EMP) + \beta_7(LAT) + \beta_8(LONG) + \\ & \beta_9(Norte) + \beta_{10}(Nordeste) + \beta_{11}(Sudeste) + \beta_{12}(Sul) \end{aligned}$$

$$Log(\phi_t) = \gamma_0 + \gamma_1(Norte) + \gamma_2(Nordeste) + \gamma_3(Sudeste) + \gamma_4(Sul)$$

É válido ressaltar que utilizamos a região Centro-Oeste como referência. Além disso, a inclusão da latitude e da Longitude na análise se deu para a verificação da dependência espacial no modelo, como sugerem Almeida Junior e Souza (2015). Partindo disso, iremos verificar as estimativas obtidas, bem como seu efeito e contribuição para a variável resposta, proporção de votos no PT, no primeiro turno das eleições de 2010.

Tabela 3: Modelo para a média com ligação Logit

Parâmetro	Estimativa	Erro padrão	Z	Valor-p
(Intercept)	2.21*	0.25	8.73	<0.0001
IDHM_R	-4.20*	0.25	-16.92	<0.0001
IDHM_L	0.0003	0.31	0.01	0.99
IDHM_E	-0.28	0.15	-1.81	0.07
T_FBSUPER	0.22*	0.10	2.15	0.03
GINI	0.77*	0.13	6.04	<0.0001
POPURB	-0.32*	0.04	-7.74	<0.0001
EMP	-1.15	0.76	-1.52	0.13
LAT	-0.01*	0.00	-6.28	<0.0001
LONG	0.01*	0.00	4.67	<0.0001
Nordeste	0.44*	0.04	10.36	<0.0001
Norte	-0.39*	0.06	-6.80	<0.0001
Sudeste	0.20*	0.03	6.21	<0.0001
Sul	0.34*	0.04	8.48	<0.0001

Das covariáveis selecionadas para o modelo IDHMR, IDHME e EMP não foram significativas, no caso das demais foi possível rejeitar $H_0 : \beta_k = 0$ ao nível de significância nominal de 5%. No que diz respeito aos efeitos das estimativas continuas verificamos que IDHMR e POPURB apresentam efeitos negativos, ou seja, quanto menor o valor desses indicadores maior o proporção de votos no PT, já T_FLSUPER e GINI retornam efeitos positivos. Quanto as regiões, constatamos que ser do Norte implica em diminuição na proporção de votos no PT quando comparado a CentroOeste, nas demais regiões verificamos efeitos positivos do voto

no PT quando comparado a região de referência, sendo o Nordeste a região que apresenta a maior contribuição. Além disso, é possível verificar que o pseudo R^2 apresenta um valor de 0.5667, apesar de não ter a mesma interpretação do que na regressão linear, essa medida também nos indica a capacidade do ajuste do modelo.

Tabela 4: Modelo para a precisão com ligação Log

Parâmetro	estimativa	Erro padrão	Z	Valor-p
(Intercept)	3.42	0.08	44.55	0.00
Nordeste	-0.46	0.09	-5.38	0.00
Norte	-1.32	0.11	-11.55	0.00
Sudeste	-0.24	0.09	-2.73	0.01
Sul	-0.02	0.09	-0.23	0.82

Apresentamos também as estimativas dos parâmetros de precisão exposta na tabela 4, assim verifica-se que o fato de ser do Sudeste e Norte quando comparado ao CentroOeste diminui a precisão da proporção de voto no PT.

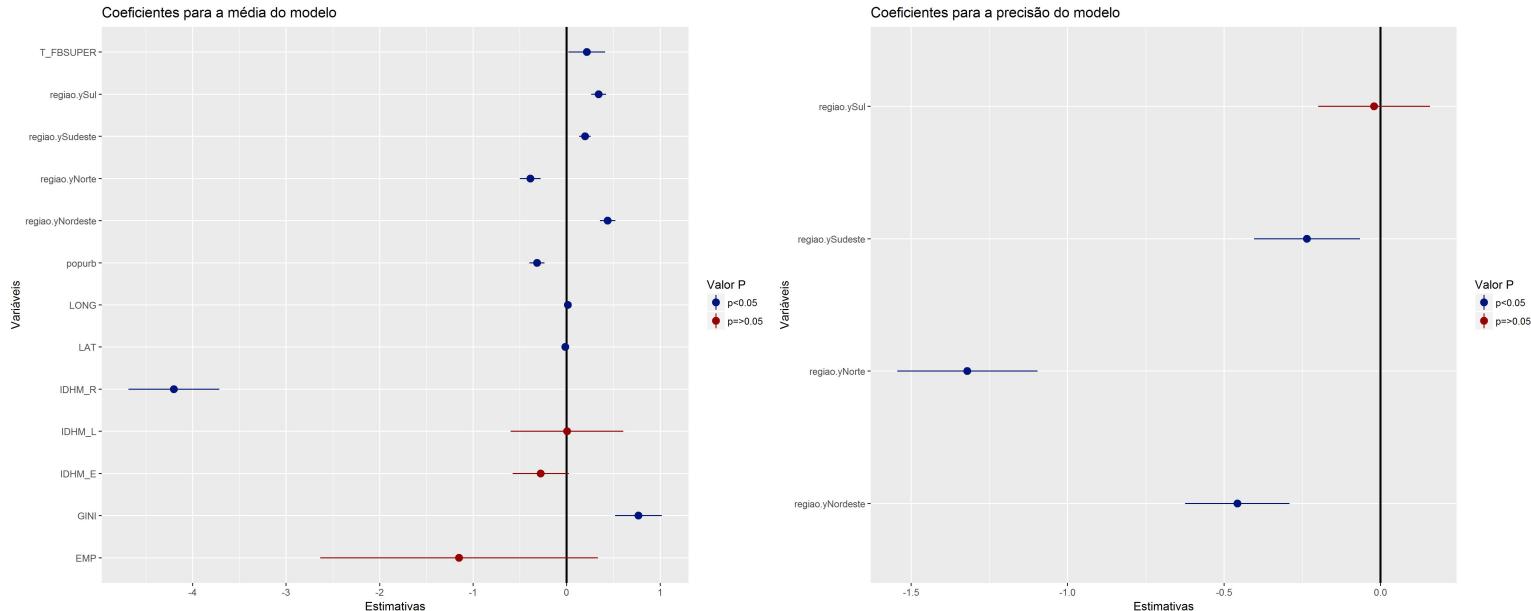


Figura 8: Efeito e intervalo de confiança das estimativas

Para facilitar a visualização dos efeitos, bem como os intervalos de confiança e a rejeição da hipótese nula, foram realizados gráficos dos coeficientes estimados expostos na figura 8. O ponto mais interessante deste gráfico é verificar a não rejeição da hipótese nula que aparece para as estimativas em vermelho, uma vez que, fica claro observar que o zero está presente no intervalo, pois ele toca a linha vertical fixada em zero.

3.2.3 Análise de resíduos

Nesta subseção nos ocupamos em realizar uma discussão sobre a adequabilidade do modelo com base nos resíduos. Sendo uma das fases mais importantes do processo de modelagem, uma vez que, as suposições que serão feitas sobre o modelo ajustado precisam ser validadas para que as inferências sejam confiáveis. Para a investigação gráfica dos resíduos utilizamos os resíduos propostos por Espinheira, Ferrari, and Cribari-Neto (2008b) chamados de standardized weighted residual 2, dados por:

$$r_i^{SW2} = \frac{(y_i - \hat{\mu}_i)}{\sqrt{\hat{\nu}_i(1 - h_{ii})}} \quad (9)$$

em que $y_i = \log(y_i/(1 - y_i))$ e $\mu_i = F(\mu_i, \phi) - F((1 - \mu_i), \phi)$, sendo $F(\cdot)$ uma função digamma.

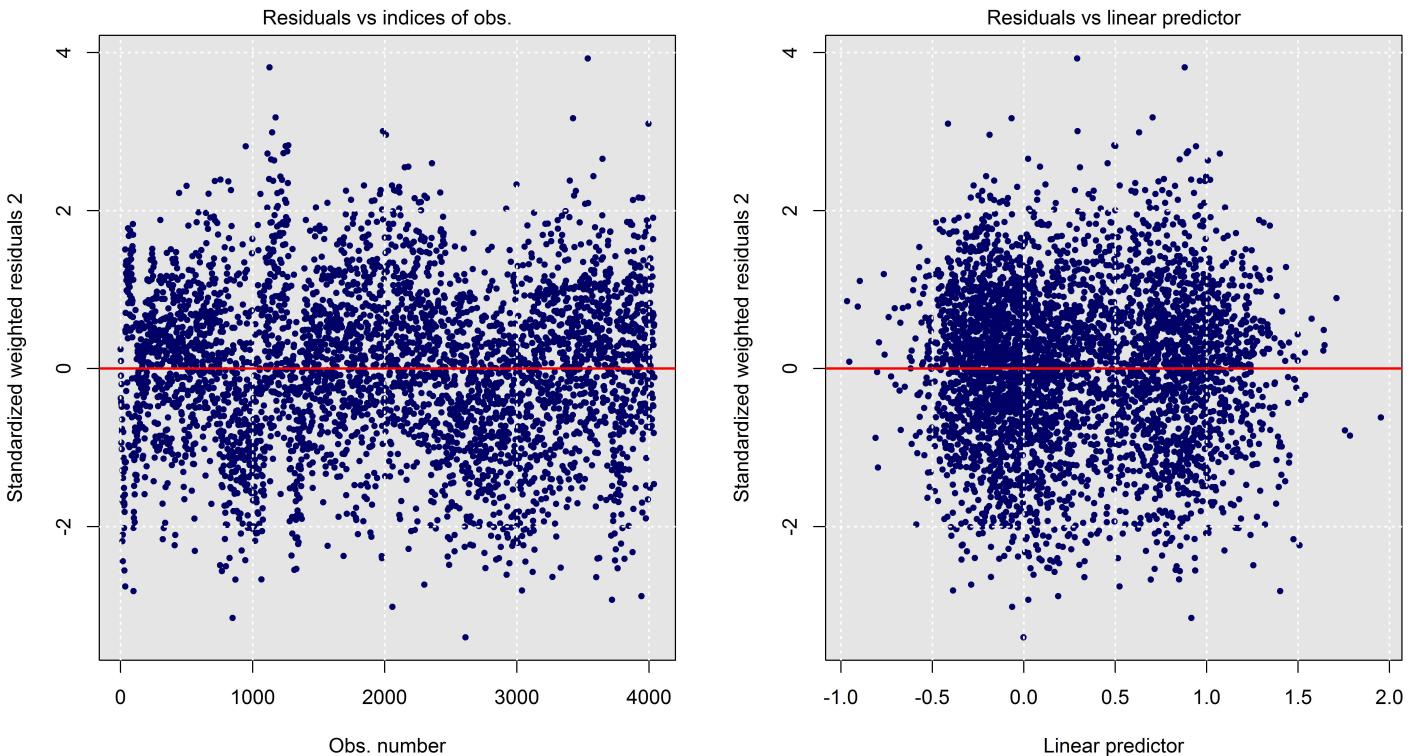


Figura 9: Resíduos de sw2

Partindo disso, pelos gráficos expostos na figura 9 pode-se notar que para gráfico o preditor linear vs os resíduos ponderados padronizados verificamos uma distribuição aparentemente aleatória em torno de zero, ou seja, não há uma tendência clara nos resíduos. Desse modo, podemos concluir que eles se organizam de forma aleatória, apontando para a não existência de homoscedasticidade. Quanto ao primeiro gráfico, mesmo que neste caso, não faça muito sentido verificar se os resíduos são independentes, foi elencada está técnica que nos garantiu independência entre as observações.

Pelo gráfico de Leverage observamos que existem alguns pontos influentes de alavancagem que precisam ser investigados e pela distância de Cook

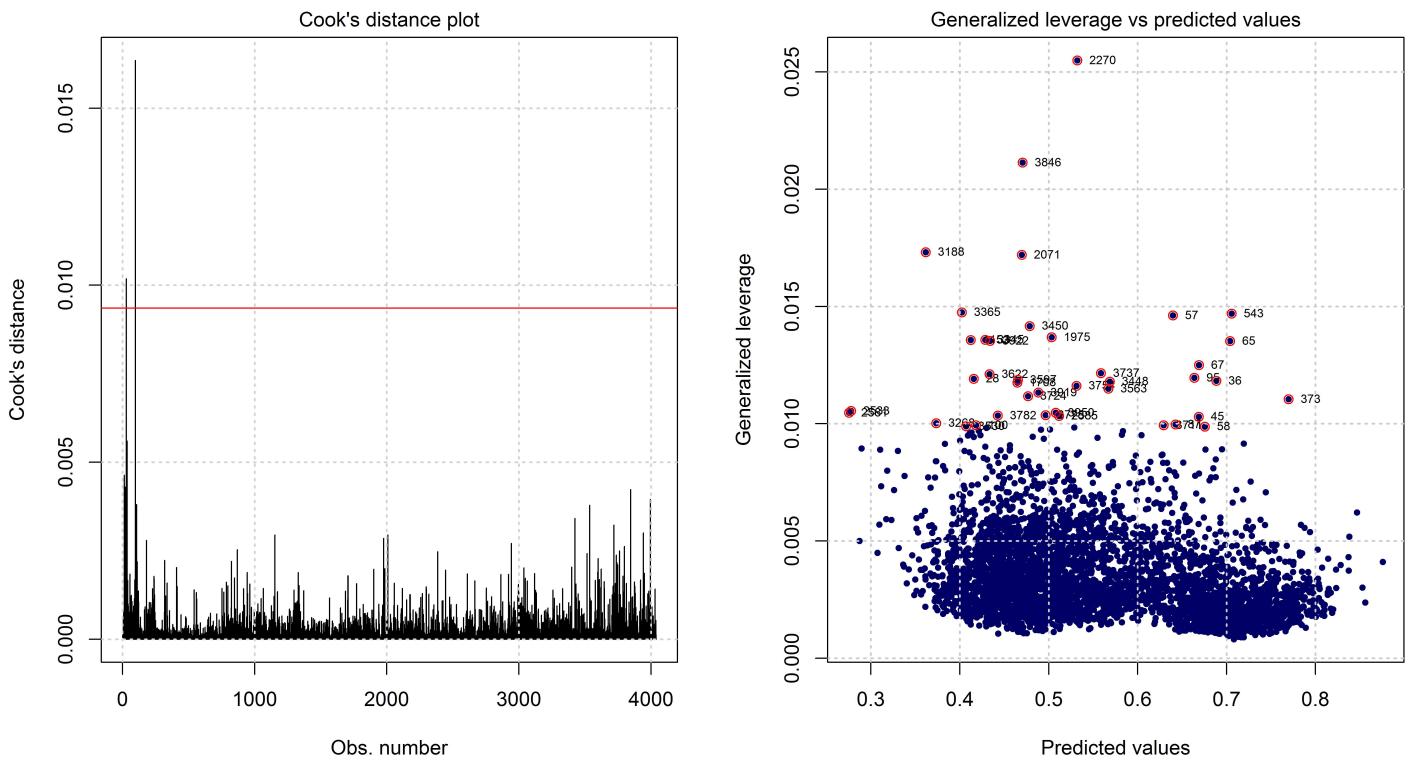


Figura 10: Distância de Cook e Leverage

observamos 2 pontos que podemos considerar como outliers. Para verificar se estas observações influenciam nas estimativas e conclusões obtidas, realizamos sua exclusão do banco e as comparamos com os resultados do modelo completo, contendo todas as observações. Os resultados obtidos foram bastante similares, desse modo, não constatamos a necessidade de sua retirada na modelagem.

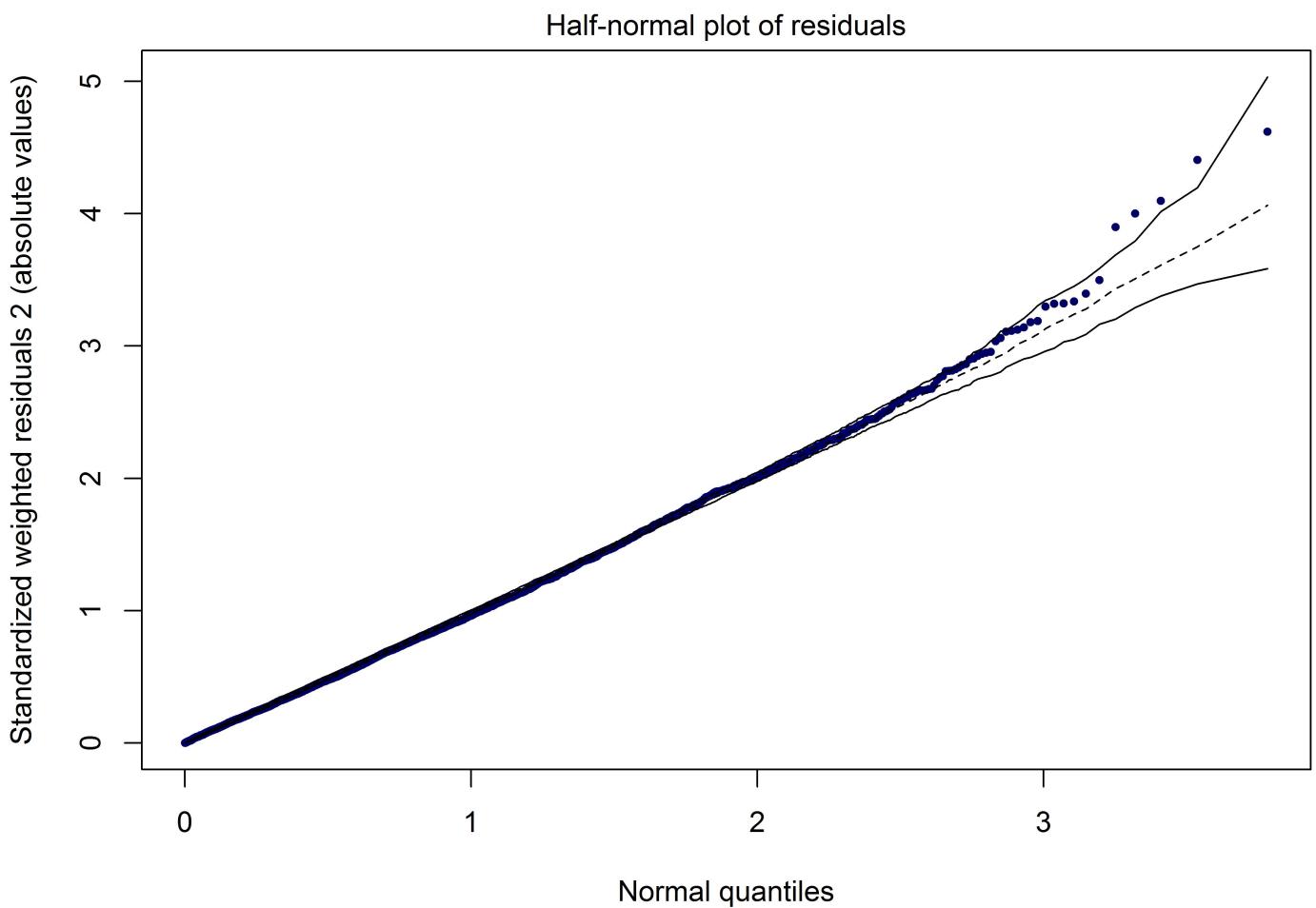


Figura 11: Envelope simulado

Aqui, temos o gráfico de envelope simulado, em que constatamos que há apenas alguns pontos que se distanciam da banda de confiança ao fim da reta teóricas apenas, porém, de um modo geral temos que o ajuste do nosso modelo foi bastante satisfatório.

4 Conclusão

Como já foi exposto, o objetivo deste trabalho foi verificar a existência da dependência espacial do voto, nas eleições presidenciais de 1998, 2002, 2006, 2010 e 2014 para o PT e PSDB, bem como, os determinantes do voto no PT em 2010. Partindo dos resultados, foi possível verificar a existência da dependência espacial do voto para ambos os partidos, principalmente na região norte e nordeste, no que diz respeito ao voto no PT e no Sul e Sudeste, no que tange ao PSDB após 2006.

Quanto ao modelo de regressão Beta referente a proporção de votos no PT no primeiro turno das eleições de 2010, verificamos a necessidade de admitir uma estrutura de regressão para o parâmetro de precisão, considerando as grandes regiões como covariáveis para tal. Partindo disso as estimativas obtidas apontaram que o IDHMR, IDHME e EMP não foram significativas para modelo. Já IDHMR e POPURB apresentam efeitos negativos, quanto a T_FLSUPER e GINI retornam efeitos positivos. Verificamos também, a existência de dependência espacial no modelo, uma vez que, as estimativas de latitude de longitude foram significativas. Por fim, no que tange as grandes regiões, constatamos que ser do Norte implica em diminuição na proporção de votos no PT quando comparado a Centro-Oeste, nas demais regiões verificamos efeitos positivos do voto no PT quando comparado a região de referência.

Referências

- Bivand, Roger S., Edzer Pebesma, and Virgilio Gomez-Rubio. Applied Spatial Data Analysis with R. Springer New York, 2013.
- Gimenes, E. R., Furriel, W. O., Borba, J., e Ribeiro, E. A. (2016). Partidarismo no Brasil: Análise longitudinal dos condicionantes da identificação partidária (2002-2014). Revista Debates, 10(2), 121.
- Abensur, Themis C., Francisco Cribari-Neto, and Tatiane A. Menezes. Impactos do Programa Bolsa Família nos resultados das eleições presidenciais no Brasil em 2006. Anais do XXXV Encontro Nacional de Economia 51 (2007).
- Dobson, Annette J., and Adrian Barnett. An introduction to generalized linear models. CRC press, 2008.
- Sileshi, G. Selecting the right statistical model for analysis of insect count data by using information theoretic measures. Bulletin of entomological research 96.5 (2006): 479-488.
- Casella, George, and Roger L. Berger. Statistical inference. Vol. 2. Pacific Grove, CA: Duxbury, 2002.
- Paula, A. Gilberto. Modelos de regressão com apoio computacional, USP, 2013.
- Agresti, Alan, and Barbara Finlay. Métodos estatísticos para as ciências sociais. 2012.
- Cameron, A. Colin, and Pravin K. Trivedi. Regression analysis of count data. Vol. 53. Cambridge university press, 2013.
- Ferrari, Silvia, and Francisco Cribari-Neto. Beta regression for modelling rates and proportions. Journal of Applied Statistics 31.7 (2004): 799-815.
- Demétrio, Clarice Garcia Borges. Modelos lineares generalizados em experimentação agronômica. USP/ESALQ, 2001.
- de Almeida Junior, Pedro M., and Tatiene C. Souza. "Estimativas de votos da presidente Dilma Rousseff nas eleições presidenciais de 2010 sob o âmbito do Bolsa Família." Ciência e Natura 37.1 (2015).
- Veiga, L. F. "Os partidos brasileiros na perspectiva dos eleitores: mudanças e continuidades na identificação partidária e na avaliação das principais legendas após 2002". Opinião Pública, Campinas, vol.13, nº 2, p. 340-365, nov.2007.
- Baquero, Marcello. Identificação Partidária e comportamento eleitoral. (Org.). Cultura(s) políticas(s) e democracia no século XXI na América Latina, 2012.
- Dalton, R.J. CitizenPolitics: Public Opinion And Political Parties In Advanced Industrial Democracies. 4th Ed. 2006.
- Peixoto, Vitor, and Lucio Rennó. "Mobilidade social ascendente e voto: as eleições presidenciais de 2010 no Brasil." Opinião Pública 17.2 (2011): 304-332.
- Fiorina, M. Retrospective voting in American national elections. NewHaven: Yale University Press, 1981.
- BRASIL, Tribunal Superior Eleitoral. "Dados estatísticos." URL <http://www.tse.jus.br/eleicoes/eleicoes-anteriores>