



## Dinâmicas Regionais e Indicadores Sociais no Engajamento Político Brasileiro

### Resumo

A pesquisa foi feita sob os dados do Estudo Eleitoral Brasileiro (ESEB) de 2014, *survey* nacional pós eleitoral realizado pelo Centro de Estudos de Opinião Pública (CESOP), que contém informações sobre a participação política dos cidadãos amostrados bem como sobre questões socioeconômicas. O objetivo da pesquisa foi relacionar tais questões com a autodeclaração de interesse em política, através de uma relação probabilística entre elas proposta por um modelo de regressão logística, e posteriormente, identificar nos resíduos deste modelo indícios de fonte de variabilidade espacial por meio da análise de dispersão espacial e ferramentas como variograma e “kriging”.

No primeiro semestre, o trabalho consistiu em estudar o questionário aplicado pelo CESOP, selecionar as variáveis de interesse, realizar análise descritiva das informações e ajustar um modelo de regressão logística adequado. O melhor modelo obtido relaciona a variável “quão interessado em política o entrevistado se autodeclara” com as variáveis “nível de escolaridade”, “quão satisfeito o entrevistado está quanto à oportunidade de acesso ao ensino superior”, “quanto o entrevistado acha que seu voto influencia no que acontece no Brasil” e “log-renda”. Quanto ao nível de escolaridade, a maior discrepância entre as probabilidades de se autodeclarar interessado em política ocorreu entre o nível “ensino universitário completo” e os demais, ou seja, a significância desta variável com relação à explicação do interesse político autodeclarado é provocada por este nível. Com relação à segunda e terceira variáveis, quanto maior a satisfação quanto à oportunidade de acesso ao ensino superior e quanto maior a influência que o entrevistado acredita que seu voto tem sob o que acontece no Brasil, maior a probabilidade de se autodeclarar interessado em política. Esta correlação positiva também acontece para a quarta e última variável explicativa: quanto maior o log-renda, mais interessado em política o entrevistado se autodeclara.

É interessante discutir sobre o fato de que embora exista relação probabilística entre as variáveis, identificada e mensurada através do modelo de regressão logística, o interesse em política, bem como diversas outras variáveis do questionário, foram obtidas através de respostas dos entrevistados sobre si mesmos. Desta forma, cada pessoa toma seus critérios para a autoavaliação de seu interesse em política, por exemplo, ou de sua classe social, satisfação com políticas de educação, entre outros. Ou seja, apesar da significância de qualquer resultado, é importante ressaltar que existe individualidade por trás de todas as respostas e nada é completamente explicado pelas variáveis coletadas e incluídas no modelo, pois inclusive sobre elas pode-se ter um grau de incerteza.

Sob os *deviance residuals* do modelo, um variograma empírico, que para cada distância  $h$  possível entre pontos, calcula a média dos quadrados das diferenças entre resíduos separados por esta distância, foi ajustado por um modelo de variograma teórico exponencial. O variograma empírico e seu ajuste estão apresentados na Figura 1.

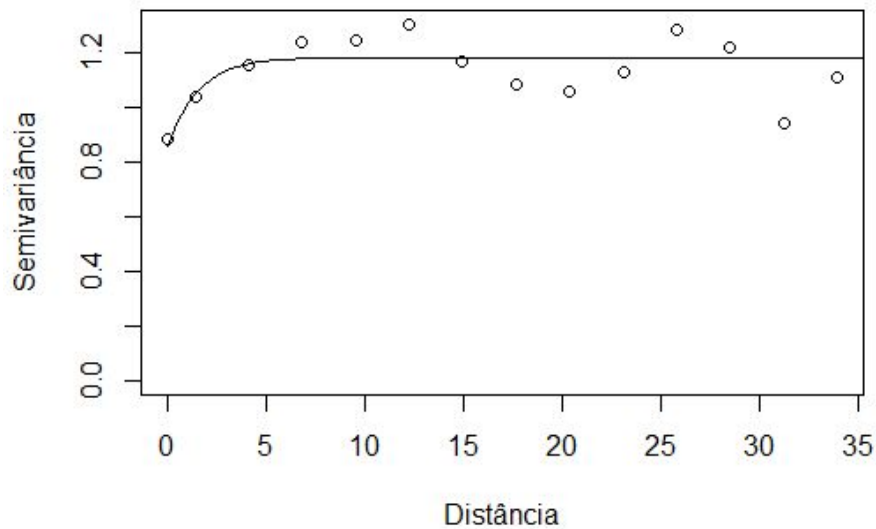


Figura 1: Variograma dos resíduos do modelo logístico utilizando todos os municípios. A distância é aproximadamente em centenas de quilômetros (projeção cartográfica do pacote *sp*).

Além disso, através dos componentes estimados da curva de ajuste (efeito pepita, patamar e dependência) e dos valores obtidos no modelo, é possível realizar uma predição suavizada dos resíduos ao longo do território cobrindo pontos com ausência de informações através do “kriging”, como é apresentado na Figura 2.

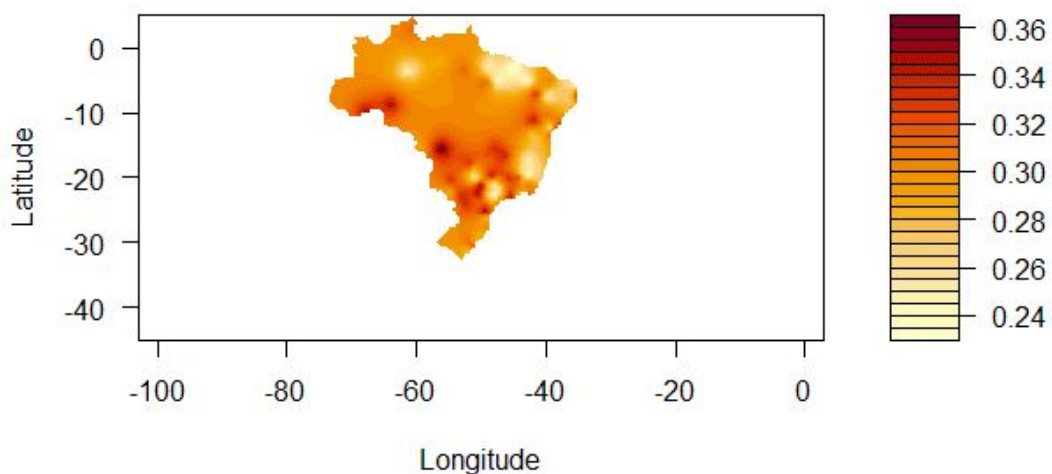


Figura 2: Predição espacial dos resíduos no território nacional

Feito a modelagem da variável de interesse e do efeito espacial nos resíduos, agora este efeito é incluído num novo modelo como uma nova variável explicativa. A ideia é de que o conhecimento de uma nova fonte de variabilidade, identificada através do espaço geográfico, implique numa diminuição da ignorância sobre a relação entre a variável resposta e as explicativas, expressa por meio da diminuição da amplitude dos resíduos.

Dados obtidos nos sites do Tribunal Superior Eleitoral (TSE) e do Instituto Brasileiro de Geografia Estatística (IBGE) com informações sobre escolaridade de eleitores e renda per capita municipal, bem como os valores preditos dos resíduos pela suavização da Figura 2, foram consolidados para serem utilizados para prever o nível de interesse político. Um novo modelo feito sob os dados do ESEB com uso restrito às variáveis “escolaridade” e “log-renda”, sem perda de suas significâncias quando comparado ao primeiro modelo, foi ajustado a fim de que a predição fosse a nível municipal. Isto foi feito levando em consideração que as variáveis “quão satisfeito o entrevistado está quanto à oportunidade de acesso ao ensino superior” e “quanto o entrevistado acha que seu voto influencia no que acontece no Brasil” são pessoais e de difícil estimação, e que elas estejam aleatoriamente distribuídas ao longo da população de um município. Além disto, algumas alterações foram feitas para corrigir problemas de granularidade, ou seja, incompatibilidade entre níveis de variáveis, a fim de que os dados pudessem ser utilizados pelo modelo para predição das probabilidades.

Dois conjuntos de valores preditos foram ajustados: Um considerando nula a predição do efeito espacial, e o outro a incrementando no cálculo. A Figura 3 apresenta à esquerda o primeiro conjunto de valores, e à direita o segundo. Na comparação é preciso atentar-se que existe diferença nas escalas de cores dos mapas. Percebe-se que o incremento da informação que gera a variabilidade espacial suaviza os valores preditos, o que não é um resultado positivo para o estudo. Entretanto, a Figura 4 apresenta a dispersão territorial do desvio padrão dos valores estimados do efeito espacial, e é possível notar que ainda nos pontos em que o desvio é mínimo, ele é alto quando comparado aos valores estimados, que giram em torno de 0,24 a 0,36.

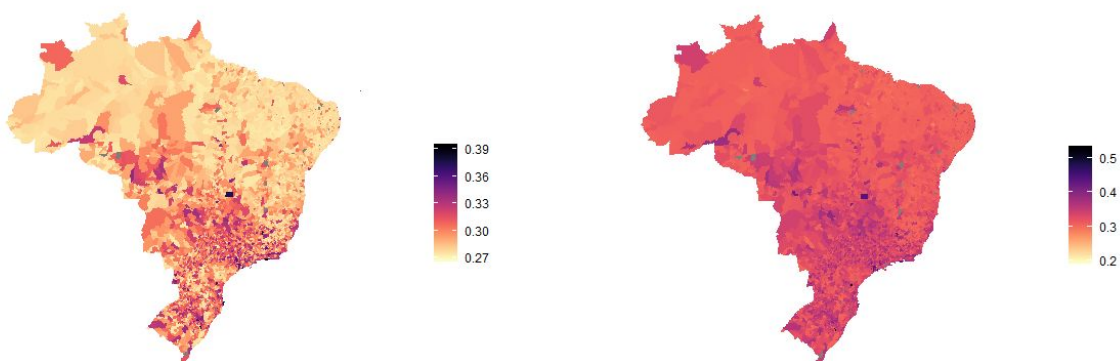


Figura 3: Dispersão territorial das probabilidades preditas de interesse municipal em política

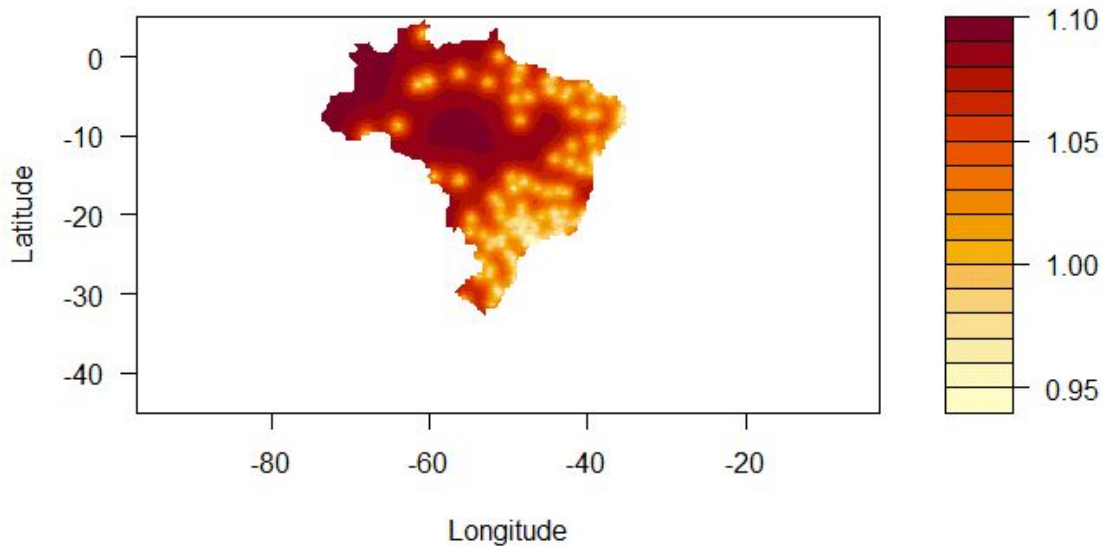


Figura 4: Dispersão territorial do desvio padrão do efeito especial ajustado

### Discussão

A metodologia proposta pode ser mais facilmente aplicada a modelos de regressão normais pelo fato de que neles os resíduos são expressos linearmente através da curva de ajuste, e após identificado um efeito espacial, os resíduos podem ser decompostos aditivamente em dois componentes, um que represente a variabilidade espacial agora conhecida, e outro que represente o erro puro restante. Desta forma, ao passo que se reconhece a existência de uma fonte de variabilidade, identificada e representada no espaço geográfico, diminui-se a ignorância quanto aos resultados que podem ser obtidos através do modelo. No caso logístico, os resíduos são função da razão de verossimilhança entre os valores reais e os ajustados, e não compõem a curva de regressão. Desta forma, a alternativa tomada foi a de incrementar a suavização dos resíduos ao longo da malha territorial como uma nova variável explicativa do modelo, e dada sua alta significância, utilizar os coeficientes estimados para a previsão de resultados. Entretanto este procedimento, bem como a alta variância do efeito espacial, podem ter influenciado para uma alta suavização de valores preditos.

Outra constatação importante é a de que incluir o efeito espacial obtido dos resíduos de um modelo linear não altera significativamente os coeficientes de regressão. Isto evita o problema de *confundimento espacial* que acontece quando uma covariável significativa representa a mesma fonte de variabilidade identificada espacialmente, que existiria em uma modelagem simultânea dos efeitos \cite{hodges}.

## Referências

- [1] A. Agresti. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, 2018.
- [2] R. S. Bivand, E. J. Pebesma e V. Gómez-Rubio. *Applied Spatial Data Analysis with R*. Springer, 2008.
- [3] A. L. V. Dias e M. T. M. Kerbauy. “Engajamento cívico e escolaridade superior: as eleições de 2014 e o comportamento político dos brasileiros”. *Revista de Sociologia e Política*, pp. 149--181, 2015.
- [4] Federal Communications Commission. “Reference points and distance computations”. *Code of Federal Regulations (Annual Edition). Title 47: Telecommunication*. 73 (208), 2016. URL: <https://www.govinfo.gov/content/pkg/CFR-2016-title47-vol4/pdf/CFR-2016-title47-vol4-sec73-208.pdf>.
- [5] R. Meneghello *et al.* Estudo Eleitoral Brasileiro. URL: <https://www.cesop.unicamp.br/por/eseb>
- [6] J. S. Hodges e B. J. Reich. “Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love”. *The American Statistician*, pp. 325--334, 2010.
- [7] Instituto Brasileiro de Geografia Estatística. \textit{Pesquisa Nacional por Amostra de Domicílios.} URL: <https://www.ibge.gov.br/estatisticas/sociais/educacao/9127-pesquisa-nacional-por-amostra-d-e-domicilios.html?edicao=18338&t=publicacoes>
- [8] S. L. Lohr. *Sampling: Design and Analysis*. Nelson Education, 2009.
- [9] E. J. Pebesma e R. S. Bivand. “Classes and methods for spatial data in R”. *R News*, 5(2), 9--13. URL: <https://cran.r-project.org/web/packages/sp/index.html>
- [10] Tribunal Superior Eleitoral. URL: <http://www.tse.jus.br/>.
- [11] W. N. Venables e B. D. Ripley. *Modern Applied Statistics with S-PLUS*. Springer, 2013.