

Prevendo o resultado da eleição presidencial brasileira de 2014

Neale Ahmed El-Dash – Sleek Data
neale.eldash@gmail.com

Introdução

Modelos de previsão eleitoral tentam prever o resultado de uma eleição. Tais previsões podem ser feitas muito antes da eleição (meses ou anos antes) e muito próximas ao dia da eleição (alguns dias antes).

Usualmente, os modelos de previsão eleitoral se dividem em dois tipos: **modelos de agregação**, os quais agregam pesquisas eleitorais, e os **modelos estruturais**, os quais incluem como preditoras variáveis econômicas e conjunturais. Esses dois tipos de modelos aspectos positivos e negativos bem distintos entre si. Os modelos de agregação podem ser vistos como dinâmicos, eles são capazes de captar todas as mudanças na intenção de voto que ocorrem durante um ciclo eleitoral. Por outro lado, os modelos estruturais conseguem utilizar informação histórica e conjuntural sobre as eleições, captando tendências que se repetem em diferentes ciclos eleitorais.

Nesse poster esses diferentes modelos serão apresentados, e discutirei como combinar esses dois tipos de modelo utilizando inferência bayesiana, permitindo que previsões sejam feitas para qualquer dia do ciclo eleitoral. Também discutirei como os modelos da literatura foram adaptados para o caso particular do Brasil.

Modelo Estrutural (Abramowitz)

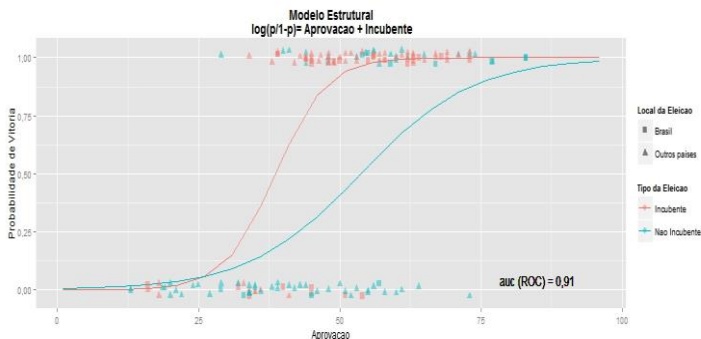
Um modelo estrutural bastante conhecido é o “time-for-change model”, de Abramowitz [1]. Nesse modelo o autor prevê o percentual de voto no candidato do governo, utilizando como variáveis explicativas o PIB americano no segundo trimestre do ano da eleição, a aprovação do atual presidente (no mês de Junho) e se o candidato do governo é o incumbente ou sucessor. A suposição desse modelo é de que a eleição presidencial é um referendo popular sobre a performance do presidente incumbente.

Implementar um modelo similar no Brasil é inviável, pois houveram apenas 6 eleições presidenciais democráticas. Nos Estados Unidos, já houveram 57 (no modelo foram utilizadas 16). Como alternativa, foi utilizado um modelo mais genérico, utilizando resultados de 157 eleições realizadas em 17 países (tabela 1), as quais incluem eleições de governadores e de presidentes.

Tabela 1 - Eleicoes utilizadas no modelo estrutural

País	Eleicoes	País	Eleicoes	País	Eleicoes
USA	93	Chile	3	Nigeria	1
Brasil	32	France	2	Paraguay	1
Canada	5	Mexico	2	Peru	1
UK	5	Bolivia	1	Russia	1
Colombia	4	Ecuador	1	Venezuela	1
Argentina	3	El Salvador	1		

Nesse modelo alternativo, a variável dependente utilizada é uma variável indicadora de vitória do candidato do governo. Foram utilizadas apenas duas variáveis explicativas: a aprovação do atual presidente e se o candidato do governo é o incumbente ou sucessor. A qualidade do ajuste do modelo logístico é bastante razoável - a área sobre a curva ROC (Receiver Operating Characteristic) é de 0,90. **Sob este modelo, a chance de vitória da Dilma (PT) na próxima eleição é de 77% (assumindo que a aprovação atual de 44% se mantera’).**



Modelo de agregação (Jackman)

A vantagem dos modelos de agregação é que com eles é possível captar as variações na intenção de voto durante o ciclo eleitoral. A possibilidade de avaliar o impacto de entrada/saída de candidatos, de escândalos, do desconhecimento dos candidatos pela maior parte da população no início da campanha eleitoral, torna esses modelos muito mais efetivos pra prever a intenção de voto durante o ciclo eleitoral.

Os modelos de agregação funcionam estimando a intenção de voto utilizando como base as n pesquisas eleitorais publicadas na mídia. Um dos modelos mais gerais foi proposto por Jackman em [2]. Nesse modelo hierárquico (bayesiano), cada pesquisa publicada é modelada como

$$y_i \sim N(\mu_i, \sigma_i^2)$$

com $i=1$ até n , onde a pesquisa i foi realizada pelo instituto j_i no dia t_j e σ_i^2 é a precisão da i -ésima pesquisa. Nesse modelo, supõem-se que a média depende de dois fatores de forma que

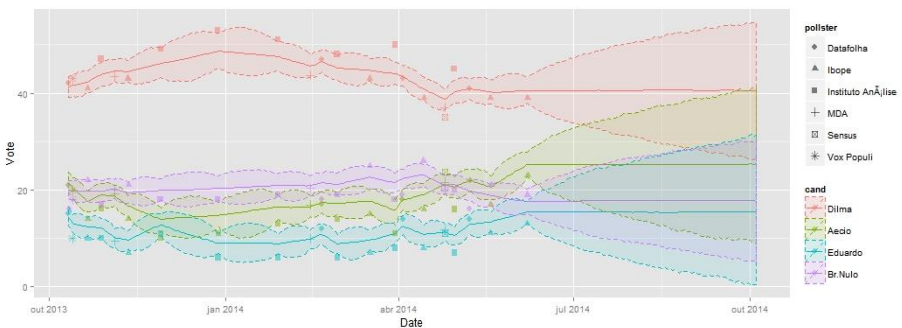
$$\mu_i = \alpha_{t_i} + \delta_{j_i}$$

onde α_{t_i} é a intenção de voto no dia t_j e δ_{j_i} efeito do instituto de pesquisa j_i que realizou a pesquisa. Incorporar o efeito do instituto que realizou a pesquisa é uma das características mais importante desse modelo, pois sabidamente os institutos utilizam diferentes metodologias as quais claramente tem impacto nas estimativas.

Também supõem-se que as diferenças de intenção de voto em dias consecutivos são originadas por um passeio aleatório, de forma que

$$\alpha_{t_i} \sim N(\alpha_{t_{i-1}}, \omega)$$

onde ω é a volatilidade do processo, ou seja, mede o quanto acreditamos que a intenção de voto pode variar. Aplicando esse modelo aos dados das pesquisas divulgadas (cenário com 3 candidatos) para a eleição presidencial brasileira de 2014, obtemos os resultados apresentados no gráfico abaixo, com projeções até o primeiro turno da eleição.



Esse modelo não foi desenvolvido pensando em fazer previsões futuras, inclusive o autor apenas o utiliza após o término da eleição. A dificuldade enfrentada pelo modelo fica clara no gráfico acima, pois as previsões de voto se mantêm constantes até a eleição. Essencialmente, o modelo supõem que nada se alterará entre a última pesquisa observada e a data da eleição. Se não houver nenhum novo “choque” no passeio aleatório, tudo se manterá igual. Claramente, quanto mais distantes estivermos da eleição, mais improvável é essa suposição.

Modelo Híbrido (Linzer)

Como foi mostrado anteriormente, os modelos estruturais são bons para prever os resultados das eleições levando em conta dados históricos e conjunturais, porém são estáticos. Por outro lado, os modelos de agregação são ótimos para estimarem a intenção de voto de forma dinâmica durante o ciclo eleitoral, porém fazem suposições muito fortes para prever corretamente o resultado de uma eleição distante no tempo.

Se o objetivo for realizar boas previsões tanto a curto quanto a longo prazo, levar em conta tanto fatores estruturais quanto os resultados de pesquisa publicadas na mídia, o ideal seria combinar os dois tipos de modelo. Em [3], Linzer propôs justamente um modelo híbrido dessa forma, baseado nos modelos estruturais e de agregação. A principal alteração com relação ao modelo de agregação do Jackman é utilizar uma priori informativa para a intenção de voto para o dia da eleição, sendo que essa priori é definida pelo modelo estrutural. Para fazer isso, é necessário utilizar um **passaio aleatório reverso** para modelar a intenção de voto em dias consecutivos. Com essa configuração, o passeio aleatório se inicia no dia da eleição, de forma que que

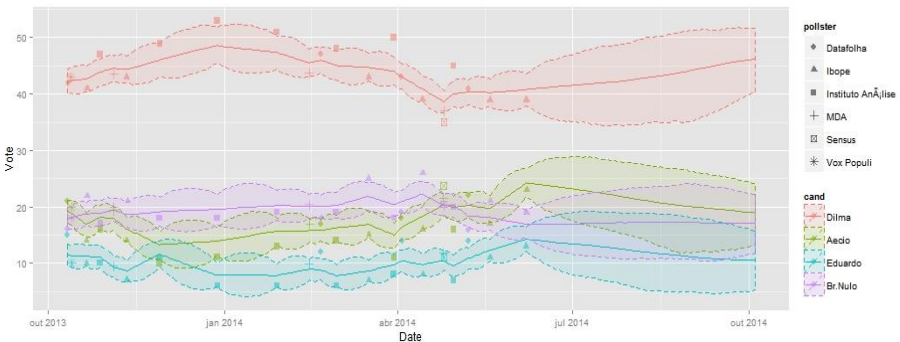
$$\alpha_{t_i} \sim N(\alpha_{t_{i+1}}, \omega) \text{ e } \alpha_T \sim N(\alpha, \sigma_\alpha^2),$$

onde T é o dia da eleição, α é a intenção de voto estimada pelo modelo estrutural e σ_α^2 é a precisão dessa estimativa. Quanto mais precisa for a priori, mais tempo e mais pesquisas serão necessárias para dominar a estimativa do modelo estrutural, permitindo ao analista controlar a “velocidade” com que as estimativas provenientes de modelos diferentes se misturam.

A maior dificuldade de implementação desse modelo se faz pelo fato de que o modelo estrutural apresentado nesse poster modela a probabilidade de vitória do candidato do governo, e a priori informativa do modelo híbrido é referente as intenções de voto para todos os candidatos. Apenas como **solução temporária (e passível de diversas críticas)** para essa questão, a probabilidade de vitória do candidato do governo foi transformada em percentual de votos utilizado uma função “empírica”, baseada nos resultados históricos das eleições presidenciais brasileiras:

$$p_{gov} = \min(p) + (\max(p) - \min(p)) * Prob.Vitória_{gov},$$

onde $\min(p)$ e $\max(p)$ representam, respectivamente, o menor e o maior percentual de votos já recebidos pelos candidatos governistas. Para os outros candidatos, assumiu-se que as proporções de votos válidos entre eles seriam iguais a pesquisa mais recente. No gráfico abaixo as previsões do modelo híbrido são apresentadas para o primeiro turno da eleição brasileira de 2014.



Referências

- [1] Abramowitz, A. I. (2008), “Forecasting the 2008 Presidential Election With the Time-for-Change Model,” PS: Political Science & Politics, 41, 691–695
- [2] Jackman, S. (2005), “Pooling the Polls Over an Election Campaign,” Australian Journal of Political Science, 40, 499–517
- [3] Drew A. Linzer (2013) Dynamic Bayesian Forecasting of Presidential Elections in the States, Journal of the American Statistical Association, 108:501, 124-134