



Modelos para tendências de intenção de voto



IME - Instituto de Matemática e Estatística

André Raz F. de Santana (andre.santana@usp.br)
Julio da Motta Singer (jmsinger@ime.usp.br)

Jane Simões de Castro (jane.castro@usp.br)
Tuany de Paula Castro (tuany.castro@usp.br)

IME - Universidade de São Paulo - Brasil

Introdução

- O trabalho se origina de artigos da revista **Carta Capital** publicados entre março e agosto de 2010;
- **Dados** de pesquisas eleitorais conduzidos por **DataFolha**, **Sensus**, **Ibope**, **Vox Populi**, entre fevereiro de 2008 e setembro de 2010;
- **Modelos polinomiais** para tendências de intenção de voto na eleição presidencial de 2010 (1º turno);

Modelo publicado pela revista Carta Capital:

- Supõe homocedasticidade;
- Supõe equi-espacamento entre pesquisas;
- Não impõe restrição de soma 100% para proporções de intenção de votos nos diferentes candidatos;
- Apresenta gráficos não compatíveis com a curva ajustada;
- permite previsões inadequadas (por exemplo, -6.9% para Serra em 03/10/2010);



Objetivo

Propor modelos alternativos que contemplem deficiências do modelo polinomial apresentado.

Regressão por mínimos quadrados ponderados

Modelo polinomial:

$$p_{ij} = \alpha_i + \beta_i t_j + \gamma_i t_j^2 + e_{ij}, \quad i = 1, 2, \quad j = 1, 2, \dots, n,$$

em que

p_{ij} : proporção de eleitores favoráveis ao i -ésimo candidato na j -ésima pesquisa de intenção de votos,

$i = 1$ corresponde à Dilma Rousseff, $i = 2$ ao José Serra,

α_i : intercepto para o candidato i .

β_i : coeficiente do termo linear associado ao candidato i ,

γ_i : coeficiente do termo quadrático associado ao candidato i ,

e_{ij} tem média zero e variância $\sigma_{ij}^2 = [p_{ij}(1 - p_{ij})]/N_j$,

N_j é o tamanho amostral da j -ésima pesquisa e

n é o número de pesquisas.

- **Incorpora**: espaçamento desigual, heterocedasticidade;

- **Não incorpora**: soma das porcentagens = 100%;

Modelo log-linear

Modelo log-linear:

$$\log(\theta_{ij}/\theta_{oj}) = \alpha_i + \beta_i t_j + \gamma_i t_j^2$$

em que

θ_{ij} : proporção de eleitores favoráveis ao i -ésimo candidato na j -ésima pesquisa, $i = D, S$ e $j = 1, 2, \dots, n$,

θ_{oj} : proporção de eleitores favoráveis a outro candidato ou indecisos na j -ésima pesquisa, $j = 1, 2, \dots, n$,

α_i : intercepto para o candidato i ,

β_i : coeficiente do termo linear associado ao candidato i ,

γ_i : coeficiente do termo quadrático associado ao candidato i ,

- **Incorpora**: espaçamento desigual, heterocedasticidade, soma das porcentagens = 100%;

Resultados

Previsões (em porcentagem) para eleição de 03 de outubro de 2010 com base em pesquisas realizadas até 6, 3, 2, 1 mês antes.

Regressão por mínimos quadrados ponderados

Previsão realizada há	Estimativa para 03/10		Intervalo de Confiança	
	Dilma	Serra	Dilma	Serra
6 meses	39.0%	22.7%	[23.2; 54.8]	[7.5; 37.8]
3 meses	29.4%	25.9%	[13.8; 44.9]	[-1.7; 53.6]
2 meses	29.8%	28.1%	[17.5; 42.0]	[5.7; 50.5]
1 mês	35.8%	27.1%	[22.3; 49.3]	[9.5; 44.7]

Modelo log linear

Previsão realizada há	Estimativa para 03/10		Intervalo de Confiança	
	Dilma	Serra	Dilma	Serra
6 meses	37.2%	26.3%	[34.7; 39.7]	[24.6; 27.8]
3 meses	32.7%	22.5%	[30.8; 34.6]	[21.3; 23.8]
2 meses	39.9%	28.9%	[38.9; 40.8]	[28.1; 29.6]
1 mês	45.5%	28.6%	[44.8; 46.1]	[28.1; 29.2]

Conclusão

- Modelo log-linear permite obter resultados mais precisos (resultado das eleições: Dilma=46.9%, Serra=32.61%).

Referências

Kutner, M.H., Neter, J., Nachtsheim, C.J. and Li, W. (2004). *Applied Linear Statistical Models, 5th ed.* New York: McGraw-Hill/Irwin.
Paulino, C.D. e Singer, J.M. (2006). *Análise de dados categorizados.* São Paulo: Blücher