

Análise de dados: eleições para o governo do estado de SP 2022 (segundo turno)

Prof. Caio Azevedo

Exemplo 5

- Vamos considerar somente os votos válidos relativos ao segundo turno das eleições para o governo do estado de SP 2022.
- Dois candidatos: Fernando Haddad e Tarcísio de Freitas.
- Base de dados: [link](#) extraídos do site [PollingData](#).
- Apresenta resultados de pesquisas de alguns institutos (votos totais e votos válidos) ocorridas no período de 14/03/2022 a 29/10/2022.
- Selicionou-se ao acaso (uniformemente) o candidato: Tarcísio e o Instituto IPEC (dentro os 3 para os quais havia 8 resultados).
- Remete ao primeiro exemplo visto ([aqui](#)).

Institutos x forma de coleta de dados

Instituto	forma de coleta de dados		
	Indefinido	Presencial	Telefônica
Atlas Intelligence	5	0	0
Badra	0	1	0
Brasmarket Inteligência	0	0	2
Datafolha	0	8	0
Equilíbrio Brasil	0	0	1
Futura	0	0	5
Ideia Big Data	0	0	2
Ipec	0	8	0
Ipesp	0	0	2
Paraná Pesquisa	0	5	0
Quaest	0	6	0
Real Time Big Data	0	6	2
Verita	0	2	0

Exemplo 5

- Vamos comparar a performance do estimador de máxima verossimilhança (emv) com o estimador pontual Bayesiano, sob a priori beta.
- Escolhemos três conjuntos de hiperparâmetros

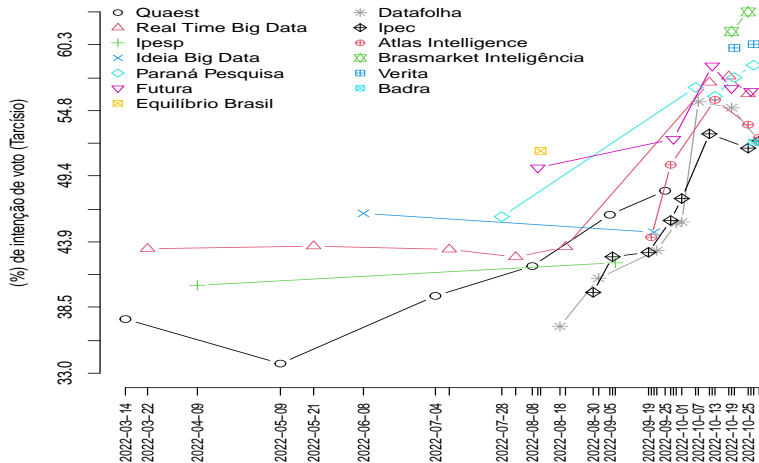
Priori	a	b
1 (otimalidade*)	$\frac{\sqrt{n}}{2}$	$\frac{\sqrt{n}}{2}$
2	$a_t = \left(\frac{1 - \tilde{\theta}_{MV}(t)}{V(\tilde{\theta}_{MV}(t))} - \frac{1}{\tilde{\theta}_{MV}(t)} \right) \tilde{\theta}_{MV}(t)^2$	
3 ($U(0,1)$)	1	1

* falaremos mais a respeito, adiante no curso; (2) significa que utilizamos as emv do resultado da pesquisa anterior

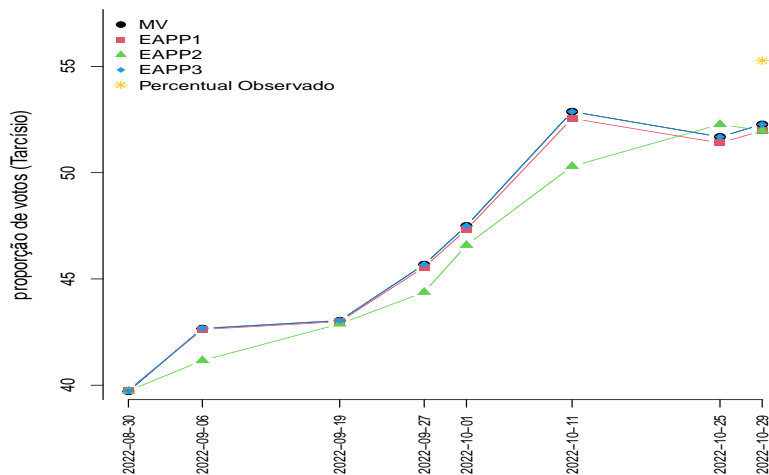
Exemplo 5

- $\tilde{\theta}_{MV}(t) = \bar{x}$: é a estimativa de mv na pesquisa feita no instante, $V(\tilde{\theta}_{MV}(t)) = \frac{\bar{x}(1-\bar{x})}{n}$: é a estimativa da variância do emv na pesquisa feita no instante, t , $t = 1, 2, 3, \dots, 7$.
- Para o primeiro instantes consideramos $a = b = 1$
- Para a priori 2, usamos a relação entre a média e a variância de uma beta com seus parâmetros (a,b) (exercício).
- Nesse caso, os hiperparâmetros variam com o tempo.

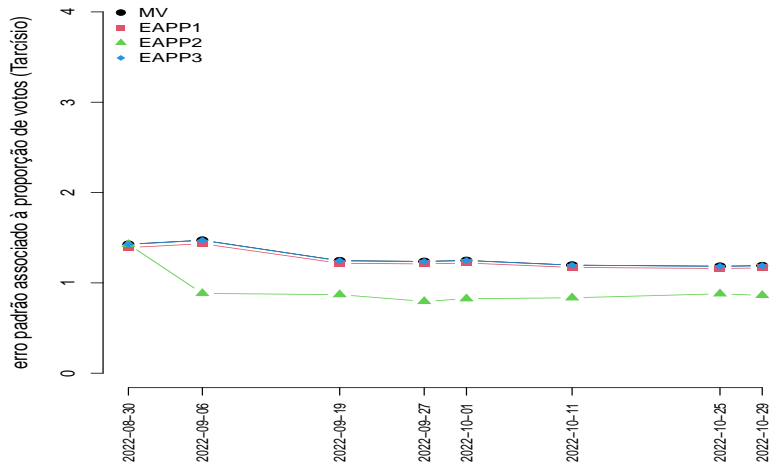
Sumários dos resultados das pesquisas



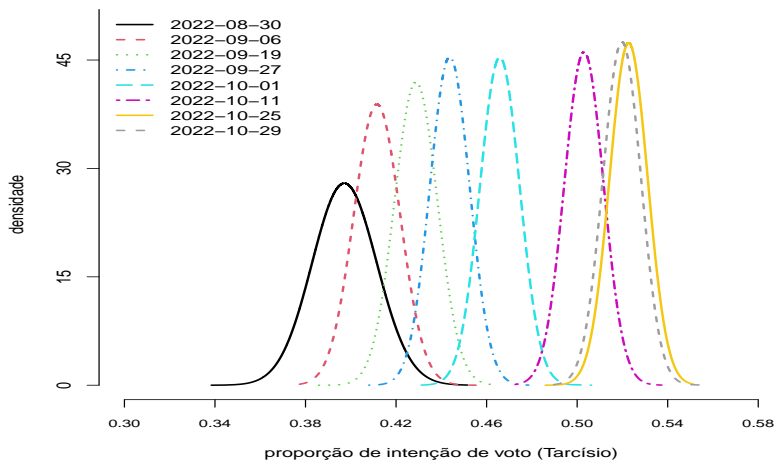
Resultados pontuais dos quatro estimadores



Resultados pontuais dos quatro estimadores



Posteriores para a priori 2



Comentários

- Observou-se uma equivalência em relação às previsões entre os quatro estimadores.
- Contudo, em termos de precisão (desvio-padrão a posteriori) o EAPP2 apresentou melhor desempenho.
- Uma abordagem bayesiana de análise mais apropriada pode ser encontrada [aqui](#).
- Uma modelagem (usando inferência frequentista ou bayesiana) mais apropriada pode ser a utilização de [modelos para dados longitudinais](#).
- Neste caso, devido ao pequeno número de pesquisas por instituto (de 1 a 8) não é possível utilizar [modelos para séries temporais](#).