



Outro indício de fraude nas eleições

[Artigos - Investimentos](#)

Escrito por Melao

Sábado, 01 Novembro 2014 12:28

Ontem recebi do amigo Victor Gabriel uma mensagem sobre o uso da Lei de Benford para identificar fraudes em eleições.

<http://gigamatematica.blogspot.com.br/2011/07/lei-de-benford.html?m=1>

*Ola melão. Li seu artigo sobre as eleições. Seria realmente válido aplicar esta lei no caso do Brasil? Ja tinha ouvido falar dessa lei? (baixe o texto postado no facebook pelo autor do estudo)*

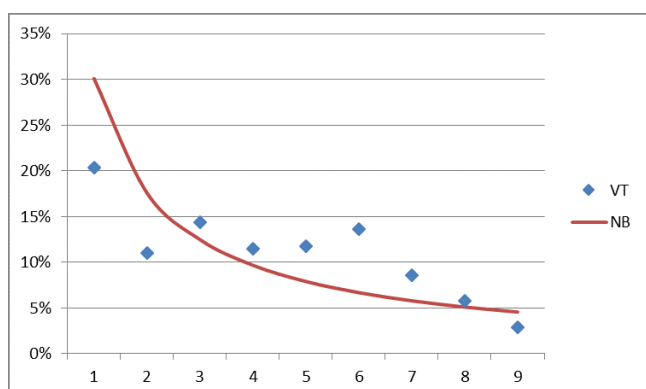
O amigo Albert Frank havia me falado sobre a Benford Law por volta de 2000, e citei esta lei em alguns artigos na Sigma Society e uma vez em minhas respostas no Ask.com. Ela é aplicada em muitas situações, mas a descrição do que é a Lei de Benford no artigo do link acima não está correta. No artigo diz: *“Dado uma amostra de números aleatórios de uma fonte de dados qualquer, qual a probabilidade do primeiro dígito ser 1? E de ser 5? E 9?”*

Não há informações suficientes no enunciado para responder, mas como as distribuições aleatórias desse tipo geralmente são uniformes, a resposta mais adequada seria 11,11% para cada. Esse será o resultado se gerar pseudoaleatórios entre 0 e 1 no Excel, multiplicar por  $10^{12}$  e depois verificar qual o primeiro algarismo.

A Lei de Benford só se aplica quando os números estão numa escala intervalar, seguem uma ordem crescente iniciada em “1” e não há lacunas. O exemplo mais comum é o número de páginas dos livros de uma biblioteca. Se todas as páginas estiverem numeradas e não faltar páginas em nenhum livro, ao escolher qualquer livro numa biblioteca e abrir numa página qualquer, há 30% de probabilidade (em vez de 11,11%) de que a o primeiro algarismo da página escolhida seja 1. E os demais algarismos têm probabilidade cada vez menor. Ao deparar com essa inexatidão logo no início, deixei de ler o artigo.

Citei na conversa o fato de que, há algum tempo, havia visto a distribuição de senhas de e-mails e PINs, e a distribuição dos primeiros algarismos não era uniforme nem era representada pela lei de Benford, sendo um exemplo de dados pseudoaleatórios que não se distribuem conforme a lei de Benford.

Como a conversa prosseguiu por algum tempo, retomei o artigo e acabei lendo mais alguns trechos. No item 4 está correta a aplicação, então comecei a me interessar. Ele me enviou uma planilha Excel com os dados brutos e analisei brevemente o caso da Bahia, em que o efeito não me pareceu tão acentuado. A Bahia é o caso mais interessante para se investigar, porque tem boa quantidade de municípios e boa quantidade de eleitores por município, de modo que a incerteza acaba não sendo tão grande.



O gráfico ao lado mostra a situação da Bahia. A curva vermelha indica a probabilidade teórica de que o primeiro algarismo seja 1, 2, 3, 4, 5, 6, 7, 8 ou 9. Os pontos azuis são os valores experimentais de frequência relativa com que foram observados os primeiros algarismos no número de votos em cada município. Há uma quantidade substancialmente menor de municípios em que o primeiro algarismo é 1 ou 2, em comparação à quantidade esperada, e há uma quantidade substancialmente maior em que o primeiro algarismo é 5 ou 6.

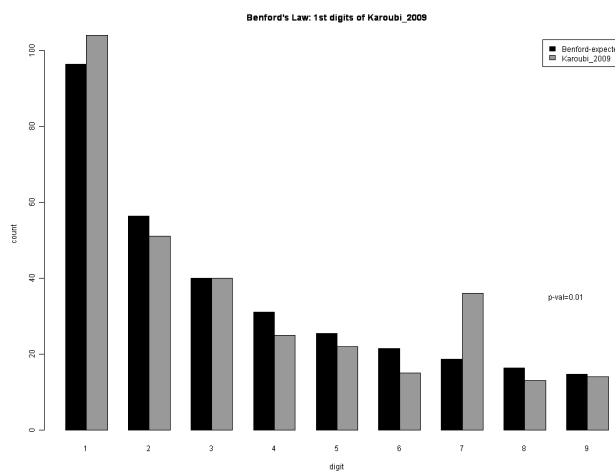
O que isso poderia indicar? Em minha opinião, quase nada. Para começar, porque a amostra com apenas 417 municípios na Bahia é muito pequena. Espera-se que destes 417, haja  $126 \pm 38$  em que o primeiro algarismo do número de votos seja 1. Para o primeiro algarismo 9, espera-se  $19 \pm 10$  municípios. A magnitude dos erros percentuais numa amostra com 126 elementos é muito grande e com 19 é ainda maior.

O total de municípios em que a votação teve 1 como primeiro algarismo foi 85, portanto apenas 32 a menos que o esperado.

Um indício nítido de fraude seria uma distribuição uniforme nos primeiros algarismos. Não é o caso. Claro que outros indícios mais sutis também poderiam se sinalizar fraude, mas não vejo evidência suficiente disso neste caso. E nas demais UF o indício me parece ainda menor. No Espírito Santo o indício poderia ser grave, não fosse pelo fato de o número de municípios ser muito pequeno, apenas 78. A expectativa era de que houvesse 30% de incidência do número 1 no primeiro algarismo entre as votações nos 78 municípios, mas foi observado apenas 7,7%. Numa amostra com 78 municípios em que o total esperado com primeiro algarismo 1 seria 23, não é tão estranho que o observado tenha sido 6. Em Roraima também houve um efeito semelhante, sendo que o número de municípios é ainda menor, portanto a tolerância ao erro é proporcionalmente muito maior.

No caso das eleições no Irã, o gráfico ao lado mostra que o indício de fraude foi muito mais nítido, porque as amostras eram muito maiores, com incertezas muito menores. Além disso, observa-se uma concentração nítida no número 7 como primeiro algarismo e uma redução em quase todos os outros primeiros algarismos. Um caso como este me parece bem mais alarmante e requer uma investigação rigorosa.

O autor do estudo que recebi, sobre as eleições no Brasil, preferiu não se identificar, mas como a pessoa que me enviou é médico e o autor do estudo faz comentários sugerindo que ele não seja estatístico, mas tem um bom domínio de Estatística, me parece razoável supor que seja médico. Minha impressão é de que o estudo dele é honesto, sem interesses partidários, mas a conclusão que ele apresenta não me parece encontrar muito respaldo nos dados.



Os 4 principais problemas que vejo são:

1. Ele deveria ter verificado a adequação à lei de Benford nas porcentagens de votos sobre o total de eleitores, em vez do número absoluto de votos. Assim atenuaria efeitos que poderiam decorrer de a distribuição nos números de eleitores nos municípios não ajustarem bem à lei de Benford, refletindo essa inadequação no número de votos.
2. Ele precisaria ter feito um estudo semelhante para todos os candidatos no primeiro e no segundo turnos e verificado em que medida se observa anomalias em cada um deles.
3. As amostras são muito pequenas tanto no número de municípios por estado quanto no número de votos em cada município, por isso seria muito importante que houvesse barras de incerteza em cada ponto azul, de modo a o Chi-quadrado (ou outro teste de aderência) levasse em consideração a distribuição de probabilidades de cada ponto azul estar numa posição diferente da verificada. Isso afeta dramaticamente a estimativa sobre a probabilidade de que o resultado observado seja significativamente diferente do esperado a um determinado nível de significância. Essa abordagem com teor mais bayesiano seria imprescindível num caso como este.
4. De maneira geral, fiquei com a impressão de que as anomalias para mais e para menos em cada algarismo se compensam, aproximadamente, e se o estudo fosse feito nos municípios do Brasil todo, em vez de ter sido nos municípios de cada estado, provavelmente grande parte das anomalias desapareceria, já que a amostra do número de municípios ficaria muito maior. A população em cada município é quase irrelevante. O fator determinante nas incertezas é o número de municípios por estado, por isso se o estudo fosse com o Brasil inteiro, embora se perdessem informações desejáveis sobre as localidades em que predominam as anomalias, a confiabilidade em que as anomalias não são meras flutuações estatísticas seria bem maior.

Como o número de votantes por município não é tão relevante, uma maneira de melhorar a sensibilidade do teste seria considerando as urnas, que são muito mais numerosas que os municípios, possibilitando reduzir a incerteza nos valores observados e detectar diferenças mais sutis. Este seria um caminho mais promissor para fazer um levantamento por localidade, porque mesmo nas UF menos populosas, há uma quantidade suficiente de urnas para que se possa ter uma incerteza razoavelmente pequena.

Portanto me parece que este estudo não fornece evidência muito forte de fraude na contagem dos votos, embora haja alguns indícios que, a meu ver, merecem atenção. Creio apenas que para garantir a imparcialidade, um estudo análogo deveria ser feito com todos os outros candidatos dos dois turnos, bem como dos candidatos a outros cargos além da presidência. O fato de uma candidata ter recebido mais votos que os demais não significa que ela deva ser o único alvo das investigações.

[Resultados do Saturno V nos últimos 4 anos](#) (339,32% em 4 anos)

Para obter mais informações técnicas sobre o Saturno V, visite nossa seção de "[artigos](#)"

Para saber mais sobre o autor, visite a seção "[Sobre o autor](#)", "[Depoimentos](#)" e "[Vídeos](#)"

Para comprar uma licença de uso do Saturno V, visite nossas seções "[Para investidores](#)" e "[Como adquirir](#)"

Para enviar elogios, dúvidas, sugestões ou críticas, preencha o [Formulário de cadastro](#)

[< Anterior](#)

[Seguinte >](#)