

# Evidências de inconformidades formais nos dados da apuração da eleição eletrônica brasileira de 2014

Sumário Executivo - Work in progress

Hugo Cesar Hoeschl *Post Doc*, Tania Cristina D'Agostini Bueno *Dra*, Gilson da Silva Paula *Eng*, Claudio Tonelli, *Adm*.

**Abstract:** O presente relatório técnico pretende demonstrar uma nova etapa de avaliação dos dados eleitorais brasileiros de 2014 considerando a aplicação da Lei de Newcomb Benford sobre os dados agrupados por Zonas Eleitorais, bem como outros aspectos.

## Introdução.

Os dados da eleição eletrônica brasileira de 2014 despertam muitas discussões e polêmicas, materializando um cenário de aplicabilidade interdisciplinar de ciência forense conjugada com novos métodos de análise de agrupamentos numéricos.

### 1. Validade da Lei de Newcomb Benford - LNB

A Lei de Newcomb Benford é a metodologia inovadora de avaliação de conformidade que detectou a fraude no famoso caso "Enron" (Bolsa de NY). Derivada de métodos observacionais, ela apresenta performances elevadas em descobertas de fraudes envolvendo "maquiagem" de números. Existe um conjunto bibliográfico, no final do texto, para maior compreensão da LNB, mas, em resumo, ela fixa probabilidades de frequências numéricas, em especial de algarismos líderes (primeiro dígito) de acordo com a seguinte distribuição:

$d$	$P(d)$	Relative size of $P(d)$
1	30.1%	
2	17.6%	
3	12.5%	
4	9.7%	
5	7.9%	
6	6.7%	
7	5.8%	
8	5.1%	
9	4.6%	

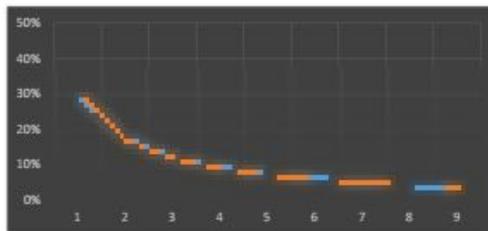
[Wikipedia Benford's Law]

A análise de cenários numéricos concretos, à luz das probabilidades fixadas pela LNB, permite identificar indícios concretos de "ofuscamentos induzidos" em conjuntos de dados, considerando que números gerados artificialmente não seguem a proporcionalidade natural prescrita pela LNB.

Vejamos a aplicação da LNB sobre os dados do Censo populacional do Brasil (2000 e 2010):

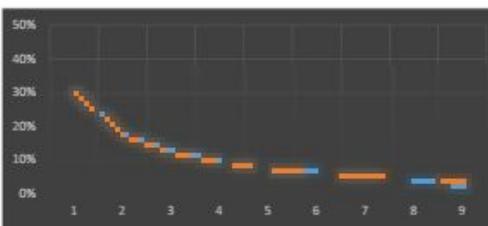
**CENSO 2000**

DIGITO	REINCIDENCIA	QTDE	DIGITO	%var/media	%var/reinc	%NBL
1	1.605	55.832.952	1	262%	29%	30,1%
2	1.008	33.827.629	2	165%	18%	17,6%
3	697	18.215.677	3	114%	13%	12,5%
4	562	13.191.923	4	92%	10%	9,7%
5	434	15.771.470	5	71%	8%	7,9%
6	407	9.282.910	6	67%	7%	6,7%
7	317	10.072.103	7	52%	6%	5,8%
8	266	6.801.463	8	43%	5%	5,1%
9	211	6.803.043	9	34%	4%	4,6%
<b>Total Geral</b>	<b>5.507</b>	<b>169.799.170</b>	<b>Média--&gt;</b>	<b>612</b>	<b>100%</b>	<b>100%</b>



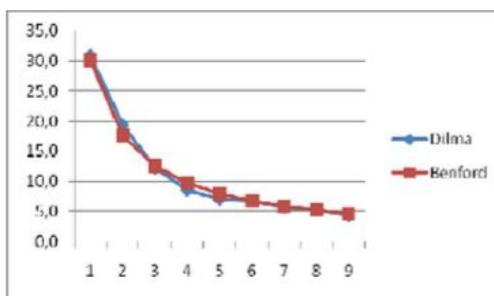
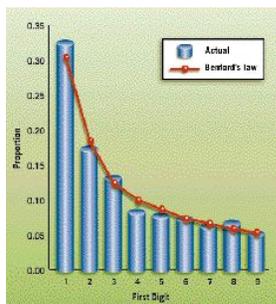
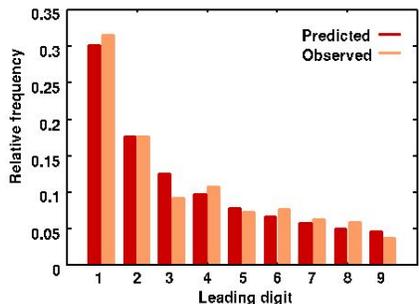
**CENSO 2010**

DIGITO	REINCIDENCIA	POPULAÇÃO	DIGITO	%var/media	%var/reinc	%NBL
1	1675	63.774.606	1	271%	30%	30,1%
2	1026	38.076.402	2	166%	18%	17,6%
3	737	21.296.183	3	119%	13%	12,5%
4	563	14.729.449	4	91%	10%	9,7%
5	438	11.493.017	5	71%	8%	7,9%
6	393	17.707.968	6	64%	7%	6,7%
7	315	9.725.975	7	51%	6%	5,8%
8	265	8.086.507	8	43%	5%	5,1%
9	153	5.842.587	9	25%	3%	4,6%
<b>Total Geral</b>	<b>5565</b>	<b>190.732.694</b>	<b>Média--&gt;</b>	<b>618</b>	<b>100%</b>	<b>100%</b>



[Dados organizados por Cláudio Tonelli]

Agora vejamos 3 situações de grupos de números analisados perante as regras de conformidade da LNB: a) Comparativo Walthoe (LNB X *Constantes da física, tabela de expansão de gases e estatísticas esportivas*); b) Censo populacional dos EUA; c) Eleição brasileira de 2010:

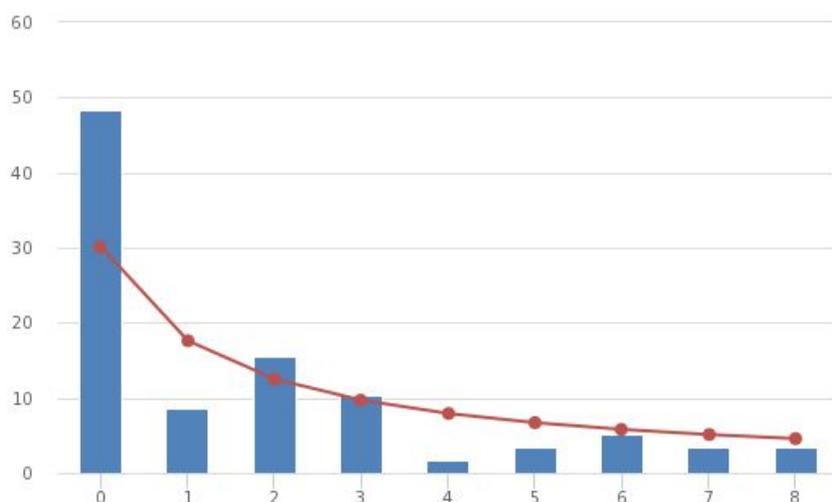


Em todos esses casos as curvas de agrupamento dos números reais seguem com exatidão a "curva ideal" desenhada pela LNB, com pequenos desvios comportados pela margem de tolerância.

## 2. Metodologia tradicional de aplicação da LNB

A metodologia tradicional de aplicabilidade da LNB prevê o comparativo da curva descendente da LNB com as curvas formadas nos casos concretos. É muito fácil fazer o comparativo. Se a curva do caso concreto não estiver seguindo a mesma curva decrescente da LNB, algo está errado.

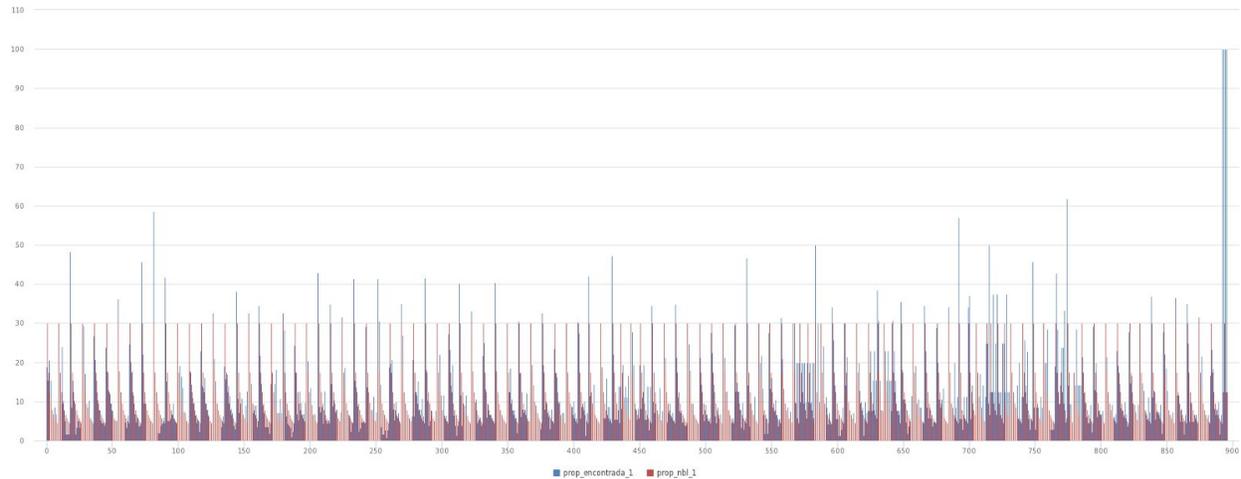
Aqui nesta imagem existe um exemplo concreto, no qual a linha **vermelha** é a "curva ideal" com os dados fixados pela LNB, enquanto as barras **azuis** são os quantitativos dos votos em branco de um estado brasileiro (ES) em 2014 (dados presidenciais):



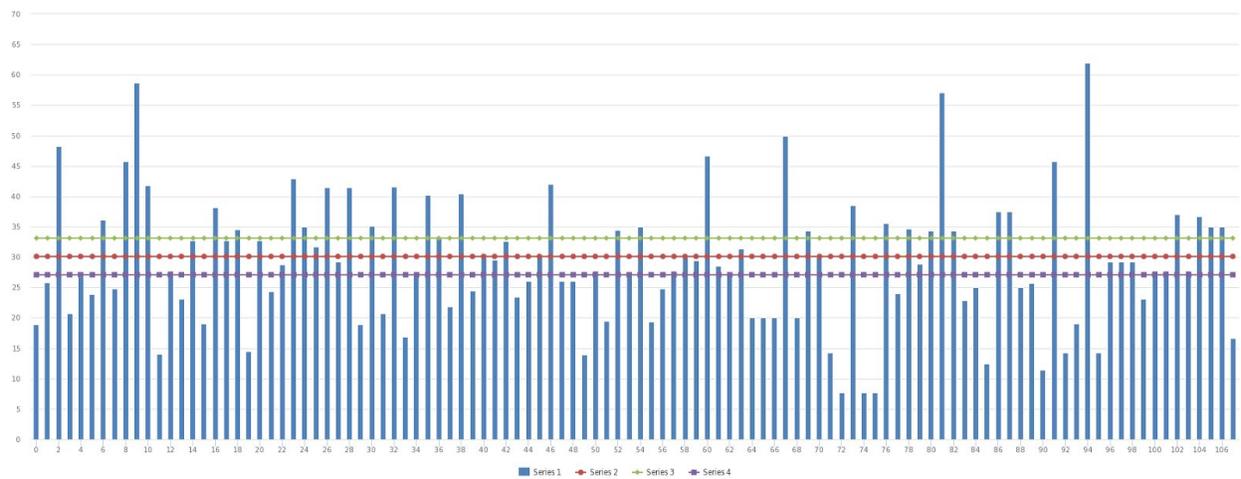
Note que existe uma flagrante inconformidade nesses dados, do ponto de vista da LNB.

## 3. Nova metodologia para aplicação da LNB

Para que os dados da eleição eletrônica brasileira de 2014 possam ser analisados de uma forma panorâmica, criamos uma nova metodologia de aplicabilidade da LNB, de acordo com os seguintes critérios: **a)** Somando todos os votos com agrupamentos pelas Zonas Eleitorais; **b)** Dividindo em grupos por estado da federação; **c)** Dividindo cada estado em 4 grupos (Branco, Nulos, Dilma, Aécio); **d)** Coletando o número líder (primeiro dígito) dos grupos; **e)** Comparando os dados concretos com os dados da LNB. Os dados coletados formam uma série que pode ser vista no seguinte gráfico, onde as barras em **azul** são os dados concretos, e as barras em **vermelho** são as probabilidades da LNB:



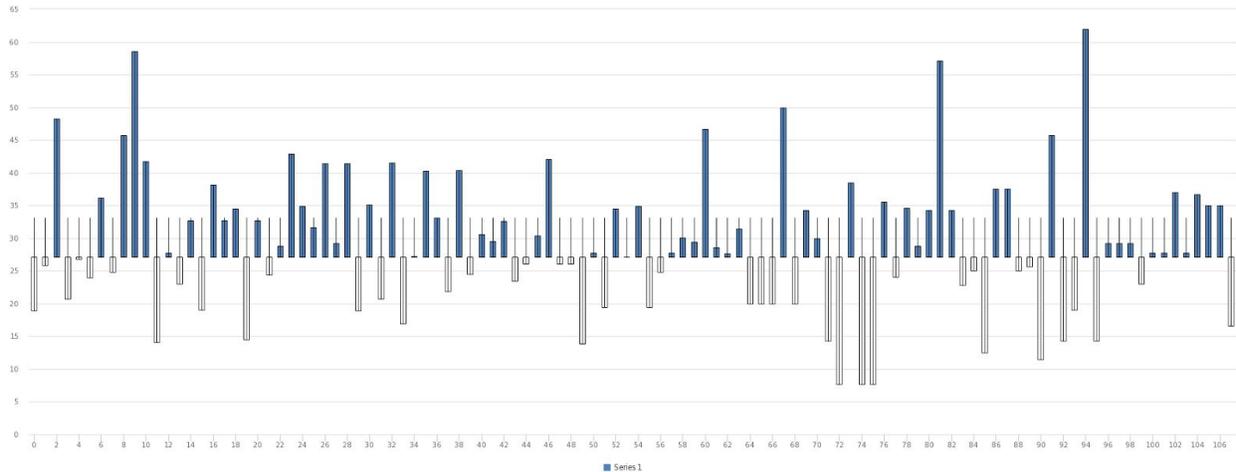
Todas as vezes que as barras não coincidem (a *mais* ou a *menos*) existe uma inconsistência. Nota-se que são muitos casos de não coincidência. Diante do grande volume de dados, optou-se por desmembrar o estudo em 2 etapas: i) Construir outro agrupamento de análise; ii) Realizar comparativos tradicionais da LNB em situações pontuais, caso a caso (veremos no item 4). Na nova metodologia, foram adicionados os seguintes passos: **f**) Novo agrupamento com os dados dos grupos derivados do número "1"; **g**) Comparativo, caso a caso, dos dados encontrados com a probabilidade da LNB. O procedimento resultou no seguinte gráfico:



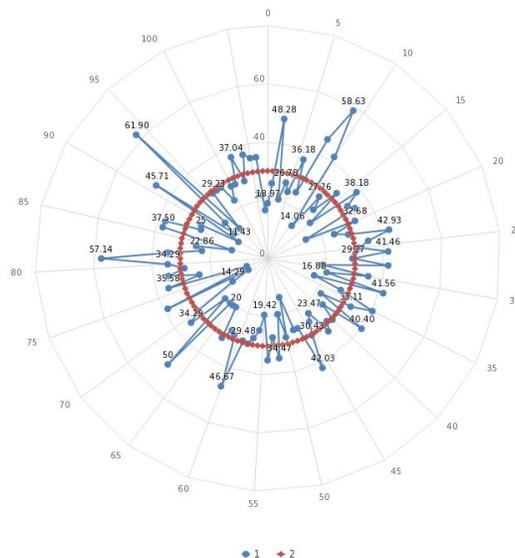
O gráfico contém as seguintes características: As barras em **azul** são os dados concretos coletados; A linha com os pontos em **vermelho** representa a taxa de probabilidade da LNB para os casos em que o algarismo líder (primeiro dígito) é o número "1" (30,103%); A linha com os pontos em **verde** representa uma taxa adicional de 10% na LNB (33,113%); A linha com os pontos em **roxo** representa um decréscimo de 10% (27,090%). Dessa forma foi criada uma *zona de conformidade*, com 10% a *mais* ou a *menos*, a fim de verificar a proporcionalidade de

inserção dos números coletados dentro da zona de conformidade. O resultado, visível no gráfico, é que somente 26.851% das séries nas quais o algarismo líder é o número "1" estão dentro da zona de conformidade.

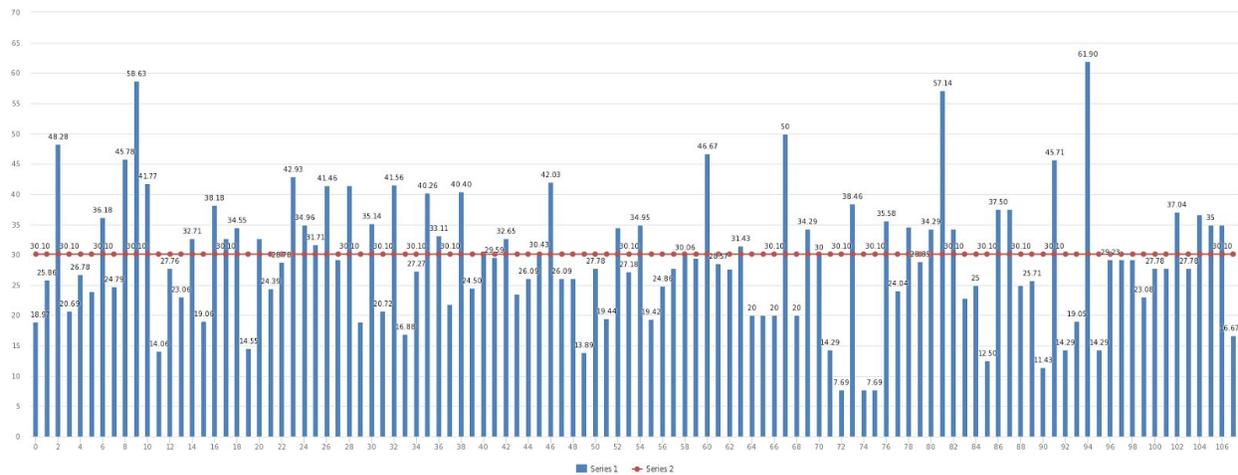
Para melhor visualização, o mesmos dados podem ser vistos de outras formas. Aqui com os valores "altos" (valores que superam a média) subindo, e com os valores "baixos" (valores que ficaram abaixo da média) descendo:



Outra forma é a visualização circular, na qual o círculo em vermelho representa a linha da LNB e os pontos em azul representam os dados concretos coletados:



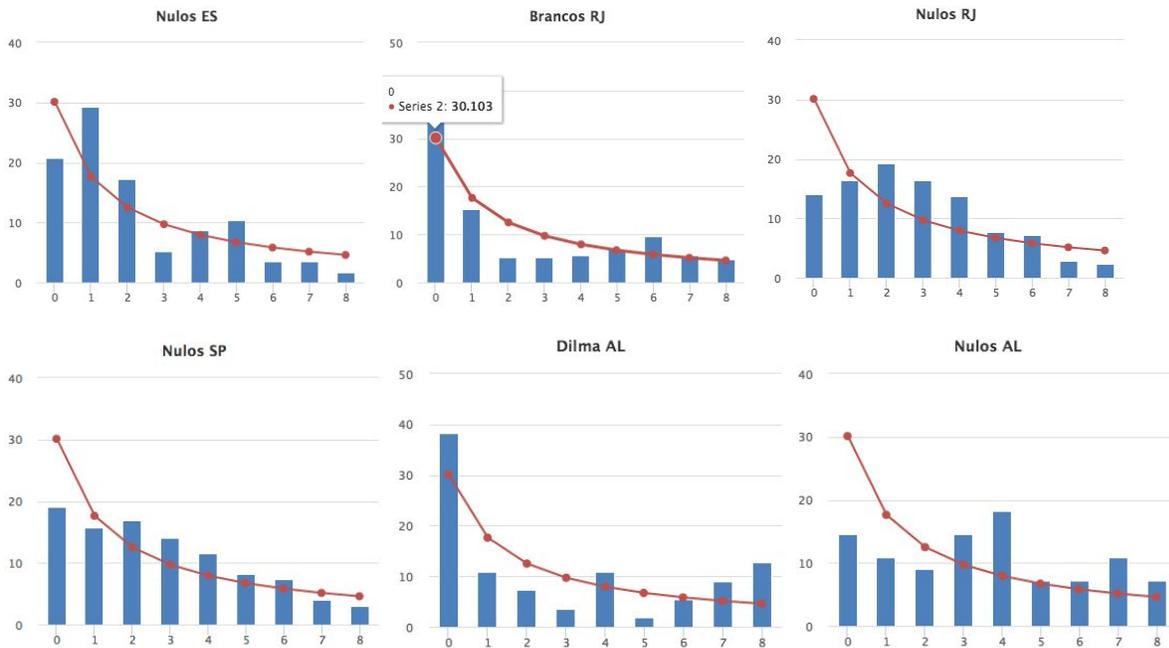
Ou com os dados dispostos em linha, com os valores apresentados caso a caso, sem os limites de tolerância de 10%:

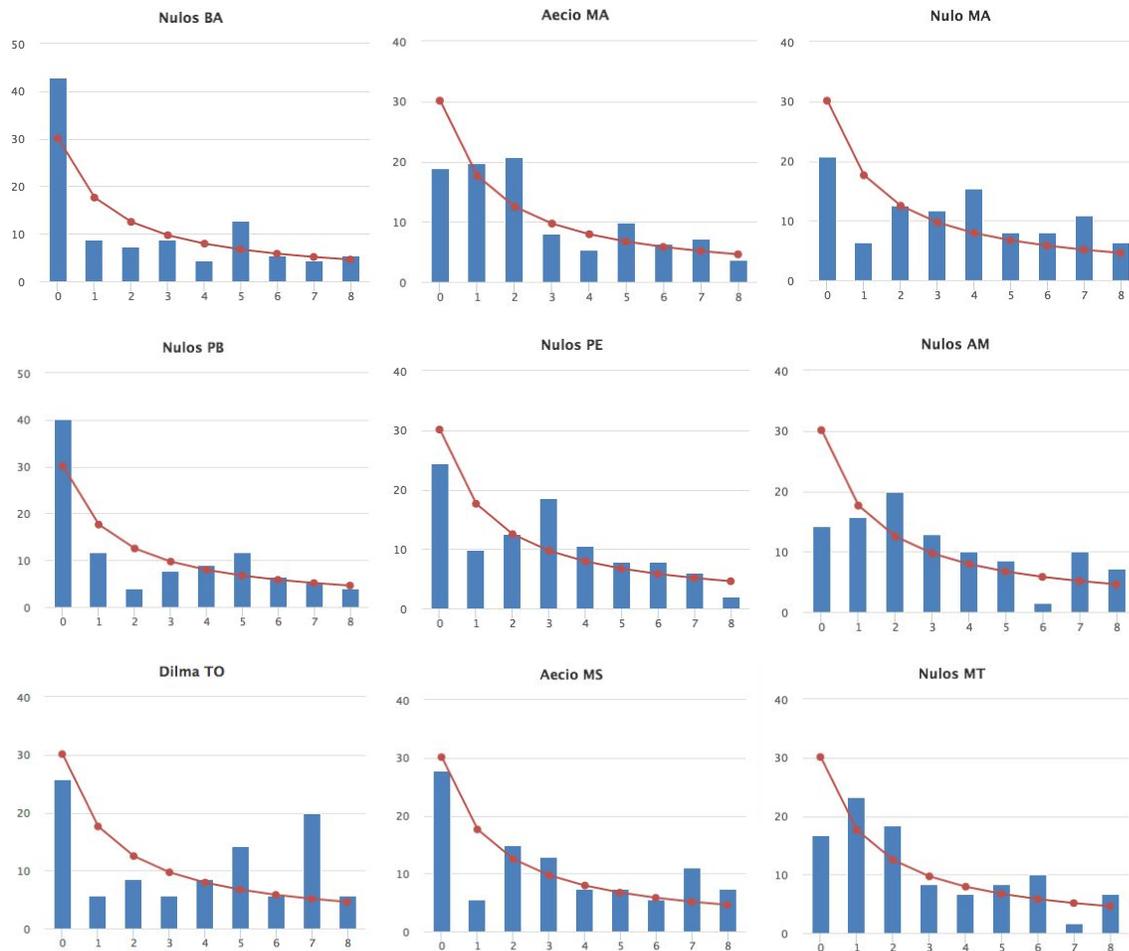


Em qualquer uma das formas de visualização o nível de inconsistência é muito elevado.

#### 4. Metodologia Tradicional da LNB

Na avaliação de conformidade da LNB, modelo tradicional, utilizando o mesmo universo referido no item "3" (ítems "a" a "e"), e comparando gráfico a gráfico com a linha ideal, existem vários gráficos que apresentam inconsistência nos dados do TSE. Destacamos aqui alguns casos, nos quais a linha em **vermelho** é a LNB, e os gráficos em **azul** são os dados reais do TSE:





O que se percebe é uma tendência forte e reafirmada de discrepância entre o *real* e o *ideal* nos dados da apuração. Essa inconsistência assume contornos mais demarcados quando da análise de votos brancos e nulos, e aparece com intensidade inclusive no comparativo entre esses dois grupos (onde era de se esperar tendências semelhantes em *brancos* e *nulos*).

### 5. Zona cinzenta de inconformidades

Além da análise pura da LNB, existem outros aspectos que disparam alertas de inconformidade na massa de dados do TSE. São eles: a) Os dados das urnas com número de eleitores acima do permitido por Lei; b) O comparativo da migração dos votos em branco; c) Emparceiramentos atípicos.

Sobre as urnas com votos acima do limite legal, vejamos a seguinte tabela:

SEÇÕES ACIMA DE 400 ELEITORES					
UF	QTD. SEÇÕES	ELEITORES APTOS	COMPARECERAM	FALTARAM	%PRESENÇA
AC	221	95,904	81,787	14,117	85.3%
AL	507	238,758	198,005	40,753	82.9%
AM	958	424,486	350,825	73,661	82.6%
AP	115	50,608	41,393	9,215	81.8%
BA	4,374	1,907,965	1,476,512	431,453	77.4%
CE	1,860	783,307	642,218	141,089	82.0%
DF	25	10,397	8,233	2,164	79.2%
ES	466	209,725	164,106	45,619	78.2%
GO	224	97,346	74,843	22,503	76.9%
MA	180	75,970	58,182	17,788	76.6%
MG	3,586	1,666,401	1,346,942	319,459	80.8%
MS	578	249,583	201,601	47,982	80.8%
MT	165	70,071	51,346	18,725	73.3%
PA	1,754	809,572	644,249	165,323	79.6%
PB	84	35,584	29,271	6,313	82.3%
PE	518	233,923	180,692	53,231	77.2%
PI	727	320,681	273,713	46,968	85.4%
PR	43	17,894	14,040	3,854	78.5%
RJ	8,197	3,799,835	2,954,507	845,328	77.8%
RN	827	365,356	311,028	54,328	85.1%
RO	79	33,887	25,892	7,995	76.4%
RR	10	4,193	3,579	614	85.4%
RS	270	115,921	94,294	21,627	81.3%
SC	257	110,244	92,130	18,114	83.6%
SE	427	196,691	162,686	34,005	82.7%
SP	7,399	3,132,864	2,518,239	614,625	80.4%
ZZ	64	33,532	23,348	10,184	69.6%
<b>Total Geral</b>	<b>33,915</b>	<b>15,090,698</b>	<b>12,023,661</b>	<b>3,067,037</b>	<b>79.7%</b>
Total de seções	428,894				
% > 400 eleitores	7.91%				

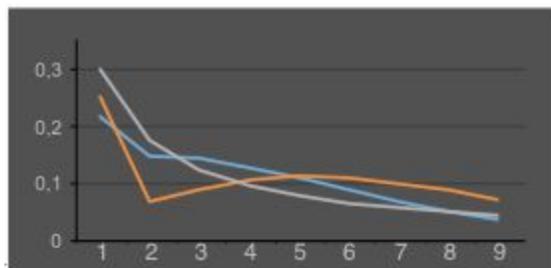
[Planilha gerada por Cláudio Tonelli]

Os diplomas legais dizem que "o quantitativo mínimo e máximo de eleitores admitidos em cada seção eleitoral é estabelecido pela Resolução -TSE n.º 14.250/88, a qual regulamentou a Lei n.º 6.966/82, e foi albergada pela Lei n.º 9.504/97. De acordo com essa legislação, é de 400 (quatrocentos) o número máximo de eleitores por seção eleitoral, sendo o número mínimo o de 50 (cinquenta) eleitores", conforme precedente judicial. A norma é muito clara, conforme o artigo 66 do Código eleitoral: "Art. 66. O juiz distribuirá os eleitores por seções, **não podendo nenhuma delas ter mais de 400 nem menos de 50 eleitores.**"

Ocorre que existe uma série de exceções, e, em função delas, o limite legal foi flexibilizado. Os dados da planilha são muito claros, existem mais de 400.000 urnas em todo o Brasil que estão acima do limite legal, totalizando um universo de mais de 15 milhões de eleitores. Somente no estado do RJ existem mais de 8.000 seções eleitorais (urnas) acima do limite legal, totalizando quase 4 milhões de votos. A existência de exceções não poderia motivar um quantitativo de tamanha expressividade.

Outro tema que se encontra em "zona cinzenta" é a redução de votos em branco entre o primeiro e o segundo turnos (mais de 50%), conforme o estudo realizado por Gilson da Silva Paula (referência em anexo). O número de votos em branco no primeiro turno foi de 4.420.489. No segundo turno foi de 1.921.819. Em outros anos isso já aconteceu, mas não com tanta intensidade. Tal percentual de migração encontra pouco respaldo em termos de probabilidades estatísticas, principalmente considerando que a tendência histórica vem sendo de crescimento, e não de redução, do voto em branco. A análise probabilística dos votos em branco utilizando a LNB mostra uma curva com elementos fora dos padrões, em especial quando utilizado o percentual de variação de votos como critério de análise.

BRANCO	1	86.606	482.189	1	98%	22%	25%	30,1%
	2	59.195	133.658	2	67%	15%	7%	17,6%
	3	57.817	173.977	3	66%	15%	9%	12,5%
	4	51.205	204.820	4	58%	13%	11%	9,7%
	5	43.285	216.425	5	49%	11%	11%	7,9%
	6	34.893	209.358	6	40%	9%	11%	6,7%
	7	27.381	191.667	7	31%	7%	10%	5,8%
	8	21.124	168.992	8	24%	5%	9%	5,1%
	9	15.637	140.733	9	18%	4%	7%	4,6%
<b>BRANCO Total</b>		<b>397.143</b>	<b>1.921.819</b>		<b>Média--&gt;</b>	<b>88.003</b>		



Outro ponto são os emparelhamentos numéricos atípicos, como os casos de seções eleitorais em sequência exata (235, 236 e 237) (17, 18 e 19) que registram exatamente o mesmo número de votos para um dos candidatos (96 no primeiro caso, 94 no segundo). Esse tipo de "match" aumenta a espessura da chamada "zona cinzenta", mostrando que outros tipos de varreduras, que busquem por este tipo de emparelhamento, possam vir a ser feitos, em larga escala.

UF	DESC. CARGO	NM PARTIDO	NM MUNICIPIO	NB LOCAL	NR_SC	NR_CN	NM NOTAVEL	QTD_VOTOS	QTD_COMPARE	QTD_APTOS	QTD_FALTOSOS	NEMLCMB
SE	PRESIDENTE	PSDB	ARACAJU	1325	235	36	AÉCIO NEVES	92	255	302	47	5
SE	PRESIDENTE	PSDB	ARACAJU	1325	237	36	AÉCIO NEVES	96	256	302	46	5
SE	PRESIDENTE	PSDB	ARACAJU	1325	236	36	AÉCIO NEVES	96	256	306	50	5
SE	PRESIDENTE	PSDB	ARACAJU	1325	235	36	AÉCIO NEVES	96	254	307	53	5
SE	PRESIDENTE	PSDB	ARACAJU	1309	229	36	AÉCIO NEVES	95	281	306	45	5
SE	PRESIDENTE	PSDB	ARACAJU	1309	228	36	AÉCIO NEVES	94	279	303	28	5
SE	PRESIDENTE	PSDB	ARACAJU	1309	225	36	AÉCIO NEVES	91	261	302	41	5
SE	PRESIDENTE	PSDB	ARACAJU	1309	222	36	AÉCIO NEVES	98	259	306	47	5
SE	PRESIDENTE	PSDB	ARACAJU	1228	184	36	AÉCIO NEVES	93	163	202	39	5
SE	PRESIDENTE	PSDB	ARACAJU	1201	179	36	AÉCIO NEVES	93	215	246	31	5
SE	PRESIDENTE	PSDB	ARACAJU	1031	62	36	AÉCIO NEVES	91	247	274	27	5
SE	PRESIDENTE	PSDB	ARACAJU	1031	60	36	AÉCIO NEVES	90	234	273	39	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1074	35	36	AÉCIO NEVES	92	285	332	47	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1066	29	36	AÉCIO NEVES	95	313	365	52	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1058	24	36	AÉCIO NEVES	90	265	338	73	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1058	23	36	AÉCIO NEVES	99	282	341	39	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1058	22	36	AÉCIO NEVES	95	274	342	68	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1058	19	36	AÉCIO NEVES	94	272	339	67	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1058	18	36	AÉCIO NEVES	94	285	342	57	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1058	17	36	AÉCIO NEVES	94	271	339	68	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1040	14	36	AÉCIO NEVES	93	278	339	61	5
SE	PRESIDENTE	PSDB	BARRA DOS COQUE	1023	5	36	AÉCIO NEVES	93	228	266	38	5
SE	PRESIDENTE	PSDB	UMBURUBÁ	1201	113	35	AÉCIO NEVES	90	295	370	75	5

Porém, a existência de uma "zona cinzenta" não significa a existência de fraude, mas demonstra a necessidade de que o sistema oficial elimine as dúvidas, com demonstrações públicas de comprovação de resultado, o que passa por comprovações *anteriores às zerézimas*. Um sistema eleitoral, qualquer que seja, está em "xeque" quando não conseguir demonstrar e comprovar os seus resultados.

*"Houve fraude eleitoral?"*. Para que seja possível afirmar, taxativamente, que houve fraude eleitoral, é necessário possuir uma prova material. Isso não existe até o momento, em função da volatilidade do cômputo de votos no âmbito das unidades coletoras (urnas, ou seções eleitorais). A mesma volatilidade que também põe em dúvida o resultado final. Porém, é possível afirmar, com grau de assertividade de 71,149%, que a aglomeração estatística dos dados agrupados por Zonas Eleitorais apresenta discrepância em relação aos parâmetros aceitáveis, no âmbito da LNB. Em alguns países, a perícia fundada na LNB pode ser considerada como prova material (caso Enron), o que ainda não ocorre perante o sistema jurídico brasileiro.

*"Como esclarecer essas dúvidas?"*. Não cabe às investigações técnico-científicas a tarefa de provar a existência de fraude. *A contrario sensu*, cabe ao sistema oficialista a comprovação de que não existem dúvidas sobre o processo de apuração. Uma eleição, para que seja considerada consistente no âmbito da conformidade de processos, precisa se constituir de procedimentos repetíveis, sob pena de invalidade epistemológica.

## 6. Conclusões e trabalhos futuros

I) Considerando a metodologia apresentada no ítem "3", o grau de consistência probabilística da eleição eletrônica brasileira de 2014 é de 26,851%;

II) A eleição brasileira de 2014, sob a ótica da Lei de Newcomb Benford, encontra-se reprovada na análise de conformidade, com grau de certeza de 73,149%;

III) A eleição brasileira de 2014 já entrou para a história como a mais duvidosa de todos os tempos;

IV) Os focos de inconsistência apontados tem potencial suficiente para alteração nos resultados em ambos os turnos;

*Trabalhos futuros*: Após realizado o trabalho de análise da LNB sobre os dados referentes aos municípios, agora foi feita a avaliação em nível de Zona Eleitoral. Uma próxima aplicação pode abordar as Seções Eleitorais, mas haverá a necessidade de tratamento especial (provavelmente com análise de 2o e 3o dígitos), em função do limite legal de 400 eleitores por urnas.

Referências operacionais:

- Todos os dados utilizados são aqueles disponibilizados pelo TSE;
- O trabalho em grupo utilizou ferramentas colaborativas *on line* (gratuitas), incluindo as atividades de coleta, análise e edição das informações;

Principais referências bibliográficas:

- <https://plus.maths.org/content/looking-out-number-one>
- [https://en.wikipedia.org/wiki/Benford%27s\\_law](https://en.wikipedia.org/wiki/Benford%27s_law)
- <http://fisicayciencia.blogspot.com.br/2007/03/son-igual-de-probables-todos-los-dgitos.html>
- [http://www.academia.edu/3596147/As\\_%C3%BAltimas\\_elei%C3%A7%C3%B5es\\_e\\_a\\_Lei\\_de\\_Benford\\_ou\\_Lei\\_do\\_Primeiro\\_D%C3%ADgito](http://www.academia.edu/3596147/As_%C3%BAltimas_elei%C3%A7%C3%B5es_e_a_Lei_de_Benford_ou_Lei_do_Primeiro_D%C3%ADgito)
- <https://plus.maths.org/content/does-iranian-election-stand-statistics?src=aop>
- <https://plus.maths.org/content/os/latestnews/jan-apr10/electstats/index?src=aop>
- <http://observatoriodaimprensa.com.br/jornal-de-debates/da-aplicacao-da-lei-de-newcom-b-benford-em-processos-eleitorais-eletronicos/>
- <http://pt.scribd.com/doc/256249069/Eleic-o-es-eleto-nicas-2014-no-Brasil#scribd>
- <http://pt.scribd.com/doc/256249091/Direito-e-Tecnologia-da-Informacao-no-cenario-internacional-e-o-desafio-da-independencia-tecnica-Liberdade-de-Expressao-X-Censural>
- <https://www.youtube.com/watch?v=hF7UOsZMQn8>
- <http://www-personal.umich.edu/~wmebane/>
- <https://franciscoalvesjunior.wordpress.com/2010/08/24/evidencias-empiricas-acerca-da-aplicacao-da-lei-de/>
- <http://www.cartacapital.com.br/blogs/carta-nas-eleicoes/soma-de-votos-nulos-brancos-e-abstencoes-e-a-maior-desde-1998-6625.html>