

Pesquisas Eleitorais e as Amostras por Cota

Nota Nº 2

Luiz Carlos da Rocha

(Continuação da Nota Nº 1)

3ª Parte – Parâmetros da margem com alternativas de Viés.

O objetivo da Nota Nº 2 é mostrar a dimensão das alterações que podem ocorrer nos valores dos parâmetros da margem, erro e nível de confiança, calculados pela metodologia padrão registrada no TSE quando se supõe a quebra do paradigma viés Zero. Os valores exibidos foram obtidos pela aproximação Normal e pelas desigualdades Chebychev/Markov, e revelarão panorama bastante distinto do divulgado de forma padronizado pelas empresas de pesquisa. Importa destacar que os resultados simulados se referem exclusivamente às populações de eleitores das grandes cidades do Brasil (grandes números).

d) Definições e Notações.

Seja p o real porcentual dos votos de certo candidato submetido a M prévias PSQ replicadas, o viés notado por v é a média dos desvios, sendo a média das prévias designada por *Média*. Na pesquisa eleitoral com amostras por cota a expectativa é ter $v \neq 0$ e vale a relação: $Média = p + v$. O conceito da margem de erro estatístico, notado por ε positivo, se refere ao intervalo $(-\varepsilon, \varepsilon)$ e significa a prefixação de um intervalo desejável contendo o *desvio* da prévia, isto é, se deseja $|\text{desvio}| \leq \varepsilon$. Associada a margem de erro é definido o Nível de Confiança (*NC*) que é a probabilidade da ocorrência $|\text{desvio}| \leq \varepsilon$. A variância das M prévias replicadas, notação *Var*, é a média dos quadrados das diferenças das prévias em relação à *Média* ou também igual à variância dos desvios em relação ao viés. Os desvios determinam o viés que por sua vez determinam a *Média* e a *Var*, que são parâmetros fundamentais no cálculo da margem de erro e do nível de confiança. É claro que a *Média* e a *Var* das prévias não são conhecidas e suas estimativas são obtidas a partir da amostra de tamanho n , mas com amostra por cota estas não são confiáveis no sentido da máxima verossimilhança; ao contrário do caso teórico, cuja seleção é aleatória.

O *erro e viés* serão notados nas tabelas em pontos percentuais e por razões práticas adotaram-se valores inteiros no intervalo de 0% a 5%. Os parâmetros foram calculados fazendo $Var=2500/n$ exatamente o mesmo valor fixado nas metodologias das pesquisas eleitorais registradas no TSE. Tal valor corresponde ao máximo da variância das prévias

eleitorais no caso ideal da amostra por seleção aleatória de tamanho n , mas com amostras por cota a Var não é limitada a $2500/n$. Todavia, para avaliar o efeito causado pelo viés nos parâmetros da margem em relação aos valores das pesquisas registradas (sem viés) o valor $2500/n$ foi mantido constante nos cálculos efetuados.

Por fim, importa mencionar que a influência do viés nos parâmetros da margem de erro vale para amostras por seleção estritamente aleatória, onde estimativas enviesadas podem ocorrer devido às falhas na mensuração do atributo e por erros não amostrais.

I – Cálculos supondo a hipótese de Normalidade

Supondo que a distribuição estatística das M prévias geradas pelas replicações tenha passado no teste da distribuição Normal, isto é, que a forma do histograma das replicações seja aderente à forma da distribuição Normal, têm-se na Tabela 1 os tamanhos das amostras que garantem nível de confiança de 95% em função do erro (ε) e viés $\pm v$, cuja notação $\pm v$ denota valores pontuais simétricos e não significa intervalo.

TABELA 1–Tamanho da amostra pela Normalidade ($NC=95\%$)

Erro ε (%)	Tamanho da amostra pela Normal com $NC = 95\%$ em função do erro ε e segundo a hipótese de Viés $\pm v$ (*)					
	Viés= Zero	Viés= $\pm 1\%$	Viés= $\pm 2\%$	Viés= $\pm 3\%$	Viés= $\pm 4\%$	Viés= $\pm 5\%$
1%	9604	VNE	VNE	VNE	VNE	VNE
2%	2401	6765	VNE	VNE	VNE	VNE
3%	1067	1691	6765	VNE	VNE	VNE
4%	600	752	1691	6765	VNE	VNE
5%	384	423	752	1691	6765	VNE

(*) VNE: Valor não existente para 95%

Fixado o erro, então para viés entre $\pm v$ sucessivos o tamanho da amostra estará entre os valores tabelados, se $\varepsilon = 3\%$ e $v = \pm 1,5\%$ então $1691 < n < 6765$. Para erro de 2% com $NC = 95\%$ e viés Zero, o tamanho da amostra é 2401 sendo essa uma das combinações preferidas das grandes empresas de pesquisas e que recentemente vêm usando tamanho igual a 2002 entrevistas com erro 2,2%. Observar como o tamanho é sensível ao viés, se o procedimento amostral gerasse $\pm 1\%$ de viés, então para manter a margem de erro 2% com 95% de confiança seria preciso uma amostra de 6765 eleitores. Logo, a declaração alardeada pelas empresas de pesquisa “a margem de erro das estimativas é 2 pontos porcentuais, para mais ou para menos, com 95% de certeza” é quase sempre falsa por causa da provável existência do viés, ainda que, as demais hipóteses admitidas na metodologia fossem verdadeiras, quais sejam, a aproximação Normal e $Var = 2500/n$.

A Tabela 2 mostra como os níveis de confiança variam em função das combinações de erro ε e viés $\pm v$ para tamanho de amostras calculadas com fixações: $NC = 95\%$, viés = Zero e $Var = 2500/n$, que são hipóteses constantes nas metodologias registradas no TSE.

TABELA 2 – Nível de Confiança pela Normalidade

Erro ε (%)	Tamanho da Amostra	Nível de Confiança (%) pela Normal em função do Erro (ε) e tamanho segundo o Viés $\pm v$					
		Viés = Zero	Viés = $\pm 1\%$	Viés = $\pm 2\%$	Viés = $\pm 3\%$	Viés = $\pm 4\%$	Viés = $\pm 5\%$
1%	9604	95,0	50,0	2,5	VNS	VNS	VNS
2%	2401	95,0	83,5	50,0	16,4	2,5	0,2
3%	1067	95,0	90,0	74,3	50,0	25,7	9,6
4%	600	95,0	92,2	83,5	68,8	50,0	31,2
5%	384	95,0	93,2	87,7	78,3	65,2	50,0

(*)VNS: Valor não significante.

Para $\varepsilon = 2\%$ e viés Zero, o tamanho 2401 garante 95% de confiança, mas esse mesmo tamanho garantiria só 83,5% de confiança caso o viés fosse $\pm 1\%$. Ainda com $\varepsilon = 2\%$ mas com viés $\pm 2\%$ o nível de confiança se reduziria a 50,0%. Para valor do viés entre sucessivos $\pm v$ então o nível de confiança NC estará entre valores tabelados, por exemplo, com erro $\varepsilon = 2\%$ e viés $v = \pm 1,5\%$ então $50,0\% < NC < 83,5\%$.

II- Cálculos usando as Desigualdades de Chebychev/Markov

Abandonada a hipótese da aproximação Normal então a margem de erro e o nível de confiabilidade podem ser avaliados pelas Desigualdades de Chebychev/Markov, suposta a aleatoriedade e conhecendo-se a variância. No caso das desigualdades o nível de confiança significa limite inferior da probabilidade do erro.

Na Tabela 3 têm-se os tamanhos de amostra calculados para as combinações de erro e viés com diferentes níveis de confiança, todos calculados com $Var = 2500/n$, onde se observam várias situações de impossibilidade. Notar que com viés diferente de Zero não se consegue confiabilidade aceitável para várias hipóteses de erro.

TABELA 3 – Tamanho da Amostra por Chebychev/Markov

Tamanho da Amostra por Chebychev/Markov em função do Erro (ε) e segundo o limite inferior do Nível de Confiança para Viés=0, 1 e 2%									
Erro ε %	95%		90%		85%		80%		
	Viés= Zero	Viés= $\pm 1\%$	Viés= Zero	Viés= $\pm 1\%$	Viés= Zero	Viés= $\pm 1\%$	Viés= Zero	Viés= $\pm 1\%$	Viés= $\pm 2\%$
1%	50000	VNE	25000	VNE	16667	VNE	12500	VNE	VNE
2%	12500	VNE	6250	VNE	4167	VNE	3125	VNE	VNE
3%	5556	VNE	2778	VNE	1852	7143	1389	3125	VNE
4%	3125	VNE	1563	4167	1042	1786	781	1136	VNE
5%	2000	10000	1000	1667	667	909	500	625	2500

VNE= Valor não existente.

Importa advertir que o tamanho da amostra calculada pelas Desigualdades em função do erro ε é sempre maior que o calculado pela aproximação Normal; fixados os demais parâmetros. Exemplo, com 95% e viés Zero para qualquer ε o tamanho pela Desigualdade é 5,2 vezes maior, por isso a preferência das empresas de pesquisa pela aproximação Normal.

Na Tabela 4 os limites inferiores dos níveis de confiança para erro ε e viés $\pm v$ foram calculados por Chebychev/Markov. Nas células onde $|v| \geq \varepsilon$ o valor Zero significa que não há garantia probabilística efetiva para NC . Na linha do erro 2% com a amostra de 12500 e com viés Zero tem-se $NC = 95\%$, mas o valor de NC se reduziria a 70% com viés = $\pm 1\%$.

TABELA 4 – Nível de Confiança por Chebychev/Markov.

Erro ε %	Tamanho da Amostra	Limite inferior do Nível de Confiança (%) por Markov em função do Erro (ε) e tamanho segundo o Viés $\pm v$					
		Viés= Zero	Viés= $\pm 1,0\%$	Viés= $\pm 2,0\%$	Viés= $\pm 3,0\%$	Viés= $\pm 4,0\%$	Viés = $\pm 5,0\%$
1%	50000	95,0	Zero	Zero	Zero	Zero	Zero
2%	12500	95,0	70,0	Zero	Zero	Zero	Zero
3%	5556	95,0	83,9	51,0	Zero	Zero	Zero
4%	3125	95,0	88,8	70,0	39,0	Zero	Zero
5%	2000	95,0	91,0	79,0	59,0	31,0	Zero

Fixado erro e tamanho, para viés assumindo valor entre sucessivos $\pm v$ então o NC estará entre valores tabelados. Exemplo, se $\varepsilon=3\%$ e $v = \pm 1,5\%$ então $51,0\% < NC < 83,9\%$.

4ª Parte – Discutindo os Cálculos.

e) As dificuldades da abordagem teórica.

As metodologias registradas no TSE calculam a margem de erro e o nível de confiança supondo não existência de viés, com $Var = 2500/n$ e usando aproximação Normal. Mas, por causa da provável existência do viés, na prática os resultados podem ser inesperados como ficou demonstrado nas tabelas exibidas. Isto sem considerar que no procedimento da amostra por cota a hipótese Normal não é testável e que a Var é uma incógnita que não respeita o valor maximal $2500/n$.

Logicamente não se tem conhecimento dos parâmetros das prévias geradas pelas M replicações PSQ, cuja ideia foi forjada como recurso teórico para organizar a argumentação. Em pesquisas eleitorais não se dispõe de replicações, não há experiências repetidas de amostras por cota em condições constantes. A rigor as replicações por cota não podem ser reproduzidas com fidelidade, mesmo fixando: datas das entrevistas, cotas, tamanho da amostra, critério de seleção das áreas e o comportamento dos entrevistadores; ainda assim as amostras teriam natureza estatística distinta por vários motivos e em especial por causa da seleção não aleatória. Mesmo que fosse possível realizar replicações PSQ, a forma da distribuição das prévias e a relação estatística entre o verdadeiro valor p a Média e a Var possivelmente iriam depender do caso particular em estudo.

Nas metodologias registradas no TSE o problema é simplificado com a suposição que as prévias por cotas atendem a teoria das amostras aleatória, a hipótese da aproximação Normal é adotada e os parâmetros são estimados por fórmulas estatísticas. Mas, nada disso tem fundamento teórico e nem há dados empíricos que justifiquem o método.

Estamos em presença da seguinte dificuldade metodológica, se por um lado a amostra estatística com seleção aleatória é impraticável, por outro lado, as metodologias registradas no TSE fazem uso simplista da equivalência com seleção aleatória oferecendo erros e níveis de confiança que carecem de fundamentações teóricas e empíricas.

Somente para ilustrar possíveis resultados decorrentes das pesquisas divulgadas têm-se na Tabela 5 os níveis de confiança (*NC*) em função do erro e viés aplicando a metodologia usual, onde se fixam os valores dos parâmetros “ $n = 2401$ e $Var = 2500/n = 1,041$ ”, que é a combinação preferida nos registros no TSE pelas grandes empresas,

Pela normalidade e admitindo Viés = $\pm 1\%$ só haveria níveis superiores a 95% para Erro $\geq 3\%$. Por outro lado, ainda com Viés = $\pm 1\%$, se a normalidade fosse falsa não haveria nível de confiança superior a 95% (para erros até 5%). Acrescente-se que nas células $\epsilon > |v|$ quando o tamanho da amostra cresce então o nível de confiança aumenta tanto pelo critério Normal quanto pelas Desigualdades.

TABELA 5 – Tamanho da Amostra $n = 2401$

Nível de Confiança (<i>NC</i>) em função do Viés e segundo o Erro ($n = 2.401$)												
Erro ($\epsilon\%$)	Normal						Chebychev/Markov					
	Viés= Zero	Viés= $\pm 1\%$	Viés= $\pm 2\%$	Viés= $\pm 3\%$	Viés= $\pm 4\%$	Viés= $\pm 5\%$	Viés= Zero	Viés= $\pm 1\%$	Viés= $\pm 2\%$	Viés= $\pm 3\%$	Viés= $\pm 4\%$	Viés= $\pm 5\%$
1%	67,3	47,5	16,2	2,5	0,2	VNS	Zero	Zero	Zero	Zero	Zero	Zero
2%	95,0	83,6	50,0	16,4	2,5	0,2	74,0	49,0	Zero	Zero	Zero	Zero
3%	99,7	97,5	83,6	50,0	16,4	2,5	88,4	77,3	44,0	Zero	Zero	Zero
4%	100,0	99,8	97,5	83,6	50,0	16,4	93,5	87,2	68,5	37,2	Zero	Zero
5%	100,0	100,0	99,8	97,5	83,6	50,0	95,8	91,8	79,8	59,8	31,8	Zero

Alternativas mais abrangentes admitiriam outras formas da distribuição das prévias e outros valores para *Var*, sendo que por Chebychev/Markov bastaria supor alternativas para a *Var*. A proposta significa optar por calcular alternativas fora do padrão metodológico das empresas de pesquisa e apesar das dificuldades de montar formulações plausíveis e das incertezas que o método incorpora, os resultados contribuiriam para o aperfeiçoamento das avaliações dos parâmetros da margem de erro das pesquisas eleitorais.