

# Pesquisa Eleitoral por Cota

## Nota Nº 1

**Luiz Carlos da Rocha**

Pres. do Conselho Federal de Estatística

### Introdução.

O objetivo das Notas é contribuir com a análise de questões da pesquisa eleitoral para que o Sistema Confe/Conre possa colaborar de forma efetiva na elaboração da Instrução Normativa das Eleições, ditada nos anos eleitorais pelo TSE. Para que a pesquisa eleitoral seja divulgada publicamente o TSE tem exigido, dentre outros dados, o registro da margem de erro e do nível de confiança das prévias. Embora a amostra por cota venha sendo aplicada em pesquisa eleitoral no Brasil há pelo menos 30 anos, o cálculo dos parâmetros têm suscitado discussões no meio acadêmico e profissional que perduram até os dias de hoje. A Nota Nº 1 conceitua procedimentos da amostra por cota e analisa a existência do *viés* nas prévias eleitorais e sua influência nos parâmetros da margem de erro.

### 1ª Parte - Amostragem por Cota

#### a) Um resumo do Procedimento

A designação “amostra por cota” é um tipo de amostragem da família dos métodos por *seleção não aleatória* e se destaca por ser largamente usado em pesquisas eleitorais na avaliação das prévias, isto é, na estimação do percentual de votos. Suas características são: a) a população alvo da pesquisa eleitoral é conceitualmente subdividida em subconjuntos em função de certos atributos dos eleitores, b) a composição (%) dos subconjuntos em relação à população alvo é denominado “cotas”, c) a participação (%) dos subconjuntos na formação da amostra é igual às cotas.

Um particular procedimento de amostra por cota é determinado quando: a população alvo é delineada, os subconjuntos são conceituados, as cotas são calculadas e a construção da amostra é detalhada. Os atributos definidores dos subconjuntos são elegíveis supondo-os explicativos na análise e interpretação dos resultados. Os atributos mais usados são: sexo, idade, instrução e classe social; todos divididos em categorias próprias e os subconjuntos são formados pelos cruzamentos dessas categorias. Exemplo, mulheres com idade entre 40 e 45 anos com grau de instrução fundamental e classe social C. O mosaico demográfico e socioeconômico definido pelos atributos e suas categorias cruzadas na população alvo é reproduzido na amostra pela aplicação das cotas, por isso as amostras por cotas são ditas “representativas da população”, mas não necessariamente dos votos. Conhecidas as prévias nos subconjuntos da amostra é possível visualizar padrões comportamentais dos eleitores.

Na entrevista são levantados os dados que identificam o subconjunto do eleitor e são registradas as preferências dos votos e quase sempre um questionário mais abrangente é aplicado contendo questões sobre: religião, uso do telefone e da internet, atividades econômicas, bens materiais, hábitos e estilo de vida. Tais informações podem enriquecer a caracterização dos subconjuntos, ou aprimorar a análise do voto, ou permitir ajustes das prévias ou viabilizar a aplicação de outros métodos de estimação. Perguntas relativas ao marketing eleitoral também são formuladas para o planejamento das campanhas eleitorais.

A escolha dos eleitores da amostra por cota não é aleatória, ou seja, não é regida por critério probabilístico, as entrevistas são realizadas em áreas geográficas definidas no plano amostral e as cotas dos eleitores são atendidas. A pesquisa é quase sempre realizada na rua e a abordagem ao eleitor é ditada por critérios subjetivos do entrevistador, sobretudo com base na aparência e na possibilidade de contato. Raramente se faz entrevista domiciliar, pois é incompatível com o processo já que compromete a rapidez e o baixo custo da pesquisa, que são vantagens relevantes da cota. Na medida em que as cotas vão sendo preenchidas a seleção se torna dirigida e o entrevistador busca na população alvo através do processo de tentativas, os tipos de eleitores que completem as cotas. Vê-se assim, que mesmo na fase da seleção dos tipos faltantes a escolha do eleitor não é intencional no sentido individual.

Ressalte-se que no planejamento da cota não se faz uso de dados individualizados dos eleitores, os subconjuntos são dimensionados com base em dados secundários de órgãos oficiais, IBGE, TSE e outros, só o número de eleitores dos subconjuntos interessa para o cálculo das cotas. O uso do subconjunto na cota difere do uso do estrato na amostragem estratificada, cujo método se utiliza do arquivo de dados individuais da população alvo. Os estratos são recortes do arquivo e geram populações separadas nas quais são selecionados os eleitores que compõem a amostra. É um procedimento mais confiável que a cota porque a seleção dos eleitores dentro dos estratos é aleatória, mas no caso eleitoral é impraticável.

A primeira decisão da amostra por cota na pesquisa eleitoral é geográfica: aonde pesquisar? Uma versão aprimorada da amostra por cota se distingue por fazer seleção aleatória das áreas da pesquisa de campo. Tal versão surgiu nos anos 50 e são conhecidas por “amostras probabilísticas por cota”, cuja sigla em inglês é PSQ. Os métodos PSQ formam nova família de amostras por cota, onde são usados modelos clássicos da amostra estatística para selecionar os locais de entrevistas: conglomerado, estratificada, sistemática e aleatória simples, aplicados em vários estágios da seleção e em função das condições impostas na prática, especialmente em função dos dados secundários disponíveis. As áreas macro da pesquisa: regiões, estados, municípios, subprefeituras, e bairros; são sucessivamente selecionadas segundo o critério da proporcionalidade das suas respectivas populações de eleitores. Ao final, dentro das áreas macro são sorteadas as áreas micros para as entrevistas de campo: quarteirões, setores censitários ou logradouros; onde os eleitores são pinçados para as entrevistas de acordo com as cotas determinadas no planejamento original.

A seleção das áreas pesquisadas da amostra por cota é aprimorado pelo PSQ, mas

ainda assim não atende a seleção probabilística do eleitor e a probabilidade dos eventos estatísticos permanece desconhecida, por isso a impossibilidade do cálculo dos parâmetros da margem de erro. Todavia a amostra estritamente estatística é impraticável na pesquisa eleitoral e ainda que pudesse ser aplicada não haveria garantia que os parâmetros da margem de erro seriam reais. Isto porque a pesquisa de opinião por seu caráter subjetivo, inconstante e intangível, introduz inevitavelmente distorções nos parâmetros da margem de erro por força da ação de fatores adversos, tais como: dúvidas, mudanças de opinião e dissimulações, como também devido às falhas e distorções no levantamento de campo. Todos esses tipos de alterações são chamados de “erros não amostrais”.

Claro que esse ambiente adverso à precisão dos resultados também está presente na amostra por cota, reforçado com a agravante falta de referência probabilística do processo. Portanto além de enfrentar as deformidades geradas pelas opiniões imprecisas e sofrer falhas no levantamento de campo, a amostra por cota por si só, sem a formulação de hipóteses convenientes torna-se impotente no cálculo da margem de erro.

## b) O Relacionamento entre a Estatística e a Amostra por Cota.

A impossibilidade do cálculo da margem de erro na pesquisa por cota está vinculada ao descumprimento da seleção aleatória, que impede prever a probabilidade de ocorrência da prévia em intervalos. Ou seja, à priori não se pode calcular a margem de erro das prévias eleitorais geradas por cota, a menos que se lance mão de hipóteses estatísticas do tipo: “a amostra por cota tem natureza estatística equivalente à amostra aleatória”. Se tal requinte teórico é admitido, então a amostra por cota se transforma num processo idealizado e sob a veracidade dessa hipótese a margem de erro e o nível de significância são calculados por fórmulas estatísticas que dependem somente do valor da prévia e do tamanho da amostra. Assim são calculados os parâmetros da margem de erro nas metodologias registradas no TSE, portanto além das causas de distorções apontadas até aqui, os parâmetros também podem sofrer deformações devido a eventual inadequação da hipótese estatística adotada.

Tais distorções podem alterar os valores teóricos da média e da variância das prévias e com isso modificar os resultados da margem de erro. Por isso convém conjecturar e analisar a influência dessas alterações, sobretudo porque os novos resultados balizam e revelam um cenário mais realista para a margem de erro e o nível de confiança das prévias eleitorais. Os exercícios de simulação serão desenvolvidos na Nota Nº 2 e Nota Nº 3.

Por se tratar de documento dirigido aos profissionais e estudantes de estatística não foram explicitados os procedimentos de cálculo dos parâmetros tabelados e tampouco se detalhou os embasamentos teóricos; da aproximação da distribuição amostral à Normal e das notáveis Desigualdades de Tchebychev/ Markov. Porém, é preciso tecer comentário sobre algumas ideias utilizadas na discussão do problema, a começar pelo procedimento amostral PSQ que será o único método considerado nesta análise; outro aspecto importante da argumentação diz respeito ao comportamento estatístico das prévias sob a condição de

amostras por PSQ, ou seja, sua distribuição estatística amostral cujo histograma é definido pelas replicações PSQ. Entende-se por replicações PSQ, a aplicação repetitiva de amostras por cota PSQ numa determinada população alvo, mantendo-se constante: os subconjuntos e suas cotas, o tamanho  $n$  das amostras, a seleção aleatória das áreas das entrevistas e o período do levantamento das informações no campo. Exemplificando, sejam M replicações PSQ gerando prévias eleitorais de candidatos a Prefeito do Rio de Janeiro usando amostras de tamanho  $n$  distribuídas proporcionalmente as populações das regiões administrativas do RJ e aplicadas em bairros selecionados com probabilidade proporcional a população, cujas entrevistas de campo dar-se-ão num só dia em locais e horários sorteados.

*#: Óbvio que a replicação PSQ é um conceito abstrato e como tal será tratada no texto.*

## 2ª Parte – O desvio e o Viés das Prévias.

### c) A existência do Viés

O **desvio** associado à prévia de um candidato é definido pela diferença entre o valor da sua prévia e o seu real percentual de votos, podendo ser positivo ou negativo. Associado as M replicações PSQ conceitua-se o **viés** como a média dos M desvios. A determinação do *desvio* pressupõe o conhecimento do verdadeiro valor do percentual de votos, que obviamente não se tem ciência a priori. Na hipótese da amostra ser aleatória sem erro de aferição e sem erros não amostrais, então a teoria estatística declara que independente do tamanho da amostra, a *média* das prévias é igual ao verdadeiro valor do percentual de voto, o que caracteriza a propriedade de não tendenciosidade. No caso teórico os desvios seriam exclusivamente de natureza amostral e na medida em que o número de repetições M cresce a soma dos desvios se anula e o viés converge para zero.

Em pesquisa de opinião e especialmente em pesquisa eleitoral a ocorrência de viés é fato naturalmente esperado com qualquer tipo de procedimento amostral, aleatório ou não. Os desvios são muitas vezes gerados pelos próprios entrevistados: respostas impensadas, mudanças naturais de opinião, dissimulações; ou são gerados pelas falhas na captação dos dados: erro no processo amostral, entrevistas mal conduzidas. Esse cenário de defeitos já seria suficientemente turbulento para distorcer os resultados das prévias mesmo que o método fosse estritamente aleatório. Mas, na amostra por cota a situação é agravada pela seleção não aleatória dos eleitores e esta representa uma relevante fonte de geração de viés. Portanto, a *média* das prévias das replicações PSQ é quase sempre igual à soma dos parâmetros: *Verdadeiro valor dos votos + viés*, isto é, a *média* embute o viés, cujo valor poderia ser estimado, teoricamente, com replicações realizadas na véspera das eleições.

Nas pesquisas eleitorais divulgadas, o tamanho da amostra é determinado de modo a satisfazer a prefixação da margem de erro e do nível de confiança, com base na Normal e no pressuposto de viés Zero. Significa que existindo viés, o que é muito provável, o tamanho da amostra calculado subestima a margem de erro ou superestima o nível de confiança.

*Logo, além das mencionadas fontes geradoras de distorção no cálculo dos parâmetros da margem de erro, acrescenta-se a inverossímil suposição do viés Zero.*