

# Pesquisas Eleitorais e as Amostras por Cota

## Nota Nº 1

**Luiz Carlos da Rocha**

### Introdução.

O objetivo das Notas é contribuir com a discussão das questões da pesquisa eleitoral, de modo que o Sistema Confe/Conre possa colaborar de forma mais efetiva nas Instruções Normativas das Eleições ditadas pelo TSE nos anos eleitorais. 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. E, embora a amostra por cota venha sendo aplicada em pesquisa eleitoral no Brasil há pelo menos 30 anos, os parâmetros do erro têm suscitado discussões no meio 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.

### 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* que se destaca por ser largamente usado na avaliação das prévias em pesquisas eleitorais, isto é, na estimação do percentual de votos. Suas características são: a) o eleitorado residente numa área geográfica (a população alvo da pesquisa eleitoral) é conceitualmente dividido 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 interpretação e análise 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 dos 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 perguntas sobre: religião; uso de internet/celular; atividades econômicas; bens materiais; hábitos e estilo de vida. Tais informações enriquecem a caracterização dos

subconjuntos e aprimoram a análise do voto, ou podem 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 e as entrevistas são realizadas em áreas geográficas definidas no plano da pesquisa. A entrevista é quase sempre realizada na rua e a abordagem ao eleitor é ditada por critérios subjetivos do entrevistador levando em conta os atributos da cota. Raramente se faz entrevista domiciliar, pois compromete a rapidez e o baixo custo da pesquisa que são vantagens relevantes da amostra por cota. No fechamento do procedimento, os tipos de eleitores que faltam para completar as cotas são buscados nas áreas geográficas da pesquisa pelo processo das tentativas.

Ressalte-se que no planejamento da cota não se faz uso de dados individualizados dos eleitores, só o número de eleitores dos subconjuntos interessa para o cálculo das cotas que são calculadas com base em dados secundários de órgãos oficiais, IBGE, TSE e outros. A função do subconjunto na amostra por cota difere do estrato na amostragem estratificada, que nesse caso utiliza o arquivo de dados individuais da população alvo. Os estratos são recortes do arquivo principal gerando novas subpopulações, nas quais se selecionam os eleitores que comporão a amostra. É um procedimento mais confiável que a cota porque a seleção é aleatória, mas no caso eleitoral esse método 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 a 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: conglomerado, estratificação e sistemática; que são métodos aplicados na seleção dos locais de entrevistas em função dos dados secundários disponíveis. As áreas macro da pesquisa: regiões, estados, municípios e distritos; são sucessivamente selecionadas segundo o critério da proporcionalidade das suas respectivas populações de eleitores. Em seguida, nas áreas macro municipais são sorteadas áreas micros do tipo regiões administrativas, subprefeituras ou bairros segundo uma natural estratificação socioeconômica e com base nas populações de eleitores; ao final são sorteadas ao acaso as áreas das pesquisas de campo: quarteirões, setores censitários ou logradouros; onde as entrevistas se darão segundo as cotas.

Embora a seleção das áreas pesquisadas por PSQ tenha caráter aleatório, ainda assim a seleção do eleitor continua sendo por cota, logo a probabilidade dos eventos estatísticos permanece desconhecida. Por isso a dificuldade da avaliação dos parâmetros da margem de erro. *Nesse texto a amostra por cota mencionada será sempre a PSQ*, que aliás é o método praticado pelas grandes empresas de pesquisa cujos resultados influenciam o eleitorado.

Na pesquisa eleitoral, a amostragem estritamente estatística é impraticável e mesmo 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, dissimulações e recusas dos entrevistados; além das distorções advindas das falhas no levantamento de campo. Todos esses tipos de perturbações são chamados na amostra aleatória de “erros não amostrais”.

Claro que esse ambiente adverso à precisão dos resultados também está presente na amostra por cota, reforçado no caso com a agravante falta de referência probabilística do processo. Portanto, além de enfrentar as imprecisões causadas pelas opiniões deformadas e de sofrer falhas nos levantamentos de campo, a amostra por cota 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.

O impedimento do cálculo da margem de erro na pesquisa por cota está vinculado ao descumprimento da seleção aleatória do eleitor que impossibilita avaliar a probabilidade de ocorrência da prévia em intervalos. Isto é, 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”.

O ponto crucial da questão é validar a natureza estatística da amostra por cota, em especial sua aderência à distribuição Normal. Se tal requinte teórico fosse verdadeiro então a amostra por cota se transformaria num processo idealizado com a margem de erro e o nível de significância sendo calculados por fórmulas estatísticas que dependeriam somente do valor da prévia e do tamanho da amostra. É exatamente assim que são calculados os parâmetros da margem de erro nas metodologias registradas no TSE. Logo, além das causas de distorções apontadas até aqui, os parâmetros calculados também sofrem 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, mesmo admitida a hipótese Normal. Por isso convém conjecturar e avaliar a influência dessas alterações, sobretudo porque, a produção dos novos resultados pode balizar e revelar um cenário mais realista para o cálculo da margem de erro e do nível de confiança das prévias eleitorais por cotas.

Por se tratar de documento dirigido aos profissionais e estudantes de estatística não foram detalhados os procedimentos de cálculo dos parâmetros tabelados e tampouco se explicitou os embasamentos teóricos; a aproximação da distribuição 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 diz respeito ao comportamento das prévias, ou seja, da sua distribuição estatística amostral sob a condição de amostras por PSQ, cujo histograma é determinado pelas replicações PSQ. Entende-se por replicações PSQ a aplicação repetitiva de amostras tipo 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, o procedimento dos entrevistadores e o período das entrevistas aos eleitores. Exemplificando, sejam  $M$  replicações PSQ gerando prévias de candidatos a Prefeito do Rio de Janeiro através de amostras de tamanho  $n$  divididas proporcionalmente às populações das regiões administrativas do RJ e aplicadas em bairros selecionados com probabilidade proporcional à população de eleitores, cujas entrevistas de campo dar-se-ão na véspera da eleição 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. Com base nas  $M$  replicações PSQ de tamanho  $n$  conceitua-se o **viés** como a média dos  $M$  desvios. O cálculo do *desvio* pressupõe o conhecimento do verdadeiro valor do percentual de votos, que obviamente só é conhecido a posteriori.

Na hipótese da amostra ser aleatória sem erro de informação do eleitor e sem erros não amostrais, a teoria estatística declara que a *média* das prévias é igual ao verdadeiro percentual de voto independente do tamanho da amostra, o que caracteriza a propriedade da não tendenciosidade. Nesse caso ideal, os desvios seriam exclusivamente de natureza amostral e na medida em que o número de repetições cresce, então o viés converge a zero e a média converge ao verdadeiro valor.

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 por falhas na captação dos dados no processo amostral e nas 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 cálculo do tamanho da amostra 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.*