

Este Guia foi elaborado e desenvolvido com a colaboração e orientação acadêmica do CESOP – Centro de Estudos de Opinião Pública da UNICAMP – Universidade Estadual de Campinas (SP), a partir de solicitação da ABEP – Associação Brasileira de Empresas de Pesquisa.

Participaram ativamente de sua formulação e elaboração a Prof. Dra. Rachel Meneguello e o Dr. Clécio da Silva Ferreira, aos quais ficamos muito agradecidos pela dedicação e pela qualidade técnica de sua contribuição, decorrente, com certeza, do profundo conhecimento acadêmico e da familiaridade de ambos com o objeto deste documento, que são as pesquisas de opinião em geral e, em especial, as pesquisas eleitorais.

O seu propósito fundamental é o de tornar transparentes os critérios e fundamentos teóricos da pesquisa eleitoral, bem como seu uso pelas empresas de pesquisa no momento de sua realização.

A ABEP agradece ao CESOP pela competente e inestimável colaboração, responsabilizando-se, porém, pelo conteúdo, forma e ordem de apresentação dos conceitos e metodologias e, em especial, pela escolha de métodos e procedimentos para aplicação prática em projetos de pesquisa, que é feita caso a caso pela empresa responsável por sua realização.

Esclarecimentos ou informações adicionais poderão ser obtidos através do e-mail abep@abep.org.

INTRODUÇÃO

Pesquisas de opinião são a melhor fonte de informação a respeito do pensamento geral de uma população sobre os temas sociais e políticos de um país. Seu papel no estado democrático é central, funciona como mecanismo de vocalização de preferências, garante a presença das distintas vontades no espaço público mais amplo e possibilita a relação de congruência entre o estado e os cidadãos, sobretudo no horizonte do distanciamento em que governantes e governados se relacionam no sistema representativo.

1. METODOLOGIAS BÁSICAS DE PESQUISA: QUALITATIVA x QUANTITATIVA

As distinções entre as duas abordagens de pesquisa são definidas tanto pelos seus métodos quanto suas metodologias. Os métodos referem-se a como os dados são coletados, a metodologia refere-se à escolha e uso da melhor abordagem para tratar de um problema teórico ou prático. Em outras palavras, o método refere-se a 'como coletar os dados', a metodologia refere-se a 'porque coletar os dados de certa forma'. A pesquisa quantitativa permite classificações e contagens das informações coletadas, possibilitando expandir seus achados para uma população mais ampla. Nestes procedimentos, os dados coletados são mais objetivos.

A pesquisa qualitativa busca uma descrição mais completa e detalhada de seu objeto. A análise de dados qualitativos permite distinções aprofundadas, dado que não necessita sistematizar as informações em classificações. No entanto, seus achados não podem ser estendidos a populações maiores com o mesmo grau de certeza que os dados quantitativos permitem.

1.1 PESQUISA QUALITATIVA

OS GRUPOS FOCAIS

O método dos grupos focais tem por base a reunião de indivíduos que interagem e que têm características ou interesses comuns articulados por um moderador. Essa interação resulta em informações sobre temas específicos.

A utilização de grupos focais foi desenvolvida originalmente para a obtenção rápida e eficiente de informações de nível individual em um contexto de dinâmica de grupo. A técnica foi usada inicialmente por pesquisadores de mercado, adotada por Robert Merton e Lazarsfeld nos anos 40 para avaliar as reações da propaganda de guerra (Merton, Fiske, and Kendall [1956] 1990).

Os grupos focais são compostos em geral por entre 7 a 10 indivíduos que não se conhecem entre si e são selecionados porque têm algumas características em comum relacionadas ao tema a ser tratado pelo grupo. O grupo funciona através da condução por um moderador de várias discussões para identificar padrões e tendências nas suas percepções sobre o tema em debate.

1.2 PESQUISA QUANTITATIVA

BREVE HISTÓRICO

Os primeiros registros de enquetes eleitorais datam da campanha presidencial norte-americana de 1824. Este foi o momento de transformação do sistema partidário americano em direção à ampliação da participação política, no qual a eleição presidencial seria pela primeira vez decidida segundo amplo voto popular.

As enquetes, ou levantamentos de opiniões, realizados com procedimentos casuais e nada científicos, ocorriam nas ocasiões de reuniões políticas nos vários estados, e resultavam da curiosidade popular sobre os possíveis resultados, bem como das

atividades dos políticos e partidos para definição dos candidatos ou conhecimento das preferências. Nas situações de campanha eleitoral, a agregação e divulgação dos registros de enquetes cabiam aos principais jornais da época, como o *The Harrisburg Pennsylvanian*, *Delaware Watch* e o *Star and North Carolina Gazette*. Não há evidências, contudo, de que os jornais conduziam, eles próprios, as enquetes.

Assim, as enquetes eleitorais adquiriram popularidade, mas sua realização manteve-se nos limites do âmbito local.

Foi apenas em 1916 que a revista *Literary Digest* realizou uma pesquisa eleitoral nacional e estimou corretamente a vitória de Woodrow Wilson naquela eleição presidencial. A partir de então, a revista passou a realizar pesquisas nacionais e, através do procedimento de envio e devolução por correio de milhares de “cartões de voto”, estimou com acerto as vitórias dos 4 seguintes presidentes da república, incluindo a primeira eleição de Franklin Roosevelt em 1932.

A eleição de 1936

A eleição norte-americana de 1936 foi marcante para a história da pesquisa eleitoral. A mesma revista *Literary Digest* conduziu uma pesquisa nacional e previu equivocadamente a vitória de Alf Landon, candidato republicano, sobre Roosevelt, candidato democrata à reeleição. Com uma amostra elaborada à base de listagem telefônica e registro de proprietários de automóveis, o jornal enviou mais de 10 milhões de “cartões de voto”, mas obteve a devolução por correio de pouco mais de 2 milhões e 300 mil votos, uma participação impressionante, mas que representava menos de 25% do total da amostra. A previsão de vitória do candidato perdedor estava errada por uma margem de 19 pontos percentuais.

Os estudos sobre as causas dessa previsão equivocada mostraram que o viés de uma amostra pouco representativa, embora gigantesca, a baixa participação proporcional da população amostrada, e a alta taxa de não-respostas foram as explicações para o erro da *Literary Digest*. Naquela mesma eleição, George Gallup, fundador em 1935 do American Institute of Public Opinion, acertou o resultado com exatidão, com uma amostra demograficamente representativa de 3.000 entrevistas. O mesmo ocorreu para Elmo Roper, que já em 1935 realizava o *Fortune Survey*, da revista *Fortune*, fundado em procedimentos científicos. Este foi o momento da introdução de métodos científicos na elaboração e realização de pesquisas.

A eleição de 1948

Apesar dos acertos recorrentes das empresas de pesquisa nas campanhas eleitorais, a eleição presidencial de 1948 mostraria as deficiências ainda existentes nas metodologias e Gallup, Roper e Crossley, falharam em prever a vitória de Harry Truman para a presidência da república. Os estudos que buscavam explicações para os erros generalizados apontavam, de um lado, as mudanças de preferências dos eleitores no período mais próximo à eleição e a incapacidade de apreensão dessas alterações pelo trabalho de campo. De fato, a empresa Gallup, por exemplo, havia realizado sua última estimativa eleitoral duas semanas antes da eleição, não levando em conta as alterações das tendências no período final da campanha.

De outro lado, os estudos apontavam como causas os problemas advindos das amostras, algumas realizadas com métodos probabilísticos, outras através de cotas. O relatório do Social Science Research Council - **Report on the Analysis of Pre-Election**

Polls and Forecasts - publicado em 1949, apresentava os argumentos estatísticos de defesa das amostras probabilísticas em relação às amostras por cotas, afirmando que este seria o procedimento que reduziria ao mínimo os erros introduzidos pelo entrevistador na seleção de respondentes. Ficava claramente estabelecida ali a polêmica sobre os procedimentos de amostragem, presente já no meio acadêmico estatístico desde o início do século 20.

No Brasil, a pesquisa de opinião foi introduzida de forma pioneira no país e na América Latina na década de 1940 com a fundação do IBOPE, em 1942. A primeira pesquisa eleitoral foi realizada em 1945, na campanha presidencial instalada com o final da ditadura do Estado Novo e início da democratização. Com 1.000 entrevistas selecionadas em pontos de fluxo na cidade de São Paulo, a pesquisa estimava o resultado da eleição em que concorreram Eduardo Gomes e Eurico Dutra, e mostrava que Eduardo Gomes tinha a preferência de São Paulo¹.

1 Pesquisa Ibope, maio de 1945. Coleção IBOPE, Arquivo Edgard Leuenroth, IFCH/UNICAMP.

I.2.1. AS BASES DA PESQUISA QUANTITATIVA

Conceitos básicos

A. INFERÊNCIA ESTATÍSTICA

Inferência é o processo pelo qual se **generalizam** informações obtidas de uma **amostra** para a **população** de onde essa amostra foi selecionada. Um exemplo típico é aquele em que se tenta **estimar** o número (ou porcentagem) de votos que um candidato vai obter nas eleições (aqui a população de interesse é aquela constituída por todos os indivíduos que vão depositar votos nas urnas) a partir dos dados da amostra selecionada numa pesquisa eleitoral.

Nesse processo de generalização, dois tipos de erros podem ser cometidos:

1) os chamados **erros amostrais**, oriundos das variações naturais existentes de amostra para amostra, e

2) os **erros não amostrais**, gerados por inúmeros fatores, como a definição inadequada da população de interesse, as respostas incorretas fornecidas por indivíduos ou unidades selecionadas, dados demográficos ou eleitorais desatualizados, questionários mal elaborados (perguntas que induzem a determinadas respostas, falta de objetividade, ordem inadequada, vocabulário inacessível etc.), entrevistadores mal treinados, ocorrências inesperadas ligadas ao tema da pesquisa, e inclusive as abstenções eleitorais, conhecidas apenas após a contagem de votos.

1 Pesquisa Ibope, maio de 1945. Coleção IBOPE, Arquivo Edgard Leuenroth, IFCH/UNICAMP.

A proporção de ocorrência de erros não amostrais pode ser reduzida através de uma especificação criteriosa da população de interesse, do treinamento apropriado dos entrevistadores, da elaboração cuidadosa de questionários etc.

A redução da proporção de ocorrência de erros amostrais pode ser obtida através de uma estratificação conveniente e do aumento do **tamanho da amostra** (designado por n), que possuem uma relação inversa entre si: **quanto mais homogênea a população, menor o tamanho da amostra necessário para a obtenção de estimativas com a mesma precisão**. Assim, uma das finalidades da estratificação é obter subpopulações mais homogêneas e conseqüentemente diminuir o tamanho requerido para a amostra. **Nesse contexto, podemos dizer que amostras representativas são aquelas em que a proporção de ocorrência dos dois tipos de erros - amostrais e não amostrais - é minimizada, ou, pelo menos, quantificada. É devido à possibilidade de ocorrência desses erros que os resultados de pesquisas realizadas através de métodos probabilísticos devem sempre incluir a margem de erro das estimativas.** Somente uma amostra selecionada através de métodos probabilísticos permite quantificar a magnitude e a proporção de ocorrências dos erros amostrais.

Em uma situação eleitoral, quando dizemos que a porcentagem esperada (dentre todos os votantes e não somente entre aqueles selecionados na amostra) de votos favoráveis a um determinado candidato "A" é de 41% com margem de erro de 2 pontos percentuais, queremos dizer que temos grande confiança que essa porcentagem populacional deve estar entre 39% e 43%.

Se a porcentagem esperada de votos favoráveis ao candidato "B" for de 40% com margem de erro de 2 pontos percentuais, o máximo que poderemos afirmar é que temos grande confiança que a diferença entre as porcentagens de votos favoráveis a esses dois candidatos esteja entre 5% (= 43% - 38%) em favor do candidato "A" e 3% (= 42% - 39%) em favor do candidato "B". Nesse caso, dizemos que há um **empate técnico** entre os dois candidatos. A interpretação de "**grande confiança**", aqui, tem uma natureza mais técnica, e essencialmente quer dizer que se repetirmos esse procedimento de estimação um grande número de vezes, em 95 de cada 100 (por exemplo), estaremos fazendo uma afirmação correta; neste caso, dizemos que o **coeficiente de confiança** do procedimento é de 95%.

O erro amostral é também chamado de margem de erro, e pode ser definido como um intervalo de possíveis resultados, consideradas as características da amostra (OLIVEIRA e GRÁCIO, 2005).

Erro amostral tolerável é a diferença tolerada, pelo pesquisador, entre o valor que a estatística acusa e o verdadeiro valor do parâmetro que se deseja estimar (Barbetta, 1999). Quanto menor o erro amostral tolerado pelo pesquisador, maior é o tamanho da amostra necessário para se atendê-lo. Considerando que o erro amostral tolerável representa o quanto o pesquisador admite errar na estimação dos parâmetros de interesse, quanto menor o erro amostral tolerado pelo pesquisador, maior será o tamanho da amostra para que se possa obter essa maior precisão da estatística.

Nível de confiança é a probabilidade de o valor do parâmetro em estudo pertencer ao intervalo centrado na estimativa estatística do parâmetro e limites determinados pelo erro amostral tolerado. Sem conhecer o tamanho N da população, o tamanho de uma

amostra aleatória simples em função de uma **variável dicotômica** é obtido através da seguinte expressão:

$$n = \frac{z_{\alpha}^2 p(1-p)}{E^2} \quad (I)$$

Onde z_{α} = valor da distribuição normal para o nível de confiança $(1-\alpha)$; p = proporção do evento na população, $0 < p < 1$; E = erro amostral tolerável.

O intervalo de confiança para a proporção p , ao nível de confiança de $(1-\alpha)$ é da forma

$$IC_{1-\alpha}(p) = \left(\hat{P} - z_{\alpha} \sqrt{\frac{P(1-P)}{n}} ; \hat{P} + z_{\alpha} \sqrt{\frac{P(1-P)}{n}} \right).$$

Considerando-se uma situação em que se conhece uma estimativa da variação populacional obtida por meio de um levantamento-piloto ou em pesquisas prévias ($p = \hat{p}$), pode-se utilizar esta estimativa como cálculo do tamanho da amostra na equação (I). Por exemplo, conhecendo-se que a estimativa para p na população em estudo seja de 0,1, para um nível de 95% de confiança ($z_{\alpha} = 1,96$) e $E = 0,025$ (ou 2,5%), pode-se trabalhar com uma amostra constituída por:

$$: \frac{1,96^2 * 0,1 * 0,9}{0,025^2} = 553 \text{ elementos.}$$

Caso não se tenha uma estimativa de p , trabalha-se com $p = 0,5$ (que reproduz uma variância populacional com a maior heterogeneidade possível, $p(1-p) < 0,25$). A fixação da proporção populacional do evento em 0,5 deve-se ao fato de ser esta a pior situação possível em termos de variabilidade populacional. Neste caso, o tamanho da amostra é dado por:

$$n = \frac{z_{\alpha}^2}{4E^2}.$$

Assim, para o mesmo nível de 95% de confiança ($z_{\alpha} = 1,96$) e mesmo erro amostral $E = 0,025$ (ou 2,5%), pode-se trabalhar com uma amostra constituída por:

$$n = \frac{1,96^2}{4 * 0,025^2} = 1537 \text{ elementos,}$$

ou seja, uma amostra três vezes maior que a primeira. Se diminuirmos o nível de confiança para 90%, o valor de n diminui para 1089. Por outro lado, mantendo-se o

nível de confiança em 95%, mas diminuindo-se o erro amostral para 2 pontos percentuais, o valor de n aumenta para 2401 indivíduos.

Com base nas constatações acima descritas, pode-se observar que o tamanho da amostra diminui em função do fato de a população ser mais homogênea (a estimativa de p se distancia de 0,5), diminui com o decréscimo do nível de confiança e aumenta com a maior precisão desejada (menor erro amostral tolerado). Portanto, quanto maior a precisão que se deseja associar à estimativa estatística, maior o tamanho amostral necessário para atendê-la.

B. O EMPATE TÉCNICO

Uma situação é considerada empate técnico quando a diferença entre os candidatos se encontra dentro das margens de erro das pesquisas, ou seja, quando há superposição dos respectivos intervalos de confiança dos candidatos.

Suponha que em determinada pesquisa a margem de erro seja de 3 pontos percentuais e o candidato **A** tenha 27% das intenções de voto e o candidato **B** tenha 30%.

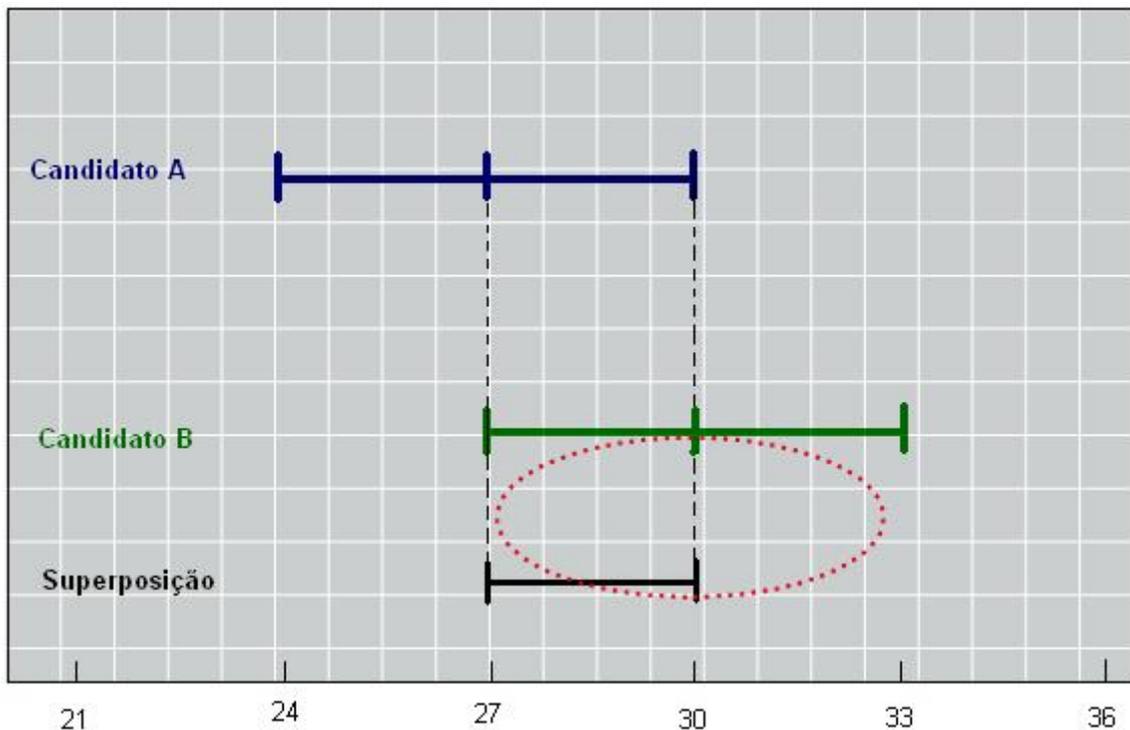
Os intervalos de confiança construídos para cada um deles são os seguintes:

Candidato A: $[27\% - 3\%; 27\%+3\%] = [24\%;30\%]$

Candidato B: $[30\% - 3\%; 30\%+3\%] = [27\%;33\%]$

Gráfico de porcentagens obtidas pelos candidatos A e B (pontos médios dos intervalos)

Obs.: os extremos são os Limites de Confiança



Quando existe uma superposição nos dois intervalos de confiança, pode-se dizer que existe um empate técnico entre os dois candidatos.

C. A AMOSTRA

Uma população pode ser infinita ou, mesmo sendo finita, pode ser muito grande, tornando-se muito difícil ou impraticável a obtenção dos dados de toda a população. Assim sendo, o pesquisador observa apenas uma parte da população, que é denominado **amostra**. Formalmente, denomina-se amostra um conjunto de objetos (ou de dados) retirados de uma população para fins de estudo da mesma. O número de objetos (ou de dados) de uma amostra é o seu tamanho e é representado geralmente por n . Uma amostra deve ser representativa da população da qual foi extraída.

O termo **amostra** refere-se também ao conjunto de dados de uma variável numa amostra de objetos. A amostragem consiste em obter os dados de uma ou mais variáveis numa amostra de objetos retirados aleatoriamente da população. Na amostragem, os dados são obtidos com rapidez e o custo operacional é bem menor do que no caso do censo.

I.2.2. A AMOSTRAGEM

A amostragem constitui um conjunto de procedimentos para a obtenção de uma amostra representativa da população-alvo e um estudo das relações existentes entre uma população e as amostras possíveis que podem ser extraídas da mesma.

A .TÉCNICAS DE AMOSTRAGEM

Num procedimento de amostragem, **as amostras extraídas devem ser representativas de uma população**, isto é, a variabilidade da população deve estar presente nas amostras e, para que isto ocorra, os objetos da população que irão constituir uma amostra devem ser escolhidos aleatoriamente.

A amostragem pode ser probabilística ou não-probabilística. Na amostragem probabilística, pode-se determinar a probabilidade de escolher um objeto da população, enquanto que na amostragem não-probabilística, esta probabilidade não pode ser calculada.

B. TIPOS DE AMOSTRAGENS PROBABILÍSTICAS AMOSTRAGEM ALEATÓRIA SIMPLES

É a mais conhecida das estratégias de amostragem. Uma amostra aleatória simples é aquela em que toda unidade (indivíduos, casos) da população estudada tem igual chance de ser incluída na amostra, e a probabilidade de uma unidade ser selecionada não é afetada pela seleção de outras unidades. Assim, a seleção é feita de forma independente.

Esta técnica de amostragem é utilizada quando a população é finita, não é demasiadamente grande e é considerada homogênea. Ela consiste em enumerar os **N** objetos de uma população e em seguida extrair uma amostra de **n** objetos da mesma utilizando bolas numeradas, fichas numeradas, números aleatórios gerados por programas de computador ou tabela de dígitos aleatórios.

Apesar de a amostra aleatória ser a base da teoria estatística, ela raramente é usada na prática. Isso ocorre porque nem sempre é possível satisfazer as condições exigidas

por este tipo de amostra, pois ela requer uma listagem completa dos indivíduos do universo a ser estudado.

AMOSTRAGEM SISTEMÁTICA

A amostra sistemática baseada em listagem é uma adaptação da amostra aleatória simples. O sorteio aleatório dos indivíduos em listagens muito extensas pode ser um processo bastante trabalhoso. Sendo assim, os pesquisadores preferem usar um procedimento sistemático para o sorteio. A amostra sistemática começa pela determinação do chamado “intervalo de amostragem”, que é a distância padrão na listagem entre os indivíduos selecionados. Por exemplo, se a lista de indivíduos da população tem 10 mil pessoas, e se desejamos extrair uma amostra de mil pessoas, o intervalo amostral será de 10. Assim, extrai-se cada décimo indivíduo da lista da população para constar na amostra. Para garantir a estrita aleatoriedade, o início da extração é determinado por sorteio. Sorteia-se um número dentro do intervalo de amostragem, neste caso um número de 1 a 10, e a partir deste número extrai-se cada décimo indivíduo da lista.

A amostragem sistemática é muito útil quando se deseja estudar uma população que vai se completando ao longo do tempo. Neste caso, para calcular o intervalo de amostragem, considera-se o número de objetos da população no momento de se extrair a amostra.

Amostragem por conglomerados

Voltando às dificuldades da amostra aleatória simples, uma das razões da pouca utilização desse tipo de amostra é que não existem listagens disponíveis para grandes populações. Sendo assim, adaptou-se a amostra aleatória simples através da introdução de múltiplos estágios de seleção, onde os estágios iniciais são os chamados conglomerados e o mesmo princípio de aleatoriedade se aplica a cada um deles. Um conglomerado é uma unidade que aglomera pessoas. Ao se sortear aleatoriamente um conglomerado, a lógica é a mesma do sorteio aleatório de indivíduos. Por exemplo, para se “amostrar” os moradores do Brasil, é preciso primeiro tirar uma amostra dos municípios do país. A amostragem por conglomerados tem como característica o fato de que a variância dentro de cada conglomerado tende a ser maior do que entre os conglomerados.

A amostra por conglomerados é utilizada quando o pesquisador busca gerar uma amostra probabilística mais eficiente quanto aos recursos de tempo e dinheiro. Em vez de uma amostragem de unidades individuais, em geral espalhadas em espaços geográficos mais amplos, o pesquisador faz a amostragem segundo grupos que já estão constituídos na população, como as vizinhanças, as escolas etc.

Amostragem estratificada

É uma amostra desenhada pelo pesquisador quando ele deseja uma amostra representativa da população segundo algumas características específicas. Os procedimentos são mais complexos quando o pesquisador busca definir subgrupos também representativos nesta amostra, e seu procedimento é combinar as estratégias de amostragem aleatória e estratificada.

A amostra estratificada foi desenvolvida como uma forma de aumentar a precisão do processo amostral, reduzindo o grau de heterogeneidade presente na amostra

aleatória simples. O sorteio dos indivíduos também é feito a partir de uma listagem, contudo, usa-se uma característica que seja importante na explicação do fenômeno em estudo para distinguir subpopulações, em que o sorteio é realizado separadamente. Um caso importante da amostragem aleatória estratificada é aquele em que o pesquisador obtém uma amostra com a mesma proporcionalidade dos estratos que compõem a população. Neste caso, a amostragem é denominada amostragem estratificada com partilha proporcional.

A amostragem estratificada tem como característica a menor variação dos dados dentro de cada estrato do que entre os estratos.

AMOSTRAGEM UTILIZANDO MÚLTIPLAS TÉCNICAS PROBABILÍSTICAS

Esta situação ocorre quando os pesquisadores utilizam 3 técnicas básicas de amostragem conjugadas entre si para gerar amostras mais complexas. Exemplo: a amostra de conglomerados múltiplos é uma técnica que envolve:

- a) um primeiro estágio de seleção aleatória de conglomerados;
- b) um segundo estágio em que as unidades de interesse são amostradas dentro dos conglomerados.

Outra situação pode ocorrer envolvendo uma amostra de quatro estágios: município, setor censitário, domicílio e indivíduo. Em cada estágio o sorteio dos elementos é feito através de procedimentos aleatórios:

- 1) O município será o primeiro estágio de amostragem. Dentro de cada município é necessário listar e sortear as diversas áreas. No Brasil, o IBGE divide o território de cada município do país nos seus chamados setores censitários, a unidade geográfica básica usada para o arrolamento da população nos censos;
- 2) A partir da lista dos setores censitários é possível obter uma representação geográfica adequada de todo o município. Dentro de cada setor censitário produz-se, então, uma lista de todos os domicílios; a partir desta lista sorteiam-se os domicílios a serem visitados;
- 3) Dentro do domicílio, os moradores são listados e opera-se um último sorteio para se determinar o indivíduo que será entrevistado.

A amostra de conglomerados não usa como primeiro estágio de amostragem apenas listagens de unidades geográficas. Dependendo da pesquisa em questão, é possível usar como conglomerado qualquer unidade que concentre pessoas. Por exemplo, para se investigar os estudantes de um município, o primeiro estágio de amostragem pode ser a escola e, dentro da escola, sorteia-se o estudante. Ou, ainda, uma organização de grande porte que possua centenas de agências pode usar a agência como seu primeiro estágio de amostragem; após ser sorteada a agência a ser pesquisada, sorteiam-se em cada agência os usuários a serem entrevistados.

C. TIPOS DE AMOSTRAGEM NÃO-PROBABILÍSTICA

Neste processo de amostragem os objetos são escolhidos através da seleção realizada por um entrevistador, e não através da probabilidade de escolher os objetos da população.

AMOSTRAGEM TIPO “BOLA DE NEVE”

Este procedimento é utilizado em algumas pesquisas de opinião. A amostra é formada por todas as pessoas que vão aparecendo onde o pesquisador está até completar o tamanho definido. Por exemplo, se um pesquisador deseja obter uma amostra de tamanho 30, ele entrevista as primeiras 30 pessoas que aparecerem.

AMOSTRAGEM INTENCIONAL

Neste procedimento, a amostra é escolhida intencionalmente com base em determinados critérios. Por exemplo, numa pesquisa sobre uma doença infecto-contagiosa escolhem-se as pessoas expostas ao risco de tal doença.

AMOSTRA POR QUOTAS (ou COTAS)

A amostra por quotas é certamente a mais usada no Brasil, devido ao seu custo e prazo de execução muito reduzidos, em comparação com os outros tipos de amostra. A amostra por quotas é um processo amostral não-aleatório, e faz parte de outra classe de desenhos amostrais, as chamadas **amostras intencionais**, em que o entrevistador intencionalmente busca certos perfis de indivíduos, um procedimento que não atende à desejada aleatoriedade estatística.

A amostra por quotas se baseia na idéia de reproduzir na amostra as características da população consideradas mais importantes para explicar um fenômeno. O procedimento é, então, estabelecer quotas de entrevistas para cada uma das combinações destas características. Smith (1983) determina condições para que se possa ignorar o plano amostral em estudo. No caso das cotas, as variáveis de controle devem conter toda a informação presente na variável de interesse.

A amostragem por cotas é semelhante à amostra estratificada, devido à tentativa de controlar características relevantes da população em estudo. No entanto, distingue-se daquele tipo de amostra pelo fato de não se basear em uma listagem dos indivíduos da população. Na amostra por quotas, os entrevistados são selecionados em um determinado local pré-estabelecido até que se preencham as quotas determinadas na amostra. Nesse último caso, é mais difícil especificar a probabilidade de cada unidade da população ser selecionada para a amostra, mas mesmo assim é comum utilizar a mesma técnica estatística empregada para amostras probabilísticas, sob a **suposição** de que todas as possíveis seleções são **equiprováveis**.

O principal problema da amostra por quotas é que pode haver outra característica da população que esteja associada com a variável de interesse e não faz parte das características das cotas, o que torna a amostra deficiente, pois parte da informação pertencente à variável de interesse pode ter sido perdida.

Em contextos onde se conhecem bem as características da população associadas a certo fenômeno, a amostragem por quotas é utilizada com bastante êxito. O melhor exemplo são as pesquisas eleitorais. As pesquisas eleitorais utilizam quotas de sexo, idade, escolaridade e região geográfica, e com isso conseguem prever corretamente, na maioria das vezes, o resultado das eleições.

Infelizmente, em poucos campos de pesquisa se conhece tão bem as características relevantes da população em estudo. As pesquisas eleitorais estão em um contexto

privilegiado, onde há grande tradição de amostra por quota. Na maioria dos contextos de pesquisa, no entanto, é impossível determinar a qualidade das pesquisas por quotas.

AMOSTRA SISTEMÁTICA DE FLUXO

A amostra sistemática de fluxo é comumente usada para pesquisas em locais limitados. Por exemplo, em uma pesquisa com os usuários do metrô pode-se selecionar os entrevistados no fluxo de pessoas que circulam nas estações; em uma pesquisa com os usuários de um serviço médico, pode-se selecionar os entrevistados no fluxo de pessoas que saem do serviço.

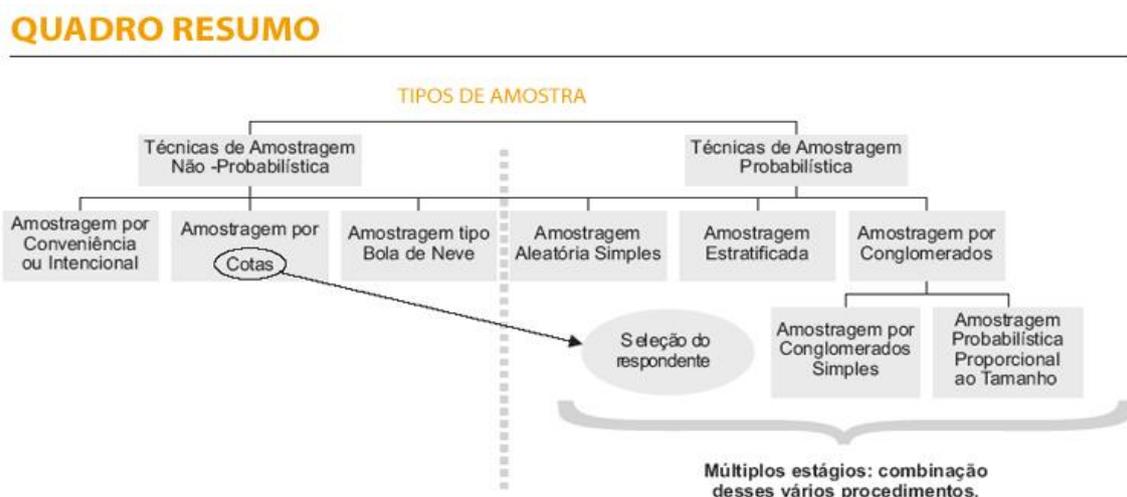
Este método de amostragem parte de um princípio semelhante ao da amostra sistemática com listagem: a extração de cada n-ésimo indivíduo que passa por um local para ser entrevistado. A determinação do intervalo de amostragem é feita a partir de informações quanto ao fluxo total de pessoas que são atendidas em cada organização em um mês, por exemplo. Assim, se uma organização atende 10 mil pessoas por mês, e se deseja extrair uma amostra de mil pessoas, o intervalo de amostragem será novamente de 10. Contudo, as mil entrevistas da amostra devem ser distribuídas pelos diferentes períodos do dia (manhã, tarde e noite, se for o caso), dias da semana e até mesmo semanas do mês proporcionalmente às estimativas do fluxo de pessoas nestes períodos.

A razão para isso se encontra na idéia de que diferenças na magnitude do fluxo de pessoas podem estar indicando diferenças no tipo de usuário que frequenta uma organização em cada período.

D. AMOSTRAGEM EM MÚLTIPLOS ESTÁGIOS

Um procedimento de amostragem pode ser realizado em várias etapas e neste caso tem-se uma amostragem em múltiplos estágios. O objetivo é combinar os diversos tipos de amostragens utilizando as vantagens de cada tipo. Numa amostragem em dois estágios, por exemplo, pode-se no primeiro estágio empregar a amostragem por conglomerados e num segundo estágio empregar a amostragem aleatória simples.

Os principais esquemas amostrais estão representados no diagrama abaixo:



2. COMO SE FAZ PESQUISA NO BRASIL

No Brasil, são geralmente realizadas utilizando amostragem em múltiplos estágios. No meio acadêmico, utiliza-se, por exemplo, o procedimento adotado pelo IBGE (Instituto Brasileiro de Geografia e Estatística): **amostragem estratificada proporcional de conglomerados, selecionados com probabilidade proporcional ao tamanho (PPT)**, da seguinte forma:

- 1) amostragem estratificada de conglomerado (com PPT) em dois estágios. Os estratos são as Unidades da Federação (UF) e a amostra é proporcional à população com 14 anos ou mais de cada UF. Os conglomerados selecionados no primeiro estágio são os municípios dentro de cada UF.
- 2) no segundo estágio, os conglomerados são os setores censitários dentro dos municípios selecionados no primeiro estágio. A medida de tamanho usada para a seleção dos conglomerados com probabilidade proporcional ao tamanho (PPT) é a população na faixa etária especificada dos municípios e setores.

Em pesquisas de mercado e de intenção de voto, adota-se, em geral, o procedimento de múltiplos estágios, com estratificação e conglomerados no(s) primeiro(s) estágio(s) (regiões, municípios, setores censitários) e em seguida, no último estágio, incorporam-se cotas por sexo, idade, grau de instrução etc, definidas segundo informações do IBGE e do Tribunal Superior Eleitoral. As cotas são utilizadas para que o entrevistador aplique entrevistas a pessoas que têm uma probabilidade de resposta baixa, evitando-se, com isto, possíveis vieses na amostra. Por exemplo, são definidas cotas de PEA (População Economicamente Ativa) e Não-PEA para impor que pessoas que trabalham e que não trabalham pertençam à amostra.

A motivação principal desse procedimento são as conseqüências do erro de não-resposta, e não somente os benefícios que traz quanto ao custo e ao tempo utilizado. O método utilizado - de estabelecer cotas dentro dos setores selecionados probabilisticamente - pode ser considerado como aproximadamente auto-ponderado e sua função é evitar as possíveis distorções que seriam introduzidas pelos entrevistadores, caso não houvesse cotas.

Embora seja amplamente conhecido o fato de que amostragem por cotas não possibilita o cálculo do erro amostral (ou margem de erro), pois ela não atende aos princípios da aleatoriedade estatística, os institutos de pesquisa adotam conscientemente o modelo de amostragem em múltiplos estágios, envolvendo cotas no último estágio, como um modelo aproximativo em relação a um modelo ideal (do ponto de vista probabilístico).

Problemas como o tempo de realização e o custo de execução da pesquisa, bem como as imposições da Lei Eleitoral, que obriga a declaração da margem de erro no momento do registro da pesquisa junto à Justiça Eleitoral, são os principais argumentos utilizados pelos institutos pela escolha do método aplicado.

Outro problema técnico, porém não muito exposto nos veículos de comunicação e pelas empresas, diz respeito à forma do cálculo das margens de erro nas pesquisas eleitorais. No primeiro turno, onde geralmente há mais de dois candidatos, a variável de interesse “candidato em quem votaria” é politômica, ou seja, possui três ou mais possibilidades de escolha, candidatos A, B, C ou outros. Dessa forma, o cálculo da margem de erro apresentado anteriormente para variáveis dicotômicas não é o apropriado, e sim, o modelo Multinomial.

O que se faz geralmente é calcular a margem de erro comparando cada categoria, ou candidato, contra as demais. No exemplo acima, supondo que se tenha K candidatos, faz-se K comparações e suas respectivas margens de erro.

O principal problema desse método é considerar a mesma margem de erro para todos os candidatos, pois cada um deve ter o seu próprio cálculo.

Os jornais impressos e telejornais compreenderam que devem informar nas reportagens eleitorais o dado da margem de erro. Entretanto, as limitações de tempo e espaço e, de modo muito especial, a complexidade dos conceitos técnicos envolvidos tornam muito difícil a sua comunicação para as distintas audiências.

Esse não é um problema exclusivamente nacional, essas mesmas dificuldades ocorrem, por exemplo, com a divulgação de eleições norte-americanas (Larson, 2003); contudo, vale lembrar que a confusão interfere na validade das informações veiculadas.

Modos alternativos de estimação dos parâmetros de proporção e seus respectivos intervalos de confiança têm sido estudados por alguns pesquisadores da área, utilizando conceitos de Inferência Bayesiana. Os quadros abaixo sistematizam os tipos de amostra de natureza probabilística e não-probabilística, segundo seus procedimentos básicos, vantagens e desvantagens.

QUADRO 1. AMOSTRAS PROBABILÍSTICAS

	AMOSTRA ALEATÓRIA SIMPLES	AMOSTRA ALEATÓRIA SISTEMÁTICA	AMOSTRA ALEATÓRIA ESTRATIFICADA		
			Proporcional	Não-Proporcional	Amostra por conglomerados
Procedimentos básicos	Listagem completa do universo; Definição de um número a cada indivíduo do universo; Através de uma tabela de números aleatórios seleção de indivíduos que irão compor a amostra.	Listagem completa do universo; Seleção do primeiro indivíduo através de um método aleatório; Seleção de cada i -ésimo indivíduo a partir do primeiro.	Divisão do universo em estratos internamente homogêneos; Seleção de indivíduos de modo aleatório, em cada estrato, As frações de amostra, em cada estrato, são proporcionais.	Idem, proporcional; Idem, proporcional; As frações em cada estrato podem ser distintas segundo a necessidade.	Divisão do universo em diversos grupos ou clusters; Seleção dos aglomerados que devem constituir a amostra; Dentro de cada aglomerado, seleção dos indivíduos da amostra, de modo aleatório.
Vantagens técnicas	Todos os indivíduos têm igual probabilidade de inclusão; Permite generalizar; Proporciona uma base para calcular o grau de disparidade entre as medidas da amostra e do universo.	Idem à amostra aleatória simples	Idênticas à simples e à sistemática; Garante a representatividade; Elimina erros de estrato.	Idêntica à simples e à sistemática; Possibilita melhor conhecimento de grupos pequenos no universo.	Idêntica à simples e à sistemática.
Vantagens econômicas	As que derivam de não investigar todo o universo.	Maior facilidade para obter a amostra;	Idênticas à simples e à sistemática; Garante a	Se os estratos forem homogêneos, torna possível	Economiza dinheiro porque concentra os entrevistadores

			representatividade; Elimina erros de estrato.	uma amostra menor e mais econômica.	em áreas próximas; Economiza tempo.
Desvantagens técnicas	Não provê um número suficiente de casos em grupos especiais; Pode haver distorções quanto à representatividade.	Nem todas as combinações têm igual probabilidade de ocorrer na amostra; Pode haver algum tipo de viés na listagem.	Idem ao item 1 anterior; Pode não prover um número suficiente de casos para estratos pequenos; Dificuldade para determinar estratos homogêneos.	Idem ao item 1 anterior; Exige tratamentos estatísticos complexos; Dificuldade para determinar estratos homogêneos.	Idem ao item 1 anterior; Exige tratamentos estatísticos complexos; Perda do caráter aleatório da amostra.
Desvantagens Econômicas	Alto custo financeiro; Alto custo de tempo.	Idem à amostra aleatória simples	Custo mais alto que na amostra simples e de conglomerado; Alto custo de tempo.	Custos mais altos que para a amostra por clusters	Mínimas

QUADRO 2. AMOSTRAS NÃO PROBABILÍSTICAS

TIPO DE AMOSTRA	BOLA DE NEVE	INSTITUCIONAL	COTAS
Procedimentos básicos	Entrevistas casuais de um número x de indivíduos (i.e. os que passam numa esquina)	Seleção de casos típicos do universo, segundo a relevância teórica	Seleção de uma cota de indivíduos em cada categoria pertinente
Vantagens e Desvantagens	Exige pessoal menos treinado e menor custo; Apresenta obstáculos à generalização		

3. QUESTIONÁRIO E MATERIAL DE APOIO

Os questionários são o instrumento central da pesquisa eleitoral, pois dele constam as estratégias para conhecimento das opiniões da população sobre o assunto pesquisado. Em geral, as perguntas de acompanhamento das campanhas eleitorais são padronizadas, de forma a possibilitar a replicação e o estudo das tendências do eleitorado.

Os questionários podem ser compostos de perguntas abertas ou fechadas. As abertas têm como vantagem possibilitar o conhecimento mais detalhado de opiniões não antecipadas pelas hipóteses da pesquisa, mas têm a desvantagem de dificultar uma análise comparativa; as perguntas fechadas têm como vantagens possibilitar a replicação do estudo e a comparação de informações, sua desvantagem está na simplicidade das respostas.

No entanto, mesmo os formatos das perguntas fechadas podem ser diversificados, ainda que para dimensionar um mesmo fenômeno, e essas diferenças podem ter implicações nos resultados.

Assim, por exemplo, quando se pergunta ao entrevistado sobre seu voto nas próximas eleições, as variações do enunciado quanto ao tempo mencionado e a forma de apresentação dos candidatos podem levar a estimativas distintas:

Exemplos:

Em quem você vai votar na eleição de Novembro?

Se a eleição fosse hoje, em que você votaria?

Destes candidatos (mostra a cartela com nomes), em quem você votaria na próxima eleição?

O mesmo ocorre com a pergunta sobre rejeição das candidaturas. Por exemplo, as formas para se medir o nível de rejeição podem solicitar a menção a um ou mais candidatos, ou ainda solicitar a rejeição em graus de intensidade.

Exemplos:

Se a eleição fosse hoje, em qual destes candidatos o (a) Sr (a) não votaria de jeito nenhum?

E em qual (is) candidato(s) o (a) Sr (a) não votaria de jeito nenhum?

Para cada um dos candidatos que eu mencionar, gostaria que o (a) Sr (a) me dissesse se votaria nele com certeza, poderia votar ou não votaria de jeito nenhum?

Nos casos acima, a primeira formulação solicita uma resposta única, ou seja, identifica qual é o candidato mais rejeitado. Na segunda pergunta, o entrevistado pode citar um ou mais candidatos (respostas múltiplas) e o resultado dá o nível de rejeição comparativo entre os vários candidatos. Na terceira formulação, identifica-se o nível de rejeição individual de cada candidato, sem qualquer comparação entre eles.

Também é importante para a pesquisa quantitativa definir a ordem das questões no questionário. Dependendo dos assuntos colocados antes da pergunta central da pesquisa, seja ela sobre a aprovação do governo, confiança no político ou intenção de voto, a resposta pode ser afetada (Patterson, K. D., Snow, G. L., Christensen, H. B. and Olsen, J. A., 2003).

Por exemplo, em um período de crise econômica, se se perguntar às pessoas o que elas pensam da economia do país antes de perguntar sua opinião sobre o Presidente, a taxa de popularidade presidencial será provavelmente mais baixa do que se a ordem das perguntas estiver invertida.

3.1. OS ANEXOS DO QUESTIONÁRIO

CARTELAS

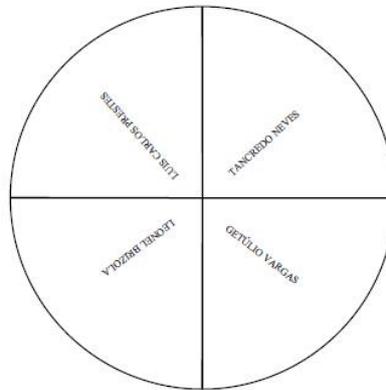
Em geral os questionários vêm acompanhados de cartelas que apresentam alternativas de respostas para cada pergunta. A apresentação dessas cartelas pelo entrevistador é um procedimento que permite ao entrevistado refletir com mais tempo e qualidade sobre sua resposta.

Escala de concordância

Discorda muito				Concorda muito		
1	2	3	4	5	6	7

Discos

Da mesma forma, nas pesquisas eleitorais as perguntas sobre preferências entre candidatos e intenção de voto vêm acompanhadas de um disco com nomes de candidatos. Esse é um procedimento utilizado para evitar que um possível ordenamento dos nomes de candidatos tenha alguma influência na escolha do entrevistado. Exemplo:



4. COLETA DE DADOS

As formas mais comuns de coleta de dados em pesquisas de opinião são:

1. as entrevistas pessoais. Esta é a técnica mais tradicional de pesquisa, mas também é a técnica de maior custo pois implica em um tempo e uma logística maiores para a coleta de dados.

2. entrevistas telefônicas. Esta uma das técnicas mais utilizadas em pesquisas de opinião, pois as entrevistas podem ser realizadas em menor tempo e com custos reduzidos, já que não implicam em deslocamento de pesquisadores. As questões aplicadas são relativamente mais simples e necessárias para uma abordagem impessoal. Sua metodologia de amostragem, no entanto, requer uma cobertura de acesso telefônico domiciliar completa, um aspecto que pode inviabilizar sua aplicação em determinadas regiões com cobertura limitada.

3. entrevistas por mala direta. Esta é a técnica menos utilizada, pois embora implique em custos muito reduzidos (os custos de impressão do questionário e os custos postais), as taxas de retorno dos questionários são sempre incertas e, em geral, muito baixas. Assim, sua utilização ocorre em pesquisas que estudam grupos populacionais menores e específicos.

4. pesquisas por internet. Esta é a técnica utilizada para pesquisas muito específicas, pois as fontes de amostragem são também específicas e limitadas (lista de assinantes, listas de consumidores, diretórios de web sites, listas de endereços eletrônicos etc). Algumas críticas a essa técnica baseiam-se na informação de que o pequeno acesso ou posse de recursos de informática presentes em muitos lugares impedem sua aplicação. Na verdade, o tamanho da cobertura e acesso ao recurso da internet conta menos para os procedimentos de pesquisa. O que conta como problema, assim como para as entrevistas por mala direta, são a técnica do auto-preenchimento e a taxa de retorno do questionário.

5. A COMPARAÇÃO ENTRE PESQUISAS

Quando em épocas de eleição, os vários órgãos de imprensa (emissoras de TV, rádio, jornais, internet e revistas) divulgam os resultados de pesquisas encomendadas a diversos institutos, é inevitável que sejam feitas comparações entre as diferentes proporções e entre as margens de erros calculadas para cada candidato.

Ocorre que, além dos erros amostrais já abordados neste texto, tais diferenças devem-se a vários fatores. Por exemplo, a evolução das tendências entre os eleitores ao longo do processo eleitoral pode levar a que pesquisas realizadas com alguma diferença no tempo reflitam as variações das preferências.

Assim, uma série de pesquisas realizadas ao longo do período eleitoral por uma mesma empresa para um mesmo cliente, utilizando a mesma metodologia, produz resultados bastante objetivos e as eventuais diferenças entre os resultados podem ser diretamente atribuídas à evolução do quadro eleitoral.

Por outro lado, as comparações entre pesquisas realizadas por diferentes empresas, ou mesmo entre pesquisas realizadas pela mesma empresa para clientes distintos, requerem um cuidado maior. Por vezes, as diferentes formas de planejamento e execução de uma pesquisa inviabilizam ou tornam inadequadas as comparações entre elas.

Outros fatores que podem explicar as diferenças entre resultados são os distintos formatos de perguntas, a forma de apresentação dos nomes dos candidatos no questionário, os planejamentos amostrais, o tempo de execução da pesquisa de campo e as formas de abordagem dos entrevistados.

5. A DIVULGAÇÃO DAS PESQUISAS

O direito ao conhecimento das opiniões é básico ao funcionamento democrático. No caso brasileiro, o histórico de restrições e liberações aos prazos de divulgação de pesquisas eleitorais é amplo.

Desde 1965, a regulação das pesquisas está presente na legislação eleitoral brasileira. No Código Eleitoral de 1965, elaborado no início do regime militar, estavam estabelecidos os 15 dias anteriores à eleição como o período de restrição para a divulgação de prévias eleitorais (lei 4737/65, art.255). Esse prazo foi ampliado com a democratização em 1985, quando a legislação estabeleceu a proibição para o período dos 21 dias anteriores à eleição de 1986 (lei 7.508/1986), e para os 30 dias anteriores ao 1º turno da eleição presidencial de 1989 e 10 dias anteriores ao 2º turno (lei 7.773/1989).

Em 1988, com base em recursos apresentados pelos meios de comunicação e de produção de pesquisas, as restrições foram suspensas, e em 1990 foram retiradas da legislação (Resolução 16.402/1990). Foi nesse momento que a legislação eleitoral avançou para o campo da regulação das informações, no sentido de dar transparência tanto sobre os agentes envolvidos no processo político, quanto sobre os parâmetros metodológicos de produção dos dados.

Nas mudanças mais recentes, a Reforma Política parcial realizada nos anos de 2005 e 2006 definiu novas regras para a realização das campanhas eleitorais e de divulgação

de pesquisas, válidas a partir das eleições municipais de 2008. Naquela lei (lei 11.300/06) foi definida a restrição da divulgação para o período dos 15 dias anteriores ao pleito. No entanto, em 8 de novembro de 2007, através da Resolução 22.623 do Tribunal Superior Eleitoral, foi estabelecida uma regulamentação aberta, definindo que pesquisas realizadas em data anterior ao dia das eleições, poderão ser divulgadas a qualquer momento, inclusive no dia das eleições. Assim, a divulgação das pesquisas está legalmente condicionada ao registro das seguintes informações no Tribunal Eleitoral com um prazo mínimo de 5 dias antes do conhecimento dos resultados:

1. o contratante da pesquisa;
2. o valor e origem dos recursos;
3. a metodologia e o período de realização da pesquisa;
4. o plano amostral e ponderação quanto a sexo, idade, grau de instrução e nível econômico do entrevistado; área física de realização do trabalho, intervalo de confiança e margem de erro;
5. o sistema interno de controle e verificação, conferência e fiscalização da coleta de dados e do trabalho de campo;
6. o questionário completo aplicado ou a ser aplicado;
7. o nome de quem pagou pela realização do trabalho;
8. documentos de comprovação do registro da empresa;
9. o nome do estatístico responsável pela pesquisa e seu registro no Conselho Regional de Estatística;
10. número do registro da empresa responsável pela pesquisa no Conselho Regional de Estatística, caso o tenha.

A atual regulamentação brasileira sobre os prazos de restrição da divulgação de pesquisas é uma das mais flexíveis em um conjunto dos países democráticos. As informações do Relatório The Freedom to Publish Opinion Polls Results, elaborado em 2003 pela ESOMAR/WAPOR (SPAGENBERG, 2003), a respeito da existência de restrições para divulgação de resultados de pesquisas eleitorais em um conjunto de 66 países, mostram que para 30 deles vigorava algum tipo de restrição, sendo que a variação dentre os mais rigorosos era de 14 e 30 dias de prazo de restrição para a divulgação de informações.

BIBLIOGRAFIA UTILIZADA

BARBETTA, P. A., Estatística aplicada às Ciências Sociais, Florianópolis: Ed. da UFSC, 1999. BUCHANAN, W., Election Predictions: an empirical assessment, Public Opinion Quarterly, 50, 2, 1986; 222-227; Cavallari, M.; Olsen, O.; Straubhaar, J, O uso de pesquisas eleitorais em decisões de voto, Opinião Pública, vol.1, 2, 1993, pp. 98-119
COMMITTEE ON ANALYSIS OF PRE-ELECTION POLLS AND FORECASTS OF THE SOCIAL SCIENCE RESEARCH COUNCIL, Report on the Analysis of Pre-Election Polls and Forecasts, The Public Opinion Quarterly, Vol. 12, No. 4. Winter, 1948-1949, pp. 599-622. Gontijo, Silvana, A Voz do Povo. O IBOPE no Brasil, Ed. Objetiva. 1996.
HOLLANDER, JOCELYN A., The Social Contexts of Focus Groups, Journal of Contemporary Ethnography 2004; 33; 602
IGO, SARAH E., "A GOLD MINE AND A TOOL FOR DEMOCRACY": GEORGE GALLUP, ELMO ROPER, AND THE BUSINESS OF SCIENTIFIC POLLING, 1935-1955