

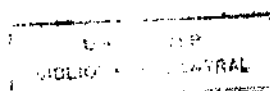
Crítica Metodológica às Pesquisas Eleitorais no Brasil

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida pelo Sr. Cristiano Ferraz e aprovada pela Comissão Julgadora.

Campinas, de de

Prof. Dr. *Favali*
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para a obtenção do Título de MESTRE em ESTATÍSTICA.



UNIDADE 78C
 TUNICAMP
 F413C
 27938
 667/96
 PREÇO R\$ 11,00
 DATA 04/07/96
 N° 00

CM-00089513-8

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Ferraz, Cristiano

F413c Crítica metodológica às pesquisas eleitorais no Brasil / Cristiano
 Ferraz -- Campinas. [S.P. :s.n.], 1996.

Orientador : José Ferreira de Carvalho

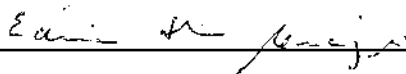
Dissertação (mestrado) - Universidade Estadual de Campinas.
 Instituto de Matemática, Estatística e Ciência da Computação.

1. Pesquisa eleitoral. 2. Opinião pública. 3. Modelos lineares
 (Estatística). 4. Amostragem (Estatística) I. Carvalho, José Ferreira de
 II. Universidade Estadual de Campinas. Instituto de Matemática,
 Estatística e Ciência da Computação. III. Título.

Tese de Mestrado defendida e aprovada em 17 de Maio de 1996
pela Banca Examinadora composta pelos Profs. Drs.



Prof (a). Dr (a). JOSÉ FERREIRA DE CARVALHO



Prof (a). Dr (a). ÉDINA SHIZUE MIAZAKI



Prof (a). Dr (a). MICHEL JEAN MARIE THIOULENT

*Aos meus pais,
Euclides e Margarida*

Agradecimentos

Gostaria de prestar meus agradecimentos a todos, colegas, professores e funcionários, que, de alguma forma, contribuíram para a concretização desta dissertação. Em especial:

- Ao Professor Dr. José Ferreira de Carvalho, que, enquanto orientador, não só contribuiu plenamente para a realização da presente dissertação, mas, fez-se presente em todo o meu processo de crescimento e amadurecimento, como Estatístico, neste mestrado.

- Ao Professor Dr. Luís Roberto Benedetti, sempre muito atencioso e pronto para ouvir e questionar, enquanto sociólogo, minhas colocações.

- Ao Professor Dr. Michel Thiollent, com quem, tendo lido seus livros e suas críticas metodológicas a respeito das Pesquisas de Opinião Pública, no Brasil, tive a oportunidade de conversar a respeito. Meus agradecimentos pela atenção e indicações bibliográficas.

- Ao Professor Dr. Ademir José Petenate, sempre disposto a momentos de discussão sobre o tema.

- Ao Centro de Estudos de Opinião Pública - CESOP - na pessoa da Coordenadora, Profa. Rachael Meneghello.

- À Lucila, minha colega de mestrado, e ao casal Edilberto e Ana Lúcia, do Instituto Nacional de Pesquisas Espaciais - INPE, que muito ajudaram na composição das imagens do trabalho.

- Ao Marcelo do Nascimento, pela parceria na criação e confecção da capa.

- Aos meus pais, Euclides e Margarida, e aos meus irmãos, Andreza, Daniela e Eugênio, que sempre acreditaram em mim. Sem o apoio deles, certamente eu não estaria aqui.

- À Carla Cláudia da Rocha Rêgo Monteiro, com quem partilhei intensamente esses dois anos de minha vida, pelos vários comentários e discussões, pelo companheirismo, compreensão e apoio, em todos os momentos, e pela ajuda na impressão e diagramação do texto.

Resumo

Com a retomada das eleições diretas pluripartidárias no país, as pesquisas eleitorais, aos poucos, foram ganhando destaque no cenário político-social. Hoje, sem dúvida alguma, estão consolidadas como mais um ator no processo de sucessão política, contribuindo, num regime democrático, para a informação do público eleitor.

Paralelamente a essa consolidação, no entanto, crescem as polêmicas e controvérsias em torno do assunto. A possibilidade de manipulação das pesquisas bem como, mesmo sem intenção, de erros de previsão, é preocupante na medida em que afeta o comportamento do eleitor.

Este trabalho estuda, do ponto de vista estatístico, os métodos comumente empregados em pesquisas eleitorais. Crítica severa é levantada contra a amostragem "por quotas", cuja base teórica é insustentável cientificamente. Os erros de levantamentos de intenção de voto são apresentados e discutidos com resultados que não surpreendem os autores deste trabalho mas que destoam do freqüentemente veiculado pela mídia. Por fim, uma alternativa de plano amostral probabilístico para pesquisas de intenção de voto é apresentado e discutido.

Sumário

Introdução:

ESTATÍSTICA E OPINIÃO PÚBLICA, 1

Capítulo 1:

UM POUCO DE HISTÓRIA, 7

Capítulo 2:

ALEATORIZAÇÃO E AMOSTRAS "REPRESENTATIVAS", 13

2.1 Amostragem por Quotas, 13

2.2 Amostras Aleatórias, 16

2.3 A Fundamentação Estatística, 21

2.4 Amostragem por Quotas versus Probabilística, 22

Capítulo 3:

PESQUISAS ELEITORAIS NO BRASIL, 33

Capítulo 4:

REFLEXÕES SOBRE O USO DE
AMOSTRAGEM PROBABILÍSTICA EM PESQUISAS ELEITORAIS, 59

4.1 Uma proposta de plano amostral probabilístico, 62

4.2 Considerações Finais, 71

Apêndice, 73

Referências Bibliográficas, 79

Introdução:

ESTATÍSTICA E OPINIÃO PÚBLICA

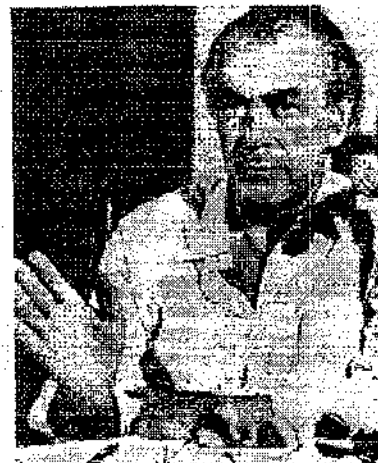
As pesquisas de opinião pública vêm cada vez mais marcando presença no cotidiano das pessoas por meio dos veículos de comunicação de massa. Abordando assuntos os mais variados, surgem como uma sistematização da preferência, gosto ou julgamento da maioria da população com respeito à temática abordada.

Do ponto de vista técnico, as pesquisas de opinião são levantamentos amostrais conduzidos por empresas especializadas, os chamados institutos de pesquisa. Por intermédio delas, tais empresas procuram investigar a opinião pública, captando as opiniões ou preferências individuais através de questionários, formulários e entrevistas - principais instrumentos de coleta de dados utilizados.

Quando o tema é eleição, um tipo particular de pesquisa de opinião se sobressai e ganha grande destaque na mídia: as pesquisas eleitorais. Através delas, o público de um modo geral recebe a notícia de quem tem a preferência do eleitor, numa tentativa de previsão sobre qual candidato será eleito. Além de grande impacto no eleitorado, tais resultados influenciam sobremaneira as estratégias de campanhas dos partidos. O candidato que está à frente nas pesquisas aproveita o fato para incorporar ao marketing de sua campanha. Os outros, ao contrário, buscam argumentos que visem a minimizar os efeitos negativos produzidos pela divulgação do resultado desfavorável. Nas eleições presidenciais de 1989, por exemplo, a primeira eleição direta para presidente no país, depois do regime militar, Fernando Collor de Melo era apontado

como o favorito em todas as pesquisas. Seus adversários contraatacavam com argumentos que iam desde a parcialidade dos institutos de pesquisa até o lançamento de dúvidas sobre as técnicas de levantamento dos dados. Não obstante a popularidade adquirida, as pesquisas de opinião, principalmente as eleitorais, são motivo de sérias controvérsias entre especialistas de diversas áreas, gerando grande debate na sociedade.

Leonel Brizola (PDT): "Utilizada de forma sistemática, a pesquisa é perigosa. É um artifício de violação do processo de formação de consciência do povo"



Ronaldo Caiado (PSD): "É omissa a metodologia das pesquisas, que têm sido feitas em 250 grandes cidades. E o resto do país? E o interior? Não me preocupa estar perto do traço zero"



regulamentação da publicação dos resultados antes da votação; a falta de objetividade em função de interesses políticos ou econômicos e os erros de previsão. Entre as partes envolvidas no debate figuram os próprios candidatos, os partidos políticos, os institutos de pesquisa, jornalistas, profissionais de marketing e propaganda, cientistas

políticos, sociólogos, estatísticos etc...

O levantamento de opinião pública é um assunto de grande importância que

pode ser criticamente estudado sob vários aspectos. Neste sentido, parece haver uma certa polarização no debate, onde destacam-se de um lado o mundo acadêmico e de

Mário Covas (PSDB): "A única pesquisa que



pretendo ganhar é a de 15 de novembro. O número de indecisos é muito grande e o que vale agora é o crescimento de contato com a população. Suspeito até mesmo quando a pesquisa me favorece. Esse instituto (referia-se ao Vox Populi) é a favor de Collor de Mello"

Luis Inácio Lula da Silva (PT): "As pesquisas,



que refletem o pensamento momentâneo, mostram uma manifestação da população em cima de bases falsas. Para mim, pesquisa de verdade é a das ruas e de porta de fábrica. Essa não erra"

outro, o profissional de pesquisa de opinião. Em relação ao aspecto metodológico das ciências sociais, por exemplo, *Pierre Bourdieu* [3], em seu texto "A Opinião Pública não Existe", faz duras críticas às sondagens de opinião, abordando questões conceituais¹. Ainda na área de ciências sociais, *Michel Thiollent* [33] critica técnicas de questionários e entrevistas, considerando o universo da pesquisa de opinião tal como é praticada pelos "homens de *marketing*, *politicólogos* e *sociólogos*". Por outro lado, o objetivo de captar a opinião majoritária numa sociedade, com respeito a qualquer assunto é,

essencialmente, um problema de ordem estatística. Em se tratando de pesquisas

¹Pierre Bourdieu põe em questão três postulados das sondagens de opinião: a) Toda pesquisa de opinião supõe que todo mundo pode ter uma opinião, ou, de outro modo, que a produção de uma opinião está ao alcance de todos; b) É suposto que todas opiniões têm o mesmo peso, a mesma força; c) Por último, no simples fato de fazer a mesma pergunta para todo mundo, acha-se implícita a hipótese de que existe um "acordo" sobre as perguntas que merecem ser feitas.

eleitorais, estas constituem uma das atividades estatísticas mais visíveis ao público. Apesar disso, pouquíssimos trabalhos na área vêm acrescentar ao debate uma visão crítica sobre o assunto. O leitor interessado encontraria apenas o livro de *Jorge de Souza* [29], "Pesquisa Eleitoral:

Críticas e Técnicas", onde, com muita propriedade, o autor expõe, de modo introdutório, alguns problemas concernentes à prática das sondagens de opinião no Brasil. Considerando todo este universo, bem como o impacto sócio-político dos levantamentos de intenção de voto na sociedade, o presente trabalho tem por objetivo contribuir para a discussão sobre o assunto, através de uma crítica metodológica de

Roberto Freire (PCB): "Pesquisa numa sociedade pouco democrática tem o objetivo de induzir, muito mais do que informar"



caráter eminentemente estatístico.

Os argumentos são apresentados nesta tese ao longo de quatro capítulos. No primeiro capítulo, uma retomada histórica de como surgiram as primeiras enquetes eleitorais. Já naquela época, algumas polêmicas foram levantadas e discutidas. No segundo, intitulado "Aleatorização e



Ulysses Guimarães (PMDB): "Não me preocupo com as pesquisas, que são superestimadas, como se decretassem vitórias e derrotas. Candidatos lires como vitoriosos, no fim, sempre perdem. Isso é quase uma regra"

Amostras Representativas", a base das técnicas utilizadas hoje em dia pelos institutos de pesquisa para a composição de suas amostras são apresentadas e

confrontadas com as idéias fundamentais em que se baseia a teoria da amostragem estatística. Ainda neste capítulo, alguns aspectos diferenciais entre o desempenho prático dos métodos de amostragem por quotas e aleatória são apresentados a partir de um experimento realizado pela *London School of Economics*, na década de 50. O terceiro capítulo traz uma análise do desempenho de pesquisas eleitorais conduzidas no Brasil, desde a retomada das eleições diretas pluripartidárias. Para tanto, dados foram colhidos junto à imprensa escrita publicada. Várias são as referências, no texto, a jornais e revistas - as fotos aqui apresentadas foram retiradas da revista *Cadernos do Terceiro Mundo*, nº 124, ano XII, 1989, cuja reportagem de capa era "Pesquisas: A nova arma do poder". Informações também foram levantadas através do CESOP - Centro de Estudos de Opinião Pública, e do arquivo Edgard Leuenroth do IFCH - Instituto de Filosofia e Ciências Humanas da UNICAMP. Por último, o quarto capítulo discute a viabilidade teórica e prática da amostragem probabilística com respeito às pesquisas de intenção de votos. Uma alternativa para o cálculo de precisões, em planos amostrais complexos, é apresentada utilizando-se a teoria de modelos lineares mistos.

Capítulo 1

UM POUCO DE HISTÓRIA

Em 1932, ano de eleições presidenciais nos Estados Unidos, uma idéia original àquela época viria a compor o cenário político-social daquele país. A revista *Literary Digest* decidiu consultar seus leitores a respeito de suas intenções de voto. Ela desejava antecipar o nome do futuro presidente. Vinte milhões de cédulas foram enviadas aos leitores. Retornaram à redação três milhões. Computados os dados, apontava-se *Franklin Roosevelt* como vencedor. Mais tarde, as urnas confirmaram a previsão. A maior *straw poll*¹ já conduzida até então havia dado certo.

O sucesso foi empolgante. A sociedade e os políticos se animaram para, na prática, utilizar o método de amostragem.

Nas eleições de 1936, a *Literary Digest* repete sua pesquisa, desta vez com dez milhões de cédulas e um retorno de dois milhões. O resultado apontava *Alfred Landon* vitorioso com 57% das intenções de voto, contra 43% de *Roosevelt*, que disputava a reeleição. Apuradas as urnas, veio a "ducha de água fria": *Roosevelt* ganhou a eleição com 62.5% dos votos.

A pesquisa falhou duplamente. Errou o nome do vencedor e cometeu um erro de previsão absurdo de 19.5 pontos percentuais.

¹Straw poll era o termo utilizado para designar pesquisas onde qualquer um pode participar respondendo. Não há nenhum tipo de controle sobre quem fará parte da amostra. Nos Estados Unidos, este tipo de pesquisa era padrão até 1933. Associada a esta prática, o número exageradamente grande de cédulas enviadas tem por trás a idéia de que quanto maior a amostra, mais fiel o resultado.

Nesta mesma eleição, no entanto, *George Gallup*, professor de jornalismo da Universidade de Princeton e mais dois pesquisadores, *Archibald Crossley* e *Elmo Roper*, também haviam realizado pesquisas eleitorais. Animados pelo sucesso do método de amostragem em 1932, utilizaram, porém, um plano diferente, chamado **amostragem por quotas**. Nele, as entrevistas eram feitas pessoalmente para eliminar o efeito tendencioso das respostas voluntárias nas *straw polls*. Tomaram uma amostra bem pequena, de apenas dez mil eleitores. *Gallup* previu a reeleição de *Roosevelt* com 55.7% . Errou em 6.8 pontos percentuais, mas, acertou no nome do vencedor !! *Crossley* e *Roper* também acertaram.

O fracasso da *Literary Digest* foi discutido sem muito aprofundamento de questões importantes de caráter estatístico. Em 1936, a teoria da amostragem ainda estava engatinhando.

Com seu sucesso, *Gallup* assumiu a cena eleitoral. Nas eleições de 1940 e 1944, *Roosevelt* foi reeleito, segundo suas previsões.

1948, porém, viria a ser um ano singular para as pesquisas eleitorais nos Estados Unidos. *Gallup* havia previsto a vitória de *Thomas Dewey* na disputa com *Harry Truman*. Ganhou *Truman*, errou *Gallup* e as pesquisas de intenção de voto ficaram desacreditadas !



O jornal *Chicago Daily Tribune* chegou a estampar em primeira página a

manchete: "*Dewey defeats Truman*". É famosa a foto de *Truman* segurando um exemplar daquele jornal, com um sorriso irônico por ter vencido a eleição. Distribuída mundialmente pela *United Press International*, vinha com o provocante título: "*Always a chance for an error !*".

Crossley, *Roper* e *Gallup* haviam estimado a percentagem de votos de *Truman* em 44.8%, 37.1% e 44.5% respectivamente. *Truman* venceu com 49.9% dos votos para presidente. *Dewey* obteve 45.1% e os demais candidatos, juntos, somaram 5%.

Considerando os erros praticados no pleito, apenas o de *Roper* - subestimou número de votos do candidato vencedor em 12.8 pontos percentuais - destoava do desempenho apresentado, nas pesquisas eleitorais de 1936, pelos mesmos pesquisadores. No entanto, as sondagens superestimaram sistematicamente o voto dos candidatos republicanos. *Gallup*, por exemplo, superestimou em 6.8 pontos percentuais a percentagem de votos de *Roosevelt* em 1936 e em 5.4 pontos percentuais o de *Truman*, em 1948. A diferença é que em 1936 este fato não teve muita importância para efeito de previsão sobre quem seria o candidato eleito. Fatalmente, os pesquisadores ignoraram o perigo inerente aos erros cometidos no passado.

Em um levantamento realizado pelo próprio *Gallup* sobre o desempenho das pesquisas de intenção de voto, cerca de 1500 pessoas foram argüídas, no período de 10 a 15 de dezembro de 1948. 75% dos entrevistados afirmaram que houve erro das pesquisas. Para as pessoas que assim opinaram foi feita ainda a seguinte questão:

"Why do you think they (the Polls) were wrong (in their predictions of the last presidential election 1948) ?"

As seguintes respostas foram computadas:

<i>"People changed their minds, people too changeable</i>	<i>12%</i>
<i>Too small sample, should have asked more people</i>	<i>11%</i>
<i>Polls Inaccuracy in interviewing - got wrong people</i>	<i>3%</i>
<i>Did not extend far enough over confidence of republican Party, Committee, etc... ..</i>	<i>9%</i>
<i>Didn't contact all underestimated Labor vote, farm, common people</i>	<i>6%</i>
<i>Criticism of people, didn't tell truth</i>	<i>10%</i>
<i>Missed Last minute Truman Trend, should have</i>	<i>7%</i>
<i>Pooled closer to election general answers: because Truman won, Dewey lost</i>	
<i>They were wrong</i>	<i>15%</i>
<i>Derogatory remarks in regards to all polls-- Wild guessers, Polls owned by GOP, not impartial</i>	
<i>Miscellaneous</i>	<i>4%</i>
<i>Don't know, no answer</i>	<i>19%"</i>

Fonte: Roper Center for Public Opinion Research, University of Connecticut, USA.

Por ordem do governo americano, o *Social Science Research Council* investigou as causas de falha da pesquisa. No relatório final, constatou-se que não foram observados os cuidados recomendados pelos princípios da teoria da amostragem estatística. Basicamente cometeram-se três erros. O primeiro: faltava aleatorização na escolha das unidades amostrais. O segundo: julgando certa a eleição de *Dewey*, *Gallup*

suspendeu as pesquisas quase um mês antes do pleito². O terceiro: fragilidade teórica do plano de amostragem por quotas utilizado.

Hoje as pesquisas eleitorais são muito difundidas em todo mundo. Seus métodos têm sido aperfeiçoados. Paralelamente ao nascimento do *American Institute of Public Opinion*, criado por *Gallup*, em 1935, foi criado o *British Institute of Public Opinion*, em 1936. Na França, em 1938, foi fundado o *IFOP - Institut Français d'Opinion Publique*. No Brasil, o primeiro instituto de pesquisa foi o *IBOPE- Instituto Brasileiro de Opinião Pública e Estatística*, criado em 1942.

Organizaram-se também movimentos de cooperação internacional entre os institutos de pesquisa, dos quais destacam-se a *WAPOR - World Association of Public Opinion Research*, e a *AAPOR - American Association of Public Opinion Research*.

²A última pesquisa foi feita em 20/10/1948 e a eleição ocorreu em 2/11/1948.

Capítulo 2

ALEATORIZAÇÃO E AMOSTRAS "REPRESENTATIVAS"

2.1 Amostragem por Quotas

O plano amostral utilizado por *Gallup*, *Crossley* e *Roper* justificava a possibilidade de extrapolações a partir do fato da amostra constituir uma verdadeira miniatura da população. Assim, seguindo o método, as pessoas entrevistadas devem corresponder proporcionalmente à composição da população de eleitores, segundo características identificáveis como sexo, idade e classe social.

Uma vez determinado o tamanho da amostra, cada entrevistador é orientado para abordar pessoas as quais julgue pertencer à sua quota que pode, por exemplo, ser de mulheres com idade entre 20 e 29 anos, de classe social alta. Repetem várias entrevistas até completar a quantidade de formulários que lhes foi atribuída. Deste modo, não se conhece, então, a probabilidade de um indivíduo da população vir a compor a amostra. Tal evento depende exclusivamente do julgamento do entrevistador. A preocupação do plano amostral era, essencialmente, a de eliminar as tendências existentes nas *straw polls*.

Com o intuito de ilustrar rapidamente a idéia de quotas, suponha uma população de eleitores onde se acredita que as variáveis sexo e classe social influenciam a intenção de voto. Sabendo que tal população é composta 51% de homens, 49% de mulheres, 10% da classe social alta, 30% da média e 60% da baixa,

uma possível amostra de 10 eleitores é a descrita na tabela 1 a seguir.

Computando as informações amostrais, seriam realizadas as seguintes extrapolações: candidato B, 50% das intenções de voto; candidato C, 30% e candidato A, 20%.

Tabela 1

Descrição de intenção de voto numa amostra por quotas

<i>Eleitor</i>	<i>Sexo</i>	<i>Classe Social</i>	<i>Intenção de voto</i>
1	Masculino	Alta	A
2	Masculino	Baixa	B
3	Masculino	Média	C
4	Masculino	Baixa	B
5	Masculino	Baixa	B
6	Feminino	Média	C
7	Feminino	Média	B
8	Feminino	Baixa	C
9	Feminino	Baixa	B
10	Feminino	Baixa	A

Como alternativas de composição de quotas, citam-se dois possíveis esquemas: quotas independentes e interrelacionadas. No primeiro caso, consideram-se apenas os totais marginais das características populacionais, sem nenhuma preocupação com as interrelações entre estas. No segundo caso, consideram-se as representações

proporcionais de todos os subgrupos correspondentes às interações entre as referidas características. As figuras 1 e 2 abaixo ilustram os dois esquemas quanto às variáveis sexo, idade e classe social.

Figura 1
Quotas independentes

<i>Sexo:</i>		<i>Idade:</i>		<i>Classe Social:</i>	
<i>Masc.</i>		20 - 29		<i>Alta</i>	
<i>Fem.</i>		30 - 44		<i>Média</i>	
		45 +		<i>Baixa</i>	

Figura 2
Quotas interrelacionadas

		<i>Classe Social:</i>					
		<i>Alta</i>		<i>Média</i>		<i>Baixa</i>	
	<i>Sexo:</i>	<i>Masc.</i>	<i>Fem.</i>	<i>Masc.</i>	<i>Fem.</i>	<i>Masc.</i>	<i>Fem.</i>
<i>Idades:</i>	20 - 29						
	30 - 44						
	45 +						

É prática comum entre os que habitualmente usam o método, estimar a variabilidade das estimativas por meio das fórmulas concernentes a amostras probabilísticas.

2.2 Amostras Aleatórias

Contrastando com a idéia de quotas, uma amostragem aleatória ou probabilística caracteriza-se essencialmente pelo conhecimento prévio da probabilidade de que cada indivíduo venha a fazer parte da amostra. Para tanto, o estatístico lança mão de processos físicos de aleatorização como, por exemplo, um simples sorteio. A importância de tais processos fica evidente a partir do modelo linear por eles gerado, como será visto nas próximas linhas.

A teoria da aleatorização para amostra aleatória simples (sem reposição) considera que cada observação é selecionada de uma população finita. Suponha Y_1, Y_2, \dots, Y_N de tal forma que cada Y_j corresponde à característica em estudo referente ao indivíduo j na população. Suponha ainda que y_i seja essa mesma característica associada ao indivíduo i observado na amostra. Assim, pode-se definir a seguinte variável aleatória de seleção (Cornfield [7]):

$$\delta_{i,j} = \begin{cases} 1, & \text{se } y_i = Y_j \\ 0, & \text{caso contrário (c.c.)} \end{cases}$$

Em palavras, $\delta_{i,j}$ assume valor um se a unidade amostral (pessoa sorteada) i corresponde ao indivíduo j na população e zero, caso contrário. Deste modo, é possível conhecer probabilidades como as seguintes:

$$\text{Prob}(\delta_{i,j} = 1) = \frac{1}{N};$$

$$\text{Prob}(\delta_{i,j} = 1 \mid \delta_{i,j'} = 1) = 0;$$

$$\text{Prob}(\delta_{i,j} = 1, \delta_{i,j'} = 1) = \frac{1}{N(N-1)}.$$

Ou seja: a probabilidade do indivíduo j da população ser o i -ésimo elemento da amostra é $\frac{1}{N}$;

A probabilidade do indivíduo j da população ser o i -ésimo elemento da amostra dado que outro indivíduo (j') já foi selecionado, como o i -ésimo elemento, é zero.

A probabilidade do indivíduo j da população ser o i -ésimo elemento da amostra e outro indivíduo (j') ser o i' -ésimo elemento selecionado para compor a amostra é $\frac{1}{N(N-1)}$.

Como próximo passo, defina os seguintes parâmetros populacionais de interesse:

$$\mu = \bar{Y} = \frac{1}{N} \sum_{j=1}^N Y_j;$$

$$\sigma^2 = S^2 = \frac{1}{N-1} \sum_{j=1}^N (Y_j - \bar{Y})^2$$

A identidade populacional dada por $Y_j = \bar{Y} + (Y_j - \bar{Y})$ é sempre válida. Cada pessoa selecionada, por sua vez, pode ser representada como função da variável

aleatória de seleção definida anteriormente, da seguinte maneira:

$$y_i = \sum_{j=1}^N \delta_{i,j} Y_j .$$

Substituindo o valor de Y_j nesta última expressão, pela identidade populacional acima, é deduzido o seguinte modelo linear para amostragem aleatória simples sem reposição:

$$y_i = \bar{Y} + \sum_{j=1}^N \delta_{i,j} (Y_j - \bar{Y})$$

$$\therefore y_i = \mu + e_i$$

Ou seja, cada observação corresponde à média (desconhecida) da população mais um erro cuja estrutura é bem conhecida. Constata-se, por exemplo, que este tem as seguintes esperança e variância:

$$E(e_i) = E\left[\sum_{j=1}^N \delta_{i,j} (Y_j - \mu) \right] = \frac{1}{N} \sum_{j=1}^N (Y_j - \mu) = 0$$

$$\begin{aligned} Var(e_i) &= E(e_i^2) = \sum_{j=1}^N \sum_{j'=1}^N (Y_j - \mu) (Y_{j'} - \mu) E(\delta_{i,j} \delta_{i,j'}) = \\ &= \frac{1}{N} \sum_{j=1}^N (Y_j - \mu)^2 = \frac{N-1}{N} \sigma^2 \end{aligned}$$

Além disso, não há independência entre as observações:

$$\begin{aligned}
 Cov(y_i, y_{i'}) &= E(e_i e_{i'}) = \sum_{j=1}^N \sum_{j'=1}^N (Y_j - \mu) (Y_{j'} - \mu) E(\delta_{ij} \delta_{i'j'}) = \\
 &= [N(N-1)]^{-1} \sum_{j \neq j'} (Y_j - \mu) (Y_{j'} - \mu) = \\
 &= [N(N-1)]^{-1} \left(\left[\sum_{j=1}^N (Y_j - \mu) \right]^2 - \sum_{j=1}^N (Y_j - \mu)^2 \right) = \\
 &= -\frac{\sigma^2}{N}
 \end{aligned}$$

De posse do modelo descrito acima, percebe-se que a média amostral é um estimador não tendencioso de μ e tem como variância $\frac{\sigma^2}{n} (1 - \frac{n}{N})$ como é mostrado a seguir.

$$E(\bar{y}) = \frac{1}{n} E\left(\sum_{i=1}^n y_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n (\mu + e_i)\right) = \mu;$$

$$Var(\bar{y}) = E(\bar{y}^2) - [E(\bar{y})]^2 = E(\bar{y}^2) - \mu^2 \quad (1)$$

$$\begin{aligned}
E(\bar{y}^2) &= \frac{1}{n^2} E\left[\left(\sum_{i=1}^n y_i\right)^2\right] = \frac{1}{n^2} E\left[\left(\sum_{i=1}^n (\mu + e_i)\right)^2\right] = \\
&= \frac{1}{n^2} E\left[\left(n\mu + \sum_{i=1}^n e_i\right)^2\right] = \\
&= \frac{1}{n^2} E\left[n^2\mu^2 + 2n\mu \sum_{i=1}^n e_i + \left(\sum_{i=1}^n e_i\right)^2\right] = \\
&= \mu^2 + \frac{1}{n^2} E\left[\left(\sum_{i=1}^n e_i\right)^2\right]
\end{aligned}$$

Agora, substituindo o resultado logo acima na equação (1), tem-se que:

$$\begin{aligned}
Var(\bar{y}) &= \frac{1}{n^2} E\left[\left(\sum_{i=1}^n e_i\right)^2\right] = \frac{1}{n^2} E\left(\sum_{i=1}^n e_i^2 + \sum_{i=1}^n \sum_{i'=1}^n e_i e_{i'}\right) = \\
&= \frac{1}{nN} \sum_{j=1}^N (Y_j - \mu)^2 - (n-1) \frac{\sigma^2}{nN} = \\
&= \frac{(N-1)}{nN} \sigma^2 - \frac{(n-1)}{nN} \sigma^2 = \\
&= \frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)
\end{aligned}$$

Assim, pode-se ver como a aleatorização gera, de fato, uma distribuição que permite a realização de inferência estatística.

2.3 A Fundamentação Estatística

A possibilidade de se ter informações sobre uma população a partir da observação de apenas uma parcela desta é tentadora. Era este o espírito do qual a revista *Literary Digest* estava imbuída quando tentou prever o resultado das eleições presidenciais nos Estados Unidos, na década de trinta. Pelas experiências narradas no primeiro capítulo, já dá para notar, no entanto, que não é qualquer parcela da população que permite obter informações relevantes sobre o todo. Com efeito, ao se provar uma xícara de café sem antes mexer bem o conteúdo, pode-se pensar que este ainda não foi adoçado, quando, em verdade, alguém pode já ter colocado açúcar antes. Se não é qualquer parcela da população que permite a realização de inferências satisfatórias, muito menos é o tamanho da amostra. Lembre-se de que a maior *straw poll* conduzida dispunha de 3 milhões de "votos" e não obteve sucesso.

O estatístico busca, à luz da teoria da amostragem, obter uma amostra onde a observação dá informações sobre parâmetros na população. Este fato muitas vezes é mal interpretado, dando lugar a explicações alusivas ao objetivo de se conseguir uma amostra "representativa"¹. Frequentemente, ao utilizar esta expressão, as pessoas têm em mente a mesma idéia que está por trás da amostragem por quotas, ou seja, a de justificar a possibilidade de extrapolações a partir do fato da amostra constituir uma réplica da população. Tal argumento, de longe, não corresponde à fundamentação da inferência estatística, cuja base é a lógica indutiva. Uma amostra pode até ser representativa, mas, em relação àquelas variáveis populacionais identificáveis, não

¹Cochran, Mosteller e Tukey [5] chamam atenção para a não existência de representatividade em nenhuma amostra particular.

quanto à variável de interesse.

Através da teoria da amostragem, é possível se ter idéia da precisão das estimativas amostrais, justificada pela existência de uma distribuição de probabilidade associada à amostra - como visto na seção 2.2. Assim, não há sentido algum no uso de fórmulas de amostras probabilísticas para estimação da variabilidade de dados provenientes de levantamentos por quotas - prática comentada na seção 2.1, ou de qualquer outro plano amostral de caráter não probabilístico.

Por fim, não é demais ressaltar: o argumento estatístico para a validade de inferências repousa, então, no processo de aleatorização associado à técnica de amostragem.

2.4 Amostragem por Quotas versus Probabilística

Apesar de tudo, a amostragem por quotas ainda é bastante utilizada, em todo mundo, para levantamentos de opinião pública. Este fato pode ser compreendido a partir dos baixos custos deste método bem como da agilidade no processamento de dados, comparado com a amostragem probabilística. Em alguns casos pode haver inclusive a impossibilidade de aleatorização às unidades amostrais. No entanto, é preciso ser cuidadoso ao optar pelo econômico e pelo fator tempo, tendo em vista as conseqüências relativas à precisão das informações obtidas. Neste sentido, com o objetivo de avaliar o desempenho deste método não-probabilístico quanto à qualidade dos dados por ele gerados, a *Division of Research Techniques* da *London School of Economics* realizou em 1952 um estudo experimental publicado por *Moser e Stuart*

[20] sob título *"An Experimental Study of Quota Sampling"*. O trabalho incluiu, antes do experimento propriamente dito, estudos preliminares a partir de dados obtidos por levantamentos passados, os quais apresentavam algumas informações relevantes. Foram comparados os resultados de dois "surveys" realizados na mesma época, verão de 1951, e cobrindo a mesma área: a amostra nacional por quota, realizada pelo *British Market Research Bureau* e o levantamento social do governo britânico, que utilizara amostra aleatória. As respostas conseguidas com os dois métodos apresentavam perfis semelhantes relativos a muitas questões. Porém, no tocante àquelas relacionadas ao status sócio-econômico do entrevistado, o quadro era acentuadamente diferente. Para se ter uma idéia, na pergunta sobre renda, o primeiro fato notável era a diferença entre as percentagens de recusa e "não sei", como resposta, nas duas amostras. Na por quota, atingia 13.5% e na aleatória, 6.6%. Com respeito à distribuição dos dados, havia uma séria divergência, como mostra o Quadro 1 abaixo.

Quadro 1

Distribuição por Renda. Amostra por quota e aleatória. (Percentagens)

<i>Intervalos de Renda</i>	<i>Homens</i>		<i>Mulheres</i>	
	<i>por Quota</i>	<i>Aleatória</i>	<i>por Quota</i>	<i>Aleatória</i>
0 a 3£	8.1	13.6	74.0	76.8
3 a 5£	17.4	11.6	19.6	15.8
5 a 7£	48.5	42.6	4.6	5.2
7 a 10£	15.0	21.3	0.9	1.6
mais que 10£	11.0	10.8	0.8	0.6
<i>Total</i>	100.0	99.9	99.9	100.0

Fonte: Moser e Stuart[20]

Os dados foram comparados com o censo de 1951 e a amostragem aleatória aproximava-se muito dos valores censitários.

Depois desta fase de estudos preliminares, realizou-se o experimento que tinha os seguintes objetivos, dentre outros:

- a) Avaliar a influência de variáveis adicionais de controle em amostras por quota de sexo, idade e classe social;
- b) Avaliar a diferença entre amostragem por quotas interrelacionadas e independentes;
- c) Levantar as diferenças de resultados obtidos com o uso de amostras por quotas e aleatória;

O delineamento experimental seguiu o seguinte esquema, repetido em três cidades: Birmingham, Bristol e Edinburgh.

<i>Organização</i>	1		2		3		4	
<i>Tipo de Quota</i>	(indep)		(interrel)		(interrel)		(indep)	
<i>Nível de Controle</i>	I	IIa	I	IIa	I	IIb	I	IIb
<i>Entrevistador</i>	1	2	1	2	1	2	1	2

As pesquisas foram realizadas pelas seguintes organizações:

- 1 - *B.B.C. Audience Research Department;*
- 2 - *British Market Research Bureau;*
- 3 - *Research Services;*
- 4 - *Market Information Services.*

Quanto aos níveis de controle, I indica o uso de quotas levando-se em consideração apenas as variáveis básicas de controle: sexo, idade e classe social. O nível IIa inclui, além das variáveis básicas, a classificação industrial das ocupações dos indivíduos (Indústria e Manufatura, Transportes e Comunicações, Distribuidores, Administração Pública e Serviços em Geral, Profissional Liberal). O nível IIb inclui, como variável adicional, localização geográfica dentro da cidade.

A partir deste planejamento, o estudo evidencia quatro esquemas de amostragem por quotas:

- (Ia) Quotas independentes;
- (Ib) Quotas interrelacionadas;
- (IIa) Quotas com controle industrial, e
- (IIb) Quotas com controle geográfico.

Em relação ao tópico entrevistador, os algarismos 1 e 2 denotam dois grupos de entrevistadores selecionados do quadro profissional das respectivas empresas.

Paralelamente às amostragens por quotas, foi tomada uma amostra aleatória em cada cidade para que fosse possível realizar comparações.

Vários são os resultados apresentados e comentados no artigo de Moser e Stuart, dentre os quais serão destacados apenas alguns com o intuito de ressaltar

aspectos diferenciais entre os métodos abordados.

Para começar, o seguinte resultado trata de comparações entre os desempenhos da amostragem aleatória, os outros tipos de amostragem por quota realizados e dados do censo de 1951, em relação às variáveis ocupação e educação:

a) Sobre a variável ocupação:

Na página seguinte, reproduz-se a tabela 13, página 365 do artigo supra citado, contendo as distribuições por ocupação para homens e mulheres

Um teste qui-quadrado de adequabilidade de ajuste foi aplicado aos dados, utilizando os valores censitários como valores esperados. Os valores das estatísticas do teste são mostrados no Quadro 2, para homens. Note, porém, que o emprego de tal teste exige que a amostra seja aleatória, o que torna os resultados relativos às amostragens por quotas apenas descritivos.

Quadro 2

Estatísticas χ^2 para o teste de adequabilidade de ajuste, para homens.

Cidades	Amostra Aleatória	Amostras por Quotas			
		Ia	Ib	IIa	IIb
Birmingham	8.46	20.42**	42.73**	10.97	40.17**
Bristol	10.24	33.35**	70.24**	16.22*	11.85
Edinburgh	13.54	11.88	10.33	28.27**	9.35

** Difere significativamente do censo a 1%.

* Difere significativamente do censo a 5%.

Fonte: Moser e Stuart[20]

TABLE 13
Distribution by Occupation, Census and Samples
Males (Aged 20 and Over*) (Percentages)

Census	Birmingham						Bristol						Bournemouth					
	Random Samples		Quota Samples		Random Samples		Random Samples		Quota Samples		Random Samples		Random Samples		Quota Samples		Random Samples	
	(a)	(b)	f _a	f _b	f _{1a}	f _{1b}	(a)	(b)	f _a	f _b	f _{1a}	f _{1b}	(a)	(b)	f _a	f _b	f _{1a}	f _{1b}
Manufacturing	47.2	52.9	56.2	26.0	48.1	23.6	31.6	25.0	28.4	14.9	14.4	37.1	24.4	27.8	25.1	26.9	23.8	28.8
Building, Contracting	1.6	1.7	5.7	3.7	6.4	4.6	1.7	3.1	5.4	10.7	8.4	9.4	7.9	6.8	6.1	7.7	7.1	10.1
Administration, Directors	1.0	1.3	0.6	3.5	5.1	1.3	3.1	3.9	3.5	3.6	0.5	3.1	1.2	2.3	2.0	2.7	1.4	2.7
Transport, Communications	8.9	9.1	5.2	16.6	8.1	15.0	13.3	13.4	14.2	11.7	16.2	8.2	13.2	14.2	8.2	17.1	10.7	11.7
Commerce, Finance	8.2	10.7	10.6	10.6	7.4	13.4	6.9	10.3	11.0	11.7	23.4	7.5	8.5	9.8	14.2	16.7	8.7	10.7
Professional and Technical	4.4	3.7	11.7	3.9	1.2	1.1	1.2	2.2	2.1	4.2	4.2	1.3	4.9	4.4	4.1	4.1	4.7	4.7
Domestic Services	4.1	3.3	2.4	1.7	2.3	2.6	4.4	4.4	4.4	4.4	4.0	4.4	4.9	4.4	4.4	4.4	4.4	4.4
Unemployed	21.1	19.1	1.4	10.7	2.3	7.0	7.0	8.4	8.3	4.9	10.0	5.0	6.7	7.4	6.1	10.2	11.4	11.4
Other, Types	6.5	6.6	8.8	10.7	6.5	8.3	7.4	16.1	10.6	17.0	10.0	12.1	7.4	8.8	10.3	10.2	9.8	13.3
Retired, not gainfully occupied	8.4	6.6	14.9	7.1	14.5	11.5	13.4	16.2	10.7	17.8	12.0	11.3	18.9	13.0	23.5	16.8	13.4	17.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Census	Birmingham						Bristol						Bournemouth					
	Random Samples		Quota Samples		Random Samples		Random Samples		Quota Samples		Random Samples		Random Samples		Quota Samples		Random Samples	
	(a)	(b)	f _a	f _b	f _{1a}	f _{1b}	(a)	(b)	f _a	f _b	f _{1a}	f _{1b}	(a)	(b)	f _a	f _b	f _{1a}	f _{1b}
Manufacturing	14.1	11.3	13.6	6.9	18.8	12.8	7.5	8.1	7.9	1.1	2.6	14.7	2.1	4.9	1.3	1.4	7.7	13.7
Building, Contracting	1.9	1.4	5.7	3.7	6.4	4.6	1.7	3.1	5.4	10.7	8.4	9.4	7.9	6.8	6.1	7.7	7.1	10.1
Administration, Directors	1.0	1.3	0.6	3.5	5.1	1.3	3.1	3.9	3.5	3.6	0.5	3.1	1.2	2.3	2.0	2.7	1.4	2.7
Transport, Communications	8.9	9.1	5.2	16.6	8.1	15.0	13.3	13.4	14.2	11.7	16.2	8.2	13.2	14.2	8.2	17.1	10.7	11.7
Commerce, Finance	8.2	10.7	10.6	10.6	7.4	13.4	6.9	10.3	11.0	11.7	23.4	7.5	8.5	9.8	14.2	16.7	8.7	10.7
Professional and Technical	4.4	3.7	11.7	3.9	1.2	1.1	1.2	2.2	2.1	4.2	4.2	1.3	4.9	4.4	4.1	4.1	4.7	4.7
Domestic Services	4.1	3.3	2.4	1.7	2.3	2.6	4.4	4.4	4.4	4.4	4.0	4.4	4.9	4.4	4.4	4.4	4.4	4.4
Unemployed	21.1	19.1	1.4	10.7	2.3	7.0	7.0	8.4	8.3	4.9	10.0	5.0	6.7	7.4	6.1	10.2	11.4	11.4
Other, Types	6.5	6.6	8.8	10.7	6.5	8.3	7.4	16.1	10.6	17.0	10.0	12.1	7.4	8.8	10.3	10.2	9.8	13.3
Retired, not gainfully occupied	8.4	6.6	14.9	7.1	14.5	11.5	13.4	16.2	10.7	17.8	12.0	11.3	18.9	13.0	23.5	16.8	13.4	17.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

* As the random samples were drawn from the electoral registers, "20 and over" should be interpreted as covering that part of the population included in the registers at the sampling date. Practically none of the random sample respondents were under 21, while 2 per cent. of the quota sample respondents were.

Notavelmente, sete, das doze amostragens por quotas, diferem significativamente dos resultados do censo, enquanto que nenhuma amostra aleatória se distancia relevantemente dos valores censitários. Observando os próprios valores da tabela 13, constata-se que a amostra por quotas tende a subestimar os valores dos trabalhadores em indústrias de manufatura.

Os resultados para as mulheres são mais gritantes, como mostra o Quadro 3, com os valores das estatísticas qui-quadrado.

Apenas uma das amostras aleatórias difere significativamente do censo, enquanto que todas - a exceção de uma - amostras por quota, diferem significativamente. Observando os dados de Bristol, podemos ver que a amostra por quota Ib (interrelacionada) estima em 13% os trabalhadores em comércio, contra 4% do censo. Por outro lado, nesta mesma cidade, a amostra por quota IIb estima as trabalhadoras de manufatura em 2% contra 14% do censo. De um modo geral, mais uma vez nota-se uma tendência da amostragem por quotas em subestimar trabalhadores em manufatura.

Quadro 3

Estatísticas χ^2 para o teste de adequabilidade de ajuste, para mulheres.

<i>Cidades</i>	<i>Amostra Aleatória</i>	<i>Amostras por Quotas</i>			
		Ia	Ib	IIa	IIb
Birmingham	5.40	16.42*	109.01**	8.38	52.75**
Bristol	19.58**	31.82**	87.61**	34.42**	36.24**
Edinburgh	9.57	32.72**	77.08**	38.99**	16.21**

** Difere significativamente do censo a 1%.

* Difere significativamente do censo a 5%.

Fonte: Moser e Stuart[20]

b) Sobre a variável educação:

Para analisar os dados referentes a educação, reproduz-se a tabela 14, página 367 do artigo de Moser e Stuart .

TABLE 14
Distribution by Age at which Full-time Education Ceased. Census and Samples.
(Aged 20 and Over.) (Percentages)

Age at which Full-time Education Ceased	Males								Females							
	Census	Random Samples		Quota Samples				Census	Random Samples		Quota Samples					
		(a)	(b)	Ia	Ib	IIa	IIb		(a)	(b)	Ia	Ib	IIa	IIb		
Below 15	81.1	78.1	76.3	69.2	68.2	74.9	70.6	77.3	76.7	79.0	63.6	66.5	74.0	69.3		
15	6.0	11.3	12.0	10.7	11.5	7.9	9.8	7.9	5.7	4.9	13.9	12.0	11.0	13.4		
16	7.2	4.9	6.5	9.5	10.1	9.8	8.8	7.0	7.2	6.5	10.6	7.8	5.1	8.5		
17-19	3.5	4.7	3.9	6.3	5.3	4.7	6.2	4.2	6.6	5.4	8.7	7.9	6.9	8.0		
20 and over	2.2	1.0	1.3	4.4	4.9	2.7	4.5	3.6	3.9	4.2	3.2	5.9	3.1	.8		
Total	100.0	100.0	100.0	100.1	100.0	100.0	99.9	100.0	100.1	100.0	100.0	100.1	100.1	100.0		

Os dados referentes a educação mostram claramente uma forte tendência nas amostragens por quota em subestimar aqueles que deixaram a escola antes dos 15 anos. É interessante como a variável controle industrial (IIa) age de maneira a inibir um pouco essa tendência. As amostras aleatórias fornecem resultados próximos aos do censo. Por outro lado, as amostras por quotas, que não utilizaram controle industrial, estão afastadas do censo em torno de dois desvios-padrões.

Moser e Stuart concluem que os dados aqui apresentados não rejeitam a hipótese de que a amostra por quota capta indivíduos com melhor nível de educação. Assim, qualquer levantamento, cuja variável educação influencie na resposta, corre risco de ser tendencioso. É necessário, portanto, algum tipo de controle desta variável,

quando se fizer uso de amostragem por quotas.

Outro fator importante diz respeito à variabilidade dos métodos de amostragem aqui comparados. A possibilidade de se ter idéia das variâncias das estimativas levantadas pela amostragem por quotas deve-se ao fato de que o experimento foi replicado. Assim, estas variâncias foram derivadas da soma de quadrados das diferenças entre réplicas com os divisores apropriados relativos ao número de observações envolvidas em cada caso. Quanto à amostra aleatória, utilizaram-se as fórmulas usuais para variância amostral de proporções e médias.

Duas variáveis serão aqui destacadas: percentagens de casas com telefone (Quadro 4) e meio de transporte usado para ir ao trabalho (Quadro 5). Em cada quadro são apresentadas três medidas de erro padrão: para as médias das amostras aleatórias (R); para as médias derivadas de amostras por quotas (Q), e o valor de duas vezes o erro-padrão para a diferença entre proporções ou médias dos dois tipos de amostragem, 2(D). Esta terceira informação permite saber se há diferença significativa, ao nível 5%, entre as médias da amostra aleatória e qualquer outra das quatro diferentes amostras por quotas.

Com respeito ao Quadro 4, estima-se que os verdadeiros valores das percentagens em cada cidade sejam 9%, 16% e 18% em Birmingham, Bristol e Edinburgh, respectivamente. Em Bristol, a amostra aleatória estimou em 16.3% a percentagem de casas com telefone, enquanto que a amostra por quotas, com variável adicional geográfica (IIb), estimou 27%, diferindo significativamente.

Quadro 4
Percentagens de casas com telefone

<i>Cidades</i>	<i>Amostra Aleatória</i>	<i>Amostras por Quotas</i>				<i>Erros-Padrões</i>		
		Ia	Ib	IIa	IIb	R	Q	2(D)
Birmingham	11.2	13.0	17.5	8.0	16.7	1.9	3.1	7.3
Bristol	16.3	19.7	17.2	12.6	27.0	2.2	3.9	8.9
Edinburgh	26.7	22.5	19.1	27.8	21.1	2.8	2.9	8.0

Fonte: Moser e Stuart[20]

Quadro 5
Meio de transporte usado para ir ao trabalho

<i>Transporte Utilizado</i>	<i>Amostra Aleatória</i>	<i>Amostras por Quotas</i>				<i>Erros-padrões</i>		
		Ia	Ib	IIa	IIb	R	Q	2(D)
A pé	25.2	26.9	20.2	27.5	26.1	2.0	2.0	5.8
Bicicleta	10.9	10.5	9.6	11.6	8.7	1.4	1.8	4.6
Moto	1.3	1.4	2.2	4.3	2.1	0.5	0.6	1.6
Carro	8.5	5.2	6.0	6.9	8.8	1.3	1.0	3.3
Público	54.1	56.0	61.9	49.7	54.4	2.4	2.7	7.3

Fonte: Moser e Stuart[20]

Com exceção de Edinburgh, nas duas outras cidades, a amostragem por quotas gerou informações com um erro-padrão maior que o da amostragem aleatória.

Quanto ao meio de transporte utilizado para ir ao trabalho, não se verifica muita diferença na magnitude dos erros-padrões dos dois tipos de amostragem. Nota-se, porém, que a amostra por quota, pelo fato de coletar seus dados nas ruas, estima um

número de pessoas que usam o carro para ir ao trabalho, menor que a amostra aleatória, cujas entrevistas são feitas em domicílio. Aparentemente, isso deixa de ocorrer quando a amostra por quota utiliza a variável geográfica adicional (IIb).

É preciso lembrar, no entanto, que todas essas observações sobre variabilidade só foram possíveis a partir da existência de réplicas no experimento. Este fato, porém, não é prática comum nos levantamentos de amostragem por quotas. Sendo assim, não há base teórica para estimar o erro amostral deste método. Não obstante essa limitação, é freqüente o uso da fórmula de erro-padrão da amostra aleatória como estimativa daquele relativo à amostra por quota. Moser e Stuart alertam para o fato de que esta prática resulta em uma considerável subestimação da real variabilidade da amostragem por quotas, o que vem a corroborar as considerações feitas na seção 2.2.

Finalizando, o trabalho deixa claro, dentre outras coisas, os aspectos de fragilidade nos planos amostrais por quotas, quando postos em prática.

Capítulo 3

PESQUISAS ELEITORAIS NO BRASIL

Com a retomada das eleições diretas pluripartidárias no país, as pesquisas eleitorais, aos poucos, foram ganhando destaque no cenário político-social. Hoje, sem dúvida alguma, estão consolidadas como mais um ator no processo de sucessão política. Tal situação provém de seu uso sistemático ao longo de vários períodos eleitorais. Desde o início, no entanto, as práticas deste tipo de pesquisa de opinião vieram acompanhadas de polêmicas e controvérsias. Temas palpitantes como a manipulação da opinião pública, influência das pesquisas nas decisões de voto e os erros de previsão têm sempre marcado presença na história das pesquisas eleitorais no Brasil.

A respeito da manipulação da opinião pública, podem-se citar vários casos que são verdadeiras tentativas baratas de burlar a opinião pública. Nas últimas eleições de 1994, por exemplo, os dois seguintes anúncios figuravam no jornal Folha de São Paulo, edição de 1º de outubro: O primeiro aponta pesquisa para deputados federal e estadual na cidade de Bragança Paulista destacando-se claramente os

— PUBLICIDADE —

PESQUISA DE OPINIÃO PÚBLICA

PESQUISA REALIZADA PELA EME-VE - PESQUISA S/C LTDA
CIDADE REALIZADA: BRAGANÇA PAULISTA - SP
PERÍODO: 26 A 28 DE SETEMBRO
PESQUISA PARA DEPUTADO FEDERAL E ESTADUAL

FEDERAL

1 - ELMAR KALIL	36,9%
MARQUINHO	15,2%
ADHEMAR DE BARROS FILHO	9,5%
CELSO RUSSOMANO	2,3%
LUIZ GUSHIKEN	1,4%
ROBERTO TUMA	1,1%
ARNALDO FARIA DE SÁ	1,1%
ALMIR AFONSO	1,1%
ROBERTO MACEDO	0,9%
DIOGO NOMEIRA	0,6%
OUTROS	3,0%
NENHUM NÃO SABEM	1,3%

ESTADUAL

1 - EDMAR CHEDID	46,0%
HAER	22,9%
JOÃO CARLOS CARVALHO	6,0%
PAULO MIGUEL	4,9%
PERCIVAL NASCIMENTO	1,4%
AFANAZIO JASADJI	1,4%
WALDIR TOLEDO	0,6%
REDINA	0,6%
MAURICIO SANTOS	0,6%
OUTROS	1,2%
NENHUM NÃO SABEM	13,0%

candidatos supostamente mais citados. O segundo caso, mais acintoso, mostra pesquisa para governo de estado e traz, em negrito, o nome de um dos candidatos¹. Infelizmente, no país, não existe nenhuma legislação que aborde este tipo de prática. Fica claro que, por

Candidato	Porcentagem
BASTOS	40%
Rossi	23%
Dirceu	10%
Medeiros	2%
Outros	7%
Brancos e Nulos	5%
Indecisos	12%

trás do fato, acha-se implícito o investimento num fenômeno conhecido tecnicamente como *Bandwagon*² e que diz respeito a outro tema muito discutido: a influência das pesquisas eleitorais nas decisões de voto do eleitor. Como consequência de tais alusões, já chegou-se inclusive a discutir a proibição da divulgação de pesquisas durante a campanha eleitoral, em 1988. Um registro sucinto e muito bem apresentado deste debate pode ser encontrado em Thiollent[34].

Ao analisar dados de pesquisas eleitorais é preciso se ter claro um critério pelo qual se caracterize o erro de previsão. Três critérios, a princípio, poderiam ser considerados:

- a) O critério de erro do nome do vencedor;
- b) O critério de erro da ordem de colocação dos candidatos;

¹ Foram para o segundo turno, os candidatos Francisco Rossi (PDT) e Mário Covas(PSDB). Este último ganhou a eleição.

²*Bandwagon* é o termo empregado para indicar a tendência, no eleitor, em dar seu voto ao candidato que aparece como favorito nas pesquisas. Em contrapartida dá-se o nome de *Underdog* ao efeito oposto, ou seja, a tendência em se apoiar o candidato mais fraco.

c) O critério de erro por estimativas fora das margens de erro consideradas, segundo certo nível de confiança.

Adotar qualquer um dos dois primeiros critérios para confirmar os acertos de previsões dos institutos de pesquisa é fechar os olhos para a devida apreciação de estudos desta natureza. O primeiro critério, por exemplo, despreza completamente o significado político das demais colocações dos candidatos concorrentes. O segundo, por sua vez, se já dá conta da limitação do critério anterior, não evidencia a gravidade de certos erros. Atente-se, por exemplo, para o caso em que o nome do candidato vencedor é correto, mas, erra-se por completo a ordem dos demais. Que efeito não terá tal fato no comportamento eleitoral dos cidadãos ? O terceiro método sim, é o que consegue suprimir todas as limitações colocadas aqui, uma vez que é coerente com os princípios estatísticos envolvidos no processo.

No Brasil, contudo, os eleitores de um modo geral, parecem contentar-se com o frágil critério de acerto no nome do vencedor, sem nenhum senso crítico das implicações sócio-políticas resultantes de tal ato.

Os casos relativos a erros de previsão dos institutos, cogitações de manipulação da opinião pública e outros fatores mais, levaram a uma crise de credibilidade das pesquisas eleitorais no país, em 1985-1986. O tempo passou, mas, as mesmas polêmicas ainda vêm à tona, principalmente nos períodos de eleição. Só para citar, o Jornal do Brasil do dia 16 de novembro de 1994 trazia a seguinte reportagem : *"Garotinho não admite derrota e ataca pesquisas"*. O então candidato ao governo do Rio de Janeiro, Anthony Garotinho, referia-se às pesquisas de boca-de-urna do IBOPE - 2o turno, que davam vitória a Marcello Alencar.

No Brasil, quatro institutos se destacam pela capacidade de gerar amostras de

cobertura nacional e pela ampla divulgação de seus trabalhos nos meios de comunicação. São eles, o IBOPE, que atua desde 1942; o Gallup, fundado em 1967; o Data-folha, fomentado em 1983 e o Vox Populi, mais novo deles, criado em 1984. Todos estes institutos - é sabido - têm em comum a utilização do método de amostragem por quotas, abordado nos capítulos anteriores, diferindo apenas quanto a alguns detalhes como variáveis de controle e gerenciamento operacional.

Em uma matéria intitulada *"Os números não mentem jamais"*, na qual figuram depoimentos de Carlos Alberto Bragança Pereira, diretor do Instituto de Matemática e Estatística da USP; Renata Nunes César, gerente do Data-folha; Örjan Olsén, ex-diretor do IBOPE e José Severo de Camargo Pereira, professor aposentado da USP e então consultor do Gallup, a revista *Superinteressante* número 09 - setembro de 1994 - descreve rapidamente como os institutos selecionam as pessoas que são abordadas por pesquisas eleitorais:

'Universo

São todos os eleitores do Brasil - cerca de 100 milhões;

Amostra

É uma 'fatia' do universo, definida segundo o número de eleitores em cada região;

Em cada Região

Calcula-se a distribuição de eleitores por estado e daí se tira, proporcionalmente, o número de entrevistas que devem ser feitas em cada um deles;

Em cada estado

As cidades são ordenadas segundo o número de eleitores que possuem. Daí se tira o número de entrevistas que devem ser feitas em cada uma;

Cada cidade

É dividida em zonas, segundo a quantidade de eleitores. As áreas onde serão feitas entrevistas são então sorteadas;

Em cada zona

A pesquisa precisa encontrar um grupo de eleitores representativos daquela área. Se 60% do eleitorado da região é da classe C, será necessário que 60% dos questionários sejam respondidos por pessoas da classe C."

Infelizmente, porém, os métodos de levantamentos de intenção de voto, nem sempre são explicitados quando da divulgação de pesquisas. O que poderia ser um elemento de respaldo teórico, contribuindo para uma cultura crítica em relação a números, índices e estatísticas, dá lugar a investimentos de puro marketing que tentam, por si só, gerar credibilidade aos institutos de pesquisa.

Em propaganda relativa às últimas eleições, o IBOPE ocupou uma página da revista *Veja*, edição de 19 de outubro de 1994, reproduzida na próxima página.

Contrastando com o marketing, em nenhum momento o instituto divulga informações relativas às técnicas de amostragem utilizadas. Isso, porém, não foi uma constante ao longo de sua história. Como pioneiro enquanto instituto de pesquisa no Brasil, o IBOPE teve de abrir caminhos nas décadas de 40 e 50. Àquela época ninguém acreditava muito em pesquisas eleitorais. Não se sabia ao certo de que tratava-se e, assim, uma certa desconfiança sempre rondava os potenciais clientes do instituto. Nas eleições presidenciais de 1950, por exemplo, nenhum dos grandes jornais da época quis comprar as pesquisas do IBOPE. Empenhados em vencer a barreira do desconhecimento, os então diretores produziram o seguinte documento,

AS URNAS CONFIRMAM: SÓ A VERDADE SE APROXIMA DO IBOPE.

CANDIDATOS	DESCRIÇÃO	PESQUISA	BOCA DE URSO	IBR	IBOPE	TSE
	20.09.94 2.000 ENTREVISTAS	01.10.94 3.000 ENTREVISTAS	3.10.94 52.000 ENTREVISTAS	12.10.94 64.782.410 ELEITORES	VOTOS NOMINAIS	VOTOS NOMINAIS
Fernando Henrique Cardoso	41	36	42	19	24	35
Collor	31	20	22	22	27	27
Costa	5	5	6	6	7	7
Guerra	5	4	4	4	5	5
Bressane	4	4	7	3	4	3
Esperdino Amorim	2	1	2	2	2	2
João Carlos	1	1	1	1	1	1
Ubirajara Ayres	0	0	0	0	0	0
Archanjo Mello	12	12	5	18		

Mais uma vez o resultado oficial das apurações confirmou os prognósticos do IBOPE. Mais uma vez a seriedade e a experiência dos seus funcionários e

diretores confirmaram o profissionalismo do IBOPE.

E mais uma vez o IBOPE está confirmado como o instituto de

pesquisa onde a verdade sempre chega na frente.

IBOPE

SÓ A VERDADE
SE APROXIMA DELE.

transcrito do original constante no arquivo do IBOPE, caderno de pesquisas especiais - volume 9, página 226 - à disposição no arquivo Edgard Leuenroth - UNICAMP. O texto preserva o modo como foram grafadas as palavras no documento original.

"COMUNICADO ESTRITAMENTE CONFIDENCIAL SOBRE AS ELEIÇÕES.

As próximas eleições de 3 de outubro despertam interesse fora do comum.

O 'IBOPE', como órgão especializado na pesquisa de opinião pública, não poderia ficar indiferente em face de acontecimento político tão importante, porisso vem, desde há um ano, fazendo repetidas sondagens, em São Paulo e Rio, além de sondagens esporádicas em outras cidades menores.

Nenhum jornal de grande circulação, excepto 'O RADICAL', quis adquirir o nosso material de pesquisa para publicação.

Ora, o 'IBOPE' não tem côr política, nem interesses pela vitoria de qualquer dos candidatos; quer apenas demonstrar que prognosticos políticos podem ser feitos, tambem no Brasil, quase com a mesma exatidão com que 'GALLUP' previu mais de 300 disputas eleitorais nos Estados Unidos.

Este comunicado tem, pois, a função de registrar, para comparação futura, perante o testemunho de nossos clientes e amigos, os prognosticos que os grandes órgãos informativos da opinião pública se recusam a divulgar.

Os resultados aqui divulgados dizem respeito a pesquisas financiadas pelo próprio 'IBOPE' interessado em adquirir mais experiencia na pesquisa política e testar seus proprios métodos.

COMO FORAM FEITAS AS PESQUIZAS

É importante salientar ao leigo que as pesquisas que fazem objeto deste comentário não constituem mero trabalho de gabinete, á base de supostos contingentes

eleitorais partidários, nem 'enquete' de tipos jornalísticos, nem mero 'palpite' de 'entendidos'. São pesquisas científicas de opinião pública em que se entrevista pessoalmente uma amostra estratificada representativa da população. Essa amostra é selecionada de forma que nela estejam representadas em proporções semelhantes, todos os grupos econômicos, raciais, sociais, culturais, políticos e de idades. constituindo, portanto, uma perfeita réplica ou miniatura do todo.

Sabe-se, por leis conhecidas em estatística, que a opinião dessa miniatura é igual a opinião do todo que ela representa, da mesma forma como um litro de água do mar contém a mesma proporção de 'H₂O' e cloreto de sódio contida no oceano onde ela foi colhida.

A OPINIÃO PÚBLICA É UM FENOMENO DINAMICO

Uma das grandes dificuldades da pesquisa de opinião pública está em que esta evolue constantemente, sob influência de diversos fatores determinantes.

Dai não ser possível estudá-la sob um ponto de vista estático. Há necessidade de registrar movimentos, como num filme de cinema, mediante várias fotografias sucessivas, até o momento dela se exteriorizar pelo ato de votar ou de assumir uma atitude definitiva e irretratável. Assim, até o momento das eleições, há sempre a possibilidade de uma variação ocasionada pela propaganda, por boatos, eventos políticos ou sociais de caráter nacional ou internacional, etc...

O Instituto Americano de Opinião Pública teve disso uma prova cabal, quando verificou que entre a última pesquisa e a eleição de Truman, doze dias depois, a opinião do eleitorado americano modificou-se um pouco em favor de Truman.

O 'IBOPE', no Brasil, tem sido mais feliz, pois, embora errando bastante nas porcentagens, tem logrado acertar nas colocações.

O 'IBOPE' prognosticou, no passado, a vitória do General Dutra em São Paulo; a vitória de Eduardo Gomes no Rio; a vitória de Adhemar e de Noveli, em São Paulo.

Estas últimas com erro insignificante.

AS PRÓXIMAS ELEIÇÕES

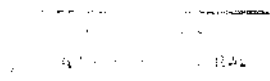
Para o próximo pleito não fizemos pesquisa nacional abrangendo todo o Brasil, pois, custaria preço incompatível com as nossas forças no momento.

Fizemos, no entanto, algumas sondagens cujos resultados apresentamos em caráter confidencial, aum circulo restrito de clientes e amigos do 'IBOPE'.

NO DISTRITO FEDERAL

(Note-se: sómente no Distrito Federal)

Pesquisa terminada em 20 de setembro -1950



P.- EM QUEM O SNR. VAI VOTAR, PARA PRESIDENTE DA REPÚBLICA ?

(...)

Segundo os sexos

	Homens	Mulheres	Total
Getulio Vargas	49,5%	47,3%	48,8%
Eduardo Gomes	28,6	31,3	29,6
Christiano Machado	13,6	11,7	12,8
João Mangabeira	2,8	0,8	2,0
Votarão em branco	1,7	2,6	2,0
Não votarão	0,6	2,6	1,4
Indecisos	3,3	3,7	3,4

..."

No referido documento, em relação ao modo como são feitas as pesquisas, pode-se notar, em primeiro lugar, que o método utilizado pelo instituto é o mesmo que Gallup utilizou na eleição presidencial de 1948, nos Estados Unidos. Em segundo lugar, invoca-se explicitamente a noção de representatividade de amostra como justificativa científica para a realização de inferência estatística - argumento já refutado no capítulo anterior. Por último, pelo menos, percebe-se a coerência das informações passadas, uma vez que não existe lei estatística que diga que a opinião das pessoas que compõem uma amostra selecionada por quotas é igual a opinião do todo. Do mesmo modo, em um litro de água do mar, a proporção de cloreto de sódio e H_2O encontrada

não pode ser a mesma do oceano de onde a amostra de água foi retirada, uma vez que tal proporção varia de local para local, num mesmo oceano.

Ainda com respeito ao documento do IBOPE, o comentário sobre o erro do Instituto Americano de Opinião Pública, mais conhecido como *Gallup Institute*, mostra o desconhecimento do conteúdo do relatório feito pelo *Social Science Research Council* a respeito do fato. Fica claro ainda que o instituto admite errar nas porcentagens de intenções de votos dos candidatos, contentando-se com o simples acerto das colocações dos mesmos³.

Apesar de tudo, quanto à eficácia dos métodos usados hoje em dia pelos institutos em geral, não é difícil encontrar entrevistas como a de Carlos Augusto Montenegro, atual diretor do IBOPE, à revista *Veja*, publicada na edição de 05 de setembro de 1990 e reproduzida abaixo:

"Veja - Em que ocasiões pode-se dizer que o IBOPE errou ?

Montenegro - Nos vinte anos em que estou no instituto, cometemos apenas um erro, que não foi técnico, mas tático. Nas eleições para prefeito em 1985, imaginamos - e nunca se deve imaginar em pesquisa - que a eleição em Fortaleza estava definida e paramos de pesquisar ali, na última semana antes da votação, concentrando nosso esforço no Recife, onde a disputa era mais acirrada. Acertamos no Recife, mas erramos em Fortaleza, onde Maria Luiza Fontenelle virou o resultado na última semana. Ou seja, demos como certo um resultado que pôde ser modificado. Erramos, sim, e assumimos isso como lição.

Nunca mais deixamos de ir até a última semana - foi assim que pegamos a virada de Luiza Erundina em 1988. "

³ Em relação ao critério de acerto de previsão, infelizmente não foi possível conseguir dados da eleição presidencial de 1950, relativos apenas ao Rio de Janeiro, então capital do país, para confrontar com os números do instituto.

Ou ainda, as seguintes explicações dadas pelo sociólogo Marcos Coimbra, diretor do Vox Populi e Antônio Manuel Teixeira Mendes, diretor executivo do Data-folha, em reportagem da revista Veja de 28 de setembro de 1994, sob título *"Susto na campanha baiana. Pesquisas conflitantes causam alvoroço"*. O assunto era a diferença entre os números do Vox Populi e Data-folha, na disputa pelo governo da Bahia.

"(...)

Coimbra admitiu o erro e anunciou que fará um novo levantamento para seu cliente, o Jornal do Brasil. 'Nosso resultado foi escalarfobético', reconhece o sociólogo. 'Instituto de pesquisa também pode errar'. Coimbra explica que, mesmo sendo feito com todo cuidado, um levantamento pode chegar a números errados em função de um somatório de pequenas falhas capazes de produzir um engano total. 'A cada 100 pesquisas, existe a chance de cinco estarem erradas', diz.

Não é a quantidade de entrevistas que determina a precisão dos índices. O que interessa é a chamada 'amostra'. 'É muito melhor você fazer cinquenta entrevistas em cinquenta cidades do que 100 entrevistas em vinte cidades', atesta Antônio Manuel Teixeira Mendes, diretor-geral do Data-folha. 'Quanto mais espalhada a amostra, melhor'.

(...)"

Na primeira entrevista apresentada, o diretor do IBOPE faz declarações contraditórias aos próprios números do instituto, como será mostrado a seguir.

Em relação a segunda reportagem, será mostrado também que, apesar da menção de Marcos Coimbra de que, em cem entrevistas há a chance de cinco estarem erradas, esse não parece ser o caso daquele ou de qualquer outro instituto aqui citado. Sobre a colocação do diretor do Data-folha, fica clara a falta de maior embasamento teórico à idéia de precisão por ele colocada.

Esses depoimentos, porém, juntamente com todo o trabalho de publicidade dos institutos, ajudam a fixar o mito - às vistas do eleitor leigo em estatística - de que as pesquisas eleitorais sempre acertam. Quando isso "eventualmente" não acontece, em

quase todos os casos, "sabe-se" exatamente o que se passou: o público mudou de opinião nos últimos momentos, antes da eleição.

A falta de transparência sobre como foram gerados os dados , já comentada brevemente em linhas anteriores, é digna de nota. Com efeito, recorrendo a matérias publicadas em revistas e jornais durante as últimas eleições, de 1994, apenas o Gallup e Data-folha apresentaram alguma preocupação em dar informações gerais sobre o método utilizado por eles, na confecção de pesquisas eleitorais. Para ilustrar, reproduzimos aqui, por exemplo, notas publicadas no jornal O Estado de São Paulo, dia 30 de março de 1994 e Jornal Folha de São Paulo, de 18 de setembro de 1994 , por ocasião de levantamentos de intenção de voto para presidente, realizados pelo Gallup e Data-folha respectivamente:

"Instituto realizou 3 mil entrevistas.

O Instituto Gallup entrevistou 3 mil eleitores entre domingo e quarta-feira, em 163 cidades de 18 Estados. As entrevistas foram feitas pessoalmente, na casa dos eleitores. A amostra foi definida por sorteios sucessivos que apontam a cidade, o bairro, o quarteirão, a casa e a pessoa a ser ouvida. As capitais são automaticamente incluídas."

"Veja o método da Pesquisa

Da redação

A pesquisa do Data-folha é um levantamento por amostragem estratificada, com sorteio aleatório dos entrevistados. O conjunto da população adulta do país é dividido em quatro sub-universos, que representam as regiões Sul, Sudeste, Nordeste e Norte/Centro-Oeste.

Em cada sub-universo os municípios são agrupados segundo localização geográfica e nível socioeconômico. Em cada grupo são sorteados municípios. Por sorteios sucessivos, chega-se ao bairro, à rua e ao indivíduo.

Entre 13 e 15 de setembro, foram entrevistados 21.074 eleitores em 657 municípios. A margem de erro é de 2 pontos percentuais para mais ou para menos.

A direção do Data-folha é exercida pelos sociólogos Antônio Manuel Teixeira Mendes e Gustavo Venturi."

A julgar pelo conteúdo das informações contidas nestas notas, parece tratar-se do uso de amostragem probabilística, pelos referidos institutos. Um fato importante a acrescentar é que, geralmente, as pesquisas declaram uma margem de erro de 3%, à exceção do Data-folha, que, às vezes, considera 2%.

Ora, de posse desses dados e novamente recorrendo a registros de pesquisas eleitorais publicados ao longo do tempo, percebe-se um quadro muito diferente daquele divulgado em marketing. Dentre os vários levantamentos apreciados, serão destacados aqui apenas alguns casos ilustrativos.

Eleições para prefeito, 1985: em São Paulo, os institutos IBOPE e Gallup apontavam Fernando Henrique Cardoso como candidato eleito, com 34.5% e 36% dos votos, respectivamente. Apuradas as urnas, Jânio Quadros venceu com 37.5%. FHC teve 34.2%. Dois erros cometidos: erraram o nome do vencedor e cometeram erro de previsão de 5.8 e 8.5 pontos percentuais, respectivos a cada instituto, na percentagem das intenções de voto em favor de Jânio Quadros. Nesta mesma eleição, os erros não pararam por aí. Basta analisar o Quadro 6, parcialmente reproduzido do artigo do sociólogo *Marcus F. Figueiredo* [8], *"As pesquisas Pré-eleitorais e as 'surpresas' políticas"*.

O IBOPE e Gallup erraram ainda absurdamente em Fortaleza e Goiânia como pode-se constatar. Surpresa também aconteceu com o sucesso do Partido dos Trabalhadores em Vitória, Espírito Santo, embora não sejam apresentados os números. Desta forma, não é verificada, na prática, a declaração de Carlos Augusto Montenegro, apresentada antes, na qual dizia que o único erro do IBOPE, durante os vinte anos em que esteve no instituto, foi nas eleições para prefeito, 1985, em Fortaleza, e por questão de tática.

Por sua vez, o Data-folha, apesar de ter números dentro da margem de erro de 3%, em sua pesquisa de 14/11 para prefeitura de São Paulo, na anterior, datada de 10/11, já apresentava margens bem maiores.

Quadro 6
Prévias Eleitorais e Resultados em Três Capitais (%)*

Cidades	Candidato	IBOPE		Gallup		Data-folha		TRE
		27/10	14/11	10/11	13/11	10/11	14/11	
São Paulo								
	Jânio	31.6	31.7	31.0	29.0	33.0	37.0	37.5
	FHC	35.2	34.1	32.0	36.0	35.0	35.0	34.2
	E.Suplicy	18.6	21.4	21.0	19.0	18.0	18.0	19.8
	Indecisos	8.4	8.5	7.0	2.0	-	-	-
Fortaleza		27/10	14/11	20/10	03/11			
	Andrade	54.3	39.5	50.0	47.0	-	-	29.7
	Alcântara	22.3	23.1	21.0	17.0	-	-	24.3
	Fontenelle	9.0	20.3	10.0	14.0	-	-	32.0
	Indecisos	-	11.4	8.0	9.0	-	-	-
Goiânia		27/10	14/11	27/10	03/11			
	D.Antunes	60.3	-	57.0	-	-	-	45.1
	D.Accorsi	11.2	-	12.0	-	-	-	40.2
	Indecisos	16.5	-	10.0	-	-	-	-

* As datas nas colunas são os dias de fechamento de campo.

Fonte: Figueiredo[8]

Apesar de todas essas constatações, o sociólogo Figueiredo insistiu em analisar os fatos sob a óptica do despreparo da sociedade para a nova fase da vida política na qual o Brasil estava se inserindo. Era a segunda vez que se davam eleições diretas pluripartidárias para cargo majoritário, no país, depois da ditadura militar. A primeira foi em 1982, para governo de estado. Ele disse:

"Mas o que houve de tão extraordinário nas últimas eleições ? Comparando-se o resultado das urnas com os prognósticos das pesquisas, parece, à primeira vista, que fomos apanhados de 'surpresa' com as vitórias de Jânio Quadros em São Paulo, de Maria Fontenelle em Fortaleza e com o sucesso do PT em Goiânia e Vitória. Por que estes fatos tornaram-se uma 'surpresa' para todos ? Que houve afinal ? (...) Erros técnicos dos institutos ou despreparo de todos nós para o fato novo chamado eleições municipais ? "

Em 1990, Jorge de Souza [29], em seu livro *"Pesquisa Eleitoral: Críticas e Técnicas"* analisou dados de pesquisas eleitorais referentes às eleições de 1986, para governador. Esses mesmos dados, provenientes de publicação no Jornal Folha de São Paulo, edição de 15 de janeiro de 1987, são apresentados aqui, no Quadro 7, para análise. A manchete era a seguinte: *"Mais uma vez as urnas deram IBOPE. Confirme."*

Quadro 7

Prognósticos IBOPE versus resultados oficiais. Eleição para governador, 1986.
(%)

	TRE	IBOPE		TRE	IBOPE
AC			PA		
F. Melo	55	59	H. Gueiros	55	63
M. Maia	32	31	J. Menezes	08	12
H. Pimenta	02	02	C. Levy	12	08
AL			PB		
F. Collor	46	50	T. Burity	56	48
G. Palmeira	38	38	M. Gadelha	34	44
R. Lessa	03	05	C. Bezerra	01	02
AM			PR		
A. Mendes	49	50	A. Dias	60	68
A. Virgílio	38	40	A. Furtado	20	19
M. Barros	03	03			
BA			PE		
W. Pires	62	61	M. Arraes	55	52
J. Marinho	28	29	J. Múcio	35	41
D. Gama	01	01			
CE			PI		
T. Jereissate	55	58	A. Silva	44	43
A. Bezerra	31	33	F. Neto	43	43
H. Coelho	03	03	N. Fonteles	03	01
ES			RJ		
M. Mauro	50	50	M. Franco	46	42
E. Alvares	31	35	D. Ribeiro	34	38
A. Vilachi	09	07	F. Gabeira	08	08

Fonte: Souza[29]

Quadro 7 (continuação)

*Prognósticos IBOPE versus resultados oficiais. Eleição para governador, 1986.
(%)*

	TRE	IBOPE		TRE	IBOPE
GO			RN		
H.Santillo	51	50	J.Faustino	45	48
M.Borges	30	33	G.Melo	47	47
D.Accorsi	07	07	A.Tinoco	01	01
MA			RS		
E.Cafeteira	70	77	P.Simon	43	47
J.Castelo	15	17	A.Pinto	24	25
M.Delta	02	02	C.Chiarelli	11	10
MT			RO		
C.Bezerra	56	52	J.Santana	40	48
F.Campos	25	29	J.Atalla	18	15
G.Barros	02	03	O.Souares	15	15
MS			SC		
M.Miranda	55	59	P.Ivo	44	51
L.Coelho	32	29	V.K.	26	25
L.Landes	02	01	Cazamiga	14	12
MG			SP		
N.Cardoso	40	36	O.Quércia	37	38
I.Franco	36	39	A.Ermirio	24	25
M.Badaró	05	05	P.Maluf	18	21
			E.Suplicy	10	09
SE					
Valadares	50	45			
Teixeira	41	37			
Magno	03	05			

Fonte: Souza[29]

Ao todo foram vinte e três pesquisas referentes a vinte e três estados. De pronto, observa-se que em Minas Gerais e Rio Grande do Norte, o IBOPE errou o nome do candidato vencedor. No Pará, dois erros graves: inverteu a ordem dos candidatos derrotados e superestimou em 8 pontos percentuais a percentagem de votos do candidato eleito.

Das vinte e três pesquisas realizadas, em dezesseis delas, há, pelo menos, uma estimativa fora da margem de erro de 3%, usualmente admitida pelos próprios institutos. Em outras palavras, o IBOPE errou em mais da metade de seus levantamentos. Para se ter idéia do absurdo, basta lembrar que se fosse utilizada uma amostra probabilística com nível de confiança 95% e a mesma margem de erro considerada, o esperado seria que, em apenas uma das pesquisas, houvesse estimativas distantes mais de três pontos percentuais para mais, ou para menos, do resultado oficial.

Para terminar a ilustração, nas últimas eleições de 1994, para governador, no segundo turno, dias antes do pleito, foram coletados os seguintes dados apresentados no Quadro 8, segundo as fontes: Jornal O Globo, edição de 15/11; Jornal Folha de São Paulo, edição 15/11; Jornal do Brasil, edição de 14/11; referindo-se respectivamente às pesquisas dos institutos IBOPE, Data-folha e Vox Populi. Os números foram convertidos em percentagem sobre os votos válidos para comparação com resultados oficiais do TSE.

Como se pode observar, o Vox Populi errou duplamente em Brasília-DF e Maranhão. Tipo do erro que contradiz qualquer cogitação de uso de amostragem probabilística: errou o nome do candidato vencedor e praticou margens de erro acima de 3 pontos percentuais. Note-se que tais erros significam inversão completa do posicionamento dos candidatos em pesquisa divulgada um dia antes do pleito ! Errou

ainda nos estados de Roraima, Santa Catarina e São Paulo.

Quadro 8

*Prognósticos de Institutos x Resultados Oficiais. Governo, 2o. turno, 1994**

Estados	TSE	IBOPE	Data-folha	Vox Populi
SÃO PAULO				
Covas (PSDB)	56.1	59.7	60.5	58.0
Rossi (PDT)	43.9	40.3	39.5	42.0
MINAS GERAIS				
Azeredo (PSDB)	58.7	54.7	58.8	56.3
Costa (PP)	41.3	45.3	41.2	43.7
BRASÍLIA-DF				
Cristóvam (PT)	53.9	52.8	53.0	46.6
Campelo (PTB)	46.1	47.2	47.0	53.4
BAHIA				
Souto (PFL)	58.6	64.3	63.9	59.0
Durval (PMN)	41.4	35.7	36.1	41.0
S. CATARINA				
Paulo (PMDB)	50.8	50.0	53.9	53.8**
Angela (PPR)	49.2	50.0	46.1	46.2**
MARANHÃO				
Samery (PFL)	50.6	51.0	-	45.8
Cafeteira (PPR)	49.4	49.0	-	54.2
RORAIMA				
Campus (PTB)	58.5	-	-	64.7
Cruz (PSDB)	41.5	-	-	35.3

* Números em percentagem sobre os votos válidos.

** Dados do Jornal do Brasil, de 12/11/94.

Fonte: Jornais O Globo, Folha de São Paulo e Jornal do Brasil.

Por sua vez, o Data-folha teve estimativas fora de controle estatístico em São Paulo, Bahia e Santa Catarina. O IBOPE não ficou atrás, teve margem de erro extrapolada em São Paulo, Bahia e Minas Gerais.

Para quem pensa que o vexame foi apenas no segundo turno, reproduz-se abaixo nota da revista *Veja*, de 19 de outubro de 1994.

"Os erros das pesquisas

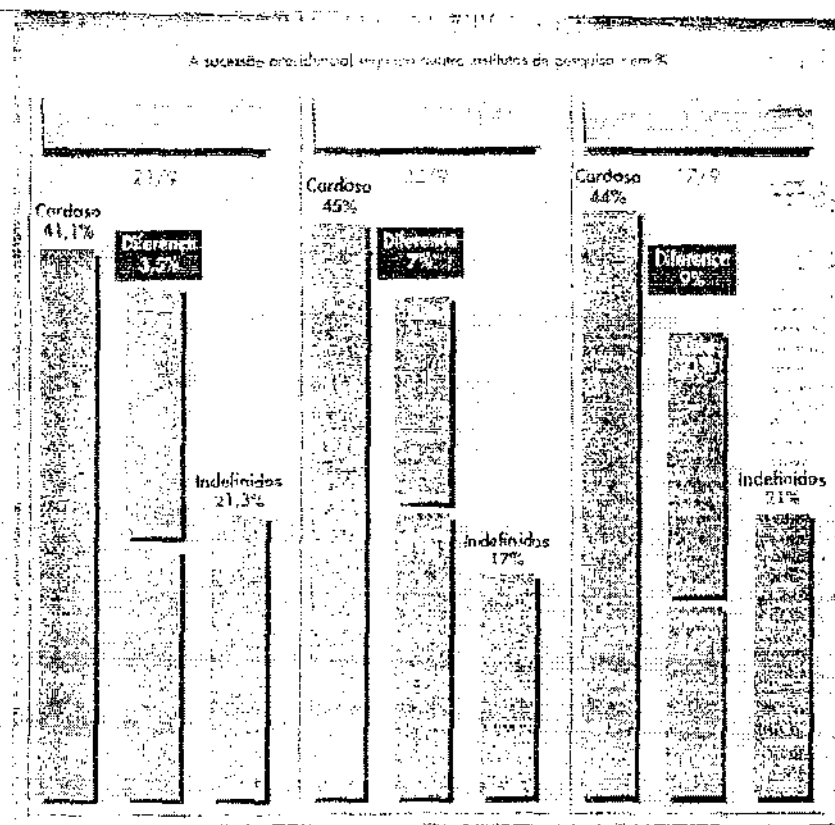
Se os institutos de pesquisa se saíram bem na disputa presidencial, fizeram um papelão em muitos Estados. Em São Paulo, a dois dias das eleições, o Gallup divulgou que Mário Covas ganharia no primeiro turno com 10 pontos de vantagem. O Vox Populi previu vitória do tucano por 8 pontos. Excluindo-se brancos e nulos, Covas teve 47% e os adversários, 53%. 'A leitura correta da pesquisa é que ele estava em queda e poderia haver segundo turno', admite o diretor do Vox Populi, Marcos Coimbra. Em Minas Gerais, todos os institutos previram a vitória de Hélio Costa. Deu segundo turno. No Rio Grande do Sul, segundo as pesquisas, Antônio Brito teria até 46% do total de votos. Por esse critério teve só 40%.

A desculpa dos pesquisadores agora é a 'leitura incorreta'. Às vésperas do pleito, Carlos Montenegro, dono do IBOPE, deu entrevista à Rádio Gaúcha, anunciando que Britto estava eleito. 'Os números estavam certos, a minha interpretação é que não', diz atualmente. Em Sergipe, houve outro tipo de problema. O senador Albano Franco, oligarca de uma família dona de TV, jornais e rádios, fez questão de divulgar até o dia da eleição uma pesquisa velha, feita dez dias antes. Dava vitória de Albano por uma margem de 18 pontos. Foi um vexame. Seu adversário Jackson Barreto, chegou em primeiro lugar para o segundo turno. No Maranhão e na Bahia, o fiasco também foi de divulgação. Os institutos previram, através de votação simulada, que muitos eleitores iriam errar na hora de votar. O IBOPE e o Data-folha fizeram essa simulação e constataram que, na urna, Roseana Sarney caía de 51% para 40%. Na Bahia, Paulo Souto caía de 43% para 34%. O mesmo aconteceu no Ceará, onde Tasso Jereissate caía de 55% para 43%. Mas nem os adversários se preocuparam em divulgar o resultado da simulação."

A respeito do comentário logo no início da reportagem, sobre os institutos terem se saído bem na disputa presidencial, o caso é que Fernando Henrique Cardoso estava tão distanciado dos demais candidatos que seria muito difícil qualquer pesquisa errar. Não obstante o fato, um detalhe não poderia escapar ao olho de um estatístico atento que tivesse tido contato com a manchete do jornal O Estado de São Paulo, dia 26 de setembro de 1994, apresentada na página seguinte.

A desculpa de variação nas amostras não existe. A observação é justamente oposta. Se os próprios institutos admitem margens de erro e níveis de confiança, a diferença entre as pesquisas deveriam corresponder a flutuações normais entre as amostras, respeitadas as margens de erro. Em outras palavras, as amostras deveriam ser responsáveis por diferenças de, no máximo três pontos percentuais para mais ou para menos sobre os parâmetros estimados. Aqui, o que se vê são diferenças absurdas entre o percentual de indecisos (indefinidos), o grupo caracterizado como "outros" e, conseqüentemente, diferença de até 9 pontos percentuais entre os votos do primeiro colocado e a soma dos demais nas pesquisas do Gallup e IBOPE.

Em resumo, o uso da amostragem por quotas em pesquisa de intenção de voto, é de precariedade considerável em relação ao rigor estatístico necessário a estudos desta natureza e é preocupante, na medida em que não traz consigo a devida transparência de como foram levantados os dados. As distorções são possíveis a partir do próprio método de amostragem. Para complicar, na prática, o problema não para por aí. Uma vez produzidos os números, entra em cena o poder da mídia através das interpretações e do modo como são divulgadas as informações.



Diferenças na última rodada de pesquisas eleitorais causa polêmica

Gallup é o único que aponta possibilidade de disputa presidencial no para o segundo turno

Na última rodada de pesquisas eleitorais, a do Instituto Gallup, liderada pelo Colômbio no quinto dia, a única que indica a possibilidade de disputa entre Fernando Henrique Cardoso (PSDB) e Luiz Inácio Lula da Silva (PT) no para o segundo turno. O Gallup apontou uma diferença de 1,5 pontos entre os dois candidatos.

Enquanto isso, os outros institutos de pesquisas apontam uma diferença de 3 pontos. Para o Vox Populi, o instituto que apoia Cardoso, a diferença, segundo pesquisa publicada sexta-

feira, é de 3 pontos. Segundo o Vox Populi, a diferença é de 3 pontos. Para o Vox Populi, o instituto que apoia Cardoso, a diferença, segundo pesquisa publicada sexta-

feira, é de 3 pontos. Segundo o Vox Populi, a diferença é de 3 pontos. Para o Vox Populi, o instituto que apoia Cardoso, a diferença, segundo pesquisa publicada sexta-

feira, é de 3 pontos. Segundo o Vox Populi, a diferença é de 3 pontos. Para o Vox Populi, o instituto que apoia Cardoso, a diferença, segundo pesquisa publicada sexta-

feira, é de 3 pontos. Segundo o Vox Populi, a diferença é de 3 pontos. Para o Vox Populi, o instituto que apoia Cardoso, a diferença, segundo pesquisa publicada sexta-

O Gallup exclui nove estudos de opinião e isso pode proporcionar um aumento no resultado. Segundo o Colômbio, a diferença é de 3 pontos. Para o Vox Populi, o instituto que apoia Cardoso, a diferença, segundo pesquisa publicada sexta-

A metodologia das pesquisas varia, além da diferença, o Colômbio faz entrevistas domiciliares e o Vox Populi o faz em locais públicos. Isso pode explicar as diferenças entre os dois institutos.

VARIACÕES NA AMOSTRA PODEM SER EXPLICAÇÃO

Enquanto isso, os outros institutos de pesquisas apontam uma diferença de 3 pontos. Para o Vox Populi, o instituto que apoia Cardoso, a diferença, segundo pesquisa publicada sexta-

Em meio a esta realidade e considerando a relevância do processo de eleição para a efetivação de uma democracia, não há como justificar o não uso de métodos científicos em levantamentos eleitorais. Cabe, então, aos estatísticos a crítica e denúncia no que diz respeito aos métodos hoje empregados pelos institutos na confecção das pesquisas eleitorais no Brasil.

Capítulo 4

REFLEXÕES SOBRE O USO DE AMOSTRAGEM PROBABILÍSTICA EM PESQUISAS ELEITORAIS

Se, por um lado, os fatos e argumentos apontam para as grandes limitações do uso de amostras por quotas em captar intenções de voto, por outro, os institutos de pesquisa se defendem invertendo a problemática, questionando que alternativas existiriam, neste caso. A utilização de métodos probabilísticos é logo descartada considerando-se as seguintes explicações:

1. *Alto custo em termos de finanças e tempo* - Uma amostra por quotas chegaria a custar até 60% do orçamento elaborado para um plano probabilístico. Além disso, a demora na captação e análise dos dados provenientes de uma amostra aleatória é incompatível com o mercado. Com o uso de quotas é possível encerrar todo o trabalho em um ou dois dias, se necessário.

2. *Dificuldade algébrica para estimar precisão dos dados, devida à complexidade dos planos amostrais* - Os planos atualmente postos em prática constam de vários estágios, antes de chegar à opinião do eleitor, como visto antes. Isto dificultaria a obtenção da precisão dos estimadores¹.

Junte-se ao dito acima, o fato de que amostragem por quotas é utilizada em

¹Mesmo quando usam fórmulas concernentes a amostras probabilísticas para estimar variabilidade numa amostragem por quotas, estas são sempre simplificadas. Elas não são consistentes com o plano executado, no caso deste ser, de fato, aleatório.

todo mundo e, tem-se então, as bases em que se apoia a prática das pesquisas de opinião pública no país.

De fato, amostragem probabilística não é isenta de limitações. Elas existem, sim, a começar pela própria natureza das variáveis em estudo: opinião pública, intenção de voto, às quais qualquer método amostral estaria sujeito - inclusive por quotas. Porém, ela possui a importante característica de permitir mensurações a respeito da variabilidade inerente à informação levantada, pelo fato desta não ter sido originada do todo e sim de uma parte deste. Em termos técnicos, as amostragens probabilísticas são as únicas que controlam erros estatísticos amostrais.

Para captar comportamentos e atitudes, faz-se necessário uma interdisciplinariedade de métodos que permitam o controle das diversas fontes de variação que influenciam o fenômeno estudado. A amostragem por quotas, por sua vez, não controla a fonte de variação proveniente do processo amostral.

No tocante aos argumentos apresentados para o não uso de um método científico na confecção de pesquisas eleitorais, as seguintes considerações devem ser feitas:

1. *Dos custos:* Para cada objetivo pretendido, existe um custo determinado. Um transplante de fígado para ser bem sucedido, por exemplo, depende das técnicas empregadas na cirurgia. A operação tem um determinado custo mínimo, sem o qual o transplante corre sérios riscos de não dar certo. Se o objetivo é saber qual candidato tem a preferência de voto do eleitor, as técnicas científicas são bem conhecidas pela teoria da amostragem estatística. A diferença é que aqui, não arcar com os custos do método implica não ter idéia da precisão das informações obtidas, enquanto que lá, dificilmente um paciente sobreviveria.

2. *Da complexidade dos planos:* Em relação a esta dificuldade, alternativas são viabilizadas a partir do uso de softwares estatísticos em conjunto com uma abordagem teórica adequada. Neste sentido, apresenta-se aqui uma nova abordagem para os planos amostrais a partir de modelos lineares mistos, como será visto ainda neste capítulo.

Quanto ao fato de que, em todo mundo, amostragem por quotas é utilizada, isso não é justificativa apresentável. Mesmo assim, não seria difícil lembrar que nas últimas eleições presidenciais na França e para primeiro ministro, na Inglaterra, todos os institutos de pesquisa erraram suas previsões. As limitações deste método existem e, é óbvio, independem de situações geográficas. O que poderia ocorrer é que, em certo país, a estabilidade política fosse de tal ordem que os riscos de erros sistemáticos no uso de um plano não probabilístico seriam suficientemente pequenos para justificar o uso de tal método, implicando inclusive em redução de custos. Mas este não parece ser o caso do Brasil.

Considerando esta mesma problemática da escolha de um método de amostragem, segundo sua natureza, *Cochran, Mosteller e Tukey* [8], já em 1954, alertavam para a necessidade de sempre se ter em mente duas questões:

- a) O pesquisador pode arcar com uma amostra probabilística ?
- b) A exposição a erros sistemáticos provenientes de uma amostra não-probabilística é suficientemente pequena para ser ignorada ?

Se a resposta for "não" para ambas as perguntas, então o levantamento não deve ser feito e, se já o foi, não deveria. Respondendo "sim" à primeira questão e "não" à segunda, então o não uso de uma amostra aleatória corresponde a falha bastante séria. A resposta "sim" às duas perguntas implica em considerar

cuidadosamente a questão econômica envolvida. Por último, se a resposta for "não" à primeira pergunta e "sim" para a segunda, sorte do pesquisador !

4.1 Uma proposta de plano amostral probabilístico

Tendo em vista as considerações anteriores, é de grande importância a existência de estudos no sentido da viabilização do uso de amostragem probabilística em pesquisas eleitorais e alternativas para controle de erros não amostrais inerentes às características da opinião pública². Assim sendo, com o intuito de contribuir nesta direção, bem como de incentivar posteriores estudos no Brasil, é apresentado a seguir um plano amostral para pesquisa eleitoral, de cunho probabilístico.

O objetivo é estimar a proporção de votos de um candidato a cargo majoritário, dando conta da precisão da informação levantada. Para o exemplo, foi considerado o caso de uma eleição municipal - para governo de estado e presidência da república, os planos são extensões naturais deste.

Imaginando a presença de "redutos eleitorais" distribuídos geograficamente segundo bairros e ainda a disponibilidade de informação a respeito do número de domicílios na cidade, por bairro, sugere-se o seguinte esquema.

1. A amostra é estratificada por bairros;
2. No bairro b , uma amostra aleatória simples (a.a.s.) de n_b ruas é selecionada sem reposição;

² Veja-se, por exemplo, o trabalho de *Noelle-Neumann* [21] em sua teoria sobre a espiral do silêncio.

3. Para a rua i sorteada no passo anterior, uma nova a.a.s., agora de m_i casas, é tomada, sem reposição;
4. Em cada uma das m_i casas selecionadas, são entrevistados todos os eleitores residentes.

Neste plano, a complexidade está na análise dentro de cada bairro. Trata-se de uma subamostragem desbalanceada em dois estágios. A abordagem clássica aplicada aqui é muito árdua. Para se ter idéia a respeito, o leitor poderá consultar, por exemplo, os livros de *Cochran* [6] e *Sukhatme & Sukhatme* [31].

Como alternativa ao desenvolvimento analítico clássico, é sugerido o entendimento de um modelo linear apropriado ao caso.

Para o esquema amostral proposto, a abordagem será via modelos lineares mistos.

Considere a cidade dividida em I bairros e que o número de casas encontradas no bairro i seja N_i . Desta forma tem-se que $N = \sum_{i=1}^I N_i$, onde N é o total de casas no município. Seja E o número de eleitores inscritos na cidade. Assim, a proporção de votos para o candidato 1, na cidade, é dada por:

$$P = \sum_{i=1}^I W_i \mu_i \quad (1)$$

com $W_i = \frac{N_i}{E}$ e μ_i = a média de votos, por casa, para o candidato 1, no bairro i .

Por outro lado, o número de votos para o candidato 1 obtidos na casa k da rua j , no bairro i , pode ser escrito como:

$$y_{ijk} = \mu_i + R_{ij} + \varepsilon_{ijk} \quad (2) ,$$

onde:

R_{ij} - o efeito aleatório de rua - é a diferença entre a média de votos, por casa, do candidato 1, na rua j (do bairro i), e a média de votos, por casa, (do candidato 1) no bairro i , com média zero e variância σ_r^2 .

ε_{ijk} - o erro aleatório - é o desvio entre o número de votos para o candidato 1, na casa k (da rua j , bairro i) e a média de votos, por casa, da rua j (no bairro i), com média zero e variância σ_e^2 .

Por sua vez, é sempre possível escrever μ_i da seguinte forma:

$$\mu_i = \mu + B_i \quad \text{onde } \mu \text{ é a média de votos (para o candidato 1), por casa, na cidade e } B_i \text{ é o efeito fixo do bairro } i.$$

Dai, reescrevendo o modelo (2), tem-se:

$$y_{ijk} = \mu + B_i + R_{ij} + \varepsilon_{ijk} \quad (3)$$

A idéia, então, é ter uma estimativa de (1) a partir da análise do modelo (3).

Para evidenciar tal possibilidade, considere a seguinte notação matricial.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (4)$$

Aqui, \mathbf{y} é um vetor unidimensional contendo os y_{ijk} 's do modelo (3). $\boldsymbol{\beta}$ é um vetor de efeitos fixos com matriz de delineamento, associada, \mathbf{X} , conhecida. \mathbf{u} é um vetor de efeitos aleatórios, onde \mathbf{Z} é a matriz de delineamento, associada, conhecida. \mathbf{e} é o vetor de erros aleatórios, aqui confundidos com os efeitos de casas, cujos elementos não são necessariamente independentes.

Considerando \mathbf{G} e \mathbf{R} as respectivas matrizes de covariâncias de \mathbf{u} e \mathbf{e} respectivamente, é possível realizar inferências não só a respeito da média de \mathbf{y} , mas também a respeito de sua variância. Seja \mathbf{V} a matriz de covariâncias de \mathbf{y} . Então:

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \quad (5)$$

Uma vez que \mathbf{V} é conhecida, uma solução das equações normais

$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, é dada por:

$$\mathbf{b} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (6)$$

onde o símbolo $(-)$ denota uma inversa generalizada.

Agora, considere a combinação linear $\boldsymbol{\lambda}'\boldsymbol{\beta}$ de particular interesse.

McLean, Sanders e Stroup[19] mostraram que:

- $\lambda'b$ é o BLUE³ de $\lambda'\beta$, caso $\lambda'\beta$ seja estimável.
- $Var(\lambda'b) = \lambda'(X'V^{-1}X)^{-1}\lambda$.

Ora, mas a proporção de votos para o candidato 1, definida em (1), é uma combinação linear de β estimável.

Para que uma combinação linear $\lambda'\beta$ seja estimável é necessário que λ pertença ao espaço gerado pelas linhas da matriz de delineamento X . Observa-se, por exemplo, que nesta matriz (aqui considerada), a primeira observação de cada linha corresponde à soma das demais observações na mesma linha. Assim, qualquer λ que satisfaça esta propriedade é estimável. Em relação à proporção de votos do candidato 1, tem-se que:

$$\begin{aligned} P &= \sum_{i=1}^I W_i \mu_i = \sum_{i=1}^I W_i (\mu + B_i) = \\ &= \frac{N}{E} \mu + \sum_{i=1}^I W_i B_i . \end{aligned}$$

Sendo $\beta' = [\mu \ B_1 \ B_2 \ \dots \ B_I]$, é fácil ver que $P = \lambda'\beta$, onde:

$$\lambda' = \left[\frac{N}{E} \quad \frac{N_1}{E} \quad \frac{N_2}{E} \quad \dots \quad \frac{N_I}{E} \right].$$

³ BLUE é a sigla de *Best Linear Unbiased Estimator*. Um estimador BLUE é dito ser o melhor no sentido de ter a menor variância dentre a classe dos estimadores lineares centrados.

Como $N = \sum_{i=1}^I N_i$, então, $P = \lambda' \beta$ é estimável.

Não obstante a constatação de estimabilidade de P , é preciso ter o conhecimento da matriz de covariâncias V - o que não acontece aqui - a fim de proceder às estimativas, como visto. Para o caso estudado, apesar de não se ter tal matriz, sabe-se qual a sua estrutura a partir do processo de aleatorização. Fixando o bairro i , a estrutura de V é, então, expressa como:

$$\begin{aligned} E(y_{ijk}, y_{ij'k'}) &= \sigma_r^2 + \sigma_e^2, \text{ se } j=j', k=k' \\ &= \sigma_r^2, \text{ se } j=j', k \neq k' \\ &= 0, \text{ se } j \neq j' \end{aligned}$$

Tal estrutura é conhecida como de Simetria Composta. Neste caso, a variância entre duas observações, numa mesma rua e mesma casa, é dada pela soma das componentes de variância de rua e casa. Se as observações forem de uma mesma rua, mas de diferentes casas, a variância será dada pelo componente de variância da rua.

Ainda assim, é necessário, no entanto, o emprego de algum método para a estimação de V .

Uma análise levando em conta toda a abordagem introdutória aqui apresentada e incorporando a informação a respeito da estrutura da matriz de covariância dos dados, pode ser viabilizada utilizando-se o PROC MIXED⁴ do SAS®.

⁴O PROC MIXED é uma generalização do PROC GLM, uma vez que este último aborda a teoria de modelos lineares padrão (de efeitos fixos), enquanto que o primeiro, incorpora análises de modelos de efeitos aleatórios, permitindo assim, a abordagem de modelos lineares mistos.

O procedimento MIXED estima V usando o método de máxima verossimilhança restrita. Para estimar $\lambda'\beta$, usa então o estimador $\lambda'b$ com V' - matriz de covariâncias de y , estimada - no lugar de V . Kackar e Harville [14,15] mostraram que procedendo desta maneira, tem-se um estimador centrado para $\lambda'\beta$.

Para exemplificar a aplicação do modelo desenvolvido nas linhas anteriores, construiu-se uma população fictícia com as seguintes características:

Quadro 9

Características, por bairros, de população fictícia gerada para testar o modelo.

Bairro	Nº de Casas	Nº de Votos para o Candidato	Nº de Eleitores
1	26	24	78
2	21	18	63
3	52	39	149
4	56	96	145
5	44	80	122
Total	199	257	557

Pelo Quadro 9, tem-se a informação de que o candidato tem, em verdade, 46% dos votos, na cidade.

Uma amostra aleatória foi, então, retirada da população, segundo o esquema proposto. Foram selecionadas 50% das ruas no bairro e 55% das casas dentro de cada rua sorteada. A amostra final é apresentada a seguir, considerando: BAIRRO - nº correspondente ao bairro onde retirou-se a amostra; RUA - nº da rua sorteada no bairro; CASA - casa sorteada na respectiva rua; NELEIT - nº de eleitores residentes na casa selecionada; N1 - nº de votos para o candidato 1, obtidos na casa.

The SAS System					
OBS	BAIRRO	RUA	CASA	NELEIT	N1
1	1	1	1	2	0
2	1	1	2	1	0
3	1	1	7	4	1
4	1	4	2	6	3
5	1	4	4	6	3
6	1	4	6	5	2
7	2	3	1	5	2
8	2	3	2	2	0
9	2	3	6	1	0
10	3	1	1	3	1
11	3	1	5	2	0
12	3	1	7	1	0
13	3	1	9	1	0
14	3	4	1	2	0
15	3	4	2	5	2
16	3	4	4	1	0
17	3	4	5	3	1
18	3	4	10	1	0
19	3	4	11	3	1
20	3	5	1	2	0
21	3	5	4	3	1
22	3	5	5	5	2
23	3	5	8	2	0
24	3	7	1	1	0
25	3	7	4	2	0
26	4	1	3	5	2
27	4	2	1	5	3
28	4	2	4	3	2
29	4	6	1	1	1
30	4	6	5	2	2
31	4	7	1	6	2
32	4	7	5	2	2
33	4	8	4	1	1
34	4	8	5	1	1
35	5	1	3	5	2
36	5	3	1	1	1
37	5	3	3	4	2
38	5	5	4	2	2
39	5	5	5	2	2
40	5	7	4	2	2
41	5	7	5	2	2
42	5	7	6	3	2

Para a análise segundo o modelo (3), utilizou-se o seguinte programa SAS®.

```

OPTIONS LS=80 PS=54 NODATE;

PROC MIXED DATA=AMOSTRA;
  CLASS BAIRRO RUA ;
  MODEL NI=BAIRRO/S;
  RANDOM RUA /SUBJECT=BAIRRO TYPE=CS;
  ESTIMATE 'P ESTIMADO' INTERCEPT .36 BAIRRO .05 .04 .09 .1 .08;

RUN;

```

Parte da saída do programa acima é apresentada a seguir.

The MIXED Procedure					
Covariance Parameter Estimates (REML)					
Cov Parm	Ratio	Estimate	Std Error	Z	Pr > Z
RUA Diagonal	0.58702759	0.27049327	0.19566498	1.38	0.1668
CS	0.00000000	0.00000000	7730682.6246	0.00	1.0000
Residual	1.00000000	0.46078459	0.12474943	3.69	0.0002

The MIXED Procedure					
Solution for Fixed Effects					
Parameter	Estimate	Std Error	DDF	T	Pr > T
INTERCEPT	1.87066849	0.35994523	26	5.20	0.0000
BAIRRO 1	-0.37066849	0.58446953	26	-0.63	0.5315
BAIRRO 2	-1.20400183	0.74407574	26	-1.62	0.1177
BAIRRO 3	-1.42218477	0.47848737	26	-2.97	0.0063
BAIRRO 4	-0.08411814	0.48628635	26	-0.17	0.8640
BAIRRO 5	0.00000000

The MIXED Procedure					
ESTIMATE Statement Results					
Parameter	Estimate	Std Error	DDF	T	Pr > T
P ESTIMADO	0.47033872	0.06254625	26	7.52	0.0000

O valor de P foi estimado em 0.47 com desvio padrão de 0.06, atingindo assim o objetivo de estimação da proporção de votos do candidato tendo idéia de sua

precisão. Para completar uma análise exploratória do método, foram retiradas 100 amostras segundo o plano amostral proposto. O programa utilizado nesta simulação encontra-se no apêndice. Os resultados estão descritos abaixo, com EST sendo o valor de P estimado e SE o respectivo desvio padrão.

The SAS System						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
EST	Estimate	100	0.4653522	0.0408993	0.3525563	0.5390210
SE	Std Error	100	0.0481277	0.0079713	0.0328964	0.0714290

O comportamento da simulação confirma que não há problemas em aplicar o método, aqui utilizado, para a estimação da variância do estimador na prática.

Assim, percebe-se que há espaço para alternativas de cálculo de variâncias dos estimadores, em planos complexos, como os de pesquisas eleitorais.

4.2 Considerações finais

Pelo que foi visto até agora, é possível afirmar que existe sim viabilidade teórica do uso de amostragem probabilística em levantamentos de intenção de voto. O que falta, então, para que tenham-se disponíveis, de fato, informações sobre margem de erro e níveis de confiança, nestas sondagens, é arcar com os custos do método, na prática. Do contrário, é preciso ser coerente e assumir o não conhecimento da precisão das informações levantadas.

Por fim, é importante salientar, mais uma vez, que pesquisa eleitoral é assunto multidisciplinar que pode e deve ser criticado de diversos pontos de vista científicos, contribuindo assim, para uma melhor qualidade das informações geradas para a sociedade. Deste modo, não é simplesmente utilizando uma metodologia estatística de amostragem que estará dado um ponto final às polêmicas e controvérsias a respeito do assunto. Contribuições de outras áreas como ciência política, sociologia, psicologia social, jornalismo, direito... são também necessárias e muito importantes.

Apêndice

O seguinte programa foi utilizado para realizar a simulação, cujo resultado encontra-se no capítulo 4.

```
%GLOBAL _DISK_;
%LET _DISK_=ON;
%GLOBAL _PRINT_;
%LET _PRINT_=OFF;

OPTIONS LS=80 PS=54 NODATE;

/* DADOS POPULACIONAIS */
data simul;
input obs bairro nrua rua ncasa casa neleit n1;
cards;
    1      1      4      1      7      1      2      0
    2      1      4      1      7      2      1      0
    3      1      4      1      7      3      4      1
    4      1      4      1      7      4      1      0
    5      1      4      1      7      5      5      2
    6      1      4      1      7      6      3      1
    7      1      4      1      7      7      4      1
    8      1      4      2      6      1      6      4
    9      1      4      2      6      2      2      0
   10      1      4      2      6      3      2      0
   11      1      4      2      6      4      1      0
   12      1      4      2      6      5      2      0
   13      1      4      2      6      6      1      0
   14      1      4      3      7      1      2      0
   15      1      4      3      7      2      2      0
   16      1      4      3      7      3      2      0
   17      1      4      3      7      4      6      3
   18      1      4      3      7      5      1      0
   19      1      4      3      7      6      6      3
   20      1      4      3      7      7      3      1
   21      1      4      4      6      1      1      0
   22      1      4      4      6      2      6      3
   23      1      4      4      6      3      2      0
   24      1      4      4      6      4      6      3
   25      1      4      4      6      5      2      0
   26      1      4      4      6      6      5      2
   27      2      3      1      6      1      5      2
   28      2      3      1      6      2      4      1
   29      2      3      1      6      3      3      1
   30      2      3      1      6      4      1      0
   31      2      3      1      6      5      4      1
   32      2      3      1      6      6      5      2
   33      2      3      2      8      1      4      1
```

34	2	3	2	8	2	3	1
35	2	3	2	8	3	3	1
36	2	3	2	8	4	5	2
37	2	3	2	8	5	3	1
38	2	3	2	8	6	3	1
39	2	3	2	8	7	5	2
40	2	3	2	8	8	1	0
41	2	3	3	7	1	5	2
42	2	3	3	7	2	2	0
43	2	3	3	7	3	2	0
44	2	3	3	7	4	2	0
45	2	3	3	7	5	1	0
46	2	3	3	7	6	1	0
47	2	3	3	7	7	1	0
48	3	8	1	9	1	3	1
49	3	8	1	9	2	1	0
50	3	8	1	9	3	3	1
51	3	8	1	9	4	2	0
52	3	8	1	9	5	2	0
53	3	8	1	9	6	3	1
54	3	8	1	9	7	1	0
55	3	8	1	9	8	2	0
56	3	8	1	9	9	1	0
57	3	8	2	4	1	5	2
58	3	8	2	4	2	1	0
59	3	8	2	4	3	3	1
60	3	8	2	4	4	4	1
61	3	8	3	3	1	6	3
62	3	8	3	3	2	4	1
63	3	8	3	3	3	2	0
64	3	8	4	11	1	2	0
65	3	8	4	11	2	5	2
66	3	8	4	11	3	6	3
67	3	8	4	11	4	1	0
68	3	8	4	11	5	3	1
69	3	8	4	11	6	2	0
70	3	8	4	11	7	2	0
71	3	8	4	11	8	4	1
72	3	8	4	11	9	1	0
73	3	8	4	11	10	1	0
74	3	8	4	11	11	3	1
75	3	8	5	8	1	2	0
76	3	8	5	8	2	3	1
77	3	8	5	8	3	1	0
78	3	8	5	8	4	3	1
79	3	8	5	8	5	5	2
80	3	8	5	8	6	2	0
81	3	8	5	8	7	3	1
82	3	8	5	8	8	2	0
83	3	8	6	6	1	4	1
84	3	8	6	6	2	4	1
85	3	8	6	6	3	3	1
86	3	8	6	6	4	6	3
87	3	8	6	6	5	3	1
88	3	8	6	6	6	6	3
89	3	8	7	5	1	1	0
90	3	8	7	5	2	2	0

91	3	8	7	5	3	2	0
92	3	8	7	5	4	2	0
93	3	8	7	5	5	2	0
94	3	8	8	6	1	3	1
95	3	8	8	6	2	2	0
96	3	8	8	6	3	4	1
97	3	8	8	6	4	5	2
98	3	8	8	6	5	4	1
99	3	8	8	6	6	2	0
100	4	11	1	3	1	2	1
101	4	11	1	3	2	2	2
102	4	11	1	3	3	5	2
103	4	11	2	4	1	5	3
104	4	11	2	4	2	5	2
105	4	11	2	4	3	2	2
106	4	11	2	4	4	3	2
107	4	11	3	4	1	3	2
108	4	11	3	4	2	1	1
109	4	11	3	4	3	2	2
110	4	11	3	4	4	2	2
111	4	11	4	5	1	5	2
112	4	11	4	5	2	5	3
113	4	11	4	5	3	1	1
114	4	11	4	5	4	2	2
115	4	11	4	5	5	2	2
116	4	11	5	9	1	3	2
117	4	11	5	9	2	1	1
118	4	11	5	9	3	1	1
119	4	11	5	9	4	3	2
120	4	11	5	9	5	3	2
121	4	11	5	9	6	5	3
122	4	11	5	9	7	5	2
123	4	11	5	9	8	1	1
124	4	11	5	9	9	1	1
125	4	11	6	5	1	1	1
126	4	11	6	5	2	1	1
127	4	11	6	5	3	1	1
128	4	11	6	5	4	2	2
129	4	11	6	5	5	2	2
130	4	11	7	5	1	6	2
131	4	11	7	5	2	4	2
132	4	11	7	5	3	3	2
133	4	11	7	5	4	3	2
134	4	11	7	5	5	2	2
135	4	11	8	5	1	2	2
136	4	11	8	5	2	2	2
137	4	11	8	5	3	5	3
138	4	11	8	5	4	1	1
139	4	11	8	5	5	1	1
140	4	11	9	7	1	5	2
141	4	11	9	7	2	2	2
142	4	11	9	7	3	2	2
143	4	11	9	7	4	1	1
144	4	11	9	7	5	1	1
145	4	11	9	7	6	1	1
146	4	11	9	7	7	1	1
147	4	11	10	5	1	3	2

148	4	11	10	5	2	3	2
149	4	11	10	5	3	2	1
150	4	11	10	5	4	5	2
151	4	11	10	5	5	1	1
152	4	11	11	4	1	1	1
153	4	11	11	4	2	6	2
154	4	11	11	4	3	2	1
155	4	11	11	4	4	3	2
156	5	8	1	3	1	1	1
157	5	8	1	3	2	3	2
158	5	8	1	3	3	5	2
159	5	8	2	6	1	2	1
160	5	8	2	6	2	3	2
161	5	8	2	6	3	6	3
162	5	8	2	6	4	1	1
163	5	8	2	6	5	2	2
164	5	8	2	6	6	1	1
165	5	8	3	4	1	1	1
166	5	8	3	4	2	2	2
167	5	8	3	4	3	4	2
168	5	8	3	4	4	4	2
169	5	8	4	7	1	2	2
170	5	8	4	7	2	2	2
171	5	8	4	7	3	1	1
172	5	8	4	7	4	6	3
173	5	8	4	7	5	2	2
174	5	8	4	7	6	3	2
175	5	8	4	7	7	6	2
176	5	8	5	5	1	2	2
177	5	8	5	5	2	2	1
178	5	8	5	5	3	2	2
179	5	8	5	5	4	2	2
180	5	8	5	5	5	2	2
181	5	8	6	6	1	4	2
182	5	8	6	6	2	1	1
183	5	8	6	6	3	5	2
184	5	8	6	6	4	2	2
185	5	8	6	6	5	4	2
186	5	8	6	6	6	2	1
187	5	8	7	6	1	2	2
188	5	8	7	6	2	3	2
189	5	8	7	6	3	4	2
190	5	8	7	6	4	2	2
191	5	8	7	6	5	2	2
192	5	8	7	6	6	3	2
193	5	8	8	7	1	2	2
194	5	8	8	7	2	3	2
195	5	8	8	7	3	6	2
196	5	8	8	7	4	2	2
197	5	8	8	7	5	2	1
198	5	8	8	7	6	4	2
199	5	8	8	7	7	2	2

;

```

/*PROGRAMA PARA RETIRADA DE AMOSTRA ALEATÓRIA ESTRATIFICADA, EM DOIS ESTÁGIOS*/

/*=====*/
data nruas;
  set simul;
  by bairro rua;
  if first.rua then nrua+1;
  if last.bairro then do;
    output;
    nrua=0;
  end;
keep bairro nrua;

/*=====*/
data rua;
  keep bairro nrua rua;
  set simul;

/*=====*/
proc freq data=rua noprint;
  tables bairro*rua/out=tab1;

/*=====*/
data newrua;
  merge nruas tab1(rename=(count=ncasa));
  by bairro;

/*=====*/
data new2rua;
  keep bairro nrua rua;
  set newrua;

/*=====*/

data finall;
  EST=.; SE=.;
run;

%macro roda(n);
  %do i=1 %to &n;
    data am2rua;
      keep bairro rua;
      retain tam npop;
      set new2rua;
      by bairro;
      if first.bairro then do;
        tam=max(1,int(0.5*nrua));
        npop=nrua;
      end;
      p=tam/npop;
      if ranuni(0)<p then do;
        output;
        tam=tam-1;
      end;
      npop=npop-1;
    end;
  %end;

```

```

/*=====*/
data amostra;
  keep bairro rua casa neleit n1;
  retain npop tam;
  merge simul am2rua(in=inrua) tab1;
  by bairro rua;
  if inrua;
  if first.rua then do;
    tam=max(int(0.55*ncasa),1);
    npop=ncasa;
  end;
  p=tam/npop;
  if ranuni(0)<p then do;
    output;
    tam=tam-1;
  end;
  npop=npop-1;

/* USO DO PROC MIXED PARA ANALISE DA AMOSTRA */

PROC MIXED DATA=AMOSTRA;
  CLASS BAIRRO RUA ;
  MODEL N1=BAIRRO/S;
  RANDOM RUA /SUBJECT=BAIRRO TYPE=CS;
  ESTIMATE 'P ESTIMADO' INTERCEPT .36 BAIRRO .05 .04 .09 .1 .08;
  MAKE 'Estimate' OUT=FINAL;

data finall;
  set finall final;
  KEEP EST SE;

%end;

%mend;

%roda(100)
proc print;run;
proc means;
run;

```

Referências Bibliográficas

- [1] AMARANTE, A. R. (1992) *Um Curso em Modelos Lineares*, Tese de Mestrado, UNICAMP.
- [2] BESSON, J. L. (1995) *A Ilusão das Estatísticas*, Unesp, São Paulo.
- [3] BOURDIEU, P. *A Opinião Pública Não Existe em Crítica Metodológica, Investigação Social e Enquete Operária*, 5ª edição, Michel Thiollent (autor e org).
- [4] BOX, G. E. P. ; HUNTER, W. G. and HUNTER, J. S. (1978) *Statistics for Experimenters*, John Wiley and Sons, New York.
- [5] COCHRAN, W. G. ; MOSTELLER, F. and TUKEY, J. W. (1954) *Principles of Sampling*, **Journal of the American Statistical Association**, Vol. 49, pp 13 - 35.
- [6] COCHRAN, W. G. (1977) *Sampling Techniques*, 3rd, John Wiley, New York.
- [7] CORNFIELD, J. (1944) *On Samples from Finite Populations*, **Journal of the American Statistical Association** .
- [8] FIGUEIREDO, M. F. (1985) *As pesquisas Pre-eleitorais e as "surpresas" políticas, em 1985: O voto em São Paulo*, Bolívar Lamounier (org), IDESP.
- [9] FIGUEIREDO, R. e CERVellini, S. (1995) *Contribuições para o Conceito de Opinião Pública*, **Opinião Pública**, pp 112 - 120.
- [10] FOLKS, J. L. (1981) *Ideas of Statistics*, John Wiley, New York.
- [11] FULLER, W. A. and BATTESE, G. E. (1973) *Transformations for Estimation of Linear Models with Nested Errors Structure*, **Journal of American Statistical Association**, Vol. 68, pp 626 - 632.

- [12] HABERMAS, J. (1990) *Historia y Crítica de la Opinión Pública*, 4ª edición, GG Mass Media.
- [13] HINKELMANN, K. and KEMPTHORNE, O. (1994) *Design and Analysis of Experiments Vol. 1: Introduction to Experimental Design*, John Wiley and Sons, Inc, New York.
- [14] KACKAR, R. N. and HARVILLE, D.A. (1984) *Unbiasedness of Two- stage Estimation and Prediction Procedures for Mixed Linear Models*, **Commun. Stat. A. Theory and Methods**, Vol. 10, pp 1249-1261.
- [15] KACKAR, R. N. and HARVILLE, D.A. (1984) *Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models*, **Journal of American Statistical Association**, Vol. 79, pp 853 - 862.
- [16] KEMPTHORNE, O. and FOLKS, L. (1971) *Probability, Statistics and Data Analysis*, The Iowa University Press, Ames.
- [17] KISH, L. (1965) *Survey Sampling*, John Wiley, New York.
- [18] KRUSKAL, W. H. and MOSTELLER, F. (1986) *Representative Sampling*, em **Encyclopedia of Statistical Sciences**, Vol. 8. (S.Kotz and N.L.Johnson, eds), John Wiley, pp 77-81.
- [19] McLEAN, R. A. ; SANDERS, W. L. and STROUP, W. W. (1991) *A Unified Approach to Mixed Linear Models*, **The American Statistician**, Vol. 45, pp 54-64.
- [20] MOSER, C. A. and STUART, A. (1953) *An experimental study of quota sampling*, **Journal of Royal Statistical Society, (A)** Vol. 116, pp 349-405.
- [21] NOELLE - NEUMANN, E. (1993) *Pesquisa Eleitoral e Clima de Opinião*, **Opinião Pública**, Vol. 1, Nº 2, pp 47-62.
- [22] PETERS, W.S. (1987) *Counting for something*, Springer-Verlag, New York.
- [23] ROBINSON, G. K. (1991) *That BLUP Is a Good Thing : The Estimation of Random Effects*, **Statistical Science** , Vol. 6 , pp 15-51.

- [24] SARNDAL, C. E. ; SWENSSON B. and WRETMAN, J. (1992) *Model Assisted Survey Sampling*, Springer-Verlag, New York.
- [25] SAS INSTITUTE Inc (1992) *SAS Technical Report* , pp 229, **SAS/STAT Softwares Changes and Enhancements**, Release 6.07, Cary, NC: SAS Institute Inc.
- [26] SEARLE, S. R. (1971) *Linear Models*, John Wiley & Sons, New York.
- [27] SEARLE, S. R. (1987) *Linear Models for Unbalanced Data*, John Wiley, New York.
- [28] SNEDECOR, G. W. and COCHRAN, W. G. (1982) *Statistical Methods*, 7rd, The Iowa State University Press, Ames.
- [29] SOUZA, J. (1990) *Pesquisa Eleitoral: Críticas e Técnicas*, Editora do Senado, Brasília-DF.
- [30] STEPHAN, F. F. and McCARTHY, P. J. (1958) *Sampling Opinions*, John Wiley and Sons, New York.
- [31] SUKHATME, P. V. and SUKHATME, B. V. (1970) *Sampling Theory of Surveys with Applications*, 2nd, Iowa State University Press, Ames.
- [32] THIOLLENT, M. (1986) *Opinião Pública e Debates Políticos*, Polis, São Paulo.
- [33] THIOLLENT, M. (1987) *Crítica Metodológica, Investigação Social e Enquete Operária*, Polis, São Paulo.
- [34] THIOLLENT, M. (1989) *Pesquisas eleitorais em debate na imprensa*, Cortez Editora e Editora Autores Associados, São Paulo.
- [35] WILLIAMS, B. (1978) *A sampler on sampling*, John Wiley, New York.

[36] WOLFINGER, R. D. ; TOBIAS, R. D. and SALL, J. (1991) *Mixed Models: A Future Direction* ; **Preceedings of the Sixteenth Annual SAS Users Group Conference**, pp 1380 - 1388.