PREVISÃO ELEITORAL POR CONTAGEM RÁPIDA NA ELEIÇÃO PRESIDENCIAL NICARAGÜENSE DE 1990

Carlos Alberto de Bragança Pereira

Julio da Motta Singer

Departamento de Estatística - USP

José Ferreira de Carvalho

Departamento de Estatística - UNICAMP

1. Introdução

Dentre as atividades desenvolvidas pela Organização dos Estados Americanos (OEA) na missão de observação das eleições nicaraguenses de 25 de fevereiro de 1990, constava um processo de previsão de resultados por contagem rápida. Por razões de cunho político, os dirigentes da OEA precisavam dispor de uma boa estimativa dos resultados finais logo após as primeiras horas da apuração. Embora existissem vários candidatos, o interesse principal estava dirigido aos candidatos da Frente Sandinista de Liberación Nacional (FSLN) e da Unión Nacional Opositora (UNO). Nesse contexto, os autores foram convidados para avaliar um sistema de previsão proposto por consultores internacionais com relação ao esquema amostral, método de estimação e administração do banco de dados.

O objetivo deste artigo é descrever os principais aspectos do trabalho de assessoria realizado durante os 12 dias de nossa permanência naquele país. Uma crítica ao planejamento amostral proposto pelo grupo de assessores internacionais da OEA está apresentada na Seção 2. Na Seção 3 indicamos os métodos que utilizamos para previsão estatística. Finalmente, na Seção 4 apresentamos os resultados do processo de previsão.

2. Análise do Planejamento Amostral

A população de 4392 juntas receptoras de votos (JRV) estava dividida em três estratos constituídos da seguinte forma:

Estrato 1: JRV's situadas nas localidades rurais de todas as nove regiões do país, com exceção de Manágua;

Estrato 2: JRV's situadas nas localidades urbanas de todas as nove regiões do país, exceto Manágua;

Estrato 3: JRV's situadas na região de Manágua (urbanas e rurais).

Em cada estrato, de forma independente, foram selecionadas JRV's seguindo o procedimento descrito abaixo:

a) com igual probabilidade de seleção, uma JRV dentre as ainda não incluídas na amostra era sorteada;

b) o número de eleitores da JRV selecionada era acumulado;

c) os passos a) e b) eram repetidos até que o número acumulado de eleitores atingisse 10% do total do estrato.

Segundo esses procedimentos foram obtidas 10 amostras exclusivas e exaustivas, das quais uma foi selecionada através de critérios subjetivos. A amostra escolhida contava com 437 JRV's, englobando cerca de 170 000 eleitores.

Evidentemente, para efeito de inferência sobre o total de votos, esse procedimento caracteriza a JRV como unidade amostral. Outra indicação nesse sentido está no fato de que os dados seriam obtidos das atas fornecidas pelas Juntas Receptoras de Votos (que também funcionam como centros de apuração) contendo informações sobre o total de votos de cada candidato, o total de votos brancos e nulos e o total de votantes. Apesar disso, o material que nos foi apresentado pelos assessores internacionais considerava o eleitor como unidade amostral.

Embora tivesse caráter aleatório, o procedimento de seleção não obedecia a nenhum dos padrões usualmente empregados em pesquisas amostrais. Isto impossibilitava a atribuição às unidades amostrais, das probabilidades de seleção necessárias para a utilização dos métodos clássicos de amostragem. Dado que, por motivos logísticos, não podíamos selecionar outra amostra utilizando métodos apropriados, decidimos examinar aquela que nos foi apresentada numa tentativa de identificação de alguma característica que sugerisse que sua utilização pudesse introduzir um possível vício nas estimativas desejadas.

Através de contatos mantidos com analistas políticos de diversas instituições e partidos envolvidos no processo eleitoral, obtivemos informações sobre as nove regiões do país, que nos sugeriram a seguinte pós-estratificação:

1: Esteli, rural

2: Esteli, urbano

3: Leon, rural

4: Leon, urbano

5: Manágua, rural

6: Manágua, urbano

7: Granada, rural

8: Granada, urbano

9: Juigalpa, rural

10: Juigalpa, urbano

11: Matagalpa, rural

12: Matagalpa, urbano

13: P. Cabezas, S. Carlos e Bluefield, rural

14: P. Cabezas, S. Carlos e Bluefield, urbano.

Como o número de eleitores inscritos (tamanho) das JRV's poderia ser uma característica importante para o processo de previsão do total de votos, comparamos suas distribuições na amostra com as correspondentes na população. Não encontramos diferenças consideráveis entre os percentis das distribuições do tamanho das JRV's na amostra e na população em qualquer dos 14 estratos. Isto nos levou à conclusão que, num processo de seleção aleatória de JRV's, uma amostra com características semelhantes àquelas da amostra escolhida teria grande chance de ser selecionada; como, por outro lado, os analistas políticos mencionados acima também não identificaram aspectos que pudessem sugerir a introdução de algum vício, acreditamos que a amostra em questão poderia ser considerada "representativa" da população.

Em processos de previsão eleitoral, deseja-se obter uma estimativa precisa (com pequena margem de erro) dos totais de votos dados aos candidatos, com uma grande confiança. Os consultores internacionais responsáveis pelo processo amostral estabeleceram que a previsão deveria apresentar erro máximo de 0,35%, com uma confiança de 99,9%. Uma análise técnica do procedimento proposto por eles indicou-nos que esses níveis não poderiam ser atingidos a partir da amostra selecionada. Esse fato está essencialmente ligado à consideração de eleitores como unidades amostrais, em vez de JRV's. Os níveis aventados só poderiam ser atingidos se os 170 000 eleitores na amostra houvessem sido selecionados aleatoriamente, independentemente de suas respectivas JRV's. Tal seleção não poderia ser efetuada, já que não há condições de selecionar cada voto individualmente. Cada JRV corresponde a um grupo (conglomerado) de eleitores. A avaliação dos níveis de precisão e de confiança efetivos para amostragem por conglomerados é difícil, porque depende do grau de homogeneidade das preferências eleitorais dos indivíduos em cada JRV, medida pela correlação intraclasse. Utilizando as informações obtidas em contatos com analistas políticos de algumas entidades envolvidas no processo eleitoral, pudemos fazer hipóteses sobre esse grau de homogeneidade. Sob essas hipóteses, concluímos que, com 99,9% de confiança, o erro de previsão deveria estar entre 1% e 5%.

Para avaliar o efeito da amostragem por conglomerados, obtivemos o número de JRV's necessário para atingir os níveis de erro e de confiança pré-estabelecidos, sob diferentes suposições sobre a correlação intraclasse. Se a correlação intraclasse for nula (isto é, se o grau de homogeneidade entre os eleitores das JRV's for nulo), o número de JRV's exigido seria 425. Para qualquer valor da correlação intraclasse acima de 0,02, precisaríamos amostrar toda a população de 4392 urnas.

3. Métodos de Previsão

Para tentar manter os erros de previsão nos níveis mais baixos, sem afetar a confiança, decidimos empregar estimadores do tipo razão e regressão, que reduzem a influência da variação dos tamanhos das JRV's. A pós-estratificação adotada também contribui para a redução do nível de erro.

A primeira técnica de previsão considerada utilizou o estimador razão baseado no método clássico de amostragem por conglomerados. A suposição básica que fizemos foi que a amostra em questão possa ser considerada equivalente a uma amostra aleatória simples de JRV's em cada estrato. Como esta suposição pode ser colocada em dúvida, devido à forma como a amostra foi selecionada, utilizamos também um segundo método de previsão baseado em modelos de superpopulações cuja suposição básica está relacionada com a distribuição dos votos por candidatos entre as JRV's dentro de cada estrato. Todos os membros do Grupo de Análise da OEA nessa missão (incluindo aí os assessores internacionais) concordaram com a suposição de que, em cada estrato, a amplitude de variação da porcentagem de votos de um determinado candidato não seria maior que 0,3 (por exemplo, que estaria entre 20% e 50% ou entre 30% e 60%, etc.) Nesse caso, o estimador utilizado foi o estimador regressão, tendo como variável independente o tamanho da JRV. Detalhes técnicos sobre os métodos de previsão empregados estão apresentados no Apêndice.

4. Resultados

Inicialmente, elaboramos programas computacionais para verificação da qualidade e consistência dos dados, já que o sistema existente não incorporava essas características, necessárias para uma análise correta. Utilizamos o sistema SAS versão 6.02, em um microcomputador "Laptop" Toshiba T1600. Todos os programas foram testados em amostras e populações simuladas.

Os resultados eram produzidos aproximadamente 10 minutos após a entrega das informações amostrais, quando não eram detectados erros nos dados de entrada; neste caso, era elaborada uma lista de erros, que era enviada ao sistema de entrada de dados, para as correções.

Na tabela abaixo, apresentamos um resumo das previsões efetuadas a partir do início da apuração (18:00 horas do dia 25 de fevereiro). As porcentagens indicadas foram calculadas em relação ao total de votos válidos (total de votantes menos abstenções menos votos brancos e nulos); os valores rotulados como Mínima e Máxima correspondem a limites de intervalos de confiança com coeficiente de confiança de 99,9%. A previsão utilizada para efeito político (numa reunião do Secretário Geral da OEA, Embaixador Baena Soares com o Presidente Daniel Ortega na madrugada do dia 26 de fevereiro) foi baseada em apenas 48% das JRV's previstas para a amostra; apesar do impressionante apoio logístico para a coleta de dados (envolvendo cerca de 400 observadores, 220 automóveis e uma sofisticada rede de telecomunicações) a informação mais completa que conseguimos obter envolveu 84% das JRV's previstas para a amostra e foi apresentada aproximadamente 44 horas após o início da apuração.

Porcentagem de Votos Previstos
Relativamente ao Total de Votos Válidos

HORAS DE APURAÇÃO	% de JRV NA AMOSTRA		MÉTODO CONGLOMERADOS UNO	MÉTODO CONGLOMERADOS FSLN	MÉTODO SUPERPOPULAÇÃO UNO	MÉTODO SUPERPOPULAÇÃO FSLN
9	48%	Mínima	51,4	37,0	54,1	37,9
		Média	55,0	40,1	55,7	39,5
		Máxima	58,6	43,2	57,3	41,1
21	74%	Mínima	53,5	36,5	55,4	37,5
		Média	56,2	39,1	56,5	38,7
		Máxima	58,9	41,7	57,7	39,9
44	84%	Mínima	54,3	35,8	55,8	37,3
		Média	56,9	38,3	56,8	38,4
		Máxima	59,6	40,9	57,9	39,5

O resultado da contagem de 81,4% das JRV's publicado no "El Nuevo Diário" de 27 de fevereiro de 1990 indicou 55,2% dos votos válidos para a UNO e 40,8% para a FSLN. Ambos os resultados estavam dentro das previsões produzidos com 48% da amostra às 3:00 horas da manhã de 26 de fevereiro de 1990.

APÊNDICE

A1 - Notação

Cada unidade amostral (JRV) é representada por um par ordenado (i, j) de índices onde $i = 1, 2, \ldots, 14$ é o número do estrato e $j = 1, 2, \ldots, N_i$ é o número da JRV do estrato. N_i é o número de JRV's do estrato i e $N = N_1 + \ldots + N_{14}$ é o número total de JRV's.

A cada JRV (i, j) corresponde o número conhecido X_{ij} de eleitores inscritos. Também associamos à JRV (i, j) o vetor (desconhecido)

$$\mathbf{Y}_{ij} = (Y_{ijA}, Y_{ijB}, Y_{ijC}, Y_{ijD}, Y_{ijE})$$

onde

$$Y_{ijA} + Y_{ijB} + Y_{ijC} + Y_{ijD} + Y_{ijE} = X_{ij},$$

Y_{ijA} é o número de abstenções associado à JRV (i, j),

Y_{ijB} é o número de votos brancos e nulos associado à JRV (i, j),

Y_{ijC} é o número de votos em favor de Violeta Chamorro associado à JRV (i, j),

Y_{ijD} é o número de votos em favor de Daniel Ortega associado à JRV (i, j) e

Y_{ijE} é o número de votos em favor de outros candidatos associado à JRV (i, j).

Nosso interesse principal está relacioando com inferências sobre o vetor de totais

$$(Y_A, Y_B, Y_C, Y_D, Y_E) = \sum_{i=1}^{14}(Y_{i\cdot A}, Y_{i\cdot B}, Y_{i\cdot C}, Y_{i\cdot D}, Y_{i\cdot E}) =$$

$$= \sum_{i=1}^{14}\sum_{j=1}^{N_i}(Y_{ijA}, Y_{ijB}, Y_{ijC}, Y_{ijD}, Y_{ijE}).$$

Note que o total de eleitores inscritos é dado por

$$X = Y_A + Y_B + Y_C + Y_D + Y_E = \sum_{i=1}^{14} X_{i\cdot} = \sum_{i=1}^{14}\sum_{j=1}^{N_i} X_{ij}$$

onde $X_{i\cdot}$ é o total de eleitores inscritos nas JRV's do estrato i.

A amostra é constituída por n_i $(\leq N_i)$ JRV's no i-ésimo estrato, totalizando $n = \sum_{i=1}^{14} n_i$ unidades amostrais.

Com o objetivo de simplificar as fórmulas, sem perda de generalidade, admitimos que a amostra é constituída pelas primeiras n_i JRV's do i-ésimo estrato, $(i = 1, \ldots, 14)$, isto é, as JRV's $(i, 1), \ldots, (i, n_i)$ formam a amostra do estrato i.

As características amostrais são representadas por:

a) $\mathbf{y}_i = (y_{iA}, y_{iB}, y_{iC}, y_{iD}, y_{iE}) = \sum_{j=1}^{n_i}(Y_{ijA}, Y_{ijB}, Y_{ijC}, Y_{ijD}, Y_{ijE})$;

os elementos desse vetor correspondem ao total de votos dos candidatos ou abstenções ou brancos/nulos nas JRV's amostradas no estrato i.

b) $\mathbf{y} = (y_{\cdot A}, y_{\cdot B}, y_{\cdot C}, y_{\cdot D}, y_{\cdot E}) = \sum_{i=1}^{14} \mathbf{y}_i$ é o vetor dos totais amostrais.

c) $x_i = \sum_{j=1}^{n_i} X_{ij}$ é o total de eleitores amostrados no estrato i.

d) $x = \sum_{i=1}^{14} x_i$ é o total de eleitores amostrados.

Por simplicidade usaremos na seqüência o índice k para indicar um elemento genérico do conjunto $\{A, B, C, D, E\}$. Por exemplo, Y_{ijk} representa um componente genérico do vetor

$$(Y_{ijA}, Y_{ijB}, Y_{ijC}, Y_{ijD}, Y_{ijE}).$$

Analogamente, definiríamos os elementos Y_k, $Y_{i\cdot k}, y_{ik}$ e $y_{\cdot k}$.

A2 - Método Clássico de Amostragem por Conglomerados utilizando o Estimador Razão

Admitimos neste caso que a seleção das n_i JRV's do estrato i foi feita ao acaso e sem reposição, ou seja, que cada JRV do estrato i tem probabilidade $\frac{n_i}{N_i}$ de ser selecionada.

Para qualquer $k \in \{A, B, C, D, E,\}$ $r_{ik} = \frac{y_{ik}}{x_i}$ é a proporção amostral de votos favoráveis à categoria k no estrato i. Assim, o previsor do total de votos da categoria k no estrato i é

$$\widehat{Y}_{i\cdot k} = r_{ik} X_{i\cdot}$$

e o previsor do total de votos da categoria k na população é simplesmente

$$\widehat{Y}_k = \sum_{i=1}^{14} \widehat{Y}_{i\cdot k}.$$

As diferenças entre os totais de votos correspondentes às diferentes categorias podem ser estimadas por

$$\Delta_{i,k,k'} = \widehat{Y}_{i\cdot k} - \widehat{Y}_{i\cdot k'} \quad \text{e} \quad \Delta_{\cdot k,k'} = \widehat{Y}_k - \widehat{Y}_{k'},$$

$k, k' \in \{A, B, C, D, E\}$.

Para os cálculos dos intervalos de previsão, as seguintes variâncias e covariâncias são consideradas:

a) $\mathrm{Var}\{\widehat{Y}_{i\cdot k}\} = X_{i\cdot}^2 \mathrm{Var}\{r_{ik}\} \cong$

$$\cong X_{i\cdot}^2 \frac{1}{x_i^2}[\mathrm{Var}\{y_{ik}\} + R_{ik}^2 \mathrm{Var}\{x_i\} - 2R_{ik}\mathrm{Cov}\{y_{ik}, x_i\}].$$

onde $R_{ik} = \frac{Y_{i\cdot k}}{X_i}$.

Esta variância pode ser estimada por

$$\mathrm{var}\{\widehat{Y}_{i\cdot k}\} = (\frac{X_{i\cdot}}{x_i})^2 n_i(1 - \frac{n_i}{N_i})[\mathrm{var}\{Y_{ijk}\} + \\ + r_{ik}^2 \mathrm{var}\{X_{ij}\} - 2\mathrm{cov}\{X_{ij}, Y_{ijk}\}]$$

onde

$$\text{var}\{Y_{ijk}\} = \frac{1}{n_i - 1}[\sum_{j=1}^{n_i} Y_{ijk}^2 - \frac{y_{ik}^2}{n_i}]$$

$$\text{var}\{X_{ij}\} = \frac{1}{n_i - 1}[\sum_{j=1}^{n_i} X_{ij}^2 - \frac{x_i^2}{N_i}]$$

$$\text{cov}\{X_{ij}, Y_{ijk}\} = \frac{1}{n_i - 1}[\sum_{j=1}^{n_i} X_{ij} Y_{ijk} - \frac{x_i y_{ik}}{n_i}]$$

b) $\text{Cov}\{\widehat{Y}_{i\cdot k}, \widehat{Y}_{i\cdot k'}\} \cong$

$$\cong (\frac{X_{i\cdot}}{x_i})^2[\text{Cov}(y_{ik}, y_{ik'}) + r_{ik} r_{ik'} \text{Var}(x_i) - r_{ik}\text{Cov}(y_{ik'}, x_i) - r_{ik'}\text{Cov}(y_{ik}, x_i)].$$

Esta covariância pode ser estimada por

$$\text{Cov}\{\widehat{Y}_{i\cdot k}, \widehat{Y}_{i\cdot k'}\} = (\frac{X_{i\cdot}}{x_i})^2 n_i (1 - \frac{n_i}{N_i})[\text{cov}\{Y_{i\cdot k}, Y_{i\cdot k'}\} +$$
$$+ r_{ik} r_{ik'} \text{var}\{X_{ij}\} - r_{ik}\text{cov}\{Y_{ijk'}, X_{ij}\} - r_{ik'}\text{cov}\{Y_{ijk}, X_{ij}\}]$$

onde

$$\text{cov}\{Y_{ijk}, Y_{ijk'}\} = \frac{1}{n_i - 1}[\sum_{j=1}^{n_i} Y_{ijk} Y_{ijk'} - \frac{y_{ik} y_{ik'}}{n_i}]$$

c) $\text{Var}\{\widehat{Y}_k\} = \sum_{i=1}^{14} \text{Var}\{\widehat{Y}_{i\cdot k}\}$ e

$\text{Var}\{\widehat{Y}_k - \widehat{Y}_{k'}\} = \text{Var}\{\widehat{Y}_k\} + \text{Var}\{\widehat{Y}_{k'}\} - 2\text{Cov}\{\widehat{Y}_k, \widehat{Y}_{k'}\}$

onde

$\text{Cov}\{\widehat{Y}_k, \widehat{Y}_{k'}\} = \sum_{i=1}^{14} \text{Cov}\{Y_{ijk}, Y_{ijk'}\}$.

Estimativas de $\text{Var}\{\widehat{Y}_k\}$ e de $\text{Var}\{\widehat{Y}_k - \widehat{Y}_{k'}\}$ são obtidas através da substituição de $\text{Var}\{\widehat{Y}_k\}$, e $\text{Cov}\{Y_{ijk}, Y_{ijk'}\}$ pelos seus correspondentes amostrais nas fórmulas acima.

A3 - Modelo de Superpopulação utilizando o Estimador Regressão

O segundo método utilizado envolveu um modelo de superpopulação. Neste caso admitimos que a cada JRV está associado um vetor aleatório $\mathbf{Y}_{ij}$ com uma distribuição de probabilidades que define o modelo de superpopulação. Admitimos também que os vetores $\mathbf{Y}_{ij}$ são independentes e dentro de cada estrato i são identicamente distribuídos.

Com esta formulação fica claro que a probabilidade de seleção das JRV's não é relevante para a construção do previsor.

Neste trabalho, para cada JRV (i,j) e $k \in \{A,B,C,D,E,\}$ consideramos o modelo

$$Y_{ijk} = \beta_{ik} X_{ij} + \varepsilon_{ijk}$$

onde β_{ik} representa a proporção esperada de votos na categoria k no estrato i, e os ε_{ijk} são tais que os vetores aleatórios

$$\mathbf{E}_{ij} = \{\varepsilon_{ijA}, \varepsilon_{ijB}, \varepsilon_{ijC}, \varepsilon_{ijD}, \varepsilon_{ijE},\}$$

são independentes e, dentro de cada estrato i, identicamente distribuídos com média zero e matriz de covariância

$$\mathrm{Var}\{\mathbf{E}_{ij}\} = \frac{X_{ij}(X_{ij}+100)}{10100} \begin{bmatrix} 4.75 & -.25 & -2 & -2 & -.5 \\ -.25 & 4.75 & -2 & -2 & -.5 \\ -2 & -2 & 24 & -16 & -4 \\ -2 & -2 & -16 & 24 & -4 \\ -.5 & -.5 & -4 & -4 & 9 \end{bmatrix}$$

Esta matriz é singular porque $Y_{ijA} + Y_{ijB} + Y_{ijC} + Y_{ijD} + Y_{ijE} = X_{ij}$ é conhecido para todo (i,j). Essencialmente, esta matriz traduz a opinião *a priori* de que a amplitude do intervalo de variação da proporção de votos esperada pela categoria k no estrato i, β_{ik}, não deveria ser maior que 0,3. Completando a opinião *a priori*, esperava-se que o vetor $(\beta_{iA}, \beta_{iB}, \beta_{iC}, \beta_{iD}, \beta_{iE},)$ não diferisse muito do vetor (0,05; 0,05; 0,4; 0,4; 0,1). Note que a matriz acima é a matriz de covariância de uma mistura obtida da composição de uma multinomial (com tamanho de amostra X_{ij}) com uma Dirichlet de ordem 4 e parâmetro (5; 5; 40; 40; 10).

Denotando a proporção de votos da categoria k na JRV (i,j) como $r_{ijk} = \frac{Y_{ijk}}{X_{ij}}$ e usando a notação

$$x_{ij} = \frac{X_{ij}}{X_{ij}+100},$$

o estimador de mínimos quadrados ponderados de β_{ik} é dado por

$$\widehat{\beta}_{ij} = \frac{\sum_{j=1}^{n_i} x_{ij} r_{ijk}}{\sum_{j=1}^{n_i} x_{ij}} .$$

(Não é difícil provar que $\widehat{\beta}_{ik}$ é um estimador não viciado de β_{ik}.)

Assim, o estimador (previsor) regressão do total de votos associados à categoria k no estrato i é dado por

$$\widetilde{Y}_{i\cdot k} = \sum_{j=1}^{n_i} Y_{ij} + \widehat{\beta}_{ik} \sum_{j=n_i+1}^{N_i} X_{ij}.$$

Para o total geral o estimador correspondente é dado por

$$\widetilde{Y}_k = \sum_{i=1}^{14} \widetilde{Y}_{i\cdot k}.$$

Consideramos aqui apenas a diferença entre os votos dos candidatos principais; isto é,

$$\widetilde{Y}_{i\cdot D} - \widetilde{Y}_{i\cdot C} \quad \text{no estrato } i$$

e

$$\widetilde{Y}_D - \widetilde{Y}_C \quad \text{na população.}$$

Para os cálculos dos intervalos de previsão e para a avaliação da qualidade de previsão, as variâncias e erros médios quadráticos apresentados na seqüência são relevantes. Para obter essas fórmulas, usamos a independência dos vetores $\mathbf{Y}_{ij}$ e a matriz de covariância de $\mathbf{E}_{ij}$, que é conhecida. Abaixo, por simplicidade, listamos a notação usada na descrição de fórmulas:

$$S_{1i} = \sum_{j=1}^{n_i} X_{ij}(X_{ij}+100), \qquad S_{2i} = \sum_{j=n_i+1}^{N_i} X_{ij}(X_{ij}+100),$$

$$S_{3i} = \sum_{j=1}^{n_i} x_{ij} = \sum_{j=1}^{n_i} \frac{X_{ij}}{X_{ij}+100}, \qquad S_{4i} = \sum_{j=1}^{n_i} X_{ij}^2,$$

$$S_{5i} = \sum_{j=1}^{n_i} X_{ij}, \qquad S_{6i} = X_{i\cdot} - S_{5i},$$

$$(K_A, K_B, K_C, K_D, K_E) = (4.75;\ 4.75;\ 24;\ 24;\ 9).$$

a) $\mathrm{Var}\{\widehat{\beta}_{ik}\} = \frac{K_k}{10100} S_{3i}^{-1}$

b) Erro médio quadrático:

$E\{(\widetilde{Y}_{i\cdot k} - Y_{i\cdot k})^2\} = \frac{K_k}{10100}[\frac{S_{6i}}{S_{3i}} + S_{2i}]$

c) $\mathrm{Var}\{\widetilde{Y}_{i\cdot k}\} = \frac{K_k}{10100}\{S_{1i} + \frac{S_{6i}}{S_{8i}}[S_{6i} + 2S_{5i}]\}$

d) Erro médio quadrático da diferença:

$$E\{[(\widetilde{Y}_{i\cdot D} - \widetilde{Y}_{i\cdot C}) - (Y_{i\cdot D} - Y_{i\cdot C})]^2\} = \frac{80}{10100}[\frac{S_{6i}^2}{S_{8i}} + S_{2i}]$$

e) $\mathrm{Var}\{\widetilde{Y}_{i\cdot D} - \widetilde{Y}_{i\cdot C}\} = \frac{80}{10100}\{S_{1i} + \frac{S_{6i}^2}{S_{8i}}[S_{6i} + 2S_{5i}]\}$.

As variâncias dos totais são obtidas através das somas das variâncias de cada estrato.

As tabelas que apresentamos a seguir descrevem os resultados obtidos pelos dois métodos, baseados na informação de 48% da amostra.

Tabela A1: Número previsto de votos

Amostragem por conglomerados utilizando o estimador razão

	UNO	FSLN	FSLN-UNO	Outros	Brancos/Nulos	Abstenções
Mínimo	749 803	538 840	-308 280	57 041	61 047	192 079
Médio	802 089	584 168	-217 920	71 181	70 107	224 544
Máximo	854 374	629 497	-127 561	85 320	79 166	257 009

Tabela A2: Número previsto de votos

Modelo de superpopulação utilizando o estimador regressão

	UNO	FSLN	FSLN-UNO	Outros	Brancos/Nulos	Abstenções
Mínimo	790 318	553 794	-278 704	56 690	60 332	210 045
Médio	813 421	576 897	-236 523	70 838	70 610	220 323
Máximo	836 524	600 000	-194 343	84 985	80 888	230 601

Bibliografia

Cassel, C.; Särndall, C. & Wretman, J.H. (1977). *Foundations of Inference in Survey Sampling*. J. Wiley, New York, 192 p.

Cochran, W.G. (1977). *Sampling Techniques*. (3rd. Edition). J. Wiley, New York, 428 p.

Kish, L. (1965). *Survey Sampling*. J. Wiley, New York, 639 p.

Royall, R.M. (1968). *An old approach to finite population sampling theory*. *JASA 63*. 1269-79.