INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA — IBGE

ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS - ENCE

RELATÓRIOS TÉCNICOS

Nº 04/89

UMA PROPOSTA DE DIMENSIONAMENTO DA AMOSTRA DO CENSO DEMOGRÁFICO DE 1990

José Carlos da Rocha C. Pinheiro - ENCE/IBGE

José Matias de Lima - ENCE/IBGE

RIO DE JANEIRO

Dezembro/89

ÌNDICE

1 - INTRODUÇÃO

2 - METODOLOGIA

2.1 - A Utilidade das Informações Coletadas

2.2 - A Desigualdade de Utilidade

2.3 - A Utilização de Frações de Amostragem Múltiplas - Partição Ótima da População

2.4 - Impactos do Uso de Frações Múltiplas sobre a Operacionalidade

3 - APLICAÇÃO AO CENSO DEMOGRÁFICO - 90

3.1 - Dados Utilizados

3.2 - Cálculo das Estimativas do Erro Amostral

3.3 - Funções Peso Consideradas

3.4 - Frações Amostrais - Abordagens de Dimensionamento Utilizadas

3.5 - Procedimentos de Modelagem

3.6 - Propostas de Dimensionamento para o CD-90

3.7 - Avaliação das Propostas de Dimensionamento

4 - CONCLUSÕES - A ESPECIFICAÇÃO DA PROPOSTA DE DIMENSIONAMENTO

5 - REFERÊNCIAS BIBLIOGRÁFICAS

6 - ANEXOS

6.1 - Variáveis Consideradas no Estudo e Estatísticas Resumo Adotadas

6.2 - Municípios Considerados no Estudo e População, segundo o Censo Demográfico de 1980

6.3 - Figuras Apresentadas no Texto

1 - INTRODUÇÃO

Avalia-se, no momento, a possibilidade de redução da fração de amostragem utilizada no Censo Demográfico (CD) do IBGE, passando dos tradicionais 25% (empregados desde a adoção de procedimentos de amostragem no censo) para 10%. O argumento seria de que, com o crescimento verificado pela população brasileira nas últimas décadas, não se justificaria manter uma fração amostral de 25%, julgada conveniente para o CD de 1960, quando foram introduzidos os procedimentos amostrais. A motivação óbvia da redução da fração é a economia de recursos que dela adviria, existindo ainda outras razões menos evidentes, tais como a simplificação dos procedimentos de coleta e expansão da amostra.

Contudo, a redução da fração de amostragem não se revelou, até o momento, um consenso entre técnicos do IBGE e usuários das informações do Censo. Argumentam, os contrários à medida, que a utilização da fração de 10% inviabilizaria o uso das informações da amostra para municípios pequenos e mesmo subdivisões de municípios médios ou grandes, pois teria impactos desastrosos sobre a precisão das estimativas.

A polêmica em torno da mudança da fração amostral origina-se no fato de que o Censo Demográfico é um levantamento de objetivos múltiplos, muitos dos quais nem mesmo podem ser previstos a priori, e cujos níveis de divulgação são extremamente variados, indo desde municípios de poucas centenas de habitantes a metrópoles do porte de São Paulo ou Rio de Janeiro. Uma fração única de amostragem certamente se revelerá inapropriada (ao menos) para um dos extremos do espectro de divulgação: ou insuficiente para os municípios pequenos, ou excessiva para os municípios grandes (ou, provavelmente, as duas coisas simultaneamente).

A tomada de decisão sobre este tema fundamental deve transcender pontos de vista particulares, de natureza subjetiva e calcados essencialmente na experiência pessoal. A questão deve ser colocada sob uma ótica objetiva, onde os diversos ganhos e contra-partidas sejam bem entendidos e preferencialmente quantificados. Certamente a decisão final deverá atentar tanto para o lado técnico, quanto para o gerencial, não devendo, entretanto, priorizar apenas um deles, em detrimento do outro.

Esse trabalho tem por objetivos: **(i)** apresentar uma análise detalhada dos possíveis ganhos e perdas associados à modificação do plano amostral do CD-80 e **(ii)** desenvolver uma proposta técnica de dimensionamento da(s) fração(ões) amostral(is) para o próximo

censo. Pretende-se, com ele, colocar a questão técnica referente à mudança da fração amostral dentro do contexto apropriado para a tomada de decisão por parte do corpo gerencial do IBGE.

O estudo foi desenvolvido sempre levando em conta aqueles que, no entender dos autores, deveriam ser os requisitos básicos de um plano amostral para o CD-90, a saber:

i) garantia de utilidade para as informações coletados, traduzida em termos da precisão mínima com que características de interesse seriam estimadas a partir delas;

ii) relativa homogeneidade de utilidade das informações para as diferentes áreas de divulgação das mesmas;

iii) operacionalidade dos procedimentos de coleta, crítica, tabulação e divulgação dos dados;

iv) custo mínimo, respeitadas as restrições quanto à utilidade, homogeneidade e operacionalidade.

Na seção 2 é descrita a metodologia utilizada no desenvolvimento do estudo. Os aspectos excessivamente técnicos dos resultados ali apresentados são omitidos, de modo a não comprometer, desnecessariamente, o entendimento por parte daqueles menos familiarizados com as técnicas estatísticas.

Apesar de elaborada para uso no dimensionamento da amostra para o CD-90, a metodologia é de caráter genérico, podendo ser facilmente adaptada a outras aplicações. A apresentação do desenvolvimento teórico da metodologia é feito respeitando a generalidade da mesma, de modo a facilitar sua utilização em outras situações.

Na seção 3 é apresentada uma aplicação da metodologia à situação particular do dimensionamento da amostra do CD-90. Faz-se uma análise detalhada dos possíveis impactos da modificação do procedimento amostral tradicional (fração de 25%), levando-se em conta diferentes possibilidades (manutenção da fração única, utilização de frações múltiplas, etc.) As propostas de dimensionamento são avaliados à luz dos quatro requisitos básicos mencionados anteriormente, buscando-se aquela que melhor os atenda de modo simultâneo.

Finalmente, na seção 4 é sugerida uma proposta de dimensionamento para o próximo Censo Demográfico, obtida como resultado do estudo descrito na seção anterior.

2 - METODOLOGIA

Nesta seção é apresentado o ferramental teórico utilizado no desenvolvimento da proposta de dimensionamento da amostra para o CD-90. Os resultados foram organizados de acordo com os requisitos básicos mencionados na introdução (utilidade, homogeneidade, operacionalidade e custos).

No item 2.1 descrevem-se medidas utilizadas para avaliar os impactos de modificações efetuadas sobre um plano amostral de referência (no caso do CD, a fração amostral única de 25%), com relação à utilidade das informações coletadas. Incluem-se, neste item, três propostas de dimensionamento de frações amostrais, que levam em conta diferentes aspectos da utilidade dos dados.

No item 2.2 aborda-se a questão da homogeneidade da utilidade das informações, considerando-se um dado nível de divulgação para os resultados (município, no caso do CD-90). Medidas de desigualdade de precisão, análogas ao índice de Ginni, utilizado na economia, são apresentadas.

No item 2.3 analisa-se a questão do dimensionamento da amostra sob a ótica do custo. São sugeridos procedimentos para o cálculo de frações e limites ótimos, levando-se em conta a minimização do tamanho global da amostra.

Por último, o item 2.4 trata, de maneira breve, dos possíveis impactos causados pela utilização de múltiplas frações de amostragem sobre a operacionalidade dos procedimentos de coleta, processamento e expansão da amostra.

2.1 - A Utilidade das Informações Coletadas

A premissa básica por detrás do desenvolvimento teórico aqui apresentado é a de que existe um plano amostral de referência, utilizado anteriormente, o qual se pretende modificar. Existem duas questões básicas a considerar:

i) quais os impactos de uma nova proposta sobre a utilidade dos dados levantados, tomando-se por base o plano amostral de referência?

ii) qual a garantia de utilidade das informações levantadas segundo o novo plano amostral?

Em (i) o que se pretende é medir a variação relativa de utilidade, quando se passa de uma proposta para outra, buscando-se, geralmente, manter uma possível perda dentro de limites toleráveis. Está implícita, neste caso, a adequação do plano de referência, cuidando-se apenas para que uma nova proposta, provavelmente mais econômica, não cause perdas excessivas com relação à situação original.

A segunda questão é mais abrangente, abrindo a possibilidade de reavaliar-se o próprio plano de referência, à luz da utilidade das informações coletadas. O plano amostral de referência não é mais tratado como uma solução adequada, da qual não se deseja afastar muito, sendo também submetido à avaliação de sua eficácia.

As duas questões supracitadas traduzem-se em enfoques distintos para a medição dos impactos sobre a utilidade dos dados. Um primeiro, associado à noção de (perda de) **Eficiência Relativa**, tratada na literatura clássica de inferência estatística, e outro no qual se procura pré-fixar um nível mínimo de precisão para as estimativas a serem obtidas a partir da amostra, como tradicionalmente é feito em **Amostragem**. Os dois enfoques são apresentados a seguir.

2.1.1. Eficiência Relativa

Classicamente, a eficiência relativa entre dois procedimentos de estimação de um determinado parâmetro é definida como a razão entre os erros quadráticos médios (EQM) correspondentes. Pinheiro e Lima (1988) apontam algumas deficiências desta medida para o contexto deste estudo:

> i) está restrita a uma única característica, quando o que se deseja é avaliar o efeito da mudança do procedimento de estimação para o conjunto de variáveis investigadas; e
>
> ii) não leva em conta a precisão absoluta dos procedimentos de estimação, dando, por exemplo, igual importância a um aumento do coeficiente de variação (CV) de 0,5% para 1%, quanto de 10% para 20% (considerando-se procedimentos não-viciados).

No mesmo trabalho, Pinheiro e Lima introduzem uma nova medida de eficiência relativa para o caso em que os dois procedimentos de estimação diferem entre si apenas pelas

frações amostrais (supostas únicas) utilizadas. Os principais resultados daquele trabalho são reproduzidos a seguir.

Suponha que, em uma pesquisa, existam r características de interesse, denotadas por $\theta_1, \cdots, \theta_r$ e que fixou-se um limite de tolerância c, acima do qual o CV de um determinado estimador $\hat{\theta}_i$ é considerado inaceitável. Seja

$$I_c(\hat{\theta}_i) = \begin{cases} 1, & \text{se} \quad CV(\hat{\theta}_i) > c \\ 0, & \text{se} \quad CV(\hat{\theta}_i) \leq c \end{cases} \qquad i = 1, \cdots, r \tag{1}$$

representando por $\hat{\theta}(f)$ o estimador da característica θ utilizando-se uma fração amostral f, define-se a **Perda de Eficiência Relativa** entre duas frações f_1 e f_2 como se segue:

$$P_c(f_1, f_2) = \sum_{i=1}^{r} |I_c\left(\hat{\theta}_i(f_1)\right) - I_c\left(\hat{\theta}_i(f_2)\right)| / r \tag{2}$$

A medida proposta corrige as duas deficiências mencionadas sobre a medida clássica (para maiores detalhes, consultar o trabalho mencionado). Pinheiro e Lima (1988) introduzem, ainda, o conceito de **Perda de Eficiência Absoluta**, associada à uma fração amostral f:

$$P_c(f) = \sum_{i=1}^{r} I_c\left(\hat{\theta}_i(f)\right) / r = P_c(1, f) \tag{3}$$

Note-se que:

$$P_c(f_1, f_2) = |P_c(f_1) - P_c(f_2)| \tag{4}$$

Adotando-se a convenção de que $f_1 \geq f_2$, o módulo pode ser omitido das expressões (2) e (4).

O principal problema associado ao uso da perda de eficiência relativa, reside na escolha do limite de tolerância c, que é variável de acordo com o usuário dos dados e a aplicação à qual eles se destinam. O problema pode ser parcialmente contornado considerando-se $P_c(f_1, f_2)$ e $P_c(f)$ como funções de c (as chamadas **curvas de perda de eficiência**

relativa e absoluta, respectivamente) e analisando o seu comportamento para valores de c dentro de uma faixa de variação razoável.

Representando-se o coeficiente de variação de $\theta_i(f)$ por $CV_i(f)$, ordenando-se estes em ordem crescente e denotando-se por $CV_{(i)}(f)$ o i-ésimo elemento de tal ordenação e por $I_c\left(\hat{\theta}_{(i)}(f)\right)$ a função indicadora (como definida em (1)) correspondente,tem-se:

$$P_c(f) = \sum_{i=1}^{r} I_c\left(\hat{\theta}_{(i)}(f)\right) \tag{5}$$

Observa-se, assim, que as funções $P(c, f)$ e $P(c, f_1, f_2)$ para f, f_1, e f_2 fixos são do tipo escada, com saltos nos pontos $CV_{(i)}(f)$, $CV_{(i)}(f_1)$ e $CV_{(i)}(f_2)$, respectivamente. As figuras 1, 2 e 3, em anexo, mostram exemplos típicos destas curvas.

A necessidade de considerar a perda de eficiência relativa (ou absoluta) como uma função do limite de tolerância constitui-se numa clara desvantagem desta medida, com relação à clássica, já que não se dispõe de um valor que sintetize a informação associada à medida, dificultando, por exemplo, a comparação direta entre frações alternativas.

Ocorre, entretanto que, se a definição prévia de um limite de tolerância constitui-se em tarefa por demais subjetiva e dependente do usuário e da aplicação específica, o estabelecimento de uma faixa de variação razoável para este limite é algo que pode ser colocado de maneira objetiva e geral. Por exemplo, dificilmente existirá uma situação (ao menos no contexto em que este estudo se desenvolveu, qual seja, dimensionamento de amostras para pesquisas de grande porte) em que se exija um CV inferior à 2%, ou se tolere um CV superior à 50%, para a estimação de uma determinada característica de interesse. Prosseguindo-se nesta linha de raciocínio, poder-se-ia atribuir uma ponderação diferenciada aos limites de tolerância compreendidos entre o mínimo exígivel e o máximo tolerável, em um exercício semelhante à especificação de **prioris** em estatística bayesiana, dando-se maior importância aos valores comumente utilizados em pesquisas por amostragem (por exemplo, 10%) e menor ênfase aos valores pouco usuais (por exemplo, 3% ou 40%). Uma função assim obtida será denominada, neste trabalho, **função peso**.

Seja w uma função real contínua. w será dita uma função peso se respeitar às duas seguintes condições:

$$i) w(c) \geq 0, \quad \forall_c \in \mathbb{R}$$

$$ii) \int_{\mathbb{R}} w(c)dc = 1 \tag{6}$$

Em outras palavras, w é uma densidade de probabilidade.

As figuras 4, 5 e 6, em anexo, apresentam exemplos de funções peso que podem ser consideradas no contexto aqui apresentado.

Observe que, na primeira função peso (fig.4) não é feita diferenciação quanto à importância dos limites de tolerância no intervalo. A terceira função (fig.6) é obtida a partir da **família de distribuições beta**, e abrange as funções do primeiro tipo, mas não as do segundo (fig.5). Neste trabalho, optou-se por utilizar funções peso do terceiro tipo mencionado, por sua grande maleabilidade e maior adequação aos objectivos do estudo.

O estabelecimento de uma função peso para os limites de tolerância permite a derivação de uma medida que sintetize a informação contida nas curvas de perda de eficiência ($P(c, f)$ ou $P(c, f_1, f_2)$), definida a seguir.

Seja w uma função peso e $P(c, f)$ uma curva de perda de eficiência absoluta, associada a uma fração f. A perda média de eficiência absoluta, ponderada pela função w, é dada por:

$$P_w(f) = \int_{\mathbb{R}} P(c, f)w(c)dc \tag{7}$$

De modo análogo, define-se a **perda média de eficiência relativa** associada a uma curva $P(c, f_1, f_2)$ e função peso w por:

$$P_w(f_1, f_2) = \int_{\mathbb{R}} P(c, f_1, f_2)w(c)dc = |P_w(f_1) - P_w(f_2)| \tag{8}$$

As figuras 7 e 8, em anexo, mostram as curvas de perda de eficiência das figuras 1 e 3, ponderadas pela função peso da figura 6.

A partir de (5) pode-se obter uma expressão bastante simples para a perda média de eficiência, apresentada a seguir.

Resultado 1

$$P_w(f) = \sum_{i=1}^{r} W\left(CV_{(i)}(f)\right)/r$$

$$P_w(f_1, f_2) = \sum_{i=1}^{r} |W\left(CV_{(i)}(f_1)\right) - W\left(CV_{(i)}(f_2)\right)|/r$$

onde $W(t) = \int_{-\infty}^{t} w(c)dc$

Um caso particular de grande interesse para o estudo desenvolvido é quando o método de seleção adotado na pesquisa é o de Amostragem Aleatória Simples Sem Reposição (AASSR). Pinheiro e Lima (1988) demonstram que, neste caso, tem-se o seguinte resultado

Resultado 2

$$CV\left(\hat{\theta}(f)\right) = \left(\frac{f_0(1-f)}{f(1-f_0)}\right)^{1/2} CV\left(\hat{\theta}(f_0)\right) \qquad \text{e}$$

$$CV_{(i)}(f) = \left(\frac{f_0(1-f)}{f(1-f_0)}\right)^{1/2} CV_{(i)}(f_0), \qquad i = 1, \cdots, r$$

com f_0 representando a fração amostral adotada no plano amostral de referência.

Desta forma, segue-se que, sob AASSR, vale o resultado apresentado a seguir

Resultado 3

$$P_w(f) = \sum_{i=1}^{r} W\left(\left(\frac{f_0(1-f)}{f(1-f_0)}\right)^{1/2} CV_{(i)}(f_0)\right)/r$$

$$P'_w(f) = \frac{dP_w(f)}{df} = -\frac{1}{2}\left(\frac{f_0}{f^3(1-f)(1-f_0)}\right)^{1/2} \times$$

$$\sum_{i=1}^{r} \frac{CV_{(i)}(f_0) w\left(\left(\frac{f_0(1-f)}{f(1-f_0)}\right)^{1/2} CV_{(i)}(f_0)\right)}{r}$$

$$P_w(0) = 1 \quad \text{e} \quad P_w(1) = 0$$

Assim, $P_w(f)$ considerada como função (contínua) de f, é estritamente decrescente no intervalo onde a f é definida e limitada (varia entre 0 e 1). A figura 9,em anexo, mostra a forma típica da função $P_w(f)$.

A perda média de eficiência pode ser utilizada na obtenção de duas estratégias de dimensionamento de frações amostrais, descritas a seguir.

A fração mínima necessária para garantir uma perda média de eficiência máxima de $\alpha \times 100\%$, com relação a uma fração de referência f_0, é definida como:

$$f_{1,w}(f_0, \alpha) = \inf\{f / P_w(f, f_0) \leq \alpha\}, \quad \alpha \in [0, 1] \tag{9}$$

Esta abordagem de dimensionamento enquadra-se dentro do primeiro enfoque mencionado no início de 2.1, ou seja, procura controlar a perda de eficiência da nova proposta, com relação ao plano amostral de referência.

Resultado 4

$$f_{1,w}(f_0, 0) = f_0; \; f_{1,w}(f_0, 1) = 0$$
$$f_{1,w}(f_0, \alpha) = \inf\{f / P_w(f) \leq \alpha + P_w(f_0)\} \leq f_0, \quad \alpha \in [0, \; 1]$$
$$f_{1,w}(f_0, \alpha) = P_w^{-1}(\alpha + P_w(f_0)), \quad \alpha \in [0, 1 - P_w(f_0)]$$

com P_w^{-1} representando a função inversa de P_w

A outra abordagem para dimensionamento da fração amostral, utilizando a perda média de eficiência, baseia-se em uma estrutura de custos para a pesquisa, que se divide em duas partes: uma associada à seleção das unidades amostrais e outra vinculada à perda de utilidade dos dados coletados para os usuários finais. Obviamente estes dois custos atuam em direções opostas, e a **fração adequada**, nesta abordagem, será a que minimiza o custo total.

Resultado 5

Seja c_1 o custo de seleção associado à variação de um ponto percentual na fração de amostragem f e c_2 o custo de utilidade, associado à variação de um ponto percentual na perda média da eficiência $P_w(f)$. Assumindo-se uma função de custo total linear, tem-se

$$f_{2,w}(c_1/c_2) = \{f/P'_w(f) = -c_1/c_2\}$$

O custo total da pesquisa, como função da fração f, é neste caso dado p

$$c_t(f) = c_1 \times f + c_2 \times P_w(f) \tag{10}$$

Esta abordagem de dimensionamento situa-se dentro do segundo enfoque citado no início de 2.1, pois não parte da hipótese de que o plano de referência é adequado. Obviamente a dificuldade maior quanto à sua utilização é o estabelecimento da relação entre os custos de seleção e utilidade (note-se que não é necessário o conhecimento individual dos custos c_1 e c_2, mas apenas da razão c_1/c_2). Contudo, esta abordagem revelou-se particularmente atraente, quando aplicada à situação prática do CD-90, como se verá na seção 3.

2.1.2. A Fração Amostral Necessária

A definição da fração amostral necessária pressupõe, de modo análogo à perda de eficiência, a especificação de um limite de tolerância para o CV de um estimador e de um percentual mínimo de características que devem ser estimadas com precisão obedecendo ao limite pré-estabelecido. Pinheiro e Lima (1988) apresentam uma formalização do conceito reproduzida a seguir.

Seja c um limite de tolerância e α o percentual mínimo de características de interesse que devem ser estimadas com CV menor ou igual a c. Então a **fração necessária** associada é dada por:

$$f(c,\alpha) = \inf\{f/P_c(f) \leq 1-\alpha\} \tag{11}$$

Pinheiro e Lima (1988), no mesmo artigo, derivam uma expressão exata para a fração necessária, quando o esquema de seleção é o de AASSR e o plano amostral de referência utiliza uma fração f_0.

Resultado 6

Sob AASSR e considerando-se um plano amostral de referência com fração f_0 tem-se:

$$f(c,\alpha) = \left((\frac{C}{CV_{(t)}(f_0)})^2 \frac{(1-f_0)}{f_0} + 1 \right)^{-1}$$

onde

$$t = \begin{cases} r \cdot \alpha, & \text{se} \quad r \cdot \alpha \in \mathbb{N} \\ [r \cdot \alpha] + 1, & \text{caso contrário} \end{cases}$$

A figura 10, em anexo, ilustra o comportamento típico da fração necessária, tomada como função do limite de tolerância c.

A dependência da fração necessária com respeito ao limite de tolerância c, faz com que esta medida esteja sujeita aos mesmos inconvenientes que as perdas de eficiência relativa e absoluta. Nada mais natural, portanto, que adotar a mesma solução encontrada para aqueles casos, ou seja, estabelecer uma função peso para os limites de tolerância (que, em geral, será diferente da utilizada para as perdas de eficiência) e calcular a média ponderada das frações necessárias, segundo esta função peso.

A fração necessária média, ponderada por uma função peso w, associada à um percentual mínimo de características de interesse α é dada por:

$$f_{3,w}(\alpha) = \int_{\mathbf{R}} f(c,\alpha) w(c) dc \tag{12}$$

Nas condições do resultado 6 tem-se o seguinte resultado

Resultado 7

Sob AASSR e considerando-se um plano amostral de referência com fração f_0, tem-se

$$f_{3,w}(\alpha) = \int_{\mathbf{R}} \frac{w(c)}{\left((\frac{C}{CV_{(t)}(f_0)})^2 \frac{(1-f_0)}{f_0} + 1 \right)} dc$$

A exemplo da fração adequada ($f_{2,w}$), a fração necessária média enquadra-se dentro da segunda abordagem de dimensionamento definida no início de 2.1, pois admite a obtenção de frações amostrais superiores àquela adotada no plano amostral de referência.

2.2 - A Desigualdade de Utilidade

Geralmente em pesquisas de grande porte, como o Censo Demográfico, estimativas de características de interesse devem ser divulgadas segundo diferentes estratos populacionais (por exemplo níveis geográficos distintos, classes sócio-econômicas, sexo, etc). Mantida uma fração amostral única, provavelmente haverá uma considerável variação de precisão das estimativas através dos diferentes estratos de divulgação.

Em geral, a precisão variará na razão direta das populações dos estratos. Existe, portanto, uma **desigualdade de utilidade** para os dados levantados, traduzida na variação da precisão das estimativas.

O propósito deste item é apresentar procedimentos para mensuração da desigualdade de precisão associada a um plano amostral, segundo uma determinada partição da população total de indivíduos investigados. No estudo efetuado sobre o dimensionamento da amostra para o CD-90, o interesse recaiu sobre a estratificação da população por nível geográfico, mais especificamente por município. Os resultados aqui apresentados são adaptações diretas dos procedimentos utilizados na economia para medir a desigualdade de renda de uma população.

Torna-se necessário, neste ponto, introduzir alguma notação adicional àquela já apresentada em 2.1, visto que foi incluída a dimensão adicional da subdivisão da população em estratos. Suponha que a população total seja particionada em p estratos. Então, define-se:

$$CV_{ij} = CV(\hat{\theta}_i) \quad \text{no j-ésimo estrato,} \quad i = 1, \cdots, r_j, \quad j = 1, \cdots, p$$

$$CV_{(i)j} = \text{ i-ésimo CV de } \hat{\theta} \text{ ordenado no estrato } j, i = 1, \cdots, r_j, \quad j = 1, \cdots, p$$

$$P_{wj}(f) = \text{ perda média de eficiência absoluta associada à fração f, no estrato } j$$

(Note-se que é admitida a possibilidade de variação do número de características de interesse por estrato).

Observe-se que, ao contrário da economia, onde apenas uma variável (a renda) está sendo considerada para fins de medição da desigualdade, têm-se aqui diversas medidas de precisão associadas a um mesmo estrato (os CV's das r_j características de interesse a serem divulgadas no estrato j). A solução natural é utilizar uma medida que, em certo sentido, sumarize o comportamento conjunto dos CV's das diversas características. Duas medidas

deste tipo foram consideradas neste estudo: a perda média de eficiência absoluta (P_{wj}) e o CV ordenado de ordem α ($CV_{(t)}j$), com $t = r_j \cdot \alpha$ se $r_j \cdot \alpha \in \mathbb{N}$ e $t = [r_j \cdot \alpha] + 1$ caso contrário. As medidas síntese da precisão associadas a um determinado estrato j serão, neste estudo, denotadas genericamente por S_j. Para o desenvolvimento que se segue, é interessante assumir a existência de uma medida de tamanho T associada a cada estrato populacional (em geral a escolha recai sobre o número de indivíduos do estrato).

Sejam $S_1, \cdots, S_p$ os valores da medida síntese de precisão associados aos estratos populacionais e $\gamma_1, \cdots, \gamma_p$ as participações percentuais (tomando-se por base a medida de tamanho T) dos estratos na população total. Tome-se a ordenação crescente dos S_j, fazendo $S_{(j)}$ representar o j-ésimo valor ordenado e $\gamma_{(j)}$ a partição percentual correspondente (os estratos são ordenados segundo os S_j e $\gamma_{(j)}$ é a participação percentual referente ao j-ésimo estrato na ordenação). A curva de Lorenz associada à medida S e à partição em estratos considerada é dada por:

$$L(y) = \sum_{j=1}^{t(y)} \frac{S_{(j)}\gamma_{(j)}}{S}, \quad y \in [0,1] \tag{13}$$

onde $t(y) = \sup\{k / \sum_{j=1}^{k} \gamma_{(j)} \leq y\}$ e $S = \sum_{j=1}^{p} S_{(j)}\gamma_{(j)}$

Na definição apresentada acima, a curva de Lorenz é uma função do tipo escada, com $p-1$ saltos, ocorrendo nos pontos $\Gamma(k) = \sum_{j=1}^{k} \gamma_{(j)}$, $k = 1, \cdots, p-1$. A exemplo do que se faz na economia, utilizaremos aqui uma versão "suavizada" da curva de Lorenz, descrita abaixo:

$$L^s(y) = \begin{cases} L(\Gamma(k)), & \text{se} \quad \exists\, k \quad \text{tal que} \quad y = \Gamma(k), \quad k = 0, \cdots, p \\ L(\Gamma(t(y))) + (y - \Gamma(t(y)))\dfrac{S_{(t(y)+1)}}{S}, & \text{caso contrário} \end{cases} \tag{14}$$

(na definição acima, convenciona-se $\gamma_{(0)} = 0, \gamma_{(p)} = 1$).

Em outras palavras, a curva "suavizada" coincide com a definição original nos pontos de salto e é obtida por interpolação linear dos valores assim definidos, para os outros pontos. A figura 11,em anexo, ilustra o comportamento da curva de Lorenz, nas suas versões normal e "suavizada".

Conforme já mencionado anteriormente, os estratos considerados na aplicação da metodologia ao CD-90, apresentada na seção 3, foram os municípios brasileiros. Duas medidas de tamanho foram utilizadas para o cálculo da desigualdade da precisão: a população do município e a medida constante $T = 1$, que atribui igual ponderação a todos estratos da população.

A curva de Lorenz associa a participação percentual dos estratos na população à participação percentual na composição do agregado das medidas síntese de precisão. Em uma situação de completa homogeneidade, as medidas de precisão deveriam coincidir em todos os estratos, e as participações destes na composição do agregado da precisão deveriam igualar as participações na composição da população. A curva correspondente a esta situação é também indicada na figura 11. Desigualdades nas precisões dos estratos traduzem-se na convexidade da curva de Lorenz, sendo esta tão mais convexa, quanto maior a desigualdade.

Diversas medidas de desigualdade baseadas na curva de Lorenz foram propostas na economia (Anand, S. (1983)), sendo a mais popular delas o índice de Ginni, definido como o dobro da área compreendida entre as curvas de Lorenz na situação de homogeneidade completa e na situação medida. Por definição o índice toma valores entre 0 e 1, crescendo com o aumento da desigualdade de precisão. Existem diversas fórmulas alternativas para o cálculo do índice de Ginni (Anand, S. (1983)) tendo-se optado neste trabalho pela seguinte:

$$G(S,T) = 2\sum_{j=1}^{p} \frac{(\Gamma(j) - \gamma(j)/2)\, S_{(j)}\gamma_{(j)}}{S} - 1 \qquad (15)$$

Quando $\gamma_{(j)} = 1/p, \quad j = 1,\cdots,p$, (ie, T $\equiv$ 1) o índice simplifica-se para sua forma mais usualmente conhecida:

$$G(S,1) = \frac{2}{P}\sum_{j=1}^{p} \frac{jS_{(j)}}{S} - \frac{p+1}{p} \qquad (16)$$

2.3 - Utilização de Frações de Amostragem Múltiplas - Partição Ótima da População

Nesta seção aborda-se a questão do uso de múltiplas frações amostrais em uma pesquisa, sob a ótica do custo de seleção. Pinheiro e Lima (1988) sugerem que o uso de frações amostrais diferenciadas, em levantamentos como o Censo Demográfico, seria de grande valia para assegurar maior homogeneidade à utilidade das informações coletadas (considerando-se subdivisões da população total para fins de estimação de características). De fato, conforme salientado na introdução, o principal argumento contra a redução da fração amostral no CD-90 é exatamente o receio de que para municípios pequenos, ou subdivisões de municípios médios e grandes, as características de interesse não possam ser estimadas dentro de um mínimo aceitável de precisão. Neste sentido, a utilização de frações amostrais diferenciadas por município (levando em conta o tamanho deste), afigura-se como uma possível solução de compromisso entre a redução de gastos (associada à redução da fração amostral) e a garantia de qualidade para as estimativas divulgadas a nível de subdivisões da população.

As possíveis vantagens e desvantagens associadas ao uso de diferentes frações de amostragem em uma pesquisa dependem, fundamentalmente:

i) da quantidade de frações a serem utilizadas;

ii) dos limites entre as subdivisões da população total onde serão utilizadas as diferentes frações e

iii) das frações a serem efetivamente utilizadas nestas subdivisões.

A primeira das questões acima é, sem dúvida, a mais delicada, pois envolve considerações de natureza gerencial, vinculadas a particularidades da pesquisa, tais como prazos acertados, supervisão de campo, treinamento da rede de coleta, etc. Não se pretende, neste trabalho, sugerir um número ótimo de frações, mas apenas apresentar resultados que auxiliem na tomada de decisão sobre esta questão.

O tratamento teórico das duas outras questões constitui o objetivo central desta seção. O número de frações será assumido como pré-fixado, sendo representado, no restante da seção, pela letra M. A partição da população, para fins de determinação das frações amostrais correspondentes, será feita com base em uma medida de tamanho T, associada a determinadas subdivisões da população total, que serão consideradas como unidade mínima para fins de aplicação da fração. No contexto do CD-90, estas subdivisões foram os municípios e a variável T a população correspondente.

Sejam

$F(t)$ = % de indivíduos da população total em subdivisões com $T \leq t$;

$f(t)$ = fração amostral associada a uma subdivisão de tamanho t.

Esta última função é obtida a partir de um dos métodos de dimensionamento apresentados em 2.1. A caracterização da fração amostral como função de T é feita a partir da especificação de um modelo. Considerações sobre a estimação do modelo (assim como da função $F(t)$) são deixadas para a seção 3, onde é apresentada a aplicação da metodologia ao CD-90.

Sejam, ainda

$t_{\min}$ = valor mínimo de T na população e

$t_{\max}$ = valor máximo de T na população

Resultado 8

Se em uma pesquisa deseja-se utilizar M frações amostrais distintas em subdivisões da população, com medida de tamanho T associada, segundo uma partição da população total definida pelos valores $t_{\min}, t_1, \cdots, t_{m-1}, t_{\max}$ da variável T, então a **fração amostral equivalente** (aquela que, se utilizada em todas as subdivisões da população, resultaria no mesmo tamanho de amostra que o correspondente às M frações) é dada por:

$$f_{eq}(t_1, \cdots, t_{m-1}) = \sum_{k=1}^{m} f(t_{k-1}, t_k) \cdot [F(t_k) - F(t_{k-1})]$$

onde $f(t_{k-1}, t_k)$ representa a fração amostral utilizada nas subdivisões populacionais compreendidas no intervalo $[t_{k-1}, t_k)$. Note que a função $f(t)$ fornece o valor da fração para um dado t, necessitando-se, portanto, de uma definição adicional para obter $f(t_{k-1}, t_k)$. Existem inúmeras alternativas, podendo-se mencionar:

$$f(t_{k-1}, t_k) = \frac{f(t_k) + f(t_{k-1})}{2}; \tag{17}$$

$$f(t_{k-1}, t_k) = f\left(\frac{t_k + t_{k-1}}{2}\right); \tag{18}$$

$$f(t_{k-1}, t_k) = \inf_{t \in [t_{k-1} t_k)} f(t); \tag{19}$$

$$f(t_{k-1}, t_k) = \sup_{t \in [t_{k-1}, t_k)} f(t) \tag{20}$$

Em (20) adota-se a solução mais custosa, mas também com maior garantia de utilidade. A solução (19) é a mais barata, mas de menor qualidade. (17) e (18) fornecem soluções intermediárias, sendo que a (17) tende a ser maior que a (18) quando f for convexa e menor quando ela for côncova. Optou-se, nesse estudo, pela solução (17), já que a análise das funções $f(t)$ para as diversas estratégias de dimensionamento consideradas indicou que a forma convexa era predominante. A solução adotada é uma alternativa intermediária, em termos de custo, entre (19) e (20), sendo, na situação particular do trabalho desenvolvido, mais conservadora, em termos de garantia de utilidade para as informações coletadas, que a opção (18).

O problema de otimização consiste em, fixado o valor de M, e conhecidas as funções f e F, encontrar os limites $t_1, \cdots, t_{m-1}$ que minimizem a fração equivalente. O algoritmo utilizado neste trabalho, consiste em uma busca da solução ótima através de uma grade fina da variável T, a partir de repetidas aplicações da solução para o caso de $M-1$ frações, até a convergência do processo.

2.4) Impactos do Uso de Frações Múltiplas sobre a Operacionalidade

A principal argumentação contrária à utilização de frações múltiplas de amostragem no CD-90 diz respeito às possíveis dificuldades operacionais que dela adviriam. Neste item procura-se analisar a questão de modo breve, com ênfase nos aspectos técnicos associados.

A utilização de frações de amostragem diferenciadas pode impactar, com relação à operacionalidade da pesquisa, em duas de suas fases: a coleta e a expansão/tabulação dos resultados.

2.4.1) Impactos sobre a Coleta

Em pesquisas de campo do porte do Censo Demográfico, usualmente adota-se o procedimento de seleção por Amostragem Sistemática. Isso faz com que as frações amostrais passíveis de utilização prática se restrinjam ao conjunto

$$\{f/\exists n \in \mathbb{N} \quad \text{tal que} \quad f = 1/n\} \tag{21}$$

Em outras palavras, as frações amostrais utilizadas em Amostragem Sistemática (ao menos na sua versão mais comumente empregada) devem ser tais que o seu inverso resulte em um número inteiro positivo. Na verdade, esta restrição diz respeito ao método de seleção e não à utilização de frações múltiplas de amostragem. Sua inclusão, aqui, prende-se ao fato de que ela se relaciona diretamente com a operacionalidade da pesquisa, devendo ser considerada na etapa de especificação das propostas de dimensionamento, descrita na seção 3.

A coleta de questionários em pesquisas como o CD, pelo próprio porte do levantamento, é feita de modo descentralizado, em geral obedecendo à partição geográfica já existente (no caso do CD, a coleta é feita a nível de setores censitários, que são subdivisões contíguas dos municípios com 200 domicílios em média). Estas unidades de coleta reportam-se a uma unidade (primária) de supervisão, que por sua vez é coordenada por outra unidade superior e assim sucessivamente, até o órgão central encarregado do levantamento. O número de estágios de supervisão depende, obviamente, da dimensão da pesquisa e da abrangência geográfica da mesma. No CD, o nível primário de supervisão é feito pelas agências (cerca de 646 em todo Brasil), que por sua vez reportam-se a delegacias (26) que dirigem-se diretamente ao órgão central.

O principal inconveniente associado ao uso de frações múltiplas, com relação à operacionalidade da fase de coleta, diz respeito à convivência de frações diferenciadas ao nível primário de supervisão. Isso ocorreria, por exemplo, quando dois municípios, nos quais se adotassem frações amostrais distintas, se reportassem a uma mesma agência. O problema é colocado apenas ao nível primário de supervisão por que, em geral, é neste que a rede de coleta apresenta maiores dificuldades de capacitação e treinamento de mão-de-obra, além de lidar diretamente com os problemas operacionais de campo (não cobertura de determinadas unidades, tendenciosidades na seleção das unidades da amostra, etc.)

A solução ideal, neste sentido, seria considerar como área mínima para aplicação de uma determinada fração, o agregado de unidades de coleta ligadas a uma mesma unidade primária de supervisão. No presente estudo esta solução não foi adotada, parte pela ausência de dados disponíveis a nível da unidade primária (utilizar agregados de municípios

tornaria impraticável o levantamento dos dados necessários aos cálculos), parte pelo receio de que tal medida pudesse impactar negativamente sobre os outros resquisitos mencionados no início da seção. A unidade mínima para aplicação das frações consideradas neste estudo é o município, conforme anteriormente mencionado. Este aspecto sem dúvida deve ser considerado quando da escolha da proposta final de dimensionamento.

2.4.2) Impactos Sobre a Expansão / Tabulação

A exemplo do que se faz na fase de coleta, na expansão/ tabulação é definido o nível de desagregação da população para o qual as estimativas serão divulgadas, a chamada área de ponderação (no caso do CD, adota-se o município, ou distritos e regiões administrativas nas grandes metrópoles).

Todos os cálculos referentes à expansão da amostra, tais como estimativas de totais e da precisão dos estimadores envolvidos, são feitos inicialmente a nível da área de ponderação, sendo posteriormente agregados para outros níveis (Mesoregiões, Unidades da Federação, etc., no caso do CD).

Assim, o uso de frações múltiplas **não tem qualquer impacto** sobre esta etapa da pesquisa, contanto que a área de ponderação esteja contida na unidade de aplicação das frações. Na aplicação desenvolvida para o CD-90, esta premissa é respeitada.

3 - APLICAÇÃO AO CENSO DEMOGRÁFICO-90

Nesta seção é apresentada uma aplicação da metodologia, descrita em 2, à situação particular que trata do dimensionamento da amostra do Censo Demográfico de 1990.

Com base nas informações provenientes do Censo Demográfico de 1980 referentes às características de pessoas e de domicílios que tiveram suas estimativas divulgadas a nível de município, é desenvolvida uma análise minuciosa dos possíveis impactos que adviriam da modificação do esquema de amostragem tradicionalmente adotado nos últimos Censos Demográficos.

O objetivo é avaliar os possíveis ganhos e perdas associados à redução da fração de amostragem quando comparado com a já tradicional fração de 25%, bem como formular uma proposta técnica de dimensionamento da amostra para o referido Censo.

3.1 - Dados Utilizados

Os dados utilizados nesta aplicação se referem às características de domicílios e de pessoas que foram investigadas no Censo Demográfico de 1980 e que tiveram suas estimativas divulgadas a nível de município (área de ponderação do CD-80).

Assim foram consideradas as características de pessoas referentes a Mão-de-Obra (pessoas de 10 anos ou mais por setor de atividade de dependência e condição de atividade, rendimento médio mensal e sexo, pessoas que trabalharam no ano de referência por setor de atividade e sexo): Dados Gerais (população residente por religião, cor, estado conjugal e sexo); Instrução(anos de estudos e sexo, curso completo por grau do curso e sexo) e Fecundidade (mulheres de 15 anos ou mais, filhos tidos e filhos vivos), o que perfaz um número superior a 150 variáveis, se for considerada a desagregação do plano tabular de divulgação a nível de município.

Quanto ao domicílio, foram consideradas as características referentes à situação, condição de ocupação, forma de abastecimento d'água, existência e uso das instalações sanitárias, escoadouro, aluguel e/ou valor da prestação da casa própria, existência e tipo de fogão, combustível usado para cozinhar, existência de telefone, iluminação elétrica, rádio, geladeira, televisão e automovéis; perfazendo um número superior a 65 variáveis.

Considerando o volume da massa de dados a ser analisado e também a grande rarefação de determinadas características na população (exemplo - a variável migração no

nível de município), optou-se por trabalhar com as variáveis a um nível mais sumarizado, considerando-se assim as informações pertinentes às características de interesse.

Desta forma, a análise é desenvolvida com base em um elenco de 20 variáveis referentes às características de domicílios e 20 referentes às características de pessoas, tendo sido adotado como "medidas resumo" a mediana e o terceiro quartil obtidos a partir dos diferentes níveis de divulgação das variáveis mencionadas. Por exemplo, a variável **Religião por Sexo** admite um número de níveis superior a 16 no plano tabular de divulgação do Censo Demográfico de 1980 (homens católicos, homens protestantes, homens da religião espiríta Kardecista, homens da religião oriental, etc ...). Para esta variável considerou-se a mediana dos valores nos seus diferentes níveis. Dessa forma são eliminados, implicitamente, os níveis muito rarefeitos na população da referida variável.

No anexo 1 é apresentada a relação das variáveis consideradas nesta aplicação, bem como as "medidas resumo" adotadas.

A não inclusão de características referentes a Migração é proposital e esta decisão foi tomada considerando a grande rarefação da variável, observada no nível de município.

Ainda com relação à agregação dos níveis de divulgação das variáveis, é importante ressaltar que, para manter sob controle a qualidade das informações (em termos de precisão) para todo o elenco de variáveis do plano tabular de divulgação, seria necessário uma amostra muito grande, na verdade praticamente um censo, considerando que existem características que são muito rarefeitas na população, por exemplo: mulheres da religião oriental ou mulheres de cor amarela, etc.

Nesta aplicação foram considerados 64 municípios abrangendo todas as regiões brasileiras, com contingentes populacionais, segundo o Censo Demográfico de 1980, variando de 1968 a 1308919 habitantes. A relação dos municípios com as respectivas populações encontra-se no anexo 2.

Dentre os municípios considerados no experimento, 21 apresentam população superior a 100.000 habitantes. Conforme sugestões apresentadas em reuniões técnicas entre membros da comissão consultiva externa para o planejamento do Censo Demográfico de 1990 e a comissão responsável pelo delineamento do plano amostral do referido censo; estes municípios foram desagregados, para fins de obtenção de estimativas do erro amostral, em partições de tamanho 50.000 habitantes.

A desagregação em partições com este contingente populacional foi considerada com base no comportamento observado, em termos populacionais, da demanda por estimativas a nível mais desagregado que a área mínima de ponderação adotada no Censo Demográfico de 1980, nos municípios mais populosos - principalmente naqueles pertencentes às principais regiões metropolitanas.

3.2 - O Cálculo das Estimativas do Erro Amostral

A avaliação da precisão dos estimadores das características investigadas a partir das amostras consideradas nesta aplicação é feita com base na análise das estimativas do erro amostral associado às estimativas obtidas sob a hipótese de um esquema de amostragem aleatória simples sem reposição (estimação de totais populacionais).

Estas estimativas são obtidas a partir dos erros amostrais referentes à amostra de 25% do Censo Demográfico de 1980 (ver Fundação Instituto Brasileiro de Geografia e Estatística. Metodologia do Censo Demográfico de 1980).

O erro amostral associado às estimativas obtidas a partir das frações de amostragem consideradas nesta aplicação, foram obtidos através da fórmula apresentada no Resultado 2 da seção 2, ou seja:

$$CV(\hat{\theta}(f)) = \left(\frac{f_0(1-f)}{f(1-f_0)}\right)^{1/2} CV(\hat{\theta}(f_0))$$

onde $CV(\hat{\theta}(f))$ é o coeficiente de variação amostral do estimador das características de interesse assumindo a fração amostral f, e f_0 é a fração de referência ($f_0 = 1/4$).

3.2.1 Cálculo do Erro Amostral nos Municípios com População Superior a 100 000 habitantes

Os municípios com população igual ou superior a 100 000 habitantes foram desagregados, para fins de cálculo das estimativas do erro amostral, em partições com população de 50 000 habitantes cada. Sob a hipótese de que o método de seleção é de amostragem aleatória simples sem reposição (estimação de totais populacionais), o erro amostral expresso em termos do coeficiente de variação amostral, nas partições definidas, pode ser obtido através da seguinte expressão:

$$CV_s(\hat{\theta}(f)) = \left(\frac{S^2(1-f)}{n_s}\right)^{1/2} / \theta \qquad (22)$$

onde: $n_s = f.50000$ é o tamanho da amostra associada à s-ésima partição no município considerado; S^2 é a variância da característica investigada, na população.

Sob a hipótese de que as variâncias populacionais (município como um todo e em uma dada partição) sejam iguais, tem-se que

$$\frac{CV(\hat{\theta}(f))}{CV_s(\hat{\theta}(f))} = \left(\frac{\frac{S^2(1-f)}{n}}{\frac{S^2(1-f)}{n_s}}\right)^{1/2} = \left(\frac{50000}{N}\right)^{1/2} = \left(\frac{1}{k}\right)^{1/2} \tag{23}$$

onde $k = N/50000$.

Portanto, a relação entre os coeficientes de variação amostral calculado a nível das partições em um determinado município, poderia ser expresso por:

$$CV_s(\hat{\theta}(f)) = (k)^{1/2} CV(\hat{\theta}(f)) \tag{24}$$

Considerando que, na desagregação em partições de 50 000 habitantes, não se levaram em conta os aspectos de contigüidade dos setores censitários do município, a hipótese de homogeneidade acima mencionada dificilmente se verificará. Assim, uma alternativa que nos parece bastante razoável é supor que: $S_s^2 = S^2/d(t)$ onde adotou-se o **fator de desconto** $d(t)$, definido por

$$d(t) = \begin{cases} 1 & \text{se} \quad t < 100000 \\ 1 + \dfrac{0.5(t - 100000)}{900000} & \text{se} \quad 100000 \leq t \leq 1000000 \\ 1,5 & \text{se} \quad t > 1000000 \end{cases} \tag{25}$$

e t representa a população do município considerado.

Dessa forma o coeficiente de variação amostral correspondente às estimativas obtidas a partir da aplicação das diferentes frações de amostragem consideradas neste estudo, à nível das partições definidas, foi avaliado através da seguinte expressão:

$$CV_s(\hat{\theta}(f)) = \left(\frac{k}{d(t)}\right)^{1/2} CV(\hat{\theta}(f)) \tag{26}$$

onde $d(t)$ é **o fator de desconto** definido anteriormente.

3.3 - Funções Peso Consideradas

Conforme mencionado em 2.1, para o cálculo das frações amostrais mínima, adequada e necessária é preciso especificar-se uma função peso que represente, quantitativamente, a importância relativa dos limites de tolerância para a precisão das estimativas.

Neste estudo, as funções peso consideradas pertencem à família de densidades Beta. Para a especificação da função peso apropriada a uma particular aplicação, a sub-rotina de cálculo desenvolvida para o trabalho solicita ao usuário o limite mínimo de tolerância (c_0), o limite máximo (c_1), o valor de importância máxima entre os limites de tolerância considerados (c_m) e o valor cuja importância relativa é a metade do máximo ($c_{m/2}$).

Assim, para o estudo desenvolvido, a função peso é completamente caracterizada pelas quatro constantes definidas no parágrafo anterior, sendo, portanto, razoável adotar-se a notação

$$w = w(c_0, c_1, c_m, c_{m/2}) \tag{27}$$

para referenciar uma função peso particular.

Na aplicação desenvolvida para o CD-90, foram utilizadas quatro funções peso diferentes, duas para os cálculos que envolvessem a perda de eficiência e duas para a derivação das frações necessárias (associadas à especificação de um limite de tolerância para a estimação de um dado percentual de características).

3.3.1 Funções Peso para o Cálculo da Perda de Eficiência

Na notação apresentada em (27), as funções peso consideradas na obtenção das diferentes perdas de eficiência foram:

$$w_{p1} = w(5, 30, 10, 20) \qquad \text{e}$$
$$w_{p2} = w(2, 40, 15, 25) \tag{28}$$

As figuras 12 e 13, em anexo, mostram os gráficos correspondentes às duas funções peso.

A primeira das funções representa uma intervenção mais direta do usuário sobre a especificação da importância dos limites de tolerância. A faixa de variação para os limites, 5 a 30, é razoavelmente estreita, sendo atribuída importância máxima a um valor relativamente baixo para o limite de tolerância, 10, tendo em vista a multiplicidade de

características e a diversidade de níveis de divulgação do CD-90. Neste sentido, esta pode ser considerada uma função peso rigorosa, em termos da exigência de precisão para as estimativas.

A segunda função está associada a uma intervenção mais branda sobre os limites de tolerância. A faixa de variação, 2 a 40, cobre praticamente todos valores que podem ser considerados razoáveis como limites de tolerância. A importância máxima é atribuída a um valor razoavelmente elevado em termos de limite tolerância, 15. Esta proposta caracteriza-se, então, por ser mais flexível que a anterior.

3.3.2 Funções Peso Para o cálculo das Frações Necessárias

Na notação de (27), as funções peso utilizadas foram:

$$w_{c1} = w(10, 40, 20, 25) \qquad \text{e}$$
$$w_{c2} = w(15, 50, 30, 38) \qquad (29)$$

À semelhança do que foi feito para a perda de eficiência, têm-se aqui duas versões para as funções peso: uma rigorosa, representada por w_{c1} e outra flexível, dada por w_{c2}.

Note-se que as funções peso associadas às frações necessárias traduzem a importância relativa dos limites de tolerância para o **máximo** coeficiente de variação de um dado percentual de variáveis. É natural, portanto, que os valores positivos destas funções estejam concentrados em limites de tolerância superiores àqueles estabelecidos para o cálculo da perda média de eficiência (w_{p1} e w_{p2}).

As figuras 14 e 15, em anexo, apresentam os gráficos das funções w_{c1} e w_{c2}, respectivamente.

3.4 Frações Amostrais - Abordagens de Dimensionamento Utilizadas

Na seção 2 foram apresentados três procedimentos alternativos para o dimensionamento da fração de amostragem. Essencialmente, o dimensionamento da fração amostral, em qualquer dos três procedimentos analisados, depende da unidade de aplicação da fração; dos coeficientes de variação de um elenco de características de interesse, obtidos a partir de um plano amostral de referência; da função peso utilizada e de determinados parâmetros associados ao procedimento específico.

No estudo desenvolvido para o CD-90, a unidade de aplicação da fração é o município; o plano amostral de referência é a fração única de 25% e as características de interesse consideradas estão descritas em 3.1; as funções peso utilizadas foram apresentadas no item anterior e os parâmetros específicos de cada procedimento são detalhados a seguir.

3.4.1 Fração Amostral Mínima

A fração amostral mínima associada a uma perda média de eficiência máxima de $\alpha \times 100\%$, com relação a uma fração de referência f_0 é definida em (9), no item 2.1.1. Na notação ali apresentada, as alternativas de dimensionamento consideradas para esta fração foram:

$$f_{1,w_{p1}}(0.25, 0.10) \quad \text{e} \quad f_{1,w_{p2}}(0.25, 0.10) \tag{30}$$

ou seja, foram utilizadas as frações mínimas correspondentes a uma perda máxima de 10% com relação à fração de 25%, associadas às duas funções peso definidas em 3.3.1.

No desenvolvimento do estudo, estas frações foram calculadas para cada município da amostra de 64 considerada no trabalho (ver 3.1). Os valores assim obtidos foram utilizados na modelagem das frações amostrais como função do tamanho do município, cujos resultados são apresentados em 3.5.

3.4.2 Fração Amostral Adequada

A fração amostral adequada associada a uma razão r entre os custos de seleção e utilidade foi definida no Resultado 5, no item 2.1.1. As alternativas de dimensionamento consideradas, segundo a notação introduzida na definição da fração, foram:

$$f_{2,w_{p1}}(1), f_{2,w_{p2}}(1), f_{2,w_{p1}}(1.5) e f_{2,w_{p2}}(1.5) \tag{31}$$

ou seja, foram utilizadas as frações adequadas correspondentes às razões 1 e 1.5 entre os custos de seleção e utilidade, associadas às funções peso definidas em 3.3.1.

A razão entre custos igual a 1 atribui igual importância ao custo da pesquisa e à utilidade das informações levantadas. Já a razão 1.5 da importância 50% maior à variação de um ponto percentual na fração amostral com relação ao mesmo incremento na perda média de eficiência absoluta (utilidade). Obviamente as frações dimensionadas segundo a razão 1 serão, para uma mesma função peso, sempre superiores àquelas dimensionadas

segundo a razão 1.5. Os resultados obtidos para os diversos municípios da amostra são apresentados no item 3.5, junto com a modelagem das frações em função do tamanho do município.

3.4.3 Fração Amostral Necessária

A fração amostral necessária média, ponderada por uma função peso w, associada a um percentual mínimo de $\alpha \times 100\%$ de características de interesse é definida em (12), no item 2.1.2. Na notação ali introduzida, as alternativas de dimensionamento consideradas foram

$$f_{3,w_{c1}}(0.75), f_{3,w_{c2}}(0.75), f_{3,w_{c1}}(0.90) e f_{3,w_{c2}}(0.90) \tag{32}$$

ou seja, foram calculadas as frações necessárias médias para 75% e 90% das características de interesse, ponderadas segundo as funções peso w_{c1} e w_{c2}, definidas em 3.3.2.

As frações necessárias para 90% das características de interesse serão sempre superiores às correspondentes a 75% das características, mantendo-se fixada a função peso. Os resultados do dimensionamento associados aos diversos municípios da amostra, bem como a modelagem das frações em função do município são apresentados em 3.5.

3.5 Procedimentos de Modelagem

Dentro da metodologia apresentada na seção 2, procedimentos de modelagem são necessários para a estimação de duas funções, $F(t)$ e $f(t)$, definidas no item 2.3.

A função $f(t)$ fornece a fração amostral correspondente a um município de tamanho t, enquanto que $F(t)$ dá o percentual de indivíduos na população que residem em municípios de tamanho menor ou igual a t. As duas funções são fundamentais na etapa de obtenção das partições ótimas da população, para o uso de um dado número de frações amostrais distintas (ver 2.3).

Neste estudo, optou-se por uma modelagem não-paramétrica das funções $f(t)$ e $F(t)$. Esta opção prende-se ao fato de que o interesse, para fins do estudo desenvolvido, concentrado na obtenção de um procedimento de cálculo que forneça, para um dado t, os valores de $f(t)$ e $F(t)$, não havendo interesse maior em procurar entender o mecanismo que relaciona t às funções f e F (para o que um modelo paramétrico seria mais conveniente).

3.5.1 Estimação da Fração Amostral em Função do Tamanho $f(t)$

A estimação do modelo, no caso da função $f(t)$, foi feita através de sucessivas suavizações dos valores obtidos a partir da amostra de 64 municípios considerada no estudo. Estas suavizações foram obtidas por meio de médias móveis, adpatadas para levar em conta o afastamento entre as abcissas dos pontos que entram no seu cálculo (de forma a dar maior peso às observações próximas).

Para todas as alternativas de dimensionamento da fração amostral consideradas (ver 3.4), foram aplicadas duas suavizações, a primeira com uma média móvel adaptada de sete termos e a segunda com cinco. Em ambas os pesos atribuídos aos valores considerados para o cálculo da média decresciam a partir do valor central. No caso da primeira suavização utilizaram-se os pesos 1/16, 2/16, 3/16, 4/16, 3/16, 2/16 e 1/16, enquanto para a segunda empregaram-se os pesos 1/9, 2/9, 3/9, 2/9 e 1/9.

As figuras 16 a 25, em anexo, apresentam os modelos ajustados, segundo a abordagem descrita acima, para as dez alternativas de dimensionamento apresentadas em 3.4. São indicados, também, os valores observados para os 64 municípios da amostra.

Ao invés do tamanho da população do município, foi utilizada como variável independente a raiz quadrada desta. Na verdade, todos procedimentos de cálculo necessários à obtenção da partição ótima da população foram feitos com base na raiz quadrada do tamanho da população do município. A razão básica para a adoção desta transformação prende-se a problemas de precisão numérica dos algoritmos computacionais utilizados. Observe-se, entretanto, que, face à natureza das grandezas envolvidas, a visualização do comportamento das funções F e f é muito mais fácil na escala transformada, que na original. Note-se, ainda, que as soluções obtidas para as partições ótimas da população são equivalentes nas escalas transformada e original.

3.5.2 Estimação da Função de Distribuição Acumulada da População do Município $F(t)$

Com relação à função F, não foram necessários procedimentos do tipo média móvel para sua estimação. A partir das projeções de população de pessoas por município, para o ano de 1989, feitas pelo Departamento de População do IBGE, foram obtidos os percentuais de indivíduos morando em municípios com população igual ou inferior a um dado valor t, para t variando de 2.000 a 12.000.000 de habitantes, segundo uma partição fina. Os valores assim encontrados representam o verdadeiro valor de $F(t)$, a menos de incorreções devido

à incerteza das projeções. A estimativa de $F(t)$ para um valor qualquer de t é obtida, no estudo, através de interpolação linear simples dos valores exatos obtidos a partir da consulta ao arquivo de projeções.

A figura 26, em anexo, mostra o gráfico de F, considerando-se a raiz do tamanho da população. A figura 27 apresenta o histograma da distribuição da variável raiz da população, segundo a partição utilizada na obtenção dos valores exatos da F. Para melhor visualização, é feito um corte em $t = 1000$ (população de 1.000.000 de habitantes).

3.6 Propostas de Dimensionamento de Amostra para o CD-90

Neste item são apresentadas as propostas de dimensionamento para o CD-90, obtidas a partir das dez alternativas de cálculo para a fração amostral descritas em 3.4, segundo diferentes quantidades de partições. As propostas de dimensionamento são especificadas através das partições da população adotadas e das frações amostrais correspondentes. Em cada caso, as partições foram obtidas de modo a minimizar a fração equivalente (ver resultado 8, item 2.3), segundo o algoritmo mencionado ao final de 2.3.

Para cada alternativa de dimensionamento da fração, foram consideradas três possibilidades: 2, 3 e 4 frações diferenciadas. Além destas, foram consideradas para fins de comparação três propostas de dimensionamento com fração única: 1/4, 1/8 e 1/10. Face ao grande número de propostas consideradas, é interessante introduzir uma notação para referenciá-las.

Uma proposta de dimensionamento com M partições da população e fração amostral dimensionada a partir do procedimento f será denotada por:

$$M/f \tag{33}$$

Se M for igual a um, f representará a fração (única) adotada.

Uma última observação sobre o cálculo das propostas de dimensionsamento refere-se às frações e limites para as partições. Apesar de terem-se utilizado as variáveis na sua forma contínua, os valores obtidos foram "arredondados" de tal forma a tornar a proposta factível em termos operacionais (desta forma, a fração 0,27 foi "arredondada" para 0.25 e assim por diante).

A análise das diversas propostas consideradas é deixada para o item 3.7.

3.6.1 Propostas com uma única Fração Amostral

Conforme mencionado na introdução deste item e seguindo a notação introduzida em (33), as propostas de dimensionamento com fração única, analisadas no estudo, foram: 1/0.25, 1/0.125 e 1/0.10 .

A primeira proposta foi considerada por ter sido aquela utilizada no último CD. A última foi incluída por ser a proposta considerada, a princípio, para utilização no próximo CD. A segunda proposta foi introduzida como uma posição intermediária entre as duas anteriores.

3.6.2 Propostas com duas Frações Amostrais

Foram obtidas propostas de dimensionamento para cada uma das dez alternativas de dimensionamento de frações apresentadas em 3.4. A tabela I apresenta os resultados obtidos, segundo a notação definida em (33) e na seção 2.

Na tabela I, a segunda coluna apresenta o limite de partição da população, lim(hab), para fins de aplicação das frações diferenciadas. A fração 1, indicada na terceira coluna, é aplicada aos municípios com população igual ou inferior ao limite, enquanto a fração 2 é aplicada aos demais municípios.

Note-se a grande estabilidade dos limites de partição: em cinco das propostas, vale 15 000, em outras três, 2 500 e apenas uma vez é superior a 30 000. Outra observação importante é que em **nenhuma** das propostas consideradas a fração 1/10 revelou-se adequada para os municípios abaixo do limite de partição. Por outro lado, apenas uma vez a fração de 1/4 foi superada nesta mesma situação; mesmo assim, para a proposta de dimensionamento baseada na fração necessária para 90% das características de interesse, com função peso rigorosa.

3.6.3 Propostas com três Frações Amostrais

De forma análoga à tabela I, a tabela II apresenta as propostas de dimensionamento para as dez alternativas de dimensionamento da fração, para três partições da população.

Na tabela II, lim 1 representa o limite da primeira partição e lim 2 o limite da segunda. A fração 1 é aplicada aos municípios de tamanho menor ou igual ao limite da primeira

Tabela I

Propostas de Dimensionamento para duas Frações

Proposta	lim(hab)	Fração 1	Fração 2
$2/f_{1w_{p1}}(0.25, 0.1)$	2500	1/6	1/8
$2/f_{1w_{p2}}(0.25, 0.1)$	200000	1/8	1/15
$2/f_{2w_{p1}}(1)$	30000	1/4	1/7
$2/f_{2w_{p2}}(1)$	15000	1/4	1/7
$2/f_{2w_{p1}}(1.5)$	2500	1/6	1/9
$2/f_{2w_{p2}}(1.5)$	2500	1/6	1/9
$2/f_{3w_{c1}}(0.75)$	15000	1/4	1/19
$2/f_{3w_{c2}}(0.75)$	15000	1/8	1/20
$2/f_{3w_{c1}}(0.90)$	15000	1/2	1/8
$2/f_{3w_{c2}}(0.90)$	15000	1/4	1/15

partição, enquanto a segunda fração é aplicada aos municípios com população entre os limites da primeirs e da segunda partição e a fração 3 aos demais municípios.

Observa-se, novamente, a grande estabilidade dos limites das partições: nenhum dos limites da primeira partição foi superior a 15 000 e sete dos limites da segunda partição foram iguais a 200 000. Novamente, com exceção da proposta baseada em $f_{3w_{c1}}(0.90)$, a fração de 1/4 mostrou-se suficiente para todas as primeiras partições, havendo menor estabilidade das frações para as segunda e terceira partições. A solução obtida para $f_{2w_{p1}}(1.5)$ é praticamente idêntica àquela encontrada para duas partições, sugerindo que, em termos da fração equivalente, não há alterações substanciais quando se passa de duas para três partições, nesta alternativa de dimensionamento.

3.6.4 Propostas com Quatro Frações Amostrais

A tabela III apresenta as propostas de dimensionamento referente a quatro partições

Tabela II

Propostas de Dimensionamento para três Frações

Proposta	lim 1(hab)	lim1 2(hab)	Fração 1	Fração 2	Fração 3
$3/f_{1w_{p1}}(0.25, 0.1)$	2500	200000	1/6	1/8	1/12
$3/f_{1w_{p2}}(0.25, 0.1)$	4000	200000	1/6	1/9	1/15
$3/f_{2w_{p1}}(1)$	15000	200000	1/3	1/6	1/9
$3/f_{2w_{p2}}(1)$	15000	200000	1/4	1/7	1/11
$3/f_{2w_{p1}}(1.5)$	2000	2500	1/5	1/6	1/9
$3/f_{2w_{p2}}(1.5)$	2500	200000	1/6	1/9	1/15
$3/f_{3w_{c1}}(0.75)$	8000	66000	1/4	1/11	1/20
$3/f_{3w_{c2}}(1.5)$	8000	66000	1/7	1/22	1/91
$3/f_{3w_{c1}}(0.90)$	15000	200000	1/2	1/8	1/20
$3/f_{3w_{c2}}(0\ 90)$	15000	200000	1/4	1/14	1/20

da população, para as alternativas de obtenção das frações consideradas.

A exemplo das tabelas anteriores, lim 1, lim 2 e lim 3 representam os limites para as partições e fração 1, fração 2, fração 3 e fração 4 as frações amostrais correspondentes.

Algumas alternativas de dimensionamento apresentam soluções idênticas às obtidas para o caso de três partições ($f_{1w_{p1}}(0.25, 0.10)$ e $f_{1w_{p2}}(0.25, 0.10)$) e mesmo para o caso de duas partições ($f_{2w_{p1}}(1.5)$). Nota-se que quatro é um número excessivo de partições, e as soluções, em geral, são muito semelhantes às de três partições.

3.7 - Avaliação das Propostas de Dimensionamento

Para fins de avaliação das propostas de dimensionamento foi necessário proceder-se a um "enxugamento" prévio das 33 propostas apresentadas em 3.6, de modo a tornar factível

Tabela III

Propostas de Dimensionamento para Quatro Frações

Proposta	lim 1 (hab)	lim 2 (hab)	lim 3 (hab)	Fração 1	Fração 2	Fração 3	Fração 4
$f_{1w_{p1}}(0.25, 0.10)$	2000	2500	200000	1/6	1/6	1/8	1/12
$f_{1w_{p2}}(0.25, 0.10)$	3000	15000	200000	1/6	1/6	1/9	1/15
$f_{2w_{p1}}(1)$	4000	15000	200000	1/3	1/4	1/6	1/9
$f_{2w_{p2}}(1)$	4000	15000	200000	1/4	1/5	1/7	1/11
$f_{2w_{p1}}(1.5)$	2000	2050	2500	1/5	1/5	1/6	1/9
$f_{2w_{p2}}(1.5)$	2500	15000	200000	1/6	1/7	1/10	1/15
$f_{3w_{c1}}(0.75)$	8000	15000	66000	1/4	1/8	1/17	1/43
$f_{3w_{c2}}(0.75)$	8000	15000	66000	1/7	1/17	1/34	1/91
$f_{3w_{c1}}(0.90)$	8000	15000	200000	1/2	1/3	1/8	1/28
$f_{3w_{c2}}(0.90)$	8000	15000	200000	1/3	1/6	1/14	1/59

a comparação entre elas. A seleção das propostas para avaliação foi feita com base nos seguintes critérios:

- eliminação das propostas em que qualquer das frações fosse inferior a 1/20 (por julgar-se ser este um limite mínimo para alteração da fração amostral, ao menos com vistas ao próximo censo);
- eliminação da proposta com número de partições maior, sempre que esta fosse equivalente à obtida para um número menor de partições; e
- escolha de uma entre as propostas equivalentes para um mesmo número de partições.

Obteve-se, assim, um conjunto reduzido de 15 propostas que, de algum modo, apresentam diferenças entre si que justifiquem sua inclusão neste item.

Estas propostas são analisadas, aqui, segundo os quatro requisitos básicos mencionados na introdução e utilizados ao longo de todo trabalho: garantia de utilidade, homogeneidade de precisão através dos municípios, operacionalidade dos procedimentos de coleta e apuração e custos.

São utilizadas para tal, as medidas descritas na metodologia e outras introduzidas neste item. Faz-se, ainda, uma tentativa de sintetizar a adequação global de uma proposta, levando-se em conta os requisitos básicos simultaneamente, através de um índice definido mais adiante no texto.

3.7.1 Custo de Seleção

A medida utilizada para avaliação do custo de seleção associado a uma dada proposta é a **fração equivalente** (f_{eq}) definida no resultado 8, do item 2.3. A tabela IV apresenta os resultados para as propostas consideradas. Inclui-se nesta tabela uma medida relativa de custo entre as propostas, e a fração equivalente relativa (f_r), definida a seguir.

A fração equivalente relativa associada à i-ésima proposta de dimensionamento é dada por:

$$f_{r,i} = \frac{f_{eq,i}}{f_{min}}, \quad i = 1, \cdots, 15 \tag{33}$$

onde f_{min} representa a menor fração equivalente entre as propostas consideradas.

A análise da tabela IV revela aspectos interessantes relacionadas ao comportamento da fração equivalente:

- para uma mesma alternativa de dimensionamento a fração equivalente decresce com o número de partições (ver por exemplo $2,3e4/f_{2w_{p2}}(1)$);

- as frações equivalentes, para uma dada alternativa de dimensionamento, parecem convergir para um valor limite, quando aumenta o número de partições (ver o mesmo exemplo que na observação anterior);

Estas observações sugerem que utilizar frações múltiplas é uma estratégia interessante para redução da fração equivalente, mas a partir de um dado ponto é pequeno o impacto marginal da inclusão de uma nova partição sobre a redução da fração equivalente.

Tabela IV

Frações Equivalentes para as Propostas de Dimensionamento

Proposta	f_{eq}	f_r
1/0.25	0.250	4.16
1/0.125	0.125	2.08
1/0.10	0.100	1.67
$2/f_{1w_{p1}}(0.25, 0.1)$	0.125	2.08
$2/f_{1w_{p2}}(0.25, 0.1)$	0.102	1.70
$2/f_{2w_{p2}}(1)$	0.156	2.60
$2/f_{3w_{c2}}(0.75)$	0.060	1.00
$2/f_{3w_{c1}}(0.90)$	0.172	2.87
$2/f_{3w_{c2}}(0.90)$	0.090	1.50
$3/f_{1w_{p1}}(0.25, 0.10)$	0.108	1.80
$3/f_{1w_{p2}}(0.25, 0.10)$	0.094	1.57
$3/f_{2w_{p2}}(1)$	0.136	2.27
$3/f_{3w_{c1}}(0.75)$	0.076	1.27
$3/f_{3w_{c1}}(0.90)$	0.142	2.37
$4/f_{2w_{p1}}(1)$	0.130	2.17

Como esperado, as propostas da menor fração equivalente estão associadas a alternativas de dimensionamento menos rigorosas em termos de precisão (por exemplo $f_{3w_{c2}}(0.75)$). Esta relação ficará mais clara com a apresentação dos resultados referentes à utilidade.

3.7.2 Utilidade

Na seção 2 não foram introduzidas medidas que sintetizassem a utilidade das in-

formações para o conjunto de subdivisões da população. O maior nível de agregação introduzido para a medida da utilidade foram as medidas síntese de precisão, S, definidas em 2.2.

Uma possibilidade de sintetizar a medida de precisão para o conjunto de municípios é utilizar a média das medidas síntese, ou ainda, a mediana destas. Face à assimetria desta variável na população de municípios, optou-se pela utilização desta última.

Existem, basicamente, duas formas para definição da mediana das medidas síntese: uma que leva em conta a participação percentual de cada município na população e outra que atribui o mesmo peso a todos municípios. Em ambas as opções as medidas síntese devem ser ordenadas, escolhendo-se o valor que deixa metade da população abaixo dele (considerando-se as populações dos municípios ordenados segundo a medida síntese) na primeira, e aquele que deixa metade dos municípios abaixo dele, na segunda. Em geral, face à correlação negativa existente entre o valor da medida síntese e o tamanho da população, a segunda medida tende a apresentar valores mais elevados que a primeira.

Tem-se aqui um problema técnico: o desconhecimento dos valores da medida síntese de precisão para o conjunto dos municípios obriga à utilização de procedimentos de estimação a partir da amostra de 64 municípios, utilizada no estudo. Para tal é necessário definir-se o peso associado a cada elemento da amostra, o que se faz a seguir.

Sejam $t_{(1)}, \cdots, t_{(64)}$ as populações dos municípios considerados na amostra ordenadas de modo crescente e sejam

$$\begin{aligned} F(t) &= \% \quad \text{de indivíduos em municípios com população} \quad \leq t; \text{ e} \\ G(t) &= \% \quad \text{de municípios com população} \quad \leq t \end{aligned} \tag{34}$$

então os pesos do i-ésimo município da amostra ordenada são dados por:

$$\begin{aligned} P_{1i} &= (F(t_{(i)} - F(t_{(i-1)}) \\ P_{2i} &= (G(t_{(i)} - G(t_{(i-1)}) \end{aligned} \qquad i = 1, \cdots, 64 \tag{35}$$

onde o primeiro peso é utilizado para cálculos que envolvam a ponderação por tamanho de população e o segundo quando não é adotada ponderação. Observe que $F(t)$ é a função F definida anteriormente para o estudo da partição ótima da população.

No estudo, as funções F e G eram desconhecidas, tendo sido estimada de modo idêntico ao descrito em 3.5, com a ressalva de que os dados utilizados referiam-se ao CD-80 (e não

à projeção para 89), já que seria inviável a obtenção de projeções de população para cada município da amostra.

Sejam então $S_1, \cdots, S_{64}$ os valores da medida síntese para cada município da amostra e $\hat{F}$ e $\hat{G}$ os estimadores das funções F e G utilizados. Ordenando-se os municípios pela medida síntese e fazendo $t_{S(i)}$ representar a população do i-ésimo município ordenado desta forma, tem-se que a mediana das medidas síntese, utilizando a ponderação por tamanho do município, é estimada por:

$$\hat{S}_p = S_{(k_p)}, \quad k_p = \inf\{k / \sum_{i=1}^{k} P_{1S(i)} \geq 0.5\} \tag{36}$$

e a mediana sem ponderação é estimada por:

$$\hat{S}_d = S_{(k_d)}, \quad k_d = \inf\{k / \sum_{i=1}^{k} P_{2S(i)} \geq 0.5\} \tag{37}$$

Foram utilizadas, aqui, quatro tipos de medida síntese: a perda média de eficiência absoluta associada à função peso $w_{p1}(P_{\dot{w}_{p_1}})$ e os CVs de ordem 50%, 75% e 90% (CV(50), CV(75) e CV(90) respectivamente).

A tabela V apresenta os resultados para os cálculos sem ponderação e a tabela VI os mesmos valores, só que com a utilização da ponderação por população. A exemplo do que se fez para a fração equivalente, calculou-se a medida síntese relativa, obtida através da divisão do valor original pelo valor mínimo, entre as propostas consideradas. Estes valores são indicados entre parênteses.

A análise das tabelas anteriores indica que:

- os valores das medidas síntese são sempre menores, quando é utilizada a ponderação por população. Isso ocorre porque, via de regra, os municípios maiores são aqueles para os quais as estimativas são obtidas com maior precisão;

- quando considerada a alternativa sem ponderação (tabela V), os menores valores das medidas síntese não estão associados aos maiores valores das frações equivalentes, como seria esperado. Isso sugere que a adoção de frações múltiplas

Tabela V

Medianas das Medidas de Precisão, sem Ponderação

Proposta	$P_{w_{c1}}$	$CV(50)$	$CV(75)$	$CV(90)$
1/0.25	26.54(1.13)	8.18(1.39)	11.75(1.07)	17.68(1.00)
1/0.125	43.73(1.87)	12.49(2.12)	17.95(1.62)	27.00(1.53)
1/0.10	50.05(2.14)	14.17(2.40)	20.35(1.84)	30.62(1.73)
$2/f_{1w_{p1}}(0.25, 0.1)$	43.73(1.87)	12.49(2.12)	17.95(1.62)	27.00(1.53)
$2/f_{1w_{p2}}(0.25, 0.1)$	43.73(1.87)	12.49(2.12)	17.95(1.62)	27.00(1.52)
$2/f_{2w_{p2}}(1)$	29.67(1.27)	8.46(1.44)	12.33(1.12)	20.87(1.18)
$2/f_{3w_{c2}}(0.75)$	54.52(2.33)	13.81(2.35)	21.37(1.94)	37.59(2.13)
$2/f_{3w_{c1}}(0.90)$	23.37(1.00)	5.88(1.00)	11.03(1.00)	21.16(1.20)
$2/f_{3w_{c2}}(0.90)$	40.03(1.71)	9.51(1.62)	16.13(1.46)	32.04(1.81)
$3/f_{1w_{p1}}(0.25, 0.1)$	41.19(1.76)	10.29(1.75)	15.19(1.38)	23.87(1.35)
$3/f_{1w_{p2}}(0.25, 0.1)$	47.06(2.01)	13.35(2.27)	19.19(1.74)	28.87(1.63)
$3/f_{2w_{p2}}(1)$	29.67(1.27)	8.46(1.44)	12.33(1.12)	20.87(1.18)
$3/f_{3w_{c1}}(0.75)$	43.98(1.88)	10.85(1.85)	18.32(1.66)	27.43(1.55)
$3/f_{3w_{c1}}(0.90)$	24.49(1.05)	5.88(1.00)	11.50(1.04)	21.54(1.22)
$4/f_{2w_{p1}}(1)$	29.88(1.28)	8.12(1.38)	11.76(1.07)	19.01(1.08)

e a partição ótima da população são estratégias poderosas para a redução do tamanho da amostra, sem perdas significativas de precisão;

• em geral, para um mesmo número das partições, a utilidade cresce com a fração equivalente.

Tabela VI

Medianas das Medidas de Precisão, Ponderação por População

Proposta	$P_{w_{p1}}$	$CV(50)$	$CV(75)$	$CV(90)$
1/0.25	9.05(1.00)	323(1.00)	7.21(1.00)	10.28(1.00)
1/0.125	19.36(2.14)	4.94(1.53)	11.01(1.53)	15.71(1.53)
1/0.10	23.99(2.65)	5.60(1.73)	12.49(1.73)	17.81(1.73)
$2/f_{1w_{p1}}(0.25, 0.1)$	19.36(2.14)	4.94(1.53)	11.01(1.53)	15.71(1.53)
$2/f_{1w_{p2}}(0.25, 0.1)$	19.59(2.16)	5.27(1.63)	11.00(2.53)	16.27(1.58)
$2/f_{2w_{p2}}(1)$	16.98(1.88)	4.57(1.41)	9.73(1.35)	14.61(1.42)
$2/f_{3w_{c2}}(0.75)$	35.69(3.94)	8.13(2.51)	15.35(2.13)	26.08(2.54)
$2/f_{3w_{c1}}(0.90)$	16.39(1.81)	4.88(1.51)	8.10(1.11)	15.32(1.49)
$2/f_{3w_{c2}}(0.90)$	29.68(3.28)	7.24(2.24)	12.10(1.68)	21.44(2.09)
$3/f_{1w_{p1}}(0.25, 0.1)$	19.36(2.14)	4.94(1.53)	11.01(1.53)	15.60(1.52)
$3/f_{1w_{p2}}(0.25, 0.1)$	21.96(2.43)	5.51(1.71)	11.77(1.63)	16.68(1.62)
$3/f_{2w_{p2}}(1)$	16.98(1.88)	4.68(1.45)	9.73(1.35)	14.52(1.41)
$3/f_{3w_{c1}}(0.75)$	27.69(3.06)	7.00(2.17)	13.17(1.83)	20.00(1.95)
$3/f_{3w_{c1}}(0.90)$	19.72(2.18)	5.22(1.62)	10.00(1.39)	15.96(1.55)
$4/f_{2w_{p1}}(1)$	14.65(1.62)	4.20(1.30)	9.31(1.29)	13.19(1.28)

- a fração única de 10% tem um desempenho medíocre, quando comparada com propostas de fração equivalente semelhantes;
- as diferenças entre as propostas ficam mais evidenciadas pela perda média de eficiência e pelo CV de 50%.

3.7.3 Homogeneidade

São utilizadas aqui as medidas de desigualdades de precisão apresentadas em 2.2. A exemplo do que foi feito em 3.7.2, são consideradas aqui quatro medidas síntese de precisão: $P_{w_{p1}}$, CV(50), CV(75) e CV(90) e utilizadas as ponderações pelas variáveis T_1 = população e $T_2 \equiv 1$ (ausência de ponderação).

As estatísticas definidas em 2.2 não podem ser diretamente utilizadas, pois não se dispõem dos valores populacionais necessários. A estimação dos índices de desigualdade é feita tomando-se a amostra como uma população de 64 municípios, substituindo os γ_j (ver 2.2) por P_{1j}, ou P_{2j} (ver (35), de 3.7.2), conforme a opção de ponderar-se por T_1, ou T_2, respectivamente e utilizando a expressão (14), de 2.2.

A tabela VII apresenta os resultados para os cálculos sem ponderação (T_2) e a tabela VIII os mesmos valores, utilizando-se a ponderação por população do município. São indicados, também, entre parênteses as medidas divididas pelo valor mínimo das propostas.

A análise das tabelas VII e VIII indica que:

- as medidas de desigualdade baseadas nos CV's de diferentes ordens têm um comportamento mais ligado ao número de partições, enquanto a medida associada à perda de eficiência parece ser influenciada também pela fração equivalente;
- os valores observados para as medidas de desigualdade variam de modo menos marcante que as medidas de utilidade;
- em geral, obtém-se um ganho de homogeneidade com a partição da população, mas este ganho é decrescente com o número de partições e até mesmo negativo, como no caso da última proposta (provavelmente a utilização de muitas partições ocasiona um problema de desigualdade de precisão nos municípios limítrofes aos limites de partição);
- as frações únicas têm desempenho idêntico, em termos de homogeneidade, quando analisadas as medidas associadas aos CV's;
- as melhores propostas, em termos de homogeneidade, são aquelas baseadas na fração necessária (f_3).

3.7.4 A Adequação Global das Propostas

Tabela VII

índices de Desigualdade de Precisão, sem Ponderação

Proposta	$P_{w_{p1}}$	$CV(50)$	$CV(75)$	$CV(90)$
1/0.25	0.29(2.64)	0.23(1.53)	0.30(1.43)	0.31(1.35)
1/0.125	0.22(2.00)	0.23(1.53)	0.30(1.43)	0.31(1.35)
1/0.10	0.20(1.82)	0.23(1.53)	0.30(1.43)	0.31(1.35)
$2/f_{1w_{p1}}(0.21)$	0.21(1.91)	0.22(1.47)	0.29(1.38)	0.31(1.35)
$2/f_{1w_{p2}}(0.25, 0.1)$	0.22(2.00)	0.23(1.53)	0.29(1.38)	0.31(1.35)
$2/f_{2w_{p2}}(1)$	0.20(1.82)	0.16(1.07)	0.24(1.14)	0.26(1.13)
$2/f_{3w_{c2}}(0.75)$	0.11(1.00)	0.17(1.13)	0.21(1.00)	0.24(1.04)
$2/f_{3w_{c1}}(0.90)$	0.26(2.36)	0.18(1.20)	0.22(1.05)	0.24(1.04)
$2/f_{3w_{c2}}(0.90)$	0.16(1.45)	0.15(1.00)	0.21(1.00)	0.23(1.00)
$3/f_{1w_{p1}}(0.25, 0.1)$	0.20(1.82)	0.20(1.33)	0.27(1.29)	0.29(1.26)
$3/f_{1w_{p2}}(0.25, 0.1)$	0.19(1.73)	0.20(1.33)	0.27(1.29)	0.29(1.26)
$3/f_{2w_{p2}}(1)$	0.20(1.82)	0.16(1.07)	0.24(1.14)	0.26(1.13)
$3/f_{3w_{c1}}(0.75)$	0.15(1.36)	0.16(1.07)	0.21(1.00)	0.27(1.17)
$3/f_{3w_{c1}}(0.90)$	0.26(2.36)	0.17(1.13)	0.21(1.00)	0.23(1.00)
$4/f_{2w_{p1}}(1)$	0.20(1.82)	0.15(1.00)	0.23(1.10)	0.26(1.13)

A tarefa de comparar as propostas a partir das diversas medidas obtidas é bastante complexa, pois, como visto, as adequações das mesmas, nos diferentes requisitos, caminham em direções opostas. Assim, em geral, uma proposta com bom desempenho em relação ao custo, não será muito adequada em termos de utilidade, e assim por diante.

Como então avaliar as propostas em termos de sua adequação conjunta com relação

Tabela VIII

índices de Desigualdade de Precisão, Ponderação por População

Proposta	$P_{w_{p1}}$	$CV(50)$	$CV(75)$	$CV(90)$
1/0.25	0.46(2.19)	0.35(1.94)	0.30(2.00)	0.36(1.89)
1/0.125	0.37(1.76)	0.35(1.94)	0.30(2.00)	0.36(1.89)
1/0.10	0.35(1.67)	0.35(1.94)	0.30(2.00)	0.36(1.89)
$2/f_{1w_{p1}}(0.25, 0.1)$	0.37(1.76)	0.34(1.89)	0.29(1.93)	0.36(1.89)
$2/f_{1w_{p2}}(0.25, 0.1)$	0.29(1.38)	0.30(1.67)	0.24(1.60)	0.30(1.58)
$2/f_{2w_{p2}}(1)$	0.35(1.67)	0.29(1.61)	0.25(1.67)	0.32(1.68)
$2/f_{3w_{c2}}(0.75)$	0.25(1.19)	0.28(1.56)	0.23(1.53)	0.30(1.58)
$2/f_{3w_{c1}}(0.90)$	0.21(1.00)	0.20(1.11)	0.15(1.00)	0.20(1.05)
$2/f_{3w_{c2}}(0.90)$	0.22(1.05)	0.24(1.33)	0.19(1.27)	0.26(1.37)
$3/f_{1w_{p1}}(0.25, 0.1)$	0.30(1.43)	0.29(1.61)	0.23(1.53)	0.30(1.58)
$3/f_{1w_{p2}}(0.25, 0.1)$	0.29(1.38)	0.30(1.67)	0.24(1.60)	0.31(1.63)
$3/f_{2w_{p2}}(1)$	0.28(1.33)	0.26(1.44)	0.20(1.33)	0.27(1.42)
$3/f_{3w_{c1}}(0.75)$	0.23(1.10)	0.25(1.39)	0.20(1.33)	0.29(1.53)
$3/f_{3w_{c1}}(0.90)$	0.21(1.00)	0.18(1.00)	0.15(1.00)	0.19(1.00)
$4/f_{2w_{p1}}(1)$	0.31(1.45)	0.27(1.50)	0.21(1.40)	0.28(1.47)

aos requisitos básicos? Uma alternativa, é sintetizar a informação das diferentes medidas em um único índice. Para fazer isso, deve-se cuidar para que haja uma certa padronização entre as medidas escolhidas para representar o desempenho da proposta em cada um dos requisitos. Uma sugestão natural é utilizar as medidas relativas, definidas para cada índice de desempenho como a razão entre ele e o valor mínimo para as propostas consideradas.

Sejam m_{ci}, $m_{ui}(S)$ e $m_{hi}(S)$ as medidas relativas são de desempenho correspondentes, respectivamente, a custos, utilidade e homogeneidade, associadas à i-ésima proposta, sendo as duas últimas baseadas na medida síntese S. Então o índice de adequação global da proposta i é definido por:

$$A_i = (\alpha_c \cdot m_{ci} + \alpha_u \cdot m_{ui}(S) + \alpha_h \cdot m_{hi}(S))^{-1} \tag{37}$$

onde α_c, α_u e α_h são pesos associados ao custo, à utilidade e à homogeneidade ($\alpha_c + \alpha_u + \alpha_h = 1$).

Tendo em vista que, por definição, as medidas relativas são sempre maiores ou iguais a um, o índice A_i toma sempre valores entre zero e um, sendo tão mais próximo de um quanto melhor o seu desempenho global ($A_i = 1 \Longleftrightarrow m_{ci} = m_{ui}(S) = m_{hi}(S) = 1$, o valor mínimo).

Se ao menos uma das medidas relativas apresentar valores excessivamente altos, A_i apresentará valores pequenos. Os pesos α são obviamente subjetivos, devendo retratar as importâncias dos diferentes requisitos na pesquisa. Neste estudo, adotou-se $\alpha_c = 3/7, \alpha_u = 3/7$ e $\alpha_h = 1/7$, incluindo uma maior importância relativa para os dois primeiros requisitos. Os índices de adequação global para as medidas síntese utilizadas foram calculados com base nas medidas sem ponderação. Os valores ponderados por população resultaram em índices de adequação praticamente idênticos aos apresentados, sendo omitidos.

A análise da tabela IX revela de forma contudente a **importância do uso de frações múltiplas de amostragem como instrumento de melhoria da adequação global** de uma pesquisa. O uso de partições da população para fins de aplicação das frações amostragem acarreta ganhos consideráveis de adequação conjunta. Observa-se, contudo, que este ganho existe até um determinado número de partições, a partir do qual parece haver um decréscimo na adequação conjunta.

O uso de três frações não parece trazer ganhos expressivos ao que já se obtinha com duas frações, sendo que quatro frações parecem implicar em uma piora da adequação global.

Na análise efetuada, não foi levado em conta o requisito de operacionalidade, devido à complexidade de quantificação do mesmo. Pode-se, contudo, fazer uma avaliação qualitativa, considerando-se os resultados da adequação global. Sem dúvida, se duas propostas oferecem a mesma adequação, a de menor número de frações deve ser preferida.

Tabela IX

Indices de Adequação Global

Proposta	$P_{w_{p1}}$	$CV(50)$	$CV(75)$	$CV(90)$
1/0.25	0.38	0.38	0.41	0.42
1/0.125	0.51	0.50	0.59	0.57
1/0.10	0.53	0.51	0.59	0.61
$2/f_{1w_{p1}}(0.25, 0.1)$	0.51	0.50	0.56	0.57
$2/f_{1w_{p2}}(0.25, 0.1)$	0.55	0.54	0.62	0.64
$2/f_{2w_{p2}}(1)$	0.52	0.53	0.57	0.56
$2/f_{3w_{c2}}(0.75)$	0.64	0.63	0.71	0.67
$2/f_{3w_{c1}}(0.90)$	0.50	0.55	0.55	0.53
$2/f_{3w_{c2}}(0.90)$	0.63	0.68	0.71	0.64
$3/f_{1w_{p1}}(0.25, 0.1)$	0.56	0.58	0.65	0.65
$3/f_{1w_{p2}}(0.25, 0.1)$	0.56	0.54	0.62	0.64
$3/f_{2w_{p2}}(1)$	0.56	0.57	0.62	0.61
$3/f_{3w_{c1}}(0.75)$	0.65	0.67	0.71	0.73
$3/f_{3w_{c1}}(0.90)$	0.55	0.62	0.62	0.59
$4/f_{2w_{p1}}(1)$	0.58	0.60	0.65	0.64

Esta avaliação aponta na direção do uso de duas frações diferenciadas, sendo acatada na próxima seção, quando se apresenta uma sugestão de dimensionamento para o CD-90.

Como última observação, nota-se a primazia das propostas dimensionadas a partir da fração necessária (f_3) sobre as demais, indicando ser este, aparentemente, o melhor método de dimensionamento, quando considerados conjuntamente os requisitos básicos da pesquisa.

4 - CONCLUSÕES - A ESPECIFICAÇÃO DA PROPOSTA DE DIMENSIONAMENTO

Nas duas seções anteriores, a questão do dimensionamento da amostra para o CD-90 foi tratada em um contexto estritamente técnico. Diversas alternativas de dimensionamento foram analisadas segundo um conjunto de medidas definidas ao longo ao texto. Ao final da seção anterior, procurou-se unificar as diferentes formas de medição da qualidade das propostas em um índice de adequação global. Este tem por objetivo auxiliar a compreensão dos diferentes desempenhos das propostas (segundo os requisitos básicos definidos na introdução), não podendo contudo ser utilizado como o instrumento único para a escolha da melhor proposta. De fato, a definição dos pesos que entram na composição do índice encerra certa dose de subjetivismo, e certamente variarão de acordo com quem analise os dados aqui apresentados.

Algumas conclusões fundamentais podem, contudo, ser extraídas do estudo realizado:

i - o uso de frações múltiplas de amostragem, em número moderado (duas, ou no máximo, três), traz grandes benefícios à qualidade global da pesquisa, permitindo uma flexibilidade que é essencial para um levantamento de natureza tão complexa e diversificada quanto o CD-90;

ii - a fração cogitada para o próximo CD(10%) é bastante inferior, em termos de desempenho global, às propostas com duas ou três frações. Em todos os dimensionamentos analisados, a fração de 10% mostrou-se insuficiente para os municípios pequenos (até 15 000 habitantes);

iii - a fração única de 25%, utilizada no último CD, foi a de pior desempenho global. Isso se deve ao fato dela ser excessiva para os municípios médios e grandes, provocando grande desigualdade de precisão e representando um gasto excessivo e, principalmente, desnecessário de recursos;

iv - houve grande estabilidade dos limites de partição das propostas de dimensionamento, sugerindo a existência de dois valores fundamentais: 15 000 habitantes, abaixo do qual se encontram os municípios com exigência de maior fração amostral, e 200 000 habitantes, separando os municípios médios dos grandes. Obviamente a classificação aqui sugerida deve ser entendida no contexto deste estudo, cujo o objetivo é o dimensionamento de frações amostrais;

v - entre os municípios pequenos, a fração de 1/4 revelou-se plenamente satisfatória, enquanto para os municípios médios esta oscilou, de acordo com a abordagem utilizada, entre 1/6 e 1/20. Para os municípios grandes, a fração de 1/15 é, quase sempre, suficiente;

vi - os ganhos de adequação global ao se passar de uma para duas frações amostrais são expressivos, mas de duas para três são insignificantes e, ao passar-se de três para quatro frações verificam-se algumas perdas. Isso parece sugerir que o número mais adequado de frações é de dois ou três. Levando-se em conta aspectos associados à operacionalidade, a escolha recai naturalmente sobre duas frações amostrais.

No momento da especificação de uma proposta de dimensionamento para o CD-90 outras considerações, que não tão somente as técnicas, devem ser feitas. Existe, neste momento, um certo impasse com relação à definição da fração de amostragem para o próximo censo, conforme mecionado diversas vezes no texto. De um lado colocam-se os defensores da fração tradicional, preocupados com a possível perda de utilidade dos dados levantados, se utilizada a fração cogitada a princípio (10%). De outro situam-se os defensores da mudança da fração, advogando a suficiência dos 10% para fins de estimação das características de interesse, nos níveis de divulgação tradicionalmente utilizados pelo IBGE, com o forte atrativo da redução de gastos.

Claramente, uma solução intermediária entre as duas posições deve ser buscada, de tal forma que a redução dos gastos se harmonize com a utilidade dos dados levantados. Uma tal solução seria a utilização de duas frações amostrais diferenciadas, uma para os municípios pequenos, para os quais os problemas de precisão são mais sentidos, e outra para os municípios médios e grandes (a hipótese de três ou mais frações foi descartada pelos argumentos apresentados anteriormente).

Entre as propostas analisadas, duas se destacam em termos de adequação conjunta aos requisitos básicos mencionados na introdução:

$2 \setminus f_{3w_{c2}}(0.75)$ - limite de 15 000 habitantes e frações 1/8 e 1/20;

$2 \setminus f_{3w_{c2}}(0.90)$ - limite de 15 000 habitantes e frações 1/4 e 1/15;

ambas possuem frações equivalentes (ver 2.3) inferiores aos 10% cogitados para o próximo censo (6% e 9% respectivamente). Analisando melhor as duas frações, verifica-se que a

primeira delas apresenta frações inferiores à totalidade das propostas de duas partições (de fato a fração de 1/8 mostrou-se insuficiente para os municípios pequenos na grande maioria das propostas). Seria, então, desaconselhável optar por esta proposta, já que, sem dúvida, seu grande atrativo é a baixa fração equivalente, mas esta não fornece garantia adequada de utilidade para as informações coletadas.

Resta, então, a segunda proposta de dimensionamento. Conforme salientado acima a fração equivalente associada é inferior aos 10% cogitados. De fato a passagem de 25% para 10% já representa uma economia enorme de recursos e talvez não faça sentido, ao menos para este censo, utilizar propostas com frações equivalente ainda inferiores a 10%. Desta forma, a proposta escolhida mantém, basicamente, a estrutura da segunda proposta citada, fazendo-se apenas uma modificação na segunda fração, no sentido de tornar a fração equivalente igual a 10%.

A proposta de dimensionamento assim obtida foi:

i - duas frações de amostragem - 1/4 e 1/12;

ii - limite entre partições - 15 000 habitantes.

As medidas de desempenho para a proposta sugerida são apresentadas mais adiante.

Observa-se, assim, que a proposta sugerida mantém praticamente intactos os índices de desempenho global da melhor das propostas consideradas no estudo, com a vantagem de utilizar uma fração equivalente igual à inicialmente cogitada para o CD-90 (com conseqüente melhora da precisão das estimativas).

Fração Equivalente - 0.10

Medidas de Precisão

	s/ponderação	c/ponderação
$P_{w_{p1}}$	38.85(1.66)	26.16(2.89)
$CV(50)$	9.10(1.55)	6.46(2.00)
$CV(75)$	14.27(1.29)	11.62(1.61)
$CV(90)$	28.40(1.61)	19.72(1.92)

Medidas de Desigualdade

	s/ponderação	c/ponderação
$P_{w_{p1}}$	0.17(1.55)	0.28(1.33)
$CV(50)$	0.15(1.00)	0.28(1.56)
$CV(75)$	0.21(1.00)	0.22(1.47)
$CV(90)$	0.24(1.04)	0.29(1.53)

Indice de Adequação Global

$P_{w_{p1}}$	0.61
$CV(50)$	0.66
$CV(75)$	0.71
$CV(90)$	0.64

Esperamos que este trabalho venha a contribuir no difícil processo de delineamento do desenho da amostra a ser adotado para o próximo Censo Demográfico. Sem dúvida, a decisão final deverá também atender a questões de natureza gerencial, que fogem ao escopo deste estudo. A colocação das diversas vantagens e desvantagens associadas às possíveis escolhas, de forma clara, objetiva e quantificada, deve, contudo, auxiliar na escolha final.

Obs.: Após a elaboração da primeira versão deste relatório, os autores tomaram conhecimento da decisão gerencial acerca do plano amostral para o Censo Demográfico de 1990. A proposta adotada foi:

i) - duas frações de amostragem - 1/5 e 1/10;

ii) - limite entre partições - 15.000 habitantes.

São incluídas aqui as medidas de desempenho correspondentes a esta última proposta:

Fração Equivalente - 0.11

Medidas de Precisão

	s/ponderação	c/ponderação
$P_{w_{p1}}$	36.95(1.58)	23.18(2.56)
$CV(50)$	9.78(1.66)	5.60(1.73)
$CV(75)$	15.11(1.37)	11.23(1.56)
$CV(90)$	25.82(1.46)	17.95(1.75)

Medidas de Desigualdade

	s/ponderação	c/ponderação
$P_{w_{p1}}$	0.17(1.55)	0.31(1.48)
$CV(50)$	0.15(1.00)	0.29(1.61)
$CV(75)$	0.23(1.10)	0.24(1.60)
$CV(90)$	0.26(1.13)	0.31(1.63)

Indice de Adequação Global

$P_{w_{p1}}$	0.58
$CV(50)$	0.60
$CV(75)$	0.64
$CV(90)$	0.63

Verifica-se, assim, que o plano amostral adotado possui adequação global inferior à proposta sugerida no trabalho, especialmente com relação às medidas $CV(50)$ e $CV(75)$. Contudo, deve-se lembrar que a decisão final, conforme mencionado ao longo do texto, não se associa apenas à análise técnica da proposta, mas também a aspectos gerenciais não contemplados aqui.

5 - REFERÊNCIAS BIBLIOGRÁFICAS

Anand, S. *Inequality and Poverty in Malaysia: Measurement and Decomposition.* Oxford University Press, 1983

Cochran, William Gemmell. *Sample Tecniques* 3 rd. ed. New York, Jo Wiley, c-1977

Abramowitz, M. & Segun A. I. *Handbook of Mathematical Functions.* Dover Publications, Inc., New York

Fundação Instituto Brasileiro de Geografia e Estatística. *Metodologia do Censo Demográfico de 1980.* Série Relatórios Metodológicos/IBGE, ISSN, 0101-2843, v.4 - Rio de Janeiro, 1983

Pinheiro, José Carlos C. & Lima, José Matias. *Avaliação dos Efeitos de Redução da Fração de Amostragem.* Série Relatórios Técnicos/IBGE, ENCE, nº 004/88 - Rio de Janeiro, 1988

ANEXO 6.1. *VARIÁVEIS CONSIDERADAS NO ESTUDO E ESTATÍSTICAS RESUMO ADOTADAS*

ANEXO 6.1. *VARIÁVEIS CONSIDERADAS NO ESTUDO E ESTATÍSTICAS RESUMO ADOTADAS*

Características de Domicílios

- Domicílios particulares ocupados duráveis (Mediana)
- Domicílios particulares ocupados rústicos (Mediana)
- Domicílios particulares permanentes (Mediana)
- Domicílios particulares permanentes alugados (3o. quartil)
- Domicílios particulares permanentes próprios (3o. quartil)
- Domicílios particulares permanentes com canalização interna (Mediana)
- Domicílios particulares permanentes sem canalização interna (Mediana)
- Domicílios particulares permanentes com instalação sanitária só do domicílio (Mediana)
- Domicílios particulares permanentes com instalação sanitária comum a mais de um domicílio (Mediana)
- Domicílios particulares permanentes sem instalação sanitária
- Domicílios particulares permanentes com equipamentos para cozinha (3o. quartil)
- Domicílios particulares permanentes segundo combustível utilizado para cozinhar (3o. quartil)
- Domicílios particulares permanentes com iluminação elétrica, com medidor
- Domicílios particulares permanentes com iluminação elétrica, sem medidor
- Domicílios particulares permanentes com telefone
- Domicílios particulares permanentes com rádio
- Domicílios particulares permanentes com geladeira
- Domicílios particulares permanentes com televisão (Mediana)
- Domicílios particulares permanentes com automóvel (Mediana)
- Domicílios particulares permanentes segundo o tempo de residência no domicílio (Mediana)

Características de Pessoas

- Pessoas economicamente ativas por setor de atividade (Mediana)
- Pessoas não economicamente ativas por setor de atividade (Mediana)
- Homens por rendimento médio mensal (3o. quartil)
- Mulheres por rendimento médio mensal (3o. quartil)
- Homens por setor de atividade (3o. quartil)
- Mulheres por setor de atividade (3o. quartil)
- Homens por religião (Mediana)
- Mulheres por religião (Mediana)
- Homens por cor (3o. quartil)
- Mulheres por cor (3o. quartil)
- Homens por estado conjugal (3o. quartil)
- Mulheres por estado conjugal (3o. quartil)
- Homens por ano de estudo (3o. quartil)
- Mulheres por ano de estudo (3o. quartil)
- Homens por grau de escolaridade completo (3o. quartil)
- Mulheres por grau de escolaridade completo (3o. quartil)
- Mulheres de 15 anos ou mais de idade que tiveram filhos
- Filhos tidos nascidos vivos
- Filhos tidos nascidos mortos
- Filhos vivos

Observação: quando não for indicada a estatística resumo é por que a característica não apresenta subdivisões em níveis, no plano tabular considerado.

ANEXO 6.2. *MUNICIPIOS CONSIDERADOS NO ESTUDO E POPULAÇÃO, SEGUNDO O CENSO DEMOGRÁFICO - 80*

ANEXO 6.2. *MUNICIPIOS CONSIDERADOS NO ESTUDO E POPULAÇÃO, SEGUNDO O CENSO DEMOGRÁFICO - 80*

Água Comprida	MG	1968	Delmiro Gouveia	AL	26812
Bento de Abreu	SP	2043	Navirai	MS	28574
Ilha de Paquetá (R.A.)	RJ	2216	Piraí	RJ	28815
Nova Araca	RS	2263	Mariana	MG	29426
Barra de São Miguel	AL	2334	Colider	MT	34638
Mairipotaba	GO	2681	Presidente Dutra	MA	39868
Afonso Cunha	MA	3428	Concordia	SC	59432
Caracol	MS	3820	Cachoeirinha	RS	63228
Victor Graeff	RS	4119	Araras	SP	65251
Santa Cruz do Arari	PA	4609	Bacabal	MA	81685
Anaurilândia	MS	7219	Caucaia	CE	94157
Vianópolis	GO	7929	Bagé	RS	100190
Dois Riachos	AL	7966	Americana	SP	122055
Duas Barras	RJ	7994	Caxias	MA	125771
Abadia dos Dourados	MG	8012	Juazeiro do Norte	CE	135687
Bernardino de Campos	SP	8989	São Caetano do Sul	SP	163030
Sumidouro	RJ	11398	Anápolis	GO	179973
Cachaoeira do Arari	PA	11463	Florianópolis	SC	187800
Petrolina de Goiás	GO	11621	São José do Rio Preto	SP	188560
Cedral	MA	12837	Uberaba	MG	199265
Tocantis	MG	13188	Juiz de Fora	MG	307820
Santana do Mundau	AL	13588	Ribeirão Preto	SP	318375
Monte Mor	SP	14008	Maceió	AL	400041
Aparecida do Taboado	MS	14070	Brasília	DF	410999
Pinheiro Machado	RS	14366	São Luiz	MA	449877
Salinópolis	PA	14415	Bangu (R.A.)	RJ	530249
Mirandópolis	SP	21534	Guarulhos	SP	532908
Piracanjuba	GO	24119	Campinas	SP	664356
Guapore	RS	24614	Curitiba	PR	1025979
Altamira do Maranhão	MA	24877	Porto Alegre	RS	1125901
São Mateus do Maranhão	MA	25005	Recife	PE	1204738
Bujaru	PA	26079	Fortaleza	CE	1308919

ANEXO 6.3. *FIGURAS APRESENTADAS NO TEXTO*

FIGURA 1 - CURVA DE PERDA DE EFICIENCIA ABSOLUTA

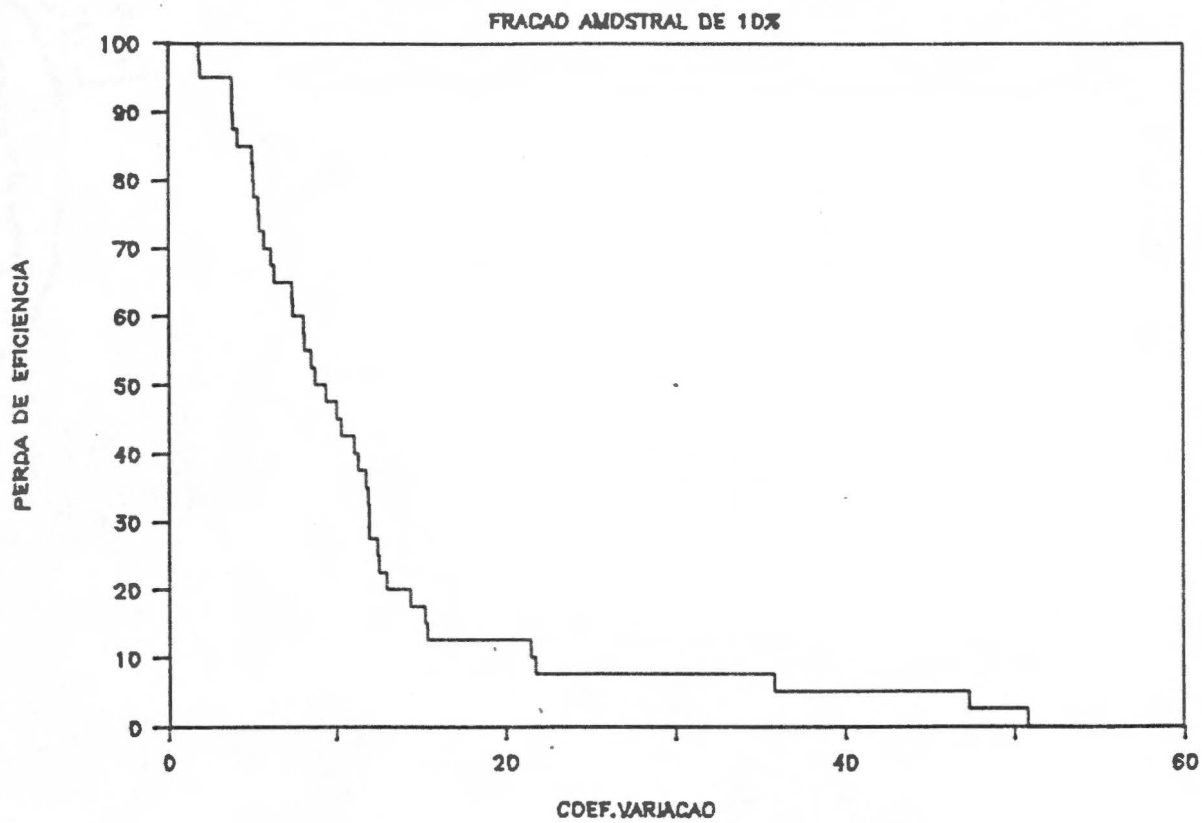


FIGURA 2 - CURVA DE PERDA DE EFICIENCIA ABSOLUTA

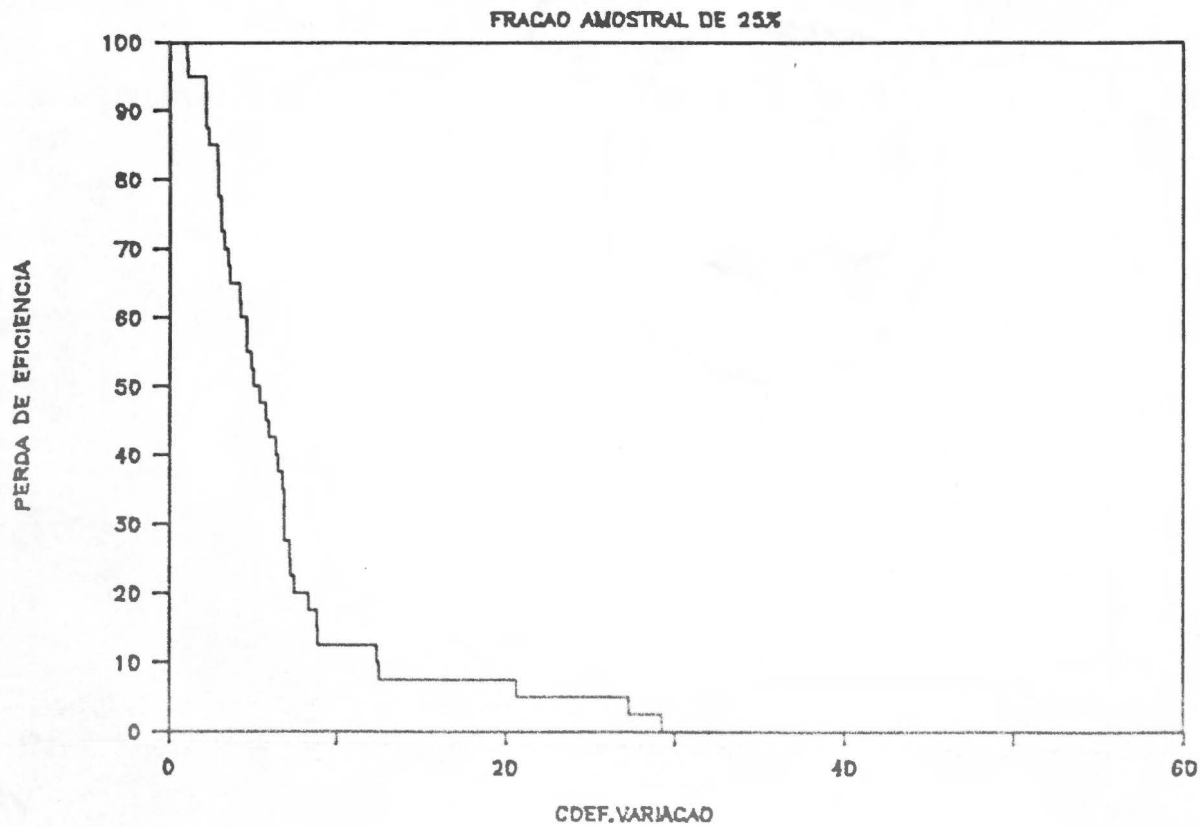


FIGURA 3 - CURVA DE PERDA DE EFICIENCIA RELATIVA

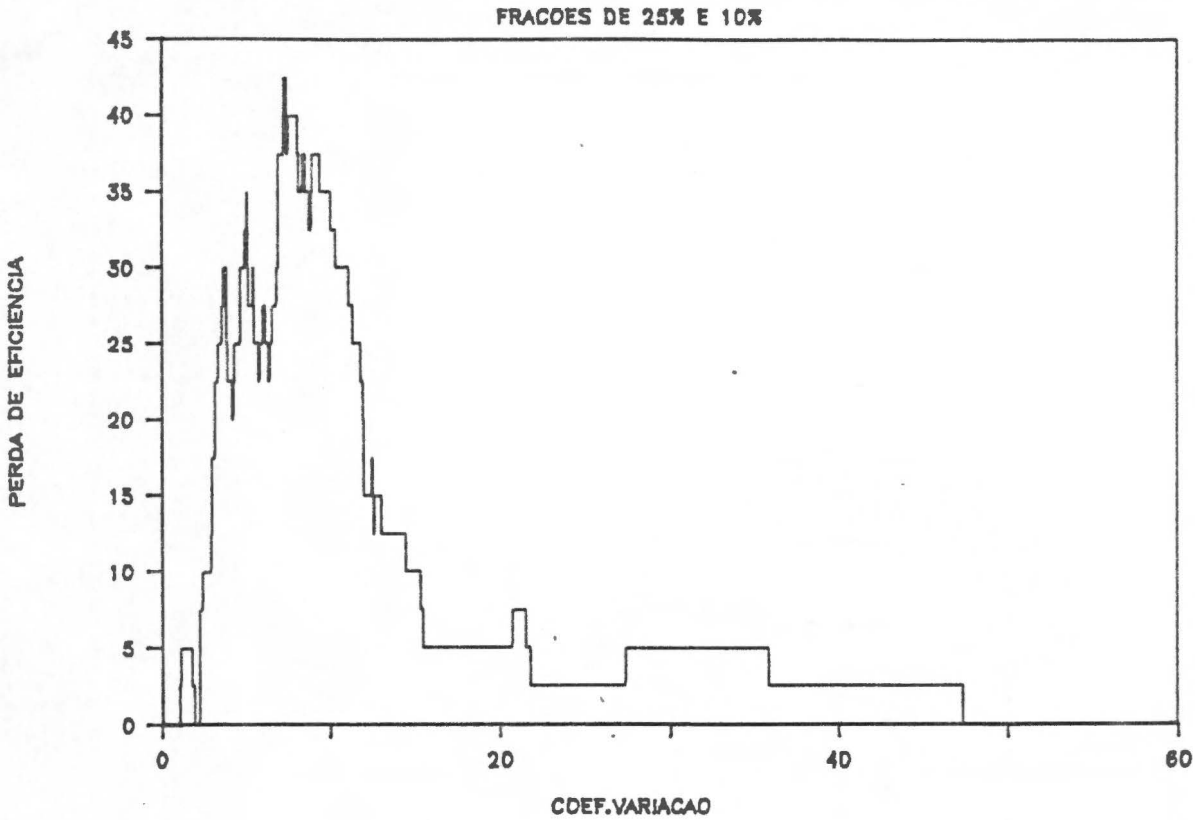


FIGURA 4 - FUNCAO PESO DO TIPO RETANGULAR

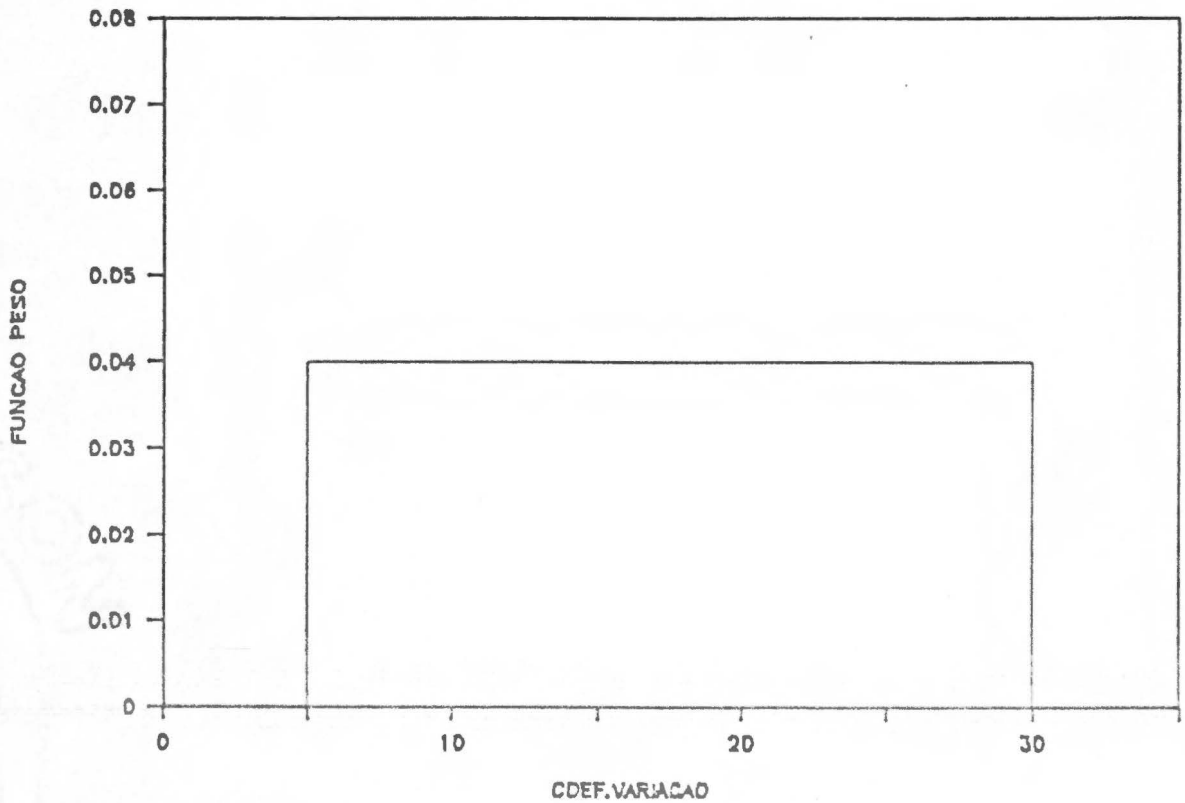


FIGURA 5 - FUNCAO PESO DO TIPO TRAPEZOIDAL

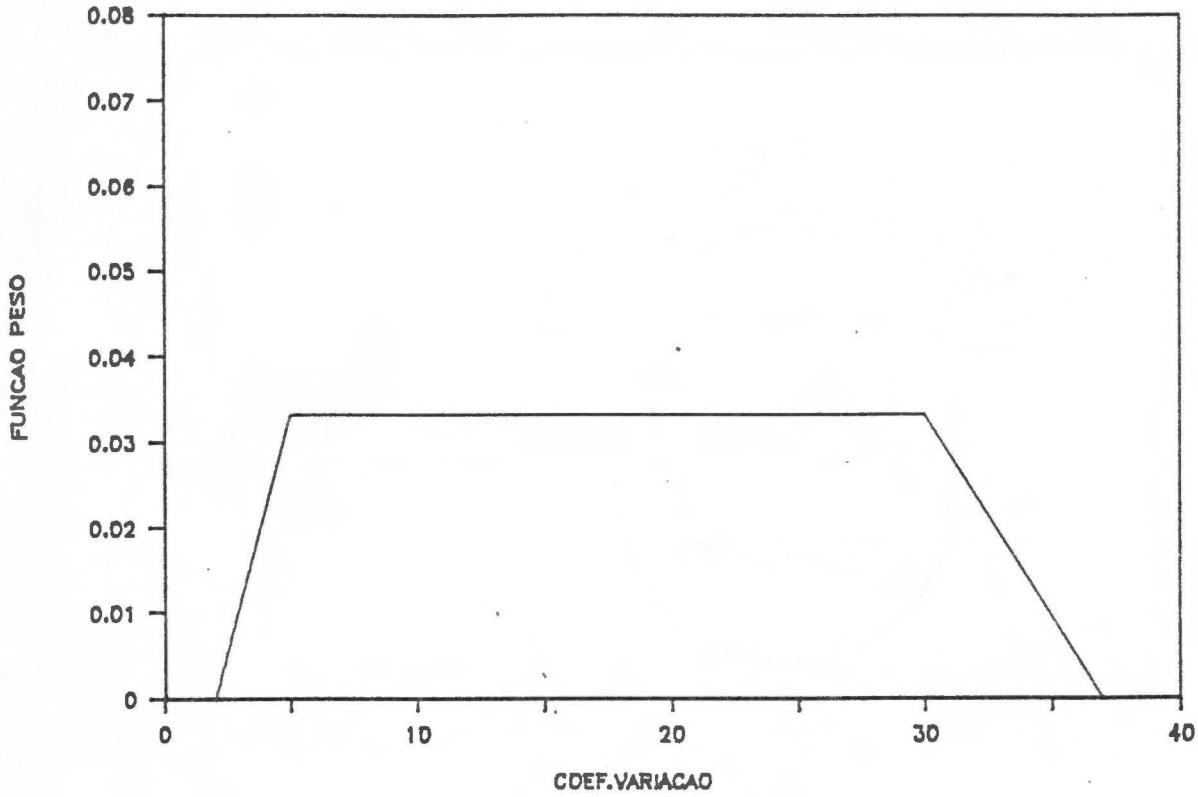


FIGURA 6 - FUNCAO PESO DO TIPO BETA

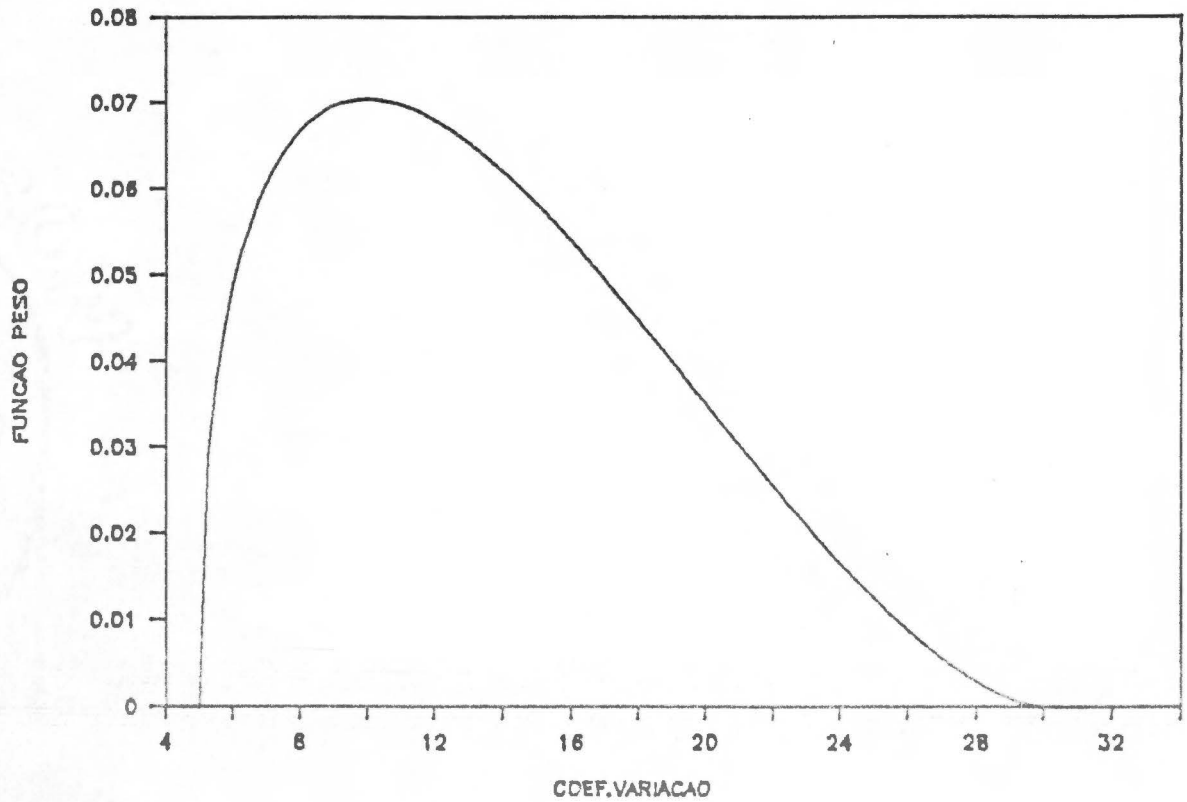


FIGURA 7 - CURVA DE EFICIENCIA ABSOLUTA PONDERADA

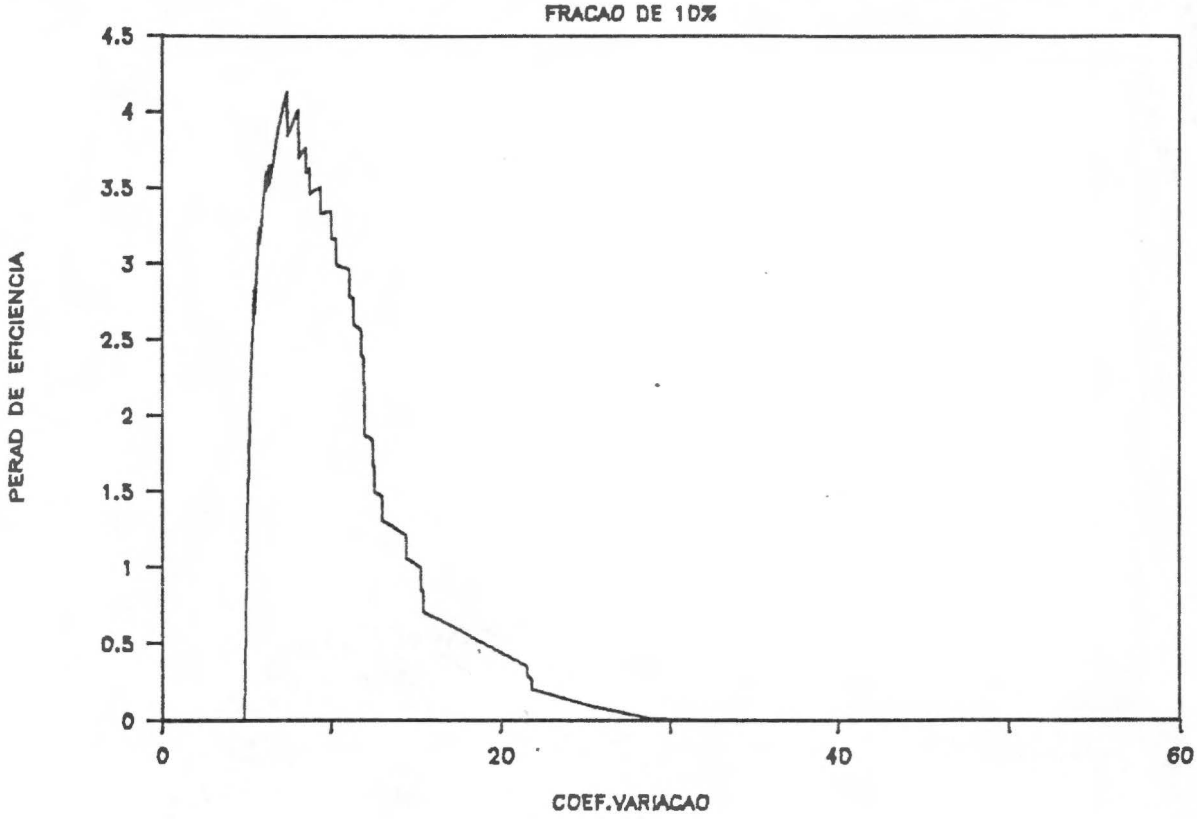


FIGURA 8 - CURVA DE EFICIENCIA RELATIVA PONDERADA

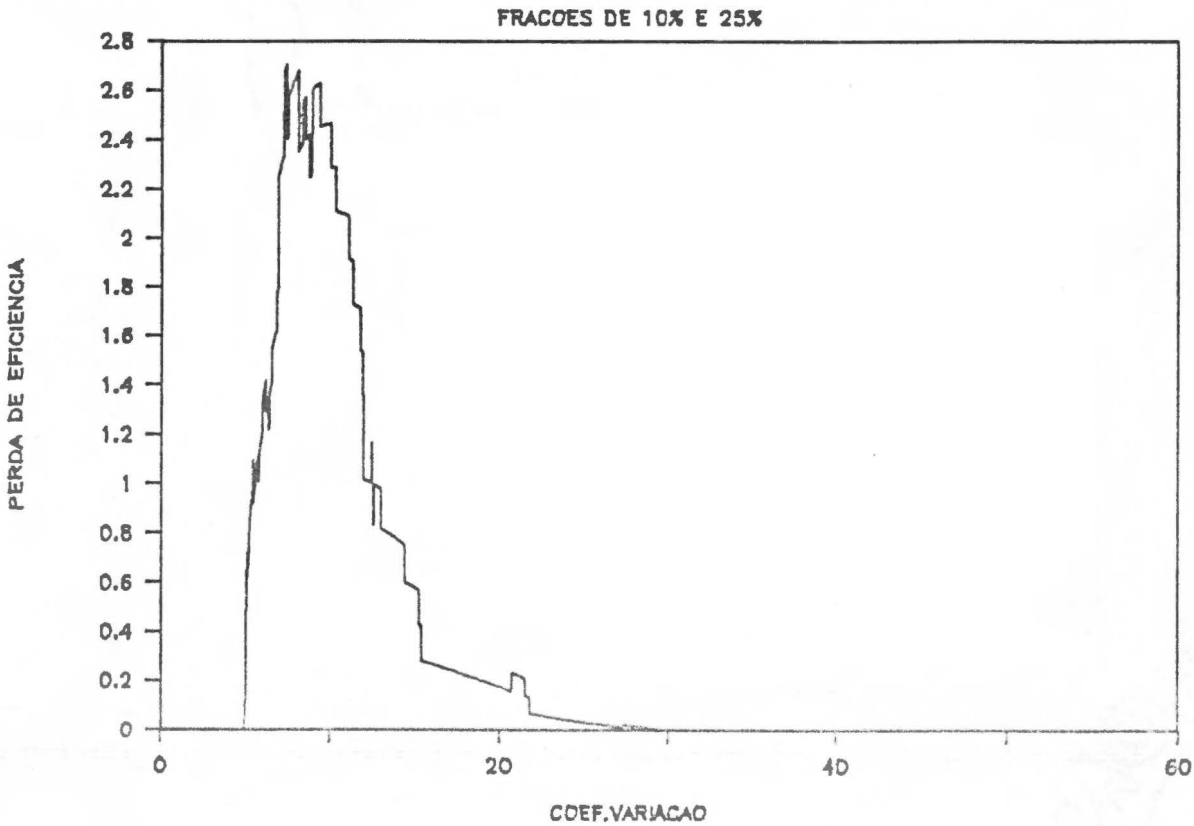


FIGURA 9 - PERDA DE EFICIENCIA EM FUNCAO DA FRACAO

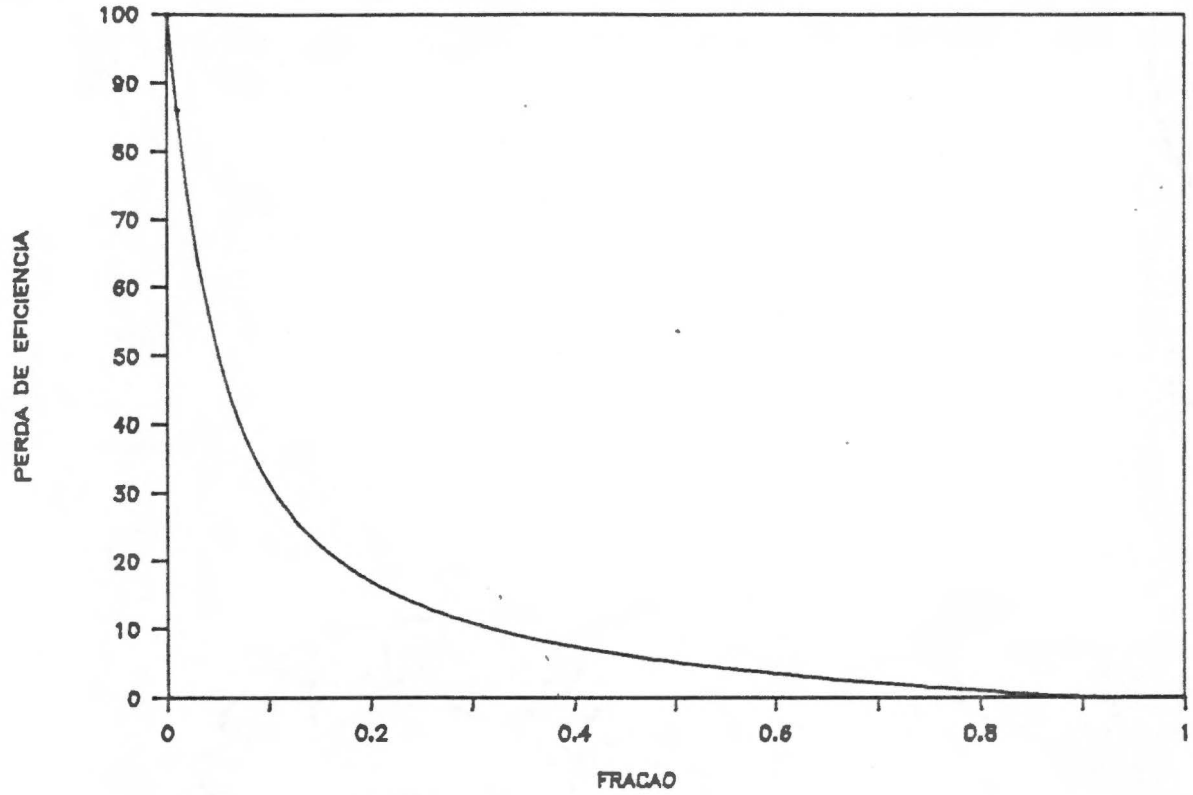


FIGURA 10 - FRACAO NECESSARIA EM FUNCAO DO CV

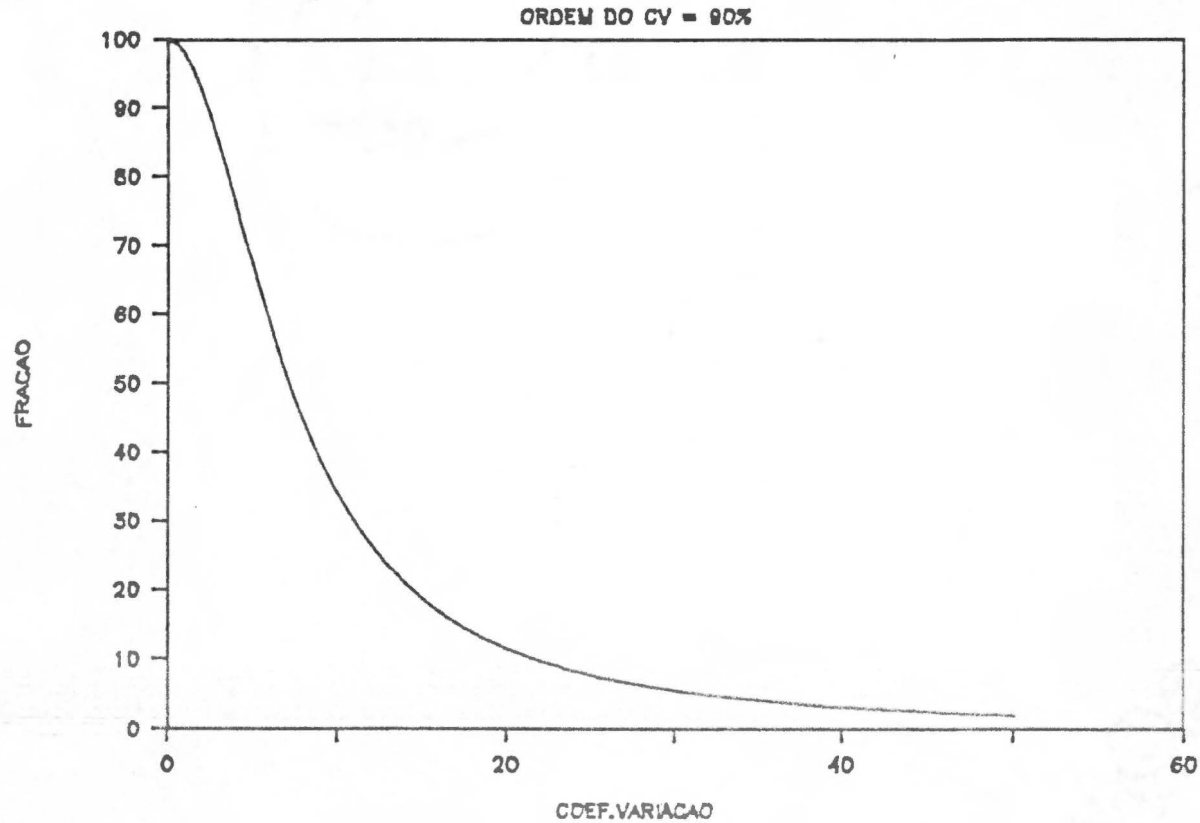


FIGURA 11 -

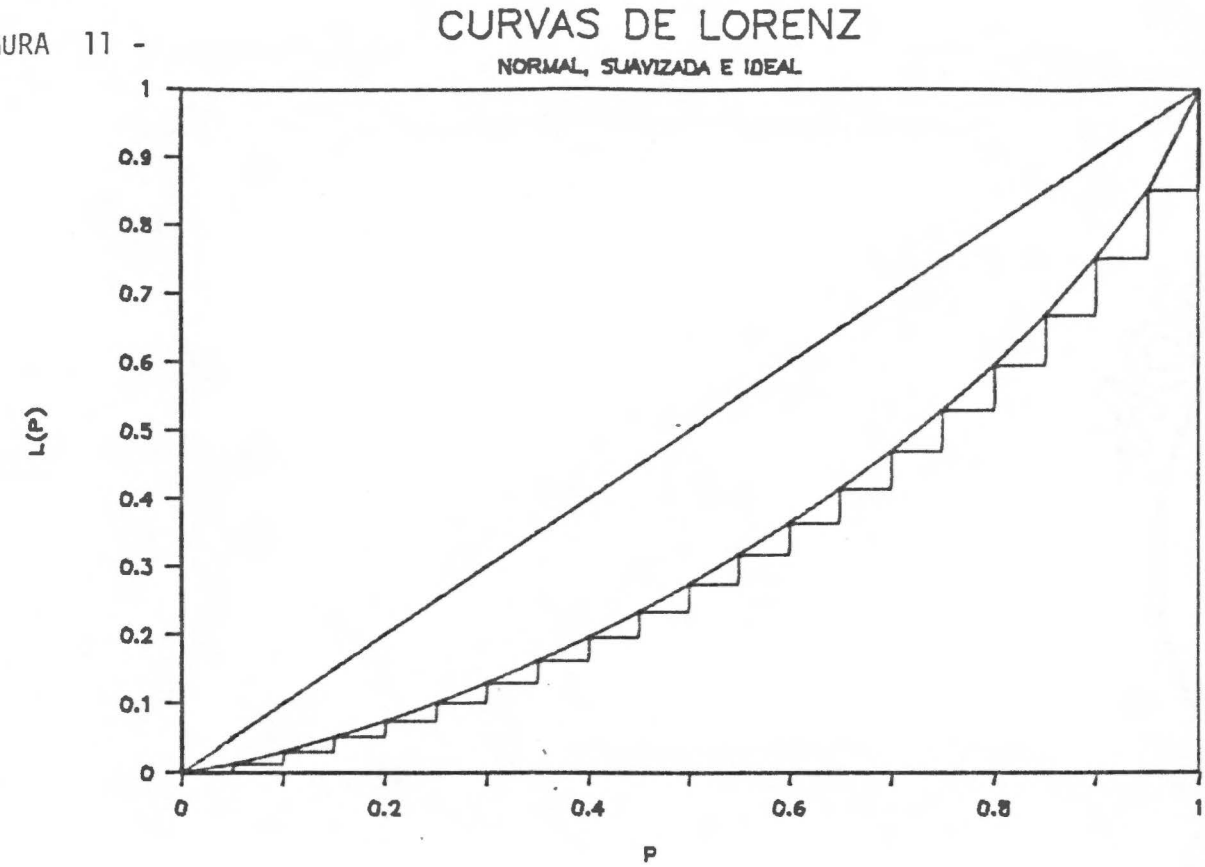


FIGURA 12 -

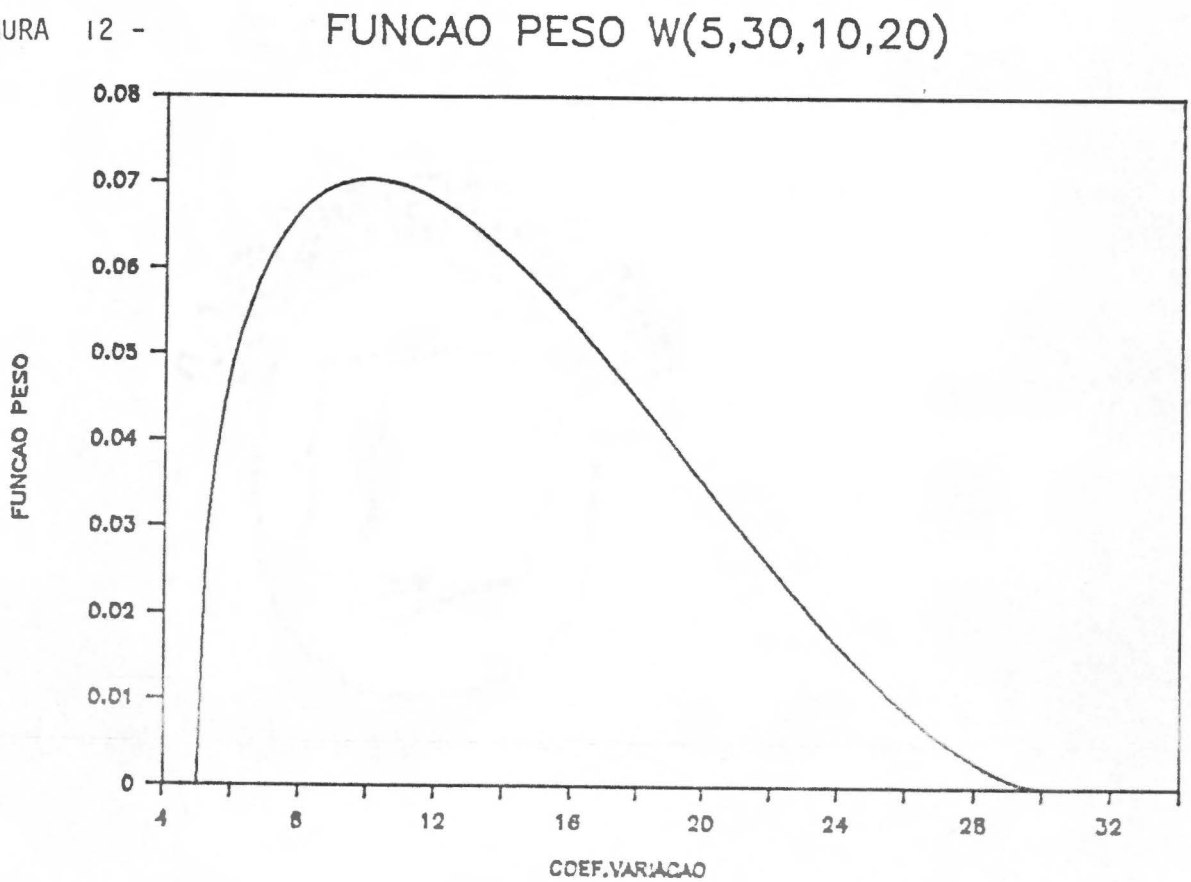


FIGURA 13 - FUNCAO PESO W(2,40,15,25)

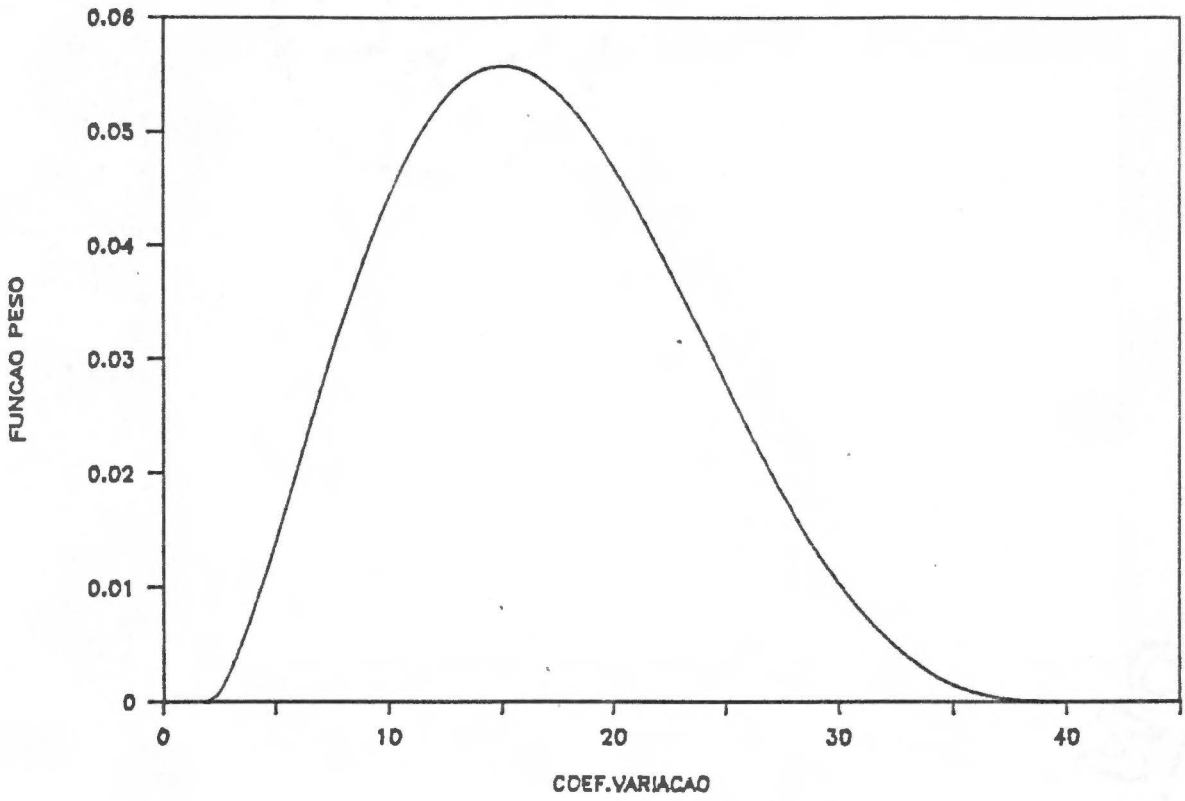


FIGURA 14 - FUNCAO PESO W(10,40,20,25)

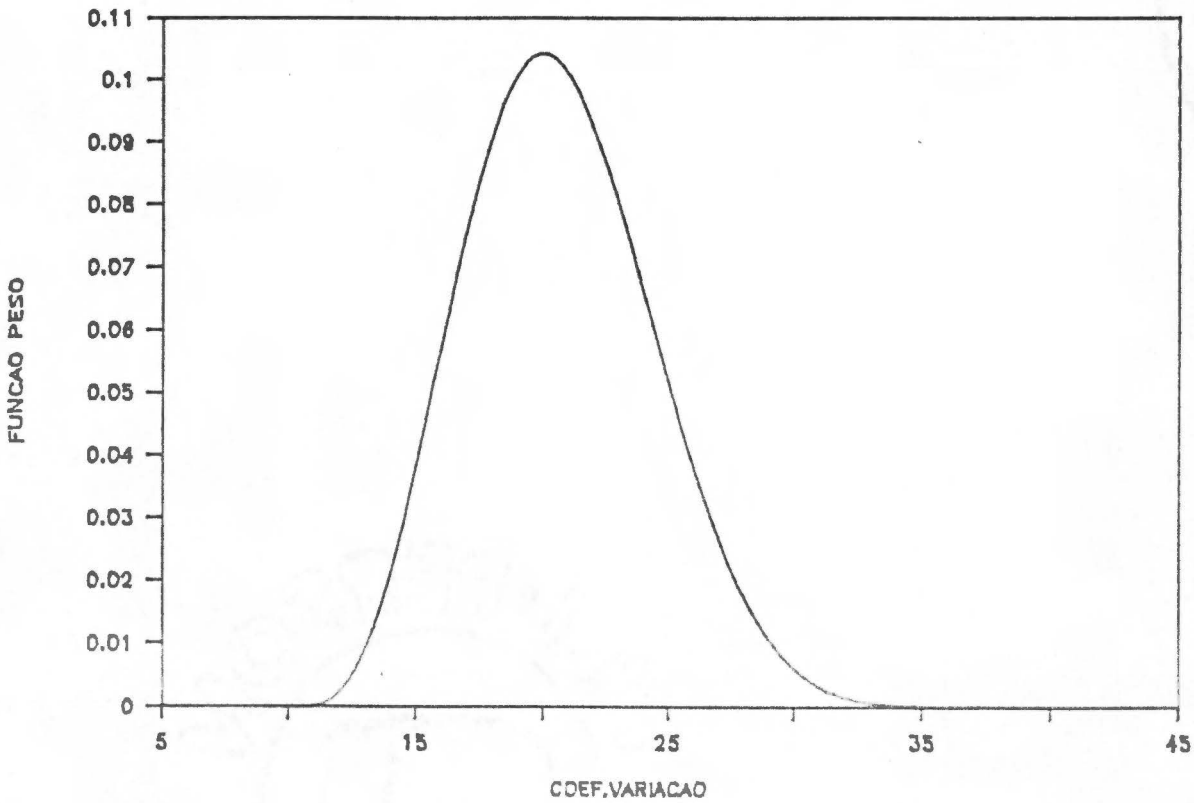


FIGURA 15 - FUNCAO PESO W(15,50,30,38)

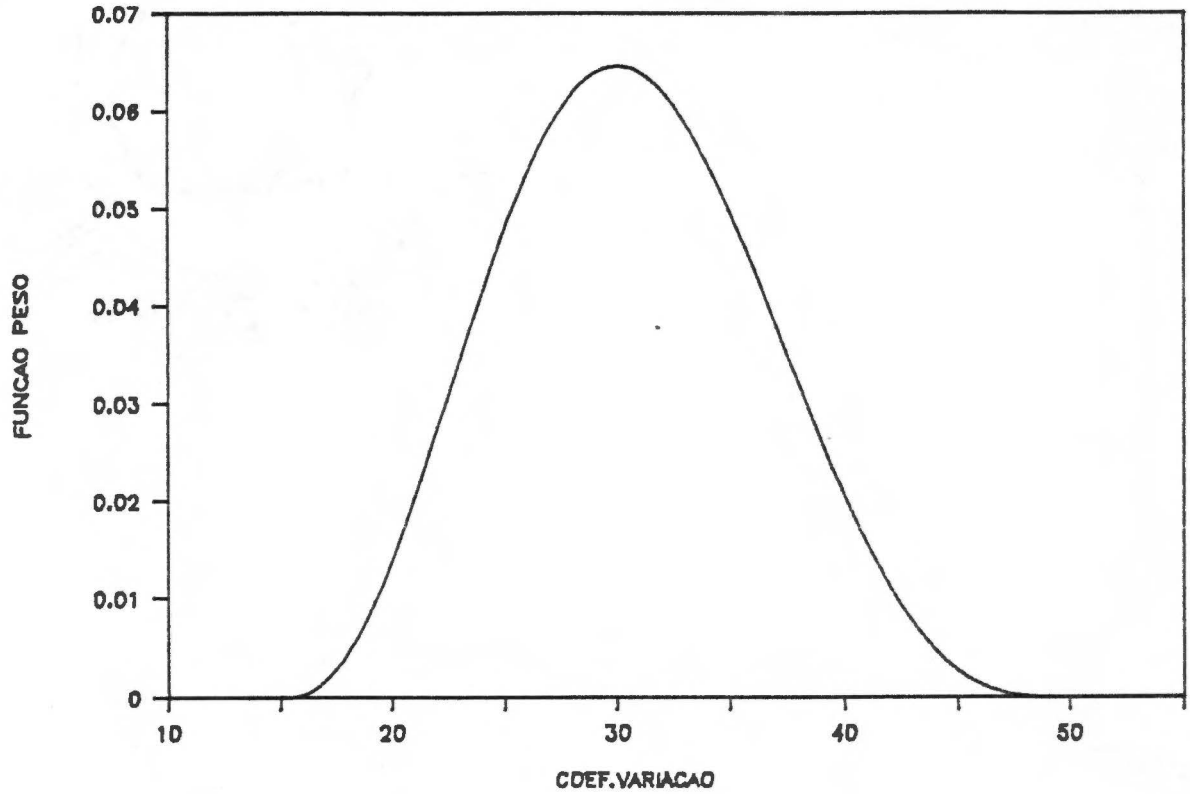


FIGURA 16 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

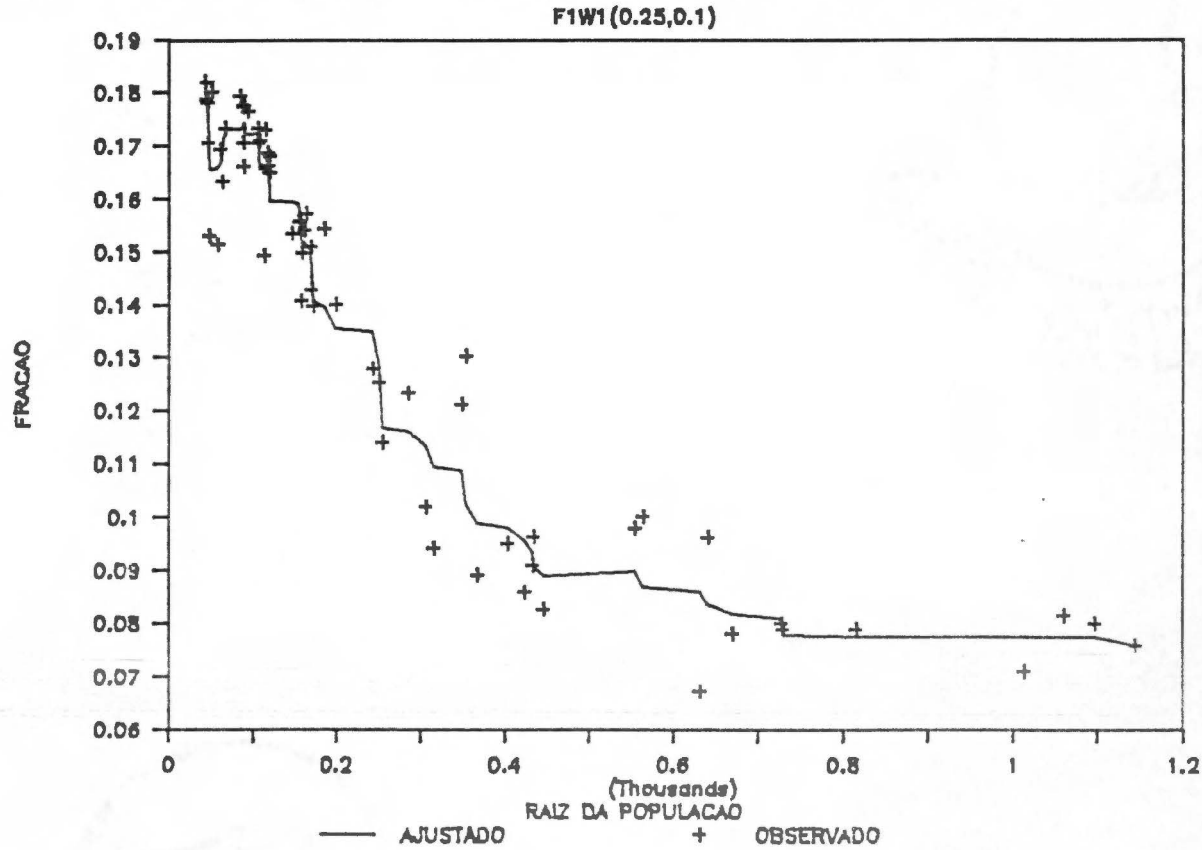


FIGURA 17 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

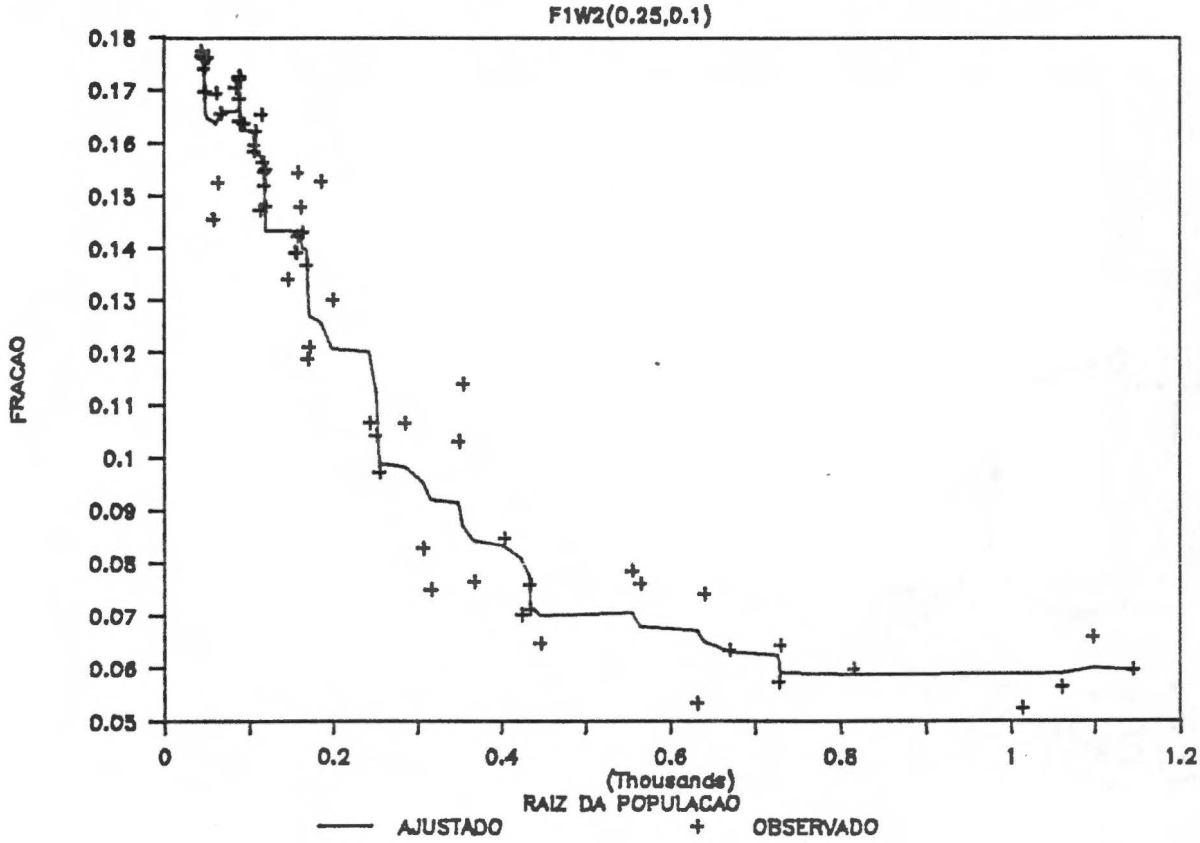


FIGURA 18 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

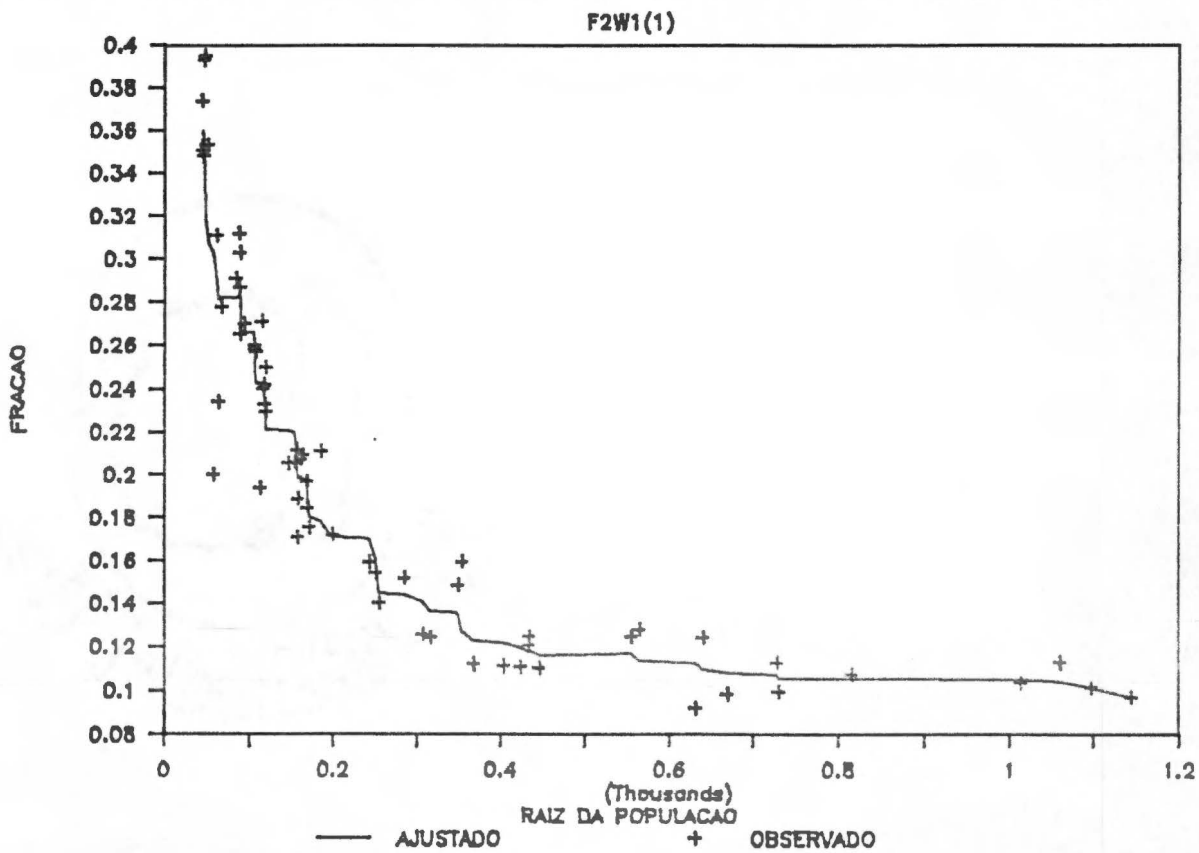


FIGURA 19 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

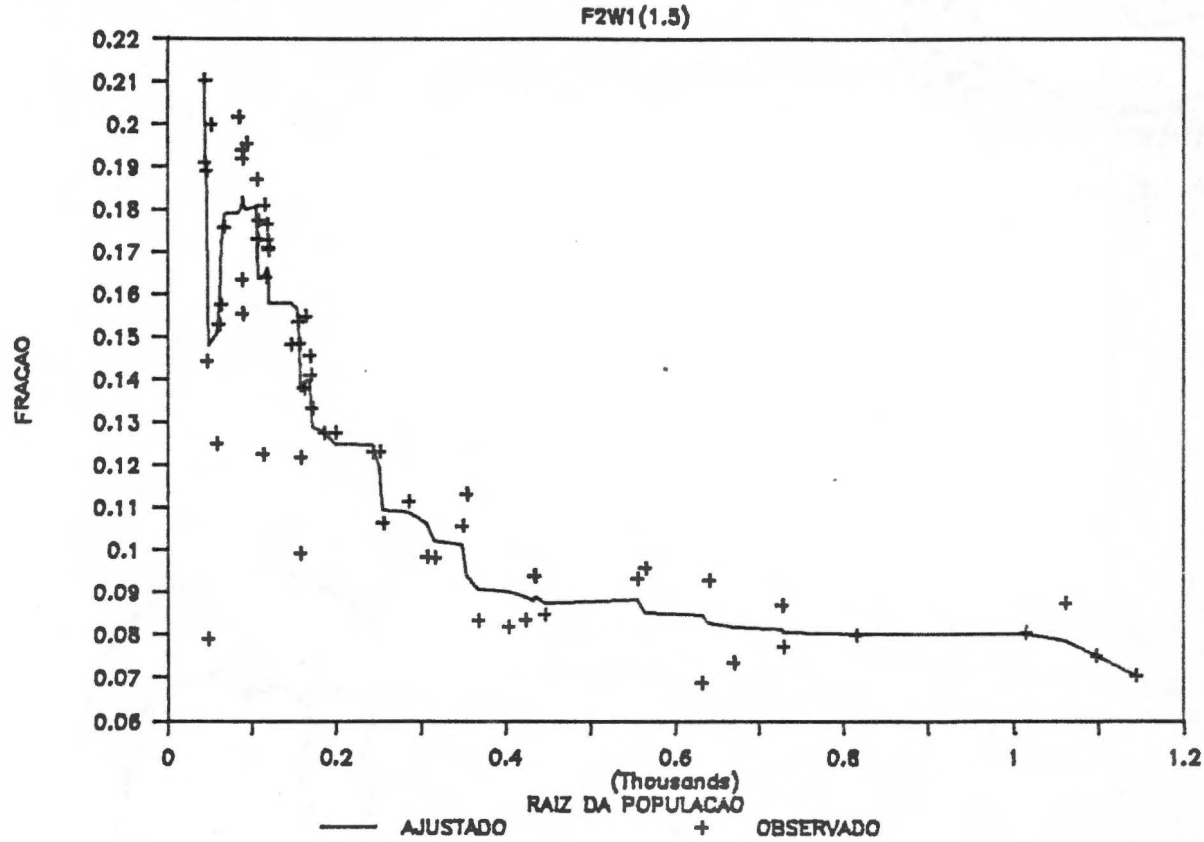


FIGURA 20 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

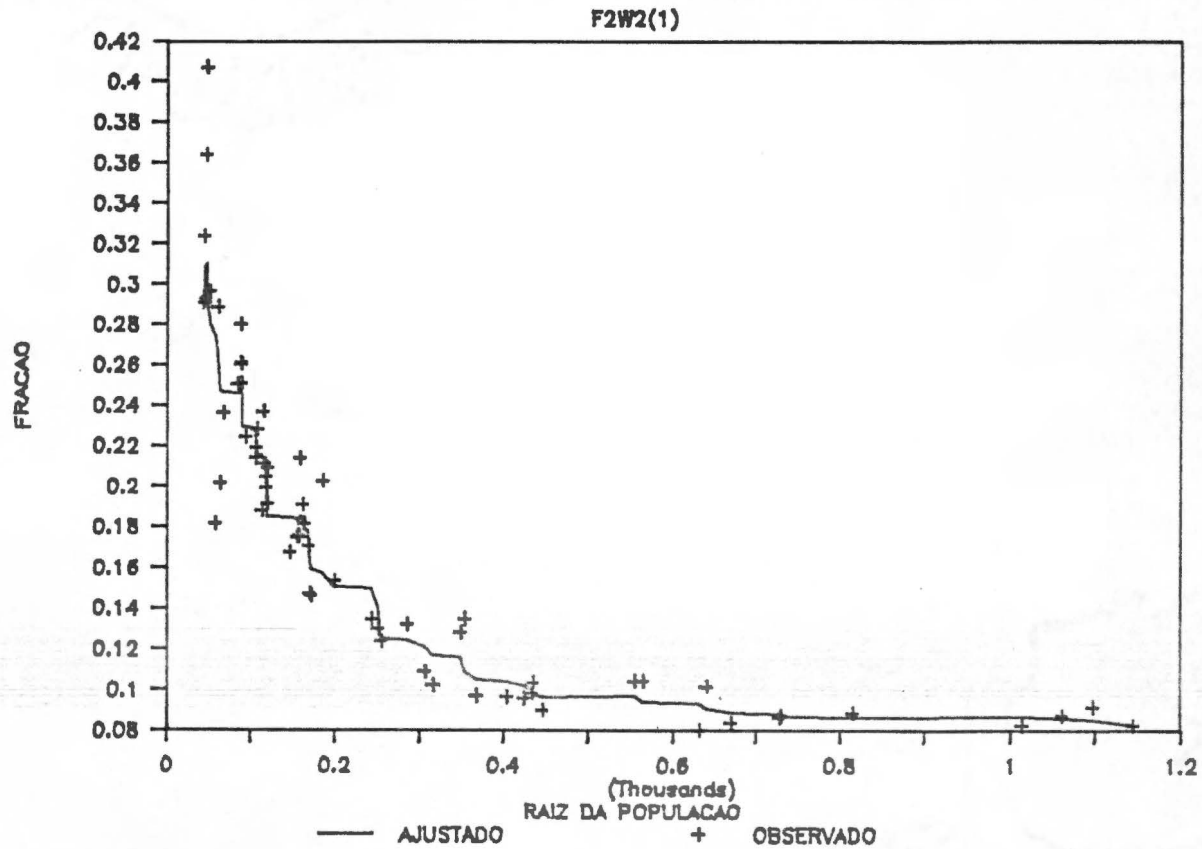


FIGURA 21 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

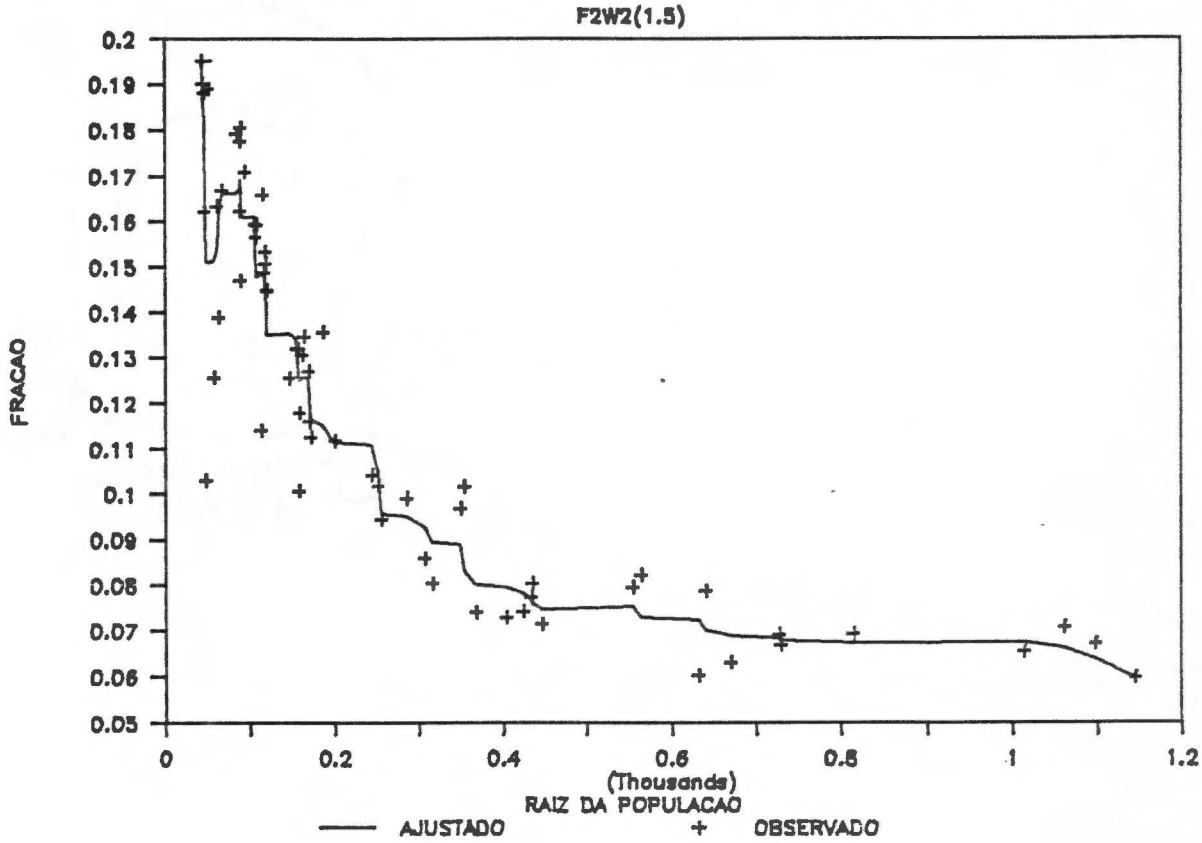


FIGURA 22 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

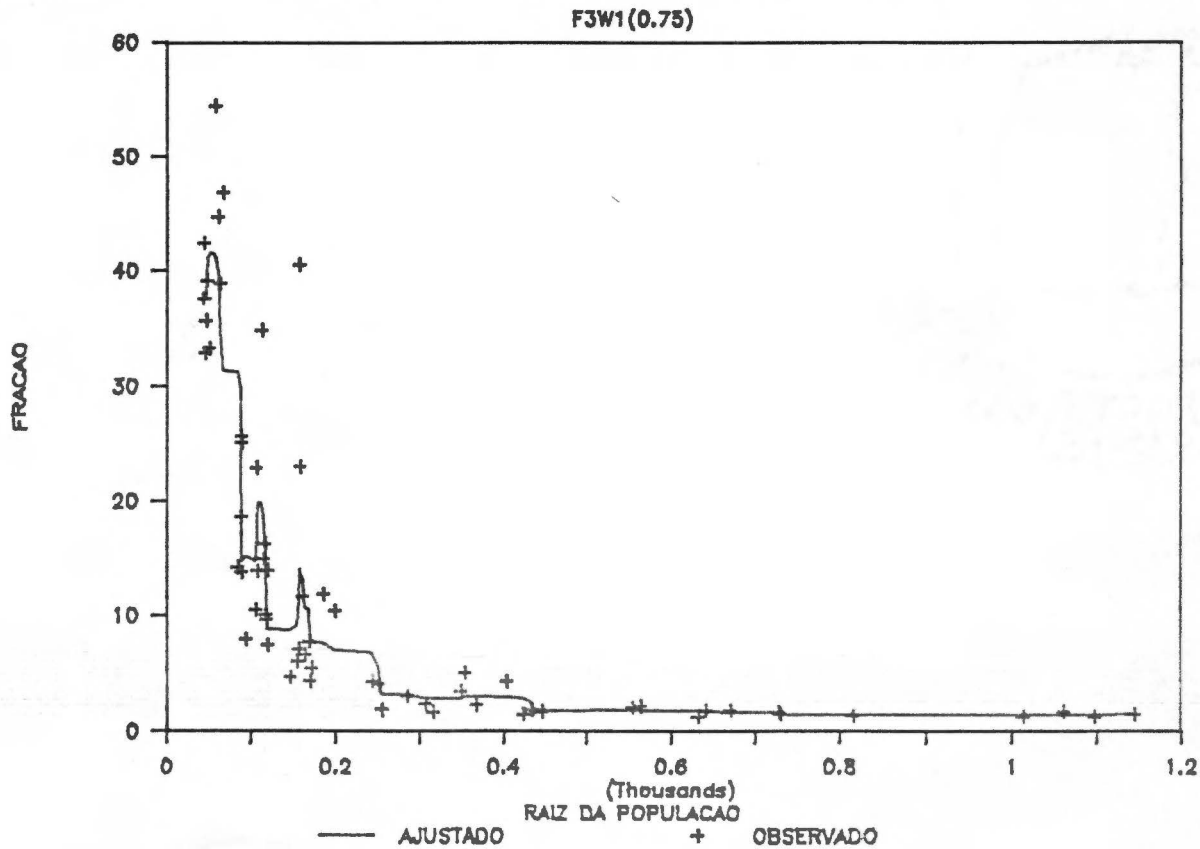


FIGURA 23 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

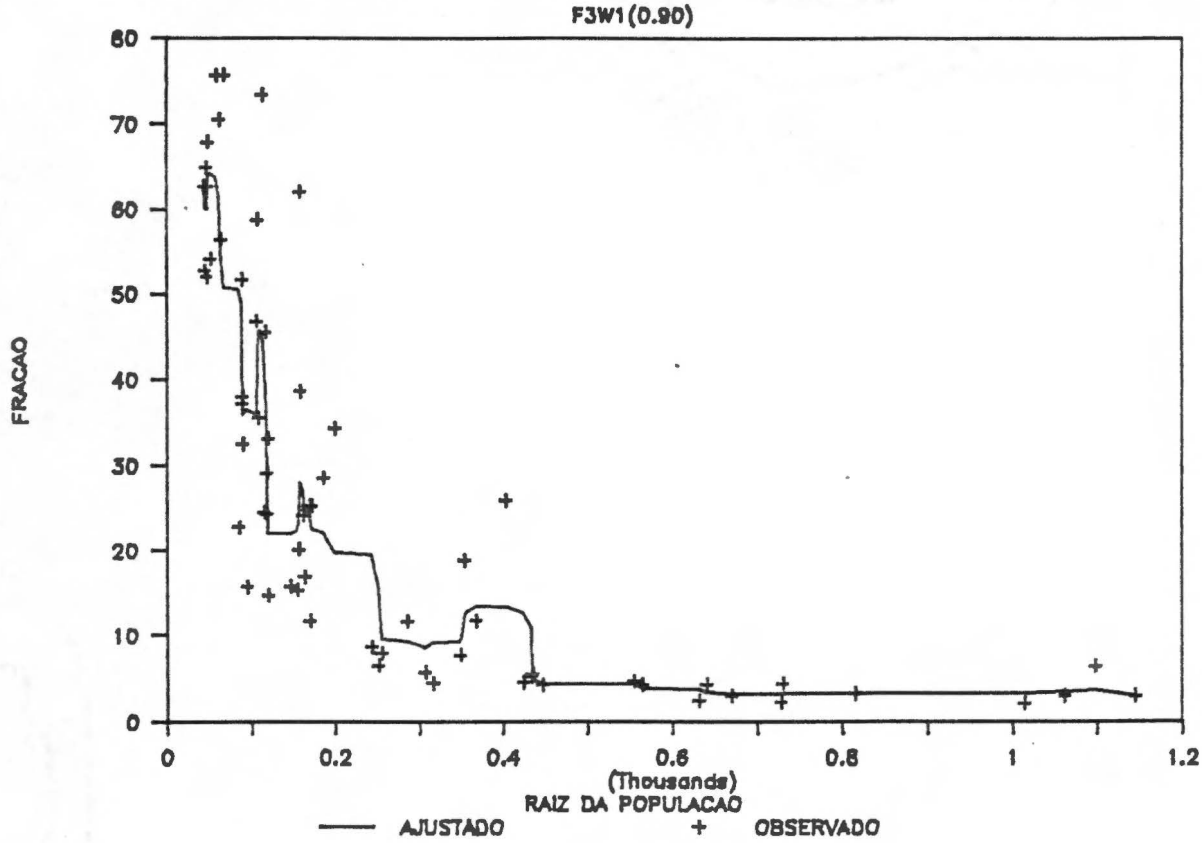


FIGURA 24 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

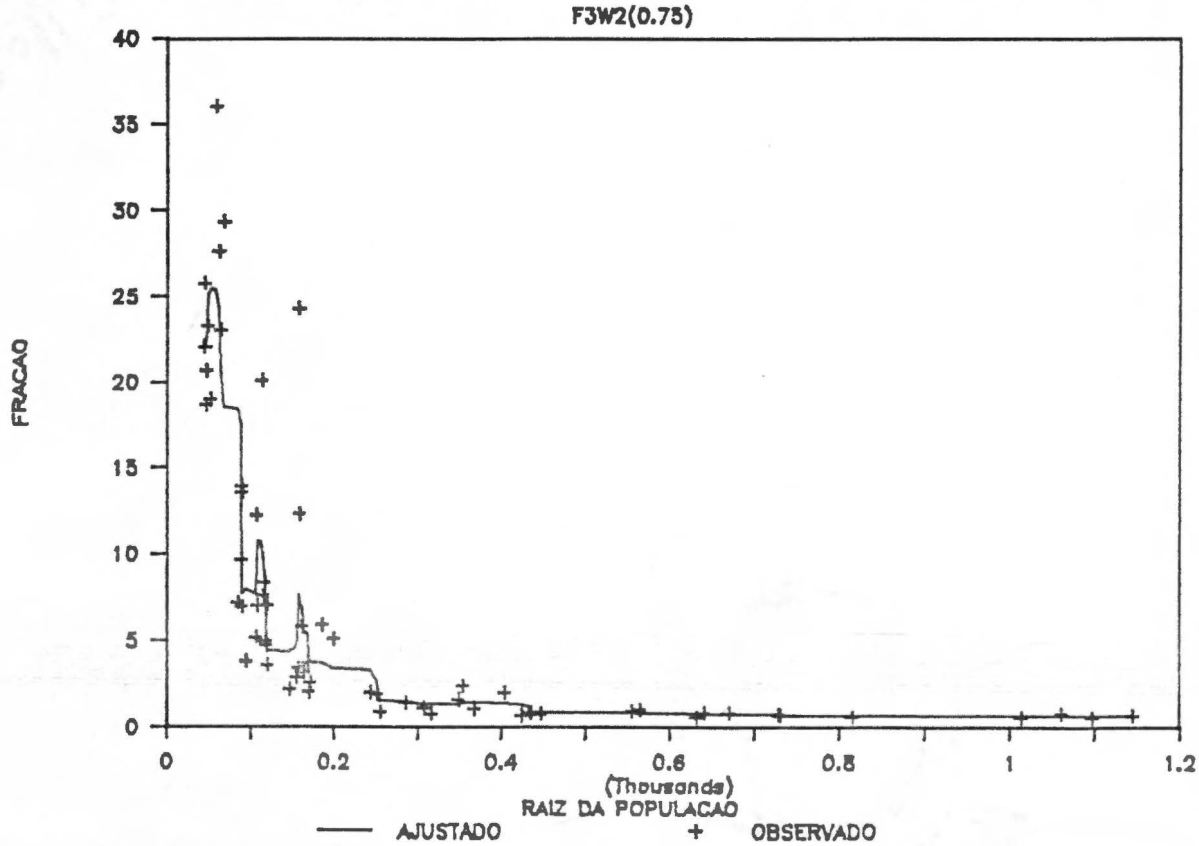


FIGURA 25 - FRACAO AMOSTRAL EM FUNCAO DO TAMANHO

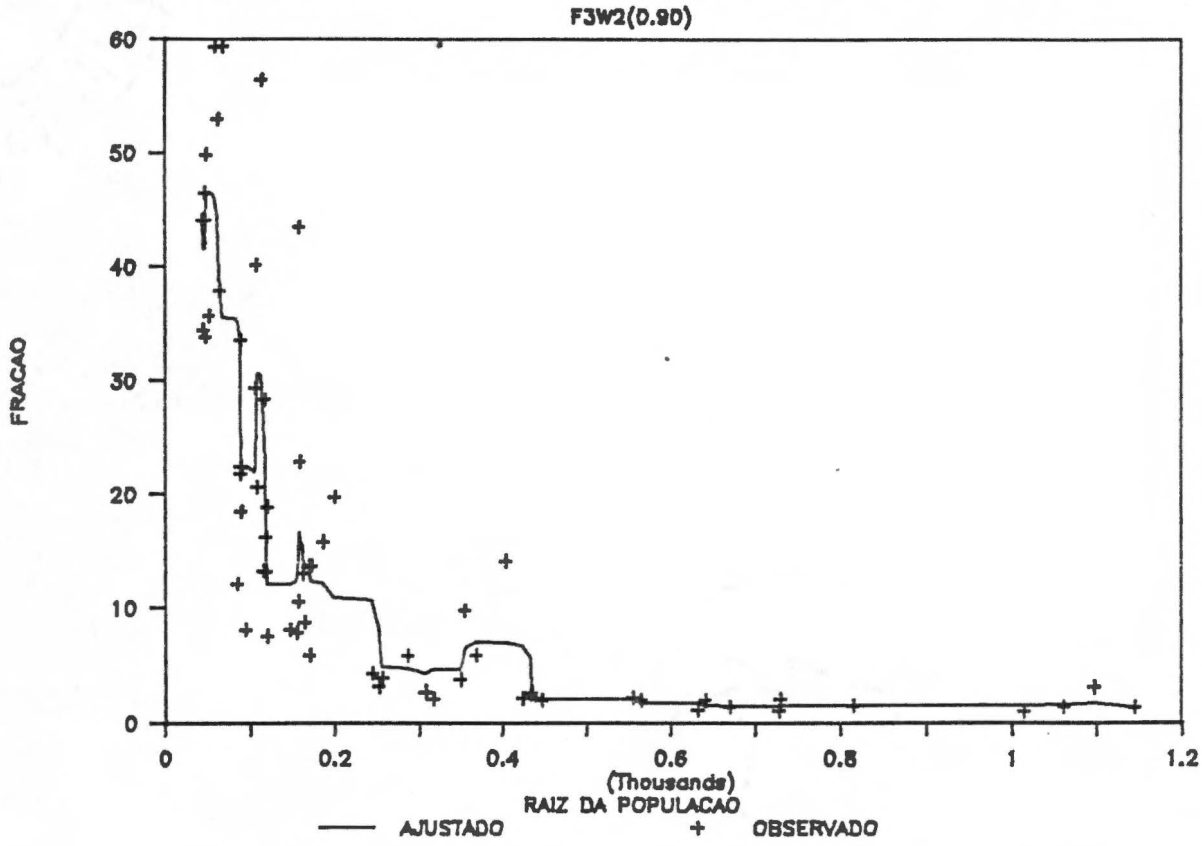


FIGURA 26 - FUNCAO DE DISTRIB. DA RAIZ DA POPULACAO

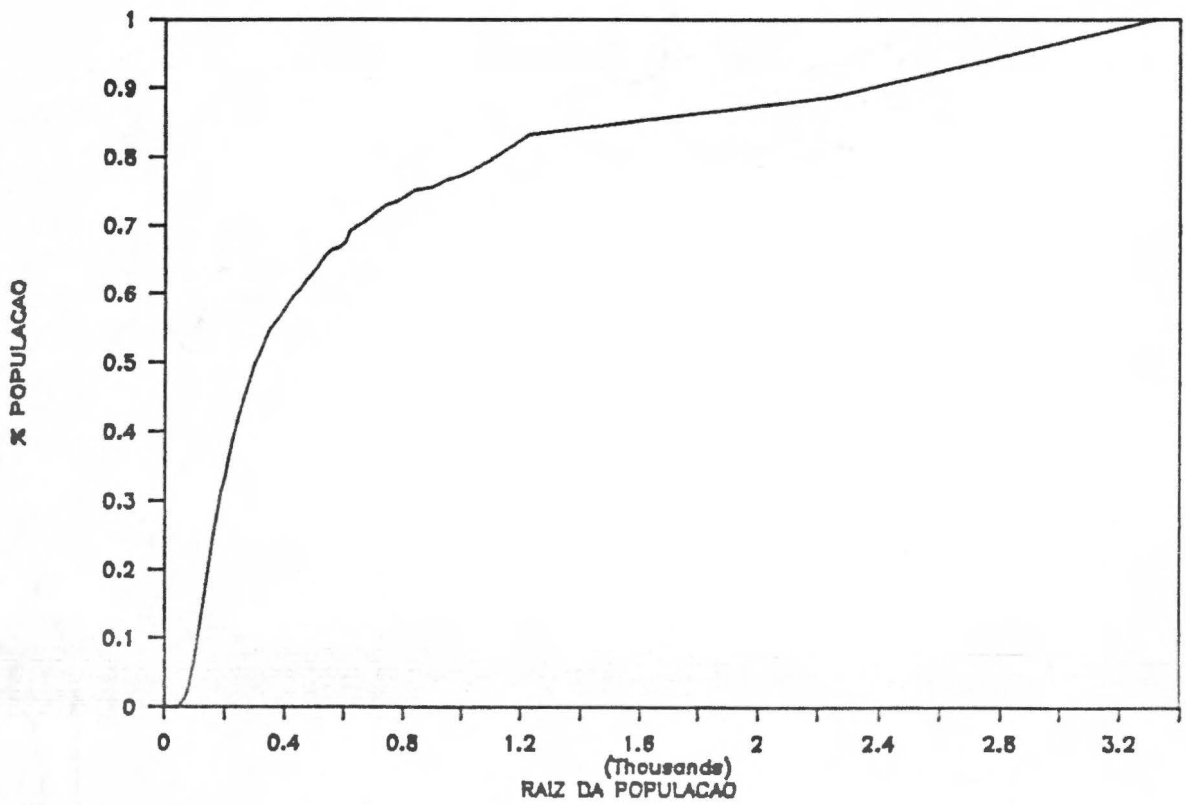
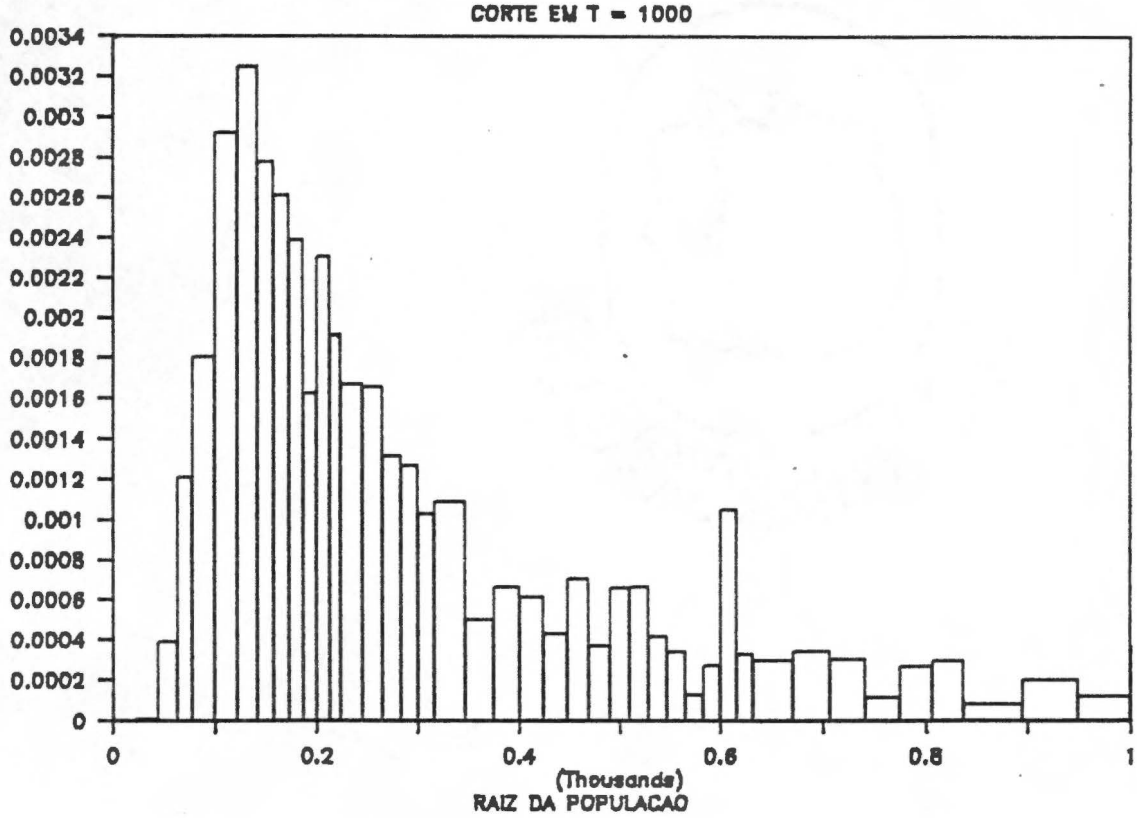


FIGURA 27 - DISTRIBUICAO DA RAIZ DA POPULACAO



NÚMEROS JÁ PUBLICADOS:

01/88 - CRÍTICA DE RAZÕES NO CENSO ECONÔMICO
Renato Martins Assunção (ENCE/IBGE)
Rosana de Freitas Castro (DEIND/IBGE)
José Carlos R.C.Pinheiro (ENCE/IBGE)

02/88 - USO DE AMOSTRAGEM EM SIMULAÇÃO DE LEGISLAÇÃO TRIBUTÁRIA
José Carlos da Rocha C. Pinheiro (ENCE/IBGE)
Manuel Martins Filho (DISUL/SERPRO)

03/88 - FORECASTING THE NUMBER OF AIDS CASES IN BRAZIL
Dani Ganeman (IME/UFRJ)
Helio S. Migon (ENCE/IBGE & IME/UFRJ)

04/88 - AVALIAÇÃO DOS EFEITOS DE REDUÇÃO DA FRAÇÃO DE AMOSTRAGEM NO CENSO DEMOGRÁFICO
José Carlos da Rocha C.Pinheiro (ENCE)
José Matias de Lima (DPE/NME)

01/89 - MIGRAÇÕES ANUAIS RURAL - URBANO - RURAL
Período 70/80
Kaizô Iwakami Beltrão (ENCE/IBGE)
Helio dos Santos Migon (ENCE/IBGE)

02/89 - CURVA DE LORENZ & ÍNDICE DE GINI PARA DISTRIBUIÇÕES DE RENDA
José Paulo Q.Carneiro (ENCE/IBGE)
Jorge Luiz Rangel Costa (CIDE)

03/89 - METODOLOGIA DE CRÍTICA DE EQUAÇÕES DE FECHAMENTO NOS CENSOS ECONÔMICOS DE 1985
José Carlos da Rocha Castelar Pinheiro (ENCE/IBGE)
Renato M. Assunção (ENCE/IBGE)

04/89 - UMA PROPOSTA DE DIMENSIONAMENTO DA AMOSTRA DO CENSO DEMOGRÁFICO DE 1990
José Carlos da Rocha Castelar Pinheiro (ENCE/IBGE)
José Matias de Lima (ENCE/IBGE)