

**HOMOGENEIDADE INTRACLASSE: estudo  
empírico sobre setores censitários como unidades  
primárias de amostragem**

**MARIA PAULA FERREIRA**

Dissertação de Mestrado  
Departamento de Epidemiologia  
da Faculdade de Saúde Pública  
da Universidade de São Paulo

Área de concentração:  
Epidemiologia

**ORIENTADORA: PROFa.  
ASSOCIADA DRA. NILZA  
NUNES DA SILVA**

São Paulo  
2003

*Aos meus pais*

## AGRADECIMENTOS

Os meus agradecimentos a Profa. Dra. Nilza Nunes da Silva pela orientação, apoio e, principalmente, paciência durante todo o processo de trabalho. Aos professores Benedito Benze e Denise Pimentel Bergamaschi e à Maria Cecília Goi Porto Alves, pelas valiosas sugestões dadas ao trabalho. À Fundação SEADE, na pessoa da Sra. Felícia Madeira, pelo apoio e cessão dos dados da Pesquisa de Condições de Vida. Aos meus colegas de Fundação, em especial a Nádya Pinheiro Dini por sua amizade e apoio durante o desenvolvimento deste trabalho. Ao meu eterno mestre Prof. Clóvis de Araújo Peres pelos preciosos conselhos e estímulos constantes. À Mitti, pela disponibilidade e interesse nas discussões sobre o trabalho. Por fim, a minha família, em especial minha irmã Maria do Carmo, por sempre acreditar em mim.

## RESUMO

Ferreira MP. **Homogeneidade intraclasse: estudo empírico sobre setores censitários como unidades primárias de amostragem.** São Paulo: 2003. [Dissertação de mestrado – Faculdade de Saúde Pública da Universidade de São Paulo]

**Objetivos.** Comparar dois métodos de estimação do coeficiente de correlação intraclasse, no contexto de inquéritos domiciliares em que o setor censitário é a unidade primária de amostragem. Avaliar o efeito da homogeneidade intra setor censitário, na amostra da Pesquisa de Condições de Vida no Estado de São Paulo, no ano de 1998. **Métodos.** Utilizando-se dois estimadores do coeficiente de correlação intraclasse foram calculadas as respectivas estimativas para 24 variáveis, em sete domínios. Também foram calculados os efeitos do delineamento e tamanhos de amostras. A fonte de dados utilizada foi a Pesquisa sobre Condições de Vida, realizada pela Fundação SEADE no Estado de São Paulo em 1998. **Resultados.** *icc* e *roh* apresentaram resultados semelhantes. Para as variáveis analisadas, a maioria das estimativas obtidas por meio de *roh* estava incluída no intervalo de confiança de 95% de *icc*. Nos sete domínios da Pesquisa de Condições de Vida, as variáveis de rendimentos são as responsáveis pelas maiores estimativas do coeficiente de correlação intraclasse. A perda de precisão comporta-se de forma semelhante em todos os domínios. Variáveis socioeconômicas, especificamente rendimentos, apresentam os maiores *deff*s, nunca inferiores a dois. **Conclusões.** A estimação do

coeficiente de correlação intraclasse por meio de *roh* apresenta resultados similares quando se utiliza *icc*. Este resultado indica que para planejamentos de amostras pode-se utilizar tanto *icc* quanto *roh*, sendo que este último apresenta uma maior facilidade de cálculo, uma vez que o *deff* é usualmente apresentado pela maioria dos pacotes computacionais de cálculo de erros amostrais. Reafirmou-se o setor censitário como unidade homogênea com relação as variáveis de rendimentos consideradas no estudo. Para a Região Metropolitana de São Paulo, os setores censitários se mostraram mais homogêneos, do que os pertencentes aos domínios do Interior, em relação às variáveis socioeconômicas, particularmente rendimentos.

**Descritores:** Coeficiente de correlação intraclasse. Planos complexos de amostragem. Efeito do delineamento. Tamanhos de amostras

## SUMMARY

Ferreira MP. **Intraclass Homogeneity: census tract as primary sampling unit a empirical study.** São Paulo: 2003. [Dissertação de mestrado – Faculdade de Saúde Pública da Universidade de São Paulo]

**Objective.** This work aims to compare two estimation methods for intraclass correlation coefficient, viewed in the context of household's surveys. The study also evaluates the homogeneity effect within census tract in the *Pesquisa de Condições de Vida* in São Paulo State. **Methodology.** Two intraclass correlation coefficient estimators were computed for 24 variables in the seven strata. The design effects and sample sizes were also computed. The data came from the *Pesquisa de Condições de Vida*, carried out by Fundação SEADE, in 1998. **Results.** The analysis of *icc* and *roh* showed similar results, with more than half of estimates of *roh* included in the *icc* 95% confidence interval. The income variables showed the greatest intraclass correlation coefficients estimates. The design effect behavior was similar in all the strata. Socioeconomic variables, particularly income variables, showed the biggest *deff*. **Conclusions.** The two estimation methods showed similar results. These suggest that in design surveys the two estimators can be used. The advantages for *roh* are two: simplicity and availability in most of statistical packages.

**Descriptors:** Intraclass correlation coefficient. Survey sampling. Design effect. Sample size.

# ÍNDICE

1 INTRODUÇÃO	1
2 REVISÃO BIBLIOGRÁFICA	10
2.1 O Coeficiente de Correlação Intraclasse	10
2.1.1 Método da ANOVA para Dados Balanceados	15
2.1.2 Método da ANOVA para Dados Não Balanceados	16
2.1.3 Considerações Gerais sobre o Método da ANOVA	18
2.2 O Coeficiente de Correlação Intraclasse em Amostragem	20
3 OBJETIVOS	25
4 METODOLOGIA	26
4.1 Fonte dos Dados	26
4.2 Estimadores	28
4.2.1 Método da ANOVA	29
4.2.2 Método KISH	30
4.2.3 Cálculo de $b_{\text{ótimo}}^*$	31
4.3 Processamento dos Dados	32
4.4 Variáveis Utilizadas no Estudo	32
5 RESULTADOS	34
5.1 Estimação do Coeficiente de Correlação Intraclasse	34
5.2 Estimação do Efeito do Delineamento	43
5.3 Estimação dos Tamanhos das Amostras	51

6 DISCUSSÃO E CONCLUSÃO	59
7 REFERÊNCIAS BIBLIOGRÁFICAS	62
ANEXO 1	A1
ANEXO 2	A3
ANEXO 3	A12



## 1 INTRODUÇÃO

As pesquisas domiciliares por amostragem constituem-se em importantes ferramentas para a obtenção de informações demográficas, socioeconômicas e epidemiológicas sobre a população em geral. Como exemplos deste tipo de levantamento tem-se, entre outros, a Pesquisa Nacional de Domicílios (PNAD), realizada anualmente pela Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), a Pesquisa de Emprego e Desemprego na Região Metropolitana de São Paulo (PED) e a Pesquisa de Condições de Vida no Estado de São Paulo (PCV), ambas realizadas pela Fundação Sistema Estadual de Análise de Dados (SEADE).

No campo da saúde pública estes levantamentos são importantes para o planejamento e avaliação de serviços de saúde, além de permitirem a realização de amplos diagnósticos de saúde. ALVES (2002) relaciona algumas vantagens dos inquéritos domiciliares para a saúde pública, citadas por diversos autores. São elas:

- *“Inclusão da parcela da população que não tem acesso ao sistema de saúde (LEBRÃO 1991);*
- *facilidade de atingir usuários de diferentes instituições prestadoras de serviços (CARVALHEIRO 1981);*
- *Superação das limitações que dados de fontes secundárias apresentam, em especial a que se refere à seleção de casos de maior gravidade, em estudos de morbidade (CESAR 1997);*
- *Possibilidade de coletar informações referentes a variáveis que se relacionam com os fenômenos sob investigação, aumentando a compreensão sobre sua*

*ocorrência e a capacidade de determinar intervenções apropriadas (WHITE 1985).”*

Nos inquéritos domiciliares utiliza-se usualmente planos complexos de amostragem, sendo a amostragem por conglomerados em dois ou mais estágios o delineamento mais freqüente. Isto é consequência do fato que, a unidade de informação – o domicílio – encontra-se normalmente agrupada em conglomerados. Domicílios localizam-se em quadras que, por sua vez, pertencem a bairros, que formam as cidades, etc. A união destes conglomerados, em algum nível, constitui a população. Assim, o procedimento da amostragem por conglomerados pode ser resumido como: *“quando os elementos da população são reunidos em grupo e, por sua vez, alguns destes são sorteados para compor a amostra, o procedimento denomina-se: amostragem por conglomerados”* (SILVA, 1998).

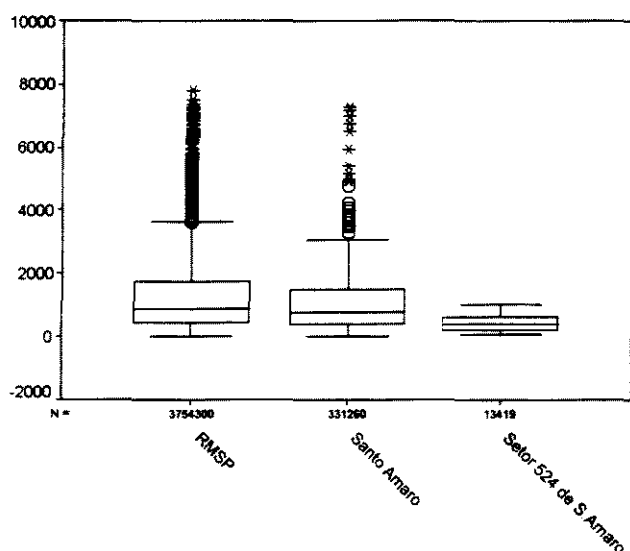
A diferença entre esse desenho amostral e a amostragem aleatória simples (AAS) é que, nesta última, todos os elementos da população possuem igual probabilidade de vir a pertencer à amostra e a seleção de cada elemento é realizada de forma independente dos demais. Já na amostragem por conglomerados, em um ou mais estágios, a probabilidade de um elemento pertencer a uma amostra é condicionada ao fato do conglomerado a que pertence ser selecionado para compor a amostra (KALTON 1983).

A grande vantagem da amostragem por conglomerados é a diminuição dos custos de obtenção da informação. A desvantagem, é a perda de precisão das estimativas em relação àquelas obtidas por meio de AAS. Pois, para uma determinada característica, os elementos pertencentes a um mesmo conglomerado são mais “parecidos” entre si, do que os pertencentes a outro conglomerado. De modo geral, a variabilidade

existente dentro do conglomerado é menor do que a variabilidade observada para o total da população de interesse, contrariando a definição teórica de conglomeração.

Assim, *“a seleção de mais um elemento de um mesmo conglomerado adiciona menos informação nova do que se este elemento fosse selecionado através de um sorteio totalmente independente”*. Ou seja, uma amostra por conglomerados em um único estágio não é tão variada quanto uma amostra aleatória simples de mesmo tamanho (LEVY e LEMESHOW 1999).

Um dos exemplos mais conhecidos sobre a similaridade existente entre elementos pertencentes a um mesmo conglomerado, refere-se aos níveis de rendimentos observados nas famílias moradoras em um mesmo setor censitário. A Figura 1 mostra um decréscimo, tanto nos níveis, quanto na variabilidade da variável renda familiar total, quando se compara a Região Metropolitana de São Paulo (RMSP), o distrito de Santo Amaro e o setor censitário de número 524 pertencente a este distrito.



**Figura 1** – Box-Plot da Renda Familiar Total, segundo regiões geográficas. 1998.  
Em reais de setembro de 1998.

Resumidamente as vantagens e desvantagens do modelo de amostragem por conglomerados estão apresentadas no Quadro 1.

**Quadro 1 – Vantagens e Desvantagens da Amostragem por Conglomerados em Relação a Amostragem por Elementos.**

VANTAGENS	DESvantagem
1. Diminuição dos custos 2. Agilidade nas operações de campo; 3. Utilização de cadastros parciais existentes para as seleções nos diferentes estágios, como por exemplo, a relação dos setores censitários.	1. Perda de precisão em relação a uma amostra de elementos, devido a possível existência de similaridade entre os elementos de um mesmo conglomerado 2. Estimadores distintos dos utilizados na AAS. 3. Necessidade de programas computacionais específicos para o processamento dos dados

Com a finalidade de mensurar a perda de precisão de uma amostragem por conglomerados, em relação a uma AAS, KISH (1965) propôs o efeito do delineamento, ou *Design Effect (Deff)*, que se constitui na razão entre a variância de um determinado estimador ( $\hat{\theta}$ ), calculado sob amostragem de conglomerados, e a variância do mesmo estimador sob amostragem aleatória simples.

$$Deff = \frac{\text{Var}(\hat{\theta})_{AC}}{\text{Var}(\hat{\theta})_{AAS}} \quad (1)$$

onde,

$\text{Var}(\hat{\theta})_{AC}$  : variância de  $\theta$  na amostra de conglomerados;

$\text{Var}(\hat{\theta})_{AAS}$  : variância de  $\theta$  na amostra aleatória simples.

Suponha uma amostra por conglomerados de tamanhos iguais ( $B$ ) em uma população de tamanho  $N$ . Dos  $A$  conglomerados da população,  $a$  são selecionados através de amostragem aleatória simples, e todos os elementos do conglomerado são selecionados para compor a amostra. O tamanho da amostra é dado por  $n = aB$  e a

fração amostral por  $f = n/N = aB/AB = a/A$ . A média amostral é  $\bar{y}_c = \frac{\sum_{\alpha} \sum_{\beta} y_{\alpha\beta}}{n} = \frac{\sum_{\alpha} \bar{y}_{\alpha}}{a}$ ,

onde  $\bar{y}_{\alpha}$  é a média do  $\alpha$ -ésimo conglomerado. A variância de  $\bar{y}_c$  é dada por:

$$V(\bar{y}_c) = \left(1 - \frac{a}{A}\right) \frac{S_a^2}{a} \quad (2)$$

onde,

$S_a^2 = \frac{\sum_{\alpha} (\bar{Y}_{\alpha} - \bar{Y})^2}{A-1}$  é a variância das médias dos conglomerados,  $\bar{Y}_{\alpha}$  é a média

populacional do  $\alpha$ -ésimo conglomerado e  $\bar{Y}$  é a média populacional.

Sob uma AAS com  $n=aB$  elementos, a média amostral é dada por  $\bar{y} = \sum_{i=1}^n y_i / n$

com variância igual a  $(1-f)S^2/n = \frac{(1-f)}{n} \cdot \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N-1}$ .

Então o efeito do desenho pode ser escrito como,

$$Deff = \frac{S_a^2 / a}{S^2 / aB} = \frac{BS_a^2}{S^2} \quad (3)$$

Assim, por exemplo, um *Deff* igual a 3,0 significa que a variância da média amostral obtida através de uma amostra por conglomerados é três vezes a variância desse estimador obtido através de uma AAS. Uma interpretação alternativa é: para se obter a mesma variância observada na amostra aleatória simples é necessário triplicar o tamanho da amostra por conglomerados.

Estimativas do *Deff* são obtidas através das estimativas do denominador e do numerador da expressão (3). Para o denominador é estimada a variância supondo-se AAS, já a estimativa para o numerador pode ser obtida por vários métodos, tais como: linearização de Taylor e replicação. (ALVES 2002)

A magnitude do *Deff* depende de como os conglomerados são constituídos. Suponha, por exemplo, que o número de conglomerados (*A*) na população seja grande e os elementos da população são distribuídos aleatoriamente nos mesmos. Então,  $S_a^2$  poderia ser interpretado como a variância das médias de *a* amostras aleatórias com *B* elementos cada, sendo aproximadamente igual a  $S^2 / B$ . Sob essas condições *Deff* tenderia a um. Porém, se os conglomerados forem mais homogêneos internamente, do que ocorreria se a alocação dos elementos fosse aleatória, as médias dos conglomerados serão mais heterogêneas entre si, implicando em  $S_a^2$  maior que  $S^2 / B$ . Assim, o *Deff* será maior do que um (KALTON 1983).

KISH (1965) estabelece a relação entre o efeito de delineamento ( $Deff$ ) e o coeficiente de correlação intraclasse ( $\rho$ ), que mensura o grau de similaridade interna dos conglomerados. A expressão (4) apresenta essa relação no caso de amostragem por conglomerados de tamanhos iguais em um estágio.

$$Deff = 1 + \rho(B - 1) \quad (4)$$

onde,

B: tamanho do conglomerado;

$\rho$  : coeficiente de correlação intraclasse.

Assim, se em uma grande população os conglomerados forem formados de forma aleatória, tem-se  $\rho \approx 0$  e  $Deff$  igual a 1, uma vez que a similaridade interna dos elementos do conglomerado é praticamente nula. Já, um valor negativo de  $\rho$  indica que os conglomerados são internamente mais heterogêneos do que ocorreria numa alocação aleatória de elementos nos conglomerados, produzindo um  $Deff < 1$ . Este resultado indica que a amostragem por conglomerados é mais precisa que uma amostragem aleatória simples. O valor máximo para  $\rho$  é um, e ocorre quando em cada conglomerado todos os elementos possuem o mesmo valor para a variável de interesse, e esses valores são diferentes entre os conglomerados. Nesse caso todos os conglomerados deveriam compor a amostra (KALTON 1983).

Na prática, valores negativos de  $\rho$  são extremamente raros. Usualmente os valores de  $\rho$  são pequenos, a maioria abaixo de 0,15. (KALTON 1983)

No caso de amostragem por conglomerados com tamanhos iguais em dois estágios o valor de  $B$  na fórmula (4) é substituído por  $b$ , correspondendo ao tamanho da amostra de elementos no segundo estágio da amostragem, ou subamostra. Para conglomerados de tamanhos distintos, KISH (1965) afirma que o tamanho médio ( $\bar{b} = n/a$ ) pode ser utilizado, desde que não existam disparidades acentuadas entre os tamanhos dos conglomerados. O número de conglomerados na amostra é representado por  $a$  e o tamanho da amostra por  $n$ .

Para KISH (1987) estimativas da correlação intraclasse ( $\rho$ ) são importantes para:

- Planejamento da amostra, no sentido de garantir a precisão das estimativas, uma vez que, para se manter a precisão desejada é necessário definir quantos conglomerados serão selecionados. E, quanto mais homogêneos internamente eles forem, maior deverá ser o número de conglomerados a compor a amostra;
- Planejamento de futuras amostras, onde se utiliza os mesmos conglomerados, mas com tamanhos amostrais distintos. Estimativas anteriores de  $\rho$  podem ser usadas com auxiliares para estes tipos de desenhos amostrais;
- Análise de subclasses, quando o tamanho da subamostra na subclasse ( $\bar{b}_c = n_c/a$ ) é muito menor que  $\bar{b} = n/a$ . Pode-se então estimar o efeito de delineamento para a subclasse através da relação:  $deff_c = 1 + roh_t(\bar{b}_c - 1)$ , onde  $deff_c$  corresponde a estimativa do efeito de delineamento para a subclasse e  $roh_t$  é uma estimativa de  $\rho$  para o total da amostra. Esta relação pode se melhorada com  $deff_c \approx 1 + k_g roh_c(\bar{b}_c - 1)$ , onde  $roh_c$  é a estimativa do coeficiente de correlação intraclasse para a variável de interesse na



subclasse e  $k_g$  é ligeiramente maior que 1 para subclasses que tendem a ser desigualmente distribuída nos conglomerados, como variáveis socioeconômicas.

Particularmente no campo da saúde pública, a divulgação dos valores das estimativas dos coeficientes de correlação intraclasse, obtidos nas diversas pesquisas domiciliares existentes, pode ser uma importante ferramenta para o planejamento de futuros inquéritos domiciliares.

Uma vez que, geralmente, os levantamentos domiciliares são baseados em amostragem por conglomerados – onde as unidades de primeiro estágio são os setores censitários criados pela Fundação IBGE – é de extrema importância a publicação de medidas que expressem o grau de similaridade desses setores, segundo renda familiar, escolaridade da população residente, classe social, etc.

Assim, no Brasil, assume particular importância o estudo de métodos de estimação do coeficiente de correlação intraclasse em amostras por conglomerados, principalmente os de tamanhos desiguais, como é o caso dos setores censitários.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 O Coeficiente de Correlação Intraclasse

A correlação intraclasse foi inicialmente utilizada em estudos sobre similaridade fraternal, como no caso da correlação entre altura de irmãos. PEARSON, em 1896, abordando este problema, propôs o coeficiente de correlação intraclasse para pares repetidos, utilizando a correlação momento-produto,  $\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$ , onde X representa a altura do primeiro irmão e Y a do segundo. Esse estimador ficou conhecido como *pairwise intraclass correlation coefficient* (BERGAMASCHI 1999).

Esta proposta parte da segmentação dos N elementos da população em M classes, cada uma com  $N_i$  elementos ( $i = 1, 2, \dots, M$ ), onde é possível calcular o grau de homogeneidade de uma certa característica de interesse Y dentro das classes. No exemplo da altura entre irmãos, os N elementos da população correspondem aos irmãos de todas as M famílias (classes),  $N_i$  representa o número de irmãos na família i e Y a altura. (FISHER 1938)

Para o cálculo do coeficiente de correlação intraclasse proposto por Pearson é necessário a criação de todos os pares de unidades distintas ( $u_{ij}, u_{ik}$ ),  $j \neq k$ , dentro de cada classe i. Onde a classe i possui  $N_i(N_i-1)$  pares de valores ( $y_{ij}, y_{ik}$ ),  $j \neq k$ , com  $y_{ij}$  e  $y_{ik}$  correspondendo aos valores observados da característica de interesse Y nos pares ( $u_{ij}, u_{ik}$ ),  $j \neq k$ . O número de pares de valores ( $y_{ij}, y_{ik}$ ) observados na população é igual

$$a \sum_{i=1}^M N_i(N_i - 1).$$

O número de pares possíveis para cada classe pode ser visualizado por meio da tabela simétrica apresentada no Quadro 2.

**Quadro 2** - Tabela simétrica representando os  $N_i(N_i-1)$  pares possíveis de valores na classe  $i$ .

	1	2	. . . . .	$N_i$
1	-	$(X_{i1}, X_{i2})$	. . . . .	$(X_{i1}, X_{iN_i})$
2	$(X_{i2}, X_{i1})$	-	. . . . .	$(X_{i2}, X_{iN_i})$
.	.	.	. . . . .	.
.	.	.	. . . . .	.
.	.	.	. . . . .	.
.	.	.	. . . . .	.
$N_i$	$(X_{iN_i}, X_{i1})$	$(X_{iN_i}, X_{i2})$	. . . . .	-

O coeficiente de correlação intraclasse proposto por Pearson é dado por:

$$\rho = \frac{\frac{1}{\sum_{i=1}^M N_i(N_i-1)} \sum_{i=1}^M \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} (Y_{ij} - \mu)(Y_{ik} - \mu)}{\sigma^2} \quad (5)$$

onde,

$$\mu = \frac{\sum_{i=1}^M (N_i - 1) \sum_{j=1}^{N_i} Y_{ij}}{\sum_{i=1}^M N_i(N_i - 1)} : \text{ é a média dos valores do vetor formado por um dos}$$

componentes dos pares de valores definidos na população, podendo ser calculado a partir da primeira ou segunda componente;

$$\sigma'^2 = \frac{\sum_{i=1}^M (N_i - 1) \sum_{j=1}^{N_i} (Y_{ij} - \mu)^2}{\sum_{i=1}^M N_i (N_i - 1)} : \text{ é a variância dos valores do vetor formado}$$

por um dos componentes dos pares de valores definidos na população, podendo ser calculado a partir da primeira ou segunda componente.

Um estimador para o coeficiente de correlação intraclasse apresentado em (5) é dado por:

$$r_i = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} (y_{ij} - \bar{y})(y_{ik} - \bar{y})}{\sum_{i=1}^m (n_i - 1) \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2} \quad (6)$$

onde,

$$\bar{y} = \frac{\sum_{i=1}^m (n_i - 1) \sum_{j=1}^{n_i} y_{ij}}{\sum_{i=1}^m n_i (n_i - 1)}$$

m: classes amostradas;

$n_i$ : número de observações em cada classe

Um resultado importante é que para classes de tamanhos iguais  $\rho$  coincide com o coeficiente de correlação apresentado nos livros de amostragem, para o caso de conglomerados de tamanhos iguais, com as classes correspondendo aos conglomerados.

Pesquisadores do IBGE (1990), inicialmente consideraram  $r_i$  nos estudos da correlação intraclasse, nos setores censitários referentes ao Censo Demográfico de 1980. Verificando, posteriormente, que esse estimador não era o mais indicado para avaliar a homogeneidade intra setor censitário.

Para DONNER (1986) o uso do estimador  $r_i$  é pouco indicado, a não ser no caso em que os conglomerados possuam tamanhos iguais.

Como uma alternativa ao cálculo da correlação intraclasse baseada na idéia de pares, ou seja, em tabelas simétricas de correlação, FISHER (1938) apresentou a correlação intraclasse como razão de componentes de variâncias, interpretando-a como uma medida da importância relativa de duas fontes de variação. Expressando assim, a correlação intraclasse como a razão entre a variância das médias de cada família e a variância de todas as observações em relação à média total. (FISHER 1938, p. 230; BERGAMASCHI 1999; SEARLE 1992).

As idéias propostas por Fisher podem ser formuladas a partir da teoria de modelos lineares. Supondo-se uma população agrupada em classes, da qual foi extraída uma amostra aleatória, onde apenas algumas classes estão representadas para uma determinada variável de interesse  $Y$ , tem-se o modelo:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (7)$$

onde,

$i=1,2,\dots,a$  e  $j=1,2,\dots,b$   $\forall i$  para dados balanceados

$j=1,2,\dots,b_i$  para dados não balanceados

$y_{ij}$  é a  $j$ -ésima observação da  $i$ -ésima classe;

$\mu$  é a média geral;

$\alpha_i$  é o efeito da  $i$ -ésima classe no valor observado da variável Y:

$\varepsilon_{ij}$  é o erro aleatório.

As terminologias “dados balanceados” e “dados não balanceados” são usuais em modelos lineares, e serão adotadas nesta seção. A primeira refere-se ao caso em que em cada classe existe o mesmo número de observações, já para a segunda tem-se um número distinto de observações em cada classe. Em amostragem o termo “dados balanceados” pode ser associado a conglomerados de tamanhos iguais e “dados não balanceados” a conglomerados de tamanhos desiguais..

No modelo (7)  $\alpha_i$  e  $\varepsilon_{ij}$  são variáveis aleatórias independentes e identicamente distribuídas, tal que:

$$E(\alpha_i) = 0 \quad \text{e} \quad \text{var}(\alpha_i) = E(\alpha_i^2) = \sigma_a^2 \quad \forall i,$$

$$\text{cov}(\alpha_i, \alpha_j) = E(\alpha_i \alpha_j) = 0 \quad \text{para } i \neq j$$

$$E(\varepsilon_{ij}) = 0 \quad \text{e} \quad \text{Var}(\varepsilon_{ij}) = E(\varepsilon_{ij}^2) = \sigma_e^2$$

$$\text{cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0 \quad \text{exceto para } i=i' \text{ e } j=j'$$

$$E \varepsilon_{ij} \text{ e } \alpha_k \text{ são independentes com } \text{cov}(\varepsilon_{ij}, \alpha_k) = 0 \quad \forall i, j \text{ e } k$$

As variâncias e covariâncias de  $y_{ij}$  podem ser expressas como:

$$\text{cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma_a^2 + \sigma_e^2 & \text{para } i=i', j=j', \\ \sigma_a^2 & \text{para } i=i', j \neq j', \\ 0 & \text{para } i \neq i'. \end{cases}$$

$\sigma_a^2$  e  $\sigma_e^2$  são chamados de componentes de variância. E esse tipo de modelo, com apenas um efeito fixo ( $\mu$ ) e os demais aleatórios ( $\alpha_i$  e  $\varepsilon_{ij}$ ), é chamado de modelo de efeitos aleatórios ou modelo aleatório (SEARLE et al. 1992).

Os componentes de variância ( $\sigma_a^2$  e  $\sigma_e^2$ ) podem ser estimados pelo método de análise de variância (ANOVA), e o coeficiente de correlação intraclassa pela relação

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}.$$

### 2.1.1 Método da ANOVA para Dados Balanceados

Seja, a partição da soma de quadrados total  $SQT = SQE + SQD$ , com

$$SQT = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2, \quad SQE = \sum_{i=1}^a b(\bar{y}_{i.} - \bar{y}_{..})^2 \quad \text{e} \quad SQD = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.})^2. \quad \text{Onde,}$$

SQE corresponde a soma de quadrados entre as classes e SQD a soma de quadrados dentro das classes.

As esperanças de SQD e SQE são dadas por:

$$E[SQD] = E\left[\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.})^2\right] = a(b-1)\sigma_e^2 \quad (8)$$

$$E[SQE] = E\left[\sum_{i=1}^a b(\bar{y}_{i.} - \bar{y}_{..})^2\right] = (a-1)(b\sigma_a^2 + \sigma_e^2) \quad (9)$$

Os estimadores não viciados para  $\sigma_c^2$  e  $\sigma_a^2$  são dados por:

$$\begin{aligned}\hat{\sigma}_c^2 &= \frac{SQD}{a(b-1)} = QMD \\ \hat{\sigma}_a^2 &= \frac{SQE/(a-1) - \hat{\sigma}_c^2}{b} = \frac{QME - QMD}{b}\end{aligned}\quad (10)$$

O estimador do coeficiente de correlação intraclassa obtido por este método é dado por:

$$icc = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_c^2} = \frac{QME - QMD}{(b-1)QMD + QME}\quad (11)$$

### 2.1.2 Método da ANOVA para Dados Não Balanceados

As somas de quadrados na análise de variância para dados não balanceados são dadas por:

$$SQE = \sum_{i=1}^a b_i (\bar{y}_i - \bar{y}_{..})^2 = \sum_{i=1}^a b_i \bar{y}_i^2 - N \bar{y}_{..}^2\quad (12)$$

$$SQD = \sum_{i=1}^a \sum_{j=1}^{b_i} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^a \sum_{j=1}^{b_i} y_{ij}^2 - \sum_{i=1}^a b_i \bar{y}_i^2\quad (13)$$

Essas expressões são as mesmas apresentadas no caso de dados balanceados, exceto pela substituição de  $b_i$  no lugar de  $b$ .



As esperanças das somas de quadrados são, respectivamente:

$$E(\text{SQE}) = (n - \sum_{i=1}^a b_i^2 / n) \sigma_a^2 + (a - 1) \sigma_e^2 \quad (14)$$

$$E(\text{SQD}) = (n - a) \sigma_e^2 \quad (15)$$

onde,  $n$  corresponde ao número total de observações

E os estimadores não viciados para  $\sigma_a^2$  e  $\sigma_e^2$  são dados por:

$$\hat{\sigma}_e^2 = \frac{\text{SQD}}{n - a} = \text{QMD} \quad \text{e} \quad \hat{\sigma}_a^2 = \frac{\text{SQE} - (a - 1)\text{QMD}}{(n - \sum_{i=1}^a b_i^2 / n)} = \frac{\text{QME} - \text{QMD}}{(n - \sum_{i=1}^a b_i^2 / n) / (a - 1)} \quad (16)$$

$$\text{onde, } \text{QME} = \frac{\text{SQE}}{a - 1}$$

Nesse caso o estimador do coeficiente de correlação intraclasse é dado por:

$$\text{icc} = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_e^2} = \frac{\text{QME} - \text{QMD}}{(n - \sum_{i=1}^a b_i^2 / n) [\frac{1}{a - 1} - 1] \text{QMD} + \text{QME}} \quad (17)$$

### 2.1.3 Considerações Gerais sobre o Método da ANOVA

Ao se utilizar a ANOVA para estimar os componentes de variância é necessário fazer algumas considerações:

- **Estimativas negativas de  $\sigma_a^2$ :** o método da ANOVA pode produzir estimativas negativas para este parâmetro, tanto para dados balanceados como não balanceados. Esse fato ocorre quando  $QME < QMD$ , e pode ser indicativo do uso de um modelo errado ou de que o verdadeiro valor da variância é igual a zero (SEARLE et al. 1992);
- **Suposição de normalidade:** o método da ANOVA não necessita dessa suposição para o cálculo dos estimadores dos componentes da variância. Esta suposição torna-se necessária para a definição dos intervalos de confiança e dos testes de hipóteses. Sob a suposição de distribuição normal,  $\alpha_i$  e  $\varepsilon_i$  são normalmente distribuídos com:

$$\begin{bmatrix} \alpha \\ \varepsilon \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 I_n & 0 \\ 0 & \sigma_e^2 I_e \end{bmatrix} \right) \quad (18)$$

- **Intervalos de confiança para  $\rho$ :** para dados não balanceados é possível estabelecer intervalos de confiança aproximados para este parâmetro, sob a suposição de normalidade (DONNER 1986; STATA 2001).

Neste caso o estimador do erro padrão de *icc* é dado por:

$$\sqrt{\text{var}(icc)} = \frac{2(1 - icc)}{g^2} (A + B + C) \quad (19)$$

onde,

$$g = \frac{(n - \sum_{i=1}^a b_i^2 / n)}{a - 1}, \quad A = \frac{\{1 + \text{icc}(g - 1)\}^2}{n - a}, \quad B = \frac{(1 - \text{icc}) \{1 + \text{icc}(2g - 1)\}}{a - 1},$$

$$C = \frac{\text{icc}^2 \left\{ \sum_{i=1}^a b_i^2 - 2n^{-1} \sum_{i=1}^a b_i^3 + n^{-2} (\sum_{i=1}^a b_i^2)^2 \right\}}{(a - 1)^2}$$

- **Dados categóricos:** GUILLIFORD (1999) utiliza o método da ANOVA para estimar os componentes de variância e a correlação intraclasse quando a variável resposta ( $y_{ij}$ ) é dicotômica.

Porém, SEARLE (1992) aconselha o uso desse método para dados categóricos, apenas quando as proporções variarem entre 0,2 e 0,8.

## 2.2 O Coeficiente de Correlação Intraclasse em Amostragem

Em amostragem, KISH (1965) utiliza a abordagem de componentes de variância para a estimação do coeficiente de correlação intraclasse em amostras por conglomerados de tamanhos iguais, em um ou dois estágios. Neste caso os conglomerados, ou unidades de primeiro estágio (UPA), correspondem as classes e os domicílios ou indivíduos, unidades de segundo estágio (USA), aos elementos dentro das classes.

Assim, pode-se supor que a população está segmentada em  $A$  conglomerados, cada um com  $B$  elementos, sendo que a união dos  $A$  conglomerados compõem a população de tamanho  $N=AB$ .

O coeficiente de correlação intraclasse (Rho) utilizado por KISH (1965, p.171) é equivalente à expressão (5), porém expresso a partir de componentes de variância.

$$\text{Rho} = \frac{\sigma_a^2 - \sigma_b^2 / (B-1)}{\sigma^2} = \frac{\frac{A-1}{A} S_a^2 - \frac{1}{B} S_b^2}{\frac{N-1}{N} S^2} \quad (20)$$

ou,

$$\text{Rho} = \frac{B}{B-1} \cdot \frac{\sigma_a^2}{\sigma^2} - \frac{1}{B-1} = 1 - \frac{S_b^2}{S^2} \cdot \frac{N}{N-1} \quad (21)$$

onde,

$$\bar{Y} = \frac{\sum_{i=1}^A \sum_{j=1}^B Y_{ij}}{AB} \quad \bar{Y}_i = \frac{\sum_{j=1}^B Y_{ij}}{B}$$

$\sigma_a^2 = \frac{1}{A} \sum_{i=1}^A (\bar{Y}_i - \bar{Y})^2$  : variância das médias dos  $A$  conglomerados da população;

$\sigma_b^2 = \frac{1}{AB} \sum_{i=1}^A \sum_{j=1}^B (Y_{ij} - \bar{Y}_i)^2$  : variância dos elementos dos conglomerados.

$$\sigma^2 = \frac{1}{AB} \sum_{i=1}^A \sum_{j=1}^B (Y_{ij} - \bar{Y})^2 = \sigma_a^2 + \sigma_b^2$$

$$S_a^2 = \frac{A}{A-1} \sigma_a^2 \quad S_b^2 = \frac{B}{B-1} \sigma_b^2 \quad = \quad S^2 = \frac{N}{N-1} \sigma^2$$

Da expressão (21) pode-se escrever:

$$\left( \frac{A-1}{A} \cdot \frac{N}{N-1} \right) S_a^2 = \frac{S^2}{B} [1 + \text{Rho}(B-1)] \quad (22)$$

Os limites de Rho são dados por:

- **Completa homogeneidade dentro dos conglomerados:** quando  $\sigma_b^2 = 0$  e

$$\sigma_a^2 = \sigma^2 \text{ implicando em } \text{Rho} = \frac{B}{B-1} \cdot \frac{\sigma_a^2}{\sigma^2} - \frac{1}{B-1} = \frac{B-1}{B-1} = 1;$$

- **Extrema heterogeneidade dentro dos conglomerados:** quando  $\sigma_b^2 = \sigma^2$

$$\text{e } \sigma_a^2 = 0 \text{ implica que } \text{Rho} = \frac{B}{B-1} \cdot \frac{\sigma_a^2}{\sigma^2} - \frac{1}{B-1} = -\frac{1}{B-1}.$$

Sob as suposições de que  $(A-1)/A$  e  $(N-1)/N$  são aproximadamente iguais a um, a relação  $\sigma^2 = \sigma_a^2 + \sigma_b^2$  torna-se:

$$S_{\cdot}^2 = S_a^2 + \frac{B-1}{B} S_b^2 \quad (23)$$

Substituindo  $S_{\cdot}^2$  na expressão (20) tem-se Roh:

$$\text{Roh} = \frac{S_a^2 - \frac{1}{B} S_b^2}{S_{\cdot}^2} = 1 - \frac{S_b^2}{S_{\cdot}^2} \quad (24)$$

Da mesma forma é possível escrever a expressão (22) substituindo-se Rho por Roh e  $S^2$  por  $S_{\cdot}^2$ :

$$S_a^2 = \frac{S_{\cdot}^2}{B} [1 + \text{Roh}(B-1)] \quad (25)$$

Para amostragem por conglomerados de tamanhos iguais em um estágio, tem-se que  $n = a \times B$ . Assim, a expressão (25) pode ser escrita como:

$$S_a^2 / a = \frac{S_{\cdot}^2}{aB} [1 + \text{Roh}(B-1)] \Rightarrow \frac{S_a^2 / a}{S_{\cdot}^2 / n} = [1 + \text{Roh}(B-1)] \quad (26)$$

Onde,  $\frac{S_a^2 / a}{S_{\cdot}^2 / n}$  corresponde ao efeito do delineamento (*Deff*).

Então Roh pode ser estimado como:

$$\text{deff} = \frac{s_a^2/a}{s^2/n} = [1 + \text{roh}(B-1)] \Rightarrow \text{roh} = \frac{\text{deff} - 1}{B-1} \quad (27)$$

Onde,  $s^2 = \frac{\sum_{i=1}^n (y_{ij} - \bar{y})^2}{n-1}$

No caso de amostragem por conglomerados de tamanhos iguais em dois estágios, com subamostras de tamanho  $b$ , pode-se estimar Roh por meio do  $\text{deff}$ , pela relação:

$$\text{deff} = \frac{s_a^2/a}{s^2/n} = [1 + \text{roh}(b-1)] \Rightarrow \text{roh} = \frac{\text{deff} - 1}{b-1} \quad (28)$$

Para amostragem por conglomerados de tamanhos distintos, KISH (1965) recomenda utilizar  $\bar{b}$ , ou seja, o tamanho médio da amostra do segundo estágio, desde que não existam disparidades acentuadas entre os tamanhos dos conglomerados. O autor destaca que, preferencialmente, o coeficiente de variação dos tamanhos dos conglomerados deve ser inferior a 0,10 e no máximo igual a 0,20.

Entre as formas de controlar esse coeficiente de variação destacam-se: estratificação por tamanhos, divisão e/ou agregação de conglomerados e sorteio dos conglomerados com probabilidade proporcional ao seu tamanho (PPT).

KISH (1965) também estabelece uma relação, entre os custos da amostragem em dois estágios e  $\text{roh}$ , para definir o tamanho da amostra de elementos dentro dos conglomerados. Assim, em uma amostra em dois estágios, o custo total pode ser

decomposto a partir da relação  $C = nc + aC_a = nc(1 + C_a/cb)$ ; onde  $n$  representa o tamanho total da amostra de elementos;  $a$  o número de conglomerados na amostra;  $C_a$  o custo associado ao conglomerado e  $c$  o custo associado ao elemento.

Assumindo que  $b = n/a$  é uma boa estimativa do tamanho da subamostra, isto é o número de elementos a serem selecionados em cada conglomerado, é possível estimar um número *ótimo* para esta subamostra por meio da relação entre os custos da amostragem e a correlação intraclasse:

$$b_{\text{ótimo}}^* = \sqrt{\frac{C_a}{c} \left( \frac{1 - \text{roh}}{\text{roh}} \right)} \quad (29)$$

Onde,

$b_{\text{ótimo}}^*$ : tamanho da amostra de elementos dentro do conglomerado;

$C_a$ : custo associado ao conglomerado;

$c$ : custo associado ao elemento;

$\text{roh}$ : estimativa do coeficiente de correlação intraclasse.

Assim, o número de conglomerados na amostra pode ser definido a partir da relação  $n = a \cdot b_{\text{ótimo}}^* \Rightarrow a = n/b_{\text{ótimo}}^*$ .



### **3 OBJETIVOS**

- Comparar dois métodos de estimação do coeficiente de correlação intraclasse, no contexto de inquéritos domiciliares, em que o setor censitário é a unidade primária de amostragem;
- Avaliar o efeito da homogeneidade intra setor censitário na amostra da Pesquisa de Condições de Vida no Estado de São Paulo, no ano de 1998.

## **4 METODOLOGIA**

### **4.1 Fonte dos dados**

A base de dados utilizada no estudo é referente aos dados coletados pela Pesquisa de Condições de Vida (PCV) no ano de 1998. Essa pesquisa é realizada de quatro em quatro anos pela Fundação Sistema Estadual de Análise de Dados (SEADE), na Região Metropolitana de São Paulo (RMSP) e municípios do Interior do Estado de São Paulo com pelo menos 50.000 habitantes em sua área urbana.

A PCV é uma pesquisa domiciliar por amostragem, que tem por objetivo captar e produzir informações que permitam avaliar as condições de vida da população paulista, através de diferentes recortes analíticos, em particular níveis de pobreza.

A tomada de campo de 1998 incluiu 103 municípios do Estado de São Paulo – 39 pertencentes a RMSP, nove à Região Metropolitana da Baixada Santista e os demais distribuídos em cinco domínios no Interior do Estado. A relação dos municípios pertencentes a cada domínio está apresentada no ANEXO I.

A amostra da pesquisa foi obtida através de uma amostragem por conglomerados em dois estágios, onde a unidade de primeiro estágio (UPA) é o setor censitário, e a de segundo estágio (USA) é o domicílio particular. Com exceção da RMSP e RM da Baixada Santista, alguns setores censitários foram agrupados ou divididos antes da coleta de dados, com a finalidade de se controlar o seu tamanho (número de domicílios). Porém, o nome setor censitário será adotado em todos os sete domínios.

Para as duas regiões metropolitanas os setores censitários foram selecionados com probabilidades proporcionais ao seu tamanho, a medida de tamanho utilizado foi o

total de domicílios particulares existentes no setor. Para os demais domínios os setores foram sorteados com igual probabilidade. Os domicílios (USA) foram selecionados a partir de uma amostra aleatória simples. A Tabela 1 apresenta para cada domínio o número de setores censitários e domicílios sorteados com as respectivas frações amostrais.

**Tabela 1** – Total de setores censitários e domicílios sorteados e frações amostrais, segundo domínio amostral. Municípios do Estado de São Paulo, 1998.

Domínios Amostrais	Total de setores censitários sorteados *	Total de domicílios sorteados	Amostra final de domicílios	Fração Amostral (f)
Aglomerado Central	91	1.751	1.261	1/251
Aglomerado Leste	106	1.784	1.084	1/580
RM da Baixada				1/266
Santistas	120	1.760	1.067	
Aglomerado Norte	110	1.739	1.114	1/257
Aglomerado Oeste	115	1.731	1.220	1/234
Vale do Paraíba	88	1.737	1.146	1/203
RMSP	508	4.500	3.628	1/1.067
Total	1.138	14.700	10.520	-

Fonte: Fundação SEADE.

\* Setores Censitários referentes ao Censo Demográfico de 1991, produzidos pela Fundação IBGE.

## 4.2 Estimadores

Para cada uma das variáveis selecionadas no estudo, foram calculadas duas estimativas do coeficiente de correlação intraclasse para os parâmetros populacionais de interesse (média ou proporções).

Assim, para cada domínio as estimativas das médias e proporções foram obtidas, por meio do estimador dado pela expressão (30). Pois, no caso de amostragem por conglomerados com tamanhos desiguais a média amostral (ou proporção) é um estimador do tipo razão (KISH 1965).

$$r = \frac{y}{x} = \frac{1}{x} \sum_{i=1}^a y_i = \frac{\sum_{i=1}^a y_i}{\sum_{i=1}^a b_i} \quad (30)$$

onde,  $a$  é o número de setores censitários (UPA) na amostra no domínio;

$x = \sum_{i=1}^a b_i = n$  é o “tamanho” da amostra;  $b_i$  é o tamanho da subamostra do  $a$ -ésimo

setor censitário;  $y_i$  é o valor da variável de interesse do elemento  $i$  no domínio. No caso em que  $Y$  é uma variável dicotômica  $y_i$  assume os valores 1, se o elemento possui a característica em estudo, e 0 em caso contrário.

A variância de  $r$  foi estimada pelo Método de Linearização de Taylor, tal que

$$\text{var}(r) = \frac{1}{x^2} [\text{var}(y) + r^2 \text{var}(x) - 2r \text{cov}(x, y)] \quad (\text{KISH 1965}).$$

É importante destacar que no cálculo das variâncias não foi utilizada a correção para população finita  $(1 - f)$ , onde  $f$  é a fração amostral. No caso da amostra da

Pesquisa de Condições de Vida, essa correção é aproximadamente um, como pode ser observado pelo tamanho das frações amostrais apresentadas na Tabela 1.

### 4.2.1 Método da ANOVA

A correlação intraclasse estimada através do método da ANOVA é dada por:

$$icc = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_b^2} = \frac{QME - QMD}{\left( \frac{n - \sum_i b_i^2 / n}{a - 1} - 1 \right) QMD + QME} \quad (31)$$

onde,  $a$  corresponde a amostra de setores censitários no domínio;  $b_i$  é o tamanho da subamostra no setor censitário  $i$ ;  $n = \sum_{i=1}^a b_i$  é o tamanho da amostra;

$$QME = \frac{SQE}{a - 1} = \frac{\sum_{i=1}^a b_i (\bar{y}_i - \bar{y}_{..})^2}{a - 1}; \quad e \quad QMD = \frac{SQD}{n - a} = \frac{\sum_{i=1}^a \sum_{j=1}^{b_i} (y_{ij} - \bar{y}_i)^2}{n - a} \quad (\text{SEARLE 1992}).$$

Os Intervalos de Confiança de 95% foram calculados como:

$$\left[ icc - 1,96\sqrt{\text{var}(icc)}; icc + 1,96\sqrt{\text{var}(icc)} \right] \quad (32)$$

Onde,

$$\sqrt{\text{var}(\text{icc})} = \frac{2(1-\text{icc})}{g^2} (A + B + C) \quad (33)$$

$$g = \frac{(n - \sum_{i=1}^a b_i^2 / n)}{a-1}, \quad A = \frac{\{1 + \text{icc}(g-1)\}^2}{n-a}, \quad B = \frac{(1-\text{icc})\{1 + \text{icc}(2g-1)\}}{a-1},$$

$$C = \frac{\text{icc}^2 \left\{ \sum_{i=1}^a b_i^2 - 2n^{-1} \sum_{i=1}^a b_i^3 + n^{-2} (\sum_{i=1}^a b_i^2)^2 \right\}}{(a-1)^2}$$

#### 4.2.2 Método KISH

O efeito de delineamento estimado a partir do método proposto por Kish (1965) é dado por:

$$\text{deff} = \frac{\text{var}(r)}{s^2/n} \quad (34)$$

com,  $s^2 = \frac{\sum_{i=1}^n (y_{ij} - \bar{y})^2}{n-1}$ . para variáveis contínuas e  $s_p^2 = p(1-p)$  para variáveis

categóricas.

O estimador da correlação intraclasse é dado por:

$$\text{roh} = \frac{\text{deff} - 1}{b-1} \quad (35)$$

onde,  $d_{eff}$  corresponde a estimativa do efeito do delineamento e  $\bar{b}$  ao número médio de elementos amostrados por setor censitário.

### 4.2.3 Cálculo do $b_{\text{ótimo}}^*$

Os tamanhos de ótimos das subamostras ( $b_{\text{ótimo}}^*$ ), considerando-se a correlação intraclasse e os custos associados aos setores censitários e aos elementos, foram obtidos a partir da expressão:

$$b_{\text{ótimo}}^{\text{icc}} = \sqrt{\frac{C_a}{c} \left( \frac{1 - \text{icc}}{\text{icc}} \right)} \quad (36)$$

O custo associado ao conglomerado é denotado por  $C_a$  e o do elemento como  $c$ . A relação entre estes dois custos ( $C_a/c$ ) é estimada como sendo igual a 20, sendo este valor o utilizado no planejamento da amostra da Pesquisa de Condições de Vida.

### **4.3 Processamento dos Dados**

Os valores de *icc* e respectivos erros padrões foram calculados através do programa computacional STATA versão 7.0 (STATA 2001).

As estimativas das variâncias, *var(r)*, foram obtidas através da rotina PROC SURVEYMEANS, do programa computacional SAS versão 8.0, que utiliza o Método de Linearização de Taylor. (NA e WATTS 2000).

Todos os demais cálculos foram realizados em planilha excel para windows 2000.

### **4.4 Variáveis Utilizadas no Estudo**

A Tabela 2 descreve as variáveis utilizadas no trabalho, no ANEXO 2 estão apresentadas as definições e conceitos associados as variáveis.



Tabela 2 – Variáveis utilizadas no estudo. Pesquisa de Condições de Vida.

1998.

Variáveis	Descrição
<i>Variáveis Quantitativas</i>	
X1. Renda domiciliar total	Valor em reais da renda domiciliar
X2. Renda domiciliar per capita	Valor em reais da renda domiciliar per capita
X3. Renda familiar total	Valor em reais da renda familiar
X4. Renda familiar per capita	Valor em reais da renda familiar per capita
X5. Anos de estudo	Número de anos de estudo: 0, 1, 2,...,17
X6. Qualidade do atendimento recebido	Escala de 0 a 10 pontos
<i>Variáveis Dicotômicas</i>	
X7. Domicílios com renda até 10 SM	Assume valor 1 em caso positivo; 0 caso contrário
X8. Domicílios com renda até 3 SM	Assume valor 1 em caso positivo; 0 caso contrário
X9. Famílias com renda até 10 SM	Assume valor 1 em caso positivo; 0 caso contrário
X10. Famílias com renda até 3 SM	Assume valor 1 em caso positivo; 0 caso contrário
X11. Existência de telefone fixo no domicílio	1 = domicílio com telefone; 0 = domicílio sem telefone
X12. Chefes com no máximo 4 anos de escolaridade	Assume valor 1 em caso positivo; 0 caso contrário
X13. Chefes com 5 a 7 anos de escolaridade	Assume valor 1 em caso positivo; 0 caso contrário
X14. Chefes com ensino fundamental completo	Assume valor 1 em caso positivo; 0 caso contrário
X15. Chefes com ensino médio completo	Assume valor 1 em caso positivo; 0 caso contrário
X16. Chefes com ensino superior completo	Assume valor 1 em caso positivo; 0 caso contrário
X17. Desempregados	1 = indivíduos de 10 anos ou mais desempregados; 0 caso contrário
X18. População economicamente ativa	1 = indivíduos de 10 anos ou mais ocupados ou desempregados; 0 caso contrário
X19. Assalariados com carteira ou do setor público	Assume valor 1 em caso positivo; 0 caso contrário
X20. Utilização de serviços de saúde	1 = indivíduos que utilizaram serviços de saúde; 0 caso contrário
X21. Portador de deficiência física ou mental	1 = portadores de deficiência física ou mental; 0 caso contrário
X22. Portador de dificuldades	1 = portadores de dificuldades; 0 caso contrário
X23. Uso de óculos ou lentes de contato	1 = usuários de óculos ou lentes de contato; 0 caso contrário
X24. Vítimas de roubos ou assaltos	1 = indivíduos que sofreram roubos ou assaltos; 0 caso contrário

## 5 RESULTADOS

### 5.1 Estimação do Coeficiente de Correlação Intraclasse

A Tabela 3 mostra que os dois estimadores (*icc* e *roh*) apresentam resultados semelhantes. Para as 24 variáveis analisadas a maioria dos valores de *roh* está incluída no intervalo de confiança de 95% de *icc*. Nos domínios do Interior, o Vale do Paraíba apresentou o menor percentual de variáveis para as quais *roh* pertence ao intervalo de confiança de *icc*, 82,6%. Todos os demais domínios do Interior apresentam percentuais superiores a este. Na Região Metropolitana de São Paulo observa-se que para 66,7 % das variáveis os valores de *roh* pertencem ao intervalo de confiança de *icc*.

**Tabela 3** –Número e porcentagem de variáveis para as quais *roh* está incluído no intervalo de confiança de 95% para *icc*. Pesquisa de Condições de Vida 1998.

Domínio	Variáveis	Variáveis onde <i>roh</i>	
		está incluído no	%
		<i>IC</i> <sub>95%</sub> de <i>icc</i>	
RMSP	24	16	66,7
Central	24	24	100,0
Leste	24	21	87,5
RM da Baixada Santista	23	21	91,3
Norte	23	22	95,7
Oeste	24	23	95,8
Vale do Paraíba	23	19	82,6

As Tabelas 4 a 10 apresentam as estimativas do coeficiente de correlação intraclasse, segundo os dois métodos de estimação, considerando-se os sete domínios.

Nos sete domínios, as variáveis de rendimentos são as responsáveis pelas maiores estimativas para o coeficiente de correlação intraclasse, independentemente do método de estimação.

Para essas variáveis as estimativas são quase sempre maiores que 0.15, que é usualmente apresentado na literatura como um valor de referência. Valores acima de 0,15 são considerados “grandes” (KALTON 1983).

Entre todos os domínios da pesquisa, a Região Metropolitana de São Paulo é a que apresenta os maiores valores do coeficiente de correlação intraclasse para essas variáveis. Para as demais variáveis, as estimativas são inferiores a 0,15 em todos os domínios amostrais.

No conjunto das 24 variáveis estudadas as estimativas de *roh* tendem a ser maiores do que as obtidas através do *icc*. Porém, quando se observam as variáveis responsáveis pelas maiores estimativas da correlação intraclasse, as de rendimentos, este resultado se altera dependendo do domínio. Enquanto para a RMSP os valores de *roh* são maiores que *icc*, o mesmo não ocorre para alguns domínios do Interior do Estado. Para essas variáveis, os domínios onde as estimativas produzidas por *roh* são menores que as produzidas por *icc* são aqueles com maiores tamanhos médios de subamostras.

Foram observadas três estimativas negativas para o coeficiente de correlação intraclasse para as variáveis: *Chefes com ensino fundamental completo* na RM da Baixada Santista, e *Portador de deficiência física e mental* nos Aglomerados Norte e Vale do Paraíba.

**Tabela 4** – Estimativas do coeficiente de correlação Intraclasse. Pesquisa de Condições de Vida – Região Metropolitana de São Paulo. 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	roh	icc	Erro Padrão icc	Intervalo de Confiança (95%)		
<i>Variáveis Contínuas</i>										
X2. Renda domiciliar per capita (Em reais)	2.804	499	5,62	518,29	0,3562	0,3582	0,0234	0,3123	—	0,4041
X4. Renda familiar per capita (Em reais)	2.923	499	5,86	510,74	0,3442	0,3543	0,0230	0,3092	—	0,3994
X3. Renda familiar total (Em reais)	2.923	499	5,86	1486,01	0,3052	0,2757	0,0218	0,2330	—	0,3184
X1. Renda domiciliar total (Em reais)	2.804	499	5,62	1506,21	0,3164	0,2756	0,0222	0,2321	—	0,3191
X5. Anos de estudo	11.315	508	22,27	6,89	0,1282	0,0985	0,0138	0,0715	—	0,1255
X6. Qualidade do atendimento recebido	3.138	487	6,44	8,06	0,1494	0,0491	0,0133	0,0230	—	0,0752
<i>Variáveis Dicotômicas</i>										
X16. Chefes com ensino superior completo	3.401	508	6,69	0,12	0,3510	0,2963	0,0205	0,2562	—	0,3364
X11. Existência de telefone fixo no domicílio	3.400	508	6,69	0,45	0,4289	0,2580	0,0197	0,2195	—	0,2965
X7. Domicílios com renda até 10 SM	2.804	499	5,62	0,66	0,2998	0,2198	0,0209	0,1788	—	0,2608
X9. Famílias com renda até 10 SM	2.923	499	5,86	0,67	0,4839	0,2149	0,0204	0,1749	—	0,2549
X23. Uso de óculos ou lentes de contato	12.690	508	24,98	0,29	0,1317	0,0995	0,0088	0,0822	—	0,1168
X15. Chefes com ensino médio completo	3.401	508	6,69	0,16	0,0780	0,0594	0,0124	0,0352	—	0,0837
X13. Chefes com 5 a 7 anos de escolaridade	3.401	508	6,69	0,35	0,0902	0,0490	0,0119	0,0258	—	0,0723
X10. Famílias com renda até 3 SM	2.923	499	5,86	0,21	0,0226	0,0325	0,0134	0,0062	—	0,0588
X8. Domicílios com renda até 3 SM	2.804	499	5,62	0,21	0,0972	0,0305	0,0138	0,0035	—	0,0575
X20. Utilização de serviços de saúde	13.000	508	25,59	0,25	0,0339	0,0303	0,0044	0,0217	—	0,0389
X24. Vítimas de roubos ou assaltos	13.002	508	25,59	0,06	0,0261	0,0290	0,0044	0,0205	—	0,0375
X19. Assalariados com carteira ou do setor público	3.910	508	7,70	0,71	0,0335	0,0272	0,0101	0,0074	—	0,0470
X17. Desempregados	6.704	508	13,20	0,18	0,0432	0,0225	0,0062	0,0103	—	0,0347
X12. Chefes com no máximo 4 anos de escolaridade	3.401	508	6,69	0,24	0,0780	0,0213	0,0105	0,0007	—	0,0419
X22. Portador de dificuldades	12.833	508	25,26	0,05	0,0184	0,0127	0,0033	0,0062	—	0,0192
X14. Chefes com ensino fundamental completo	3.401	508	6,69	0,14	0,0137	0,0071	0,0098	0,0000	—	0,0264
X18. População economicamente ativa	10.786	508	21,23	0,62	0,0074	0,0036	0,0032	0,0000	—	0,0100
X21. Portador de deficiência física ou mental	13.002	508	25,59	0,01	0,0026	0,0032	0,0027	0,0000	—	0,0085

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

**Tabela 5** – Estimativas do coeficiente de correlação Intraclasse. Pesquisa de Condições de Vida – Aglomerado Central. 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	roh	icc	Erro Padrão icc	Intervalo de Confiança (95%)		
<i>Variáveis Contínuas</i>										
X4. Renda familiar per capita (Em reais)	1.280	91	14,07	406,86	0,1846	0,1965	0,0320	0,1338	—	0,2592
X2. Renda domiciliar per capita (Em reais)	1.217	91	13,37	415,95	0,1828	0,1927	0,0330	0,1281	—	0,2573
X3. Renda familiar total (Em reais)	1.280	91	14,07	1.205,42	0,1796	0,1921	0,0326	0,1281	—	0,2561
X1. Renda domiciliar total (Em reais)	1.217	91	13,37	1.232,02	0,1769	0,1861	0,0325	0,1225	—	0,2497
X6. Qualidade do atendimento recebido	1.091	91	11,99	8,22	0,0541	0,0520	0,0200	0,0129	—	0,0911
X5. Anos de estudo	4.640	91	50,99	7,04	0,0517	0,0465	0,0262	0,0000	—	0,0979
<i>Variáveis Dicotômicas</i>										
X11. Existência de telefone fixo no domicílio	1.486	91	16,33	0,44	0,3135	0,2661	0,0368	0,1939	—	0,3383
X16. Chefes com ensino superior completo	1.488	91	15,58	0,11	0,1773	0,1845	0,0306	0,1245	—	0,2445
X9. Famílias com renda até 10 SM	1.280	91	14,07	0,70	0,1605	0,1565	0,0294	0,0988	—	0,2142
X7. Domicílios com renda até 10 SM	1.217	91	13,37	0,69	0,1552	0,1532	0,0294	0,0955	—	0,2109
X8. Domicílios com renda até 3 SM	1.217	91	13,37	0,20	0,1145	0,1023	0,0240	0,0553	—	0,1493
X15. Chefes com ensino médio completo	1.488	91	15,58	0,15	0,0992	0,0996	0,0220	0,0564	—	0,1428
X10. Famílias com renda até 3 SM	1.280	91	14,07	0,21	0,1125	0,0992	0,0234	0,0533	—	0,1451
X12. Chefes com no máximo 4 anos de escolaridade	1.488	91	15,58	0,25	0,0931	0,0793	0,0197	0,0408	—	0,1178
X23. Uso de óculos ou lentes de contato	5.202	91	57,16	0,27	0,0711	0,0620	0,0117	0,0390	—	0,0850
X13. Chefes com 5 a 7 anos de escolaridade	1.488	91	15,58	0,35	0,0763	0,0599	0,0173	0,0260	—	0,0938
X14. Chefes com ensino fundamental completo	1.488	91	15,58	0,13	0,0315	0,0327	0,0138	0,0057	—	0,0597
X19. Assalariados com carteira ou do setor público	1.532	91	16,84	0,75	0,0273	0,0247	0,0247	0,0000	—	0,0730
X17. Desempregados	2.516	91	27,65	0,14	0,0222	0,0212	0,0086	0,0044	—	0,0380
X20. Utilização de serviços de saúde	5.318	91	58,44	0,22	0,0276	0,0209	0,0057	0,0097	—	0,0321
X22. Portador de dificuldades	5.247	91	57,66	0,05	0,0132	0,0176	0,0053	0,0073	—	0,0279
X24. Vítimas de roubos ou assaltos	5.318	91	58,44	0,04	0,0113	0,0148	0,0048	0,0054	—	0,0242
X18. População economicamente ativa	4.443	91	48,82	0,57	0,0146	0,0135	0,0051	0,0035	—	0,0235
X21. Portador de deficiência física ou mental	5.318	91	58,44	0,01	0,0053	0,0026	0,0030	0,0000	—	0,0084

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

**Tabela 6** – Estimativas do coeficiente de correlação Intraclasse. Pesquisa de Condições de Vida – Aglomerado Leste. 1998

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	roh	icc	Erro Padrão icc	Intervalo de Confiança (95%)		
<i>Variáveis Contínuas</i>										
X2. Renda domiciliar per capita (Em reais)	1.039	106	9,8	465,4	0,2916	0,3855	0,0437	0,2999	—	0,4711
X4. Renda familiar per capita (Em reais)	1.100	106	10,38	453,77	0,2857	0,3670	0,0405	0,2876	—	0,4464
X1. Renda domiciliar total (Em reais)	1.039	106	9,8	1.364,16	0,1971	0,2651	0,0391	0,1885	—	0,3417
X3. Renda familiar total (Em reais)	1.100	106	10,38	1.329,29	0,1988	0,2626	0,0386	0,1869	—	0,3383
X6. Qualidade do atendimento recebido	1.239	106	11,69	8,47	0,0407	0,0518	0,0187	0,0152	—	0,0884
X5. Anos de estudo	4.534	106	42,77	6,91	0,0378	0,0327	0,0261	0,0000	—	0,0839
<i>Variáveis Dicotômicas</i>										
X16. Chefes com ensino superior completo	1.424	106	12,66	0,12	0,2049	0,2312	0,0336	0,1654	—	0,2970
X11. Existência de telefone fixo no domicílio	1.421	106	13,41	0,50	0,2750	0,2265	0,0333	0,1613	—	0,2917
X7. Domicílios com renda até 10 SM	1.039	106	9,8	0,65	0,1410	0,1387	0,0298	0,0804	—	0,1970
X9. Famílias com renda até 10 SM	1.100	106	10,38	0,66	0,1480	0,1345	0,0288	0,0781	—	0,1909
X12. Chefes com no máximo 4 anos de escolaridade	1.424	106	12,66	0,25	0,1029	0,0792	0,0200	0,0400	—	0,1184
X23. Uso de óculos ou lentes de contato	4.990	106	47,08	0,30	0,0827	0,0763	0,0134	0,0500	—	0,1026
X19. Assalariados com carteira ou do setor público	1.597	106	15,07	0,71	0,0790	0,0565	0,0166	0,0240	—	0,0890
X13. Chefes com 5 a 7 anos de escolaridade	1.424	106	12,66	0,36	0,0521	0,0531	0,0171	0,0197	—	0,0865
X15. Chefes com ensino médio completo	1.424	106	12,66	0,14	0,0475	0,0404	0,0156	0,0099	—	0,0709
X8. Domicílios com renda até 3 SM	1.039	106	9,8	0,19	0,0384	0,0362	0,0191	0,0000	—	0,0736
X10. Famílias com renda até 3 SM	1.100	106	10,38	0,20	0,0464	0,0362	0,0183	0,0004	—	0,0720
X17. Desempregados	2.668	106	25,17	0,15	0,0505	0,0280	0,0094	0,0097	—	0,0463
X24. Vítimas de roubos ou assaltos	5.159	106	48,67	0,05	0,0192	0,0212	0,0059	0,0097	—	0,0327
X20. Utilização de serviços de saúde	5.159	106	48,67	0,25	0,0227	0,0210	0,0058	0,0096	—	0,0324
X14. Chefes com ensino fundamental completo	1.424	106	12,66	0,13	0,0355	0,0200	0,0131	0,0000	—	0,0457
X22. Portador de dificuldades	5.052	106	47,66	0,05	0,0241	0,0165	0,0052	0,0062	—	0,0268
X18. População economicamente ativa	4.343	106	40,97	0,61	0,0020	0,0040	0,0040	0,0000	—	0,0118
X21. Portador de deficiência física ou mental	5.159	106	48,67	0,02	0,0036	0,0017	0,0031	0,0000	—	0,0078

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

**Tabela 7 – Estimativas do coeficiente de correlação Intraclasse. Pesquisa de Condições de Vida – Região Metropolitana da Baixada Santista. 1998.**

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	roh	icc	Erro Padrão icc	Intervalo de Confiança (95%)		
<i>Variáveis Contínuas</i>										
X4. Renda familiar per capita (Em reais)	1.096	120	9,13	476,73	0,2431	0,2527	0,0299	0,1941	—	0,3113
X3. Renda familiar total (Em reais)	1.096	120	9,13	1.387,22	0,2270	0,2417	0,0351	0,1730	—	0,3104
X2. Renda domiciliar per capita (Em reais)	1.032	120	8,6	483,94	0,2355	0,2403	0,0357	0,1703	—	0,3103
X1. Renda domiciliar total (Em reais)	1.032	120	8,6	1.407,87	0,2184	0,2301	0,0352	0,1612	—	0,2990
X6. Qualidade do atendimento recebido	1.072	120	8,93	8,11	0,1013	0,0513	0,0187	0,0147	—	0,0879
X5. Anos de estudo	4.235	120	35,29	7,26	0,0508	0,0393	0,0236	0,0000	—	0,0856
<i>Variáveis Dicotômicas</i>										
X11. Existência de telefone fixo no domicílio	1.336	120	11,13	0,45	0,2925	0,2451	0,0332	0,1800	—	0,3102
X9. Famílias com renda até 10 SM	1.096	120	9,13	0,65	0,1995	0,1936	0,0322	0,1305	—	0,2567
X7. Domicílios com renda até 10 SM	1.032	120	8,6	0,64	0,2050	0,1932	0,0329	0,1286	—	0,2578
X16. Chefes com ensino superior completo	1.283	120	10,69	0,09	0,1699	0,1720	0,0285	0,1162	—	0,2278
X12. Chefes com no máximo 4 anos de escolaridade	1.283	120	10,69	0,26	0,0945	0,0784	0,0204	0,0383	—	0,1184
X15. Chefes com ensino médio completo	1.283	120	10,69	0,18	0,0770	0,0739	0,0200	0,0347	—	0,1131
X23. Uso de óculos ou lentes de contato	4.747	120	39,56	0,31	0,0861	0,0725	0,0123	0,0484	—	0,0966
X19. Assalariados com carteira ou do setor público	1.268	120	10,57	0,73	0,0858	0,0648	0,0199	0,0258	—	0,1038
X10. Famílias com renda até 3 SM	1.096	120	9,13	0,20	0,0720	0,0500	0,0203	0,0103	—	0,0897
X8. Domicílios com renda até 3 SM	1.032	120	8,6	0,19	0,0738	0,0499	0,0212	0,0084	—	0,0914
X13. Chefes com 5 a 7 anos de escolaridade	1.283	120	10,69	0,33	0,0541	0,0363	0,0047	0,0271	—	0,0456
X20. Utilização de serviços de saúde	4.903	120	40,86	0,23	0,0422	0,0334	0,0075	0,0187	—	0,0481
X22. Portador de dificuldades	4.819	120	40,16	0,06	0,0157	0,0193	0,0057	0,0080	—	0,0306
X24. Vítimas de roubos ou assaltos	4.903	120	40,86	0,07	0,0142	0,0190	0,0057	0,0079	—	0,0301
X18. População economicamente ativa	4.044	120	33,7	0,56	0,0188	0,0184	0,0062	0,0000	—	0,0306
X17. Desempregados	2.259	120	18,83	0,18	0,0113	0,0110	0,0084	0,0000	—	0,0275
X21. Portador de deficiência física ou mental	4.903	120	40,86	0,02	0,0049	0,0078	0,0042	0,0000	—	0,0161
X14. Chefes com ensino fundamental completo*	1.283	120	10,69	0,14	-	-	-	-	—	-

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

Estimativas negativas do coeficiente de correlação intraclasse.

**Tabela 8 – Estimativas do coeficiente de correlação Intraclasse. Pesquisa de Condições de Vida – Aglomerado Norte. 1998**

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	roh	icc	Erro Padrão icc	Intervalo de Confiança (95%)		
<i>Variáveis Contínuas</i>										
X2. Renda domiciliar per capita (Em reais)	1.055	110	9,59	409,36	0,1858	0,2832	0,0392	0,2064	—	0,3600
X4. Renda familiar per capita (Em reais)	1.139	110	10,35	396,72	0,1828	0,2820	0,0350	0,2134	—	0,3506
X1. Renda domiciliar total (Em reais)	1.055	110	9,59	1.251,21	0,1717	0,2507	0,0375	0,1772	—	0,3242
X3. Renda familiar total (Em reais)	1.139	110	10,35	1.211,54	0,1668	0,2499	0,0367	0,1780	—	0,3218
X6. Qualidade do atendimento recebido	1.255	110	11,41	8,49	0,0819	0,0659	0,0203	0,0262	—	0,1056
X5. Anos de estudo	4.404	110	40,04	6,99	0,0404	0,0350	0,0208	0,0000	—	0,0758
<i>Variáveis Dicotômicas</i>										
X11. Existência de telefone fixo no domicílio	1.364	110	12,4	0,52	0,2986	0,2246	0,0330	0,1599	—	0,2893
X16. Chefes com ensino superior completo	1.317	110	11,97	0,10	0,1579	0,1889	0,0304	0,1293	—	0,2485
X7. Domicílios com renda até 10 SM	1.055	110	9,59	0,68	0,1207	0,1194	0,0276	0,0654	—	0,1734
X9. Famílias com renda até 10 SM	1.139	110	10,35	0,70	0,1178	0,1183	0,0265	0,0663	—	0,1703
X23. Uso de óculos ou lentes de contato	4.851	110	44,1	0,31	0,0743	0,0623	0,0114	0,0400	—	0,0846
X19. Assalariados com carteira ou do setor público	1.461	110	13,28	0,77	0,0666	0,0558	0,0173	0,0219	—	0,0897
X15. Chefes com ensino médio completo	1.317	110	11,97	0,14	0,0480	0,0453	0,0167	0,0126	—	0,0779
X17. Desempregados	2.488	110	22,62	0,16	0,0560	0,0433	0,0116	0,0205	—	0,0661
X13. Chefes com 5 a 7 anos de escolaridade	1.317	110	11,97	0,37	0,0427	0,0379	0,0158	0,0069	—	0,0689
X12. Chefes com no máximo 4 anos de escolaridade	1.317	110	11,97	0,27	0,0500	0,0329	0,0152	0,0030	—	0,0627
X14. Chefes com ensino fundamental completo	1.317	110	11,97	0,13	0,0323	0,0277	0,0146	0,0000	—	0,0564
X10. Famílias com renda até 3 SM	1.139	110	10,35	0,21	0,0421	0,0257	0,0167	0,0000	—	0,0585
X20. Utilização de serviços de saúde	5.005	110	45,5	0,27	0,0372	0,0245	0,0063	0,0121	—	0,0369
X8. Domicílios com renda até 3 SM	1.055	110	9,59	0,20	0,0364	0,0215	0,0174	0,0000	—	0,0556
X24. Vítimas de roubos ou assaltos	5.005	110	45,5	0,04	0,0135	0,0194	0,0056	0,0084	—	0,0304
X22. Portador de dificuldades	4.931	110	44,83	0,06	0,0163	0,0125	0,0047	0,0032	—	0,0218
X18. População economicamente ativa	4.231	110	38,46	0,59	0,0064	0,0119	0,0119	0,0000	—	0,0351
X21. Portador de deficiência física ou mental*	5.005	110	45,5	0,01	-	-	-	-	—	-

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

Estimativas negativas do coeficiente de correlação intraclasse.



**Tabela 9** – Estimativas do coeficiente de correlação Intraclasse. Pesquisa de Condições de Vida – Aglomerado Oeste, 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	roh	icc	Erro Padrão icc	Intervalo de Confiança (95%)		
<i>Variáveis Contínuas</i>										
X2. Renda domiciliar per capita (Em reais)	1.162	115	10,1	395,54	0,1248	0,1736	0,0307	0,1134	—	0,2338
X4. Renda familiar per capita (Em reais)	1.238	115	10,77	383,52	0,1242	0,1719	0,0300	0,1131	—	0,2307
X1. Renda domiciliar total (Em reais)	1.162	115	10,1	1141,29	0,1249	0,1610	0,0297	0,1028	—	0,2192
X3. Renda familiar total (Em reais)	1.238	115	10,77	1106,74	0,1246	0,1598	0,0290	0,1029	—	0,2167
X6. Qualidade do atendimento recebido	1.123	115	9,77	8,38	0,1500	0,0923	0,0244	0,0446	—	0,1400
X5. Anos de estudo	4.406	115	38,31	7,39	0,0243	0,0236	0,0204	0,0000	—	0,0635
<i>Variáveis Dicotômicas</i>										
X16. Chefes com ensino superior completo	1.381	115	12,01	0,13	0,1956	0,1839	0,0292	0,1267	—	0,2411
X11. Existência de telefone fixo no domicílio	1.447	115	12,58	0,45	0,1871	0,1599	0,0273	0,1065	—	0,2133
X7. Domicílios com renda até 10 SM	1.162	115	10,1	0,71	0,1139	0,1099	0,0253	0,0603	—	0,1595
X9. Famílias com renda até 10 SM	1.238	115	10,77	0,72	0,1188	0,1089	0,0245	0,0608	—	0,1570
X15. Chefes com ensino médio completo	1.381	115	12,01	0,17	0,0371	0,0478	0,0164	0,0156	—	0,0800
X19. Assalariados com carteira ou do setor público	1.418	115	12,33	0,72	0,0586	0,0478	0,0160	0,0164	—	0,0792
X13. Chefes com 5 a 7 anos de escolaridade	1.381	115	12,01	0,32	0,0472	0,0471	0,0164	0,0151	—	0,0791
X12. Chefes com no máximo 4 anos de escolaridade	1.381	115	12,01	0,29	0,0515	0,0410	0,0157	0,0102	—	0,0717
X23. Uso de óculos ou lentes de contato	4.819	115	41,9	0,32	0,0504	0,0391	0,0391	0,0000	—	0,1158
X18. População economicamente ativa	4.215	115	36,65	0,59	0,0307	0,0264	0,0071	0,0125	—	0,0403
X8. Domicílios com renda até 3 SM	1.162	115	10,1	0,24	0,0327	0,0132	0,0153	0,0000	—	0,0432
X10. Famílias com renda até 3 SM	1.238	115	10,77	0,25	0,0343	0,0128	0,0144	0,0000	—	0,0410
X21. Portador de deficiência física ou mental	4.967	115	43,19	0,02	0,0170	0,0116	0,0046	0,0025	—	0,0207
X17. Desempregados	2.468	115	21,46	0,16	0,0221	0,0113	0,0077	0,0000	—	0,0265
X20. Utilização de serviços de saúde	4.966	115	43,18	0,24	0,0093	0,0097	0,0044	0,0011	—	0,0183
X14. Chefes com ensino fundamental completo	1.381	115	12,01	0,10	0,0083	0,0087	0,0120	0,0000	—	0,0322
X24. Vítimas de roubos ou assaltos	4.967	115	43,19	0,04	0,0123	0,0084	0,0042	0,0002	—	0,0166
X22. Portador de dificuldades	4.859	115	42,25	0,05	0,0075	0,0058	0,0039	0,0000	—	0,0135

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

Tabela 10 – Estimativas do coeficiente de correlação Intraclasse. Pesquisa de Condições de Vida – Vale do Paraíba. 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	roh	icc	Erro Padrão icc	Intervalo de Confiança (95%)		
<i>Variáveis Contínuas</i>										
X4. Renda familiar per capita (Em reais)	1.163	88	13,22	419,57	0,2124	0,2509	0,0295	0,1940	—	0,3078
X2. Renda domiciliar per capita (Em reais)	1.119	88	12,72	427,79	0,2213	0,2494	0,0393	0,1723	—	0,3265
X1. Renda domiciliar total (Em reais)	1.119	88	12,72	1274,28	0,1725	0,1955	0,0351	0,1268	—	0,2642
X3. Renda familiar total (Em reais)	1.163	88	13,22	1249,51	0,1633	0,1955	0,0346	0,1373	—	0,2537
X5. Anos de estudo	4.759	88	54,08	7,05	0,0738	0,0501	0,0234	0,0042	—	0,0960
X6. Qualidade do atendimento recebido	1.168	88	13,27	8,30	0,0398	0,0246	0,0153	0,0000	—	0,0545
<i>Variáveis Dicotômicas</i>										
X16. Chefes com ensino superior completo	1.388	88	15,77	0,11	0,1547	0,1839	0,0292	0,1267	—	0,2411
X11. Existência de telefone fixo no domicílio	1.459	88	16,58	0,40	0,2546	0,1755	0,0301	0,1164	—	0,2346
X7. Domicílios com renda até 10 SM	1.119	88	12,72	0,67	0,1894	0,1470	0,0304	0,0873	—	0,2067
X9. Famílias com renda até 10 SM	1.163	88	13,22	0,67	0,1768	0,1445	0,0298	0,0862	—	0,2028
X12. Chefes com no máximo 4 anos de escolaridade	1.388	88	15,77	0,22	0,0855	0,0737	0,0193	0,0359	—	0,1115
X15. Chefes com ensino médio completo	1.388	88	15,77	0,18	0,0944	0,0591	0,0174	0,0250	—	0,0932
X23. Uso de óculos ou lentes de contato	5.327	88	60,53	0,30	0,0798	0,0511	0,0105	0,0306	—	0,0716
X8. Domicílios com renda até 3 SM	1.119	88	12,72	0,21	0,0794	0,0415	0,0180	0,0062	—	0,0768
X10. Famílias com renda até 3 SM	1.163	88	13,22	0,22	0,0703	0,0381	0,0171	0,0046	—	0,0716
X13. Chefes com 5 a 7 anos de escolaridade	1.388	88	15,77	0,33	0,0410	0,0288	0,0134	0,0026	—	0,0550
X17. Desempregados	2.537	88	28,83	0,18	0,0249	0,0242	0,0089	0,0067	—	0,0417
X19. Assalariados com carteira ou do setor público	1.527	88	17,35	0,77	0,0389	0,0220	0,0121	0,0000	—	0,0458
X22. Portador de dificuldades	5.387	88	61,22	0,05	0,0144	0,0195	0,0056	0,0086	—	0,0304
X20. Utilização de serviços de saúde	5.483	88	62,31	0,23	0,0212	0,0159	0,0049	0,0062	—	0,0256
X14. Chefes com ensino fundamental completo	1.388	88	15,77	0,17	0,0110	0,0142	0,0114	0,0000	—	0,0365
X24. Vítimas de roubos ou assaltos	5.482	88	62,3	0,06	0,0128	0,0130	0,0045	0,0042	—	0,0218
X18. População economicamente ativa	4.524	88	51,41	0,56	0,0046	0,0055	0,0038	0,0000	—	0,0130
X21. Portador de deficiência física ou mental*	5.482	88	62,3	0,02	-	-	-	-	—	-

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.  
Estimativas negativas do coeficiente de correlação intraclasse.

## 5.2 Estimação do Efeito de Delineamento

As estimativas do efeito do delineamento estão apresentadas nas tabelas 11 a 17. Estas tabelas mostram que, para o conjunto das variáveis estudadas, a perda de precisão em relação a uma amostra aleatória simples, ocorre segundo um mesmo padrão em todos os sete domínios da pesquisa.

Como o esperado as variáveis socioeconômicas, especificamente – *Renda domiciliar total*, *Renda domiciliar per capita*, *Renda familiar total*, *Renda familiar per capita* – apresentam os maiores efeitos de delineamento. Para essas variáveis os *deff's* nunca foram inferiores a dois.

Os valores do *deff* para as variáveis de rendimentos na Região Metropolitana de São Paulo, são inferiores ou iguais aos observados nos demais domínios, com exceção do Aglomerado Oeste. Isto ocorre apesar dos valores grandes para o coeficiente de correlação intraclasse destas variáveis de rendimentos (Tabela 4). Este fato pode ser explicado pelo maior número de setores censitários na amostra da Região Metropolitana de São Paulo e o pequeno tamanho das subamostras ( $\bar{b}$  em torno de 6).

Para a maioria das variáveis consideradas, o Aglomerado Oeste apresenta os menores valores para o efeito de delineamento, porém esse domínio possui a terceira maior amostra de setores censitários (115), sendo menor apenas que a RMSP (508) e Região Metropolitana da Baixada Santista (120).

A perda de precisão associada a variável *Utilização de serviços de Saúde* é da ordem de 2 na maioria dos domínios. Os maiores valores são observados para o

Aglomerado Central, Região Metropolitana da Baixada Santista e Aglomerado Norte.

**Tabela 11** – Estimativas do efeito de delineamento. Pesquisa de Condições de Vida – Região Metropolitana de São Paulo. 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	deff
<i>Variáveis Contínuas</i>					
1. Renda domiciliar total (Em reais)	2.804	499	5,62	1.506,21	2,45
2. Renda domiciliar per capita (Em reais)	2.804	499	5,62	518,29	2,63
3. Renda familiar total (Em reais)	2.923	499	5,86	1.486,01	2,47
4. Renda familiar per capita (Em reais)	2.923	499	5,86	510,74	2,66
5. Anos de estudo	11.315	508	22,27	6,89	3,73
6. Qualidade do atendimento recebido	3.138	487	6,44	8,06	1,81
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	2.804	499	5,62	0,66	2,37
8. Domicílios com renda até 3 SM	2.804	499	5,62	0,21	1,45
9. Famílias com renda até 10 SM	2.923	499	5,86	0,67	3,33
10. Famílias com renda até 3 SM	2.923	499	5,86	0,21	1,11
11. Existência de telefone fixo no domicílio	3.400	508	6,69	0,45	3,44
12. Chefes com menos de 4 anos de escolaridade	3.401	508	6,69	0,24	1,44
13. Chefes com 5 a 7 anos de escolaridade	3.401	508	6,69	0,35	1,51
14. Chefes com ensino fundamental completo	3.401	508	6,69	0,14	1,08
15. Chefes com ensino médio completo	3.401	508	6,69	0,16	1,44
16. Chefes com ensino superior completo	3.401	508	6,69	0,12	3,00
17. Desempregados	6.704	508	13,20	0,18	1,53
18. População economicamente ativa	10.786	508	21,23	0,62	1,15
19. Assalariados com carteira ou do setor público	3.910	508	7,70	0,71	1,22
20. Utilização de serviços de saúde	13.000	508	25,59	0,25	1,83
21. Portador de deficiência física ou mental	13.002	508	25,59	0,01	1,06
22. Portador de dificuldades	12.833	508	25,26	0,05	1,45
23. Uso de óculos ou lentes de contato	12.690	508	24,98	0,29	4,16
24. Vítimas de roubos ou assaltos	13.002	508	25,59	0,06	1,64

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

**Tabela 12** – Estimativas do efeito de delineamento. Pesquisa de Condições de Vida – Aglomerado Central. 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	deff
<i>Variáveis Contínuas</i>					
1. Renda domiciliar total (Em reais)	1.217	91	13,37	1.232,02	3,19
2. Renda domiciliar per capita (Em reais)	1.217	91	13,37	415,95	3,26
3. Renda familiar total (Em reais)	1.280	91	14,07	1.205,42	3,35
4. Renda familiar per capita (Em reais)	1.280	91	14,07	406,86	3,41
5. Anos de estudo	4.640	91	50,99	7,04	3,58
6. Qualidade do atendimento recebido	1.091	91	11,99	8,22	1,59
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.217	91	13,37	0,69	2,92
8. Domicílios com renda até 3 SM	1.217	91	13,37	0,20	2,42
9. Famílias com renda até 10 SM	1.280	91	14,07	0,70	3,10
10. Famílias com renda até 3 SM	1.280	91	14,07	0,21	2,47
11. Existência de telefone fixo no domicílio	1.486	91	16,33	0,44	5,81
12. Chefes com menos de 4 anos de escolaridade	1.418	91	15,58	0,25	2,36
13. Chefes com 5 a 7 anos de escolaridade	1.418	91	15,58	0,35	2,11
14. Chefes com ensino fundamental completo	1.418	91	15,58	0,13	1,46
15. Chefes com ensino médio completo	1.418	91	15,58	0,15	2,45
16. Chefes com ensino superior completo	1.418	91	15,58	0,11	3,58
17. Desempregados	2.516	91	27,65	0,14	1,59
18. População economicamente ativa	4.443	91	48,82	0,57	1,70
19. Assalariados com carteira ou do setor público	1.532	91	16,84	0,75	1,43
20. Utilização de serviços de saúde	5.318	91	58,44	0,22	2,59
21. Portador de deficiência física ou mental	5.318	91	58,44	0,01	1,31
22. Portador de dificuldades	5.247	91	57,66	0,05	1,75
23. Uso de óculos ou lentes de contato	5.202	91	57,16	0,27	4,99
24. Vítimas de roubos ou assaltos	5.318	91	58,44	0,04	1,65

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

**Tabela 13** – Estimativas do efeito de delineamento. Pesquisa de Condições de Vida – Aglomerado Leste. 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	deff
<i>Variáveis Contínuas</i>					
1. Renda domiciliar total (Em reais)	1.039	106	9,80	1.364,16	2,73
2. Renda domiciliar per capita (Em reais)	1.039	106	9,80	465,40	3,57
3. Renda familiar total (Em reais)	1.100	106	10,38	1.329,29	2,86
4. Renda familiar per capita (Em reais)	1.100	106	10,38	453,77	3,68
5. Anos de estudo	4.534	106	42,77	6,91	2,58
6. Qualidade do atendimento recebido	1.239	106	11,69	8,47	1,44
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.039	106	9,80	0,65	2,24
8. Domicílios com renda até 3 SM	1.039	106	9,80	0,19	1,34
9. Famílias com renda até 10 SM	1.100	106	10,38	0,66	2,39
10. Famílias com renda até 3 SM	1.100	106	10,38	0,20	1,43
11. Existência de telefone fixo no domicílio	1.421	106	13,41	0,50	4,41
12. Chefes com menos de 4 anos de escolaridade	1.342	106	12,66	0,25	2,20
13. Chefes com 5 a 7 anos de escolaridade	1.342	106	12,66	0,36	1,61
14. Chefes com ensino fundamental completo	1.342	106	12,66	0,13	1,41
15. Chefes com ensino médio completo	1.342	106	12,66	0,14	1,55
16. Chefes com ensino superior completo	1.342	106	12,66	0,12	3,39
17. Desempregados	2.668	106	25,17	0,15	2,22
18. População economicamente ativa	4.343	106	40,97	0,61	1,08
19. Assalariados com carteira ou do setor público	1.597	106	15,07	0,71	2,11
20. Utilização de serviços de saúde	5.159	106	48,67	0,25	2,08
21. Portador de deficiência física ou mental	5.159	106	48,67	0,02	1,17
22. Portador de dificuldades	5.052	106	47,66	0,05	2,13
23. Uso de óculos ou lentes de contato	4.990	106	47,08	0,30	4,81
24. Vítimas de roubos ou assaltos	5.159	106	48,67	0,05	1,92

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

**Tabela 14** – Estimativas do efeito de delineamento. Pesquisa de Condições de Vida – Região Metropolitana da Baixada Santista. 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	deff
<i>Variáveis Contínuas</i>					
1. Renda domiciliar total (Em reais)	1.032	120	8,60	1.407,87	2,66
2. Renda domiciliar per capita (Em reais)	1.032	120	8,60	483,94	2,79
3. Renda familiar total (Em reais)	1.096	120	9,13	1.387,22	2,85
4. Renda familiar per capita (Em reais)	1.096	120	9,13	476,73	2,98
5. Anos de estudo	4.235	120	35,29	7,26	2,74
6. Qualidade do atendimento recebido	1.072	120	8,93	8,11	1,80
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.032	120	8,60	0,64	2,56
8. Domicílios com renda até 3 SM	1.032	120	8,60	0,19	1,56
9. Famílias com renda até 10 SM	1.096	120	9,13	0,65	2,62
10. Famílias com renda até 3 SM	1.096	120	9,13	0,20	1,59
11. Existência de telefone fixo no domicílio	1.336	120	11,13	0,45	3,96
12. Chefes com menos de 4 anos de escolaridade	1.283	120	10,69	0,26	1,92
13. Chefes com 5 a 7 anos de escolaridade	1.283	120	10,69	0,33	1,52
14. Chefes com ensino fundamental completo*	1.283	120	10,69	0,14	0,97
15. Chefes com ensino médio completo	1.283	120	10,69	0,18	1,75
16. Chefes com ensino superior completo	1.283	120	10,69	0,09	2,65
17. Desempregados	2.259	120	18,83	0,18	1,20
18. População economicamente ativa	4.044	120	33,70	0,56	1,62
19. Assalariados com carteira ou do setor público	1.268	120	10,57	0,73	1,82
20. Utilização de serviços de saúde	4.903	120	40,86	0,23	2,68
21. Portador de deficiência física ou mental	4.903	120	40,86	0,02	1,19
22. Portador de dificuldades	4.819	120	40,16	0,06	1,62
23. Uso de óculos ou lentes de contato	4.747	120	39,56	0,31	4,32
24. Vítimas de roubos ou assaltos	4.903	120	40,86	0,07	1,56

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

Estimativas negativas do coeficiente de correlação intraclasse, impossibilitando o cálculo de  $deff_{ANOVA}$ .

**Tabela 15 – Estimativas do efeito de. Pesquisa de Condições de Vida – Aglomerado Norte. 1998.**

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	deff
<i>Variáveis Contínuas</i>					
1. Renda domiciliar total (Em reais)	1.055	110	9,59	1251,21	2,46
2. Renda domiciliar per capita (Em reais)	1.055	110	9,59	409,36	2,60
3. Renda familiar total (Em reais)	1.139	110	10,35	1.211,54	2,56
4. Renda familiar per capita (Em reais)	1.139	110	10,35	396,72	2,71
5. Anos de estudo	4.404	110	40,04	6,99	2,57
6. Qualidade do atendimento recebido	1.255	110	11,41	8,49	1,85
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.055	110	9,59	0,68	2,04
8. Domicílios com renda até 3 SM	1.055	110	9,59	0,20	1,31
9. Famílias com renda até 10 SM	1.139	110	10,35	0,70	2,10
10. Famílias com renda até 3 SM	1.139	110	10,35	0,21	1,39
11. Existência de telefone fixo no domicílio	1.364	110	12,40	0,52	4,40
12. Chefes com menos de 4 anos de escolaridade	1.317	110	11,97	0,27	1,55
13. Chefes com 5 a 7 anos de escolaridade	1.317	110	11,97	0,37	1,47
14. Chefes com ensino fundamental completo	1.317	110	11,97	0,13	1,35
15. Chefes com ensino médio completo	1.317	110	11,97	0,14	1,53
16. Chefes com ensino superior completo	1.317	110	11,97	0,10	2,73
17. Desempregados	2.488	110	22,62	0,16	2,21
18. População economicamente ativa	4.231	110	38,46	0,59	1,24
19. Assalariados com carteira ou do setor público	1.461	110	13,28	0,77	1,82
20. Utilização de serviços de saúde	5.005	110	45,50	0,27	2,65
21. Portador de deficiência física ou mental*	5.005	110	45,50	0,01	0,90
22. Portador de dificuldades	4.931	110	44,83	0,06	1,71
23. Uso de óculos ou lentes de contato	4.851	110	44,10	0,31	4,20
24. Vítimas de roubos ou assaltos	5.005	110	45,50	0,04	1,60

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

Estimativas negativas do coeficiente de correlação intraclasse, impossibilitando o cálculo de deff<sub>ic</sub>.



**Tabela 16** – Estimativas do efeito de delineamento. Pesquisa de Condições de Vida – Aglomerado Oeste. 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	deff
<i>Variáveis Contínuas</i>					
1. Renda domiciliar total (Em reais)	1.162	115	10,10	1.141,29	2,14
2. Renda domiciliar per capita (Em reais)	1.162	115	10,10	395,54	2,14
3. Renda familiar total (Em reais)	1.238	115	10,77	1.106,74	2,22
4. Renda familiar per capita (Em reais)	1.238	115	10,77	383,52	2,21
5. Anos de estudo	4.406	115	38,31	7,39	1,91
6. Qualidade do atendimento recebido	1.123	115	9,77	8,38	2,32
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.162	115	10,10	0,71	2,04
8. Domicílios com renda até 3 SM	1.162	115	10,10	0,24	1,30
9. Famílias com renda até 10 SM	1.238	115	10,77	0,72	2,16
10. Famílias com renda até 3 SM	1.238	115	10,77	0,25	1,33
11. Existência de telefone fixo no domicílio	1.447	115	12,58	0,45	3,17
12. Chefes com menos de 4 anos de escolaridade	1.381	115	12,01	0,29	1,57
13. Chefes com 5 a 7 anos de escolaridade	1.381	115	12,01	0,32	1,52
14. Chefes com ensino fundamental completo	1.381	115	12,01	0,10	1,09
15. Chefes com ensino médio completo	1.381	115	12,01	0,17	1,41
16. Chefes com ensino superior completo	1.381	115	12,01	0,13	3,15
17. Desempregados	2.468	115	21,46	0,16	1,45
18. População economicamente ativa	4.215	115	36,65	0,59	2,10
19. Assalariados com carteira ou do setor público	1.418	115	12,33	0,72	1,66
20. Utilização de serviços de saúde	4.966	115	43,18	0,24	1,39
21. Portador de deficiência física ou mental	4.967	115	43,19	0,02	1,72
22. Portador de dificuldades	4.859	115	42,25	0,05	1,31
23. Uso de óculos ou lentes de contato	4.819	115	41,90	0,32	3,06
24. Vítimas de roubos ou assaltos	4.967	115	43,19	0,04	1,52

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

**Tabela 17** – Estimativas do efeito de delineamento. Pesquisa de Condições de Vida – Vale do Paraíba, 1998.

Variáveis	Número de casos analisados	Setores censitários	$\bar{b}$	$r^*$	deff
<i>Variáveis Contínuas</i>					
1. Renda domiciliar total (Em reais)	1.119	88	12,72	1.274,28	3,02
2. Renda domiciliar per capita (Em reais)	1.119	88	12,72	427,79	3,59
3. Renda familiar total (Em reais)	1.163	88	13,22	1.249,51	2,99
4. Renda familiar per capita (Em reais)	1.163	88	13,22	419,57	3,59
5. Anos de estudo	4.759	88	54,08	7,05	4,92
6. Qualidade do atendimento recebido	1.168	88	13,27	8,30	1,49
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.119	88	12,72	0,67	3,22
8. Domicílios com renda até 3 SM	1.119	88	12,72	0,21	1,93
9. Famílias com renda até 10 SM	1.163	88	13,22	0,67	3,16
10. Famílias com renda até 3 SM	1.163	88	13,22	0,22	1,86
11. Existência de telefone fixo no domicílio	1.459	88	16,58	0,40	4,97
12. Chefes com menos de 4 anos de escolaridade	1.388	88	15,77	0,22	2,26
13. Chefes com 5 a 7 anos de escolaridade	1.388	88	15,77	0,33	1,61
14. Chefes com ensino fundamental completo	1.388	88	15,77	0,17	1,16
15. Chefes com ensino médio completo	1.388	88	15,77	0,18	2,39
16. Chefes com ensino superior completo	1.388	88	15,77	0,11	3,29
17. Desempregados	2.537	88	28,83	0,18	1,69
18. População economicamente ativa	4.524	88	51,41	0,56	1,23
19. Assalariados com carteira ou do setor público	1.527	88	17,35	0,77	1,66
20. Utilização de serviços de saúde	5.483	88	62,31	0,23	2,30
21. Portador de deficiência física ou mental*	5.482	88	62,30	0,02	0,85
22. Portador de dificuldades	5.387	88	61,22	0,05	1,87
23. Uso de óculos ou lentes de contato	5.327	88	60,53	0,30	5,75
24. Vítimas de roubos ou assaltos	5.482	88	62,30	0,06	1,78

\* Para as variáveis contínuas  $r$  refere-se a média, para as variáveis dicotômicas  $r$  refere-se a proporção.

Estimativas negativas do coeficiente de correlação intraclasse, impossibilitando o cálculo de  $deff_{ANOVA}$ .

### 5.3 Estimação dos Tamanhos das Amostras

As Tabelas 18 a 23 mostram os tamanhos previstos para número de setores censitários (a) e para a subamostra ( $b_{\text{ótimo}}^{\text{icc}}$ ), considerando-se a relação entre coeficiente de correlação intraclasse e o custo associado a amostragem.

Entre os resultados apresentados, destaca-se o fato de que, para a Região Metropolitana de São Paulo, o valor médio observado para a subamostra ( $\bar{b}$ ) é inferior ao valor ótimo para quase todas as variáveis estudadas.

Para os demais domínios observa-se que, dado os tamanhos das amostras da pesquisa (n), o tamanho observado para  $\bar{b}$  é superior aos valores ótimos, quando se considera as variáveis de rendimentos.

Assim, nesses domínios o aumento da precisão da amostra pode ocorrer através do aumento do número de setores censitários selecionados, podendo-se obter amostras mais precisas sem aumentar o número total de domicílios na amostra.

**Tabela 17** – Estimativas do número ótimo de elementos a serem amostrados nos setores censitários, fixando-se o tamanho da amostra. Pesquisa de Condições de Vida – Região Metropolitana de São Paulo. 1998.

Variáveis	Tamanho da amostra (n)	Número de setores censitários na amostra	$\bar{b}$	$a = n / b^{\text{icc ótimo}}$	$b^{\text{icc ótimo}}$
<i>Variáveis Contínuas</i>					
1. Renda domiciliar total (Em reais)	2.804	499	5,62	387	7,25
2. Renda domiciliar per capita (Em reais)	2.804	499	5,62	468	5,99
3. Renda familiar total (Em reais)	2.923	499	5,86	403	7,25
4. Renda familiar per capita (Em reais)	2.923	499	5,86	484	6,04
5. Anos de estudo	11.315	508	22,27	836	13,53
6. Qualidade do atendimento recebido	3.138	487	6,44	159	19,68
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	2.804	499	5,62	333	8,43
8. Domicílios com renda até 3 SM	2.804	499	5,62	111	25,21
9. Famílias com renda até 10 SM	2.923	499	5,86	342	8,55
10. Famílias com renda até 3 SM	2.923	499	5,86	120	24,4
11. Existência de telefone fixo no domicílio	3.400	508	6,69	449	7,58
12. Chefes com menos de 4 anos de escolaridade	3.401	508	6,69	136	25
13. Chefes com 5 a 7 anos de escolaridade	3.401	508	6,69	178	19,09
14. Chefes com ensino fundamental completo	3.401	508	6,69	64	52,75
15. Chefes com ensino médio completo	3.401	508	6,69	199	17,07
16. Chefes com ensino superior completo	3.401	508	6,69	496	6,85
17. Desempregados	6.704	508	13,20	227	29,5
18. População economicamente ativa	10.786	508	21,23	144	74,86
19. Assalariados com carteira ou do setor público	3.910	508	7,70	146	26,76
20. Utilização de serviços de saúde	13.000	508	25,59	514	25,28
21. Portador de deficiência física ou mental	13.002	508	25,59	165	78,81
22. Portador de dificuldades	12.833	508	25,26	325	39,51
23. Uso de óculos ou lentes de contato	12.690	508	24,98	943	13,45
24. Vítimas de roubos ou assaltos	13.002	508	25,59	503	25,87

**Tabela 18** – Estimativas do número ótimo de elementos a serem amostrados nos setores censitários, fixando-se o tamanho da amostra, segundo domínios amostrais. Pesquisa de Condições de Vida – Aglomerado Central. 1998.

Variáveis	Tamanho da (n)	Número de setores censitários na amostra	$\bar{b}$	$a = n / b^{\text{icc}}_{\text{ótimo}}$	$b^{\text{icc}}_{\text{ótimo}}$
<i>Variáveis Contínuas</i>					
1. Renda domiciliar total (Em reais)	1.217	91	13,37	130	9,35
2. Renda domiciliar per capita (Em reais)	1.217	91	13,37	133	9,15
3. Renda familiar total (Em reais)	1.280	91	14,07	140	9,17
4. Renda familiar per capita (Em reais)	1.280	91	14,07	142	9,04
5. Anos de estudo	4.640	91	50,99	229	20,25
6. Qualidade do atendimento recebido	1.091	91	11,99	57	19,09
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.217	91	13,37	116	10,51
8. Domicílios com renda até 3 SM	1.217	91	13,37	92	13,25
9. Famílias com renda até 10 SM	1.280	91	14,07	123	10,38
10. Famílias com renda até 3 SM	1.280	91	14,07	95	13,48
11. Existência de telefone fixo no domicílio	1.486	91	16,33	200	7,43
12. Chefes com menos de 4 anos de escolaridade	1.418	91	15,58	100	14,21
13. Chefes com 5 a 7 anos de escolaridade	1.418	91	15,58	82	17,38
14. Chefes com ensino fundamental completo	1.418	91	15,58	58	24,24
15. Chefes com ensino médio completo	1.418	91	15,58	96	14,81
16. Chefes com ensino superior completo	1.418	91	15,58	149	9,54
17. Desempregados	2.516	91	27,65	83	30,35
18. População economicamente ativa	4.443	91	48,82	116	38,23
19. Assalariados com carteira ou do setor público	1.532	91	16,84	54	28,12
20. Utilização de serviços de saúde	5.318	91	58,44	174	30,62
21. Portador de deficiência física ou mental	5.318	91	58,44	61	87,64
22. Portador de dificuldades	5.247	91	57,66	157	33,39
23. Uso de óculos ou lentes de contato	5.202	91	57,16	334	15,56
24. Vítimas de roubos ou assaltos	5.318	91	58,44	146	36,47

**Tabela 19** – Estimativas do número ótimo de elementos a serem amostrados nos setores censitários, fixando-se o tamanho da amostra, segundo domínios amostrais. Pesquisa de Condições de Vida – Aglomerado Leste. 1998.

Variáveis	Tamanho da (n)	Número de setores censitários na amostra	$\bar{b}$	$a = n / b_{\text{ótimo}}^{\text{icc}}$	$b_{\text{ótimo}}^{\text{icc}}$
<i>Variáveis Quantitativas</i>					
1. Renda domiciliar total	1.039	106	9,8	139	7,45
2. Renda domiciliar per capita	1.039	106	9,8	184	5,65
3. Renda familiar total	1.100	106	10,38	147	7,49
4. Renda familiar per capita	1.100	106	10,38	187	5,87
5. Anos de estudo	4.534	106	42,77	186	24,32
6. Qualidade do atendimento recebido	1.239	106	11,69	65	19,14
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.039	106	9,8	93	11,14
8. Domicílios com renda até 3 SM	1.039	106	9,8	45	23,09
9. Famílias com renda até 10 SM	1.100	106	10,38	97	11,34
10. Famílias com renda até 3 SM	1.100	106	10,38	48	23,06
11. Existência de telefone fixo no domicílio	1.421	106	13,41	172	8,26
12. Chefes com menos de 4 anos de escolaridade	1.342	106	12,66	90	14,94
13. Chefes com 5 a 7 anos de escolaridade	1.342	106	12,66	97	13,89
14. Chefes com ensino fundamental completo	1.342	106	12,66	46	29,49
15. Chefes com ensino médio completo	1.342	106	12,66	63	21,47
16. Chefes com ensino superior completo	1.342	106	12,66	171	7,85
17. Desempregados	2.668	106	25,17	101	26,35
18. População economicamente ativa	4.343	106	40,97	62	70,57
19. Assalariados com carteira ou do setor público	1.597	106	15,07	87	18,28
20. Utilização de serviços de saúde	5.159	106	48,67	169	30,54
21. Portador de deficiência física ou mental	5.159	106	48,67	48	107,74
22. Portador de dificuldades	5.052	106	47,66	146	34,54
23. Uso de óculos ou lentes de contato	4.990	106	47,08	321	15,56
24. Vítimas de roubos ou assaltos	5.159	106	48,67	170	30,39

**Tabela 20** – Estimativas do número ótimo de elementos a serem amostrados nos setores censitários, fixando-se o tamanho da amostra, segundo domínios amostrais. Pesquisa de Condições de Vida – Região Metropolitana da Baixada Santista. 1998.

Variáveis	Tamanho da amostra (n)	Número de setores censitários na amostra	$\bar{b}$	$a = n / b_{\text{icc}}^{\text{ótimo}}$	$b_{\text{icc}}^{\text{ótimo}}$
<i>Variáveis Quantitativas</i>					
1. Renda domiciliar total	1.032	120	8,60	126	8,18
2. Renda domiciliar per capita	1.032	120	8,60	130	7,95
3. Renda familiar total	1.096	120	9,13	138	7,92
4. Renda familiar per capita	1.096	120	9,13	143	7,69
5. Anos de estudo	4.235	120	35,29	192	22,11
6. Qualidade do atendimento recebido	1.072	120	8,93	56	19,23
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.032	120	8,60	113	9,14
8. Domicílios com renda até 3 SM	1.032	120	8,60	53	19,51
9. Famílias com renda até 10 SM	1.096	120	9,13	120	9,13
10. Famílias com renda até 3 SM	1.096	120	9,13	56	19,50
11. Existência de telefone fixo no domicílio	1.336	120	11,13	170	7,85
12. Chefes com menos de 4 anos de escolaridade	1.283	120	10,69	87	14,72
13. Chefes com 5 a 7 anos de escolaridade	1.283	120	10,69	61	21,16
14. Chefes com ensino fundamental completo *	1.283	120	10,69	-	-
15. Chefes com ensino médio completo	1.283	120	10,69	81	15,78
16. Chefes com ensino superior completo	1.283	120	10,69	132	9,74
17. Desempregados	2.259	120	18,83	71	31,75
18. População economicamente ativa	4.044	120	33,70	124	32,70
19. Assalariados com carteira ou do setor público	1.268	120	10,57	45	28,12
20. Utilização de serviços de saúde	4.903	120	40,86	204	24,05
21. Portador de deficiência física ou mental	4.903	120	40,86	97	50,42
22. Portador de dificuldades	4.819	120	40,16	151	31,86
23. Uso de óculos ou lentes de contato	4.747	120	39,56	297	16,00
24. Vítimas de roubos ou assaltos	4.903	120	40,86	153	32,13

\* Estimativas negativas do coeficiente de correlação intraclasses, impossibilitando o cálculo de  $b_{\text{icc}}^{\text{ótimo}}$ .

**Tabela 21** – Estimativas do número ótimo de elementos a serem amostrados nos setores censitários, fixando-se o tamanho da amostra, segundo domínios amostrais. Pesquisa de Condições de Vida – Aglomerado Norte. 1998.

Variáveis	Tamanho da (n)	Número de setores censitários na amostra	$\bar{b}$	$a = n / b_{\text{ótimo}}^{\text{icc}}$	$b_{\text{ótimo}}^{\text{icc}}$
<i>Variáveis Quantitativas</i>					
1. Renda domiciliar total	1.055	110	9,59	136	7,73
2. Renda domiciliar per capita	1.055	110	9,59	148	7,11
3. Renda familiar total	1.139	110	10,35	147	7,75
4. Renda familiar per capita	1.139	110	10,35	160	7,14
5. Anos de estudo	4.404	110	40,04	188	23,48
6. Qualidade do atendimento recebido	1.255	110	11,41	75	16,84
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.055	110	9,59	87	12,14
8. Domicílios com renda até 3 SM	1.055	110	9,59	35	30,14
9. Famílias com renda até 10 SM	1.139	110	10,35	93	12,21
10. Famílias com renda até 3 SM	1.139	110	10,35	41	27,54
11. Existência de telefone fixo no domicílio	1.364	110	12,40	164	8,31
12. Chefes com menos de 4 anos de escolaridade	1.317	110	11,97	57	23,15
13. Chefes com 5 a 7 anos de escolaridade	1.317	110	11,97	57	23,00
14. Chefes com ensino fundamental completo	1.317	110	11,97	47	27,73
15. Chefes com ensino médio completo	1.317	110	11,97	59	22,34
16. Chefes com ensino superior completo	1.317	110	11,97	147	8,94
17. Desempregados	2.488	110	22,62	118	21,03
18. População economicamente ativa	4.231	110	38,46	104	40,83
19. Assalariados com carteira ou do setor público	1.461	110	13,28	79	18,40
20. Utilização de serviços de saúde	5.005	110	45,50	177	28,20
21. Portador de deficiência física ou mental *	5.005	110	45,50	-	-
22. Portador de dificuldades	4.931	110	44,83	124	39,81
23. Uso de óculos ou lentes de contato	4.851	110	44,10	280	17,35
24. Vítimas de roubos ou assaltos	5.005	110	45,50	157	31,81

\* Estimativas negativas do coeficiente de correlação intraclasse, impossibilitando o cálculo de  $b_{\text{ótimo}}^{\text{icc}}$ .



**Tabela 22** – Estimativas do número ótimo de elementos a serem amostrados nos setores censitários, fixando-se o tamanho da amostra, segundo domínios amostrais. Pesquisa de Condições de Vida – Aglomerado Oeste. 1998.

Variáveis	Tamanho da (n)	Número de setores censitários na amostra	$\bar{b}$	$a = n / b^{\text{icc}}_{\text{ótimo}}$	$b^{\text{icc}}_{\text{ótimo}}$
<i>Variáveis Quantitativas</i>					
1. Renda domiciliar total	1.162	115	10,10	114	10,21
2. Renda domiciliar per capita	1.162	115	10,10	119	9,76
3. Renda familiar total	1.238	115	10,77	121	10,26
4. Renda familiar per capita	1.238	115	10,77	126	9,81
5. Anos de estudo	4.406	115	38,31	153	28,76
6. Qualidade do atendimento recebido	1.123	115	9,77	80	14,02
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.162	115	10,10	91	12,73
8. Domicílios com renda até 3 SM	1.162	115	10,10	30	38,62
9. Famílias com renda até 10 SM	1.238	115	10,77	97	12,79
10. Famílias com renda até 3 SM	1.238	115	10,77	32	39,24
11. Existência de telefone fixo no domicílio	1.447	115	12,58	141	10,25
12. Chefes com menos de 4 anos de escolaridade	1.381	115	12,01	66	20,98
13. Chefes com 5 a 7 anos de escolaridade	1.381	115	12,01	67	20,64
14. Chefes com ensino fundamental completo	1.381	115	12,01	31	44,24
15. Chefes com ensino médio completo	1.381	115	12,01	49	28,18
16. Chefes com ensino superior completo	1.381	115	12,01	151	9,12
17. Desempregados	2.468	115	21,46	59	41,83
18. População economicamente ativa	4.215	115	36,65	155	27,15
19. Assalariados com carteira ou do setor público	1.418	115	12,33	71	19,95
20. Utilização de serviços de saúde	4.966	115	43,18	110	45,14
21. Portador de deficiência física ou mental	4.967	115	43,19	120	41,28
22. Portador de dificuldades	4.859	115	42,25	83	58,44
23. Uso de óculos ou lentes de contato	4.819	115	41,90	217	22,17
24. Vítimas de roubos ou assaltos	4.967	115	43,19	102	48,58

**Tabela 23** – Estimativas do número ótimo de elementos a serem amostrados nos setores censitários, fixando-se o tamanho da amostra, segundo domínios amostrais. Pesquisa de Condições de Vida – Vale do Paraíba. 1998.

Variáveis	Tamanho da (n)	Número de setores censitários na amostra	$\bar{b}$	$a = n / b_{\text{ótimo}}^{\text{icc}}$	$b_{\text{ótimo}}^{\text{icc}}$
<i>Variáveis Quantitativas</i>					
1. Renda domiciliar total	1.119	88	12,72	123	9,07
2. Renda domiciliar per capita	1.119	88	12,72	144	7,76
3. Renda familiar total	1.163	88	13,22	113	10,26
4. Renda familiar per capita	1.163	88	13,22	150	7,73
5. Anos de estudo	4.759	88	54,08	244	19,48
6. Qualidade do atendimento recebido	1.168	88	13,27	41	28,16
<i>Variáveis Dicotômicas</i>					
7. Domicílios com renda até 10 SM	1.119	88	12,72	104	10,77
8. Domicílios com renda até 3 SM	1.119	88	12,72	52	21,49
9. Famílias com renda até 10 SM	1.163	88	13,22	107	10,88
10. Famílias com renda até 3 SM	1.163	88	13,22	52	22,48
11. Existência de telefone fixo no domicílio	1.459	88	16,58	151	9,69
12. Chefes com menos de 4 anos de escolaridade	1.388	88	15,77	88	15,86
13. Chefes com 5 a 7 anos de escolaridade	1.388	88	15,77	54	25,79
14. Chefes com ensino fundamental completo	1.388	88	15,77	43	31,97
15. Chefes com ensino médio completo	1.388	88	15,77	80	17,31
16. Chefes com ensino superior completo	1.388	88	15,77	149	9,29
17. Desempregados	2.537	88	28,83	89	28,37
18. População economicamente ativa	4.524	88	51,41	75	59,97
19. Assalariados com carteira ou do setor público	1.527	88	17,35	51	29,81
20. Utilização de serviços de saúde	5.483	88	62,31	156	35,20
21. Portador de deficiência física ou mental *	5.482	88	62,30	-	-
22. Portador de dificuldades	5.387	88	61,22	170	31,69
23. Uso de óculos ou lentes de contato	5.327	88	60,53	276	19,27
24. Vítimas de roubos ou assaltos	5.482	88	62,30	141	38,96

\* Estimativas negativas do coeficiente de correlação intraclasses, impossibilitando o cálculo de  $b_{\text{ótimo}}^{\text{icc}}$ .

## 6 DISCUSSÃO E CONCLUSÃO

Neste estudo, foram calculadas estimativas do coeficiente de correlação intraclasse, que são relevantes no planejamento de futuras amostras que utilizem o setor censitário como unidade primária de amostragem.

Foram utilizados dois métodos de estimação. No primeiro estimou-se o efeito de delineamento (*deff*) e através dele a correlação intraclasse (*roh*), e no segundo, estimou-se diretamente o coeficiente de correlação intraclasse (*icc*), através de componentes de variância.

A comparação entre esses dois métodos mostrou que as estimativas obtidas por meio de *roh* são compatíveis às obtidas, quando se utilizam componentes de variância (*icc*) para estimar a correlação intraclasse. Das 24 variáveis estudadas a maioria dos valores gerados por *roh* estava incluída no intervalo de confiança de 95% de *icc*.

A seu favor, *roh* tem a facilidade de cálculo, uma vez que o *deff* é usualmente apresentado pela maioria dos programas computacionais de cálculo de erros para planos de amostragem por conglomerados. Entre esses programas destacam-se Wes Var, STATA (STATISTICS DATA ANALYSIS), SAS, Epi Info (módulo CSAMPLE) entre outros, sendo o Epi Info de domínio público.

Porém, *roh* é afetado pelo tamanho do conglomerado. Assim, KISH (1965) aconselha o uso de *roh* para conglomerados de tamanhos desiguais, como no presente estudo, quando os tamanhos dos conglomerados não tiveram muita variabilidade.

Quanto ao *icc*, DONNER (1982) analisando dados para hipertensão, tabagismo, consumo de bebidas alcoólicas e obesidade observou valores mais altos para *icc* quando os conglomerados eram menores (por exemplo, domicílios ou casais) e valores menores de *icc* quando os conglomerados eram maiores (por exemplo, regiões). GUILLIFORD (1999) estudando *icc* para um grande número de variáveis constatou que os valores tendem a ser maiores para pequenos conglomerados, tais como domicílios.

As variáveis – *Portador de deficiência física e mental, Portador de dificuldades e Vítimas de roubos ou assaltos* – indicam a ocorrência de eventos “raros” na população, o que poderia gerar instabilidade nas estimativas de *icc*. Porém foram mantidas no estudo, já que foram obtidas estimativas compatíveis para os dois estimadores (SEARLE et al. 1992).

Em relação à Pesquisa de Condições de Vida (PCV) observou-se que para as variáveis selecionadas, a amostragem por conglomerados, utilizando o setor censitário como unidade primária de amostragem, se mostrou menos precisa do que a amostragem aleatória simples, para todos os domínios da amostra.

Este é o resultado esperado, uma vez que, KISH (1965) destaca que a variância de amostras por conglomerados, particularmente em pesquisas sociais, é geralmente maior do que uma amostra de elementos.

Para os dois métodos de estimação constatou-se que o coeficiente de correlação intraclasse pode variar substancialmente entre diferentes tipos de variáveis, quando se considera o setor censitário como unidade primária de amostragem. Destaca-se o fato, que em todos os domínios verificou-se esse mesmo padrão de comportamento, especialmente a homogeneidade dos setores censitários em relação às variáveis

socioeconômicas, especificamente rendimentos. Apesar disso, o Aglomerado Oeste e o Aglomerado Central apresentaram as menores estimativas para essas variáveis.

O achado de que o setor censitário é unidade homogênea com relação as variáveis de rendimentos foi descrita em estudo realizado pela Fundação IBGE (1990). Nesse trabalho, foram calculadas estimativas do coeficiente de correlação intraclasse e do efeito do delineamento utilizando-se os dados da amostra do Censo Demográfico de 1980, para as regiões metropolitanas do Brasil, considerando-se variáveis socioeconômicas.

A perda de precisão comporta-se de forma semelhante em todos os domínios. As variáveis socioeconômicas, especificamente os rendimentos, apresentam os maiores *deff*'s, nunca inferiores a dois.

A perda de precisão da amostra da Região Metropolitana de São Paulo, relativamente as variáveis de rendimentos, é inferior ou igual à observada para os demais domínios, apesar das maiores estimativas do coeficiente de correlação intraclasse na Região Metropolitana de São Paulo. Este fato pode estar relacionado ao grande tamanho da amostra dos setores censitários na RMSP.

Para os domínios do Interior do Estado as amostras da PCV poderiam ganhar em precisão se houvesse um aumento no número de setores censitários na amostra, sem aumento no total de domicílios pesquisados. Para a Região Metropolitana de São Paulo o aumento de precisão da amostra requer um aumento no número de domicílios investigados (n).

## 7 REFERÊNCIAS

1. Alves MCGP. Técnicas de replicação em análise de dados de inquéritos domiciliares. São Paulo; 2002 [Tese de Doutorado – Faculdade de Saúde Pública - USP].
2. An A, Watts D. SAS<sup>R</sup> Procedures for analysis of sample survey data. **Proceedings of Section on Survey Research Methods, American Statistical Association.** 2000; 120-129.
3. Bergamaschi DP. Correlação intraclasse de Pearson para pares repetidos - comparação entre dois estimadores. São Paulo; 1999 [Tese de Doutorado – Faculdade de Saúde Pública - USP].
4. Bussab, WO, Dini NP. Pesquisa de emprego e desemprego SEADE/DIEESE: regiões homogêneas da Grande São Paulo. **São Paulo em Perspectiva.** 1985; 1(3): 5-11.
5. Cochran WG. **Sampling Techniques.** 3<sup>rd</sup> ed. New York: John Wiley & Sons; 1977.
6. Ching TH. Coeficiente de correlação intraclasse: planejamento com alocação ótima e aplicação no estudo de confiabilidade de medidas. São Paulo; 1995 [Dissertação de Mestrado – Instituto de Matemática e Estatística - USP].

7. Cordeiro R. Efeito do desenho em amostragem de conglomerado para estimar a distribuição de ocupações entre trabalhadores. **Rev Saúde Pública**. 2001; 35(1): 10-15.
8. Donner A. An empirical study of cluster randomization. **Int J Epidemiol**. 1982; 11: 283-286.
9. Donner A. A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. **Int Stat Rev**. 1986; 54: 67-82.
10. Donner A, Wells G. A comparison of confidence interval methods for the intraclass correlation coefficient. **Biometrics**. 1986; 42: 401-412.
11. Ferrinho P, Valli T, Croeneveld T, Buch E, Coetzee D. The effects of cluster sampling in na African urban setting. **Central African Journal of Medicine**. 1992; 38(8): 324-330.
12. Fisher R. **Statistical methods for research workers**. 7<sup>rd</sup> ed. Londres: Oliver and Boyd. 1938; (Biological monographs and manuals, 5)
13. Fundação IBGE. **Efeito de Conglomeração da malha setorial do Censo Demográfico de 1980**. Rio de Janeiro; 1990 (Textos para discussão, 32).
14. Fundação IBGE. **GR-701 Manual de delimitação dos setores e das zonas de trabalho do 1990 – Base Operacional**. Rio de Janeiro; 1990.
15. Fundação IBGE. **CD-1.09 Manual do Recenseador**. Rio de Janeiro; 2000.

16. Fundação SEADE. **Pesquisa de Condições de Vida 1990: plano amostral**. São Paulo; 1990. [Documento para uso interno da Fundação SEADE].
17. Fundação SEADE. **Pesquisa de Condições de Vida 1994: plano amostral**. São Paulo; 1994. [Documento para uso interno da Fundação SEADE].
18. Fundação SEADE. **Pesquisa de Condições de Vida 1998: plano amostral**. São Paulo; 1998. [Documento para uso interno da Fundação SEADE].
19. Fundação SEADE. **Pesquisa de Condições de Vida 1998: primeiros resultados**. São Paulo: Fundação SEADE; 1999.
20. Gulliford MC, Ukoumunne OC; Chinn S. Components of Variance and Intraclass Correlations for the Design of Community-based Surveys and Intervention Studies. *American Journal of Epidemiology*. 1999; 149(9): 876-883.
21. UK Office for National Statistics. **Health Survey for England: the health of young people '95-97**. 1997. Available from <<http://www.official-documents.co.uk/document/doh/survey97>>.
22. Kalton G. **Introduction to survey sampling**. Newbury Park: Sage Publications; 1983. (Quantitative Applications in the Social Science, 35).
23. Kish L. **Survey sampling**. New York: John Wiley & Sons; 1965.
24. Kish L. Deffs: why, when and How? A review. **Proceedings of Section on Survey Research Methods, American Statistical Association**. 1989; 209-211.
25. Kish L. **Leslie Kish selected papers**. New York. John Wiley & Sons; 2003.



26. Levy PS, Lemeshow S. **Sampling populations: methods and applications**. 3<sup>rd</sup> ed. New York: John Wiley & Sons; 1999.
27. Särndal CE, Swenssom B, Wretman J. **Model assisted survey sampling**. New York: Springer-Verlag, 1992.
28. SAS Institute Inc. **SAS/STAT™ User's Guide**. 6.03 Edition. Cary, 1994.
29. SAS Institute Inc. **The system SAS for windows 8.0** [programa de computador]. Cary, 1999.
30. Searle SR. **Linear models**. New York: John Wiley & Sons; 1971.
31. Searle SR, Casella G, McCulloch CE. **Variance components**. New York: John Wiley & Sons; 1992.
32. Silva NN. **Amostragem Probabilística**. São Paulo: EDUSP, 1998.
33. STATA CORP. **Stata statistical software: release 7.0** [programa de computador]. College Station: Stata Corporation; 2001.

## ANEXO 1

## MUNICÍPIOS QUE COMPÕEM O UNIVERSO DA PCV

- **Aglomerado Central:** Avaré, Bauru, Botucatu, Itapetininga, Itapeva, Itu, Jaú, Salto, Sorocaba, Tatuí, Votorantim.
- **Aglomerado Norte:** Araraquara, Barretos, Bebedouro, Franca, Jaboticabal, Matão, Mococa, Ribeirão Preto, São Carlos, São João da Boa Vista, Sertãozinho.
- **Aglomerado Leste:** Americana, Araras, Atibaia, Bragança Paulista, Campinas, Campo Limpo Paulista, Hortolândia, Indaiatuba, Itapira, Itatiba, Jundiaí, Leme, Limeira, Mogi-Guaçu, Mogi-Mirim, Piracicaba, Pirassununga, Rio Claro, Santa Bárbara d'Oeste, Sumaré, Valinhos, Várzea Paulista.
- **Aglomerado Oeste:** Araçatuba, Assis, Birigüi, Catanduva, Fernandópolis, Lins, Marília, Ourinhos, Presidente Prudente, São José do Rio Preto, Tupã, Votuporanga.
- **Vale do Paraíba:** Caçapava, Cruzeiro, Guaratinguetá, Jacareí, Lorena, Pindamonhangaba, São José dos Campos, Taubaté.

- **RM da Baixada Santista:** Bertioga, Cubatão, Guarujá, Itanhaém, Mongaguá, Peruíbe, Praia Grande, Santos, São Vicente.
- **RM de São Paulo:** Arujá, Barueri, Biritiba-Mirim, Caieiras, Cajamar, Carapicuíba, Cotia, Diadema, Embu, Embu-Guaçu, Ferraz de Vasconcelos, Francisco Morato, Franco da Rocha, Guararema, Guarulhos, Itapecerica da Serra, Itapevi, Itaquaquecetuba, Jandira, Jquitituba, Mairiporã, Mauá, Mogi das Cruzes, Osasco, Pirapora do Bom Jesus, Poá, Ribeirão Pires, Rio Grande da Serra, Salesópolis, Santa Isabel, Santana de Parnaíba, Santo André, São Bernardo do Campo, São Caetano do Sul, São Paulo, Suzano, Taboão da Serra, Vargem Grande Paulista.

## ANEXO 2

### GLOSSÁRIO DE VARIÁVEIS

#### *Anos de Estudo – XE9802*

Indicam o grau de escolaridade dos indivíduos de sete anos e mais, tomando-se por base a última série de ensino regular concluída com aprovação, que varia entre zero e 17 anos de escolaridade. Para os indivíduos que concluíram o ensino supletivo de 1º ou 2º graus, foi atribuída escolaridade compatível com o ensino formal (8 e 11 anos de escolaridade, respectivamente).

#### *Adequação da Infra-Estrutura Urbana*

Distingue as moradias segundo a condição de adequação da infra-estrutura urbana no entorno do domicílio, tomando por base as ligações às redes públicas de água, lixo e esgotamento sanitário. Com exceção da fossa séptica (que tem tratamento adequado dos dejetos), as demais alternativas às redes públicas de abastecimento de água, de coleta do lixo e do esgoto foram consideradas impróprias, uma vez que a proximidade entre as moradias situadas em áreas urbanas não elimina o risco de contaminação dos reservatórios de água domésticos e daqueles destinados ao abastecimento coletivo. Mesmo um poço, por exemplo, pode ter a água contaminada pela fossa negra (ou rudimentar) de um domicílio vizinho:

**adequado:** quando a família vive em moradia onde existem as três ligações às redes públicas ou fossa séptica como alternativa à rede pública de esgotamento sanitário;

**inadequado:** quando a família vive em moradia onde se utiliza, regularmente, pelo menos uma alternativa imprópria - esgoto despejado em fossa negra, a céu aberto, em riachos ou em represas, que se infiltra na terra ou que atinge diretamente as fontes de água; lixo queimado, jogado em terreno baldio ou enterrado, que pode ser levado pela chuva até os córregos; água recolhida de poço ou bica, que não tem distribuição controlada pelo poder público para garantir sua salubridade.

### ***Condição de Salubridade da Moradia***

Indica o grau de salubridade da moradia considerando-se: falta de ventilação, falta de sol, excesso de umidade e janelas funcionando mal ou impedidas. Tem-se duas classificação para as moradias quanto ao seu grau de salubridade:

**adequadas:** quando a moradia não apresenta nenhuma das características citadas anteriormente;

**inadequadas:** quando a moradia apresenta pelo menos uma das características citadas anteriormente.

### ***Desempregados***

São os indivíduos que se encontram numa situação involuntária de não-trabalho, por falta de oportunidade de trabalho ou que exercem trabalhos irregulares com desejo de mudança. Engloba três situações:

**desemprego aberto:** pessoas sem trabalho nos últimos 7 dias e que procuraram trabalho de maneira efetiva nos 30 dias anteriores ao da entrevista;

**desemprego oculto pelo trabalho precário:** pessoas que realizam trabalhos precários – algum trabalho remunerado ocasional de auto-ocupação ou trabalho

não-remunerado em ajuda a negócios de parentes – e que procuraram substituir este trabalho nos 30 dias anteriores ao da entrevista ou que, não tendo procurado neste período, o fizeram sem êxito até doze meses atrás;

**desemprego oculto pelo desalento:** pessoas que não possuem trabalho e nem procuraram nos 30 dias anteriores ao da entrevista, por desestímulos do mercado de trabalho ou por circunstâncias fortuitas, mas apresentaram procura efetiva nos últimos doze meses.

### ***Domicílio***

É o local de moradia, constituído por um ou mais cômodos, com entrada independente e separação. Entende-se, por entrada independente, o acesso direto à moradia, sem passagem por cômodos destinados à moradia de outras pessoas, e, por separação, o local de moradia que é limitado por paredes, muro, cerca, etc., além de ser coberto por um teto, o que permite às famílias ou às pessoas que o habitam isolarem-se das demais.

### ***Domicílios Localizados em Cortiços***

Edificação não localizada em favela, que apresenta condições de salubridade e conservação geralmente insatisfatórias (independente da forma – casarão, cômodos de quintal, etc.), ocupada por várias famílias, sem vínculos de parentesco, que convivem em espaço onde há evidente insuficiência de equipamentos hidráulicos e sanitários, falta de privacidade, ventilação e insolação; o uso de equipamentos sanitários (pia de cozinha e/ou banheiro e/ou tanque de lavar roupas) é dividido entre os moradores.

### ***Domicílios Localizados em Favelas***

Edificação construída, inteira ou parcialmente, com material não apropriado a uma habitação (barraco isolado); ou edificação que integra agrupamento de moradias sem o loteamento observado na maioria das áreas urbanas (ou seja, sem delimitação das ruas e das calçadas, com distâncias variáveis entre as moradias, etc.), independente do material utilizado na sua construção e do número de domicílios que formam o agrupamento (moradia em favela).

### ***Família***

Conjunto de pessoas residentes em um domicílio ligadas por laços de parentesco (consangüinidade, adoção ou afinidade), assim como a pessoa que vive só. Portanto, os grupos conviventes – pessoas sem laços de parentesco (consangüinidade, adoção ou afinidade) que residem num mesmo domicílio – não são considerados família.

Os pensionistas, empregados domésticos e parentes do empregado doméstico moradores do domicílio não são incluídos como membros da família.

A identificação de uma ou mais famílias em um domicílio é obtida a partir da ordenação das relações nucleares (casal), relações primárias (pai, mãe, filho, irmão) e relações secundárias (tio, sobrinho, primo, etc.) e estas famílias são classificadas como:

**família principal:** família única que reside no domicílio ou, quando existe mais de uma família, aquela cujo chefe é ao mesmo tempo chefe do domicílio;

**família(s) secundária(s):** família(s) que divide(m) a moradia com a família principal.

### ***Morador***

Tipos de morador abrangidos pela pesquisa:

**presente:** morador que reside de forma permanente ou habitual no domicílio pesquisado;

**ausente:** morador que reside de forma permanente ou habitual no domicílio pesquisado, mas que, no momento da entrevista, está ausente por um período máximo de seis meses, por motivos de viagem a negócio ou passeio, estudo, doença, detenção sem sentença definitiva ou por outros motivos;

**não morador presente:** morador que não reside de forma fixa ou permanente no domicílio pesquisado e que tem outro domicílio onde mora de forma habitual, para o qual retorna regularmente no mínimo uma vez a cada seis meses. Esta pessoa, no momento da entrevista, encontra-se residindo temporariamente no domicílio pesquisado por razões de estudo, trabalho, tratamento de saúde, etc.

### ***Ocupados***

São os indivíduos que, nos sete dias anteriores ao da entrevista, possuem trabalho remunerado exercido regularmente, com ou sem procura de trabalho; ou que, neste período, possuem trabalho remunerado exercido de forma irregular, desde que não tenham procurado trabalho diferente do atual; ou possuem trabalho não-remunerado de ajuda em negócios de parentes, sem procura de trabalho.



***PEA –População Economicamente Ativa***

População Economicamente Ativa: corresponde à parcela da População em Idade Ativa (PIA) que está ocupada ou desempregada.

***Períodos de Referência***

A PCV utiliza períodos de referência diferenciados de acordo com o objetivo de cada questão investigada:

**últimos sete dias:** referem-se aos sete dias imediatamente anteriores ao dia da entrevista;

**últimos 30 dias:** referem-se aos 30 dias imediatamente anteriores ao dia da entrevista;

**últimos 12 meses:** correspondem aos 12 meses que antecedem o dia da entrevista;

***Portador de Deficiência Física ou Mental***

Foram considerados como portadores de deficiência os indivíduos que declararam possuir os seguintes tipos de deficiência: cegueira, surdez, paralisia permanente de perna e/ou braço e/ou de algum dos lados do corpo, falta de algum membro, retardamento mental, síndrome de Down ou autismo.

***Portador de Dificuldades***

Foram considerados como portadores de dificuldades, os indivíduos portadores de lesões por esforços repetitivos ou outras deficiências e os indivíduos não deficientes, que declararam possuir dificuldade de visão e/ou locomoção e/ou audição.

### ***Posse de Convênio Médico***

Indica a disponibilidade de recursos próprios para assistência médica. Considera todo e qualquer convênio ou plano de saúde por pagamento antecipado, independentemente da personalidade jurídica do prestador de serviços, compreendendo, portanto, empresas de medicina de grupo, cooperativas médicas, instituições de seguro-saúde e planos de autogestão.

### ***Qualidade do Atendimento Recebido***

Nota entre 0 e 10 atribuída pelos usuários de serviços de saúde nos 30 dias anteriores ao da entrevista, que expressa o seu grau de satisfação com a qualidade do atendimento.

### ***Registro em Carteira***

Identifica a formalização ou não do contrato de trabalho do assalariado, do setor privado ou público, e do empregado doméstico:

**com registro:** assalariado ou empregado doméstico com contrato de trabalho regido pela CLT – Consolidação das Leis de Trabalho, com a devida anotação na carteira profissional;

**estatutário:** assalariado do setor público, cujo contrato de trabalho é regido pelo Estatuto do Funcionalismo Público, não estando, portanto, sujeito a assinatura da carteira profissional;

**sem registro:** assalariado ou empregado doméstico com vínculo empregatício caracterizado pela legislação trabalhista vigente, mas sem carteira de trabalho assinada.

### ***Renda Domiciliar per Capita***

Quociente entre a renda familiar total ampliada e o número de membros do domicílio. A renda familiar per capita média é calculada como a soma das rendas familiares per capita dividida pelo total de famílias.

### ***Renda Domiciliar Total***

Soma dos rendimentos auferidos pelos membros do domicílio no mês anterior ao da pesquisa, provenientes das seguintes fontes: rendimentos brutos do trabalho principal; rendimentos brutos do trabalho adicional; aposentadorias e pensões previdenciárias e alimentícias; auxílios; renda de desempregados e inativos; outras rendas; renda de aluguéis – ampliada; e rendimentos financeiros.

### ***Renda Familiar per Capita***

Quociente entre a renda familiar total ampliada e o número de membros da família. A renda familiar per capita média é calculada como a soma das rendas familiares per capita dividida pelo total de famílias.

### ***Renda Familiar Total***

Soma dos rendimentos auferidos pelos membros da família no mês anterior ao da pesquisa, provenientes das seguintes fontes: rendimentos brutos do trabalho

principal; rendimentos brutos do trabalho adicional; aposentadorias e pensões previdenciárias e alimentícias; auxílios; renda de desempregados e inativos; outras rendas; renda de aluguéis – ampliada; e rendimentos financeiros.

### ***Setor censitário***

É a unidade territorial básica de coleta dos Censos Demográficos e Pesquisas Domiciliares por amostragem realizados pela Fundação IBGE. Constituem-se em áreas contínuas situadas em um único quadro urbano ou rural, com seus contornos respeitando os limites territoriais legalmente definidos. Possuem em média 300 domicílios. (IBGE 1990)

### ***Utilização de Serviços de Saúde***

Expressa o resultado da procura de atendimento, indicando a proporção de indivíduos que, nos 30 dias anteriores ao da entrevista, foram atendidos no mesmo dia em que se deslocaram até um estabelecimento de saúde.

### ***Valores Monetários***

Em reais de setembro de 1998. Nesse período o salário mínimo correspondia a R\$120,00.

### ***Vítimas de roubos ou assaltos***

Foram considerados como vítimas de roubos ou assaltos os indivíduos que declararam ter sofrido esse tipo de violência nos últimos 12 meses anteriores a pesquisa.

**ANEXO 3**  
**COMPONENTES DE VARIÂNCIA**

**Tabela 24** – Estimativas dos componentes de variância para as variáveis contínuas utilizadas no estudo, segundo domínios amostrais.

Pesquisa de Condições de Vida. 1998

Variáveis	RMSP		Agglomerado Central		Agglomerado Leste		RMSantos		Agglomerado Norte		Agglomerado Oeste		Vale do Paraíba	
	Entre Setores ( $\hat{\sigma}_a^c$ )	Dentro do Setor ( $\hat{\sigma}_c^c$ )	Entre Setores ( $\hat{\sigma}_a^c$ )	Dentro do Setor ( $\hat{\sigma}_c^c$ )	Entre Setores ( $\hat{\sigma}_a^c$ )	Dentro do Setor ( $\hat{\sigma}_c^c$ )	Entre Setores ( $\hat{\sigma}_a^c$ )	Dentro do Setor ( $\hat{\sigma}_c^c$ )	Entre Setores ( $\hat{\sigma}_a^c$ )	Dentro do Setor ( $\hat{\sigma}_c^c$ )	Entre Setores ( $\hat{\sigma}_a^c$ )	Dentro do Setor ( $\hat{\sigma}_c^c$ )	Entre Setores ( $\hat{\sigma}_a^c$ )	Dentro do Setor ( $\hat{\sigma}_c^c$ )
1. Renda domiciliar total	1.087.834	2.858.614	293.954	1.279.164	587.337	1.628.053	577.014	1.930.885	441.940	1.320.633	243.154	1.267.543	373.415	1.536.769
2. Renda domiciliar per capita	194.222	348.047	46.034	192.830	123.092	196.230	85.892	271.600	68.694	173.870	44.349	211.118	74.393	223.910
3. Renda familiar total	1.063.302	2.792.867	292.956	1.232.105	560.712	1.574.413	593.040	1.860.197	419.286	1.258.684	231.136	1.215.567	365.171	1.502.693
4. Renda familiar per capita	186.679	340.271	45.310	185.242	112.634	194.277	88.081	260.533	65.034	165.607	41.866	201.646	73.102	218.208
5. Anos de estudo	3,95513	36,01555	3,40001	69,73322	2,56337	75,79015	3,70183	90,51309	2,52655	69,63998	1,93184	79,87442	2,32054	44,00687
6. Qualidade do atendimento recebido	0,31956	6,18997	0,31056	5,65763	0,25249	4,62365	0,32942	6,08920	0,28750	4,07596	0,51817	5,09501	0,12131	4,80853

**Tabela 25** – Estimativas dos componentes de variância para as variáveis dicotômicas utilizadas no estudo, segundo domínios amostrais.

Pesquisa de Condições de Vida. 1998

Variáveis	RMSP		Agglomerado Central		Agglomerado Leste		RMSantos		Agglomerado Norte		Agglomerado Oeste		Vale do Paraíba	
	Entre Setores	Dentro do Setor	Entre Setores	Dentro do Setor	Entre Setores	Dentro do Setor	Entre Setores	Dentro do Setor	Entre Setores	Dentro do Setor	Entre Setores	Dentro do Setor	Entre Setores	Dentro do Setor
	$(\hat{\sigma}_a^c)$	$(\hat{\sigma}_e^c)$	$(\hat{\sigma}_a^c)$	$(\hat{\sigma}_e^c)$	$(\hat{\sigma}_a^c)$	$(\hat{\sigma}_e^c)$	$(\hat{\sigma}_a^c)$	$(\hat{\sigma}_e^c)$	$(\hat{\sigma}_a^c)$	$(\hat{\sigma}_e^c)$	$(\hat{\sigma}_a^c)$	$(\hat{\sigma}_e^c)$	$(\hat{\sigma}_a^c)$	$(\hat{\sigma}_e^c)$
7. Domicílios com renda até 10 SM	0,04930	0,17504	0,03285	0,18154	0,03170	0,19676	0,04444	0,18561	0,02586	0,19063	0,02263	0,18331	0,03277	0,19010
8. Domicílios com renda até 3 SM	0,00499	0,15852	0,01600	0,01436	0,00550	0,14656	0,00778	0,14809	0,00346	0,15729	0,00240	0,17889	0,00690	0,15925
9. Famílias com renda até 10 SM	0,04780	0,17458	0,03292	0,17734	0,03030	0,19494	0,00441	0,18380	0,02498	0,18627	0,02178	0,17817	0,03189	0,18871
10. Famílias com renda até 3 SM	0,00541	0,16099	0,01627	0,14777	0,00573	0,15229	0,00792	0,15053	0,00432	0,16370	0,00238	0,18306	0,00651	0,16443
11. Existência de telefone fixo no domicílio	0,06368	0,18311	0,06582	0,18155	0,05681	0,19396	0,06079	0,18726	0,05627	0,19424	0,03968	0,20843	0,04128	0,19933
12. Chefes com menos de 4 anos de escolaridade	0,00562	0,17562	0,01700	0,17171	0,01561	0,17423	0,01621	0,17656	0,00706	0,18937	0,00891	0,19616	0,01265	0,15909
13. Chefes com 5 a 7 anos de escolaridade	0,01181	0,21525	0,01416	0,21403	0,01367	0,21639	0,00953	0,21338	0,00845	0,22332	0,00971	0,20684	0,00643	0,21391
14. Chefes com ensino fundamental completo*	0,00837	0,11640	0,00386	0,11334	0,00263	0,11417	-0,00006	0,11859	0,00287	0,11052	0,00089	0,08680	0,00273	0,13948
15. Chefes com ensino médio completo	0,00847	0,12338	0,01086	0,11898	0,00488	0,11248	0,01089	0,13554	0,00452	0,11802	0,00351	0,13927	0,00908	0,13603
16. Chefes com ensino superior completo	0,03211	0,07538	0,01749	0,07959	0,02554	0,07869	0,01458	0,06918	0,01838	0,07352	0,02149	0,08945	0,01788	0,07711
17. Desempregados	0,00328	0,14265	0,00253	0,11645	0,00364	0,12625	0,00162	0,14673	0,00578	0,12789	0,00150	0,13134	0,00361	0,14544
18. População economicamente ativa	0,00084	0,23441	0,00332	0,24239	0,00095	0,23605	0,00453	0,24214	0,00287	0,23946	0,00642	0,23639	0,00136	0,24502
19. Assalariados com carteira ou do setor público	0,00557	0,19954	0,00462	0,18274	0,01157	0,19324	0,01268	0,18310	0,00996	0,16847	0,00831	0,19536	0,00390	0,17329
20. Utilização de serviços de saúde	0,00575	0,18365	0,00359	0,16820	0,00396	0,18462	0,00594	0,17184	0,00480	0,19088	0,00179	0,18245	0,00281	0,17396
21. Portador de deficiência física ou mental *	0,00004	0,01279	0,00003	0,10314	0,00003	0,02028	0,00013	0,01671	-0,00003	0,01460	0,00025	0,02103	-0,00001	0,01704
22. Portador de dificuldades	0,00059	0,04582	0,00081	0,04527	0,00077	0,04575	0,00117	0,05941	0,00072	0,05732	0,00027	0,04688	0,00096	0,04801
23. Uso de óculos ou lentes de contato	0,02061	0,18648	0,01224	0,18527	0,01605	0,19435	0,01555	0,19899	0,01344	0,20221	0,00847	0,20819	0,01082	0,20085
24. Vítimas de roubos ou assaltos	0,00174	0,05817	0,00058	0,03858	0,00094	0,04344	0,00118	0,06074	0,00081	0,04105	0,00034	0,04016	0,00068	0,05188

\* Estimativas negativas.