

Ministério do Planejamento, Orçamento e Gestão
Instituto Brasileiro de Geografia e Estatística – IBGE
Diretoria de Pesquisas
Departamento de Indústria

Textos para discussão
Diretoria de Pesquisas
número 5

Estudos para definição da amostra da Pesquisa Industrial Mensal de Emprego e Salário

Ana Maria Lima de Farias¹
Doutora em Engenharia Elétrica

Rio de Janeiro
2001

¹ Consultora contratada pelo convênio Rede de Pesquisas e Desenvolvimento de Políticas Públicas – Rede IPEA – Projeto PNUD BRA/97/013.

Instituto Brasileiro de Geografia e Estatística - IBGE
Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro, RJ - Brasil

Textos para discussão. Diretoria de Pesquisas, ISSN 1518-675X

Divulga estudos e outros trabalhos técnicos desenvolvidos pelo IBGE ou em conjunto com outras instituições, bem como resultantes de consultorias técnicas e traduções consideradas relevantes para disseminação pelo Instituto. A série está subdividida por unidade organizacional e os textos são de responsabilidade de cada área específica.

ISBN 85-240-0857-1

© IBGE. 2001

Impressão

Gráfica Digital/Centro de Documentação e Disseminação de Informações - CDDI/IBGE, em 2000.

Capa

Gerência de Criação/CDDI

Farias, Ana Maria Lima de

Estudos para definição da amostra da pesquisa industrial mensal de emprego e salário / Ana Maria Lima de Farias. – Rio de Janeiro : IBGE, Departamento de Indústria, 2001.

59p. - (Textos para discussão. Diretoria de Pesquisas, ISSN 1518-675X; n.5)

Inclui bibliografia.
ISBN 85-240-0857-1

1. Levantamentos industriais – Brasil. 2. Amostragem (Estatística). 3. Brasil – Indústrias – Estatística. I. IBGE. Departamento de Indústria. II. Título. III. Série.

Gerência de Biblioteca e Acervos Especiais CDU 311.21:338.45(81)
IBGE/RJ/2001-11 ECO

Impresso no Brasil / Printed in Brazil

Sumário

Apresentação	5
1 Introdução	7
2 Os Indicadores	7
3 Unidade de Investigação e População-Alvo	7
4 Cadastro e População de Referência	8
4.1 Análise do cadastro intermediário	9
4.2 População de referência e cadastro final	9
5 Domínios de Análise	11
5.1 Proposta inicial	11
5.2 Proposta alternativa	12
6 Desenho da Amostra	13
6.1 Estratos finais definidos segundo Hidiroglou (1986)	17
6.2 Estratos finais definidos pelo método de Lavallée-Hidiroglou (1988)	18
6.3 Cálculo do tamanho da amostra para as diversas propostas	20
7 A Amostra da PIMES	27
8. Seleção e Rotação da Amostra	31
8.1 Algoritmo de seleção e rotação da amostra	32
8.2 Exemplo	36
8.3 Algumas observações sobre a operacionalização do processo	38
8.4 Simulação	39
Referências	53
Apêndice	54

Apresentação

A reestruturação do Sistema de Estatísticas Econômicas, em implementação no IBGE desde 1994, tem por objetivo melhorar a capacidade de resposta da Instituição à crescente e progressivamente mais diversificada demanda de informações, através do aumento de eficiência quanto à qualidade, tempo e custo.

Este sistema tem por base um modelo de produção estatística integrado e ágil, composto por pesquisas, articuladas entre si, desenhadas a partir de um cadastro central de empresas. O sistema é organizado por segmentos econômicos (indústria, comércio, serviços, construção, etc.), que formam subsistemas específicos. Cada subsistema, por sua vez, possui um núcleo central, as pesquisas estruturais anuais e conjunturais, e um conjunto de pesquisas satélites, cobrindo temas específicos ou aprofundando as abordagens setoriais.

Dentro desta linha, o IBGE atualmente encontra-se em fase de consolidação das pesquisas estruturais anuais e de reformulação/implementação das pesquisas mensais.

No subsistema das estatísticas industriais, a Pesquisa Industrial Mensal de Emprego e Salário (PIMES) é a primeira pesquisa conjuntural a ser reformulada sob a égide do projeto de reformulação. A pesquisa mantém objetivos semelhantes aos da atual Pesquisa Industrial Mensal de Dados Gerais (PIM-DG), ou seja, o acompanhamento da evolução de curto prazo do emprego e do salário na indústria.

O objetivo deste texto é apresentar os estudos feitos para definição do desenho da amostra da PIMES, documentando todas as análises e decisões efetuadas ao longo do processo.

Silvio Sales de Oliveira Silva
Chefe do Departamento de Indústria

1 Introdução

Este texto tem como objetivo documentar todos os estudos feitos para a definição da amostra da nova Pesquisa Industrial Mensal de Emprego e Salário (PIMES), que virá a substituir a atual Pesquisa Industrial Mensal - Dados Gerais (PIM-DG). Nessa reformulação, o objetivo central da pesquisa continua o mesmo: fornecer estimativas de indicadores conjunturais de emprego e salário na indústria. No entanto, para caracterizar mais especificamente os objetivos da pesquisa e as mudanças sendo feitas com relação à pesquisa atual, é necessário especificar a população-alvo, a população de referência, a unidade de investigação e os domínios de análise para os quais se pretende produzir as estimativas. Cada um desses tópicos será abordado nas seções seguintes.

2 Os Indicadores

As variáveis atualmente investigadas são (i) Pessoal ocupado na produção (POP); (ii) Admissões; (iii) Desligamentos; (iv) Número de horas pagas na produção; (v) Valor dos salários contratuais pagos ao POP; (vi) Valor das horas extras pagas ao POP; (vii) Valor da folha de pagamento; (viii) Valor da produção. Para essas variáveis, são estimados apenas índices mensais e anuais, sem a divulgação dos totais envolvidos.

Na nova proposta, as variáveis passam a ser: (i) Pessoal ocupado total; (ii) Admissões; (iii) Desligamentos; (iv) Número de horas pagas na produção; (v) Valor da folha de pagamentos; serão produzidas estimativas de totais e os índices mensais e anuais.

3 Unidade de Investigação e População-Alvo

A unidade de investigação da PIMES será a unidade local (UL). Tal decisão foi tomada em virtude da necessidade da pesquisa produzir estimativas por região e atividade industrial. A população alvo é definida como o conjunto de todas as ULs industriais (seções C e D da CNAE - Classificação Nacional de Atividades Econômicas) ativas no ano da pesquisa.

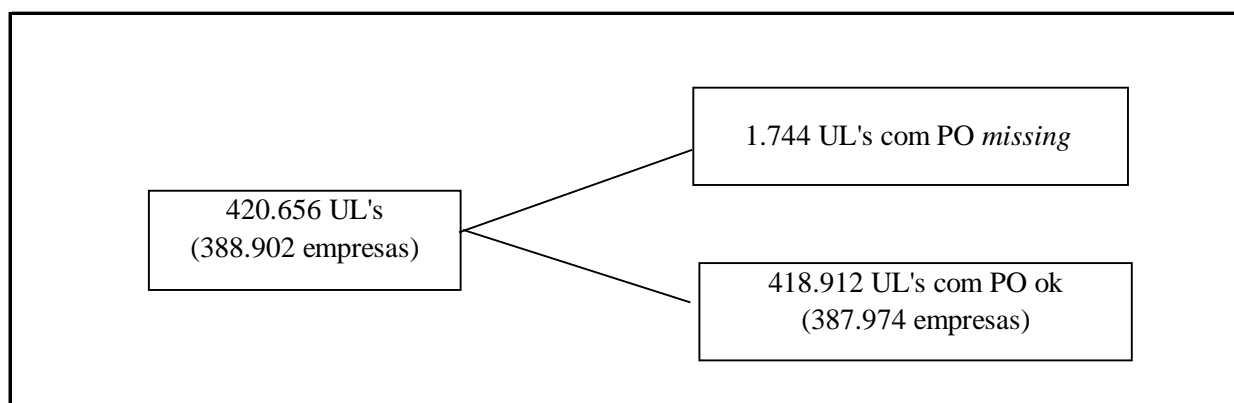
4 Cadastro e População de Referência

Duas possibilidades foram consideradas para o cadastro básico de seleção (CBS) da amostra da PIMES: o cadastro central de empresas (CEMPRE) do IBGE e o cadastro de informantes da Pesquisa Industrial Anual (PIA). A segunda alternativa significaria tratar a amostra da PIMES como uma subamostra da PIA.

A grande vantagem de considerar a amostra da PIMES como uma subamostra da PIA seria garantir uma alta taxa de informantes em atividade (“vivos”); no entanto, embora essa pudesse ser uma vantagem em termos operacionais, é provável que um viés fosse introduzido nas estimativas, uma vez que estaríamos considerando apenas informantes “vivos”. Além disso, como as unidades de investigação da PIMES e da PIA são distintas (na PIA, a unidade de investigação é a empresa), a possibilidade de unificação de procedimentos fica bastante reduzida. Sendo assim, decidiu-se trabalhar com o CEMPRE, o que garante uma maior atualização do cadastro de seleção. [Para maiores detalhes sobre as vantagens e desvantagens de cada alternativa, ver Farias(1999).]

Uma versão preliminar do cadastro básico de seleção da amostra da PIMES foi fornecida pelo Diretoria de Informática, constando de 420.656 observações, correspondentes às ULs industriais de 389.022 empresas industriais, conforme solicitação do Departamento de Indústria (DEIND). Dessas 420.656 ULs, 1.744 não tinham informação sobre o Pessoal Ocupado (PO). Como essa informação é imprescindível para os estudos do desenho amostral, optou-se por excluir estas ULs; sendo assim, os estudos seguintes se baseiam no cadastro formado pelas 418.912 ULs restantes, que correspondem a 387.974 empresas (ver figura 1).

Figura 1
 Cadastro básico de seleção da amostra da PIMES



4.1 Análise do cadastro intermediário

Das ULs remanescentes, 261.598 (62,45%) pertencem a empresas com PO < 5, ou seja, a empresas que estão fora do âmbito da Pesquisa Industrial Anual. Dessas, 14.074 têm PO nulo, tanto para a empresa quanto para a UL. As 247.524 ULs pertencentes a empresas com PO não nulo, mas menor que 5, pertencem a 245.712 empresas mas 452 têm PO nulo.

Para as 157.314 ULs pertencentes a empresas do âmbito da PIA, 4.926 têm PO = 0, enquanto 12.904 têm PO < 5. Na figura 2 da página seguinte temos o esquema do cadastro intermediário.

4.2 População de referência e cadastro final

No total, há 399.460 (247.072 + 12.904 + 139.484) ULs com PO não nulo empregando 5.573.042 pessoas. No entanto, as 259.976 ULs com PO menor que 5 empregam apenas 509.246 pessoas, correspondendo a 9,14% do pessoal total. Sendo assim, a exemplo do que já é feito na PIA, a população de referência da PIMES será definida como o conjunto das ULs com 5 ou mais pessoas empregadas. Logo, o cadastro final para seleção da amostra da PIMES será formado pelas 139.484 ULs industriais com PO ≥ 5. Nas tabelas 1 e 2 temos a distribuição dessas ULs de acordo com a Unidade da Federação e CNAE a 2 dígitos.

Figura 2
Esquema do cadastro intermediário de informantes da PIMES

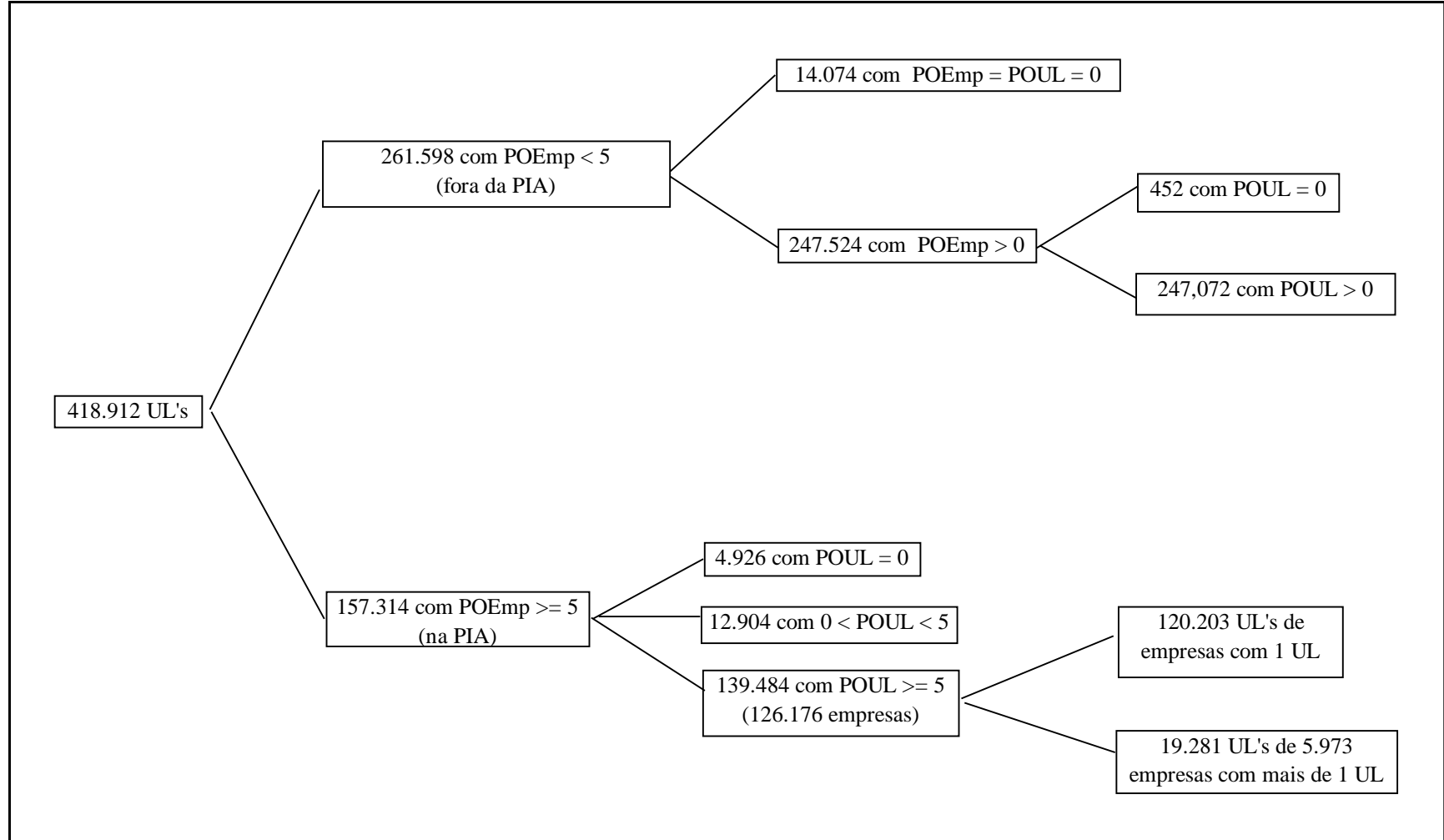


Tabela 1
Distribuição das ULs por Unidade da Federação

Unidade da Federação Código	Sigla	Frequência Simples		Frequência Acumulada	
		Absoluta	%	Absoluta	%
11	RO	880	0,63	880	0,63
12	AC	151	0,11	1.031	0,74
13	AM	705	0,51	1.736	1,24
14	RR	70	0,05	1.806	1,29
15	PA	1.425	1,02	3.231	2,32
16	AP	81	0,06	3.312	2,37
17	TO	274	0,20	3.586	2,57
21	MA	670	0,48	4.256	3,05
22	PI	567	0,41	4.823	3,46
23	CE	2.974	2,13	7.797	5,59
24	RN	983	0,70	8.780	6,29
25	PB	1.213	0,87	9.993	7,16
26	PE	3.290	2,36	13.283	9,52
27	AL	569	0,41	13.852	9,93
28	SE	605	0,43	14.457	10,36
29	BA	3.368	2,41	17.825	12,78
31	MG	16.847	12,08	34.672	24,86
32	ES	2.914	2,09	37.586	26,95
33	RJ	10.803	7,74	48.389	34,69
35	SP	50.135	35,94	98.524	70,63
41	PR	11.157	8,00	109.681	78,63
42	SC	9.388	6,73	119.069	85,36
43	RS	13.539	9,71	132.608	95,07
50	MS	973	0,70	133.581	95,77
51	MT	1.688	1,21	135.269	96,98
52	GO	3.356	2,41	138.625	99,38
53	DF	859	0,62	139.484	100,00
Total Brasil		139.484	100,00		

5 Domínios de Análise

Os domínios de análise estão baseados na localização geográfica e na atividade industrial desenvolvida na UL.

5.1 Proposta inicial

A proposta inicial para os domínios de análise considerava a atividade econômica definida a partir da CNAE, trabalhando-se com o nível de 2 dígitos (divisão), o que totaliza 27 divisões. Para a localização geográfica, foi considerado o seguinte detalhamento:

Tabela 2
Distribuição das ULs por CNAE 2 dígitos

CNAE	Frequência Simples		Frequência Acumulada	
	Absoluta	%	Absoluta	%
10	64	0,05	64	0,05
11	57	0,04	121	0,09
13	333	0,24	454	0,33
14	3.330	2,39	3.784	2,71
15	24.048	17,24	27.832	19,95
16	176	0,13	28.008	20,08
17	5.057	3,63	33.065	23,71
18	16.961	12,16	50.026	35,87
19	5.189	3,72	55.215	39,59
20	8.669	6,22	63.884	45,80
21	2.207	1,58	66.091	47,38
22	8.009	5,74	74.100	53,12
23	385	0,28	74.485	53,40
24	5.698	4,09	80.183	57,49
25	6.187	4,44	86.370	61,92
26	11.542	8,27	97.912	70,20
27	3.133	2,25	101.045	72,44
28	11.405	8,18	112.450	80,62
29	6.298	4,52	118.748	85,13
30	371	0,27	119.119	85,40
31	2.430	1,74	121.549	87,14
32	1.029	0,74	122.578	87,88
33	1.321	0,95	123.899	88,83
34	2.981	2,14	126.880	90,96
35	728	0,52	127.608	91,49
36	11.584	8,30	139.192	99,79
37	292	0,21	139.484	100,00
Total Indústria	139.484	100,00		

- Regiões Norte e Centro-Oeste; Região Nordeste; Minas Gerais; Rio de Janeiro; São Paulo; Região Sudeste; Região Sul.

A diferença fundamental entre essa proposta e a versão atual da PIM-DG diz respeito ao estado do Espírito Santo: na versão atual, esse estado está incluído no estrato denominado Complemento Brasil, junto com as regiões Norte e Centro-Oeste, para o qual não se divulgam as estimativas. Na nova versão, esse estado apenas complementa a região Sudeste.

5.2 Proposta alternativa

Uma segunda proposta considerada para os domínios de análise consiste em se trabalhar com alguns agrupamentos de divisões, conforme exibido na tabela 3, e

desmembrar os estados da região Sul e alguns da região Nordeste, de modo que as regiões consideradas passam a ser:

- Regiões Norte e Centro-Oeste; Ceará; Pernambuco; Bahia; Região Nordeste; Minas Gerais; Rio de Janeiro; São Paulo; Região Sudeste; Paraná; Santa Catarina; Rio Grande do Sul.

Nas tabelas 4 e 5 temos a distribuição das ULs nos estratos naturais para essas duas propostas.

6 Desenho da Amostra

Das variáveis existentes no cadastro de seleção, optou-se por utilizar o número de pessoas ocupadas (PO) para definir o desenho da amostra.

Para permitir a estimação dos indicadores para os domínios de análise desejados, a população foi dividida em estratos naturais construídos a partir dos cruzamentos da localização geográfica e da atividade econômica. Considerando, então, as duas propostas de domínios de análise anteriores, os dois conjuntos de estratos naturais são definidos a partir dos seguintes conjuntos de regiões geográficas:

Proposta inicial

- Regiões Norte e Centro-Oeste; Região Nordeste; Minas Gerais; Espírito Santo; Rio de Janeiro; São Paulo; Região Sul

Proposta alternativa

- Regiões Norte e Centro-Oeste; Região Nordeste, exclusive CE, PE, BA; Ceará; Pernambuco; Bahia; Minas Gerais; Espírito Santo; Rio de Janeiro; São Paulo; Paraná; Santa Catarina; Rio Grande do Sul

Dada a forte assimetria da distribuição da variável PO, decidiu-se dividir a população em cada estrato natural em um estrato certo (*take-all*) e um ou mais estratos amostrados (*take-some*) [ver Hidiroglou(1986) e Lavallée-Hidiroglou(1988)]. A questão que se coloca, agora, é a definição desses estratos finais, construídos a partir da variável PO.

Tabela 3
Classificação de atividades para a PIMES

Divisões da CNAE (2 dígitos)	Agrupamentos de divisões	
10 Extração de carvão mineral 11 Extração de petróleo e serviços correlatos 13 Extração de minerais metálicos 14 Extração de minerais não metálicos	10+11+13+14	Indústrias Extrativas
15 Fabr. de produtos alimentícios e bebidas 16 Fabr. de produtos do fumo	15+16	Fabricação de alimentos, bebidas e produtos do fumo
17 Fabr. de produtos têxteis	17	Fabricação de produtos têxteis
18 Confeção de artigos do vestuário e acessórios	18	Confeção de artigos do vestuário e acessórios
19 Preparação de couros e fabr. de artefatos de couro, artigos de viagem e calçados	19	Indústria do calçado, incl. preparação de artigos de couro
21 Fabr. de produtos de celulose, papel e produtos de papel 22 Edição, impressão e reprodução de gravações	21+22	Indústria do papel e gráfica
23 Fabr. de coque, refino de petróleo, elaboração de combustíveis nucleares e produção de álcool	23	Coque, refino de petróleo, combustíveis nucleares e álcool
24 Fabr. de produtos químicos	24	Fabr. de produtos químicos
25 Fabr. de artigos de borracha e plástico	25	Fabr. de produtos de borracha e plástico
26 Fabr. de produtos de minerais não metálicos	26	Fabricação de produtos de minerais não metálicos
27 Metalurgia básica	27	Metalurgia básica
28 Fabr. de produtos de metal, excl. máquinas e equipamentos	28	Fabr. de produtos de metal, excl. máquinas e equipamentos
29 Fabr. de máquinas e equipamentos 30 Fabr. de máquinas para escritório e equipamentos de informática	29+30	Fabr. de máquinas e equipamentos, excl. elétricos, eletrônicos, de precisão e de comunicação
31 Fabr. de máquinas, aparelhos e materiais elétricos 32 Fabr. de material eletrônico e de aparelhos e equipamentos de comunicação 33 Fabr. equip. instr. médico-hospitalares, instr. precisão e óticos, equip. para automação industrial, cronômetros, relógios	31+32+33	Fabr. de máq. e aparelhos elétricos, eletrônicos, de precisão e de comunicação
34 Fabr. e montagem de veículos automotores, reboques e carrocerias 35 Fabr. de outros equipamentos de transporte	34+35	Fabricação de meios de transporte
20 Fabr. de produtos de madeira 36 Fabr. de móveis e indústrias diversas 37 Reciclagem	20+36+37	Fabricação de outros produtos da ind. de transformação

Tabela 4
Distribuição das ULs segundo a proposta inicial para os domínios de análise

CNAE-2	Região							Total Brasil
	Norte e Centro-Oeste	Nordeste	MG	ES	RJ	SP	Sul	
10	1	3	11		4	6	39	64
11	15	17	1	2	18	1	3	57
13	61	27	158	9	28	30	20	333
14	198	461	663	286	263	785	674	3.330
15	2.549	4.308	3.759	495	1.935	5.846	5.156	24.048
16	10	57	11	1	11	18	68	176
17	167	562	606	42	256	2.252	1.172	5.057
18	975	1.860	2.517	492	1.716	5.821	3.580	16.961
19	179	350	810	44	206	1.581	2.019	5.189
20	1.975	573	496	171	225	1.254	3.975	8.669
21	76	154	142	14	185	1.107	529	2.207
22	695	807	789	111	944	3.212	1.451	8.009
23	46	68	27	7	19	167	51	385
24	272	641	595	40	664	2.513	973	5.698
25	266	467	408	66	460	3.208	1.312	6.187
26	895	1.440	1.533	587	876	3.474	2.737	11.542
27	112	207	460	35	255	1.456	608	3.133
28	494	609	1.275	137	842	5.388	2.660	11.405
29	167	267	380	64	368	3.413	1.639	6.298
30	40	22	24	4	16	208	57	371
31	86	130	173	14	154	1.404	469	2.430
32	80	40	89	5	88	579	148	1.029
33	62	77	121	8	150	705	198	1.321
34	235	164	315	52	171	1.348	696	2.981
35	73	57	52	10	129	289	118	728
36	721	836	1.391	215	790	3.967	3.664	11.584
37	12	35	41	3	30	103	68	292
Total Indústria	10.462	14.239	16.847	2.914	10.803	50.135	34.084	139.484

Tabela 5
Distribuição das ULs segundo a proposta alternativa para os domínios de análise

CNAE Agrupada	Norte e Centro-Oeste	CE	PE	BA	Nordeste excl.CE, PE,BA	Nordeste	MG	ES	RJ	SP	Sudeste	PR	SC	RS	Sul	Total Brasil
10+11+13+14	275	61	74	153	220	508	833	297	313	822	2.265	283	190	263	736	3.784
15+16	2.559	687	1.140	1.039	1.499	4.365	3.770	496	1.946	5.864	12.076	1.779	1.196	2.249	5.224	24.224
17	167	138	122	104	198	562	606	42	256	2.252	3.156	316	535	321	1.172	5.057
18	975	743	339	341	437	1.860	2.517	492	1.716	5.821	10.546	1.120	1.581	879	3.580	16.961
19	179	127	48	62	113	350	810	44	206	1.581	2.641	245	171	1.603	2.019	5.189
21+22	771	170	218	258	315	961	931	125	1.129	4.319	6.504	733	452	795	1.980	10.216
23	46	5	31	9	23	68	27	7	19	167	220	35	8	8	51	385
24	272	125	165	184	167	641	595	40	664	2.513	3.812	393	149	431	973	5.698
25	266	85	113	130	139	467	408	66	460	3.208	4.142	374	328	610	1.312	6.187
26	895	266	351	295	528	1.440	1.533	587	876	3.474	6.470	1.000	863	874	2.737	11.542
27	112	44	49	60	54	207	460	35	255	1.456	2.206	196	138	274	608	3.133
28	494	118	145	154	192	609	1.275	137	842	5.388	7.642	811	594	1.255	2.660	11.405
29+30	207	63	79	79	68	289	404	68	384	3.621	4.477	433	440	823	1.696	6.669
31+32+33	228	54	86	51	56	247	383	27	392	2.688	3.490	304	161	350	815	4.780
34+35	308	43	50	50	78	221	367	62	300	1.637	2.366	296	177	341	814	3.709
20+36+37	2.708	245	280	399	520	1.444	1.928	389	1.045	5.324	8.686	2.839	2.405	2.463	7.707	20.545
Total Indústria	10.462	2.974	3.290	3.368	4.607	14.239	16.847	2.914	10.803	50.135	80.699	11.157	9.388	13.539	34.084	139.484

6.1 Estratos finais definidos segundo Hidiroglou (1986)

Hidiroglou (1986) propôs um método para estratificação de uma população assimétrica em um estrato certo (*take-all*) e outro estrato amostrado (*take-some*), do qual uma amostra aleatória simples é retirada sem reposição. O algoritmo proposto apresenta o valor ótimo de corte que minimiza o tamanho da amostra para um dado coeficiente de variação (CV) do estimador do total da variável de análise.

Nas tabelas 6 e 7 temos o tamanho total da amostra, considerando-se os valores de 10% e 15% para o CV, para as duas propostas de estratos naturais.

Tabela 6
Tamanho da amostra pelo método de Hidiroglou (1986)
Proposta inicial para os estratos naturais

Região	CV=10%			CV=15%		
	Est. certo	Est. amostrado	Total	Est. certo	Est. amostrado	Total
Norte e Centro-Oeste	535	417	952	337	327	664
Nordeste	665	470	1.135	452	369	821
Minas Gerais	668	590	1.258	423	447	870
Espírito Santo	221	190	411	153	142	295
Rio de Janeiro	557	420	977	371	330	701
São Paulo	1.377	1.298	2.675	837	958	1.795
Sul	1.066	885	1.951	676	672	1.348
Total Brasil	5.089	4.270	9.359	3.249	3.245	6.494

Tabela 7
Tamanho da amostra pelo método de Hidiroglou (1986)
Proposta alternativa para os estratos naturais

Região	CV=10%			CV=15%		
	Est. certo	Est. amostrado	Total	Est. certo	Est. amostrado	Total
Norte e Centro-Oeste	459	362	821	287	282	569
Nordeste excl. CE,PE,BA	310	197	507	221	157	378
Ceará	225	157	382	169	116	285
Pernambuco	250	173	423	177	135	312
Bahia	259	178	437	182	140	322
Minas Gerais	588	238	826	376	395	771
Espírito Santo	171	176	347	123	120	243
Rio de Janeiro	491	376	867	323	289	612
São Paulo	1.167	1.198	2.365	680	873	1.553
Paraná	473	414	887	324	290	614
Santa Catarina	445	295	740	303	239	542
Rio Grande do Sul	600	477	1.077	389	366	755
Total Brasil	5.438	4.241	9.679	3.554	3.402	6.956

Os tamanhos de amostra são bastante altos, mesmo para CV=15%, o que nos leva a excluir a possibilidade de adoção desse método.

6.2 Estratos finais definidos pelo método de Lavallée-Hidiroglou (1988)

Lavallée e Hidiroglou propuseram, em 1988, um algoritmo iterativo para decompor uma população assimétrica em um estrato certo e um dado número de estratos amostrados, de modo a minimizar o tamanho total da amostra, dados o coeficiente de variação desejado para o estimador do total da variável de análise e o método de alocação da amostra nos estratos amostrados.

Nas tabelas 8 e 9 temos o tamanho da amostra obtido para o total de 4 estratos (3 amostrados + 1 certo), CV=10% e alocação de Neyman nos estratos amostrados. Por questões numéricas do algoritmo, só foram feitos os cálculos para os estratos naturais com população maior que 19.

Tabela 8
Tamanho da amostra pelo método de Lavallée-Hidiroglou (1988)
CV=10% - Alocação de Neyman – Proposta inicial para os estratos naturais

CNAE-2	Estrato 1		Estrato 2		Estrato 3		Est. certo	Total	
	Pop.	Am.	Pop.	Am.	Pop.	Am.		Pop.	Am.
Norte e Centro-Oeste	6.961	34	2.584	47	762	55	117	10.424	252
Nordeste	10.582	43	2.738	59	759	81	140	14.219	321
Minas gerais	12.078	45	3.589	55	1.005	72	152	16.824	328
Espírito Santo	1.580	11	870	20	329	24	58	2.837	116
Rio de Janeiro	7.430	39	2.472	48	700	63	133	10.735	281
São Paulo	38.688	96	9.175	131	2.101	151	146	50.110	525
Sul	26.387	84	6.219	100	1.334	114	141	34.081	442
Total Brasil	103.706	352	27.647	460	6.990	560	887	139.230 ²	2.265

Embora haja uma forte redução no tamanho da amostra, uma desvantagem da aplicação desse método em uma pesquisa mensal é a variabilidade dos limites dos estratos, o que pode dificultar o processo mensal de crítica e controle da amostra. O próximo passo, então, consistiu em analisar os limites resultantes do algoritmo para tentar unificar a estratificação. Na tabela 10 temos as estatísticas-resumo dos limites superiores para a proposta inicial de estratos naturais e na tabela 11, as estatísticas referentes aos limites para a proposta alternativa. Nas tabelas A1 a A6 do Apêndice são dados os limites superiores de todos os estratos finais.

² Por razões de ordem numérica, só consideramos estratos com população maior que 19.

Tabela 9

Tamanho da amostra pelo método de Lavallée-Hidiroglou (1988)
CV=10% - Alocação de Neyman – Proposta alternativa para os estratos naturais

CNAE-2	Estrato 1		Estrato 2		Estrato 3		Est. certo	Total	
	Pop.	Am.	Pop.	Am.	Pop.	Am.		Pop.	Am.
Norte e Centro-Oeste	7.287	32	2.431	44	654	46	90	10.462	212
Nordeste excl. CE,PE,BA	3.221	20	942	24	336	36	85	4.584	163
Ceará	1.898	12	707	16	274	25	90	2.969	144
Pernambuco	2.149	12	734	21	322	29	85	3.290	147
Bahia	2.129	13	845	22	298	25	87	3.359	149
Minas Gerais	12.477	44	3.382	52	875	65	113	16.847	275
Espírito Santo	1.654	9	871	17	318	19	64	2.907	115
Rio de Janeiro	7.658	34	2.391	43	624	53	111	10.784	240
São Paulo	39.024	80	9.043	110	1.961	122	107	50.135	419
Paraná	7.986	38	2.435	46	601	49	100	11.122	235
Santa Catarina	6.797	30	1.994	36	488	44	101	9.380	210
Rio Grande do Sul	9.904	44	2.786	56	747	69	94	13.531	265
Total Brasil	102.184	368	28.561	487	7.498	582	1.127	139.370 ³	2.574

Tabela 10

Estatísticas-resumo dos limites superiores dos estratos definidos pelo método de Lavallée-Hidiroglou (1988) – CV=10% - Proposta inicial para os estratos naturais

Estatística	Estrato 1	Estrato 2	Estrato 3
Média	20,67	93,03	449,48
Mediana	17	65	284
Mínimo	6	10	24
Máximo	73	580	3537
Desvio padrão	11,78	90,11	491,40

Tabela 11

Estatísticas-resumo dos limites superiores dos estratos definidos pelo método de Lavallée-Hidiroglou (1988) – CV=10% - Proposta alternativa para os estratos naturais

Estatística	Estrato 1	Estrato 2	Estrato 3
Média	19,41	81,93	395,11
Mediana	17	61	270
Mínimo	7	13	43
Máximo	69	473	2714
Desvio padrão	9,62	65,92	382,72

A análise dos resultados gerais para os limites dos estratos levou às seguintes propostas para os estratos finais:

³ Por razões de ordem numérica, só consideramos estratos com população maior que 19.

<u>Proposta 1</u>	<u>Proposta 2</u>
[5,30)	[5,30)
[30,100)	[30,100)
[100,500)	[100,400)
≥ 500 (estrato certo)	≥ 400 (estrato certo)

A escolha de PO=30 para limite do primeiro estrato (e não PO=20) foi tomada para se manter uma analogia com a PIA e também porque as diferenças nos tamanhos de amostra resultaram muito pequenas.

6.3 Cálculo do tamanho da amostra para as diversas propostas

O desenho amostral considerado será o de amostragem aleatória estratificada nos estratos naturais, com alocação de Neyman nos estratos finais. Os estratos naturais são definidos pelos cruzamentos de localização geográfica e atividade econômica e os estratos finais pelo porte da UL, medido através do seu pessoal ocupado. De acordo com as discussões anteriores, serão consideradas as seguintes possibilidades para as três variáveis definidoras do desenho amostral:

- Localização geográfica
 - Regiões Norte e Centro-Oeste; Região Nordeste; Minas Gerais; Espírito Santo; Rio de Janeiro; São Paulo; Região Sul.
 - Regiões Norte e Centro-Oeste; Região Nordeste exclusive CE, PE, BA; Ceará; Pernambuco; Bahia; Minas Gerais; Espírito Santo; Rio de Janeiro; São Paulo; Paraná; Santa Catarina; Rio Grande do Sul.
- Atividade econômica
 - CNAE a 2 dígitos, num total de 27 divisões.
 - Agrupamentos de CNAE a 2 dígitos, num total de 16 agrupamentos (tabela 3)
- Classes de PO
 - [5, 30) [30, 100) [100, 500) ≥ 500
 - [5, 30) [30, 100) [100, 400) ≥ 400

O tamanho da amostra para o estrato amostrado em cada estrato natural foi calculado de forma que o CV para o estimador do total de pessoal ocupado fosse de 10% ou 15%. Assim, o tamanho da amostra em cada estrato natural é dado por:

$$n = N_C + \frac{\left(\sum_{h=1}^3 N_h S_h \right)^2}{c^2 Y_A^2 + \sum_{h=1}^3 N_h S_h^2} \quad (1)$$

onde

- n é o tamanho da amostra no estrato natural em questão;
- N_C é o tamanho da população no estrato certo do estrato natural em questão;
- N_h é o tamanho da população no estrato amostrado final h, h=1,2,3;
- S_h^2 é a variância populacional do PO no estrato amostrado final h, h=1,2,3;
- Y_A é o total populacional do PO no estrato amostrado do estrato natural em questão;
- c é o CV pré-fixado para o estimador do PO total na parte amostrada do estrato natural em questão.

Definido o tamanho total da amostra para cada estrato natural, o tamanho da amostra n_h em cada estrato amostrado final é dado por:

$$n_h = n \times \frac{N_h S_h}{\sum_h N_h S_h} \quad (2)$$

segundo a alocação ótima de Neyman.

Os tamanhos de amostra calculados por (1) ou por (2), quando fracionários, foram sempre arredondados para o inteiro imediatamente maior. Além disso, para evitar problemas operacionais com amostras muito pequenas, foi arbitrado um valor mínimo de 5 ULs para o tamanho da amostra em cada estrato final. Então, quando o valor encontrado para n_h pela expressão (2) foi menor que 5, tomou-se o tamanho da amostra igual a 5, caso $N_h > 5$. Sempre que $N_h \leq 5$, tomou-se o tamanho da amostra no estrato igual ao tamanho da população, o que equivale a incluir na amostra, com certeza, todas as ULs do estrato final em questão. Nas tabelas que se seguem, o valor obtido por (2), arredondado, será denotado Am. Exata ou n1, enquanto o tamanho final,

estipulado obedecendo a restrição de tamanho mínimo de 5 ULs, será denotado Am. Final ou n.

Vamos considerar inicialmente o estrato certo definido pelo corte de 500 empregados. Na tabela 12 temos os resultados referentes à proposta inicial de atividade econômica, que considera as 27 divisões da CNAE, com as diferentes propostas para localização geográfica sendo apresentadas nas partes (a) e (b). Na tabela 13 temos os resultados análogos para os agrupamentos de atividade econômica da tabela 3.

Analisando os resultados, decidiu-se fixar o coeficiente de variação teórico em 10%, principalmente pelo fato de que também serão estimados totais para outras variáveis, além da variável de análise PO.

O tamanho da amostra obtido para o detalhamento máximo da atividade econômica e das regiões geográficas, 6.303, foi considerado grande, dadas as restrições orçamentárias e de pessoal.

Como o agrupamento de divisões foi acordado pelos usuários internos da pesquisa (CEE – Coordenação das Estatísticas Econômicas, DEIND – Departamento de Indústria, DECNA – Departamento de Contas Nacionais, etc), decidiu-se mantê-lo, junto com o detalhamento maior para a localização geográfica.

Nas tabelas 14 e 15 temos os resultados para o estrato certo definido pelo corte de 400 empregados. Em todos os casos, há um aumento no tamanho total da amostra em virtude do aumento do tamanho do estrato certo, que passa de 1.199 ULs para 1.690 ULs.

Tabela 12
Tamanho da amostra – Amostra aleatória estratificada com alocação de Neyman
Atividade econômica: CNAE 2 dígitos (27 divisões)

(a) Proposta inicial para as regiões geográficas

Região	População	CV=10%					CV=15%				
		Estrato amostrado		Est. certo	Total		Est. amostrado		Est. Certo	Total	
		Am. exata	Am. final	PO >= 500	Am. exata	Am. final	Am. exata	Am. final	PO >= 500	Am. exata	Am. final
Norte e Centro-Oeste	10.462	390	467	61	451	528	218	373	61	279	434
Nordeste	14.239	423	480	177	600	657	240	375	177	417	552
Minas Gerais	16.847	456	501	99	555	600	247	377	99	346	476
Espírito Santo	2.914	263	311	12	275	323	160	253	12	172	265
Rio de Janeiro	10.803	426	473	79	505	552	241	373	79	320	452
São Paulo	50.135	510	540	496	1.006	1.036	261	381	496	757	877
Sul	34.084	504	548	275	779	823	259	390	275	534	665
Total Brasil	139.484	2.972	3.320	1.199	4.171	4.519	1.626	2.522	1.199	2.825	3.721

(b) Proposta alternativa para as regiões geográficas

Região	População	CV=10%					CV=15%				
		Estrato amostrado		Est. certo	Total		Est. amostrado		Est. Certo	Total	
		Am. exata	Am. final	PO >= 500	Am. exata	Am. final	Am. exata	Am. final	PO >= 500	Am. exata	Am. final
Norte e Centro-Oeste	10.462	390	467	61	451	528	218	373	61	279	434
Nordeste excl. CE,PE,BA	4.607	322	379	70	392	449	190	315	70	260	385
Ceará	2.974	274	323	41	315	364	168	276	41	209	317
Pernambuco	3.290	289	344	48	337	392	175	291	48	223	339
Bahia	3.368	304	372	18	322	390	184	307	18	202	325
Minas Gerais	16.847	456	501	99	555	600	247	377	99	346	476
Espírito Santo	2.914	263	311	12	275	323	160	253	12	172	265
Rio de Janeiro	10.803	426	473	79	505	552	241	373	79	320	452
São Paulo	50.135	510	540	496	1.006	1.036	261	381	496	757	877
Paraná	11.157	433	470	65	498	535	241	353	65	306	418
Santa Catarina	9.388	379	426	88	467	514	216	342	88	304	430
Rio Grande do Sul	13.539	441	498	122	563	620	245	373	122	367	495
Total Brasil	139.484	4.487	5.104	1.199	5.686	6.303	2.546	4.014	1.199	3.745	5.213

Tabela 13
Tamanho da amostra – Amostra aleatória estratificada com alocação de Neyman
Atividade econômica: 16 agrupamentos de divisões

(a) Proposta inicial para as regiões geográficas

Região	População	CV=10%					CV=15%				
		Estrato amostrado		Est. certo	Total		Est. amostrado		Est. Certo	Total	
		Am. exata	Am. final	PO >= 500	Am. exata	Am. final	Am. exata	Am. final	PO >= 500	Am. exata	Am. final
Norte e Centro-Oeste	10.462	299	333	61	360	394	159	245	61	220	306
Nordeste	14.239	315	332	177	492	509	167	245	177	344	422
Minas Gerais	16.847	339	350	99	438	449	175	245	99	274	344
Espírito Santo	2.914	205	239	12	217	251	122	203	12	134	215
Rio de Janeiro	10.803	318	334	79	397	413	169	242	79	248	321
São Paulo	50.135	352	366	496	848	862	172	245	496	668	741
Sul	34.084	356	370	275	631	645	173	246	275	448	521
Total Brasil	139.484	2.184	2.324	1.199	3.383	3.523	1.137	1.671	1.199	2.336	2.870

(b) Proposta alternativa para as regiões geográficas

Região	População	CV=10%					CV=15%				
		Estrato amostrado		Est. certo	Total		Est. amostrado		Est. Certo	Total	
		Am. exata	Am. final	PO >= 500	Am. exata	Am. final	Am. exata	Am. final	PO >= 500	Am. exata	Am. final
Norte e Centro-Oeste	10.462	299	333	61	360	394	159	245	61	220	306
Nordeste excl. CE,PE,BA	4.607	259	289	70	329	359	146	232	70	216	302
Ceará	2.974	220	258	41	261	299	127	219	41	168	260
Pernambuco	3.290	229	267	48	277	315	136	227	48	184	275
Bahia	3.368	232	271	18	250	289	135	221	18	153	239
Minas Gerais	16.847	339	350	99	438	449	175	245	99	274	344
Espírito Santo	2.914	205	239	12	217	251	122	203	12	134	215
Rio de Janeiro	10.803	318	334	79	397	413	169	242	79	248	321
São Paulo	50.135	352	366	496	848	862	172	245	496	668	741
Paraná	11.157	322	342	65	387	407	169	248	65	234	313
Santa Catarina	9.388	285	303	88	373	391	154	232	88	242	320
Rio Grande do Sul	13.539	329	345	122	451	467	168	244	122	290	366
Total Brasil	139.484	3.389	3.697	1.199	4.588	4.896	1.832	2.803	1.199	3.031	4.002

Tabela 14

Tamanho da amostra – Amostra aleatória estratificada com alocação de Neyman
 Atividade econômica: CNAE 2 dígitos (27 divisões)
 (a) Proposta inicial para as regiões geográficas

Região	População	Amostra - CV=10%				
		Estrato amostrado		Estrato certo	Total	
		Am. exata	Am. final	>= 400	Am. exata	Am. final
Norte e Centro-Oeste	10.462	384	461	85	469	546
Nordeste	14.239	415	464	219	634	683
Minas Gerais	16.847	436	476	144	580	620
Espírito Santo	2.914	272	318	18	290	336
Rio de Janeiro	10.803	414	462	113	527	575
São Paulo	50.135	465	493	706	1.171	1.199
Sul	34.084	472	512	405	877	917
Total Brasil	139.484	2.858	3.186	1.690	4.548	4.876

(b) Proposta alternativa para as regiões geográficas

Região	População	Amostra - CV=10%				
		Estrato amostrado		Estrato certo	Total	
		Am. exata	Am. final	>= 400	Am. exata	Am. final
Norte e Centro-Oeste	10.462	384	461	85	469	546
Nordeste excl. CE,PE,BA	4.607	319	373	84	403	457
Ceará	2.974	279	319	49	328	368
Pernambuco	3.290	296	346	58	354	404
Bahia	3.368	306	366	28	334	394
Minas Gerais	16.847	436	476	144	580	620
Espírito Santo	2.914	272	318	18	290	336
Rio de Janeiro	10.803	414	462	113	527	575
São Paulo	50.135	465	493	706	1.171	1.199
Paraná	11.157	416	452	99	515	551
Santa Catarina	9.388	370	413	116	486	529
Rio Grande do Sul	13.539	414	473	190	604	663
Total Brasil	139.484	4.371	4.952	1.690	6.061	6.642

Tabela 15

Tamanho da amostra – Amostra aleatória estratificada com alocação de Neyman
 Atividade econômica: 16 agrupamentos de divisões
 (a) Proposta inicial para as regiões geográficas

Região	População	Amostra - CV=10%				
		Estrato amostrado		Estrato certo	Total	
		Am. exata	Am. final	>= 400	Am. exata	Am. final
Norte e Centro-Oeste	10.462	289	322	85	374	407
Nordeste	14.239	293	316	219	512	535
Minas Gerais	16.847	313	322	144	457	466
Espírito Santo	2.914	214	248	18	232	266
Rio de Janeiro	10.803	301	319	113	414	432
São Paulo	50.135	319	333	706	1.025	1.039
Sul	34.084	329	340	405	734	745
Total Brasil	139.484	2.058	2.200	1.690	3.748	3.890

(b) Proposta alternativa para as regiões geográficas

Região	População	Amostra - CV=10%				
		Estrato amostrado		Estrato certo	Total	
		Am. exata	Am. final	>= 400	Am. exata	Am. final
Norte e Centro-Oeste	10.462	289	322	85	374	407
Nordeste excl. CE,PE,BA	4.607	250	280	84	334	364
Ceará	2.974	223	254	49	272	303
Pernambuco	3.290	233	265	58	291	323
Bahia	3.368	230	266	28	258	294
Minas Gerais	16.847	313	322	144	457	466
Espírito Santo	2.914	214	248	18	232	266
Rio de Janeiro	10.803	301	319	113	414	432
São Paulo	50.135	319	333	706	1.025	1.039
Paraná	11.157	307	327	99	406	426
Santa Catarina	9.388	279	294	116	395	410
Rio Grande do Sul	13.539	307	325	190	497	515
Total Brasil	139.484	3.265	3.555	1.690	4.955	5.245

7 A Amostra da PIMES

Diante dos resultados obtidos, o desenho da amostra da PIMES contemplaria os seguintes aspectos:

- Unidade de investigação: Unidade Local
- Variável de Análise: Pessoal Ocupado
- Tipo de amostragem: Amostragem aleatória estratificada com alocação de Neyman
- Coeficiente de variação: 10%
- Estratos naturais: 192 cruzamentos de localização geográfica e atividade econômica.
 - * Localização geográfica: Regiões Norte e Centro-Oeste; Região Nordeste exclusive CE, PE, BA; Ceará; Pernambuco; Bahia; Minas Gerais; Espírito Santo; Rio de Janeiro; São Paulo; Paraná; Santa Catarina; Rio Grande do Sul.
 - * Atividade econômica: 16 agrupamentos de divisões da tabela 3.
- Estratos finais: Classes de PO definidas por:
 - $5 \leq PO < 30$
 - $30 \leq PO < 100$
 - $100 \leq PO < 500$
 - $PO \geq 500$ (estrato certo)

No entanto, uma nova análise dos agrupamentos da CNAE foi feita e, devido ao fato de o agrupamento “Fabricação de Outros Produtos da Indústria de Transformação” ter a segunda maior participação no PO total (ver tabela 16), decidiu-se isolar a divisão “Fabricação de Produtos de Madeira”. Além disso, decidiu-se separar também a divisão “Fabricação de Produtos do Fumo” em virtude da forte sazonalidade da produção e também do espalhamento geográfico das ULs.

Trabalhando-se com esses 18 agrupamentos da CNAE e as 12 regiões geográficas propostas, o tamanho total da amostra seria de 5.275 ULs distribuídas pelas regiões e agrupamentos de CNAE conforme exibido nas tabelas 17 e 18.

Tabela 16
Participação dos agrupamentos de CNAE no PO total

Agrupamento da CNAE	Pessoal Ocupado	
	Total	%
15	933.052	18,41
← 20+36+37	476.756	9,41
18	369.932	7,30
34+35	334.468	6,60
21+22	330.504	6,52
29+30	317.978	6,28
28	298.735	5,90
26	284.888	5,62
24	279.678	5,52
17	255.221	5,04
31+32+33	253.634	5,01
19	252.094	4,98
25	244.274	4,82
27	190.317	3,76
10+11+13+14	118.676	2,34
23	105.987	2,09
16	20.902	0,41
Total da Indústria	5.067.096	100,00

Uma análise da tabela 18 revela que a amostra para a divisão Fumo (CNAE=16) é bastante grande (130) quando comparada com o tamanho da população (176). Dados esse resultado e a já citada peculiaridade dessa atividade, decidiu-se incluir na amostra todas as ULs de tal divisão, ou seja, a divisão Fumo passa a ser considerada como um estrato certo.

Tabela 17
Tamanho da amostra da PIMES por região geográfica

Região	População	Amostra – CV=10%				
		Est. amostrado		Est. Certo (>= 500)	Total	
		Am. exata	Am. final		Am. exata	Am. final
Norte e Centro-Oeste	10.462	335	365	61	396	426
Nordeste excl. CE,PE,BA	4.607	295	321	70	365	391
Ceará	2.974	249	279	41	290	320
Pernambuco	3.290	261	294	48	309	342
Bahia	3.368	273	304	18	291	322
Minas Gerais	16.847	375	386	99	474	485
Espírito Santo	2.914	231	258	12	243	270
Rio de Janeiro	10.803	353	365	79	432	444
São Paulo	50.135	392	406	496	888	902
Paraná	11.157	356	375	65	421	440
Santa Catarina	9.388	314	334	88	402	422
Rio Grande do Sul	13.539	365	389	122	487	511
Total Brasil	139.484	3.799	4.076	1.199	4.998	5.275

Tabela 18
Tamanho da população e da amostra por agrupamento de CNAE

Grupos de CNAE	[5, 30)			[30, 100)			[100, 500)			>=500	Total		
	N	n1	n	N	n1	n	N	n1	n		N	n1	n
10+11+13+14	3.082	106	106	562	58	64	113	59	61	27	3.784	250	258
15	19.943	110	110	2.529	48	60	1.306	111	111	270	24.048	539	551
16	81	31	46	47	34	42	41	35	35	7	176	107	130
17	3.838	71	75	733	39	59	383	96	96	103	5.057	309	333
18	14.530	148	148	1.984	60	64	402	61	66	45	16.961	314	323
19	3.780	77	84	920	53	62	416	70	70	73	5.189	273	289
20	7.249	157	157	1.138	55	61	261	44	46	21	8.669	277	285
21+22	8.440	100	100	1.186	42	60	520	95	95	70	10.216	307	325
23	135	26	49	67	26	33	127	61	62	56	385	169	200
24	3.942	66	71	1.138	53	63	553	92	92	65	5.698	276	291
25	4.498	75	77	1.199	53	60	445	70	74	45	6.187	243	256
26	9.563	122	122	1.580	63	67	368	67	71	31	11.542	283	291
27	2.368	68	78	471	44	54	242	76	78	52	3.133	240	262
28	9.465	124	124	1.459	50	60	440	62	66	41	11.405	277	291
29+30	4.728	67	69	1.332	51	61	532	80	80	77	6.669	275	287
31+32+33	3.445	61	69	828	47	58	418	84	86	89	4.780	281	302
34+35	2.644	71	76	643	47	60	320	70	70	102	3.709	290	308
36+37	10.028	135	135	1.417	59	64	406	69	69	25	11.876	288	293
Total da Indústria	111.759	1.615	1.696	19.233	882	1.052	7.293	1.302	1.328	1.199	139.484	4.998	5.275

A análise dos resultados da PIA mostra uma taxa de morte de empresas na ordem de 20%, o que revela uma desatualização do cadastro. De modo a evitar que a representatividade da amostra da PIMES fique prejudicada em função de tal desatualização, decidiu-se também aumentar em 20% o tamanho da amostra em cada estrato natural. Nas tabelas 19 e 20 temos a distribuição da amostra final da PIMES por região e agrupamento de CNAE respectivamente. Tal amostra, de 5.862 ULs, está distribuída em 787 estratos finais, dos quais 527 são amostrados, 151 são estratos certos e 109 são estratos definidos gerencialmente como certos (fumo ou tamanho mínimo da amostra igual a 5).

Tabela 19
Tamanho da amostra final da PIMES por região geográfica

Região	População	Amostra – CV=10%				
		Est. amostrado		Est. Certo (>= 500)	Total	
		Am. exata	Am. final		Am. exata	Am. final
Norte e Centro-Oeste	10.462	390	412	61	451	473
Nordeste excl. CE,PE,BA	4.607	359	370	70	429	440
Ceará	2.974	286	309	41	327	350
Pernambuco	3.290	301	326	48	349	374
Bahia	3.368	324	344	18	342	362
Minas Gerais	16.847	438	445	99	537	544
Espírito Santo	2.914	261	281	12	273	293
Rio de Janeiro	10.803	414	421	79	493	500
São Paulo	50.135	463	472	496	959	968
Paraná	11.157	419	429	65	484	494
Santa Catarina	9.388	383	389	88	471	477
Rio Grande do Sul	13.539	461	465	122	583	587
Total Brasil	139.484	4.499	4.663	1.199	5.698	5.862

Tabela 20
Tamanho da população e da amostra por agrupamento de CNAE

Grupos de CNAE	[5, 30)			[30, 100)			[100, 500)			>=500	Total		
	N	n1	n	N	N1	n	N	n1	n	N	N	n1	n
10+11+13+14	3.082	127	127	562	70	72	113	66	68	27	3.784	290	294
15	19.943	131	131	2.529	56	61	1.306	134	134	270	24.048	591	596
16	81	81	81	47	47	47	41	41	41	7	176	176	176
17	3.838	83	85	733	47	60	383	115	114	103	5.057	348	362
18	14.530	175	175	1.984	71	71	402	72	76	45	16.961	363	367
19	3.780	89	94	920	64	67	416	80	78	73	5.189	306	312
20	7.249	187	187	1.138	65	67	261	52	53	21	8.669	325	328
21+22	8.440	124	124	1.186	50	62	520	111	111	70	10.216	355	367
23	135	26	49	67	26	33	127	69	70	56	385	177	208
24	3.942	79	81	1.138	62	68	553	106	106	65	5.698	312	320
25	4.498	86	87	1.199	61	65	445	83	86	45	6.187	275	283
26	9.563	145	145	1.580	76	77	368	77	81	31	11.542	329	334
27	2.368	79	88	471	49	56	242	90	89	52	3.133	270	285
28	9.465	147	147	1.459	59	65	440	68	72	41	11.405	315	325
29+30	4.728	76	78	1.332	60	66	532	92	92	77	6.669	305	313
31+32+33	3.445	71	78	828	53	61	418	96	97	89	4.780	309	325
34+35	2.644	84	87	643	54	63	320	81	81	102	3.709	321	333
36+37	10.028	160	160	1.417	69	72	406	77	77	25	11.876	331	334
Total da Indústria	111.759	1.950	2.004	19.233	1.039	1.133	7.293	1.510	1.526	1.199	139.484	5.698	5.862

8. Seleção e Rotação da Amostra

A exemplo do que já é feito nas Pesquisas Anuais da Indústria e do Comércio, o cadastro de seleção da PIMES será extraído, a cada ano, do cadastro central com as informações mais atualizadas. Uma seleção independente da amostra a cada ano, no entanto, poderia provocar variações muito bruscas nos índices. Sendo assim, será adotado um mecanismo de seleção e rotação da amostra que permitirá que as UL's dos estratos amostrados permaneçam por um número máximo esperado de rodadas da pesquisa. Tal mecanismo é totalmente análogo ao empregado na PIA, mas utiliza os Números Aleatórios Permanentes (NAP) [ver Ohlsson (1995)] das ULs, uma vez que a unidade de investigação da PIMES é a UL.

Definido o tempo esperado de permanência na amostra das ULs dos estratos amostrados (na PIA, esse tempo é de 4 anos para as pequenas empresas), o mesmo

algoritmo usado na PIA [ver Silva et al.(1998)] definirá a nova amostra da PIMES, assegurando a coordenação negativa das amostras em anos adjacentes. A diferença fundamental será o fato de que, em vez de ser feita em um único instante de tempo, a rotação será distribuída ao longo do ano, de modo a evitar variações bruscas nos índices devidas simplesmente à substituição de parte da amostra.

No que segue, estaremos descrevendo o procedimento a ser aplicado em cada um dos estratos amostrados finais da amostra da PIMES que são definidos pelos cruzamentos de Região Geográfica, agrupamentos da CNAE a 2 dígitos e as 3 classes de PO dadas por [5,30), [30,100) e [100,500).

8.1 Algoritmo de seleção e rotação da amostra

A seguir descrevem-se as etapas do procedimento de seleção e rotação da amostra nos estratos amostrados.

- As ULs do cadastro são estratificadas de acordo com os critérios de estratificação definidos no plano amostral, sendo a estratificação refeita a cada ano, considerando as informações mais atualizadas disponíveis.
- Cada UL, ao entrar no cadastro do IBGE, recebe um número aleatório permanente (NAP) que não será modificado enquanto a UL permanecer ativa no cadastro. Esses NAPs já foram gerados quando da montagem do cadastro para seleção das amostras das Pesquisas Anuais da Indústria (PIA) e do Comércio (PAC).
- Em cada estrato amostrado final, as ULs são ordenadas crescentemente segundo os NAPs a elas associados. Obtém-se, assim, a população ordenada de números aleatórios $A_{(1)}, A_{(2)}, \dots, A_{(N)}$, onde N aqui representa o tamanho da população do estrato final em questão. Partindo da população ordenada, determinam-se as posições das ULs, ou seja, os postos P_1, P_2, \dots, P_N das ULs segundo os números aleatórios.
- Em cada rodada da pesquisa, calcula-se novamente o tamanho da amostra para todos os estratos, usando as informações mais atualizadas.
- As posições inicial e final das ULs a serem incluídas na amostra ao longo do ano são dadas por:

$$\left\lfloor (r-1) \times \frac{n}{T} \right\rfloor \bmod N + 1 \quad (3)$$

$$\text{Final} = \text{Início} + n - 1 \quad (4)$$

onde

$\lfloor \cdot \rfloor$ representa o menor inteiro maior ou igual a \cdot ;

r é a rodada da pesquisa;

n é o tamanho da amostra no estrato final em questão;

T é o limite máximo de rodadas que se espera que as ULs permaneçam na amostra;

N é o tamanho da população no estrato final em questão;

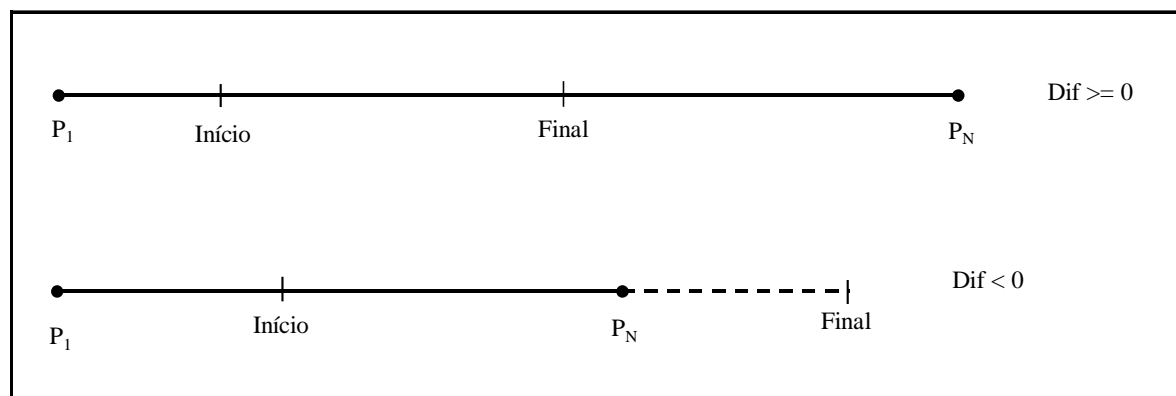
mod é a função módulo, que retorna o resto da divisão.

- Caso o valor de Final seja maior que o tamanho N da população, o procedimento consiste em completar a amostra com o número necessário de elementos do início do estrato (menores NAPs). Então, calcula-se a diferença

$$\text{Dif} = \text{Final} - N = \text{Início} + n - 1 - N = n + (\text{Início} - N - 1) \quad (5)$$

que controla a disponibilidade de ULs para a rotação (ver figura 3).

Figura 3
Ilustração do processo de rotação – Parte I



- A regra final de inclusão das ULs na amostra é a seguinte (ver figura 4):

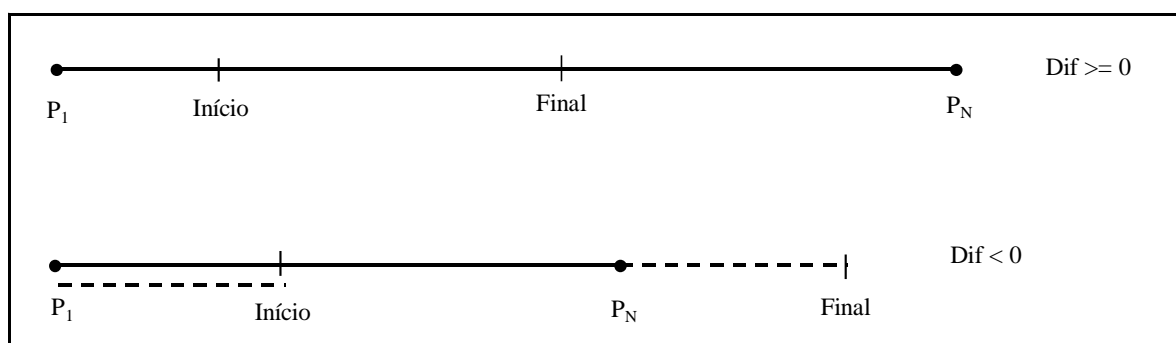
(ii) se $\text{Dif} \leq 0$ incluem-se as ULs para as quais

$$\text{Início} \leq P_i \leq \text{Final} \quad (6)$$

(ii) se $Dif > 0$ (isto é, $Final > N$) incluem-se as ULs para as quais

$$\left\{ \begin{array}{l} \text{Início} \leq P_i \leq \text{Final} \\ \text{ou} \\ P_i \leq Dif \end{array} \right. \quad (7)$$

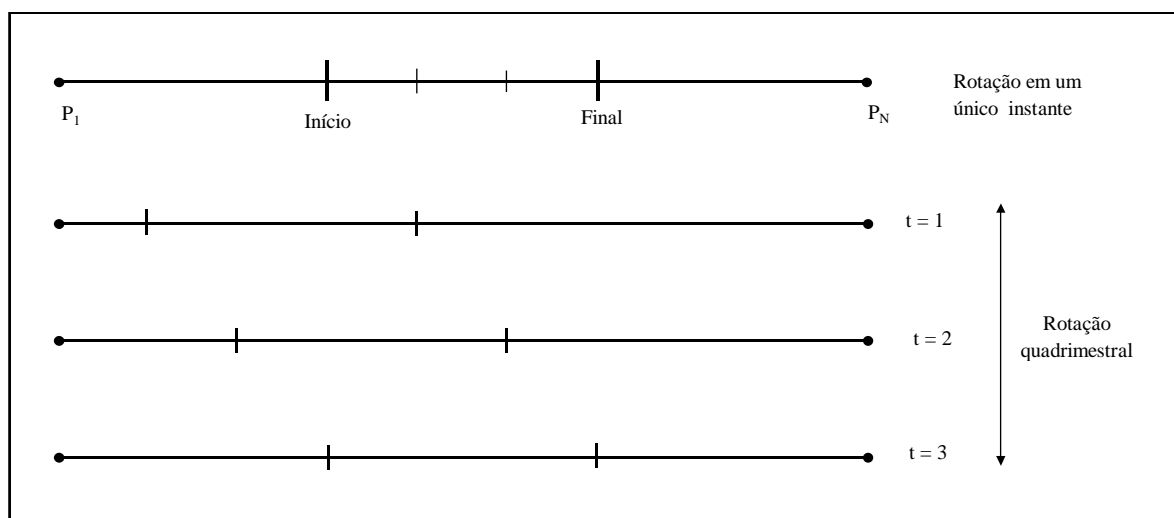
Figura 4
Ilustração do processo de rotação – Parte II



Até aqui, o processo é idêntico ao da PIA. Então, no instante t_1^r onde é iniciada a r -ésima rodada da pesquisa, tem-se, com o procedimento acima, todas as ULs que serão incluídas na amostra ao longo do ano. Na primeira rodada da pesquisa ($r = 1$), são incluídas as ULs com os n primeiros postos. A partir da primeira rodada, espera-se, em média, uma rotação de $\left(\frac{1}{T} \times 100\right)\%$ da amostra, isto é, $\left(\frac{1}{T} \times 100\right)\%$ da amostra é substituída ao longo do ano.

A diferença na PIMES é que, a partir da segunda rodada, a substituição de parte da amostra não será feita em um único instante, mas, sim, distribuída por trimestre ou por quadrimestre (ver figura 5).

Figura 5
Ilustração do processo de rotação – Parte III



Vamos formalizar, agora, a regra de inclusão das ULs, supondo que a substituição será feita em cada trimestre. Consideremos, então, a r -ésima rodada da pesquisa, $r \geq 2$. A substituição de parte da amostra será feita em 4 instantes:

- $t=1$ instante da rotação;
- $t=2$ primeiro mês do segundo trimestre;
- $t=3$ primeiro mês do terceiro trimestre;
- $t=4$ primeiro mês do quarto trimestre.

Para definir as ULs que entrarão no trimestre t da r -ésima rodada, definem-se:

$$Início_t^r = \left[\left(r - 2 + \frac{t}{4} \right) \times \frac{n}{T} \right] \bmod N + 1 \quad r = 2, 3, \dots; \quad t = 1, 2, 3, 4 \quad (8)$$

$$Final_t^r = Início_t^r + n - 1 \quad (9)$$

$$Dif_t^r = Final_t^r - N = n - (N - Início_t^r + 1) \quad (10)$$

e selecionam-se as ULs com posto P_i tal que:

$$(i) \text{ se } Dif_t^r \leq 0, \quad \text{Início}_t^r \leq P_i \leq \text{Final}_t^r \quad (11)$$

$$(ii) \text{ se } Dif_t^r > 0 \quad \left\{ \begin{array}{l} \text{Início}_t^r \leq P_i \leq N \\ \text{ou} \\ P_i \leq Dif_t^r \end{array} \right. \quad (12)$$

Caso se opte por fazer a rotação por quadrimestre, o procedimento é análogo: a substituição será feita em 3 instantes de tempo ($t=1,2,3$) e nas fórmulas, em vez de $\frac{t}{4}$, teremos $\frac{t}{3}$.

8.2 Exemplo

Consideremos um estrato final onde o tamanho da população é $N = 22$ e de onde se pretende tirar uma amostra de tamanho $n = 11$. Suponhamos que se pretenda manter cada UL por no máximo 3 anos, isto é, $T = 3$. Suponhamos, também, que a população permaneça fixa, de modo que a cada rodada temos o mesmo tamanho de população e amostra. É claro que esse exemplo é uma simplificação da realidade, uma vez que, na prática, teremos, a cada ano, tamanhos de população e amostra diferentes.

1ª rodada

No início da pesquisa ($r = 1$), são selecionadas as ULs com os 11 primeiros postos.

2ª rodada ($r = 2$)

$t=1$

$$\text{Início}_1^2 = \left\lfloor \left(2 - 2 + \frac{1}{4} \right) \times \frac{11}{3} \right\rfloor \bmod 22 + 1 = \lfloor 0,92 \rfloor \bmod 22 + 1 = 1 \bmod 22 + 1 = 1 + 1 = 2$$

$$\text{Final}_1^2 = 2 + 11 - 1 = 12 \quad \Rightarrow \quad 2 \leq P_i \leq 12$$

$t=2$

$$\text{Início}_2^2 = \left\lfloor \left(2 - 2 + \frac{2}{4} \right) \times \frac{11}{3} \right\rfloor \bmod 22 + 1 = \lfloor 1,83 \rfloor \bmod 22 + 1 = 2 \bmod 22 + 1 = 3$$

$$\text{Final}_2^2 = 3 + 11 - 1 = 13 \quad \Rightarrow \quad 3 \leq P_i \leq 13$$

$t=3$

$$\text{Início}_3^2 = \left\lfloor \left(2 - 2 + \frac{3}{4} \right) \times \frac{11}{3} \right\rfloor \bmod 22 + 1 = \lfloor 2,75 \rfloor \bmod 22 + 1 = 3 \bmod 22 + 1 = 3 + 1 = 4$$

$$\text{Final}_3^2 = 4 + 11 - 1 = 14 \quad \Rightarrow \quad 4 \leq P_i \leq 14$$

t=4

$$\text{Início}_4^2 = \left\lfloor \left(2 - 2 + \frac{4}{4} \right) \times \frac{11}{3} \right\rfloor \bmod 22 + 1 = \lfloor 3,67 \rfloor \bmod 22 + 1 = 4 \bmod 22 + 1 = 4 + 1 = 5$$

$$\text{Final}_4^2 = 5 + 11 - 1 = 15 \quad \Rightarrow \quad 5 \leq P_i \leq 15$$

3ª rodada (r = 3)

$$t=1 \quad 6 \leq P_i \leq 16$$

$$t=2 \quad 7 \leq P_i \leq 17$$

$$t=3 \quad 8 \leq P_i \leq 18$$

$$t=4 \quad 9 \leq P_i \leq 19$$

4ª rodada (r = 4)

$$t=1 \quad 10 \leq P_i \leq 20$$

$$t=2 \quad 11 \leq P_i \leq 21$$

$$t=3 \quad 12 \leq P_i \leq 22$$

$$t=4 \quad 12 \leq P_i \leq 22$$

5ª rodada (r = 5)

$$t=1 \quad \text{Início}_1^5 = 13 \quad \text{Final}_1^5 = 23 \quad \text{Dif}_1^5 = 1 \quad \Rightarrow \quad 13 \leq P_i \leq 22 \quad \text{ou} \quad P_i \leq 1$$

$$t=2 \quad \text{Início}_2^5 = 14 \quad \text{Final}_2^5 = 24 \quad \text{Dif}_2^5 = 2 \quad \Rightarrow \quad 14 \leq P_i \leq 22 \quad \text{ou} \quad P_i \leq 2$$

$$t=3 \quad \text{Início}_3^5 = 15 \quad \text{Final}_3^5 = 25 \quad \text{Dif}_3^5 = 3 \quad \Rightarrow \quad 15 \leq P_i \leq 22 \quad \text{ou} \quad P_i \leq 3$$

$$t=4 \quad \text{Início}_4^5 = 16 \quad \text{Final}_4^5 = 26 \quad \text{Dif}_4^5 = 4 \quad \Rightarrow \quad 16 \leq P_i \leq 22 \quad \text{ou} \quad P_i \leq 4$$

6ª rodada (r = 6)

$$t=1 \quad \text{Início}_1^6 = 17 \quad \text{Final}_1^6 = 27 \quad \text{Dif}_1^6 = 5 \quad \Rightarrow \quad 17 \leq P_i \leq 22 \quad \text{ou} \quad P_i \leq 5$$

$$t=2 \quad \text{Início}_2^6 = 18 \quad \text{Final}_2^6 = 28 \quad \text{Dif}_2^6 = 6 \quad \Rightarrow \quad 18 \leq P_i \leq 22 \quad \text{ou} \quad P_i \leq 6$$

$$t=3 \quad \text{Início}_3^6 = 19 \quad \text{Final}_3^6 = 29 \quad \text{Dif}_3^6 = 7 \quad \Rightarrow \quad 19 \leq P_i \leq 22 \quad \text{ou} \quad P_i \leq 7$$

$$t=4 \quad \text{Início}_4^6 = 20 \quad \text{Final}_4^6 = 30 \quad \text{Dif}_4^6 = 8 \quad \Rightarrow \quad 20 \leq P_i \leq 22 \quad \text{ou} \quad P_i \leq 8$$

Na figura 6 ilustra-se o processo de rotação para o exemplo.

Figura 6
Ilustração do processo de rotação para o exemplo

r=1		1 2 3 4 5 6 7 8 9 10 11
r=2	t=1	2 3 4 5 6 7 8 9 10 11 12
	t=2	3 4 5 6 7 8 9 10 11 12 13
	t=3	4 5 6 7 8 9 10 11 12 13 14
	t=4	5 6 7 8 9 10 11 12 13 14 15
r=3	t=1	6 7 8 9 10 11 12 13 14 15 16
	t=2	7 8 9 10 11 12 13 14 15 16 17
	t=3	8 9 10 11 12 13 14 15 16 17 18
	t=4	9 10 11 12 13 14 15 16 17 18 19
r=4	t=1	10 11 12 13 14 15 16 17 18 19 20
	t=2	11 12 13 14 15 16 17 18 19 20 21
	t=3	12 13 14 15 16 17 18 19 20 21 22
	t=4	12 13 14 15 16 17 18 19 20 21 22
r=5	t=1	1 13 14 15 16 17 18 19 20 21 22
	t=2	1 2 14 15 16 17 18 19 20 21 22
	t=3	1 2 3 15 16 17 18 19 20 21 22
	t=4	1 2 3 4 16 17 18 19 20 21 22
r=6	t=1	1 2 3 4 5 17 18 19 20 21 22
	t=2	1 2 3 4 5 6 18 19 20 21 22
	t=3	1 2 3 4 5 6 7 19 20 21 22
	t=4	1 2 3 4 5 6 7 8 20 21 22

8.3 Algumas observações sobre a operacionalização do processo

- 1) Para alguns estratos finais, a diferença entre o tamanho da população e o tamanho da amostra é pequena. Com esse processo de rotação, é possível que, nesses estratos, algumas ULs tenham que sair e entrar na amostra em um momento posterior no mesmo ano. Operacionalmente, isso pode ser bastante complicado. Sugere-se, então, que os estratos onde a diferença entre os tamanhos da população e amostra seja menor que 5 sejam definidos como estratos certos.
- 2) A rotação da amostra deverá ser feita anualmente, em algum instante no qual se tenha o cadastro o mais atualizado possível (RAIS mais recente e incorporação das FACs da PIA).

- 3) Neste instante, usando as fórmulas (6) a (10), são identificadas as ULs que entrarão na amostra e o trimestre do ingresso. Comparando com a amostra da rodada anterior, é possível identificar as ULs que deixarão a amostra e as ULs que entrarão na amostra. As primeiras devem ser notificadas que deixarão de participar da pesquisa por um certo tempo mas poderão retornar ao painel de informantes em um momento futuro. As ULs novas têm que ser notificadas e treinadas para responder o questionário.
- 4) A questão da retirada e inclusão de ULs na amostra tem que ser tratada com bastante cuidado. Fazer a exclusão e a inclusão simultaneamente pressupõe que os “novos” informantes responderão o questionário a contento na primeira entrevista e essa pode ser uma suposição errônea. Talvez valha a pena manter os “velhos” informantes por mais 1 ou 2 meses, garantindo, assim, a completitude da amostra. Isso certamente acarretará um aumento no serviço de coleta, já que teremos uma amostra maior por 1 ou 2 meses na época de cada rotação. Essa sobrecarga de trabalho de coleta pode acarretar na necessidade de se fazer a rotação em 3 instantes de tempo ao longo do ano, em vez de trimestralmente.
- 5) Em termos de armazenamento da amostra no microcomputador, é necessário manter, não só a amostra atual, mas também a amostra da rodada anterior, uma vez que a substituição é feita por partes.
- 6) Caso se decida pela utilização de painéis comparáveis no processo de estimação, um cuidado especial é necessário. As ULs novas terão que ter seus dados levantados a partir do mês anterior ao previsto para sua inclusão na amostra, isto é, levantam-se os dados a partir do mês j mas a UL só entra no cálculo dos indicadores no mês $j+1$.

8.4 Simulação

Para avaliar o impacto do processo de rotação, foi feito um estudo utilizando cadastros de dois anos consecutivos: 1997 e 1998. Com base no cadastro de 1997, selecionou-se o que seria a primeira amostra da pesquisa e no cadastro de 1998 aplicou-se o processo de rotação para gerar a amostra da segunda rodada. Nesse estudo, considerou-se também como certo gerencial o estrato para o qual a diferença entre os tamanhos da população e da amostra era menor que 5 (ver observação 1). Essa decisão acarretou um aumento de 116 ULs na amostra de 1997 e de 107 ULs na amostra de 1998 (ver tabela 21).

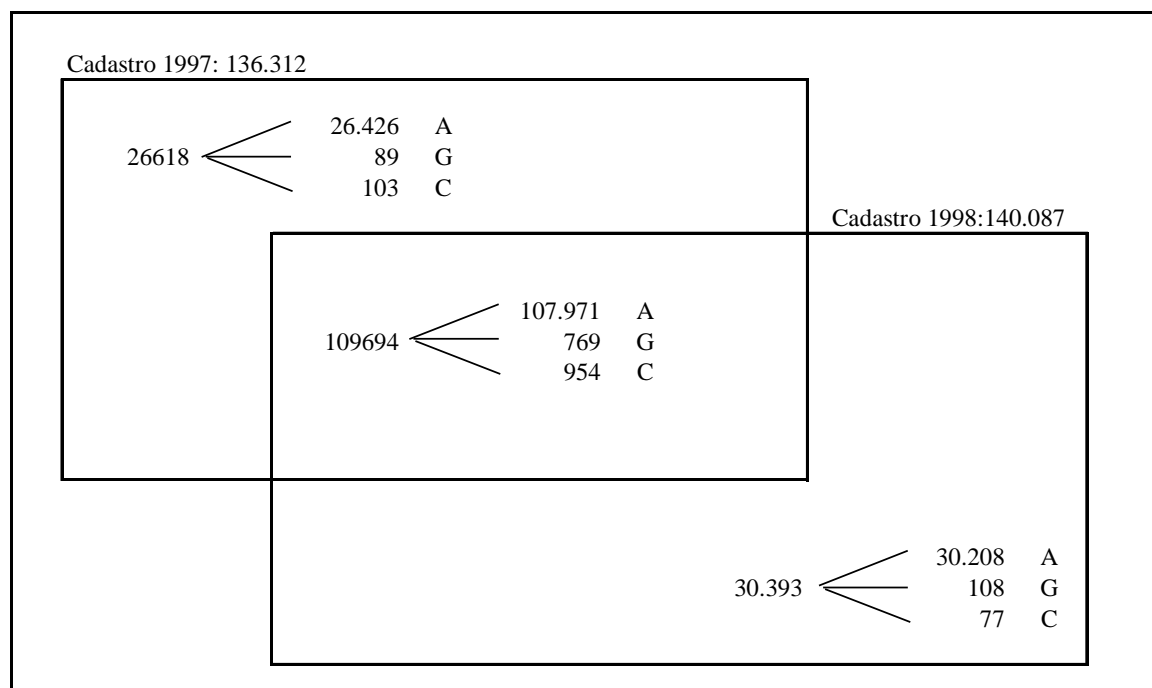
Tabela 21
Tamanho da população e da amostra nos cadastros de 1997 e 1998

	Cadastro 1997	Cadastro 1998
População	136.312	140.087
Amostra exata	5.018	4.918
Amostra com estratos gerenciais	5.291	5.203
Amostra com estratos gerenciais + 20%	5.822	5.763
Amostra com estratos gerenciais + 20% + novos estratos gerenciais	5.938	5.870
Estratos amostrados	472	479
Estratos gerenciais	165	158
Estratos certos	151	153

A composição desses dois cadastros está ilustrada na figura 7 a seguir, onde utilizou-se a seguinte notação:

- A estrato amostrado
- G estrato certo gerencial
- C estrato certo

Figura 7
Composição dos cadastros de 1997 e 1998



Pode-se observar que na parte comum dos dois cadastros houve um remanejamento das ULs dentro dos estratos finais, detalhada na tabela 22. Das 107.971 ULs dos estratos amostrados do cadastro de 1998, 107.546 (99,6%) pertenciam aos estratos amostrados no ano anterior; das 769 ULs dos estratos certos gerenciais, 604 (78,5%) também pertenciam a esse tipo de estrato em 1997, enquanto das 954 ULs do

estrato certo, 851 (89,2%) também eram do estrato certo em 1997. Por outro lado, 26.618 ULs que estavam no cadastro de 1997 não aparecem no cadastro de 1998; é interessante notar que essas mortes correspondem aproximadamente a 20% do total de ULs, o que é um indicativo que a folga de 20% dada no tamanho da amostra é razoável.

Tabela 22
Remanejamento das ULs comuns aos 2 cadastros

		Cadastro 1997			Total
		Amostrado	Gerencial	Certo	
Cadastro 1998	Amostrado	107.546	210	215	107.971
	Gerencial	138	604	27	769
	Certo	93	10	851	954

O tamanho da amostra inicial da pesquisa, baseada no cadastro de 1997, é de 5.938 ULs, das quais 1.196 pertencem ao estrato certo, 913 aos estratos certos gerenciais e 3.829 aos estratos amostrados. O tamanho da amostra para a segunda rodada é de 5.870 ULs, das quais 1.031 pertencem ao estrato certo, 877 aos estratos certos gerenciais e 3.962 aos estratos amostrados (ver tabela 23).

Tabela 23
Distribuição das amostras e do número de estratos segundo o tipo de estrato

	Tipo de estrato			Total
	Amostrado	Gerencial	Certo	
1ª amostra (1997)	3.829 (472)	913 (165)	1.196 (151)	5.938
2ª amostra (1998)	3.962 (479)	877 (158)	1.031 (153)	5.870

O mecanismo de rotação se aplica à parte dos 479 estratos amostrados e vamos considerar a rotação distribuída por quadrimestre (3 instantes de tempo no ano). Assim, a cada quadrimestre teremos uma “nova” amostra de 3.962 ULs, mais as 1.908 (877 + 1.031) ULs dos estratos certos. O tempo esperado de permanência das ULs na amostra considerado nesse exemplo foi de 3 anos ($T=3$).

Com relação às ULs que entrarão na amostra ao longo de todo o ano (1º, 2º ou 3º quadrimestre), temos a seguinte situação, ilustrada na figura 8: 2.572 (118 + 74 + 1.314 + 432 + 290 + 192 + 152) ULs são novas e 4.139 (408 + 241 + 1.807 + 94 + 25 + 685 + 879) já pertenciam à amostra anterior. Da amostra anterior, por sua vez, 1.799 (30,1%) deixam de participar do painel.

Vamos analisar, agora, as amostras de períodos consecutivos, para mensurar o impacto de superposição de entrevistas, caso seja necessário manter a amostra antiga até a adaptação dos novos informantes (ver observação 4). Na figura 9 temos os resultados dessas análises. O impacto forte, com maior taxa de sobreposição de amostras, ocorre no momento da troca de cadastros, ou seja, no primeiro quadrimestre, quando deverão ser visitadas 1.850 ULs a mais.

Se a rotação fosse feita em um único instante, teríamos a situação ilustrada na figura 10.

Nas figuras 11 e 12 ilustra-se o procedimento de rotação trimestral, tomando como 4 anos o tempo esperado de permanência das ULs na amostra ($T=4$). Novamente, o maior impacto é no momento da renovação do cadastro. Nos outros instantes, há uma maior distribuição da sobrecarga de trabalho, o que é mais conveniente em termos operacionais, uma vez que há uma maior uniformidade no volume de trabalho.

Analisados os resultados anteriores, decidiu-se optar por esse último esquema de rotação, caracterizado pelos seguintes parâmetros: tempo esperado de permanência na amostra $T=4$ anos e rotação distribuída ao longo dos 4 trimestres. A escolha do primeiro parâmetro se baseou principalmente no fato de que, a cada ano, já há uma rotação natural em função da mudança do cadastro. A rotação trimestral foi escolhida para se ter uma distribuição mais regular da carga de trabalho da equipe de coleta.

Figura 8
 ULs das amostras – rotação quadrimestral
 Tempo esperado de permanência na amostra: 3 anos

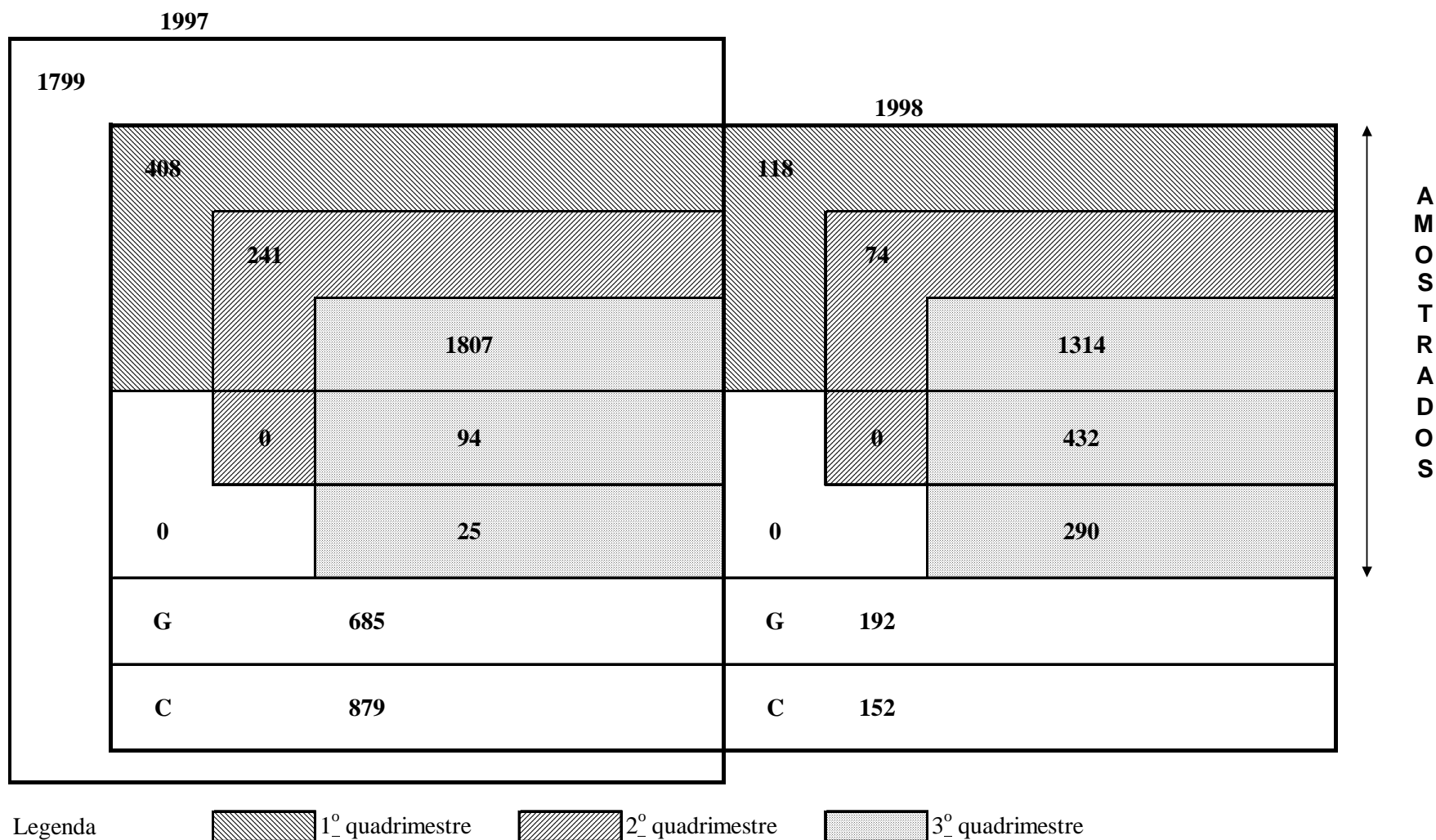


Figura 9

ULs das amostras para rotação efetuada quadrimestralmente

Tempo esperado de permanência na amostra: 3 anos

Amostra 1997: 5938

1918 (1799 + 94 + 25)	Amostra 1-1998: 5870	
2456 (408 + 241 + 1807)	1506 (118 + 74 + 1314)	A
685	192	G
879	152	C

Número de ULs a mais

$$152+192+1506 = 1850$$

Amostra 1-1998: 5870

526 (408 + 118)	Amostra 2-1998: 5870	
3436 (241 + 1807 + 74 + 1314)	526 (432 + 94)	A
877 (685 + 192)	0	G
1031 (879 + 152)	0	C

Número de ULs a mais

$$526$$

Amostra 2-1998: 5870

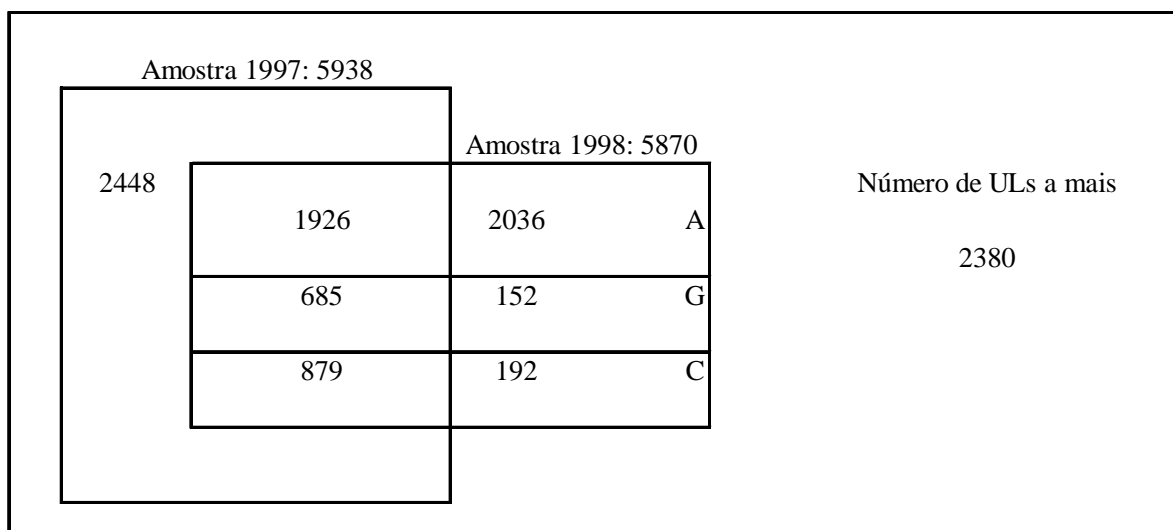
315 (241 + 74)	Amostra 3-1998: 5870	
3647 (1807 + 94 + 1314 + 432)	315 (290 + 25)	A
877	0	G
1031	0	C

Número de ULs a mais

$$315$$

Figura 10

ULs das amostras para rotação efetuada em um único momento
Tempo esperado de permanência na amostra: 3 anos



Para auxiliar a compreensão do processo de rotação, vamos considerar dois estratos finais específicos, correspondentes à região Nordeste exclusive Ceará, Pernambuco e Bahia, classe de PO de 5 a 30 e atividades Metalurgia Básica e Produtos de Metal, exclusive máquinas e equipamentos, respectivamente. Nas figuras 13 e 14 temos o esquema da população e os tamanhos das amostras nos dois estratos, bem como o posto das ULs que devem entrar na amostra em cada trimestre de 1998. Nas tabelas 24 e 25 temos a relação das ULs dos dois cadastros e das amostras. Da primeira pode-se ver que uma empresa nova, correspondente à observação 49, não entra na amostra, devido ao fato de o seu NAP ser menor que o da primeira observação a ser incluída na amostra. Da segunda tabela, constata-se que uma empresa nova, correspondente à observação 107, entra na amostra por apenas um trimestre. Embora seja uma situação operacionalmente inconveniente, não há maneira de evitar esse tipo de problema, dada a aleatoriedade dos NAPs.

Figura 11
 ULs das amostras – Rotação trimestral – Tempo esperado de permanência na amostra: 4 anos

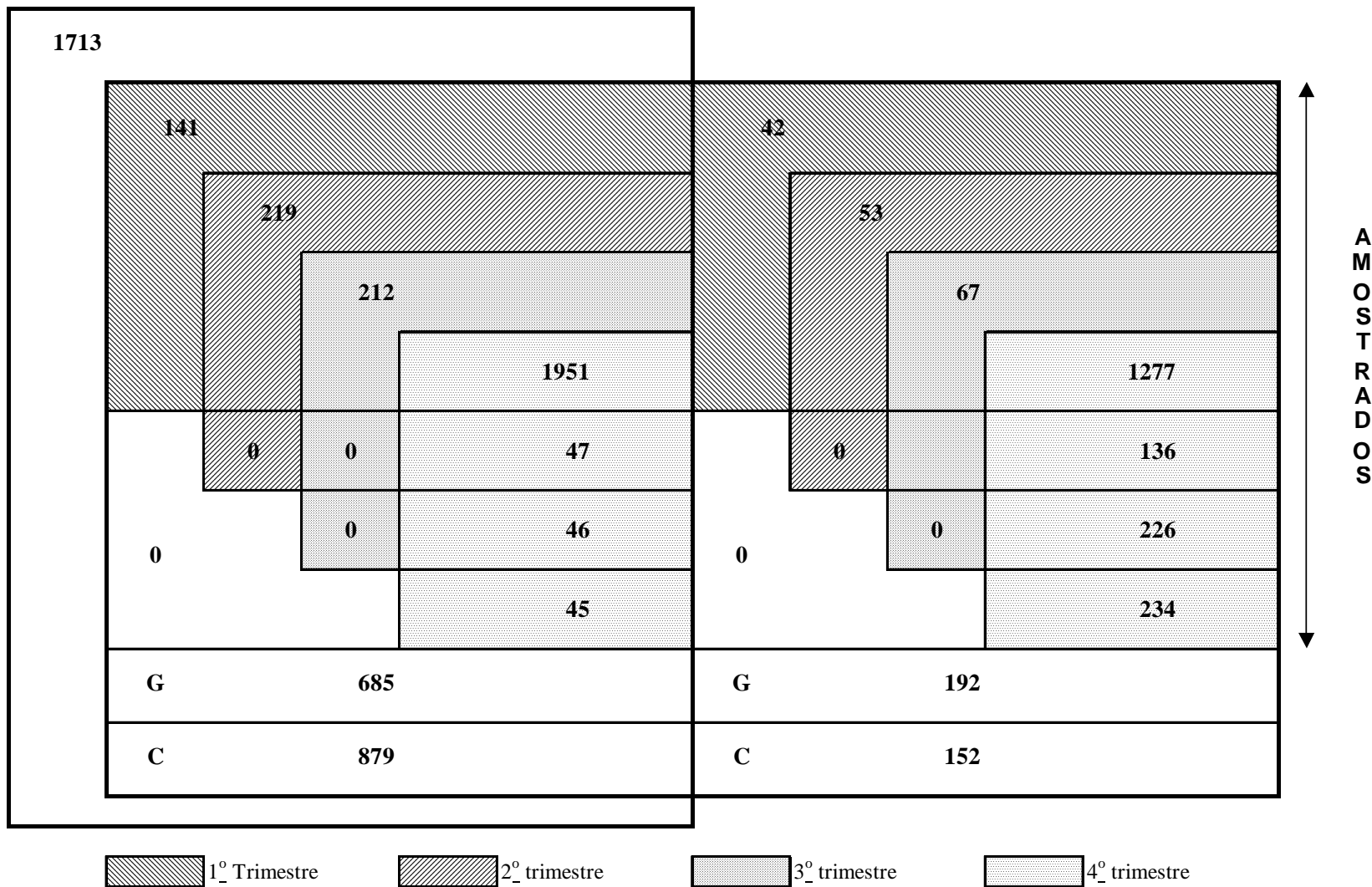


Figura 12
 ULs das amostras para rotação efetuada trimestralmente
 Tempo esperado de permanência na amostra: 4 anos

Amostra 1997: 5938			
1851 (1713 + 47 + 46 + 45)		Amostra 1-1998: 5870	
2523 (141 + 219 + 212 + 1951)		1439	A
685		192	G
879		152	C
			Número de UL's a mais 1783 (1439+192+152)
Amostra 1-1998: 5870			
183 (141 + 42)		Amostra 2-1998: 5870	
3779 (219+212+1951+53+67+1277)		183	A
877		0	G
1031		0	C
			Número de UL's a mais 183
Amostra 2-1998: 5870			
272 (219+53)		Amostra 3-1998: 5870	
3690 (212+1951+47+67+1277+136)		272	A
877		0	G
1031		0	C
			Número de UL's a mais 272
Amostra 3-1998: 5870			
279 (212+67)		Amostra 4-1998: 5870	
3683 (1951+47+46+1277+136+226)		279	A
877		0	G
1031		0	C
			Número de UL's a mais 279

Figura 13
Nordeste exclusive CE,PE,BA – Metalurgia Básica – [5,30)

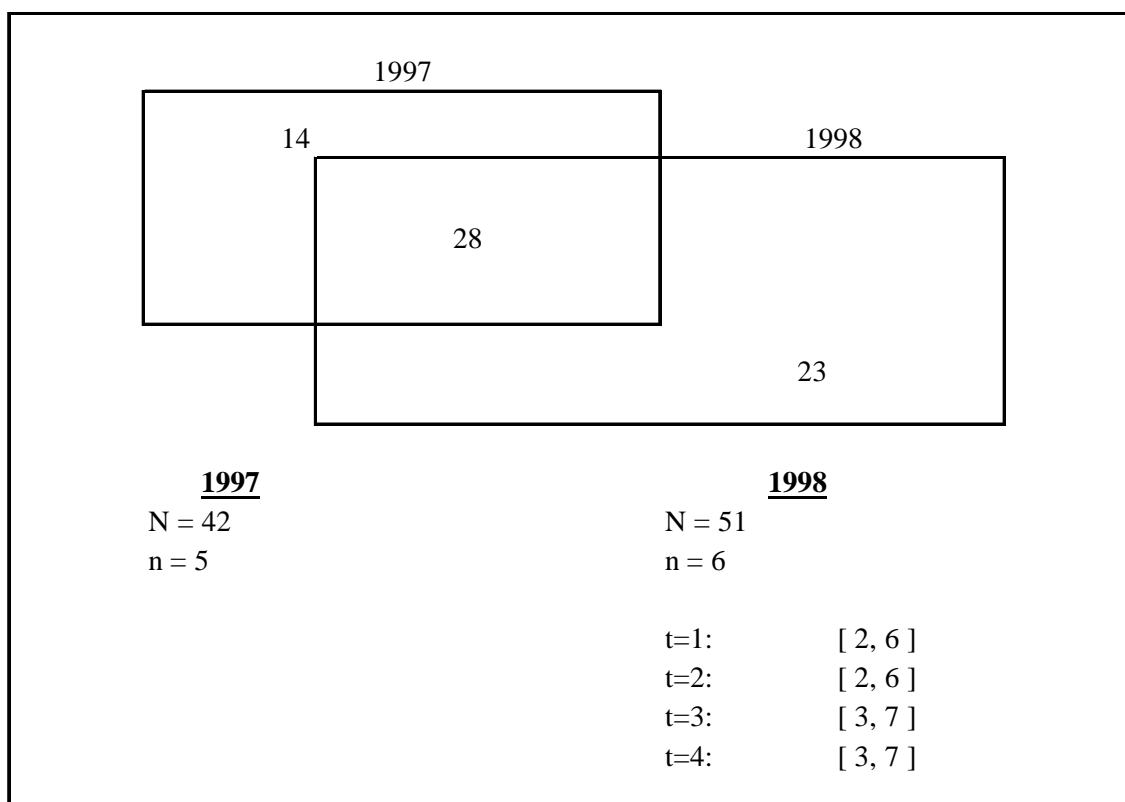


Figura 14
Nordeste exclusive CE,PE,BA – Prod. Metálicos excl. Máq.e Equip. – [5,30)

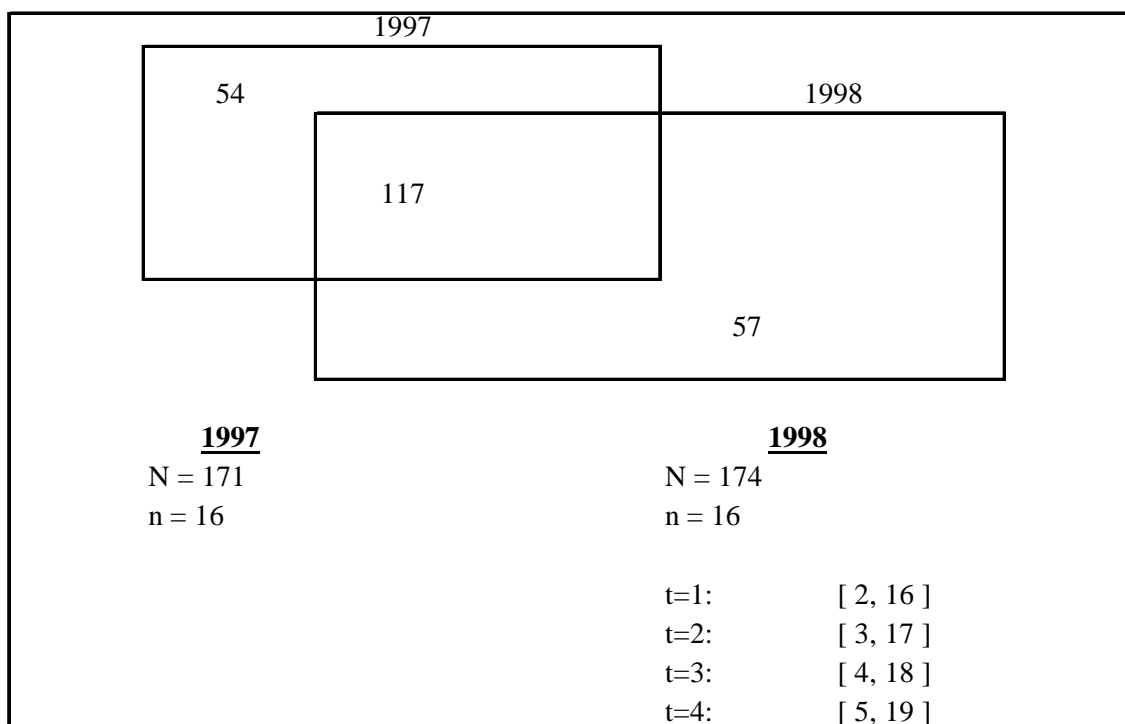


Tabela 24
Nordeste exclusive CE,PE,BA – Metalurgia Básica – [5,30]

Obs	Empresa	UL	Aleat	Obs	Empresa	UL	Aleat
60	69577666	1	0,01783	49	24366361	1	0,02733
58	63410427	1	0,03357	5	887309	1	0,04920
61	70041082	1	0,06845	56	41145798	1	0,06201
50	35141381	1	0,07481	61	70041082	1	0,06845
38	10937886	1	0,07769	18	2166039	1	0,07480
7	1005101	1	0,11205	50	35141381	1	0,07481
22	5804836	1	0,15103	4	799182	1	0,09574
26	8223687	2	0,17272	22	5804836	1	0,15103
15	1950539	1	0,18073	26	8223687	2	0,17272
47	23600430	1	0,21119	47	23600430	1	0,21119
30	8490716	1	0,24135	13	1759484	1	0,27749
44	12924452	1	0,24661	23	5804836	2	0,28826
52	35581529	1	0,26291	9	1324669	1	0,31278
13	1759484	1	0,27749	39	11889292	1	0,35841
9	1324669	1	0,31278	17	2165005	1	0,39565
8	1117740	1	0,34608	63	70323944	1	0,40327
54	40982076	1	0,35245	31	9093386	1	0,40456
39	11889292	1	0,35841	42	12574927	1	0,42890
31	9093386	1	0,40456	45	13414156	1	0,44915
24	7159809	1	0,42605	65	97464689	1	0,45336
42	12574927	1	0,42890	34	10288892	1	0,45392
45	13414156	1	0,44915	64	97432579	1	0,45412
65	97464689	1	0,45336	14	1794492	1	0,46096
34	10288892	1	0,45392	16	2021305	1	0,49123
64	97432579	1	0,45412	36	10744209	1	0,50791
43	12636429	1	0,45448	21	2891966	1	0,52760
2	386643	1	0,49374	59	69385227	1	0,52771
40	11895190	1	0,54029	48	24102360	1	0,53481
37	10745255	1	0,55264	19	2428701	1	0,53873
3	526052	1	0,60821	40	11895190	1	0,54029
28	8424210	1	0,62726	37	10745255	1	0,55264
6	993944	1	0,64417	25	7628886	1	0,58509
62	70111521	1	0,65201	3	526052	1	0,60821
35	10306322	1	0,75828	28	8424210	1	0,62726
32	9222928	1	0,77806	6	993944	1	0,64417
46	23513013	1	0,78159	62	70111521	1	0,65201
29	8471476	1	0,84029	32	9222928	1	0,77806
53	40924227	1	0,85682	46	23513013	1	0,78159
33	9364977	1	0,86906	10	1422866	1	0,78210
55	41002882	1	0,89812	1	207452	1	0,78967
51	35361351	1	0,92118	20	2638660	1	0,83626
27	8341752	1	0,96743	29	8471476	1	0,84029
				12	1748000	1	0,84904
				53	40924227	1	0,85682
				33	9364977	1	0,86906
				55	41002882	1	0,89812
				51	35361351	1	0,92118
				41	12171716	1	0,92668
				27	8341752	1	0,96743
				57	41199886	1	0,96875
				11	1610480	1	0,98626

Legenda

	Cadastro 97
	Cadastro 98
	Cadastros 97 e 98

Tabela 25
Nordeste exclusive CE,PE,BA – Prod. de Metal, excl. máq. e Equip. – [5,30)

Obs	Empresa	UL	Aleat	Obs	Empresa	UL	Aleat
116	10952638	1	0,00587	116	10952638	1	0,00587
199	41188871	1	0,01095	107	10330264	1	0,00713
83	8518433	1	0,01503	199	41188871	1	0,01095
226	74030354	1	0,01538	224	70132170	1	0,01183
112	10751253	1	0,01559	31	1281002	1	0,01193
201	41499286	1	0,01613	83	8518433	1	0,01503
152	23502404	1	0,01717	112	10751253	1	0,01559
178	32892648	1	0,01829	201	41499286	1	0,01613
147	15595242	1	0,02025	152	23502404	1	0,01717
69	5806070	1	0,02262	211	69577666	1	0,01783
127	12513248	1	0,02297	54	1962146	1	0,01943
221	70102025	1	0,02347	59	2231854	1	0,01999
183	35370584	1	0,03240	147	15595242	1	0,02025
159	24218505	1	0,03789	69	5806070	1	0,02262
39	1679962	1	0,05788	127	12513248	1	0,02297
202	41518408	1	0,07205	221	70102025	1	0,02347
9	485473	1	0,08192	183	35370584	1	0,03240
173	32839318	1	0,08434	159	24218505	1	0,03789
90	8822108	1	0,09403	39	1679962	1	0,05788
228	86779741	1	0,10156	143	15056021	1	0,05978
145	15121817	1	0,10317	62	2703480	1	0,06341
129	12740064	1	0,10492	9	485473	1	0,08192
81	8467110	1	0,12362	90	8822108	1	0,09403
191	40761652	1	0,13576	228	86779741	1	0,10156
114	10835536	1	0,13624	145	15121817	1	0,10317
141	13382619	1	0,14099	55	2105317	1	0,10438
130	12827952	1	0,14256	129	12740064	1	0,10492
97	9169046	1	0,14517	50	1925398	1	0,10647
36	1517563	1	0,14589	63	2800037	1	0,11807
48	1843210	1	0,15701	195	41005208	1	0,12609
66	5489869	1	0,16533	108	10330777	1	0,13247
103	9280116	1	0,16722	114	10835536	1	0,13624
162	24306862	1	0,17010	141	13382619	1	0,14099
153	23505373	1	0,17391	36	1517563	1	0,14589
179	35149632	1	0,17788	142	13946272	1	0,14729
64	5233705	1	0,18622	35	1503052	1	0,15437
96	9135419	1	0,19473	48	1843210	1	0,15701
217	70009808	1	0,21368	66	5489869	1	0,16533
67	5633623	1	0,21419	103	9280116	1	0,16722
192	40775280	1	0,21752	45	1755670	1	0,1681
109	10332880	1	0,219	17	864988	1	0,16821
220	70092044	1	0,22128	162	24306862	1	0,1701
77	8209942	1	0,22601	153	23505373	1	0,17391
210	69424018	1	0,22782	179	35149632	1	0,17788
...				...			

Legenda

	Cadastro 97
	Cadastro 98
	Cadastros 97 e 98

Na tabela 26 apresenta-se o número de empresas envolvidas nas amostras, considerando as possibilidades de sobrepor ou não os velhos e novos informantes. Como já visto, o impacto forte ocorre no momento da troca de cadastro mas, mesmo assim, o número de empresas envolvidas não é muito grande, sendo possível o processamento nas Divisões de Pesquisas (DIPEQ) regionais. Como essa sobreposição torna maior a chance de completude da amostra, decidiu-se por antecipar em 1 mês a coleta de dados dos novos informantes.

Tabela 26
Número de empresas envolvidas nas amostras

UF	Am. 97	Sem sobreposição				Com sobreposição			
		Trim. 1	Trim. 2	Trim. 3	Trim. 4	Trim. 1	Trim. 2	Trim. 3	Trim. 4
11	26	24	24	24	24	34	25	25	24
12	4	4	4	5	4	5	4	5	5
13	82	82	81	83	80	108	83	85	85
15	60	63	61	62	62	83	63	63	64
16	2	1	1	1	2	2	1	1	2
17	4	9	8	8	8	10	9	8	8
21	73	63	63	63	63	90	64	67	34
22	40	43	44	47	47	55	45	48	50
23	340	341	341	342	341	439	351	353	359
24	78	84	81	79	82	108	84	86	86
25	101	96	98	98	95	128	100	103	101
26	342	325	325	323	324	426	335	337	343
27	72	70	70	71	71	87	71	72	73
28	38	37	38	38	37	49	38	38	39
29	323	327	327	327	329	445	335	340	351
31	483	483	484	482	481	636	501	513	504
32	277	290	290	290	288	375	299	305	306
33	450	423	425	425	426	571	439	449	447
35	799	754	754	754	755	981	778	785	779
41	446	446	445	447	444	577	464	471	467
42	406	416	418	418	417	523	431	442	437
43	504	498	499	502	506	642	515	527	528
50	36	38	40	40	41	52	41	42	42
51	55	62	64	62	62	76	65	68	64
52	138	145	141	141	148	191	150	150	157
53	30	33	33	32	29	45	33	33	32
Total Brasil	5.209	5.157	5.159	5.164	5.166	6.738	5.324	5.416	5.417

9 Conclusão

Nesse texto apresentou-se o estudo realizado para definir o desenho amostral da Pesquisa Industrial Mensal de Emprego e Salário a ser realizada pelo Departamento de Indústria do IBGE. O desenho amostral acordado é o de amostragem aleatória estratificada com alocação de Neyman nos estratos finais. Os estratos naturais são definidos pelos cruzamentos de Região Geográfica e Atividade Econômica e os estratos finais são definidos pelo número de pessoas ocupadas. Todas as ULs com 500 ou mais pessoas ocupadas entram na amostra com probabilidade um (estrato certo) e na parte amostrada de cada estrato natural toma-se uma amostra de modo a garantir um coeficiente de variação de 10% para o estimador do total da variável de análise, que é o Pessoal Ocupado (PO).

As regiões geográficas consideradas são: Norte e Centro-Oeste, Nordeste exclusive Ceará, Pernambuco e Bahia, Ceará, Pernambuco, Bahia, Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul.

As atividades econômicas são definidas pelos seguintes agrupamentos da CNAE a 2 dígitos: Indústria Extrativa (10+11+13+14), Alimentos e Bebidas (15), Fumo (16), Têxteis (17), Vestuário (18), Couros (19), Madeira (20), Papel e Gráfica (21+22), Coque e Petróleo (23), Prod. Químicos (24), Borracha e plástico (25), Minerais não metálicos (26), Metalurgia básica (27), Produtos de metal, exclusive máquinas e equipamentos (28), Máquinas e equipamentos, exclusive eletrônicos, elétricos, de precisão e de comunicação (29+30), Equipamentos elétricos, eletrônicos, de precisão e de comunicação (31+32+33), Meios de transporte (34+35), Outros produtos da ind. de transformação (36+37).

O tamanho mínimo da amostra em cada estrato final foi definido como 5 e, assim, os estratos finais com população menor que 5 foram definidos como estratos certos gerenciais. Todas as ULs da divisão Fumo também foram incluídas com certeza na amostra, definindo novos estratos certos gerenciais. Para contornar problemas devidos à desatualização cadastral, o tamanho da amostra em cada estrato natural amostrado foi aumentado em 20%. Finalmente, definiu-se como estrato certo gerencial todo aquele em que a diferença entre os tamanhos da população e da amostra fosse menor que 5.

A seleção e a rotação da amostra serão feitas com base nos Números Aleatórios Permanentes. A cada ano a amostra será rotacionada, de modo que cada UL permaneça, em média, 4 anos no painel. A substituição das ULs, devida ao mecanismo de rotação, será feita trimestralmente, mantendo-se os velhos e novos informantes simultaneamente por 1 mês.

Referências

- FARIAS, A.M.L. (1999) *Relatório das Atividades de Consultoria*, Relatório Interno DEIND/IBGE, dezembro.
- HIDIROGLOU, M.A. (1986) The construction of a self-representing stratum of large units in survey design. *The American Statistician*, **40**, 27-31.
- LAVALLÉE, P. e HIDIROGLOU, M.A. (1988) On the stratification of skewed populations. *Survey Methodology*, **14**, 33-43.
- OHLSSON, E. (1995) Coordination of Samples using Permanent Random Numbers. In Cox, Binder, Chinnappa, Christianson, Colledge e Kott (eds.) *Business Survey Methods*, New York, Wiley, p. 153-169.
- SILVA, P.LN. et al. (1998) *Planejamento Amostral para as Pesquisas Anuais da Indústria e do Comércio*. Textos para Discussão nº 92, outubro de 1998, Rio de Janeiro: IBGE.

APÊNDICE

Limites superiores dos estratos – Lavallée-Hidiroglou (1988)

CV=10% - Alocação de Neyman

Tabela A1

Limites superiores do estrato 1 – Proposta inicial para os estratos naturais

CNAE-2	Norte e Centro-Oeste	Nordeste	Minas Gerais	Espírito Santo	Rio de Janeiro	São Paulo	Sul
10							20
11							
13	14	10	34		14	11	8
14	12	19	15	11	14	14	13
15	28	54	25	18	28	41	40
16		20					25
17	13	34	32	11	18	41	30
18	13	19	16	13	18	22	25
19	13	26	16	10	14	25	41
20	23	15	10	10	11	18	24
21	13	20	21		19	34	34
22	13	15	12	8	22	28	20
23	52	71	14			53	40
24	13	29	19	8	29	41	27
25	16	23	17	10	26	35	29
26	16	21	18	14	16	24	19
27	16	15	43	11	23	31	18
28	12	13	22	10	17	28	21
29	14	24	21	11	18	40	31
30	12	11	10			22	9
31	12	18	19		16	42	34
32	46	9	14		16	37	12
33	11	12	10		15	25	15
34	11	11	41	9	16	73	33
35	15	14	10		11	22	13
36	14	15	16	11	14	23	22
37		6	9		7	9	8
Média	17,48	20,96	19,33	11,00	17,36	30,79	23,50
Mediana	13	18	16,5	11	16	28	23
Mínimo	11	6	9	8	7	9	8
Máximo	52	71	43	18	29	73	41
Desvio padrão	10,70	14,30	9,50	2,51	5,46	14,14	10,01

Tabela A2
Limites superiores do estrato 2 – Proposta inicial para os estratos naturais

CNAE-2	Norte e Centro-Oeste	Nordeste	Minas Gerais	Espírito Santo	Rio de Janeiro	São Paulo	Sul
10							77
11							
13	59	50	164		42	28	18
14	29	60	43	28	40	40	43
15	154	580	131	70	156	267	272
16		70					108
17	45	200	176	25	75	239	180
18	38	82	60	37	76	103	138
19	36	124	55	24	41	117	227
20	87	52	26	23	31	71	118
21	47	72	79		63	163	156
22	45	64	38	17	110	156	94
23	251	551	65			243	192
24	42	129	65	18	124	231	120
25	45	83	54	21	114	170	124
26	50	71	66	39	52	116	95
27	62	68	212	43	128	147	73
28	35	39	102	29	68	137	105
29	36	94	78	31	69	229	168
30	28	28	25			70	22
31	31	67	72		64	238	167
32	276	23	42		46	183	38
33	29	37	26		43	98	58
34	25	26	256	16	43	504	187
35	65	43	25		33	87	35
36	46	44	58	28	44	109	103
37		10	19		15	19	17
Média	67,87	106,68	80,71	29,93	67,14	156,88	112,88
Mediana	45	67	62,5	28	57,5	142	106,5
Mínimo	25	10	19	16	15	19	17
Máximo	276	580	256	70	156	504	272
Desvio padrão	67,44	143,74	62,84	13,65	36,94	103,51	67,24

Tabela A3

Limites superiores do estrato 3 – Proposta inicial para os estratos naturais

CNAE-2	Norte e Centro-Oeste	Nordeste	Minas Gerais	Espírito Santo	Rio de Janeiro	São Paulo	Sul
10							265
11							
13	287	285	636		128	73	41
14	102	212	171	95	151	181	164
15	882	3537	795	308	881	1741	1691
16		236					408
17	187	928	757	90	332	1242	951
18	167	517	338	147	432	712	866
19	115	644	250	73	152	615	1272
20	491	205	110	68	108	352	726
21	194	243	279		228	876	672
22	208	336	189	58	525	899	465
23	696	2019	326			824	866
24	189	502	291	56	548	1335	496
25	170	318	222	54	487	1021	607
26	225	346	344	158	235	752	617
27	255	344	1005	280	770	831	395
28	135	171	489	101	339	938	628
29	112	339	327	110	297	1453	980
30	86	69	77			241	71
31	95	284	299		276	1153	707
32	1045	65	142		145	810	191
33	112	118	95		151	427	237
34	77	80	1478	36	167	2761	942
35	287	132	85		166	385	108
36	198	197	302	95	209	695	646
37		24	51		52	60	46
Média	274,57	486,04	377,42	115,27	308,14	849,04	579,15
Mediana	189	284	295	95	231,5	817	612
Mínimo	77	24	51	36	52	60	41
Máximo	1045	3537	1478	308	881	2761	1691
Desvio padrão	259,27	750,61	339,26	80,03	217,81	594,54	402,14

Tabela A4

Limites superiores do estrato 1 – Proposta alternativa para os estratos naturais

Gupo da CNAE	Norte e Centro-Oeste	Nordeste excl. CE,PE,B A	CE	PE	BA	MG	ES	RJ	SP	PR	SC	RS
10+11+13+14	19	20	9	12	16	23	13	18	14	14	12	11
15+16	28	37	22	24	18	25	18	29	41	28	23	30
17	13	35	29	21	11	32	11	18	41	19	26	14
18	13	17	20	13	12	16	13	18	22	16	21	14
19	13	20	29	15	8	16	10	14	25	14	12	56
21+22	17	14	13	16	13	19	8	24	30	23	27	23
23	52			25		14			53			
24	13	18	12	17	31	19	8	29	41	17	12	24
25	16	14	15	18	12	17	10	26	35	18	29	21
26	16	18	12	17	18	18	14	16	24	15	16	14
27	16	7	7	11	11	43	11	23	31	13	20	15
28	12	11	13	10	10	22	10	17	28	12	15	23
29+30	19	9	21	13	15	21	11	18	40	30	25	26
31+32+33	26	9	11	14	9	19	9	19	42	21	22	20
34+35	17	14	11	12	8	38	11	20	69	18	18	32
20+36+37	22	16	11	11	12	16	11	15	23	24	23	17
Média	19,50	17,27	15,67	15,56	13,60	22,38	11,20	20,27	34,94	18,80	20,07	22,67
Mediana	16,5	16	13	14,5	12	19	11	18	33	18	21	21
Mínimo	12	7	7	10	8	14	8	14	14	12	12	11
Máximo	52	37	29	25	31	43	18	29	69	30	29	56
Desvio padrão	9,84	8,60	6,91	4,57	5,78	8,35	2,54	4,82	13,48	5,43	5,68	11,11

Tabela A5

Limites superiores do estrato 2 – Proposta alternativa para os estratos naturais

Gupo da CNAE	Norte e Centro-Oeste	Nordeste excl. CE,PE,B A	CE	PE	BA	MG	ES	RJ	SP	PR	SC	RS
10+11+13+14	76	61	21	29	53	114	34	67	42	36	34	26
15+16	153	281	113	147	86	131	70	158	267	139	122	171
17	45	203	153	80	34	176	25	75	239	83	141	50
18	38	78	83	39	30	60	37	76	103	56	97	49
19	36	83	150	42	17	55	24	41	117	44	39	304
21+22	71	47	42	52	45	89	19	108	170	114	123	112
23	251			121		65			243			
24	42	67	35	60	136	65	18	124	231	53	34	95
25	45	43	44	56	29	54	21	114	170	61	125	76
26	50	53	31	51	55	66	39	52	116	57	71	50
27	62	16	13	36	24	212	43	128	147	41	71	52
28	35	27	32	26	26	102	29	68	137	33	58	116
29+30	64	21	63	35	44	78	31	69	227	143	117	110
31+32+33	107	23	28	51	19	75	17	68	233	74	107	84
34+35	68	37	25	28	17	242	26	71	473	67	61	195
20+36+37	85	50	25	25	35	59	27	51	112	114	99	67
Média	76,75	72,67	57,20	54,88	43,33	102,69	30,67	84,67	189,19	74,33	86,60	103,80
Mediana	63	50	35	46,5	34	76,5	27	71	170	61	97	84
Mínimo	35	16	13	25	17	54	17	41	42	33	34	26
Máximo	251	281	153	147	136	242	70	158	473	143	141	304
Desvio padrão	55,67	73,11	46,30	34,49	31,43	58,76	13,39	33,58	99,31	36,50	36,23	72,35

Tabela A6

Limites superiores do estrato 3 – Proposta alternativa para os estratos naturais

Gupo da CNAE	Norte e Centro-Oeste	Nordeste excl. CE,PE,B A	CE	PE	BA	MG	ES	RJ	SP	PR	SC	RS
10+11+13+14	341	236	60	84	234	580	135	326	190	113	128	86
15+16	883	1680	635	1105	447	806	308	883	1749	723	787	931
17	187	821	693	275	136	757	90	332	1242	367	741	213
18	167	377	402	165	109	338	147	432	712	270	576	217
19	115	339	719	137	53	250	73	152	615	159	139	1439
21+22	318	196	158	192	190	428	94	520	1156	526	480	493
23	696			804		326			824			
24	189	201	124	209	429	291	56	548	1335	201	116	351
25	170	143	140	158	111	222	54	487	1021	245	482	326
26	225	205	112	200	174	344	158	235	752	488	405	238
27	255	66	43	131	120	1005	280	770	831	130	303	270
28	135	91	102	93	81	489	101	339	938	150	251	582
29+30	203	67	197	109	153	332	110	297	1456	588	587	543
31+32+33	502	69	92	230	81	326	47	304	1329	307	497	376
34+35	292	94	63	80	44	1435	69	302	2714	301	252	874
20+36+37	519	200	86	86	125	336	100	252	780	658	540	389
Média	324,81	319,00	241,73	253,63	165,80	516,56	121,47	411,93	1102,75	335,07	418,93	488,53
Mediana	240	200	124	161,5	125	341	100	332	979,5	288	480	376
Mínimo	115	66	43	80	44	222	47	152	190	113	116	86
Máximo	883	1680	719	1105	447	1435	308	883	2714	723	787	1439
Desvio padrão	218,46	422,14	243,85	284,82	121,49	331,26	77,62	201,22	571,67	197,72	214,65	353,51