

UMA ROTINA DE REDIMENSIONAMENTO DE AMOSTRAS PARA ESTIMATIVAS DE SAFRAS ATRAVÉS DO STATISTICAL ANALYSIS SOFTWARE (SAS)ⁱ

Lilian C. Anefalosⁱⁱ
José R. Vicente²
Lúcio Fagundes²

1 - INTRODUÇÃO

As decisões do poder público e de vários setores da economia podem ser fortemente influenciadas por variações nas quantidades produzidas de produtos agrícolas. Tanto o abastecimento interno, como as exportações e a estabilidade dos preços dependem da magnitude das safras agrícolas, e medidas destinadas a minorar eventuais efeitos indesejáveis dependem de boas estimativas sobre a produção.

Parte das estatísticas agrícolas do Estado de São Paulo baseia-se, desde meados da década de 50, em levantamentos efetuados por amostragemⁱⁱⁱ. Os dados são coletados diretamente junto a produtores rurais, e apresentam uma série de vantagens sobre os provenientes de outros tipos de levantamentos^{iv}.

Atualmente, questionários dos elementos que compõem a amostra são levados ao campo nos meses de setembro, novembro, fevereiro, abril e junho, e são preenchidos por técnicos da Coordenadoria de Assistência Técnica Integral (CATI), e o número de imóveis visitados é de 3.622.

As alterações na composição das explorações agrícolas e em sua distribuição espacial fazem com que a amostra necessite ser reestimada periodicamente^v. Além disso, o número de questionários não respondidos e os erros de preenchimento também costumam aumentar com o tempo em que a amostra permanece sem ser substituída^{vi}. Apesar dessa necessidade de substituição periódica, a última troca foi efetuada em 1986.

Várias são as dificuldades que vêm impedindo que essa substituição ocorra com a frequência desejada; a principal é a inexistência de cadastros novos com a qualidade necessária. Dentre os demais fatores, pode ser citada a carência de um programa de computador de execução simples e que permita as necessárias simulações de erros de amostragem e a introdução de novos produtos e regiões^{vii}.

O presente estudo teve como objetivo desenvolver e testar uma rotina, executável no

software estatístico SAS, para permitir a reestimação periódica de amostras para previsão de safras.

2 - METODOLOGIA

O programa proposto para efetuar redimensionamentos de amostras foi dividido em três partes^{viii}. Na parte I, a rotina classifica os dados já existentes dos imóveis rurais, por estrato de área^{ix} e por Divisão Regional Agrícola (DIRA)^x; basicamente, áreas (ou pés) das culturas selecionadas e os respectivos fatores de expansão^{xi}. Em seguida, são efetuados os cálculos das médias das áreas (ou pés) e de suas respectivas variâncias, por produto, por estrato e por DIRA, empregando a PROC MEANS do SAS.

De posse desses valores, é possível estimar as áreas totais (expandidas) de cada cultura, por DIRA e por estrato:

$$T_{ijk} = \frac{N_{jk}}{n_{jk}} \sum x_{ijk} \quad (1)$$

onde, N_{jk} é o número de imóveis rurais existentes no k-ésimo estrato da j-ésima DIRA, n_{jk} é o número de elementos na amostra, no k-ésimo estrato da j-ésima DIRA, e, x_{ijk} é a área (ou o número de pés) da i-ésima cultura, informada em cada questionário, do k-ésimo estrato da j-ésima DIRA.

A variância total do estrato, considerando-se a correção para população finita, é dada por (CAMARGO, 1988):

$$s_{T_{ijk}}^2 = N_{jk} (N_{jk} - n_{jk}) s_{x_{ijk}}^2 \quad (2)$$

onde, $s_{x_{ijk}}^2$ é a variância da média das áreas (ou do número de pés), para a cultura i, DIRA j, e estrato k.

As áreas e as variâncias totais de cada cultura, para o Estado, são obtidas pela soma das respectivas parcelas dos estratos e das DIRAs, também efetuado através da PROC MEANS.

Com essas estimativas, a Parte II do programa efetua a primeira aproximação de amostras específicas para cada cultura, calculando, inicialmente,

os erros de amostragem da amostra anterior ou amostra piloto (e_i)⁴ e estimando o número de elementos n_{2i} ⁵, necessário para a obtenção do nível de erro e_2 ⁶ desejado (STEVENS, 1951):

$$n_{2i} = \frac{e_i}{e_2} \sqrt{\frac{S_{T_i}^2}{T_i}} \cdot n_i \quad (4)$$

onde, $S_{T_i}^2$ ⁹ é a variância total, T_i ¹⁰ é a área (ou número de pés) total para a i -ésima cultura e, n_i é o número de elementos da amostra anterior.

Esses resultados podem ser analisados pelo pesquisador interessado, que fixa como desejar erros de amostragem pretendidos para cada cultura. Esse passo pode ser suprimido, e a rotina irá dimensionar amostras para cada cultura, assumindo erros de amostragem de 1% a 10% do total das áreas (ou pés). Qualquer que seja a alternativa escolhida, o número de elementos (nas amostras por produto) é estimado, na Parte III do programa, por (KISH, 1965 e PINO, não publ.):

$$n_{2i} = \frac{n_i}{1 + \left(\frac{n_i}{N}\right)} \quad (5)$$

com

$$n_i = \frac{\sigma_i^2}{V_i^2} \quad (6)$$

onde V_i^2 ¹³ é a variância desejada para a área da i -ésima cultura, e σ_i^2 ¹⁴ é a variância dessa mesma área na população, que é geralmente desconhecida e pode ser calculada a partir da estimativa da variância na amostra anterior:

$$\sigma_i^2 = n'' \cdot S_{T_i}^2 \quad (7)$$

sendo,

$$n'' = \frac{n_i}{1 - \frac{n_i}{N}} \quad (8)$$

onde N é o número total de imóveis rurais. Após estimadas as amostras, por cultura, são alocadas nos estratos das DIRAs, proporcionalmente às suas contribuições para as variâncias totais, através da Partilha de Neyman (COCHRAN, 1963 e CAMPOS & PIVA, 1974):

$$n_{ijk} = n_{2i} \frac{N_{jk} \cdot Sx_{ijk}}{\sum_{j,k} N_{jk} \cdot Sx_{ijk}} \quad (9)$$

Por último, é calculada amostra conciliatória, segundo o(s) nível(is) de erro determinado(s), com base nos valores médios de todas as amostras (COCHRAN, 1963), por estrato e por DIRA.

Para testar a rotina acima descrita, foram utilizados como amostras piloto, dados do ano agrícola 1990/91, referentes a doze culturas: algodão, amendoim das águas e da seca, arroz, café, cana-de-açúcar, feijão das águas, da seca e de inverno, laranja, milho e soja. Além do dimensionamento com erros de amostragem entre 1% e 10%, para as áreas colhidas de todas essas culturas, tentou-se uma amostra conciliatória combinando diversos níveis de erros. Como tamanho de amostra possível de ser levantada, definiu-se o número de, aproximadamente, 5.000 elementos. O sistema de referência empregado foi o cadastro do INCRA, composto por 305.704 imóveis rurais acima de 3,0 hectares, atualizado até 1985, que também é a base da amostra que vem sendo levada a campo.

3 - RESULTADOS E DISCUSSÃO

Para o ano agrícola 1990/91, os erros de amostragem obtidos nos cálculos finais das áreas das culturas consideradas foram: algodão 10,6%; amendoim das águas 18,5%; amendoim da seca 24,3%; arroz 15,2%; café 8,7%; cana-de-açúcar 7,4%; feijão das águas 11,1%; feijão da seca 15,8%; feijão de inverno 14,4%; laranja 8,1%; milho 4,9%; e soja 8,6%.

Através da aproximação descrita em (3), observa-se que amostras com erros de 1% para o total de área colhida (ou pés) só seriam possíveis para cana-de-açúcar, milho, soja, café e laranja (Tabela 1); os outros produtos teriam que ser levantados censitariamente, já que as amostras necessárias para tal nível de erro seriam superiores ao tamanho da população (mais de 305.000 imóveis). Na prática, todavia, os tamanhos de todas essas amostras possíveis tornam-as inviáveis.

Para as culturas do algodão e do feijão das águas, amostras com erros de 5% para as áreas colhidas deveriam ter aproximadamente 13.000 elementos, enquanto que amostras da dimensão da atualmente levada a campo propiciariam erros de

amostragem por volta de 10%. O amendoim, tanto das águas como da seca, devido à grande retração de área observada em relação aos níveis do início da década de 70, tornou-se problemático para levantamentos amostrais: erros da ordem de 10% para a área colhida necessitariam de amostras entre, aproximadamente, 10 e 17 mil elementos. Processo de retração de área semelhante ocorreu com a cultura do arroz, o que obrigaria o levantamento de quase 7 mil imóveis para erro de amostragem próximo a 10%; amostras pouco maiores que as do arroz levariam aos mesmos níveis de erros para o feijão da seca. Com a cana-de-açúcar, o aumento de área ocorrido durante as décadas de 70 e 80 fez com que fossem suficientes cerca de 6 mil elementos na amostra para a obtenção de erros da ordem de 5%; com 1,5 mil imóveis seria possível a obtenção de erros de 10%. O milho, que era o produto de distribuição mais uniforme pelo Estado, necessitaria de menos de 3.000 imóveis amostrados para obter-se erro de 5% para a área colhida; com cerca de 700 elementos uma amostra já deveria permitir erros de 10% para essa cultura. Para a soja, erros de 5% só seriam possíveis com amostras de 8,5 mil imóveis, sendo necessários cerca de 2,1 mil para erros máximos de 10%, sempre considerando a área colhida; amostras dessa grandeza também seriam necessárias para os mesmos níveis de erro considerando-se o número de pés de café em produção. O feijão de inverno, que ganhou importância em São Paulo desde o final da década de 70, e apresenta área oscilando em torno de 100 mil hectares, obrigaria a levantar mais de 22 mil imóveis rurais para a obtenção de erros de amostragem de 5% e quase 6 mil para 10% de erro. A laranja, apesar do grande aumento de área observado a partir do início da década de 70, ainda necessita de amostras da ordem de 7 mil elementos para erros de amostragem de 5% para os pés em produção, enquanto que menos de 2 mil imóveis seriam suficientes para erros de 10%.

TABELA 1 - Número Aproximado de Elementos de Amostras por Produto, para Vários Níveis de Erro¹

Produto	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Algodão	322.575	80.644	35.842	20.161	12.903	8.960	6.583	5.040	3.982	3.226
Amendoim das águas	974.586	243.647	108.287	60.912	38.983	27.072	19.890	15.228	12.032	9.746
Amendoim da seca	1.688.619	422.155	187.624	105.539	67.545	46.906	34.462	26.385	20.847	16.886
Arroz	660.529	165.132	73.392	41.283	26.421	18.348	13.480	10.321	8.155	6.605
Cana-de-açúcar	147.405	36.851	16.378	9.213	5.896	4.095	3.008	2.303	1.820	1.474

Adotando-se o critério de 5% de erro para as áreas (ou pés) de todas as culturas consideradas, a amostra conciliatória resultante teria cerca de 18 mil elementos (Tabela 2). AS DIRAs de Sorocaba (DIRA 4) e São José do Rio Preto (DIRA 8) teriam mais de 3.000 elementos de amostragem e o estrato 7 (30,1 a 50,0 hectares) teria mais de 4.000. Se o nível de erro fosse elevado para 10%, para todas as culturas, o número de imóveis da amostra seria de 5.087, com as duas DIRAs citadas tendo cerca de 900 elementos e o estrato 7 (imóveis de 30,1 a 50,0 hectares), mais de 1.150 (Tabela 3)^{xii}.

Partindo de um erro de amostragem máximo de 10% para a área (ou pés) colhida de cada cultura, ou procurando-se amostras individuais com cerca de 5 a 6 mil elementos (onde isso foi possível com erro menor que 10%), efetuaram-se reestimativas, partilhas e amostras conciliatórias, obtendo-se amostra com 6.358 elementos (Tabela 4). Outra simulação foi efetuada eliminando-se dos cálculos a cultura do amendoim da seca (Tabela 5) e das águas e da seca (Tabela 6). Nesse último caso, a amostra conciliatória obtida ficou com cerca de 5.026 elementos, com as DIRAs de Sorocaba (1.177) e Campinas (1.012) recebendo o maior número de elementos; entre os estratos, o único com mais de 1.000 elementos foi o de número 8 (imóveis de 50,1 a 100,0 hectares).

4 - RESUMO, CONCLUSÕES E CONSIDERAÇÕES FINAIS

Foi desenvolvida uma rotina em SAS capaz de efetuar redimensionamentos de amostras para previsão e estimativa de safras.

Uma amostra conciliatória com 5.026 elementos foi obtida, para erros de amostragem entre, aproximadamente, 4% e 10% para as áreas (ou pés) totais de dez das principais culturas do Estado de

Feijão das águas	333.192	83.298	37.021	20.824	13.328	9.255	6.800	5.206	4.113	3.332
Feijão da seca	710.407	177.602	78.934	44.400	28.416	19.734	14.498	11.100	8.770	7.104
Milho	69.775	17.444	7.753	4.361	2.791	1.938	1.424	1.090	861	698
Soja	213.098	53.274	23.678	13.319	8.524	5.919	4.349	3.330	2.631	2.131
Feijão de inverno	560.665	140.166	62.296	35.042	22.427	15.574	11.442	8.760	6.922	5.607
Café	216.219	54.055	24.024	13.514	8.649	6.006	4.413	3.378	2.669	2.162
Laranja	178.957	44.739	19.884	11.185	7.158	4.971	3.652	2.796	2.209	1.790

¹Cálculo efetuado através do método descrito em STEVENS (1951).

Fonte: Resultados da pesquisa.

TABELA 2 - Distribuição da Amostra Conciliatória com 5% de Erro de Amostragem para os Doze Produtos

Estrato	DIRA										Total
	2	3	4	5	6	7	8	9	10	11	
3	3	19	10	25	3	2	3	2	34	2	103
4	14	4	343	261	123	9	57	23	56	7	897
5	25	4	930	476	144	37	1.166	818	174	66	3.839
6	78	431	395	307	421	12	468	360	222	89	2.783
7	5	26	414	300	231	111	939	33	1.148	957	4.163
8	384	64	828	299	540	47	339	37	235	293	3.065
9	2	15	279	196	445	35	357	36	81	242	1.688
10	2	2	36	26	187	9	41	12	17	75	407
11	2	28	49	46	414	12	17	19	16	22	626
12	2	13	30	17	192	139	16	12	61	67	549
13	2	2	93	8	73	7	15	24	65	68	358
Total	519	607	3.410	1.960	2.774	419	3.416	1.376	2.108	1.887	18.476

Fonte: Resultados da pesquisa.

TABELA 3 - Distribuição da Amostra Conciliatória com 10% de Erro de Amostragem para os Doze Produtos

Estrato	DIRA										Total
	2	3	4	5	6	7	8	9	10	11	
3	2	6	4	8	2	2	2	2	10	2	40
4	5	2	91	69	34	3	16	7	16	3	246

5	8	2	246	124	38	10	332	233	46	18	1.056
6	22	116	105	80	114	4	129	105	61	24	759
7	3	8	110	79	61	32	265	10	321	269	1.156
8	104	18	222	78	141	13	89	10	65	78	819
9	2	5	74	51	119	10	94	10	23	68	456
10	2	2	10	7	49	3	11	4	5	21	116
11	2	9	14	12	115	4	5	6	5	6	178
12	2	4	8	5	52	41	5	4	17	19	158
13	2	2	25	3	20	3	4	7	18	20	104
Total	153	175	909	516	745	126	952	398	587	528	5.087

Fonte: Resultados da pesquisa.

TABELA 4 - Distribuição da Amostra Conciliatória com Erros de Amostragem de 4% para Milho, 6% para Cana-de-açúcar e Laranja, 7% para Soja e Café, 9% para Algodão e Feijão de Inverno e 10% para os Demais Produtos

Estrato	DIRA										Total
	2	3	4	5	6	7	8	9	10	11	
3	2	7	5	18	3	2	2	2	11	2	55
4	5	5	106	97	39	5	20	8	18	4	306
5	9	6	289	206	77	20	364	241	56	35	1.304
6	23	117	126	134	142	7	162	106	69	55	940
7	3	10	133	133	111	38	302	10	328	288	1.355
8	104	19	259	121	220	25	128	17	72	102	1.067
9	2	7	97	85	163	18	124	12	26	82	616
10	2	2	14	16	74	7	18	6	6	23	169
11	2	9	18	23	128	8	8	7	6	11	221
12	2	4	11	10	66	45	7	5	18	21	190
13	2	2	31	5	30	5	9	10	22	21	137
Total	156	189	1.090	847	1.051	180	1.146	425	631	644	6.358

Fonte: Resultados da pesquisa.

TABELA 5 - Distribuição da Amostra Conciliatória com Erros de Amostragem de 4% para Milho, 6% para Cana-de-açúcar e Laranja, 7% para Soja e Café, 9% para Algodão e Feijão de Inverno e 10% para os Demais Produtos, Excluindo Amendoim da Seca

Estrato	DIRA										Total
	2	3	4	5	6	7	8	9	10	11	
3	2	8	6	19	3	2	2	2	12	2	58
4	5	5	115	106	42	5	22	8	20	4	332
5	10	6	316	224	84	22	97	88	61	38	945
6	25	127	137	145	155	8	129	8	58	60	852
7	3	11	144	144	121	19	123	11	186	133	894
8	113	21	282	132	240	27	140	19	40	111	1.124
9	2	8	105	92	155	20	135	13	17	46	592
10	2	2	16	17	80	7	20	7	6	15	172
11	2	10	20	25	101	8	9	7	6	11	199
12	2	5	12	10	72	8	8	5	19	11	152
13	2	2	34	6	32	5	10	10	24	6	132
Total	168	204	1.187	922	1.082	130	693	179	450	437	5.452

Fonte: Resultados da pesquisa.

TABELA 6 - Distribuição da Amostra Conciliatória com Erros de Amostragem de 4% para Milho, 6% para Cana-de-açúcar e Laranja, 7% para Soja e Café, 9% para Algodão e Feijão de Inverno e 10% para os Demais Produtos, Excluindo Amendoim da Seca e das Águas

Estrato	DIRA										Total
	2	3	4	5	6	7	8	9	10	11	
3	2	8	6	21	3	2	2	2	13	2	61
4	6	5	127	116	13	5	24	8	22	4	329
5	10	7	347	247	92	24	107	22	67	42	963
6	27	140	150	160	65	8	82	9	35	65	741
7	3	12	159	159	110	20	111	12	63	75	723
8	124	23	183	145	241	30	152	20	43	91	1.053
9	2	8	116	101	138	22	148	14	19	34	602
10	2	2	17	19	85	7	22	7	7	6	174
11	2	11	22	28	35	9	9	8	6	11	140
12	2	5	13	11	33	8	9	5	21	7	114
13	2	2	37	6	21	6	11	11	26	3	125
Total	183	222	1.177	1.012	835	141	675	119	322	342	5.026

Fonte: Resultados da pesquisa.

São Paulo.

A rotina elaborada é simples e com alterações mínimas pode-se alocar elementos em outras estratificações, por exemplo, incorporando novas DIRAs como Barretos, São Carlos e Vale do Para-

napanema, criadas recentemente pela Secretaria de Agricultura e Abastecimento do Estado de São Paulo. Similarmente, basta que os dados de um novo cadastro sejam distribuídos pelas caselas da estratificação desejada para que o mesmo possa ser utilizado.

NOTAS

LITERATURA CITADA

- CAMARGO, Milton N. **Amostra para previsão e estimativas de safras agrícolas do Estado de São Paulo em vigor a partir de junho de 1981**. São Paulo, IEA, 1988. 75p. (Relatório de Pesquisa, 27/88).
- CAMARGO FILHO, Waldemar P. et alii. **Estatísticas da produção agrícola no Estado de São Paulo**. São Paulo, IEA, 1990. 1v.
- CAMPOS, Humberto & PIVA, Luiz H.O. Dimensionamento de amostra para estimativa e previsão de safra no Estado de São Paulo. **Agricultura em São Paulo**, SP, **21(3)**:65-88, 1974.
- COCHRAN, Willian G. **Técnicas de amostragem**. Rio de Janeiro, Fundo de Cultura, 1963. 55-5p.
- KISH, Leslie. **Survey sampling**. New York, John Wiley, 1965. 643p.
- PINO, Francisco A. **Apontamentos do curso de amostragem probabilística**. (não publ.).
- ____ & CASER, Denise V. **Análise de erros não amostrais em São Paulo**. São Paulo, IEA, 1984a. 25p. (Relatório de Pesquisa, 10/84).
- ____ & _____. **Falta de resposta em levantamentos por amostragem: um estudo de caso**. São Paulo, IEA, 1984b. 25p. (Relatório de Pesquisa, 8/84).
- _____. et alii. Algumas aplicações do SAS em economia agrícola. **Informações Econômicas**, SP, **22(12)**:25-30, dez. 1992.
- SANTOS, Vera L.F. et alii. Dimensionamento de amostra para levantamento da citricultura paulista. **Pesquisa Agropecuária Brasileira**, Brasília, **22(1)**:15-21, jan. 1987.
- SCHATTAN, Salomão. **Obtenção de estatísticas agrícolas pelo método de amostragem**. São Paulo. DPV/DER, 1953. 39p.
- STEVENS, Wilfred L. **Estimativa e previsão de safras através de um levantamento por amostragem**. São Paulo, DPV/DER, 1951. 13p.
- VICENTE, José R.; CAMARGO, Milton N.; PE-RES, Rubens F. Dimensionamento de amostras para levantamentos agrícolas pela estratificação de eficiência máxima. **Pesquisa Agropecuária Brasileira**, Brasília, **24(4)**:393-397, abr. 1989.
- _____. et alii. **A bananicultura paulista em 1980: tópicos da produção e dimensionamento de amostras para previsão de safras**. São Paulo, IEA, 1987. 30p. (Relatório de Pesquisa, 2/87).

UMA ROTINA DE REDIMENSIONAMENTO DE AMOSTRAS PARA ESTIMATIVAS DE SAFRAS ATRAVÉS DO STATISTICAL ANALYSIS SOFTWARE (SAS)

Anexo 1

Rotina para Redimensionamento de Amostras para Previsão de Safras

```

/* Parte I:
Conversao de arquivos ascii em SasDataSet,
criando os arquivos: &arq_sai, para o Metodo
descrito em STEVENS (1951) e &arq_fim para
Partilha de Neyman (COCHRAN, 1965) */

options ls=66 ps=80 nodate nonumber nocenter;
title;

libname drive 'c:\amostra';
filename nome1 'c:\amostra\objfat91.txt';
filename nome2 'c:\amostra\obj91001.txt';

%let arq1=objfat91;
%let arq2=obj91001;
%let arq=obj1;

/* I.1 - Arquivo 1 - &arq1 */

data &arq1 (drop=calculo codprod fat_cor area_ncr);
  infile nome1;
  input calculo 1-1 dira 2-4 estrato 5-6 delega 7-8
  codprop 9-11 codprod 12-19 produto 12-15
  fat_exp 20-24 fat_cor 25-31 area_ncr 32-39
  area_cor 40-48;
run;

/* I.2 - Arquivo 2 - &arq2 */

data &arq2 (drop=codarea area_01 ano);
  infile nome2;
  input dira 1-2 estrato 3-4 delega 5-6 codprop 7-9
  codarea 10-16 area_01 17-27 ano 28-29 mes
  30-31;
run;
proc sort data=&arq1;
  by dira estrato delega codprop;run;

proc sort data=&arq2;
  by dira estrato delega codprop;run;

/* I.3 - Arquivos parciais - obj{produto}, a partir
dos arquivos 1 e 2 */

data obj113;
  merge &arq2 (where=(mes=2)) &arq1 (where=
  (produto=113));
  by dira estrato delega codprop;
  produto=113;
  output;
run;

data obj102;
  merge &arq2 (where=(mes=4)) &arq1 (where=
  (produto=102));
  by dira estrato delega codprop;
  produto=102;
  output;
run;

data obj305;
  merge &arq2 (where=(mes=4)) &arq1 (where=
  (produto=305));
  by dira estrato delega codprop;
  produto=305;
  output;
run;

data obj101;
  merge &arq2 (where=(mes=6)) &arq1 (where=
  (produto=101));
  by dira estrato delega codprop;
  produto=101;
  output;
run;

data obj103;
  merge &arq2 (where=(mes=6)) &arq1 (where=
  (produto=103));

```

```

by dira estrato delega codprop;
produto=103;
output;
run;

data obj105;
merge &arq2 (where=(mes=6)) &arq1 (where=
(produto=105));
by dira estrato delega codprop;
produto=105;
output;
run;

data obj114;
merge &arq2 (where=(mes=6)) &arq1 (where=
(produto=114));
by dira estrato delega codprop;
produto=114;
output;
run;

data obj124;
merge &arq2 (where=(mes=6)) &arq1 (where=
(produto=124));
by dira estrato delega codprop;
produto=124;
output;
run;

data obj126;
merge &arq2 (where=(mes=6)) &arq1 (where=
(produto=126));
by dira estrato delega codprop;
produto=126;
output;
run;

data obj133;
merge &arq2 (where=(mes=9)) &arq1 (where=
(produto=133));
by dira estrato delega codprop;
produto=133;
output;
run;

data obj109;
merge &arq2 (where=(mes=11)) &arq1 (where= (-
produto=109));
by dira estrato delega codprop;
produto=109;
output;
run;

data obj311;
merge &arq2 (where=(mes=11)) &arq1 (where= (-
produto=311));
by dira estrato delega codprop;
produto=311;
output;
run;

/* I.4 - Arquivo - &arq - com todos obj{produto}
parciais */

data &arq;
set obj113 obj102 obj305 obj101 obj103 obj105
obj114 obj124 obj126 obj133 obj109 obj311;
if area_cor=. then area_cor=0;
if fat_exp=. then fat_exp=0;
output;
run;

/* I.5 - Obtencao do fator de expansao maximo
(fat_exp) por &ordena (produto dira estrato) */

%let arq_sai=obj2;
%let arq_fim=obj3;
%let ordena=produto dira estrato;
%let var_mean=fat_exp;
%let estat=max;

%let nome_est=%str(max=fat1_exp);

%macro estat;
proc sort data=&arq; by &ordena; run;

proc means data=&arq &estat noprint;
var &var_mean;
by &ordena;
output out=&arq_sai &nome_est;
run;
%mend estat;
%estat

proc sort data=&arq_sai; by &ordena; run;

data &arq_fim (drop=_type__freq_);

```

```

merge &arq &arq_sai;
  by &ordena;
output;
run;

/* I.6 - Criacao de novas variaveis: totais por
&ordena a partir do arquivo &arq */

%let arq=obj3;
%let arq_sai=obj4;
%let arq_fim=obj5;
%let var_mean=area_cor fat1_exp;
%let estat= N sum var;
%let nome_est=%str(N=nar NFE
  sum=ax NH_EXP var=VAX vnh);

%estat

%let var_drop= nar vnh;

data &arq_fim (drop=_type__freq_ &var_drop);
set &arq_sai;
  ar_exp= ax * nh_exp / nfe;
  vx_exp= nh_exp * (nh_exp - nfe) * vax/nfe;
  nh_sx= nh_exp * sqrt(vx_exp);
output;
run;

/* Arquivos finais: &arq_sai (STEVENS) e &arq_fim
(NEYMAN) */

%let arq=obj5;
%let arq_sai=objv1_91;
%let arq_fim=objv_91;
%let ordena=produto;
%let var_mean=nfe Nh_exp Vx_exp ar_exp nh_sx;
%let estat=sum;
%let nome_est=%str(sum=nfe_tot Nh_tot Vx_tot
  ar_tot nh_sxtot);

%estat

proc sort data=&arq_sai; by &ordena; run;

data &arq_fim (drop=_type__freq_);
merge &arq &arq_sai;
  by &ordena;
output;
run;

```

Informações Econômicas, SP, v.24, n.5, maio 1994.

/* Parte II:

1a. aproximacao para no. de elementos da nova amostra pelo Metodo descrito em STEVENS (1951) */

/* II.1 - Criacao do arquivo &arq_st - calculo do erro amostral */

```

%let arq_st=stevens;
%let inicio=1;
%let fim=10;
%let num=%eval(&fim-&inicio+1);

data &arq_st;
set &arq_sai;
  erro_am= sqrt(vx_tot)*100/ar_tot;
  array n_{&num} n_&inicio-n_&fim;
  do i=&inicio to &fim;
    n_{i} = nfe_tot * (erro_am/i)**2;
  end;
output;
run;

```

/* II.2 - Criacao de tabela contendo os resultados segundo STEVENS */

```

proc format;
  value prod 101="algodao"    102="amend.
                                aguas"
            103="amend. seca" 105="arroz"
            305="cafe"       109="cana"
            113="feijao aguas" 133="feijao inverno"
            114="feijao seca"  311="laranja"
            124="milho"       126="soja";
run;

proc tabulate data=&arq_st formchar=" - - - - "
  noseps format=10.;

title1 "Numero Aproximado de Elementos de Amos-
  tras por Produto (STEVENS,1951)";
class produto;
var n_&inicio-n_&fim;
format produto prod.;
table produto=" ", n_1 n_2 n_3 n_4 n_5 n_6 n_7

```

```

n_8 n_9 n_10 / box="produto" rts=20;
keylabel sum=" ";
label N_1="1%" N_2="2%" N_3="3%"
      N_4="4%" N_5="5%" N_6="6%"
      N_7="7%" N_8="8%" N_9="9%"
      N_10="10%";
run;

```

/* Parte III:

Calculos para o redimensionamento - KISH (1965)
*/

/* III.1 - Alocacao dos elementos da nova amostra por produto atraves da Partilha de Neyman (COCHRAN, 1951).

Considerou-se N total fixo para todos os produtos (=305704), garantindo no minimo 2 elementos por casela na amostra final. */

```
options ps=60 ls=120 nodate nonumber;
```

```
%let N=305704;
```

```
%let arq=amostra;
```

```
data &arq (keep=produto dira estrato pn_&inicio-pn_&fim);
```

```
set &arq_&fim;
```

```
array V_{&num} V_&inicio-V_&fim;
```

```
array ni_{&num} ni_&inicio-ni_&fim;
```

```
array nn_{&num} nn_&inicio-nn_&fim;
```

```
array pn_{&num} pn_&inicio-pn_&fim;
```

```
do i=&inicio to &fim;
```

```
  nii = nfe_tot/(1-nfe_tot/&N);
```

```
  V_{i} = (ar_tot * i/100)**2;
```

```
  ni_{i} = nii * Vx_tot / V_{i};
```

```
  nn_{i} = ni_{i}/(1+ ni_{i}/&N);
```

```
  pn_{i} = nn_{i} * nh_sx / nh_sxtot;
```

```
  if pn_{i}<2 then pn_{i}=2;
```

```
end;
```

```
run;
```

/* III.2 - Amostra Conciliatoria:

Geracao da tabela conciliatoria para &erro % erro amostragem para a variavel &var_tab por dira e estrato (&ordena) */

```
/* Para 1% erro */
```

```
%let erro=1;
```

```
title1 "Amostra Conciliatoria - &erro% erro amos-  
tragem para os produtos";
```

```
/* Criacao do arquivo &arq, criando a variavel  
var_tab, a partir do erro de amostragem adotado  
(%erro) */
```

```
%macro erro;
```

```
data &arq;
```

```
set &arq;
```

```
var_tab=pn_&erro;
```

```
%mend erro;
```

```
%erro
```

```
/* Criacao do arquivo &arq_sai, contendo as medias as  
variaveis &var_mean, por &ordena, a partir do arquivo  
&arq (%estat) */
```

```
%let arq_sai=concil;
```

```
%let ordena=dira estrato;
```

```
%let var_mean=var_tab;
```

```
%let estat=mean;
```

```
%let nome_est=%str(mean=&var_mean);
```

```
%estat
```

```
/* Criacao da tabela conciliatoria, contendo as  
variaveis dira, estrato e var_tab, a partir do arquivo  
&arq_sai (%tabela) */
```

```
%macro tabela;
```

```
proc tabulate data=&arq_sai formchar=" - - - - "
```

```
  noseps format=8.;
```

```
class &ordena;
```

```
var var_tab;
```

```
table estrato=" " all="total", var_tab=" " *sum=" "
```

```
  *(dira all="total") / box="estrato rts=10;
```

```
run;
```

```
%mend tabela;
```

```
%tabela
```

```
/* Para 5% erro */
```

```
%let erro=5;
```

```
title1 "Amostra Conciliatoria - &erro% erro amos-  
tragem para os produtos";
```

```
%erro
```

```
%estat
```

```
%tabela
```

```

/* Para 10% erro */
%let erro=10;
title1 "Amostra Conciliatoria - &erro% erro amos-
tragem para os produtos";
%erro
%estat
%tabela

```

```

/* Outras Simulacoes com diferentes erros amostra-
gens para cada produto */

```

```

/* 1a. simulacao */

```

```

%let erro1=5;
%let n_prod1=demais;
%let erro2=10;
%let prod2=102,103;
%let n_prod2=amendoim;

title1 "Amostra Conciliatoria - &n_prod2 &erro2%,
&n_prod1 &erro1%";

```

```

data &arq;
set &arq;
if produto in (&prod2) then var_tab=pn_&erro2;
else var_tab=pn_&erro1;
run;
%estat
%tabela

```

```

/* 2a. simulacao */

```

```

%let prod2=113,114,133;
%let n_prod2=feijao;

title1 "Amostra Conciliatoria - &n_prod2 &erro2%,
&n_prod1 &erro1%";

```

```

data &arq;
set &arq;
if produto in (&prod2) then var_tab=pn_&erro2;
else var_tab=pn_&erro1;
run;

```

```

%estat
%tabela

```

```

/* 3a. simulacao */

```

```

%let prod_del=102,103;

```

```

%let n_del=excluindo amendoim;
%let prod2=113,114,133;
%let n_prod2=feijao;

```

```

title1 "Amostra Conciliatoria - &n_prod2 &erro2%,
&n_prod1 &erro1%, &n_del";

```

```

data &arq;
set &arq;
if produto in (&prod_del) then var_tab=.;
else if produto in (&prod2) then
var_tab=pn_&erro2;
else var_tab=pn_&erro1;
run;
%estat
%tabela

```

```

/* 4a. simulacao */

```

```

%let erro1=4;
%let prod1=124;
%let n_prod1=milho;

%let erro2=6;
%let prod2=109,311;
%let n_prod2=cana,laranja;

```

```

%let erro3=7;
%let prod3=126,305;
%let n_prod3=soja,cafe;

```

```

%let erro4=9;
%let prod4=101,113;
%let n_prod4=algodao,feijao inverno;

```

```

%let erro5=10;
%let n_prod5=demais produtos;

```

```

title1 "Amostra Conciliatoria - &n_prod1 &erro1%,
&n_prod2 & erro2%,";
title2 "&n_prod3 &erro3%, &n_prod4 &erro4%,
&n_prod5 &erro5%";

```

```

data &arq;
set &arq;
if produto=&prod1 then var_tab=pn_&erro1;
else if produto in (&prod2) then
var_tab=pn_&erro2;

```

```

else if produto in (&prod3) then
  var_tab=pn_&erro3;
else if produto in (&prod4) then
  var_tab=pn_&erro4;
else var_tab=pn_&erro5;
run;
%estat
%tabela

/* 5a. simulacao */

%let prod_del=103;
%let n_del=excluindo amendoim seca;

title1 "Amostra Conciliatoria - &n_prod1 &erro1%,
&n_prod2 &erro2%,";
title2 "&n_prod3 &erro3%, &n_prod4 &erro4%,
&n_prod5 &erro5%,&n_del";

data &arq;
set &arq;
if produto=&prod_del then var_tab=.;

```

```

else if produto=&prod1 then var_tab=pn_&erro1;
else if produto in (&prod2) then
  var_tab=pn_&erro2;
else if produto in (&prod3) then
  var_tab=pn_&erro3
else if produto in (&prod4) then
  var_tab=pn_&erro4;
else var_tab=pn_&erro5;
run;

%estat
%tabela

/* 6a. simulacao */

%let prod_del=102,103;
%let n_del=excluindo amendoim seca e das aguas;

title1 "Amostra Conciliatoria - &n_prod1 &erro1%,
&n_prod2 &erro2%,";
title2 "&n_prod3 &erro3%, &n_prod4 &erro4%,
&n_prod5 &erro5%,&n_del";

data &arq;
set &arq;
if produto in (&prod_del) then var_tab=.;
else if produto=&prod1 then var_tab=pn_&erro1;
else if produto in (&prod2) then
  var_tab=pn_&erro2;
else if produto in (&prod3) then
  var_tab=pn_&erro3;
else if produto in (&prod4) then
  var_tab=pn_&erro4;
else var_tab=pn_&erro5;
run;

%estat
%tabela

```

Anexo 2

Descrição dos Códigos do Anexo 1

I. Criação dos arquivos SASDATASET:

=> Arquivos de dados utilizados:

- objfat91.txt : contém dados dos imóveis rurais amostrados, que informaram a existência das culturas selecionadas, em cada levantamento;
- obj91001.txt : contém informações sobre todos os informantes de cada levantamento.

=> Descrição dos Códigos utilizados no IEA dos 12 Produtos Estudados:

(número do produto + código da informação coletada, e mês do levantamento.)

- 1130112: área plantada em alqueires - feijão das águas (fev.)
- 3050223: número de pés em produção - café (abr.)
- 1020161: área colhida em alqueires - amendoim das águas (abr.)
- 1010161: alqueires plantados - algodão (jun.)
- 1030161: alqueires realmente colhidos - amendoim da seca (jun.)
- 1050112: alqueires plantados - arroz (jun.)
- 1140161: alqueires realmente colhidos - feijão da seca (jun.)
- 1240161: alqueires realmente colhidos - milho em grão (jun.)
- 1260161: alqueires realmente colhidos - soja (jun.)
- 1330112: área plantada - feijão de inverno (set.)
- 1090169: alqueires a serem cortados no próx. ano agrícola - cana (nov.)
- 3110235: no. pés já existentes (total) - laranja (nov.)

=> Descrição das variáveis:

I.1. Arquivo 1:

- &arq1: define o nome do arquivo 1, referido como objfat91, através do comando macro %let;
- cálculo: número do levantamento utilizado: 1 (intenção de plantio) a 5 (final da colheita);
- DIRA: Divisão Regional Agrícola do Estado de São Paulo (total de 10 regiões: DIRAS 2 a 11);
- estrato: estratificação dos imóveis rurais para o Estado de São Paulo, de acordo com sua área, (total de 10 estratos);
- delega: Delegacias Regionais Agrícolas (DAs), ou

seja, subdivisões das DIRAS;

- codprop: código da propriedade, ordenado dentro da DIRA, estrato e DA;
- codprod: código do produto (número do mesmo + informação coletada);
- produto: número do produto considerado no estudo;
- fat_exp: fator de expansão por produto, DIRA e estrato;
- fat_cor: fator de correção (razão entre a área do imóvel existente no cadastro e a efetivamente informada no questionário) por produto, DIRA e estrato;
- area_ncr: área cultivada não corrigida, conforme informada no questionário, por produto, DIRA e estrato;
- area_cor: área cultivada, corrigida através do fator de correção, por produto, DIRA e estrato.

I.2. Arquivo 2:

- &arq2: define o nome do arquivo 2, referido como obj91001, através do comando macro %let;
- cod_area: código de área do imóvel rural;
- area_01: área atual do imóvel rural;
- ano/mes: ano e mês do levantamento utilizado.

I.3. Arquivos parciais obj+{número do produto}:

- a partir dos arquivos originais &arq1 e &arq2.

I.4. Arquivo &arq:

- &arq: define o nome do arquivo, referido como obj1, através do comando macro %let, composto por todos os produtos, a partir dos arquivos parciais.

I.5. Obtenção do fator de expansão máximo:

- &arq_sai: define o nome do arquivo de saída, referido como obj2, gerado a partir da proc means;
- &arq_fim: define o nome do arquivo final, referido como obj3, a partir da combinação de &arq e &arq_sai;
- &ordena: contém as variáveis que definirão os limites da proc means para obtenção do fator máximo;
- &var_mean: variável que será utilizada na proc

means, denominada fat_exp;
 - &estat: nome da estatística utilizada nessa proc means;
 - &nome_estat: nome da variável que conterá a estatística, denominada fat_exp1;
 - %macro estat: criação do modelo para cálculo de estatísticas, através da proc means, de acordo com as variáveis macro definidas através do comando %let;
 - %estat: processa o modelo para cálculo de estatísticas.

I.6. Criação de novas variáveis:

- &arq_sai: define o novo nome do arquivo de saída, referido como obj3, a partir da proc means, utilizando as mesmas variáveis definidas na variável &ordena;
 - &arq_fim: define o novo nome do arquivo final, referido como obj4, a partir da combinação de &arq e &arq_sai;
 - &var_mean: novas variáveis que serão utilizadas na proc means, area_cor e fat1_exp;
 - &estat: nome das estatísticas utilizadas nessa proc means;
 - &nome_estat: nomes das variáveis que conterão as estatísticas;
 - %estat: processa novamente o modelo para cálculo de estatísticas;
 - nfe: número de imóveis do levantamento por produto, dira e estrato;
 - nh_exp: número de imóveis expandidos por produto, dira e estrato;
 - vax: variância da área cultivada por produto, dira e estrato;
 - vax/nfe: variância da média;
 - &var_drop: variáveis que não farão parte do &arq_fim;
 - vx_exp: variância original expandida por produto, dira e estrato;
 - ar_exp: área média expandida por produto, dira e estrato;
 - nh_sx: no. imóveis x desvio padrão por produto, dira e estrato.

Para os arquivos finais:

- &arq: define o arquivo fonte, obj5, para ser utilizado na proc means;
 - &arq_sai: define novo nome do arquivo de saída, denominado objv1_91, a partir da proc means, utilizando as mesmas variáveis definidas na variável &ordena;
 - &arq_fim: define o novo nome do arquivo final,

denominado objv_91, a partir da combinação de &arq e &arq_sai;
 - &ordena: contém a variável que definirá os limites da proc means;
 - &var_mean: novas variáveis que serão utilizadas na proc means, nfe nh_exp vx_exp, ar_exp e nh_sx;
 - &estat: nome das estatísticas utilizadas nessa proc means;
 - &nome_estat: nomes das variáveis que conterão as estatísticas;
 - %estat: processa novamente o modelo para cálculo de estatísticas;
 - nh_tot: (nh_exp) total por produto;
 - Vx_tot: (Vx_exp) total por produto;
 - ar_tot: (ar_exp) total por produto;
 - nh_sxtot: (nh_sx) total por produto.

II. Método de STEVENS (1951):

II.1. Criação do arquivo:

- &arq_st: define o novo nome do arquivo, denominado stevens;
 - &inicio: define o erro amostral inicial;
 - &fim: define o erro amostral final;
 - &num: define o número de erros amostrais utilizados;
 - n_{i}: número de elementos da nova amostra, considerando i% de erro amostragem desejado (i variando de &inicio até &fim), definido a partir do comando array;
 - erro_am: erro de amostragem do levantamento.

II.2. Criação da tabela:

- através da proc tabulate.

III. Cálculos para o Redimensionamento:

III.1. Partilha de Neyman:

- &N: contém o número total de imóveis do cadastro, definido através do comando macro %let;
 - &arq: define o nome do arquivo, denominado amostra, que será utilizado nos cálculos;
 - &arq_fim: define o nome do arquivo fonte já criado, denominado objv_91;
 - V_{i}: variância desejada com i% de erro amostragem (i de 1% a 10%);
 - pn_{i}: número de elementos para a nova amostra,

segundo a Partilha de Neyman, com $i\%$ de erro amostragem (i de 1% a 10%).

III.2. Amostra Conciliatória:

- &erro: define o valor do erro amostral a ser utilizado na conciliatória;
- %erro: processa o modelo para criação da variável var_tab de acordo com o erro adotado;
- %tabela: processa o modelo para criação da

tabela conciliatória, contendo as variáveis DIRA, estrato e var_tab, conforme o erro amostral adotado;

- &erro1, &erro2: definem os valores dos diferentes erros amostrais adotados para cada produto na conciliatória;
- &prod1, &prod2: definem os números dos produtos que conterão os respectivos &erro1 e &erro2 em determinadas simulações da conciliatória;
- &n_prod1, &n_prod2: definem os nomes dos produtos, a partir de &prod1 e &prod2, respectivamente;
- prod_del: define o(s) número(s) do(s) produto(s) que serão excluídos em determinadas simulações;
- n_del: define o(s) nome(s) do(s) produto(s) excluídos.

¹Este trabalho é parte integrante do projeto SPTC 16-011/93. Versão preliminar foi apresentada no III Congresso Brasileiro do Grupo de Usuários do SAS, Londrina, PR, de 15 a 18 de março de 1994. Recebido em 28/04/94. Liberado para publicação em 16/05/94.

²Engenheiro Agrônomo, Pesquisador Científico do IEA.

³Os trabalhos pioneiros de STEVENS (1951) e SCHATTAN (1953) descrevem o sistema, apresentam e discutem seus primeiros resultados; um histórico dos levantamentos de previsão de safras pode ser visto em CAMARGO FILHO et alii (1990).

⁴A esse respeito, além dos trabalhos de STEVENS (1951) e SCHATTAN (1953), ver COCHRAN (1963).

⁵Exemplos são os trabalhos de CAMPOS & PIVA (1974) e CAMARGO (1988).

⁶Sobre falta de respostas nesses levantamentos, ver PINO & CASER (1984b); os mesmos autores descreveram os tipos de erros mais encontrados nos questionários (PINO & CASER, 1984a).

⁷Alguns estudos desenvolvidos no Instituto de Economia Agrícola (IEA) empregaram (ou desenvolveram) programas ou rotinas de dimensionamento de amostras para situações específicas; pode-se citar, por exemplo, SANTOS et alii (1987), VICENTE et alii (1987), VICENTE, CAMARGO; PERES (1989) e PINO et alii (1992).

⁸A rotina completa é apresentada no Anexo 1; uma descrição dos códigos utilizados na mesma está no Anexo 2.

⁹Os estratos de área utilizados são os propostos por STEVENS (1951): o estrato 3 engloba imóveis rurais com áreas entre 3,1 e 5,0 hectares; o estrato 4, imóveis de 5,1 a 10,0 hectares; o estrato 5, de 10,1 a 20,0 hectares; o estrato 6, de 20,1 a 30,0 hectares; o estrato 7, de 30,1 a 50,0 hectares; o estrato 8, de 50,1 a 100,0 hectares; o estrato 9, de 100,1 a 200,0 hectares; o estrato 10, de 200,1 a 300,0 hectares; o estrato 11, de 300,1 a 500,0 hectares; o estrato 12, de 500,1 a 1.000,0 hectares; e o estrato 13, engloba imóveis rurais com áreas maiores que 1.000,1 hectares. Os imóveis com áreas até 3,0 hectares não são considerados nos levantamentos, devido à sua reduzida importância econômica.

¹⁰Regionalização adotada pela Secretaria de Agricultura e Abastecimento do Estado de São Paulo, em que a DIRA 2 corresponde a Registro; a DIRA 3, a São José dos Campos; a DIRA 4, a Sorocaba; a DIRA 5, a Campinas; a DIRA 6, a Ribeirão Preto; a DIRA 7, a Bauru; a DIRA 8, a São José de Rio Preto; a DIRA 9, a Araçatuba; a DIRA 10, a Presidente Prudente; e a DIRA 11, a Marília.

¹¹O fator de expansão de cada estrato, em cada DIRA, é a razão entre o número de imóveis do cadastro e o número de elementos da amostra.

¹²Aceitando-se um erro de 10% para amendoim das águas e da seca e de 5% para os demais produtos, a amostra conciliatória obtida teria cerca de 13.000 elementos. Excluindo-se os amendoins, e com 10% de erro para os feijões e 5% para os demais produtos, o tamanho de amostra obtido foi de 8.648.