

# **Análise de Erros Padrão Simuladas para os Resultados do Trabalho de Inquérito Agrícola 2002 Baseados numa Amostra de 94 Distritos para Moçambique**

David J. Megill  
Consultor de Amostragem

Abril de 2002

## **1. Introdução**

Em Março de 2002 o consultor trabalhou com o Instituto Nacional de Estatística (INE) e o Ministério de Agricultura e Desenvolvimento Rural (MADER) no estudo de erros padrão para o Censo Agro-Pecuário (CAP) 1999-2000, e uma análise de erros padrão simulados para o Trabalho de Inquérito Agrícola (TIA) 2002, baseado em desenhos de amostragem alternativos. A metodologia e resultados deste estudo estão documentados no relatório “Análise do Grau de Precisão para os Resultados do Censo Agro-Pecuário 1999-2000 de Moçambique, e Revisão dos Planos de Amostragem para o Trabalho de Inquérito Agrícola 2002” (Megill, Março de 2002). Aquele relatório deve ser usado como referência junto com este para completar a documentação deste estudo.

A conclusão do análise prévio foi de que a eficiência da amostra para o TIA 2002 aumentaria quase 20 por cento se usássemos uma amostra bi-etápica para o inquérito em vez de uma amostra de três etapas com 66 distritos seleccionados na primeira etapa. Ao mesmo tempo, uma amostra bi-etápica implicaria aumentar o número de distritos seleccionados a 132, dado a maior dispersão dos segmentos seleccionados. Isto aumentaria os custos e complicaria a logística operativa do trabalho do campo. Por isso o consultor foi contratado a fazer uma análise da precisão e eficiência de outra amostra alternativa para o TIA 2002, baseado em um desenho de três etapas, mas com 94 distritos seleccionados na primeira etapa. Dado que este desenho vai dispersar mais a amostra comparado com uma amostra de 66 distritos, a hipótese é que vai melhorar a eficiência estatística da amostra. Também vai ser necessário de estudar a aumento em custos correspondentes para fazer o trabalho de campo em 94 distritos. Logo vamos saber se a ganho em eficiência vai justificar os custos correspondentes.

## **2. Desenho Alternativo para a Amostra do TIA 2002, Baseado em 94 Distritos Amostrais**

Quando o Christopher Hill, *Senior Census Advisor* da FAO (*Food and Agriculture Organization*) das Nações Unidas, elaborou o desenho preliminar da amostra para o TIA 2002 com 66 distritos seleccionadas na primeira etapa, êle fez a selecção em três grupos: dois de 33 distritos e um de 28 distritos (dado que todos os distritos já tinham sido seleccionados em quatro estratos). O desenho preliminar de três etapas com 66 distritos foi baseado nos primeiros dois grupos de 33 distritos amostrais cada. O novo desenho alternativo envolve a adição do último

grupo de 28 distritos para completar a amostra representativa de 94 distritos dentro dos 26 estratos (província por zona agro-ecológica) definidos no relatório prévio. A atribuição destes 94 distritos amostrais por estrato está apresentada no Quadro 1.

Aumentar o número de distritos amostrais na primeira etapa implica diminuir o número de segmentos seleccionados por distrito amostral na segunda etapa; estes segmentos correspondem às unidades primárias de amostragem (UPAs) definidas para o CAP. Isto também vai modificar um pouco a atribuição de explorações agro-pecuárias seleccionadas por estrato, mas tentamos manter a atribuição da amostra o mais similar possível àquela da amostra preliminar.

Seleccionando 6 segmentos amostrais por distrito, vamos ter uma amostra total de 564 segmentos dentro dos 94 distritos amostrais, comparado com 528 segmentos no desenho preliminar. Mantendo uma amostra de 8 pequenas explorações dentro de cada segmento, isto também vai aumentar o tamanho total da amostra de 4,224 a 4,512 pequenas explorações.

Temos que tomar isto em conta quando comparamos os resultados simulados dos dois desenhos. Também seria possível seleccionar 7 pequenas explorações amostrais por segmento baixo a nova alternativa, que implicaria uma amostra total de 3,948 pequenas explorações, mas isso não seria recomendável a não ser que o custo do número total de questionários seja um problema; isto diminuiria um pouco a precisão dos resultados. O Quadro 1 apresenta a atribuição de pequenas explorações para a amostra alternativa de 94 distritos, como também para a amostra preliminar de 66 distritos, para que se podem ser comparadas.

No Quadro 1 se pode ver que dentro de 13 dos 26 estratos todos os distritos são seleccionados, isto é, os distritos são auto-representados na amostra; estes 13 estratos contêm aproximadamente 33 por cento dos agregados familiares rurais de Moçambique. Nestes estratos o desenho da amostra seria bi-etápica. No desenho da amostra preliminar com 66 distritos amostrais, sómente quatro estratos tinham todos os distritos seleccionados; uma amostra do mesmo tamanho (8 segmentos) foi seleccionada na primeira etapa dentro de cada um destes distritos. Neste caso os factores de expansão variam por distrito dentro do estrato. Para obter uma amostra auto-ponderada dentro de cada estrato, se recomenda seleccionar a amostra de UPAs (segmentos) sistematicamente dentro de cada estrato (província por zona agro-ecológica) inteiro. As UPAs do CAP na base de amostragem devem ser ordenadas por distrito para manter uma estratificação implícita por distrito. Neste caso a número de segmentos seleccionados por distrito vai variar um pouco dependendo do número de agregados familiares dentro do distrito; isto é, a atribuição da amostra seria proporcional ao tamanho do distrito.

Quadro 1. Atribuição da Amostra por Estrato para o Desenho Alternativo do TIA 2002 com 94 Distritos Amostrais: Distritos, Segmentos e Pequenas Explorações

Província	Estrato	No. Distritos CAP	No. Distritos TIA	No. Segmentos TIA	No. Peq. Explor. TIA, 94 Distritos	No. Peq. Explor. TIA, 66 Distritos
Niassa	0101	5	3	18	144	128
	0106	11	3	18	144	128
Cabo Delgado	0205	2	2	12	96	128
	0206	2	2	12	96	128
	0210	13	6	36	288	256
Nampula	0306	3	3	18	144	128
	0310	15	9	54	432	384
	0311	2	2	12	96	128
Zambézia	0404	3	3	18	144	128
	0407	5	5	30	240	256
	0411	9	6	36	288	256
Tete	0502	3	3	18	144	128
	0508	3	3	18	144	128
	0512	7	3	18	144	128
Manica	0603	5	3	18	144	128
	0609	3	3	18	144	128
	0614	2	2	12	96	128
Sofala	0709	4	3	18	144	128
	0711	5	3	18	144	128
	0712	3	3	18	144	128
Inhambane	0813	11	6	36	288	256
	0814	3	3	18	144	128
Gaza	0913	3	3	18	144	128
	0914	5	3	18	144	128
	0915	4	3	18	144	128
Maputo-Provincia	1014	7	6	36	288	256
<b>Moçambique</b>		138	94	564	4,512	4,224

Podemos usar um caso extremo para ilustrar a vantagem de fazer a selecção sistemática de segmentos dentro do estrato inteiro. O estrato 0609, que corresponde à zona agro-ecológica 09 na província de Manica, tem três distritos, todos incluídos na amostra de 94 distritos para o TIA; o distrito menor neste estrato é Macossa, com 2,106 agregados familiares, e o distrito maior é Mossurize, com 27,138 agregados familiares. Se seleccionamos 6 segmentos, ou 48 pequenas explorações, dentro de cada distrito, o factor de expansão para as pequenas explorações na amostra do TIA em Macossa seria aproximadamente 43.875; para Mossurize, o factor de expansão seria 565.375, ou 12.9 vezes o factor para Macossa. Esta variabilidade nos factores de expansão diminuiria a eficiência da amostra e aumentaria a variância das estimativas. Se seleccionáramos os segmentos em duas etapas dentro do estrato inteiro, todas as pequenas explorações teriam o mesmo factor de expansão, 283.352. Ao mesmo tempo, o número de segmentos seleccionados dentro dos distritos nestes estratos com amostra bi-etápica vai variar de acordo ao seu tamanho. Para ilustrar este efeito da diferença dos tamanhos de distritos, Quadro 2 apresenta o número de segmentos seleccionados em cada distrito nos 13 estratos com uma amostra bi-etápica para esta amostra alternativa de 94 distritos para o TIA. Pode-se ver que o número de segmentos seleccionados por distrito varia entre 1 e 12 de acordo aos tamanhos dos distritos correspondentes.

Quadro 2. Número Total de Agregados Familiares por Distrito, e Atribuição de Segmentos e Pequenas Explorações Amostrais dentro dos Estratos com Amostra Bi-etápica, Desenho Alternativo de 94 Distritos

Província/ Estrato	Código do Distrito	Nome do Distrito	Número de Agregados Familiares no Distrito	Número de Segmentos (UPAs) Seleccionados	Número de Pequenas Explorações Seleccionadas
Cabo Delgado 0205	11	Mueda	22,543	8	64
	12	Muidumbe	14,789	4	32
Cabo Delgado 0206	03	Balama	27,617	5	40
	13	Namuno	39,542	7	56
Nampula 0306	06	Malema	33,421	6	48
	16	Murupula	29,445	5	40
	21	Ribáuè	32,950	7	56
Nampula 0311	02	Angoche	60,673	6	48
	12	Moma	65,055	6	48

Quadro 2. Número Total de Agregados Familiares por Distrito, e Atribuição de Segmentos e Pequenas Explorações Amostrais dentro dos Estratos com Amostra Bi-etápica, Desenho Alternativo de 94 Distritos (Continuação)

Província/ Estrato	Código do Distrito	Nome do Distrito	Número de Agregados Familiares no Distrito	Número de Segmentos (UPAs) Seleccionados	Número de Pequenas Explorações Seleccionadas
Zambézia 0404	05	Gurué	46,561	5	40
	10	Milange	90,793	10	80
	15	Namarroi	22,911	3	24
Zambézia 0407	02	Alto Molócue	45,440	7	56
	04	Gilé	32,748	4	32
	06	Ile	54,564	8	64
	08	Lugela	25,374	3	24
	13	Morrumbala	59,525	8	64
Tete 0502	02	Angónia	61,348	12	96
	07	Macanga	11,276	2	16
	12	Tsangano	25,173	4	32
Tete 0508	05	Chifunde	12,100	7	56
	09	Marávia	12,378	7	56
	13	Zumbu	7,137	4	32
Manica 0609	04	Guro	11,177	5	40
	06	Macossa	2,106	1	8
	08	Mossurize	27,138	12	96
Manica 0614	05	Machaze	14,928	8	64
	10	Tambara	7,788	4	32
Sofala 0712	03	Caia	17,825	8	64
	04	Chemba	11,859	5	40
	10	Maríngue	12,998	5	40
Inhambane 0814	02	Funhalouro	6,321	5	40
	08	Mabote	8,250	5	40
	12	Panda	11,170	8	64
Gaza 0913	01	Cidade Xai-xai	19,604	4	32
	09	Mandlacaze	38,495	8	64
	12	Xai-xai	35,890	6	48

Para obter uma amostra auto-ponderada (com o mesmo factor de expansão) dentro de cada

estrato usando amostragem bi-etápica, foi necessário tomar em conta as taxas de amostragem diferenciais por distrito dentro do estrato. Como se explicou no relatório prévio, os factores de expansão para o CAP dentro de um estrato (província por zona agro-ecológica) variam de um distrito ao outro baseado nas diferenças das razões  $M_h/n_h$  (quer dizer, o número total de agregados familiares no distrito dividido pelo número de UPAs seleccionadas no distrito para o CAP). Para controlar esta variabilidade nas taxas de amostragem por distrito no CAP, as UPAs da amostra bi-etápica nos 13 estratos para o desenho alternativo do TIA foram seleccionadas proporcionalmente ao factor  $M_h/n_h$  para cada distrito. Um ficheiro Excel foi usado para fazer esta selecção de UPAs sistematicamente com probabilidade proporcional a  $M_h/n_h$  dentro de cada estrato.

A base de amostragem usada para seleccionar a amostra bi-etápica para o análise simulado nos 13 estratos só inclui as UPAs do CAP com dados disponíveis. Se depois deste estudo se decidem usar esta amostra de 94 distritos para o TIA 2002, este ficheiro Excel para a selecção bi-etápica pode ser modificado para incluir as UPAs do CAP que estão faltando dados. Dado que as UPAs dentro de cada estrato estão ordenados por código de distrito e código de UPA, é fácil identificar as UPAs omitidas. Por exemplo, se na fôlha Excel a UPA 003 na base para um estrato é seguido por UPA 005, quer dizer que a UPA 004 está faltando. Neste caso, se deve inserir uma linha entre UPAs 003 e 005 na fôlha Excel, e copiar a informação para a UPA 004. Neste caso, seria necessário modificar a medida de tamanho para o distrito correspondente,  $M_h/n_h$ , para que a nova UPA seja incluída no denominador  $n_h$  (número de UPAs seleccionadas para o CAP dentro do distrito). Por exemplo, se dentro de um distrito particular originalmente foram seleccionados 16 UPAs para o CAP, mas estão faltando dados do CAP para uma destas UPAs, sómente 15 UPAs vão aparecer na base de amostragem do TIA para este estudo de simulação, e a medida de tamanho  $M_h/n_h$  no ficheiro Excel seria baseado num  $n_h$  de 15. Depois de inserir uma linha para a UPA do CAP omitida, seria necessário calcular a nova medida de tamanho  $M_h/n_h$  para este distrito, mudando o valor de  $n_h$  a 16. Neste caso a suma das medidas de tamanho dentro do distrito vai ser a mesma que antes,  $M_h$ , o número de agregados dentro do distrito. Por esta razão a medida de tamanho acumulada para o estrato e o intervalo de selecção vão ser iguais aos valores originais, e se pode usar o mesmo arranque aleatório que aparece na fôlha Excel. Seguindo a selecção sistemática proporcional ao tamanho das UPAs para o TIA, é possível que a UPA seleccionada dentro do distrito com a base modificada vai ser diferente (mas em muitos casos seria o mesmo). Para os outros distritos (com base original completa), a mesma amostra de UPAs seleccionada para este análise simulado pode ser usada para o próprio TIA. Se é decido usar esta amostra de 94 distritos para o TIA, este procedimento vai reduzir muito o trabalho para a selecção bi-etápica dentro dos 13 estratos; a maior parte da selecção já está feita.

No caso dos 13 estratos em que a amostra seria seleccionada em três etapas, se usou a amostra de 6 UPAs por distrito seleccionadas por Chris Hill para o TIA. Para esta análise de erros padrão simulados, dados do CAP não estavam disponíveis para 17 dos segmentos seleccionados para o TIA, que representam um pouco mais de 3 por cento da amostra de 564 segmentos. Treze dos segmentos amostrais que faltam dados são do estrato 0411 em Zambézia; douze destes são de dois distritos não cobertos pelo CAP por causa das inundações. O Quadro 3 apresenta uma lista dos 17 segmentos (UPAs seleccionadas para o CAP) sem dados, por estrato. No caso do

desenho de amostragem preliminar baseado em 66 distritos amostrais, estavam faltando dados do CAP para 13 dos segmentos seleccionados (um pouco mais de 2 por cento). Entretanto, em geral os resultados deste estudo de simulação devem medir bem os erros padrão e efeitos do desenho aproximados baseado no desenho alternativo do TIA. Os resultados para Zambézia devem ser interpretados com mais cuidado, dado que os segmentos que estão faltando dados do CAP representam 15 por cento da amostra alternativa do TIA para esta província.

Quadro 3. Lista de Segmentos Seleccionados para o TIA 2002 Sem Dados do CAP

Província	Estrato	Código de Distrito	Código de UPA do CAP
Niassa	0101	03	031
Zambézia	0411	09	222
		09	227
		09	232
		09	243
		09	248
		09	253
		11	334
		14	399
		14	401
		14	405
		14	412
		14	423
		14	427
Tete	0512	04	067
Manica	0603	09	144
Gaza	0915	04	053

### **3. Procedimentos Usados para Gerar os Factores de Expansão e a Codificação do Desenho de Amostragem para o Análise de CENVAR**

Para elaborar a análise de CENVAR dos erros padrão simulados baseada no desenho alternativo de 94 distritos amostrais para o TIA 2002, foi necessário gerar as variáveis de desenho no ficheiro de CENVAR. Estas variáveis incluem os códigos de estrato e UPA, a probabilidade de selecção média da primeira etapa para cada estrato (para o factor de ajuste por população finita), e o factor de expansão para cada exploração agro-pecuária. Usamos uma metodologia similar àquela usada para as análises de CENVAR prévias para calcular os erros padrão baseados no

desenho preliminar de 66 distritos amostrais e no desenho alternativo bi-etápico. No caso da amostra alternativa de 94 distritos amostrais, usamos um desenho de tres etapas para os 13 estratos (província por zona agro-ecológica) maiores, e um desenho bi-etápico para os 13 estratos menores (com todos os distritos seleccionados).

A amostra para as pequenas explorações foi desenhada para ser auto-ponderada (com probabilidades iguais) dentro de cada estrato. Então usamos a seguinte fórmula para estimar o factor de expansão para as pequenas explorações:

$$W_{h(p)} = \frac{M_h \& M_{h(m)}}{m_{h(p)}} ,$$

onde:

$W_{h(p)}$  = factor de expansão para as pequenas explorações seleccionadas no estrato h para o TIA 2002

$M_h$  = número total de agregados familiares dentro do estrato h, estimado das projecções demográficas baseadas no Recenseamento Geral da População e Habitação (RGPH) de 1997

$M_{h(m)}$  = número total de médias explorações no estrato h, estimado dos dados do CAP

$m_{h(p)}$  = número de pequenas explorações seleccionadas com dados do CAP no estrato h na amostra alternativa de 94 distritos amostrais

Dado que todas as médias explorações seriam seleccionadas dentro dos segmentos amostrais, suas probabilidades de selecção seriam baseadas no tamanho do segmento (UPA do CAP), como explicado no relatório prévio. Como não temos medida de tamanho para as UPAs do CAP, um factor de expansão aproximado foi usado para as médias explorações na estimação para o CAP, baseado na estimativa da média do número de agregados por UPA dentro do distrito. Então um factor de expansão aproximado também foi calculado para as médias explorações na amostra do TIA 2002, baseado na razão entre os factores de expansão do CAP para as médias e pequenas explorações amostrais dentro de cada distrito. Podemos expressar este factor de expansão assim:

$$W_{dh(m)} = W_{h(p)} \times \frac{W_{Cdh(m)}}{W_{Cdh(p)}} ,$$

onde:



$W_{dh(m)} =$  factor de expansão para as médias explorações seleccionadas no distrito d do estrato h para o TIA 2002

$W_{h(p)} =$  factor de expansão para as pequenas explorações seleccionadas no estrato h para o TIA 2002, especificada previamente

$W_{Cdh(m)} =$  factor de expansão do CAP para as médias explorações no distrito d do estrato h

$W_{Cdh(p)} =$  factor de expansão do CAP para as pequenas explorações no distrito d do estrato h

Dada a alta taxa de amostragem para os distritos na primeira etapa, foi necessário de incluir na análise de CENVAR um ajuste por população finita, baseada na probabilidade média da primeira etapa dentro de cada estrato. No caso dos 13 estratos com amostra de tres etapas para o TIA 2002, os distritos foram seleccionados com probabilidade proporcional ao número de agregados familiares no distrito. Então a probabilidade média da primeira etapa dentro do estrato seria igual à proporção de agregados familiares nos distritos seleccionados dentro do estrato (isto é, o número de agregados familiares dentro dos distritos seleccionados no estrato dividido pelo número total de agregados no estrato inteiro). No caso dos 13 estratos com amostra bi-etápica para o TIA 2002, não sabemos o número total de UPAs no estrato. Por isso, foi necessário estimar o número de UPAs dentro de cada distrito indirectamente usando a razão entre os factores de expansão do CAP para as pequenas e médias explorações dentro do distrito, que pode ser expressado como o seguinte:

$$\frac{W_{Cdh(p)}}{W_{Cdh(m)}} = \frac{\frac{M_d}{n_d \times 8}}{\frac{M_d}{n_d \times \bar{M}_{di}}} = \frac{\bar{M}_{di}}{8},$$

onde:

$M_d =$  número total de agregados familiares no distrito d

$n_d =$  número de UPAs seleccionadas no distrito d para o CAP

$\bar{M}_{di} =$  número médio de agregados familiares por UPA no distrito d

Agora podemos usar esta fórmula para calcular o número médio de agregados familiares por UPA dentro do distrito:

$$\bar{M}_{di} = 8 \times \frac{W_{Cdh(p)}}{W_{Cdh(m)}}$$

Para estimar o número total de UPAs dentro do distrito, podemos dividir o número total de agregados no distrito ( $M_d$ ), pela estimativa do tamanho médio da UPA dentro do distrito ( $\bar{M}_{di}$ ), assim:

$$N_d = \frac{M_d}{\bar{M}_{di}}$$

Logo estimamos a probabilidade média da primeira etapa para o estrato da seguinte maneira:

$$p_{1h} = \frac{n_h}{\sum_{d \in h} N_d} = \frac{n_h}{N_h},$$

onde:

$p_{1h}$  = probabilidade de selecção da primeira etapa para as UPAs (segmentos) no estrato h com amostra bi-etápica para o TIA 2002

$n_h$  = número de UPAs seleccionadas no estrato (província por zona agro-ecológica) h para o TIA 2002

$N_h$  = número total aproximado de UPAs no estrato h para o TIA 2002 (a suma dos valores de  $N_d$  para os distritos dentro do estrato)

Para a análise de CENVAR, é necessário gerar códigos apropriados para o estrato e UPA que refletem o desenho da amostra alternativa de 94 distritos amostrais. Para todos os estratos podemos usar o código de quatro dígitos para a província e zona agro-ecológica. Mas a definição de UPA vai ser diferente para os 13 estratos com amostra de três etapas e os 13 estratos com amostra bi-etápica. Usamos um código de UPA de três dígitos. Para os estratos com amostra bi-etápica, copiamos o código de UPA usado para o CAP. Mas no caso dos estratos com amostra de três etapas, geramos o seguinte código de UPA: "0" + código de distrito.

#### **4. Análise dos Erros Padrão Simulados para a Amostra de 94 Distritos para o TIA**

Depois da selecção de 564 segmentos nos 94 distritos amostrais, os dados do CAP para 537

destes segmentos (excluindo os 17 segmentos sem dados) foram usados para o análise de erros padrão simulados baseado neste desenho alternativo do TIA com 94 distritos amostrais. O pacote CENVAR foi usado para calcular os erros padrão; a metodologia de estimação usada por CENVAR está descrita no relatório prévio.

Os resultados deste análise simulado das medidas de precisão e eficiência de amostragem para estimativas de área total por cultura estão apresentados no Anexo I. Comparamos estes resultados com as medidas de precisão simuladas para a amostra baseada em 66 distritos amostrais apresentados no Anexo II do relatório prévio, para determinar o efeito do desenho alternativo na eficiência da amostra. Pode-se ver que em geral os CVs e efeitos do desenho baixaram consideravelmente comparados com os resultados correspondentes da amostra de 66 distritos. Como no análise prévio, examinamos os resultados de CENVAR para as 12 culturas que tinham coeficientes de variação (CVs) menores ou iguais a 15 por cento a nível nacional, baseados na amostra de 66 distritos: milho, arroz, mandioca, pepino, mapira, batata doce, abóbora, feijão nhemba, mexoeira, feijão boer, feijão jugo, e amendoim. Para estas 12 culturas com melhor precisão, a média dos efeitos do desenho baixou de 4.1 a 2.9. A razão entre estes dois indicadores ( $2.9/4.1$ ), 0.71, quer dizer que usando esta amostra de 94 distritos se pode baixar o tamanho da amostra por 29 por cento para obter o mesmo nível de precisão que se vai obter do desenho do TIA baseado na selecção de 66 distritos na primeira etapa. É interessante que por coincidência a média dos efeitos do desenho para estas 12 culturas baseado no análise simulado para um desenho bi-etápico, apresentada no relatório prévio, resultou em quase o mesmo valor de 2.9. O Quadro 4 compara os resultados dos efeitos do desenho para estas três amostras alternativas. Pode-se ver neste quadro que para algumas culturas o efeito do desenho é mais baixo para a amostra bi-etápica, e para outras é mais baixo para a amostra de 94 distritos, dado a variabilidade de amostragem, ainda que a média é quase o mesmo baixo estas duas alternativas.

Quadro 4. Comparação dos Efeitos do Desenho (DEFF) para as 12 Culturas com Maior Precisão, Baseado no Análise de Simulação para as Três Amostras Alternativas para o TIA 2002

Cultura	Efeito do Desenho (DEFF)		
	Amostra de 66 Distritos	Amostra Bi-etápica	Amostra de 94 Distritos
Milho	4.77	2.81	3.18
Arroz	6.21	4.63	3.74
Mandioca	6.89	3.64	3.75
Pepino	1.82	2.97	4.72
Mapira	3.24	3.12	2.47
Batata Doce	2.11	2.10	1.99
Abóbora	3.66	1.87	2.75
Feijão Nhemba	3.06	2.43	2.17
Mexoeira	3.20	2.60	2.22
Feijão Boer	5.39	3.24	3.06
Feijão Jugo	3.05	2.41	1.94
Amendoim	5.88	3.48	3.01
Média	4.11	2.94	2.92

Também se pode notar no Anexo I que 13 culturas (incluindo quiabo) têm um CV menor que 15 por cento a nível nacional, comparado com 12 culturas para a amostra de 66 distritos, e 17 para a amostra bi-etápica. Isto também ilustra que a eficiência da amostra de 94 distritos seria entre aquelas das outras duas alternativas, tomando em conta a variabilidade de amostragem. A nível de cultura por província, encontramos 20 casos em que o CV baixa a menos de 15 por cento quando se vai de 66 a 94 distritos amostrais, comparado com 16 casos quando se vai de uma amostra de 66 distritos a uma amostra bi-etápica.

A conclusão deste estudo comparativo entre as três amostras alternativas para o TIA é que aumentando a número de distritos seleccionados de 66 a 94, e usando uma amostra bi-etápica nos 13 estratos menores, o ganho em eficiência de amostragem é similar àquela obtida de uma amostra bi-etápica dispersa em 132 distritos. Isto é parcialmente devido à amostra bi-etápica nos 13 estratos menores, que representam um terço dos agregados familiares rurais. Outra razão desta similaridade em eficiência é que para a amostra de três etapas os distritos são seleccionados com probabilidade proporcional ao tamanho; por isso a maioria dos distritos adicionais para a amostra bi-etápica seriam as mais pequenas, com sómente um ou dois segmentos seleccionados. Entretanto, é importante notar que no cálculo dos erros padrão simulados para a amostra alternativa bi-etápica, não incluímos factores de ajuste à variância por população finita na primeira etapa. Ao mesmo tempo, factores de ajuste foram usados no análise

simulado para a amostra de 94 distritos, inclusive para os 13 estratos com amostra bi-etápica, como explicado na Secção 3. A média dos factores de ajuste por população finita na primeira etapa para estes 13 estratos mais pequenos foi 0.936, que representaria uma redução das CVs de aproximadamente 3.2 por cento para estes estratos. Isto é, se tivéssimos incluído factores de ajuste por população finita na primeira etapa no cálculo de erros padrão simulados para a amostra alternativa bi-etápica, os CVs e os efeitos do desenho correspondentes diminuiriam um pouco para esta alternativa. Dado que os resultados de cada amostra são sujeitos a variabilidade amostral também, não podemos concluir que a eficiência da amostra de 94 distritos seria igual àquela da amostra bi-etápica, mas o ganho em eficiência comparado com a amostra de 66 distritos é similar.

Outro comentário importante sobre a amostra alternativa de 94 distritos é que no caso de alguns estratos com amostra de três etapas, quase todos os distritos foram seleccionados com probabilidade proporcional ao tamanho na primeira etapa, e em alguns casos a amostra resultante não vai ser auto-ponderada para as pequenas explorações dentro do estrato. Para ilustrar este problema, podemos examinar o estrato 0709 em Sofala, onde três dos quatro distritos foram seleccionados. O Quadro 5 apresenta o número total de agregados familiares nos distritos do estrato 0709.

Quadro 5. Número Total de Agregados Familiares para os Quatro Distritos no Estrato 0709 de Sofala, Intervalo de Selecção da Primeira Etapa, e Distritos Seleccionados

Código de Distrito	Nome do Distrito	Número Total de Agregados Familiares	Seleccionados para o TIA (S)
06	Chibabava	15,899	S
08	Gorongosa	15,716	S
12	Muanza	2,861	
13	Nhamatanda	30,321	S
Intervalo de Selecção para 3 Distritos		21,599	

Pode-se ver no Quadro 5 que o distrito 13 de Sofala, Nhamatanda, tem um total de 30,321 agregados familiares; então sua medida de tamanho é maior que o intervalo de selecção (21,599). Por isso este distrito seria auto-representado, e pode ser seleccionado duas vezes usando uma selecção sistemática proporcional ao tamanho; neste caso, seria necessário dobrar o número de segmentos seleccionados neste distrito a 12 para manter a amostra auto-representada. Se tratamos este distrito como auto-representado e seleccionamos 6 segmentos, os factores de expansão para este distrito seriam diferentes, e também complicaria a metodologia de estimação das variâncias. Por esta razão os resultados do análise simulado não são exactos para este estrato e os outros com quase todos os distritos seleccionados. Para resolver este problema, se recomenda seleccionar o resto dos distritos nestes estratos, e usar uma amostra bi-etápica como

nos outros 13 estratos menores. Isto significaria seleccionar um distrito mais nos seguintes estratos: 0709, 0915 e 1014. Neste caso a amostra aumentaria a 97 distritos, mas não seria necessário aumentar o número de segmentos seleccionados nestes três estratos. Por exemplo, para o estrato 0709, seria possível seleccionar uma amostra bi-etápica de 18 segmentos entre os quatro distritos (em vez de aumentar a amostra a 24 segmentos).

Outra conclusão importante é que se decidem usar a amostra alternativa de 94 distritos, seria muito recomendável modificar esta alternativa para incluir todos os distritos nos estratos 0709, 0915 e 1014, e seleccionar uma amostra bi-etápica nestes estratos (para uma amostra total de 97 distritos). No caso do estrato 0711, um dos distritos seria auto-representado, e também seria melhor seleccionar todos os cinco distritos neste estrato e usar uma amostra bi-etápica (isto é, incluindo os dois distritos restantes); este caso não seria tão crítico como os outros três estratos, e esta opção aumentaria a amostra a 99 distritos.