

Recomendações para o Desenho da Amostra para o Trabalho de Inquérito Agrícola (TIA) 2003

David J. Megill
Consultor de Amostragem

Fevereiro de 2002

1. Introdução

O desenho da amostra para o Trabalho de Inquérito Agrícola (TIA) 2002 está documentado no relatório “Análise de Erros Padrão Simulados para os Resultados do Trabalho de Inquérito Agrícola 2002 Baseados numa Amostra de 80 Distritos para Moçambique” (Megill, Abril de 2002). O estudo da precisão simulada baseada no desenho da amostra do TIA em 80 distritos foi feita usando um análise de CENVAR com os dados do Censo Agro-Pecuário (CAP). A conclusão deste estudo foi que muitos dos coeficientes de variação (CVs), ou erros relativos de amostragem, para a estimativa de área total por cultura a nível de província eram altas. Mesmo a nível nacional sómente 12 culturas tinham CVs menores de 15 por cento. Agora que os dados do TIA 2002 já estão disponíveis, podemos fazer mais análises de CENVAR para estudar a precisão dos resultados do próprio inquérito. Isso é importante não sómente para medir a precisão dos resultados do inquérito para os usuários, mas também para estudar a possibilidade de melhorar a precisão e eficiência da amostragem para o TIA no futuro.

De acordo ao estudo da precisão simulada, a eficiência estatística da amostra de 80 distritos para o TIA 2002 melhorou-se muito comparado com a amostra original de 66 distritos. Mesmo assim, recomenda-se aumentar um pouco mais a dispersão e tamanho da amostra para o TIA 2003 para melhorar ainda mais a eficiência da amostragem. Sómente assim podemos melhorar a precisão dos resultados a nível de província. Dado que a amostra do TIA 2003 também vai depender dos recursos disponíveis para levar a cabo o inquérito e a logística operativa correspondente, o Ministério de Agricultura e Desenvolvimento Rural (MADER) tem que decidir a possibilidade de aumentar o número de distritos seleccionados para a amostra baseado nestas recomendações. Outra consideração importante é a sustentabilidade da amostra do TIA para o futuro, porque seria necessário continuar com este desenho de amostragem cada ano para manter a comparabilidade dos resultados anuais.

O propósito deste relatório é de documentar o desenho da amostra recomendada para o TIA 2003, baseado em 94 distritos amostrais e uma média de 7 distritos seleccionados por distrito. Uma consideração importante é a sobre-posição máxima dos segmentos seleccionados para o TIA 2002 e TIA 2003, para melhorar a precisão de estimativas de diferenças anuais e assegurar consistência nos resultados cada ano. Este relatório também descreve o nível de precisão que pode-se esperar para os resultados do TIA 2003 baseado na amostragem recomendada.

2. Atribuição da Amostra Recomendada para o TIA 2003 Baseada em 94 Distritos

Para aumentar a eficiência da amostragem para o TIA 2003, será necessário dispersar mais a amostra comparado com o desenho de 80 distritos amostrais para o TIA 2002. Estudos de simulação da precisão para os resultados do TIA (usando dados do CAP) foram feitos para as alternativas de 66, 80 e 94 distritos amostrais, como também para uma amostra bi-etápica, que inclui UPAs seleccionadas em 132 distritos (quase todos). A conclusão destes estudos foi que a amostra bi-etápica proporcionaria uma eficiência estatística máxima, mas também implicaria um maior custo para o trabalho de campo e complicaria a logística operativa. Comparando os resultados da simulação para as outras alternativas estudadas para a amostra do TIA, encontramos que a eficiência da amostragem melhorou significativamente quando aumentamos o número de distritos amostrais de 80 a 94.

O desenho de uma amostra de 94 distritos para o TIA está descrito no relatório “Análise de Erros Padrão Simuladas para os Resultados do Trabalho de Inquérito Agrícola 2002 Baseados numa Amostra de 94 Distritos para Moçambique” (Megill, Abril de 2002). Para este estudo de simulação, uma média de 6 segmentos foram seleccionados por distrito e 8 pequenas explorações por segmento, que resultou em uma amostra total de 564 segmentos e 4.512 pequenas explorações. Para este estudo queríamos um tamanho de amostra comparável à alternativa de 66 distritos amostrais. Mas para o TIA 2003, outra maneira de melhorar a precisão dos resultados a nível de província seria de aumentar um pouco o tamanho da amostra. Por isso recomenda-se aumentar o número de segmentos seleccionados por distrito a 7, que também seria consistente com o desenho para o TIA 2002 em 80 distritos amostrais. Comparando o tamanho da amostra do TIA 2002 com o tamanho correspondente recomendado para o TIA 2003, o número de segmentos seleccionados aumentaria de 560 a 658, e o número de pequenas explorações seleccionadas aumentaria de 4.480 a 5.264, ou seja um aumento de 17,5 por cento.

A amostra de 94 distritos amostrais recomendada para o TIA 2003 incluiria os mesmos 80 distritos amostrais incluídos na amostra do TIA 2003, e 14 novos distritos identificados no Quadro 1. Estes distritos são os mesmos que aqueles incluídos no estudo de precisão simulado para a amostra de 94 distritos (mas a amostra de segmentos é diferente). O Quadro 1 também indica o tipo de amostra (bi-etápica ou tri-etápica) nos estratos correspondentes aos novos distritos. Cinco estratos tri-etápicos na amostra do TIA 2002 passariam a ser bi-etápicos na amostra recomendada para o TIA 2003. A lista dos 14 novos distritos foi usada para estimar os custos relacionados à inclusão destes distritos na amostra do TIA 2003.

A atribuição da amostra de segmentos e pequenas explorações recomendada para o TIA 2003 está apresentada no Quadro 2. Este quadro também mostra a atribuição da amostra do TIA 2002 para propósitos de comparação.

Quadro 1. Lista dos 14 Distritos Adicionais Incluídos no Amostra do TIA 2003, e Tipo de Amostragem nos Estratos Correspondentes

| Província | Estrato | Nome do Distrito | Tipo de Amostragem no Estrato |
|--------------|---------|-------------------|-------------------------------|
| Niassa | 0101 | Lago | Tri-etápico |
| Cabo Delgado | 0210 | Ancuabe | Tri-etápico |
| Nampula | 0306 | Malema | Bi-etápico |
| Nampula | 0310 | Mossuril | Tri-etápico |
| Nampula | 0310 | Nampula | Tri-etápico |
| Zambézia | 0411 | Chinde | Tri-etápico |
| Tete | 0508 | Chifunde | Bi-etápico |
| Sofala | 0709 | Gorongosa | Tri-etápico |
| Sofala | 0712 | Caia | Bi-etápico |
| Inhambane | 0813 | Homoine | Tri-etápico |
| Inhambane | 0814 | Mabote | Bi-etápico |
| Gaza | 0913 | Cidade de Xai-Xai | Bi-etápico |
| Gaza | 0915 | Chicualacuala | Tri-etápico |
| Maputo | 1014 | Marracuene | Tri-etápico |

Quadro 2. Atribuição da Amostra por Estrato Recomendado para o TIA 2003 com 94 Distritos Amostrais: Distritos, Segmentos e Pequenas Explorações

| Província | Estrato | No. Distritos CAP | No. Distritos TIA | No. Segmentos TIA | Número de Pequenas Explorações Amostra TIA 2003 | Número de Pequenas Explorações Amostra TIA 2002 |
|----------------------|---------|-------------------|-------------------|-------------------|---|---|
| Niassa | 0101 | 5 | 3 | 21 | 168 | 112 |
| | 0106 | 11 | 3 | 21 | 168 | 168 |
| Cabo Delgado | 0205 | 2 | 2 | 14 | 112 | 112 |
| | 0206 | 2 | 2 | 14 | 112 | 112 |
| | 0210 | 13 | 6 | 42 | 336 | 280 |
| Nampula | 0306 | 3 | 3 | 21 | 168 | 112 |
| | 0310 | 15 | 9 | 63 | 504 | 392 |
| | 0311 | 2 | 2 | 14 | 112 | 112 |
| Zambézia | 0404 | 3 | 3 | 21 | 168 | 168 |
| | 0407 | 5 | 5 | 35 | 280 | 280 |
| | 0411 | 9 | 6 | 42 | 336 | 280 |
| Tete | 0502 | 3 | 3 | 21 | 168 | 168 |
| | 0508 | 3 | 3 | 21 | 168 | 112 |
| | 0512 | 7 | 3 | 21 | 168 | 168 |
| Manica | 0603 | 5 | 3 | 21 | 168 | 168 |
| | 0609 | 3 | 3 | 21 | 168 | 168 |
| | 0614 | 2 | 2 | 14 | 112 | 112 |
| Sofala | 0709 | 4 | 3 | 21 | 168 | 112 |
| | 0711 | 5 | 3 | 21 | 168 | 168 |
| | 0712 | 3 | 3 | 21 | 168 | 112 |
| Inhambane | 0813 | 11 | 6 | 42 | 336 | 280 |
| | 0814 | 3 | 3 | 21 | 168 | 112 |
| Gaza | 0913 | 3 | 3 | 21 | 168 | 112 |
| | 0914 | 5 | 3 | 21 | 168 | 168 |
| | 0915 | 4 | 3 | 21 | 168 | 112 |
| Maputo- Província | 1014 | 7 | 6 | 42 | 336 | 280 |
| Moçambique | | 138 | 94 | 658 | 5,264 | 4,480 |

No caso dos estratos que aparecem no Quadro 2 com o mesmo número de distritos na amostra para o TIA 2003 e na base do CAP, a amostra é bi-etápica; os segmentos nestes estratos foram seleccionados sistematicamente na primeira etapa. Pode-se ver no Quadro 2 que 13 dos 26 estratos teriam uma amostra bi-etápica para o TIA 2003. Os outros 13 estratos teriam uma amostra tri-etápica, com distritos seleccionados na primeira etapa. No caso da amostra do TIA 2002 em 80 distritos, sómente 8 dos 26 estratos tinham amostra bi-etápica. Então cinco estratos adicionais passam a ser bi-etápicas para o TIA 2003, requerindo uma nova amostra de segmentos.

3. Sobreposição das Amostras de Segmentos para TIA 2002 e TIA 2003

Para propósitos de eficiencia estatística e operativa, é recomendável ter uma máxima sobreposição da amostra para os inquéritos anuais do TIA. A correlação entre as amostras anuais vai melhorar a precisão das estimativas de diferenças dos dados do TIA para dois anos consecutivos. Uma metodologia que facilitaria a operação do campo para o TIA 2003 seria de manter na amostra do TIA 2003 a maioria dos segmentos seleccionados para o TIA 2002. Em cada segmento seleccionado seria necessário actualizar a listagem, e seleccionar uma nova amostra de explorações. Para o futuro, também pode-se considerar uma rotação parcial (por exemplo, aproximadamente 20 por cento) dos segmentos amostrais do TIA de um ano ao outro, para evitar que os mesmos agregados familiares sejam inqueridos várias vezes.

Para manter o número máximo de segmentos seleccionados para o TIA 2002 na amostra para o TIA 2003, primeiro foi necessário examinar o tipo de amostragem em cada estrato. No caso dos 13 estratos com amostragem tri-etápica para o TIA 2003, podemos usar os mesmos sete segmentos seleccionados para o TIA 2002 em cada distrito incluído em ambos inquéritos; os distritos novos para o TIA 2003 requerem uma nova amostra de segmentos. Os oito estratos com amostra bi-etápica para o TIA 2002 continuam sendo bi-etápicas para o TIA 2003, e também podemos manter os mesmos segmentos na amostra.

No caso dos cinco estratos com amostragem tri-etápica para o TIA 2002 e bi-etápica para o TIA 2003, foi necessário fazer uma nova selecção de segmentos, usando amostragem sistemático proporcional à razão M_h/n_h (o número total de agregados familiares no distrito dividido pelo número de UPAs seleccionadas no distrito para o CAP). Esta metodologia de amostragem para os estratos bi-etápicas, descrito no relatório “Análise de Erros Padrão Simuladas para os Resultados do Trabalho de Inquérito Agrícola 2002 Baseados numa Amostra de 94 Distritos para Moçambique”, foi necessário para obter uma amostra auto-ponderada para as pequenas explorações dentro do estrato. Nestes estratos com amostragem bi-etápica o número de segmentos seleccionados varia por distrito de acordo a seu tamanho.

O Quadro 3 apresenta a distribuição dos segmentos seleccionados para o TIA 2003 de acordo ao tipo de estrato, e a sobre-posição com os segmentos seleccionados para o TIA 2002. Pode-se ver neste quadro que um total de 520 dos 560 segmentos na amostra do TIA 2002 seriam mantidos na amostra para o TIA 2003. Este nível de sobre-posição vai assegurar uma boa correlação entre os resultados do TIA 2002 e TIA 2003. Um total de 138 segmentos novos foram seleccionados para o TIA 2003 para completar a amostra total de 658 segmentos.

Quadro 3. Distribuição dos Segmentos Seleccionados para o TIA 2003, de Acordo ao Tipo

de Estrato e Inclusão no TIA 2002

| Tipo de Amostragem no Estrato | Número de Segmentos Incluídos no TIA 2002 e 2003 | Número de Segmentos Novos no TIA 2003 |
|------------------------------------|--|---------------------------------------|
| Tri-etápico 2002, Tri-etápico 2003 | 336 | 63 |
| Bi-etápico 2002, Bi-etápico 2003 | 154 | 0 |
| Tri-etápico 2002, Bi-etápico 2003 | 30* | 75 |
| Total | 520 | 138 |

* Estes 30 segmentos na amostra independente seleccionada para o TIA 2003 coincidiram com segmentos seleccionados para o TIA 2002 nestes estratos com amostragem tri-etápica para o TIA 2002 e bi-etápica para o TIA 2003. Esta sobre-posição depende da amostra individual.

4. Comparação da Eficiência das Amostras para TIA 2002 e TIA 2003

A precisão das estimativas baseadas no desenho da amostra recomendada para o TIA 2003 deve ser significativamente melhor que a precisão para os resultados correspondentes do TIA 2002, dado a maior dispersão da amostra em 94 distritos e também o aumento no tamanho da amostra de 560 a 658 segmentos amostrais, ou seja 17.5 por cento. Os resultados dos estudos de precisão simulada para as amostras alternativas do TIA de 80 e 94 distritos amostrais podem ser usados para examinar o melhoramento na precisão das estimativas do TIA 2003 comparados com os resultados correspondentes do TIA 2002. Um resultado destes estudos foi que para as 12 culturas principais, a média dos efeitos do desenho para a estimativa de área total foi 3.56 para a amostra de 80 distritos e 2.92 para o desenho alternativo de 94 distritos amostrais. O efeito do desenho para uma estimativa do inquérito é definido como a variância da estimativa baseado no desenho da amostra complexa (estratificada multi-etápica) dividida pela variância correspondente de uma amostra aleatória simples do mesmo tamanho; é uma medida da eficiência relativa do desenho da amostra.

Para estimar o efeito neto do desenho recomendado para o TIA 2003 na precisão das estimativas do próximo inquérito comparado com os resultados correspondentes dos dados do TIA 2002, podemos usar a seguinte relação para a variância:

$$var(\hat{X})_{03} = var(\hat{X})_{02} \times \frac{DEFF_{03}}{DEFF_{02}} \times \frac{n_{02}}{n_{03}} = var(\hat{X})_{02} \times \frac{2.92}{3.56} \times \frac{560}{658} = 0.698 \times var(\hat{X})_{02} ,$$

onde:

$var(\hat{X})_{03} =$ variância aproximada da estimativa de um total (como área em culturas) baseada no desenho de amostragem recomendado para o TIA 2003

$var(\hat{X})_{02} =$ variância da estimativa de um total (como área em culturas) da amostra para o TIA 2002 de 80 distritos amostrais, calculada dos dados do TIA 2002 ou do estudo de simulação

$DEFF_{03} =$ efeito do desenho para a amostra de 94 distritos recomendada para o TIA 2003, obtida do estudo de simulação; a média de $DEFF_{03}$ calculada no estudo de simulação para as estimativas de área total para as 12 culturas principais foi 2.92

$DEFF_{02} =$ efeito do desenho para a amostra de 80 distritos para o TIA 2002, calculado dos dados do TIA 2002 ou do estudo de simulação; a média de $DEFF_{02}$ calculada no estudo de simulação para as estimativas de área total para as 12 culturas principais foi 3.56

$n_{02} = 560 =$ número de segmentos seleccionados para a amostra do TIA 2002

$n_{03} = 658 =$ número de segmentos seleccionados para a amostra do TIA 2003

O erro padrão da estimativa é igual ao raiz quadrado da variância, e pode ser expressada assim:

$$s(\hat{X})_{03} = \sqrt{0.698 \times var(\hat{X})_{02}} = 0.836 \times s(\hat{X})_{02} ,$$

onde:

$s(\hat{X})_{03} =$ erro padrão aproximado da estimativa de um total (como área em culturas) baseado no desenho de amostragem recomendado para o TIA 2003

$s(\hat{X})_{02} =$ erro padrão da estimativa de um total (como área em culturas) da amostra para o TIA 2002 de 80 distritos, calculado dos dados do TIA 2002 ou do estudo de simulação

A última expressão indica que em geral o novo desenho recomendado para o TIA 2003 pode reduzir os erros padrão (e as CVs) das estimativas do inquérito por mais de 16 por cento

comparado com a precisão dos resultados correspondentes do TIA 2002. Estes cálculos servem sómente para ilustrar o aumento em precisão que pode-se esperar em geral; os resultados reais podem variar para as diferentes culturas e províncias.

5. Considerações para Melhorar a Eficiência da Amostra do TIA 2003 para Fruteiras, Bovinos ou Outras Características Menos Frequentes

Agora que os dados do TIA 2002 estão disponíveis, é muito importante seguir com os análises de CENVAR para examinar a precisão das estimativas mais importantes. Como foi indicado na secção prévia, o novo desenho recomendado para o TIA 2003 deve melhorar o nível de precisão das estimativas em geral, reduzindo as CVs por uma média de mais de 16 por cento. Mas no caso de algumas características importantes a nível de província, como por exemplo algumas espécies de pecuária (como bovinos) e fruteiras que se encontram com menos frequência em várias províncias, pode ser que o nível de precisão que vai resultar do desenho da amostra para o TIA 2003 não seja satisfatória. As fruteiras que estão recebendo atenção especial são banana, citrinos, mangueiras e ananás. Neste caso a única maneira de aumentar o nível de precisão seria através de uma estratégia de amostragem especial.

Para aumentar a precisão de certas características importantes como fruteiras ou bovinos, respeitando a amostra recomendada de 94 distritos para o TIA 2003, a metodologia mais eficaz seria de identificar as explorações com estas características dentro dos segmentos seleccionados, através da listagem. Seria necessário incluir na listagem perguntas específicas para identificar as explorações com as culturas ou pecuária de interesse especial. Seria eficiente estabelecer um critério mínimo para a área ou árvores de fruteiras, ou o número de bovinos (que já temos na definição de médias explorações), para limitar o número de entrevistas adicionais. Neste caso, pode-se tratar as explorações identificadas com estas características na listagem como se tratam as médias e grandes explorações na amostra do TIA 2002. Isto é, estas explorações podem ser incluídas com certeza dentro dos segmentos seleccionados. Para propósitos de estimação, os factores de expansão para estas explorações seriam calculadas da mesma forma que os factores para as médias explorações; teriam a mesma probabilidade de selecção do segmento.

Ao mesmo tempo é importante examinar o efeito desta metodologia nos custos do inquérito, dado que vai aumentar o número de entrevistas. Por exemplo, se a maioria das pequenas explorações dentro de um segmento seleccionado tem as características de interesse (fruteiras ou bovinos), o número de entrevistas para o TIA 2003 neste segmento pode ser muito alto. Neste caso, pode-se estabelecer um número máximo (por exemplo, 10) explorações com as características especiais que seriam seleccionados na amostra para o segmento. Se a listagem identifica mais explorações com as características especiais que este mínimo, seria necessário seleccionar este número máximo das explorações listadas. Por exemplo, se 35 pequenas explorações com as fruteiras de interesse são listadas dentro do segmento, pode-se seleccionar 10 pequenas explorações desta base de 35, à parte da selecção de 8 pequenas explorações para a amostra básica. Este procedimento também requeriria um tratamento especial para calcular os factores de expansão baseado nas probabilidades de selecção.

Uma alternativa que pode-se considerar para limitar os custos relacionados a esta estratégia de amostragem especial seria de limitar este procedimento de listagem e selecção de explorações com características de interesse a certas províncias onde seria necessário melhorar a precisão das estimativas correspondentes.

Considerando o aumento nos custos do inquérito e na complexidade de estimação, sómente deve-se considerar esta estratégia de amostragem especial se a informação para as fruteiras e outras características de interesse (como bovinos) tem a importancia correspondente. Se o MADER está considerando a possibilidade de usar esta alternativa para a amostragem do TIA 2003, os seguintes passos são recomendados:

- (1) Deve-se primeiro fazer análises de CENVAR para as estimativas de fruteiras (banana, citrinos, mangueiras e ananás), bovinos e outras características consideradas importantes, usando com os dados do TIA 2002.
- (2) Identificar estimativas de fruteiras ou bovinos por província considerados importantes que não tem precisão satisfatório (por exemplo, CV superior a 20 por cento para um total).
- (3) Estimar os custos da estratégia de amostragem especial para melhorar a precisão das estimativas consideradas mais importantes.
- (4) Especificar a metodologia da listagem necessária para usar esta estratégia de amostragem especial. A listagem deve incluir perguntas para identificar as explorações com características especiais, e os manuais devem ter instruções de como tratar a selecção destas explorações.