

Modificação da Metodologia para o Cálculo de Factores de Expansão para os Dados do Trabalho de Inquérito Agrícola 2002

David J. Megill
Consultor de Amostragem

Fevereiro de 2003

1. Introdução

A metodologia preliminar de estimação para os dados do Trabalho de Inquérito Agrícola (TIA) 2002 está documentada no relatório “Cálculo de Factores de Expansão para os Dados do Trabalho de Inquérito Agrícola 2002” (Megill, Dezembro de 2002). Os factores de expansão correspondentes foram calculados usando um ficheiro de Excel. Depois de produzir quadros preliminares dos resultados do TIA 2002 usando estes factores de expansão, fizemos uma avaliação da estimativa total ponderada de explorações, usando os dados do Censo Agro-Pecuário (CAP) e do II Recenseamento Geral da População e Habitação (RGPH) de 1997. Basado neste estudo elaboramos uma metodologia mais confiável para estimar as projecções de agregados familiares para 2002 usados para calcular os factores de expansão das pequenas explorações na amostra do TIA 2002. Usamos esta metodologia para calcular novos factores de expansão para as pequenas explorações.

O propósito deste relatório é de avaliar os factores de expansão preliminares do TIA 2002 e especificar modificações recomendadas para o cálculo dos factores de expansão finais para as pequenas explorações. Os factores de expansão para as médias e grandes explorações especificados no relatório prévio foram basados no desenho da amostra, e não dependem das projecções. Por isso não é necessário modificar os factores de expansão para as médias e grandes explorações especificados antes. Este relatório deve ser acompanhado com o relatório “Cálculo de Factores de Expansão para os Dados do Trabalho de Inquérito Agrícola 2002” para completar a documentação da metodologia de estimação para o TIA 2002.

2. Comparação do Número Total de Explorações do TIA 2002 com Resultados do CAP

Uma comparação da estimativa preliminar do número total ponderado de explorações do TIA 2002 (3.790.000) com a estimativa correspondente do CAP (3.076.000) indicou que os factores de expansão preliminares do TIA podem estar sobre-estimando os totais correspondentes. Também revisamos quadros dos dados do RGPH 1997 e as projecções de população elaboradas pelo Instituto Nacional de Estatística (INE). Depois de uma discussão interna em MADER, identificamos duas fontes de tendenciosidade possíveis nos factores de expansão preliminares das pequenas explorações na amostra do TIA 2002. O primeiro factor é relacionado à estimativa prévia que usamos para a projecção do número total de agregados familiares em 2002 por estrato. O factor de crescimento aplicado ao número de agregados familiares do RGPH 1997 foi baseado na razão entre a projecção da população para 2002 e a população recenseada em 1997.

Uma investigação no INE indicou que as projecções de população tomam em conta a omissão estimada do censo. O uso dos dados do RGPH 1997 não ajustados no denominador do factor de crescimento resultou em uma sobre-estimação na projecção do total de agregados familiares.

A outra fonte de tendenciosidade na projecção do número total de agregados familiares é a diferença significativa entre a média de pessoas por agregado familiar estimada dos dados do TIA 2002 (4,9) e a cifra correspondente do RGPH 1997 (4,1 a nível nacional). Outros inquéritos também indicam tamanho médio de agregados familiares mais alto que o censo, inclusive o Inquérito aos Agregados Familiares (IAF) de 2002-03, que também mediu uma média de 4,9 pessoas por agregado familiar (a nível nacional). Isto indica que o conceito de agregado familiar poderia ser aplicado de uma forma um pouco diferente no censo e nos inquéritos, resultando em mais agregados familiares no RGPH 1997.

No caso do CAP, a média de pessoas por agregado familiar (5,3) também era mais alto que a média correspondente do censo, que poderia introduzir uma tendenciosidade similar nos dados ponderados do CAP. Ao mesmo tempo, encontramos uma tendenciosidade no sentido contrário que afeitaram os factores de expansão para o CAP. O número de explorações na base de amostragem do RGPH 1997 usado para calcular os factores de expansão para o CAP foi estimado usando a informação do censo sobre machambas. Nas áreas rurais, os dados do censo indicam que sómente 92 por cento dos agregados familiares em áreas rurais têm machambas, comparado com mais de 99 por cento de acordo aos dados do TIA 2002. Esta tendenciosidade reduziu o efeito da média elevada de pessoas por agregado familiar nos dados do CAP. O efeito neto das duas tendenciosidades opostas resultou em estimativas ponderadas do CAP consistentes com as novas estimativas ponderadas do TIA 2002, como se pode ver mais tarde no Quadro 1.

Basado neste estudo elaboramos uma nova metodologia para calcular o factor de expansão para as pequenas explorações na amostra do TIA 2002 que pode reduzir a tendenciosidade correspondente. Este novo factor é baseado na projecção da população em 2002 por estrato (do análise demográfico do INE) e no número pessoas nas pequenas explorações seleccionadas.

3. Novo Factor de Expansão para as Pequenas Explorações

Para compreender melhor a metodologia de ponderação, podemos primeiro estimar a projecção do número total de agregados familiares em cada estrato, usando a seguinte fórmula:

$$\hat{M}_{02h} = M_{97h} \times \frac{\hat{P}_{02h}}{\hat{P}_{97h}} \times \frac{\bar{p}_{97h}}{\bar{p}_{TIA02h}} = M_{97h} \times \frac{\hat{P}_{02h}}{\hat{P}_{97h}} \times \frac{\frac{\hat{P}_{97h}}{M_{97h}}}{\frac{p_{TIA02h}}{m_{TIA02h}}} = \frac{\hat{P}_{02h}}{\bar{p}_{TIA02h}},$$

onde:

\hat{M}_{02h} = projecção do número total de agregados familiares dentro do estrato h para o período do TIA 2002

M_{97h} = número de agregados familiares da base do RGPH 1997 para o estrato h

\hat{P}_{02h} = projecção da população total de 2002 para o estrato h, baseado no análise demográfico do INE, ajustada para omissão censal

\hat{P}_{97h} = projecção da população total de 1997 para o estrato h, baseado no análise demográfico do INE, ajustada para omissão censal

$\bar{p}_{97h} = \frac{\hat{P}_{97h}}{M_{97h}}$, média de pessoas por agregado familiar para o estrato h derivado dos dados do RGPH 1997, ajustado para omissão censal

$\bar{p}_{TIA02h} = \frac{p_{TIA02h}}{m_{TIA02h}}$, estimativa da média de pessoas por agregado familiar para o estrato h derivado dos dados do TIA 2002, assumindo que a amostra dentro do estrato é auto-ponderado

p_{TIA02h} = número total de pessoas (não ponderado) nos agregados familiares seleccionados para o TIA 2002 no estrato h

m_{TIA02h} = número total de agregados familiares seleccionados para o TIA 2002 no estrato h

No caso do TIA 2002, as pequenas explorações são auto-ponderadas dentro de cada estrato (quer dizer, têm o mesmo factor de expansão básico dentro do estrato). Para estimar a parte da projecção de população para 2002 que corresponde às pequenas explorações (\hat{P}_{02h}), temos que subtrair a população nas médias e grandes explorações, estimada dos dados ponderados do TIA 2002. Usando a expressão para a projecção de agregados familiares no estrato para 2002, \hat{M}_{02h} , podemos simplificar o cálculo do factor de expansão para as pequenas explorações dentro do estrato h assim:

$$W_{h(p)} = \frac{\hat{P}_{02h} \& \hat{P}_{02h(mg)}}{m_{TIA02h}^{)}} \cdot \frac{\frac{p_{TIA02h}^{)}}{m_{TIA02h}^{)}}}{\frac{\hat{P}_{02h} \& \hat{P}_{02h(mg)}}{p_{TIA02h}^{)}}},$$

onde:

$\hat{M}_{02h(mg)}$ = estimativa do número total de médias e grandes explorações no estrato h, baseada nos dados ponderados do TIA 2002

$m_{TIA02h}^{)}$ = número de explorações pequenas seleccionadas no estrato h para o TIA 2002

$\hat{P}_{02h(mg)}$ = estimativa da população total nas médias e grandes explorações no estrato h, derivada dos dados ponderados do TIA 2002

$p_{TIA02h}^{)}$ = número de pessoas nas pequenas explorações seleccionadas no estrato h para o TIA 2002

O componente da população aproximado correspondente às médias e grandes explorações ($\hat{P}_{02h(mg)}$) pode ser estimada multiplicando o número ponderado de médias e grandes explorações pela média de pessoas por agregado familiar. A estimativa da média de pessoas por agregado familiar para as médias explorações nos dados do TIA 2002 é aproximadamente 9, quase o dobro da estimativa correspondente para as explorações pequenas. Dado o número elevado de pessoas para estas explorações e o número relativamente baixo de médias explorações na amostra para alguns estratos, usamos os dados para todas as explorações na amostra dentro do estrato, incluindo as pequenas, para estimar a média de pessoas por agregado familiar usada no cálculo do componente $\hat{P}_{02h(mg)}$. Este foi um pequeno ajuste para estabilizar os factores, mais não deve afeitar muito os resultados finais.

Dado que não temos informação sobre pessoas para todas as explorações pequenas na amostra (nem para os agregados familiares sem machamba), podemos estimar o componente $p_{TIA02h}^{)}$ assim:

$$p_{TIA02h}^{)} = m_{TIA02h}^{)} \times \bar{p}_{TIA02h}^{)},$$

onde:

$m_{TIA02h}^{)} =$ número de pequenas explorações (e agregados familiares sem machamba) seleccionados no estrato h =
(número de distritos na amostra do TIA 2002 para estrato h) x 7 x 8

$\bar{p}_{TIA02h}^{)} =$ estimativa da média de pessoas por agregado familiar dos dados para as pequenas explorações na amostra para o estrato h que completaram o inquérito

Como indicado no relatório prévio, também é necessário ajustar os factores de expansão para as pequenas explorações a nível de segmento para tomar em conta as pequenas explorações não inqueridas. Usando a mesma simbologia, o factor de expansão final ($W'_{hij(p)}$) para as pequenas explorações dentro de cada segmento amostral pode ser expressada da seguinte maneira:

$$W'_{hij(p)} = W_{h(p)} \times \frac{m_{hij(p)}}{m_{hij(p)}^{)}$$

onde:

$m_{hij(p)} =$ número de pequenas explorações (excluindo agregados familiares sem actividades agro-pecuárias) seleccionadas no segmento amostral j do distrito amostral i no estrato h

$m'_{hij(p)} =$ número de pequenas explorações com questionários do TIA completados no segmento amostral j do distrito amostral i no estrato h

4. Resultados da Nova Metodologia de Estimação para o TIA 2002

Aplicando esta nova metodologia para calcular os factores de expansão para as pequenas explorações na amostra do TIA 2002, o número ponderado de explorações diminuiu significativamente, devido principalmente à diferença entre a média de pessoas por agregado familiar dos dados do TIA 2002 e a cifra correspondente do RGPH 1997. Quadro 1 apresenta a nova estimativa ponderada do número total de explorações e a população total nos estratos de Moçambique incluídas no TIA 2002, e a distribuição por tipo de exploração.

Quadro 1. Estimativa do Número Total de Explorações e População por Tipo, Basado nos

Novos Factores de Expansão para as Pequenas Explorações, TIA 2002

Tipo de Exploração/Base	Número Ponderado de Explorações	População Ponderada
Pequenas	3.092.669	15.232.042
Médias	34.982	315.538
Grandes nas UPAs	2.534	22.853
Grandes na Lista do Distrito	520	4.690
Total TIA 2002	3.130.705	15.575.124

Os dados do TIA 2002 não incluem informação sobre o número de pessoas nas grandes explorações; a população ponderada no Quadro 1 para estas explorações são basadas na média de pessoas por agregado familiar para as médias explorações, que é aproximadamente 9. Esta população foi estimada para propósitos de referencia e para estimar a população ponderada total. A estimativa da média de pessoas por agregado familiar para as pequenas explorações é 4,9. A projecção demográfica do INE para a população total nos estratos incluídos no TIA 2002 é 15.552.886, que coincide aproximadamente com a população ponderada do TIA 2002, devido à metodologia de estimação. Pode-se ver também que o número total ponderado de explorações do CAP é mais consistente com estes resultados do TIA 2002.