

Análise de Erros Padrão Simulados para os Resultados do Trabalho de Inquérito Agrícola 2002 Baseados numa Amostra de 80 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). Depois uma análise adicional foi feita para estimar a precisão simulada para outra amostra alternativa para o TIA, baseada numa amostra de 94 distritos. A metodologia e resultados deste estudo estão documentados no relatório “Análise de Erros Padrão Simulados para os Resultados do Trabalho de Inquérito Agrícola 2002 Baseados numa Amostra de 94 Distritos para Moçambique” (Megill, Abril de 2002). Os três relatórios devem ser usados juntos como referência para completar a documentação deste estudo.

A conclusão das análises prévias 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; a amostra de 94 distritos resultou numa melhora similar da precisão. Ao mesmo tempo, uma amostra bi-etápica (dispersado em 132 distritos) ou de três etapas concentrado em 94 distritos 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 uma quarta amostra alternativa para o TIA 2002, baseada em um desenho de três etapas, mas com 80 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. Ao mesmo tempo, o desenho baseado em 80 distritos amostrais deve ter menos eficiência estatística que a amostra baseada em 94 distritos, ou o desenho bi-etápico. Mas também é muito importante tomar em conta os custos de fazer o trabalho de campo em 80 distritos, comparado com os custos correspondentes para os outros desenhos. Logo pode-se determinar qual alternativa teria a melhor combinação de eficiência e custos.

2. Desenho Alternativo para a Amostra do TIA 2002, Baseado em 80 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 de distritos em três grupos: dois grupos de 33 distritos cada 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, distribuídos em 26 estratos (província por zona agro-ecológica); este desenho está descrito no primeiro relatório. Para obter a amostra de 80 distritos, primeiro seleccionamos estes mesmos 66 distritos amostrais; logo Hill seleccionou uma sub-amostra de 14 distritos adicionais do último grupo de 28 distritos amostrais. A atribuição destes 80 distritos amostrais por estrato está apresentada no Quadro 1.

Para manter um tamanho similar na amostra total de explorações agro-pecuárias, também foi necessário ajustar o número de segmentos seleccionados por distrito; estes segmentos correspondem às unidades primárias de amostragem (UPAs) definidas para o CAP. No caso da primeira alternativa com 66 distritos amostrais, seleccionamos 8 segmentos por distrito, e para a amostra de 94 distritos seleccionamos 6 segmentos por distrito. No caso do desenho alternativo com 80 distritos amostrais, decidimos seleccionar 7 segmentos por distrito, ou seja um total de 560 segmentos, comparado com 564 para o desenho de 94 distritos amostrais. Isto também implica que a atribuição de explorações agro-pecuárias seleccionadas por estrato é um pouco diferente para cada desenho; o número de pequenas explorações seleccionadas para cada desenho também é apresentado no Quadro 1. Temos que tomar em conta essas diferenças quando comparamos os resultados simulados dos três desenhos.

No Quadro 1 se pode ver que para o desenho alternativo de 80 distritos amostrais, todos os distritos são seleccionados dentro de 8 dos 26 estratos; isto é, os distritos são auto-representados na amostra para estes 8 estratos, que contêm aproximadamente 24 por cento dos agregados familiares rurais de Moçambique. Nestes estratos o desenho da amostra seria bi-etápica. Como foi explicado no relatório para o desenho alternativo de 94 distritos amostrais, para obter uma amostra auto-ponderada para as pequenas explorações dentro de cada um destes estratos, se recomenda seleccionar a amostra de segmentos (UPAs) sistematicamente dentro de cada estrato (província por zona agro-ecológica) inteiro. Por isso o número de segmentos seleccionados dentro dos distritos nestes estratos com amostra bi-etápica vai variar de acordo ao tamanho do distrito. Para ilustrar este efeito da diferença dos tamanhos de distritos, Quadro 2 apresenta o número de segmentos seleccionados em cada distrito nos 8 estratos com uma amostra bi-etápica para esta amostra alternativa de 80 distritos para o TIA. Pode-se ver que o número de segmentos seleccionados por distrito varia entre 1 e 14 de acordo aos tamanhos dos distritos correspondentes; este procedimento resulta em uma atribuição da amostra proporcional ao tamanho do distrito dentro de cada estrato com amostra bi-etápica.

Quadro 1. Atribuição da Amostra por Estrato para o Desenho Alternativo do TIA 2002 com 80 Distritos Amostrais: Distritos, Segmentos e Pequenas Explorações

Província	Estrato	No. Distritos CAP	No. Distritos TIA	No. Segmentos TIA	No. Peq. Explor. TIA, 80 Distritos	No. Peq. Explor. TIA, 94 Distritos	No. Peq. Explor. TIA, 66 Distritos
Niassa	0101	5	2	14	112	144	128
	0106	11	3	21	168	144	128
Cabo Delgado	0205	2	2	14	112	96	128
	0206	2	2	14	112	96	128
	0210	13	5	35	280	288	256
Nampula	0306	3	2	14	112	144	128
	0310	15	7	49	392	432	384
	0311	2	2	14	112	96	128
Zambézia	0404	3	3	21	168	144	128
	0407	5	5	35	280	240	256
	0411	9	5	35	280	288	256
Tete	0502	3	3	21	168	144	128
	0508	3	2	14	112	144	128
	0512	7	3	21	168	144	128
Manica	0603	5	3	21	168	144	128
	0609	3	3	21	168	144	128
	0614	2	2	14	112	96	128
Sofala	0709	4	2	14	112	144	128
	0711	5	3	21	168	144	128
	0712	3	2	14	112	144	128
Inhambane	0813	11	5	35	280	288	256
	0814	3	2	14	112	144	128
Gaza	0913	3	2	14	112	144	128
	0914	5	3	21	168	144	128
	0915	4	2	14	112	144	128
Maputo-Provincia	1014	7	5	35	280	288	256
Moçambique		138	80	560	4,480	4,512	4,224

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 80 Distritos Amostrais

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	9	72
	12	Muidumbe	14,789	5	40
Cabo Delgado 0206	03	Balama	27,617	5	40
	13	Namuno	39,542	9	72
Nampula 0311	02	Angoche	60,673	7	56
	12	Moma	65,055	7	56
Zambézia 0404	05	Gurué	46,561	6	48
	10	Milange	90,793	12	96
	15	Namarroi	22,911	3	24
Zambézia 0407	02	Alto Molócue	45,440	7	56
	04	Gilé	32,748	6	48
	06	Ile	54,564	8	64
	08	Lugela	25,374	5	40
	13	Morrumbala	59,525	9	72
Tete 0502	02	Angónia	61,348	13	104
	07	Macanga	11,276	2	16
	12	Tsangano	25,173	6	48
Manica 0609	04	Guro	11,177	6	48
	06	Macossa	2,106	1	8
	08	Mossurize	27,138	14	112
Manica 0614	05	Machaze	14,928	9	72
	10	Tambara	7,788	5	40

Nos estratos com amostra bi-etápica, as UPAs do CAP na base de amostragem foram ordenadas por distrito para manter uma estratificação implícita por distrito. 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 para o CAP. 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 8 estratos para o desenho alternativo de 80 distritos amostrais para o TIA foram seleccionadas proporcionalmente ao factor M_h/n_h para cada distrito. O ficheiro Excel TIA4SAMP.XLS 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 a análise simulada nos 8 estratos só inclui as UPAs do CAP com dados disponíveis. Se depois deste estudo se decidem usar esta amostra de 80 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, como explicado em detalhe no relatório sobre a alternativa de 94 distritos amostrais. Usando este procedimento vai reduzir muito o trabalho para a selecção bi-etápica dentro dos 8 estratos; a maior parte da selecção já está feita.

Preparando para a selecção da amostra para o TIA 2002, Christopher Hill seleccionó 8 segmentos (UPAs do CAP) dentro de cada distrito, dividido em dois grupos: um de 6 segmentos amostrais e outro de 2. No caso dos 18 estratos em que a amostra seria seleccionada em três etapas, se usou primero o grupo de 6 segmentos por distrito seleccionados por Hill para o TIA. Para seleccionar o sétimo segmento amostral dentro de cada distrito nestes estratos, um segmento foi seleccionado aleatoriamente do segundo grupo de dois segmentos amostrais; esta selecção está documentada na fôlha "UPA7" do ficheiro Excel TIA4FRAM.XLS.

Para esta análise de erros padrão simulados para a amostra alternativa de 80 distritos, dados do CAP não estavam disponíveis para 20 dos segmentos seleccionados para o TIA, que representam 3.6 por cento da amostra de 560 segmentos. Quinze dos segmentos amostrais que faltam dados são do estrato 0411 em Zambézia; 14 destes são de dois distritos não cobertos pelo CAP por causa das inundações. O Quadro 3 apresenta uma lista dos 20 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); para a amostra alternativa de 94 distritos amostrais, estavam faltando dados do CAP para 17 segmentos. Entretanto, em geral os resultados deste estudo de simulação para o desenho alternativo de 80 distritos 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 16 por cento da amostra alternativa do TIA para esta província.

Quadro 3. Lista de Segmentos Sem Dados do CAP, Seleccionados para o TIA 2002, Amostra Alternativa de 80 Distritos

Província	Estrato	Código de Distrito	Código de UPA do CAP
Nampula	0310	09	015
Zambézia	0411	09	222
		09	227
		09	232
		09	243
		09	248
		09	253
		09	255
		11	334
		14	399
		14	401
		14	405
		14	412
		14	423
		14	427
		14	430
Tete	0508	09	151
		13	227
	0512	04	067
Manica	0603	09	144

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

Para elaborar a análise de CENVAR dos erros padrão simulados baseada no desenho alternativo de 80 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 a análise de CENVAR prévia para calcular os erros padrão simulados baseados no desenho alternativo de 94 distritos amostrais, explicado no relatório prévio.

A metodologia usada para calcular as probabilidades da primeira etapa e os factores de expansão

para as pequenas e médias explorações dentro de cada estrato foi explicada em detalhe no relatório “Análise de Erros Padrão Simuladas para os Resultados do Trabalho de Inquérito Agrícola 2002 Baseados numa Amostra de 94 Distritos para Moçambique.” Também é necessário gerar códigos apropriados para o estrato e UPA que refletem o desenho da amostra alternativa de 80 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 18 estratos com amostra de três etapas e os 8 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 80 Distritos para o TIA

Depois da selecção de 560 segmentos nos 80 distritos amostrais, os dados do CAP para 540 destes segmentos (excluindo os 20 segmentos sem dados) foram usados para a análise de erros padrão simulados baseado neste desenho alternativo do TIA. Para cada segmento amostral, foram seleccionados 8 pequenas explorações e todas as médias explorações que encontraram no segmento durante o CAP. O pacote CENVAR foi usado para calcular os erros padrão; a metodologia de estimação usada por CENVAR está descrita 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.”

Os resultados desta análise simulada das medidas de precisão e eficiência de amostragem para estimativas de área total por cultura estão apresentados no Anexo I. Podemos comparar estes resultados com as medidas de precisão simuladas para a amostra preliminar baseada em 66 distritos amostrais, e também com os resultados da análise similar para o desenho alternativo de 94 distritos amostrais, apresentados nos anexos dos relatórios prévios, para avaliar os efeitos dos desenhos alternativos na eficiência da amostra.

Como se esperava, em geral os CVs e efeitos do desenho da amostra de 80 distritos amostrais baixaram comparados com os resultados correspondentes do desenho baseado em 66 distritos amostrais. Como nas análises prévias, comparamos 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 (para a amostra de 66 distritos) a 3.6. A razão entre estes dois indicadores ($3.6/4.1$), 0.87, significa que usando esta amostra de 80 distritos se pode baixar o tamanho da amostra por 13 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. Na análise prévia para a amostra de 94 distritos, a média dos efeitos do desenho para estas mesmas 12 culturas foi 2.9, indicando um aumento de 29 por cento na eficiência de amostragem comparado com o desenho de 66 distritos. Pode-se ver que o ganho em eficiência de amostragem para a amostra de 80 distritos foi um pouco menos da metade do ganho correspondente para a alternativa de 94 distritos. Entretanto, é importante tomar em conta

também a variabilidade de amostragem nestes resultados, e a atribuição da amostra um pouco diferente para cada desenho apresentada no Quadro 1.

O Quadro 4 compara os resultados dos efeitos do desenho para as quatro amostras alternativas. Pode-se ver neste quadro que às vezes o efeito do desenho para uma cultura é mais baixo para um desenho, e às vezes para outra alternativa, dado a variabilidade de amostragem.

Quadro 4. Comparação dos Efeitos do Desenho (DEFF) para as 12 Culturas com Maior Precisão, Baseado na Análise de Simulação para as Quatro Amostras Alternativas para o TIA 2002

Cultura	Efeito do Desenho (DEFF)			
	Amostra de 66 Distritos	Amostra Bi-etápica	Amostra de 94 Distritos	Amostra de 80 Distritos
Milho	4.77	2.81	3.18	3.03
Arroz	6.21	4.63	3.74	5.93
Mandioca	6.89	3.64	3.75	5.29
Pepino	1.82	2.97	4.72	4.21
Mapira	3.24	3.12	2.47	2.73
Batata Doce	2.11	2.10	1.99	2.18
Abóbora	3.66	1.87	2.75	2.53
Feijão Nhemba	3.06	2.43	2.17	1.98
Mexoeira	3.20	2.60	2.22	3.63
Feijão Boer	5.39	3.24	3.06	3.70
Feijão Jugo	3.05	2.41	1.94	2.79
Amendoim	5.88	3.48	3.01	4.76
Média	4.11	2.94	2.92	3.56

Também se pode notar no Anexo I que 12 culturas têm um CV menor que 15 por cento a nível nacional (incluindo quiabo em vez de mexoeira, que tem um CV de 15.7 por cento), comparado com 13 culturas para a amostra de 94 distritos, e 17 para a amostra bi-etápica. Isto também ilustra que a eficiência da amostra de 80 distritos seria entre aquelas das outras duas alternativas de três etapas (de 66 e 94 distritos, respectivamente), tomando em conta a variabilidade de amostragem. A nível de cultura por província, encontramos 12 casos em que o CV baixa a menos de 15 por cento quando se vai de 66 a 80 distritos amostrais, comparado com 20 casos quando se vai de 66 a 94 distritos amostrais.

A conclusão principal deste estudo comparativo entre a precisão simulada para as amostras alternativas de 66, 80 e 94 distritos amostrais para o TIA 2002 é que os CVs e efeitos do desenho

vão baixar quando se aumenta o número de distritos na amostra. Este é o resultado esperado, dado que quando aumentamos o número de distritos na amostra estamos dispersando mais a amostra de explorações agro-pecuárias. Em realidade, cada desenho alternativo usa uma combinação de três etapas de amostragem para os estratos maiores e duas etapas para os estratos menores. Usamos amostra bi-etápica em 4 estratos no caso da amostra de 66 distritos; em 8 estratos para o desenho de 80 distritos amostrais; e em 13 estratos para a alternativa de 94 distritos amostrais. A precisão seria maximizada (dado um tamanho similar da amostra) usando uma amostra bi-etápica para todos os estratos. Como foi explicado no primeiro relatório, a alternativa bi-etápica resultaria na selecção de UPAs dentro de 132 distritos diferentes. Ao mesmo tempo, os custos do trabalho de campo também aumentam com o número de distritos na amostra; a amostra bi-etápica teria os custos mais altos.

Para decidir qual desenho de amostragem alternativo se deve usar para o TIA 2002, primeiro vai ser necessário determinar os custos aproximados para cada um. Logo se pode comparar a precisão dos resultados e eficiência da amostra usando os diferentes desenhos, apresentados nos anexos dos relatórios. Se a decisão final é de usar a amostra de 80 distritos, é importante examinar os CVs simulados das estimativas de área total para as diferentes culturas a nível de província, apresentados no Anexo I, para determinar se esta precisão seria suficiente. A única maneira de melhorar a precisão baixo esta alternativa seria aumentando o tamanho da amostra (por exemplo, aumentando o número de segmentos por distrito de 7 a 8. Alternativamente, se pode seleccionar mais distritos.