

Plano de Amostragem e Metodologia de Estimação para o Inquérito Especial de Caju 2017

David J. Megill
Consultor de Amostragem
Instituto de Fomento de Caju (INCAJU)

Setembro de 2018

1. Antecedentes

O Instituto de Fomento de Caju (INCAJU) e o Ministério de Agricultura e Segurança Alimentar (MASA) conduziram o Inquérito Especial de Caju 2017 (IECAJU) para obter dados confiáveis sobre a produção de caju, número de cajueiros e características das explorações com cajueiros nas províncias de Moçambique mais importantes para a produção de caju. O IECAJU cobre as seguintes províncias: Cabo Delgado, Nampula, Zambézia, Manica, Sofala, Inhambane, Gaza, Maputo Província e Maputo Cidade.

Uma amostra bietápica estratificada foi usada para o IECAJU. A base de amostragem para este inquérito foi elaborado da base de dados do Censo Agro-Pecuário (CAP) II de 2009. Identificamos todas as unidades primárias de amostragem (UPAs) do CAP II com pelo menos 3 explorações com cajueiros, para elaborar a base de amostragem do IECAJU, e seleccionamos uma amostra de 300 UPAs para o inquérito. Em cada distrito também foi elaborada uma lista especial de grandes explorações com mais de 500 cajueiros, para incluir no inquérito com uma probabilidade de 1.

O propósito deste relatório é de descrever o plano de amostragem do IECAJU e a metodologia para calcular os ponderadores para os dados do inquérito. Este plano de amostragem foi elaborado com colaboração do Dr. Domingos Diogo, Consultor para o IECAJU.

2. Base de Amostragem e Estratificação

O CAP II de 2009 foi usado como base de amostragem para seleccionar as UPAs na primeira etapa de amostragem para o IECAJU. As áreas de enumeração (AEs) do Recenseamento Geral da População e Habitação (RGPH) de 2007 foram definidas como UPAs para o CAP II. Estas UPAs foram estratificadas por distrito dentro de cada província, e uma amostra de 3.501 UPAs foi seleccionada para o CAP II na primeira etapa, usando amostragem sistemática com probabilidade proporcional ao tamanho (PPT) dentro de cada distrito. Uma listagem de todos os agregados familiares foi feito dentro de cada UPA seleccionada, e na segunda etapa foram seleccionadas 10 pequenas explorações por UPA. Todas as médias explorações identificadas na listagem foram seleccionadas com uma probabilidade de 1 na segunda etapa para o CAP II. Uma base especial foi elaborada para as grandes explorações dentro de cada distrito, e todas estas explorações foram incluídas no CAP II. Depois do trabalho de campo para o CAP II

captaram dados para 34.882 pequenas explorações, 4.472 médias explorações e 884 grandes explorações.

Para estimar o número total de cajueiros e a produção total de caju, é importante incluir na amostra do IECAJU todas as explorações com cajueiros consideradas grandes. Neste caso definimos como grandes explorações aquelas com mais de 500 cajueiros. Para este propósito foi necessário elaborar uma lista de todas as grandes explorações com mais de 500 cajueiros para cada distrito. Esta base é independente da base de amostragem do CAP II. Nas UPAs seleccionadas para o IECAJU quando encontraram algumas grandes explorações de cajueiros que aparecem na base de grandes explorações, foi necessário evitar duplicação.

Dado que o âmbito do IECAJU é limitado a explorações com cajueiros, examinamos a distribuição de cajueiros na base do CAP II. Estes resultados estão apresentados no Quadro 1.

Quadro 1. Distribuição de cajueiros na base do CAP II por província: número de AEs e explorações com cajueiros, e número de árvores de caju na base do CAP II por província (não ponderado), e número ponderado de explorações com cajueiros.

Província	No. UPAS com cajueiros	No. explorações pequenas e médias com cajueiros	Explorações médias com cajueiros	No. de árvores de caju	No. ponderado de explorações com cajueiros
Niassa	89	207	-	532	13.126
Cabo Delgado	321	1.278	141	85.155	108.796
Nampula	576	3.150	163	130.469	412.202
Zambézia	442	2.096	15	23.075	305.057
Tete	16	23	-	87	1.689
Manica	81	352	22	18.466	29.862
Sofala	215	992	14	17.836	86.709
Inhambane	306	4.055	205	180.205	221.105
Gaza	209	1.534	31	32.803	133.546
Maputo Província	139	757	5	6.887	56.849
Maputo Cidade	90	283	4	3.235	11.475
Total	2.484	14.727	600	498.750	1.380.417

Também usamos os dados do CAP II para tabular estimativas ponderadas do número total de cajueiros por província, com os erros amostrais correspondentes; estes resultados estão apresentados no Quadro 2.

Quadro 2. Estimativas ponderadas do número total de árvores de caju dos dados do CAP II por província, erros padrão e intervalos de confiança de 95%

Domínio	Estimativa	Erro padrão	Intervalo de confiança de 95%		CV	Efeito do desenho	No. explorações com cajueiros na amostra do CAP II
			Limite inferior	Limite superior			
Moçambique	38.052.902	1.375.749,8	35.355.092	40.750.712	0,036	3,29	14.726
Província							
Niassa	30.205	4.189,6	21.989	38.420	0,139	0,97	207
Cabo Delgado	4.785.248	748.256,7	3.317.936	6.252.560	0,156	4,67	1.278
Nampula	15.471.211	892.948,6	13.720.161	17.222.260	0,058	2,69	3.150
Zambézia	3.233.676	311.088,5	2.623.639	3.843.712	0,096	3,01	2.096
Tete	4.139	1.302,3	1.585	6.692	0,315	0,64	23
Manica	1.248.553	215.939,1	825.102	1.672.004	0,173	1,00	352
Sofala	1.302.350	164.225,3	980.308	1.624.392	0,126	2,21	992
Inhambane	8.566.962	444.421,5	7.695.463	9.438.461	0,052	2,06	4.054
Gaza	2.912.165	404.767,0	2.118.428	3.705.903	0,139	3,91	1.534
Maputo provincia	405.128	58.795,8	289.831	520.425	0,145	1,12	757
Maputo Cidade	93.266	15.607,1	62.661	123.871	0,167	0,55	283

Pode-se ver no Quadro 2 que o número de cajueiros varia muito por província. As províncias com maior concentração de cajueiros são Nampula, Inhambane, Cabo Delgado e Zambézia. Dado que o número de cajueiros para as províncias de Niassa e Tete é relativamente baixo, decidiu-se excluir estas duas províncias do âmbito do IECAJU. No Quadro 2 também pode-se notar que para a maioria das províncias os coeficientes de variação (CVs) são entre 12% e 17%, indicando uma variabilidade de amostragem relativamente elevada.

Para o desenho da amostra também foi necessário definir o número mínimo de cajueiros por exploração para o IECAJU. Usamos os dados do CAP II para examinar a distribuição das pequenas e médias explorações por número de cajueiros. Esta distribuição está apresentada no Quadro 3. Pode-se ver neste quadro que 19% das explorações com caju na amostra do CAP II só têm um ou dois cajueiros. Dado a produção mínima de caju nestas explorações, foi decidido de usar o critério mínimo de 3 cajueiros para considerar uma exploração eligível para o IECAJU.

Também é importante examinar o número de cajueiros na amostra do CAP II por unidade primária de amostragem (UPA), para determinar o critério mínimo para as UPAs que vamos incluir na base de amostragem. O Quadro 4 apresenta a distribuição de UPAs por número de cajueiros na amostra do CAP II. Pode-se ver neste quadro que se excluímos as UPAs do CAP com 1 ou 2 cajueiros na amostra, isto representa 9.5% do total. Por isso decidimos que o critério mínimo para incluir as UPAs do CAP II na base de amostragem para o IECAJU é 3 cajueiros na amostra.

Quadro 3. Distribuição das pequenas e médias explorações na amostra do CAP II por número de cajueiros

No. cajueiros	No. explorações pequenas/ médias	Percentagem	Percentagem acumulado
1	1.615	11,0	11,0
2	1.186	8,1	19,0
3	1.064	7,2	26,3
4	773	5,2	31,5
5	739	5,0	36,5
6	523	3,6	40,1
7	439	3,0	43,1
8	347	2,4	45,4
9	182	1,2	46,7
10	903	6,1	52,8
11-15	1.145	7,8	60,6
16-20	999	6,8	67,4
21-25	491	3,3	70,7
26-30	764	5,2	75,9
31-35	234	1,6	77,5
36-40	381	2,6	80,1
41-45	121	0,8	80,9
46-50	610	4,1	85,0
51-60	349	2,4	87,4
61-70	211	1,4	88,8
71-80	199	1,4	90,2
81-90	105	0,7	90,9
91-100	347	2,4	93,3
101-150	390	2,6	95,9
151-200	242	1,6	97,6
201-300	190	1,3	98,9
301-400	66	0,4	99,3
401-500	39	0,3	99,6
501-1000	55	0,4	99,9
1001-1500	4	0,0	100,0
1501-2000	4	0,0	100,0
Moçambique	14.717	100,0	100,0

Quadro 4. Distribuição de UPAs do CAP II por número de cajueiros na amostra

No. de cajueiros na amostra	No. UPAs do CAP II	% UPAs	% cumulado de UPAs
1	135	5,4	5,4
2	100	4,0	9,5
3	87	3,5	13,0
4	56	2,3	15,3
5	61	2,5	17,7
6	54	2,2	19,9
7	49	2,0	21,9
8	39	1,6	23,4
9	34	1,4	24,8
10	49	2,0	26,8
11	36	1,4	28,2
12	34	1,4	29,6
13	23	0,9	30,5
14	26	1,0	31,6
15	31	1,2	32,8
16	25	1,0	33,8
17	21	0,8	34,7
18	18	0,7	35,4
19	18	0,7	36,1
20	38	1,5	37,6

Para aumentar a eficiência da amostra do IECAJU, também é importante estratificar as UPAs do CAP II por nível de concentração de cajueiros em cada província. Primeiro dividimos as UPAs do CAP II com cajueiros para cada província em 3 estratos: alto (25% de UPAs com mais cajueiros), médio (de 40 a 75%) e baixo (40% com menos cajueiros). Logo distribuimos as UPAs seleccionadas por estrato pelos factores 3-2-1. Neste caso 50% das UPAs seriam seleccionadas no estrato alto, um terço no estrato médio, e um sexto no estrato baixo. Exactamente metade da amostra vai ser do estrato alto, mas as proporções para os outros estratos podem variar um pouco destas frações. A base de amostragem final para o IECAJU inclui 2,183 UPAs do CAP II. A distribuição destas UPAs por província e estrato está apresentada mais adiante no Quadro 6, que também inclui a distribuição da amostra final.

Para propósito de estatificação das explorações por tamanho na listagem, definimos como pequenas explorações aquelas que tem 100 ou menos cajueiros. Classificamos como médias explorações aquelas com 101 a 500 cajueiros. Como indicamos antes, as grandes explorações seriam aquelas com mais de 500 cajueiros. Todas as grandes explorações na base especial foram incluídas na amostra. Quando encontraram outras grandes explorações dentro das UPAs seleccionadas que não aparecem na base especial, foram incluídas com probabilidade de 1 na segunda etapa, juntamente com as médias explorações.

3. Tamanho e Distribuição da Amostra para o IECAJU

O tamanho da amostra para um inquérito particular é função da precisão requerida para as estimativas principais para cada domínio, bem como pelas limitações operacionais e de recursos. A precisão dos resultados do inquérito depende tanto do erro de amostragem, que pode ser medido por meio de estimativas de variâncias, e o erro não amostral de outras fontes, como erros de resposta, medida, codificação e processamento. O erro de amostragem é inversamente proporcional à raiz quadrada do tamanho da amostra. Por outro lado, o erro não amostral pode aumentar com o tamanho da amostra, uma vez que é mais difícil de controlar a qualidade de uma operação grande. Por isso, é importante que o tamanho da amostra seja manejável para fins de controle de qualidade e operacional.

Dado os objetivos do IECAJU, os recursos disponíveis e considerações de qualidade dos dados, foi decidido de seleccionar uma amostra de 300 UPAs na primeira etapa de amostragem. Uma listagem de todos os agregados familiares dentro de cada UPA foi feita para identificar todas as explorações que têm 3 ou mais cajueiros, classificadas como pequenas (com 100 ou menos cajueiros) ou médias (com entre 101 e 500 cajueiros). Na segunda etapa 10 pequenas explorações foram seleccionadas dentro de cada UPA, para uma amostra total de 3.000 pequenas explorações. O número total de entrevistas depende do número de médias encontradas nas UPAs seleccionadas e o número de grandes explorações na base separada.

Para determinar a distribuição das UPAs na amostra por província, primeiro examinamos uma distribuição proporcional ao raiz quadrado do número de explorações com cajueiros em cada província, estimado dos dados do CAP II. Esta distribuição aumenta o número de UPAs na amostra para as províncias menores e diminui um pouco a amostra para as províncias maiores, comparado com uma distribuição proporcional correspondente. Logo ajustamos o número de UPAs amostrais por província para ter uma amostra mínima de 24 UPAs para o estrato combinado de Maputo Província e Maputo Cidade, e uma amostra máxima de 56 UPAs para Nampula. A distribuição da base de amostragem e a amostra de UPAs por província está apresentada no Quadro 5.

Quadro 5. Distribuição da base de amostragem e a amostra de UPAS e pequenas explorações de caju por província para o IECAJU

Província	No. UPAS com cajueiros no CAP II	No. ponderado de explorações com caju (CAP II)	Distribuição de UPAs ajustado na amostra do IECAJU	Amostra de pequenas explorações de caju para o IECAJU
Cabo Delgado	321	108.796	32	320
Nampula	576	412.202	56	560
Zambézia	442	305.057	48	480
Manica	81	29.862	28	280
Sofala	215	86.709	28	280
Inhambane	306	221.105	48	480
Gaza	209	133.546	36	360
Maputo Cidade e Província	229	68.325	24	240
Moçambique	2.484	1.365.601	300	3.000

Como indicado acima, as UPAs na base de amostragem do CAP II para cada província foram estratificadas por nível de concentração de cajueiros. Definimos os seguintes estratos: alto (25% de UPAs com mais cajueiros na província), médio (de 40 a 75%) e baixo (40% com menos cajueiros). Para cada província 50% das UPAs na amostra foram atribuídas ao estrato alto, um terço das UPAs para o estrato médio e um sexto das UPAs para o estrato baixo. Assim estamos concentrando a amostra nos estratos com mais cajueiros para melhorar a eficiência da amostragem. O Quadro 6 apresenta a distribuição final das UPAs seleccionadas por estrato dentro de cada província.

Quadro 6. Distribuição da base de amostragem do CAP II de UPAs com cajueiros e amostra de UPAs para o IECAJU, por província e estrato de concentração de cajueiros

Província	Estrato	Nível de concentração de cajueiros	UPAs na base do CAP II	UPAs na amostra do IECAJU
Cabo Delgado	21	Baixo	119	5
	22	Médio	103	11
	23	Alto	74	16
Nampula	31	Baixo	225	9
	32	Médio	198	19
	33	Alto	139	28
Zambézia	41	Baixo	157	8
	42	Médio	133	16
	43	Alto	96	24
Manica	61	Baixo	25	5
	62	Médio	22	9
	63	Alto	15	14
Sofala	71	Baixo	70	5
	72	Médio	63	9
	73	Alto	43	14
Inhambane	81	Baixo	122	8
	82	Médio	107	16
	83	Alto	76	24
Gaza	91	Baixo	78	6
	92	Médio	69	12
	93	Alto	48	18
Maputo Província/Cidade	101	Baixo	83	4
	102	Médio	68	8
	103	Alto	50	12
Total			2.183	300

4. Selecção da Amostra de UPAs para o IECAJU

Um código de estrato baixo, médio ou alto por província foi gerado para as 2.183 UPAs com cajueiros na base do CAP II. A distribuição das UPAs por estrato está apresentada no Quadro 6. A base de amostragem foi ordenada por província, estrato, distrito, posto administrativo, localidade, bairro e código de AE. Logo seleccionamos a amostra de UPAs dentro de cada estrato usando amostragem aleatória sistemática com probabilidades iguais. O número de UPAs seleccionadas em cada estrato está especificado no Quadro 6.

Uma listagem de todos os agregados familiares foi feito dentro de cada UPA seleccionada para identificar as explorações pequenas e médias com cajueiros. Uma amostra de 10 pequenas explorações foi seleccionada da listagem usando amostragem aleatória sistemática. Todas as médias explorações identificadas na listagem de cada UPA foram seleccionadas. As grandes explorações encontradas na listagem também foram seleccionadas, depois de verificar se estão incluídas na lista de grandes explorações da província para evitar duplicação. A base inteira de grandes explorações para cada província foi incluída na amostra do IECAJU com probabilidade de 1.

5. Cálculo de Ponderadores para o IECAJU

Para que as estimativas da amostra de um inquérito sejam representativas da população, é necessário multiplicar os dados por um ponderador. O ponderador básico para cada exploração com cajueiros na amostra seria igual ao inverso da sua probabilidade de selecção (calculado multiplicando as probabilidades em cada etapa de amostragem). Dado que os ponderadores são diferentes para cada tipo de exploração (pequeno, médio e grande), o cálculo dos ponderadores é especificado individualmente para cada tipo de exploração com cajueiros.

5.1. Ponderadores básicos para as Pequenas Explorações

Dado que as UPAs para o IECAJU foram seleccionadas das UPAs na amostra do CAP II, o primeiro componente da probabilidade é a probabilidade de selecção da UPA no CAP II; as UPAs foram seleccionadas com PPT dentro de cada distrito. O segundo componente é a probabilidade de selecção da subamostra de UPAs para o IECAJU dentro do estrato de concentração de cajueiros. O terceiro componente é a probabilidade de selecção das 10 pequenas explorações com cajueiros dentro de cada UPA. Então a probabilidade total pode ser expressada da seguinte maneira:

$$p_{hi(p)} = p_{CAPhi} \times \frac{n_{lh}}{n_{CAPh}} \times \frac{m_{hi(p)}}{M_{hi(p)}},$$

Onde:

$p_{hi(p)}$ = probabilidade de selecção das pequenas explorações com cajueiros na i-éssima UPA seleccionada no estrato h (província por estrato de concentração de cajueiros) para o IECAJU

p_{CAPhi} = probabilidade de selecção no CAP II da i-éssima UPA seleccionada no estrato h

n_{lh} = número de UPAs seleccionadas no estrato h para o IECAJU

n_{CAPh} = número de UPAs na base do CAP II para o estrato h

$m_{hi(p)}$ = número de pequenas explorações com cajueiros seleccionadas na i-éssima UPA na amostra do estrato h (geralmente igual a 10)

$M_{hi(p)}$ = número de pequenas explorações com cajueiros listadas na i-éssima UPA na amostra do estrato h

O ponderador básico de amostragem é calculado como o inverso da probabilidade de selecção. Baseando-se na fórmula acima para a probabilidade, o ponderador básico para as pequenas explorações com cajueiros pode ser expressada da maneira seguinte:

$$W_{hi(p)} = \frac{n_{CAPh} \times M_{hi(p)}}{p_{CAPhi} \times n_{lh} \times m_{hi(p)}},$$

onde:

$W_{hi(p)}$ = ponderador básico para as pequenas explorações com cajueiros seleccionadas na i-éssima UPA na amostra do estrato h

Também é necessário ajustar o ponderador básico para tomar em conta a taxa de resposta das pequenas explorações dentro de cada UPA. Neste caso o ponderador ajustado pode ser calculada da seguinte maneira:

$$W'_{hi(p)} = W_{hi(p)} \times \frac{m_{hi(p)}}{m'_{hi(p)}}$$

onde:

$W'_{hi(p)}$ = ponderador ajustado para as pequenas explorações com cajueiros seleccionadas na i-éssima UPA na amostra do estrato h

$m'_{hi(p)}$ = número de pequenas explorações entrevistadas para o IECAJU na i-éssima UPA na amostra do estrato h

5.2. Ponderadores para as Médias Explorações

Dado que todas as médias explorações listadas na UPA foram seleccionadas para o IECAJU, estas explorações têm a mesma probabilidade de selecção da UPA, calculada usando a seguinte fórmula:

$$p_{hi(m)} = p_{CAPhi} \times \frac{n_{lh}}{n_{CAPh}},$$

Onde:

$p_{hi(m)}$ = probabilidade de selecção de médias explorações com cajueiros na i-éssima UPA seleccionada no estrato h

O ponderador básico para as médias explorações é o inverso desta probabilidade de selecção:

$$W_{hi(m)} = \frac{n_{CAPh}}{p_{CAPhi} \times n_{Ih}},$$

onde:

$W_{hi(m)}$ = ponderador básico para as médias explorações com cajueiros seleccionadas na i-éssima UPA na amostra do estrato h

Também é necessário ajustar o ponderador básico para tomar em conta a taxa de resposta das médias explorações dentro de cada UPA. Neste caso o ponderador ajustado pode ser calculada da seguinte maneira:

$$W'_{hi(m)} = W_{hi(m)} \times \frac{M_{hi(m)}}{M'_{hi(m)}},$$

onde:

$W'_{hi(m)}$ = ponderador ajustado para as médias explorações com cajueiros seleccionadas na i-éssima UPA na amostra do estrato h

$M_{hi(m)}$ = número total de médias explorações com cajueiros listadas para o IECAJU na i-éssima UPA na amostra do estrato h

$M'_{hi(m)}$ = número de médias explorações com cajueiros entrevistadas para o IECAJU na i-éssima UPA na amostra do estrato h

5.3. Ponderadores para as Grandes Explorações

Uma base de amostragem de todas as grandes explorações de cajueiros foi elaborada em cada distrito para o IECAJU. Todas estas grandes explorações com cajueiros foram incluídas na amostra do IECAJU com uma probabilidade igual a 1, isto é, todas as grandes explorações com cajueiros são auto-representadas. Portanto, o ponderador básico para estas grandes explorações com cajueiros é igual a 1.

É importante ajustar o ponderador básico para tomar em conta as grandes explorações agrícolas não inquiridas ou não contactadas dentro do distrito. O ponderador final $W_{d(g)}$ para uma grande exploração agrícola em cada distrito foi calculado usando a seguinte formula:

$$W'_{d(g)} = \frac{M_{d(g)}}{M'_{d(g)}},$$

onde:

$W'_{d(g)}$ = Ponderador ajustado para as grandes explorações com cajueiros no distrito d

$M_{d(g)}$ = Número total de grandes explorações com cajueiros na base de amostragem para o distrito d

$M'_{d(g)}$ = Número total de grandes explorações com cajueiros com entrevista completa no distrito d

Na listagem dentro da amostra de AEs para o IECAJU, encontraram algumas explorações com cajueiros que atendam aos critérios de grandes explorações, mas que não aparecem na lista de grandes explorações do distrito. É importante distinguir estas grandes explorações com cajueiros daquelas da lista de grandes explorações previamente identificadas no distrito, dado que as probabilidades de selecção são diferentes. No caso de grandes explorações com cajueiros encontradas na listagem dentro de cada AE (e que não aparecem na lista das grandes explorações do distrito), o ponderador será o mesmo que aquele especificado acima para as médias explorações.

5.4. Ajuste final dos ponderadores para cada província

Depois de calcular os ponderadores para todas as explorações pequenas, médias e grandes de acordo às fórmulas especificadas acima, estes ponderadores foram usados para estimar o número total de explorações com cajueiros, o número total de cajueiros e outros indicadores. Deve-se notar que sómente as explorações com 3 ou mais cajueiros foram elegíveis para o IECAJU. Quando comparamos o número total ponderado de explorações do IECAJU por província com o número total ponderado de explorações com 3 ou mais cajueiros dos dados do CAP II, notamos uma sub-estimação nos resultados do IECAJU para todas as províncias com a excepção de Cabo Delgado. Dado que a amostra do CAP II foi muito maior do que a amostra correspondente do IECAJU e as estimativas do número total de explorações com 3 ou mais cajueiros do CAP II eram mais confiáveis, usamos os dados do CAP II de 2009 para fazer uma projecção do número total de explorações com cajueiros em 2017 para ajustar os ponderadores preliminares do IECAJU. Neste caso incluímos todas as explorações com cajueiros do CAP II, incluindo explorações com 1 ou 2 cajueiros. O factor de ajuste para todos os ponderadores de cada província foi calculado usando a seguinte fórmula:

$$A_p = \frac{\hat{E}_{p(CAP)}}{\hat{E}_{p(I)}}$$

onde:

A_p = factor de ajuste para todos os ponderadores preliminares do IECAJU para a província p

$\hat{E}_{p(CAP)}$ = projecção do número total de explorações com cajueiros em 2017 para a província p , usando como base os dados do CAP II

$\hat{E}_{p(I)}$ = estimativa do número total de explorações com cajueiros para a província p dos dados do IECAJU, usando os ponderadores preliminares descritos acima

Para fazer a projecção do número total de explorações com cajueiros em 2017 para cada província, usamos a estimativa do CAP II (2009) do número total ponderado de explorações com cajueiros como base, e assumimos uma taxa de crescimento de 5% para todas as províncias com a excepção de Cabo Delgado, onde usamos um factor de ajuste de 1 para manter os ponderadores preliminares do IECAJU. O Quadro 7 apresenta as estimativas do número total de explorações do IECAJU usando os ponderadores preliminares para cada província, a projecção do número total de explorações em 2017, e o factor de ajuste para os ponderadores de cada província.

Quadro 7. Factores de ajuste para ponderadores do IECAJU por província

Província	No. total de explorações, IECAJU, ponderadores preliminares	Projecção de explorações com cajueiros 2017	Factor de ajuste para ponderadores do IECAJU
Cabo Delgado	78.169	78.169	1,080462
Nampula	364.199	370.616	1,078505
Zambézia	151.431	183.913	1,343562
Manica	18.355	21.930	1,240507
Sofala	37.489	46.267	1,308836
Inhambane	104.681	213.710	2,049038
Gaza	67.893	100.338	1,510388
Maputo total	23.307	32.168	1,710758
Total	845.524	1.047.111	

6. Tipos de Estimativas dos Dados do IECAJU

A maioria das estimativas a serem derivadas dos dados do IECAJU serão tipos de totais ou razões. A estimativa de um total pode ser expressada da seguinte maneira:

$$\hat{Y} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_j W_{hij} \times y_{hij}$$

onde:

L = número de estratos

y_{hij} = valor da variável y para a j-ésima exploração dentro da i-ésima UPA amostral no estrato h

Este total é simplesmente a soma dos dados ponderados para todas as explorações na amostra para os estratos correspondentes.

A estimativa de um rácio seria calculada da seguinte maneira:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}},$$

onde \hat{Y} e \hat{X} são estimativas de totais para as variáveis y e x, respectivamente, calculados como especificado antes.

No caso de uma amostra estratificada de várias etapas, como o IECAJU, as médias e proporções são tipos de rácios. No caso de uma média, a variável x no denominador da razão seria igual a 1 para cada unidade de análise (por exemplo, exploração), então o denominador seria simplesmente a soma dos ponderadores. No caso de uma proporção, a variável x também seria igual a 1 para todas as unidades, e a variável y seria igual a 1 ou 0, dependendo se a unidade tem ou não a característica de interesse.

7. Cálculo de Erros Amostrais nos resultados do IECAJU

Na publicação dos resultados do IECAJU, é importante incluir um anexo sobre a exatidão dos dados do inquérito. Este anexo deve incluir estimativas dos erros amostrais para os indicadores mais importantes do inquérito e também uma descrição dos erros não amostrais.

O erro padrão, ou raiz quadrado da variância, é usado para medir o erro de amostragem, mas também pode incluir a parte variável dos erros não amostrais. O estimador da variância deve tomar em conta os diferentes aspectos do desenho de amostragem, como a estratificação e a conglomeração. Os programas de SPSS (Complex Samples) e Stata usam um estimador de variância que toma em conta o plano de amostragem. Estes programas usam um estimador de variância linearizado de série Taylor. Para a estimativa de um total, a variância é calculada por SPSS e Stata usando a seguinte fórmula:

Estimador de variância para um total

$$V(\hat{Y}) = \sum_{h=1}^L \left[(1-f_h) \times \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right)^2 \right],$$

onde:

f_h = probabilidade media de selecção das UPAs dentro do estrato h

n_h = número de UPAs seleccionadas para o estrato h

\hat{Y}_{hi} = total ponderado da variável y para a i-ésima UPA do estrato h

$\hat{Y}_h = \sum_{i=1}^{n_h} \hat{Y}_{hi}$ = total ponderado da variável y para o estrato y

O factor $(1-f_h)$ na fórmula da variância é importante no caso do IECAJU, porque representa a correção para uma população finita. Isto é especialmente crítico no caso do estrato das grandes explorações auto-ponderadas, onde a probabilidade de selecção é 1. Neste caso esta correção elimina o componente da variância deste estrato.

Para a estimativa de um rácio, a variância é calculada por SPSS e Stata usando a seguinte fórmula:

Estimador de variância para um rácio

$$V(\hat{R}) = \frac{1}{\hat{X}^2} \times \left[V(\hat{Y}) + \hat{R}^2 \times V(\hat{X}) - 2 \times \hat{R} \times COV(\hat{X}, \hat{Y}) \right],$$

onde:

$$COV(\hat{X}, \hat{Y}) = \sum_{h=1}^L \left[(1-f_h) \times \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} \left(\hat{X}_{hi} - \frac{\hat{X}_h}{n_h} \right) \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right) \right]$$

$V(\hat{Y})$ e $V(\hat{X})$ são calculados usando a fórmula para a variância da estimativa de um total, especificada anteriormente.

O erro padrão, que é uma medida do erro de amostragem, é calculado como a raiz quadrada da variância.

Para calcular os erros de amostragem usando o módulo de Complex Samples de SPSS, foi necessário gerar algumas variáveis adicionais nos ficheiros de dados do IECAJU para refletir bem a natureza do desenho da amostra no cálculo dos erros de amostragem. No caso da amostra de UPAs para a selecção das pequenas e médias explorações, um código de estrato de 3 dígitos foi criado. Os primeiros 2 dígitos correspondem ao código da província; para a combinação de Maputo Província e Maputo Cidade, usamos o código de província 10. O último dígito do código de estrato corresponde ao nível de concentração de cajueiros (1, 2 ou 3) descrito acima na secção sobre a base de amostragem. Estes estratos estão especificados no Quadro 6. Para as grandes explorações na base especial, usamos o código de estrato 999 para classificar o estrato auto-representado. Para especificar a UPA (conglomerado) para os estratos bi-etápicos, usamos o código de UPA que aparece na base de dados. Estes códigos de UPA foram copiados numa nova variável UPA2. No caso do estrato auto-representado de grandes explorações, usamos o valor de 999 para a UPA. Para a correção pela população finita, criamos uma nova variável fpc, que corresponde à probabilidade de selecção média das explorações dentro de cada estrato.

Para preparar a aplicativa de Complex Samples de SPSS para os dados do IECAJU, foi necessário gerar um ficheiro de especificação do plano de amostragem (csplan), onde indicamos as variáveis para o estrato, UPA, fpc e ponderador. O software Stata usa o mesmo estimador da variância que Complex Samples de SPSS. Se usam Stata para calcular os erros de amostragem, podem usar estas mesmas variáveis de amostragem para especificar o plano de amostragem para a análise usando o comando SVY.

Usando Complex Samples de SPSS, medidas de precisão foram calculadas para os seguintes indicadores por província, e a nível nacional:

1. Número total de explorações com cajueiros
2. Quantidade produzida de castanha bruta em toneladas
3. Quantidade vendida de castanha bruta em toneladas
4. Quantidade oferecida de castanha bruta em toneladas
5. Quantidade consumida de castanha bruta em toneladas
6. Quantidade processada de castanha bruta em toneladas
7. Número total de cajueiros
8. Número total de cajueiros em crescimento
9. Número total de cajueiros em produção
10. Número total de cajueiros velhos (não produtivos)

Estes resultados estão apresentados no Anexo A. Para cada indicador, os quadros no Anexo A incluem o valor da estimativa, erro padrão (medida do erro de amostragem), coeficiente de variação, intervalo de confiança de 95%, efeito do desenho e número de explorações na amostra para cada categoria.

Para estudar o nível de precisão das diferentes estimativas do IECAJU, podemos examinar os coeficientes de variação (CVs) nos quadros do Anexo A, que representam os erros de amostragem relativos. No caso das estimativas à nível nacional, pode-se ver que os CVs são

menores de 10% para a maioria das estimativas, que indica uma boa precisão. No caso das estimativas à nível de província, os CVs são mais elevados, especialmente para províncias menores. As estimativas com CVs acima de 20% devem ser usados com cautela.

Uma razão de encontrar erros de amostragem relativamente elevadas para algumas províncias é devido à variabilidade nos dados de CAP II, que foi a única base de amostragem disponível para o IECAJU. Esta base consiste de uma amostra aleatória de UPAs com explorações agropecuárias dentro de cada distrito, mas não foi desenhada para culturas específicas. Para o IECAJU foi possível concentrar a amostra nas UPAs do CAP II onde encontraram mais cajueiros em 2010, mas dado a natureza dinâmica da agricultura, encontraram diferenças depois de 7 anos. Mesmo assim, os resultados do IECAJU são bem representativos, dado a listagem actualizada de explorações com cajueiros em cada UPA seleccionada, como também a cobertura compreensiva das grandes explorações com cajueiros.

ANEXO A. Estimativas de Indicadores do IECAJU por Província e Nível Nacional, com Erros Padrão, Coeficientes de Variação, Intervalos de Confiança de 95%, Efeitos de Amostragem (DEFF), e Número de Observações

Quadro A1. Número total de explorações com cajueiros

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	1,047,111	49,549	0.047	949,562	1,144,659	-	4,536
Província							
Cabo Delgado	78,169	9,405	0.120	59,652	96,685	5.32	957
Nampula	370,616	34,875	0.094	301,957	439,275	22.10	1,077
Zambézia	183,913	14,614	0.079	155,143	212,683	6.13	578
Manica	21,930	3,358	0.153	15,320	28,541	2.28	354
Sofala	46,267	5,669	0.123	35,107	57,427	3.16	321
Inhambane	213,710	26,747	0.125	161,053	266,366	18.29	635
Gaza	100,338	11,554	0.115	77,592	123,085	6.40	445
Maputo combinado	32,168	6,675	0.208	19,026	45,310	6.22	169

Quadro A2. Número total de árvores de caju

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	42,410,726	3,716,721	0.088	35,093,528	49,727,924	1.41	4,527
Província							
Cabo Delgado	4,788,607	1,257,715	0.263	2,312,513	7,264,700	4.40	956
Nampula	18,044,773	2,676,738	0.148	12,775,014	23,314,532	1.33	1,077
Zambézia	4,838,058	1,536,587	0.318	1,812,941	7,863,174	0.75	578
Manica	921,428	198,110	0.215	531,404	1,311,452	0.41	354
Sofala	1,830,971	782,373	0.427	290,695	3,371,247	2.94	319
Inhambane	8,632,726	1,305,105	0.151	6,063,335	11,202,116	3.14	635
Gaza	3,019,389	587,605	0.195	1,862,557	4,176,221	0.97	439
Maputo combinado	334,774	78,798	0.235	179,642	489,907	0.91	169

Quadro A3. Número total de árvores de caju em crescimento

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	5,626,047	546,977	0.097	4,549,201	6,702,893	1.22	4,527
Província							
Cabo Delgado	622,016	123,565	0.199	378,751	865,280	1.53	956
Nampula	2,086,015	335,703	0.161	1,425,110	2,746,921	2.29	1,077
Zambézia	1,054,927	319,504	0.303	425,912	1,683,942	0.74	578
Manica	205,443	44,897	0.219	117,052	293,833	0.41	354
Sofala	417,649	209,949	0.503	4,317	830,980	1.98	319
Inhambane	927,446	136,906	0.148	657,916	1,196,976	1.12	635
Gaza	222,482	45,349	0.204	133,203	311,761	0.32	439
Maputo combinado	90,070	47,351	0.526	0	183,291	1.13	169

Quadro A4. Número total de árvores de caju em produção

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	30,747,418	3,016,367	0.098	24,809,025	36,685,811	1.21	4,527
Província							
Cabo Delgado	3,798,249	1,059,461	0.279	1,712,462	5,884,035	4.48	956
Nampula	13,959,825	2,302,115	0.165	9,427,596	18,492,053	1.09	1,077
Zambézia	3,481,961	1,225,294	0.352	1,069,696	5,894,226	0.75	578
Manica	648,217	145,450	0.224	361,865	934,568	0.34	354
Sofala	1,131,538	401,403	0.355	341,286	1,921,791	2.64	319
Inhambane	5,724,798	927,883	0.162	3,898,052	7,551,544	2.80	635
Gaza	1,865,284	360,928	0.193	1,154,717	2,575,851	1.13	439
Maputo combinado	137,546	37,605	0.273	63,513	211,580	0.68	169

Quadro A5. Número total de árvores de caju velhos (não produtivos)

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	6,037,261	586,790	0.097	4,882,034	7,192,487	1.07	4,527
Província							
Cabo Delgado	368,342	104,737	0.284	162,143	574,541	0.57	956
Nampula	1,998,933	284,839	0.142	1,438,163	2,559,703	2.33	1,077
Zambézia	301,169	41,793	0.139	218,891	383,448	0.35	578
Manica	67,769	14,581	0.215	39,063	96,474	0.15	354
Sofala	281,784	177,987	0.632	0	632,192	0.83	319
Inhambane	1,980,482	335,717	0.170	1,319,548	2,641,416	2.97	635
Gaza	931,623	324,364	0.348	293,040	1,570,207	0.55	439
Maputo combinado	107,158	25,727	0.240	56,508	157,808	1.16	169

Quadro A6. Quantidade produzida de castanha bruta em toneladas

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	140,570.69	12,782.48	0.091	115,405.52	165,735.87	1.54	4,536
Província							
Cabo Delgado	11,402.41	2,476.45	0.217	6,526.97	16,277.85	1.77	957
Nampula	75,740.91	11,032.01	0.146	54,021.94	97,459.89	1.51	1,077
Zambézia	9,977.62	1,686.39	0.169	6,657.60	13,297.65	3.48	578
Manica	5,817.85	1,277.88	0.220	3,302.06	8,333.63	0.18	354
Sofala	7,077.18	3,048.27	0.431	1,075.99	13,078.37	5.20	321
Inhambane	15,018.68	2,755.45	0.183	9,593.95	20,443.41	4.68	635
Gaza	11,668.41	2,413.06	0.207	6,917.76	16,419.06	0.85	445
Maputo combinado	3,867.63	2,893.03	0.748	0.00	9,563.21	1.63	169

Quadro A7. Quantidade vendida de castanha bruta em toneladas

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	88,656.19	9,983.90	0.113	69,000.65	108,311.72	1.58	4,536
Província							
Cabo Delgado	7,837.01	2,281.35	0.291	3,345.65	12,328.36	1.67	957
Nampula	54,350.21	8,967.59	0.165	36,695.51	72,004.91	1.67	1,077
Zambézia	5,666.54	1,187.83	0.210	3,328.03	8,005.06	3.06	578
Manica	4,406.91	1,068.50	0.242	2,303.33	6,510.48	0.13	354
Sofala	4,389.87	2,416.12	0.550	0.00	9,146.55	4.70	321
Inhambane	6,504.22	1,936.31	0.298	2,692.16	10,316.28	2.11	635
Gaza	5,454.09	1,383.94	0.254	2,729.50	8,178.68	2.79	445
Maputo combinado	47.34	34.00	0.718	-19.59	114.28	1.46	169

Quadro A8. Quantidade oferecida de castanha bruta em toneladas

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	7,395.70	1,146.48	0.155	5,138.60	9,652.80	1.19	4,536
Província							
Cabo Delgado	301.14	79.14	0.263	145.34	456.93	1.20	957
Nampula	2,689.97	539.67	0.201	1,627.51	3,752.44	5.81	1,077
Zambézia	777.28	326.18	0.420	135.13	1,419.43	2.16	578
Manica	150.23	38.81	0.258	73.82	226.63	0.88	354
Sofala	327.27	87.68	0.268	154.65	499.89	0.97	321
Inhambane	1,271.27	274.46	0.216	730.94	1,811.61	3.59	635
Gaza	1,713.53	899.89	0.525	0.00	3,485.16	0.84	445
Maputo combinado	165.01	127.25	0.771	0.00	415.52	1.36	169

Quadro A9. Quantidade consumida de castanha bruta em toneladas

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	20,848.77	1,767.64	0.085	17,368.78	24,328.77	2.23	4,536
Província							
Cabo Delgado	882.75	177.03	0.201	534.23	1,231.27	4.41	957
Nampula	6,883.30	822.00	0.119	5,265.01	8,501.59	9.01	1,077
Zambézia	1,183.69	151.81	0.128	884.83	1,482.55	4.81	578
Manica	649.70	137.38	0.211	379.23	920.17	1.70	354
Sofala	1,505.66	567.55	0.377	388.32	2,623.00	1.58	321
Inhambane	6,699.91	1,111.17	0.166	4,512.32	8,887.50	6.07	635
Gaza	2,708.99	899.40	0.332	938.33	4,479.65	0.84	445
Maputo combinado	334.78	99.41	0.297	139.07	530.48	2.34	169

Quadro A10. Quantidade processada de castanha bruta em toneladas

Domínio	Valor	Erro padrão	CV	Intervalo de confiança de 95%		DEFF	Número de observações (explorações)
				Inferior	Superior		
Moçambique	3,158.86	2,675.26	0.847	-2,107.98	8,425.70	1.40	4,536
Província							
Cabo Delgado	10.91	4.49	0.411	2.08	19.75	0.37	957
Nampula	84.00	42.21	0.502	0.90	167.11	3.53	1,077
Zambézia	55.41	34.90	0.630	0.00	124.12	2.56	578
Manica	24.52	12.55	0.512	0.00	49.22	0.13	354
Sofala	15.27	6.29	0.412	2.88	27.65	0.69	321
Inhambane	162.85	53.86	0.331	56.80	268.89	0.07	635
Gaza	84.67	37.34	0.441	11.15	158.18	0.56	445
Maputo combinado	2,721.23	2,673.85	0.983	0.00	7,985.31	1.41	169