

**Inquérito aos Agregados Familiares sobre
Despesas e Receitas (IDR) 1999-2001**

Relatório sobre o Plano de Amostragem

Por:
Ana Paula J. Machado

Luanda, Dezembro de 2000

ÍNDICE

INTRODUÇÃO.....	3
ÂMBITO GEOGRÁFICO.....	3
BASE DE AMOSTRAGEM E UNIDADE DE AMOSTRAGEM.....	4
DESENHO DA AMOSTRA.....	5
TAMANHO DA AMOSTRA.....	5
LISTAGEM DOS AGREGADOS FAMILIARES	8
SELECÇÃO DA AMOSTRA.....	9
A) Área Urbana	9
AMOSTRA DE CONGLOMERADOS	9
AMOSTRA DE AGREGADOS FAMILIARES	11
<i>Substituição das Famílias.....</i>	<i>11</i>
AMOSTRA DE COMUNAS	12
AMOSTRA DE ALDEIAS	13
AMOSTRA DE AGREGADOS FAMILIARES	14
DISTRIBUIÇÃO TEMPORAL DA AMOSTRA DE CONGLOMERADOS.....	14
PROBABILIDADES DE SELECÇÃO	14
A) ÁREA URBANA.....	15
B) ÁREA RURAL.....	15
PESO DE AMOSTRAGEM OU FACTOR DE EXPANSÃO DA AMOSTRA.....	16
A) ÁREA URBANA.....	17
B) ÁREA RURAL.....	17
ESTIMADORES.....	18
A) ÁREA URBANA.....	18
B) ÁREA RURAL.....	19

Introdução

O Instituto Nacional de Estatística de Angola (INE) irá realizar um Inquérito aos Agregados Familiares sobre Despesas e Receitas (IDR) no período de Fevereiro de 2000 à Fevereiro de 2001, com o objectivo principal de fornecer aos utilizadores informação sobre as receitas e as despesas dos agregados familiares nas províncias de Luanda, Lunda Norte, Cabinda, Benguela, Namibe, Huíla e Cunene.

Os resultados do inquérito irão permitir aos utilizadores quantificar as receitas e conhecer a estrutura das despesas das famílias, relacionar esta informação as características demográficas, de educação e de ocupação da população. Estes resultados irão contribuir ainda para análise do comportamento socioeconómico das famílias e permitirão a elaboração de uma nova estrutura de ponderação para o Índice de Preços no Consumidor (IPC) em cada uma das províncias do inquérito.

O inquérito será dirigido aos agregados familiares por amostragem, pelo que será necessário cuidar de vários aspectos entre os quais se destaca a preparação da base de amostragem e o plano do desenho da amostra a ser utilizado.

Âmbito Geográfico

Devido a questões de ordem financeira e de segurança político - militar, achou-se conveniente realizar o inquérito num conjunto de cidades que de acordo com os objectivos do IDR e as características a analisar, apresentassem características que retratassem a situação actual do país. Assim sendo, o inquérito será realizado nas províncias/ cidades abaixo indicadas:

Quadro 1 – Distribuição das Províncias por Domínios de Estudo

Província	Domínio de Estudo	Caracterização
Luanda	Toda a província	Urbano e Rural
Cabinda	Cidade de Cabinda	Urbano
Benguela	Cidades de Benguela, Catumbela e Lobito	Urbano
Namibe	Toda a província	Urbano e Rural
Lunda Norte	Cidades do Dundo, Chitato e Luachimo	Urbano

Huíla	Cidade do Lubango	Urbano
Cunene	Toda a província	Urbano e Rural

Base de Amostragem e Unidade de Amostragem

Devido ao grande crescimento da população nas capitais de províncias, e a desactualização da cartografia existente em todas as províncias onde se irá realizar o IDR, foi implementada nestas províncias uma operação de contagem rápida das habitações e actualização cartográfica que permitisse a formação da base de amostragem.

A operação de contagem de habitações e actualização cartográfica consistiu num trabalho de campo em que os técnicos do INE e pessoal contratado para o efeito, com base nos planos cartográficos das cidades, realizaram uma contagem rápida do número de habitações em cada quarteirão, assinalando este nº no mapa e delimitando os limites físicos de cada quarteirão. Após este trabalho foram constituídos os conglomerados com uma média de 100 habitações seguindo um conjunto de regras previamente elaboradas. As secções são constituídas por parte de um, um ou mais quarteirões contíguos.

Para a área rural será utilizada a base de amostragem existente no INE e utilizada para os inquéritos realizados junto aos agregados familiares. Esta base de amostragem é constituída a partir da Classificação da Divisão Política e Administrativa (DPA) e contém informação sobre a população.

A Divisão Política e Administrativa da República de Angola está constituída por províncias, cada uma das quais está dividida em Municípios, e estes por sua vez subdivididos em Comunas. Cada comuna é constituída por várias aldeias ou bairros.

O IDR terá como unidade estatística de amostragem o agregado familiar. A unidades de análise serão o agregado familiar e o indivíduo, de acordo com as secções do questionário.

Desenho da Amostra

O plano amostral escolhido para o IDR utiliza um desenho com duas fases de selecção na área urbana e três na área rural. As amostras serão independentes para cada uma das províncias. Para as províncias do Namibe e Cunene a amostra será independente por área urbana e rural.

Área urbana:

A unidade de amostragem da 1ª fase (UPA) será o conglomerado, que deverá ser seleccionado com probabilidade proporcional ao seu número de habitações. A unidade de selecção da 2ª fase (USA) será o agregado familiar, cuja selecção deverá ser aleatória.

Área rural:

A unidade de amostragem da 1ª fase (UPA) será a comuna, que deverá ser seleccionada com probabilidade proporcional a sua população. A unidade da 2ª fase será a aldeia (USA), cuja selecção deverá ser aleatória. A unidade da 3ª fase (UTA) será o agregado familiar, cuja selecção deverá ser, também, aleatória.

Tamanho da Amostra

O tamanho da amostra para um inquérito é determinado pelo orçamento e recursos disponíveis. Uma amostra demasiado grande implica maiores recursos financeiros para a sua implementação, enquanto uma amostra muito pequena diminui a fiabilidade dos resultados. Por outro lado a dimensão da amostra depende do nível de precisão e grau de confiança pretendidos para as estimativas a obter por domínio de estudo. A precisão dos resultados do inquérito depende dos erros amostral e não amostrais. O erro amostral é medido pelo estimador da variância. Os erros não amostrais poderão ser avaliados e controlados através de estudos de validação, reentrevistas e controle de qualidade dos dados.

O erro amostral será menor quanto maior for o tamanho da amostra, ou seja, inversamente proporcional ao tamanho da amostra. Por outro lado os erros não amostrais tendem a aumentar com o tamanho da amostra, porque é torna-se mais difícil controlar a qualidade dos dados de uma amostra grande. Portanto, é

fundamental encontrar um tamanho da amostra adequado ao orçamento, e que permita qualidade e controle operacional.

As variáveis mais apropriadas para determinar o tamanho (dimensionar) da amostra do IDR, ao nível dos agregados familiares são as variáveis receitas e/ou despesas, uma vez que um dos objectivos do inquérito se refere a quantificação dos gastos monetários de consumo efectuados pelos agregados. As estimativas das variáveis utilizadas para o dimensionamento da amostra do IDR foram obtidas com base na informação disponível a partir do *Inquérito Prioritário sobre as Condições de Vida nos Domicílios (IPCVD)* realizado em 1995 e os Resultados do *Inquérito aos Agregados Rurais* realizado pelo Ministério da Agricultura e Desenvolvimento Rural em 1997, para a área urbana e rural de cada província, respectivamente..

Para a área urbana, a variável receitas não foi consistente, tendo sido adoptada a variável despesas, para dimensionar a amostra.

Para a dimensão da amostra estipulou-se o grau de confiança em 95%.

A expressão utilizada foi a seguinte:

$$n = \frac{z_{\alpha} s^2}{d^2} \times deff$$

onde

n - é a dimensão da amostra para cada domínio de estudo, ou seja, o nº de agregados familiares a inquirir em cada área de estimação.

$z_{\alpha} = 1.96 \cong 2$ - é o valor da estatística z para um nível de confiança de 95% da distribuição normal.

d - é o erro absoluto admitido.

s^2 - é a estimativa da variância da população para uma variável específica

$Deff = 2,3$ - é o factor de ajustamento para o método de amostragem utilizado

A dimensão da amostra para a província de Luanda foi fixada em 2568 agregados familiares, para a variável média das despesas do agregado de modo a garantir um erro de amostragem não superior a 5%.

Quadro 2 – Cálculo do Tamanho da Amostra para a Província de Luanda

Indicadores	Prevalência Estimada (NKz)	Erro Admitido (%)
Média de despesas por agregado	299,162,647.00	4.8
Variância da Média de despesas por agregado	46,194,751,815,285,800.00	
Média de despesas em alimentação	218,100,756.50	4.9
Variância da Média de despesas em alimentação por agregado	23,485,250,464,562,200.00	

A dimensão da amostra para a área rural da província do Cunene foi fixada em 780 agregados familiares, para a variável rendimentos de modo a garantir um erro máximo de 5.4%. A amostra para a área rural da província do Namibe foi fixada em 570 agregados familiares.

Para as restantes províncias o IDR será realizado apenas na a área urbana e a dimensão da amostra para cada província foi fixada em 1200 agregados familiares.

O tamanho da amostra permite produzir estimativas para os indicadores de cada província e ainda separadamente para a área urbana e rural da província do Cunene e Namibe.

O tamanho total da amostra é de 10610 agregados familiares distribuídos pelos 877 conglomerados, conforme indicado no quadro abaixo:

Quadro 3 – Distribuição dos Conglomerados e Agregados Familiares por Províncias e Áreas

Província	Agregados Familiares na Amostra	Área Urbana		Área Rural		
		Agregados Familiares na Amostra	Número de conglomerados na Amostra	Agregados Familiares na Amostra	N.º de Comunas na Amostra	N.º de Aldeias na Amostra
Total	10610	9300	746	1310	23	131
Luanda	2700	2700	225			
Cabinda	1200	1200	100			
Benguela	1200	1200	100			
Huíla	1200	1200	100			
Cunene	1580	830	41	740	13	74
Namibe	1530	960	80	570	10	57
Lunda Norte	1200	1200	100			

Listagem dos Agregados Familiares

Nem sempre uma habitação corresponde a um agregado familiar (AF), subestimando assim, o número de agregados familiares. Para evitar a subestimação dos AF será realizada a operação de listagem dos agregados familiares. A operação de listagem tem como finalidade determinar o nº de AF nas secções/aldeias seleccionadas para a amostra, a partir da qual serão seleccionados os agregados familiares. Para efectuar a operação de listagem em cada conglomerado (secção/aldeia), o registador deverá utilizar um croqui, a fim de facilitar o seu trabalho de campo. Para esta actividade deve-se primeiro percorrer toda a área a listar, fazendo o reconhecimento dos seus limites, segundo o percurso estabelecido (esquina noroeste, e sempre com a mão direita junto a parede exterior das habitações).

A operação de listagem será independente por conglomerado, durante a qual a cada AF será atribuído um nº de ordem, que deve ser inscrito na listagem conjuntamente com o nome do chefe do AF e endereço da habitação. Este nº de ordem será a base para identificação da selecção dos AF no conglomerado. Para o efeito, serão utilizados formulários simples para a recolha da informação necessária. A operação de listagem permitirá obter ponderadores mais exactos para a amostra.

Seleccção da Amostra

Notação:

A) Área Urbana

Para cada província tem-se:

- ❖ N_i - é o número de habitações no conglomerado i ;
- ❖ N'_i - é o número de agregados familiares no conglomerado i , obtidos a partir da operação de listagem;
- ❖ N - é o número de habitações na área urbana da província;
- ❖ s - é o número de conglomerados a seleccionar na área urbana da província;
- ❖ I_1 - é o intervalo de amostragem para a selecção sistemática dos conglomerados;
- ❖ $n_i=12$ - é o número de agregados a seleccionar no conglomerado i ;
- ❖ n'_i - é o número de agregados familiares entrevistados no conglomerado i ; $n'_i \leq 12$.

Amostra de Conglomerados

A selecção dos conglomerados será feita de forma independente em cada província com probabilidade proporcional ao número de habitações nele existente, usando a selecção sistemática.

Para identificação das unidades da amostra, durante o processo de realização do inquérito, quer nas actividades de campo, quer na garantia do controle e da qualidade, quer para o processo de estimação, serão utilizados os códigos geográficos, sendo os três primeiros correspondentes a Divisão Política Administrativa de Angola:

- ❖ Província - 2 dígitos

- ❖ Município - 2 dígitos
- ❖ Comuna - 2 dígitos
- ❖ Área – 1 dígito (Urbana =1, Rural =2)
- ❖ Zona - 2 dígitos
- ❖ Aldeias/Bairros – 2 dígitos
- ❖ Conglomerado - 3 dígitos

O procedimento utilizado para a selecção dos conglomerados é o descrito abaixo:

1. Prepara-se para cada província uma lista com todos os seus conglomerados;
2. Regista-se o número de habitações - N_i - obtido da operação de contagem de habitações para cada conglomerado;
3. Calcula-se o número acumulado de habitações. O último número acumulado será igual ao número de habitações da área urbana na província;
4. Determina-se o intervalo de selecção $I_1 = \frac{N}{s}$, dividindo o número total de habitações pelo número de conglomerados a seleccionar no respectivo domínio de estimação;
5. Selecciona-se um número aleatório de arranque A entre 1 e I_1 , inclusive, guardando as casas decimais, no caso do intervalo de amostragem não ser um nº inteiro;
6. O 1º conglomerado seleccionado será o que apresentar um número acumulado mais próximo de A , sem excedê-lo;
7. Os restantes conglomerados a seleccionar, obtêm-se, adicionando o intervalo de selecção ao nº de arranque A .

Amostra de Agregados Familiares

A selecção dos AF será feita sistematicamente nos formulários de registo de AF de cada conglomerado seleccionado para a amostra.

O intervalo de selecção sistemática é obtido pelo quociente entre o número de AF listados no conglomerado e o número de AF que se pretende seleccionar.

$$I_2 = \frac{N_i}{12}$$

Selecciona-se um número aleatório, $A_1 \leq I_2$, que será o número de ordem do primeiro AF seleccionado. Para a selecção dos restantes no conglomerado i , basta adicionar o intervalo I_2 ao número aleatório A_1 .

Substituição das Famílias

Poderão surgir casos em que se impõe a substituição dos AF devido as situações abaixo indicadas:

- a) A habitação está ocupada por um AF diferente daquele que consta do formulário do Registo;
- b) Não identificação do AF devido a deficiências na listagem;
- c) Impossibilidade de contacto por estar ausente durante o período de realização do inquérito;
- d) Recusa.

No caso a) o AF deve ser substituído pelo AF que esteja a viver actualmente na habitação.

Nos restantes casos haverá substituição do AF. Para manter em 12 o nº de AF a serem inquiridos por conglomerado e a fim de evitar recorrer a substituição dos AF de forma desorganizada, optou-se por disponibilizar para cada conglomerado seleccionado seis AF suplentes seleccionados, também, aleatoriamente. Os seis

AF suplentes ou de reserva só deverão ser utilizados em caso de absoluta necessidade.

A probabilidade de selecção de um AF suplente não será utilizada para o cálculo dos ponderadores por ser aproximadamente zero.

B) Área Rural

Para cada província tem-se:

- ❖ C - é a estimativa do total da população na área rural da província;
- ❖ C_i - é a estimativa do total da população na área rural da comuna i ;
- ❖ c - é o n.º de comunas a seleccionar em cada província;
- ❖ $a = 6$ - é o n.º de aldeias a seleccionar em cada comuna seleccionada;
- ❖ A_i - é o n.º total de aldeias na comuna i ;
- ❖ A_{ij} - é o n.º de AF na aldeia j da comuna i ;
- ❖ I_1 - é o intervalo sistemático de selecção das comunas;
- ❖ I_2 - é o intervalo sistemático de selecção das aldeias;
- ❖ I_3 - é o intervalo sistemático de selecção dos AF;
- ❖ $n_{ij} = 10$ - é o n.º de AF a seleccionar na aldeia j da comuna i ;
- ❖ n'_{ij} - é o n.º de AF entrevistados na aldeia j da comuna i : $n'_{ij} \leq 10$.

Amostra de Comunas

A selecção das comunas nas províncias do Namibe e Cunene será independente da área urbana e ser proporcional a estimativa da população nela existente, utilizando o procedimento seguinte:

1. prepara-se para cada província uma lista com todas as suas comunas, as respectivas estimativas da população e os totais acumulados dessas estimativas;

2. calcula-se o intervalo de selecção $I_1 = \frac{C}{c}$, determinado pela estimativa total da população da área rural da comuna, dividido pelo n.º de comunas a seleccionar na respectiva província. É necessário manter as casas decimais;
 3. selecciona-se um n.º aleatório de arranque $A_1 \in [1, I_1]$
 4. selecciona-se a 1ª comuna que apresentar um acumulado igual ou maior a A_1
 5. para a selecção das restantes comunas adiciona-se o intervalo de selecção, ou seja:
 - a) para a 2ª comuna será $A_1 + I_1$;
 - b) para a 3ª comuna será $A_1 + I_1 + I_1 + \dots$
- (As comunas seleccionadas serão aquelas que apresentarem um acumulado mais próximo do n.º calculado em 5);
6. repete-se a operação descrita em 5), até obter as “c” comunas a seleccionar por província.

Amostra de Aldeias

Em cada comuna as aldeias foram ordenadas em ordem sequencial do seu código geográfico.

Em cada comuna seleccionada serão seleccionadas aleatoriamente 6 aldeias de forma sistemática.

O intervalo de selecção na 2ª etapa será :

$$I_2 = \frac{A_i}{6}$$

O n.º aleatório de arranque será $A_2 \in [1, I_2]$

A selecção das 6 aldeias será efectuada de modo idêntico ao descrito no ponto anterior.

Amostra de Agregados Familiares

A selecção dos AF será feita sistematicamente na lista preparada dos AF de cada aldeia seleccionada para a amostra. Para o efeito, para cada aldeia deve-se ter, os seus AF, associados a um número entre 1 e A_{ij} . O intervalo de selecção sistemática é obtido pelo quociente entre o número de AF listados na aldeia e o número de AF que se pretendem a inquirir na aldeia - $n_{ij} = 10$.

$$I_3 = \frac{A_{ij}}{n_{ij}}$$

Tal como na área urbana, na área rural, para além dos 10 AF que se pretende inquirir, serão seleccionados seis AF suplementes. O procedimento será similar ao descrito no caso urbano.

Distribuição Temporal da Amostra de Conglomerados

O IDR foi planeado para ser realizado no período de um ano. Foram criados períodos de trabalho de 12 dias, sendo 8 dias para a recolha da informação, 2 dias para correcção, revisão e controle dos dados e os restantes 2 dias para repouso da equipa de campo.

Conhecidas as secções e aldeias (conglomerados) seleccionadas em cada província, proceder-se-á a sua distribuição pelos 24 períodos de trabalho de duração do inquérito, assim todas os conglomerados estarão associadas a um período de trabalho. Este procedimento permite conhecer os conglomerados a serem inquiridos em cada período.

Probabilidades de Selecção

A) Área Urbana

- A probabilidade de selecção do conglomerado i é dada por:

$$P_i = s \times \frac{N_i}{N} \quad \text{Se } N_i < I_1$$

$$P_i = 1 \quad \text{Se } N_i \geq I_1$$

- A probabilidade de selecção do AF j , do conglomerado i é dada por:

$$P_{ij} = s \times \frac{n_i}{N} \quad \text{Se } P_i = 1$$

$$P_{ij} = \frac{n_i}{N'_i} \quad \text{Se } P_i < 1$$

Dentro do conglomerado todos os agregados têm igual probabilidade de selecção.

- A probabilidade final de selecção de cada AF é dada pelo produto das probabilidades de selecção das unidades em cada uma das etapas. Assim teremos :

$$P_i \times P_{ij} = s \times \frac{N_i}{N} \times \frac{n_i}{N'_i} \quad \text{Se } P_i < 1$$

$$P_i \times P_{ij} = s \times \frac{n_i}{N} \quad \text{Se } P_i = 1$$

B) Área Rural

Para a área rural de cada província tem-se que :

- A probabilidade de selecção da comuna i é dada por:

$$P_i = c \times \frac{C_i}{C} \quad \text{Se } C_i < I_1$$
$$P_i = 1 \quad \text{Se } C_i \geq I_1$$

- A probabilidade de selecção da aldeia j, da comuna i é dada por:

$$P_{ij} = a \times \frac{1}{A_i}$$

- A probabilidade de selecção do AF k da comuna i, da aldeia j é dada por:

$$P_{ijk} = \frac{n_{ij}}{A_{ij}}$$

- A probabilidade final de selecção de cada AF é dada por:

$$P_i \times P_{ij} \times P_{ijk} = c \times \frac{C_i}{C} \times \frac{a}{A_i} \times \frac{n_{ij}}{A_{ij}} \quad \text{Se } C_i < I_1$$

$$P_i \times P_{ij} \times P_{ijk} = \frac{a}{A_i} \times \frac{n_{ij}}{A_{ij}} \quad \text{Se } C_i \geq I_1$$

Assume-se que a probabilidade de selecção de todos os membros de um AF é igual a do AF, tanto na área urbana como rural.

Peso de Amostragem ou Factor de Expansão da Amostra

O peso de amostragem para cada AF é dado pelo recíproco da sua probabilidade de selecção.

O factor de expansão não depende de j e de k, ou seja, qualquer AF dentro do conglomerado i, ou aldeia j, tem o mesmo valor como peso de amostragem.

A) Área Urbana

$$w_i = w_{ij} = \frac{1}{P_i \times P_{ij}}$$

B) Área Rural

$$w_{ij} = w_{ijk} = \frac{1}{P_i \times P_{ij} \times P_{ijk}}$$

O tratamento das entrevistas não realizadas dos AF seleccionados para a amostra será feito com o ajuste do factor de expansão, considerando apenas as entrevistas realizadas

Área Urbana

$$w'_i = w_i \times \frac{n_i}{n'_i}$$

A) Área Rural

$$w'_{ij} = w_{ij} \times \frac{n_{ij}}{n'_{ij}}$$

Estimadores

O estimador para o total de uma variável Y é dado por:

A) Área Urbana

$$\hat{Y} = \sum_i w'_i Y_i \quad (1)$$

onde:

\hat{Y} é o total estimado para a variável Y;

w'_i é o peso de amostragem ajustado do conglomerado i;

$Y_i = \sum_i^{n'_i} Y_{ij}$ total da variável y na amostra conglomerado i

B) Área Rural

$$\hat{Y} = \sum_i \sum_j w'_{ij} Y_{ij} \quad (1)$$

onde:

\hat{Y} é o total estimado para a variável Y;

w'_{ij} é o peso de amostragem ajustado da aldeia j da comuna i;

Y_{ij} é o total da variável Y, na amostra, da aldeia j, comuna i.

$$Y_{ij} = \sum_k^{n'_{ij}} y_{ijk}$$

- O estimador dum dado ratio é dado pela expressão seguinte :

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

\hat{R} é o estimador para o ratio de 2 variáveis, Y/X;

\hat{Y} é o estimador do total para a variável Y, dado pela fórmula (1);

\hat{X} é o estimador do total para a variável X, dado pela fórmula (1).