

### Tamanho da amostra por concelho

O tamanho da amostra foi desenhado segundo os objectivos de precisão fixados de forma independente para cada um dos 22 concelhos. O termo representatividade foi expresso em termos técnicos estatísticos pelo coeficiente de variação (CV) associa a determinado nível de confiança  $\alpha$ . Neste caso, o CV pode ser exprimido da seguinte forma :

$$CV_{\alpha} = \frac{u_{\alpha} \sqrt{V(\hat{\theta})}}{\theta}$$

Em que  $\theta$  é o parâmetro estimado na população de referência, em IMC-2011 foi a taxa de desemprego estimada para a população de 15 anos ou mais e  $V(\hat{\theta})$  designa a variância do estimador fornecido pela amostra.

Dado que a *representatividade* foi considerada  $CV_{\alpha}$  é a mesma para cada concelho a formula para calculo do tamanho mínimo de amostra que satisfaça os objectivos fixados simplifica no caso de  $\theta$  ser a percentagem P.

$$n_0 = \frac{u_{\alpha}^2}{CV_{\alpha}^2} \times \frac{1-P}{P} \times deff_{\Pi}$$

O ajustamento utilizado para a população finita N é  $n = \frac{n_0}{1 + \frac{n_0}{N}}$ . Nesta expressão, o valor de  $deff_{\Pi}$  é

de la estrutura de sondagem (*Design Effect*) sobre o tamanho da amostra. E foi considerado de 0.61 com base nos dados do QUIBB 2006.

**Tabela 1 tamanho da amostra em termos de Agregados familiares, IMC-2011**

CONCELHO	N RGPH 2010	n
Ribeira Grande	4 551	440
Paul	1 637	375
Porto Novo	3 936	433
São Vicente	19 962	475
Ribeira Brava - São Nicolau	2 064	394
Tarrafal - S.Nicolau	1 277	353
Sal	6 796	454
Boa Vista	2 601	410
Maio	1 905	388
Tarrafal	4 285	437
Santa Catarina	8 608	461
Santa Cruz	5 664	448
Praia	33 239	480
São Domingos	2 693	412
São Miguel	3 544	428
São Salvador do Mundo	1 701	379
São Lourenço dos Orgãos	1 454	365
Ribeira Grande - Santiago	1 557	371
Mosteiros	2 220	399
São Filipe	5 097	444
Santa Catarina - Fogo	1 115	339
Brava	1 587	373
Total	117 493	9 058

## **Tipo de amostragem**

A amostragem do IMC-2011 é do tipo probabilístico a uma etapa realizada de forma independente, aleatória e sem reposição ao nível dos agregados familiares em cada concelho, estratificado implicitamente por Distrito de Recenseamento (DR) com afectação implícita e proporcional ao número de agregados por DR

## **Método de selecção dos agregados familiares**

O método de selecção dos agregados independentemente em cada concelho será uma tiragem sistemática circular no ficheiro de agregado familiar previamente ordenado segundo : meio (urbano/rural) e de seguida segundo o código agregado, ID\_Menage dentro de cada meio de residência. Como o código do Agregado familiar começa pelo código do Distrito de Recenseamento (DR), esta estratégia proporciona uma estratificação implícita segundo o distrito de recenseamento com uma alocação da amostra proporcional ao tamanho de cada DR. Este tipo de estratégia é conhecido como sendo óptimo no interior de cada domínio principal de análise (concelho). A estrutura da amostra respeita os critérios duma amostra probabilística em cada concelho e simplesmente aleatória sem reposição ao nível dos agregados familiares, estratificada por DR com alocação proporcional ao número de agregado por DR.

## **Tecnicidade de selecção : Tiragem sistemática circular**

A tiragem sistemática permite seleccionar facilmente uma amostra de tamanho fixo  $n$ , de uma população de tamanho  $N$  não seja necessariamente múltiplo de  $n$ , com uma probabilidade igual a  $n/N$  para cada unidade da população (neste caso os agregados) e de realizar uma estratificação implícita. O método circular será particularmente importante no momento de rotação parcial da amostra a partir de uma lista de agregados actualizada Para realizar a selecção sistemática circular o ficheiro deve estar previamente ordenado segundo o meio de residência, o código de identificação do agregado e os agregados sequenciados de 1 a  $N$  dentro de cada concelho.

## **Taxa de rotação de amostra**

Na concepção do IMC-2011 deve-se fazer a rotação da amostra segundo uma taxa de 1/3 ou cerca de 33,3%, podendo assim a amostra ser um múltiplo de 3 a fim de guardar uma amostra de tamanho fixo de 1/3 em todos os trimestres.

## **Frequência de rotação**

A frequência recomendada é semestral. Com uma rotação de 33,3% ao fim de seis meses o processo guarda sempre uma amostra de 33% que permanece durante um ano, permitindo ter dados para melhorar a precisão das comparações anuais para os inquéritos anuais.

## **Frequência de actualização da base de amostragem.**

No processo de amostragem de IMC-2011 a base de amostragem deve ser actualizada a cada dezoito meses.

## Estimação e ponderação

### Ponderação inicial

#### 1. Auto ponderação em cada concelho

Como a estrutura do IMC é probabilística independente em cada Concelho a probabilidade de selecção de cada agregado é o mesmo e é dado  $n_c/N_c$ , e portanto o peso ou factor de ponderação é :  $W_c = \frac{N_c}{n_c}$  em que  $N_c$  = número de agregado do concelho e  $n_c$  = tamanho da amostra no concelho.

Assim, a estrutura de sondagem proposta permite obter uma propriedade extremamente importante, seja a Auto ponderação dos agregados em cada concelho. Esta propriedade simplifica consideravelmente a estimação em cada concelho e torna a precisão muito eficaz.

### Estimador de base

O estimador de base em cada concelho é o estimador do total duma variável qualquer  $Y$  para o conjunto de agregados é dado por  $\hat{Y}_c = \sum_{i=1}^{n_c} W_c y_{ci}$ . A ponderação inicial  $W_c$  será guardada constante tão longe que possível até que não se faça a rotação parcial da amostra.

**Tabela 2: Ponderação inicial da amostra**

CONCELHO	N RGPB 2010	n final	Ponderação inicial
Ribeira Grande	4.551	441	10,32
Paul	1.637	375	4,37
Porto Novo	3.936	435	9,05
São Vicente	19.962	477	41,85
Ribeira Brava - São Nicolau	2.064	396	5,21
Tarrafal - S.Nicolau	1.277	354	3,61
Sal	6.796	456	14,90
Boa Vista	2.601	411	6,33
Maio	1.905	390	4,88
Tarrafal	4.285	438	9,78
Santa Catarina	8.608	462	18,63
Santa Cruz	5.664	450	12,59
Praia	33.239	480	69,25
São Domingos	2.693	414	6,50
São Miguel	3.544	429	8,26
São Salvador do Mundo	1.701	381	4,46
São Lourenço dos Orgãos	1.454	366	3,97
Ribeira Grande - Santiago	1.557	372	4,19
Mosteiros	2.220	399	5,56
São Filipe	5.097	444	11,48
Santa Catarina - Fogo	1.115	339	3,29
Brava	1.587	375	4,23
Total	117.493	9.084	

## Ponderação na presença de rotação

Deve-se distinguir dois casos

### **Caso I : ponderação se a rotação parcial teve lugar após a actualização da base de amostragem**

A base de amostragem tem um conjunto de agregados que provem de dois períodos  $N_{c1}$  et  $N_{c2}$ .

A rotação de  $(1/3)$  de amostra é realizada de maneira aleatória por um processo sistemático de  $n_c/3$  agregados de entre os  $N_{c2}$  disponíveis.

Por isso teremos duas amostras no período 2 disponíveis para análise:

- A amostra do painel de tamanho  $2n_c/3$ .
- A nova amostra de tamanho  $n_c/3$ .

Na primeira rotação cada agregado da amostra de painel tem a probabilidade de pertencer a amostra definida por  $\frac{1}{W_{c1}} \times \frac{2}{3}$

Para a estimação do total no período 2 proveniente do painel, cada agregado terá o peso seguinte, chamado peso do painel :  $W_{c2}^p = W_{c1} \times \frac{3}{2}$

A nova amostra obtida no período 2 fornece também um estimador. Nesta estimação, cada agregado familiar recebe o peso seguinte, chamado peso de rotação:  $W_{c2}^r = \frac{N_{c2}}{n_c} \times 3$

Os dois estimadores serão adicionados num único, utilizando a amostra completa de tamanho  $n_c$  de agregados no período 2. Uma forma racional de se fazer a agregação das duas estimações é utilizar a média ponderada pelo tamanho relativo de cada amostra proveniente, ou seja  $2/3$  para amostra de painel e  $1/3$  para amostra de rotação. Assim, os agregados de amostra de painel terão no estimador agregado o peso de  $W_{c2}^a = W_{c1}$ . O que corresponde à primeira rotação após a actualização da base a :  $W_{c2}^a = \frac{N_{c1}}{n_c}$ .

Os agregados da amostra de rotação tem no estimador agregado o peso de  $W_{c2}^a = \frac{N_{c2}}{n_c}$ .

A aplicação deste procedimento de estimação na segunda rotação com actualização da base de amostragem, em que ter-se-ia o tamanho  $N_{c3}$  conduzirá a considerar três amostras em vez de duas, cada tamanho de amostra de  $n_c/3$ , um estimador agregado através da média ponderada de três estimadores, e por amostra da segunda rotação, em que o peso terá a seguinte expressão:

$$W_{c2}^a = \frac{N_{c3}}{n_c}$$

### **Caso II : ponderação se a rotação parcial teve lugar sem actualização da base de amostragem**

Neste caso a mesma teoria de estimação será utilizada se não se realizou a actualização da base de amostragem, com a simplificação adicional já que  $N_{c2}=N_{c1}$ . Assim a amostra de rotação tem o mesmo peso no estimador agregado que no estimador do painel

#### O estimador de painel de variação trimestral na população

De uma maneira geral, o estimador « natural » no trimestre t tem a seguinte formula de estimação :

$$\widehat{Y}_{c,t} = \sum_{i=1}^{n_c} W_{c,t}^a y_{ci,t}$$

No trimestre seguinte t+1, teremos o segundo estimador :

$$\widehat{Y}_{c,t+1} = \sum_{i=1}^{n_c} W_{c,t+1}^a y_{ci,t+1}$$

A primeira maneira de estimar a variação trimestral será evidentemente de considerar a diferença entre os dois estimadores agregados em trimestre t+1 et t :

$$\widehat{Y}_{c,t+1} - \widehat{Y}_{c,t}$$

É necessário ver que este estimador incorpora dois efeitos : o efeito de rotação e o efeito de crescimento da população entre os dois trimestres, chamado efeito da base de amostragem.

Um estimador mais eficaz da variação trimestral, eliminando estes dois efeitos é utilizar somente amostra de painel de dois semestre consecutivos, ou seja a amostra de 2/3 (66,6%) da amostra de tamanho  $n_c$ , se houve rotação, e a amostra de 100% do tamanho de amostra de  $n_c$ , se não houve rotação. Para cada agregado  $i$  do painel, mede-se a diferença  $y_{ci,t+1}^d = y_{ci,t+1} - y_{ci,t}$  .

É preciso agora aplicar a esta diferença um peso apropriado, seja o peso painel dado por

- $W_c^p = W_{c,t} \times \frac{3}{2}$ , se houver rotação,
- Ou simplesmente  $W_c^p = W_{c,t}$  se não houver rotação.

Este estimador é denominado estimado painel da variação trimestral

Assim, do ponto de vista prático, deve-se guardar no ficheiro de dados duas variáveis de ponderação:

- O peso painel definido acima,
- O peso corrente utilizado no estimador do trimestre t+1, definido anteriormente para o caso I e II.

Estas duas variáveis de ponderação devem ser actualizadas sempre que se realize o Inquérito tendo em conta a rotação ou não da amostra com e a presença ou não de actualização da base de sondagem.

### Ponderação na presença de rotação

Resumindo as principais características do sistema de ponderação que implica dois pesos :

- O peso corrente de um agregado da amostra é dado por  $W_c^a = \frac{N_{ct}}{n_c}$ 
  - Em que  $N_{ct}$  é o tamanho da base de amostragem em cada concelho no trimestre t em que os agregados são seleccionados pela primeira vez.
  - $n_c$  é o tamanho da amostra constante na amostra do concelho estabelecido inicialmente.
- O peso painel para um agregado que permanece no trimestre seguinte t+1, ou seja o seu peso corrente do período precedente t, multiplicado por 3/2 e houver rotação parcial da amostra de um agregado sobre três (1/3) da amostra no trimestre t+1.