

Probabilidade de sondagem (seleção dos elementos) e coeficiente de ponderação

O inquérito de gasto e satisfação de turistas realizado em 2011 (IGST-2011) visava, de entre outros, dois objetivos fundamentais:

- 1) Obter dados fiáveis e consistentes para aferir sobre a estrutura de gasto turístico a fim de estabelecer um perfil atualizado do turismo em Cabo Verde
- 2) Obter elementos para a montagem de um sistema de cálculo da conta satélite do turismo.

Os dados recolhidos que ajudam a realizar estes objetivos são, na sua maioria, de carácter quantitativo. E mais, tendo em conta que a repartição da amostra de turistas não é proporcional aos estratos considerados (tipo de voos: regulares e não regulares, época de estudo: época alta e época baixa), deve-se utilizar os pesos de sondagem/amostragem em toda as análises que utilizam aos dados contínuos desta pesquisa para melhor garantir a representatividade atual da amostra ao nível nacional e ao nível do domínio de estudo.

O quadro 1 mostra a taxa de sondagem inicialmente proposta para esta pesquisa segundo o domínio de estudo e tipo de voo.

Quadro 1: Repartição da amostra do IGST-2011, por domínio de estudo e estratos

Estratos	Domínios de estudo			
	Sal	Boa Vista	Praia	S.Vicente
Charter	8%	8%	10%	10%
Regular	12%	12%	10%	10%
Total	10%	10%	10%	10%

Fonte:INE,ISGT-2011

A taxa de sondagem efetiva da pesquisa variou em torno da taxa ou fração de sondagem de 10% previamente estipulada. Na grande maioria dos voos pesquisados obedeceu-se a taxa de amostragem fixa de 10% dos passageiros a embarcar por voo.

Quadro 2: Resumo de turistas previsto a inquirir por época e a taxa de sondagem, efetivamente, inquirida, IGST-2011

Época	Total embarcado	Total inquirido	%	Taxa de sondagem (%)
Alta	9321	947	53,4	10,2
Baixa	8055	826	46,6	10,3
Total	17376	1773	100	10,2

Fonte:INE,IGST-2011

Probabilidade seleção de um turista, e coeficiente de ponderação

Dado o objetivo geral da pesquisa e, com intuito de facilitar o cálculo dos ponderadores (fator de extrapolação ou elevação) da amostragem, as probabilidades de seleção dos turistas em cada fase da amostragem, serão calculados por nacionalidade dos turistas (as mais frequentes e para grupos de países com certa similitude), para a época alta e para a época baixa. O agrupamento dos países foi metodologicamente escolhido par facilitar a comparabilidade com outros estudos.

Do ponto de vista duma sondagem por conglomerados a varias etapas, os voos são considerados as unidades primárias de seleção (UPS). No IGST todos os voos internacionais com pelo menos 30 passageiros foram selecionados. Como a pesquisa foi feita em todos os aeroportos de ondem partem voos internacionais, a escolha do aeroporto não constitui fase de seleção e por isso a sua escolha foi exaustiva com probabilidade **igual à unidade**.

Relativamente à escolha de voos, também era previsto ser realizado **em todos os voos (probabilidade de seleção dos voos seria igual à unidade)**, partindo de hipótese que todos os voos internacionais tivessem pelo menos 30 passageiros. Verificou-se que durante a época de realização da pesquisa houve voos que partiram dos aeroportos internacionais com menos de 30 passageiros.

Quadro 3: Capacidade do voo (marcados e previstos na data de entrevista)

Época	Com >=30 passageiros	Com <30 passageiros	Total
Alta	88	4	92
Baixa	73	8	81
Total	161	12	173

Fonte:INE,IGST-2011

Dado que um voo era selecionado se tivesse pelo menos 30 passageiros, o quadro 3 pode ajudar a estimar a probabilidade de seleção de um voo em cada uma das épocas. Assim, para o $i^{\text{ésimo}}$ voo (UPS) do estrato h , as notações são as seguintes:

h : estrato: época da realização da pesquisa

$h=1$: Época Alta

$h=2$: Época Baixa

P_{1hi} : Probabilidade de sondagem na primeira etapa do $i^{\text{ésimo}}$ voo da época h

P_{2hi} : Probabilidade de sondagem na segunda etapa (escolha de turistas em cada voo)
 $i^{\text{ésimo}}$ voo escolhido do estrato h

Seja n_h o numero total de voos (UPS) com pelo menos 30 passageiros na época h , M_h o numero total de voos na época h . A probabilidade de selecionar um voo com pelo menos 30 passageiros na época h ($h=1; h=2$) pode ser corrigido por aproximada por:

$$P_{1h} = \frac{n_h}{M_h} \quad \text{dai,} \quad P_{11} = \frac{88}{92} \quad \text{e} \quad P_{12} = \frac{73}{81}$$

Na segunda etapa, selecionou-se b_{hi} de turistas a partir dos L_{hi} turistas que embarcam no $i^{\text{ésimo}}$ voo do estrato (época) h . Portanto :

$$P_{2hi} = \frac{b_{hi}}{L_{hi}}$$

Nesta pesquisa, em todos os voos selecionados (amostrados) foram entrevistados uma taxa

constante de 10% ou seja 1/10. Portanto $P_{2hi} = \frac{b_{hi}}{L_{hi}} = \frac{1}{10}$

Exemplo, se um voo tivesse 120 passageiros selecionou-se os turistas numa taxa (probabilidade iguais de) 12/120 ou seja 1/10.

- 1) A probabilidade global para selecionar um turista no $i^{\text{ésimo}}$ voo e por época h (Alta e Baixa), pode ser aproximada como :

- 2) Para época alta, $P_{1i} = P_{11i} \times P_{21i} = \frac{88}{92} \times \frac{1}{10}$

1) Para época Baixa

$$P_{1i} = P_{12i} \times P_{22i} = \frac{73}{91} \times \frac{1}{10}$$

A componente principal do peso de sondagem é o inverso da probabilidade global para seleccionar um turista:

$$W_{hi} = \frac{1}{P_{hi}} \text{ para cada uma das épocas.}$$

Cálculo do fator de ponderação ajustado (modificado)

Com os dados da Agencia de Aviação Civil (AAC) relativo às estatísticas de voos internacionais procura-se preferencialmente encontrar o número total de pessoas embarcados por voos (charter e regular), por aeroporto, ou na pior das hipóteses o total por mês. Se for possível encontrar estes dados, o fator de extrapolação poderá ser calculado da seguinte forma.

Portanto a extrapolação será feita por nacionalidade. Para o efeito, criou-se uma variável que agrupe as nacionalidades segundo a frequência e que acumule cerca de 80% dos turistas, independentemente da época.

$$f_{hj} = A_h * \frac{W_{hj} * T_{hj}}{n_{hj}}$$

Em que,

f_j: é o factor de extrapolação em que se multiplicara por cada registo correspondentes aos indivíduos de 15 anos ou mais na amostra da nacionalidade j.

W_h: é o peso ou a percentagem da sondagem dos turistas não residentes inquiridos na amostra segundo a época h (h=1: época baixa, h=2: época alta)

T_h: é o total de pessoas efetivamente embarcadas (observados) por voo fornecido pela AAC para o mês e ano de estudo segundo o nível de detalhe supracitado.

n_h : é o total de turistas por cada nacionalidade com entrevistas validas (incluindo os não residentes).