



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

## » Utilização de informação auxiliar no processo de estimação

Uma aplicação no contexto das empresas «

Rita Sousa, Conceição Ferreira  
DMSI/ME

 30/03/2012



## Índice



- Introdução
- IUTICE
- Motivação
- Metodologia atual do IUTICE
- Método da calibração
- Ajustamento às atuais dimensões
- Principais resultados
- Conclusões
- Considerações finais



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL





## Introdução



- » No contexto das empresas existem fatores gerados pela sua demografia que dão origem a alterações frequentes de estrutura.
- » As alterações na actividade económica bem como na dimensão da empresa provocam diferenças significativas nos universos de referência, entre os momentos de seleção das amostras e de apuramento de resultados.
- » A utilização de informação auxiliar pode constituir uma mais-valia na correção dos desenhos amostrais e no aumento de precisão das estimativas.



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



## IUTICE

Inquérito à Utilização de Tecnologias da Informação  
e da Comunicação nas Empresas



- » Este estudo baseia-se numa aplicação ao Inquérito à Utilização de Tecnologias da Informação e da Comunicação nas Empresas (IUTICE).
- » A informação produzida por este inquérito resulta de diferentes módulos específicos sobre a utilização das Tecnologias da Informação e da Comunicação (TIC) :
  - ✓ acesso e utilização de Internet;
  - ✓ transmissão e partilha electrónica de informação;
  - ✓ sistemas de segurança;
  - ✓ recursos e despesas nas TIC.



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL





## IUTICE

Inquérito à Utilização de Tecnologias da Informação e da Comunicação nas Empresas



» A população alvo são as empresas sedeadas em território nacional, com 1 ou mais trabalhadores, cuja atividade económica se insere nas secções da CAE-Rev.3: C, D e E, F, G, H, I, J, K, L, M, N, S.

» Os resultados desta operação estatística baseiam-se numa amostra aleatória estratificada, selecionada pelas seguintes variáveis de estratificação:

- ✓ atividade económica;
- ✓ número de pessoas ao serviço;
- ✓ volume de negócios.



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



## Motivação



» A amostra do IUTICE garante representatividade em diferentes agrupamentos da CAE Rev. 3 e nos escalões do Número de Pessoas ao Serviço (NPS), de acordo com as recomendações do Eurostat.

» A diversidade de indicadores anuais obedece a muitas desagregações além das consideradas na seleção da amostra, o que origina estimativas com níveis de precisão nem sempre concordantes com os padrões desejáveis.

» Neste estudo avaliam-se diferentes metodologias de ajustamento, por utilização de informação auxiliar no processo de estimação.



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL





## Aplicação prática ao IUTICE



- » Esta análise é efetuada com base nos dados da amostra do IUTICE recolhida em 2010.
- » A base de recolha é constituída por 5524 unidades respondentes para um universo de referência de cerca de 270000 empresas.
- » As variáveis de análise são os principais indicadores, relativos à utilização das tecnologias de informação e da comunicação:
  - ✓ utilização de computador;
  - ✓ acesso à internet;
  - ✓ existência de *website*.



## Metodologia atual do IUTICE



- » A alteração da atividade económica e da própria dimensão da empresa são fatores que levam à mudança frequente de estrato, no intervalo de tempo que separa os momentos de seleção e de apuramento.
- » O método atualmente utilizado no IUTICE consiste em recorrer à Informação Empresarial Simplificada (IES) para reclassificar as empresas no estrato final.
- » Esta metodologia já utiliza informação auxiliar, sendo os pesos amostrais recalculados de acordo com a nova distribuição das empresas nos estratos finais.





## Metodologia atual do IUTICE



- » A atual metodologia do IUTICE classifica as empresas nos estratos de acordo com a informação atual, dando origem a novos pesos amostrais  $d_k^*$ .
- » O total de uma dada variável de interesse  $Y$  é estimado com os pesos amostrais ajustados, pelo usual estimador de *Horvitz-Thompson*:

$$\hat{\tau}_{y,HT}^* = \sum_{h=1}^H \sum_{k \in s_h} d_k^* y_k$$

- » Para uma amostra aleatória estratificada, a variância deste estimador é dada por:

$$\text{Var}(\hat{\tau}_{y,HT}^*) = N^2 \sum_{h=1}^H W_h^2 (1 - f_h) \frac{S_h^2}{n_h}$$

- ✓ Em que  $W_h = \frac{N_h}{N}$  corresponde à percentagem de empresas no universo do estrato  $h$
- ✓ E  $f_h = \frac{n_h}{N_h}$  à fração de amostragem no estrato  $h$ .



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



## Métodos alternativos



- » Como alternativa à atual metodologia do IUTICE são testados 2 métodos que ajustam os pesos amostrais iniciais, mantendo a estratificação inicial.
- » A pós-estratificação e a calibração são técnicas que permitem aumentar a eficiência dos estimadores, ajustando os pesos amostrais iniciais de acordo os totais conhecidos da população.
- » O ajustamento dos pesos amostrais iniciais permite aproximar o desenho inicial da amostra à realidade do momento de apuramento.



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL





## Método de calibração



- » A calibração do desenho amostral consiste em recorrer a informação auxiliar para ajustar os pesos amostrais iniciais, obtendo-se totais marginais idênticos aos conhecidos da população.
- » O ajustamento dos pesos amostrais iniciais, face aos totais marginais, é feito por minimização de uma dada função distância, que permite obter novos ponderadores tão próximo quanto possível dos iniciais.



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



## Método de calibração



- » **Pesos amostrais ajustados, no indivíduo  $k$ , como solução de:**

$$w_k : \text{Min}_{w_k} \sum_{k \in S} d_k G(w_k / d_k)$$

- ✓ sob a restrição para o total marginal da população:  $X = \sum_{k \in S} w_k x_k$
- ✓ com  $G$  como função distância
- ✓ com  $d_k$  como peso amostral inicial do indivíduo  $k$

- » **Os pesos amostrais ajustados são obtidos pelo método “raking ratio”, que corresponde a aplicar a seguinte função distância:**

$$G(r) = r \log(r) - r + 1, \quad r = \frac{w_k}{d_k}.$$

- ✓ quando convergente, opta-se pelo o método “logit” (“raking ratio” com limites)



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL





## Método de calibração



- » Neste método, o total de uma dada variável de interesse  $Y$  é estimado pelo seguinte estimador:

$$\hat{t}_{y, CAL} = \sum_{h=1}^H \sum_{k \in s_h} w_k y_k$$

- ✓ Em que  $w_k$  é o peso amostral ajustado pelo método de calibração

- » A variância deste estimador é estimada pelo método de reamostragem *Jackknife*:

$$\hat{V}ar(\hat{t}_{y, CAL}) = \frac{n-1}{n} \sum_{i=1}^n \sum_{h=1}^H (\hat{t}_{y_{hi}, CAL} - \hat{t}_{y_h, CAL})^2$$

- ✓ Em que  $\hat{t}_{y_{hi}, CAL}$  corresponde à estimativa do total de  $Y$  no estrato  $h$  da subamostra  $i$



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



## Ajustamento às atuais dimensões



- » Neste método, o ajustamento dos pesos amostrais iniciais é feito por totais conhecidos da população, de acordo com o estrato atual, obtendo-se os seguintes pesos amostrais ajustados:

$$w_k = d_k \times \frac{N_g}{\hat{N}_g}$$

Em que:

- ✓  $k \in s_h$
- ✓  $g$  representa o estrato atual de acordo com a informação atualizada das empresas
- ✓  $N_g$  e  $\hat{N}_g$  são, respetivamente, as dimensões atual e inicial do estrato  $g$  no universo



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



## Ajustamento às atuais dimensões



- » Para este método, consideramos o estimador pós-estratificado para estimar o total de uma dada variável de interesse  $Y$ :

$$\hat{t}_{y, POST} = \sum_{g=1}^G \sum_{h=1}^H \sum_{i \in S_{gh}} \frac{N_h}{n_h} \frac{N_g}{\hat{N}_g} y_i$$

- » Para uma amostra aleatória estratificada, a variância do estimador pós-estratificado é dada por:

$$Var(\hat{t}_{y, POST}) = \sum_{g=1}^G \sum_{h=1}^H N_h \left( \frac{N_h - n_h}{n_h} \right) \left[ \frac{N_{hg}}{N_h} S_{y, hg}^2 + \frac{N_{hg}}{N_h} \left( 1 - \frac{N_{hg}}{N_h} \right) (\mu_{hg} - \mu_g)^2 \right]$$

- ✓ Em que  $N_{hg}$  corresponde ao número de empresas no universo que pertenciam ao estrato  $h$  e que agora pertencem ao estrato  $g$



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



## Resultados



- » Analisando os principais indicadores de utilização das TIC, apresentam-se algumas medidas descritivas para os valores dos coeficientes de variação das estimativas nas diferentes metodologias.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Sem informação auxiliar	0	0.020	1.70	3.861	5.570	25.45
Metodologia atual	0	0.000	1.39	3.567	4.815	22.60
Com calibração	0	0.015	1.70	3.849	5.575	25.56
Ajustamento às dimensões atuais	0	0.000	1.24	3.903	5.215	24.80



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

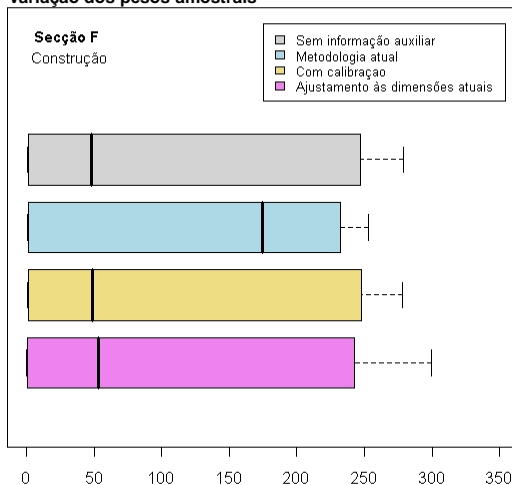




## Secção F Atividades da Construção



### Variação dos pesos amostrais



» A metodologia atual permite reduzir a dispersão dos pesos amostrais utilizados no cálculo das estimativas.



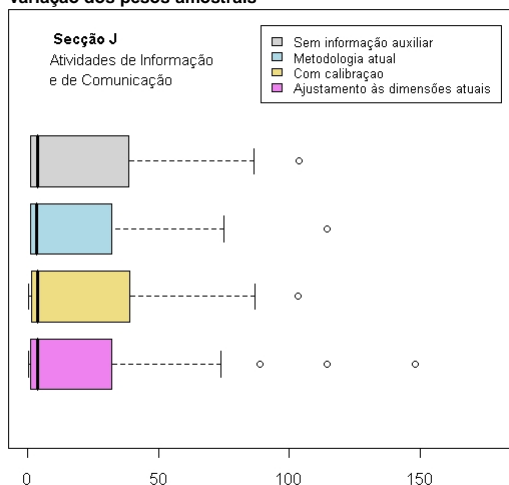
INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



## Secção J Atividades de Informação e de Comunicação



### Variação dos pesos amostrais



» O mesmo se pode concluir para os pesos amostrais das empresas do sector das TIC.



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL



## Principais resultados



- » A utilização de informação auxiliar apresenta ganhos de precisão face à não utilização da informação atualizada das empresas. No entanto, a metodologia atual é a que revela, em geral, melhores resultados.

Questão	Dimensão Empresa	Sem informação auxiliar			Metodologia atual			Calibração			Ajustamento às atuais dimensões		
		Estimativa	Desvio padrão	CV(%)	Estimativa	Desvio padrão	CV(%)	Estimativa	Desvio padrão	CV(%)	Estimativa	Desvio padrão	CV(%)
Utiliza computador?	Muito pequena	141904,62	3063,47	2,16	142266,22	3030,34	2,13	141930,10	3066,56	2,16	143204,14	3069,91	2,14
	Pequena	37511,79	515,95	1,38	38037,91	393,74	1,04	37430,24	514,29	1,37	35410,20	438,08	1,24
	Média	5670,70	0,62	0,01	5683,00	0,00	0,00	5681,47	0,58	0,01	5564,81	0,00	0,00
	Grande	810,95	0,24	0,03	794,00	0,00	0,00	793,33	0,15	0,02	771,04	0,00	0,00
Acede à Internet?	Muito pequena	120311,51	3079,56	2,56	120601,52	3054,26	2,53	120321,63	3082,21	2,56	121392,23	3053,10	2,52
	Pequena	36147,62	615,11	1,70	36664,55	510,14	1,39	36070,02	613,67	1,70	34065,87	536,58	1,58
	Média	5638,33	31,88	0,57	5659,23	23,27	0,41	5648,78	32,18	0,57	5533,75	32,85	0,59
	Grande	810,95	0,24	0,03	794,00	0,00	0,00	793,33	0,15	0,02	771,04	0,00	0,00
Tem Website?	Muito pequena	40269,20	2174,51	5,40	39400,91	2143,94	5,44	40255,54	2176,03	5,41	40450,27	2113,02	5,22
	Pequena	19033,09	1089,07	5,72	18879,75	1002,88	5,31	18970,20	1087,15	5,73	17832,69	1082,33	6,07
	Média	3983,00	124,73	3,13	4273,20	114,68	2,68	3978,29	126,93	3,19	3899,46	127,54	3,27
	Grande	756,90	1,73	0,23	746,17	0,90	0,12	746,32	1,61	0,22	715,02	1,40	0,20



## Conclusões



- » No contexto dos inquéritos às empresas, principalmente os de periodicidade anual, a informação atualizada via fontes administrativas pode contribuir para melhorar a precisão das estimativas.
- » Os métodos de ajustamento têm a vantagem de fazer corresponder as estimativas dos totais aos conhecidos da população, apresentando ganhos pontuais de precisão.
- » Ainda assim, a metodologia atual, por utilizar informação auxiliar individualizada (empresa a empresa), corrigindo o estrato de cada unidade estatística, permite obter estimativas mais precisas.
- » Em geral, os métodos de ajustamento dos pesos amostrais apresentam vantagens face à não utilização de informação auxiliar. Nesta aplicação essa vantagem não é tão evidente pelo facto de utilizarmos na correção as mesmas variáveis que servem de base à estratificação.





## Considerações finais



- » Os métodos de ajustamento estudados não revelaram ganhos muito significativos porque a informação atualizada das empresas baseou-se nas variáveis de estratificação.
- » Atualmente, no âmbito das empresas, existem algumas variáveis auxiliares disponibilizadas pela IES que apresentam ainda algum défice de cobertura.
- » Quando esta informação administrativa melhorar a cobertura, do universo das empresas, será interessante testar o ajustamento dos pesos amostrais iniciais por totais de outras variáveis disponíveis, como por exemplo o valor das transações comerciais.



## Referências



- ❑ COCHRAN, W. G. (1977) *Sampling techniques*. 3rd Edition, New York, Springer.
- ❑ DEVILLE, J.-C. (1999) *Variance estimation for complex statistics and estimators: linearization and residual techniques*, *Survey Methodology*, 25(2).
- ❑ DEVILLE, J.-C., SÄRNDAL, C.-E (1992) Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418), 376-382.
- ❑ LUMLEY, T. (2008). Survey analysis in R, [www.faculty.washington.edu/tlumley/survey/](http://www.faculty.washington.edu/tlumley/survey/), (acedido em 18 de Julho de 2011).
- ❑ LUMLEY, T. (2010) *Complex surveys: a guide to analysis using R*, Wiley.
- ❑ LUMLEY, T. (2011) Survey: analysis of complex survey samples. *R package version 3.26-1*.
- ❑ SÄRNDAL, C.-E., SWENSSON, B., WRETMAN, J.,H. (1992) *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- ❑ SÄRNDAL, C.-E. (2007) The calibration approach in survey theory and practice, *Survey Methodology*, 33(2), 99-119.
- ❑ SHAO, J., TU, D. (1995) *The jackknife and bootstrap*. New York: Springer-Verlag.
- ❑ SMILHILY, M., STORM, H. (2010) ICT usage in enterprises1 - 2009, *Eurostat Publications*, Issue 1.
- ❑ WOLTER, K.M. (1985) *Introduction to variance estimation*, New York: Springer-Verlag.





INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

» Utilização de informação auxiliar no processo de  
estimação

Uma aplicação no contexto das empresas «

Obrigada

Rita Sousa, Conceição Ferreira  
DMSI / ME

 30/03/2012



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

