



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

»» Confidencialidade: problemática e tendências metodológicas

Pedro Campos ««

INE/ Departamento de Metodologia e Sistemas de Informação



SEMINÁRIO
TRATAMENTO DO SEGREDO ESTATÍSTICO

INE, 12-10-2012



Confidencialidade: problemática e tendências metodológicas



- Introdução e problemática
- Metodologias
- Tendências metodológicas





Confidencialidade: problemática e tendências metodológicas



- Introdução e problemática
- Metodologias
- Tendências metodológicas





Conceitos e princípios fundamentais

- **Segredo estatístico/confidencialidade**
Visa salvaguardar a privacidade dos cidadãos e garantir a confiança no Sistema Estatístico Nacional
- **Autoridade estatística**
Obrigatoriedade e gratuidade de fornecimento de informação às autoridades estatísticas para a produção de estatísticas oficiais
- **Microdados e Macrodados**





Conceitos e princípios fundamentais (cont.)

- **Qualidade vs segredo**

As estatísticas oficiais devem respeitar os padrões nacionais e internacionais de qualidade estatística

- **Métodos de Controlo de divulgação estatística**

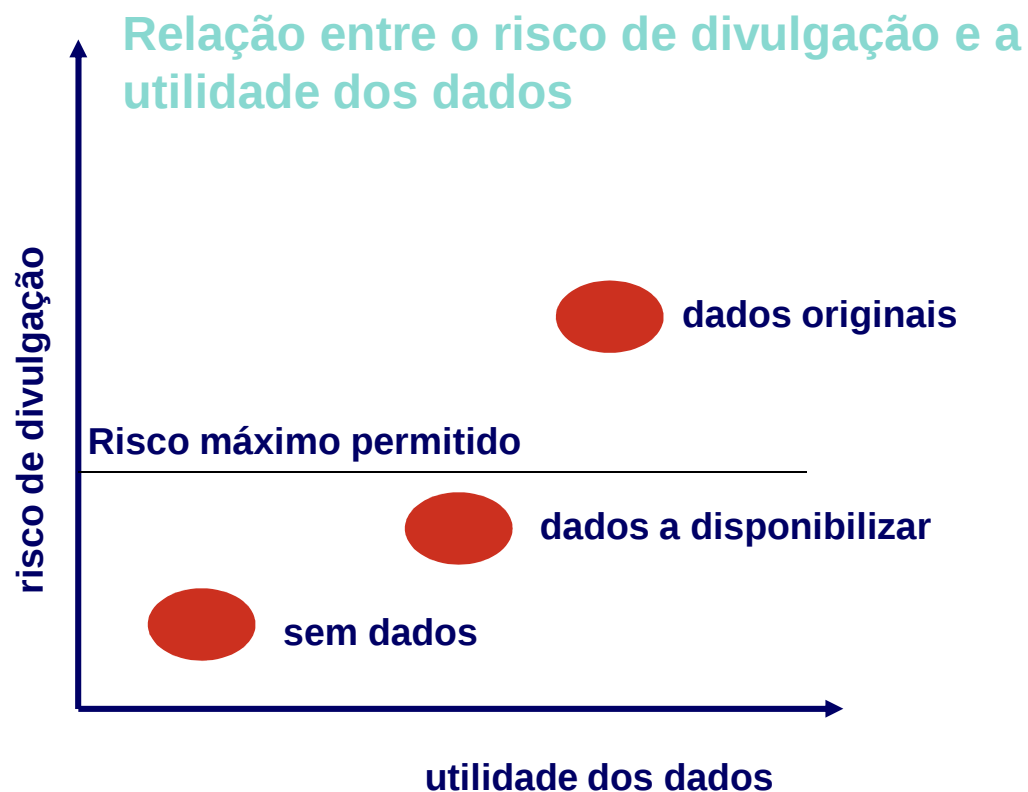
(statistical disclosure control)

- têm como objetivo a difusão dos dados agregados ou a preparação dos microdados,
- As técnicas utilizadas de acordo com os critérios pré-definidos, procuram minimizar o risco de identificação das unidades estatísticas a que se referem ao mesmo tempo que preservam algumas das propriedades estatísticas dos microdados (por exemplo, totais), passando o responsável pela difusão dos dados a ter um papel de gestor de risco.



Problemática

Risco e perda de informação



- A perda de informação aumenta à medida que diminui o risco de identificação;
- A utilidade dos dados é tanto menor quanto maior for a perda de informação;
- O controlo de divulgação para proteção individual deve ser feito de modo a encontrar-se um compromisso entre essas duas quantidades.



O tratamento de dados pessoais deve processar-se de forma transparente e no estrito respeito pela reserva da vida privada, bem como pelos direitos, liberdades e garantias fundamentais.

Princípio geral da Lei da Proteção de Dados Pessoais, (Lei nº 67/98, Artigo 2º)





- Problemas de confidencialidade podem existir em diversas áreas:
 - Sigilo Bancário
 - Segredo Estatístico
 - Segredo Profissional
 - Saúde
 - Videovigilância
 - ...



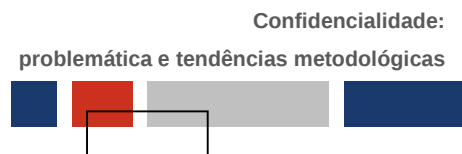


Confidencialidade: problemática e tendências metodológicas



- Introdução e problemática
- **Metodologias**
- Tendências metodológicas





Metodologias de Protecção da Confidencialidade

Podem dividir-se em:

- ✓ Métodos para **macrodados** (dados agregados em tabelas)
- ✓ Métodos para **microdados** (dados individualizados)



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Macrodados

Dados agregados de respondentes individuais (indivíduos, entidades económicas, etc.) representados sob a forma de tabelas estatísticas.



Frequência



Magnitude





Macrodados

Exemplo: tabela de frequências

Local de residência (NUTS - 2002)		Casos notificados de doenças de declaração obrigatória (N.º) por Local de residência (NUTS - 2002) e Doenças de declaração obrigatória; Anual (2)							
		Período de referência dos dados (1)							
		2010							
		Doenças de declaração obrigatória							
		Total	Tuberculose respiratória	Outras salmoneloses	Febre escarotodular	Parotidite epidémica	Hepatite por vírus B	Outros casos	
		N.º	N.º	N.º	N.º	N.º	N.º	N.º	
Portugal	PT	3132	1736	206	140	140	47	863	
Continente	1	3055	1695	206	140	139	47	828	
Região Autónoma dos Açores	2	45	...	0	0	...	0	26	
Região Autónoma da Madeira	3	32	...	0	0	...	0	9	

Fonte: INE





Macrodados

Exemplo: tabela de magnitude

Localização geográfica (NUTS - 2002)	Actividade económica (CAE Rev. 3)	Volume de negócios (Série 2007-2009 - €) das empresas por Localização geográfica (NUTS - 2002) e Actividade económica (CAE Rev. 3); Anual	
		Período de referência dos dados	
		2009	
		€	
Portugal	P Total	335887311762	
	T Pesca e aquicultura	369639375	
	Indústrias extractivas	1132609949	
	Indústrias transformadoras	70629520911	
	Electricidade, gás, vapor, água quente e fria e ar frio	16374371094	
	Captação, tratamento e distribuição de água; saneamento, gestão de resíduos e despoluição	2492033107	
	Construção	34581351963	
	Comércio por grosso e a retalho; reparação de veículos automóveis e motociclos	127481379662	
	Transportes e armazenagem	16567937606	
	Alojamento, restauração e similares	9542589961	
	Actividades de informação e de comunicação	13668455221	
	Actividades imobiliárias	6067777135	
	Actividades de consultoria, científicas, técnicas e similares	11756903662	
	Actividades administrativas e dos serviços de apoio	9976855449	
	Educação	1501542147	
	Actividades de saúde humana e apoio social	10288230257	
	Actividades artísticas, de espectáculos, desportivas e recreativas	1686247431	
	Outras actividades de serviços	1769866832	

Fonte: INE



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

TRATAMENTO DO SEGREDO ESTATÍSTICO

INE, 12-10-2012

Algumas metodologias de controlo de divulgação estatística para **macrodados**:

- Regra da frequência mínima
- Regra (n.k)
- Regra p%
- Arredondamento
- ...



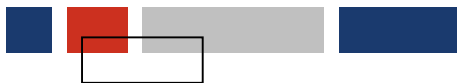
Registos que contêm informação de **unidades estatísticas**



A unidade estatística é a unidade básica de observação (*o inquirido*) num inquérito estatístico;

Constitui o objeto de análise ao qual estão associadas características estatísticas.

- 
- Família
 - Indivíduo
 - Empresa
 - Instituição
 - ...



Microdados

Algumas metodologias de controlo de divulgação estatística para **microdados**:

- Eliminação de variáveis
- Recodificação
- Supressão local
- Microagregações numéricas
- ...



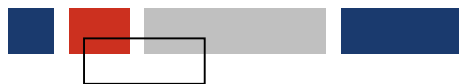


Confidencialidade: problemática e tendências metodológicas



- Introdução e problemática
- Metodologias
- **Tendências metodológicas**





Tendências metodológicas

Macrodados

Utilização preferencial da regra de sensibilidade $p\%$ em detrimento da regra (n,k) , pois tem em conta todos os inquiridos de uma célula e não apenas os n maiores.

Questão: dificuldade na transição definição de um valor ótimo para óptimo.





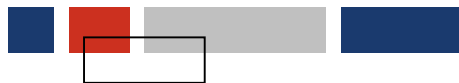
Tendências metodológicas

Macrodados vs microdados

Maior aproximação entre as abordagens micro e macro:
o tratamento dos dados deve ser feito ao nível da
unidade estatística de análise e não da célula

Início de uma nova perspetiva de análise dos métodos
de controlo de divulgação estatística





Tendências metodológicas

Microdados

Ficheiros PUF

Importância do trabalho com este tipo de ficheiros de microdados públicos **ao nível do SEE**. Em alguns países, esse trabalho já está em curso e noutros os pedidos de PUFs tendem a crescer.

Duas abordagens:

- Prevalência dada à estrutura do ficheiro com requisitos mínimos para o conteúdo da informação. Neste sentido devem-se treinar os utilizadores potenciais no uso de PUFs;
- Prevalência dada ao conteúdo de informação, no sentido de estimular o uso científico de PUF (como é um caso dos EUA).

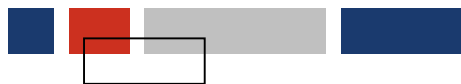




Tendências metodológicas

- **Difusão de dados espaciais através de mapas (O'Keefe, 2012),**
- Nas áreas da epidemiologia e investigação na saúde, são utilizados mapas para representar fenómenos.
- Nos mapas, os casos notificados de doenças são representados através de pontos, linhas e áreas
- Esses objetos estão sujeitos a tratamento da confidencialidade
- Existem técnicas que combinam pontos, linhas e áreas em métodos não perturbativos para tratamento da confidencialidade





Carta da Confidencialidade do SEE

Linhas gerais

- Maximização da publicação de totais;
- Adoção de standards e princípios destinando-se a evitar “ruídos” na validação dos dados nos quadros finais;
- Incentivo ao envio de informação de confidencial dos EM para o Eurostat;
- Assegurar aos EM que, através da adoção desta carta, os seus dados estarão seguros.



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

TRATAMENTO DO SEGREDO ESTATÍSTICO
INE, 12-10-2012

SEMINÁRIO



Carta da Confidencialidade do SEE

Framework Partnership Agreement (FPA)

- Será lançado um Framework Partnership Agreement (FPA) para harmonização do SDC.
- Será feita uma seleção de parceiros com base numa chamada limitada aos membros do SEE e poderá preparar-se uma ESSnet focalizada em resolução de problemas relacionada com harmonização, standardização e implementação de acções.





Tendências metodológicas

Harmonização no acesso a microdados

- **Baixo risco**, em que os dados são agregados ao nível de NUTS I, por exemplo, ou nos casos em que os dados contêm poucas variáveis e não pode ser ligado a outros registos individuais;
- **Baixo impacto**, nos casos em que os dados não são sensíveis. Estas são as situações comuns nos inquéritos que tratam matérias que não causam embaraço aos indivíduos ou empresas;
- **Alto risco**, correspondendo a dados de um nível geográfico (ou classificação económica) muito desagregados, o que faz com que as variáveis apresentem informação muito detalhada;
- **Alto impacto**: neste caso, os valores obtidos para cada indivíduo ou empresa são muito sensíveis, envolvendo doenças ou grande quantidade de opiniões pessoais em tópicos sensíveis.





Tendências metodológicas

Acesso remoto

- O acesso remoto a microdados é um recurso on-line seguro, em que os investigadores se ligam a um servidor através de uma palavra-chave ou outros dispositivos seguros, onde os dados e os programas estão localizados;
- Alguns institutos de estatística possuem sistemas de acesso remoto num RDC (Research Data Center), que permite o acesso a investigadores autenticados e credenciados (CBS; Instituto Sueco de estatística - **MONA II**; Luxembourg Income Study - **LISSY**)

