

# **Soluções de divulgação externa de informação estatística e preservação do segredo estatístico**

**Filipa Duarte de Carvalho**

**Instituto Superior de Economia e Gestão - UTL**

**Centro de Investigação Operacional - FCUL**

**Seminário sobre o princípio do segredo estatístico - 13 de Janeiro de 2005**

# Sumário da apresentação

1. Tipos de dados e técnicas de protecção
2. Problema de protecção de dados confidenciais
3. Métodos de resolução do problema
4. Resultados práticos
5. Trabalho em curso

# Microdados:

informação individual sobre pessoas, empresas etc.

área de residencia = **X**

idade > **85**

passado criminal = **tem cadastro criminal**

# Protecção de microdados

- Modificação dos dados
- Supressão local
- Agregação → **Tabelas**

# Tabelas

	Reg. A	Reg. B	Reg. C	Total
Act. I	11	47	58	116
Act. II	1	<b>X</b>	33	49 <b>X = 15</b>
Act. III	2	31	20	53
Total	14	93	111	218

Willenborg, L. & de Waal, T. (1996)

# Tabelas

	Reg. A	Reg. B	Reg. C	Total
Act. I	11	<del>47</del>	<del>58</del>	116
Act. II	1	<del>X</del>	<del>33</del>	49 $X \mu [0,48]$
Act. III	2	31	20	53
Total	14	93	111	218

**custo = valor da informação suprimida = 138**

# Tabelas

	Reg. A	Reg. B	Reg. C	Total	
Act. I	11	47	58	116	
Act. II	<del>X</del>	<del>X</del>	33	49	$X \mu [13,16]$
Act. III	<del>X</del>	<del>X</del>	20	53	
Total	14	93	111	218	

**custo = valor da informação suprimida = 34**

# Tabelas

	Reg. A	Reg. B	Reg. C	Total
Act. I	11	<del>47</del>	<del>58</del>	116
Act. II	1	<del>X</del>	<del>33</del>	49
Act. III	2	31	20	53
Total	14	93	111	218

$x \mu [0,48]$

informação suprimida = 138

	Reg. A	Reg. B	Reg. C	Total
Act. I	11	47	58	116
Act. II	<del>X</del>	<del>X</del>	33	49
Act. III	<del>X</del>	<del>33</del>	20	53
Total	14	93	111	218

$x \mu [13,16]$

informação suprimida = 34



# Problema da Supressão

Encontrar um conjunto de dados não confidenciais, de valor tão pequeno quanto possível, cuja supressão garanta a protecção de todos os dados confidenciais (também suprimidos).

**O problema da supressão é NP-difícil !**

# **Variantes do problema da supressão**

- **Protecção exacta**
- **Protecção intervalar**

# Soluções para a protecção exacta

Heurísticas:

**Tabelas protegidas com pequena perda de informação**

Abordagem Lagrangeana:

**Avaliação da qualidade das tabelas protegidas**

Formulação do problema:

**Tabelas protegidas com a perda mínima de informação**

# Soluções para a protecção intervalar

Heurísticas:

**Tabelas protegidas com pequena perda de informação**

Formulação parcial do problema:

**Avaliação da qualidade das tabelas protegidas**

Formulação do problema:

**Tabelas protegidas com a perda mínima de informação**

# Resultados práticos

**380 tabelas estatísticas**

**20 linhas × 10 colunas**

•  
•  
•

**200 linhas × 200 colunas**

# Resultados práticos (380 tabelas)

**Protecção exacta:** Protecção óptima em todos os casos

**Protecção intervalar:**

Protecção óptima para 300 tabelas

Protecção para 80 tabelas com perda de informação que não excede 4% da mínima

# Trabalho em curso

## Método exacto baseado em cortes inteiros

(trabalho conjunto: Almeida, M.T., Carvalho, F.D.)

## Algoritmos genéticos

(trabalho conjunto: Almeida, M.T., Schutz, G., Carvalho, F.D.)

# Teses

**Carvalho, F.D.**, O problema da supressão na protecção de informação confidencial: formalizações e algoritmos.

*Tese de Doutoramento*, Universidade Técnica de Lisboa (2002).

**Carvalho, F.D.**, Optimização em redes na protecção de informação em tabelas estatísticas bidimensionais.

*Tese de Mestrado*, Universidade Técnica de Lisboa (1995).

**Carvalho, F.D.**, Statistical disclosure in two dimensional tables.

*Master Thesis*, Erasmus University, Rotterdam (1992).



# Trabalhos publicados

Almeida, M.T., Carvalho, F.D. (2004), [Exact disclosure prevention in two-dimensional statistical tables](#). Aceite para publicação em *Computers and Operations Research*.

Carvalho, F.D., Almeida, M.T. (2000), [Lower-bounding procedures for the 2-dimensional cell suppression problem](#). *European Journal of Operational Research* 123 pp 29-41.

Carvalho, F.D., Almeida, M.T. (1998), [Heuristic Methods for the Cell Suppression Problem in General Statistical Tables](#). *Proceedings of the Conference Statistical Data Protection '98*. Lisbon, Portugal.

Carvalho F.D., Dellaert N., Osório, M. (1994), [Statistical Disclosure in Two-Dimensional Tables: general tables](#). *Journal of the American Statistical Association* 89, pp 1547-1557.

# Monografias

- L. Willenborg and T. De Waal. *Elements of statistical disclosure control*. Springer-Verlag, 2001.
- L. Willenborg and T. De Waal. *Statistical disclosure control in practice*. Springer-Verlag, 1996.
- P. Doyle et al. *Confidentiality, disclosure, and data access. Theory and practical applications for statistical agencies*. North-Holland, 2001