

# Parte VI:

## Medidas de Interesse

## Exceções

# Modelo Suporte/Confiança

- **Modelo Suporte/Confiança** (*Agrawal et al, 1993*):
  - Determinar todas as regras que tenham suporte e confiança maiores ou iguais aos valores mínimos especificados.
- **Problemas:**
  - O número de regras gerado costuma ser extremamente volumoso. Identificar as regras realmente úteis e interessantes torna-se uma tarefa difícil.
  - O modelo gera regras redundantes, ilusórias ou até mesmo contraditórias.

# Modelo Suporte/Confiança

- **Entrada:**
  - Base de Dados: POF-FGV (registro das compras mensais de 422 famílias cariocas em Junho de 1998)
  - SupMin = 3%
  - ConfMin = 60%
- **Resultados:**
  - Foram mineradas **8.469** regras de associação envolvendo 2 itens!

# Modelo Suporte/Confiança

- **Grupo I:** antecedente e conseqüente muito populares.

R: Cenoura  $\Rightarrow$  Batata Inglesa      s: 70,38%      c: 91,38%

Sup(Cenoura) = 77,01%

Sup(Batata Inglesa) = 81,75%

- **Grupo II:** antecedente pouco freqüente e conseqüente muito freqüente (**obs:** 80% das regras mineradas).

R: Filé de Viola  $\Rightarrow$  Açúcar Refinado s: 7,58%      c: 86,49%

Sup(Filé de Viola) = 8,77%

Sup(Açúcar Refinado) = 86,49%

- **Grupo III:** antecedente e conseqüente pouco freqüentes.

R: Strogonoff de Frango (caixa)  $\Rightarrow$  Lasanha (caixa)      s: 3,32% c: 77,78%

Sup(Strogonoff) = 4,27%

Sup(Lasanha) = 14,45 %

# Dependência entre Itens

| <i>Id</i> | <i>Regra de Associação</i>                  | <i>Sup<sub>A</sub></i> | <i>Sup<sub>B</sub></i> | <i>Sup</i> | <i>Conf</i> |
|-----------|---|------------------------|------------------------|------------|-------------|
| R1        | Filé de Viola $\Rightarrow$ Açúcar Refinado | 8,77%                  | 86,49%                 | 7,58%      | 86,49%      |

- A confiança da regra indica que 86,49% dos clientes que compram filé de viola, também compram açúcar refinado.
- A probabilidade de qualquer cliente comprar açúcar refinado é de 86.49%.
- Os dois produtos são **independentes**.

$$\text{Sup}(B) = \text{Conf}(A \Rightarrow B)$$

# Dependência entre Itens

| <i>Id</i> | <i>Regra de Associação</i>               | <i>Sup<sub>A</sub></i> | <i>Sup<sub>B</sub></i> | <i>Sup</i> | <i>Conf</i> |
|-----------|--|------------------------|------------------------|------------|-------------|
| R2        | Banana Nanica $\Rightarrow$ Banana Prata | 12,09%                 | 76,07%                 | 7,35%      | 60,78%      |

- A confiança da regra indica que 60,78% dos clientes que compram banana nanica, também compram banana prata.
- A probabilidade de qualquer cliente comprar banana prata é de 76.07%. Portanto clientes que compram banana nanica têm menor chance de comprar banana prata.
- Os dois produtos possuem **dependência negativa**

$$\text{Sup}(B) > \text{Conf}(A \Rightarrow B)$$

# Dependência entre Itens

| <i>Id</i> | <i>Regra de Associação</i>                                    | <i>Sup<sub>A</sub></i> | <i>Sup<sub>B</sub></i> | <i>Sup</i> | <i>Conf</i> |
|-----------|---|------------------------|------------------------|------------|-------------|
| R3        | Milho Verde em Conserva $\Rightarrow$<br>Ervilhas em Conserva | 32,94%                 | 37,91%                 | 27,01%     | 82,01%      |

- A confiança da regra indica que 82,01% dos clientes que compram milho verde, também compram ervilhas.
- A probabilidade de qualquer cliente comprar ervilhas é de 37.91%. Portanto clientes que compram milho verde têm uma chance muito maior de comprar ervilhas.
- Os dois produtos possuem **dependência positiva**.

$$\text{Sup}(B) < \text{Conf}(A \Rightarrow B)$$

# Medidas Objetivas

- **Lift( $A \Rightarrow B$ )** : indica o quanto mais freqüente torna-se B quando A ocorre.

$$\text{Lift}(A \Rightarrow B) = \text{Conf}(A \Rightarrow B) \div \text{Sup}(B)$$

**Ex 1:** R: Filé de Viola  $\Rightarrow$  Açúcar Refinado

$$\text{Lift}(R) = 0,8649 \div 0,8649 = \mathbf{1}$$

**Ex 2:** R: Banana Nanica  $\Rightarrow$  Banana Prata

$$\text{Lift}(R) = 0,6078 \div 0,7607 = \mathbf{0,80}$$

**Ex 3:** R: Milho Verde em Conserva  $\Rightarrow$  Ervilhas em Conserva

$$\text{Lift}(R) = 0,8201 \div 0,3791 = \mathbf{2,21}$$



# Exceções

- **Proposta:**
  - Mineração de exceções negativas e positivas.
  - Usuários podem explorar um conjunto de regras de associação, descobrindo se elas tornam-se significativamente **mais fortes** ou **mais fracas** em algum subconjunto da base de dados.

# Exceções – Base da AIDS

- **Fonte:** Ministério da Saúde
- Casos de AIDS diagnosticados no Brasil entre 1980 e 2001.
- 172.563 tuplas e 10 atributos. (pacientes com idade superior a 13 anos).
- **Entrada:**
  - $R: (Transmissão\ Sexual = "Não") \Rightarrow (Drogas = "Sim")$   
Sup = 13,05%, Conf = 88,09%
  - $A = \{Sexo, Idade, Região\}$
- **Saída:**
  - 41 exceções negativas e 5 exceções positivas.

# Exceções – Base da AIDS

- **Exceção Negativa:**

*(Transmissão Sexual = "Não")  $\Rightarrow^s$  (Drogas = "Sim") [ (Sexo = "F"),  
(Região = "Norte) ]*

# Exceções – Base da AIDS

| Microsoft Access - [F_NORTE_DROGAS : Consulta seleção]                    |        |    |                  |                    |            |        |  |
|---|--------|----|------------------|--------------------|------------|--------|--|
| Arquivo Editar Exibir Inserir Formatar Registros Ferramentas Janela Ajuda |        |    |                  |                    |            |        |  |
| SEXO  | REGIAO | UF | IDADE            | TRANSMISSAO_SEXUAL | TRANSFUSAO | DROGAS |  |
| F   | NORTE  | PA | 35-39 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | RR | 25-29 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | PA | ACIMA DE 50 ANOS | NAO                | SIM        | NAO    |  |
| F   | NORTE  | PA | 40-49 ANOS       | NAO                | NAO        | SIM    |  |
| F   | NORTE  | PA | 40-49 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | PA | 35-39 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | RO | 20-24 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | RO | 30-34 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | AM | ACIMA DE 50 ANOS | NAO                | SIM        | NAO    |  |
| F   | NORTE  | AP | 25-29 ANOS       | NAO                | NAO        | SIM    |  |
| F   | NORTE  | RO | 25-29 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | RO | 20-24 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | PA | 25-29 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | PA | 40-49 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | TO | 35-39 ANOS       | NAO                | NAO        | SIM    |  |
| F   | NORTE  | PA | 35-39 ANOS       | NAO                | SIM        | NAO    |  |
| F   | NORTE  | PA | 30-34 ANOS       | NAO                | NAO        | SIM    |  |
| F   | NORTE  | PA | 20-24 ANOS       | NAO                | NAO        | SIM    |  |

# Exceções – Base do Censo

- **Fonte:** UCI Repository of Machine Learning Databases
- Censo da cidade de Washington, ano de 1990
- 48.842 tuplas e 14 atributos.
- **Entrada:**
  - $R: (IncomeClass = ">50K")$   
Sup = 24,08%
  - $A = \{Education, Gender, HoursPerWeek, Age\}.$
- **Saída:**
  - 20 exceções positivas.

# Exceções – Base do Censo

- **Exceções Positivas:**

$\Rightarrow^{+s}(IncomeClass = ">50K") [(Education = "Doctorate")]$   
 $MF^+ = 0,6702$

$\Rightarrow^{+s}(IncomeClass = ">50K") [(Education = "Masters")]$   
 $MF^+ = 0,5682$

$\Rightarrow^{+s}(IncomeClass = ">50K") [(HoursPerWeek = "56-60")]$   
 $MF^+ = 0,4474$

$\Rightarrow^{+s}(IncomeClass = ">50K") [(HoursPerWeek = "51-55")]$   
 $MF^+ = 0,4428$

$\Rightarrow^{+s}(IncomeClass = ">50K") [(Age = "46-50")]$   
 $MF^+ = 0,3961$

$\Rightarrow^{+s}(IncomeClass = ">50K") [(Age = "51-55")]$   
 $MF^+ = 0,3934$

# Exceções – Regras Híbridas

- **Base de Dados da POF (1999)**

- **Regras Híbridas Positivas:**

$\{cerveja\ em\ lata\} \Rightarrow^{+s+c} \{salame\ e\ salaminho\} [(Membros=1)]$

$\{neston\} \Rightarrow^{+s+c} \{farinha\ láctea\} [(Membros=5)]$

$\{leite\ longa\ vida\} \Rightarrow^{+s+c} \{pão\ francês\} [(Cidade="Porto\ Alegre")]$

- **Regras Híbridas Negativas:**

$\{neston\} \Rightarrow^{-s-c} \{farinha\ láctea\} [(Membros=2)]$

$\{leite\ longa\ vida\} \Rightarrow^{-s-c} \{pão\ francês\} [(Renda="2,5\ a\ 4,99\ Sal.\ Min.")]$

$\{leite\ longa\ vida\} \Rightarrow^{-s-c} \{pão\ francês\} [(Cidade="Recife")]$