



Medidas de Associação e Correlação

Estatística Aplicada

Ana Maria Nogales Vasconcelos

Maria Teresa Leão Costa



Medidas de Associação para Variáveis Categorizadas



Variáveis Categorizados

◆ Uma **variável categorizada** é aquela na qual a escala de medida consiste em um conjunto de categorias.

Exemplos:

- Opinião sobre horário eleitoral – *a favor ou contra*
- Grau de instrução: *analfabeto, fundamental, médio, superior*
- Diagnóstico de uma doença - *presente ou ausente*
- Faixa de Renda - *0 a 2 S.M, 2 a 4 S.M, > 4 S.M*

Variáveis categorizadas:

- variáveis qualitativas
- variáveis quantitativas organizadas em categorias (classes)



Tabelas de Contingência

- ◆ Mostra a frequência conjunta de duas variáveis categorizadas



Tabelas de Contingência

- ◆ Mostra a frequência conjunta de duas variáveis categorizadas



Tabelas de Contingência

- ◆ Mostra a frequência conjunta de duas variáveis categorizadas

Exemplo

Tipo de Conta: C C P P C C P P C P
Sexo: M F F M M M F M M F
(C=conta corrente, P=Poupança; M=Masc., F=Fem.)



Tabelas de Contingência

- ◆ Mostra a frequência conjunta de duas variáveis categorizadas

Exemplo

Tipo de Conta: C C P P C C P P C P
Sexo: M F F M M M F M M F
(C=conta corrente, P=Poupança; M=Masc., F=Fem.)

Tipo de Conta	Sexo		Total
	Masc.	Fem.	
Corrente	4	1	5
Poupança	2	3	5
Total	6	4	10



Notação

f_{ij} → frequência observada da linha i
coluna j

$f_{i\cdot}$ → frequência total da linha i

$f_{\cdot j}$ → frequência total da coluna j



Tabelas de Contingência

- ◆ Mostra a frequência conjunta de duas variáveis categorizadas

Exemplo

Tipo de Conta: C C P P C C P P C P
Sexo: M F F M M M F M M F
(C=conta corrente, P=Poupança; M=Masc., F=Fem.)

Tipo de Conta	Sexo		Total
	Masc.	Fem.	
Corrente	4	1	5
Poupança	2	3	5
Total	6	4	10

Diagrama de contagem da Tabela de Contingência:

- f_{ij} é o valor na célula i,j , apontado por uma seta de cima para baixo.
- $f_{.j}$ é a soma das frequências da coluna j , apontada por uma seta de esquerda para direita.
- $f_{i.}$ é a soma das frequências da linha i , apontada por uma seta de cima para baixo.



Tabelas de Contingência

- ◆ Para analisar tabelas de contingência podemos usar as seguintes freqüências relativas, expressas em percentagem:
 - *% da linha*
 - *% da coluna*
 - *% do total*
- ◆ Ajuda a estudar relações entre variáveis



Tabelas de Contingência (% da Linha)

Tipo De Conta	Sexo		Total
	Masc.	Fem.	
Corrente	4	1	5
Poupança	(80)	(20)	(100)
Total	2	3	5
	(40)	(60)	(100)
	6	4	10
	(60)	(40)	(100)

(Freq. da célula) (100)
Freq. Total da Linha

$$(3/5)(100) = 60\%$$



Tabelas de Contingência (% da Coluna)

Tipo De Conta	Sexo		Total
	Masc.	Fem.	
Corrente	4	1	5
Poupança	(67)	(25)	(50)
Total	2	3	5
	(33)	(75)	(50)
	6	4	10
	(100)	(100)	(100)

(Freq. da célula) (100)
Freq. Total da Coluna

$$(3/4)(100) = 75\%$$

Tabelas de Contingência (% do Total)

Tipo De Conta	Sexo		Total
	Masc.	Fem.	
Corrente	4	1	5
Poupança	(40)	(10)	(50)
Total	2	3	5
	(20)	(30)	(50)
	6	4	10
	(60)	(40)	(100)

(Freq. da célula) (100)
Freq. Total

$$(3/10)(100) = 30\%$$



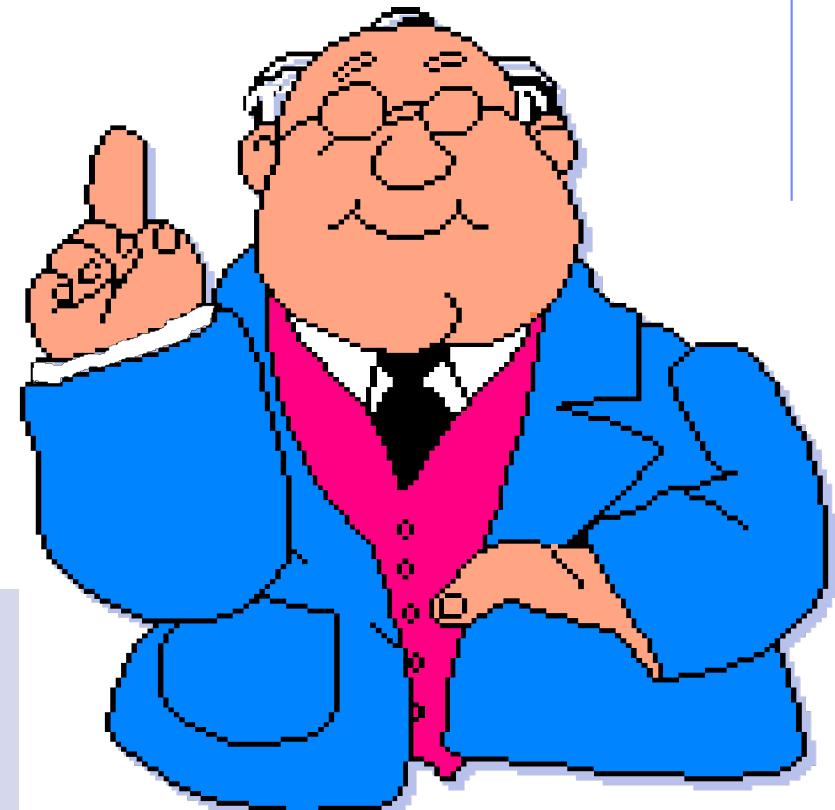
Qual Porcentagem usar?

- ◆ Calcular a % na direção da ***variável explicativa***

- ◆ Se a variável explicativa está linha, usar total da linha

Exemplo:

- ◆ sexo é a variável explicativa
 - ‘Explica’ tipo de conta
 - *% da linha*





Para Pensar...

Em uma pesquisa de mercado para Visa, deseja-se analisar os dados sobre uso de cartão de crédito e renda anual.



Renda (x1000):	12	20	32	45	72	46	18	55	...
Uso:	S	N	N	S	S	S	N	S	...

(Categorias de renda: $< \$25\,000$; $\geq \$25\,000$;
Categorias de uso de cartão de crédito: S=sim, N=não)



Representação Tabular

Renda	Uso		Total
	Não	Sim	
< 25	200	100	300
≥ 25	100	400	500
Total	300	500	800



Representação Tabular

Variável Explanatória

Renda

		Uso		Total
		Não	Sim	
< 25	< 25	200	100	300
	≥ 25	100	400	500
Total		300	500	800



Representação Tabular

Variável Explanatória

Renda

		Uso		Total
		Não	Sim	
< 25	Total	200	100	300
	≥ 25	100	400	500
Total		300	500	800

Percentagem da linha

$(400/500)(100) = 80\%$



Representação Tabular

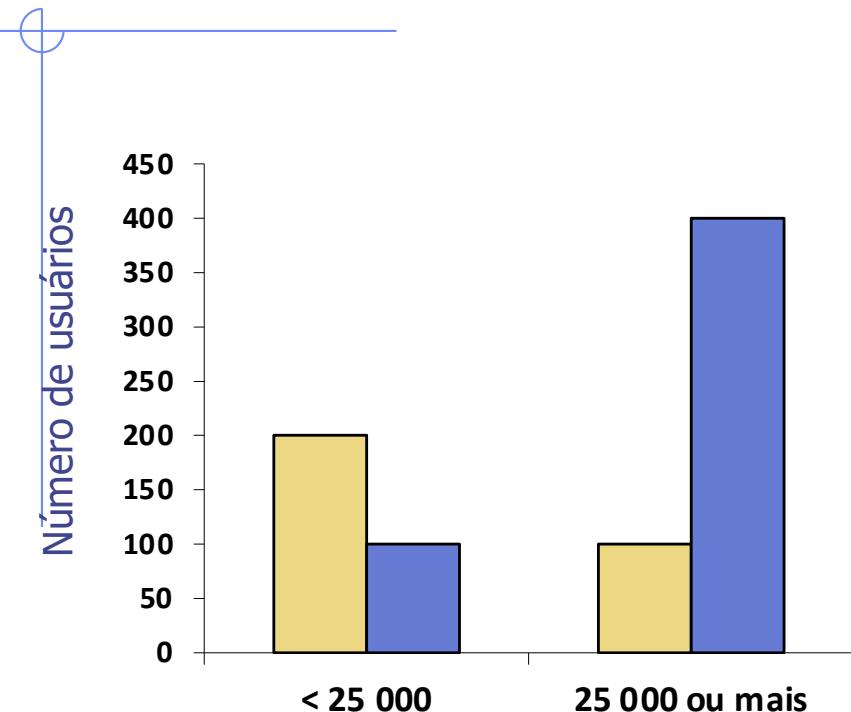
Variável Explanatória	Renda	Uso		Total
		Não	Sim	
< 25	< 25	200 (67)	100 (33)	300 (100)
	≥ 25	100 (20)	400 (80)	500 (100)
Total		300 (37,5)	500 (62,5)	800 (100)

Percentagem da linha

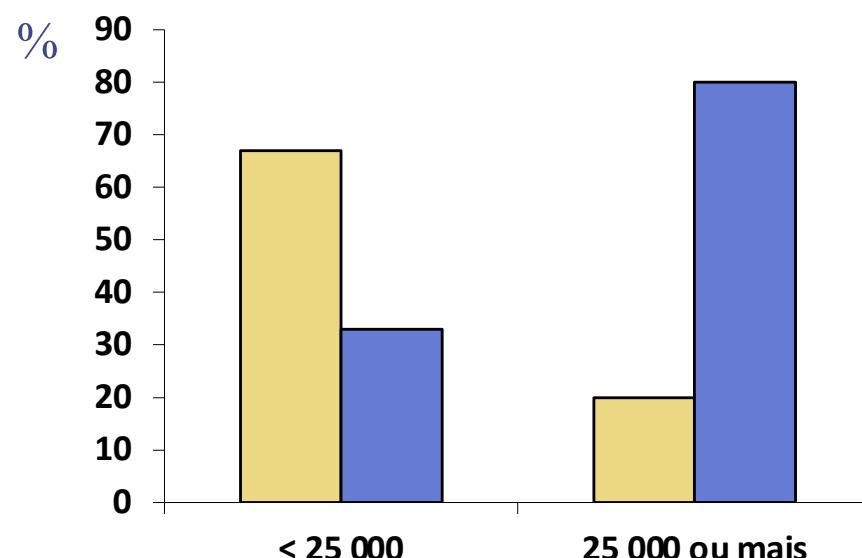
$$(400/500)(100) = 80\%$$



Representação Gráfica



■ Não usa cartão ■ Usa cartão



■ Não usa cartão ■ Usa cartão



Associação

■ Deseja-se saber o grau de associação entre duas variáveis de modo que se possa prever melhor o resultado de uma delas quando conhecemos a realização de outra.



Medidas de Associação



Frequência Esperada

- ◆ Supondo que *não existe associação entre renda anual e o uso de cartão de crédito*, qual a frequência esperada de pessoas com renda < 25000 que deveriam usar cartão de crédito ?

**proporção de usuários
de cartão de crédito** \times

$$\left(\frac{f_{\cdot j}}{n} \right)$$

**Total de pessoas
com renda <25000**

$$f_{i \cdot}$$

$$0,625 \times 300 = 187,5$$



Frequência Esperada

$$e_{ij} = \frac{f_{\cdot j}}{n} \times f_{i\cdot} = \frac{f_{\cdot j} \times f_{i\cdot}}{n}$$

■ onde:

n → tamanho da amostra

$f_{i\cdot}$ → frequência total da linha i

$f_{\cdot j}$ → frequência total da coluna j



Frequência Esperada

Renda	Uso		Total
	Não	Sim	
< 25	200 (112,5)	100 (187,5)	300
≥ 25	100 (187,5)	400 (312,5)	500
Total	300	500	800



Frequência Esperada

Renda	Uso		Total
	Não	Sim	
< 25	200 (112,5) 87,5	100 (187,5) -87,5	300
≥ 25	100 (187,5) -87,5	400 (312,5) 87,5	500
Total	300	500	800



Qui-quadrado (χ^2)

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

Observe que: $0 \leq \chi^2 < \infty$



Exemplo

Renda	Uso		Total
	Não	Sim	
< 25	200 (112,5)	100 (187,5)	300
≥ 25	87,5 -87,5	-87,5 87,5	500
Total	300	500	800

$$\begin{aligned}\chi^2 &= \frac{(200-112,5)^2}{112,5} + \frac{(100-187,5)^2}{187,5} + \\ &+ \frac{(100-187,5)^2}{187,5} + \frac{(400-312,5)^2}{312,5} = \\ &= 68,06 + 40,83 + 40,83 + 24,50 \\ &= 174,22\end{aligned}$$



Coeficiente de Contingência

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

$$e \quad 0 \leq C \leq \sqrt{\frac{t-1}{t}}$$

onde t = min (r, s)



Exemplo

$$C = \sqrt{\frac{174,22}{174,22 + 800}} = 0,42$$

$$t = \min(2, 2) = 2$$

$$\text{então } 0 \leq C \leq \sqrt{\frac{2-1}{2}} = 0,707$$