

# *Parte V*

## *Relação entre variáveis*

- Como medir e testar a significância da associação entre duas variáveis qualitativas (Capítulo 12)
- Como estudar a correlação entre duas variáveis quantitativas (Capítulo 13)
- Como construir modelos para o relacionamento entre duas variáveis (Capítulo 13)

# Estatística Aplicada às Ciências Sociais

## Sexta Edição

Pedro Alberto Barbetta

Florianópolis: Editora da UFSC, 2006

## Cap. 12 – Análise de dados categorizados

# Comparação entre amostras

- Comparação entre amostras em que a variável-resposta é qualitativa (categórica).
  - Exemplo: comparação entre amostras de homens e de mulheres quanto ao tabagismo (*fumante* ou *não-fumante*)
- Como testar, com base em amostras, se as populações que geraram as amostras têm a mesma distribuição de probabilidades em termos das categorias da variável-resposta? (Teste de homogeneidade qui-quadrado)

# Análise de associação

- Existe *associação* entre duas variáveis se o conhecimento de uma altera a probabilidade de algum resultado da outra.
  - Exemplo: **Clima** (chuvoso ou ensolarado) e **ir à praia** (sim ou não)
- Como testar se existe associação entre duas variáveis qualitativas? (Teste de associação qui-quadrado)
- Como medir o grau de associação descrita pelos dados amostrais? (Coeficientes de associação)

# Teste de associação qui-quadrado

(equivalente ao teste de homogeneidade qui-quadrado)

- Exemplo 12.1: Existe associação entre sexo (homem ou mulher) e tabagismo (fumante ou não-fumante)?
- Hipóteses:
  - $H_0$ : *Sexo e tabagismo* são variáveis *independentes* na população em estudo.
  - $H_1$ : Existe *associação* entre as variáveis *sexo* e *tabagismo*, na população em estudo.

ou:

$$H_0: \pi_h = \pi_m \quad \text{e} \quad H_1: \pi_h \neq \pi_m$$

$\pi_h$  = probabilidade de um homem ser fumante

$\pi_m$  = probabilidade de uma mulher ser fumante

## Exemplo 12.1: os dados

Tabagismo	Sexo		Total (%)
	masculino	feminino	
fumante	92	38	130
(%)	(46)	(38)	(43,3)
não-fumante	108	62	170
(%)	(54)	(62)	(56,7)
Total	200	100	300
(%)	(100)	(100)	(100)

## Exemplo 12.1: os cálculos

$$E = \frac{(\text{total da linha}) \times (\text{total da coluna})}{(\text{total geral})}$$

Tabagismo	Sexo		Total
	masculino	feminino	
fumante	$E = \frac{130 \times 200}{300} = 86,67$	$E = \frac{130 \times 100}{300} = 43,33$	130
não fumante	$E = \frac{170 \times 200}{300} = 113,33$	$E = \frac{170 \times 100}{300} = 56,67$	170
Total	200	100	300

## Exemplo 12.1: os cálculos

Estatística do teste:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Tabagismo	Sexo	
	masculino	feminino
fumante	$\frac{(92 - 86,67)^2}{86,67} = 0,328$	$\frac{(38 - 43,33)^2}{43,33} = 0,656$
não fumante	$\frac{(108 - 113,33)^2}{113,33} = 0,251$	$\frac{(62 - 56,67)^2}{56,67} = 0,501$

$$\chi^2 = 0,328 + 0,656 + 0,251 + 0,501 = 1,74$$



# Distribuição de referência

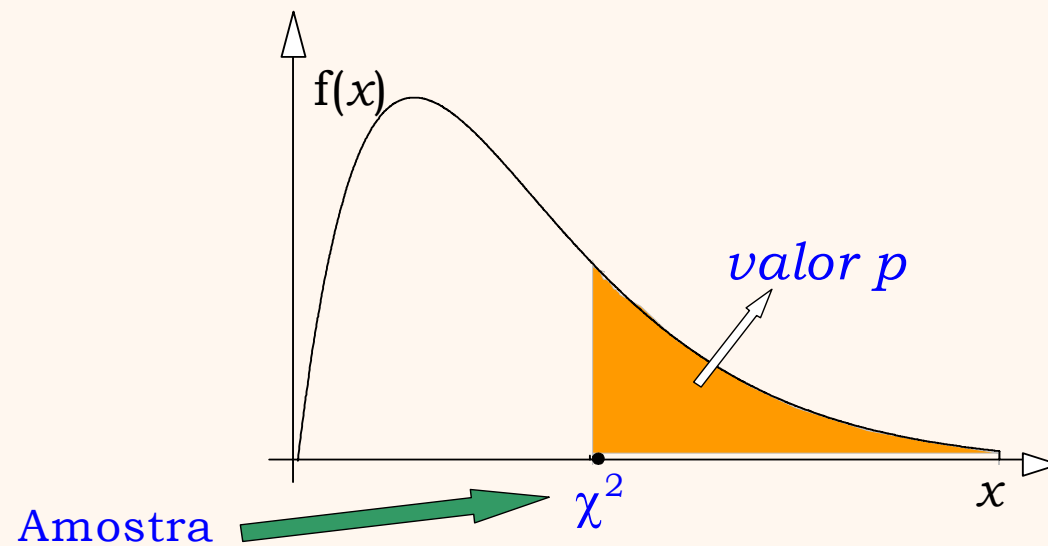
- Sob  $H_0$ , os possíveis valores de  $\chi^2$  seguem uma distribuição qui-quadrado com graus de liberdade:

$$gl = (\ell - 1) \cdot (c - 1)$$

onde  $\ell$  é o número de linhas e  $c$  é o número de colunas da tabela.

- Supõe-se que as amostras sejam razoavelmente grandes (todas as frequências esperadas pelo menos iguais a cinco).

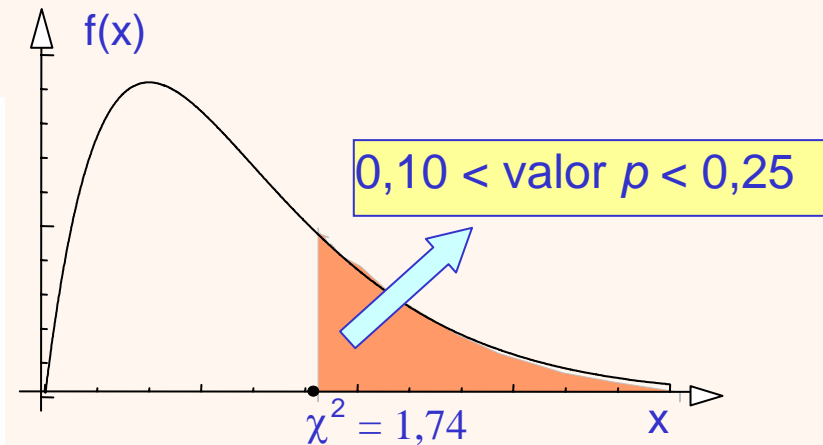
# Valor-p



Ver Tabela 6 no apêndice do livro

## Exemplo 12.1: uso da Tabela Qui-quadrado

dados observados		Área na cauda superior			
	<i>gl</i>	0,250	0,100	0,050	...
$\chi^2 = 1,74$	1	1,32	2,71	3,84	...
	...	...	...	...	...



Valor  $p$  entre 0,10 e 0,25. Portanto, o teste aceita  $H_0$  ao nível de significância de 5%.

## Correção em tabelas 2 x 2

- No cálculo do qui-quadrado, subtrai-se 0,5 unidades na magnitude da diferença entre as frequências observada e esperada, em cada casela.

$$\chi^2 = \sum \frac{(|O - E| - 0,5)^2}{E}$$

Refazer o Exemplo 12.1

# Medidas de associação

- Um coeficiente de associação descreve, em termos das amostras observadas, o quanto os dados de duas variáveis se mostram associados.
- É uma medida descritiva da(s) amostra(s). Não é uma inferência como o teste qui-quadrado!
- Em geral, os coeficientes geram valores entre 0 (independência) e 1 (associação perfeita)

# Coeficiente de contingência

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

# Coeficiente de contingência modificado

$$C^* = \sqrt{\frac{k \cdot \chi^2}{(k-1) \cdot (n + \chi^2)}}$$

onde  $k$  é o menor valor entre  $\ell$  (número de linhas da tabela) e  $c$  (número de colunas da tabela).

## Exemplo: coeficiente de contingência modificado

Amostra A		
Tabagismo	Sexo	
	homem	mulher
fumante	80 (40%)	40 (40%)
não-fumante	120 (60%)	60 (60%)
Total	200 (100%)	100 (100%)

Amostra B		
Tabagismo	Sexo	
	homem	mulher
fumante	200 (100%)	0 (0%)
não-fumante	0 (0%)	100 (100%)
Total	200 (100%)	100 (100%)

$$C^* = \sqrt{\frac{(2) \cdot (0)}{(2-1) \cdot (0+300)}} = 0$$

→ *Independência!*

$$C^* = \sqrt{\frac{(2) \cdot (300)}{(2-1) \cdot (300+300)}} = 1$$

→ *Associação perfeita!*

Ver, no livro, outros coeficientes de associação.