



SME0803 Visualização e Exploração de Dados

Associação entre variáveis qualitativas

Prof. Cibeles Russo

cibele@icmc.usp.br

Baseado em

Murteira, B. J. F., Análise Exploratória de Dados. McGraw-Hill, Lisboa, 1993.

Notas de aula de Análise Exploratória de Dados do Mário de Castro, ICMC-USP, 2010.

Tabelas de contingência (frequências absolutas)

Sejam $x \in x_1, \dots, x_k$ e $y \in y_1, \dots, y_m$, $1 < k \leq n$ e $1 < m \leq n$.

Seja f_{ij} a **frequência absoluta** do par (x_i, y_j) , $i = 1, \dots, k$, $j = 1, \dots, m$.

Tabelas de contingência (frequências absolutas)

Sejam $x \in x_1, \dots, x_k$ e $y \in y_1, \dots, y_m$, $1 < k \leq n$ e $1 < m \leq n$.

Seja f_{ij} a **frequência absoluta** do par (x_i, y_j) , $i = 1, \dots, k$, $j = 1, \dots, m$.

Tabela de contingência (*contingency table*) ou **tabela de dupla entrada**: tabela com os diferentes pares (x_i, y_j) e suas frequências f_{ij} .

Distribuição conjunta de frequências absolutas x e y .

$x \backslash y$	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}	\dots	f_{1j}	\dots	f_{1m}	$f_{1\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}	\dots	f_{ij}	\dots	f_{im}	$f_{i\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}	\dots	f_{kj}	\dots	f_{km}	$f_{k\bullet}$
Total	$f_{\bullet 1}$	\dots	$f_{\bullet j}$	\dots	$f_{\bullet m}$	n

Distribuição conjunta de frequências absolutas x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}	\dots	f_{1j}	\dots	f_{1m}	$f_{1\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}	\dots	f_{ij}	\dots	f_{im}	$f_{i\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}	\dots	f_{kj}	\dots	f_{km}	$f_{k\bullet}$
Total	$f_{\bullet 1}$	\dots	$f_{\bullet j}$	\dots	$f_{\bullet m}$	n

em que $f_{i\bullet} = \sum_{j=1}^m f_{ij}$ para $i = 1, \dots, k$

Distribuição conjunta de frequências absolutas x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}	\dots	f_{1j}	\dots	f_{1m}	$f_{1\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}	\dots	f_{ij}	\dots	f_{im}	$f_{i\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}	\dots	f_{kj}	\dots	f_{km}	$f_{k\bullet}$
Total	$f_{\bullet 1}$	\dots	$f_{\bullet j}$	\dots	$f_{\bullet m}$	n

em que $f_{i\bullet} = \sum_{j=1}^m f_{ij}$ para $i = 1, \dots, k$ e $\sum_{i=1}^k f_{i\bullet} = n$,

Distribuição conjunta de frequências absolutas x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}	\dots	f_{1j}	\dots	f_{1m}	$f_{1\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}	\dots	f_{ij}	\dots	f_{im}	$f_{i\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}	\dots	f_{kj}	\dots	f_{km}	$f_{k\bullet}$
Total	$f_{\bullet 1}$	\dots	$f_{\bullet j}$	\dots	$f_{\bullet m}$	n

em que $f_{i\bullet} = \sum_{j=1}^m f_{ij}$ para $i = 1, \dots, k$ e $\sum_{i=1}^k f_{i\bullet} = n$,

$$f_{\bullet j} = \sum_{i=1}^k f_{ij} \text{ para } j = 1, \dots, m$$

Distribuição conjunta de frequências absolutas x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}	\dots	f_{1j}	\dots	f_{1m}	$f_{1\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}	\dots	f_{ij}	\dots	f_{im}	$f_{i\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}	\dots	f_{kj}	\dots	f_{km}	$f_{k\bullet}$
Total	$f_{\bullet 1}$	\dots	$f_{\bullet j}$	\dots	$f_{\bullet m}$	n

em que $f_{i\bullet} = \sum_{j=1}^m f_{ij}$ para $i = 1, \dots, k$ e $\sum_{i=1}^k f_{i\bullet} = n$,

$f_{\bullet j} = \sum_{i=1}^k f_{ij}$ para $j = 1, \dots, m$ e $\sum_{j=1}^m f_{\bullet j} = n$

Distribuição conjunta de frequências absolutas x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}	\dots	f_{1j}	\dots	f_{1m}	$f_{1\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}	\dots	f_{ij}	\dots	f_{im}	$f_{i\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}	\dots	f_{kj}	\dots	f_{km}	$f_{k\bullet}$
Total	$f_{\bullet 1}$	\dots	$f_{\bullet j}$	\dots	$f_{\bullet m}$	n

em que $f_{i\bullet} = \sum_{j=1}^m f_{ij}$ para $i = 1, \dots, k$ e $\sum_{i=1}^k f_{i\bullet} = n$,

$f_{\bullet j} = \sum_{i=1}^k f_{ij}$ para $j = 1, \dots, m$ e $\sum_{j=1}^m f_{\bullet j} = n$ e $\sum_{i=1}^k \sum_{j=1}^m f_{ij} = n$.

Tabelas de contingência (frequências absolutas)

Distribuição marginal de x (em destaque)

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}	\dots	f_{1j}	\dots	f_{1m}	$f_{1\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}	\dots	f_{ij}	\dots	f_{im}	$f_{i\bullet}$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}	\dots	f_{kj}	\dots	f_{km}	$f_{k\bullet}$
Total	$f_{\bullet 1}$	\dots	$f_{\bullet j}$	\dots	$f_{\bullet m}$	n

Tabelas de contingência (frequências absolutas)

Distribuição marginal de y (em destaque)

x \ y	y					Total
	y_1	...	y_j	...	y_m	
x_1	f_{11}	...	f_{1j}	...	f_{1m}	$f_{1\bullet}$
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_i	f_{i1}	...	f_{ij}	...	f_{im}	$f_{i\bullet}$
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_k	f_{k1}	...	f_{kj}	...	f_{km}	$f_{k\bullet}$
Total	$f_{\bullet 1}$...	$f_{\bullet j}$...	$f_{\bullet m}$	n

Tabelas de contingência (frequências relativas)

Seja f_{ij}^* a **frequência relativa** do par (x_i, y_j) , $i = 1, \dots, k$, $j = 1, \dots, m$.

Tabelas de contingência (frequências relativas)

Seja f_{ij}^* a **frequência relativa** do par (x_i, y_j) , $i = 1, \dots, k$, $j = 1, \dots, m$.

Distribuição conjunta de frequências relativas (*contingency table*) ou **tabela de dupla entrada de frequências relativas**: tabela com os diferentes pares (x_i, y_j) e suas frequências relativas f_{ij}^* .

Distribuição conjunta de frequências relativas de x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}^*	\dots	f_{1j}^*	\dots	f_{1m}^*	$f_{1\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}^*	\dots	f_{ij}^*	\dots	f_{im}^*	$f_{i\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}^*	\dots	f_{kj}^*	\dots	f_{km}^*	$f_{k\bullet}^*$
Total	$f_{\bullet 1}^*$	\dots	$f_{\bullet j}^*$	\dots	$f_{\bullet m}^*$	1

Distribuição conjunta de frequências relativas de x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}^*	\dots	f_{1j}^*	\dots	f_{1m}^*	$f_{1\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}^*	\dots	f_{ij}^*	\dots	f_{im}^*	$f_{i\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}^*	\dots	f_{kj}^*	\dots	f_{km}^*	$f_{k\bullet}^*$
Total	$f_{\bullet 1}^*$	\dots	$f_{\bullet j}^*$	\dots	$f_{\bullet m}^*$	1

em que $f_{i\bullet}^* = \sum_{j=1}^m f_{ij}^*$ para $i = 1, \dots, k$

Distribuição conjunta de frequências relativas de x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}^*	\dots	f_{1j}^*	\dots	f_{1m}^*	$f_{1\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}^*	\dots	f_{ij}^*	\dots	f_{im}^*	$f_{i\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}^*	\dots	f_{kj}^*	\dots	f_{km}^*	$f_{k\bullet}^*$
Total	$f_{\bullet 1}^*$	\dots	$f_{\bullet j}^*$	\dots	$f_{\bullet m}^*$	1

em que $f_{i\bullet}^* = \sum_{j=1}^m f_{ij}^*$ para $i = 1, \dots, k$ e $\sum_{i=1}^k f_{i\bullet}^* = 1$,

Distribuição conjunta de frequências relativas de x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}^*	\dots	f_{1j}^*	\dots	f_{1m}^*	$f_{1\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}^*	\dots	f_{ij}^*	\dots	f_{im}^*	$f_{i\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}^*	\dots	f_{kj}^*	\dots	f_{km}^*	$f_{k\bullet}^*$
Total	$f_{\bullet 1}^*$	\dots	$f_{\bullet j}^*$	\dots	$f_{\bullet m}^*$	1

em que $f_{i\bullet}^* = \sum_{j=1}^m f_{ij}^*$ para $i = 1, \dots, k$ e $\sum_{i=1}^k f_{i\bullet}^* = 1$,

$$f_{\bullet j}^* = \sum_{i=1}^k f_{ij}^* \text{ para } j = 1, \dots, m$$

Distribuição conjunta de frequências relativas de x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}^*	\dots	f_{1j}^*	\dots	f_{1m}^*	$f_{1\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}^*	\dots	f_{ij}^*	\dots	f_{im}^*	$f_{i\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}^*	\dots	f_{kj}^*	\dots	f_{km}^*	$f_{k\bullet}^*$
Total	$f_{\bullet 1}^*$	\dots	$f_{\bullet j}^*$	\dots	$f_{\bullet m}^*$	1

em que $f_{i\bullet}^* = \sum_{j=1}^m f_{ij}^*$ para $i = 1, \dots, k$ e $\sum_{i=1}^k f_{i\bullet}^* = 1$,

$f_{\bullet j}^* = \sum_{i=1}^k f_{ij}^*$ para $j = 1, \dots, m$ e $\sum_{j=1}^m f_{\bullet j}^* = 1$

Distribuição conjunta de frequências relativas de x e y .

x \ y	y					Total
	y_1	\dots	y_j	\dots	y_m	
x_1	f_{11}^*	\dots	f_{1j}^*	\dots	f_{1m}^*	$f_{1\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	f_{i1}^*	\dots	f_{ij}^*	\dots	f_{im}^*	$f_{i\bullet}^*$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	f_{k1}^*	\dots	f_{kj}^*	\dots	f_{km}^*	$f_{k\bullet}^*$
Total	$f_{\bullet 1}^*$	\dots	$f_{\bullet j}^*$	\dots	$f_{\bullet m}^*$	1

em que $f_{i\bullet}^* = \sum_{j=1}^m f_{ij}^*$ para $i = 1, \dots, k$ e $\sum_{i=1}^k f_{i\bullet}^* = 1$,

$f_{\bullet j}^* = \sum_{i=1}^k f_{ij}^*$ para $j = 1, \dots, m$ e $\sum_{j=1}^m f_{\bullet j}^* = 1$ e $\sum_{i=1}^k \sum_{j=1}^m f_{ij}^* = 1$.

Distribuição conjunta de frequências relativas de x e y .

$x \backslash y$	y_1	...	y_j	...	y_m	Total
x_1	f_{11}/n	...	f_{1j}/n	...	f_{1m}/n	$f_{1\bullet}/n$
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_i	f_{i1}/n	...	f_{ij}/n	...	f_{im}/n	$f_{i\bullet}/n$
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_k	f_{k1}/n	...	f_{kj}/n	...	f_{km}/n	$f_{k\bullet}/n$
Total	$f_{\bullet 1}/n$...	$f_{\bullet j}/n$...	$f_{\bullet m}/n$	1

Distribuição conjunta de frequências relativas de x e y .

$x \backslash y$	y_1	...	y_j	...	y_m	Total
x_1	f_{11}/n	...	f_{1j}/n	...	f_{1m}/n	$f_{1\bullet}/n$
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_i	f_{i1}/n	...	f_{ij}/n	...	f_{im}/n	$f_{i\bullet}/n$
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_k	f_{k1}/n	...	f_{kj}/n	...	f_{km}/n	$f_{k\bullet}/n$
Total	$f_{\bullet 1}/n$...	$f_{\bullet j}/n$...	$f_{\bullet m}/n$	1

Tabelas de contingência (frequências relativas)

Distribuição marginal de x (em destaque)

x \ y	y						Total
	y_1	...	y_j	...	y_m		
x_1	f_{11}/n	...	f_{1j}/n	...	f_{1m}/n	$f_{1\bullet}/n$	
\vdots	\vdots	...	\vdots	...	\vdots	\vdots	
x_i	f_{i1}/n	...	f_{ij}/n	...	f_{im}/n	$f_{i\bullet}/n$	
\vdots	\vdots	...	\vdots	...	\vdots	\vdots	
x_k	f_{k1}/n	...	f_{kj}/n	...	f_{km}/n	$f_{k\bullet}/n$	
Total	$f_{\bullet 1}/n$...	$f_{\bullet j}/n$...	$f_{\bullet m}/n$	1	

Tabelas de contingência (frequências relativas)

Distribuição marginal de y (em destaque)

x \ y	y					Total
	y_1	...	y_j	...	y_m	
x_1	f_{11}/n	...	f_{1j}/n	...	f_{1m}/n	$f_{1\bullet}/n$
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_i	f_{i1}/n	...	f_{ij}/n	...	f_{im}/n	$f_{i\bullet}/n$
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_k	f_{k1}/n	...	f_{kj}/n	...	f_{km}/n	$f_{k\bullet}/n$
Total	$f_{\bullet 1}/n$...	$f_{\bullet j}/n$...	$f_{\bullet m}/n$	1

Tabelas de contingência (frequências relativas)

Distribuição **condicional** de y dado $x = x_i$ (em destaque)

x \ y	y					Total
	y_1	...	y_j	...	y_m	
x_1	$f_{11}/f_{1\bullet}$...	$f_{1j}/f_{1\bullet}$...	$f_{1m}/f_{1\bullet}$	1
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_i	$f_{i1}/f_{i\bullet}$...	$f_{ij}/f_{i\bullet}$...	$f_{im}/f_{i\bullet}$	1
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_k	$f_{k1}/f_{k\bullet}$...	$f_{kj}/f_{k\bullet}$...	$f_{km}/f_{k\bullet}$	1

Tabelas de contingência (frequências relativas)

Distribuição **condicional** de x dado $y = y_j$ (em destaque)

x \ y	y				
	y_1	...	y_j	...	y_m
x_1	$f_{11}/f_{\bullet 1}$...	$f_{1j}/f_{\bullet j}$...	$f_{1m}/f_{\bullet m}$
\vdots	\vdots	...	\vdots	...	\vdots
x_i	$f_{i1}/f_{\bullet 1}$...	$f_{ij}/f_{\bullet j}$...	$f_{im}/f_{\bullet m}$
\vdots	\vdots	...	\vdots	...	\vdots
x_k	$f_{k1}/f_{\bullet 1}$...	$f_{kj}/f_{\bullet j}$...	$f_{km}/f_{\bullet m}$
Total	1	...	1	...	1

Associação entre variáveis qualitativas

As tabelas de distribuição condicional também são tabelas de frequências relativas. Qual usar?

- (a) Relação causal bilateral ($x \leftrightarrow y$): Tabela de frequências relativas (dividido por n).

Associação entre variáveis qualitativas

As tabelas de distribuição condicional também são tabelas de frequências relativas. Qual usar?

- (a) Relação causal bilateral ($x \leftrightarrow y$): Tabela de frequências relativas (dividido por n).
- (b) Relação causal unilateral ($x \rightarrow y$): Distribuição condicional de y dado x .

Associação entre variáveis qualitativas

As tabelas de distribuição condicional também são tabelas de frequências relativas. Qual usar?

- (a) Relação causal bilateral ($x \leftrightarrow y$): Tabela de frequências relativas (dividido por n).
- (b) Relação causal unilateral ($x \rightarrow y$): Distribuição condicional de y dado x .
- (b) Relação causal unilateral ($y \rightarrow x$) Distribuição condicional de x dado y .

Associação entre variáveis qualitativas

Obs 1: Em (b) temos k distribuições condicionais de y . Quanto mais semelhantes forem estas distribuições, mais fraca é a associação entre x e y .

Associação entre variáveis qualitativas

Obs 1: Em (b) temos k distribuições condicionais de y . Quanto mais semelhantes forem estas distribuições, mais fraca é a associação entre x e y .

Obs 2: Em (c) é usual trocar os nomes, de modo que x ocupe as linhas e y ocupe as colunas da tabela de contingências.

Adaptação do conceito de independência

Para avaliar a associação entre duas variáveis qualitativas, precisamos entender a ideia de independência.

x e y são independentes se, e somente se,

$$f_{ij} = \frac{f_{i\bullet} f_{\bullet j}}{n}, j = 1, \dots, m \text{ e } i = 1, \dots, k.$$

De forma equivalente,

$$\frac{f_{ij}}{n} = \frac{f_{i\bullet}}{n} \frac{f_{\bullet j}}{n}, j = 1, \dots, m \text{ e } i = 1, \dots, k.$$

Obs: Adaptação do conceito de independência entre as variáveis aleatórias discretas X e Y : $P(X = a, Y = b) = P(X = a)P(Y = b)$.

Exemplo

Considere novamente os dados da CompanhiaMB, especificamente as variáveis grau de instrução e estado civil. Como avaliar se existe associação entre elas? Considere a tabela de contingência.

grau de instrução \ estado civil	estado civil	
	casado	solteiro
ensino_fundamental	5	7
ensino_medio	12	6
superior	3	3

Exemplo

Como seria essa tabela sob independência? Primeiro obtemos os totais por linhas e colunas:

estado civil grau de instrução	estado civil		Total
	casado	solteiro	
ensino_fundamental	5	7	12
ensino_medio	12	6	18
superior	3	3	6
Total	20	16	36

Exemplo

E então fazemos $f_{ij}^{ind} = \frac{f_{i\bullet} \cdot f_{\bullet j}}{n}$, $j = 1, \dots, m$ e $i = 1, \dots, k$.

estado civil \ grau de instrução	estado civil		Total
	casado	solteiro	
ensino_fundamental	6,7	5,3	12
ensino_medio	10	8	18
superior	3,3	2,7	6
Total	20	16	36

Medidas de associação: Qui-quadrado de Pearson

Baseado nas diferenças entre as frequências absolutas observadas (f_{ij}) e as frequências calculadas supondo independência entre x e y ($f_{ij}^{ind} = f_{i\bullet} f_{\bullet j} / n$):

$$Q^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(f_{ij} - f_{ij}^{ind})^2}{f_{ij}^{ind}}$$

$Q^2 = 0 \implies$ ausência de associação entre x e y

$Q^2 > 0 \implies$ comparar com o quantil de uma v.a. com distribuição

$$\chi^2_{(k-1)(m-1)}$$

Coeficientes de associação

Coeficiente de Contingência:

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

O valor máximo de C depende de k e m .

Coeficiente de Tschuprow: (ver

https://en.wikipedia.org/wiki/Tschuprow%27s_T)

$$T = \sqrt{\frac{Q^2}{n\sqrt{(k-1)(m-1)}}$$

Obs: $0 \leq T \leq 1$. $T = 0$ indica independência.

Exemplo

Considere o caso particular em que

Distribuição conjunta de frequências absolutas x e y .

$x \backslash y$	y					Total
	y_1	\dots	y_i	\dots	y_k	
x_1	f_{11}	\dots	0	\dots	0	f_{11}
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_i	0	\dots	f_{ii}	\dots	0	f_{ii}
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
x_k	0	\dots	0	\dots	f_{kk}	f_{kk}
Total	f_{11}	\dots	f_{ii}	\dots	f_{kk}	n

Exemplo

Considere o caso particular em que

Distribuição conjunta de frequências absolutas x e y .

x \ y	y					Total
	y_1	...	y_i	...	y_k	
x_1	f_{11}	...	0	...	0	f_{11}
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_i	0	...	f_{ii}	...	0	f_{ii}
\vdots	\vdots	...	\vdots	...	\vdots	\vdots
x_k	0	...	0	...	f_{kk}	f_{kk}
Total	f_{11}	...	f_{ii}	...	f_{kk}	n

Exercício: Provar que, neste caso, $Q^2 = n(k - 1)$. Logo, $T = 1$.

Apresente outros exemplos nos quais $T = 1$.

Exemplo: estado civil x grau de instrução

$Q^2 = 1.91$ (exercício: desenvolva os cálculos ou use o R).

Comparar com o quantil da $Q^2_{(3-1)(2-1)}$, supondo $\alpha = 0.05$ temos $q_c = 5.99 \implies$ Não existe associação entre estado civil e grau de instrução. Obs: Um teste mais adequado nesse caso seria o Teste Exato de Fisher.

$$C = \sqrt{\frac{Q^2}{Q^2 + n}} = 0.22$$

$T = \sqrt{\frac{Q^2}{n\sqrt{(k-1)(m-1)}}} = 0.19$, que também indica associação fraca entre estado civil e instrução.

Exemplo: variável qualitativa x variável qualitativa

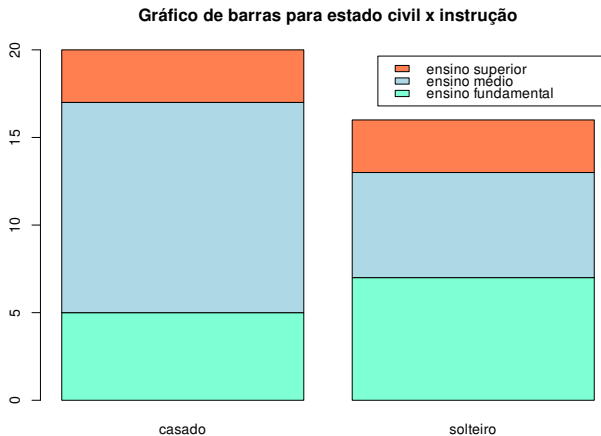


Figura: Gráfico de barras empilhadas de estado civil e grau de instrução para os dados da Companhia MB. Fonte: Elaborado pela autora.

Exemplo: variável qualitativa x variável qualitativa

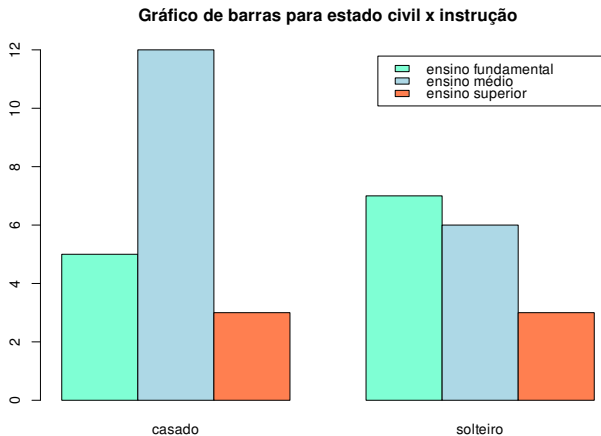


Figura: Gráfico de barras adjacentes de estado civil e grau de instrução para os dados da Companhia MB. Fonte: Elaborado pela autora.

Exemplo: variável qualitativa x variável qualitativa

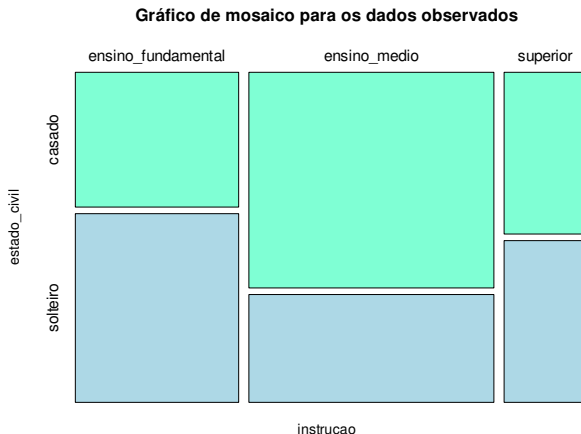


Figura: Gráfico de mosaico de estado civil e grau de instrução para os dados da Companhia MB. Fonte: Elaborado pela autora.

Exemplo: variável qualitativa x variável qualitativa



Figura: Gráfico de mosaico de estado civil e grau de instrução sob independência. Fonte: Elaborado pela autora.