

Tabelas Bidimensionais

Unidade I Parte 4

Teste de Independência para Dados Ordinais

O Problema ...

Deseja-se verificar se existe associação entre duas variáveis categorizadas *ordinais* X e Y .

Neste caso, podemos não só avaliar a *intensidade da associação*, mas também o *sentido da associação* (positivo ou negativo).

Exemplo:

X - nível de instrução (*fundamental, médio e superior*)

Y - nível de satisfação com o produto (*insatisfeito, indiferente e satisfeito*)

Associação positiva - quanto maior o nível de instrução, maior a satisfação com o produto.

Associação negativa - quanto maior o nível de instrução, menor a satisfação com o produto

Sejam

$u_1 \leq u_2 \leq \dots \leq u_I$ os escores para as linhas

e sejam

$v_1 \leq v_2 \leq \dots \leq v_J$ os escores para as colunas.

Os escores tem o mesmo ordenamento que os níveis das categorias a são ditos ser monótonos.

Para os escores escolhidos, o coeficiente de correlação momento-produto de Pearson entre X e Y é igual a,

$$r = \frac{\sum_{i,j} u_i v_j n_{ij} - \left(\sum u_i n_{i+} \right) \left(\sum v_j n_{+j} \right) / n}{\sqrt{\left[\sum u_i^2 n_{i+} - \frac{\left(\sum u_i n_{i+} \right)^2}{n} \right] \left[\sum v_j^2 n_{+j} - \frac{\left(\sum v_j n_{+j} \right)^2}{n} \right]}}$$

Teste de Hipóteses

H_0) Independência entre X e Y ($\rho = 0$) ;

H_1) Dependência entre X e Y ($\rho \neq 0$).

A estatística do teste é,

$$M^2 = (n-1) r^2$$

que, para amostras grandes, tem *distribuição aproximadamente qui-quadrado* com 1 g.l. .

Ou ainda,

$$M = \sqrt{n-1} r$$

que, para amostras grandes, tem *distribuição aproximadamente $N(0,1)$* sob H_0

Exemplo

Consumo de Álcool	Mal formação congênita		Total
	Ausente	Presente	
0	17066	48	17114
< 1	14464	38	14502
1-2	788	5	793
3-5	126	1	127
≥ 6	37	1	38
Total	32481	93	32574

Fonte: B. I. Graubard and E. L. Korn, *Biometrics* 43:471-476(1987)
Reimpresso com a permissão da Sociedade de Biometria

Escolha dos Escores

- **Escores igualmente espaçados** frequentemente fornecem uma composição razoável quando o rótulo das categorias não sugerem algumas escolhas óbvias, tal como as categorias (*liberal, moderado, conservador*) para filosofia política.
- **Ponto médio das classes** de uma variável quantitativa categorizada.
- **“Midrank”** :
 - procedimento alternativo evita a responsabilidade da seleção dos escores e usa os dados para formá-los automaticamente.
 - Atribui “**rank**” (postos) aos indivíduos e os usa como escores das categorias;
 - Para todos os indivíduos em uma categoria atribui-se a média dos postos (**midranks**) que seriam aplicados para um ordenamento completo da amostra de **1** a **n**.