USP – ICMC – SME0811 - Dados Categorizados
Prova única – 1º/2012 – S O L U Ç Ã O

1. (a) Em uma tabela 2×2, se $\pi_{ij} = 1/4$, $i, j = 1, 2$, as variáveis X e Y são independentes. Apresente uma outra tabela 2×2 em que as variáveis X e Y são independentes.

   (b) Para uma tabela genérica $I \times J$, $I \geq 2$, $J \geq 2$ e $I \neq J$, apresente probabilidades π_{ij}, $i = 1, \ldots, I$ e $j = 1, \ldots, J$, de modo que X e Y sejam independentes.

   Solução. (a) Uma solução é obtida tomando $a > 0$, $b > 0$,

$$\pi_{11} = \pi_{21} = \frac{a}{2(a+b)} \quad \text{e} \quad \pi_{12} = \pi_{22} = \frac{b}{2(a+b)}.$$

   Notamos que a razão de chances é igual a 1, de modo que as variáveis são independentes. Como exemplo numérico, para $a = 5$ e $b = 3$ obtemos $\pi_{11} = \pi_{21} = 5/16$ e $\pi_{12} = \pi_{22} = 3/16$.

   (b) Em uma tabela $I \times J$ se tomarmos $\pi_{ij} = \frac{1}{IJ}$ notamos que para todo $i \in \{1, \ldots, I\}$ a distribuição condicional de $Y|X = i$ tem probabilidades $(1/J, \ldots, 1/J)$. Logo, as variáveis são independentes. Também pode ser verificado que $\pi_{ij} = \pi_{i+} \times \pi_{+j}$, $\forall\, i, j$.

2. X e Y são condicionalmente independentes dada Z. X e Z são marginalmente independentes. (a) Prove que X é conjuntamente independente de Y e Z. (b) Prove que X e Y são marginalmente independentes.

   Solução. (a) Escrevemos

$$\begin{aligned}
&P(X = i, Y = j, Z = k) \\
&= P(X = i|Y = j, Z = k)P(Y = j, Z = k), \text{ pela regra do produto} \\
&= P(X = i|Z = k)P(Y = j, Z = k), \text{ pois } (X \coprod Y)|Z \\
&= P(X = i)P(Y = j, Z = k), \text{ pois } X \coprod Z,
\end{aligned}$$

   $\forall\, i, j, k$. Logo, $X \coprod (Y, Z)$.

   (b) Escrevemos

$$\begin{aligned}
P(X = i, Y = j) &= \sum_k P(X = i, Y = j, Z = k) \\
&= \sum_k P(X = i)P(Y = j, Z = k), \text{ utilizando a letra (a)} \\
&= P(X = i) \sum_k P(Y = j, Z = k) = P(X = i)P(Y = j),
\end{aligned}$$

   $\forall\, i, j$. Logo, $X \coprod Y$.
   De uma forma mais direta, da letra (a) temos que $X \coprod (Y, Z)$. Logo, $X \coprod Y$.

3. Dados foram coletados de três variáveis binárias A, C e M com níveis {Sim: 1, Não: 2} indicando a presença ou a ausência das substâncias A, C, e M como resultado de uma análise. Os resultados do teste de bondade do ajuste de alguns modelos com a estatística G^2 são mostrados abaixo. Para cada modelo ajustado também são apresentadas estimativas das frequências esperadas. (a) Selecione um modelo que ajuste bem os dados. (b) Utilizando o modelo do item (a), apresente estimativas das razões de chances parciais para cada par de variáveis e descreva os resultados obtidos.

Bondade do ajuste

Modelos	gl	G2	valor-p
(A,C,M)	4	843.827	< 0.001
(M,AC)	3	843.827	< 0.001
(AM,CM)	2	187.754	< 0.001
(AC,AM,CM)	1	0.374	0.541

Frequencias esperadas estimadas

			Modelos			
M	C	A	(A,C,M)	(M,AC)	(AM,CM)	(AC,AM,CM)
Sim	Sim	Sim	540.0	611.2	909.24	910.38
Nao	Sim	Sim	740.2	837.8	438.84	538.62
Sim	Nao	Sim	282.1	210.9	45.76	44.62
Nao	Nao	Sim	386.7	289.1	555.16	455.38
Sim	Sim	Nao	90.6	19.4	4.76	3.62
Nao	Sim	Nao	124.2	26.6	142.16	42.38
Sim	Nao	Nao	47.3	118.5	0.24	1.38
Nao	Nao	Nao	64.9	162.5	179.84	279.62

Solução. (a) Observamos que para o modelo de homogeneidade marginal, representado por (AC, AM, CM), temos $G^2 = 0,374$ com 1 g.l. ($p =$ 0,541). Adotando um nível de significância de 5%, concluímos que o ajuste por este modelo é satisfatório.

(b) Apresentamos estimativas das razões de chances θ. Para $M \in \{1, 2\}$,

$$\widehat{\theta}_{AC} = \frac{910,38 \times 1,38}{3,62 \times 44,62} = 7,78.$$

Tanto na presença quanto na ausência da substância M, a chance de A estar presente quando C está presente é estimada como 7,8 vezes a chance de A estar presente quando C não está presente[1].

Para $C \in \{1, 2\}$,

$$\widehat{\theta}_{AM} = \frac{910,38 \times 42,38}{3,62 \times 538,62} = 19,79.$$

Para $A \in \{1, 2\}$,

$$\widehat{\theta}_{CM} = \frac{910,38 \times 455,38}{44,62 \times 538,62} = 17,25.$$

[1]Procure escrever frases semelhantes para os outros dois pares

4. Em uma tabela de contingências 4×4 para dados pareados as frequências são $n_{ii} = 5$ para todo i, $n_{i,i+1} = 15$ para $i = 1, 2, 3$, $n_{41} = 15$ e $n_{ij} = 0$ para os demais i e j. Afirma-se que a associação entre as variáveis é forte, mas a concordância não é forte. Apresente valores de medidas que justifiquem esta afirmação.

Solução. A tabela é representada abaixo.

```
-----------------
            Y
X   --------------
     1   2  3  4
-----------------
1    5  15   0   0
2    0   5  15   0
3    0   0   5  15
4   15   0   0   5
-----------------
```

Nesta tabela X e Y têm a mesma distribuição marginal (homogeneidade marginal) com probabilidades $(1/4, 1/4, 1/4, 1/4)$. Supondo independência entre X e Y e levando em conta que $n = 80$, as frequências esperadas são dadas por $\frac{1}{4} \times \frac{1}{4} \times 80 = 5$. A estatística X^2 de Pearson é calculada como

$$X^2 = \sum_{i=1}^{4} \sum_{j=1}^{4} \frac{(n_{ij} - 5)^2}{5} = \frac{1}{5} \sum_{i=1}^{4} \sum_{j=1}^{4} (n_{ij} - 5)^2 = 120.$$

Como os graus de liberdade são $3 \times 3 = 9$, o valor de X^2 é alto e a hipótese de independência é rejeitada, indicando associação entre X e Y (de fato, o valor-p é da ordem de 10^{-21}).

Notando que na tabela as estimativas de diversas razões de chance são infinitas, isto indica uma forte associação entre X e Y.

Em relação à concordância, a probabilidade de concordância observada é estimada como sendo $4 \times \frac{1}{16} = 1/4$. Por outro lado, a probabilidade de concordância ao acaso (na situação de independência) tem estimativa $4 \times \frac{1}{4} \times \frac{1}{4} = 1/4$. Sendo assim, a estimativa da medida de concordância κ é $\widehat{\kappa} = 0$, indicando fraca concordância entre X e Y.