USP – ICMC – SME0811 - Análise de Dados Categorizados

1ª prova – 1º/2018 – S O L U Ç Ã O

1. As variáveis X e Y são categóricas, ambas com dois níveis. Prove que se X e Y são independentes, então (i) a diferença de proporções é igual a 0, (ii) o risco relativo é igual a 1 e (iii) a razão de chances é igual a 1.

   Solução. Denotamos $\pi_{ij} = P(X = i, Y = j)$. Independência entre X e Y significa que $\pi_{ij} = \pi_{i+}\,\pi_{+j}$, para $i, j = 1, 2$. Calculamos
   (i) $D_{12} = \pi_{1|1} - \pi_{1|2} = \dfrac{\pi_{11}}{\pi_{1+}} - \dfrac{\pi_{21}}{\pi_{2+}} \overset{\text{indep.}}{=} \dfrac{\pi_{1+}\,\pi_{+1}}{\pi_{1+}} - \dfrac{\pi_{2+}\,\pi_{+1}}{\pi_{2+}} = \pi_{+1} - \pi_{+1} = 0.$
   De outra forma, como X e Y são independentes, as distribuições condicionais são as mesmas de modo que $\pi_{1|1} - \pi_{1|2} = 0$.
   (ii) $RR_{12} = \pi_{1|1}/\pi_{1|2} = \dfrac{\pi_{11}/\pi_{1+}}{\pi_{21}/\pi_{2+}} = \dfrac{\pi_{11}\,\pi_{2+}}{\pi_{1+}\pi_{21}} \overset{\text{indep.}}{=} \dfrac{\pi_{1+}\,\pi_{+1}\,\pi_{2+}}{\pi_{1+}\,\pi_{2+}\,\pi_{+1}} = 1.$
   De outra forma, como X e Y são independentes, as distribuições condicionais são as mesmas de modo que $\pi_{1|1}/\pi_{1|2} = 1$.
   (iii) $RC_{12} = \dfrac{\pi_{11}\,\pi_{22}}{\pi_{12}\,\pi_{21}} \overset{\text{indep.}}{=} \dfrac{\pi_{1+}\,\pi_{+1}\,\pi_{2+}\,\pi_{+2}}{\pi_{1+}\,\pi_{+2}\;\pi_{2+}\,\pi_{+1}} = 1.$
   De outra forma, como X e Y são independentes, as distribuições condicionais são as mesmas de modo que $\{\pi_{1|1}/(1 - \pi_{1|1})\}/\{\pi_{1|2}/(1 - \pi_{1|2})\} = 1$.

2. A Tabela 1 refere-se a um ensaio clínico com 457 pacientes que sofreram trauma provocado por hemorragia cerebral. Os pacientes foram avaliados após receberem um tratamento. Classifique as variáveis e apresente a estimativa de uma medida de associação que você considerar adequada. A associação entre as variáveis é forte?

   Solução. Pelo enunciado, vemos que "Tratamento" é a variável explicativa e "Avaliação" é a variável resposta. As duas variáveis são ordinais e estão apresentadas na Tabela 1 com níveis em ordem crescente.
   Como as variáveis são ordinais, a medida de associação γ pode ser utilizada. Primeiro calculamos o número de pares concordantes (C), dado por
   $C = 21 \times (54 + \cdots + 49 + \cdots + 41) + 44 \times (64 + 31 + 58 + 41) + 47 \times (31 + 41) + 14 \times (49 + 58 + 41) + 54 \times (58 + 41) + 64 \times 41 = 28199$,
   e o número de pares discordantes (D), dado por
   $D = 44 \times (14 + 4) + 47 \times (14 + 54 + 4 + 59) + 30 \times (14 + \cdots + 4 + \cdots + 58) + 54 \times 4 + 64 \times (4 + 49) + 31 \times (4 + 49 + 58) = 20818$.
   A estimativa é $\widehat{\gamma} = (C - D)/(C + D) = 0,151$. A associação é positiva (uma dosagem mais alta tende a levar a uma melhor recuperação) e fraca.

3. Os dados de um ensaio clínico com 103 crianças sobre o papel preventivo da vitamina C sobre resfriados encontram-se na Tabela 2.
   Utilizando as comparações que você considerar apropriadas para este estudo e com base em estimativas pontuais, apresente suas conclusões e interpretações.

   Solução. O estudo é um ensaio clínico. Portanto, as três formas de comparar proporções vistas na questão 1 podem ser utilizadas. Com os dados da Tabela 2, obtemos a distribuição condicional de Y (indicador de resfriado; $Y = 1$: sem) para (i) $X = 1$ (vitamina C) com probabilidades $p_{1|1} = 21/57 = 0,368$ e $p_{2|1} = 1 - p_{1|1} = 0,632$ e (ii) $X = 2$ (placebo) com probabilidades $p_{1|2} = 11/46 = 0,239$ e $p_{2|2} = 1 - p_{1|2} = 0,761$. As estimativas são $\widehat{D}_{12} = p_{1|1} - p_{1|2} = 0,129$, $\widehat{RR}_{12} = p_{1|1}/p_{1|2} = 1,541$ e $\widehat{RC}_{12} = n_{11}n_{22}/(n_{12}n_{21}) = 21 \times 35/(11 \times 36) = 1,856$. Como $\widehat{D}_{12} > 0$, $\widehat{RR}_{12} > 1$ e $\widehat{RC}_{12} > 1$, os dados indicam que o uso de vitamina C tem efeito preventivo sobre a ocorrência de resfriados. Com base na estimativa do risco relativo, para quem usa vitamina C, a probabilidade de não se resfriar é cerca de 50% maior em relação a quem não usa.

4. Em uma tabela $I \times 2$ com probabilidades $(\pi_{1|i}, \pi_{2|i} = 1 - \pi_{1|i})$ e contagens (n_{i1}, n_{i2}), para $i = 1, \ldots, I$, suponha que as linhas são independentes. Afirma-se que a função verossimilhança é proporcional a $\prod_{i=1}^{I} \pi_{1|i}^{n_{i1}}(1 - \pi_{1|i})^{n_{i+} - n_{i1}}$. Justifique esta afirmativa, indicando também a distribuição das contagens em cada linha.

   Solução. O número de níveis da variável Y é 2. Portanto, para a contagem n_{i1}, a distribuição binomial com parâmetros n_{i+} e $\pi_{1|i}$ pode ser adotada. Tomando n_{i1} no conjunto $\{0, 1, \ldots, n_{i+}\}$, a probabilidade é dada por $\binom{n_{i+}}{n_{i1}} \pi_{1|i}^{n_{i1}}(1 - \pi_{1|i})^{n_{i+} - n_{i1}}$, para $i = 1, \ldots, I$. Como as linhas da tabela são independentes, o modelo probabilístico resultante é o produto de binomiais, com função verossimilhança $L = \prod_{i=1}^{I} \binom{n_{i+}}{n_{i1}} \pi_{1|i}^{n_{i1}}(1 - \pi_{1|i})^{n_{i+} - n_{i1}} = \prod_{i=1}^{I} \binom{n_{i+}}{n_{i1}} \prod_{i=1}^{I} \pi_{1|i}^{n_{i1}}(1 - \pi_{1|i})^{n_{i+} - n_{i1}} \propto \prod_{i=1}^{I} \pi_{1|i}^{n_{i1}}(1 - \pi_{1|i})^{n_{i+} - n_{i1}}$, sendo que a constante de proporcionalidade $(= \prod_{i=1}^{I} \binom{n_{i+}}{n_{i1}})$ não envolve as probabilidades $\pi_{1|i}$.

Tabela 1: Dados para a questão 2.

	Avaliação				
Tratamento	Estado vegetativo	Incapacidade severa	Incapacidade leve	Boa recuperação	Total
Dosagem baixa	21	44	47	30	142
Dosagem média	14	54	64	31	163
Dosagem alta	4	49	58	41	152

Tabela 2: Dados para a questão 3.

	Resfriado		
Grupo	Sem	Com	Total
Vitamina C	21	36	57
Placebo	11	35	46