Testes de hipótese para tabelas de contingência: parte 3 (mais sobre testes de independência/homogeneidade e medidas de concordância)

Prof. Caio Azevedo

Exemplo 7: efeitos de certos fatores na sobrevivência de recém nascidos

- Os dados em questão foram extraídos de Agresti (1990) e estão relacionados à um grupo de gestantes fumantes classificadas segundo alguns fatores:
  - idade: < 30 ou 30 ou $+$.
  - número de cigarros consumidos por dia: < 5 ou 5 ou $+$.
  - tempo de gestação: ≤ 260 dias ou > 260 dias.
- Objetivo: saber se existe relação entre quantidade de cigarros consumidos e a sobrevivência do recém nascido em função das combinações dos níveis dos outros fatores: idade e tempo de gestação.

Dados observados

idade	Duração da gestação	N. de cigarros	Sobrevivência		
			Não	Sim	Total
<30	≤ 260	< 5	50	315	365
		$5+$	9	40	49
	>260	< 5	24	4012	4036
		$5+$	6	459	465
$30+$	≤ 260	< 5	41	147	188
		$5+$	4	11	15
	>260	<5	14	1594	1608
		$5 +$	1	124	125

Comentários

- Vamos considerar que cada um dos totais (por linha) foi fixado. Assim teremos um produto de binomiais.
- Queremos avaliar se existe dependência entre número de cigarros consumidos e a sobrevivência do recém-nascido em função dos outros fatores.
- Por enquanto, vamos nos concentrar apenas no fator idade.

Dados observados desconsiderando-se os fatores

N. de cigarros	Sobrevivência		
	Não	Sim	Total
< 5	129	6068	6197
5 ou $+$	20	634	654

Teste de qui-quadrado para independência: $q_h = 2,212$, p-valor $= 0,1369$. Assim, não rejeitamos a hipótese de independência entre a quantidade de cigarros e a sobrevivência do recém nascido.

Dados observados (considerando apenas o fator idade)

idade	N. de cigarros	Sobrevivência		
		Não	Sim	Total
<30	< 5	74	4327	4401
	5+	15	499	514
30+	< 5	55	1741	1796
	5+	5	135	140

Comentários

- Perguntas:
  - As variáveis "n. de cigarros" e sobrevivência são dependentes em (dentro de) cada estrato?
  - As " estruturas de dependências" são do mesmo tipo?
- Podemos realizar testes de independência para cada estrato (formado pelos dois grupos de idade), usando o teste de qui-quadrado ou o teste para a razão de chances.
- Podemos também quantificar a dependência utilizando as medidas de associação vistas anteriormente.

Relembrando (para cada nível da variável idade)

		Sobrevivência		
		Não	Sim	Total
n. de cigarros	< 5	$N_{11}(\theta_{11})$	$N_{12}(\theta_{12})$	$n_{1.}$
	$5+$	$N_{21}(\theta_{21})$	$N_{22}(\theta_{22})$	$n_{2.}$
Total	-	$n_{.1}$	$n_{.2}$	$n_{..}$

Para cada estrato, queremos testar se $H_0 : \theta_{11} = \theta_{21}$ vs $H_1 : \theta_{11} \neq \theta_{21}$, ou de modo equivalente, se $H_0 : \pi = 1$ vs $H_1 : \pi \neq 1$, em que $\pi = \dfrac{\frac{\theta_{11}}{1-\theta_{11}}}{\frac{\theta_{21}}{1-\theta_{21}}}$ é a razão de chances.

Relembrando

- Temos, sob H_0, que: $N_{11} \sim \text{binomial}(n_{1.}, \theta)$ e $N_{21} \sim \text{binomial}(n_{2.}, \theta)$, em que $N_{11} \perp N_{21}$.
- Além disso, sob H_0, $Z = N_{11} + N_{21} \sim \text{binomial}(n_{1.} + n_{2.}, \theta)$.
- Assim, sob H_0, $N_{11}|Z = z \sim \text{hipergeométrica}(n_{..}, n_{1.}, z = n_{.1})$.
- Dessa forma, sob H_0, $\mathcal{E}(N_{11}|Z = z) = \dfrac{n_{1.}z}{n_{..}} = \mathcal{E}_{\pi_1}$ e $\mathcal{V}(N_{11}|Z = z) = \dfrac{n_{1.}n_{2.}z(n_{..} - z)}{n_{..}^2(n_{..} - 1)} = \mathcal{V}_{\pi_1}$.

Teste de Mantel Haneszel (MH) (para cada estrato)

- Pode-se demonstrar, que se $Q_{MH} = \frac{\left(N_{11} - \mathcal{E}_{\pi_1}\right)^2}{\mathcal{V}_{\pi_1}}$, então $Q_{MH} \approx \chi^2_1$, sob H_0, para $n_{1.}$, $n_{2.}$, $n_{.1}$ e $n_{.2}$ suficientemente grandes.
- Idade: < 30
  - Teste de MH: $q_{MH} = 3,96$ (p-valor = 0,0466).
  - Teste de qui-quadrado: $q_H = 3,29$ (p-valor = 0,0695).
- Idade: 30 ou $+$.
  - Teste de MH: $q_{MH} = 0,01$ (p-valor = 0,9163)
  - Teste de qui-quadrado: $q_H = 0,01$ (p-valor = 0,9350)

Generalizando (para cada nível da variável idade)

		Sobrevivência		
		Não	Sim	Total
n. de cigarros	< 5	$N_{g11}(\theta_{g11})$	$N_{g12}(\theta_{g12})$	$n_{g1.}$
	$5+$	$N_{g21}(\theta_{g21})$	$N_{g22}(\theta_{g22})$	$n_{g2.}$
Total	-	$n_{g.1}$	$n_{g.2}$	$n_{g..}$

Para $g = 1, 2$; 1 (idade < 30), 2 (idade 30 ou $+$).

Teste de Mantel Haenszel para G estratos

- Considere que tenhamos $g = 1, ..., G$ estratos.
- Queremos testar se

  $H_0 : \pi_1 = \pi_2 = ... = \pi_G = 1$ vs H_1 : há pelo menos uma diferença.
- A estatística de Mantel Haenszel para testar tais hipótses é dada por:

$$Q_{MH} = \frac{\left(\sum_{g=1}^{G} N_{g11} - \sum_{g=1}^{G} \mathcal{E}_{\pi_{g1}}\right)^2}{\sum_{g=1}^{G} \mathcal{V}_{\pi_{g1}}}$$

Teste de Mantel Haenszel para G estratos

- Analogamente ao caso de um único estrato, temos que, sob H_0,

$$\mathcal{E}(N_{g11}|Z_g = z_g) = \frac{n_{g1.}z_g}{n_{g..}} = \mathcal{E}_{\pi_{g1}}$$

e

$$\mathcal{V}(N_{g11}|Z_g = z_g) = \frac{n_{g1.}n_{g2.}z_g(n_{g..} - z_g)}{n_{g..}^2(n_{g..} - 1)} = \mathcal{V}_{\pi_{g1}}.$$

- Além disso, sob H_0, $Q_{MH} \approx \chi_1^2$ para $n_{g1.}$, $n_{g2.}$, $n_{g.1}$ e $n_{g.2}, g = 1, ..., G$, suficientemente grandes.

Estimador de Mantel Haenszel para a razão de chances comum

- O estimador (de Mantel Haenszel) da razão de chances comum entre os G estratos é dado por:

$$\widehat{\pi}_{MH} = \frac{\sum_{g=1}^{G} N_g \widehat{\pi}_g}{\sum_{g=1}^{G} N_g}$$

em que $N_g = \dfrac{N_{g21}\left(n_{g1.} - N_{g11}\right)}{n_{g..}}$.

- A variância assintótica de $\widehat{\pi}_{MH}$ é dada por

$$\mathcal{V}_A(\widehat{\pi}_{MH}) = \pi^2 \frac{\sum_{g=1}^{G} \omega_g^{-1} a_g}{\left(\sum_{g=1}^{G} a_g\right)^2}$$

Estimador de Mantel Haenszel para a razão de chances comum

- em que π é a "verdadeira" razão de chances comum,
  $\omega_g = (n_{g1.}\theta_{g11}(1-\theta_{g11}))^{-1} + (n_{g2.}\theta_{g21}(1-\theta_{g21}))^{-1}$ e
  $a_g = \frac{n_{g1.}n_{g2.}(1-\theta_{g11})\theta_{g21}}{n_{g1.}}$, g=1,...,G.
- Para $n_{g1.}$, $n_{g2.}$, $n_{g.1}$ e $n_{g.2}$, suficientemente grandes, $g = 1, ..., G$, temos que $\widehat{\pi}_{MH} \approx N(\mathcal{E}(\widehat{\pi}_{MH}) = \pi, \mathcal{V}_A(\widehat{\pi}_{MH}))$.

Voltando ao exemplo em questão, considerando apenas os grupos definidos pela idade

- Resultados para o logaritmo natural (η) das razões de chances.

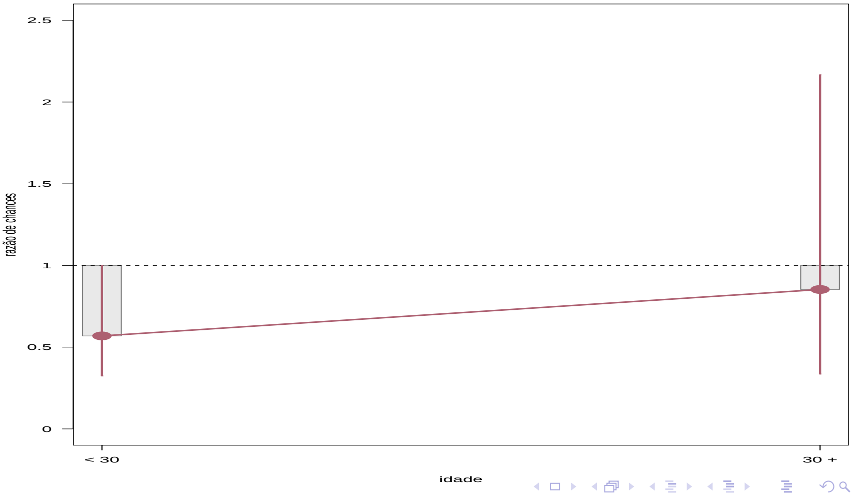
grupo	Estimativa	EP	Estat. do test	p-valor
< 30	-0,564	0,283	-1,992	0,0232
$30 +$	-0,159	0,456	-0,349	0,3646

- Resultados para as razões de chances.

grupo	Estimativa	IC (95%)
< 30	0,569	[0,327 ; 0,991]
$30 +$	0,853	[0,349 ; 2,084]

- Teste de MH : $q_{MH} = 2,835$ e $p - valor = 0,0922$.

Razões de chance estimadas e respectivos IC (95%)



Considerando os quatro grupos: idade $\times$ duração da gestação

- Resultados para o logaritmo natural (η) das razões de chances.

grupo	Estimativa	EP	Estat. do test	p-valor
< 30 & $<= 260$	-0,349	0,391	-0,892	0,186
< 30 & > 260	-0,782	0,444	-1,761	0,039
$30+$ & $<=260$	-0,265	0,583	-0,455	0,325
$30+$ & > 260	0,085	0,863	0,099	0,461

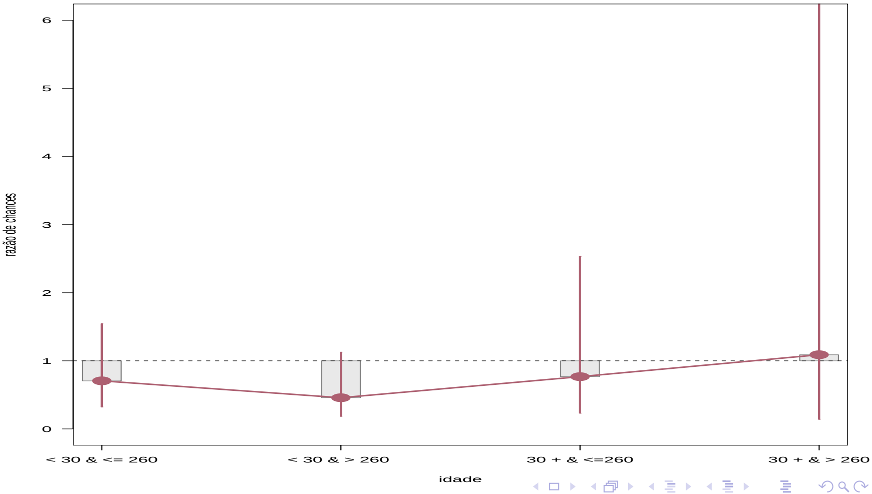
Considerando os quatro grupos: idade $\times$ duração da gestação

- Resultados para as razões de chances.

grupo	Estimativa	IC (95%)
< 30 & $<= 260$	0,705	[0,328 ; 1,518]
< 30 & > 260	0,458	[0,192 ; 1,092]
$30+$ & $<=260$	0,767	[0,245 ; 2,405]
$30+$ & > 260	1,089	[0,201 ; 5,908]

- Teste de MH : $q_{MH} = 2,191$ e $p - valor = 0,1388$.

Razões de chance estimadas e respectivos IC (95%)



Comentários

- Em ambos os casos: dois grupos (definidos pela idade) e quatro grupos (definidos pela idade com a duração da gestação) os testes não concordam entre si.
- Aparentemente, há associação entre o número de cigarros consumidos e a sobrevivência do recém-nascido para o estrato idade < 30 (no primeiro caso) e para o estrato idade < 30 e duração da gestação > 260 (no segundo caso).
- Utilizar outras técnicas para dirimir esta (aparente) dúvida: refazer os testes através de simulações, inferência bayesiana, **modelos de regressão para dados binomiais (regressão logística)** .

Voltando ao Exemplo 1: comparação de métodos de detecção de cárie

		Risco de cárie segundo o método convencional			
		Baixo	Médio	Alto	Total
Risco de cárie segundo o método simplificado	Baixo	11	5	0	16
	Médio	14	34	7	55
	Alto	2	13	11	26
Total	-	27	52	18	97

Queremos verificar o grau de concordância entre os métodos.

Relembrando

		Risco de cárie segundo o método convencional			
		Baixo	Médio	Alto	Total
Risco de cárie segundo o método simplificado	Baixo	θ_{11}	θ_{12}	θ_{13}	$\theta_{1.}$
	Médio	θ_{21}	θ_{22}	θ_{23}	$\theta_{2.}$
	Alto	θ_{31}	θ_{32}	θ_{33}	$\theta_{3.}$
Total	-	$\theta_{.1}$	$\theta_{.2}$	$\theta_{.3}$	$\theta_{..} = 1$

- $\theta_{i.} = \sum_{j=1}^{3} \theta_{ij}, i = 1, 2, 3$, $\theta_{.j} = \sum_{i=1}^{3} \theta_{ij}, j = 1, 2, 3$ e $\theta_{..} = \sum_{i=1}^{3} \sum_{j=1}^{3} \theta_{ij} = 1.$

Coeficiente de concordância κ de Cohen

- A idéia do coeficiente κ, diferentemente do coeficiente τ de Kendall, é comparar a proporção (total) de concordância conjunta com a proporção (total) esperada de concordância conjunta.
- Sejam:
  - (Proporção (total) de concordância conjunta): $\theta_o = \sum_{i=1}^{r} \theta_{ii}$, em que r é número de linhas/colunas.
  - (Proporção (total) esperada de concordância sob independência): $\theta_e = \sum_{i=1}^{r} \theta_{i.}\theta_{.i}$.
- O coeficiente de concordância κ de Cohen é definido como $\kappa = \dfrac{\theta_o - \theta_e}{1 - \theta_e}$.

Concordância absoluta

		Risco de cárie segundo o método convencional			
		Baixo	Médio	Alto	Total
Risco de cárie segundo o método simplificado	Baixo	θ_{11}	0	0	$\theta_{1.}$
	Médio	0	θ_{22}	0	$\theta_{2.}$
	Alto	0	0	θ_{33}	$\theta_{3.}$
Total	-	$\theta_{.1}$	$\theta_{.2}$	$\theta_{.3}$	$\theta_{..} = 1$

- Ou seja $\theta_{i.} = \theta_{.j} = \theta_{ij}, \forall i = j$ (concordância absoluta).
- Claramente, a independência é incompatível com essa situação.

Coeficiente de concordância κ de Cohen

- Lembrando: o coeficiente κ mensura a concordância conjunta (e não a concordância marginal).
- Porque comparar as proporções observadas de concordância conjunta com as proporções esperadas de concordância conjunta sob independência? Já vimos que a independência e a concordância plena são incompatíveis.
- Quanto menor (maior) for a quantidade de pares concordantes menor (maior) será o coeficiente κ (podendo ser negativo).
- Assim, quanto maior for a concordância entre os métodos, esperamos que θ_o seja não apenas diferente, mas sim, maior do que do que θ_e.

Estimadores associados ao Coeficiente de concordância κ

- Estimador : $\widehat{\kappa} = \frac{\widehat{\theta}_o - \widehat{\theta}_e}{1 - \widehat{\theta}_e}$, em que $\widehat{\theta}_o = \sum_{i=1}^{r} \widehat{\theta}_{ii}$ e $\widehat{\theta}_e = \sum_{i=1}^{r} \widehat{\theta}_{i.}\widehat{\theta}_{.i}$.
- Em que $\widehat{\theta}_{ii}, \widehat{\theta}_{i.}$ e $\widehat{\theta}_{.i}$ são os respectivos estimadores de MV, ou seja $\widehat{\theta}_{ii} = \frac{N_{ii}}{n_{..}}, \widehat{\theta}_{i.} = \frac{N_{i.}}{n_{..}}, \widehat{\theta}_{.i} = \frac{N_{.i}}{n_{..}}$.
- Estimador do erro-padrão assintótico de $\widehat{\kappa}$: $\widehat{EP}(\widehat{\kappa}) = \frac{\sqrt{A+B-C}}{\left(1-\widehat{\theta}_e\right)\sqrt{n_{..}}}$, em que

$$
\begin{aligned}
A &= \sum_{i=1}^{r} \widehat{\theta}_{ii} \left[1 - (\widehat{\theta}_{i.} + \widehat{\theta}_{.i})(1 - \widehat{\kappa})\right]^2 \\
B &= (1 - \widehat{\kappa})^2 \sum_{i=1}^{r} \sum_{j \neq i} \widehat{\theta}_{ij}(\theta_{.i} + \theta_{j.})^2 \\
C &= \left[\widehat{\kappa} - \widehat{\theta}_e(1 - \widehat{\kappa})\right]^2
\end{aligned}
$$

Coeficiente de concordância κ ponderado

- O coeficiente κ, da forma como o apresentamos, é mais apropriado quando as categorias são nominais. Além disso, ele não leva em consideração pares discordantes.
- Quando as categorias são ordinais, é importante considerar alguma possível "distância" entre elas.
- Nesse caso, é mais apropriado utilizar o coeficiente de concordância κ ponderado (κ_w).
- Pesos lineares (ou "igualmente espaçados"): $w_{ij} = 1 - \frac{|i-j|}{r-1}$.
- Pesos quadráticos (ou de "Fleiss-Cohen"): $w_{ij} = 1 - \frac{(i-j)^2}{(r-1)^2}$.

Exemplos de valores para os pesos: tabela $3\times$ 3

- Pesos lineares

$$\begin{bmatrix} 1 & 0,5 & 0 \\ 0,5 & 1 & 0,5 \\ 0 & 0,5 & 1 \end{bmatrix}.$$

- Pesos quadráticos

$$\begin{bmatrix} 1 & 0,5 & 0 \\ 0,5 & 1 & 0,75 \\ 0 & 0,75 & 1 \end{bmatrix}.$$

Estimadores associados ao coeficiente de concordância κ_w

- Estimador : $\widehat{\kappa}_w = \dfrac{\widehat{\theta}_{o_w} - \widehat{\theta}_{e_w}}{1 - \widehat{\theta}_{e_w}}$, em que $\widehat{\theta}_{o_w} = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\widehat{\theta}_{ij}$ e $\widehat{\theta}_{e_w} = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij}\widehat{\theta}_{i.}\widehat{\theta}_{.j}$.
- Estimador do erro-padrão assintótico de $\widehat{\kappa}_w$: $\widehat{EP}(\widehat{\kappa}_w) = \frac{\sqrt{A-B}}{\left(1-\widehat{\theta}_e\right)\sqrt{n_{..}}}$, em que

$$
\begin{aligned}
A &= \sum_{i=1}^{r}\sum_{j=1}^{r}\widehat{\theta}_{ij}\left[w_{ij} - (w_{i.} + w_{.j})(1-\widehat{\kappa}_w)\right]^2 \\
B &= \left[\widehat{\kappa}_w - \widehat{\theta}_{e_w}(1-\widehat{\kappa}_w)\right]^2
\end{aligned}
$$

e $w_{i.} = \sum_{j=1}^{r} w_{ij}\widehat{\theta}_{.j}$ e $w_{.j} = \sum_{i=1}^{r} w_{ij}\widehat{\theta}_{i.}$.

- Note que, diferentemente do coeficiente κ, κ_w leva em consideração mais classificações (concordantes e algumas discordantes).

Inferência

- Temos, para $n_{ij} \forall i, j$, suficientemente grandes, que as distribuições tanto $\widehat{\kappa}$ quanto $\widehat{\kappa}_w$ aproximam-se de normais com suas respectivas médias (valores verdadeiros) e variâncias (como apresentadas anteriormente).
- Assim, podemos testar hipóteses de interesse e construir intervalos de confiança, ambos assintóticos.
- Inferências com base em reamostragem, também são possíveis.
- Os coeficientes κ e κ_w deve ser utilizados em tabelas quadradas, somente.

Inferência

- De acordo com Landis and Koch (1977), podemos classificar os valores dos coeficiente κ e κ_w, de acordo com

Valor	Magnitude da concordância
$<$ 0,0	nula (aleatório)
0,00 - 0,20	muito leve
0,21 - 0,40	leve
0,41 - 0,60	moderada
0,61 - 0,80	substancial
0,81 - 1,00	quase perfeita

Resultados

- Resultados

Coeficiente	Estimativa	EP	IC (95 %)
κ	0,296	0,083	[0,134; 0,458]
κ_w (pesos lineares)	0,372	0,077	[0,221 ; 0,522]
κ_w (pesos quadráticos)	0,471	0,077	[0,320 ; 0,623]

Aparentemente, há uma concordância de leve a moderada entre os métodos (no sentido postulado pelos coeficientes κ e κ_w).