Correlação Tetracórica

Antonio Eduardo Gomes
aegomes@unb.br

A correlação tetracórica é uma medida de correlação entre duas variáveis dicotômicas. Ela assume que as variáveis são contínuas, mas que foram medidas de maneira binária.

Dados dois itens (1 e 2), suponha que os eventos $\{U_1 = 0\}$ e $\{U_2 = 0\}$ sejam equivalentes, respectivamente, aos eventos $\{X \leq x_0\}$ e $\{Y \leq y_0\}$, sendo X e Y variáveis aleatórias contínuas tais que $(X, Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_X^2, \rho)$.

Suponha que observamos as frequências conjuntas de U_1 e U_2 dadas na tabela de contingência abaixo.

Table:

$U_1 \backslash U_2$	0	1	Total
0	n_{00}	n_{01}	$n_{00} + n_{01}$
1	n_{10}	n_{11}	$n_{10} + n_{11}$
Total	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n

Definindo $\gamma = (x_0 - \mu)/\sigma_X$ e $\tau = (y_0 - \mu)/\sigma_Y$, estimamos γ e τ, respectivamente, por

$$\widehat{\gamma} = \Phi^{-1}\left(\frac{n_{00} + n_{01}}{n}\right) \quad \text{e} \quad \widehat{\tau} = \Phi^{-1}\left(\frac{n_{00} + n_{10}}{n}\right).$$

Como $((X-\mu)/\sigma_X, (Y-\mu)/\sigma_Y) \sim N_2(0,0,1,1,\rho)$, estimamos ρ através de $\widehat{\rho}$ que satisfaz a relação

$$\frac{n_{11}}{n} = \int_{\widehat{\gamma}}^{\infty} \int_{\widehat{\tau}}^{\infty} \frac{1}{2\pi\sqrt{(1-\widehat{\rho}^2)}} \exp\left[\frac{-(x^2 - 2\widehat{\rho}xy + y^2)}{2(1-\widehat{\rho}^2)}\right] dy\, dx\,.$$

O valor de $\widehat{\rho}$ pode ser aproximado pelo coeficiente de associação de Yule:

$$Q_1 = \frac{n_{00}n_{11} - n_{01}n_{10}}{n_{00}n_{11} + n_{01}n_{10}}\,.$$

Exemplo: Considere os dados da tabela abaixo:

Table:

$U_1 \backslash U_2$	0	1	Total
0	11	29	40
1	9	1	10
Total	20	30	50

As estimativas via máxima verossimilhança foram $\widehat{\rho} = -0,8075$, $\widehat{\gamma} = 0,8416$ e $\widehat{\tau} = -0,2533$.
Para os dados acima, temos $Q_1 = -0,9191$.