Notes on sample sizes required for Receiver Operator Characteristic (ROC) curves

13 November 2018

1 Area under the curve of ROC

$$W = \frac{1}{n_A \times n_N} \sum_{1}^{n_A} \sum_{1}^{n_N} S(\chi_A, \chi_N)$$

2 Standard error of the area under the curve of ROC

$$SE(W) = \sqrt{\frac{\theta(1 - \theta) + (n_A - 1)(Q_1 - \theta^2) + (n_N - 1)(Q_2 - \theta^2)}{n_A \times n_N}}$$

where

$$\theta = \text{area under the curve}$$

$$n_A = \text{number of cases}$$

$$n_N = \text{number of non-cases}$$

$$Q_1 = \frac{\theta}{2 - \theta}$$

$$Q_2 = \frac{2\theta^2}{1 + \theta}$$

3 Sample size given a specified confidence interval

$$N \;=\; \frac{Z^2_{\alpha/2} \;\times\; VF}{L^2}$$

$$where:$$

$$Z_{\frac{\alpha}{2}} \;=\; \text{1.96 for a 95\% CI}$$

$$L \;= \text{0.05 (desired half-width of the CI)}$$

$$VF \;=\; \text{variance function}$$

$$N \;=\; \text{number of cases}$$

The variance function (VF) is calculated as follows:

$$VF \;=\; 0.0099 \;\times\; e^{-A \;\times\; A/2} \;\times\; \left[(5 \;\times\; A^2 \;+\; 8) \;+\; \frac{(A^2 \;+\; 8)}{k} \right]$$

$$where:$$

$$\begin{aligned} A \;&=\; \text{parameter from the binomial distribution} \\ &=\; \phi^{-1}(AUC) \;\times\; 1.414 \end{aligned}$$

4 Sample size for detecting differences between areas under two ROC curves

$$n = \left[\frac{Z_\alpha \sqrt{2V_1} + Z_\beta \sqrt{V_1 + V_2}}{\delta} \right]^2$$

where

$$Z_\alpha = 1.645 \text{ for a 5\% one-sided test of significance}$$

$$Z_\beta = 0.84, 1.28, \text{ or } 1.645 \text{ for 80\%, 90\%, or 95\% power}$$

$$\delta = \theta_2 - \theta_1$$

$$V_1 = Q_1 + Q_2 - 2\theta_1^2 \text{ where } Q_1 = \frac{\theta_1}{2 - \theta_1} \text{ and } Q_2 = \frac{2\theta_1^2}{1 + \theta_1}$$

$$V_2 = Q_1 + Q_2 - 2\theta_2^2 \text{ where } Q_1 = \frac{\theta_2}{2 - \theta_2} \text{ and } Q_2 = \frac{2\theta_2^2}{1 + \theta_2}$$