Details on the Sample size calculator for logistic regression

We require estimates of the probabilities of an event at the mean and mean + standard deviation of the predictor. Let these be p_m and p_s respectively. Hsieh (1989) derive the sample size to demonstrate the effect of the predictor on the event as the expression

$$n = \frac{\left[z_\alpha + z_\beta \exp(-\tau^2/4)\right]^2 (1 + 2p_m\delta)/(p_m\tau^2)}{(1 - R^2)}$$

Where z_α and z_β are the right tail quantiles of the standard normal for a false positive rate of α and a false negative rate of β,

τ is the log odds ratio given by $\ln\left[\frac{p_s/(1-p_s)}{p_m/(1-p_m)}\right]$,

$\delta = \frac{\left[1 + (1+\tau^2)\exp(5\tau^2/4)\right]}{\left[1 + \exp(-\tau^2/4)\right]}$, and

R is the multiple correlation coefficient between the predictor and other covariates in the model. Use $R = 0$ when there are no additional covariates being considered.

References:

1) Hsieh, F Y (1989) Sample size tables for logistic regression. Statistics in Medicine. 8:795-802.
2) Agresti, Alan (2002) Categorical Data Analysis. 2nd edition. John Wiley and Sons, Inc., Hoboken, New Jersey.

Return to Calculator at this link.