Methodology for Exact Confidence Interval for Risk Difference (SAS 9.2)

An exact $100(1-\alpha)\%$ confidence interval for the risk difference can be obtained in SAS 9.2 by specifying an EXACT RISKDIFF statement in PROC FREQ.

Denote the proportion difference by $d = p_1 - p_2$. For a 2×2 table with row totals n_1 and n_2, the joint probability function (product of two independent binomials) can be expressed in terms of the table cell frequencies and the parameters d and p_2 as follows:

$$f(n_{11}, n_{21}; n_1, n_2, d, p_2) = \begin{bmatrix} n_1 \\ n_{11} \end{bmatrix} (d + p_2)^{n_{11}} (1 - d - p_2)^{n_1 - n_{11}} \times \begin{bmatrix} n_2 \\ n_{21} \end{bmatrix} (p_2)^{n_{21}} (1 - p_2)^{n_2 - n_{21}}$$

When constructing confidence limits for the proportion difference, the parameter of interest is d and p_2 is a nuisance parameter.

Denote the observed value of the proportion difference by $d_0 = \hat{p}_1 - \hat{p}_2$
The $100(1-\alpha)\%$ confidence limits for d (denoted d_L and d_U) are computed as

$$d_L = \sup(d_*: P_U(d_*) > \alpha/2)$$
$$d_U = \inf(d_*: P_L(d_*) > \alpha/2)$$

Where

$$P_U(d_*) = \sup_{p_2}\left(\sum_{A,D(a)\geq d_0} f(n_{11}, n_{21}; n_1, n_2, d_*, p_2)\right)$$
$$P_L(d_*) = \sup_{p_2}\left(\sum_{A,D(a)\leq d_0} f(n_{11}, n_{21}; n_1, n_2, d_*, p_2)\right)$$

The set A includes all 2×2 tables with row sums equal to n_1 and n_2, and $D(a)$ denotes the value of the proportion difference $(p_1 - p_2)$ for table a in A. To compute $P_U(d_*)$, the sum includes probabilities of those tables for which $D(a) \geq d_0$, where d_0 is the observed value of the proportion difference. For a fixed value of d_*, $P_U(d_*)$ is taken to be the maximum sum over all possible value of p_2. Details can be found in Santner and Snell (1980) and Agresti and Min (2001).

Essentially, this method determines the greatest lower confidence limit and smallest upper confidence limit such that, with row totals fixed, the sum of tables probabilities where risk differences are as extreme or more extreme than the observed risk difference are no more than $\alpha/2$ in either direction.

The confidence limits are conservative for small samples because this is a discrete problem; the confidence coefficient is not exactly $1 - \alpha$ but is at least $1 - \alpha$ (Agresti, 1992).