Program SG_ASO.EXE and EQUIV_ASO.EXE

Considerer $(x_1, y_1, x_2, y_2) \sim M(n; p_{11}, p_{12}, p_{21}, p_{22})$ for data in Table below. The probability of an experimental result like the one in the Table 1 is:

$$P(x_1, y_1, x_2, y_2)=n!(x_1!y_1!x_2!y_2!)^{-1} p_{11}^{x_1} p_{12}^{y_1} p_{21}^{x_2} p_{22}^{y_2} . \qquad \textbf{(1)}$$

Table 1
CROSS-SECTIONAL STUDY
Presentation of results (probabilities) in a problem of comparison between two proportions (illness vs. risk factor) when only a sample of n observations exists (multinomial case)

		Illness		
		YES	*NO*	
Exposed to the	*NO*	x_1 (p_{11})	y_1 (p_{12})	n_1 (q)
risk factor	*YES*	x_2 (p_{21})	y_2 (p_{22})	n_2 (1–q)
		a_1	a_2	n (1)

Considerer $p_1 = p_{11} / (p_{11}+p_{12})$ and $p_2 = p_{21} / (p_{21}+p_{22})$, where p_2 (p_1) is the prevalence of an illness in the group of YES exposed (NOT exposed) to a risk factor. The aim is to perform inferences about $d = p_2-p_1$. So, a reparametrization of the model (1) can be performed on p_1, p_2 and $q=p_{11}+p_{12}$:

$$P(x_1, y_1, x_2, y_2) = \left[\binom{n}{n_1} q^{n_1}(1-q)^{n_2}\right] \times \left[\binom{n_1}{x_1}\binom{n_2}{x_2} p_1^{x_1}(1-p_1)^{y_1} p_2^{x_2}(1-p_2)^{y_2}\right], \qquad \textbf{(2)}$$

Under H_δ: $d=\delta$ (where $-1<\delta<+1$), if $p_1=p$ then $p_2=p+\delta$ and so (2) is:

$$P(x_1, y_1, x_2, y_2 \mid \delta) = \left[\binom{n}{n_1} q^{n_1}(1-q)^{n_2}\right] \times \left[\binom{n_1}{x_1}\binom{n_2}{x_2} p^{x_1}(1-p)^{y_1} (p+\delta)^{x_2}(1-p-\delta)^{y_2}\right],$$

where p and q are two nuisance parameters taking the values:

$$0 \le q \le 1, \ \max\{0; -\delta\} \le p \le \min\{1; 1-\delta\}.$$

For a critical region CR formed by a set of values (x_1, y_1, x_2, y_2), the error α of the test will be $\alpha(p, q|\delta) = \Sigma_{CR} P(x_1, y_1, x_2, y_2 \mid \delta)$, and the size of the test will be $\alpha^*(\delta) = \mathrm{Max}_{p,q}\ \alpha(p, q|\delta)$. There are several ways for obtaining the CR, but the one that provides the generally most powerful test is that based on the order given by the Z-pooled statistic with the Yates' continuity correction (Z_Y). The exact p-value $\alpha^* = \mathrm{Max}_{\delta\in H}\ \alpha^*(\delta)$ of the observed data (x_1, y_1, x_2, y_2) depends on the null hypothesis H and alternative hypothesis K to be demonstrated.

The asymptotic p-values (see expressions bellow) are based on the Z-pooled statistics with the Pirie and Hamdan's continuity correction (Z_{PH}) or without continuity correction (Z_0).

In both cases: if $\hat{d} = \hat{p}_2 - \hat{p}_1 \in$ H, where $\hat{p}_i = x_i / n_i$, then p-value = 1.

There are two programs:

Program SG_ASO.EXE

- Case **SG**: H: $d \leq \delta$ vs. K: $d > \delta$ (Superiority Generalized). In particular:
  - ∗ When $\delta = 0$: is the classic case **S** of Superiority.
  - ∗ When $\delta < 0$: is the case **NI** of Non-Inferiority.
  - ∗ When $\delta > 0$: is the case **SS** of Substantial-Superiority.
- Case **IG**: H: $d \geq \delta$ vs. K: $d < \delta$ (Inferiority Generalized).

Program EQUIV_ASO.EXE

- Case **SG2**: H: $d = \delta$ vs. K: $d \neq \delta$ (two-tailed SG).
- Case **PE**: H: $|d| \geq \Delta$ vs. K: $|d| < \Delta$ (Practice Equality or Equivalence) (Δ>0).
- Case **SD**: H: $|d| \leq \Delta$ vs. K: $|d| > \Delta$ (Substantially Difference) (Δ>0).

∗∗∗

The Z-pooled two-tailed statistic for H: $d=\delta$ is:

$$Z_c = \begin{cases} \dfrac{|\hat{d}-\delta| - c}{s(\delta)} \times f & \text{if } |\hat{d}-\delta| > c \\ 0 & \text{if } |\hat{d}-\delta| \leq c \end{cases} \quad \text{where } s(\delta) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{(\hat{p}+\delta)(1-\hat{p}-\delta)}{n_2}} \text{ and } f = \sqrt{\frac{n-1}{n}},$$

where $\hat{p}$ is the maximum likelihood estimator for p under H: $d=\delta$, and c is a continuity correction: $c = n/2n_1n_2$, $c = 1/2n_1n_2$ and $c = 0$ for Z_Y, Z_{PH} and Z_0 respectively. The asymptotic p-values are:

$$P_{SG} = F\left\{\frac{-\hat{d}+\delta+c}{s(\delta)} \times f\right\}, \quad P_{IG} = F\left\{\frac{\hat{d}-\delta+c}{s(\delta)} \times f\right\}$$

$$P_{PE}(\Delta) = \max_{\delta=-\Delta,\,+\Delta} \left[F\left\{\frac{|\hat{d}|+\Delta+c}{s(\delta)} \times f\right\} - F\left\{\frac{-|\hat{d}|+\Delta-c}{s(\delta)} \times f\right\} \right]$$

$$P_{SD}(\Delta) = F\left\{\frac{-\left|\hat{d}+\Delta\right|+c}{\max_{\delta=-\Delta,+\Delta} s(\delta)}\times f\right\} + F\left\{\frac{-\left|\hat{d}-\Delta\right|+c}{\max_{\delta=-\Delta,+\Delta} s(\delta)}\times f\right\}$$

$$P_{SG2}(\delta) = \begin{cases} 2\times F\left\{\frac{-\left|\hat{d}-\delta\right|+c}{s(\delta)}\times f\right\} & \text{if } 2\delta-1\leq \hat{d}\leq 2\delta+1 \\ F\left\{\frac{-\left|\hat{d}-\delta\right|+c}{s(\delta)}\times f\right\} & \text{other wise,} \end{cases}$$

where $F(\cdot)$ refers to the distribution function of a standard normal random variable *z*.

In the program:

- $\hat{d}$ = Delta (ML);
- $\hat{p}$ = p(ML);
- $\hat{q} = n_1/n$ = q(ML);

where ML ≡ maximum likelihood.

∗∗∗

For more details see:

Martín Andrés, A.; Tapia Garcia, J. M. and del Moral Ávila, M.J. (2005). Unconditional inferences on the difference of two proportions in cross-sectional studies. *Biometrical Journal 47 (2),* 177-187.

Martín Andrés, A.; Tapia Garcia, J. M. and Del Moral Ávila, M.J. (2008). Two-tailed unconditional inferences on the difference of two proportions in cross-sectional studies. *Communications in Statistic - Simulation and Computation 37 (3)*, 455-465.