Confidence Interval for Matched Pairs Difference of Proportions (Fleiss, 1981; McNemar 1947)

Consider the following motivating example:

A study was undertaken to evaluate whether a telephone interview to evaluate functional status of n patients with a stable chronic disorder agreed with the assessment at a clinic visit several days before. The data are as follows:

	Telephone Interview		
Clinical Visit	Mild/Moderate	Severe	
Mild/Moderate	n_{11}	n_{12}	$n_{11}+n_{12}$
Severe	n_{21}	n_{22}	$n_{21}+n_{22}$
	$n_{11}+n_{21}$	$n_{12}+n_{22}$	n

We are interested in calculating a $100*(1-\alpha)\%$ confidence interval for $p_{tel}-p_{clin}$, where p_{tel} is the proportion of patients indicating "severe" in the telephone interview, and p_{clin} is the proportion of patients indicating "severe" at the clinic visit. We can observe that estimates of these proportions are $\hat{p}_{tel}=\frac{n_{12}+n_{22}}{n}$ and $\hat{p}_{clin}=\frac{n_{21}+n_{22}}{n}$.

Since the same n individuals provide data for both the clinic visit and the telephone interview, the proportions indicating "severe" are correlated (i.e. not independent). Thus we must take this correlation into account when calculating the standard error for use in our confidence interval.

Derivation of standard error using a single multinomial distribution

Assume that n is fixed and that the distribution of individuals in the 4 cells is multinomial. That is, $(n_{11}, n_{12}, n_{21}, n_{22})$ has a multinomial distribution with probability mass function

$$f(n_{11}, n_{12}, n_{21}, n_{22}) = \frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!}(\pi_{11})^{n_{11}}(\pi_{12})^{n_{12}}(\pi_{21})^{n_{21}}(\pi_{22})^{n_{22}}$$

where $n=n_{11}+n_{12}+n_{21}+n_{22}$ and $\pi_{11}+\pi_{12}+\pi_{21}+\pi_{22}=1$ (i.e. the probabilities of being selected for each table cell add up to 1).

We can write $\hat{p}_{tel}-\hat{p}_{clin}=\frac{n_{12}+n_{22}}{n}-\frac{n_{21}+n_{22}}{n}=\frac{n_{12}-n_{21}}{n}=\frac{n_{12}}{n}-\frac{n_{21}}{n}$.
Now, a variance estimate can be given as

$$Var(\hat{p}_{tel}-\hat{p}_{clin}) = Var(\frac{n_{12}}{n}-\frac{n_{21}}{n}) = \frac{1}{n^2}Var(n_{12})+\frac{1}{n^2}Var(n_{21})-2(\frac{1}{n})(\frac{1}{n})Cov(n_{12},n_{21})$$

$$= \frac{1}{n}[\pi_{12}(1-\pi_{12})+\pi_{21}(1-\pi_{21})]-\frac{2}{n^2}(-n\pi_{12}\pi_{21})$$

Estimating π_{12} and π_{21} by their sample proportions n_{12}/n and n_{21}/n we get:

$$= \frac{1}{n}[\frac{n_{12}}{n}(\frac{n-n_{12}}{n})+\frac{n_{21}}{n}(\frac{n-n_{21}}{n})]+\frac{2}{n}(\frac{n_{12}}{n})(\frac{n_{21}}{n})$$

$$= \frac{n*n_{12}-n_{12}^2+n*n_{21}-n_{21}^2+2n_{12}*n_{21}}{n^3}$$

$$= \frac{(n_{12}+n_{21})}{n^2} - \frac{(n_{12}-n_{21})^2}{n^3}$$

And thus the standard error is:

$$SE(\hat{p}_{tel} - \hat{p}_{clin}) = \frac{1}{n}\sqrt{(n_{12}+n_{21}) - \frac{(n_{12}-n_{21})^2}{n}} \tag{1}$$

Derivation of standard error by examining change in response

Let us examine the change in responses from the clinic visit to the telephone interview. Define X_i, $(i = 1, \ldots, n)$ as follows:

$X_i = 1$ if individual i indicated mild/moderate on the clinic visit and severe on the telephone
$X_i = 0$ if individual i indicated the same response on both the clinic visit and the telephone
$X_i = -1$ if individual i indicated severe on the clinic visit and mild/moderate on the telephone

Then we have that $X_1, \ldots, X_n$ are independent and identically distributed with the following distribution:

Clinic	Telephone	X_i	Frequency	Probability
Mild/Moderate	Severe	1	n_{12}	$\frac{n_{12}}{n}$
Mild/Moderate Severe	Mild/Moderate Severe	0	$n_{11}+n_{22}$	$\frac{n_{11}+n_{22}}{n}$
Severe	Mild/Moderate	-1	n_{21}	$\frac{n_{21}}{n}$

First, note that the sample mean of the X_i's is $\bar{X} = \frac{n_{12}*1+(n_{11}+n_{12})*0+n_{21}*(-1)}{n} = \frac{n_{12}-n_{21}}{n} = \hat{p}_{tel} - \hat{p}_{clin}$.

Now, when we use $\bar{X}$ as an estimate of $E(X_i)$, we have:

$$Var(\bar{X}) = Var(\frac{1}{n}\Sigma X_i) = \frac{1}{n^2}\Sigma Var(X_i) = \frac{1}{n}[\frac{n_{12}}{n}(1-\bar{X})^2 + \frac{(n_{11}+n_{22})}{n}(0-\bar{X})^2 + \frac{n_{21}}{n}(-1-\bar{X})^2]$$

$$= \frac{1}{n}[\frac{n_{12}}{n}(1-\frac{n_{12}-n_{21}}{n})^2 + \frac{(n_{11}+n_{22})}{n}(0-\frac{n_{12}-n_{21}}{n})^2 + \frac{n_{21}}{n}(-1-\frac{n_{12}-n_{21}}{n})^2]$$

$$= \frac{1}{n^2}[n_{12} - \frac{2*n_{12}*(n_{12}-n_{21})}{n} + \frac{n_{12}*(n_{12}-n_{21})^2}{n^2} + \frac{(n_{11}+n_{22})(n_{12}-n_{21})^2}{n^2} + n_{21} + \frac{2*n_{21}*(n_{12}-n_{21})}{n} + \frac{n_{21}*(n_{12}-n_{21})^2}{n^2}]$$

$$= \frac{1}{n^2}[(n_{12}+n_{21}) + \frac{2*n_{21}*(n_{12}-n_{21}) - 2*n_{12}*(n_{12}-n_{21})}{n} + \frac{(n_{11}+n_{22}+n_{21}+n_{12})(n_{12}-n_{21})^2}{n^2}]$$

$$= \frac{1}{n^2}[(n_{12}+n_{21}) + \frac{2*(n_{21}-n_{12})(n_{12}-n_{21})}{n} + \frac{n*(n_{12}-n_{21})^2}{n^2}]$$

$$= \frac{(n_{12}+n_{21})}{n^2} - \frac{2(n_{12}-n_{21})^2}{n^3} + \frac{(n_{12}-n_{21})^2}{n^3}$$

$$= \frac{(n_{12}+n_{21})}{n^2} - \frac{(n_{12}-n_{21})^2}{n^3}$$

And thus the standard error is the same as (1):

$$SE(\hat{p}_{tel} - \hat{p}_{clin}) = \frac{1}{n}\sqrt{(n_{12}+n_{21}) - \frac{(n_{12}-n_{21})^2}{n}} \tag{2}$$

Confidence intervals under the null and alternative hypotheses

Both formulas (1) and (2) for the standard error have no underlying assumption about whether n_{12} and n_{21} are equal. Each of these two numbers represent the number of each type of discordant pair. Thus a $100 * (1-\alpha)\%$ confidence interval for $p_{tel} - p_{clin}$ can be given as:

$$(\frac{n_{12}-n_{21}}{n}) \pm Z_{1-\alpha/2} * \frac{1}{n}\sqrt{(n_{12}+n_{21}) - \frac{(n_{12}-n_{21})^2}{n}}$$

Under the null hypothesis for McNemar's Test, we are assuming that the probabilities of each type of discordant pair are equal (e.g. $H_0 : \pi_{12} = \pi_{21} = \pi$). The derivation of the standard error under the single multinomial case thus becomes:

$$\begin{aligned} Var(\hat{p}_{tel} - \hat{p}_{clin}) = Var(\frac{n_{12}}{n} - \frac{n_{21}}{n}) &= \frac{1}{n^2}Var(n_{12}) + \frac{1}{n^2}Var(n_{21}) - 2(\frac{1}{n})(\frac{1}{n})Cov(n_{12}, n_{21}) \\ &= \frac{1}{n}[\pi_{12}(1-\pi_{12}) + \pi_{21}(1-\pi_{21})] - \frac{2}{n^2}(-n\pi_{12}\pi_{21}) \\ &= \frac{1}{n}[2\pi - 2\pi^2] + \frac{2\pi^2}{n} = \frac{2\pi}{n} \end{aligned}$$

Estimating π by a pooling of the sample proportions $(n_{12} + n_{21})/2n$ we get:

$$\begin{aligned} &= \frac{2(\frac{n_{12}+n_{21}}{2n})}{n} \\ &= \frac{n_{12}+n_{21}}{n^2} \end{aligned}$$

With standard error

$$SE(\hat{p}_{tel} - \hat{p}_{clin}) = \frac{\sqrt{n_{12}+n_{21}}}{n} \tag{3}$$

Under the second derivation of the standard error using change in response, using 0 as an estimate of $E(X_i)$ will result in the same expression for the standard error as (3). The confidence interval under H_0 then reduces to:

$$(\frac{n_{12}-n_{21}}{n}) \pm Z_{1-\alpha/2} * \frac{1}{n}\sqrt{n_{12}+n_{21}}$$

Here, $Z_{1-\alpha/2}$ represents the $(1-\alpha/2)$th quantile of a standard normal distribution.

Numerical Example

Consider the motivating example with actual cell counts:

	Telephone Interview		
Clinical Visit	Mild/Moderate	Severe	
Mild/Moderate	106	28	134
Severe	30	36	66
	136	64	200

Here, $\hat{p_{tel}} = 64/200 = 0.32$ and $\hat{p_{clin}} = 66/200 = 0.33$. Under the null hypothesis, the standard error is estimated to be $\frac{\sqrt{28+30}}{200} = 0.0380788655$. Under the alternative hypothesis, the standard error is estimated to be $\frac{1}{200}\sqrt{(28+30) - \frac{(28-30)^2}{200}} = 0.0380722996$. The confidence interval under the null hypothesis is thus $-0.01 \pm 1.96 * 0.0380788655 = (-0.084635, 0.0646346)$ and the confidence interval under the alternative hypothesis is thus $-0.01 \pm 1.96 * 0.0380722996 = (-0.0846217, 0.0646217)$.