Matched pairs binary data

First	Second Survey		
survey	Approve	Disapprove	Total
Approve	794	150	944
Disapprove	86	570	656
Total	880	720	1600

	Cases		
Controls	Exposed	Unexposed	Total
Exposed	27	29	56
Unexposed	3	4	7
Total	30	33	63

Dependence

- Matched binary can arise from
  - Measuring a response at two occassions
  - Matching on case status in a retrospective study
  - Matching on exposure status in a prospective or cross-sectional study
- The pairs on binary observations are dependent, so our existing methods do not apply
- We will discuss the process of making conclusions about the marginal probabilities and odds

Notation

	time 2				time 2		
time 1	Yes	No	Total	time 1	Yes	No	Total
Yes	n_{11}	n_{12}	n_{1+}	Yes	π_{11}	π_{12}	π_{1+}
no	n_{21}	n_{22}	n_{2+}	no	π_{21}	π_{22}	π_{2+}
Total	n_{+1}	n_{+2}	n	Total	π_{+1}	π_{+2}	1

- We assume that the $(n_{11}, n_{12}, n_{21}, n_{22})$ are multinomial with n trials and probabilities $(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$
- π_{1+} and π_{+1} are the marginal probabilities of a yes response at the two occasions
- $\pi_{1+} = P(\text{Yes} \mid \text{Time 1})$
- $\pi_{+1} = P(\text{Yes} \mid \text{Time 2})$

Marginal homogeneity

- Marginal homogeneity is the hypothesis $H_0 : \pi_{1+} = \pi_{+1}$
-
- The obvious estimate of
- Under H_0 a consistent estimate of the variance is
- Therefore

  follows an asymptotic χ^2 distribution with 1 degree of freedom

McNemar's test

- The test from the previous page is called McNemar's test
- Notice that only the discordant cells enter into the test
  - n_{12} and n_{21} carry the relevant information about wheth or not π_{1+} and π_{+1} differ
  - n_{11} and n_{22} contribute information to estimating the magnitude of this difference

Example

$\frac{(80-150)^2}{86+150} = 17.36$

P-value = 3×10^{-5}.

Hence we reject the null hypothesis and conclude that there is evidence to suggest a change in opinion between the two polls

In R

```
mcnemar.test(matrix(c(794, 86, 150, 570), 2),
                correct = FALSE)
```

The `correct` option applies a continuity correction

Estimation

- Let $\hat{\pi}_{ij} = n_{ij}/n$ be the sample proportions
- $d = \hat{\pi}_{1+} - \hat{\pi}_{+1} = (n_{12} - n_{21})/n$ estimates the difference in the marginal proportions
- The variance of d is
- $\frac{d-(\pi_{1+}-\pi_{+1})}{\hat{\sigma}_d}$ follows an asymptotic normal distribution
- Compare σ_d^2 with what we would use if the proportions were independent

Example

$d = 944/1600 - 880/1600 = .59 - .55 = .04$

$\hat{\pi}_{11} = .50$**,** $\hat{\pi}_{12} = .09$**,** $\hat{\pi}_{21} = .05$**,** $\hat{\pi}_{22} = .36$

$\hat{\sigma}_d^2 = \{.59(1 - .59) + .55(1 - .55) - 2(.50 \times .36 - .09 \times .05)\}/1600$

$\hat{\sigma}_d = .0095$

95% CI - $.04 \pm 1.96 \times .0095 = [.06, .02]$

Note ignoring the dependence yields $\hat{\sigma}_d = .0175$

Relationship with CMH test

Each subject's (or matched pair's) responses can be represented as one of four tables.

	Response	
Time	Yes	No
First	1	0
Second	1	0

	Response	
Time	Yes	No
First	1	0
Second	0	1

	Response	
Time	Yes	No
First	0	1
Second	1	0

	Response	
Time	Yes	No
First	0	1
Second	0	1

Result

- McNemar's test is equivalent to the CMH test where subject is the stratifying variable and each 2×2 table is the observed zero-one table for that subject
- This representation is only useful for conceptual purposes

Estimating the marginal odds ratio

- The marginal odds ratio is

- The maximum likelihood estimate of the margina *log* odds ratio is

- The asymptotic variance of this estimator is

Example In the approval rating example the marginal OR compares the odds of approval at time 1 to that at time 2

$\hat{\theta} = \log(944 \times 720/880 \times 656) = .16$

Estimated standard error = $.039$

CI for the log odds ratio = $.16 \pm 1.96 \times .039 = [.084, .236]$