Matching

The aim of matching is remove confounding by matching subjects to be similar on a potential confounder. Doing so eliminates (or reduces) confounding, as well as reducing variability thereby increasing power.

Recall a paired t-test with two independent samples:

$$t = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}} = \frac{\bar{X}_1 - \bar{X}_2}{\frac{s_D}{\sqrt{n}}}$$

with n-1 degrees of freedom and standard error:

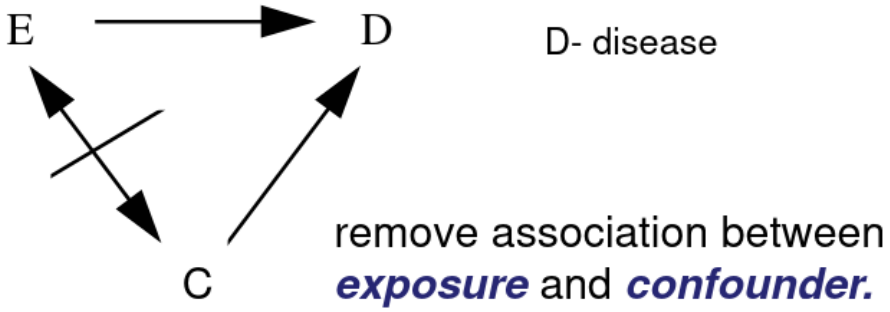
$$s_D = \sqrt{\frac{\sum (D_i - \bar{D})^2}{n-1}}$$

The test is inversely related to variance.

Types of Matching

- Matched Pairs (covered today)
- Categorical Matching (unmatched analysis, stratified or regression)
  - Stratify cases, then find equal number of controls for each category (or equal multiple).
- Caliper Matching
  - Only for continuous variables
  - Similar to categorical but not the same
- Nearest Neighbor
  - Select 'closest' control as match
  - May have minimum match criteria

Matching in Follow-Up Study



Remove Confounding (C) in the study sample between Exposed (E) and unexposed by matching on the potential confounders.

There are 4 possible combinations of outcomes in exposed and unexposed groups:

Concordant pairs

Exposed	Unexposed
Diseased	Diseased
Diseased	Not diseased
Not diseased	Diseased
Not diseased	Not diseased

Discordant pairs

Corncordant pairs have the same outcome between pairs, and opposite in discordant pairs.

An example presentation for matching 2x2 tables:

	Exposed	Not Exposed	
Diseased	15 (a)	10 (b)	25
Not diseased	35 (c)	40 (d)	75
	50	50	100 people

		Unexposed		
		Diseased	Not diseased	
Exposed	Diseased	5 (a')	10 (b')	15 (a=a'+b')
	Not diseased	5 (c')	30 (d')	35 (c=c'+d')
		10 (b=a'+c')	40 (d=b'+d')	50 pairs

Notice the column and cell totals now equal the value of cells a,b,c,and d in the original table.

In the paired follow-up study table we would calculate the Relative Risk Ratio:

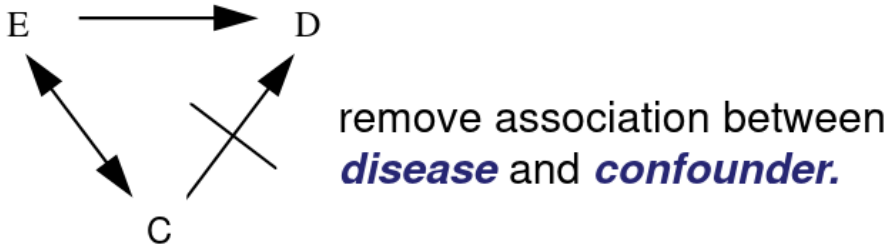
$$\hat{RR} = \frac{\hat{p}_E}{\hat{p}_{\bar{E}}} = \frac{(a'+b')/n}{(a'+c')/n} = \frac{a'+b'}{a'+c'} = \frac{\text{number exposed with disease}}{\text{number unexposed with disease}}$$

If we take the Risk Ratio of both the above tables, we find they are both the same (1.5).

Also, the confidence interval for RR in a follow up would be:

$$RR \times \exp\left(\pm z\sqrt{\frac{b'+c'}{(a'+b')(a'+c')}}\right)$$

Matching in Case-Control Study



Remove Confounding (C) in the sample study between cases and controls by matching on potential confounders where for each case we select a control with the same values for the confounding variables.

For case control studies we set up our pairs differently:

		Controls (Not diseased)		
		Exposed	Not exposed	
Cases (Diseased)	Exposed Not exposed	a' c'	b' d'	a'+b' = a c'+d' = b
		a'+c' = c	b'+d' = d	# pairs

And this also result in a different estimate of Odds Ratio:

$$\hat{O}R = \frac{b'}{c'}$$

And likewise the confidence interval would be

OR x exp(+/- Z*sqrt(1/b' + 1/c')

ORs in Case-Control studies will not be the same if ignoring matching. This is unlike the situation for RR, where the point estimate is the same whether considering the matching or not.

The McNemar Test

The McNemar Test is a non-parametric test for paired nominal data. It is a chi-square distribution and can be used for retrospective case-control or follow-up studies. It assumes:

- The two groups are mutually exclusive
- A random sample

H0: The proportion of some disease is the same in participants with exposure and those without exposure (RR=1)
Ha: The proportion of some disease is not the same in participants with exposure and those without exposure (RR != 1)

$$\chi^2_{McNemar} = \frac{(b' - c')^2}{b' + c'}$$

with df = 1

Matched Analyses and Mantel-Haenszel

Mantel-Haenszel methods applied to strata established by matched sets are equivalent to the conventional matched methods. Works for Cohort (Follow-Up) and case-control studies. When the matched pair is a stratum, we carry out the MH method on each pair.

So from the example above, if we calculated 50 strata for 50 matched pairs, we would end up with 50 tables:

		Unexposed		
		Diseased	Not diseased	
Exposed	Diseased Not diseased	5 (a') 5 (c')	10 (b') 30 (d')	15 35
		10	40	50 pairs

	Exposed	Not Exposed	
Diseased Not diseased	1 0	1 0	2 0
	1	1	2

– – – → 5 tables

	Exposed	Not Exposed	
Diseased Not diseased	0 1	1 0	1 1
	1	1	2

→ 5 tables

	Exposed	Not Exposed	
Diseased Not diseased	1 0	0 1	1 1
	1	1	2

– – – → 10 tables

	Exposed	Not Exposed	
Diseased Not diseased	0 1	0 1	0 2
	1	1	2

→ 30 tables

And then we apply the mRR and MH to each of the 50 strata to obtain the McNemar result.

$$\hat{mRR} = \frac{\sum_{strata} an_0 / n}{\sum_{strata} bn_1 / n} = \frac{5\left(\frac{1\bullet 1}{2}\right)+10\left(\frac{1\bullet 1}{2}\right)+5\left(\frac{0\bullet 1}{2}\right)+30\left(\frac{0\bullet 1}{2}\right)}{5\left(\frac{1\bullet 1}{2}\right)+10\left(\frac{0\bullet 1}{2}\right)+5\left(\frac{1\bullet 1}{2}\right)+30\left(\frac{0\bullet 1}{2}\right)}$$

$$\chi^2_{MH} = \frac{\left[\sum_{strata} \frac{ad-bc}{n}\right]^2}{\sum_{strata} \frac{n_0 n_1 m_0 m_1}{n^2 (n-1)}} = \frac{\left[5\left(\frac{0}{2}\right)+10\left(\frac{1}{2}\right)+5\left(-\frac{1}{2}\right)+30\left(\frac{0}{2}\right)\right]^2}{5\left(\frac{0}{4}\right)+10\left(\frac{1}{4}\right)+5\left(\frac{1}{4}\right)+30\left(\frac{0}{4}\right)}$$

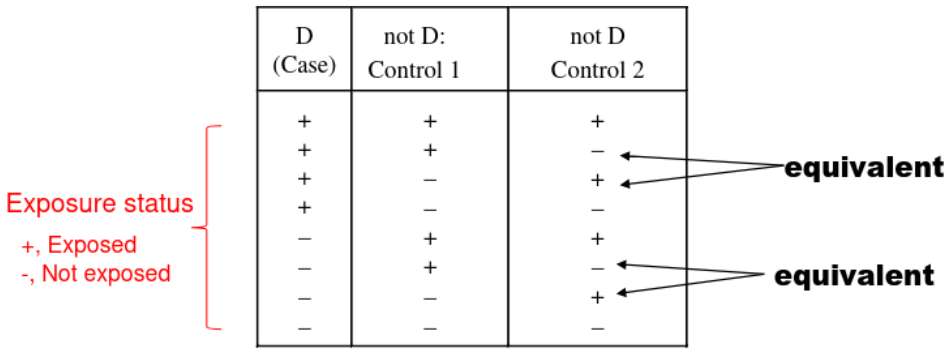
Testing Interaction

It is possible to test interaction between a matching variable and exposure (whether a matching variable modified the effect of E on D).

We obtain an OR for each subgroup of the interaction variable and conduct a formal test to see whether the two ORs are equal.

R-1 Matching

Match one index to R referent subjects. We create a strata for every matched set. A stratum includes a matched set of R + 1 subjects.



In a case control study:

- Index = case
- Referent = control
- Stratum:

	E	Not E
D	Y	1-Y
Not D	X	R-X

- Exposure status:

$$Y = \begin{cases} 1 \text{ if } E \mid D \\ 0 \text{ if } \bar{E} \mid D \end{cases}$$

X = # of exposed controls;
R = # of controls/cases

- We can then use the following formulas for estimate of OR and chi-squared:

$$\hat{OR} = \frac{\sum_{sets} Y\left(R - X\right)}{\sum_{sets} X\left(1 - Y\right)}$$

$$\chi^2_{MH} = \frac{\left[R\sum_{sets} Y - \sum_{sets} X\right]^2}{(1+R)\sum_{sets}\left(X+Y\right) - \sum_{sets}\left(X+Y\right)^2} \quad \text{with} \quad df = 1$$

McNemar's test cannot be used in R-1 matching.

When to Match?

- How hard is it to obtain subjects?
- How hard is it to match?
- How much does matching buy?
  - How influential are the confounders? How stronly correlated are the matched pairs?
  - Is variation within the pairs small relative to variation between pairs?
- What do we gain?
  - Eliminate/reduce confounding
  - Reduce variability and increase power

Issues with Matching

- Matched studies use restriction sampling
- The exposed and the cases will not typically represent the general population
  - Ex. a case-control study of smoking and lung cancer
- The cases will on average be more male and older than the general population and also may not reflect racial distribution
- We may not be able to generalize results

Code

```
### Matched pairs can be analyzed using the Mantel-Haenszel method.
### Each matched pair is treated as a separate stratum
mantelhaen.test(pairs$exposed, pairs$diseased, pairs$match, correct=FALSE)

# table of pairs:
table(exposed.d , unexposed.d)
```

```
### Analyze data with mcnemar.test()
mcnemar.test(table(exposed.d , unexposed.d), correct=FALSE)
mcnemar.test(table(exposed.d , unexposed.d))

### Matched sets can be analyzed using the Mantel-Haenszel method.
### Each matched set is treated as a separate stratum
mantelhaen.test(Rto1$smoke, Rto1$casecon, Rto1$casenum, correct=FALSE)
mantelhaen.test(Rto1$smoke, Rto1$casecon, Rto1$casenum)
```

Revision #10
Created 3 October 2022 13:56:56 by Elkip
Updated 3 October 2022 16:14:06 by Elkip