Class Notes: STAT 501
Nonparametrics & Log-Linear Models
Multivariate Matching
McNemar Test

Da Kuang

University of Pennsylvania

Contents

1 Matching

1.1 Motivation

In an observational study, subjects in the treatment group may not be comparable to those in the control group. Differences in the outcome between the groups are then not necessarily attributable to a treatment effect but instead may arise from differences in confounders.

A popular approach to controlling for confounders is to construct matched pairs through optimal multivariate matching (Paul R. Rosenbaum. Design of Observational Studies).

1.2 Eliminating the Confounding by Matching

Matching is usually performed by matching a fixed ratio of controls to treatment subjects, most commonly pair (one-to-one) matching. Sometimes one-to-two or one-to-three matching. But we rarely have more than one-to-three matching. It is because that the computation cost increase but only gives us marginal improvement.Alternative forms of matching may bring desirable features such as variable ratio matching and full matching.

For instance, suppose we have 100 treatment patient and 200 control patients, then we select 100 control out of the 200 to match treatment patients. We usually only focus on the matched groups but sometime it is beneficial to do some extra calculation on the unmatched group. The unmatched subjects can also be used to reduce confounding bias.

After constructing the match, we check if the observed covariates are balanced so that they have similar distributions, i.e., whether the observed confounders are balanced.

The absolute standardized difference ($|D|$) is usually used to check covariate balance

and evaluate the match quality.

The numerator of $|D|$ is the absolute difference between the treated and the (matched) control in covariate means, and the denominator is the pooled standard deviation before the match, the square root of the average of the two groups' sample variances. Smaller absolute standardized difference implies closer covariate proximity.

For instance, a $|D| < 0.1$ indicates that the absolute mean difference is within 10% of the pooled standard deviation.

1.3 Assess the Matching By Hypothesis Testing

To assess covariate balance, formal tests can also be used, namely the Wilcoxon rank sum test for continuous variables and Fisher's exact test for binary variables.

If the observed confounders are balanced, then, assuming that there are no unmeasured confounders, outcome analysis can be constructed in straightforward ways as in a matched pair randomized trial.

McNemar's test can be used for binary outcomes.

Wilcoxon signed rank test for continuous outcomes.

2 Example: Odds Ratio in Paired and Unpaired Data

Suppose we have the following data.

id	group	age	mortality
A	treatment	30	no
B	treatment	30	no
C	treatment	40	yes
D	treatment	50	yes
a	control	30	no
b	control	30	yes
c	control	45	yes
d	control	50	no
e	control	60	yes
f	control	60	yes

Figure 1: Example Data

2.1   Review: 2×2 **Contingency Table**

The data can be represented by the following contingency table.

	mortality		
	yes	no	total
treatment	2	2	4
control	4	2	6
total	6	4	10

Figure 2: Contingency Table

The odds ratio is the ratio of two odds

$$\theta = \frac{\frac{\mathbf{P}(\text{Yes}|\text{Treatment})}{\mathbf{P}(\text{No}|\text{Treatment})}}{\frac{\mathbf{P}(\text{Yes}|\text{Treatment})}{\mathbf{P}(\text{No}|\text{Treatment})}} = \frac{\mathbf{P}(\text{Yes}|\text{Treatment})\mathbf{P}(\text{No}|\text{Treatment})}{\mathbf{P}(\text{No}|\text{Treatment})\mathbf{P}(\text{Yes}|\text{Treatment})}$$

In the sample,

• Proportion of mortality in the treatment group: $2/4 = 0.5$.

- Proportion of mortality in the control group: $4/6 = 0.67$.
- The estimated odds in the treatment group: $2/2 = 1$.
- The estimated odds in the control group: $4/2 = 2$.

Therefore, the estimated odds ratio is

$$\widehat{\theta} = \frac{2 \times 2}{4 \times 2} = \frac{1}{2}$$

2.2 Matching for Paired Data

To eliminate the effect of age for mortality, we can pair the data as follows, where treatment id and matched id are matched by age.

pair	age	treatment id	matched control id	age
1	30	A	a	30
2	30	B	b	30
3	40	C	c	45
4	50	D	d	50

Figure 3: Matching Data

The mortality of the matched data is as follows.

pair	mortality	treatment id	matched control id	mortality
1	no	A	a	no
2	no	B	b	yes
3	yes	C	c	yes
4	yes	D	d	no

Figure 4: Mortality of Matched Data

The frequency of mortality in each group can be represented by the following table.

	matched control		
treatment	yes	no	total
yes	1 (C-c)	1 (D-d)	2
no	1 (B-b)	1 (A-a)	2
total	2	2	4

Figure 5: Contingency Table for Paired Data

After matching, in the paired sample,

- Proportion of mortality in the treatment group: $2/4 = 0.5$. (Sum of first row)
- Proportion of mortality in the control group: $2/4 = 0.5$. (Sum of first column)

The odds ratio is

$$\lambda = \frac{\mathbf{P}[(\text{Yes}|\text{Treatment})\&(\text{No}|\text{Matched Control})]}{\mathbf{P}[(\text{No}|\text{Treatment})\&(\text{Yes}|\text{Matched Control})]}$$

The estimated odds ratio is

$$\widehat{\lambda} = \frac{1(\text{from pair D-d})}{1(\text{from pair B-b})} = 1$$

3 Odds Ratio of Matching Paired Data

3.1 Set-up

Let X be the treatment variable.

- $X = 1$: treatment group
- $X = 0$: control group

Let Y be the outcome variable. $Y = 1$ or $Y = 0$.

In the j-th pair, let $(Y_{j,t}, Y_{j,mc})$ be the outcome status of the treated individual and the matched individual in the control group.

The event status can be one of the four possibilities $(0,0), (0,1), (1,0), (1,1)$.

	Outcome	
Pair	Y_t	Y_{mc}
1	1	0
2	0	1
$\vdots$	$\vdots$	$\vdots$

Figure 6: Paired Data List

The paired data can be represented into the contingency table.

	Matched control group		
Treatment group	$Y = 1$	$Y = 0$	Total
$Y = 1$	n_{11}	n_{10}	$n_1 = n_{11} + n_{10}$
$Y = 0$	n_{01}	n_{00}	$n_0 = n_{01} + n_{00}$
Total	$n_{.1} = n_{11} + n_{01}$	$n_{.0} = n_{10} + n_{00}$	N

Figure 7: Contingency Table for Paired Data

Each frequency in the table represents number of pairs. For instance, "n_{10}" means the number of pairs where the treated individuals have $Y = 1$ while the matched individual in the control group have $Y = 0$.

N is the total number of pairs. $N = n_{11} + n_{10} + n_{01} + n_{00}$.

The proportion of $Y = 1$

- in the treatment group is $(n_{11} + n_{10}/N)$.
- in the matched control group is $(n_{11} + n_{01}/N)$.

Hence, the difference of the proportions is $(n_{10} - n_{01})/N$.

3.2 Odds Ratio

Use probabilities to represent paired groups.

	Matched control group		
Treatment group	$Y = 1$	$Y = 0$	Total
$Y = 1$	π_{11}	π_{10}	$\pi_1 = \pi_{11} + \pi_{10}$
$Y = 0$	π_{01}	π_{00}	$\pi_0 = \pi_{01} + \pi_{00}$
Total	$\pi_{.1} = \pi_{11} + \pi_{01}$	$\pi_{.0} = \pi_{10} + \pi_{00}$	1

Figure 8: Probability in Contingency Table

In the table,

- $\pi_{10} := \mathbf{P}(Y_{j,t} = 1, Y_{j,mc} = 0)$
- $\pi_{01} := \mathbf{P}(Y_{j,t} = 0, Y_{j,mc} = 1)$.
- $\pi_1 := \mathbf{P}(Y_{j,t} = 1)$.
- $\pi_{.1} := \mathbf{P}(Y_{j,mc} = 1)$.

The odds ratio $\lambda = \pi_{10}/\pi_{01}$ measures the association between the outcome and the treatment.

An estimate of the odds ratio is $\widehat{\lambda} = n_{10}/n_{01}$.

3.3 Hypothesis Testing

The null hypothesis is that the proportion of $Y = 1$ is the same in both treatment group and the matched control group.

$$H_0 : \pi_1 = \pi_{.1} \iff H_0 : \pi_{10} = \pi_{01} \iff H_0 : \lambda = 1$$

The alternative hypothesis is

$$H_1 : \pi_{1} \neq \pi_{.1}$$

The test statistic T can be seen as approximately sampled from χ_1^2 distribution for large sample.

$$T = \frac{(n_{10} - n_{01})^2}{n_{10} + n_{01}}$$

When consider continuity correction

$$T = \frac{(|n_{10} - n_{01}| - 1)^2}{n_{10} + n_{01}}$$

Reject H_0 if $T > \chi_{\alpha,1}^2$.

3.4   Example

Johnson and Johnson (1972), in a study that was interested in testing the theory that the tonsils protect the body against invasion of the lymph nodes by a Hodgkin's disease virus, obtained tonsillectomy data on 85 Hodgkin's cases and a sibling of each case.

The data showed 41 tonsillectomies among the Hodgkin's cases and 33 tonsillectomies among the siblings. The pairing of a case with sibling means that the rates for the two groups are not independent.

The pairing of a case with sibling means that the rates for the two groups are not independent. The pairing should be taken into account in the analysis in order to achieve the best chance of detecting a departure from the null hypothesis.

H_0 : Hodgkin's cases and their siblings have the same rates of tonsillectomy.

The null hypothesis can be reduced to

$$H_0 : \text{The odds of tonsillectomy is the same.}$$

| | Sibling | | |
Hodgkin's Patients	Tonsillectomy	No Tonsillectomy	Total
Tonsillectomy	26	15	41
No Tonsillectomy	7	37	44
Total	33	52	85

Figure 9: Contingency Table

```
1  da=matrix(c(26,7,15,37),2)
2  teststat=(15-7)^2/(15+7)
3  pchisq(teststat,1,lower.tail=F)
4  # [1] 0.08808151
5  mcnemar.test(da,correct=F)
6  # McNemar's Chi-squared test
7  #
8  # data: da
9  # McNemar's chi-squared = 2.9091, df = 1, p-value = 0.08808
10 pchisq((abs(15-7)-1)^2/(15+7),1,lower.tail=F)
11 # [1] 0.135593
12 mcnemar.test(da)
13 # McNemar's Chi-squared test with continuity correction
14 #
15 # data: da
16 # McNemar's chi-squared = 2.2273, df = 1, p-value = 0.1356
```

4 Exact Test for Small Samples

If there is no association between tonsillectomy and Hodgkin's disease, then the probability is 1/2 that a patient-sibling pair falls in the upper-right cell and 1/2 that it falls in the lower-left cell, given that the pair falls off the main diagonal. Since the pairs are independent, the ratio 15/22 can be compared with 1/2 by a binomial test.

$$n_{10} \sim \text{Binomial}(n^*, p), \text{ where}$$

$$n^* = n_{10} + n_{01}$$

$$p = \frac{\pi_{10}}{\pi_{10} + \pi_{01}}$$

The null hypothesis is $H_0 : p = 1/2$.

If n^* if large, normal approximation gives

$$T = \frac{n_{10} - \frac{n^*}{2}}{\sqrt{n^*\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}} = \frac{n_{10} - n_{01}}{\sqrt{n_{10} + n_{01}}}$$

If the alternative hypothesis is two-sided,

```
1  nstat=(15-7)/sqrt(15+7)
2  pv=2*pnorm(nstat,lower.tail=F)
3  pexact=2*pbinom(14,22,prob=1/2,lower.tail=F)
4  c(pv,pexact)
5  # [1] 0.08808151 0.13380051
```

If the alternative hypothesis is one-sided,

H_1 : the rate of tonsillectomy is higher in patients with Hodgkin's disease

```
1  pbinom(14, size=22, prob=1/2, lower.tail=F)
2  # [1] 0.06690025
3  binom.test(15,22,1/2,alternative="g")
4  # Exact binomial test
5  #
6  # data: 15 and 22
7  # number of successes = 15, number of trials = 22, p-value = 0.0669
```

```
 8  # alternative hypothesis: true probability of success is greater than 0.5
 9  # 95 percent confidence interval:
10  # 0.4845433 1.0000000
11  # sample estimates:
12  # probability of success
13  # 0.6818182
14  library(BSDA)
15  x=c(rep(1,15),rep(0,7))
16  y=c(rep(0,15),rep(1,7))
17  cbind(x,y)
18  SIGN.test(x,y,alternative="g")
19  # Dependent-samples Sign-Test
20  #
21  # data: x and y
22  # S = 15, p-value = 0.0669
23  # alternative hypothesis: true median difference is greater than 0
24  # 95 percent confidence interval:
25  # 0.1687213 Inf
26  # sample estimates:
27  # median of x-y
28  # 1
29  #
30  # Achieved and Interpolated Confidence Intervals:
31  #
32  # Conf.Level L.E.pt U.E.pt
33  # Lower Achieved CI 0.9331 1.0000 Inf
34  # Interpolated CI 0.9500 0.1687 Inf
35  # Upper Achieved CI 0.9738 -1.0000 Inf
```

5 Example

A study concerns the approval ratings of a prime minister in two surveys. In the first survey, ratings were obtained on 1600 citizens and then in a second survey, six months later, the same citizens were resurveyed.

- In the 1st survey, 944 indicated approval of the Prime Minister's performance in office.
- In the 2nd survey, 880 indicated approval.

There each subject has two observation. So that we have 1600 pairs of data out of 1600 people.

	Second survey	
First survey	Approve	Disapprove
Approve	794	150
Disapprove	86	570

Figure 10: Contingency Table

The null hypothesis is

$$H_0 : \text{the approval rate does not change.}$$

It can be reduced to

$$H_0 : \text{the odds of approval does not change.}$$

It means that the probability of going from approval to disapproval is the same as the probability of going from disapproval to approval.

In the following code, we identify that 4 more percentage of people going from approval to disapproval. McNemar test tells us that it is significant.

```
 1  Performance=
 2    matrix(c(794, 86, 150, 570),
 3          nrow = 2,
 4          dimnames = list("1st␣Survey" = c("Approve", "Disapprove"),
 5                          "2nd␣Survey" = c("Approve", "Disapprove")))

 7  (794+150)/1600 # [1] 0.59
 8  (794+86)/1600 # [1] 0.55

10  # the proportion of going from approval to disapproval is higher than
11  # the proportion of going from disapproval to approval.
12  (150-86)/1600 # [1] 0.04

14  Performance
15  # 2nd Survey
16  # 1st Survey Approve Disapprove
17  # Approve 794 150
18  # Disapprove 86 570
19  mcnemar.test(Performance,correct=F)
20  # McNemar's Chi-squared test
21  #
22  # data: Performance
23  # McNemar's chi-squared = 17.356, df = 1, p-value = 3.099e-05

25  n10=150
26  n01=86
27  tstat=(n10-n01)^2/(n10+n01)
28  pchisq(tstat,1,lower.tail=F)
29  # [1] 3.099293e-05

31  mcnemar.test(Performance)
```

```
32  # McNemar's Chi-squared test with continuity correction
33  #
34  # data: Performance
35  # McNemar's chi-squared = 16.818, df = 1, p-value = 4.115e-05
36  tstat=(abs(n10-n01)-1)^2/(n10+n01)
37  pchisq(tstat,1,lower.tail=F)
38  # [1] 4.114562e-05
```

kappa test focus on the agreement while the McNemar Test focus on the disagreement.

```
1  library(psych)
2  cohen.kappa(Performance)
3  # Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels =
        levels)
4  #
5  # Cohen Kappa and Weighted Kappa correlation coefficients and confidence
        boundaries
6  # lower estimate upper
7  # unweighted kappa 0.66 0.7 0.73
8  # weighted kappa 0.66 0.7 0.73
9  #
10 # Number of subjects = 1600
```