Lecture 26: Conditional Logistic Models for Matched Pairs

Dipankar Bandyopadhyay, Ph.D.

BMTRY 711: Analysis of Categorical Data Spring 2011

Division of Biostatistics and Epidemiology

Medical University of South Carolina

Conditional Logistic Regression

Purpose

1. Eliminate unwanted nuisance parameters

2. Use with sparse data

- Suppose, we can group our covariates into J unique combinations
- and as such, we can form j (2×2) tables
- Think of each of the j stratum as a matched pair (or matched set if R:1 matching used)

Conditional Logistic Regression, Matched Pairs

Question: When would you have small strata sample sizes y_{j++} ?

Answer: Matched pairs study

- Matched pairs studies have become increasingly popular in biostatistics and epidemiology.

- For example, in a matched case-control study, you select a case, and then try to match a control to the case, you may match on variables like age, sex, race, etc. The case and control make up a 'strata'.

- Alternatively, unlike the matched case-control, you could match individuals by age, sex, race, etc., and then give them two different treatments, and see if the patients respond or don't respond to treatment; this would be a matched prospective study.

- Going one-step further, you can do a cross-over trial, in which you give the same patient two different drugs, and see if the patient responds or doesn't respond to the treatment; here the strata consists of the two binary measurements on the same subject.

- In general, matching yields more precision than an unmatched study, but matching can be complicated (See Klienbaum & Klien Chp 8)

- In these matched studies, besides treatment (or exposure), there may be other covariates of interest as well.

- These 'other' covariates of interest cannot be matching variables, or their effects will drop out of the conditional likelihood.

- In these matched studies, the stratum sample sizes are $y_{j++} = 2$.

- Which is small

Example

- With data like this, you have $j = 1, ..., J$ strata (2×2) tables with one subject on each treatment, or, equivalently, one case and one control. The total sample size for each table is $y_{j++} = 2$.

- For ease of notation and exposition, we will consider a prospective matched pairs study, although since we will be conditioning on both margins of the (2×2), the methods can be used for prospective or case-control studies.

- In a clinical trial (prospective study), you match two patients in the same hospital on age, sex, race, etc., and give one patient treatment 1 and the other patient treatment 2. The bernoulli outcome could be the recurrence of cancer. The resulting data look like that on the following page.

		OUTCOME					OUTCOME	
Pair	TRT	SUCC	FAIL		Pair	TRT	SUCC	FAIL
----	----	----	----		----	----	----	----
1	NEW	1	0		16	NEW	1	0
	PLAC	0	1			PLAC	1	0
2	NEW	0	1		17	NEW	0	1
	PLAC	0	1			PLAC	1	0
3	NEW	1	0		18	NEW	1	0
	PLAC	0	1			PLAC	1	0
4	NEW	1	0		19	NEW	1	0
	PLAC	0	1			PLAC	1	0
5	NEW	1	0		20	NEW	0	1
	PLAC	1	0			PLAC	1	0
6	NEW	1	0		21	NEW	0	1
	PLAC	0	1			PLAC	1	0
7	NEW	1	0		22	NEW	0	1
	PLAC	0	1			PLAC	0	1

8	NEW	0	1	23	NEW	0	1
	PLAC	0	1		PLAC	1	0
9	NEW	1	0	24	NEW	1	0
	PLAC	0	1		PLAC	1	0
10	NEW	1	0	25	NEW	1	0
	PLAC	0	1		PLAC	1	0
11	NEW	0	1	26	NEW	0	1
	PLAC	1	0		PLAC	1	0
12	NEW	1	0	27	NEW	1	0
	PLAC	0	1		PLAC	1	0
13	NEW	0	1	28	NEW	1	0
	PLAC	1	0		PLAC	0	1
14	NEW	1	0	29	NEW	0	1
	PLAC	0	1		PLAC	1	0
15	NEW	1	0	30	NEW	1	0
	PLAC	0	1		PLAC	1	0

Again, the jth (2×2) table (j^{th} Strata) of cell counts looks like:

TABLE j (or stratum 'j')

Variable (Y)

		1	2	
Variable (X)	1	Y_{j11}	Y_{j12}	$Y_{j1+} = 1$
	2	Y_{j21}	Y_{j22}	$Y_{j2+} = 1$
		Y_{j+1}	Y_{j+2}	$Y_{j++} = 2$

However, the following are the only unique combinations of data possible

	Y		
	Success	Failure	
New	1	0	1
X			
Placebo	1	0	1

	Y		
	Success	Failure	
New	0	1	1
X			
Placebo	0	1	1

	Y		
	Success	Failure	
New	1	0	1
X			
Placebo	0	1	1

	Y		
	Success	Failure	
New	0	1	1
X			
Placebo	1	0	1

- Each matched pair form a strata, and, you can never have more than 2 subjects in a stratum.

- For this prospective study, the rows in stratum j are independent Bernoulli random variables (Binomial with sample size 1), with the subject on the new treatment (row 1),

$$Y_{j11} \sim Bern(p_{j1})$$

where

$$\text{logit}(p_{j1}) = \alpha_j + \beta$$

and α_j is the effect of the jth stratum (the matching variables), and the subject on placebo (row 2) has

$$Y_{j21} \sim Bern(p_{j2})$$

where

$$\text{logit}(p_{j2}) = \alpha_j$$

- Note, if this was a matched case-control study, then the we can still use the above (prospective study) model, in which we rewrite the logits as

$$\text{logit}(p_{j1}) = \alpha_j^* + \beta,$$

and

$$\text{logit}(p_{j2}) = \alpha_j^*$$

Here, the intercept α_j^* is not the true α_j. However, in our minds, these intercepts, α_j^*, are nuisance parameters that we do not want to estimate, anyway. Also, β is still the log-odds ratio of interest.

- Now, using conditional likelihood theory we know that, if we further condition on the columns (so that we have a (2×2) table with both margins fixed), the conditional likelihood is only a function of the odds ratio for the (2×2) table.

- For this (2×2) table, the log-odds ratio is

$$\text{logit}(p_{j1}) - \text{logit}(p_{j2}) = \\ [\alpha_j + \beta] - \alpha_j = \beta$$

- Thus, we can form a conditional likelihood, eliminating the unwanted nuisance parameters α_j, (or α_j^*) by conditioning on both margins.

Conditional Likelihood for Matched Pairs

- Now, we will show that, for a matched pairs study, the conditional likelihood has a particularly simple form.

- We can even use unconditional logistic regression to estimate the parameters.

- With both margins fixed, we only need to consider one random variable in the (2×2) table, and, again, we will consider

$$Y_{j11}$$

- Since, the rows $Y_{j1+} = Y_{j2+} = 1$ are fixed by design (one patient on each treatment), the possible table configurations are

		Y		
		Success	Failure	
	New	1	0	1
X				
	Placebo	1	0	1
		2	0	2

		Y		
		Success	Failure	
	New	0	1	1
X				
	Placebo	0	1	1
		0	2	2

		Y		
		Success	Failure	
	New	1	0	1
X				
	Placebo	0	1	1
		1	1	2

		Y		
		Success	Failure	
	New	0	1	1
X				
	Placebo	1	0	1
		1	1	2

- Now, we condition on both margins, although, given both row totals equal 1, we only need to condition on the first column total Y_{j+1}, since the second column total equals $2 - Y_{j+1}$.

- The conditional likelihood is made up of the products over strata of

$$P[Y_{j11} = y_{j11}|Y_{j+1}]$$

  (For ease of notation, we have dropped y_{j1+} and y_{j2+} since they are fixed by design).

- We know that there are only four possible tables for each stratum.

- First, let's consider the situation when either the patient on treatment 1 or the patient or treatment 2 succeeds, but not both, i.e., the first column total $Y_{j+1} = 1$.

- Then, we could see either of the following (2×2) tables

```
TRT          SUCC     FAIL
-------      ------  ----
NEW            1        0   1
PLAC           0        1   1
               1        1   2


TRT          SUCC     FAIL
-------      ------  ----
NEW            0        1   1
PLAC           1        0   1
               1        1   2
```

- Now, given both margins are $(1, 1)$, Y_{j11} can take on 2 possible values, 0 or 1.

- Suppose we let

$$\pi = P[y_{j11} = 1 | Y_{j+1} = 1]$$

- Now, the definition of a bernoulli variable is one that can only take on values 0 or 1, so $\{Y_{j11} | y_{j+1} = 1\}$ is Bernoulli with probability distribution

$$\pi^{y_{j11}}(1 - \pi)^{1 - y_{j11}}$$

- Then, we need to determine π.

The non-central hypergeometric

- Using the non-central hypergeometric,

$$
\pi = P[Y_{j11} = y_{j11} | Y_{j+1} = 1] =
$$

$$
\frac{\dbinom{y_{j+1}}{y_{j11}} \dbinom{y_{j++} - y_{j+1}}{y_{j1+} - y_{j11}} (OR_j)^{y_{j11}}}{\sum_\ell \dbinom{y_{j+1}}{\ell} \dbinom{y_{j++} - y_{j+1}}{y_{j1+} - \ell} (OR_j)^{\ell}} =
$$

$$
\frac{\dbinom{1}{y_{j11}} \dbinom{1}{1 - y_{j11}} e^{(\beta) y_{j11}}}{\sum_{\ell=0}^{1} \dbinom{1}{\ell} \dbinom{1}{1 - \ell} e^{(\beta)\ell}}
$$

- Then, the probability $Y_{j11} = 1$ is

$$\pi = P[Y_{j11} = 1 | Y_{j+1} = 1] =$$

$$\frac{\begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{(\beta)1}}{\begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1-0 \end{pmatrix} e^{(\beta)0} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1-1 \end{pmatrix} e^{(\beta)1}} = \frac{e^{\beta}}{1+e^{\beta}}$$

- Then

$$\pi = P[Y_{j11} = 1 | Y_{j+1} = 1] = \frac{e^\beta}{1 + e^\beta}$$

or, equivalently, the log-odds ratio (β) of interest actually equals the logit of the conditional probability:

$$\text{logit}(\pi) = \text{logit}\{P[Y_{j11} = 1 | Y_{j+1} = 1]\} = \beta$$

- Then, the distribution of

$$\{Y_{j11} | Y_{j+1} = 1\} \sim Bern\left(\frac{e^\beta}{1 + e^\beta}\right),$$

which, again is not a function of α_j,

- Then, the contribution to the conditional likelihood from this stratum is

$$L_j(\beta) = \left(\frac{e^\beta}{1 + e^\beta}\right)^{y_{j11}} \left(\frac{1}{1 + e^\beta}\right)^{(1 - y_{j11})}$$

- Next, consider the case when both the treatment and placebo subjects have a successful response, $Y_{j+1} = 2$,

```
TRT          SUCC     FAIL
-------      ------- ----
NEW            1        0    1
PLAC           1        0    1
               2        0    2
```

- We see that the conditional probability

$$P[y_{j11} = 0 | Y_{j+1} = 2] = 0,$$

$$P[y_{j11} = 1 | Y_{j+1} = 2] = 1,$$

and

$$P[y_{j11} = 2 | Y_{j+1} = 2] = 0,$$

- This conditional distribution is said to be degenerate, all the probability is located at $Y_{j11} = 1$. Also, this conditional distribution is not a function of the parameter of interest β.

- Next, consider the case when both the treatment and placebo subjects fail, $Y_{j+1} = 0$,

```
TRT         SUCC    FAIL
-------     ------- ----
NEW           0        1    1
PLAC          0        1    1
              0        2    2
```

- We see that the conditional probability

$$P[y_{j11} = 0|Y_{j+1} = 0] = 1,$$

$$P[y_{j11} = 1|Y_{j+1} = 0] = 0,$$

and

$$P[y_{j11} = 2|Y_{j+1} = 0] = 0,$$

- This conditional distribution is also degenerate, all the probability is located at $Y_{j11} = 0$. Also, this conditional distribution is not a function of the parameter of interest β.

- Then, only strata with 1 in each margin contribute any information about β when conditioning on both margins:

```
TRT         SUCC    FAIL

-------     ------- ----
NEW           1        0   1
PLAC          0        1   1
              1        1   2


TRT         SUCC    FAIL

-------     ------- ----
NEW           0        1   1
PLAC          1        0   1
              1        1   2
```

- If this was a matched case-control study, we still have that only tables with one in each margin contribute, (only the tables with one subject exposed, and one not exposed), contribute to the conditional likelihood,

```
              Case    Control
    Exposed    1          0     1
Not Exposed    0          1     1
               1          1
```

```
              Case    Control
    Exposed    0          1     1
Not Exposed    1          0     1
               1          1
```

- Pairs like these are sometimes called 'discordant pairs'.
- Note, in these discordant pairs, there is a 1 in every margin:

$$Y_{j+1} = Y_{j+2} = 1 \quad \text{and} \quad Y_{j1+} = Y_{j2+} = 1,$$

- Alternatively, we can also write the discordant pairs as the pairs with

$$Y_{j11} \neq Y_{j21}.$$

Conditional Likelihood

- In other words, for a prospective or case-control study, the conditional likelihood can be reduced to a product over strata with 1 in every margin ($Y_{j+1} = 1$ and $Y_{j1+} = 1$), or, equivalently. strata with $Y_{j11} \neq Y_{j21}$.

$$
\begin{aligned}
L^c(\beta) &= \prod_{j:Y_{j1+}=Y_{j+1}=1} L_j(\beta) \\
&= \prod_{j:Y_{j1+}=Y_{j+1}=1} \left(\frac{e^\beta}{1+e^\beta}\right)^{y_{j11}} \left(\frac{1}{1+e^\beta}\right)^{1-y_{j11}} \\
&= \left(\frac{e^\beta}{1+e^\beta}\right)^{n_{10}} \left(\frac{1}{1+e^\beta}\right)^{n_{01}}
\end{aligned}
$$

- Where

$$n_{10}$$

  is the number of matched pairs with the following (2×2) table (case exposed, control unexposed) or (new treatment successful, old treatment fail)

```
            D        not D
      E     1          0
  not E     0          1
```

- Note that the subscript 10, in n_{10} is a pneumonic for the first row of the table containing a 1 in the first cell (a success) and a 0 in the second cell (a failure)

- And

$$n_{01}$$

  is the number of matched pairs with the following (2×2) table (case unexposed, control exposed), (new treatment fails, old treatment succeeds) :

```
            D        not D
      E     0          1
  not E     1          0
```

The Conditional Maximum likelihood estimator (CMLE)

- The conditional likelihood is proportional to a binomial likelihood with sample size $n^* = n_{10} + n_{01}$.

$$\left(\frac{e^\beta}{1 + e^\beta} \right)^{n_{10}} \left(\frac{1}{1 + e^\beta} \right)^{n^* - n_{10}} = \pi^{n_{10}} (1 - \pi)^{n^* - n_{10}},$$

where

$$\pi = P[y_{j11} = 1 | Y_{j+1} = 1] = \frac{e^\beta}{1 + e^\beta}$$

- Applying the results maximum likelihood results for a single binomial sample with success probability π, and n_{10} success out of $n^* = n_{10} + n_{01}$ trials, the CMLE of π is

$$\widehat{\pi} = \frac{n_{10}}{n^*} = \frac{n_{10}}{n_{10} + n_{01}}$$

- Since

$$\pi = P[y_{j11} = 1 | Y_{j+1} = 1] = \frac{e^\beta}{1 + e^\beta} \ ,$$

as in logistic regression,

$$\beta = \mathsf{logit}(\pi)$$

- The CMLE of the log-odds ratio $\beta = \text{logit}(\pi)$ is

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \text{logit}(\widehat{\pi}) \\[2mm]
&= \log\left(\frac{n_{10}}{n^* - n_{10}}\right) \\[2mm]
&= \log\left(\frac{n_{10}}{n_{01}}\right)
\end{aligned}
$$

- Then, the CMLE of the OR is

$$
\widehat{OR}_{CMLE} = e^{\hat{\boldsymbol{\beta}}} = \frac{n_{10}}{n_{01}}
$$

- Now, you can show that the Mantel-Haenszel estimate of the common odds ratio also equals

$$
\widehat{OR}_{MH} = \widehat{OR}_{CMLE} = \frac{n_{10}}{n_{01}}
$$

- Thus, as we discussed before, the Mantel-Haenzsel estimator works well when the strata sizes are small.

Test for $\beta = 0$

- Since

$$\{Y_{j11}|Y_{j+1} = 1\} \sim Bern\left(\frac{e^\beta}{1 + e^\beta}\right),$$

and the sum of independent Bernoullis is Binomial, we have that n_{10} given $(n_{10} + n_{01})$ is binomial:

$$n_{10} = \sum_{j:Y_{j+1}=Y_{j1+}=1} Y_{j11} \sim Bin\left[(n_{10} + n_{01}), \frac{e^\beta}{1 + e^\beta}\right]$$

- In particular, we have a binomial sample of size $(n_{10} + n_{01})$ with n_{10} successes, and success probability

$$\pi = \frac{e^\beta}{1 + e^\beta}$$

Conditional Confidence Interval for OR

- A 95% confidence interval for $\beta = log(OR)$ is

$$\hat{\boldsymbol{\beta}} \pm 1.96\sqrt{\widehat{Var}[\hat{\boldsymbol{\beta}}]}$$

- However, since

$$\hat{\boldsymbol{\beta}} = \text{logit}(\widehat{\pi}) = \log\left(\frac{n_{10}}{n_{01}}\right)$$

the confidence interval can be equivalently written as

$$\text{logit}(\widehat{\pi}) \pm 1.96\sqrt{\widehat{Var}[\text{logit}(\widehat{\pi})]}$$

- Using the fact that, logit$(\widehat{\pi})$ can be formed by using the logit for a single binomial sample, in large samples, for the binomial, the variance estimate of the logit is

$$\widehat{Var}(\text{logit}(\widehat{\pi})) = \frac{1}{n_{10}} + \frac{1}{n_{01}}$$

- Then, a large sample 95% confidence interval for $\beta = log(OR)$ is

$$\text{logit}(\widehat{\pi}) \pm 1.96\sqrt{\frac{1}{n_{10}} + \frac{1}{n_{01}}}$$

and we can exponentiate the endpoints to get a large sample confidence interval for the odds ratio:

$$\exp\left[\text{logit}(\widehat{\pi}) \pm 1.96\sqrt{\frac{1}{n_{10}} + \frac{1}{n_{01}}}\right]$$

Testing

- Suppose you want to test

$$H_0 : \beta = \log(OR) = 0$$

- This is equivalent to testing

$$H_0 : \pi = \frac{e^\beta}{1 + e^\beta} = \frac{e^0}{1 + e^0} = .5$$

- We can base the test statistic for this null on the distribution of n_{10}.

$$Z = \frac{n_{10} - E(n_{10}|H_0 : \beta = 0)}{\sqrt{Var(n_{10}|H_0 : \beta = 0)}} \sim N(0, 1)$$

- Under the null,

$$n_{10} \sim Bin\left[(n_{10} + n_{01}), .5\right]$$

- In particular, under the null

$$H_0 : \beta = 0$$

or equivalently,

$$H_0 : \pi = \frac{e^\beta}{1 + e^\beta} = \frac{e^0}{1 + e^0} = .5$$

$$E(n_{10}|H_0 : \beta = 0) = (n_{10} + n_{01}).5$$

and

$$
\begin{aligned}
Var(n_{10}|H_0 : \beta = 0) &= (n_{10} + n_{01}).5(1 - .5) \\
&= (n_{10} + n_{01}).25
\end{aligned}
$$

- Putting these in Z, we get

$$
\begin{aligned}
Z &= \frac{n_{10} - E(n_{10}|n_{10}+n_{01})}{\sqrt{Var(n_{10}|n_{10}+n_{01})}} \\
&= \frac{n_{10} - (n_{10}+n_{01}).5}{\sqrt{(n_{10}+n_{01}).25}} \\
&= \frac{.5[n_{10}-n_{01}]}{.5\sqrt{(n_{10}+n_{01})}} \\
&= \frac{[n_{10}-n_{01}]}{\sqrt{(n_{10}+n_{01})}}
\end{aligned}
$$

or

$$
Z^2 = \frac{[n_{10} - n_{01}]^2}{n_{10} + n_{01}} \sim \chi_1^2
$$

which has popularly been come to known as McNemar's Statistic for matched pairs.

- For matched pairs, if we treat each pair as a strata, and use the Mantel-Haenzsel test for conditional independence between Exposure and Disease given strata, we also get McNemar's test:

$$Z^2 = \frac{[n_{10} - n_{01}]^2}{n_{10} + n_{01}}$$

(2×2) table for matched pairs

- Often you will see matched pair data summarized in a (2×2) table as follows:
  Matched Case-Control study

		CASE		
		UNEXPOSED	EXPOSED	
	UNEXPOSED	n_{11}	n_{10}	n_{1+}
CONTROL				
	EXPOSED	n_{01}	n_{00}	n_{0+}
	total	n_{+1}	n_{+0}	n

Matched Pairs Clinical Trial

		NEW TREATMENT		
		FAILURE	SUCCESS	
	FAILURE	n_{11}	n_{10}	n_{1+}
PLACEBO				
	SUCCESS	n_{01}	n_{00}	n_{0+}
	total	n_{+1}	n_{+0}	n

- Then, the estimated odds ratio is obtained as the ratio of the off diagonal elements:

$$\widehat{OR} = \frac{n_{10}}{n_{01}}.$$

and McNemar's chi-square (=Mantel-Haenszel), is

$$Z^2 = \frac{[n_{10} - n_{01}]^2}{n_{10} + n_{01}} \sim \chi_1^2$$

- Exact $p-$value can be obtained by using one sample binomial test for $H_o : \beta = 0$ ($H_o : \pi = .5$) when there are n_{10} successes in $n_{10} + n_{01}$ trials.

Above Example using SAS Proc Freq

```
data pair;
 input pair trt out @@;
/* out : 1=S, 0=F */
/* trt : 1=New, 0=Plac */

cards;
 1 1 1 1 0 0 2 1 0 2 0 0 3 1 1 3 0 0
 4 1 1 4 0 0 5 1 1 5 0 1 6 1 1 6 0 0
 7 1 1 7 0 0 8 1 0 8 0 0 9 1 1 9 0 0
10 1 1 10 0 0 11 1 0 11 0 1 12 1 1
12 0 0 13 1 0 13 0 1 14 1 1 14 0 0
15 1 1 15 0 0 16 1 1 16 0 1 17 1 0
17 0 1 18 1 1 18 0 1 19 1 1 19 0 1
20 1 0 20 0 1 21 1 0 21 0 1 22 1 0
22 0 0 23 1 0 23 0 1 24 1 1 24 0 1
25 1 1 25 0 1 26 1 0 26 0 1 27 1 1
27 0 1 28 1 1 28 0 0 29 1 0 29 0 1
30 1 1 30 0 1
;
```

```
proc freq;
 table pair*trt*out / cmh;
run;


/* Proc Freq Output */


SUMMARY STATISTICS FOR TRT BY OUT
CONTROLLING FOR PAIR


Cochran-Mantel-Haenszel Statistics (Based on Table Scores)


Statistic    Alternative Hypothesis     DF       Value        Prob
--------------------------------------------------------------------
    1         Nonzero Correlation        1        0.474        0.491
    2         Row Mean Scores Differ     1        0.474        0.491
    3         General Association        1        0.474        0.491


Estimates of the Common Relative Risk (Row1/Row2)
                                                 95%
Type of Study     Method              Value   Confidence Limits
--------------------------------------------------------------------
Case-Control      Mantel-Haenszel     1.375    0.5531    3.4184
```

- With a little data reformatting (to be seen later), we have

```
table_                                     Cumulative      Cumulative
pattern     Frequency      Percent      Frequency         Percent
---------------------------------------------------------------------
0*0                 3        10.00              3           10.00
0*1                 8        26.67             11           36.67
1*0                11        36.67             22           73.33
1*1                 8        26.67             30          100.00
```

- Where "0*0" represents 3 matched pairs had no successes (i.e., treatment = failure and placebo = failure)

- (refer back to slides 6 & 7, pairs 2, 8 and 22)

- Note that the CMLE is

$$\widehat{OR}_{CMLE} = \frac{n_{10}}{n_{01}} = \frac{11}{8} = 1.375$$

Using Proc Logistic for Conditional Logistic Regression

```
proc logistic descending;
 class pair;
  model out = pair trt ;
  exact trt / estimate = both /*both = logor & or */;
 run;
```

```
/* Conditional Logistic Regression Output */

Exact Conditional Analysis

          Conditional Exact Tests


                                    --- p-Value ---
Effect    Test             Statistic   Exact      Mid

trt       Score              0.4737    0.6476    0.5755
          Probability        0.1442    0.6476    0.5755
```

```
                    Exact Parameter Estimates


                                  95% Confidence
Parameter      Estimate              Limits              p-Value

trt              0.3185          -0.6857       1.3706       0.6476



                       Exact Odds Ratios


                                  95% Confidence
Parameter      Estimate              Limits              p-Value

trt               1.375          0.504        3.938        0.6476
```

Transforming data in SAS for McNemar's

```
data trt(drop=trt);
 set pair;
 where trt=1;
 rename out=trtout;
run;

data plac(drop=trt);
 set pair;
 where trt=0;
 rename out=placout;
run;

data new;
 merge trt plac;
   by pair;

proc print noobs;
 run;
```

```
   /* DATASET NEW */
PAIR      TRTOUT      PLACOUT


   1         1            0
   2         0            0
   3         1            0
   4         1            0
   5         1            1
   6         1            0
   7         1            0
   8         0            0
   9         1            0
  10         1            0
  11         0            1
  12         1            0
...
  29         0            1
  30         1            1
```

McNemar's

We will use Proc Freq to get the data for McNemar' Stat,

```
proc freq order=freq;
 table trtout*placout/agree;
run;
```

```
TRTOUT        PLACOUT

Col Pct   |        1|        0|   Total
---------+--------+--------+
      1  |       8 |      11 |       19
         |   26.67 |   36.67 |   63.33
         |   42.11 |   57.89 |
         |   50.00 |   78.57 |
---------+--------+--------+
      0  |       8 |       3 |       11
         |   26.67 |   10.00 |   36.67
         |   72.73 |   27.27 |
         |   50.00 |   21.43 |
---------+--------+--------+
Total            16       14       30
              53.33    46.67   100.00
```

```
Statistics for Table of trtout by placout

     McNemar's Test
------------------------
Statistic (S)      0.4737
DF                      1
Pr > S             0.4913
```

- Then, the estimated odds ratio is obtained as the ratio of the off diagonal elements:

$$\widehat{OR} = \frac{n_{10}}{n_{01}} = \frac{11}{8} = 1.375$$

and McNemar's chi-square (=Mantel-Haenszel = Conditional Score Stat), is

$$Z^2 = \frac{[n_{10} - n_{01}]^2}{n_{10} + n_{01}} = \frac{(11 - 8)^2}{11 + 8} = 0.4737$$

Exact $p-$values

- Proc Logistic using the EXACT option calculated the exact $p-$value

- Proc Freq calculated the large sample approximation

- If you would like to calculate the exact binomial probabilities, you can refer back to lectures 02 or 25

```
data pvalues;
 input outcome $ count;
 p_largesample = 1 - probchi(0.4737,1);
cards;
1succ 11
2fail 8
;
run;
proc print data=pvalues;
 var p_largesample;
run;
proc freq data=pvalues;
 tables outcome / binomial;
 weight count;
 exact binomial;
run;
```

```
Obs     p_largesample

 1         0.49129
 2         0.49129 <- prints twice due to 2 data lines



The FREQ Procedure

        Binomial Proportion
        for outcome = 1succ
---------------------------------------
   Test of H0: Proportion = 0.5


ASE under H0                      0.1147
Z                                 0.6882
One-sided Pr >  Z                 0.2456
Two-sided Pr > |Z|                0.4913


Exact Test
One-sided Pr >=  P                0.3238
Two-sided = 2 * One-sided         0.6476


Sample Size = 19
```