Chapter 3 Sets of 2 × 2 Tables

3.1 Introduction

- Stratified analysis: one strategy for examining the association between two variables while adjusting for the effects of others.

Center	Treatment	Yes	No	Total
1	Test	29	16	45
1	Placebo	14	31	45
Total		43	47	90
2	Test	37	8	45
2	Placebo	24	21	45
Total		61	29	90

- Explanatory variables
- Research sites or hospitals in multi-center study
- Is there an association between the row and column variables in the tables, and what is the strength of that association?
- Investigating overall association instead of association in just one table

3.2 Mantel-Haenszel Test

Let $h = 1,2,\ldots,q$ index the strata, and consider the following table as representative of q 2×2 tables:

	Yes	No	Total
Group 1	n_{h11}	n_{h12}	n_{h1+}
Group 2	n_{h21}	n_{h22}	n_{h2+}
Total	n_{h+1}	n_{h+2}	n_h

Under the null hypothesis of no treatment difference:

$$E\left\{n_{h11} \middle| H_0\right\} = \frac{n_{h1+}n_{h+1}}{n_h} = m_{h11}$$

$$V\left\{n_{h11} \middle| H_0\right\} = \frac{n_{h1+}n_{h2+}n_{h+1}n_{h+2}}{n_h^2\left(n_h - 1\right)} = v_{h11}$$

(from hypergeometric distribution)

Mantel-Haenszel Statistic (1959)

$$Q_{MH} = \frac{\left\{ \sum_{h=1}^{q} n_{h11} - \sum_{h=1}^{q} m_{h11} \right\}^2}{\sum_{h=1}^{q} v_{h11}}$$

$$= \frac{\left\{ \sum_{h=1}^{q} \left(n_{h1+} n_{h2+} / n_h \right) \left(p_{h11} - p_{h21} \right) \right\}^2}{\sum_{h=1}^{q} v_{h11}}$$

$Q_{MH} \approx \chi^2(1)$ for large overall row sample sizes

Mantel-Fleiss criterion for sample size:

$$\min\left\{\left[\sum_{h=1}^{q} m_{h11} - \sum_{h=1}^{q}\left(n_{h11}\right)_L\right],\ \left[\sum_{h=1}^{q}\left(n_{h11}\right)_U - \sum_{h=1}^{q}\left(m_{h11}\right)\right]\right\} > 5$$

where $\left(n_{h11}\right)_L = \max\left(0,\ n_{h1+} - n_{h+2}\right)$

and $\left(n_{h11}\right)_U = \min\left(n_{h+1},\ n_{h1+}\right)$

3.2.1 Respiratory Data Example

```
data respire;
       input center treatment $ response $ count @@;
       n_response=(response='y');
       datalines;
       1 test    y 29 1 test    n 16
       1 placebo y 14 1 placebo n 31
       2 test    y 37 2 test    n 8
       2 placebo y 24 2 placebo n 21
       ;
run;

proc freq order=data;
  weight count;
  tables center*treatment*response /
       nocol nopct chisq cmh(mf);
run;
```

Output 3.1 Table 1 Results

Table 1 of treatment by response
Controlling for center=1

treatment / Frequency / Row Pct	response: y	n	Total
test	29 64.44	16 35.56	45
placebo	14 31.11	31 68.89	45
Total	43	47	90

Statistic	DF	Value	Prob
Chi-Square	1	10.0198	0.0015
Likelihood Ratio Chi-Square	1	10.2162	0.0014
Continuity Adj. Chi-Square	1	8.7284	0.0031
Mantel-Haenszel Chi-Square	1	9.9085	0.0016
Phi Coefficient		0.3337	
Contingency Coefficient		0.3165	
Cramer's V		0.3337	

Fisher's Exact Test

Cell (1,1) Frequency (F)	29
Left-sided Pr <= F	0.9997
Right-sided Pr >= F	0.0015
Table Probability (P)	0.0011
Two-sided Pr <= P	0.0029

Output 3.2 Table 2 Results

```
Table 2 of treatment by response
   Controlling for center=2

treatment     response

Frequency
Row Pct    y          n            Total
--------------------------------
test             37          8        45
              82.22      17.78
--------------------------------
placebo          24         21        45
              53.33      46.67
--------------------------------
Total            61         29        90
```

Statistics for Table 2 of treatment by response
Controlling for center=2

Statistic	DF	Value	Prob
Chi-Square	1	8.5981	0.0034
Likelihood Ratio Chi-Square	1	8.8322	0.0030
Continuity Adj. Chi-Square	1	7.3262	0.0068
Mantel-Haenszel Chi-Square	1	8.5025	0.0035
Phi Coefficient		0.3091	
Contingency Coefficient		0.2953	
Cramer's V		0.3091	

Fisher's Exact Test

Cell (1,1) Frequency (F)	37
Left-sided Pr <= F	0.9993
Right-sided Pr >= F	0.0031
Table Probability (P)	0.0025
Two-sided Pr <= P	0.0063

Output 3.3 Summary Statistics

```
              Summary Statistics for treatment by response
                          Controlling for center

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

    Statistic    Alternative Hypothesis    DF        Value       Prob
    ---------------------------------------------------------------
        1        Nonzero Correlation        1      18.4106     <.0001
        2        Row Mean Scores Differ     1      18.4106     <.0001
        3        General Association        1      18.4106     <.0001

                Mantel-Fleiss Criterion      36.0000
```

- CMH(MF) option in PROC FREQ gives the Mantel-Fleiss Criterion. Values <5 signify sample sizes inadequate for the assumption of an asymptotic chi-square distribution for Q_{MH}

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control	Mantel-Haenszel	4.0288	2.1057	7.7084
(Odds Ratio)	Logit	4.0286	2.1057	7.7072
Cohort	Mantel-Haenszel	1.7368	1.3301	2.2680
(Col1 Risk)	Logit	1.6760	1.2943	2.1703
Cohort	Mantel-Haenszel	0.4615	0.3162	0.6737
(Col2 Risk)	Logit	0.4738	0.3264	0.6877

Breslow-Day Test for
Homogeneity of the Odds Ratios

Chi-Square	0.0002
DF	1
Pr > ChiSq	0.9900

Total Sample Size = 180

- The magnitude of the difference in proportions (test – placebo) of the responses can be obtained through the GLM procedure

- Response variable `n_response` is dichotomized as 1 for Yes, 0 for No

- Center and Treatment as explanatory variables. ESTIMATE statement provides the estimated difference in proportions, averaged across center.

```
proc glm;
    class center treatment;
    freq count;
    model n_response = center treatment;
    estimate ‘direction’ treatment -1 1;
run;
```

- The output provides a difference in proportions of 0.311. The standard error given by the ESTIMATE statement is incorrect, as it assumes homogeneous variance for the groups.

- The correct variance v_d is given as a weighting (across center) of the continuity-corrected variances for the difference of proportions within each center. Details provided on page 53 of the text.

- A 100*(1-α)% confidence interval for the difference of proportion d (adjusted for center) can be obtained as

$$d \pm z_{1-\alpha/2} (v_d)^{1/2}$$

3.2.2 Health Policy Data Example

- Favorable or Unfavorable response tabulated by residence (Urban, Rural) and stress (Low, High)

Table 3.3 Health Policy Opinion Data

Residence	Stress	Favorable	Unfavorable	Total
Urban	Low	48	12	60
Urban	High	96	94	190
	Total	144	106	250
Rural	Low	55	135	190
Rural	High	7	53	60
	Total	62	188	250

Table 3.4 Pooled Health Policy Opinion Data

Stress	Favorable	Unfavorable	Total
Low	103	147	250
High	103	147	250
Total	206	294	500

When pooling by residence, there is clearly no association between stress and favorable response (Table 3.4). However, within each type of residence, there is a higher proportion of favorable responses for people with low stress than with high stress (Table 3.3).

```
data stress;
        input region $ stress $ outcome $ count @@;
        datalines;
urban low f 48 urban low u 12
urban high f 96 urban high u 94
rural low f 55 rural low u 135
rural high f 7 rural high u 53
;
proc freq order=data;
        weight count;
        tables region*stress*outcome /chisq cmh nocol nopct;
run;
```

Output 3.7 Summary Statistics

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

 Statistic     Alternative Hypothesis   DF        Value      Prob

          1    Nonzero Correlation      1         23.0502    <.0001

          2    Row Mean Scores Differ   1         23.0502    <.0001

          3    General Association      1         23.0502    <.0001
```

Note that d=0.2335. The favorable proportion is 0.2335 higher, on average, for the low stress groups than high stress, across regions.

3.2.3 Soft Drink Example

- Study participants tried a new soft drink for a week then were asked if they would switch from their preferred soft drinks to this new soft drink. The data are tabulated by gender and nationality of origin.

Table 3.5 Soft Drink Data

		Switch?		
Gender	Country	Yes	No	Total
Male	American	29	6	35
Male	British	19	15	34
Total		48	21	69
Female	American	7	23	30
Female	British	24	29	53
Total		31	52	83

```
data soft;
        input gender $ country $ question $ count @@;
        datalines;
male American y 29 male American n  6
male British  y 19 male British  n 15
female American y  7 female American n 23
female British  y 24 female British  n 29
;

ods graphics on;
proc freq order=data;
   weight count;
   tables gender*country*question / riskdiff(cl=(wald) correct) measures
        cmh plots=(riskdiffplot oddsratioplot(logbase=2));
run;
ods graphics off;
```

The SAS code uses the PLOTS option to create plots for the risk differences and for the odds ratios (on the log base 2 scale). The ODS GRAPHICS statements must be specified to obtain these plots.

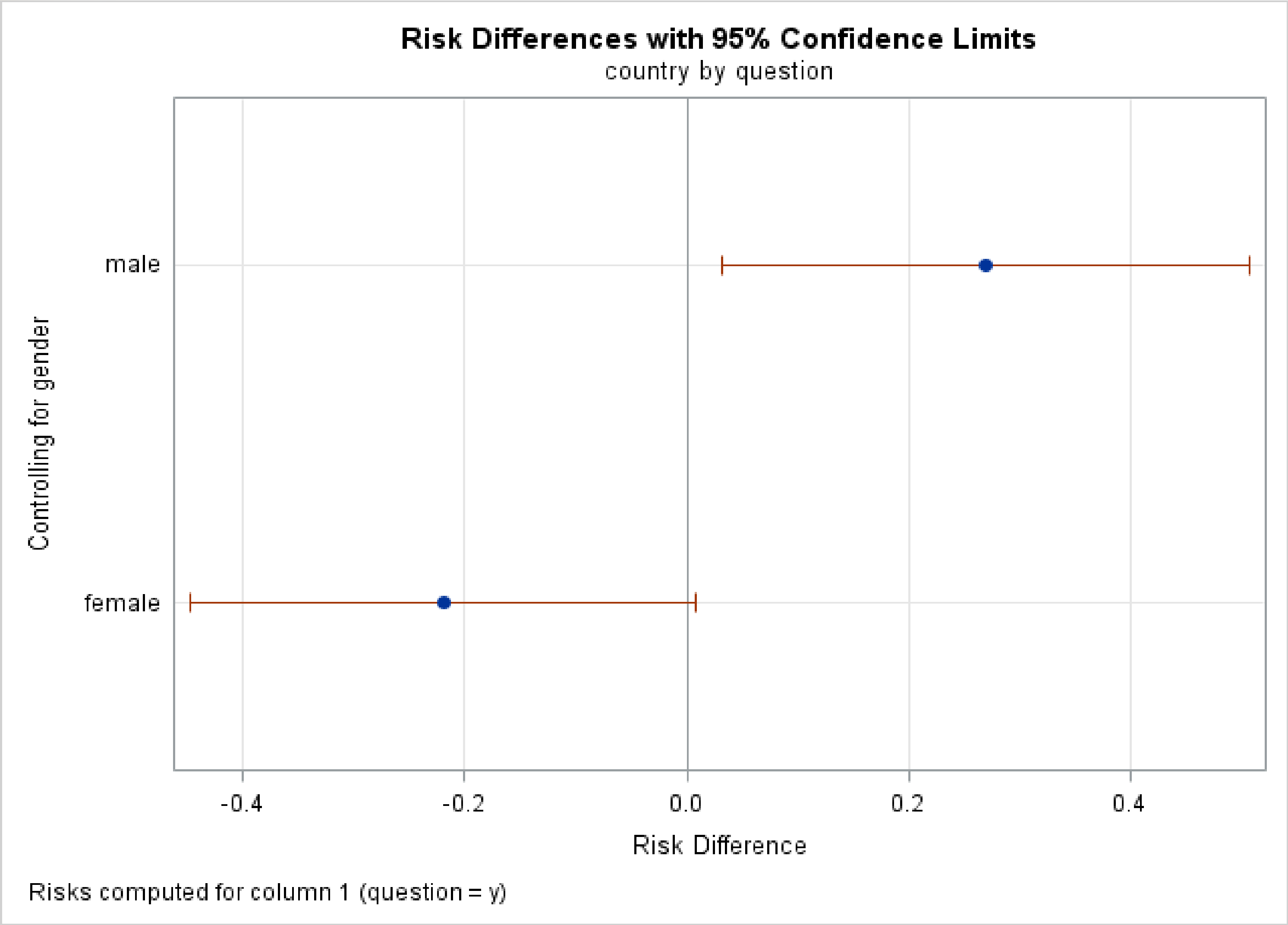
Output 3.10 Summary Statistics

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

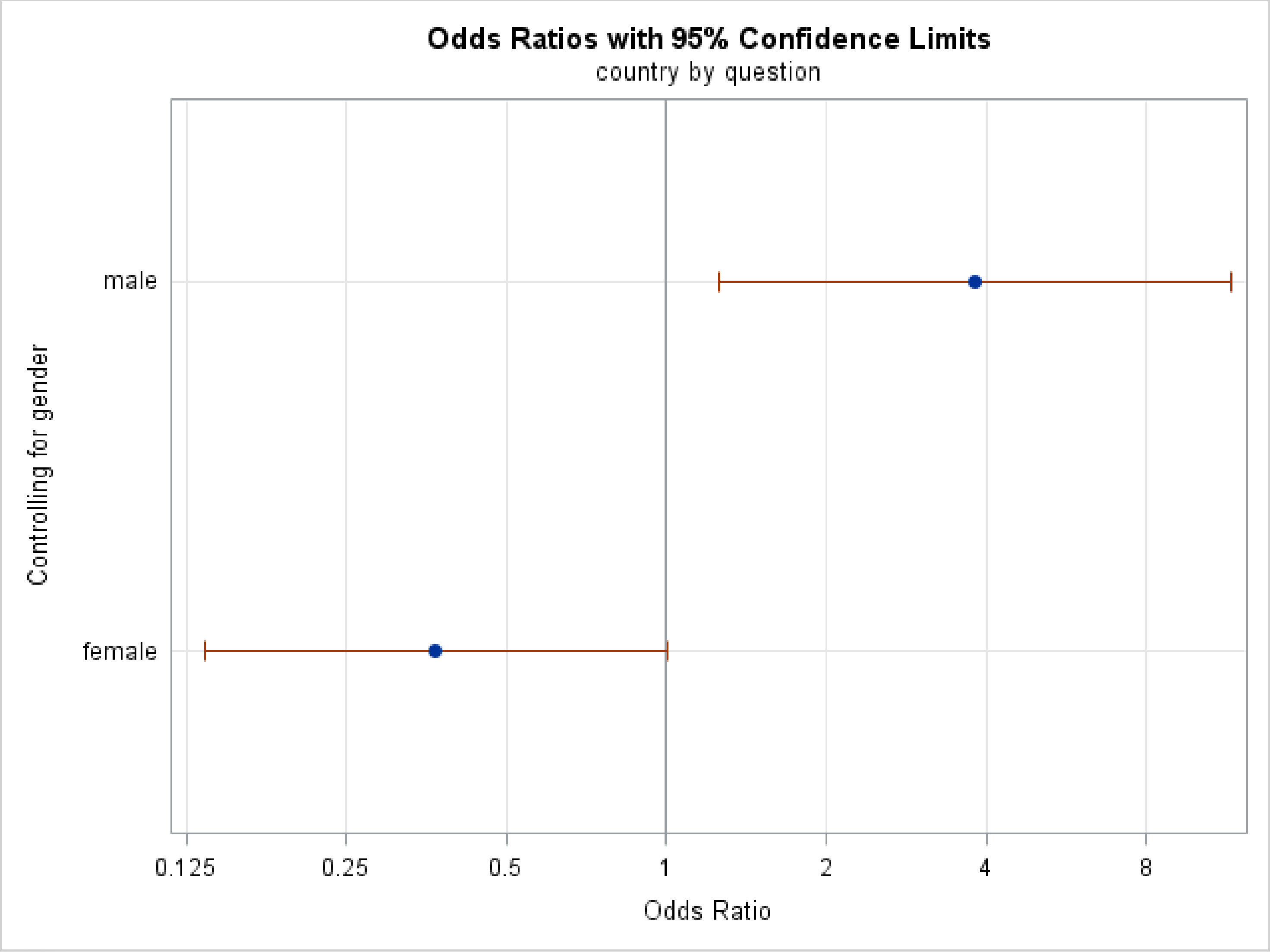
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0243	0.8762
2	Row Mean Scores Differ	1	0.0243	0.8762
3	General Association	1	0.0243	0.8762

There is insufficient evidence to conclude that country is associated with a willingness to switch sodas, after adjusting for gender (p=0.8762).

Output 3.11 Direction of Association for Soft Drink Data- Proportion Differences



Output 3.12 Direction of Association for Soft Drink Data- Odds Ratios



3.3 Measures of Association

- For the hth stratum:

$$OR_h = \frac{p_{h1}/(1-p_{h1})}{p_{h2}/(1-p_{h2})} = \frac{n_{h11}n_{h22}}{n_{h12}n_{h21}}$$

- OR_h estimates ψ_h, the population odds ratio for the hth stratum.

- If the ψ_h are homogeneous, the Mantel-Haenszel estimator for the common odds ratio ψ is

$$\hat{\psi}_{MH} = \sum_{h=1}^{q} \frac{n_{h11}n_{h22}}{n_h} \Bigg/ \sum_{h=1}^{q} \frac{n_{h12}n_{h21}}{n_h}$$

A 100(1-α)% confidence interval is given as

$$(\hat{\psi}_{MH}\exp(-z_{\alpha/2}\hat{\sigma}), \hat{\psi}_{MH}\exp(z_{\alpha/2}\hat{\sigma}))$$

where

$$\hat{\sigma}^2 = \widehat{var}[\log(\hat{\psi}_{MH})]$$

$$= \frac{\sum_h (n_{h11}+n_{h22})(n_{h11}n_{h22})/n_h^2}{2\left(\sum_h n_{h11}n_{h22}/n_h\right)^2}$$

$$+ \frac{\sum_h [(n_{h11}+n_{h22})(n_{h12}n_{h21}) + (n_{h12}+n_{h21})(n_{h11}n_{h22})]/n_h^2}{2\left(\sum_h n_{h11}n_{h22}/n_h\right)\left(\sum_h n_{h12}n_{h21}/n_h\right)}$$

$$+ \frac{\sum_h (n_{h12}+n_{h21})(n_{h12}n_{h21})/n_h^2}{2\left(\sum_h n_{h12}n_{h21}/n_h\right)^2}$$

This is the Robins, Breslow, and Greenland (1986) estimator for the log of the Mantel-Haenszel common odds ratio.

- Another estimator of ψ is the logit estimator, a weighted regression estimate defined as

$$\hat{\psi}_L = \exp\left\{ \sum_{h=1}^{q} w_h f_h \Big/ \sum_{h=1}^{q} w_h \right\} = \exp\left\{ \bar{f} \right\}$$

where f_h = log OR_h and

$$w_h = \left\{ \frac{1}{n_{h11}} + \frac{1}{n_{h12}} + \frac{1}{n_{h21}} + \frac{1}{n_{h22}} \right\}^{-1} = \left(1/v_{fh} \right)$$

where v_{fh} is the estimator for variance of f_h

- A $100(1 - \alpha)\%$ confidence interval for $\hat{\psi}_L$ is

$$\exp\left\{ \bar{f} \pm z_{\alpha/2} \left[\sum_{h=1}^{q} w_h \right]^{-1/2} \right\}$$

3.3.1 Homogeneity of Odds Ratios

- Breslow-Day test from SAS PROC FREQ. Determine

$$\hat{\psi}_{MH} = \sum_{h=1}^{q} \frac{n_{h11}n_{h22}}{n_h} \Big/ \sum_{h=1}^{q} \frac{n_{h12}n_{h21}}{n_h}$$

and solve $\frac{m_{h11}(n_h - n_{h1+} - n_{h+1} + m_{h11})}{(n_{h1+} - m_{h11})(n_{h+1} - m_{h11})} = \hat{\psi}_{MH}$

for m_{h11}, $m_{h12} = (n_{h1+} - m_{h11})$, $m_{h21} = (n_{h+1} - m_{h11})$, and $m_{h22} = (n_h - n_{h1+} - n_{h+1} + m_{h11})$ as estimated expected counts under $H_0 : \psi_h = \psi$ for homogeneity.

Determine:

$$Q_{BD} = \sum_{h=1}^{q} \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(n_{hij} - m_{hij})^2}{m_{hij}} \approx \chi^2(q-1) \text{ distribution}$$

- Other tests for homogeneity include Logit Chi-Square

$$Q_W = \sum_{h=1}^{q} \left(f_h - \bar{f} \right)^2 w_h, \text{ where } f_h = \log\left\{ \tfrac{n_{h11} n_{h22}}{n_{h12} n_{h21}} \right\},$$

$$\bar{f} = \frac{\sum_{h=1}^{q} w_h f_h}{\sum_{h=1}^{q} w_h} \text{ and } w_h = \left(\tfrac{1}{n_{h11}} + \tfrac{1}{n_{h12}} + \tfrac{1}{n_{h21}} + \tfrac{1}{n_{h22}} \right)^{-1} = \left(1/v_{f_h} \right).$$

$$\text{For } q = 2: \; Q_W = (f_1 - f_2)^2 / (v_{f_1} + v_{f_2}).$$

- Also, log-likelihood ratio statistic and score statistic from logistic regression can be used
- Pseudo-homogeneity $Q_{PH} = \sum_{h=1}^{q} \frac{(n_{h11} - m_{h11})^2}{v_{h11}} - Q_{MH}$,

where $m_{h11} = \frac{n_{h1+} n_{h+1}}{n_h}$, $v_{h11} = \frac{n_{h1+} n_{h2+} n_{h+1} n_{h+2}}{n_h^2 (n_h - 1)}$

3.3.2 Coronary Artery Disease Data Example

Gender	ECG	Disease	No Disease	Total
Female	< 0.1 ST segment depression	4	11	15
Female	≥ 0.1 ST segment depression	8	10	18
Male	< 0.1 ST segment depression	9	9	18
Male	≥ 0.1 ST segment depression	21	6	27

```
data ca;
    input gender  $  ECG  $  disease  $  count;
    datalines;
female   <0.1     yes      4
female   <0.1     no       11
female   >=0.1    yes      8
female   >=0.1    no       10
male     <0.1     yes      9
male     <0.1     no       9
male     >=0.1    yes      21
male     >=0.1    no       6
;
proc freq;
        weight count;
        tables gender*disease / nocol nopct chisq
        tables gender*ECG*disease / nocol nopct cmh chisq measures;
run;
```

Output 3.16 Stratified Analysis

Summary Statistics for ECG by disease
Controlling for gender

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	4.5026	0.0338
2	Row Mean Scores Differ	1	4.5026	0.0338
3	General Association	1	4.5026	0.0338

Output 3.17 Odds Ratios

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control	Mantel-Haenszel	2.8467	1.0765	7.5279
(Odds Ratio)	Logit	2.8593	1.0807	7.5650
Cohort	Mantel-Haenszel	1.6414	1.0410	2.5879
(Col1 Risk)	Logit	1.5249	0.9833	2.3647
Cohort	Mantel-Haenszel	0.6299	0.3980	0.9969
(Col2 Risk)	Logit	0.6337	0.4046	0.9926

Breslow-Day Test for
Homogeneity of the Odds Ratios

Chi-Square	0.2155
DF	1
Pr > ChiSq	0.6425

Total Sample Size = 78

3.4 Exact Confidence Limits for Common Odds Ratios for 2 × 2 Tables

- In SAS 9.3, PROC FREQ can give you exact confidence limits for the average odds ratio in a set of 2 × 2 tables through the EXACT COMOR statement.

- An analog to the Breslow-Day test for homogeneity of the odds ratio is Zelen's test. An exact p-value for Zelen's test can be obtained in PROC FREQ by adding EQOR to the EXACT statement (EXACT COMOR EQOR).

- Consider the following data:

Location	Program	Good	Not Good	Total
Downtown	Office	12	5	17
Downtown	Home	3	5	8
	Total	15	10	25
Satellite	Office	6	1	7
Satellite	Home	1	3	4
	Total	7	4	11

- We are interested in computing an odds ratio comparing good results for office vs. home programs. The sample sizes are small. Therefore, we will request exact confidence limits.

- The MF in CMH(MF) requests the Mantel-Fleiss criterion. The EXACT COMOR EQOR statement requests the exact confidence limits for the common odds ratio and Zelen's test.

```
data exercise;
      input location $ program $ outcome $ count @@;
      datalines;
downtown office  good 12 downtown  office not 5
downtown home    good 3  downtown  home   not 5
satellite office good 6  satellite office not 1
satellite home   good 1  satellite home   not 3
;
run;

proc freq order=data;
   weight count;
   tables location*program*outcome / cmh(mf);
   exact comor eqor;
run;
```

Output 3.19 Mantel-Fleiss Criterion

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

Statistic    Alternative Hypothesis         DF    Value   Prob
---------------------------------------------------------------
    1        Nonzero Correlation             1    5.5739 0.0182
    2        Row Mean Scores Differ          1    5.5739 0.0182
    3        General Association             1    5.5739 0.0182

    Mantel-Fleiss Criterion 4.6545 Warning: Criterion < 5
```

Output 3.20 Exact Confidence Limits

```
         Common Odds Ratio
 ------------------------------------
 Mantel-Haenszel Estimate      5.8421

 Asymptotic Conf Limits
 95% Lower Conf Limit          1.3012
 95% Upper Conf Limit         26.2296

 Exact Conf Limits
 95% Lower Conf Limit          1.0486
 95% Upper Conf Limit         33.3124
```

- Output 3.20 displays the exact estimate of the common odds ratio comparing office vs. home programs. Those persons participating in the office program had roughly 6 times the odds of a good test outcome vs. those in a home program. However, note that the exact confidence limits are very wide.

- The output also shows that the exact method provides a more conservative picture than the inappropriate asymptotic method.

- The exact confidence limits for the common odds ratio could also have been obtained through an exact logistic regression analysis using PROC LOGISTIC where outcome is the response, program and location are predictors, and an EXACT statement is used to obtain the exact odds ratio for program while conditioning on location.

<u>Output 3.21</u> Zelen's Test Results

```
      Tests for Homogeneity
         of Odds Ratios
------------------------------
Breslow-Day Chi-Square  0.6970
DF                           1
Pr > ChiSq              0.4038

Zelen's Exact Test (P)  0.4917
Exact Pr <= P           1.0000
```

•The exact p-value for Zelen's test is 1, indicating we do not have enough evidence to reject homogeneity of the odds ratio for program (by location).