Chapter 4: Sets of $2 \times r$ and $s \times 2$ Tables

4.1 Introduction

1) Sets of $2 \times r$ tables in which the column variable is ordinally scaled

- Investigating response variable with multiple ordered outcomes for a combined set of strata

- Comparing new treatment and placebo on extent of patient improvement rated minimal, moderate, substantial

2) Sets of $s \times 2$ tables in which the row variable is ordinally scaled

- Interested in trend of proportions across ordered groups for combined set of strata

- Comparing proportion of successful outcomes for different dosage levels of new drug

Statistical tests of no association between two groups and a response variable relative to the alternative of a location shift for the population represented by one group relative to that represented by the other

1. Let $i = 1, 2, \ldots, n$ index n patients with eligibility for random assignment to either of two groups. Let y_{ji} represent the response of patient i if assigned to group j where $j = 1, 2$. Consider the null hypothesis H_0: $y_{1i} = y_{2i} = y_i$ for all i (i.e., no association between groups and response)

2. Let $u_i = 1$ if patient i is assigned to group 1 and let $u_i = 0$ if i is assigned to group 2. With simple random sampling without replacement as the method for assigning n_1 patients to group 1 and $(n - n_1) = n_2$ patients to group 2 (i.e., equal probabilities for all possible $\frac{n!}{n_1!n_2!}$ random partitions of the patients into two groups with sample sizes n_1 and n_2),

$$E\{u_i\} = (n_1 / n)$$

$$Var\{u_i\} = (n_1 n_2 / n^2)$$

$$Cov\{u_i, u_{i'}\} = n_1 n_2 / n^2 (n-1).$$

3. Let $\bar{y}_1 = \left(\sum_{i=1}^{n} u_i y_i / n_1 \right)$ and let $\bar{y}_2 = \left(\sum_{i=1}^{n} (1-u_i) y_i / n_2 \right)$.

These statistics are the sample means for group1 and group 2.

Also,

$$(\bar{y}_1 - \bar{y}_2) = \sum_{i=1}^{n} y_i \{(n / n_1 n_2) u_i - (1/ n_2)\}$$

4. Under H_0 and conditional on the $\{y_i\}$,

$$E\{\bar{y}_1\} = \sum_{i=1}^{n} (y_i / n_1)(n_1 / n) = \left(\sum_{i=1}^{n} y_i / n \right) = \bar{y},$$

$$Var\{\bar{y}_1\} = \left\{ \sum_{i=1}^{n} (y_i^2 n_2 / n_1 n^2) - \sum_{i \neq i'}^{n} y_i y_{i'} n_2 / n_1 n^2 (n-1) \right\}$$

$$= \left(n_2 /(n-1) n_1 n \right) \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2$$

$$= \left(s^2 / n_1 \right) \left(1 - (n_1 / n) \right) = v_{\bar{y}}$$

Here $\bar{y}$ is the finite population mean for the $\{y_i\}$ and s^2 is the finite population variance; $(1 - (n_1/n))$ is the finite population correction for sampling variance of $\bar{y}_1$.

5. Let $Q = \left(\bar{y}_1 - \bar{y}\right)^2 / v_{\bar{y}}$. For sufficiently large sample sizes n_1 and n_2 (e.g., both ≥ 10), Q approximately has the chi-squared distribution with df = 1 under H_0. Note that

$$\left(\bar{y}_1 - \bar{y}\right) = \left\{\left(n_1 + n_2\right)\bar{y}_1 - \left(n_1\bar{y}_1 + n_2\bar{y}_2\right)\right\} / n =$$

$$n_2((\bar{y}_1 - \bar{y}_2) / n)$$

and so $Q = \left(\bar{y}_1 - \bar{y}_2\right)^2 / \left\{(1/n_1) + (1/n_2)\right\} s^2$.

6. When n_1 and n_2 are not large, evaluation of Q is possible through its exact distribution relative to the ($n!/\ n_1!\ n_2!$) possible realizations for the u_i under the method of randomized allocation of patients to groups.

7. The test statistic Q in (5) corresponds to

a. $(n - 1)/n$ times the Pearson chi-square statistic for all y_i as 0 or 1 for a dichotomous response

b. the Wilcoxon rank sum test when the y_i are ranks with midranks for ties

c. the Cochran-Armitage trend test when the y_i are consecutive integers or other natural scores for ordered categories

4.2 The $2 \times r$ Table

- Before discussing strategies for assessing association in sets of $2 \times r$ tables, consider first the single $2 \times r$ table with ordinal outcome

	Improvement			
Treatment	None	Some	Marked	Total
Active	13	7	21	41
Placebo	29	7	7	43
Total	42	14	28	84

- Define mean for Active group as

$$\bar{f}_1 = \sum_{j=1}^{3} \frac{a_j n_{1j}}{n_{1+}},$$

where $\boldsymbol{a} = \{a_j\} = (a_1, a_2, a_3)$ are a set of scores reflecting response levels

$$\text{Then } E\left\{ \bar{f}_1 \middle| H_0 \right\} = \sum_{j=1}^{3} \left(a_j \frac{n_{1+} n_{+j}}{n_{1+} n} \right) = \sum_{j=1}^{3} a_j \frac{n_{+j}}{n} = \mu_a$$

and

$$V\left\{ \bar{f}_1 \middle| H_0 \right\} = \frac{n - n_{1+}}{n_{1+}(n-1)} \sum_{j=1}^{3} \left(a_j - \mu_a \right)^2 \left(\frac{n_{+j}}{n} \right) = \frac{(n - n_{1+}) v_a}{n_{1+}(n-1)}$$

- Mean score statistic:

$$Q_S = \frac{\left(\bar{f}_1 - \mu_a\right)^2}{\left\{\left(n-n_{1+}\right)/\left[n_1+\left(n-1\right)\right]\right\}v_a},$$

and since $\bar{f}_1 \approx$ normally distributed, then $Q_S \approx$ distributed chi-square with 1 df.

Also, $Q_S = \dfrac{(\bar{f}_1 - \bar{f}_2)^2}{\left\{1/n_{1+} + 1/n_{2+}\right\}\left\{nv_a/(n-1)\right\}}$

- By taking advantage of the ordinality of the response variable, Q_S can test H_0 : *No association* vs. H_1 : *Location shifts* with fewer degrees of freedom

- ```
  proc freq data=arth order=data;
    weight count;
    tables treat*response / chisq nocol nopct;
  run;
  ```
  (See “Mantel-Haenszel χ^2” line of output)

  or

  ```
  proc freq data=arth order=data;
    weight count;
    tables treat*response / cmh nocol nopct;
  run;
  ```
  (See “Row Mean Scores Differ” line of output)
  ```

Mean Score Statistic

Table of treat by response

treat / Frequency / Row Pct	none	some	marked	Total
active	13 31.71	7 17.07	21 51.22	41
placebo	29 67.44	7 16.28	7 16.28	43
Total	42	14	28	84

Statistics for Table of treat by response

Statistic	DF	Value	Prob
Chi-Square	2	13.0550	0.0015
Likelihood Ratio Chi-Square	2	13.5298	0.0012
Mantel-Haenszel Chi-Square	1	12.8590	0.0003
Phi Coefficient		0.3942	
Contingency Coefficient		0.3668	
Cramer's V		0.3942	

Sample Size = 84

4.3 Sets of 2 × r Tables

- After first considering the single 2 × r table with ordinal outcome, now extend methodology to assess association in sets of 2 × r tables
- Let the following table be representative of q 2 × r tables, $h = 1,2,\ldots, q$

Level of Column Variable

	1	2	…	r	Total
Group 1	n_{h11}	n_{h12}	…	n_{h1r}	n_{h1+}
Group2	n_{h21}	n_{h22}	…	n_{h2r}	n_{h2+}
Total	n_{h+1}	n_{h+2}	…	n_{h+r}	n_h

- For rheumatoid arthritis data in table in textbook, $r = 3$ and $q = 2$
- n_{hij} represents number of patients in hth stratum who received ith treatment and had jth response
- Suppose $\{a_{hj}\}$ is a set of scores for response levels in hth stratum. Then sum of strata scores for 1st treatment is:

$$f_{+1+} = \sum_{h=1}^{2} \sum_{j=1}^{3} a_{hj} n_{h1j} = \sum_{h=1}^{2} n_{h1+} \bar{f}_{h1},$$

where $\bar{f}_{h1} = \sum_{j=1}^{3} \frac{a_{hj} n_{h1j}}{n_{h1+}}$ is mean score for Group 1

- Under null hypothesis:

$$E\left\{ f_{+1+} \middle| H_0 \right\} = \sum_{h=1}^{2} n_{h1+} \mu_h = \mu_*$$

and variance

$$V\left\{ f_{+1+} \middle| H_0 \right\} = \sum_{h=1}^{2} \frac{n_{h1+}(n_h - n_{h1+})}{(n_h - 1)} v_h = v_*,$$

where finite subpopulation mean $\mu_h = \sum_{j=1}^{3} \frac{a_{hj} n_{h+j}}{n_h}$

and variance for hth stratum $v_h = \sum_{j=1}^{3} \frac{\left(a_{hj} - \mu_h\right)^2 n_{h+j}}{n_h}$

- Extended Mantel-Haenszel mean score statistic:

$$Q_{S\,MH} = \frac{\left(f_{+1+} - \mu_*\right)^2}{v_*},$$

and since f_{+1+} is $\approx$ normally distributed if sample sizes n_{+i+} are sufficiently large, then $Q_{SMH} \approx$ distributed chi-square with 1 d.f.

Also, $Q_{SMH} = \left[\sum_{h=1}^{2} \left\{ \frac{n_{h1+} n_{h2+}}{(n_{h1+} + n_{h2+})} \left(\bar{f}_{h1} - \bar{f}_{h2} \right) \right\} \right]^2 / v_*$

4.3.1 Choosing Scores

- Integer scores

  $a_j = j$ for $j = 1, 2, \ldots, r$

  Useful when response levels are ordered categories that can be viewed as equally spaced and when response levels correspond to discrete counts

- Standardized midranks (or modified ridit scores)

$$a_j = \frac{2\left[\sum_{k=1}^{j} n_{+k}\right] - n_{+j} + 1}{2(n+1)}$$

The $\{a_j\}$ are constrained to lie between 0 and 1

Advantage over integer scores is they require no scaling of response levels other than that implied by relative ordering

- Logrank scores

$$a_j = 1 - \sum_{k=1}^{j} \left(\frac{n_{+k}}{\sum_{m=k}^{r} n_{+m}} \right)$$

Useful when distribution thought to be L-shaped, and there is greater interest in treatment differences for response levels with higher values than lower values

4.3.2 Analyzing the Arthritis Data

Gender = Female

Response

Treatment	None	Some	Marked	Total
Active	6	5	16	27
Placebo	19	7	6	32
Total	25	12	22	59

Gender = Male

Response

Treatment	None	Some	Marked	Total
Active	7	2	5	14
Placebo	10	0	1	11
Total	17	2	6	25

```
proc freq data=arth order=data;
    weight count;
    tables gender*treat*response / cmh nocol nopct;
    tables gender*treat*response / cmh nocol nopct
       scores=modridit;
run;
```

- Table Scores

  Row Mean Scores Differ Statistic = 14.63 ($p < 0.001$)

- Modified Ridit Scores

  Row Mean Scores Differ Statistic = 15.00 ($p < 0.001$)

4.3.3 Rank Statistics for Ordered Data

Example: Rheumatoid Arthritis Data

		Improvement		
Sex	Treatment	None	Some	Marked
Female	Placebo	19	7	6
Female	Active	6	5	16
Male	Placebo	10	0	1
Male	Active	7	2	5

$$\text{Mann-Whitney Estimator}: g_h = \sum_{j=1}^{3} p_{hAj} \left\{ \left(\sum_{k=1}^{j} p_{hPk} \right) - 0.5 p_{hPj} \right\}$$

$$= \Pr(A > P) + 0.5 \Pr(A = P)$$

$$= \sum_{j} \Pr(A = j) \{ \Pr(P \le j) - 0.5 \Pr(P = j) \}$$

Somer's *D* Statistics from PROC FREQ (using MEASURES option):

$$D = \frac{\{\Pr(A > P) - \Pr(A < P)\}}{\Pr(A > P) + \Pr(A < P) + \Pr(A = P)}$$

	Somer's *D*	Std. Err.
Females:	0.4664	0.1235
Males:	0.3961	0.1638

Transformation of Somer's *D* Index:

g_h = (Somer's *D* + 1) / 2

g_F = (0.4664 + 1) / 2 = 0.7332

g_M = (0.3961 + 1) / 2 = 0.6981

$SE(g_h) = [SE(\text{Somer's } D)]/2$

$$SE(g_F) = 0.12352 / 2 = 0.0618 \rightarrow v_F = 0.0618^2 = 0.003813$$

$$SE(g_M) = 0.16375 / 2 = 0.0819 \rightarrow v_M = 0.0819^2 = 0.006708$$

Test of Homogeneity: $Q_H = (g_\mathrm{F} - g_\mathrm{M})^2 / (v_\mathrm{F} + v_\mathrm{M})$

$= (0.7332 - 0.6981)^2 / (0.003813 + 0.006708)$

$= 0.1174$ ($p = 0.73$, d.f. = 1)

If homogeneous, common estimator is

$$\bar{g} = \left[\sum_{h=F}^{M} (g_h / v_h)\right] \Big/ \left[\sum_{h=F}^{M} (1/v_h)\right]$$

$$= \frac{\{(0.7332/0.0038) + (0.6981/0.0067)\}}{\{(1/0.0038) + (1/0.0067)\}}$$

$$= 0.72046$$

with its estimate of variance

$$v_{\bar{g}} = \left\{ \sum_{g=F}^{M} \left(1/v_h\right) \right\}^{-1}$$

$$= \left\{ \left(1/0.0038\right) + \left(1/0.0067\right) \right\}^{-1}$$

$$= 0.002431$$

and hypothesis test of common Mann-Whitney estimator = 1/2

$$Q_{\bar{g}} = \left(\bar{g} - 0.5\right)^2 / v_{\bar{g}}$$

$$= \left(0.72 - 0.5\right)^2 / 0.002431$$

$$= 19.99\left(p < 0.0001,\ \text{d.f.} = 1\right)$$

If sample sizes within strata are not large, you can use

$$\tilde{g} = \sum_{h=F}^{M} w_h g_h \Big/ \sum_{h=F}^{M} w_h$$

where $w_h = n_{h1} n_{h2} / (n_{h1} + n_{h2})$

The variance of $\tilde{g}$ can be calculated as

$$v_{\tilde{g}} = \sum_{h=F}^{M} w_h^2 (\text{s.e.}(g_h))^2 \Big/ \left[\sum_{h=F}^{M} w_h^2 \right]$$

PROC IML code in textbook computes these values in SAS

Applications of Exact Methods for Association between Groups and an Ordered Categorical Variable

Example: Patient Response status for rheumatoid arthritis – Wilcoxon Ranks (Source: Koch, et al (1982, Biometrics, 563-595))

Response	Active	Placebo	Total
Excellent	5	2	7
Good	11	4	15
Moderate	5	7	12
Fair	1	7	8
Poor	5	12	17
Total	27	32	59

```
proc freq order=data;
   weight count;
   table resp*treat / nocol norow nopct scores=modridit;
   exact mhchi scorr;
run;
```

```
                                                        Prob         Prob
Statistic                      DF        Value   (Asymptotic)   (Exact)
-------------------------------------------------------------------------
Chi-Square                      4      11.9300      0.0179
Likelihood Ratio Chi-Square     4      12.6678      0.0130
MH Chi-Square (Mod. Ridits)     1       8.7284      0.0031         0.0028 ⇐
```

```
Spearman Correlation Coefficient
--------------------------------
Correlation (r)              0.3879
ASE                          0.1185
95% Lower Conf Bound         0.1558
95% Upper Conf Bound         0.6201

  Test of H0: Correlation = 0

ASE under H0                 0.1188
Z                            3.2646
One-sided Pr >  Z            0.0005
Two-sided Pr > |Z|           0.0011

Exact Test
One-sided Pr >=  r           0.0014
Two-sided Pr >= |r|          0.0028 ⇐
```

4.3.4 Colds Example

Gender	Residence	Periods with Colds			Total
		0	1	2	
Female	Urban	45	64	71	180
Female	Rural	80	104	116	300
Total		125	168	187	480
Male	Urban	84	124	82	290
Male	Rural	106	117	87	310
Total		190	241	169	600

Is there an association between residence (urban or rural) and number of periods with colds (0,1, or 2), controlling for gender?

- Number of periods with colds can be considered an ordinal variable in which the levels are equally spaced

- The usual ANOVA strategy for interval-scaled response is not appropriate because periods with colds may not be normally distributed with homogenous variance

- An extended Mantel-Haenszel analysis is more appropriate

```
proc freq data=colds order=data;
   weight count;
   tables gender*residence*per_cold / all nocol nopct;
run;
```

Summary Statistics for residence by per_cold
Controlling for gender

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.7379	0.3903
2	Row Mean Scores Differ	1	0.7379	0.3903
3	General Association	2	1.9707	0.3733

$Q_{SMH} = 0.7379$ with p-value 0.3903. There appears to be no association between residence and number of periods with cold for these data, controlling for gender.

- You can compute a weighted difference of means which serves as a distance measure (effect size), similarly to as in Chapter 3:

$$d = \frac{\sum_h w_h(\bar{f}_{h1} - \bar{f}_{h2})}{\sum_h w_h}$$

$$v_d = \sum_h w_h^2 \left\{ \frac{v_{h1}}{n_{h1+}} + \frac{v_{h2}}{n_{h2+}} \right\} \Big/ \left(\sum_h w_h \right)^2$$

where

$$v_{hi} = \sum_{j=1}^{r} (a_{hj} - \bar{f}_{hi})^2 n_{hij} \Big/ (n_{hi+} - 1)$$

and

$$w_h = (n_{h1+} n_{h2+} / n_h)$$

You can use the GLM procedure for computing d, although you must use the formula on the prior slide for computing v_d

```
proc glm;
   class gender residence;
   freq count;
   model per_cold = gender residence;
   estimate 'd' residence 1 -1;
run;
```

Here, $d = 0.0416$ and v_d is calculated to be 0.0023

4.4 The $s \times 2$ Table

Father's Usage	Risk Perception	Adolescent Usage		Total
		No	Yes	
No	Minimal	59	25	84
No	Moderate	169	29	198
No	Substantial	196	9	205
Yes	Minimal	11	8	19
Yes	Moderate	33	11	44
Yes	Substantial	22	2	24

- Is there a discernable trend in proportions of adolescent usage over levels of risk perception? Does usage decline with higher risk perception?

Father's Usage	Risk Perception	Adolescent Usage		Total
		No	Yes	
No	Minimal	59	25	84
No	Moderate	169	29	198
No	Substantial	196	9	205

- $$\bar{f} = \sum_{i=1}^{3} c_i \bar{f}_i \left(\frac{n_{i+}}{n}\right) = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{c_i a_j n_{ij}}{n},$$

  where $\boldsymbol{c} = (c_1, c_2, c_3)$ represents scores for the groups and $\boldsymbol{a} = (a_1, a_2)$ represents scores for the columns

- Then $E\left\{\bar{f}\,|H_0\right\} = \sum_{i=1}^{3} c_i \left(\frac{n_{i+}}{n}\right) \sum_{j=1}^{2} a_j \left(\frac{n_{+j}}{n}\right) = \mu_c \mu_a$

and

$$V\left\{\bar{f}\,|H_0\right\} = \sum_{i=1}^{3} (c_i - \mu_c)^2 \left(\frac{n_{i+}}{n}\right) \sum_{j=1}^{2} \frac{(a_j - \mu_a)^2 \left(\frac{n_{+j}}{n}\right)}{(n-1)}$$

$$= \frac{v_c v_a}{(n-1)}$$

- For large samples, $\bar{f}$ has approximate normal distribution. Thus, the correlation statistic is calculated as follows

$$Q_{CS} = \frac{\left(\bar{f} - E\left\{\bar{f}|H_0\right\}\right)^2}{V\left\{\bar{f}|H_0\right\}}$$

$$= \frac{(n-1)\left[\sum_{i=1}^{3}\sum_{j=1}^{2}(c_i-\mu_c)(a_j-\mu_a)n_{ij}\right]^2}{\left[\sum_{i=1}^{3}(c_i-\mu_c)^2 n_{i+}\right]\left[\sum_{j=1}^{2}(a_j-\mu_a)^2 n_{+j}\right]}$$

$$= (n-1)r_{ac}^2,$$

where r_{ac} is the Pearson correlation coefficient.

Thus, Q_{CS} is ≈ chi-square with 1 d.f.

```
data tobacco;
   length risk $11. ;
   input f_usage $ risk $ usage $ count @@;
   datalines;
no  minimal     no    59  no  minimal     yes  25
no  moderate    no   169  no  moderate    yes  29
no  substantial no   196  no  substantial yes   9
yes minimal     no    11  yes minimal     yes   8
yes moderate    no    33  yes moderate    yes  11
yes substantial no    22  yes substantial yes   2
;
run;


 proc freq;
    weight count;
    tables f_usage*risk*usage / cmh chisq measures trend;
 run;
```

Results for No Father's Usage

Statistics for Table 1 of risk by usage
Controlling for f_usage=no

Statistic	DF	Value	Prob
Chi-Square	2	34.9217	<.0001
Likelihood Ratio Chi-Square	2	34.0684	<.0001
Mantel-Haenszel Chi-Square	1	34.2843	<.0001
Phi Coefficient		0.2678	
Contingency Coefficient		0.2587	
Cramer's V		0.2678	

Cochran-Armitage Trend Test

Statistics for Table 1 of risk by usage
Controlling for f_usage=no

Cochran-Armitage Trend Test

Statistic (Z)	5.8613
One-sided Pr > Z	<.0001
Two-sided Pr > |Z|	<.0001

Sample Size = 487

Statistics for Table 1 of risk by usage
Controlling for f_usage=no

Statistic	Value	ASE
Gamma	-0.5948	0.0772
Kendall's Tau-b	-0.2477	0.0395
Stuart's Tau-c	-0.1863	0.0339
Somers' D C|R	-0.1484	0.0267
Somers' D R|C	-0.4135	0.0628
Pearson Correlation	-0.2656	0.0439
Spearman Correlation	-0.2602	0.0415
Lambda Asymmetric C|R	0.0000	0.0000
Lambda Asymmetric R|C	0.0709	0.0211
Lambda Symmetric	0.0580	0.0169
Uncertainty Coefficient C|R	0.0908	0.0290
Uncertainty Coefficient R|C	0.0339	0.0112
Uncertainty Coefficient Symmetric	0.0493	0.0161

4.5 Sets of $s \times 2$ Tables

4.5.1 Correlation Statistic

- Extended Mantel-Haenszel correlation statistic for the association of two variables that were ordinal for a combined set of strata

$$Q_{CSMH} = \frac{\left\{ \sum_{h=1}^{q} n_h \left(\bar{f}_h - E\{ \bar{f}_h | H_0 \} \right) \right\}^2}{\sum_{h=1}^{q} n_h^2 V\{ \bar{f} | H_0 \}}$$

$$= \frac{\left\{ \sum_{h=1}^{q} n_h \sqrt{v_{hc} v_{ha}}\, r_{ca,h} \right\}^2}{\sum_{h=1}^{q} \left[n_h^2 v_{hc} v_{ha} / (n_h - 1) \right]}$$

- $Q_{CSMH} \approx$ follows chi-square distribution with 1 df when combined strata sample sizes are sufficiently large

$$\sum_{h=1}^{q} n_h \geq 40$$

4.5.2 Analysis of Smokeless Tobacco Data

```
proc freq;
 weight count;
 tables f_usage*risk*usage / cmh;
 tables f_usage*risk*usage / cmh scores=modridit;
run;
```

Evaluate "Nonzero Correlation" statistic on Mantel-Haenszel output (Q_{CSMH})

Results for Combined Tables

Summary Statistics for risk by usage
Controlling for f_usage

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	40.664	<0.0001
2	Row Mean Scores Differ	2	41.058	<0.0001
3	General Association	2	41.058	<0.0001

Cochran-Mantel-Haenszel Statistics (Modified Ridit Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	39.305	<0.0001
2	Row Mean Scores Differ	2	41.083	<0.0001
3	General Association	2	41.058	<0.0001

Total Sample Size = 574

4.5.3 Pain Data Analysis

	Diagnosis I		Diagnosis II	
	Adverse Effects		Adverse Effects	
Treatment	No	Yes	No	Yes
Placebo	26	6	26	6
Dosage 1	26	7	12	20
Dosage 2	23	9	13	20
Dosage 3	18	14	1	31
Dosage 4	9	23	1	31

```
proc freq order=data;
      weight count;
      tables diagnosis*treatment*response / cmh;
run;
```

Summary Statistics for treatment by response
Controlling for diagnosis

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	71.726	<0.0001
2	Row Mean Scores Differ	4	74.531	<0.0001
3	General Association	4	74.531	<0.0001

Summary of Extended Mantel-Haenszel Statistics

Table Dimensions	Statistic	DF	Corresponding PROC FREQ MH Label
2×2	Q_{MH}	1	Nonzero Correlation Row Mean Scores Differ General Association
$2 \times r$	Q_{SMH}	1	Nonzero Correlation Row Mean Scores Differ
$s \times 2$	Q_{CSMH}	1	Nonzero Correlation

4.6 Relationships between Sets of Tables

- Transpose rows and columns of previous table, and analyze as two 2 × r tables

Diagnosis	Adverse Effects	Placebo	Dose 1	Dose 2	Dose 3	Dose 4
I	No	26	26	23	18	9
I	Yes	6	7	9	14	23
II	No	26	12	13	1	1
II	Yes	6	20	20	31	31

```
proc freq order=data;
      weight count;
      tables diagnosis*response*treatment / cmh;
run;
```

Transposed Analysis

Summary Statistics for response by treatment
Controlling for diagnosis

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	71.726	<0.0001
2	Row Mean Scores Differ	1	71.726	<0.0001
3	General Association	4	74.531	<0.0001

Original Analysis

Summary Statistics for treatment by response
Controlling for diagnosis

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	71.726	<0.0001
2	Row Mean Scores Differ	4	74.531	<0.0001
3	General Association	4	74.531	<0.0001

4.7 Exact Analysis of Association for the $s \times 2$ Table

Mice Surviving Exposure to Vibrio Vulnificus

Hours	Carbenicillin	Cefotaxime	Total	Ranks	Logranks
0-6	1	1	2	1.5	0.909
6-12	3	1	4	4.5	0.709
12-18	5	1	6	9.5	0.334
18-24	1	0	1	13	0.234
24-30	1	2	3	15	-0.099
30-48	0	2	2	17.5	-0.433
48-72	1	1	2	19.5	-0.933
72-96	0	1	1	21	-1.433
>96	0	1	1	22	-2.433
Total	12	10	22		

```
proc sort data=mice;
   by LogRank;
run;

proc freq;
   weight count;
   tables LogRank*Treatment / norow nocol nopct scorout chisq;
   tables LogRank*Treatment / noprint scores=rank scorout chisq;
   exact mhchi;
run;
```

Row Scores

LogRank	Score
-2.433	-2.433
-1.433	-1.433
-0.933	-0.933
-0.433	-0.433
-0.099	-0.099
0.234	0.234
0.334	0.334
0.709	0.709
0.909	0.909

MH Chi-Square Test for Logrank Scores

```
          Mantel-Haenszel Chi-Square Test
          -------------------------------
          Chi-Square                  4.0569
          DF                               1
          Asymptotic Pr > ChiSq       0.0440
          Exact      Pr > ChiSq       0.0367
```

MH Chi-Square Test for Rank Scores

```
          Mantel-Haenszel Chi-Square Test
                   (Rank Scores)
          -------------------------------
          Chi-Square                  3.5118
          DF                               1
          Asymptotic Pr > ChiSq       0.0609
          Exact      Pr > ChiSq       0.0625
```