Chapter 5: The $s \times r$ Table

- Extends concepts in previous chapters to address tests of association for general $s \times r$ table

5.2 Association

5.2.1 Tests for General Association

- Example

Party	Neighborhood			
	Bayside	Highland	Longview	Sheffeld
Democrat	221	160	360	140
Independent	200	291	160	311
Republican	208	106	316	97

- General $s \times r$ table

Group	Response Variable Categories				Total
	1	2	…	r	
1	n_{11}	n_{12}	…	n_{1r}	n_{1+}
2	n_{21}	n_{22}	…	n_{2r}	n_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
s	n_{s1}	n_{s2}	…	n_{sr}	n_{s+}
Total	n_{+1}	n_{+2}	…	n_{+r}	n

H_0 : Distributions of response are comparable to that for randomized groups

$$\Pr\{n_{ij}\} = \frac{\prod_{i=1}^{s} n_{i+}! \prod_{j=1}^{r} n_{+j}!}{n! \prod_{i=1}^{s} \prod_{j=1}^{r} n_{ij}!}$$

- Pearson chi-square statistic for general association

$$Q_P = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{\left(n_{ij} - m_{ij}\right)^2}{m_{ij}}$$

where $m_{ij} = E\{n_{ij} \mid H_0\} = \dfrac{n_{i+} n_{+j}}{n}$

- If sample size is sufficiently large (all expected counts $m_{ij} \geq 5$), $Q_P \approx$ chi-square distribution with $(s - 1)(r - 1)$ degrees of freedom. In case of 2×2 table, $r = 2$ and $s = 2$ so that Q_P has 1 d.f.

- Just as for 2 × 2 tables, randomization statistic Q can be written

$$Q = \tfrac{n-1}{n} Q_P$$

and is also ≈ chi-square with $(s - 1)(r - 1)$ d.f.

- Covariance structure under H_0 is

$$Cov\left\{n_{ij}, n_{i'j'} \middle| H_0\right\} = \frac{m_{ij}(n\delta_{ii'} - n_{i'+})(n\delta_{jj'} - n_{+j'})}{n(n-1)}$$

where $\delta_{kk'} = 1$ if $k = k'$ and $\delta_{kk'} = 0$ if $k \neq k'$

- The randomization statistic Q is computed from the quadratic form:

$$Q = (\boldsymbol{n}-\boldsymbol{m})'\boldsymbol{A}'(\boldsymbol{AVA}')^{-1}\boldsymbol{A}(\boldsymbol{n}-\boldsymbol{m})$$

where $\boldsymbol{n} = \left(n_{11}, n_{12}, \ldots, n_{1r}, \ldots, n_{s1}, \ldots, n_{sr}\right)'$ is

vector of observed frequencies, ***m*** is corresponding vector of expected frequencies, ***V*** is covariance matrix, and ***A*** is matrix containing scores

- Usual choice for $\boldsymbol{A}$ for testing general association is:

$$\boldsymbol{A} = \left[\mathbf{I}_{(r-1)}, \mathbf{0}_{(r-1)}\right] \otimes \left[\mathbf{I}_{(s-1)}, \mathbf{0}_{(s-1)}\right]$$

where $\mathbf{I}_{(j-1)}$ is the $(j-1) \times (j-1)$ identity matrix, $\mathbf{0}_{(j-1)}$ is a $(j-1)$ vector of 0's, and $\otimes$ denotes left-hand Kronecker product (matrix on left multiplies each element of matrix on right)

- To calculate statistic *Q* for table of political party and neighborhood data included previously:

```
ods graphics on;
proc freq data=neighbor;
   weight count;
   tables party*neighborhood /
     plots=mosaicplot chisq cmh nocol nopct;
run;
ods graphics off;
```

(See “General Association” results)

- PLOTS=MOSAIC requests mosaic plot

- A mosaic plot uses tiles proportional to cell frequencies to provide a visual representation of the contingency table

  - Horizontal space represents relative count size of the variable on the x-axis and the vertical space represents relative count size of the variable on the y-axis

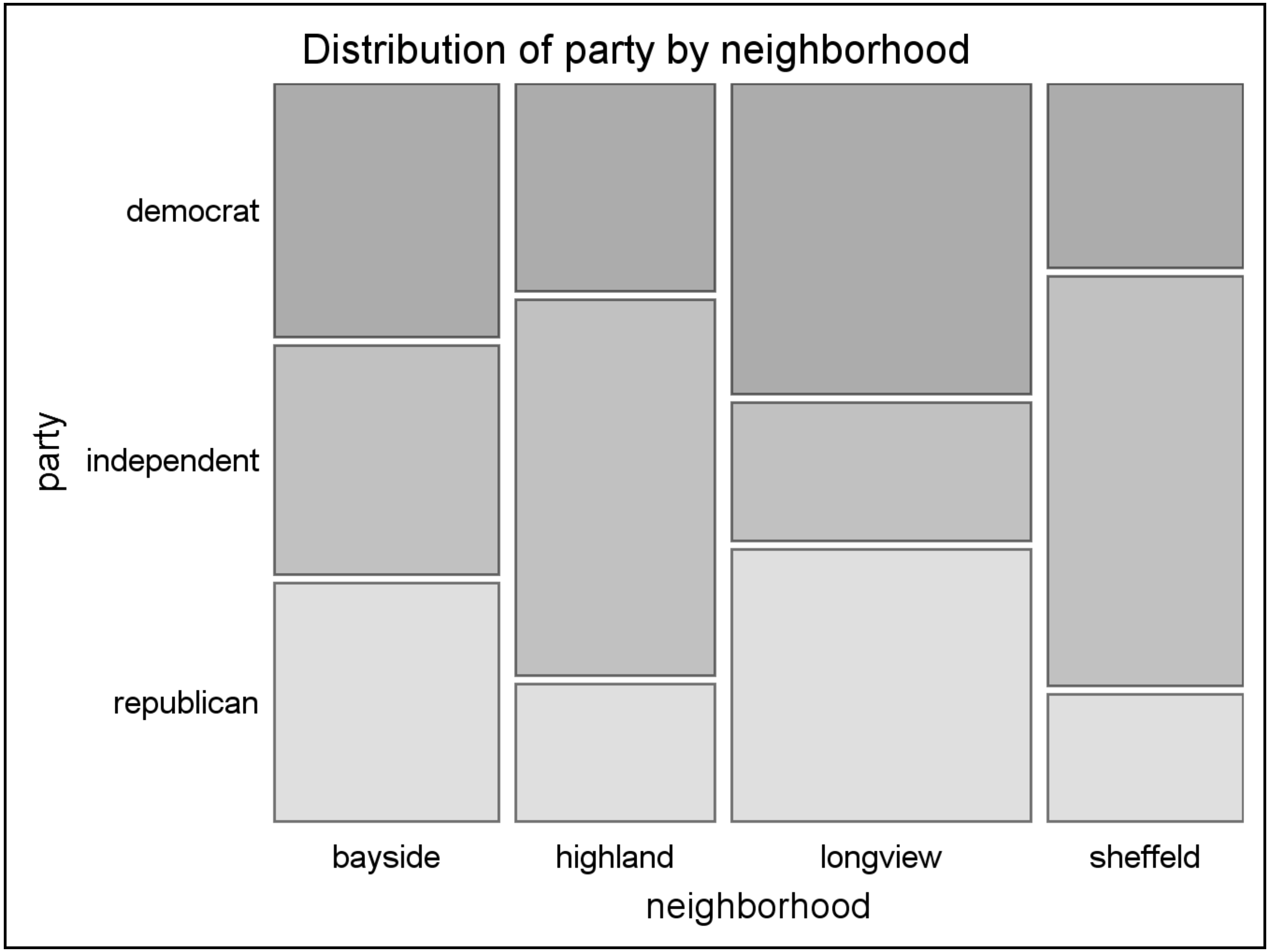
Frequency Table

Table of party by neighborhood

party — neighborhood

Frequency	bayside	highland	longview	sheffeld	Total
democrat	221	160	360	140	881
independent	200	291	160	311	962
republican	208	106	316	97	727
Total	629	557	836	548	2570

Mosaic Plot



Distribution of party by neighborhood
party
democrat
independent
republican
bayside
highland
longview
sheffeld
neighborhood

Pearson Chi-Square

STATISTICS FOR TABLE OF PARTY BY NEI_HOOD

Statistic	DF	Value	Prob
Chi-Square	6	273.919	0.001
Likelihood Ratio Chi-Square	6	282.327	0.001
Mantel-Haenszel Chi-Square	1	0.812	0.367
Phi Coefficient		0.326	
Contingency Coefficient		0.310	
Cramer's V		0.231	

Randomization *Q*

SUMMARY STATISTICS FOR PARTY BY NEI_HOOD

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.812	0.3674
2	Row Mean Scores Differ	2	13.894	0.0010
3	General Association	6	273.812	<.0001

5.2.2 Mean Score Test

Treatment	Hours of Relief				
	0	1	2	3	4
Placebo	6	9	6	3	1
Standard	1	4	6	6	8
Test	2	5	6	8	6

- Now $\boldsymbol{A}$ is chosen to assign scores to response levels and compare resulting linear functions of scores for $(s - 1)$ groups to their expected values

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}' & \mathbf{0}' & \cdots & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0}' & \boldsymbol{a}' & \cdots & \mathbf{0}' & \mathbf{0}' \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & \boldsymbol{a}' & \mathbf{0}' \end{bmatrix}$$

where $\boldsymbol{a}' =$ (0 1 2 3 4) for the above table if actual values were used as scores. With this choice of $\boldsymbol{A}$, the randomization Q becomes the mean score statistic Q_S

- The mean score statistic can be written as:

$$Q_S = \frac{(n-1)\sum_{i=1}^{s} n_{i+}(\bar{f}_i - \mu_{\mathbf{a}})^2}{n v_a}$$

where $\bar{f}_i = \sum_{j=1}^{r} \frac{a_j n_{ij}}{n_{i+}}$ with expected value

$\mu_a = E\left\{ \bar{f}_i \middle| H_0 \right\} = \sum_{j=1}^{r} \frac{a_j n_{+j}}{n}$, and variance

$$v_a = \sum_{j=1}^{r} \left(a_j - \mu_a\right)^2 \left(\frac{n_{+j}}{n}\right)$$

- ```
  proc freq data=pain;
     weight count;
     tables treatment*hours / cmh nocol nopct;
  run;
  ```
  (See “Row Mean Scores Differ” results)

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	8.0668	0.0045
2	Row Mean Scores Differ	2	13.7346	0.0010
3	General Association	8	14.4030	0.0718
  ```

7.2 Kruskal-Wallis Test

- Kruskal-Wallis test (1952) is generalization of two-sample Wilcoxon-Mann-Whitney test to three or more groups:

  H_0: Distribution of response variable is same in multiple independently sampled populations

  vs H_1: Location difference among populations

- Kruskal-Wallis test can be used whenever one-way analysis of variance model is appropriate

- For small samples: can use EXACT statement in NPAR1WAY procedure, or StatXact package (Mehta and Patel 1991)

- For larger samples (at least 5 obs per group): p-value approximated using asymptotic chi-square distribution with $s - 1$ degrees of freedom (s is number of groups), which is Mantel-Haenszel mean score statistic with rank scores

- Example: Data from study of anecubital vein cortisol levels at time of delivery in pregnant women. Are median cortisol levels different among three groups of women?

Group I: elective C-section
Group II: emergency C-section (induced labor)
Group III: vaginal or C-section (spontaneous)

Group I		Group II		Group III	
Patient	Level	Patient	Level	Patient	Level
1	262	1	465	1	343
2	307	2	501	2	772
3	211	3	455	3	207
4	323	4	355	4	1048
5	454	5	468	5	838
6	339	6	362	6	687
7	304				
8	154				
9	287				
10	356				

- To compare the three groups using the Kruskal-Wallis test, compute the Mantel-Haenszel mean score statistic with rank scores:

```
proc freq;
    tables group*cortisol / noprint cmh2
                            scores=rank;
run;
```

- Results show the three groups of subjects differ with respect to cortisol level (row mean score chi-square = 9.232, 2 df, p = 0.010)
- Can also use NPAR1WAY and WILCOXON option

5.2.3 Correlation Test

Treatment	Washability			
	Low	Med	High	Total
Water	27	14	5	46
Standard	10	17	26	53
Super	5	12	50	67

- Now A is chosen to assign scores both to levels of response variable and grouping variable to obtain Q_{CS}

$$\boldsymbol{A} = [\boldsymbol{a}' \otimes \boldsymbol{c}'] = [a_1 c_1, \ldots, a_r c_1, \ldots, a_r c_s]$$

where $\boldsymbol{a'} = (a_1, a_2, \ldots, a_r)$ are scores for response levels, and $\boldsymbol{c'} = (c_1, c_2, \ldots, c_s)$ are scores for levels of grouping variable, and $\boldsymbol{A}$ has dimension $1 \times sr$

```
proc freq order=data;
   weight count;
   tables treatment*washability / chisq cmh nocol nopct;
run;
```

```
        Summary Statistics for treatment by washability

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

Statistic    Alternative Hypothesis    DF       Value      Prob
---------------------------------------------------------------
    1        Nonzero Correlation        1     50.6016    <.0001
    2        Row Mean Scores Differ     2     52.7786    <.0001
    3        General Association        4     54.7560    <.0001
```

5.3 Exact Tests for Association

Type of Car	Advertising Source				
	TV	Magazine	Paper	Radio	Total
Sedan	4	0	0	2	6
Sporty	0	3	3	4	10
Utility	5	5	2	2	14

- Sample size does not meet requirements for usual tests of association via Pearson or randomization χ^2

- ```
  proc freq data=market;
     weight count;
     table car*AdSource / norow nocol nopct;
     exact fisher pchi lrchi;
  run;
  ```
  ```

Exact Test Results

```
    Pearson Chi-Square Test
__________________________________
Chi-Square                 11.5984
DF                               6
Asymptotic Pr >  ChiSq      0.0716
Exact      Pr >= ChiSq      0.0664

 Likelihood Ratio Chi-Square Test
__________________________________
Chi-Square                 16.3095
DF                               6
Asymptotic Pr >  ChiSq      0.0122
Exact      Pr >= ChiSq      0.0272

       Fisher's Exact Test
__________________________________
Table Probability (P)    2.545E-05
Pr <= P                     0.0473

        Sample Size = 30
```

5.3.2 Test of Correlation

- Exact *p*-values are available for the correlation test for the case when both the rows and columns of a table are ordinally scaled.

Dose in Mg	Response				
	Poor	Fair	Good	Excellent	Total
25	1	1	1	0	3
50	1	2	1	1	5
75	0	0	2	2	4
100	0	0	7	0	7

- Since both rows and columns can be considered ordinal, the type of association is linear and the correlation MH statistic is suitable. However, note that there are several zero/small cells.

- Compute exact *p*-value (sum of the exact *p*-values associated with the tables where the test statistic is larger than the one we observe):

```
proc freq data=disorder;
   weight count;
   tables dose*outcome / nocol norow nopct measures;
     exact mhchi;
run;
```

Exact Results for Correlation MH

```
Mantel-Haenszel Chi-Square Test
------------------------------
Chi-Square                  3.9314
DF                               1
Asymptotic Pr >  ChiSq      0.0474
Exact      Pr >= ChiSq      0.0488

        Sample Size = 19
```

5.4.1 Ordinal Measures of Association

- Can be useful when data do not lie on an obvious scale, but are ordinal in nature

- Examples:
  - Spearman rank correlation coefficient is produced by substituting ranks as variable values for the Pearson correlation coefficient
  - Gamma, Kendall's tau-*b*, Stuart's tau-*c*, and Somers' D statistics are based on concordant and discordant pairs

- To compute, request MEASURES option:

```
proc freq order=data;
   weight count;
   tables treatment*washability/measures noprint nocol nopct cl;
run;
```

Measures of Association

Statistics for Table of treatment by washability

Statistic	Value	ASE	95% Confidence Limits	
Gamma	0.6974	0.0636	0.5728	0.8221
Kendall's Tau-b	0.4969	0.0553	0.3885	0.6053
Stuart's Tau-c	0.4803	0.0545	0.3734	0.5872
Somers' D C|R	0.4864	0.0542	0.3802	0.5926
Somers' D R|C	0.5077	0.0572	0.3956	0.6197
Pearson Correlation	0.5538	0.0590	0.4382	0.6693
Spearman Correlation	0.5479	0.0596	0.4311	0.6648
Lambda Asymmetric C|R	0.2588	0.0573	0.1465	0.3711
Lambda Asymmetric R\C	0.2727	0.0673	0.1409	0.4046
Lambda Symmetric	0.2663	0.0559	0.1567	0.3759
Uncertainty Coefficient C|R	0.1668	0.0389	0.0906	0.2431
Uncertainty Coefficient R|C	0.1609	0.0372	0.0880	0.2339
Uncertainty Coefficient Symmetric	0.1638	0.0380	0.0893	0.2383

5.4.2 Exact Tests for Ordinal Measures of Association

- Can also test whether a particular measure is equal to zero. In the case of correlation coefficients, we can produce exact *p*-values.

	Parental Interference			
Grades	Low	Medium	High	Total
1 – 2	3	1	0	4
3 – 4	3	2	1	6
5 – 6	1	3	2	6

```
proc freq order=data data=soccer;
   weight count;
   tables grades*degree / nocol nopct norow;
   exact scorr;
run;
```

Spearman Correlation Coefficient and Hypothesis Test for Spearman's Rank Test

```
Statistics for Table of grades by degree

    Spearman Correlation Coefficient
    ────────────────────────────────
    Correlation (r)            0.4878
    ASE                        0.1843
    95% Lower Conf Limit       0.1265
    95% Upper Conf Limit       0.8491

      Test of H0: Correlation = 0

    ASE under H0               0.1872
    Z                          2.6055
    One-sided Pr >  Z          0.0046
    Two-sided Pr > |Z|         0.0092

    Exact Test
    One-sided Pr >=  r         0.0354
    Two-sided Pr >= |r|        0.0637

            Sample Size = 16
```

5.5 Observer Agreement

- Important to evaluate observer agreement to understand possible contributions to measurement error, and as part of evaluation of testing new instruments and procedures

- Contingency table: rows are ratings of one observer and columns are ratings of another observer. Cells on diagonal represent cases where observers agree

- Diagnostic classification of multiple sclerosis patients into one of four diagnostic classes:

New Orleans Neurologist	Winnipeg Neurologist			
	1	2	3	4
1	38	5	0	1
2	33	11	3	0
3	10	14	5	6
4	3	7	3	10

- Suppose π_{ij} is probability of subject being classified in *i*th category by first observer and j^{th} category by second observer. Then:

$$\Pi_0 = \sum \pi_{ii}$$

is the probability that observers agree. If ratings are independent, probability of agreement is

$$\Pi_e = \sum \pi_{i+}\pi_{+i}$$

So $\Pi_o - \Pi_e$ is amount of agreement beyond that expected by chance. The *kappa coefficient* (Cohen 1960) is defined as

$$\kappa = \frac{\Pi_0 - \Pi_e}{1 - \Pi_e},$$

$\kappa = 1$ when there is perfect agreement, $\kappa = 0$ when agreement equals that expected by chance

- If interested in taking into account those disagreements just one category away, there is a weighted form of kappa statistic which allows assigning weights, or scores, to various categories

Weighted κ

$$\kappa_w = \frac{\sum\sum w_{ij}\pi_{ij} - \sum\sum w_{ij}\pi_{i+}\pi_{+j}}{1-\sum\sum_{ij} w_{ij}\pi_{i+}\pi_{+j}},$$

where w_{ij} represents weights with values between 0 and 1. One possible set of weights is

$$w_{ij} = 1 - \frac{\left|\text{score}(i) - \text{score}(j)\right|}{\text{score}(dim) - \text{score}(1)}$$

where score(i) is the score for ith row, score (j) is score for jth column, and *dim* is dimension of $s \times s$ table.

```
ods graphics on;
proc freq;
   weight count;
   tables no_rater*w_rater / agree norow nocol nopct;
run;
ods graphics off;
```

Winnipeg Data

Table of no_rater by w_rater

no_rater Frequency	w_rater 1	2	3	4	Total
1	38	5	0	1	44
2	33	11	3	0	47
3	10	14	5	6	35
4	3	7	3	10	23
Total	84	37	11	17	149

Kappa Statistics

STATISTICS FOR TABLE OF NO_RATER BY W_RATER

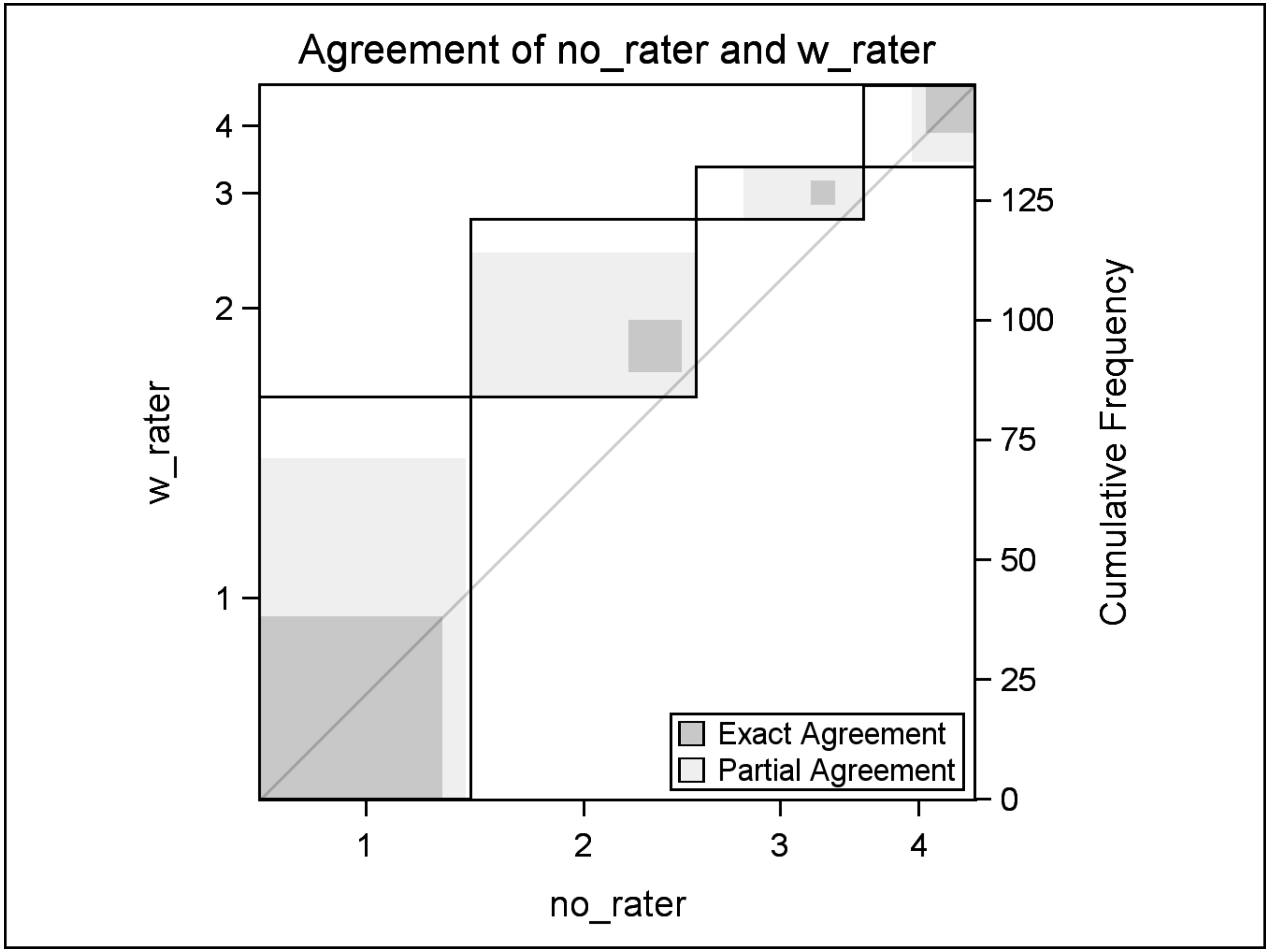
Test of Symmetry

Statistic = 46.749 DF = 6 Prob = 0.001

Kappa Coefficients

Statistic	Value	ASE	95% Confidence Bounds	
Simple Kappa	0.208	0.050	0.109	0.307
Weighted Kappa	0.380	0.052	0.278	0.481

Agreement Plot



5.5.2 Exact *p*-values for the Kappa Statistic

Example: Two raters evaluating the same 24 people:

Rater One	Rater Two			
	1	2	3	4
1	4	0	1	0
2	0	2	6	1
3	1	0	2	1
4	0	2	1	3

```
proc freq data=pilot;
  weight count;
  tables rater1*rater2 / norow nocol nopct;
  exact kappa;
run;
```

Exact Results for Kappa Test

```
Statistics for Table of rater1 by rater2

     Simple Kappa Coefficient
  --------------------------------
  Kappa (K)                 0.2989
  ASE                       0.1286
  95% Lower Conf Limit      0.0469
  95% Upper Conf Limit      0.5509

       Test of H0: Kappa = 0

  ASE under H0              0.1066
  Z                         2.8032
  One-sided Pr >  Z         0.0025
  Two-sided Pr > |Z|        0.0051

  Exact Test
  One-sided Pr >=  K        0.0088
  Two-sided Pr >= |K|       0.0088
```

5.6 Test for Ordered Differences

- You may be interested in testing against an ordered alternative: Are mean scores strictly increasing (or decreasing) across the levels of the row variable?

- Jonckheere-Terpstra test is designed to test the null hypothesis that the distribution of the ordered responses is the same across the various rows of a table
  - Detects whether there are differences in

$$d_1 \le d_2 \le \cdots \le d_s \quad \text{or} \quad d_s \ge d_{s-1} \ge \cdots \ge d_1$$

where d_i represents the *i*th group effect

Jonckheere-Terpstra Test

	Ordered Categories for Response				
Ordered Groups	$j = 1$	2	…	r	Total
$i = 1$	n_{11}	n_{12}	…	n_{1r}	n_{1+}
2	n_{21}	n_{22}	…	n_{2r}	n_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
s	n_{s1}	n_{s2}	…	n_{sr}	n_{s+}
Total	n_{+1}	n_{+2}	…	n_{+r}	n

$$\text{Let } U_{ii'} = \sum_{j=1}^{r} n_{ij} \left\{ \sum_{j'=j}^{r} n_{i'j'} - 0.5 n_{i'j} \right\}$$

$$= \text{Mann - Whitney Statistic for } i < i'$$

$J = \sum_{1 \le i < i' \le s} \sum U_{ii'}$ to test null hypothesis H_0 of equality of distributions for the s groups with power against a progressive shift of the distributions (i.e., as i increases, the higher categories become progressively more likely and the lower ones become less likely, or vice versa.)

For situations with sufficiently large sample sizes (e.g., all $n_{i+} \geq 10$),

$$Z = \left\{ J - E\left(J \middle| H_0 \right) \right\} \Big/ \left\{ Var\left(J \middle| H_0 \right) \right\}^{1/2}$$

approximately has the standard normal distribution with expected value 0 and variance 1. When sample sizes are not large, exact p-values can be used to evaluate J. Computations for J are possible through the JT option of PROC FREQ.

Computations for $s \times r$ Table

Operation	Severity of Symptoms			Total
	Moderate	Slight	None	
gre	16	38	53	107
v + h	12	40	58	110
v + a	13	23	68	104
v + d	7	28	61	96

The following statements request the analysis for the dumping syndrome data.

```
data operate ;
   input trt $ severity $ count @@ ;
datalines;
gre moderate 16 gre slight 38 gre none 53
v+h moderate 12 v+h slight 40 v+h none 58
v+a moderate 13 v+a slight 23 v+a none 68
v+d moderate  7 v+d slight 28 v+d none 61
;

proc freq data=operate order=data ;
   weight count ;
   tables trt*severity / scores=rank norow nocol
                           nopct jt;
run ;
```

```
Statistics for Table of trt by severity

     Jonckheere-Terpstra Test
    ------------------------------
    Statistic            35697.0000
    Z                        2.5712
    One-sided Pr >   Z       0.0051
    Two-sided Pr > |Z|       0.0101

         Sample Size = 417
```

- Exact version of the Jonckheere-Terpstra test is available by specifying the option JT in an EXACT statement