

# Chapter 10

## Fourfold Contingency Tables, II



Chapter 10 of *The Measurement of Association* continues the discussion of fourfold ( $2 \times 2$ ) contingency tables initiated in Chap. 9, but concentrates on symmetrical  $2 \times 2$  contingency tables, where each marginal frequency total is equal to  $N/2$ . In the same way that  $2 \times 2$  contingency tables are special cases of  $r \times c$  contingency tables, symmetrical  $2 \times 2$  contingency tables are special cases of fourfold tables. Symmetrical  $2 \times 2$  tables provide additional insight into the relationships among various measures of association.

Included in Chap. 10 are exact and Monte Carlo permutation statistical methods applied to Pearson's  $\phi^2$ , Tschuprov's  $T^2$ , Cramér's  $V^2$ , Pearson's  $r_{xy}$  product-moment correlation coefficient, Leik and Gove's  $d_N^c$  measure of nominal association, Goodman and Kruskal's  $t_a$  and  $t_b$  asymmetric measures, Kendall's  $\tau_b$  and Stuart's  $\tau_c$  measures, Somers'  $d_{yx}$  and  $d_{xy}$  asymmetric measures, simple percentage differences,  $D_x$  and  $D_y$ , Yule's  $Y$  measure of nominal association, and Cohen's unweighted and weighted  $\kappa$  measures of chance-corrected inter-rater agreement.

Also included in Chap. 10 are some extensions to multiple  $2 \times 2$  contingency tables and  $2 \times 2 \times 2$  contingency tables, including the Mantel–Haenszel test for combined  $2 \times 2$  contingency tables, Cohen's kappa measure of chance-corrected inter-rater agreement, McNemar's and Cochran's  $Q$  tests, Fisher's exact test for  $2 \times 2 \times 2$  and  $2 \times 2 \times 2 \times 2$  contingency tables, and tests for interactions in  $2 \times 2 \times 2$  and  $2 \times 2 \times 2 \times 2$  contingency tables.

### 10.1 Symmetrical Fourfold Tables

A symmetrical fourfold contingency table is a  $2 \times 2$  contingency table in which  $N$  is even and each marginal frequency total is equal to  $N/2$ . To illustrate the analysis of symmetrical fourfold contingency tables, consider the general layout of a  $2 \times 2$  table,

**Table 10.1** Notation for variables  $x$  and  $y$  with categories dummy-coded 0 and 1

$x$	$y$		Total
	0	1	
0	$a$	$b$	$a + b$
1	$c$	$d$	$c + d$
Total	$a + c$	$b + d$	$N$

**Table 10.2** Example  $2 \times 2$  contingency data for variables  $x$  and  $y$  with categories dummy-coded 0 and 1

$x$	$y$		Total
	0	1	
0	4	2	6
1	2	4	6
Total	6	6	12

such as given in Table 10.1, and an example  $2 \times 2$  frequency table, such as given in Table 10.2, where each marginal frequency total is equal to  $N/2 = 12/2 = 6$ .

### 10.1.1 Statistics $\phi^2$ , $T^2$ , and $V^2$

For the frequency data given in Table 10.2, Pearson’s chi-squared test statistic is given by

$$\chi^2 = N \left( \sum_{i=1}^r \sum_{j=1}^c \frac{O_{ij}^2}{R_i C_j} - 1 \right),$$

where  $O_{ij}$  is the observed cell frequency for  $i, j = 1, 2$ ,  $R_i$  denotes a row total for  $i = 1, 2$ , and  $C_j$  denotes a column total for  $j = 1, 2$ . Thus, for the frequency data given in Table 10.2,

$$\chi^2 = 12 \left[ \frac{4^2 + 2^2 + 2^2 + 4^2}{(6)(6)} - 1 \right] = 1.3333.$$

Then, Pearson’s  $\phi$  measure of association is given by

$$\phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{1.3333}{12}} = \pm 0.3333$$

and  $\phi^2 = (0.3333)^2 = 0.1111$ . Alternatively, using the notation given in Table 10.1,

$$\phi = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} = \frac{(4)(4) - (2)(2)}{\sqrt{(6)(6)(6)(6)}} = +0.3333.$$

Tschuprov’s measure of nominal association is

$$T^2 = \frac{\chi^2}{N\sqrt{(r-1)(c-1)}} = \frac{1.3333}{12\sqrt{(2-1)(2-1)}} = 0.1111$$

and  $T = \sqrt{T^2} = \sqrt{0.1111} = 0.3333$ . Also, Cramér’s measure of nominal association is

$$V^2 = \frac{\chi^2}{N[\min(r-1, c-1)]} = \frac{1.3333}{12[\min(2-1, 2-1)]} = 0.1111$$

and  $V = \sqrt{V^2} = \sqrt{0.1111} = 0.3333$ . Thus, Pearson’s  $\phi$ , Tschuprov’s  $T$ , and Cramér’s  $V$  are equivalent for a symmetrical  $2 \times 2$  contingency table.

### 10.1.2 Pearson’s $r_{xy}$ Correlation Coefficient

Next, consider Pearson’s product-moment correlation coefficient given by

$$r_{xy} = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{\sqrt{\left[ N \sum_{i=1}^N x_i^2 - \left( \sum_{i=1}^N x_i \right)^2 \right] \left[ N \sum_{i=1}^N y_i^2 - \left( \sum_{i=1}^N y_i \right)^2 \right]}}$$

The binary-coded (0, 1) data listed in Table 10.3 were obtained from the frequency data given in Table 10.2, where Objects 1 through 4, coded (0, 0), represent the four

**Table 10.3** Example dummy-coded (0, 1) values from the  $2 \times 2$  contingency table in Table 10.2

Object	Variable	
	$x$	$y$
1	0	0
2	0	0
3	0	0
4	0	0
5	0	1
6	0	1
7	1	0
8	1	0
9	1	1
10	1	1
11	1	1
12	1	1

objects in row 1 and column 1 of Table 10.2; Objects 5 and 6, coded (0, 1), represent the two objects in row 1 and column 2; Objects 7 and 8, coded (1, 0), represent the two objects in row 2 and column 1; and Objects 9 through 12, coded (1, 1), represent the four objects in row 2 and column 2 of Table 10.2.

For the binary-coded data listed in Table 10.3,

$$N = 12, \quad \sum_{i=1}^N x_i = \sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i = \sum_{i=1}^N y_i^2 = 6, \quad \sum_{i=1}^N x_i y_i = +4,$$

Pearson's product-moment correlation coefficient is

$$\begin{aligned} r_{xy} &= \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{\sqrt{\left[ N \sum_{i=1}^N x_i^2 - \left( \sum_{i=1}^N x_i \right)^2 \right] \left[ N \sum_{i=1}^N y_i^2 - \left( \sum_{i=1}^N y_i \right)^2 \right]}} \\ &= \frac{(12)(+4) - (6)(6)}{\sqrt{[(12)(6) - 6^2][(12)(6) - 6^2]}} = +0.3333, \end{aligned}$$

and  $r_{xy}^2 = (+0.3333)^2 = 0.1111$ .

### 10.1.3 Regression Coefficients

For the binary-coded data listed in Table 10.3, the slope (unstandardized regression coefficient) of the regression line with variable  $y$  the dependent variable is

$$b_{yx} = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N \sum_{i=1}^r x_i^2 - \left( \sum_{i=1}^r x_i \right)^2} = \frac{(12)(+4) - (6)(6)}{(12)(6) - 6^2} = +0.3333$$

and the standardized regression coefficient with variable  $x$  the dependent variable is

$$\hat{\beta}_{yx} = b_{yx} \left( \frac{s_x}{s_y} \right) = +0.3333 \left( \frac{0.5222}{0.5222} \right) = +0.3333.$$

Also the unstandardized regression coefficient with variable  $x$  the dependent variable is

$$b_{xy} = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N \sum_{i=1}^r y_i^2 - \left( \sum_{i=1}^r y_i \right)^2} = \frac{(12)(+4) - (6)(6)}{(12)(6) - 6^2} = +0.3333$$

and the standardized regression coefficient with variable  $y$  the dependent variable is

$$\hat{\beta}_{xy} = b_{xy} \left( \frac{s_x}{s_y} \right) = +0.3333 \left( \frac{0.5222}{0.5222} \right) = +0.3333 .$$

Thus it is demonstrated that  $\phi = T = V = r_{xy} = b_{yx} = b_{xy} = \hat{\beta}_{yx} = \hat{\beta}_{xy}$  for a symmetrical  $2 \times 2$  contingency table.

#### 10.1.4 Leik and Gove's $d_N^c$ Statistic

Leik and Gove's  $d_N^c$  test statistic for two nominal-level variables is described in detail in Chap. 4, Sect. 4.9. As noted by Leik and Gove, for symmetrical  $2 \times 2$  contingency tables,  $d_N^c$  is equivalent to the traditional chi-squared-based measures such as Pearson's  $\phi^2$ , Tschuprov's  $T^2$ , and Cramér's  $V^2$  [15, p. 291]. Test statistic  $d_N^c$  is based on three  $r \times c$  contingency tables: one  $r \times c$  contingency table containing the observed cell frequency values, a second  $r \times c$  contingency table containing the expected cell frequency values, and a third  $r \times c$  contingency table containing the maximized cell frequency values. Here, the observed values of concordant pairs,  $C$ ; discordant pairs,  $D$ ; pairs tied on variable  $x$ ,  $T_x$ ; pairs tied on variable  $y$ ,  $T_y$ ; and pairs tied on both variables  $x$  and  $y$ ,  $T_{xy}$ , are indicated without primes, the expected values of concordant pairs,  $C$ ; discordant pairs,  $D$ ; pairs tied on variable  $x$ ,  $T_x$ ; pairs tied on variable  $y$ ,  $T_y$ ; and pairs tied on both variables  $x$  and  $y$ ,  $T_{xy}$ , are indicated with a single prime ( $'$ ), and the maximized values of concordant pairs,  $C$ ; discordant pairs,  $D$ ; pairs tied on variable  $x$ ,  $T_x$ ; pairs tied on variable  $y$ ,  $T_y$ ; and pairs tied on both variables  $x$  and  $y$ ,  $T_{xy}$ , are indicated with double primes ( $''$ ).

Consider  $d_N^c$  for a symmetrical  $2 \times 2$  contingency table, where

$$d_N^c = \frac{T_y' - T_y}{T_y' - T_y''} = \frac{T_x' - T_x}{t_x' - T_x''} = \frac{T_{xy}' - T_{xy}}{T_{xy}' - T_{xy}''} = \frac{(C' + D') - (C + D)}{(C' + D') - (C'' + D'')} .$$

**Table 10.4** Observed values for a  $2 \times 2$  contingency table with categories dummy-coded 0 and 1

$x$	$y$		Total
	0	1	
0	4	2	6
1	2	4	6
Total	6	6	12

For the observed data given in Table 10.2 on p. 578, replicated in Table 10.4 for convenience, the observed values of  $C$ ,  $D$ ,  $T_x$ ,  $T_y$ , and  $T_{xy}$  are

$$C = ad = (4)(4) = 16 ,$$

$$D = bc = (2)(2) = 4 ,$$

$$T_x = ab + cd = (4)(2) + (2)(4) = 16 ,$$

$$T_y = ac + bd = (4)(2) + (2)(4) = 16 ,$$

$$\begin{aligned} T_{xy} &= \frac{1}{2}[(a)(a-1) + (b)(b-1) + (c)(c-1) + (d)(d-1)] \\ &= \frac{1}{2}[(4)(3) + (2)(1) + (2)(1) + (4)(3)] = 14 , \end{aligned}$$

and

$$\begin{aligned} C + D + T_x + T_y + T_{xy} &= 16 + 4 + 16 + 16 + 14 \\ &= \frac{N(N-1)}{2} = \frac{12(12-1)}{2} = 66 . \end{aligned}$$

Next, consider the expected values for the observed data in Table 10.4, given in Table 10.5, where

$$E_{11} = E_{12} = E_{21} = E_{22} = \frac{(6)(6)}{12} = 3 .$$

**Table 10.5** Expected values for the  $2 \times 2$  contingency table data in Table 10.4

$x$	$y$		Total
	0	1	
0	3	3	6
1	3	3	6
Total	6	6	12

For the expected cell values given in Table 10.5,

$$C' = ad = (3)(3) = 9 ,$$

$$D' = bc = (3)(3) = 9 ,$$

$$T'_x = ab + cd = (3)(3) + (3)(3) = 18 ,$$

$$T'_y = ac + bd = (3)(3) + (3)(3) = 18 ,$$

$$\begin{aligned} T''_{xy} &= \frac{1}{2}[(a)(a-1) + (b)(b-1) + (c)(c-1) + (d)(d-1)] \\ &= \frac{1}{2}[(3)(2) + (3)(2) + (3)(2) + (3)(2)] = 12 , \end{aligned}$$

and

$$\begin{aligned} C' + D' + T'_x + T'_y + T''_{xy} &= 9 + 9 + 18 + 18 + 12 \\ &= \frac{N(N-1)}{2} = \frac{12(12-1)}{2} = 66 . \end{aligned}$$

Finally, consider the maximized cell frequencies for the data in Table 10.4, given in Table 10.6. For the maximized values given in Table 10.6,

$$C'' = ad = (6)(6) = 36 ,$$

$$D'' = bc = (0)(0) = 0 ,$$

$$T''_x = ab + cd = (6)(0) + (0)(6) = 0 ,$$

$$T''_y = ac + bd = (6)(0) + (0)(6) = 0 ,$$

$$\begin{aligned} T''_{xy} &= \frac{1}{2}[(a)(a-1) + (b)(b-1) + (c)(c-1) + (d)(d-1)] \\ &= \frac{1}{2}[(6)(5) + (6)(5)] = 30 , \end{aligned}$$

**Table 10.6** Maximized values for the  $2 \times 2$  contingency table data in Table 10.4

x	y		Total
	0	1	
0	6	0	6
1	0	6	6
Total	6	6	12

and

$$\begin{aligned} C'' + D'' + T_x'' + T_y'' + T_{xy}'' &= 36 + 0 + 0 + 0 + 30 \\ &= \frac{N(N-1)}{2} = \frac{12(12-1)}{2} = 66. \end{aligned}$$

Then, Leik and Gove's  $d_N^c$  measure is

$$d_N^c = \frac{T_y' - T_y}{T_y' - T_y''} = \frac{18 - 16}{18 - 0} = 0.1111,$$

or

$$d_N^c = \frac{T_x' - T_x}{T_x' - T_x''} = \frac{18 - 16}{18 - 0} = 0.1111,$$

or

$$d_N^c = \frac{T_{xy}' - T_{xy}}{T_{xy}' - T_{xy}''} = \frac{12 - 14}{12 - 30} = 0.1111,$$

or

$$d_N^c = \frac{(C' + D') - (C + D)}{(C' + D') - (C'' + D'')} = \frac{(9 + 9) - (16 + 4)}{(9 + 9) - (36 - 0)} = 0.1111.$$

Thus it is demonstrated that  $\phi^2 = T^2 = V^2 = r_{xy}^2 = d_N^c$  for a symmetrical  $2 \times 2$  contingency table.

### 10.1.5 Goodman and Kruskal's $t_a$ and $t_b$ Statistics

Goodman and Kruskal's  $t_a$  and  $t_b$  measures of nominal association are discussed in Chap. 4, Sect. 4.3. Consider the notation for a  $2 \times 2$  contingency table given in Table 10.7. For the frequency data given in Table 10.4 on p. 582, Goodman and Kruskal's asymmetric measure of association with variable  $a$  the dependent

**Table 10.7** Notation for a  $2 \times 2$  contingency data for variables  $a$  and  $b$  with dummy (0, 1) coding

$b$	$a$		Total
	0	1	
0	$n_{11}$	$n_{12}$	$n_{1.}$
1	$n_{21}$	$n_{22}$	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$N$



variable is

$$t_a = \frac{N \sum_{j=1}^c \sum_{i=1}^r \frac{n_{ij}^2}{n_{.j}} - \sum_{i=1}^r n_{i.}^2}{N^2 - \sum_{i=1}^r n_{i.}^2}$$

$$= \frac{12 \left( \frac{4^2 + 2^2 + 2^2 + 4^2}{6} \right) - 6^2 - 6^2}{12^2 - 6^2 - 6^2} = 0.1111$$

and Goodman and Kruskal's asymmetric measure with variable  $b$  the dependent variable is

$$t_b = \frac{N \sum_{i=1}^r \sum_{j=1}^c \frac{n_{ij}^2}{n_{i.}} - \sum_{j=1}^c n_{.j}^2}{N^2 - \sum_{j=1}^c n_{.j}^2}$$

$$= \frac{12 \left( \frac{4^2 + 2^2 + 2^2 + 4^2}{6} \right) - 6^2 - 6^2}{12^2 - 6^2 - 6^2} = 0.1111 .$$

### 10.1.6 Kendall's $\tau_b$ Statistic

Kendall's  $\tau_b$  measure of ordinal association is detailed in Chap. 5, Sect. 5.4. For the frequency data given in Table 10.4 on p. 582, the number of concordant pairs is

$$C = ad = (4)(4) = 16 ,$$

the number of discordant pairs is

$$D = bc = (2)(2) = 4 ,$$

the number of pairs tied on variable  $x$  but not tied on variable  $y$  is

$$T_x = ab + cd = (4)(2) + (2)(4) = 16 ,$$

and the number of pairs tied on variable  $y$  but not tied on variable  $x$  is

$$T_y = ac + bd = (4)(2) + (2)(4) = 16 .$$

Then, Kendall's  $\tau_b$  measure is

$$\begin{aligned}\tau_b &= \frac{C - D}{\sqrt{(C + D + T_x)(C + D + T_y)}} \\ &= \frac{16 - 4}{\sqrt{(16 + 4 + 16)(16 + 4 + 16)}} = +0.3333 .\end{aligned}$$

Alternatively, following the notation given in Table 10.1,

$$\tau_b = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} = \frac{(4)(4) - (2)(2)}{\sqrt{(6)(6)(6)(6)}} = +0.3333 .$$

### 10.1.7 Stuart's $\tau_c$ Statistic

Stuart's  $\tau_c$  measure of ordinal association is discussed in Chap. 5, Sect. 5.5 and is given by

$$\tau_c = \frac{2mS}{N^2(m - 1)} ,$$

where  $m$  is the minimum number of rows or columns. For the frequency data given in Table 10.4 on p. 582,  $m = \min(r, c) = \min(2, 2) = 2$ , the number of concordant pairs is

$$C = ad = (4)(4) = 16 ,$$

the number of discordant pairs is

$$D = bc = (2)(2) = 4 ,$$

Kendall's  $S$  is

$$S = C - D = 16 - 4 = +12 ,$$

and Stuart's  $\tau_c$  measure is

$$\tau_c = \frac{2mS}{N^2(m - 1)} = \frac{2(2)(+12)}{12^2(2 - 1)} = +0.3333 .$$

### 10.1.8 Somers' $d_{yx}$ and $d_{xy}$ Statistics

Somers'  $d_{yx}$  and  $d_{xy}$  asymmetric measures of ordinal association are discussed in Chap. 5, Sect. 5.7. For the frequency data given in Table 10.4 on p. 582, Somers' asymmetric measure of association with variable  $y$  the dependent variable is

$$d_{yx} = \frac{C - D}{C + D + T_y} = \frac{16 - 4}{16 + 4 + 16} = +0.3333$$

and Somers' asymmetric measure with variable  $x$  the dependent variable is

$$d_{xy} = \frac{C - D}{C + D + T_x} = \frac{16 - 4}{16 + 4 + 16} = +0.3333 .$$

Alternatively,

$$d_{yx} = \frac{ad - bc}{(a + c)(b + d)} = \frac{(4)(4) - (2)(2)}{(6)(6)} = +0.3333$$

and

$$d_{xy} = \frac{ad - bc}{(a + b)(c + d)} = \frac{(4)(4) - (2)(2)}{(6)(6)} = +0.3333 .$$

### 10.1.9 Percentage Differences

Percentage differences are discussed in Chap. 9, Sect. 9.10. For the frequency data given in Table 10.4 on p. 582, the percentage difference for variable  $x$  is

$$D_x = \left| \frac{a}{a + b} - \frac{c}{c + d} \right| = \left| \frac{4}{6} - \frac{2}{6} \right| = |0.6667 - 0.3333| = 0.3333$$

and the percentage difference for variable  $y$  is

$$D_y = \left| \frac{a}{a + c} - \frac{b}{b + d} \right| = \left| \frac{4}{6} - \frac{2}{6} \right| = |0.6667 - 0.3333| = 0.3333 .$$

### 10.1.10 Yule's $Y$ Statistic

Yule's  $Y$  measure of nominal association is discussed in Chap. 9, Sect. 9.6. For the frequency data given in Table 10.4 on p. 582, Yule's coefficient of colligation is

$$Y = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} = \frac{\sqrt{(4)(4)} - \sqrt{(2)(2)}}{\sqrt{(4)(4)} + \sqrt{(2)(2)}} = +0.3333.$$

### 10.1.11 Cohen's $\kappa$ Statistic

Cohen's unweighted kappa measure of inter-rater agreement is discussed in Chap. 4, Sect. 4.5, and Cohen's linear and quadratic weighted kappa measures of inter-rater agreement are discussed in Chap. 6, Sect. 6.5. For the frequency data given in Table 10.4 on p. 582, let  $O_{ii}$  for  $i = 1, 2$  denote the observed cell frequencies on the principal diagonal and  $E_{ii}$  for  $i = 1, 2$  denote the expected cell frequencies on the principal diagonal. Then, Cohen's unweighted chance-corrected coefficient of inter-rater agreement is

$$\kappa = \frac{\sum_{i=1}^r O_{ii} - \sum_{i=1}^r E_{ii}}{N - \sum_{i=1}^r E_{ii}} = \frac{(4 + 4) - (3 + 3)}{12 - (3 + 3)} = +0.3333.$$

Cohen's weighted kappa measure of inter-rater agreement for  $b = 2$  judges and  $c$  categories is given by

$$\kappa_w = 1 - \frac{N \sum_{i=1}^c \sum_{j=1}^c w_{ij} n_{ij}}{\sum_{i=1}^c \sum_{j=1}^c w_{ij} R_i C_j}, \quad (10.1)$$

where  $n_{ij}$  denotes the observed cell frequencies,  $w_{ij}$  denotes the cell weights,  $R_i$  and  $C_j$  denote the observed row and column marginal frequency totals for  $i, j = 1, \dots, c$ , and

$$N = \sum_{i=1}^c \sum_{j=1}^c n_{ij}$$

denotes the table frequency total. For Cohen’s unweighted kappa measure of inter-rater agreement, the cell disagreement “weights” are given by

$$w_{ij} = \begin{cases} 0 & \text{if } i = j , \\ 1 & \text{otherwise ,} \end{cases}$$

and for Cohen’s weighted kappa measure of inter-rater agreement the cell disagreement weights are given by

$$w_{ij} = \begin{cases} 0 & \text{if } i = j , \\ |i - j| & \text{otherwise ,} \end{cases}$$

for linear weighting, and

$$w_{ij} = \begin{cases} 0 & \text{if } i = j , \\ (i - j)^2 & \text{otherwise ,} \end{cases}$$

for quadratic weighting. For the frequency data given in Table 10.4 on p. 582, Cohen’s linear-weighted kappa measure of inter-rater agreement is  $\kappa_w = +0.3333$  and Cohen’s quadratic-weighted kappa measure of inter-rater agreement is  $\kappa_w = +0.3333$ .

## 10.2 Inter-relationships Among the Measures

The inter-relationships among the various measures for a symmetrical 2×2 contingency table can be summarized as follows. The Pearson product-moment correlation coefficient,  $r_{xy}$ ; the unstandardized slopes of the two regression lines,  $b_{yx}$  and  $b_{xy}$ ; Yule’s coefficient of colligation,  $Y$ ; Pearson’s mean-square contingency coefficient,  $\phi$ ; Tschuprov’s  $T$  measure; Cramér’s  $V$  measure; Kendall’s  $\tau_b$  measure; Stuart’s  $\tau_c$  measure; Somers’  $d_{yx}$  and  $d_{xy}$  asymmetric measures; the two percentage differences,  $D_x$  and  $D_y$ ; and Cohen’s  $\kappa$  unweighted and weighted measures of chance-corrected inter-rater agreement are all equivalent measures, i.e.,

$$r_{xy} = b_{yx} = b_{xy} = Y = \phi = T = V = \tau_b = \tau_c = d_{yx} = d_{xy} \\ = D_x = D_y = \kappa = \kappa_w .$$

Also, Pearson’s squared product-moment correlation coefficient,  $r_{xy}^2$ ; Pearson’s mean-squared contingency coefficient,  $\phi^2$ ; Tschuprov’s  $T^2$  measure; Cramér’s  $V^2$  measure; Leik and Gove’s  $d_N^c$  measure; and Goodman and Kruskal’s  $t_b$  and  $t_a$

measures of association are all equivalent measures, i.e.,

$$r_{xy}^2 = \phi^2 = T^2 = V^2 = d_N^c = t_b = t_a .$$

### 10.2.1 Notational Inconsistencies

Measures of association for  $2 \times 2$  contingency tables in particular, and  $r \times c$  contingency tables in general, can be very confusing. First, some measures are denoted by uppercase Latin letters, e.g., Yule's  $Q$  and  $Y$ , Tschuprov's  $T^2$ , and Cramér's  $V^2$ ; some measures are denoted by lowercase Latin letters, e.g., Somers'  $d_{yx}$  and  $d_{xy}$ , Leik and Gove's  $d_N^c$ , and Goodman and Kruskal's  $t_a$  and  $t_b$ ; and some measures are denoted by lowercase Greek letters, e.g., Pearson's  $\phi$ , Kendall's  $\tau_b$ , and Cohen's  $\kappa$ . While it would be preferable to reserve Greek letters for population parameters that are being estimated by sample statistics and Latin letters for sample statistics, once symbols are in common use it is difficult to standardize usage.<sup>1</sup> Second, certain measures of association appear as squared, whereas others do not. In particular, for the  $2 \times 2$  case, the non-squared symbols  $t_b$  and  $t_a$  for Goodman and Kruskal's asymmetric measures of nominal association are equivalent to Pearson's symmetric measures  $\phi^2$  and  $r_{xy}^2$ . Third, some measures norm between 0 and 1 for  $2 \times 2$  contingency tables, e.g., Goodman and Kruskal's  $t_a$  and  $t_b$ ; others norm between  $-1$  and  $+1$ , e.g., Kendall's  $\tau_b$  and Cramér's  $V$ ; and still others norm between 0 and  $\infty$ , e.g., the odds ratio. Finally, some measures identify the two variables as  $x$  and  $y$ , e.g., Somers'  $d_{yx}$  and  $d_{xy}$ , while others identify the two variables as  $a$  and  $b$ , e.g., Kendall's  $\tau_a$  and  $\tau_b$ .

## 10.3 Extended Fourfold Contingency Tables

In some cases, measures of association have been introduced to analyze fourfold tables that have either been extended to analyze a series of  $2 \times 2$  contingency tables or redesigned to consider multidimensional contingency tables with two categories in each dimension. In this section a small number of such measures are considered, including the Mantel–Haenszel test, McNemar's  $Q$  test, Cochran's  $Q$  test, Cohen's chance-corrected measure of inter-rater agreement, Fisher's exact probability test for  $2 \times 2 \times 2$  contingency tables, and tests for interactions in  $2 \times 2 \times 2$  and  $2 \times 2 \times 2 \times 2$  contingency tables

<sup>1</sup>It was not too many years ago that while  $\mu_x$  and  $\sigma_x^2$  denoted the population mean and variance, respectively,  $\hat{\mu}_x$  and  $\hat{\sigma}_x^2$  denoted the unbiased sample-estimated population mean and variance. The *American Psychological Association* presently recommends using  $M$  for the sample mean instead of the conventional  $\bar{x}$ .

**Table 10.8** General layout of a 3-way contingency table with  $r = 2$  rows,  $c = 2$  columns, and  $S$  strata

Stratum		Column 1	Column 2	Total	Stratum total	Table total
1	Row 1	$n_{111}$	$n_{121}$	$n_{1.1}$		
	Row 2	$n_{211}$	$n_{221}$	$n_{2.1}$		
	Total	$n_{.11}$	$n_{.21}$		$n_{..1}$	
2	Row 1	$n_{112}$	$n_{122}$	$n_{1.2}$		
	Row 2	$n_{212}$	$n_{222}$	$n_{2.2}$		
	Total	$n_{.12}$	$n_{.22}$		$n_{..2}$	
⋮	⋮	⋮	⋮	⋮	⋮	
S	Row 1	$n_{11S}$	$n_{12S}$	$n_{1.S}$		
	Row 2	$n_{21S}$	$n_{22S}$	$n_{2.S}$		
	Total	$n_{.1S}$	$n_{.2S}$		$n_{..S}$	
Row 1 Total		$n_{1.}$	$n_{2.}$	$n_{1..}$		
Row 2 Total		$n_{2.}$	$n_{2.}$	$n_{2..}$		
Column Total		$n_{.1.}$	$n_{.2.}$			
Table Total						$n_{...}$

### 10.4 The Mantel–Haenszel Test

The Mantel–Haenszel test, developed by Nathan Mantel and William Haenszel in 1959, is a test of significance for  $S$  combined  $2 \times 2$  contingency tables.<sup>2</sup> Suppose that a treatment is compared with a control in each of  $S$  strata, where the outcome is binary: success or failure. Of interest is whether or not the treatment increases the probability of success.

Let  $n_{ijk}$  denote the cell frequency for  $i, j = 1, 2$  discrete categories and  $k = 1, \dots, S$  discrete strata for a  $2 \times 2 \times S$  contingency table. Table 10.8 illustrates a three-way contingency table with  $r = 2$  rows,  $c = 2$  columns, and  $S$  strata. Denote by a dot ( $\cdot$ ) the partial sum of all rows, all columns, or all strata, depending on the position of the ( $\cdot$ ) in the subscript list. If the ( $\cdot$ ) is in the first subscript position, the sum is over all rows; if the ( $\cdot$ ) is in the second subscript position, the sum is over all columns; and if the ( $\cdot$ ) is in the third subscript position, the sum is over all strata. Thus,  $n_{i.}$  denotes the marginal frequency total of the  $i$ th row,  $i = 1, 2$ , summed over all columns and strata;  $n_{.j.}$  denotes the marginal frequency total of the  $j$ th column,  $j = 1, 2$ , summed over all rows and strata;  $n_{..k}$  denotes the marginal frequency total of the  $k$ th stratum,  $k = 1, \dots, S$ , summed over all rows and columns; and  $n_{...}$  denotes the table frequency total. The Mantel–Haenszel statistical model, under the null hypothesis, states that the  $S$   $2 \times 2$  contingency tables are independent and the marginal frequency totals for each of the  $2 \times 2$  contingency tables are fixed [17].

<sup>2</sup>The test is often called the Cochran–Mantel–Haenszel test as William Cochran presented essentially the same test in an earlier paper [5].

Then, the probability for the  $n_{11k}$  frequency of each of the  $2 \times 2$  contingency tables under the null hypothesis is the hypergeometric point probability value given by

$$p(n_{11k} | n_{1.k}, n_{.1k}, n_{.k}) = \binom{n_{.k}}{n_{11k}} \binom{n_{.k}}{n_{12k}} \binom{n_{.k}}{n_{1.k}}^{-1} \\ = \frac{n_{1.k}! n_{2.k}! n_{.k}! n_{.k}!}{n_{.k}! n_{11k}! n_{12k}! n_{21k}! n_{22k}!}, \quad (10.2)$$

where  $n_{..k} = n_{11k} + n_{12k} + n_{21k} + n_{22k}$ ,  $n_{2.k} = n_{.k} - n_{1.k}$ ,  $n_{.2k} = n_{.k} - n_{.1k}$ , and  $k = 1, \dots, S$ .

The test statistic of interest is given by

$$T = \sum_{k=1}^S n_{11k},$$

where the summation is over only one cell since for any  $2 \times 2$  contingency table with fixed marginal frequency totals the entry in any one cell determines the entries in the remaining three cells.

Under the null hypothesis ( $H_0$ ) of the model in Eq. (10.2), the mean and variance of test statistic  $T$  are given by

$$E[T | H_0] = \sum_{k=1}^S \frac{n_{1.k} n_{.1k}}{n_{.k}}$$

and

$$\text{VAR}(T | H_0) = \sum_{k=1}^S \frac{n_{1.k} n_{2.k} n_{.1k} n_{.2k}}{(n_{.k})^2 (n_{.k} - 1)},$$

respectively. The Mantel–Haenszel test statistic, corrected for continuity, is given by

$$M = \frac{\left( |T - E[T | H_0]| - \frac{1}{2} \right)^2}{\text{VAR}(T | H_0)}.$$

The Mantel–Haenszel test statistic,  $M$ , is approximately distributed as Pearson's chi-squared with one degree of freedom as  $N \rightarrow \infty$ .<sup>3</sup>

---

<sup>3</sup>The symbol  $M$  for the Mantel–Haenszel test should not be confused with the symbol  $M$  for the number of possible, equally-likely arrangements of the observed data under the Fisher–Pitman permutation model.



### 10.4.1 Example Analysis

Consider the example data set given in Table 10.9 with  $r = 2$  rows,  $c = 2$  columns,  $S = 3$  strata, and  $n_{...} = 74$  total observations. For the data listed in Table 10.9, the observed value of test statistic  $T$  is

$$T_o = \sum_{k=1}^S n_{11k} = 2 + 2 + 4 = 8.00 ,$$

the expected value of  $T$  under the null hypothesis is

$$E[T|H_0] = \sum_{k=1}^S \frac{n_{1.k} n_{.1k}}{n_{..k}} = \frac{(3)(7)}{32} + \frac{(4)(4)}{24} + \frac{(5)(5)}{18} = 2.7118 ,$$

the variance of  $T$  is

$$\begin{aligned} \text{VAR}(T|H_0) &= \sum_{k=1}^S \frac{n_{1.k} n_{2.k} n_{.1k} n_{.2k}}{(n_{..k})^2 (n_{..k} - 1)} \\ &= \frac{(3)(29)(7)(25)}{(32)^2 (32 - 1)} + \frac{(4)(20)(4)(20)}{(24)^2 (24 - 1)} + \frac{(5)(13)(5)(13)}{(18)^2 (18 - 1)} = 1.7272 , \end{aligned}$$

**Table 10.9** General layout of a 3-way contingency table with  $r = 2$  rows,  $c = 2$  columns, and  $S = 3$  strata

Stratum		Column 1	Column 2	Total	Stratum total	Table total
1	Row 1	2	1	3		
	Row 2	5	24	29		
	Total	7	25		32	
2	Row 1	2	2	4		
	Row 2	2	18	20		
	Total	4	20		24	
3	Row 1	4	1	5		
	Row 2	1	12	13		
	Total	5	13		18	
Row 1 Total		8	4	12		
Row 2 Total		8	54	62		
Column Total		16	58			
Table Total						74

and the observed Mantel–Haenszel test statistic is

$$M_o = \frac{\left(|T_o - E[T|H_0]| - \frac{1}{2}\right)^2}{\text{VAR}(T|H_0)} = \frac{\left(|8.00 - 2.7118| - \frac{1}{2}\right)^2}{1.7272} = 13.2742. \quad (10.3)$$

Mantel and Haenszel's  $M$  test statistic is approximately distributed as Pearson's chi-squared with one degree of freedom. For the observed value of  $M_o = 13.2742$  the approximate chi-squared probability value is  $P = 0.2691 \times 10^{-3}$ .

In Eq. (10.3),  $E[T|H_0]$ ,  $\text{VAR}(T|H_0)$ , and the correction factor, are all invariant under permutation, leaving only variable  $T$ . Thus, for the data listed in Table 10.9 the approximate Monte Carlo resampling probability value based on  $L = 1,000,000$  random arrangements of the observed data under the null hypothesis is

$$\begin{aligned} P(M \geq M_o|H_0) &= \frac{\text{number of } M \text{ values} \geq M_o}{L} \\ &= P(T \geq T_o|H_0) = \frac{\text{number of } T \text{ values} \geq T_o}{L} = \frac{2,555}{1,000,000} \\ &= 0.2555 \times 10^{-2}. \end{aligned}$$

### 10.4.2 Measures of Effect Size

Two types of measures of effect size have been proposed to represent the strength of a treatment effect [32]. One type, designated the  $d$ -family, is based on one or more measures of the differences between groups or levels of an independent variable. Representative of the  $d$ -family is Cohen's  $d$ , which calculates the effect size by the number of standard deviations separating the means of the groups or levels [8]. The second type of measure of effect size, designated the  $r$ -family, represents some sort of correlation between the independent variables. Measures in the  $r$ -family are typically measures of correlation or association, the most prominent being Pearson's squared product-moment correlation coefficient. Since the Mantel–Haenszel test is based on a  $2 \times 2 \times S$  contingency table, the  $d$ -family is not applicable.

The  $r$ -family measures of effect size contains two types of measures: putative maximum-corrected and chance-corrected. Maximum-corrected measures of effect size standardize the observed test statistic value by the maximum possible value of the test statistic. Maximum-corrected measures of effect size are bounded between 0 and 1 and are interpretable as the proportion of the maximum possible value of the test statistic. On the other hand, chance-corrected measures of effect size standardize the observed test statistic value by the expected value of the test statistic. Chance-corrected measures of effect size can attain a maximum value of +1, but may be less than 0 when the test statistic value is less than expected by chance and are

interpretable as the proportion above, or below, what is expected by chance. In 2010 Berry, Johnston, and Mielke developed two measures of effect size for the Mantel–Haenszel test statistic: a maximum-corrected and a chance-corrected measure of effect size [2].

### Maximum-Corrected Measure of Effect Size

Let  $M_o$  and  $T_o$  denote the observed values of  $M$  and  $T$ , respectively. Then, the maximum-corrected measure of effect size is given by  $M_o$  divided by the maximum possible value of  $M$ . The maximum value of  $T$  for an observed  $2 \times 2 \times S$  contingency table is given by

$$T_{\max} = \sum_{k=1}^S \min(n_{1.k}, n_{.1k}) ,$$

where  $\min(n_{1.k}, n_{.1k})$  is the maximum value of  $n_{11k}$  in the  $k$ th of  $S$   $2 \times 2$  contingency tables. Thus, the maximum value of  $M$  is given by

$$M_{\max} = \frac{\left( |T_{\max} - E[T|H_0]| - \frac{1}{2} \right)^2}{\text{VAR}(T|H_0)}$$

and the maximum-corrected measure of effect size for  $M$  is given by the observed value of  $M$  divided by the maximum value of  $M$ , i.e.,

$$\text{ES}_M = \frac{M_o}{M_{\max}} .$$

For the frequency data given in Table 10.2 on p. 578, the maximum value of  $T$  is

$$T_{\max} = \sum_{k=1}^S \min(n_{1.k}, n_{.1k}) = 3 + 4 + 5 = 12.00 ,$$

the maximum value of  $M$  is

$$M_{\max} = \frac{\left( |T_{\max} - E[T|H_0]| - \frac{1}{2} \right)^2}{\text{VAR}(T|H_0)} = \frac{\left( |12.00 - 2.7118| - \frac{1}{2} \right)^2}{1.7272} = 44.7162 ,$$

and the maximum-corrected measure of effect size is

$$\text{ES}_M = \frac{M_o}{M_{\max}} = \frac{13.2742}{44.7162} = 0.2969 ,$$

indicating that  $M_o = 13.2742$  accounts for approximately 30% of the maximum value of  $M$ , given the observed row, column, and stratum marginal frequency distributions, {12, 62}, {16, 58}, and {32, 24, 18}, respectively.

### Chance-Corrected Measure of Effect Size

A chance-corrected measure of effect size for the Mantel–Haenszel test may be given by statistic  $M$ , standardized by the expected value of  $M$ . Thus, the chance-corrected measure is given by

$$ES_C = \frac{M - E[M|H_0]}{M_{\max} - E[M|H_0]} = 1 - \frac{M_{\max} - M}{M_{\max} - 1},$$

where  $E[M] = 1$  since the mean of a chi-squared distribution is equal to the degrees of freedom and  $M$  is approximately distributed as chi-squared with one degree of freedom. For the frequency data given in Table 10.2, the chance-corrected measure of effect size is

$$ES_C = 1 - \frac{44.7162 - 13.2742}{44.7162 - 1} = +0.2808,$$

indicating that  $M_o = 13.2742$  accounts for approximately 28% above what is expected by chance. In general, chance-corrected measures of effect size, such as  $ES_C$ , tend to slightly smaller values than maximum-corrected measures, such as  $ES_M$ , for the same set of data [2, pp. 398–399].

## 10.5 Cohen's Kappa Measure

In 1960 Jacob Cohen introduced statistic kappa, an unweighted, chance-corrected measure of inter-rater agreement between two judges for a set of  $c$  disjoint, unordered categories [6]. In 1968 Cohen expanded kappa to include weighting for measuring the agreement between two judges for a set of  $c$  disjoint, ordered categories [7]. Unweighted kappa is discussed more completely in Chap. 4, Sect. 4.5, and weighted kappa is discussed in detail in Chap. 6, Sect. 6.5. Whereas unweighted kappa for categorical data did not distinguish among magnitudes of disagreement, weighted kappa for ordinal-level data incorporated the magnitude of each disagreement and provided partial credit for disagreements when agreement was not complete [16]. Weighted kappa is easily extended to interval-level data [3]. The usual approach is to assign weights to each disagreement pair with larger weights indicating greater disagreement. In the cases of both, unweighted and weighted kappa, kappa is equal to +1 when perfect agreement between the two judges occurs, 0 when agreement is equal to that expected under independence, and

negative when agreement is less than expected by chance. Unweighted kappa and weighted kappa are conventionally designated as  $\kappa$  and  $\kappa_w$ , respectively. Two forms of weighting are popular for weighted kappa: linear weighting, in which category disagreement weights progress outward linearly from the agreement diagonal, and quadratic weighting, in which category disagreement weights progress outward geometrically from the agreement diagonal. In keeping with the theme of this chapter—fourfold contingency tables— $\kappa$  and  $\kappa_w$  are extended to multiple judges with  $c = 2$  categories.

Consider first  $b = 2$  judges and  $c = 2$  categories. A generalized calculation formula that applies to both unweighted and weighted kappa for  $b = 2$  judges and  $c$  categories is given by

$$\kappa = 1 - \frac{N \sum_{i=1}^c \sum_{j=1}^c w_{ij} n_{ij}}{\sum_{i=1}^c \sum_{j=1}^c w_{ij} R_i C_j}, \tag{10.4}$$

where  $n_{ij}$  denotes the observed cell frequencies,  $w_{ij}$  denotes the cell weights,  $R_i$  and  $C_j$  denote the observed row and column marginal frequency totals for  $i, j = 1, \dots, c$ , and

$$N = \sum_{i=1}^c \sum_{j=1}^c n_{ij}$$

denotes the table frequency total.

Given a  $c \times c$  agreement table with  $N$  objects cross-classified by the ratings of two independent judges into  $c$  disjoint categories, an exact permutation test generates all  $M$  possible, equally-likely arrangements of the  $N$  objects in the  $c^2$  cells, while preserving the total number of objects in each category, i.e., the marginal frequency distributions. For each arrangement of cell frequencies with fixed marginal frequency distributions, the kappa statistic,  $\kappa$ , and the exact point probability,  $p(n_{ij} | n_{i.}, n_{.j}, N)$ , are calculated, where

$$p(n_{ij} | R_i, C_j, N) = \frac{\left( \prod_{i=1}^c R_i! \right) \left( \prod_{j=1}^c C_j! \right)}{N! \prod_{i=1}^c \prod_{j=1}^c n_{ij}!}$$

is the conventional hypergeometric probability of a  $c \times c$  contingency table.

Let  $\kappa_0$  denote the value of the observed weighted kappa statistic and  $M$  denote the total number of distinct cell frequency arrangements of the  $N$  objects in the  $c \times c$  agreement table, given fixed marginal frequency totals. Then the exact probability value of  $\kappa_0$  under the null hypothesis is given by

$$P(\kappa_0|H_0) = \sum_{k=1}^M \Psi(\kappa_k) p(n_{ij}|R_i, C_j, N),$$

where

$$\Psi(\kappa_k) = \begin{cases} 1 & \text{if } \kappa_k \geq \kappa_0, \\ 0 & \text{otherwise.} \end{cases}$$

When  $M$  is very large, exact permutation analyses quickly become impractical and Monte Carlo resampling procedures become necessary. Let  $L$  denote a random sample of all  $M$  possible values of  $\kappa$ . Then, under the null hypothesis the resampling approximate probability value for the observed value of  $\kappa$ ,  $\kappa_0$  is given by

$$P(\kappa_0) = \frac{1}{L} \sum_{l=1}^L \Psi_l(\kappa),$$

where

$$\Psi_l(\kappa) = \begin{cases} 1 & \text{if } \kappa \geq \kappa_0, \\ 0 & \text{otherwise.} \end{cases}$$

To calculate Cohen's unweighted kappa with Eq.(10.4) on p. 597, the cell disagreement "weights" are given by

$$w_{ij} = \begin{cases} 0 & \text{if } i = j, \\ 1 & \text{otherwise.} \end{cases}$$

To calculate Cohen's weighted kappa with linear weighting, the cell disagreement weights are given by

$$w_{ij} = \begin{cases} 0 & \text{if } i = j, \\ |i - j| & \text{otherwise.} \end{cases}$$

To calculate Cohen’s weighted kappa with quadratic weighting, the cell disagreement weights are given by

$$w_{ij} = \begin{cases} 0 & \text{if } i = j , \\ (i - j)^2 & \text{otherwise .} \end{cases}$$

Thus, as demonstrated, for  $b = 2$  judges and  $c = 2$  categories, the cell disagreement weights are the same for unweighted kappa ( $\kappa$ ) and weighted kappa ( $\kappa_w$ ) with either linear or quadratic weighting.

### 10.5.1 Example 1

To illustrate the application of Cohen’s unweighted kappa with  $b = 2$  judges and  $c = 2$  categories, consider the frequency data given in Table 10.10, where  $b = 2$  independent judges have each assigned  $N = 123$  observations to  $c = 2$  disjoint, unordered categories labeled Pro and Con. Assign the number 1 to the categories labeled “Pro” and the number 2 to the categories labeled “Con.” Then following Eq. (10.4) on p. 597,

$$\begin{aligned} \kappa &= 1 - \frac{N \sum_{i=1}^c \sum_{j=1}^c w_{ij} n_{ij}}{\sum_{i=1}^c \sum_{j=1}^c w_{ij} R_i C_j} \\ &= 1 - \frac{123[(0)(42) + (1)(23) + (1)(18) + (0)(40)]}{(0)(65)(60) + (1)(65)(63) + (1)(58)(60) + (0)(58)(63)} \\ &= +0.3343 , \end{aligned}$$

indicating approximately 33% agreement between the two judges above that expected by chance.

For the frequency data given in Table 10.10, there are only

$$\begin{aligned} M &= \min(a + b, a + c) - \max(0, a - d) + 1 \\ &= \min(65, 60) - \max(0, 42 - 40) + 1 = 60 - 2 + 1 = 59 \end{aligned}$$

**Table 10.10** Example 2×2 contingency table for  $b = 2$  independent judges and  $c = 2$  disjoint categories

Judge 1	Judge 2		Total
	Pro	Con	
Pro	42	23	65
Con	18	40	58
Total	60	63	123

possible, equally-likely arrangements in the reference set of all permutations of the cell frequencies in Table 10.10 given the observed row and column marginal frequency distributions, {65, 58} and {60, 63}, respectively, making an exact permutation analysis possible. If the  $M = 59$  possible arrangements of the frequency data given in Table 10.10 occur with equal chance, the exact probability value of  $\kappa$  under the null hypothesis is the sum of the hypergeometric point probability values associated with  $\kappa = +0.3343$  or greater.

Table 10.11 lists the  $n_{11}$  cell frequency values, unweighted kappa values, and associated hypergeometric probability values for the frequency data given in Table 10.10, where the  $n_{11}$  cell values associated with  $\kappa$  values equal to or greater

**Table 10.11** Listing of the  $M = 59$  possible arrangements of cell frequencies, unweighted kappa values, and associated hypergeometric probability values for the data given in Table 10.10

$n_{11}$	Kappa	Probability	$n_{11}$	Kappa	Probability
2	-0.9648	$0.2825 \times 10^{-32}$	32	$+0.9505 \times 10^{-2}$	0.1425
3	-0.9328	$0.3441 \times 10^{-29}$	33	$+0.4198 \times 10^{-1}$	0.1287
4	-0.8998	$0.1520 \times 10^{-26}$	34	$+0.7446 \times 10^{-1}$	0.1022
5	-0.8673	$0.3462 \times 10^{-24}$	35	+0.1069	$0.7033 \times 10^{-1}$
6	-0.8349	$0.4760 \times 10^{-22}$	36	+0.1394	$0.4370 \times 10^{-1}$
7	-0.8024	$0.4333 \times 10^{-20}$	37	+0.1719	$0.2349 \times 10^{-1}$
8	-0.7699	$0.2775 \times 10^{-18}$	38	+0.2044	$0.1106 \times 10^{-1}$
9	-0.7374	$0.1305 \times 10^{-16}$	39	+0.2368	$0.4552 \times 10^{-2}$
10	-0.7049	$0.4660 \times 10^{-15}$	40	+0.2693	$0.1635 \times 10^{-2}$
11	-0.6725	$0.1295 \times 10^{-13}$	41	+0.3018	$0.5113 \times 10^{-3}$
12	-0.6400	$0.2855 \times 10^{-12}$	42*	+0.3343	$0.1388 \times 10^{-3}$
13	-0.6075	$0.5078 \times 10^{-11}$	43*	+0.3667	$0.3259 \times 10^{-4}$
14	-0.5750	$0.7388 \times 10^{-10}$	44*	+0.3992	$0.6595 \times 10^{-5}$
15	-0.5426	$0.8888 \times 10^{-9}$	45*	+0.4317	$0.1145 \times 10^{-5}$
16	-0.5101	$0.8928 \times 10^{-7}$	46*	+0.4642	$0.1697 \times 10^{-6}$
17	-0.4776	$0.7548 \times 10^{-7}$	47*	+0.4966	$0.2135 \times 10^{-7}$
18	-0.4451	$0.5410 \times 10^{-6}$	48*	+0.5291	$0.2262 \times 10^{-8}$
19	-0.4127	$0.3306 \times 10^{-5}$	49*	+0.5616	$0.2004 \times 10^{-9}$
20	-0.3802	$0.1732 \times 10^{-4}$	50*	+0.5941	$0.1470 \times 10^{-10}$
21	-0.3477	$0.7814 \times 10^{-4}$	51*	+0.6265	$0.8822 \times 10^{-12}$
22	-0.3152	$0.3047 \times 10^{-3}$	52*	+0.6590	$0.4275 \times 10^{-13}$
23	-0.2828	$0.1031 \times 10^{-2}$	53*	+0.6915	$0.1645 \times 10^{-14}$
24	-0.2503	$0.3034 \times 10^{-2}$	54*	+0.7240	$0.4921 \times 10^{-16}$
25	-0.2178	$0.7788 \times 10^{-2}$	55*	+0.7564	$0.1114 \times 10^{-17}$
26	-0.1853	$0.1747 \times 10^{-1}$	56*	+0.7889	$0.1842 \times 10^{-19}$
27	-0.1529	$0.3433 \times 10^{-1}$	57*	+0.8214	$0.2115 \times 10^{-21}$
28	-0.1204	$0.5913 \times 10^{-1}$	58*	+0.8539	$0.1563 \times 10^{-23}$
29	$-0.8792 \times 10^{-1}$	$0.8941 \times 10^{-1}$	59*	+0.8863	$0.6507 \times 10^{-26}$
30	$-0.5545 \times 10^{-1}$	0.1188	60*	+0.9188	$0.1122 \times 10^{-28}$
31	$-0.2297 \times 10^{-1}$	0.1387			



than the observed value of  $\kappa = +0.3343$  are indicated with asterisks. Because there is only one degree of freedom, it is sufficient to list the cell frequency values for only one cell,  $n_{11}$ . For the frequency data given in Table 10.10, the exact upper-tail hypergeometric probability value of the observed  $\kappa$  value is

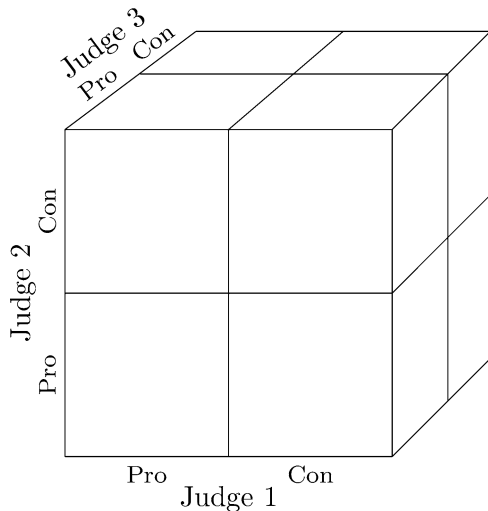
$$P = 0.1388 \times 10^{-3} + 0.3259 \times 10^{-4} + \dots + 0.6507 \times 10^{-26} + 0.1122 \times 10^{-28} = 0.1793 \times 10^{-3}.$$

### 10.5.2 Example 2

Although weighted and unweighted kappa were originally formulated to compare only two judges, both  $\kappa$  and  $\kappa_w$  can be generalized to accommodate multiple judges [25]. However, with multiple judges an exact permutation analysis becomes impractical except for very small sample sizes; therefore, a Monte Carlo resampling permutation analysis is preferred when analyzing agreement data from multiple judges. The analysis for  $b$  multiple judges may be conceptualized as a  $b$ -way contingency table with  $c = 2$  categories on each axis. Figure 10.1 illustrates a  $2 \times 2 \times 2$  contingency table with  $b = 3$  judges and  $c = 2$  disjoint, unordered categories labeled Pro and Con.

To illustrate the application of Cohen's kappa with multiple judges and  $c = 2$  disjoint categories, consider the frequency data given in Table 10.12, where  $b = 3$  judges have independently assigned  $N = 254$  observations to  $c = 2$  categories labeled Pro and Con. A generalized calculation formula that applies to

**Fig. 10.1** Graphic depiction of a  $2 \times 2 \times 2$  contingency table with  $b = 3$  independent judges and  $c = 2$  disjoint categories



**Table 10.12** Example  
 $2 \times 2 \times 2$  contingency table for  
 $b = 3$  independent judges and  
 $c = 2$  disjoint categories

Judge 1	Judge 2	Judge 3	
		Pro	Con
Pro	Pro	42	23
	Con	18	40
Con	Pro	41	29
	Con	33	28

both unweighted and weighted kappa for  $b = 3$  judges and  $c$  categories is given by

$$\kappa = 1 - \frac{N^2 \sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^c w_{ijk} n_{ijk}}{\sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^c w_{ijk} R_i C_j S_k}, \tag{10.5}$$

where  $n_{ijk}$  denotes the observed cell frequencies,  $w_{ijk}$  denotes the cell weights,  $R_i$ ,  $C_j$ , and  $S_k$  denote the observed row, column, and slice marginal frequency totals for  $i, j, k = 1, \dots, c$ , and

$$N = \sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^c n_{ijk}$$

denotes the table frequency total.

Given a  $c \times c \times c$  agreement table with  $N$  objects cross-classified by  $b = 3$  independent judges, an exact permutation test involves generating all possible, equally-likely arrangements of the  $N$  objects to the  $c^3$  cells, while preserving the observed row, column, and slice marginal frequency distributions,  $\{123, 131\}$ ,  $\{135, 119\}$ , and  $\{134, 120\}$ , respectively. For each arrangement of cell frequencies, the kappa statistic,  $\kappa$ , and the exact hypergeometric point probability value under the null hypothesis,  $p(n_{ijk} | R_i, C_j, S_k, N)$ , are calculated, where

$$p(n_{ijk} | R_i, C_j, S_k, N) = \frac{\left( \prod_{i=1}^c R_i! \right) \left( \prod_{j=1}^c C_j! \right) \left( \prod_{k=1}^c S_k! \right)}{(N!)^2 \prod_{i=1}^c \prod_{j=1}^c \prod_{k=1}^c n_{ijk}!}$$

[20].

If  $\kappa_o$  denotes the value of the observed kappa test statistic, the exact probability value of  $\kappa_o$  under the null hypothesis is given by

$$P(\kappa_o|H_0) = \sum_{l=1}^M \Psi_l(n_{ijk}|R_i, C_j, S_k, N) ,$$

where

$$\Psi_l(n_{ijk}|R_i, C_j, S_k, N) = \begin{cases} p(n_{ijk}|R_i, C_j, S_k, N) & \text{if } \kappa \geq \kappa_o , \\ 0 & \text{otherwise ,} \end{cases}$$

and  $M$  denotes the total number of possible, equally-likely arrangements in the reference set of all permutations of cell frequencies in Table 10.12 given the observed marginal frequency distributions. When  $M$  is very large, as is typical with multi-way contingency tables, exact tests are impractical and Monte Carlo resampling becomes necessary, where a random sample,  $L$ , of the  $M$  possible arrangements of cell frequencies provides for a comparison of  $\kappa$  test statistics calculated on the  $L$  random tables with the  $\kappa$  test statistic calculated on the observed table.

**Unweighted Kappa**

Unweighted kappa and weighted kappa, with either linear or quadratic weighting, yield the same result when analyzing agreement data for  $b = 2$  judges and  $c = 2$  categories. For  $b > 2$  judges and  $c = 2$  categories, unweighted kappa and weighted kappa usually yield different results, but weighted kappa with linear weighting and weighted kappa with quadratic weighting yield the same result. For the frequency data given in Table 10.12, assign the number 1 to the categories labeled “Pro” and the number 2 to the categories labeled “Con.” Then the cell disagreement “weights” for unweighted kappa are given by

$$w_{ijk} = \begin{cases} 0 & \text{if } i = j = k , \\ 1 & \text{otherwise .} \end{cases}$$

Following Eq.(10.5) on p. 602, Cohen’s unweighted kappa coefficient is  $\kappa = +0.1862$ , indicating approximately 19% agreement among the  $b = 3$  judges above that expected by chance. If  $\kappa_o$  denotes the observed value of  $\kappa$ , the approximate Monte Carlo resampling probability value based on  $L = 1,000,000$  random arrangements of the cell frequencies, given the observed row, column, and slice marginal frequency distributions, {123, 131}, {135, 119}, and {134, 120},

respectively, is

$$P(\kappa \geq \kappa_o | H_0) = \frac{\text{number of } \kappa \text{ values } \geq \kappa_o}{L} = \frac{2,250}{1,000,000} = 0.0023 .$$

### Weighted Kappa

For the frequency data given in Table 10.12, assign the number 1 to the categories labeled “Pro” and the number 2 to the categories labeled “Con.” Then the linear cell disagreement weights are given by

$$w_{ijk} = |i - j| + |i - k| + |j - k|$$

and the quadratic cell disagreement weights are given by

$$w_{ijk} = (i - j)^2 + (i - k)^2 + (j - k)^2$$

for  $i, j, k = 1, \dots, c$ . Table 10.13 lists the eight cell indices and the associated linear and quadratic weights for a  $2 \times 2 \times 2$  agreement table, demonstrating that with  $c = 2$  categories, the linear and quadratic weights are identical.

Following Eq. (10.5) on p. 602, Cohen’s weighted kappa with linear weighting is  $\kappa_w = +0.0342$ , indicating approximately 3% agreement among the  $b = 3$  judges above that expected by chance. If  $\kappa_o$  denotes the observed value of  $\kappa_w$ , the approximate Monte Carlo resampling probability value based on  $L = 1,000,000$  random arrangements of the cell frequencies, given the observed row, column, and slice marginal frequency distributions, {123, 131}, {135, 119}, and {134, 120}, respectively, is

$$P(\kappa_w \geq \kappa_o | H_0) = \frac{\text{number of } \kappa_w \text{ values } \geq \kappa_o}{L} = \frac{190,610}{1,000,000} = 0.1906 .$$

**Table 10.13** Cells, linear weights, and quadratic weights for  $b = 3$  independent judges and  $c = 2$  disjoint categories

Cell	Weight	
	Linear	Quadratic
111	0	0
112	2	2
121	2	2
122	2	2
211	2	2
212	2	2
221	2	2
222	0	0

Because with  $c = 2$  categories the linear and quadratic weights are the same, the results are identical with quadratic weighting, i.e.,  $\kappa_w = +0.0342$  and  $P = 0.1906$ .

### 10.5.3 Example 3

For this third example of Cohen's chance-corrected measure of inter-rater agreement, consider  $b = 4$  judges who independently assign  $N = 76$  observations to  $c = 2$  disjoint, unordered categories labeled Pro and Con. The frequency data are given in Table 10.14.

A generalized calculation formula that applies to both unweighted and weighted kappa for  $b = 4$  judges and  $c$  categories is given by

$$\kappa = 1 - \frac{N^3 \sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^c \sum_{l=1}^c w_{ijkl} n_{ijkl}}{\sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^c \sum_{l=1}^c w_{ijkl} R_i C_j S_k L_l}, \tag{10.6}$$

where  $n_{ijkl}$  denotes the observed cell frequencies,  $w_{ijkl}$  denotes the cell weights,  $R_i$ ,  $C_j$ ,  $S_k$ , and  $L_l$  denote the observed row, column, slice, and level marginal frequency totals for  $i, j, k, l = 1, \dots, c$ , and

$$N = \sum_{i=1}^c \sum_{j=1}^c \sum_{k=1}^c \sum_{l=1}^c n_{ijkl}$$

denotes the table frequency total.

Given a  $c \times c \times c \times c$  agreement table with  $N$  objects cross-classified by  $b = 4$  independent judges, an exact permutation test involves generating all possible, equally-likely arrangements of the  $N$  objects to the  $c^4$  cells, while preserving the

**Table 10.14** Example  $2 \times 2 \times 2 \times 2$  contingency table for  $b = 4$  independent judges and  $c = 2$  disjoint categories

Judge 1	Judge 2	Judge 3	Judge 4	
			Pro	Con
Pro	Pro	Pro	5	2
		Con	4	1
	Con	Pro	7	3
		Con	9	2
Con	Pro	Pro	8	4
		Con	1	9
	Con	Pro	7	3
		Con	3	8

observed row, column, slice, and level marginal frequency distributions, {33, 43}, {34, 42}, {39, 37}, and {44, 32}, respectively. For each arrangement of cell frequencies, the kappa statistic,  $\kappa$ , and the exact hypergeometric point probability value under the null hypothesis,  $p(n_{ijkl}|R_i, C_j, S_k, L_l, N)$ , are calculated, where

$$p(n_{ijkl}|R_i, C_j, S_k, L_l, N) = \frac{\left(\prod_{i=1}^c R_i!\right) \left(\prod_{j=1}^c C_j!\right) \left(\prod_{k=1}^c S_k!\right) \left(\prod_{l=1}^c L_l!\right)}{(N!)^3 \prod_{i=1}^c \prod_{j=1}^c \prod_{k=1}^c \prod_{l=1}^c n_{ijkl}!}$$

[20].

If  $\kappa_0$  denotes the value of the observed kappa test statistic, the exact probability value of  $\kappa_0$  under the null hypothesis is given by

$$P(\kappa_0|H_0) = \sum_{l=1}^M \Psi_l(n_{ijkl}|R_i, C_j, S_k, L_l, N) ,$$

where

$$\Psi_l(n_{ijkl}|R_i, C_j, S_k, L_l, N) = \begin{cases} p(n_{ijkl}|R_i, C_j, S_k, L_l, N) & \text{if } \kappa \geq \kappa_0 , \\ 0 & \text{otherwise ,} \end{cases}$$

and  $M$  denotes the total number of possible, equally-likely arrangements in the reference set of all permutations of cell frequencies in Table 10.14 given the row, column, slice, and level observed marginal frequency distributions, {33, 43}, {34, 42}, {39, 37}, and {44, 32}, respectively. When  $M$  is very large, as is typical with multi-way contingency tables, exact tests are impractical and Monte Carlo resampling becomes necessary, where a random sample,  $L$ , of the  $M$  possible arrangements of cell frequencies provides for a comparison of  $\kappa$  test statistics calculated on the  $L$  random tables with the  $\kappa$  test statistic calculated on the observed table.

### Unweighted Kappa

For the frequency data given in Table 10.14, assign the number 1 to the categories labeled “Pro” and the number 2 to the categories labeled “Con.” Then the cell disagreement “weights” for unweighted kappa are given by

$$w_{ijkl} = \begin{cases} 0 & \text{if } i = j = k = l , \\ 1 & \text{otherwise .} \end{cases}$$

Following Eq. (10.6) on p. 605, Cohen's unweighted kappa coefficient is  $\kappa = +0.0561$ , indicating approximately 6% agreement among the  $b = 4$  judges above that expected by chance. If  $\kappa_o$  denotes the observed value of  $\kappa$ , the approximate Monte Carlo resampling probability value based on  $L = 1,000,000$  random arrangements of the cell frequencies, given the observed marginal frequency distributions, is

$$P(\kappa \geq \kappa_o | H_0) = \frac{\text{number of } \kappa \text{ values } \geq \kappa_o}{L} = \frac{9,475}{1,000,000} = 0.0095 .$$

### Weighted Kappa

For the frequency data given in Table 10.14, assign the number 1 to the categories labeled "Pro" and the number 2 to the categories labeled "Con." Then the linear cell disagreement weights are given by

$$w_{ijkl} = |i - j| + |i - k| + |i - l| + |j - k| + |j - l| + |k - l|$$

and the quadratic cell disagreement weights are given by

$$w_{ijkl} = (i - j)^2 + (i - k)^2 + (i - l)^2 + (j - k)^2 + (j - l)^2 + (k - l)^2$$

for  $i, j, k, l = 1, \dots, c$ .

Table 10.15 lists the 16 cell indices and the associated linear and quadratic weights for a  $2 \times 2 \times 2 \times 2$  agreement table. Note that for  $c = 2$  categories, the linear and quadratic weights are identical.

**Table 10.15** Cells, linear weights, and quadratic weights for  $b = 4$  independent judges and  $c = 2$  disjoint categories

Cell	Weight	
	Linear	Quadratic
1111	0	0
1112	3	3
1121	3	3
1122	4	4
1211	3	3
1212	4	4
1221	4	4
1222	3	3
2111	3	3
2112	4	4
2121	4	4
2122	3	3
2211	4	4
2212	3	3
2221	3	3
2222	0	0

Following Eq. (10.6) on p. 605, Cohen's weighted kappa with linear weighting is  $\kappa_w = +0.0654$ , indicating approximately 7% agreement among the  $b = 4$  judges above that expected by chance. If  $\kappa_o$  denotes the observed value of  $\kappa_w$ , the approximate Monte Carlo resampling probability value based on  $L = 1,000,000$  random arrangements of the cell frequencies, given the observed row, column, slice, and level marginal frequency distributions, {33, 43}, {34, 42}, {39, 37}, and {44, 32}, respectively, is

$$P(\kappa_w \geq \kappa_o | H_0) = \frac{\text{number of } \kappa_w \text{ values } \geq \kappa_o}{L} = \frac{3,967}{1,000,000} = 0.0040 .$$

Because, with  $c = 2$  categories, the linear and quadratic weights are the same, the results are identical to those obtained with quadratic weighting, i.e.,  $\kappa_w = +0.0654$  and  $P = 0.0040$ .

## 10.6 McNemar's and Cochran's $Q$ Tests for Change

In 1947 Quinn McNemar proposed a test for change over  $k = 2$  time periods [18]. In 1950 William Cochran developed a test for change for  $k \geq 2$  time periods [4]. For  $k = 2$ , Cochran's  $Q$  test for related proportions is identical to McNemar's  $Q$  test for related proportions. The McNemar and Cochran  $Q$  tests are described in detail in Chap. 4, Sects. 4.6 and 4.7, respectively.

### 10.6.1 McNemar's $Q$ Test for Change

Represent a  $2 \times 2$  contingency table as in Table 10.16. Then, McNemar's test for change is given by

$$Q = \frac{(B - C)^2}{B + C} ,$$

where  $B$  and  $C$  represent the two cells of change, i.e., Pro to Con and Con to Pro.

**Table 10.16** Notation for a  $2 \times 2$  cross-classification for McNemar's test for change

Time 1	Time 2		Total
	Pro	Con	
Pro	A	B	A + B
Con	C	D	C + D
Total	A + C	B + D	N



**Illustration**

To illustrate the calculation of probability values for McNemar's  $Q$  test for change, consider the frequency data given in Table 10.17, where  $N = 9$  subjects have been recorded as either Pro or Con on a specified issue at Time 1 and again on the same issue at Time 2. For the frequency data given in Table 10.17, the observed value of McNemar's  $Q$  test statistic is

$$Q = \frac{(B - C)^2}{B + C} = \frac{(5 - 1)^2}{5 + 1} = 2.6667 .$$

The exact probability value of an observed value of  $Q$ , under the null hypothesis, is given by the sum of the hypergeometric point probability values associated with the  $Q$  values equal to or greater than the observed value of  $Q$ . For the frequency data given in Table 10.17, there are only

$$M = \min(A + B, A + C) - \max(0, A - D) + 1$$

$$= \min(7, 3) - \max(0, 2 - 1) + 1 = 3 - 1 + 1 = 3$$

possible, equally-likely arrangements in the reference set of all permutations of cell frequencies given the two cell frequencies of change, 5 and 1, and only two  $Q$  values are equal to or greater than the observed value of  $Q = 2.6667$ . The exact upper-tail probability of the observed  $Q$  value is  $P = 0.9167$ , i.e., the sum of the hypergeometric point probability values associated with values of  $Q = 2.6667$  or greater.

More specifically, Table 10.18 displays the complete reference set of three possible  $2 \times 2$  contingency tables given the row and column marginal frequency distributions,  $\{7, 2\}$  and  $\{3, 6\}$ , respectively. For Table A in Table 10.18,  $Q = 2.0000$  and the associated hypergeometric point probability value is  $p = 0.0833$ . For

**Table 10.17** Example frequency data for McNemar's test for change with  $N = 9$  subjects

	Time 2		Total
	Pro	Con	
Time 1 Pro	2	5	7
Con	1	1	2
Total	3	6	9

**Table 10.18** Three possible cell arrangements given the marginal frequency distributions  $\{7, 2\}$  and  $\{3, 6\}$ ,  $Q$  values, and hypergeometric point probability values

Table	Frequency	$Q$	Probability
A	1 6		
	2 0	2.0000	0.0833
B	2 5		
	1 1	2.6667	0.5000
C	3 4		
	0 2	4.0000	0.4167

Table B in Table 10.18, the observed table,  $Q = 2.6667$  and the associated hypergeometric point probability value is  $p = 0.5000$ . And for Table C in Table 10.18,  $Q = 4.0000$  and the associated hypergeometric point probability value is  $p = 0.4167$ . Thus, the cumulative hypergeometric probability value for  $Q = 2.6667$  is the sum of the hypergeometric point probability values associated with values of  $Q = 2.6667$  or greater; in this case, the probability values associated with  $Q = 2.6667$  and  $Q = 4.0000$ , i.e.,  $P = 0.5000 + 0.4167 = 0.9167$ .

McNemar's  $Q$  test statistic is approximately distributed as chi-squared with 1 degree of freedom. While no responsible researcher would knowingly fit a chi-squared distribution function to only three possible outcomes, small samples, such as in Table 10.17, sometimes occur inadvertently. Suppose a researcher is employed by a national food service provider and begins with a reasonable, but small sample of subjects. As the research analysis proceeds, an interest develops in a subset of subjects composed of only women, breast-feeding their first child, and residing on a Native American reservation. Such unplanned small samples are relatively common and are not suitable for a conventional analysis. The chi-squared value for the observed data in Table 10.17 is  $\chi^2 = 0.3214$  and the probability value is  $P = 0.5708$ , which, as expected, is far removed from the exact probability value of  $P = 0.9167$ .

### Example

A more realistic example illustrating McNemar's  $Q$  test for change is given in Table 10.19, where  $N = 70$  subjects were recorded as either Pro or Con on a specified issue at Time 1 and again on the same issue at Time 2. At Time 1, 40 of the 70 subjects were in favor of the issue and 30 subjects were opposed. At Time 2, 50 subjects were in favor and 20 were opposed. Of those subjects that changed, seven changed from Pro to Con and 17 changed from Con to Pro. For the frequency data given in Table 10.19, McNemar's test statistic is

$$Q = \frac{(B - C)^2}{B + C} = \frac{(7 - 17)^2}{7 + 17} = \frac{100}{24} = 4.1667 .$$

**Table 10.19** Example frequency data for McNemar's test for change with  $N = 70$  subjects

Time 1	Time 2		Total
	Pro	Con	
Pro	33	7	40
Con	17	13	30
Total	50	20	70

For the frequency data given in Table 10.19, there are only

$$M = \min(A + B, A + C) - \max(0, A - D) + 1$$

$$= \min(40, 50) - \max(0, 33 - 13) + 1 = 40 - 20 + 1 = 21$$

possible, equally-likely arrangements in the reference set of all permutations of cell frequencies given the observed row and column marginal frequency distributions,  $\{40, 30\}$  and  $\{50, 20\}$ , respectively, making an exact permutation analysis possible. Since  $M = 21$  is a reasonably small number of arrangements, it will be illustrative to list the 21 sets of cell frequencies, McNemar's  $Q$  values, and the associated hypergeometric point probability values in Table 10.20, where the rows with hypergeometric probability values associated with  $Q$  values equal to or greater than the observed value of  $Q = 4.1667$  are indicated with asterisks.

If the  $M = 21$  possible arrangements of the frequency data given in Table 10.19 occur with equal chance, the exact probability of  $Q$  under the null hypothesis is the sum of the hypergeometric point probability values associated with  $Q = 4.1667$  or greater. For the frequency data given in Table 10.19, the exact upper-tail probability

**Table 10.20** Cell frequencies, McNemar's  $Q$  values, and exact hypergeometric point probability values for  $M = 21$  possible arrangements of the observed data in Table 10.19

Table	Cell frequencies				$Q$	Probability
	$n_{11}$	$n_{12}$	$n_{21}$	$n_{22}$		
1	20	20	30	0	2.0000	$0.8515 \times 10^{-6}$
2	21	19	29	1	2.0833	$0.2433 \times 10^{-4}$
3	22	18	28	2	2.1739	$0.3047 \times 10^{-3}$
4	23	17	27	3	2.2727	$0.2225 \times 10^{-2}$
5	24	16	26	4	2.3809	$0.1064 \times 10^{-1}$
6	25	15	25	5	2.5000	$0.3541 \times 10^{-1}$
7	26	14	24	6	2.6316	$0.8512 \times 10^{-1}$
8	27	13	23	7	2.7778	0.1513
9	28	12	22	8	2.9412	0.2020
10	29	11	21	9	3.1250	0.2043
11	30	10	20	10	3.3333	0.1573
12	31	9	19	11	3.5714	$0.9227 \times 10^{-1}$
13	32	8	18	12	3.8462	$0.4019 \times 10^{-1}$
14*	33	7	17	13	4.1667	$0.1379 \times 10^{-1}$
15*	34	6	16	14	4.5455	$0.3448 \times 10^{-2}$
16*	35	5	15	15	5.0000	$0.6305 \times 10^{-3}$
17*	36	4	14	16	5.5556	$0.8210 \times 10^{-4}$
18*	37	3	13	17	6.2500	$0.7309 \times 10^{-5}$
19*	38	2	12	18	7.1429	$0.4167 \times 10^{-6}$
20*	39	1	11	19	8.3333	$0.1350 \times 10^{-7}$
21*	40	0	10	20	10.0000	$0.1856 \times 10^{-9}$
Sum						1.0000

of the observed value of  $Q$  value is

$$\begin{aligned}
 P &= 0.1379 \times 10^{-1} + 0.3448 \times 10^{-2} + 0.6305 \times 10^{-3} + 0.8210 \times 10^{-4} \\
 &\quad + 0.7309 \times 10^{-5} + 0.4167 \times 10^{-6} + 0.1350 \times 10^{-7} + 0.1856 \times 10^{-9} \\
 &= 0.0180 .
 \end{aligned}$$

For comparison, the value of chi-squared for the frequency data given in Table 10.19 is  $\chi^2 = 5.6058$  and with 1 degree of freedom, the probability value is  $P = 0.0179$ , which compares favorably with the exact probability value of  $P = 0.0180$ .

### 10.6.2 Cochran's $Q$ Test for Change

Cochran's  $Q$  test for  $k \geq 2$  treatments can be considered an extension of McNemar's  $Q$  test for  $k = 2$  treatments or time periods. Cochran's  $Q$  test is described more completely in Chap. 4, Sect. 4.7.

Cochran's  $Q$  test for the analysis of  $k$  treatment conditions (columns) and  $N$  subjects (rows) is given by

$$Q = \frac{(k-1) \left( k \sum_{j=1}^k C_j^2 - A^2 \right)}{kA - B}, \quad (10.7)$$

where

$$C_j = \sum_{i=1}^N x_{ij}$$

is the number of 1s in the  $j$ th of  $k$  columns,

$$R_i = \sum_{j=1}^k x_{ij}$$

is the number of 1s in the  $i$ th of  $N$  rows,

$$A = \sum_{i=1}^N R_i, \quad B = \sum_{i=1}^N R_i^2,$$

and  $x_{ij}$  denotes the cell entry of either 0 or 1 associated with the  $i$ th of  $N$  rows and the  $j$ th of  $k$  columns. The null hypothesis stipulates that each of the

$$M = \prod_{i=1}^N \binom{k}{R_i}$$

distinguishable arrangements of 1s and 0s within each of the  $N$  rows occurs with equal probability, given that the values of  $R_1, \dots, R_N$  are fixed [21].

**Example**

To illustrate Cochran's  $Q$  test for change, consider the binary data listed in Table 10.21 consisting of the responses (1 or 0) for  $N = 9$  subjects evaluated over  $k = 3$  time periods, where a 1 indicates success on a prescribed task and a 0 indicates failure. For the binary data listed in Table 10.21,

$$\sum_{j=1}^k C_j^2 = 1^2 + 8^2 + 5^2 = 90 ,$$

$$A = \sum_{i=1}^N R_i = 2 + 2 + 2 + 2 + 1 + 1 + 1 + 1 + 2 = 14 ,$$

$$B = \sum_{i=1}^N R_i^2 = 2^2 + 2^2 + 2^2 + 2^2 + 1^1 + 1^2 + 1^2 + 1^2 + 2^2 = 24 ,$$

**Table 10.21** Successes (1) and failures (0) of  $N = 9$  subjects on a series of  $k = 3$  time periods

Subject	Time			$R_i$
	1	2	3	
1	0	1	1	2
2	0	1	1	2
3	0	1	1	2
4	0	1	1	2
5	0	1	0	1
6	0	1	0	1
7	1	0	0	1
8	0	1	0	1
9	0	1	1	2
$C_j$	1	8	5	14

and, following Eq. (10.7) on p. 612, the observed value of Cochran's  $Q$  is

$$Q = \frac{(k-1) \left( k \sum_{j=1}^k C_j^2 - A^2 \right)}{kA - B} = \frac{(3-1)[(3)(90) - 14^2]}{(3)(14) - 24} = 8.2222 .$$

For the binary data listed in Table 10.21, there are only

$$M = \prod_{i=1}^N \binom{k}{R_i} = \binom{3}{1}^4 \binom{3}{2}^5 = (3^4)(3^5) = 19,683$$

possible, equally-likely arrangements in the reference set of all permutations of the observed binary data, making an exact permutation analysis feasible. Based on  $M = 19,683$  arrangements of the observed data, there are 312  $Q$  values equal to or greater than the observed value of  $Q = 8.2222$ . If  $Q_o$  denotes the observed value of  $Q$ , the exact upper-tail probability value of the observed data is

$$P(Q \geq Q_o | H_0) = \frac{\text{number of } Q \text{ values } \geq Q_o}{M} = \frac{312}{19,683} = 0.0159 .$$

For comparison, under the null hypothesis Cochran's  $Q$  is approximately distributed as chi-squared with  $k - 1$  degrees of freedom. The approximate probability of  $Q = 8.2222$  with  $k - 1 = 3 - 1 = 2$  degrees of freedom is  $P = 0.0164$ .

## 10.7 Fisher's Exact Probability Test

Fisher's exact probability test was independently developed by R.A. Fisher, Joseph Irwin, and Frank Yates in the early 1930s [11, 14, 34]. Characteristically, Fisher's exact test is applied to  $2 \times 2$  contingency tables, but can be generalized and extended to more complex contingency tables. The eponymous exact test for  $2 \times 2$  tables and several extensions are detailed in Chap. 4, Sects. 4.11 and 4.12. In this chapter on fourfold contingency tables, only  $2 \times 2$  and  $2 \times 2 \times 2$  contingency tables are considered.

### 10.7.1 Analysis of $2 \times 2$ Contingency Tables

Consider a  $2 \times 2$  contingency table containing  $N$  cases, where  $x_o$  denotes the observed frequency of any cell and  $r$  and  $c$  represent the row and column marginal frequency totals, respectively, corresponding to  $x_o$ . Table 10.22 illustrates the notation for a  $2 \times 2$  contingency table.

**Table 10.22** Example notation for a 2×2 contingency table

	A <sub>1</sub>	A <sub>2</sub>	Total
B <sub>1</sub>	$x$	$r - x$	$r$
B <sub>2</sub>	$c - x$	$N - r - c + x$	$N - r$
Total	$c$	$N - c$	$N$

**Table 10.23** Example 2×2 contingency table for Fisher's exact test

	A <sub>1</sub>	A <sub>2</sub>	Total
B <sub>1</sub>	13	2	15
B <sub>2</sub>	7	8	15
Total	20	10	30

Given the notation in Table 10.22, Fisher's exact test for 2×2 contingency tables is given by

$$P = \sum_{x=a}^b p(x|r, c, N) ,$$

where  $a = \max(0, r + c - N)$ ,  $b = \min(r, c)$ , and the hypergeometric point probability value is given by

$$p(x|r, c, N) = \binom{c}{x} \binom{N - c}{r - x} \binom{N}{r}^{-1} = \frac{r! (N - r)! c! (N - c)!}{N! x! (r - x)! (N - r - c + x)!} .$$

To illustrate Fisher's exact probability test for a multi-way contingency table, consider the 2×2 contingency table given in Table 10.23 where  $x_0 = 13$ ,  $r = 15$ ,  $c = 20$ , and  $N = 30$ .

For the frequency data given in Table 10.23, there are only

$$\begin{aligned} M &= \min(r, c) - \max(0, r + c - N) + 1 \\ &= \min(15, 20) - \max(0, 15 + 20 - 30) + 1 = 15 - 5 + 1 = 11 \end{aligned}$$

possible, equally-likely arrangements in the reference set of all permutations of cell frequencies given the observed row and column marginal frequency distributions, {15, 15} and {20, 10}, respectively, making an exact permutation analysis possible. Table 10.24 lists the  $M = 11$  possible values of  $x$  and associated hypergeometric point probability values to nine decimal places.

The exact probability value is obtained by summing all the hypergeometric point probability values equal to or less than the hypergeometric point probability value of the observed table, indicated with asterisks in Table 10.24. Thus,

$$P = 0.022488756 + 0.002498751 + 0.000099950 = 0.025087457$$

**Table 10.24** Probability values for  $M = 11$  possible arrangements of cell frequencies in Table 10.23, given the marginal frequency distributions  $\{15, 15\}$  and  $\{20, 10\}$

$x$	$p(x r, c, N)$
5	0.000099950
6	0.002498751
7	0.022488756
8	0.097451274
9	0.227386307
10	0.300149925
11	0.227386307
12	0.097451274
13*	0.022488756
14*	0.002498751
15*	0.000099950
Total	1.000000000

for the upper tail of the distribution, i.e., the sum of the hypergeometric point probability values associated with  $x = 13, 14,$  and  $15$ . Since the probability distribution is symmetric in this case, the exact hypergeometric probability value is twice the probability of the upper tail, i.e.,  $P = 2(0.0251) = 0.0502$ .

### 10.7.2 Analysis of $2 \times 2 \times 2$ Contingency Tables

Analyses of multi-way contingency tables are more complex than simple two-way tables; see Chap. 4, Sect. 4.12. For a two-way contingency table the degrees of freedom are given by  $df = (r - 1)(c - 1)$ , where  $r$  denotes the number of rows and  $c$  denotes the number of columns. Thus, in the case of a  $2 \times 2$  contingency table the degrees of freedom are  $(2 - 1)(2 - 1) = 1$  and only one cell frequency need be permuted over its range. In the  $2 \times 2$  example above, the chosen cell  $(A_1 B_1)$  was designated as  $x$  in Table 10.22.

For multi-way contingency tables the degrees of freedom are given by

$$df = \prod_{i=1}^r c_i - \sum_{i=1}^r (c_i - 1) - 1,$$

where  $r$  denotes the number of dimensions and  $c_i$  denotes the number of categories in each dimension,  $i = 1, \dots, r$  [24, p. 309]. Thus, for a  $2 \times 2 \times 2$  contingency table with  $c = 2$  disjoint categories in each of  $r = 3$  dimensions,

$$df = 2^3 - 3(2 - 1) - 1 = 4.$$

Consider a  $2 \times 2 \times 2$  contingency table where  $n_{ijk}$  denotes the cell frequency of the  $i$ th row,  $j$ th column, and  $k$ th slice for  $i, j, k = 1, 2$ . Let  $A = n_{1..}$ ,  $B = n_{.1.}$ ,



$C = n_{..1}$ , and  $N = n_{...}$  denote the observed marginal frequency totals of the first row, first column, first slice, and entire table, respectively, such that  $1 \leq A \leq B \leq C \leq N/2$ . Also, let  $w = n_{111}$ ,  $x = n_{112}$ ,  $y = n_{121}$ , and  $z = n_{211}$  denote four cell frequencies of the  $2 \times 2 \times 2$  contingency table. Then, the probability for any specified  $w, x, y$ , and  $z$  is given by

$$p(w, x, y, z|A, B, C, N) = \frac{[A!(N - A)! B!(N - B)! C!(N - C)!]}{\times [(N!)^2 w! x! y! z! (A - w - x - y)! (B - w - x - z)! (C - w - y - z)! (N - A - B - C + 2w + x + y + z)!]^{-1}}$$

[26].

The bounds for  $w, x, y$ , and  $z$  are

$$\begin{aligned} 0 &\leq w \leq M_w, \\ 0 &\leq x \leq M_x, \\ 0 &\leq y \leq M_y, \end{aligned}$$

and

$$L_x \leq z \leq M_z,$$

respectively, where  $M_w = A$ ,  $M_x = A - w$ ,  $M_y = A - w - x$ ,  $M_z = \min(B - w - x, C - w - y)$ , and  $L_z = \max(0, A + B + C - N - 2w - x - y)$ . If  $w_o, x_o, y_o$ , and  $z_o$  denote the values of  $w, x, y$ , and  $z$  in the observed contingency table, then Fisher's exact probability value for a  $2 \times 2 \times 2$  contingency table is given by

$$P = \sum_{w=0}^{M_w} \sum_{x=0}^{M_x} \sum_{y=0}^{M_y} \sum_{z=L_z}^{M_z} p(w, x, y, z|A, B, C, N) \psi(w, x, y, z),$$

where

$$\psi(w, x, y, z) = \begin{cases} 1 & \text{if } p(w, x, y, z) \leq p(w_o, x_o, y_o, z_o), \\ 0 & \text{otherwise.} \end{cases}$$

To illustrate Fisher's exact probability test, consider the  $2 \times 2 \times 2$  contingency table given in Table 10.25 where  $N = 75$  and the observed values of  $w, x, y$ , and  $z$  are  $w_o = 13, x_o = 8, y_o = 4$ , and  $z_o = 18$ . For the frequency data given in Table 10.25 there are  $M = 77,910$  possible arrangements in the reference set of all permutations of cell frequencies given the observed row, column, and

**Table 10.25** Example  
 $2 \times 2 \times 2$  contingency table for  
 Fisher's exact test

Judge 1	Judge 2	Judge 3	
		Pro	Con
Pro	Pro	13	8
	Con	4	11
Con	Pro	18	5
	Con	9	7

slice marginal distributions,  $\{44, 31\}$ ,  $\{44, 31\}$ , and  $\{44, 31\}$ , respectively, making an exact permutation analysis feasible. Fisher's exact probability is the sum of the hypergeometric point probability values equal to or less than the probability value associated with the observed contingency table; in this case, there are 2,991 tables with probability values equal to or less than the probability value of the observed table, i.e.,  $p = 0.1743 \times 10^{-4}$ , yielding  $P = 0.0384$ .

## 10.8 Contingency Table Interactions

It is occasionally necessary to test the independence among multiple classification variables, each of which consists of two mutually exclusive classes, e.g., a  $2 \times 2 \times 2$  or  $2^3$  contingency table. In this section exact permutation procedures are described for analyzing interactions in  $2 \times 2 \times 2$  and  $2 \times 2 \times 2 \times 2$  contingency tables.

### 10.8.1 Analysis of $2 \times 2 \times 2$ Contingency Tables

Mielke, Berry, and Zelterman provided a procedure for determining the exact global probability value obtained from an examination of all possible arrangements of the eight cell frequencies of a  $2 \times 2 \times 2$  contingency table, conditioned on the observed marginal frequency totals [26]. An alternative approach that is not as computationally intensive and, quite possibly, more fruitful is to examine the first- and second-order interactions of a  $2 \times 2 \times 2$  table when the observed marginal frequency totals are considered to be fixed [22]. This approach was first proposed by Bartlett [1] and has been discussed by Darroch [9, 10], Haber [12, 13], Odoroff [27], Plackett [29], Pomar [30], Simpson [33], and Zachs and Solomon [35]. In this section an algorithm is described that computes the exact probability values of the three first-order (two-variable) interactions and the single second-order (three-variable) interaction.

The logic on which the algorithm is based was apparently first developed by Lambert Adolphe Jacques Quetelet to calculate binomial probability values in 1846 [31]. Beginning with a small arbitrary initial value, a simple recursion procedure generates relative frequency values for all possible  $2 \times 2 \times 2$  contingency

tables, given the observed marginal frequency totals. The desired exact probability value is obtained by summing the relative frequency values equal to or less than the observed relative frequency value and dividing the resultant sum by the unrestricted relative frequency total.

Consider a sample of  $N$  independent observations arranged in a  $2 \times 2 \times 2$  contingency table. Let  $n_{ijk}$  denote the observed cell frequency of the  $i$ th row,  $j$ th column, and  $k$ th slice, and let  $p_{ijk}$  denote the corresponding cell probability for  $i, j, k = 1, 2$ . Also let  $n_{.jk}, n_{i.k}, n_{ij.}, n_{1..}, n_{.j.}, n_{..k}$ , and  $n_{...}$  indicate the observed marginal frequency totals of the  $2 \times 2 \times 2$  contingency table, and let the corresponding marginals over  $p_{ijk}$  be indicated by  $p_{.jk}, p_{i.k}, p_{ij.}, p_{1..}, p_{.j.}, p_{..k}$ , and  $p_{...}$ , respectively, for  $i, j, k = 1, 2$ . Because the categories are mutually exclusive and exhaustive,  $n_{...} = N$  and  $p_{...} = 1$ .

Let  $r$  denote the number of dimensions and  $c_i$  denote the number of categories in each dimension,  $i = 1, \dots, r$ . Then for a  $2 \times 2 \times 2$  contingency table there are

$$\prod_{i=1}^r c_i - \sum_{i=1}^r (c_i - 1) - 1$$

$$= (2)(2)(2) - [(2 - 1) + (2 - 1) + (2 - 1)] - 1 = 8 - 3 - 1 = 4$$

degrees of freedom and, consequently, four interaction terms to be considered: three first-order and one second-order. Following Bartlett, the null hypotheses for the three first-order interactions are

$$H_0: p_{.11}p_{.22} = p_{.12}p_{.21} ,$$

$$H_0: p_{1.1}p_{2.2} = p_{1.2}p_{2.1} ,$$

and

$$H_0: p_{11.}p_{22.} = p_{12.}p_{21.}$$

[1]. The null hypothesis for the second-order interaction is

$$H_0: p_{111}p_{122}p_{212}p_{221} = p_{112}p_{121}p_{211}p_{222}$$

[1, 13, 28].

For simplicity, set  $x = n_{111}$ ,  $a = n_{.11}$ ,  $b = n_{1.1}$ ,  $c = n_{11.}$ ,  $A = n_{1..}$ ,  $B = n_{.1.}$ ,  $C = n_{..1}$ , and  $N = n_{...}$ . The point probability of  $x$  is given by

$$P(x|a, b, c, A, B, C, N) = [A! (N - A)! B! (N - B)! C! (N - C)!]$$

$$\times [(N!)^2 x! (a - x)! (b - x)! (c - x)! (A - b - c + x)!]$$

$$(B - a - c + x)! (C - a - b + x)! (N - A - B - C + a + b + c - x)!]^{-1} .$$

If  $H(k)$ , given  $a, b, c, A, B, C$ , and  $N$ , is a recursively defined positive function, then solving the recursive relation  $H(k + 1) = H(k) \times g(k)$  yields

$$g(k) = \frac{(a - k)(b - k)(c - k)(N - A - B - C + a + b + c - k)}{(k + 1)(A - b - c + k + 1)(B - a - c + k + 1)(C - a - b + k + 1)},$$

which may be used to enumerate the distribution of  $P(k|a, b, c, A, B, C, N)$ ,  $v \leq k \leq w$ , where

$$v = \max(0, b + c - A, a + c - B, a + b - C),$$

$$w = \min(a, b, c, N - A - B - C + a + b + c),$$

and where  $H(v)$  is initially set to some small value, such as  $10^{-20}$ . The total over the completely enumerated distribution may be found by

$$T = \sum_{k=v}^w H(k).$$

The exact second-order interaction probability value is found by

$$P = \sum_{k=v}^w \frac{H(k)I_k}{T},$$

where

$$I_k = \begin{cases} 1 & \text{if } H(k) \leq H(x), \\ 0 & \text{otherwise.} \end{cases}$$

### A 2×2×2 Contingency Table Example

Table 10.26 depicts a 2×2×2 contingency table based on  $N = 76$  responses to a question (Yes, No) classified by gender (Female, Male), in two elementary school grades (First, Fourth).

**Table 10.26** Cross-classification of yes/no responses, categorized by gender and elementary school grade

Grade	Gender			
	Female		Male	
	Yes	No	Yes	No
First	10	4	2	16
Fourth	6	11	15	12

Table 10.27 provides the cell frequencies for Grade by Gender, conditioned on Response. The first-order interaction probability value associated with the cell frequencies in Table 10.27 is

$$\begin{aligned}
 P(a|a + b, a + c, N) &= \binom{a + c}{a} \binom{b + d}{b} \binom{N}{a + b}^{-1} \\
 &= \binom{31}{14} \binom{45}{18} \binom{76}{32}^{-1} = \frac{32! 44! 31! 45!}{76! 14! 18! 17! 27!} = 0.8134 .
 \end{aligned}$$

Table 10.28 provides the cell frequencies for Gender by Response, conditioned on Grade. The first-order interaction probability value associated with the cell frequencies in Table 10.28 is

$$\begin{aligned}
 P(a|a + b, a + c, N) &= \binom{a + c}{a} \binom{b + d}{b} \binom{N}{a + b}^{-1} \\
 &= \binom{33}{16} \binom{43}{15} \binom{76}{31}^{-1} = \frac{31! 45! 33! 43!}{76! 16! 15! 17! 28!} = 0.2496 .
 \end{aligned}$$

Table 10.29 provides the cell frequencies for Grade by Response, conditioned on Gender. The first-order interaction probability value associated with the cell

**Table 10.27** Grade by Gender, conditioned on Response

Grade	Gender		Total
	Female	Male	
First	14	18	32
Fourth	17	27	44
Total	31	45	76

**Table 10.28** Gender by Response, conditioned on Grade

Gender	Response		Total
	Yes	No	
Female	16	15	31
Male	17	28	45
Total	33	43	76

**Table 10.29** Grade by Response, conditioned on Gender

Grade	Response		Total
	Yes	No	
First	12	20	32
Fourth	21	23	44
Total	33	43	76

frequencies in Table 10.29 is

$$\begin{aligned}
 P(a|a+b, a+c, N) &= \binom{a+c}{a} \binom{b+d}{b} \binom{N}{a+b}^{-1} \\
 &= \binom{33}{12} \binom{43}{20} \binom{76}{32}^{-1} = \frac{32! 44! 33! 43!}{76! 12! 20! 21! 23!} = 0.4830.
 \end{aligned}$$

The second-order interaction probability value for the frequency data given in Table 10.29 is  $P = 0.9036 \times 10^{-3}$  and the global probability of a table this extreme or more extreme than the observed table in Table 10.29 is  $P = 0.4453 \times 10^{-2}$  [26].

### 10.8.2 Analysis of $2 \times 2 \times 2 \times 2$ Contingency Tables

Utilizing the recursion procedure presented in the previous example, it is possible to analyze a  $2 \times 2 \times 2 \times 2$  or  $2^4$  contingency table [23]. The conditional probability value of a  $2 \times 2 \times 2 \times 2$  contingency table is a special case of the conditional probability of an  $r$ -way contingency table as defined in Mielke and Berry [20]. Zelterman, Chan, and Mielke [36] provided an algorithm for the exact global probability value obtained from an examination of all possible arrangements of the 16 cell frequencies of a  $2 \times 2 \times 2 \times 2$  contingency table, conditioned on the observed marginal frequency totals. An alternative approach is to examine the first-, second-, and third-order interactions in a  $2 \times 2 \times 2 \times 2$  table when the observed marginal frequency totals are considered to be fixed.

Let  $r$  denote the number of dimensions and  $c_i$  denote the number of categories in each dimension,  $i = 1, \dots, r$ , then for a  $2 \times 2 \times 2 \times 2$  contingency table there are

$$\begin{aligned}
 \prod_{i=1}^r c_i - \sum_{i=1}^r (c_i - 1) - 1 \\
 &= (2)(2)(2)(2) - [(2-1) + (2-1) + (2-1) + (2-1)] - 1 \\
 &= 16 - 4 - 1 = 11
 \end{aligned}$$

degrees of freedom and, consequently, 11 interaction terms to be considered: six first-order and four second-order, and one third-order. In this section, a procedure is described for computing the exact probability values of the six first-order (two-variable) interactions, the four second-order (three-variable) interactions, and the single third-order (four-variable) interactions for a  $2 \times 2 \times 2 \times 2$  contingency table.

Following Mielke [19], let  $p_{i_1 i_2 i_3 i_4}$  denote the probability of cell  $i_1 i_2 i_3 i_4$  in a  $2 \times 2 \times 2 \times 2$  contingency table, where the index  $i_j = 1$  or  $2$  for  $j = 1, 2, 3, 4$ . The six null hypotheses of no first-order interactions for a  $2 \times 2 \times 2 \times 2$  contingency table are

$$H_0: p_{1100}p_{2200} = p_{1200}p_{2100} ,$$

$$H_0: p_{1010}p_{2020} = p_{1020}p_{2010} ,$$

$$H_0: p_{1001}p_{2002} = p_{1002}p_{2001} ,$$

$$H_0: p_{0110}p_{0220} = p_{0120}p_{0210} ,$$

$$H_0: p_{0101}p_{0202} = p_{0102}p_{0201} ,$$

and

$$H_0: p_{0011}p_{0022} = p_{0012}p_{0021} ,$$

where the usual summation convention is employed. Thus,  $p_{0101}$  is the sum over indices  $i_1$  and  $i_3$ . The four null hypotheses of no second-order interaction for a  $2 \times 2 \times 2 \times 2$  contingency table are

$$H_0: p_{1110}p_{2210}p_{1220}p_{2120} = p_{1120}p_{2220}p_{1210}p_{2110} ,$$

$$H_0: p_{1101}p_{2201}p_{1202}p_{2102} = p_{1102}p_{2202}p_{1201}p_{2101} ,$$

$$H_0: p_{1011}p_{2021}p_{1022}p_{2012} = p_{1012}p_{2022}p_{1021}p_{2011} ,$$

and

$$H_0: p_{0111}p_{0221}p_{0122}p_{0212} = p_{0112}p_{0222}p_{0121}p_{0211} .$$

The null hypothesis of no third-order interaction for a  $2 \times 2 \times 2 \times 2$  contingency table is given by

$$\begin{aligned} H_0: p_{1111}p_{2211}p_{1221}p_{2121}p_{1122}p_{2222}p_{1212}p_{2112} \\ = p_{1112}p_{2212}p_{1222}p_{2122}p_{1121}p_{2221}p_{1211}p_{2111} . \end{aligned}$$

Table 10.30 contains data from a  $2 \times 2 \times 2 \times 2$  contingency table based on  $N = 1,356$  responses classified on four dichotomous variables:  $A$ ,  $B$ ,  $C$ , and  $D$ . The first-, second-, and third-order interaction exact probability values associated with the data listed in Table 10.30 are given in Table 10.31.

**Table 10.30** Example data for a  $2 \times 2 \times 2 \times 2$  contingency table

Variable				Frequency
A	B	C	D	
1	1	1	1	187
1	1	1	2	15
1	1	2	1	42
1	1	2	2	40
1	2	1	1	256
1	2	1	2	42
1	2	2	1	34
1	2	2	2	62
2	1	1	1	177
2	1	1	2	14
2	1	2	1	30
2	1	2	2	63
2	2	1	1	194
2	2	1	2	27
2	2	2	1	52
2	2	2	2	121
<i>N</i>				1,356

**Table 10.31** Interactions and associated exact hypergeometric probability values for the data listed in Table 10.30

Interaction	Probability
A×B	$0.3822 \times 10^{-9}$
A×C	$0.4891 \times 10^{-3}$
A×D	$0.8690 \times 10^{-4}$
B×C	0.2181
B×D	$0.5475 \times 10^{-5}$
C×D	1.0000
A×B×C	0.4491
A×B×D	$0.2792 \times 10^{-1}$
A×C×D	0.7999
B×C×D	$0.4021 \times 10^{-2}$
A×B×C×D	$0.6517 \times 10^{-1}$

## 10.9 Coda

Chapter 10 applied exact and Monte Carlo permutation statistical methods to measures of association for symmetrical  $2 \times 2$  contingency tables. Included in Chap. 10 were discussions of Pearson’s  $\phi$ , Tschuprov’s  $T^2$ , and Cramér’s  $V^2$  coefficients of contingency, Pearson’s product-moment correlation coefficient, Leik and Gove’s  $d_N^c$  measure, Goodman and Kruskal’s  $t_a$  and  $t_b$  asymmetric measures of nominal association, Kendall’s  $\tau_b$  and Stuart’s  $\tau_c$  measures of ordinal association, Somers’  $d_{yx}$  and  $d_{xy}$  asymmetric measures of ordinal association, Yule’s  $Y$  measure of nominal association, simple percentage differences, and Cohen’s unweighted and weighted  $\kappa$  measures of inter-rater agreement.



Chapter 10 concluded with an examination of extensions to multiple  $2 \times 2$  contingency tables and  $2 \times 2 \times 2$  contingency tables, including the Mantel–Haenszel test for combined  $2 \times 2$  contingency tables, Cohen’s chance-corrected measure of inter-rater agreement, McNemar’s and Cochran’s  $Q$  tests for change, Fisher’s exact test for  $2 \times 2 \times 2$  and  $2 \times 2 \times 2 \times 2$  contingency tables, and tests for interactions in  $2 \times 2 \times 2$  and  $2 \times 2 \times 2 \times 2$  contingency tables.

## References

1. Bartlett, M.S.: Contingency table interactions. *Suppl. J. R. Stat. Soc.* **2**, 248–252 (1935)
2. Berry, K.J., Johnston, J.E., Mielke, P.W.: Maximum-corrected and chance-corrected measures of effect size for the Mantel–Haenszel test. *Psychol. Rep.* **107**, 393–401 (2010)
3. Berry, K.J., Mielke, P.W.: A generalization of Cohen’s kappa agreement measure to interval measurement and multiple raters. *Educ. Psychol. Meas.* **48**, 921–933 (1988)
4. Cochran, W.G.: The comparison of percentages in matched samples. *Biometrika* **37**, 256–266 (1950)
5. Cochran, W.G.: Some methods for strengthening the common  $\chi^2$  test. *Biometrics* **10**, 417–452 (1954)
6. Cohen, J.: A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960)
7. Cohen, J.: Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* **70**, 213–220 (1968)
8. Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Erlbaum, Hillsdale, NJ (1988)
9. Darroch, J.N.: Interactions in multi-factor contingency tables. *J. R. Stat. Soc. B Meth.* **24**, 251–263 (1962)
10. Darroch, J.N.: Multiplicative and additive interaction in contingency tables. *Biometrika* **61**, 207–214 (1974)
11. Fisher, R.A.: The logic of inductive inference (with discussion). *J. R. Stat. Soc.* **98**, 39–82 (1935)
12. Haber, M.: Sample sizes for the exact test of “no interaction” in  $2 \times 2 \times 2$  tables. *Biometrics* **39**, 493–498 (1983)
13. Haber, M.: A comparison of tests for the hypothesis of no three-factor interaction in  $2 \times 2 \times 2$  contingency tables. *J. Stat. Comp. Sim.* **20**, 205–215 (1984)
14. Irwin, J.O.: Tests of significance for differences between percentages based on small numbers. *Metron* **12**, 83–94 (1935)
15. Leik, R.K., Gove, W.R.: Integrated approach to measuring association. In: Costner, H.L. (ed.) *Sociological Methodology*, pp. 279–301. Jossey Bass, San Francisco, CA (1971)
16. Maclure, M., Willett, W.C.: Misinterpretation and misuse of the kappa statistic. *Am. J. Epidemiol.* **126**, 161–169 (1987)
17. Mantel, N., Haenszel, W.: Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer I* **22**, 719–748 (1959)
18. McNemar, Q.: Note on the sampling error of the differences between correlated proportions and percentages. *Psychometrika* **12**, 153–157 (1947)
19. Mielke, P.W.: Some exact and nonasymptotic analyses of discrete goodness-of-fit and  $r$ -way contingency tables. In: Johnson, N.L., Balakrishnan, N. (eds.) *Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz*, pp. 179–192. Wiley, New York (1997)
20. Mielke, P.W., Berry, K.J.: Cumulant methods for analyzing independence of  $r$ -way contingency tables and goodness-of-fit frequency data. *Biometrika* **75**, 790–793 (1988)

21. Mielke, P.W., Berry, K.J.: Nonasymptotic inferences based on Cochran's  $Q$  test. *Percept. Motor Skill* **81**, 319–322 (1995)
22. Mielke, P.W., Berry, K.J.: Exact probabilities for first-order and second-order interactions in  $2 \times 2 \times 2$  contingency tables. *Educ. Psychol. Meas.* **56**, 843–847 (1996)
23. Mielke, P.W., Berry, K.J.: Exact probabilities for first-order, second-order, and third-order interactions in  $2 \times 2 \times 2 \times 2$  contingency tables. *Percept. Motor Skill* **86**, 760–762 (1998)
24. Mielke, P.W., Berry, K.J.: *Permutation Methods: A Distance Function Approach*, 2nd edn. Springer-Verlag, New York (2007)
25. Mielke, P.W., Berry, K.J., Johnston, J.E.: Resampling probability values for weighted kappa with multiple raters. *Psychol. Rep.* **102**, 606–613 (2008)
26. Mielke, P.W., Berry, K.J., Zelterman, D.: Fisher's exact test of mutual independence for  $2 \times 2 \times 2$  cross-classification tables. *Educ. Psychol. Meas.* **54**, 110–114 (1994)
27. Odoroff, C.L.: A comparison of minimum logit chi-square estimation and maximum likelihood estimation in  $2 \times 2 \times 2$  and  $3 \times 2 \times 2$  contingency tables: Tests for interaction. *J. Am. Stat. Assoc.* **65**, 1617–1631 (1970)
28. O'Neill, M.E.: A comparison of the additive and multiplicative definitions of second-order interaction in  $2 \times 2 \times 2$  contingency tables. *J. Stat. Comp. Sim.* **15**, 33–50 (1982)
29. Plackett, R.L.: A note on interactions in contingency tables. *J. R. Stat. Soc. B Meth.* **24**, 162–166 (1962)
30. Pomar, M.I.: Demystifying loglinear analysis: Four ways to assess interaction in a  $2 \times 2 \times 2$  table. *Sociol. Persp.* **27**, 111–135 (1984)
31. Quetelet, L.A.J.: *Lettres à S. A. R. le Duc Régnant de Saxe-Cobourg et Gotha, sur la Théorie des Probabilités Appliquée aux Sciences Morales et Politiques*. Hayez, Bruxelles (1846). [English translation, *Letters Addressed to H.R.H. the Grand Duke of Saxe Coburg and Gotha on the Theory of Probabilities as Applied to the Moral and Political Sciences*, by O.G. Downes and published by Charles & Edwin Layton, London, 1849]
32. Rosenthal, R.: Parametric measures of effect size. In: Cooper, H., Hedges, L.V. (eds.) *The Handbook of Research Synthesis*, pp. 231–234. Russell Sage, New York (1994)
33. Simpson, E.H.: The interpretation of interaction in contingency tables. *J. R. Stat. Soc. B Meth.* **13**, 238–241 (1951)
34. Yates, F.: Contingency tables involving small numbers and the  $\chi^2$  test. *Suppl. J. R. Stat. Soc.* **1**, 217–235 (1934)
35. Zachs, S., Solomon, H.: On testing and estimating the interaction between treatments and environmental conditions in binomial experiments: The case of two stations. *Commun. Stat. Theor. M* **5**, 197–223 (1976)
36. Zelterman, D., Chan, I.S., Mielke, P.W.: Exact tests of significance in higher dimensional tables. *Am. Stat.* **49**, 357–361 (1995)