Chapter 37
Phi Tests for Nominal Data

1 General Purpose

Nominal data are the simplest type of data. Unlike ordinal data (Chap. 9) and continuous data (Chaps. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, and 34), they are assumed not to have a stepping function. Examples are genders, age classes, family names. Of nominal data the simplest versions are the bifurcated data (binary data, dichotomous data, yes no data). Chi-square tests can be used for analysis, but they do not provide levels of association, which may be clinically rather relevant. As an example, males and females may be assessed for successful exams.

	success	yes	no
males		50(a)	25(b)
females		50(c)	25(d)

The value of $[(a \times d) - (b \times c)]$ can be used to estimate the level of association. In the above example the level of association is 0. The gender does not give the faintest prediction of the chance of a successful exam.

	success	yes	no
males		50(a)	0(b)
females		0(c)	50(d)

In the above example the level of association equals 1 (100 %). The outcome predicts the chance of a successful exam with 100 % certainty. Phi's, otherwise called Cramer's V's, are used to calculate the precise level of association, being between −1 and 1, and can be easily tested for statistical significance with the help of a chi-square test.

T.J. Cleophas, A.H. Zwinderman, *Clinical Data Analysis on a Pocket Calculator*,
DOI 10.1007/978-3-319-27104-0_37

2 Schematic Overview of Type of Data File

Predictor (binary)	outcome (binary)
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.

3 Primary Scientific Question

In a data sample of binary variables, what is the level of association between a binary predictor and another binary outcome variable.

4 Data Example

In a hospital many patients tend to fall out of bed. We wish to find out whether one department performs better than the other.

	fall out of bed	yes	no
department 1		15(a)	20(b)
department 2		15(c)	5(d)

$$\begin{aligned}\text{phi} &= (ad-bc)/\sqrt{(a+b)(c+d)(a+c)(b+d)} = \\ &= (300–75)/\sqrt{(35 \times 20 \times 30 \times 25)} = \\ &= 0.31\end{aligned}$$

We can predict the chance of falling out of bed from the department by $0.31 = 31\ \%$. Knowing the department, you will be 31 % certain about the chance of falling out of bed. Is this level of association statistically significant? this means it 31 % significantly more than 0 %. This of course depends on the magnitude of the samples.

With very small samples it is hard to obtain statistical significance. The underneath chi-square equation is adequate for statistical testing.

$$\begin{aligned}\text{chi-square} &= \text{phi}^2 \times \text{n}, \text{ where } \text{n} = (\text{a}+\text{b}+\text{c}+\text{d}) \\ &= 0.31^2 \times 55 \\ &= 5.29\end{aligned}$$

The underneath chi-square table shows areas under the curve in the top row, and (df) degrees of freedom in the left-end column, and furthermore plenty of chi-square values. The table shows, that a chi-square value of 5.29 with 1 df (degree of freedom), $((2–1) \times (2–1) = 1)$ is between 3.841 and 6.635. This means that the p-value is between 0.05 and 0.01, and, thus, < 0.05. The association is significantly better than an association of zero, no association at all, at $p < 0.05$.

Chi-squared distribution				
	Two-tailed *P*-value			
df	0.10	0.05	0.01	0.001
1	2.706	3.841	6.635	10.827
2	4.605	5.991	9.210	13.815
3	6.251	7.851	11.345	16.266
4	7.779	9.488	13.277	18.466
5	9.236	11.070	15.086	20.515
6	10.645	12.592	16.812	22.457
7	12.017	14.067	18.475	24.321
8	13.362	15.507	20.090	26.124
9	14.684	16.919	21.666	27.877
10	15.987	18.307	23.209	29.588
11	17.275	19.675	24.725	31.264
12	18.549	21.026	26.217	32.909
13	19.812	22.362	27.688	34.527
14	21.064	23.685	29.141	36.124
15	22.307	24.996	30.578	37.698
16	23.542	26.296	32.000	39.252
17	24.769	27.587	33.409	40.791
18	25.989	28.869	34.805	42.312
19	27.204	30.144	36.191	43.819
20	28.412	31.410	37.566	45.314
21	29.615	32.671	38.932	46.796
22	30.813	33.924	40.289	48.268
23	32.007	35.172	41.638	49.728
24	33.196	36.415	42.980	51.179
25	34.382	37.652	44.314	52.619
26	35.536	38.885	45.642	54.051

(continued)

Chi-squared distribution				
	Two-tailed *P*-value			
df	0.10	0.05	0.01	0.001
27	36.741	40.113	46.963	55.475
28	37.916	41.337	48.278	56.892
29	39.087	42.557	49.588	58.301
30	40.256	43.773	50.892	59.702
40	51.805	55.758	63.691	73.403
50	63.167	67.505	76.154	86.660
60	74.397	79.082	88.379	99.608
70	85.527	90.531	100.43	112.32
80	96.578	101.88	112.33	124.84
90	107.57	113.15	124.12	137.21
100	118.50	124.34	135.81	149.45

5 Conclusion

Nominal data are the simplest type of data. Unlike ordinal data (Chap. 9) and continuous data (Chaps. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, and 34), they have no stepping function. Of nominal data the simplest versions are the bifurcated data (binary data). Chi-square tests can be used for analysis (Chap. 38), but they do not provide levels of association, which may be clinically rather relevant. Phi's, otherwise called Cramer's V's, are used to calculate the precise level of association, and can be additionally tested for statistical significance with the help of a chi-square test.

6 Note

More background, theoretical and mathematical information of binary data is given in Statistics applied to clinical studies 5th edition, Chap. 3, Springer Heidelberg Germany, 2012, from the same authors.