Chapter 47
Hierarchical Loglinear Models for Higher Order Cross-Tabs

1 General Purpose

The Pearson chi-square test is traditionally used for analyzing two dimensional contingency tables, otherwise called cross-tabs or interaction matrices (Chap. 38). It can anwer questions like: is the risk of falling out of bed different between the departments of surgery and internal medicine (Chaps. 37 and 38). The analysis is, however, very limited, because only the interaction between the two variables, e.g., (1) falling out of bed (yes, no) and (2) department (one or the other) is assessed. In contrast, in an observational data set we may be interested in the effects of the two variables separately:

1. is there a significant difference between the numbers of patients falling out of bed and the patients who don't (the main effect of variable 1),
2. is there a difference between the numbers of patients being in one department and those being in the other (the main effect of variable 2).

The Pearson test is unable to answer such questions. Hierarchical loglinear modeling is a pretty novel methodology adequate for the purpose, but not yet widely available. In SPSS versions 16–23 it is not in the menu, but only accessible through syntax commands.

In order to simultaneously analyze, in a 2×2 cross tab, the effects of the main variable in addtion to their interaction, ANOVA (analysis of variance) might be considered. In ANOVA with two predictor factors and one outcome, outcome observations are often modeled as a linear combination of:

1 the grand mean
2 the main effect of the first predictor
3 the main effect of the second predictor
4 the interaction effect of the first and the second predictor.

T.J. Cleophas, A.H. Zwinderman, *Clinical Data Analysis on a Pocket Calculator*,
DOI 10.1007/978-3-319-27104-0_47

However, ANOVA requires continuous outcome variables and contingency tables consist of counted data (numbers of responders, numbers of yes answers), like numbers of patients falling out of bed. With cell-counts data, like interaction matrices, traditional ANOVA is impossible, because the outcome-observations must be modeled as the product of the above 4 effects, rather than their linear add-up sum. The trick is to transform the multiplicative model into a linear model using

logarithmic transformation (ln = natural logarithm is always used).

Outcome = 1*2*3*4 (* = symbol of multiplication)
Log outcome = log 1 + log 2 + log 3 + log 4

2 Schematic Overview of Type of Data File

Treatment modality	Outcome
(1 and 2)	(1 and 2)
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.

3 Primary Scientific Question

Can hierarchical loglinear modeling simultaneously assess the effects of the main variables in addition to their interaction?

4 Data Example

A simple 2×2 contingency table is given with two treatment groups as row variable and the presence of sleeplessness as column variable. A loglinear analysis is given underneath. Loglikelihood ratio tests are used for the computations (see also the Chap. 46).

	column	1	2	
row	1	50	150	200
	2	90	60	150
		140	210	350

All counts have to be logarithmically transformed (ln 50 = 3.912 etc.).

	column	1	2	
row	1	3.912	5.011	5.298
	2	4.500	4.049	5.011
		4.942	5.347	5.848

4.1 *First Order Effects*

Is there a significant main effect of the column variable (is the number of sleepy people significantly different from that of non-sleepy people). Expected log frequencies log(350/2) = 5.165. The loglikelihood ratio (LLR) chi-square test is used for testing.

In this test −2 loglikelihood ratio should be larger than $z^2 = 2^2 = 4$ in order to obtain statistical significance at a $p < 0.05$ level, and z^2 = one-degree-of-freedom chi-square value as explained in the Chap. 46).

$$-2\ LLR_{column} = 2 * (140*(4.942-5.165) + 210*(5.347-5.165))$$
$$= 140.0,$$

* = symbol of multiplication.

The underneath chi-square table has an upper row with areas under the curve (p-values), a left-end column with degrees of freedom (df), and, furthermore, a whole lot of chi-square values.

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
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

A chi-square value of 140, one degree of freedom, means that p is much <0.001.

Is there a significant main effect of the row variable (is the numbers of treatments in group 1 significantly different from that of group 2). Expected log frequencies = log (350/2) = 5.165.

$$-2\ \mathrm{LLR}_{\mathrm{row}} = 2 * (200*(5.298-5.165)+150*(5.011-5.165)) = 7.0,$$

A chi-square value of 7.0 and 1df, means that $p < 0.01$.

4.2 Second Order Effects

Is there a significant interaction between the row and column variable. The loglikelihood ratio (LLR) chi-square test is again used for testing.

$$-2\ \mathrm{LLR}_{\mathrm{column\ x\ row}} = 2 * [(200*(5.298-5.165)+150*(5.011-5.165) +140*(4.942-5.165)+210*(5.347-5.165)] = 21.0,$$

1 df, $p < 0.001$.

The traditional Pearson chi-square test for "row x column" is similarly very significant, although with a larger chi-square value. We will use the pocket calculator method (Chap. 38).

$$\text{Pearson chi-square}_{\mathrm{column\ x\ row}} = [(50*60 - 90*150)^2 * 350]/(140*210*150*200) = 43.75,$$

1 df, $P < 0.0001$.

* = symbol of multiplication.

The Pearson chi-square value is larger than the above second order log likelihood ratio test. This is, because the former does not account first order effects. In general, if you account more, then you will prove less.

5 Conclusion

The above example shows that logarithmic transformation of multiplicative models for analyzing contingency tables can readily provide first and second statistics with the help of log likelihood ratio tests. Also, in practice, higher order contingency tables do exist. E.g, we may want to know, whether variables like ageclass, gender, and other patient characteristics interact with the variables (1) and (2). Calculations are, of course, increasingly complex, and a pocket calculator assessment is impossible. In Chap. 52, SPSS for starters and second levelers 2nd edition, Springer

Heidelberg Germany, 2015, from the same authors, examples of third, and fourth order hierarchical loglinear models are given.

6 Note

More background, theoretical and mathematical information of hierarchical log linear models is given in SPSS for starters and second levelers 2nd edition, Chap. 52, Springer Heidelberg Germany, 2015, from the same authors.