Multiple factor interactions. Role of effect modifier factors.

Lucette Carter (Université Paris 10 – Nanterre)

This paper is focused on the analysis of categorical data and the related contingency table (of two or more dimensions), with emphasis on multiple factor interaction. It seems very important to draw the attention of the students on how the relationship between two or more factors is affected by the introduction of a supplementary factor.

Although the paper is focused on a course on categorical data for 4th year students in psychology, it is suggested that the basic idea of the modifier effect of a factor on the association between other factors can already be introduced in high school statistical cursus (Carter,L. and Piednoir, J.L Statistiquement Vôtre, 2002)

As an introduction to the analysis of multidimensional tables, some simple examples exhibit how a two factor association varies when the level of a third factor is taken into account. A recent study on the relationship between ethnicity (black or white) and depression for women in United States (Gazmararian *et al*, AEP,vol5**,** ***6****, 1995*) shows the role of two « effect modifiers » (socioeconomic and marital status). The odds of depression for black women is twice that of white women, but taking SSE and marital status into account, the authors note the surprising result that the odds of depression for black women remains significantly higher (that the odds for white women) only for the nonpoor and married strata.

The Simpson paradox illustrates an extreme case when the introduction of a third factor modifies not only the intensity of the relationship between two factors (as in the above example) but also the direction of the relationship. The well known example below (Radelet, M., 1981) concerns a study of homicides in United States distributed following the sentence (death or other sentence), the ethnicity of the accused and of the victim.

« Victim : Black »

	Death sentence	Other sentence
Accused : Black	6	97
Accused : White	0	9

« Victim : White »

	Death sentence	Other sentence
Accused : Black	11	52
Accused : White	19	132

In each case (for black and white victim), the death sentence frequency is higher when the accused is black. But when the ethnicity of the victim is not taken into account, the relationship between the two other factors is reversed, the death sentence frequency being higher for black accused. The paradox is easily explained by the fact that death sentence frequency is higher when the victim is white and that there is a strong positive association between the ethnicity of victims and accused. The analysis of odds-ratio (marginal and conditionnal) can be usefully presented at this stage as they give a clear summary of the paradox and its explanation.

Examples in epidemiology and social sciences (issued from students memoirs and from various WEB sources) constitute a natural introduction to the systematic analysis of the structure of contingency tables. The various models based on marginal and conditionnal independance are presented with the corresponding goodness of fit tests. At this stage, the students are encouraged to « guess » which model is adequate by explorating the data (calculating odds-ratio, for example). Emphasis is given on the Bartlett model for which the association between two factors does not vary with the level of a third factor (Christensen, R. *Log-linear models*, 1990).

At this stage, the more systematic approach of the log linear method, following the hierarchical principle, follows. Analyses involve a trade-off between simplicity of the model and increasing the goodness of fit. Practical examples show that the type II error for the adequacy of the Bartlett model can be very high. The students should be encouraged to increase the level of the type I error of the test and to diversify the data analyses before conclusion.

Examples can be presented to illustrate problems that can arise if statistical methods, related to the log-linear methods, are applied without inspecting necessary hypothesis such as the absence of certain multiple effects. For example, the Mantel-Haenszel chi-square statistic produces a test that measures the overall association between the 2 first factors over the levels of a third factor. But this test is often presented (and applied) without verifying the necessary condition of the absence of the 3-way interaction factor in the log-linear model.