

# Association hypotheses based on generalized odds ratios

## *Ipotesi di associazione basate su odds ratio generalizzati*

Manuela Cazzaro

Dipartimento di Metodi Quantitativi per l'Economia  
Università di Milano-Bicocca  
manuela.cazzaro@unimib.it

**Riassunto:** Questo lavoro è dedicato all'analisi delle relazioni che si instaurano tra le ipotesi di associazione uniforme tra le variabili coinvolte in una tabella di contingenza basate sui 16 diversi tipi di odds ratio generalizzati. L'ipotesi considerata è appena meno restrittiva di quella di indipendenza ed usa un solo parametro per descrivere l'associazione. Il contributo originale di questo lavoro, a conoscenza dell'autrice, è il teorema relativo alle implicazioni tra ipotesi di associazione uniforme basate su odds ratio diversi. Con questa presentazione si vuole incentivare l'utilizzo di vincoli simultanei su diversi tipi di odds ratio in quei modelli che gestiscono l'associazione tra variabili categoriali.

**Keywords:** contingency table, odds ratio, uniform association.

## 1. Introduction

In the analysis of the dependence among the categorical variables involved in a contingency table, the *generalized odds ratios* are set in a worth position. They are cell probability cross ratios. The odds ratios (o. r.) are an analytical association index: for each table of  $I \times J$  dimensions, in fact,  $(I - 1)(J - 1)$  generalized o. r. can be obtained. In particular, there exist 16 different kinds of generalized o. r., Douglas *et al.* (1990), that can be defined by partitioning rectangular subarrays of a table into 4 sets of adjacent cells. The most famous of these measures is the *local* odds ratio. Other kind of o. r. can be considered according to the nature or the role of the categorical variables involved in the contingency table: nominal or ordinal, dependent or independent. Odds ratios that make no difference between the role of the variables are called 'symmetric'. Different association hypotheses among the variables can be tested constraining opportunely these o. r. This work focuses its attention on the most parsimonious hypothesis, even if restrictive, of *uniform* association. This hypothesis reduces drastically the number of the parameters that measure the dependence among the variables in a table. In fact, it is assumed that all the  $(I - 1)(J - 1)$  o. r. are equal to the same parameter  $k$ ,  $k \in \mathbb{R}$ . The  $k = 1$  case is the trivial hypothesis of stochastic independence in which, constraining any of the o. r. equal to 1, implies any other o. r. equal to 1. In this work a complete analysis on the uniform association hypothesis based on all the 16 o. r. will be presented.

## 2. Implications among uniform association hypotheses

In this paragraph the existing relationships among the uniform association hypotheses applied to the categorical variables involved in a two way contingency table based on the

---

<sup>0</sup>This research is supported by *Finanziamento Giovani Ricercatori 2001*, Università di Milano-Bicocca.

different o. r. will be described. A generalized odds ratio is defined according to the way the cells of a table are aggregated. In the literature 4 different approaches are known: *local (L)*, *global (G)*, *continuation (C)* and *reverse continuation (R)*. Choosing a row, say  $r$ , and column, say  $c$ , approach 16 different kind of o. r. can be constructed.

Let's introduce some notations that will become useful in order to establish a unified representation for the different generalized odds ratios. Let  $A$  and  $B$  be the row and the column categorical variable of a contingency table and let  $a_i$ ;  $i = 1; \dots; I$  and  $b_j$ ;  $j = 1; \dots; J$  be the number of levels of  $A$  and  $B$  respectively. Let's consider two subsets of levels  $A_i$  and  $\bar{A}_i$  for the variable  $A$  and  $B_j$  and  $\bar{B}_j$  for the variable  $B$ . The joint probability of  $\Pr(A \in A_i; B \in B_j)$  is defined as  $\Pr(A_i B_j)$ . The generalized odds ratios will depend on the cut point  $i$ ;  $i = 1; \dots; I - 1$  and  $j$ ;  $j = 1; \dots; J - 1$  used to select the subsets of the row and the column variable levels respectively. The cut point are fixed according with the row and the column approach chosen to aggregate the cells. A generalized odds ratio  $\mu_{ij}^{rc}$  can be defined as:

$$\mu_{ij}^{rc} = \frac{\Pr(A_{i;r} B_{j;c}) \Pr(\bar{A}_{i;r} \bar{B}_{j;c})}{\Pr(\bar{A}_{i;r} B_{j;c}) \Pr(A_{i;r} \bar{B}_{j;c})}; \quad i = 1; \dots; I - 1 \quad j = 1; \dots; J - 1$$

where  $A_{i;r} = fa_{ig}$ ; for  $r = L; C$  and  $A_{i;r} = fa_{im}$ ;  $m \leq ig$ ; for  $r = G; R$  while  $\bar{A}_{i;r} = fa_{i+1g}$ ; for  $r = L; R$  and  $\bar{A}_{i;r} = fa_{im}$ ;  $m > ig$ ; for  $r = G; C$ . The same definition holds for the  $B_{j;c}$  and  $\bar{B}_{j;c}$  subsets. For example, if  $r = L$  and  $c = L$  the probability  $\Pr(A_{i;r} B_{j;c})$  will indicate the probability of a single cell, if  $r = L$  and  $c = C$  or  $r = R$  and  $c = L$  then  $\Pr(A_{i;r} B_{j;c})$  will describe the probabilities' sum of the cells of a partition of a row or column vector respectively of the table, if  $r = G$  and  $c = G$  the probability  $\Pr(A_{i;r} B_{j;c})$  will indicate the probabilities' sum of a rectangular or square cells' block of the table. The  $(I - 1)(J - 1)$  generalized odds ratios of the same type obtained from a  $I \times J$  table can be expressed as a vector  $\mu^{rc} = (\mu_{ij}^{rc}; i = 1; \dots; I - 1; j = 1; \dots; J - 1)$  where  $r$  and  $c$  are the approaches already described. The name of each of the 16 generalized odds ratios will be formed of two parts representing the  $r$  and  $c$  used approaches. If  $r = c$  it is adopted the name just one time.

Let's define the hypothesis of uniform association based on a generalized odds ratio as  $U(k; \mu^{rc})$  where  $\mu^{rc}$  is the chosen odds ratio. In this way, for example, the property 'the local-global odds ratios satisfy the uniform association hypothesis' (all the local-global o. r. are equal to a constant  $k$ ), will be equivalent to the notation  $U(k; \mu^{L-G})$ . Let's introduce a useful definition concerning the theorem that will establish the relationships:

**Definition:** Two odds ratios are 'conformable' if: (i) they have at least a common approach between the row and column approaches, (ii) the couple of non common approaches is not of Continuation and Reverse Continuation kind. For the transitive property this definition could be extended to more than two generalized odds ratios.

**Example:** any couple of o. r. chosen among the local, the continuation-local and the global-local o. r. are conformable, like any of two o. r. among the local-reverse continuation, reverse continuation and global-reverse continuation o. r.. On the contrary, any of two o. r. among the local, continuation and reverse continuation are not conformable.

According to this definition, 16 different sets of three o. r., containing different o. r., with a common approach could be constructed. The sets of three conformable o. r. are:

$$\begin{aligned} & \textcircled{\mu^L; \mu^{L-C}; \mu^{L-G}}; \textcircled{\mu^L; \mu^{L-R}; \mu^{L-G}}; \textcircled{\mu^L; \mu^{C-L}; \mu^{G-L}}; \textcircled{\mu^L; \mu^{R-L}; \mu^{G-L}}; \\ & \textcircled{\mu^{G-L}; \mu^{G-C}; \mu^G}; \textcircled{\mu^{G-L}; \mu^{G-R}; \mu^G}; \textcircled{\mu^{L-G}; \mu^{C-G}; \mu^G}; \textcircled{\mu^{L-G}; \mu^{R-G}; \mu^G}; \\ & \textcircled{\mu^{C-L}; \mu^C; \mu^{C-G}}; \textcircled{\mu^{C-L}; \mu^{C-R}; \mu^{C-G}}; \textcircled{\mu^{L-C}; \mu^C; \mu^{G-C}}; \textcircled{\mu^{L-C}; \mu^{R-C}; \mu^{G-C}}; \end{aligned}$$

$$^{\odot} \mu^{R-L}; \mu^{R-C}; \mu^{R-G}{}^a; \quad ^{\odot} \mu^{R-L}; \mu^R; \mu^{R-G}{}^a; \quad ^{\odot} \mu^{L-R}; \mu^{C-R}; \mu^{G-R}{}^a; \quad ^{\odot} \mu^{L-R}; \mu^R; \mu^{G-R}{}^a;$$

It is important to remember that, following the literature, the uniform association hypothesis based on one particular odds ratio does not imply any restriction (as uniform association) on other kind of odds ratios. Using the definition just mentioned it is now possible to fix some 'hierarchies' among different uniform association hypotheses based on different odds ratios. Let's now focus on the following theorem:

**Theorem:** For each set of three conformable odds ratios, if any of two odds ratios satisfy the uniform association hypothesis with the same constant  $k$ , then also the remaining odds ratio satisfies the uniform association hypothesis with the same  $k$  ( $k \neq 1$ ).

**Proof.** A complete proof will require the analysis of each of the 16 relationships highlighted in the theorem. Only three particular cases will be presented as outlines of all the implications.

$$\text{a) } U(k; \mu^L) \setminus U(k; \mu^{C-L}) \Rightarrow U(k; \mu^{G-L}):$$

The following drawing represents the 'relevant' structure of cells and aggregation of cells in the table for the calculation of the three o. r. considered. In particular a; b; c; d represent single cells, on the contrary e and f some partitions of a vertical 'vector' of the  $I \times J$  table of which the  $j$  and  $j + 1$  columns are considered. Let's associate to each cell or set of cells the probability identified with the corresponding capital letter: the a cell will have A probability, the set of cells f will have probability equal to F; obtained as sum of the single probability of the cells contained in the f vector, etc.

|   |   |
|---|---|
| a | b |
| d | c |
| f | e |
| : | : |

To prove the relationship of point a), it is necessary to demonstrate that:

$$\begin{aligned} AC &= kBD \\ A(C + E) &= kB(D + F) \quad ; \Rightarrow (A + D)E = kF(B + C); \text{ but from the last two} \\ DE &= kFC \end{aligned}$$

hypotheses it follows  $AC + AE + DE = kBD + kBF + kFC$  and  $AC + E(A + D) = kF(B + C) + kBD$ : Now if  $AC = kBD$ , then it follows that  $E(A + D) = kF(B + C)$ : Naturally the relationship of point a) applies also to:  $U(k; \mu^L) \setminus U(k; \mu^{G-L}) \Rightarrow U(k; \mu^{C-L})$  and  $U(k; \mu^{G-L}) \setminus U(k; \mu^{C-L}) \Rightarrow U(k; \mu^L)$ ; following the same structure of the proof. Let's consider now another set of implications:

$$\text{b) } U(k; \mu^{R-L}) \setminus U(k; \mu^{R-C}) \Rightarrow U(k; \mu^{R-G}):$$

In this case the relevant structure of the table is composed from the generic  $j \geq 1, j$  columns and from the set of columns from the  $j + 1$  to the  $J$ . The letters and the structure of the drawing respect the fixed conventions:

|   |   |     |
|---|---|-----|
| a | b | c : |
|   |   | :   |
| f | e | d : |

To prove the relationship of point b), it is necessary to demonstrate that:

$$\begin{aligned} AE &= kFB \\ A(E + D) &= kF(B + C) \quad ; \Rightarrow (A + B)D = kC(F + E); \text{ but from the last two} \\ BD &= kEC \end{aligned}$$

hypotheses it follows  $AE + AD + BD = kFB + kFC + kEC$  and  $AE + D(A + B) = kFB + kC(F + E)$ : Now if  $AE = kFB$ , then it follows that  $(A + B)D = kC(F + E)$ : Naturally the relationship of point b) applies also to the other two couples of o. r. toward the third one, following the same structure of the proof.

Let's consider the last series of implications:

$$c) U(k; \mu^{L-G}) \setminus U(k; \mu^{C-G}) = U(k; \mu^G):$$

To prove this relationship it is necessary to refer to a more complicated table structure but that consider again the generic  $j \geq 1$ ,  $j$  columns and the set of columns from the  $j + 1$  to  $J$ : The usual meaning of the letters is still valid:

|   |   |     |
|---|---|-----|
| a | b | c : |
| f | e | d : |
| i | h | g : |
|   |   | :   |

To prove the relationship of point c), it is necessary to demonstrate that:

$$\begin{aligned} (A + B) D &= kC (F + E) &= & (A + B + F + E) G = \\ (A + B) (D + G) &= kC (F + E + I + H) &= & (A + B + F + E) G = \\ (F + E) G &= kD (I + H) &= & k(I + H) (C + D); \end{aligned}$$

but from the last two hypotheses it follows  $(A + B) D + (A + B) G + (F + E) G = kC (I + H) + kC (F + E) + kD (I + H)$  and  $(A + B) D + G (A + B + F + E) = k (C + D) (I + H) + kC (F + E)$ : Now if  $(A + B) D = kC (F + E)$ , then it follows that  $(A + B + F + E) G = k (I + H) (C + D)$ :

Also in this context the proof expresses for the point c) can be applied to the other two couples of o. r. toward the third one. The structure of the a) case can be applied to all the sets of o. r. that have the Local approach in common, the proofs relating to the b) case hold for all the sets of o. r. that have the Continuation or the Reverse Continuation approach in common, finally the proofs presented for the c) case have the same structure for all the sets of o. r. that involve the Global approach. ¥

### 3. Conclusions

One of the tasks of this work is to highlight the intrinsic relationships among the uniform association hypotheses based on different odds ratios as it was already proposed in the literature, Yanagimoto (1972), Douglas *et al.* (1990), about the positive association hypotheses in ordinal contingency tables based on the generalized odds ratios posed greater or equal to one. In this work, the relationships among different generalized odds ratios fixed all equal to a constant  $k$  have been verified. Therefore it has been highlighted the meaning of simultaneous uniform constraints on couple of odds ratios and as these constraints imply the satisfaction of the uniform association hypothesis based on a third odds ratio. Next step in the research will be toward those models for the analysis of contingency tables that constrain more than one kind of odds ratio simultaneously, as this approach is not deeply treated in literature yet. A further extension of this research would investigate the possibility of constraining also non conformable odds ratios.

### References

- Douglas R., Fienberg S.E., Lee M.T., Sampson A.R. and Whitaker L.R. (1990) *Positive dependence concepts for ordinal contingency tables, Lecture notes-monograph series: Topics in statistical dependence*, Institute of Mathem. Statistics, Hayward.
- Yanagimoto T. (1972) Families of positively dependent random variables, *Annals of the Institute of Statistical Mathematics*, 24, 559–573.