Xie, Yu. 1991. "Model Fertility Schedules Revisited: The Log-Multiplicative Model Approach." *Social Science Research* 20:355–68.
———. 1992. "The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables." *American Sociological Review* 57:380–95.
———. 1994. "Log-Multiplicative Models for Discrete-Time, Discrete-Covariate Event History Data." Pp. 301–40 in *Sociological Methodology*, edited by Peter V. Marsden. Cambridge, MA: Blackwell Publishers.
Xie, Yu, and Ellen Efron Pimentel. 1992. "Age Patterns of Martial Fertility: Revisiting the Coale-Trussell Method." *Journal of the American Statistical Association* 87: 977–84.
Yamaguchi, Kazuo. 1987. "Models for Comparing Mobility Tables: Toward Parsimony and Substance." *American Sociological Review* 52:482–94.

COMMENT: THE ESSENTIAL TENSION BETWEEN PARSIMONY AND ACCURACY

*Yu Xie**

Goodman and Hout present a new set of powerful statistical methods for comparing two-way associations across a third dimension. Providing a unifying framework that encompasses earlier attempts proposed by Yamaguchi (1987) and Xie (1992), this paper is highly significant and as such will have a long-term impact on the way multidimensional contingency tables are analyzed in the future.

The main message of Goodman and Hout's article is an important one and is worth reiterating here. For a three-way cross-tabulation of row (R), column (C), and layer (L), indexed respectively by subscripts i (1, ... I), j (1, ... J), and k (1, ... K), the expected frequency under the saturated model can be written as

$$F_{ijk} = \tau_0 \tau_i^R \tau_j^C \tau_k^L \tau_{ik}^{RL} \tau_{jk}^{CL} \tau_{ij}^{RC} \tau_{ijk}^{RCL}, \qquad (1)$$

where the τ parameters are multiplicative effects subject to usual normalization constraints. Terms τ_0, τ_i^R, τ_j^C, and τ_k^L are commonly called "marginal" or "main" effects, τ_{ik}^{RL}, τ_{jk}^{CL}, and τ_{ij}^{RC} "two-way interactions," and τ_{ijk}^{RCL} "three-way interactions." Goodman and Hout's key idea is to constrain equation (1) to the following form (their equation 6):

$$F_{ijk} = \tau_0 \tau_i^R \tau_j^C \tau_k^L \tau_{ik}^{RL} \tau_{jk}^{CL} \tau_{ij}^{RC} \exp(\psi_{ij} \phi_k). \qquad (2)$$

That is, Goodman and Hout add constraints to the three-way interactions among R, C, and L. Whereas the saturated model (equation 1) uses up $(I-1)(J-1)(K-1)$ degrees of freedom for three-way interactions, the Goodman-Hout model (equation 2) consumes only $(I-1)(J-1)+(K-2)$ degrees of freedom. These implied constraints are noteworthy for two reasons: (1) they may yield far more parsimonious models than the saturated model, with relatively little deterioration in goodness-of-fit; and (2) they afford researchers an easy method to systematically examine the variation in the $R \times C$ two-way association across the third dimension L.

*University of Michigan

Much of the idea can be traced to Goodman's earlier work (notably Goodman 1986), although the Goodman and Hout paper can be viewed as directly responding to Xie (1992).[1] Xie's model is of the form

$$F_{ijk} = \tau_0 \tau_i^R \tau_j^C \tau_k^L \tau_{ik}^{RL} \tau_{jk}^{CL} \exp(\psi_{ij} \phi_k). \tag{3}$$

Note that, without the last term $\exp(\psi_{ij}\phi_k)$, the other terms τ_0, τ_i^R, τ_j^C, τ_k^L, τ_{ik}^{RL}, and τ_{jk}^{CL} constitute the null hypothesis of conditional independence of R and C given L. Compared with the saturated model (equation 1), the Goodman-Hout model (equation 2) constrains τ_{ijk}^{RCL} to be $\exp(\psi_{ij}\phi_k)$, whereas the Xie model constrains $\tau_{ij}^{RC}\tau_{ijk}^{RCL}$ to be $\exp(\psi_{ij}\phi_k)$. Several commonalities emerge between the Goodman-Hout model and the Xie model. First, for most applications, both begin with the conditional independence model as the baseline null model.[2] Second, both are more parsimonious than the saturated model. Third, both are designed to describe the variation in the $R \times C$ two-way association over the dimension of L—i.e., assuming three-way interactions. Finally, the specification of the variation in the $R \times C$ two-way association (i.e., three-way interactions) is identical in both models.

The last point of comparison merits some elaboration. Let $\theta_{ij|k}$ denote the odds ratio for four adjacent cells between R and C $(i, j; i+1, j; i, j+1; i+1, j+1)$, given L (i.e., equation 2 in Goodman and Hout). Given their model specification (equation 2), Goodman and Hout show that

$$\ln \theta_{ij|k} = \mu_{ij} + \mu'_{ij}\phi_k, \tag{4}$$

where μ_{ij} is the logged cross-product of τ_{ij}^{RC}, and μ'_{ij} is the logged cross-product of $\exp(\psi_{ij})$:

$$\mu_{ij} = \ln \tau_{ij}^{RC} - \ln \tau_{(i+1)j}^{RC} - \ln \tau_{i(j+1)}^{RC} + \ln \tau_{(i+1)(j+1)}^{RC},$$

$$\mu'_{ij} = \psi_{ij} - \psi_{(i+1)j} - \psi_{i(j+1)} + \psi_{(i+1)(j+1)}.$$

Equation (4) is analogous to a regression with an intercept term (μ_{ij}) and a slope term (μ'_{ij}), hence the phrase "regression-type approach." The μ_{ij} term, consisting of τ_{ij}^{RC} parameters, establishes the baseline pattern of the $R \times C$ two-way association, whereas the μ'_{ij} term establishes the typical pattern of *deviation* in the $R \times C$ two-way association from the baseline association. For the kth layer, the actual deviation is the product of the pattern of deviation (μ'_{ij}) and the strength of deviation (ϕ_k). As shown in Goodman and Hout (their equation 9), the contrasts in log-odds ratios between layers k and k' are given by

[1]See Xie (1992:382) for acknowledgment of Goodman (1986).

[2]The regression-type model could also start with the null three-way association model (i.e., $\exp(\psi_{ij}\phi_k) = 1$ in equation 2), which corresponds to setting ϕ_k to any constant in the multiplicative approach (equation 3) and thus alters the multiplicative specification.

$$\ln \theta_{ij|k} - \ln \theta_{ij|k'} = \mu'_{ij}(\phi_k - \phi_{k'}). \tag{5}$$

Being a special case of the model proposed by Goodman and Hout, Xie's model changes equation (4) to

$$\ln \theta_{ij|k} = \mu'_{ij}\phi_k. \tag{6}$$

It is easy to see that equation (5) also holds true for Xie's model. Thus, as far as cross-layer comparison in the $R \times C$ two-way association is concerned, Goodman and Hout's regression-type model has the same properties as Xie's multiplicative model. The real difference between the two is that the multiplicative model does not contain the term describing the baseline association between R and C—i.e., τ_{ij}^{RC}. This difference was anticipated by Xie (1992:392) with the statement that "[one possible extension of the multiplicative model is] to decompose two-way association parameters into two parts: (a) those that do not vary across tables and (b) those that vary across tables."

Given the tradeoff between the greater parsimony of Xie's multiplicative approach and the lesser restrictiveness of Goodman and Hout's regression-type approach, which approach is preferable will depend on substantive applications. Empirical as well as conceptual considerations should be taken into account. Empirically, the researcher is always faced with the essential tension between parsimony on the one hand and accuracy on the other hand. By parsimony I mean models with few parameters. By accuracy I mean the ability to reproduce observed data, measured by goodness-of-fit statistics. Not allowing for a free baseline pattern, the multiplicative model is more parsimonious but describes observed data with less accuracy than the regression-type model. Whether improvement in goodness-of-fit as a result of allowing for additional parameters in the regression-type model costs too many extra degrees of freedom is an empirical question and should be addressed empirically. This difficulty is well illustrated in the three-nation mobility example reanalyzed by Goodman and Hout, where the multiplicative model (Model M4$^{\times}$) appears superior to the regression-type model (Model M4*) according to the BIC criterion before further constraints were applied.

It is also a conceptual issue whether a baseline association pattern should be established apart from layers. When the intercept is present in the regression-type model, the basic structure of log-odds ratios is contributed by two $R \times C$ pattern matrices (τ_{ij}^{RC} and ψ_{ij}). This presents a challenge to the researcher who wants to impose a common structure—e.g., uniform association—on the conditional two-way tables at all layers. For example, to test the hypothesis that uniform association applies to each layer k, both the τ_{ij}^{RC} terms and the ψ_{ij} terms must be constrained to the uniform association pattern. Goodman and Hout explicitly discuss different specifications for τ_{ij}^{RC} and ψ_{ij}, both across models and within a single model. For the case of Goodman's (1979) association model II (which was renamed as the *RC* model by Goodman (1981) and is now commonly referred to by that name), I make a link to the *RC(m)* model extensively discussed in Goodman (1986) and Becker and Clogg (1989) and briefly referred to in Xie (1992:392). To see this, let us parameterize both τ_{ij}^{RC} and ψ_{ij} according to the *RC* model,

$$\ln \tau_{ij}^{RC} = \lambda_i^R \lambda_j^C,$$

$$\psi_{ij} = \gamma_i^R \gamma_j^C, \qquad (7)$$

where λ_i^R and λ_j^C denote respectively row and column scores (to be estimated) for the baseline pattern, and γ_i^R and γ_j^C denote respectively row and column scores (to be estimated) for the deviation pattern. Given these constraints in (7), it follows that equation (4) can be rewritten as

$$\ln \theta_{ij|k} = [(\lambda_i^R - \lambda_{i-1}^R)(\lambda_j^C - \lambda_{j-1}^C)] + [(\gamma_i^R - \gamma_{i-1}^R)(\gamma_j^C - \gamma_{j-1}^C)]\phi_k.$$

At each layer, the basic log-odds ratio structure can be seen as a two-dimensional association model $RC(2)$, with one dimension invariant across layers and another dimension lowered or strengthened by a multiplier across layers.

In my multiplicative model, parallel structures across layers are guaranteed, since this feature is part of the model. At each layer, the log-odds-ratio structure is the same ($\mu'_{ij}\phi_k$), decomposable into a pattern component (μ'_{ij}, function of i and j,) and a level component (ϕ_k, function of k). In this case, the pattern of deviation in the $R \times C$ association across layers is the same as the pattern of the $R \times C$ association at each layer. While this constraint can be unduly restrictive in some applications, the multiplicative model may be preferred in comparative mobility research, due to its close ties to the long-standing theoretical concern with "common fluidity" in the literature. Generalizing the useful concept of common fluidity by assuming a common pattern with possibly varying strengths, the multiplicative model introduces some deviation from commonality. A generalization of the multiplicative model, the regression-type model introduces further deviation from commonality, as shown in Goodman and Hout's analysis of the three-nation data. In brief, the two specifications represent two different conceptualizations of the cross-national differences in the intergenerational mobility process.

If the column (dependent) variable is dichotomous, the intercept (μ_{ij}) and slope (μ'_{ij}) terms are only functions of i. In this case, the regression-type model can be very powerful. Xie's (1994) paper on event-history models is based on this idea. The payoff is especially high when there is a *natural* reference group (or a theoretical concept) for the intercept component. In studies of age patterns of fertility, for example, the intercept can be typical age patterns of natural fertility. The slope is a typical age pattern of fertility limitation. The combination is tantamount to the so-called Coale-Trussell model (Coale and Trussell 1974; Xie 1991). Xie and Pimentel (1992) illustrate an application of a model similar to that proposed by Goodman and Hout, except importing estimates of a reference age pattern, that of natural fertility, from an external source. In the fertility example, there is a strong theoretical prior for equating the baseline age pattern to that of natural fertility, resulting in an easy interpretation for the strength parameter (m, equivalent to ϕ in equation 2).

To conclude, I wish to emphasize that the statistical framework proposed by Goodman and Hout is very powerful; as with other useful statistical frameworks, fruitful application of it requires attentiveness to theoretical concerns, flexibility in implementation, and patient searches for parsimonious and meaningful specifications.

REFERENCES

Becker, Mark P., and Clifford C. Clogg. 1989. "Analysis of Sets of Two-way Contingency Tables Using Association Models." *Journal of the American Statistical Association* 84:142–51.

Coale, Ansley J., and T. James Trussell. 1974. "Model Fertility Schedules: Variations in the Age Structure of Childbearing in Human Populations." *Population Index* 40:185–258.

Goodman, Leo A. 1979. "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories." *Journal of the American Statistical Association* 74:537–52.

———. 1981. "Association Models and Canonical Correlation in the Analysis of Cross-Classifications Having Ordered Categories." *Journal of the American Statistical Association* 76:320–34.

———. 1986. "Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-Linear Models Approach in the Analysis of Contingency Tables." *International Statistical Review* 54:243–309.

Xie, Yu. 1991. "Model Fertility Schedules Revisited: The Log-Multiplicative Model Approach." Social Science Research 20:355–68.

———. 1992. "The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables." *American Sociological Review* 57:380–95.

———. 1994. "Log-Multiplicative Models for Discrete-Time, Discrete-Covariate Event History Data." Pp. 301–40 in *Sociological Methodology 1994*, edited by Peter V. Marsden. Cambridge, MA: Blackwell Publishers.

Xie, Yu, and Ellen Efron Pimentel. 1992. "Age Patterns of Marital Fertility: Revising the Coale-Trussell Method." *Journal of the American Statistical Association* 87:977–84.

Yamaguchi, Kazuo. 1987. "Models for Comparing Mobility Tables: Toward Parsimony and Substance." *American Sociological Review* 52:482–94.