An Alternative Purging Method: Controlling the Composition-Dependent Interaction in an Analysis of Rates

Yu Xie
Population Studies Center, University of Michigan, 1225 South University Avenue, Ann Arbor, Michigan 48104

The purging method controlling for the composition-group interaction developed by Clogg and his associates has proven useful in demographic research. This article proposes an alternative method, partial CD purging, that controls the interaction between composition and the dependent variable. The purged rates from this new method are invariant to changes in the marginal distribution of composition, but those from the earlier purging method are not. Mathematical relationships between the proposed method and other techniques are also explored.

An important step has been taken by Clogg and his associates (Clogg, 1978; Clogg and Eliason, 1988; Clogg and Shockey, 1985; Clogg, Shockey, and Eliason, 1987) in a series of papers centered around the idea of purging the confounding effects of composition in crude rates. The purging method espoused in these papers provides a convenient way to bridge the gap between statistical models and descriptive rates (for a review, see Hoem, 1987). Communication between demographers versed in statistical modeling and other researchers is thus facilitated.

There are several variations on the purging method. Their essential feature is the elimination of either the partial or the marginal composition-group interaction, when the three-factor interaction is absent, in a cross-classification of counts by composition, group, and dependent variables. When the three-factor interaction is present, purging is still possible by concurrently adjusting for the three-factor interaction.

In this article I propose an alternative purging method that controls the partial interaction between composition and the dependent variable. The rates resulting from this alternative purging method differ in general from the purged rates of either partial or marginal composition-group purging. For some applications at least, the former have desirable properties.

What to Purge, CG or CD?

Following Clogg (1978), Clogg and Eliason (1988), and Clogg, Shockey, and Eliason (1987), let C, G, and D denote the composition, group, and dependent variables, in a cross-classified table of $C \times G \times D$.\textsuperscript{1} The categories of C, G, and D are indexed as $i (i = 1, \ldots , I)$, $j (j = 1, \ldots , J)$, and $k (k = 1, \ldots , K)$, respectively. The purging method proposed by Clogg and his associates eliminates the confounding effects of the partial or marginal CG interaction as well as of the three-factor CGD interaction when it is present. An alternative method is to purge the partial CD interaction and the three-factor CGD interaction. This alternative is reasonable because the relationship between G and D cannot be confounded by C once the CD and CGD interaction terms are purged.

Copyright © 1989 Population Association of America
It is well known that for the C variable to be confounding, both the CG and CD interaction terms must be present. Using linear regression as an analogy, bias due to an omitted variable is present only when the omitted variable both is correlated with the independent variable of interest and affects the dependent variable. For a three-way contingency table of $C \times G \times D$, the table is collapsible over C if either CG and CGD interactions or CD and CGD interactions are nil (Bishop, Fienberg, and Holland, 1975:39). When the table is collapsible in the dimension of C, information about C is irrelevant. Hence the GD association is not confounded by the C variable. Thus partial CD purging is an acceptable alternative method of purging, in an analogous way to partial CG purging. Rates resulting from the two methods, however, can be very different. As I will show, the partial CD purged rates are invariant to changes in the marginal distribution of C, but the partial CG purged rates are not. This property makes partial CD purging preferable to CG purging at least for some applications.

Borrowing Clogg and Eliason's (1988) notation, a cell frequency $F_{ijk}$ can be described by the following multiplicative model:

\[
F_{ijk} = \tau \tau_i^C \tau_j^G \tau_k^D \tau_{ik}^{CD} \tau_{jk}^{GD} \tau_{ik}^{CGD},
\]

where the $\tau$ parameters are subject to the usual normalizations (Clagg, 1978). For simplicity, I only consider the case with no three-factor interaction, where $\tau^{CGD} = 1$. The partial CG purging method first obtains purged frequencies by dividing cell frequencies by $\tau^{CG}$ ($F_{ijk}/\tau^{CG}$) and then calculates rates based on the purged frequencies. Similarly, the CD purging method divides observed frequencies by $\tau^{CD}$ to obtain purged frequencies ($F_{ijk}/\tau^{CD}$) and then calculates rates from the purged frequencies.

In the following example, I will show that adjusted rates from the CD purging method are invariant to changes in the marginal distribution of composition, whereas those from the CG purging method are not. Invariance to the marginal distribution of composition can be desirable for a number of reasons. First, the invariance means that composition is statistically controlled as an exogenous variable. Everything else being equal, the CD purged rates, like structural parameters, do not vary as composition changes. This is especially useful in comparative (or trend) analysis (e.g., see Hauser and Grusky, 1988). Second, the invariance is already an old friend in the familiar framework of logit analysis. As I will show, there is a close relationship between logit analysis and partial CD purging. Third, for samples that are stratified on variable C, the marginal distribution of C in the sample is generally different from that in the population, depending on the particular sampling design. It is preferable to have measures that are not affected by sampling design.

**An Example With Hypothetical Data**

I demonstrate the invariance property of the CD purging method by using the hypothetical data shown in Table 1. Case 1 is a simple $2 \times 2 \times 2$ three-way table with no three-factor interaction. In case 2, I alter the table by doubling the frequencies of the cells that are related to $C_b$ thus changing the marginal distribution of variable C. The changes in the parameters are displayed in Table 2.

As shown in Table 2, cases 1 and 2 are identical except for the main effect, $\tau$, and the marginal effect, $\tau^C$. The other two marginal parameters and all of the two-way association parameters are identical. Case 3 was constructed in such a way that $\tau^C$ is the same as in case 1 while $\tau^G$ and $\tau^{CG}$ are different. I apply both the partial CG purging and the partial CD purging methods to each of the three cases and obtain the purged rates reported in Table 3.

As expected, the crude rates differ across cases 1, 2, and 3. What is especially interesting is that the CG purged rates in case 2 differ from those in case 1, whereas the CD purged rates are identical in the two cases. Recall that the only difference between the two cases is