Natalie A. Sheridan

From:	Document Delivery - Olin Library <okudd@cornell.edu>
Sent:	Wednesday, June 1, 2022 2:30 PM
To:	Mathematics Library
Cc:	Natalie A. Sheridan
Subject:	Document Delivery Request from Olin, TN1690723

ScanIt Request for - Patron: John Abowd
TN: 1690723

ScanIt
Olin/Kroch/Uris Libraries

The following document has been requested through ScanIt.

Location: Mathematics Library (Malott Hall)
Call #: QA276.A1 A51
Title: Proceedings of the Statistical Computing Section.
Vol#: 1976
Issue#: OCLC Number/Unique Identifier:
Date: 1976

Article: Statistical disclosure in publication hierarchies
Author: L. Cox
Pages: Approximately 130-136

1.) Please scan this document according to the scanning procedures. Save the PDF file as 1690723.pdf

2.) Reply to this message. Please indicate the number of pages scanned: ___

3.) If you are unable to fill this request, please check one of the reasons below when replying:

___In use
___In process
___Lost
___Non-circulating
___Not owned
___On order
___At bindery
___Lack volume or issue
___Not on shelf
___On reserve
___Poor condition
___Not found as cited
___Other (please explain)

Thank you!

STATISTICAL DISCLOSURE IN PUBLICATION HIERARCHIES

Lawrence H. Cox, U.S. Bureau of the Census

THE HIERARCHY

The term publication hierarchy will denote a collection of aggregate statistical cells compiled from a source of basic data records. In the computation of the cells, certain fields in the data records will be parameter fields whose values determine the manner in which the set of data records is to be partitioned into subsets, each subset defining the extent of an aggregate cell, while other fields will be data fields over which prescribed computations are performed on each subset so defined to produce the value of the cell. Each field in the record format may be a parameter field, a data field or an ignored field, depending upon the statistic being tabulated. For example, the field "number of employees" would be a parameter field if the statistic "total payroll" were being tabulated by state, county and county-part for various size classifications of manufacturing establishments in a given industry, size defined according to total employment; it would appear as a data field if the statistic "total number of employees" were tabulated strictly according to geographic parameters; and it would be ignored if "total sales" were being tabulated strictly geographically. Without loss of generality, we henceforth assume that both the parameter and the data fields are fixed.

The publication hierarchy will in general not be strictly hierarchical in that some hierarchy cells may be directly disaggregated in more than one unique way (a disaggregation of a hierarchy cell is direct if the aggregate cell is the only non-empty hierarchy cell which can be formed as a union of a combination of its disaggregates). The geographic hierarchy in the preceding illustration is strictly hierarchical. Statistical disclosure analysis in strict hierarchies is quite straightforward and will emerge as a special application of the techniques described in this paper.

STATISTICAL DISCLOSURE ANALYSIS AND CELL SUPPRESSION

The goal of statistical disclosure analysis in a publication hierarchy is to provide confidentiality to the respondents and mask the values of all sensitive publication cells, referred to as "protecting" the sensitive cells. The criteria determining whether a cell is sensitive or not will depend on the application, but, in general, the concepts of threshold and dominance are fundamental to the definition. For frequency count data, a cell which contains fewer than a prescribed threshold value may be considered as sensitive, since the number of respondents represented by this cell is so small that the probability of one respondent's being identified and further inferences about its attributes made is not statistically negligible. This rule applies equally well to quantitative data, such as total sales, for, if the number of respondents is small, each may subtract its contribution from the cell value to obtain a better estimate of the contributions of the others. Disclosure is more subtle in quantitative data, however, as even though the number of respondents may be high, if the total contribution of a small subset of the respondents dominates the cell value, so that the total contribution of the remaining respondents is a small fraction of the cell value, then knowledge of the identities of the respondents in the cell may lead to the disclosure of information about some of the cell respondents by other respondents or other knowledgeable parties. For quantitative data, sensitive cells may be defined in terms of cell dominance by the "n-respondent, k%" rule which states that any cell in which n or fewer respondents (for relatively small n) contribute a total of k% or greater to the cell value is sensitive. Threshold rules are, therefore, dominance rules with k = 100.

Once an unambiguous definition of sensitive cell has been accepted, the key concepts in statistical disclosure analysis are estimation and protection. A sensitive cell must be protected so that only acceptable estimates of its value may be inferred through computation in the hierarchy. For each sensitive cell X, two numbers, $L(X)$ and $U(X)$, are, on the basis of statistical and subject matter considerations, determined, from which an acceptable estimate of the value $V(X)$ of the sensitive cell X is defined to be an estimate which does not penetrate the interval $L(X) \leq V(X) \leq U(X)$. $L(X)$ and $U(X)$ are called the bounds of equivocation of X. If X is a frequency count and n is the threshold below which $V(X)$ lies, then the values $L(X) = 0$ and $U(X) = n+1$ are generally chosen. If X is an aggregate of quantitative data and the "n-respondent, k% rule of dominance is applied, then typically $L(X) = pV(X)$ and $U(X) = (1+q)V(X)$, for p and q functions of k and the dominant portion $D(X)$ of $V(X)$, $0 \leq p$, q and $p \leq 1$. As the hierarchy becomes more complex, the number of estimates of the value of each sensitive cell increases in both number and in subtlety. Statistical disclosure analysis in a publication hierarchy requires a cell protection mechanism which operates in concert with techniques which generate a maximum of structural information about the hierarchy.

Foremost among protection techniques in disclosure analysis are random rounding, random perturbations, rolling-up and cell suppression. In random rounding, the cell value is rounded up or down to some predetermined base randomly according to its residue modulo the base, so that if the base is 5 and the cell value is 13, then the cell value will be rounded to 10 with probability 0.4 and will be rounded to 15 with probability 0.6 depending upon the outcome of a random event (such as choosing an integer from the uniform distribution of the integers from 1 to 10 and rounding down if the number if less than 5 and up otherwise). Random perturbation is a similar mechanism, whereby a random range r is chosen and an integer chosen from the uniform distribution of the integers between -r and r is added to the cell value. Both techniques have the property that the expected value of aggregates of rounded or perturbed cells is the aggregate value of the original cells. Random rounding and random perturbations are generally applicable only to frequency count data due to their limited ability to mask the values of large cells. Rolling-up is the substitution of an aggregate for a set of its disaggregates, such as publishing the total value for two states in lieu of publishing either. Rolling-up may be viewed as a specialized of cell suppression, which is the deletion from publication of all sensitive cells togther with the fewest additional cells possible to maintain the required level of protection for all sensitive cells. Cell suppression requires a sound mechanism forperforming complementary disclosure analysis in a single table (the suppression of additional cells in the table to protect sensitive cells), as well as a means of representing and controlling all linear relationships involving the values of sensitive cells in the hierarchy. The remainder of this article will deal with the problems of disclosure analysis in a hierarchy assuming cell suppression as the protection mechanism.[1]

LINEAR ESTIMATION IN A HIERARCHY

In the simplest case, namely when the hierarchy is trivial (i.e., there are no linear relationships between the cell values), to provide adequate protection to all sensitive cells it suffices to suppress only those sensitive cells. In the next most complex case, namely in a strictly hierarchical situation, where X is sensitive and appears only as direct disaggregate of T, $X+X'+X''+X''' = T$, it suffices to suppress X and either T or a combination of X', X'', X''' such that the aggregate value of all suppressed cells is greater than or equal to U(X). In general, the latter approach is preferred, so that an aggregate would be given publication priority over its disaggregates in a publish-suppress decision. If, in addition, X is disaggregated, $X_1+X_2+\ldots X_k = X$, than it suffices to suppress a combination of the X_j such that the aggregate value of the published (unsuppressed) cells is less than or equal to L(X).

In a statistical hierarchy of aggregates, the cell values are related by a system of linear equations representing all cell disaggregations in the hierarchy. The relationships between the cells in the hierarchy may be represented by a lattice, in which a downward line from an upper cell to a lower cell indicates that the lower cell is a direct disaggregate of the upper cell. Equivalently, the subset of the date file consisting of those data records from which the value of the lower cell was computed is a subset of the subset of the data file corresponding to the upper cell. The lattice is, therefore, a graphical representation of the partial ordering of the subsets of the data file defined by those relationships between the parameter fields which define the cells. For example, Figure 1 is the lattice representation of a network comprising all states and their counties and places, and all SMSA's (Standard Metropolitan Statistical Areas) in the United States. This lattice is three-dimensional (as indicated by the dotted lines), which is equivalent to the statement that the statistical tables in the network are generically three-dimensional tables (e.g., a state may be directly disaggregated either by its counties, by its places or by portions of those SMSA's which intersect the state, with none of these disaggregations being a refinement of any other).[2]

The subset-set-superset information represented by the lattice is necessary to construct the equations between the values of the cells defined by the hierarchy (i.e., to construct all cell tables which can logically be formed in the hierarchy) and to define the order of processing of the disclosure analysis from higher level tables and cells to those at lower levels of aggregation. Complementary disclosure analysis is complete if every sensitive cell is adequately protected within each logical cell table in which it appears, and is consistent if adequate protection is afforded all sensitive cells across all logical tables. These conditions must be met if disclosure analysis is to be truly effective, as must the competing goal of minimizing over-suppression of cells (i.e., suppressing more cells than necessary and/or suppressing more significant cells when less significant cells would suffice). To control the disclosure process and evaluate the resulting final publication suppression pattern, a capability for determining best (or neat-best) linear estimates of the values of sensitive cells is necessary.

Once a suppression pattern throughout the hierarchy has been tentatively chosen, it must be evaluated for completeness and consistency. The original system of equations defined by the hierarchy is now replaced by another system of equations S, obtained from the original system by substituting the value of each published cell for its corresponding variable in each equation in which the variable appears. An obvious, and the most precise, means of obtaining estimates of the values of sensitive cells which may be inferred from S is to solve the linear programs:

Maximize x subject to the constraints of S, all variables ≥ 0
Minimize x subject to the constraints of S, all variables ≥ 0

for each sensitive cell X and its corresponding variable x to obtain, respectively, the best linear upper and lower estimates of the value V(X) of X. Additional suppressions may be made or the table released for publication on the basis of the results of this analysis. In a relatively small hierarchy or one containing few suppressions, this may be feasible, but in a large and complex tabulation hierarchy such as a census or major survey, the computational enormity of this undertaking renders this approach impractical.

A commonly employed technique is to estimate the value of each sensitive cell from above in each line (equation) in which it appears in the logical tables (the system $\underline{S}$) by bounding V(X) from above by the difference between the line total (if it is unsuppressed) and the total value of the published cells in the line. The minimum of these estimates is then taken to be the best upper estimate of V(X). Similarly, lower bounds on V(X) are given by the sum of the value of all published cells in a line for which V(X) is a marginal total. The maximum of these lower bounds is then taken to be the best lower estimate of V(X). This amounts to examining only the immediate supersets and subsets in the lattice of the set corresponding to a sensitive cell X. This approach, which we shall refer to as the <u>line-estimation</u> technique, does not in general yield best linear estimates and has serious shortcomings which we shall illustrate.

The tables in Example 1 represent one statistical table (Table (1.00)) containing 16 non-negative internal cells, 8 of which have been suppressed, and 8 unsuppressed marginal totals. Tables (1.11)-(1.22) represent disaggregations of two of the rows and two of the columns of Table (1.00). Using the line-estimation technique, the following estimates of the values of the eight suppressed cells are obtained:

$$2 \le x_{11} \le 9 \qquad 1 \le x_{22} \le 8$$
$$2 \le x_{12} \le 8 \qquad 1 \le x_{23} \le 4$$
$$0 \le x_{13} \le 4 \qquad 1 \le x_{24} \le 6$$
$$0 \le x_{41} \le 11 \qquad 1 \le x_{44} \le 6$$

It is easily demonstrated that these bounds are far from optimal. In (1.00), add the equations defined by the first and second rows and subtract from this the sum of the equations defined by the second and third columns. The result is the equation: $x_{11} + x_{24} = 5$. Therefore, $2 \le x_{11} \le 4$ and $1 \le x_{24} \le 3$. These estimates are still inoptimal. In (1.12), add the equations corresponding to the first and second columns and subtract from this the sum of the equations corresponding to the second and third rows. The result is the equation $x''_{11} - x''_{23} = 2$, which implies that $x''_{11} \ge 2$. As $x_{11} = x''_{11} + 2 + x''_{14}$, then $x_{11} \ge 4$ and hence $x_{11} = 4$ and $x_{24} = 1$, so that $x_{41} = 7$ and $x_{44} = 5$.

The remaining equations from (1.00) are

$$(2.1) \qquad x_{12} + x_{13} = 5$$
$$(2.2) \qquad x_{12} + x_{22} = 8$$
$$(2.3) \qquad x_{13} + x_{23} = 4$$
$$(2.4) \qquad x_{22} + x_{23} = 7$$

As $x_{12} \ge 2$ then $x_{22} \le 6$ by (2.4) and $x_{13} \ge 3$ by (2.2) so that $x_{23} \le 1$ by (2.3). As $x_{13} \ge 0$, then $x_{12} \le 5$ by (2.1) and $x_{23} \le 4$ by (2.3), so $x_{22} \ge 3$ by (2.4).

Without any additional disaggregation information, we conclude $x_{11} = 4$, $x_{24} = 1$, so that $x_{41} = 7$ and $x_{44} = 5$ and

$$2 \le x_{12} \le 5$$
$$0 \le x_{13} \le 3$$
$$3 \le x_{22} \le 6$$
$$1 \le x_{23} \le 4$$

These estimates are best possible in the hierarchy defined by Example 1.

In a similar vein, Example 2 represents a logical table with published marginal totals and 12 out of its 25 internal cells suppressed. Without any disaggregation information for these suppressed cells, it is still possible to precisely determine x_{11} and thereby improve the estimates of the other variables. $x_{11} = 13$, as can be seen by subtracting the sum of the equations corresponding to the second, third and fourth columns from the sum of the equations corresponding to the first, second and third rows.

Underlying these algebraic manipulations and illustrated by their conclusions is that the effect of cell suppression in a hierarchy on each sensitive cell is to publish various aggregates and disaggregates of the cell whose values can be taken as upper and lower estimates of the value of the suppressed sensitive cell; and that these estimiating cell combinations are not necessarily restricted to simple unions of cells along a particular line containing the sensitive cell. Algebraically, this is manifest. From a subject matter analysis perspective, however, it is not clear that, in Example 1, the aggregate $x_{12} + x_{24}$ is effectively published, especially if x_{11} represents the total sales in millions of dollars of furniture stores (SIC 5712) in the state of New York and x_{24} represents the total sales in millions of floor covering stores (SIC 5713) in the state of Pennsylvania. Further, if a few companies were represented in sales in each industry in the respective states to the extent that the total contribution of some of these companies was dominant in the aggregate, then the aggregate cell would also be sensitive and, as its value is effectively published, it would require protection as well. This may be accomplished by suppressing the cells corresponding to x_{14} and/or x_{21}, for example.

The completeness and consistency of the disclosure analysis is measured by the extent to which it produces best possible linear estimates of the values of sensitive cells and of all potentially sensitive aggregates of suppressed cells. This information is keyed to records constructed for each suppressed publication cell and, ideally, is applied iteratively until best possible linear estimates are obtained. It must be emphasized that the global problem, namely that of obtaining best estimates of the value of each sensitive cell in the hierarchy and the choice of those additional suppressions to be made to guarantee that these estimates are all acceptable, is logically equivalent to complex iteration of a potentially vast number of many-dimensional linear programs, beginning with the system of constraints defined by assigning the true value to each variable except those corresponding to the sensitive cells, choosing additional suppressions one or a few at a time, and replacing the true values of these cells by variables, thereby creating a larger system of constraints, solving the linear program over these constraints, and repeating this process until a collection of linear programs as previously described results whose solutions generate only acceptable estimates of the values of all sensitive cells. These programs must also generate and analyze all potentially sensitive combinations of suppressed cells. In the hierarchies corresponding to Examples 1 and 2, this may be a feasible suppression-estimation approach. In general, however, a local array or table analysis and suppression strategy which is adequate in most cases must suffice. One ad hoc technique for analyzing the suppression pattern in a logical array and identifying certain sensitive combinations of cells is described below. These techniques offer a significant improvement over the simple line estimation approach and may be expanded to sets of arrays whenever it is feasible to do so.

As each suppressed cell in the hierarchy is suppressed, a corresponding record is created in a random access or core resident file which indicates that best upper and lower bounds of the value of this cell computed thus far in the analysis, together with the acceptable bounds of equivocation for the cell and the identities and contributions of the t largest contributors to the cell value, for predetermined $t \geq n$. Once the complementary disclosure analysis has been tentatively completed on the basis of line estimates for a logical table at a particular level of aggregation, programs employing appropriate linear estimation techniques will be applied to the disclosure pattern generated for this logical table. Additional suppressions as necessary will be made and the array reanalyzed until all sensitive cells are deemed adequately protected. As with complementary disclosure analysis, the proprocessing flow is from higher levels of aggregation to lower levels. Estimates of the values of suppressed cells made at higher levels of aggregation will be employed and refined as the analysis proceeds to lower levels. A second analytical pass made in reverse order (from lower to higher levels of aggregation) would further refine these estimates, such as was done in the analysis of Example 1.

The ad hoc technique employed to generate upper and lower estimates of the values of suppressed cells and identify sensitive combinations of cells in a logical array will be illustrated for a two-dimensional table. We shall refer to Table (1.00) of Example 1 for concreteness. Note that the analysis in Example 1 employs this technique twice, at two successive levels of aggregation. The straightforward line estimation technique reveals that the value of each suppressed cell in the array is estimated from above by the values of two combinations of cells, namely the sum of the values of all suppressed cells in its row and column, respectively. If the marginal total for either of these lines is published, then, by subtraction, so is the value of the aggregate which estimates the value of each suppressed cell from above. This aggregate is tested for sensitivity and additional suppressions are made on this line, if necessary. Indeed, this test is made for each line during the complementary suppression process before each candidate cell for suppression is suppressed. For purpose of this illustration, assume all row and column marginal totals have been published. To attempt to improve, say, the column estimate thus obtained of, say, the variable x_{11}, the equation for this variable's row (row 1) is employed. To eliminate all variables other than the object variable (x_{11}) from this equation, subtract from this row equation the sum of the column equations corresponding to columns containing the other variables in this equation. The result is a derived equation in which the object variable appears with a coefficient of +1 and other variables appear with a coefficient of -1. Since all variables in the array are assumed to be non-negative, then the net constant (-3) appearing on the opposite side of the new derived equation ($x_{11} - x_{22} - x_{23} = -3$) may yield an improved lower bound for the object variable; and, in any case, may exhibit useful information about the relationships between the variables. To obtain upper bounds, equations involving only variables with positive coefficients are necessary (i.e., cell combinations involving the object cell which are effectively published). To obtain such equations, all variables appearing with coefficient -1 in the derived equation must be eliminated from the derived equation (e.g., x_{22} and x_{23}). To do this, add the row equation of each row containing one of these variables with coefficient -1 to the derived equation. The new derived equation ($x_{11} + x_{24} = 5$) yields an upper bound for each variable appearing in it and exhibits another cell combination which is effectively published. This procedure is then iterated, with equations corresponding to columns not containing the object variable but which contain variables with coefficient +1 in the derived equation being subtracted from the derived equation and equations corresponding to rows containing variables with coefficient -1 in the new derived equation being added to the new derived equation, with sensitive cell combinations and upper and lower bounds on variables being noted as they arise. The procedure is terminated when an equation involving only variables from the equation corresponding to the column containing the object variable is derived. In this example, the row equation $x_{22} + x_{23} + x_{24} = 8$ is added to the derived equation $x_{11} - x_{22} - x_{23} = -3$ to yield the new derived equation $x_{11} + x_{24} = 5$. The column equation $x_{24} + x_{44} = 6$ is then subtracted from the new derived equation, yielding

$x_{11} - x_{44} = -1$, to which the row equation $x_{41} + x_{44} = 12$ is added to produce the final derived equation $x_{11} + x_{41} = 1$. In general, the final derived equation will either be the original column equation for the object variable, or a refinement of this equation. The same process is applied to attempt to improve the row estimate of the object variable, with column equations being added and row equations being subtracted, beginning with the column containing the object variable, the process terminating with an equation involving only variables from the row containing the object variable. If all marginal totals are not published, then some of the constants which would otherwise result are replaced by interval estimates of these marginal totals obtained as thir process was applied at one higher level of aggregation.

This procedure can be visualized geometrically. The variables x_{12}, x_{13}, x_{22}, x_{23} form a closed suppression pattern within the first two rows in that they include all suppressions in second and third columns. Removing this pattern from the first two rows yields the cells x_{11} and x_{24} and the resulting equation $x_{11} + x_{24} = 5$. In Example 2, this imbalance is even more pronounced as the residue so obtained consists of precisely x_{11}.

The advantage this procedure has over the technique of solving linear programs on a logical table by logical table basis is that the variables and cell combinations to be investigated need not be specified in applying the ad hoc technique and that the ad hoc technique may be applied on a row by row and column by column basis so that the same computations need not be repeated. In the linear programming approach, many linear programs must be solved over the same set of constraints. Moreover, in the ad hoc approach, each derived equation may be examined for sensitivity of cell combinations, which a general purpose linear programming package would not be geared to do. In applying the linear programming technique, one must predict the cell combinations of interest (e.g., the objective function $x_{11} + x_{24}$ must be maximized and minimized over the constraints to yield the final result). The linear programming approach does, however, yield the best linear estimates of the values of suppressed cells, whereas in general the ad hoc approach may not. Neither technique is guaranteed to produce all sensitive cell combinations, although one could apply a modification of the ad hoc technique to exhaust the various row and column combinations. Under investigation is a program that combines the estimation power of linear programming techniques with the ad hoc technique's computational simplicity and its ability to generate appropriate cell combinations.[3]

THE AUTOMATED SYSTEM

An automated system to perform complementary disclosure analysis has been designed by the author and is being developed for implementation on a Univac 1110. The system will generate the lattice which defines the cells from the geographic and subject-matter codes. The lattice will control the aggregation of the cells, the construction of the logical tables and the inter-table disclosure analysis which maintains consistency between estimates of the values of suppressed cells. Minimum suppression algorithms will be applied to each logical table in tandem with estimation techniques as described in this paper to effect the intra-table complementary disclosure analysis and, to the extent possible, assure that only acceptable estimates of the values of suppressed cells may be made within each logical table. Testing of this system is planned for Spring 1977. Based upon these test results, the decision to implement this system in the 1977 Economic Censuses will be made by the U.S. Bureau of the Census. Research into techniques of disclosure analysis is on-going at the Census Bureau.

---

[1] For further details on cell suppression strategies, see Cox, L., Disclosure Analysis and Cell Suppression, Proceedings of the Social Statistics Section, American Statistical Association Annual Meeting, 1975, pp. 380-382.

[2] See Cox, L., Applications of Lattice Theory to Coding and Decoding, Proceedings of the ACSM AUTO-CARTO II (Second International Conference on Automated Cartography), 1975, (Proceedings in publication)

[3] An approach to the problem of identifying certain single cells in the hierarchy whose precise values are effectively published is described in Menezes, O.J., Testing for Confidentiality and Residual Disclosure in the Census of Mines, Forestry and Manufactures, Ontario Statistical Center, Ontario, 1975 (unpublished). Menezes employs matrix rank computations to identify variables in large systems of linear equations which can be solved for by matrix inversion (such as x_{11} in Example 2). The remaining "free" variables (i.e., all other suppressed cells whose precise values are not computable) are not estimated, however, and available upper and lower bound information is not employed in the computation, as was done in Example 1.

EXAMPLE 1

(1.00)

x_{11}	x_{12}	x_{13}	3	12
2	x_{22}	x_{23}	x_{24}	10
5	3	8	2	18
x_{41}	2	4	x_{44}	18
18	13	16	11	58

(1.11)

1	2	0	1	4
x'_{21}	x'_{22}	x'_{23}	1	4
x'_{31}	x'_{32}	x'_{33}	1	3
x_{11}	x_{12}	x_{13}	3	11

(1.12)

x''_{11}	0	2	x''_{14}	x_{11}
0	x''_{22}	x''_{23}	0	2
x''_{31}	x''_{32}	1	0	5
0	0	x''_{43}	x''_{44}	x_{41}
4	4	5	5	18

(1.21)

1	1	1	1	4
1	x'''_{22}	x'''_{23}	x'''_{24}	4
0	x'''_{32}	x'''_{33}	x'''_{34}	2
2	x_{22}	x_{23}	x_{24}	10

(1.22)

1	1	1	0	3
1	0	x^{IV}_{23}	x^{IV}_{24}	x_{24}
0	0	1	1	2
0	1	x^{IV}_{43}	x^{IV}_{44}	x_{44}
2	2	4	3	11

EXAMPLE 2

x_{11}	x_{12}	x_{13}	1	2	16
2	x_{22}	x_{23}	x_{24}	1	14
2	x_{32}	1	x_{34}	3	12
x_{41}	2	5	0	x_{45}	12
x_{51}	4	1	3	x_{55}	12
16	12	11	11	16	66

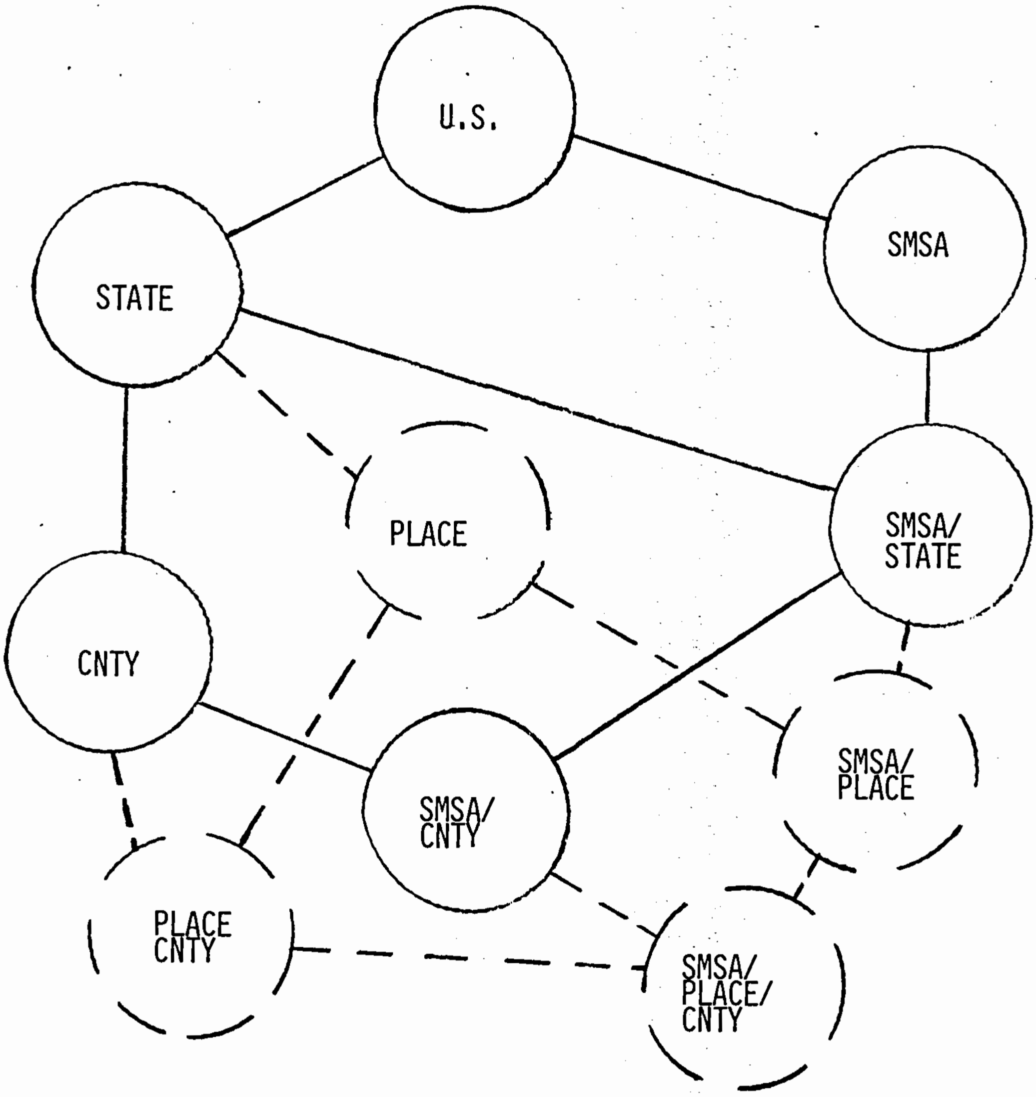
U.S.
STATE
SMSA
PLACE
SMSA/
STATE
CNTY
SMSA/
PLACE
SMSA/
CNTY
PLACE
CNTY
SMSA/
PLACE/
CNTY

FIGURE 1