AVOIDING DISCLOSURE IN TABULATIONS

Richard Bell, Social Security Administration

The various types of disclosure in tabulations are discussed. Appropriate examples, taken from tables appearing at the end of this paper, are presented.

The classification which follows represents an effort to develop a comprehensive and logical description of different types of disclosure. Suggestions for improvement will be welcomed.

Disclosure will be studied both for tabulations involving count (frequency) data and for those containing quantity (magnitude) data.

1. Exact disclosure
   a. Count data: A marginal total equals one of its detail cells; this detail cell is defined as narrowly as possible from the records upon which the tabulation is based.
   Table 1: All beneficiaries in County B are black.

   b. Magnitude data:
   (i) A quantity corresponds to a cell with only one member.
   Table 2: Total sales for the single establishment in Industry B is $125,000,000.

   (ii) A quantity assumes its maximum or minimum possible value.
   Table 3: If the maximum possible payment under the program is $190, then each person in State B receives precisely $190.

2. Approximate disclosure
   a. Count data: A detail cell is zero; the disclosure is not exact.
   Table 1: No beneficiary in County C is white.
   Table 4: The age of each beneficiary in County B is restricted to the interval (65, 69).

   b. Magnitude data:
   (i) The value of a quantity corresponding to an individual cell member is restricted to an interval (L, U) where the lower and upper limits are determined by such relationships as the following from published data and logical operations:

$$U = M;$$
$$U = m + N(A - m);$$
$$U = T - m(N - 1);$$
$$L = m;$$
$$L = M - N(M - A);$$
$$L = T - M(N - 1),$$

   where

   N is the number of members in the cell,

   A is the published, average value of the quantity for the cell,

   T is the published, total value of the quantity for the cell (note T = AN),

   M and m are the maximum and minimum possible values, respectively, for any member in the cell.

   (If two or more distinct values are available for either U or L, select the largest of the possible lower limits for L and the smallest of the possible upper limits for U, respectively.)

   Table 2: Total sales for each establishment in Industry C of table 2 is between 0 and $125,000,000.

   Table 3: Monthly benefit for each of the four beneficiaries in State A cannot exceed $632.

$$L = m = 0$$
$$U = m + N(A - m) = 4(158).$$

   Table 3: If the maximum possible payment under the program is $192, then each person in State B receives at least $120.

$$U_1 = m + N(A - m) = 0 + 36(190) = 6840$$
$$U_2 = M = 192$$
$$U = \text{minimum}(6840, 192) = 192$$
$$L_1 = m = 0$$
$$L_2 = M - N(M - A) = 192 - 72 = 120$$
$$L = \text{maximum}(0, 120) = 120.$$

   (ii) Information about other characteristics associated with the same cell is used to estimate the value of a quantity corresponding to an individual cell member.
   Table 2: If it is known from another source that all five establishments in Industry C have about the same number of employees, then total sales of $25,000,000 can be estimated for each member of the industry. Similarly, if one of the five has 80% of the employment, then the estimate for this "largest" establishment of $100,000,000 in total sales is reasonable.

3. Probability-based disclosure
   The probability that a given member of a total cell with T members belongs to a particular detail cell with D members is D/T.

   Table 1: Assign a probability of 28/30 to the event that a person known to be a

beneficiary in County C is black.

4. Internal disclosure
A member (or a coalition of members) of a group uses his own (their own), as well as published data, to obtain confidential information about others in the group.

a. Count data:
(i) Exact disclosure: The difference between the values of a marginal total and one of its detail cells is equal to the number of members of a coalition not belonging to the detail cell; the detail cell is as narrowly defined as possible.

(ii) Approximate disclosure: The difference between the values of a marginal total and the sum of a proper subset of detail cells is equal to the number of members of a coalition not belonging to the proper subset (equivalently, all members of a detail cell also belong to a coalition); but the disclosure is not exact.

(iii) Probabilistic disclosure: Define the following

S = the published number of members in the total cell,

D = the published number of members in the detail cell,

C = the coalition size,

X = the number of coalition members also belonging to the detail cell.

The probability is $\frac{D-X}{S-C}$ that another member of the marginal total, but outside the coalition, lies in the detail cell.

Table 5: The black worker in County A knows that all the other workers in his county are white. A black worker in County B deduces that the probability is 65/66 that another worker, unknown to him, is white; the coalition of two black workers in County B knows that all other workers in County B are white. The black worker in County C knows he is the only black worker in his county.

b. Magnitude data:
(i) Exact disclosure: After a coalition of size C adjusts a published figure by means of its own data, the revised value of the characteristic involves either type of disclosure described in 1b. (Equivalently, a quantity is published for a cell of size C + R where one of the following conditions hold:
(1) R = 1
(2) The revised value of the published figure, obtained by adjusting for the contribution of the coalition, is a maximum or a minimum.)

(ii) Approximate disclosure: After a coalition adjusts a published figure by means of its own data, the value of a quantity corresponding to an individual cell member is restricted to an interval as described in 2b.

Table 2: If one of the establishments in Industry C has total sales = S, then total sales for each of its competitors must be less than \$125,000,000 - S.

Table 6: If the maximum possible benefit for each of the beneficiaries is \$140, then a person receiving \$40 in County A can deduce that each of the other two members of his cell must receive between \$120 and \$140. Although it would be impossible for a user, not belonging to the cell for County B, to restrict the payment amount to either person in that county to any interval smaller than (0, 140), either beneficiary can readily compute the payment to the other person by use of the published cell.

5. Indirect disclosure
Any of the above types of disclosure is derived by algebraic manipulation and/or logical operations.

a. Count data: Neither table 7 nor table 8 provides sensitive information directly. However, by combining information from both tables, it is seen that all men over 75 with medical coverage have hospital coverage; all women with medical coverage but without hospital coverage are under 65.

Table 9: By subtraction, it follows that there are no workers of race other than white or black in Industry A and that all workers in Industry C are white.

b. Magnitude data: Suppose Industry A consists of two disjoint sub-industries A1 and A2 and that the following information is available from various tables:

Industry	Number of companies	Total sales
A	5	\$200,000,000
A1	4	\$150,000,000

By subtraction, the one company belonging to Industry A2 has total sales of \$50,000,000.

Table 1: Number of beneficiaries by county and race

County	Race			
	Total	White	Black	Other
A	40	15	20	5
B	30	0	30	0
C	30	0	28	2

Table 2: Total sales, by industry

Industry	Number of establishments	Total sales
A	18	$450,000,000
B	1	125,000,000
C	5	125,000,000

Table 3: Average monthly benefits, by State

State	Number of beneficiaries	Average monthly benefit
A	4	$158
B	36	190

Table 4: Number of beneficiaries, by county and age

County	Age class				
	Total	Under 65	65-69	70-74	75 and over
A	37	3	15	11	8
B	4	0	4	0	0

Table 5: Race of workers by county

County	Total	White	Black	Other
A	94	93	1	0
B	67	65	2	0
C	103	101	1	1

Table 6: Number of beneficiaries and average payment amount

County	Number of beneficiaries	Average payment amount
A	3	$100
B	2	70

Table 7: Number of persons with hospital and medical coverage, by age and sex

Age	Hospital and medical coverage		
	Total	Male	Female
Total.....	9,593	4,633	4,960
Under 65.....	3,534	1,714	1,820
65-74.....	3,147	1,517	1,630
Over 75.....	2,912	1,402	1,510

Table 8: Number of persons with medical coverage, by age and sex

Age	Medical coverage		
	Total	Male	Female
Total.....	9,609	4,640	4,969
Under 65.....	3,548	1,719	1,829
65-74.....	3,149	1,519	1,630
Over 75.....	2,912	1,402	1,510

Table 9: Race of workers by industry

Industry	Total	Male	Female
Total.....	400	328	62
A.....	60	30	30
B.....	236	194	32
C.....	104	--	--