BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-88/08

DOCUMENTATION OF THE SAMPLING AND ESTIMATION
PROCEDURES FOR THE LAW ENFORCEMENT AGENCY SURVEY

by

Carma R. Hogue
Statistical Research Division
Bureau of the Census
Room 3130, F.O.B. #4
Washington, D.C. 20233 U.S.A.

Recommended by: Nash J. Monsour

Report completed: March 1988

Report issued: March 1988

Documentation of the Sampling and Estimation
Procedures for the Law Enforcement Agency Survey

The Law Enforcement Agency Survey, which began in 1987, is a first-time
survey of State and local law enforcement agencies conducted for the Bureau of
Justice Statistics. This survey is intended to give national estimates of
administrative and management statistics (including employment and workload
measures) in State and local law enforcement agencies. The sample was
designed with the major variable of interest being the total number of sworn
employees in these agencies. Details of the sample design are given in
section 1. Details of the estimation of totals are given in section 2 with
details of the variance estimation given in section 3. Issues for future
surveys are discussed in section 4.

1.   Sample Design

The 1985 Director Survey of Law Enforcement Agencies was used as the
sampling frame. The Directory Survey, which was sent to all State and local
law enforcement agencies, verified the existence of these agencies, their
jurisdiction, and level of government, and obtained current data on agency
employment (total employees and sworn personnel).

All State law enforcement agencies as well as all 287 city and township
law enforcement agencies with at least 135 sworn employees were included in
the sample with certainty. The remainder of the approximately 16,000 law
enforcement agencies are grouped into six strata. The primary sampling unit
is a county-area. At the second stage, all county law enforcement agencies
and a sample of city and township agencies were selected from the county-areas
selected at the first stage. The goal was to select about 3000 agencies with
a coefficient of variation of .01 on the variable "Total Sworn Employees."

The first-stage units, county-areas, are the same as the units used in
the 1983 Survey of Local Jails. The sample design for this survey is
described in a memorandum for the record dated June 1984, written by Blanche
Hurwitz and entitled "National Survey of Local Jails." As such, there are 5
strata defined by the average daily inmate population (ADP) of the jail(s) in
the county-area. There were 301 county-areas that were out-of-scope in the
Survey of Local Jails since they contained no jails. These county-areas were
placed into a separate stratum. All of these first-stage units were included
with certainty. The stratum definitions and sample sizes are given in Table 1.

Table 1. Stratum Designations and Sizes

Stratum	ADP	No. of Counties (M_h)	No. of Counties in Sample (m_h)	First-stage Weight (w_h)
1	100+	45	45	1.0000
2	50-99	314	157	2.0000
3	21-49	656	164	4.0000
4	8-20	878	125	7.0240
5	0-7	833	85	9.8000
6	No jails	301	301	1.0000
		3027	877	

We decided to use the county-areas from the Survey of Local Jails because of
time constraints in the sampling and pre-survey processing weeks. Also,
processing costs may be reduced by using the same county-areas in both
surveys. It was also desirable to do this for BJS's analysis purposes. We
believe that the correlation between a measure of size of the jail population
in a county-area and the size of the law enforcement agencies in an area
should be relatively high.

In the second stage of sampling, a systematic sample of law enforcement
agencies was drawn from each stratum. After the certainty units (all county

law enforcement agencies and large municipal agencies) were removed, the remaining law enforcement agencies were arranged in order of decreasing total sworn employees within each stratum. Then, a systematic sample was selected from each stratum using the random start and take every given in Table 2.

Table 2. Second-Stage Sampling Specifications for
Municipal and Township Agencies

Stratum	$\hat{N}_h$	N_h'	$n_{h.}^*$	Random Start	Take Every (r_h)
Certainty	289	289	289	1.000	1.000
1	4097	4097	799	4.895	5.128
2	2061	970	305	2.385	3.180
3	2646	623	275	0.618	2.266
4	2155	289	143	1.128	2.021
5	1238	119	77	1.475	1.545
6	264	264	147	1.714	1.796
	12750	6651	2035		

The total number of law enforcement agencies for each stratum was not known and had to be estimated from the known number of noncertainty agencies in the sample from each stratum, N_h', and from the first-stage sampling weight, $w_h = M_h/m_h$, where M_h is the total number of county-areas in stratum h and m_h is the number of county-areas selected in the sample from stratum h. Therefore, $\hat{N}_h = \left(w_h N_h' \big/ \sum_{h=1}^{L_{NC}} w_h N_h' \right) N$

where   N is the known total number of agencies in the noncertainty strata, and

L_{NC} is the number of noncertainty strata (=6).

For strata 1 and 6, $\hat{N}_h = N_h'$, since they are first-stage certainty strata.

There were 967 county agencies and 289 municipal certainty agencies in the sample, thus leaving 1746 agencies to be selected at random from the six strata. Using the variable 'Total Sworn Employees' and $\hat{N}_h$, the 1746 agencies

were optimally allocated to the strata. Table 2 gives the results of that allocation. The resulting relative standard error was about .012 for an estimate of total sworn employees from local law enforcement agencies and about .010 for the same variable for State and local law enforcement agency estimates. The relative standard errors for total employment estimates were about the same.

In Texas, during the followup of nonrespondents, it was discovered that the sampling frame was not complete. Central offices, which had responded to the 1985 survey, were included in the sampling frame, but several smaller constable offices, which should be in the current survey, were not included in the sampling frame. The nonrespondents and the extra constable offices were ordered within county by size in a serpentine fashion (highest to lowest in the first county, lowest to highest in the next, etc.). A systematic sample of one-fourth of the offices was selected for the survey.

In summary, the sample was designed to give 289 municipal certainty agencies, 967 county agencies, 105 State certainty agencies, and 1746 noncertainty municipal agencies. The final sample included 312 municipal certainty agencies, 967 county agencies, 64 State certainty agencies, and 1735 noncertainty municipal agencies. The discrepancy in the number of State agencies was due mainly to finding some State agencies to be out-of-scope for this survey (predominantly special purpose State agencies such as ABC Boards, fish and game wardens, etc.). (The discrepancy in the number of noncertainty municipal agencies was due largely to getting only 138 rather than 147 agencies in stratum 6. This discrepancy in the number of noncertainty agencies is due to the fact that some State agencies were included in stratum 6.) This error was found after the sampling was done. Time did not permit us to evaluate the effect of this problem, though it does not bias the estimates.

2. Estimation

2.1 Totals

In order to estimate totals (employees, sworn employees, full-time, part-time, etc.), the following equation should be used: (The first term comes from the certainties; the second term is from the county law enforcement agencies; and the third piece is from noncounty, noncertainty agencies. If an estimate for the county agencies is desired, only county certainties and the second term would be added. If a municipal total is desired, the third term is added to municipal certainties. If a State and local estimate is desired, terms 2 and 3 are added to State, county, and township certainties.)

$$x' = X_C' + \sum_{h=1}^{L_{NC}} w_h \sum_{i=1}^{m_h} x_{hi} + \sum_{h=1}^{L_{NC}} w_h r_h \sum_{i=1}^{m_h} \sum_{j=1}^{n^*_{hi}} a_{hij} \, x_{hij} \qquad (1)$$

where

X_C' is the total of variable X for all law enforcement agencies that are in the sample with certainty, i.e., first- and second-stage weights are 1.00.

L_{NC} is the number of noncertainty strata (= 6 for this sample).

w_h is the first-stage sampling weight, as given in Table 1 = M_h/m_h (= 1 for strata 1 and 6).

M_h is the number of county-areas in stratum h.

m_h is the number of county-areas in the sample for stratum h.

r_h is the second-stage sampling weight for stratum h, as given in Table 2.

n^*_{hi} = number of noncounty law enforcement agencies in sample from stratum h, county i.

x_{hi} = the total value of the characteristic for all county law enforcement agencies in stratum h, county area i. (In some counties, more than one county agency exist in the county. x_{hi} is the total of all agencies in the county.)

x_{hij} is the value of the characteristic for stratum h, county area i, noncounty law enforcement agency j. (For counts of agencies, x_{hij} is either 0 or 1 for the characteristic of interest.)

$a_{hij} = \begin{cases} 4.0 \text{ if the agency is a subsampled agency in Texas} \\ 1.0 \text{ otherwise} \end{cases}$

2.2 Year-to-Year Change

In the future, if year-to-year changes are desired, the following equation should be used.

$$\hat{D} = \frac{x' - y'}{y'} \tag{2}$$

where x' is as defined in equation (1) for the current year;

y' is as defined in equation (1) for the previous year.

2.3 Ratios

The ratio of one estimated total to the other, e.g., the ratio of male to female police officers, may be estimated as follows:

$$\hat{R} = \frac{x'}{z'} \tag{3}$$

where x' is as defined in equation (1) for characteristic X; and

z' is similarly defined for characteristic Z.

2.4 Means

For estimates of means, the total of the variable of interest should be divided by the total number of law enforcement agencies in the universe.

$$\bar{x} = x'/N \qquad\qquad (4)$$

where x' is given in equation (1), and

           N is the total number of law enforcement agencies.

2.5 Proportions (Percentages)

In order to get the proportion of a total in various subcategories, the following equation should be used:

$$P_k = y_k' / y_.' \qquad\qquad (5)$$

where y_k' is estimated using equation (1) for subcategory k, and

$$y_.' = \sum_{k=1}^{K} y_k' \qquad \text{or the total of all subcategories for the variable of interest (where there are K subcategories).}$$

Percentages are obtained by multiplying P_k by 100.

3. Variance Estimation

3.1 Totals

The first term of the estimator given in equation (1) involves only law enforcement agencies taken with certainty. Therefore, that term does not contribute to the variance. The second term of the estimator comes from county law enforcement agencies which were taken with certainty at the second stage. The variance of this part of the estimator can be estimated as follows:

$$s_{x_1}^2 = \hat{Var}\left(\sum_{h=1}^{L_{NC}} w_h \sum_{i=1}^{m_h} x_{hi} \right) = \sum_{h=1}^{L_{NC}} m_h w_h (w_h - 1) s_h^2 \qquad (6)$$

where

$$s_h^2 = \left[\sum_{i=1}^{m_h} x_{hi}^2 - \left(\sum_{i=1}^{m_h} x_{hi} \right)^2 / m_h \right] / (m_h - 1)$$

All other variables were defined in equation (1).

The contribution to the variance from the third part involves the two stages of selecting jurisdictions and then sampling law enforcement agencies from among all of the agencies in those selected jurisdictions. Since most of the noncertainty jurisdictions would have either 0 or 1 agency selected for the sample, the clustering would be minimal and the sample could be viewed as a stratified random selection of agencies from the universe of agencies in the stratum. As such the estimate of the variance from the noncounty law enforcement agencies is

$$s_{x_2}^2 = \hat{Var}\left(\sum_{h=1}^{L_{NC}} w_h r_h \sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}^*} a_{hij} x_{hij} \right) = \sum_{h=1}^{L_{NC}} n_{h\cdot}^* w_h r_h (w_h r_h - 1) s_h^{2*} \qquad (7)$$

where

$$s_h^{2*} = \left[\sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}^*} a_{hij}^2 x_{hij}^2 - \left(\sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}^*} a_{hij} x_{hij} \right)^2 / n_{h\cdot}^* \right] / (n_{h\cdot}^* - 1)$$

$n_{h\cdot}^*$ = number of noncounty law enforcement agencies in stratum h.

All other variables were defined in equation (1). The estimated variance, $s_{x'}^2$, of the estimate given in equation (1) is obtained by adding estimates of variance from equations (6) and (7):

$$s_{x'}^2 = s_{x_1}^2 + s_{x_2}^2 \qquad (8)$$

3.2 Year-to-Year Change

The estimate of change, $\hat{D}$, given in equation (2) can be rewritten as

$$\hat{D} = \frac{x'}{y'} - 1$$

and the variance of $\hat{D}$ is then simply the variance of the ratio x'/y' is estimated as follows:

$$s_{\hat{D}}^2 = \left(\frac{x'}{y'}\right)^2 \left[s_{x'}^2/x'^2 + s_{y'}^2/y'^2 - 2\, s_{x'y'}/x'y' \right] \qquad (9)$$

where   x' and y'   are the totals for the current and previous years and are estimated using equation (1).

$s_{x'}^2$ and $s_{y'}^2$ are estimated using equation (8).

$$s_{x'y'} = s_{x_1 y_1} + s_{x_2 y_2}$$

$$s_{x_1 y_1} = \sum_{h=1}^{L_{NC}} m_h w_h (w_h - 1) \left[\sum_{i=1}^{m_h} x_{hi} y_{hi} - \left(\sum_{i=1}^{m_h} x_{hi} \right) \left(\sum_{i=1}^{m_h} y_{hi} \right) \right] / (m_h - 1)$$

$$s_{x_2 y_2} = \sum_{h=1}^{L_{NC}} n_{h\cdot}^{*} w_h r_h (w_h r_h - 1) \left[\sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}^{*}} x_{hij} y_{hij} - \right.$$

$$\left(\sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}^{*}} x_{hij} \right) \left(\sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}^{*}} y_{hij} \right) / n_{h\cdot}^{*} \left. \right] / (n_{h\cdot}^{*} - 1)$$

3.3 Ratios

The variance of $\hat{R}$ as defined in equation (3) can be estimated as in equation (9) with an estimate, z', for a different characteristic, Z, substituted for y'.

$$s_R^2 = \left(\frac{x'}{z'}\right)^2 \left[s_{x'}^2/x'^2 + s_{z'}^2/z'^2 - 2\, s_{x'z'}/x'z' \right] \tag{10}$$

All variables have been defined previously.

3.4 Means

The variance of the estimates of means can be estimated as in equation (11).

$$s_{\bar{x}}^2 = s_{x'}^2 / N^2 \tag{11}$$

where $s_{x'}^2$ is given in equation (8).

3.5 Proportion

The variance of P_k can be estimated as in equation (12).

$$s_{P_k}^2 = P_k^2 \left[s_{y_k'}^2/y_k'^2 + s_{y_.'}^2/y_.'^2 - 2 s_{y_k'y_.'}/y_k' y_.' \right] \tag{12}$$

where

$s_{y_k'}^2$ is estimated as in equation (8) with sums over only those units in category k;

$s_{y_.'}^2$ is estimated as in equation (8) with sums across all categories;

$s_{y_k'y_.'} = s_{y_{k1}'y_{.1}'} + s_{y_{k2}'y_{.2}'}$

$$s_{y'_{k1}y'_{\cdot 1}} = \sum_{h=1}^{L_{NC}} m_h w_h (w_h - 1) \, s_{hy'_{k1}y'_{\cdot 1}}$$

$$s_{y'_{k2}y'_{\cdot 2}} = \sum_{h=1}^{L_{NC}} n_{h\cdot}^{*} \, w_h r_h (w_h r_h - 1) \, s_{hy'_{k2}y'_{\cdot 2}}^{*}$$

where

$$s_{hy'_{k1}y'_{\cdot 1}} = \left[\sum_{i=1}^{m_h} y_{khi} y_{\cdot hi} - \left(\sum_{i=1}^{m_h} y_{khi} \right)\left(\sum_{i=1}^{m_h} y_{\cdot hi} \right)/m_h \right]/(m_h - 1)$$

$$\cdot \; s_{hy'_{k2}y'_{\cdot 2}}^{*} = \left[\sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}^{*}} a_{hij} y_{khij} y_{\cdot hij} - \left(\sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}^{*}} a_{hij} y_{khij} \right)\left(\sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}^{*}} a_{hij} y_{\cdot hij} \right) \right.$$

$$\left. /n_{h\cdot}^{*} \right] \; / \; (n_{h\cdot}^{*} - 1)$$

All other variables have been defined previously.

3.6 Other Estimates of Variability

The estimated standard error of an estimate is the square root of the estimated variance of the estimate. The estimated relative standard error is the estimated standard error of the estimate of interest divided by the estimate of interest.

4. Issues for Future Surveys

Studies should be conducted to determine whether we want to continue to have this survey connected to the Survey of Local Jails. We should study the advantages and disadvantages of using the same county areas for both samples. Data from the 1985 Census of Law Enforcement Agencies can be used to

determine the effect of stratifying on total law enforcement agency employment within the county rather than average daily jail population.

The estimates of totals and ratios should be printed out for all variables of interest. Each estimate's standard error and relative standard error should also be printed out. The table in the appendix shows the equations for estimating the variable of interest and the appropriate variance.

Appendix

	Variable	Equation #	Variance equation #
I1.	Size of jurisdiction	1	8
I2.	Size of agency	1	8
I3.	Type of agency	1	8
OD1.	Law enforcement functions	1 and 5	8 and 12
OD2.	Calls resulting in a dispatch	1	8
OD3.	Agency participation in 911	1 and 5	8 and 12
OD4.	Agencies operating a lockup	1 and 5	8 and 12
OD6.	Maximum holding time	1 and 5	8 and 12
OD7.	Maximum capacity	1 and 5	8 and 12
OD8.	Average daily population	1 and 5	8 and 12
OD9.	Total number of admissions to lockups	1	8
ED1.	Equipment provided	1 and 5	8 and 12
ED2.	Sidearm issued	1 and 5	8 and 12
ED3.	Vehicles operated	1	8
ED4.	Vehicle take home policy	1 and 5	8 and 12
ED5.	Off-duty car use	1 and 5	8 and 12
ED6.	Computer usage	1 and 5	8 and 12
PD1.	Personnel totals and means	1 and 4	8 and 11
PD2.	Personnel by race	1 and 4	8 and 11
PD3.	Average salary by rank	3	10
ED1.	Mean operating expenditures	4	11
ED2.	Capital expenditures by agency type	3	10
ED3.	Capital expenditure ranges	1	8

	CJ-44 Variables:		
PP1.	Residency requirement	1 and 5	8 and 12
PP1a.	Geographical boundaries	1 and 5	8 and 12
PP2.	Type of patrol shift	1	8
PP3.	Incentive/differential pay	1 and 5	8 and 12
PP4.	Educational requirements	1	8
PP5.	Classroom/field training hours	1 and 3	8 and 10
PP5a.	Average cost of training	3	10
PP5b.	Range of training costs	1	8
PP6.	Collective bargaining	1	8
PP7.	Program staffing	1	8
PP8.	Written directives	1	8
PP9.	Civil litigation cases by type	1	8