DOCUMENTATION OF THE SAMPLING AND ESTIMATION
PROCEDURES FOR THE SURVEY OF FELONY CONVICTIONS

by

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

Documentation of the Sampling and Estimation
Procedures for the Survey of Felony Convictions

The Survey of Felony Convictions was conducted in order to gather much needed national data on the sentencing of convicted felons. The National Judicial Reporting Program (NJRP) is sponsored by the Bureau of Justice Statistics with the pretest of the program being conducted by the Bureau of the Census. Census will investigate the feasibility of developing national estimates of felony sentences by type of crime and by length and type of sentence. In particular, Census will be examining the courts to determine the avaтlability of the records, the condition of the records, automation of the various offices, etc. Census will test different methods of data collection: collection of computerized data, on-site compilation of data from court records, and mail canvass reporting.

A two-stage cluster sample was selected in order to yield a relative standard error (coefficient of variation) of about .015 on both a national estimate and an estimate for the top 75 jurisdictions or county areas (based on 1985 population) of the percentage of convictions receiving a prison sentence. The first stage was a stratified systematic sample of jurisdictions. At the second stage, all general jurisdiction courts handling felony cases within a jurisdiction were included in the sample, but a systematic sample of convictions was selected for each of eight types of crime: homicide, rape, robbery, burglary, assault, larceny, drugs, and all other felonies.

On May 1, 1987, specifications for selecting the sample were distributed. A revision to the specifications (based on experiences in York Co., Pennsylvania) was issued in a memorandum dated September 23, 1987.

Sampling and estimation procedures are documented in this paper. Sources of the data are given in section 1; first-stage sampling specifications are given in section 2; and second-stage sampling procedures are described in section 3. Section 4 describes the estimation of totals and ratios. Variance estimation is discussed in section 5. In section 6, considerations for future redesigns are discussed. The appendix contains definitions of the terminology used in this report.

1. Sources of Data

The data for the first-stage sample selection came from the 1985 Directory Survey of General Jurisdiction Courts which was collected by Governments Division. The 1985 Survey of Courts, a first-time mail canvass of all general jurisdiction courts in the country, was designed to obtain data on workload for felonies and other criminal cases that were filed with the courts. Total numbers of convictions, acquittals, dismissals, etc. were also recorded along with data on office automation.

Three major problems exist with basing the sample on the data collected from the survey. First, in the totals given for felony convictions, dispositions, and filings by some courts, misdemeanors were also included. Sometimes, the inclusion of misdemeanors doubled the expected felony count (as in York, Pennsylvania). Only actual site-visits enable us to determine the percentage of felony counts that were misdemeanors.

The second problem was that some of the data given in the counts are defendant-based (as requested on the form); some are charge-based; and others are indictment-based. Examples of these three types are given below for the case of an indictment in which two defendants have three charges each:

(1) Defendant-based = 2 cases

(2) Charge-based = 6 cases

(3) Indictment-based = 1 case

In the Top 75 jurisdictions, three included misdemeanors in their counts; seven reported indictment-based data; six reported charge-based data; one reported a mixture of defendant- and indictment-based data; two reported a mixture of defendant- and charge-based; and two reported a mixture of indictment- and charge-based data. When the survey is repeated, we will attempt to adjust the data received from the courts using the data currently being collected from our sample. In the next total canvass of courts, persons providing the data should be encouraged to give their best estimates of defendant-based felonies.

The third problem was that some courts could not provide the number of convictions (reporting total dispositions or, failing that, filings). Prior to sampling, missing convictions data had to be imputed. Of the 3109 jurisdictions, 816 of them were missing convictions data. Of the 75 largest jurisdictions, 16 were missing convictions. Separate regressions on dispositions were done on the Top 75 and the balance in order to obtain values for the missing convictions. (In imputing for the Top 75, Los Angeles, California was left out of the regression since it dominated the regression.) Since some dispositions data were also missing, the following procedure was followed:

1. Felony filings counts were imputed from population counts;

2. Felony dispositions counts were imputed from felony filings; and

3. Felony convictions counts were imputed from felony dispositions.

For the second stage of sampling, the Bureau of Justice Statistics (BJS) provided data based on past studies in the nation's largest cities. These previous studies were done by the National Association of Criminal Justice Planners (by Mark Cunniff) and by PROMIS. BJS provided an approximate distribution of convictions by crime type. They also provided conservative estimates of the percentages of convictions that are sentenced to prison, by crime type. (Percentages above 50 percent were rounded down and those below 50 percent were rounded up.) The second-stage sampling rates were based on these data, which are given in Table 1.

Table 1. Distribution of Convictions and Sentencing Percentages by Crime Type

Crime Type	Estimated Percentages of Convictions	Estimated Sentencing Percentages
Homicide	2.0%	80%
Rape	2.0%	60%
Robbery	10.5%	60%
Assault	5.5%	25%
Burglary	17.5%	30%
Larceny	13.5%	15%
Drugs	14.0%	20%
Other	35.0%	15%

2. First Stage of Sampling

The sample was designed to yield 100 jurisdictions to give a relative standard error of .03 on estimates of the number of dispositions for the 75 largest jurisdictions. The sample size of 100 jurisdictions was prescribed by the Bureau of Justice Statistics and is expected to yield a relative standard error of about 9 percent on a national estimate of total dispositions. The jurisdictions were stratified on the number of convictions in 1985. Mark Cunniff (NACJP) had selected 31 jurisdictions that he wanted to survey. Those 31 jurisdictions include 22 from the Top 75 jurisdictions and 9 from the balance of the jurisdictions. All of these jurisdictions were treated as

certainty jurisdictions. Other jurisdictions in the Top 75 were added to certainty if their total number of convictions in 1985 exceeded 5500.

The noncertainty jurisdictions were stratified as follows: the remainder of the Top 75, balance-large (1500-5499 felony convictions), balance-medium (250-1499 felony convictions), and balance-small (< 250 felony convictions). The optimum allocation (based on dispositions) to the 4 strata is given in Table 2.

Table 2. First-Stage Stratification of the
Survey of Felony Convictions

Stratum	Number of Jurisdictions (M)	Sample Number of Jurisdictions (m)	First-Stage Weight (W)
Certainty:			
(1) Top 75	31	31	1.0000
(2) Balance	10	10	1.0000
Noncertainty:			
(3) Top 75	44	23	1.9130
Balance: (4) Large	86	13	6.6154
(5) Medium	438	9	48.6667
(6) Small	2500	14	178.5714
	3109	100	

The file was sorted within each of the strata with jurisdictions that are automated listed first by region, and within region, by the number of convictions. A systematic sample using the prescribed take rate was selected within each of the four noncertainty strata. One of the 10 certainty cases in the balance of jurisdictions refused to participate in the survey. Jurisdiction number 12105 with similar characteristics was substituted for jurisdiction 47037. The selected jurisdictions are given in the May 1, 1987 memorandum entitled "Sample for the Survey of Felony Convictions." In that memorandum the selected units are identified by their FIPS codes.

3. Second Stage of Sampling

At the second stage of sampling, all courts within the selected first-stage jurisdictions were selected for inclusion in the sample. A systematic sample of homicides, rapes, robberies, etc. was selected from each jurisdiction's court records. For each crime category, a sample size was calculated using a coefficient of variation of .013 for the top 75 estimate of the percentage of convictions receiving a prison sentence, and a coefficient of variation of about .015 for the corresponding national estimate. The number of convictions per crime type was not known. An estimate of the percentage distribution was available from previous data. The distribution, which is given in Table 1, was applied to the total number of convictions in 1985 in order to get estimates of convictions by crime type. The following equation yielded the sample size needed to obtain the prescribed coefficients of variation.

$$n_h = \frac{\Delta p_h q_h / CV^2}{1 + (\frac{\Delta p_h q_h}{CV^2} - 1)/\hat{N}_h} \tag{1}$$

where p_h is an estimate of the percentage of convictions of crime type h that receive prison sentences. (These are given in Table 1.)

q_h is $1-p_h$.

$\hat{N}_h$ is the estimated number of convictions in crime type h.

Δ is the design effect to account for the clustering in the jurisdictions. (It was assumed to be 2.25 for this design.)

CV   is the prescribed coefficient of variation.

The overall second-stage sampling rate, f_h, for each group is calculated by dividing n_h by $\hat{N}_h$. Stratum sampling rates for the second stage are calculated by dividing f_h by the first-stage selection probability, $p_j = m_j/M_j$ where m_j is the number of sample jurisdictions and M_j is the total number of jurisdictions in stratum j. Ideally, the inverse of f_h/p_j gives the take rate

for crime type h, stratum j. (The rate was rounded conservatively so that it would yield an adequate sample size.) In each jurisdiction's court, a systematic sample of each crime type is taken.

Preliminary second-stage take rates were calculated in this manner, but after GOVS examined records from their first test site in York, Pennsylvania, they found that a substantial portion of the total convictions count actually contained misdemeanors. (In some cases, the original charge was a misdemeanor; in other cases, the original charge was a felony but was reduced to a misdemeanor.) If this trend continued across the country, the preliminary sampling rates would have yielded samples much smaller than the samples needed to meet the error requirements. Consequently, the estimated number of convictions per crime type was decreased by 20 percent and used this as the value of $\hat{N}_h$. New take rates were calculated and these appear in Table 3.

Table 3. Take Rates and Random Starts
for Second Stage, by Crime Type

1st Stage Stratum	Homicide	Rape	Robbery	Assault	Burglary	Larceny	Drugs	Other
			Take Every					
Certainties:								
Top 75	2	2	8	5	15	18	15	47
Balance	11	7	36	24	69	87	72	226
Noncertainties:								
Top 75	1	1	4	3	7	9	8	24
Bal: Large	1	1	5	3	10	13	10	34
Med.	1	1	1	1	1	1	1	4
Small	1	1	1	1	1	1	1	1
			Start With					
Certainties:								
Top 75	2	1	2	1	13	1	3	12
Balance	7	2	30	14	39	18	5	55
Noncertainties:								
Top 75	1	1	3	1	3	8	5	7
Bal: Large	1	1	1	1	3	8	5	5
Med.	1	1	1	1	1	1	1	1
Small	1	1	1	1	1	1	1	1

Mark Cunniff was responsible for collecting data from 40 (his original 31 plus 9 noncertainty jurisdictions) of the 100 jurisdictions. For his purposes, he needed to collect more data than the take rates yielded. He sampled much more heavily, and in the certainty jurisdictions, we used his full sample (appropriately weighting the selected units). In the 9 noncertainty jurisdictions, he provided a subsample of his sample, thus yielding the take rates we prescribed.

Governments Division/Field Division took a systematic sample in the other 60 counties using the prescribed take rates. In York, Pennsylvania, the jurisdiction selected for pretest operations, the preliminary "balance-large" stratum take rates were used since we found 400 (after eliminating misdemeanor charges, misdemeanor convictions, and duplicate defendants) felonies instead of the predicted 1500 felonies. (York, Pennsylvania was originally sampled in the "large" stratum during the first stage.) Another jurisdiction, Marlboro, South Carolina, was changed to take all felony cases because of the large number of misdemeanors included with the felonies. The revised second-stage weights indicate that we should have taken 1 in 4 of the "other" category, but homicide, rape, etc. were all sampled at the correct rates in Marlboro. In Shelby, Tennessee, which is in the certainty stratum, GOVS decided to use a sample that had been selected by a member of the Tennessee Sentencing Commission. The sample was a systematic selection of every third felony conviction. Sampling was not done within crime type, but theoretically the sample should yield a 1-in-3 sample for each type. Since this rate does not meet the requirements for a top 75 certainty for rape and homicide, GOVS will take all rapes and homicides with certainty using supplemental data collection. A weight of 1.00 will be used for rapes and homicides, while a weight of 3.00 will be used for all other crime types.

4. Estimation

4.1 Totals

Jurisdiction estimates of total by crime type are calculated by summing the value of the characteristic over all records selected in the sample and multiplying that by the take every, as given in equation (2).

$$\hat{X}_{hji} = TE_{hji} \sum_{k=1}^{n_{hji}} X_{hjik} \qquad (2)$$

where

X_{hjik} = value of X for defendant k in jurisdiction i of stratum j in crime category h;

n_{hji} = number of defendants in the sample in stratum j, jurisdiction i, crime class h;

TE_{hji} = the second-stage take-every for jurisdiction i of stratum j in category h.

In order to estimate a national total of any characteristic for a crime-type category, $\hat{X}_h$, equation (3) should be used. The first-stage weight is multiplied by each jurisdiction total and summed.

$$\hat{X}_h = \sum_{j=1}^{6} \sum_{i=1}^{m_j} W_j \hat{X}_{hji} \qquad (3)$$

where

$\hat{X}_{hji}$ = weighted value of X recorded in jurisdiction i of stratum j in crime category h as calculated in equation (2).

m_j = number of selected sample jurisdictions in stratum j (= the total number of jurisdictions in the certainty strata);

W_j = first-stage sampling weight, inverse of the probability of selection at the first stage. These first-stage weights are given in Table 2 for strata 1 through 6.

In order to make an estimate of total for the Top 75 only, all sums would involve only jurisdictions in the Top 75. The appropriate weights would be W_1 for jurisdictions in stratum 1 and W_3 for jurisdictions in stratum 3.

For a single jurisdiction, a total across all categories is obtained by summing jurisdiction totals across all 8 crime categories. In order to obtain a national total over all categories, $\hat{X}_h$ would be calculated for each category and summed across all categories to yield $\hat{X}$. Top 75 estimates would be done like the national estimates.

4.2 Ratios

In order to estimate ratios for single jurisdictions, like the ratio of felony convictions that receive a prison sentence, equation (4) should be used.

$$r_{hji} = \frac{\hat{Y}_{hji}}{\hat{Z}_{hji}} \tag{4}$$

where $\hat{Y}_{hji}$ and $\hat{Z}_{hji}$ are estimated using equation (2) for the two totals of interest. For example, $\hat{Y}_{hji}$ would be the number of convicted felons of crime type h receiving a prison sentence, and $\hat{Z}_{hji}$ would be the estimated total number of convicted felons.

In order to make an estimate for the Top 75, totals involve only sums using sample units from the 75 largest jurisdictions. The national estimate of a ratio is given in equation (5).

$$r_h = \hat{Y}_h / \hat{Z}_h \tag{5}$$

where $\hat{Y}_h$ and $\hat{Z}_h$ are defined as the characteristics of interest and estimated as in equation (3).

In order to make estimates of the average length of prison sentence, the above estimates, r_h and r_{hji} can be used with $\hat{Y}_{hji} = TE_{hji} \sum_{k=1}^{n_{hji}} Y_{hjik}$ with Y_{hjik} being the prison sentence for defendant R of jurisdiction i of stratum j in crime category h and with $\hat{Z}_{hji}$ being $TE_{hji} n_{hji}$.

In order to estimate the proportion of felony convictions, for example, that are of crime-type category h, equation (6) is used for the estimate:

$$\hat{P}_h = \hat{X}_h / \sum_{l=1}^{8} \hat{X}_h \qquad (6)$$

where $\hat{X}_h$ is estimated using equation (2).

The Top 75 estimates are calculated in a manner similar to equation (6), using totals for the Top 75 only. For individual jurisdictions, the proportion of felony convictions, for example, of crime type j is estimated by

$$\hat{P}_{hji} = \hat{X}_{hji} / \sum_{h=1}^{8} \hat{X}_{hji} \qquad (7)$$

where $\hat{X}_{hji}$ was defined in equation (2).

4.3 Summary of the Estimation Formulas:

Total for crime class -

jurisdiction:
$$\hat{X}_{hji} = TE_{hji} \sum_{k=1}^{n_{hji}} X_{hjik}$$

Top 75:
$$\hat{X}_{hT} = \sum_{j=1,3} W_j \hat{X}_{hji}$$

national:
$$\hat{X}_h = \sum_{j=1}^{6} W_j \hat{X}_{hji}$$

Total over all crime classes -

jurisdiction:
$$\hat{X}_{\cdot ji} = \sum_{h=1}^{8} \hat{X}_{hji}$$

Top 75:
$$\hat{X}_{\cdot T} = \sum_{h=1}^{8} \hat{X}_{hT}$$

national:
$$\hat{X} = \sum_{h=1}^{8} \hat{X}_h$$

Ratio for crime class -

jurisdiction:
$$r_{hji} = \hat{Y}_{hji}/\hat{Z}_{hji}$$

Top 75:
$$r_{hj} = \hat{Y}_{hT}/\hat{Z}_{hT}$$

national:
$$r_h = \hat{Y}_h/\hat{Z}_h$$

Ratio over all crime classes −

jurisdiction: $\quad \hat{R}_{ji} = \hat{Y}_{\cdot ji}/\hat{Z}_{\cdot ji}$

Top 75: $\quad \hat{R}_T = \hat{Y}_{\cdot T}/\hat{Z}_{\cdot T}$

national: $\quad \hat{R} = \hat{Y}/\hat{Z}$

Proportion of cases in crime class −

jurisdiction: $\quad \hat{P}_{hji} = \hat{X}_{hji}/\hat{X}_{\cdot ji}$

Top 75: $\quad \hat{P}_{hT} = \hat{X}_{hT}/\hat{X}_{\cdot T}$

national: $\quad \hat{P}_h = \hat{X}_h/\hat{X}$

5. Variance Estimation

5.1 Variance Estimates of Totals

5.1.1 Jurisdiction Estimates

For a single jurisdiction, the variance of the estimate of total (given in section 4) is

$$S^2_{hjiX} = \frac{N_{hji}}{n_{hji}} \frac{(N_{hji} - n_{hji})}{(N_{hji}-1)} \sum_{k=1}^{N_{hji}} (X_{hjik} - \overline{X}_{hji})^2 \qquad (8)$$

where $\quad \overline{X}_{hji} = \sum_{k=1}^{N_{hji}} X_{hjik}/N_{hji}$

$X_{hjik} =$ the value of characteristic X for defendant k in crime class h, stratum j, jurisdiction i.

N_{hji} = the total number of defendants in stratum j, jurisdiction i, crime class h.

n_{hji} = number of defendants in the sample in crime class h, stratum j, jurisdiction i.

X_{hjik} = value of the characteristic X for defendant k of jurisdiction i, stratum j, crime class h. X_{hjik} usually has a value of 0 or 1. When the characteristic X is number of years of prison sentence, X_{hjik} = 0 if defendant k was not sentenced and the number of years of the sentence otherwise.

The variance is estimated by s^2_{whjiX} in (9), assuming n_{hji} is large.

$$s^2_{whjiX} = TE_{hji}(TE_{hji}-1) \left[\sum_{k=1}^{n_{hji}} X^2_{hjik} - \left(\sum_{k=1}^{n_{hji}} X_{hjik} \right)^2 / n_{hji} \right] \qquad (9)$$

For an estimate of a jurisdiction total across all crime classes, the following equation gives the variance:

$$S^2_{\cdot jiX} = \sum_{h=1}^{8} S^2_{hjiX} \qquad (10)$$

where S^2_{hjiX} is given in (8).

The estimated variance is given by

$$s^2_{w \cdot jiX} = \sum_{h=1}^{8} s^2_{whjiX} \qquad (11)$$

5.1.2 Top 75 Estimates

Estimates for the Top 75 are made from the Top 75 certainties and the balance of Top 75 noncertainty stratum. The variance arises from two sources: (1) variation due to sampling convictions within a jurisdiction (which arises from both certainty and noncertainty jurisdictions) and (2) variations between noncertainty jurisdictions.

For the Top 75 certainty jurisdictions, the total variance is given by

$$\sum_{i=1}^{M_1} S_{h1iX}^2 \tag{12}$$

This variance is estimated by

$$s_{wh1X}^2 = \sum_{i=1}^{m_1} s_{wh1iX}^2 \tag{13}$$

where $M_1 = m_1$ is the number of jurisdictions in the Top 75 certainty stratum, and

s_{wh1iX}^2 is defined in equation (9) for j=1.

An estimate of the variance for the noncertainty stratum is given by the sum of the between jurisdiction and within jurisdiction components. The within component for the noncertainty stratum is estimated like equation (13) and is given in equation (14).

$$s_{wh3X}^2 = W_3 \sum_{i=1}^{m_3} s_{wh3iX}^2 \tag{14}$$

where m_3 is the number of selected Top 75 noncertainty jurisdictions;

W_3 is the first-stage weight for stratum 3, and

s_{wh3iX}^2 is calculated using equation (9) for stratum 3.

For Top 75 estimates of total, the within component of variance is the sum of the certainty and noncertainty within components. The between

jurisdiction component for the Top 75 noncertainty jurisdictions is given in equation (15).

$$s^2_{bhX} = W_3(W_3-1) \, s^2_{h3X} \tag{15}$$

where

$$s^2_{hjX} = m_j \left[\sum_{i=1}^{m_j} TE^2_{hji} \, X^2_{hji} - \left(\sum_{i=1}^{m_j} TE_{hji} \, X_{hji} \right)^2 / m_j \right] / (m_j-1) \tag{15a}$$

s^2_{h3X} is equation (15a) for stratum 3.

All other variables have been defined previously.

In summary, the estimate of variance for a Top 75 estimate of total within a crime type h is the sum of the within components of variance from the certainty and noncertainty strata and the between component.

In order to estimate the variance of a total across all crime types, the variances for each crime type are added. (Sampling was done independently within each crime type.)

5.1.3 National Totals

Estimates for variances of national totals are calculated in a manner similar to the variance estimates for the Top 75 totals. All sums are over all jurisdictions or strata rather than over the Top 75 jurisdictions. All strata are included; the Top 75 has only one noncertainty stratum.

As with the Top 75 certainties, the variance for the certainty jurisdictions is estimated by

$$s^2_{whCX} = \sum_{j=1}^{2} \sum_{i=1}^{m_j} s^2_{whjiX} \tag{16}$$

All variables are as defined in equation (9).

From page 303 of Cochran (1977), the variance for the noncertainty strata is estimated by equation (17), assuming n_{hji} is large.

$$s_{hNCX}^2 = \sum_{j=3}^{6} W_j(W_j-1)\, s_{hjX}^2 + \sum_{j=3}^{6} W_j \sum_{i=1}^{m_j} s_{whjiX}^2 \tag{17}$$

where s_{hjX}^2 is given in equation (15a);

s_{whjiX}^2 is given in equation (9); and

W_j is the first-stage weight.

The estimated total variance is calculated by adding equations (16) for the certainty strata and (17) for the noncertainty strata. The variances of totals over all crime types are obtained by adding the variances for each crime type since sampling was done independently within each crime type.

5.2   Variances of Ratios

5.2.1 Jurisdiction Estimates

The variance of the ratio as given in equation (5) can be estimated by

$$v(r_{hji}) = \frac{1}{Z_{hji}^2}\left(s_{whjiY}^2 + r_{hji}^2\, s_{whjiZ}^2 - 2\, r_{hji}\, s_{whjiYZ}\right) \tag{18}$$

where s_{whjiY}^2 and s_{whjiZ}^2 are defined as in equation (9) with Y, the numerator of R, and Z, the denominator of R, substituted for X.

$$s_{whjiYZ} = TE_{hji}(TE_{hji}-1)[\sum_{k=1}^{n_{hji}} Y_{hjik} Z_{hjik} -$$

$$(\sum_{k=1}^{n_{hji}} Y_{hjik})(\sum_{k=1}^{n_{hji}} Z_{hjik})/n_{hji}] \tag{18a}$$

All other variables were previously defined.

The variance of a ratio of two variables over all crime classes is estimated by

$$v(\hat{R}_{ji}) = \frac{1}{Z_{\cdot ji}^2}(s_{w\cdot jiY}^2 + \hat{R}_{ji}^2 \ s_{w\cdot jiZ}^2 - 2\hat{R}_{ji} \ s_{w\cdot jiYZ}) \tag{19}$$

where $s_{w\cdot jiY}^2$ and $s_{w\cdot jiZ}^2$ are defined as in equation (11).

$$s_{w\cdot jiYZ} = \sum_{h=1}^{8} s_{whjiYZ} \tag{19a}$$

s_{whjiYZ} is given in equation (18a).

5.2.2 Top 75 Estimates

Within a crime class, the variance of the ratio of Y to Z for two Top 75 estimates of totals can be estimated as follows. Again, there are two major components of variance; the within component, which arises from sampling convictions within a jurisdiction; and a between component, which arises from sampling jurisdictions and does not receive a contribution from the certainty jurisdictions. The within component is estimated as follows:

$$s^2_{whTr} = \sum_{i=1}^{m_1} \frac{1}{\hat{Z}^2_{hT}} \left(s^2_{wh1iY} + r^2_{hT} \, s^2_{wh1iZ} - 2 \, r_{hT} \, s_{wh1iYZ}\right)$$

$$+ W_3 \sum_{i=1}^{m_3} \frac{1}{\hat{Z}^2_{hT}} \left(s^2_{wh3iY} + r^2_{hT} \, s^2_{wh3iZ} - 2 \, r_{hT} \, s_{wh3iYZ}\right) \tag{20}$$

where s^2_{wh3iY} and s^2_{wh3iZ} are given in equation (9). s_{wh3iYZ} is given in equation (18a). These estimates come from the Top 75 noncertainty jurisdictions.

s^2_{wh1iY} and s^2_{wh1iZ} are given in equation (9). s^2_{wh1iYZ} is given in equation (18a). These estimates are for the Top 75 certainty jurisdictions.

r_{hT} and $\hat{Z}_{hT}$ are defined in section 4.2. They are Top 75 estimates.

All other variables were defined previously.

The total variance is estimated by the following:

$$v(r_{hT}) = \{W_3(W_3-1)[s^2_{h3Y} + r^2_{hT} \, s^2_{h3Z} - 2r_{hT}s_{h3YZ}]\}/\hat{Z}^2_{hT} + s^2_{whTr} \tag{21}$$

where s^2_{h3Y} and s^2_{h3Z} are given in equation (15a), but for stratum 3;

$$s_{hjYZ} = m_j \left[\sum_{i=1}^{m_j} TE^2_{hji} \, Y_{hji} Z_{hji} \right.$$

$$\left. - \left(\sum_{i=1}^{m_j} TE_{hji} \, Y_{hji}\right)\left(\sum_{i=1}^{m_j} TE_{hji} \, Z_{hji}\right)/m_j \right]/(m_j-1) \tag{21a}$$

s_{h3YZ} is s_{hjYZ} for stratum 3.

All other variables have been defined previously.

For a ratio of totals over all crime classes, the following equation is used to get an estimate of the within component of variance.

$$s^2_{wTR} = \frac{1}{\hat{Z}^2_{.T}} \left\{ \sum_{i=1}^{m_1} \left(s^2_{w \cdot 1iY} + \hat{R}^2_T s^2_{w \cdot 1iZ} - 2\hat{R}_T s_{w \cdot 1iYZ} \right) \right.$$

$$\left. + W^2_3 \sum_{i=1}^{m_3} \left(s^2_{w \cdot 3iY} + \hat{R}^2_T s^2_{w \cdot 3iZ} - 2 \hat{R}_T s_{w \cdot 3iYZ} \right) \right\} \tag{22}$$

where

$s^2_{w \cdot 1iY}, s^2_{w \cdot 1iZ}, s^2_{w \cdot 3iY}, s^2_{w \cdot 3iZ}$ are defined as in equation (11);

$s_{w \cdot 3iYZ}$ and $s_{w \cdot 1iYZ}$ are defined in equation (19a) with $j=1$ for the Top 75 certainty jurisdictions and $j=3$ for the Top 75 noncertainty jurisdictions; and

$$\hat{R}_T = \hat{Y}_T / \hat{Z}_T$$

All other variables have been defined previously.

The total variance can be estimated by

$$v(\hat{R}_T) = \frac{1}{\hat{Z}^2_{.T}} \left\{ W_3(W_3-1) \sum_{h=1}^{8} \left(s^2_{h3Y} + \hat{R}^2_T s^2_{h3Z} - 2\hat{R}_T s_{h3YZ} \right) \right.$$

$$\left. - W_3 (W_3-1) \sum_{i=1}^{m_3} \left(s^2_{w \cdot 3iY} + \hat{R}^2_T s^2_{w \cdot 3iZ} - 2 \hat{R}_T s_{w \cdot 3iYZ} \right) \right\} + s^2_{wTR} \tag{23}$$

where $s^2_{h3Y}, s^2_{h3Z},$ and s_{h3YZ} were defined in equation (21);

$s^2_{w \cdot 3iY}, s^2_{w \cdot 3iZ},$ and $s_{w \cdot 3iYZ}$ were discussed in equation (22); and s^2_{wTR} is given in equation (22).

All other variables have been defined previously.

5.2.3 <u>National Estimates</u>

National estimates are made in the same way as the Top 75 estimates of variance, but all sums extend over all jurisdictions rather than just the Top 75 jurisdictions.

The variance of r_h is estimated by the following for each crime type:

$$v(r_h) = \{ \sum_{j=1}^{2} \sum_{i=1}^{m_j} (s_{whjiY}^2 + r_h^2 s_{whjiZ}^2 - 2 r_h s_{whjiYZ})$$

$$+ \sum_{j=3}^{6} W_j \sum_{i=1}^{m_j} (s_{whjiY}^2 + r_h^2 s_{whjiZ}^2 - 2 r_h s_{whjiYZ})$$

$$+ \sum_{j=3}^{6} W_j (W_j - 1) [s_{hjY}^2 + r_h^2 s_{hjZ}^2 - 2 r_h s_{hjYZ}]\}/\hat{Z}_h^2 \qquad (24)$$

where s_{whjiY}^2 and s_{whjiZ}^2 are defined in equation (9);

s_{whjiYZ} is defined in equation (18a);

s_{hjY}^2 and s_{hjZ}^2 are defined in equation (15a); and

s_{hjYZ} is defined in equation (21a).

The variance of $\hat{R}$ over all crime types is estimated by

$$v(\hat{R}) = \{ \sum_{j=1}^{2} \sum_{i=1}^{m_j} (s^2_{w \cdot jiY} + \hat{R}^2 s^2_{w \cdot jiZ} -2 \hat{R} s_{w \cdot jiYZ})$$

$$+ \sum_{j=3}^{6} W_j \sum_{i=1}^{m_j} (s^2_{w \cdot jiY} + \hat{R}^2 s^2_{w \cdot jiZ} -2 \hat{R} s_{w \cdot jiYZ})$$

$$+ \sum_{j=3}^{6} W_j(W_j-1) \sum_{h=1}^{8} (s^2_{hjY} + \hat{R}^2 s^2_{hjZ} -2 \hat{R} s_{hjYZ})\}/\hat{Z}^2 \qquad (25)$$

where $s^2_{w \cdot jiY}$ and $s^2_{w \cdot jiZ}$ are defined as in equation (11); and

$s_{w \cdot jiYZ}$ is defined in equation (19a).

All other variables have been defined previously.

5.3 Variance Estimates of Proportions

5.3.1 Jurisdiction Estimates

For a single jurisdiction, the variance of the estimated proportion of cases in a crime class can be estimated by

$$v(\hat{P}_{hji}) = ((1-2 \hat{P}_{hji}) s^2_{whjiX} + \hat{P}_{hji}^2 s^2_{w \cdot jiX})/\hat{X}^2_{\cdot ji} \qquad (26)$$

where s^2_{whjiX} is defined in equation (9);

$\hat{X}_{\cdot ji}$ and $\hat{P}_{hji}$ are defined in section 4.3; and

$s^2_{w \cdot jiX}$ is defined in equation (11).

5.3.2 Top 75 Estimates

The Top 75 estimates of the variance of the proportion of cases in a crime class can be estimated by

$$v(\hat{P}_{hT}) = \{ \sum_{i=1}^{m_1} ((1-2\,\hat{P}_{hT})\, s^2_{wh1iX} + \hat{P}^2_{hT}\, s^2_{w\cdot 1iX})$$

$$+ W_3 \sum_{i=1}^{m_3} ((1-2\,\hat{P}_{hT})\, s^2_{wh3iX} + \hat{P}^2_{hT}\, s^2_{w\cdot 3iX})$$

$$+ W_3(W_3-1)\, [(1-2\,\hat{P}_{hT})\, s^2_{h3X} + \hat{P}^2_{hT} \sum_{h=1}^{8} s^2_{h3X}]\}/\hat{X}^2_{\cdot T} \qquad (27)$$

where

$\hat{X}_{\cdot T}$ and $\hat{P}_{hT}$ are defined in section 4.3;

s^2_{wh1iX} and s^2_{wh3iX} are defined in equation (9);

$s^2_{w\cdot 1iX}$ and $s^2_{w\cdot 3iX}$ are defined in equation (11); and

s^2_{h3X} is defined in equation (15a).

5.3.3 National Estimates

National estimates of the variances of proportions are calculated in a manner similar to the Top 75 estimates of variance.

$$v(\hat{P}_h) = \{ \sum_{j=1}^{2} \sum_{i=1}^{m_j} ((1-2\,\hat{P}_h)\, s^2_{whjiX} + \hat{P}^2_h\, s^2_{w\cdot jiX})$$

$$+ \sum_{j=3}^{6} W_j \sum_{i=1}^{m_j} ((1-2\,\hat{P}_h)\, s^2_{whjiX} + \hat{P}^2_h\, s^2_{w\cdot jiX})$$

$$+ \sum_{j=3}^{6} W_j(W_j-1)\, [(1-2\,\hat{P}_h)\, s^2_{hjX} + \hat{P}^2_h \sum_{h=1}^{8} s^2_{hjX}]\}/\hat{X}^2 \qquad (28)$$

where

$\hat{X}$ and $\hat{P}_h$ are defined in section 4.3;

s_{whjiX}^2 is defined in equation (9);

$s_{w \cdot jiX}^2$ is defined in equation (11); and

s_{hjX}^2 is defined in equation (15a).

5.4 Standard Errors and Relative Standard Errors

In all cases, the standard error is the square root of the estimated variance. The relative standard error is the standard error divided by the appropriate estimate.

6. Considerations for Future Redesigns

In designing the full sample for the survey of felons sentenced in 1988, the 1986 pretest data should be used to update our estimates of the percentage of convictions that receive sentences per crime type. The pretest data should also be used to give a better estimate of the percentage of "felony convictions" that are actually misdemeanors. For the pretest, we assumed that 20 percent of the "convictions" were actually misdemeanors. For 1988, more accurate estimates should be available. Separate estimates should be made for the Top 75 and the balance. For 1988, we should probably try to update the 1985 Directory Survey of General Jurisdiction Courts. If a new sample is to be selected every 5-10 years, an update to the universe listing (including updates to filings, dispositions, and convictions) should be done prior to sampling. Definitions of exactly what we want (the number of defendant-based

felony convictions, dispositions, and filings) should be clearly stated on the
questionnaire. If possible, data on convictions by crime type should be
sought. From the pretest, we should also examine the methods of reporting
convictions (defendant-based, charge-based, indictment-based, or a mixture) to
investigate the effects of the mixed reporting on developing the sample
design.

For the sample design, the following data will be needed to make the
required adjustments to the 1985 data:

(1)   Total misdemeanors by jurisdiction;

(2)   Total felonies by jurisdiction;

(3)   National and top 75 estimates of the percentage of convictions
      receiving prison sentences, by crime class;

(4)   National and top 75 distributions of the percentage of convictions
      by crime class;

(5)   For all 100 jurisdictions, 1985 convictions, filings, and
      dispositions (including the imputed values, properly flagged), and
      the 1986 convictions, along with filings and dispositions if they
      were collected; and

(6)   1986 major estimates and their relative standard errors, by crime
      class, for the Top 75 and for the national estimates.

If the research shows that adjustments to the 1985 data are needed or if more
noncertainty strata are needed, sums and sums of squares will be needed for
each stratum.

References

Cochran, William G. (1977).   Sampling Techniques, third edition, John Wiley &
      Sons Inc., New York

APPENDIX

Criminal Justice Terminology

(All definitions are from the Dictionary of Criminal Justice Data Terminology,
Second edition, Bureau of Justice Statistics, SEARCH Group Inc., 1981).

conviction –
The judgment of a court, based on the verdict of a jury or judicial officer, or on the guilty plea or nolo contendre plea of the defendant, that the defendant is guilty of the offense(s) with which he or she has been charged.

defendant disposition –
The class of prosecutorial or judicial actions which terminate or provisionally halt proceedings regarding a given defendant in a criminal case after charges have been filed in court.

felony –
A criminal offense punishable by death, or by incarceration in a prison facility.

filing –
The initiation of a criminal case by formal submission to the court of a charging document, alleging that one or more named persons have committed one or more specified criminal offenses.

misdemeanor –
An offense punishable by incarceration, usually in a local confinement facility, for a period of which the upper limit is prescribed by statute in a given jurisdiction, typically limited to a year or less.