STANDARD ERROR ESTIMATION FOR THE NATIONAL CRIME SURVEY

David Bateman and Paul Bettin, U.S. Bureau of the Census

I. Introduction

The primary purpose of this paper is to present the method used to estimate standard errors for the National Crime Survey. Formulas for estimating both within and between-PSU variances are also illustrated. Section II describes the sample design and the estimation procedure. Section III describes the standard error estimators used for various types of estimates.

It was our original intent that this paper would emphasize the numerical results as compared to the method of computing the results. However, because of the complexity of the task, we have been unable to complete the computer programming necessary to produce these results. We intend to write another paper presenting these results when they become available.

The standard errors that are presently being used for analysis purposes for this survey were obtained by using data from other surveys to estimate the design effects expected for NCS. Standard errors were then produced by applying these design effects to the expected simple random sample standard errors.

II. Nature of the Survey, Sample Design, and Estimation Procedure for the National Crime Survey

A. Purpose of the Survey

The National Crime Survey is a sample survey conducted quarterly by the U.S. Bureau of the Census for the Law Enforcement Assistance Administration of the U.S. Department of Justice. The purpose of the survey is to ascertain the extent to which persons aged 12 and over and households have been victimized by certain types of crime and to study the character and nature of criminal incidents and their victims.

B. Sample Design

The sample design is a stratified multistage cluster design. The primary sampling units (PSU's) were formed from counties or groups of contiguous counties, using every county in the entire United States. Approximately 1,930 PSU's were so created and form the universe from which the sample PSU's were selected. These PSU's were the same ones created for the Current Population Survey (CPS). Prior to selecting the sample PSU's, the 1,930 PSU's were grouped into 376 strata. One hundred and fifty-six of the strata consist of only one PSU and these type PSU's are called self-representing (SR). The remaining 220 strata were formed by combining PSU's with similar characteristics such as geographic region, population density, rate of growth in the 1960-1970 decade, proportion nonwhite, principal industry, etc. These characteristics were selected because they tend to be related to a large number of demographic characteristics. The strata were formed so that their 1970 population sizes were approximately equal. From these 220 strata, one PSU was selected per stratum with probability proportionate to the size of PSU. These type PSU's are called non-self-representing (NSR). The 376 selected PSU's comprise the first stage of sampling (220 NSR PSU's and 156 SR PSU's).

The objectives of the remaining stages of sampling were to obtain a self-weighting probability sample of approximately 75,000 households. This is obtained by adjusting accordingly the rate of sampling within each PSU. The sample of households within a selected PSU is selected in two stages. The first stage involves the selection of enumeration districts (ED's), geographic areas used for the 1970 Census that usually have well-defined boundaries and contain, on the average, about 300 households. ED's are selected systematically with probability proportionate to their 1970 population. Before selection the ED's are arranged in a predetermined geographic manner. The selected ED's are then subdivided into segments or compact clusters of about four housing units and a sample of these segments is then taken. Wherever possible, the segments were formed from the list of addresses compiled during the 1970 Census. Where the list of addresses was incomplete or inaccurate, area sampling methods were used to select the sample. The address lists were used in about two-thirds of the ED's, these being in primarily urban areas. Area sampling techniques were applied to the remaining ED's.

Those units built after the 1970 Census was conducted that were not included in the above sampling process were sampled primarily from a list of new construction building permits issued from permit issuing offices in the sampled areas. This sampling is an independent sampling operation. Those areas that are not permit issuing are sampled for new construction by means of a sample of area segments. The resulting sample of new construction units is a small part of the total sample, increasing, however, as the decade progresses.

C. Rotation Scheme

Because of the concern for respondent fatigue, a rotation scheme is used for this survey. The sample of 75,000 households is divided into six groups or rotations. Once the rotation is fully operative, households in each rotation group will be interviewed once every six months for three and one-half years, the initial interview being only for purposes of bounding, i.e., establishing a time frame to avoid recording duplicative reports on subsequent visits. Each rotation group is further divided into six panels. One-sixth of each rotation group, or one panel, will be interviewed each month during the six month period. Additional samples of 75,000 households selected in the above manner will be similarly assigned to rotation groups and panels for subsequent rotation into sample. One rotation group will enter the sample every six months and the corresponding rotation group from a previous sample will be phased out.

The assignments of rotation group numbers and panel numbers to the 75,000 sample housing units were made to complete segments of housing units. The assignments were made with the following objectives:

1. Each panel should be a systematic one-sixth subsample of the 75,000 housing units.
2. Each rotation group should be a systematic one-sixth subsample of the 75,000 sample housing units, as well as a one-sixth subsample of the sample interviewed each month-i.e., each panel.

3. It should be possible to combine rotation groups and panels to form subsamples for variance estimation purposes.

D. Estimation Procedure

The estimation procedure makes extensive use of available auxiliary data on characteristics of the survey population correlated to those characteristics to be investigated in the survey in order to obtain more reliable estimates. This auxiliary data is used in the adjustments for households eligible for interview, but which were not interviewed, and in the various stages of ratio estimation employed. The estimation procedure is performed on a quarterly basis to produce quarterly estimates of total crime victimizations and of the rate of crime victimization. Sample data from eight months of interviewing are required to produce a quarterly estimate. For example, the months of Februrary through September are required to produce an estimate for the first quarter of a year.

This interviewing scheme utilizing a six-month recall period enables one to obtain extremely reliable results from a relatively small sample size. Biases can occur due to telescoping (placing the crime victimizations in more recent months than when they actually occurred), however, it is felt that these biases are offset by using this interviewing scheme.

The estimation procedure is made up of several stages that are described below:

1. Adjustments for noninterviews: The estimates for a quarter are produced by first inflating the sample data results by the reciprocal of the probability of selection. This would include the sample data from all eight months required to produce a quarterly estimate. Those sample units that are eligible for interview, but which were not interviewed, are accounted for by computing the following ratio adjustment factor in cells where it is thought that the interviewed and non-interviewed sample units are positively correlated.

$$\text{Ratio} = \frac{\text{Inflated Interviewed Sample Units+ Inflated Non-interviewed Sample Units}}{\text{Inflated Interviewed Sample Units}}$$

These ratios are then applied to the inflated sample data for the interviewed units in the respective noninterview cells.

For person characteristics noninterview adjustment factors were calculated for both persons missed within enumerated households and for persons missed because of entire households being missed. For household characteristics, noninterview adjustment factors were calculated for missed households only.

2. First Stage Ratio Estimation: The purpose of the first-stage ratio estimate is to reduce the contribution to the variance arising from the sampling of primary sampling units (PSU's) in nonself-representing (NSR) strata. These ratios are based on 1970 Census data and are applied only to sample data from the NSR PSU's. The ratios are computed in 48 cells determined by region, race and residence. The ratios take the form of the 1970 Census counts of persons in NSR strata for the region, race and residence cells divided by the sample estimates of persons in these cells. These sample estimates of persons are obtained by inflating the 1970 population of the NSR PSU's selected for the sample to the stratum levels.

3. Second Stage Ratio Estimation: The second stage ratio estimation procedure brings the distribution of the sample population into closer agreement with the entire population with respect to characteristics of age, sex, and color. Estimates are ratioed to independent totals prepared each month by carrying forward the most recent census data to take account of subsequent aging of the population, mortality, and migration between the U.S. and other countries. This ratio estimation is performed in 72 age, sex, color cells.

4. "Principal Person" and Third Stage of Ratio Estimation for Estimates of Household Victimization

The first and second stage ratio estimation is performed on a person basis and applied to sample estimates of persons. The estimation procedure for estimates of households involves, in addition to the two stages of ratio estimation for persons, a third stage of ratio estimation. This third stage adjusts the sample estimate of occupied housing units in four residence, tenure categories to independently derived current estimates of occupied housing units in each of four categories. The four categories are Urban-owner, Urban-renter, Rural-owner, and Rural-renter.

These independent current estimates of occupied housing units are obtained by a regression estimation procedure using past months of CPS data on the number of occupied housing units. The sample estimate of occupied housing units for each of the third stage cells is derived by means of a "principal person" type of estimation procedure. This involves designating one person in a household as a "principal person" and effectively obtaining an estimate of the number of these "principal persons" by using the first two stages of ratio estimation. This estimate is then used as the estimate of occupied housing units for the third stage of ratio estimation. The "principal person" is defined as:

i. The wife in a husband-wife household, and
ii. The head of a household in non-husband-wife household.

The estimate of the number of households experiencing a household crime victimization then takes the form of multiplying the third stage ratio of the independent total of occupied housing units to the sample estimate of occupied housing units times the sample estimate for the specific household crime. The sample estimate for the specific household crime victimization is also derived using a "principal person" type estimation procedure.

The "principal person" is thought to produce better estimates than the use of household heads since coverage is generally better for females than for males.

E. Estimation Formulas for Quarterly Estimates of Levels and Rates

1. Quarterly Estimates of Level

a. The estimation formula for personal victimization events is:

$$V''_P = \sum_j \sum_a \left\{ \frac{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot V'_{jac_{NSR}} + V'_{ja_{SR}}}{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot z_{2jac_{NSR}} + z_{2ja_{SR}}} \right\} \cdot Z_{2ja}$$

where

V'=estimate of personal crime victimizations for a given crime characteristic for a given month of interview obtained by inflating the sample data by the reciprocal of the probability of selection and adjusting for the noninterviewed sample units.

Notation for subscripts:

j=1,...,8 and denotes the month of interview for the eight months used to obtain the quarterly estimate.

a refers to one of the 72 age, sex, color cells used in the second stage ratio estimation process.

c refers to one of the 48 color, residence cells used in the first stage ratio estimation process.

SR refers to data collected in SR PSU's.

NSR refers to data collected in NSR PSU's.

Notation for variables:

Z_1 = 1970 Census count of total persons. These counts refer to NSR strata only and are the independent totals used in the first stage ratio estimation process.

z_1 = estimate of Z_1 using 1970 Census data for the sample NSR PSU's inflated to the NSR strata level.

Z_2 = independent estimate of total persons used in the second stage ratio estimation process.

z_2 = estimate of Z_2 from the sample obtained by inflating the sample data by the reciprocal of the probability of selection and adjusting for noninterviewed sample units.

b. Household Victimization Estimates

The estimation formula for household victimization events is:

$$V''_H = \sum_r \frac{Z_{3r}}{z_{3r}} \cdot \left\{ \sum_j \sum_a \left[\frac{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot V'_{jacr_{NSR}} + V'_{jar_{SR}}}{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot z_{2jac_{NSR}} + z_{2ja_{SR}}} \right] \cdot Z_{2ja} \right\}$$

where, besides the notation described in II.E.1.a above, the additional subscripts and variables are:

Additional notation for subscripts:

r = refers to one of the four residence, tenure cells used in the third stage ratio estimation process.

Z_3 = independent estimate of total housing units used in the third stage ratio estimation process.

z_3 = estimation of Z_3 obtained by the "principal persons" estimation procedure explained above.

Note that the a^{th} age, sex, color cell applicable to a household is determined by the age, sex, and color of the "principal person" defined for that household.

2. Quarterly Estimates of Rates

a. Personal Victimization Rate Estimates

The estimate of a victimization rate for personal victimizations is the ratio of the estimated number of personal crime victimizations for the crime of interest divided by the estimate of the number of persons in the class of interest.

$$V''_{R_P} = \frac{V''_P}{\sum_j \sum_a \omega_j \left\{ \frac{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot Y'_{jac_{NSR}} + Y'_{ja_{SR}}}{\sum_c \frac{z_{1c}}{z_{1c}} \cdot z_{2jac_{NSR}} + z_{2ja_{SR}}} \right\} Z_{2ja}} = \frac{V''_P}{Y''_P}$$

where, besides the notation explained in E.1 above,

Y'_j is the estimated number of persons in the class of interest for the j^{th} month of interview obtained by inflating the sample data by the reciprocal of the probability of selection and adjusting for noninterviewed sample cases, and

ω_j is a weight factor required to adjust the estimate to a U.S. level:

ω_1 and ω_8 = 1/3, ω_2 and ω_7 = 2/3, and $\omega_3, \omega_4, \omega_5$, and ω_6 = 1.

b. Household Victimization Rate Estimates

The estimate of a victimization rate for household victimizations is the ratio of the estimated number of household crime victimizations for the crime of interest divided by the estimate of the number of households in the class of interest. The estimation formula is:

$$V''_{R_H} = \frac{V''_H}{\sum_r \frac{Z_{3r}}{z_{3r}} \left\{ \sum_j \sum_a \left[\frac{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot Y'_{jacr_{NSR}} + Y'_{jar_{SR}}}{\sum_c \frac{Z_{1c}}{z_{1c}} z_{2jac_{NSR}} + z_{2ja_{SR}}} \right] Z_{2ja} \right\}} = \frac{V''_H}{Y''_H}$$

3. Analysis that is presently underway indicates that the third stage of ratio estimation for household victimizations is not very effective. Therefore, there is a strong possibility that we will drop the third stage of ratio estimation; and hence, the estimation formulas for household victimization will be identical to those for personal victimizations. For this reason, and for purposes of keeping the paper within size limitations, the following formulas are given for personal victimizations only.

F. Estimation Formulas for Annual or Semi-Annual Estimates of Levels and Rates

1. Personal Victimization

The estimation formulas for levels of annual or semi-annual victimizations are:

Annual Estimates

$V''_{A_P} = \sum_{i=1}^{4} V''_{iP}$ where the sum is over all four quarters of the year.

Semi-Annual Estimates

$V''_{S_P} = \sum_{i=1}^{2} V''_{iP}$ where the sum is over the two quarters making up the semi-annual period.

2. The estimation formulas for annual or semi-annual rates is:

Annual Estimates

$$V''_{AR_P} = 4V''_{A_P} \Big/ \sum_{i=1}^{4} Y''_{Pi}$$

Semi-Annual Estimates

$$V''_{SR_P} = 2V''_{S_P} \Big/ \sum_{i=1}^{2} Y''_{Pi}$$

G. Estimation Formulas for Quarter-to-Quarter Change in Levels and Rates*

Note, the t and t-1 notation denote the two quarters that are being compared where t denotes the most recent quarter.

1. The estimation formula for quarter-to-quarter change in levels is:

$$V''_{QC_P} = \left(V''_{tP} - V''_{(t-1)P}\right) \Big/ V''_{(t-1)P}$$

2. The estimation formula for quarter-to-quarter change in rates is:

$$V''_{QRC_P} = \left(V''_{R_{Pt}} - V''_{R_{P(t-1)}}\right) \Big/ V''_{R_{P(t-1)}}$$

H. Estimation Formula for Change in Annual Estimation of Personal Victimization Levels and Rates

Note, the t and t-1 notation denotes the t^{th} year and $(t-1)^{th}$ year respectively.

1. The estimation formula for annual change in levels is:

$$V''_{AC_P} = \left(V''_{A_{tP}} - V''_{A_{(t-1)P}}\right) \Big/ V''_{A_{(t-1)P}}$$

2. The estimation formula for annual change in rates is:

$$V''_{ARC_P} = \left(V''_{AR_{tP}} - V''_{AR_{(t-1)P}}\right) \Big/ V''_{AR_{(t-1)P}}$$

III. Standard Error Estimation

A. Introduction

1. Components of the Standard Errors

An analytical statement of the standard error for characteristics of the National Crime Survey (NCS) can be expressed as the sum of several components-one for each stage of sampling and components with and without various stages of ratio estimation. Whereas the most important standard error estimates are the standard errors on the final estimates, we intend to estimate additionally the following for research purposes:

a. The "between PSU" component of the standard error of an estimate. This component arises because of the selection of a sample of primary sampling units (PSU's) out of each nonself-representing (NSR) stratum.

b. The "within PSU" component of the standard error of an estimate. This component arises because of the selection and interview of a sample of housing units within each selected PSU.

c. The standard error on "unbiased" estimates. All ratio estimation steps will be omitted for both person and household crime characteristics.

d. The standard error on estimates that include only the first stage of ratio estimation. Thus, the second stage ratio estimation procedure will be omitted for person characteristics and the second and third stage ratio estimation procedure will be omitted for household characteristics.

e. The standard error on household estimates that include only the first and second stages of ratio estimation. The third stage of ratio estimation will be omitted.

2. Various Standard Error Estimators

Section B below will illustrate some of the various standard error estimators for which we intend to obtain estimates. The formulas will be given only for estimates of the levels and rates of personal victimization for the reasons stated above. Standard errors will be calculated for the following:

a. Quarterly estimates of victimization levels for both person and household characteristics. As a by-product we will obtain:

i. "Between-PSU" component of the standard error.

ii. "Within-PSU" component of the standard error.

iii. Standard error estimates without the various stages of ratio estimation described in III.A.1 above.

b. Quarterly estimates of victimization rates for both person and household characteristics.

c. Annual and semi-annual estimates of victimization levels for both person and household characteristics. As a by-product we will again obtain the following:

i. "Between-PSU" component of the standard error.

ii. "Within-PSU" component of the standard error.

iii. Standard error estimates without the various stages of ratio estimation described in III.A.1 above.

d. Annual and semi-annual estimates of victimization rates for both person and household characteristics.

e. Change from any given quarter to another quarter for both victimization rates and levels.

f. Change from any given annual or semi-annual period to another annual or semi-annual period for both victimization rates and levels.

B. Method of Standard Error Estimation

1. Taylor Expansion Method vs. Balanced Repeated Replication Method

There are two essential means by which standard errors can be calculated for this survey. The Taylor expansion method or the balanced repeated replication method. It was our decision to use the Taylor expansion method. The basic design was a one PSU per stratum design; therefore, strata were combined so that each pair of sample PSU's were grouped together in a "combined" stratum. This enabled us to use a method of standard error estimation that was proposed by Keyfitz 1/ and more recently generalized by Tepping 2/ and Woodruff 3/. These individuals showed that consistent estimates of the standard error for complex non-linear estimators are provided by relatively simple quadratic functions of the observations in each stratum.

The decision to use a Taylor Expansion Method was decided upon a basis of efficiency. If we only needed standard errors on the final estimates we most probably would have gone to a balanced replication method because of its computer programming simplicity. A major emphasis, however, was on obtaining information on the components of the standard errors that have been described previously. We felt that the most efficient way to obtain both standard errors on final estimates and the components information was to use a Taylor series expansion method.

2. Keyfitz Method Applied to Taylor Expansion

In order to apply the Keyfitz method to estimate the variance of the Taylor expansion of an estimate, half-sample estimates need to be obtained within each stratum or group of strata. The procedure for obtaining the required half-sample estimates differs for SR strata and NSR strata. In general, the half-samples should be formed such that the half-sample estimates are independent and the expected values of the half-sample estimates within a stratum are approximately equal.

a. Formation of the half-samples in SR strata

Half-samples are formed within each of the 156 SR strata. Recall that at the time of sample selection, each sample cluster of about four housing units was assigned to a rotation group and a panel which determined the month of interview. The possible combinations of the rotation group numbers and the eight months of interview used to obtain a quarterly estimate are divided into two groups. One half-sample consists of those sample clusters with the rotation group and month of interview assignments in one group and the second half-sample consists of the remaining sample. The table in c. below illustrates the composition of the half-samples.

The half-samples are so formed to offset possible biases due to the rotation pattern and the month of interview. The existence of rotation group bias means that the expected value of estimates based on each rotation group will not be equal. This is compensated for by assigning half of each rotation group to one half-sample and the remaining half of the rotation group to the other half-sample. There is also a bias attributable to the month of interview due to telescoping as mentioned previously. These biases are offset in the formation of the half-samples by assigning the months of interview to half-samples such that interview months with an expected negative bias are combined with interview months with an expected positive bias. The net result of this is an approximately zero bias for each half-sample.

b. Formation of the half-samples in NSR strata

Half-samples are formed within each of 110 pairs of NSR strata. Since only one PSU is selected per NSR stratum, it is necessary to pair like strata for purposes of estimating total NSR variances.

In order to estimate both total NSR variance and "within" NSR variance, two techniques are employed for forming half-samples. One technique is that described above for SR PSU's. This technique, when applied to NSR PSU's, gives an estimate of the "within-PSU" component of the standard error for NSR PSU's.

The second technique is used to provide estimates of the total variance contributed by the NSR strata. In this technique each PSU in a collapsed pair is a half-sample. However, in this case, the expected values of the resulting half-sample estimates are not necessarily equal. Therefore, weights are required such that when they are applied to each half-sample, the expected value of the weighted half-sample values are approximately equal and the estimate of variance that is obtained from the weighted half-sample data is approximately equal to the variance we are trying to estimate.

Define these weights as $P_{s1}=\sqrt{\frac{N_1}{N_2}}$ and $P_{s2}=\sqrt{\frac{N_2}{N_1}}$, reference 4/.

The 1 and 2 denotes the two half-samples, s denotes the s^{th} pair of strata, N_1 denotes the population of the first stratum in a pair, and N_2 is the population of the other stratum in a pair.

c. Illustration of half-sample assignments for computing "within" SR and NSR variances

Let g=1,...,6, represent six rotation groups and j=1,...,8 represent the eight months of interview that are used to obtain a quarterly estimate.

Table of half-sample assignments:

Values of g	Possible combinations of g and j, i.e., gj: Half Sample 1	Half-Sample 2
1	11,12,14,17,18	13,15,16
2	23,25,26	21,22,24,27,28
3	31,32,34,37,38	33,35,36
4	43,45,46	41,42,44,47,48
5	51,52,54,57,58	53,55,56
6	63,65,66	61,62,64,67,68

3. Standard Error Estimators

The formulas below are given for the variance of an estimate. The standard error would be obtained by taking the square root of the variance.

<u>Additional Notation Required for the Standard Error Estimates</u>

i denotes the i^{th} characteristic for which a standard error will be calculated.

s refers to the s^{th} stratum or pair of strata in which half-sample estimates are to be obtained, call this the s^{th} Keyfitz cluster.

V_i(NSR) Best available estimate of the number of personal crime victimizations for NSR PSU's.

V_i(SR) Same as above except for SR PSU's.

For SR PSU's total variance and NSR PSU's "within" variance:

$\Delta V'_{igjas} = +V'_{igjas1}$ if gj is an element of half-sample 1.

$= -V'_{igjas2}$ if gj is an element of half-sample 2.

$\Delta z_{2gjas} = +z_{2gjas1}$ if gj is an element of half-sample 1.

$\Delta z_{2gjas} = -z_{2gjas2}$ if gj is an element of half-sample 2.

s_1 refers to half-sample 1 in the s^{th} Keyfitz cluster.

s_2 refers to half-sample 2 in the s^{th} Keyfitz cluster.

For NSR PSU's total variances:

$$\Delta V'_{igjas} = P_{s2} \cdot V'_{igjas1} - P_{s1} \cdot V'_{igjas2}$$

$$\Delta z_{1cs} = P_{s2} \cdot z_{1cs1} - P_{s1} \cdot z_{1cs2}$$

$$\Delta z_{2gjas} = P_{s2} \cdot z_{2gjas1} - P_{s1} \cdot z_{2gjas2}$$

P_{s1} and P_{s2} are the weights described in section III.B.2.

$\gamma_{ja} = \dfrac{Z_{2ja}}{E(z_{2ja})}$. This term results from the Taylor expansion of the estimate.

a. Quarterly Personal Victimization Levels

i. Variance of unbiased estimate

$$\hat{\sigma}^2_{up} = \sum_{s}^{\substack{SR\\PSU's}} \left(\sum_g \sum_j \sum_a \Delta V'_{igjas}\right)^2 + \sum_{s}^{\substack{NSR\\PSU's}} \left(\sum_g \sum_j \sum_a \Delta V'_{igjas}\right)^2$$

ii. Variance of estimate using only first stage ratio estimate

$$\hat{\sigma}^2_{1p} = \sum_{s}^{\substack{SR\\PSU's}} \left(\sum_g \sum_j \sum_a \Delta V'_{igjas}\right)^2 + \sum_{s}^{\substack{NSR\\PSU's}} \left\{\sum_g \sum_j \sum_a \Delta V'_{igjas} - \sum_j \sum_a \sum_c (\Delta z_{1cs}) \frac{V(NSR)_{ijac}}{Z_{1c}}\right\}^2$$

iii. "Within PSU" variance estimate for unbiased estimate

$$\hat{\sigma}^2_{w_{up}} \doteq \sum_{s}^{\substack{SR\\PSU's}} \left\{\sum_g \sum_j \sum_a \Delta V'_{igjas}\right\}^2 + \sum_{s}^{\substack{NSR\\PSU's}} \left\{\sum_g \sum_j \sum_a \Delta V'_{igjas}\right\}^2$$

iv. "Within PSU" variance estimate for final estimate (includes second stage ratio estimates)

$$\hat{\sigma}^2_{w_{fp}} = \sum_{s}^{\substack{SR\\PSU's}} \left\{\sum_g \sum_j \sum_a (\gamma_{ja}) \left[\Delta V'_{igjas} - (\Delta z_{2gjas}) \frac{V(NSR)_{ija} + V(SR)_{ija}}{Z_{2ja}}\right]\right\}^2$$

$$+ \sum_{s}^{\substack{NSR\\PSU's}} \left\{\sum_g \sum_j \sum_a (\gamma_{ja}) \left[\Delta V'_{igjas} - (\Delta z_{2gjas}) \frac{V(NSR)_{ija} + V(SR)_{ija}}{Z_{2ja}}\right]\right\}^2$$

v. Variance of final estimates

$$\hat{\sigma}^2_{fp} = \sum_{s}^{\substack{SR\\PSU's}} \left\{\sum_g \sum_h \sum_a (\gamma_{ja}) \left[\Delta V'_{igjas} - (\Delta z_{2gjas}) \frac{V(NSR)_{ija} + V(SR)_{ija}}{Z_{2ja}}\right]\right\}^2$$

$$+ \sum_{s}^{\substack{NSR\\PSU's}} \left\{\sum_g \sum_j \sum_a (\gamma_{ja}) \left[\Delta V'_{igjas} - (\Delta z_{2gjas}) \left(\frac{V(NSR)_{ija} + V(SR)_{ija}}{Z_{2ja}}\right)\right]\right.$$

$$\left. + \sum_j \sum_a \sum_c (\gamma_{ja}) (\Delta z_{1cs}) \left[\frac{\gamma_{ja}\left(V(NSR)_{ija} + V(SR)_{ija}\right)}{Z_{1c}} \frac{Z(NSR)_{2jac}}{Z_{2ja}} - \frac{V(NSR)_{ijac}}{Z_{1c}}\right]\right\}^2$$

b. Quarterly Personal Victimization Rates

i. Variance of unbiased estimate

In the following formula $\sum_s$ means to sum across both SR and NSR Keyfitz clusters. The subclusters for NSR PSU's are defined as for estimating total variance.

b. Quarterly Personal Victimization Rates

i. Variance of unbiased estimate

$$\hat{\sigma}^2_{R_{up}} = \sum_s \left\{ \frac{\sum_g \sum_j \sum_a \left(\Delta V'_{igjas}\right)}{\sum_j \sum_a \left[\left(Y(NSR)_{bja}+Y(SR)_{bja}\right) w_j\right]} - \frac{\sum_j \sum_a \left[V(NSR)_{ija}+V(SR)_{ija}\right]}{\left[\sum_j \sum_a \left(Y(NSR)_{ija}+Y(SR)_{bja}\right) w_j\right]^2} \left[\sum_g \sum_j \sum_a (\Delta y'_{bgjas})\right] \right\}^2$$

ii. Standard error on final estimate

$$\sigma^2_{F_P} = \sum_s \left\{ \sum_g \sum_j \sum_a \frac{\gamma_{ja}}{\kappa_{ob}} (\Delta V'_{igjas}) - \sum_g \sum_j \sum_a \frac{\kappa_{1i}}{(\kappa_{ob})^2} (\Delta y'_{bgjas}) \gamma_{ja}\omega'_j + \sum_c \frac{\kappa_{ob}\kappa_{2ic}-\kappa_{1i}\kappa_{3bc}}{(\kappa_{ob})^2} (\Delta z_{1cs}) \right.$$

$$\left. + \sum_g \sum_j \sum_a \frac{\kappa_{ob}\kappa_{4ija}+\kappa_{1i}\kappa_{bja}}{(\kappa_{ob})^2} (\Delta z_{2gjas}) \right\}^2$$

Besides the notation described previously, the additional notation used in the above estimator is:

$\Delta y'_{bgjas}$ is defined similarly to $\Delta V'_{igjas}$.

$$\kappa_{ob} = \sum_j \sum_a \left\{Y(NSR)_{bja} + Y(SR)_{bja}\right\} \gamma_{ja}\omega_j$$

$$\kappa_{1i} = \sum_j \sum_a \left\{V(NSR)_{ija} + V(SR)_{ija}\right\} \gamma_{ja}$$

$$\kappa_{2ic} = \sum_j \sum_a \left[\frac{\gamma_{ja}(-Z_{2ja})V(NSR)_{ijac}+Z(NSR)_{2jac}\gamma^2_{ja}\{V(NSR)_{ija}+V(SR)_{ija}\}}{Z_{1c} \quad Z_{2ja}}\right]$$

$$\kappa_{3bc} \quad \sum_j \sum_a \left[\frac{\gamma_{ja}(-Z_{2ja})Y(NSR)_{bjac}+Z(NSR)_{2jac}\gamma^2_{ja}\{Y(NSR)_{bja}+Y(SR)_{bja}\}}{Z_{1c} \quad Z_{2ja}}\right] \omega_j$$

$$\kappa_{4ija} = \left[-\left\{V(NSR)_{ija}+V(SR)_{ija}\right\} \Big/ Z_{2ja}\right] \gamma^2_{ja}$$

$$\kappa_{5bja} = \left[\left\{Y(NSR)_{bja}+Y(SR)_{bja}\right\} \Big/ Z_{2ja}\right] \gamma^2_{ja}\omega_j$$

$Y_b(SR)$ and $Y_b(NSR)$ denote the best available estimate of the number of persons having a specified characteristic.

y'_b unbiased estimate of the number of persons having a specified characteristic.

$Z(NSR)_2$ best available estimate of total persons.

b subscript denotes the b^{th} base characteristic; i.e., the kind of person or household for which crime characteristics are being measured.

c. Annual or Semi-annual Personal Victimization Levels

i. Standard error of final estimate

$$\hat{\sigma}^2_{FA_P} = \sum_{s}^{\substack{SR \\ PSU's}} \left\{ \sum_h \sum_g \sum_j \sum_a (\gamma_{hja}) \left[\Delta V'_{ihgjas} - (\Delta z_{2hgjas}) \quad \frac{V(NSR)_{ihja}+V(SR)_{ihja}}{Z_{2hja}}\right]\right\}^2$$

$$+ \sum_{s}^{\substack{NSR \\ PSU's}} \left\{ \sum_h \sum_g \sum_j \sum_a (\gamma_{hja}) \left[\Delta V'_{ihgjas} - (\Delta z_{2hgjas}) \frac{V(NSR)_{ihja}+V(SR)_{ihj}}{Z_{2hja}}\right.\right.$$

$$\left.\left. + \sum_h \sum_j \sum_a \sum_c (\gamma_{hja})(\Delta z_{1cs}) \frac{\gamma_{hja}\left(V(NSR)_{ihja}+V(SR)_{ihja}\right)\left(Z(NSR)_{2hjac}\right)}{Z_{1c} \quad Z_{2hja}} - \frac{V(NSR)_{ihjac}}{Z_{1c}}\right]\right\}^2$$

d. Annual or Semi-annual Personal Victimization Rate

1. Standard error of final estimate

$$\hat{\sigma}^2_{FAR_p} = \sum_{s}^{\text{All PSU's}} M \left\{ \sum_h \sum_g \sum_j \sum_a \frac{\gamma_{hja}}{\left(\sum_h K_{obh}\right)} \left(\Delta V'_{ihgjas}\right) - \sum_h \sum_g \sum_j \sum_a \frac{\left(\sum_h K_{1ih}\right)}{\left(\sum_h K_{obh}\right)^2} \gamma_{hja}\omega_j (\Delta y'_{bhgjas}) + \sum_h \sum_c \Delta z_{1cs} \frac{\left(\sum_h K_{obh}\right) K_{2ihc} - \left(\sum_h K_{1ih}\right) K_{3bhc}}{\left(\sum_h K_{obh}\right)^2} + \sum_h \sum_g \sum_j \sum_a \frac{\left(\sum_h K_{obh}\right) K_{41hja} + \left(\sum_h K_{1ih}\right) K_{5bhja}}{\left(\sum_h K_{obh}\right)^2} (\Delta z_{2hgjas}) \right\}^2$$

M=16 for annual variance estimator.
M=4 for a semi-annual variance estimator.

e. Quarter-to-Quarter Change in Personal Victimization Levels

Let $L(f_p)_{ts}$ be a Keyfitz cluster total for a certain quarter, i.e., quantities inside the brackets in III.B.3.a.v prior to squaring the quantity.

$$\hat{\sigma}^2_{X''_{QC_p}} = \sum_s \left\{ \frac{1}{K_{(t-1)\,1i}} L(f_p)_{ts} - \frac{K_{t1i}}{K_{(t-1)1i}^{\;2}} L(f_p)_{(t-1)s} \right\}^2$$

f. Quarter-to-Quarter Change in Personal Victimization Rates

Define $L(FR_p)_{ts}$ similarly to $L(f_p)_{ts}$ except in this case the Keyfitz cluster total applies to rates.

$$\hat{\sigma}^2_{X''_{QCR_p}} = \sum_s \left\{ \frac{K_{(t-1)ob}}{K_{(t-1)2i}} L(FR_p)_{ts} - \frac{(K_{t1i})(K_{(t-1)ob})^2}{(K_{tob})(K_{(t-1)1i})} L(FX_p)_{(t-1)s} \right\}^2$$

g. Year-to-Year Change in Personal Victimization Rates

Let $L(FA_p)_{ths}$ be the Keyfitz cluster totals given in III.B.3.c.i for the h^{th} quarter of the t^{th} year, then

$$\hat{\sigma}^2_{X''_{AC_p}} = \sum_s \left\{ \sum_h \frac{1}{K_{(t-1)hi}} L(FA_p)_{ths} - \sum_h \frac{K_{th1i}}{(K_{(t-1)h1i})^2} L(FA_p)_{(t-1)hs} \right\}^2$$

h. Year-to-Year Change in Personal Victimization Rate

Let $L(FAR_p)$ be a ratio estimated Keyfitz cluster total given in III.B.3.d.i for the h^{th} quarter of the t^{th} year, then

$$\hat{\sigma}^2_{X_{AR_p}} = \sum_s^{\text{All PSU's}} \left\{ \sum_h \frac{\sum_h K_{(t-1)hob}}{\sum_h K_{(t-1)h1i}} L(FAR_p)_{ths} - \sum_h \frac{\sum_h K_{th1i}}{\sum_h K_{thob}} \cdot \frac{\sum_h K_{(t-1)hob}^{\;2}}{\sum_h K_{(t-1)h1i}^{\;2}} L(FAR_p)_{(t-1)hs} \right\}^2$$

4. Generalization of the Standard Error

Every estimate made from the survey process has its own standard error. It is, however, impractical to compute an estimate of the standard error for every sample estimate. Therefore, standard errors are estimated for a representative subset of the sample estimates. These standard errors are then generalized to be applicable to all estimates made from the survey.

More than one generalized curve will be fitted to the estimates of the standard errors produced. At a minimum, there will be one table for personal crimes and one table for household crimes. There could also be separate curves for the various tabulation categories such as for crime victimizations, victim events, and crime incidents. The number of curves produced will depend on the degree of difference in the standard errors computed for the various categories.

The curves that are fitted are of the form of $\sigma^2_x \doteq x^2(a+\frac{b}{x})$, where x is the estimate of level, σ^2_x is the estimate of variance from the curve, and a and b are constants. The procedure for determining a and b is a process which minimizes the sum of the squared differences between the observed values, call $\frac{\hat{\sigma}^2_x}{\hat{x}}$, and the predicted values $a+\frac{b}{x}$ divided by the predicted values $a+\frac{b}{x}$. Similar type curves will be fitted to the computed variances of rates and other computed variances of changes in levels and rates.

REFERENCES

1. Keyfitz, Nathan, "Estimates of Sampling Variance when Two Units are Selected from Each Stratum," Journal of the American Statistical Association 52 (December 1957), pp. 503-510.

2. Tepping, Benjamin J., "Variance Estimation in Complex Surveys," Proceedings of the Social Statistics Section 1968 ASA, pp. 11-18.

*The estimators illustrated in this section would hold for change from a given quarter to the succeeding quarter, for change from a given quarter to the same quarter one year hence, and for changes in semi-annual data.

3. Woodruff, Ralph S., "A Simple Method for Approximating the Variance of a Complicated Estimate," Journal of the American Statistical Association 66 (June 1971), pp. 411-413.

4. Shapiro, Gary, "Optimum Weights for Use in Keyfitz Estimates of NSR Variances," Internal Census Bureau memorandum to Anthony G. Turner, July 19, 1966.

APPENDIX

1. Derivation of the variance estimator for estimates of the level of personal victimization

$$V''_P = \sum_j \sum_a \left\{ \frac{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot V'_{(NSR)_{jac}} + V'_{(SR)_{ja}}}{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot z_{(NSR)_{2jac}} + z_{(SR)_{ja}}} \right\} \cdot Z_{2ja}$$

is the estimation formula given in paragraph II.E.1.a for estimating the level of personal victimization.

Applying the theory presented in references 2/ and 3/, the variance of V''_P is approximately equal to the variance of the following linear form:

$$(1)\quad \sum_j \sum_a \sum_c V'_{(NSR)_{jac}} \left.\frac{\partial V''_P}{\partial V'_{(NSR)_{jac}}}\right|_E + \sum_j \sum_a V'_{(SR)_{ja}} \cdot \left.\frac{\partial V''_P}{\partial V'_{(SR)_{ja}}}\right|_E$$

$$+ \sum_j \sum_a \sum_c z_{1c} \cdot \left.\frac{\partial V''_P}{\partial z_{1c}}\right|_E + \sum_j \sum_a \sum_c z_{(NSR)_{2jac}} \cdot \left.\frac{\partial V''_P}{\partial z_{(NSR)_{2jac}}}\right|_E$$

$$+ \sum_j \sum_a z_{(SR)_{ja}} \cdot \left.\frac{\partial V''_P}{\partial z_{(SR)_{ja}}}\right|_E$$

, where each of the partial derivatives are evaluated at their expected values.

Taking the partial derivatives and evaluating these at their expected value gives:

$$(2)\quad \sum_j \sum_a \Bigg\{ (\gamma_{ja}) \Bigg[\left[V'_{(NSR)_{ja}} + V'_{(SR)_{ja}} \right] + \sum_c z_{1c} \left(\frac{-z_{2ja} V_{(NSR)_{jac}} + Z_{(NSR)_{2jac}} \left(V_{(NSR)_{ja}} + V_{(SR)_{ja}} \right)}{Z_{1c} Z_{2ja}} \right)$$

$$- \sum_c z(NSR)_{2jac} \left(\frac{V_{(NSR)_{ja}} + V_{(SR)_{ja}}}{Z_{2ja}} \right)$$

$$- z_{(SR)_{2ja}} \left(\frac{V_{(NSR)_{ja}} + V_{(SR)_{ja}}}{Z_{2ja}} \right) \Bigg] \Bigg\} = L(f_p)$$

In deriving (2) it is assumed that $E\left(\frac{Z_{1c}}{z_{1c}}\right) \doteq 1$ and that $\gamma_{ja} = E\left(\frac{Z_{2ja}}{z_{2ja}}\right)$.

Also, $V = E(V')$. These V values are approximated from the sample data using the first and second stage combined ratio estimate.

A theorem by Nathan Keyfitz, reference 1/, is then applied for purposes of estimating the variance of this linear form by means of a sum of squares of half-sample differences.

This sum of squares is given in paragraph III.B.3.a.v of the personal victimization rates.

2. Derivation of the variance estimator for estimates of rates.

$$V''_{R_P} = \frac{V''_P}{\sum_j \sum_a \omega_j \left\{ \frac{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot Y'(NSR)_{jac} + Y'(SR)_{ja}}{\sum_c \frac{Z_{1c}}{z_{1c}} \cdot z(NSR)_{2jac} + z(SR)_{2ja}} \right\} Z_{2ja}}$$

is the estimation formula given in paragraph II.E.2.a for estimating the rate of personal victimization.

The manner of obtaining a linear form with variance approximately equal to the variance of V''_{R_P} is similar to the procedure followed for obtaining (2). However, in this instance, partial derivatives involving the Y'variables need to be included and these partial derivatives are much more complex.

The linear form obtained for V''_{R_P} is as follows:

$$(3) \quad \sum_j \sum_a \left\{ \frac{Y_{ja}}{K_{0b}} (V'_{ja}) - \frac{K_{1i}}{(K_{0b})^2} (Y'_{ja})(\omega_j) + \frac{K_{0b}K_{4ja} + K_{1i}K_{5ja}}{(K_{0b})^2} z_{2ja} \right\} + \sum_c \left\{ \frac{K_{0b}K_{2c} - K_{1c}K_{3c}}{(K_{0b})^2} z_{1c} \right\} = L(FR_p)$$

Keyfitz theory is then applied to produce the variance estimator given in paragraph III.B.3.b.ii.