USERS' GUIDE TO THE NATIONAL CRIME VICTIMIZATION SURVEY (NCVS) GENERALIZED VARIANCE FUNCTIONS

G. Lance Couzens, BS
Bonnie Shook-Sa, MAS
Philip Lee, BS
Marcus Berzofsky, DrPH

RTI International
3040 East Cornwallis Road
Research Triangle Park, NC 27709

CONTENTS

INTRODUCTION

The National Crime Victimization Survey (NCVS), sponsored by the Bureau of Justice Statistics (BJS), is used to estimate the frequency and characteristics of criminal victimization in the United States. When calculating NCVS estimates, researchers must take into account the complex stratified, multistage sample design and resulting analysis weights. Stratification, clustering, and variation in analysis weights all affect the variances of survey parameters, and inappropriately accounting for these factors during estimation can lead to invalid results (Cochran, 1977).

Two broad methods exist for calculating variances of estimates from complex sample designs: direct variance estimation, and variance estimation by way of generalized variance functions (GVFs). Variances can be estimated directly using specialized software in conjunction with design elements present in the data, although this method depends on certain data structures that are not the default for publicly available NCVS datasets and assumes that users have access to such software and the knowledge to use it. A Direct Variance User's Guide was developed to provide guidance on how to appropriately estimate variances for the NCVS (Shook-Sa, Couzens, & Berzofsky, 2015) using this method. Alternatively, variances can be approximated using GVFs, which are models of the relationships between estimates, their characteristics (i.e., crime type, year, and subpopulation size), and their variances. Using these formulas, users can calculate variance estimates without knowledge of the sample design and without specialized software. This is the method traditionally used for NCVS variance estimation.

This guide provides documentation for users estimating variances with GVFs and incorporates guidance on the three key NCVS estimate types found in most BJS reports: victimization totals, proportions, and rates. It includes descriptions of calculating the estimates themselves as well as of conducting significance testing between pairs of estimates. *Section 1* summarizes the data structure of the NCVS PUFs. *Section 2* summarizes how the GVFs were developed, compares GVF with direct variance estimation, and provides guidance on when each method is appropriate. *Section 3* outlines how to apply the GVFs to calculate variance estimates for the NCVS, and *Section 4* provides detailed examples for various types of estimates. *Appendix A* contains the GVF parameters for 1993 through 2012, and *Appendix B* provides the

2-year correlations by crime type for 1993 through 2012, which are needed for pooled-year estimation.

User's Guide Terms and Notation

The following two tables define frequently used terms and outline the notation in this NCVS user's guide.

Terms Used in This Guide	
Generalized variance functions (GVFs)	GVFs use formulas and parameters from models of direct variance estimates that allow users to calculate approximations of variances without knowledge of the sample design. This is the method of variance estimation traditionally used for the National Crime Victimization Survey (NCVS) and the method documented in this user's guide.
Taylor series linearization (TSL)	TSL is the recommended method of direct variance estimation for the NCVS. TSL provides variance estimates of linear and nonlinear statistics through Taylor series expansions of linear functions.
Personal crimes	The NCVS defines personal crimes as those committed against an individual. The NCVS classifies personal crimes into five crime types: rape/sexual assault, robbery, aggravated assault, simple assault, and personal theft. Classifications are based on the *V4529* variable on the NCVS public use file (PUF).
Property crimes	The NCVS defines property crimes as those committed against households. The NCVS classifies property crimes into three crime types: household burglary, motor vehicle theft, and theft. Classifications are based on the *V4529* variable on the NCVS PUF.
Primary sampling units (PSUs)	In a household survey with a cluster sample design,* PSUs are typically geographic areas that are the first level of selection in the sample. On the NCVS PUF, the actual PSUs are not available because of disclosure concerns. Instead, PSUs are approximated with the pseudostratum† (*V2117*) and half sample§ (*V2118*) codes.
Series victimizations	Series victimizations occur when an NCVS respondent recalls at least six criminal incidents of a similar nature but cannot recall the dates and other details of the individual incidents well enough to report them separately. Each series victimization is represented as a single record on the incident-level file, and the number of incidents linked to that series is recorded. Weight adjustments are needed to appropriately account for series victimizations when computing NCVS estimates. The series count (*serieswgt*) and series-adjusted victimization weight (*newwgt*) are included on the modified analysis files for years before 2014, and the series-adjusted victimization weight will be included on the PUF starting in 2014.

*Cluster sampling methods group sampling units into subgroups called clusters. Rather than selecting sampling units directly, analysts select clusters for inclusion in the sample. Cluster sampling is used when a list of sampling units is not available at the time the sample is selected and when the analyst needs to control the geographic distribution of sampled units for logistical data collection purposes (Cochran, W. G. [1977]. *Sampling techniques*. New York, NY: Wiley.).

†Pseudostrata are available on the NCVS PUFs and approximate the true sampling strata.

§Half sample codes are available on the NCVS PUFs and approximate the true PSUs.

	Notations Used in This Guide
i	Represents National Crime Victimization Survey (NCVS) households. These can be identified on the household-level file on the basis of the household identification number (*IDHH*).
j	Represents NCVS individual respondents within households *i*. These can be identified on the person-level file on the basis of the person identification number (*IDPER*).
k	Represents reporting periods (i.e., *YEARQ*) for households *i* and individual respondents *j*.
l	Represents victimization records for respondent *j* in household *i* and reporting period *k*. Each record on the NCVS incident-level file is associated with a victimization record *l*.
D	Represents one or more domain characteristics of interest in the calculation of NCVS estimates. For victimization totals and proportions, domains can be defined on the basis of crime types (e.g., violent crimes, property crimes), characteristics of victims (e.g., age, sex, household income), or characteristics of the victimizations (e.g., victimizations reported to police, victimizations committed with a weapon present). Domains could also be a combination of all of these types of characteristics. For the calculation of victimization rates, domains are defined on the basis of characteristics of the victims.
A_a	Represents level *a* of covariate *A*. Covariate *A* is defined in the calculation of victimization proportions and represents the characteristic for which the analyst wants to obtain the distribution of victimizations in domain *D*.
C	Crime type *C* is defined in the calculation of NCVS victimization rates. It is the personal or property crime for which the analyst wants to obtain a victimization rate.

SECTION 1. DATA STRUCTURE

NCVS public use files (PUFs) are available through the National Archive of Criminal Justice Data within the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan. NCVS data, which are available in SAS, SPSS, Stata, R, and ASCII plain text formats, are organized in four files: an address-level file,[1] a household-level file, a person-level file, and an incident-level file.[2] Furthermore, since the NCVS employs a 6-month retroactive reference period for reporting crimes, data are available in two file structures: collection-year files and data-year files. Collection-year files contain incidents based on when the interviews were conducted (as opposed to when the incidents actually occurred). Data-year files contain incidents based on when they occurred, regardless of when the interviews were conducted. Different sets of survey weights are provided for each type of file. Under the collection-year approach, only 12 months of interviews are needed for annual estimation. With the data-year approach, annual estimates cannot be made until 18 months of interviews have been conducted, making collection-year-based estimation more timely. For most outcomes the difference between the collection-year and data-year estimates is not statistically or substantively different. Therefore, collection-year data are more commonly used in BJS analyses and reports. Although this user's guide will introduce both collection- and data-year weights, the examples will focus on collection-year estimation.

1.1   Household-Level File

The household-level file contains one record per reporting period for each sampled household in the NCVS. It contains data from the household screening interview, which assesses whether a household experienced any property crimes during the previous 6 months. Additionally, the household-level file contains characteristics of the household's surrounding area, such as the census region and Metropolitan Statistical Area (MSA) status, and the characteristics of the principal and reference persons within the household.

---

[1]   The address-level file is not used in the calculation of NCVS estimates and is therefore not further documented in *Section 1*.

[2] ICPSR includes both an incident file and an incident extract file. The incident extract file includes the associated household- and person- level variables, eliminating the need to merge these variables from the person- and household-level files. This document assumes that the analyst is using the incident extract file.

Because households stay in the NCVS sample for 3 years—reporting seven times at 6-month intervals—most households appear on the annual household file more than once. For this reason, both the household identification number (*IDHH*) and the year and quarter indicator (*YEARQ*) must be used to uniquely identify households by reporting period when merging the household file with other NCVS data files. Household-level estimates for the collection year use the collection-year household weight (*WGTHHCY*), whereas estimates for the data year are based on the data-year household weight (*WGTHHDY*). The household file is most commonly used in the calculation of property victimization rates (see **Section 3.1.3** for more details).

1.2   Person-Level File

The person-level file contains data for each household member aged 12 or older in responding NCVS households. Each record corresponds to a sampled person within a reporting period. Data come from the personal screening interviews, which are administered to all eligible and participating household members. The screening interview determines whether a person experienced a personal victimization during the previous 6 months.

Like the household file, the annual person-level file contains multiple records for most respondents. The person identification number (*IDPER*) must be used in conjunction with the year and quarter indicator (*YEARQ*) to uniquely identify persons within reporting periods when merging with other NCVS data files. The household identification number (*IDHH*) is available on the person-level file to facilitate merging with corresponding household-level data. Person-level estimates for the collection year use the collection-year person weight (*WGTPERCY*), whereas estimates for the data year are based on the data-year person weight (*WGTPERDY*). The person file is most commonly used in the calculation of personal victimization rates (see **Section 3.1.3** for more details).

1.3   Incident-Level File

In most cases, the incident-level file contains one record for each victimization reported by NCVS respondents. It contains both property crimes reported by the household respondent (i.e., household burglary, motor vehicle theft, and theft) and personal crimes reported by any NCVS respondent (i.e., rape/sexual assault, robbery, aggravated assault, simple assault, and personal theft). In addition, the incident-level file contains the corresponding household- and

person-level variables from the household and person files. The incident file contains information on crime type as well as on details of each victimization as drawn from the incident report (e.g., persons present, victim-offender relationship, weapon use). If the respondent reports six or more criminal incidents of a similar nature but cannot recall specific details of each one, the incidents are collapsed into a single record on the incident-level file and the total victimization count is recorded. These types of victimizations are called series victimizations.

Because respondents can report multiple incidents in a single interview, the incident file contains multiple records for some respondents and no records for respondents who did not report any criminal victimizations. The incident file can be linked to the household file with the household identification number (*IDHH*) and the year and quarter indicator (*YEARQ*). For personal crimes, the incident file can be linked to the person file with the person identification number (*IDPER*) and the year and quarter indicator (*YEARQ*). Estimates based on the incident file for the collection year are based on the collection-year victimization weight (*WGTVICCY*), whereas estimates for the data year are based on the data-year victimization weight (*WGTVICDY*).[3] Victimization weights are used to calculate victimization totals, proportions, and property and personal victimization rates (see ***Section 3.1*** for more details).

---

[3]   The victimization weights available on the NCVS public use files before 2014 do not account for series victimizations. A modified series-adjusted victimization weight (*newwgt*) can be derived and used in analyses that include series victimizations.

SECTION 2. NCVS VARIANCE ESTIMATION

When calculating estimates from a complex survey sample such as the one used in the NCVS, it is important to account for the corresponding changes in precision that result from the sample design. In other words, when estimating a population characteristic based on a sample, a level of uncertainty is incurred from sample variation. This uncertainty is often expressed through measures of precision such as standard errors (SEs) or confidence intervals, both of which are functions of the variance, transformed to the same scale as the estimate itself.

Typically, the clustering employed at the survey sampling stage results in precision reductions at the estimation stage, leading to larger SEs and wider confidence intervals (although survey coverage and implementation costs are improved). GVFs allow end users of NCVS data to approximate the change in precision for a given estimate by accounting for the characteristics that are known to affect its variance. The relationship between these characteristics and the estimated variance is modeled by the GVF formulas; the following sections provide some background on how the GVFs are developed and how GVF-estimated variances differ from those estimated directly using survey-appropriate software and pseudo design variables provided in NCVS PUFs.

2.1   Development of the NCVS GVFs

GVFs are models of the relationship between estimates, their characteristics (i.e., crime type, year, and subpopulation size), and their variances. In practice, they are estimate type-specific formulas that users can implement with simple software (e.g., Excel) and without knowledge of the NCVS design. Historically, GVFs have been used in conjunction with the NCVS because public use datasets are not designed for direct variance estimation across all common estimate types. Although the variances associated with percentages and totals may be estimated using only the incident file, and all the information necessary for direct variance estimation for rates is provided across three separate datasets, no clear guidance on how to construct an analytic file useful for all types of estimates, or comparisons with GVF-based estimates, was available until recently (Shook-Sa et al. [2015] includes this information). Even

with the newly-available guidance, GVFs remain useful in a number of scenarios—particularly when the user cannot employ software designed for survey estimation.

Each year the U.S. Census Bureau estimates new parameters to be used in conjunction with the GVF formulas. These parameters originate from regression models fit over variances expressed as functions of crime statistics (rates, percentages, totals, and means) and subpopulation sizes (Ash et al., 2008). Because these models are fit over a wide range of estimates using various combinations of subpopulations and detailed crime outcomes, they provide good approximations for most types of estimates.

2.2   GVFs vs. Direct Variance Estimation

In contrast to GVFs, direct variance estimation makes use of specialized software and sample design characteristics present in the analysis data for calculation of standard errors and confidence intervals, as well as for significance testing. In the case of the NCVS, variances may be estimated directly using pseudo-design variables in conjunction with one of a collection of specific methods designed to account for complex survey designs. Perhaps the easiest and most efficient method is Taylor Series Linearization (TSL), the preferred method for direct variance estimation with the NCVS and the one described in the NCVS Direct Variance User's Guide (Shook-Sa et al., 2015).

GVFs and direct variance estimation can be used in many of the same scenarios, and the choice of which to use in these situations will be dictated largely by ease and availability of resources. In terms of the ease with which each can be implemented, GVFs are clearly simpler in many ways, primarily because direct variance estimation requires a certain data structure that is not the default for NCVS PUFs. Specifically, direct variance estimation requires the input of a single dataset containing pseudo design variables, victimization variables, and any characteristics needed to define analysis domains.

For estimation of rates, victimization summaries must be calculated and merged onto either person or household files, depending on whether person or property crime estimates are needed. Additionally, care must be taken to ensure that the relationship between household or person weights and victimization weights is correct and that any adjustment factors built into the

victimization weight are taken into account. Estimation of victimization totals and certain types of percentages (where the analysis goal is to estimate what percentage of incidents have a certain characteristic) is often simpler, as such estimates require only incident datasets. Even in these instances, however, special care must be taken to ensure that all PSUs in the analysis domain are represented, even if no persons or households within those PSUs reported being victimized. This entails creation of so-called "dummy" records in the incident dataset.

For these reasons, ad-hoc or one-time analyses using direct variance estimation may require a considerable amount of effort that would not be required if GVFs were used instead. In some instances, though—particularly when many estimates are needed at a given time or on a recurring basis—direct variance estimation may require less effort, as variances are calculated along with point estimates. Generally speaking, economy of scale tips the ease-of-use balance in favor of direct variance estimation when the volume or frequency of analysis increases.

Although GVFs are simpler to implement in many cases, in some situations they are less appropriate—or impossible—to use. Because GVFs are models of an underlying relationship between point estimates and their variances, any characteristic of the required analysis not accounted for in the creation of the estimation sample over which the models are fit can lead to incorrect results. This is most likely to occur when estimates are calculated over very specific domains that were probably not included when the GVF models were estimated. In situations like these, GVFs may not be the most appropriate option. In other scenarios, they cannot be used at all. A very common analytic circumstance in which GVFs cannot be used is regression. There is no mechanism for estimating the variance of model parameters with NCVS GVFs. This is a significant limitation, as many analyses of even moderate complexity incorporate multivariable estimation.

SECTION 3. CALCULATING GVF ESTIMATES

GVF formulas are available for the estimation of variances corresponding to each of the three key NCVS estimate types: totals, percentages, and rates. The following sections detail how these formulas should be used and what inputs are required.

3.1   Key NCVS Estimates

The three key types of victimization estimates supported by the NCVS are totals, percentages, and rates. Because the relationship between an estimate and its variance differs across these estimate types, there are three distinct GVF formulas. The following sections describe how these estimates and their corresponding variances should be calculated.

3.1.1   Victimization Totals

Victimization totals are numbers of criminal victimizations with certain characteristics (e.g., the estimated number of simple assaults committed against females, the estimated number of household burglaries in the Northeast census region). Victimization totals are calculated from the incident file using the series-adjusted victimization weight (*newwgt*).[4,5] The estimated victimization total for domain D, $\hat{t}_D$, is

$$\hat{t}_D = \sum_{ijkl \in D} v_{ijkl},$$

where v_{ijkl} is the series-adjusted victimization weight (*newwgt*) for household i, respondent j, reporting period k, and victimization l. The series-adjusted victimization weights are summed for victimizations l with the domain characteristics of interest D for all households i, respondents j, and reporting periods k present on the incident file.

---

[4]   *newwgt* is defined as $s_l*WGTVICCY$, where s_l is the number of series victimizations represented by victimization record l (and $s_l = 1$ for nonseries victimization records). See ***Appendix A*** for more details on the calculation of *newwgt*.

[5]   If series victimizations are excluded from estimates, the unadjusted victimization weight (*WGTVICCY*) can be used in the calculation of totals. This applies to the calculation of victimization proportions as well (see ***Section 3.2.2***). Analysts should clearly document whether reported estimates include or exclude series victimizations.

To estimate the variance of a given total $\hat{t}_D$, the following formula should be used:

$$\hat{V}_t(\hat{t}_D; a, b, c) = a\hat{t}_D{}^2 + b\hat{t}_D + c\hat{t}_D\sqrt{\hat{t}_D},$$

where the GVF parameters a, b, and c are those provided annually by the U.S. Census Bureau (see **Appendix A**).

GVF-based z-scores for significance testing between two totals may be calculated via the following formula:

$$z = \frac{|\hat{t}_1 - \hat{t}_2|}{\sqrt{\hat{V}_{t_1} + \hat{V}_{t_2} - 2\hat{\rho}\sqrt{\hat{V}_{t_1}\hat{V}_{t_2}}}},$$

where $\hat{\rho}$ is the crime-specific estimated 2-year correlation (see **Appendix B**). Note that for across-domain comparisons within a single year, comparisons between two pooled-year estimates, comparisons between nonconsecutive years for the same crime type, and comparisons across crime types, $\hat{\rho} = 0$.

3.1.2   Victimization Percentages

Victimization percentages describe characteristics of victimizations or victims (e.g., the estimated percentage of robberies reported to the police, the estimated distribution of property crime victims by household income). Victimization percentages are calculated from the incident file using the series-adjusted victimization weight (*newwgt*). The estimated victimization percentage for domain D across level a of variable A, $\hat{p}_{A_a,D}$, is

$$\hat{p}_{A_a,D} = 100 * \left(\sum_{ijkl \in A_a,D} v_{ijkl} \right) \Big/ \widehat{N}_D,$$

where v_{ijkl} is the series-adjusted victimization weight (*newwgt*) for household i, respondent j, reporting period k, and victimization l. $\widehat{N}_D$ is the estimated population size of the entire domain D:

$$\widehat{N}_D = \sum_{ijkl \in D} v_{ijkl}.$$

In the numerator, series-adjusted victimization weights are summed for victimizations l with the domain characteristics of interest D associated with level a of covariate A for all households i, respondents j, and reporting periods k present on the incident file. In the denominator, victimization weights are summed for victimizations l with the domain characteristics of interest D for all households i, respondents j, and reporting periods k present on the incident file.

To estimate the variance of a given percentage $\hat{p}_{A_a,D}$, the following formula should be used:

$$\widehat{V}_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right) = b\frac{\hat{p}_{A_a,D}\left(100 - \hat{p}_{A_a,D}\right)}{\widehat{N}_D} + c\frac{\hat{p}_{A_a,D}\left(\sqrt{100\hat{p}_{A_a,D}} - \hat{p}_{A_a,D}\right)}{\sqrt{\widehat{N}_D}},$$

where the GVF parameters b and c are those provided annually by the U.S. Census Bureau (see **Appendix A**).

GVF-based z-scores for significance testing between two percentages may be calculated via the following formula:

$$z = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\widehat{V}_{p_1} + \widehat{V}_{p_2} - 2\hat{\rho}\sqrt{\widehat{V}_{p_1}\widehat{V}_{p_2}}}},$$

where $\hat{\rho}$ is the crime-specific estimated 2-year correlation (see **Appendix B**). Note that for across-domain comparisons within a single year, comparisons between two pooled-year estimates, comparisons between non-consecutive years for the same crime type, and comparisons across crime types, $\hat{\rho} = 0$.

3.1.3 Victimization Rates

Victimization rates are estimated numbers of victimizations per 1,000 persons or households in the population (e.g., the rate of simple assaults against females per 1,000 women,

the rate of thefts reported to the police per 1,000 households).[6] The estimated victimization rate for crime C in domain D, $\hat{r}_{C,D}$ is

$$\hat{r}_{C,D} = n_p * \left(\sum_{ijkl \in C,D} v_{ijkl} \right) \Big/ \widehat{N}_D,$$

where v_{ijkl} is the series-adjusted victimization weight (*newwgt*) for household i, respondent j (if crime C is a personal crime), reporting period k, and victimization l. n_p is the number of persons or households represented in the denominator of the rate ($n_p = 1,000$ is standard for BJS-published rates). $\widehat{N}_D$ is the estimated population size of the domain D:

$$\widehat{N}_D = \sum_{ijk \in D} w_{ijk},$$

where w_{ijk} is the person weight (*WGTPERCY*) for household i, respondent j, and reporting period k (if crime C is a personal crime) or the household weight (*WGTHHCY*) for household i and reporting period k (if crime C is a property crime).

To estimate the variance of a given rate $\hat{r}_{C,D}$, the following formula should be used:

$$\widehat{V}_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right) = b \frac{\hat{r}_{C,D}(n_p - \hat{r}_{C,D})}{\widehat{N}_D} + c \frac{\hat{r}_{C,D}\left(\sqrt{n_p * \hat{r}_{C,D}} - \hat{r}_{C,D}\right)}{\sqrt{\widehat{N}_D}},$$

where the GVF parameters b and c are those provided annually by the U.S. Census Bureau (see ***Appendix A***).

GVF-based z-scores for significance testing between two rates may be calculated via the following formula:

$$z = \frac{|\hat{r}_1 - \hat{r}_2|}{\sqrt{\widehat{V}_{r_1} + \widehat{V}_{r_2} - 2\hat{\rho}\sqrt{\widehat{V}_{r_1}\widehat{V}_{r_2}}}},$$

---

[6] Standard BJS-published rates are estimated as being among 1,000 persons or households, but alternate rates (e.g., the number of victimizations among 100,000 persons) may be estimated as well.

where $\hat{\rho}$ is the crime-specific estimated 2-year correlation (see ***Appendix B***). Note that for across-domain comparisons within a single year, comparisons between two pooled-year estimates, comparisons between nonconsecutive years for the same crime type, and comparisons across crime types, $\hat{\rho} = 0$.

3.2   Determining GVF Parameters

The first step in estimating the variance of a point estimate is selecting the correct GVF parameters. To do this, the user must establish whether the estimate is for a personal or household crime; whether the estimate will be based on a single year or multiple, pooled years; and whether it will be an overall or domain-specific estimate (i.e., among a particular subclass of the population). GVF parameters are provided for all collection years from 1993 to the present (see ***Appendix A***) and are organized into four sets within year. Each set contains the parameters a, b, and c, and the sets are defined as follows:

Set 1. Overall Person Crime Estimates
Set 2. Person Crime Domain Estimates
Set 3. Overall Property Crime Estimates
Set 4. Property Crime Domain Estimates

Selection of the correct set corresponding to the desired estimate takes care of overall vs. domain and personal vs. property crime concerns. All that remains is whether or not the estimate is to be based on a single or multiple, pooled years of data. For single-year estimates, nothing beyond selection of the correct set is required. For pooled-year estimates, however, some calculations are required to arrive at the final set of parameters to be used with the appropriate GVF formula.

3.2.1   Pooled-Year Estimates

To calculate GVF parameters for pooled-year estimates, data users must account for the correlation between years. Correlation coefficients are provided separately for key personal and property crimes and for all year pairings from 1993/1994 to the present (see ***Appendix B***). Regardless of the number of years contributing to a pooled-year estimate, only the correlation coefficient between the two most recent years is required. The following formulas may be used to arrive at a_p, b_p, and c_p values for pooled-year GVF estimates:

$$a_p = \left(\frac{1}{n_Y} \sum_{y \in Y} a_y \right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y^2} \right),$$

$$b_p = \left(\frac{1}{n_Y} \sum_{y \in Y} b_y \right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y} \right),$$

and

$$c_p = \left(\frac{1}{n_Y} \sum_{y \in Y} c_y \right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y^{\frac{3}{2}}} \right),$$

where n_Y is the number of years in the set of years Y, and $\hat{\rho}$ is the correlation coefficient corresponding to the two most recent years in the set Y. The parameters a_p, b_p, and c_p, once calculated, can be used in the GVF formulas from *Section 3.2*.

SECTION 4. GVF EXAMPLES

4.1   Totals

Example 1. Estimate the Total Number of Violent Victimizations, 2012

Step 1: Determine appropriate GVF formula:

$$V_t(\hat{t}_D; a, b, c) \ = \ a\hat{t}_D{}^2 + b\hat{t}_D + c\hat{t}_D\sqrt{\hat{t}_D}$$

Step 2: Determine appropriate GVF parameters:

- Year of interest: 2012
- Are aggregated parameters needed? No
- Personal or property crime? Personal
- Overall or domain estimate? Overall

a = -0.00025343
b = 6,439
c = 3.704

Step 3: Calculate estimate

- Determine appropriate formula for estimate type:

$$\hat{t}_D \ = \ \sum_{ijkl \in D} v_{ijkl}$$

File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newcrime = 1 and V4022 ne 1

SAS code:

```
libname ncvs "<Location of Incident File>";

proc means data = ncvs.NCVS_Incident sum;
      where year = 2012 and newcrime = 1 and V4022 ne 1;
var newwgt;
run;
```

Total estimate: 6,842,593

Step 4: Calculate GVF standard error:

$$\widehat{SE}_t = \sqrt{-0.00025343 * (6{,}842{,}593)^2 + 6439 * 6{,}842{,}593 + 3.704 * 6{,}842{,}593 * \sqrt{6{,}842{,}593}}$$
$$= 313{,}834.091$$

Example 2. Compare the Total Number of Violent Victimizations in 2011 With That in 2012

Note: The 2012 estimate and standard error were calculated in Example 1. First, calculate estimate and standard error for 2011, following the same steps as in Example 1:

Step 1: Determine appropriate GVF formula:

$$V_t(\hat{t}_D; a, b, c) = a\hat{t}_D{}^2 + b\hat{t}_D + c\hat{t}_D\sqrt{\hat{t}_D}$$

Step 2: Determine appropriate GVF parameters:

- Year of interest: 2011
- Are aggregated parameters needed? No
- Personal or property crime? Personal
- Overall or domain estimate? Overall

a = -0.00060211
b = 2,439
c = 9.511

Step 3: Calculate estimate

- Determine appropriate formula for estimate type:

$$\hat{t}_D = \sum_{ijkl \in D} v_{ijkl}$$

File: Incident-level
Weight: newwgt
Subpopulation: year = 2011 and newcrime = 1 and V4022 ne 1
SAS code:

```
libname ncvs "<Location of Incident File>";

proc means data = ncvs.NCVS_Incident sum;
   where year = 2011 and newcrime = 1 and V4022 ne 1;
   var newwgt;
run;
```

Total estimate: 5,812,523

Step 4: Calculate GVF standard error:

$$\widehat{SE}_t = \sqrt{-0.00060211 * (5{,}812{,}523)^2 + 2439 * 5{,}812{,}523 + 9.511 * 5{,}812{,}523 * \sqrt{5{,}812{,}523}}$$

$$= 356{,}534.327$$

Next, compare 2011 and 2012 estimates:

Step 1: Calculate z-score:

$$z = \frac{|\hat{t}_1 - \hat{t}_2|}{\sqrt{\hat{V}_{t_1} + \hat{V}_{t_2} - 2\hat{\rho}\sqrt{\hat{V}_{t_1}\hat{V}_{t_2}}}}$$

- Correlation years: 2011–2012
- Crime type: Crimes of violence
- $\hat{\rho} = 0.11$

$$z = \frac{|5{,}812{,}523 - 6{,}842{,}593|}{\sqrt{356{,}534.3267^2 + 313{,}834.091^2 - 2 * 0.11 * \sqrt{356{,}534.3267^2 * 313{,}834.091^2}}}$$

$$= 2.298$$

Conclusion: Because 2.30 > 1.96, the 2011 and 2012 violent victimization totals are significantly different at the 5% level.

Example 3. Total and average number of household burglaries, 2010–2012

Step 1: Determine appropriate GVF formula:

$$V_t(\hat{t}_D; a, b, c) = a\hat{t}_D^{\,2} + b\hat{t}_D + c\hat{t}_D\sqrt{\hat{t}_D}$$

Step 2: Determine appropriate GVF parameters:

- Years of interest: 2010–2012
- Are aggregated parameters needed? Yes
- Personal or property crime? Property
- Overall or domain estimate? Overall
- $\hat{\rho} = 0.14$

Parameter	2010	2011	2012
a	-0.00020353	-0.00028723	-0.00020722
b	2450	4182	2345
c	2.035	2.809	2.116

Let n_Y be the number of pooled years. Then the aggregated, or pooled, parameters a_p, b_p, and c_p are calculated as follows:

$$a_p = \left(\frac{1}{n_Y}\sum_{y \in Y} a_y\right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y^2}\right)$$

$$= \left(\frac{-0.00020353 - 0.00028723 - 0.00020722}{3}\right) * \left(\frac{3 + 0.14 * (3*(3) - 4)}{3^2}\right)$$

$$= \frac{-0.00069798}{3} * \frac{3.7}{9}$$

$$= -0.00023266 * 0.411111111$$
$$= -0.000095649$$

$$b_p = \left(\frac{1}{n_Y}\sum_{y \in Y} b_y\right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y}\right)$$

$$= \left(\frac{2,450 + 4,182 + 2,345}{3}\right) * \left(\frac{3 + 0.14 * (3*(3) - 4)}{3}\right)$$

$$= \frac{8977}{3} * \frac{3.7}{3}$$

$$= 2,992.33333 * 1.233333333$$

$$= 3,690.54$$

$$c_p = \left(\frac{1}{n_Y}\sum_{y \in Y} c_y\right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y^{\frac{3}{2}}}\right)$$

$$= \left(\frac{2.035+2.809+2.116}{3}\right) * \left(\frac{3+0.14*(3*(3)-4)}{3^{1.5}}\right)$$

$$= \left(\frac{2.035+2.809+2.116}{3}\right) * \left(\frac{3.7}{5.196152423}\right)$$

$$= \frac{6.96}{3} * \frac{3.7}{5.196152423}$$

$$= 2.32 * 0.712065332$$

$$= 1.652$$

Step 3: Calculate estimate

- Determine appropriate formula for estimate type:

$$\hat{t}_D = \sum_{ijkl \in D} v_{ijkl}$$

File: Incident-level
Weight: newwgt
Subpopulation: year ge 2010 and year le 2012 and newoff = 6 and V4022 ne 1
SAS code:

```
libname ncvs "<Location of Incident File>";

proc means data = ncvs.NCVS_Incident sum;
   where year ge 2010 and year le 2012 and newoff = 6 and V4022 ne 1;
   var newwgt;
run;
```

Total Estimate: 10,554,559

Step 4: Calculate GVF standard error for the total:

$$\sqrt{V_t(\hat{t}_D; a, b, c)} =$$

$$= \sqrt{-0.000095649 * (10{,}554{,}559)^2 + 3{,}690.54 * 10{,}554{,}559 + 1.65199157 * 10{,}554{,}559 * \sqrt{10{,}554{,}559}}$$

$$= 291{,}449.33$$

Step 5: Calculate estimate and GVF standard error for the annual average:

The estimates above are the total and SE(total) for the total number of household burglaries from 2010 through 2012. To calculate the average number of household burglaries each year, divide the total and SE(total) by the number of years (3).

Average number of household burglaries: 3,518,186.33
SE(average number of household burglaries):   97,149.78

Example 4. Average Household Burglaries, 2007–2009 Compared With 2010–2012

Note: The 2010–2012 estimate and standard error were calculated in Example 3. First, calculate the estimate and standard error for 2007–2009, following the same steps as in Example 3:

Average number of household burglaries: 3,484,566.33
SE(average number of household burglaries):   112,251.30

Next, compare 2007–2009 and 2010–2012 average annual estimates:

Step1: Calculate z-score:

$$z = \frac{|\hat{t}_1 - \hat{t}_2|}{\sqrt{\hat{V}_{t_1} + \hat{V}_{t_2} - 2\hat{\rho}\sqrt{\hat{V}_{t_1}\hat{V}_{t_2}}}}$$

- Correlation years: N/A (although $\hat{\rho}$ is accounted for in the two pooled estimates, it is assumed to be zero in the covariance term of the z-score equation when comparing two multiyear estimates)
- Crime type: Crimes of violence
- $\hat{\rho} = 0$

Step 2: Determine significance:

$$z = \frac{|\,3,518,186.33 - 3,484,566.33\,|}{\sqrt{97,149.78^2 + 112,251.30^2 - 2*(0)*\sqrt{(97,149.78^2)*(112,251.30^2)}}}$$

$$= \frac{33,620}{148,453.4721}$$

$$= 0.226 < 1.96$$

Conclusion: Because 0.226 < 1.645, the 2007–2009 and 2010–2012 total annual estimates are not significantly different at either the 5% or 10% levels.

4.2   Percentages

Example 5. Percentage of 2012 Property Crimes Committed Against Households With Income Less Than $25,000

Step 1: Determine appropriate GVF formula:

$$V_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right) = b\frac{\hat{p}_{A_a,D}\left(100 - \hat{p}_{A_a,D}\right)}{\widehat{N}_D} + c\frac{\hat{p}_{A_a,D}\left(\sqrt{100\hat{p}_{A_a,D}} - \hat{p}_{A_a,D}\right)}{\sqrt{\widehat{N}_D}}$$

Step 2: Determine appropriate GVF parameters:

- Year of interest: 2012
- Are aggregated parameters needed? No
- Personal or property crime? Property
- Overall or domain estimate? Domain

b = 2,810
c = 2.057

Step 3: Calculate estimate

$$\hat{p}_{A_a,D} = 100 * \left(\sum_{ijkl \in A_a,D} v_{ijkl}\right) \Big/ \widehat{N}_D$$

$$\widehat{N}_D = \sum_{ijkl \in D} v_{ijkl}$$

Numerator:
File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newcrime = 3 and V4022 ne 1 and hincome in (1 2 3)

Denominator:
File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newcrime = 3

SAS code:

```
libname ncvs "<Location of Incident File>";
```

Numerator:
```
proc means data = ncvs.NCVS_Incident sum;
    where year = 2012 and newcrime = 3 and V4022 ne 1 and hincome in (1 2 3);
    var newwgt;
run;
```

Denominator:

```
proc means data = ncvs.NCVS_Incident sum;
    where year = 2012 and newcrime = 3;
    var newwgt;
run;
```

Numerator: 5,042,004
Denominator: 19,751,046

$$\hat{p}_{A_a,D} = \frac{100*5,042,004}{19,751,046} = 25.53$$

Step 4: Calculate GVF standard error:

$$\sqrt{V_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right)} =$$

$$= \sqrt{2,810 * \frac{25.5 * (100 - 25.5)}{19,751,046} + 2.057 * \frac{25.5 * \left(\sqrt{100 * 25.5} - 25.5\right)}{\sqrt{19,751,046}}}$$

$$= \sqrt{0.270 + 0.295}$$
$$= \sqrt{0.565}$$
$$= 0.752$$

Example 6. Percentage of Property Crimes Committed Against Households With Income Less Than $25,000 Vs. Those With Income Greater Than or Equal to $25,000, 2012

Note: The 2012 estimated percentage of property crimes committed against households with income less than $25,000 was calculated in Example 5. First, calculate the estimate and standard error for the percentage of property crimes committed against households with income greater than or equal to $25,000, following the same steps as in Example 5.

Step 1: Determine appropriate GVF formula:

$$V_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right) = b\frac{\hat{p}_{A_a,D}\left(100 - \hat{p}_{A_a,D}\right)}{\widehat{N}_D} + c\frac{\hat{p}_{A_a,D}\left(\sqrt{100\hat{p}_{A_a,D}} - \hat{p}_{A_a,D}\right)}{\sqrt{\widehat{N}_D}}$$

Step 2: Determine appropriate GVF parameters:

- Year of interest: 2012
- Are aggregated parameters needed? No
- Personal or property crime? Property
- Overall or domain estimate? Domain

b = 2,810
c = 2.057

Step 3: Calculate estimate

- Determine appropriate formula for estimate type:

$$\hat{p}_{A_a,D} = 100 * \left(\sum_{ijkl \in A_a,D} v_{ijkl} \right) \Big/ \widehat{N}_D$$

$$\widehat{N}_D = \sum_{ijkl \in D} v_{ijkl}$$

Numerator:
File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newcrime = 3 and V4022 ne 1 and hincome in (4 5 6 7)

Denominator:
File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newcrime = 3

SAS code:
```
libname ncvs "<Location of Incident File>";
```

Numerator:
```
proc means data = ncvs.NCVS_Incident sum;
     where year = 2012 and newcrime = 3 and V4022 ne 1 and hincome in (4 5 6 7);
var newwgt;
run;
```

Denominator:
```
proc means data = ncvs.NCVS_Incident sum;
     where year = 2012 and newcrime = 3;
var newwgt;
run;
```

Numerator: 10,171,090
Denominator: 19,751,046

$$\frac{10,171,090}{19,751,046} * 100 = 51.5$$

Step 4: Calculate GVF standard error:

$$\sqrt{V_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right)} =$$

$$= \sqrt{2{,}810 * \frac{51.5 * (100 - 51.5)}{19{,}751{,}046} + 2.057 * \frac{51.5 * \left(\sqrt{100 * 51.5} - 51.5\right)}{\sqrt{19{,}751{,}046}}}$$

$$= \sqrt{0.355 + 0.483}$$
$$= \sqrt{0.838}$$
$$= 0.915$$

Next, compare 2012 estimates for households with income less than $25,000 to the estimate for households with income greater than or equal to $25,000:

Step1: Calculate z-score:

$$z = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\widehat{V}_{p_1} + \widehat{V}_{p_2} - 2\hat{\rho}\sqrt{\widehat{V}_{p_1}\widehat{V}_{p_2}}}}$$

- Correlation years: n/a, $\hat{\rho} = 0$ for estimates within the same year
- Crime type: n/a
- $\hat{\rho} = 0$

$$z = \frac{|25.5 - 51.5|}{\sqrt{0.752^2 + 0.916^2 - 2*(0)*\sqrt{0.752^2 * 0.916^2}}} = \frac{26.0}{\sqrt{1.40}} = \frac{26.0}{1.19} = 21.8$$

Conclusion: Because 21.8 > 1.96, the 2012 estimates for households with income less than $25,000 and greater than or equal to $25,000 are significantly different at the 5% level.

Example 7. Percentage of Violent Victimizations Reported to Police, 2010–2012

Step 1: Determine appropriate GVF formula:

$$V_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right) = b\frac{\hat{p}_{A_a,D}\left(100 - \hat{p}_{A_a,D}\right)}{\widehat{N}_D} + c\frac{\hat{p}_{A_a,D}\left(\sqrt{100\hat{p}_{A_a,D}} - \hat{p}_{A_a,D}\right)}{\sqrt{\widehat{N}_D}}$$

Step 2: Determine appropriate GVF parameters:

- Years of interest: 2010–2012
- Are aggregated parameters needed? Yes
- Personal or property crime? Personal

- Overall or domain estimate? Overall
- $\hat{\rho} = 0.11$

Parameter	2010	2011	2012
b	1,384	2,439	6,439
c	13.696	9.511	3.704

Let n_Y be the number of pooled years. Then the aggregated, or pooled, parameters a_p, b_p, and c_p are calculated as follows:

$$
\begin{aligned}
b_p &= \left(\frac{1}{n_Y}\sum_{y \in Y} b_y\right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y}\right) \\
&= \left(\frac{1,384+2,439+6,439}{3}\right) * \left(\frac{3+0.11*(3*(3)-4)}{3}\right) \\
&= \left(\frac{10,262}{3}\right) * \left(\frac{3.55}{3}\right) \\
&= 3,420.667 * 1.183 \\
&= 4,047.79
\end{aligned}
$$

$$
\begin{aligned}
c_p &= \left(\frac{1}{n_Y}\sum_{y \in Y} c_y\right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y^{\frac{3}{2}}}\right) \\
&= \left(\frac{13.696+9.511+3.704}{3}\right) * \left(\frac{3+0.11*(3*(3)-4)}{3^{1.5}}\right) \\
&= \left(\frac{26.911}{3}\right) * \left(\frac{3.55}{3^{1.5}}\right) \\
&= 8.970 * 0.683 \\
&= 6.129
\end{aligned}
$$

Step 3: Calculate estimate

$$
\hat{p}_{A_a,D} = 100 * \left(\sum_{ijkl \in A_a,D} v_{ijkl}\right) \Big/ \hat{N}_D
$$

$$
\hat{N}_D = \sum_{ijkl \in D} v_{ijkl}
$$

Numerator:
File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newcrime = 1 and V4022 ne 1 and notify = 1

Denominator:
File: Incident-level
Weight: newwgt

Subpopulation: year = 2012 and newcrime = 1

SAS code:

```
libname ncvs "<Location of Incident File>";
```

Numerator:

```
proc means data = ncvs.NCVS_Incident sum;
    where year = 2012 and newcrime = 1 and V4022 ne 1 and notify = 1;
    var newwgt;
run;
```

Denominator:

```
proc means data = ncvs.NCVS_Incident sum;
    where year = 2012 and newcrime = 1;
    var newwgt;
run;
```

Numerator: 3,024,368.44
Denominator: 6,931,633.61

$$\frac{100*3,024,368}{6,931,634} = 43.6$$

Step 4: Calculate GVF standard error:

$$\sqrt{V_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right)} =$$
$$= \sqrt{4{,}048 * \frac{43.6*(100-43.6)}{6{,}931{,}634} + 6.13 * \frac{43.6*\left(\sqrt{100*43.6}-43.6\right)}{\sqrt{6{,}931{,}634}}}$$
$$= \sqrt{1.44 + 2.28}$$
$$= \sqrt{3.72}$$
$$= 1.93$$

Example 8. Percentage of Violent Victimizations Reported to Police: Males Compared With Females, 2010–2012

First, calculate the estimates for males:

Step 1: Determine appropriate GVF formula:

$$V_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right) = b\frac{\hat{p}_{A_a,D}\left(100 - \hat{p}_{A_a,D}\right)}{\widehat{N}_D} + c\frac{\hat{p}_{A_a,D}\left(\sqrt{100\hat{p}_{A_a,D}} - \hat{p}_{A_a,D}\right)}{\sqrt{\widehat{N}_D}}$$

Step 2: Determine appropriate GVF parameters:

- Years of interest: 2010–2012
- Are aggregated parameters needed? Yes
- Personal or property crime? Personal
- Overall or domain estimate? Domain
- $\hat{\rho} = 0.11$

Parameter	2010	2011	2012
b	1,960	2,309	3,893
c	17.278	12.916	8.530

Let n_Y be the number of pooled years. Then the aggregated, or pooled, parameters a_p, b_p, and c_p are calculated as follows:

$$b_p = \left(\frac{1}{n_Y} \sum_{y \in Y} b_y \right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y} \right)$$

$$= \left(\frac{17.278 + 12.916 + 8.530}{3} \right) * \left(\frac{3 + 0.11 * (3*(3) - 4)}{5.196} \right)$$

$$= \left(\frac{8,162}{3} \right) * \left(\frac{3.55}{3} \right)$$

$$= 2,720.667 * 1.183$$

$$= 3,219.456$$

$$c_p = \left(\frac{1}{n_Y} \sum_{y \in Y} c_y \right) * \left(\frac{n_Y + \hat{\rho}(3n_Y - 4)}{n_Y^{\frac{3}{2}}} \right)$$

$$= \left(\frac{13.696 + 9.511 + 3.704}{3} \right) * \left(\frac{3 + 0.11 * (3*(3) - 4)}{3^{1.5}} \right)$$

$$= \left(\frac{38.724}{3} \right) * \left(\frac{3.55}{5.196} \right)$$

$$= 12.908 * 0.683$$

$$= 8.819$$

Step 3: Calculate estimate:

$$\hat{p}_{A_a,D} = 100 * \left(\sum_{ijkl \in A_a,D} v_{ijkl} \right) \Big/ \hat{N}_D$$

$$\hat{N}_D = \sum_{ijkl \in D} v_{ijkl}$$

Numerator:
File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newcrime = 1 and V4022 ne 1 and notify = 1 and gender = 1

Denominator:
File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newcrime = 1 and gender = 1

SAS code:

```
libname ncvs "<Location of Incident File>";
```

Numerator:
```
proc means data = ncvs.NCVS_Incident sum;
      where year = 2012 and newcrime = 1 and V4022 ne 1 and notify = 1 and gender = 1;
      var newwgt;
run;
```
Denominator:
```
proc means data = ncvs.NCVS_Incident sum;
      where year = 2012 and newcrime = 1 and gender = 1;
      var newwgt;
run;
```

Numerator: 1,565,448.96
Denominator: 3,770,672.41

$$\frac{100 * 1,565,449}{3,770,672} = 4.15$$

Step 4: Calculate GVF standard error:

$$\sqrt{V_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right)} =$$

$$= \sqrt{3,219 * \frac{41.5 * (100 - 41.5)}{3,770,672} + 8.82 * \frac{41.5 * \left(\sqrt{100 * 41.5} - 41.5\right)}{\sqrt{3,770,672}}}$$

$$= \sqrt{2.07 + 4.32}$$

$$= \sqrt{6.39}$$

$$= 2.53$$

The estimates for females are calculated similarly:

Numerator: 1,458,919
Denominator: 3,160,961

$$\frac{100*1{,}458{,}919}{3{,}160{,}961} = 46.2$$

$$\sqrt{V_p\left(\hat{p}_{A_a,D}, \widehat{N}_D; b, c\right)} = 2.74$$

Compare males and females:

- $\hat{\rho} = 0$ (comparison is between two pooled-year estimates)

$$z = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\widehat{V}_{p_1} + \widehat{V}_{p_2} - 2\hat{\rho}\sqrt{\widehat{V}_{p_1}\widehat{V}_{p_2}}}} = \frac{|41.5 - 46.2|}{\sqrt{2.53^2 + 2.74^2 - 2*(0)*\sqrt{2.53^2 * 2.74^2}}} = \frac{4.7}{\sqrt{13.9}} = \frac{4.7}{3.73} = 1.26$$

Conclusion: Because $1.26 < 1.96$, the estimates for males and females are not significantly different at the 5% level.

4.3 Rates

Example 9. Rate of Simple Assault, 2012

Step 1: Determine appropriate GVF formula:

$$V_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right) = b\frac{\hat{r}_{C,D}\left(1000 - \hat{r}_{C,D}\right)}{\widehat{N}_D} + c\frac{\hat{r}_{C,D}\left(\sqrt{1000\hat{r}_{C,D}} - \hat{r}_{C,D}\right)}{\sqrt{\widehat{N}_D}}$$

Step 2: Determine appropriate GVF parameters:

- Year of interest: 2012
- Are aggregated parameters needed? No
- Personal or property crime? Personal
- Overall or domain estimate? Overall

$b = 6{,}439$
$c = 3.704$

Step 3: Calculate estimate

- Determine appropriate formula for estimate type:

$$\hat{r}_{C,D} = 1000 * \left(\sum_{ijkl \in C,D} v_{ijkl}\right)\bigg/ \widehat{N}_D$$

$$\widehat{N}_D = \sum_{ijk \in D} w_{ijk}$$

Numerator:
File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newoff = 4 and V4022 ne 1

Denominator:
File: Person-level
Weight: wgtpercy
Subpopulation: year = 2012

SAS code:
```
libname ncvs "<Location of Incident and Person Files>";
```

Numerator:
```
proc means data = ncvs.NCVS_Incident sum;
        where year = 2012 and newoff = 4 and V4022 ne 1;
var newwgt;
run;
```

Denominator:
```
proc means data = ncvs.NCVS_Person sum;
        where year = 2012;
var wgtpercy;
run;
```

Numerator: 4,757,902
Denominator: 261,996,322

$$\frac{4,757,902}{261,996,322} * 1,000 = 18.2$$

Step 4: Calculate GVF standard error:

$$\sqrt{V_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right)} =$$

$$= \sqrt{6,439 * \frac{18.2*(1,000-18.2)}{261,996,322} + 3.704 * \frac{18.2*(\sqrt{1,000*18.2}-18.2)}{\sqrt{261,996,322}}}$$

$$= \sqrt{0.438 + 0.485}$$

$$= \sqrt{0.923}$$

$$= 0.961$$

Example 10. Rate of Simple Assault Vs. Rate of Aggravated Assault, 2012

NOTE: The rate of simple assault for 2012 was calculated in Example 9. First, repeat steps for calculation of aggravated assault in 2012.

Step 1: Determine appropriate GVF formula:

$$V_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right) \;=\; b\,\frac{\hat{r}_{C,D}\left(1000 - \hat{r}_{C,D}\right)}{\widehat{N}_D} + c\,\frac{\hat{r}_{C,D}\left(\sqrt{1000\hat{r}_{C,D}} - \hat{r}_{C,D}\right)}{\sqrt{\widehat{N}_D}}$$

Step 2: Determine appropriate GVF parameters:

- Year of interest: 2012
- Are aggregated parameters needed? No
- Personal or property crime? Personal
- Overall or domain estimate? Overall

b = 6,439
c = 3.704

Step 3: Calculate estimate

- Determine appropriate formula for estimate type:

$$\hat{r}_{C,D} \;=\; 1000 * \left(\sum_{ijkl \in C,D} v_{ijkl}\right) \Big/ \widehat{N}_D$$

$$\widehat{N}_D \;=\; \sum_{ijk \in D} w_{ijk}$$

Numerator:
File: Incident-level
Weight: newwgt
Subpopulation: year = 2012 and newoff = 3 and V4022 ne 1

Denominator:
File: Person-level
Weight: wgtpercy
Subpopulation: year = 2012

SAS code:
```
libname ncvs "<Location of Incident and Person Files>";
```

Numerator:
```
proc means data = ncvs.NCVS_Incident sum;
        where year = 2012 and newoff = 3 and V4022 ne 1;
var newwgt;
run;
```

Denominator:
```
proc means data = ncvs.NCVS_Person sum;
        where year = 2012;
var wgtpercy;
run;
```

Numerator = 996,106
Denominator = 261,996,322

$$\frac{996,106}{261,996,322} * 1,000 = 3.80$$

Step 4: Calculate GVF standard error:

$$\sqrt{V_r(\hat{r}_{C,D}, \widehat{N}_D; b, c)} =$$

$$= \sqrt{6,439 * \frac{3.80 * (1,000-3.80)}{261,996,322} + 3.704 * \frac{3.80*(\sqrt{1,000*3.80}-3.80)}{\sqrt{261,996,322}}}$$

$$= \sqrt{0.0931 + 0.0503}$$

$$= \sqrt{0.143}$$

$$= 0.378$$

Finally, compare rates of simple and aggravated assault:

Step1: GVF Comparison Formulas:

$$z = \frac{|\hat{r}_1 - \hat{r}_2|}{\sqrt{\widehat{V}_{r_1} + \widehat{V}_{r_2} - 2\hat{\rho}\sqrt{\widehat{V}_{r_1}\widehat{V}_{r_2}}}}$$

Step 2: Specify $\hat{\rho}$

- Correlation years: n/a, $\hat{\rho} = 0$ for estimates within the same year
- Crime type: n/a
- $\hat{\rho} = 0$

Step 3: Determine significance:

$$z = \frac{|\hat{r}_1 - \hat{r}_2|}{\sqrt{\hat{V}_{r_1} + \hat{V}_{r_2} - 2\hat{\rho}\sqrt{\hat{V}_{r_1}\hat{V}_{r_2}}}} = \frac{|18.2 - 3.80|}{\sqrt{0.961^2 + 0.378^2 - 2*(0)*\sqrt{0.961^2*0.378^2}}} = \frac{14.4}{\sqrt{1.07}} = \frac{14.4}{1.03} = 14.0$$

Conclusion: Because $14.0 > 1.96$, the 2012 rates of simple and aggravated assault are significantly different at the 5% level.

Example 11. Rate of Property Crimes Reported to Police, 2010–2012

Step 1: Determine appropriate GVF formula:

$$V_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right) = b\frac{\hat{r}_{C,D}\left(1000 - \hat{r}_{C,D}\right)}{\widehat{N}_D} + c\frac{\hat{r}_{C,D}\left(\sqrt{1000\hat{r}_{C,D}} - \hat{r}_{C,D}\right)}{\sqrt{\widehat{N}_D}}$$

Step 2: Determine appropriate GVF parameters:

- Years of interest: 2010–2012
- Are aggregated parameters needed? Yes
- Personal or property crime? Property
- Overall or domain estimate? Overall
- $\hat{\rho} = 0.47$

Parameter	2010	2011	2012
b	2,450	4,182	2,345
c	2.035	2.809	2.116

Let Y = the number of pooled years. Then the aggregated, or pooled, parameters b_p and c_p are calculated as follows:

$$b_p = \left(\frac{\sum_{i=1}^{Y} b_i}{Y}\right) * \left(\frac{Y + \hat{\rho} * (3Y - 4)}{Y}\right)$$

$$= \left(\frac{2,450+4,182+2,345}{3}\right) * \left(\frac{3+0.47*(3*(3)-4)}{3}\right)$$

$$= \left(\frac{8,977}{3}\right) * \left(\frac{5.35}{3}\right)$$

$$= 2,992.333*1.783$$

$$= 5,336$$

$$c_p = \left(\frac{\sum_{i=1}^{Y} c_i}{Y}\right) * \left(\frac{Y + \hat{\rho} * (3Y - 4)}{Y^{1.5}}\right)$$

$$= \left(\frac{2.035+2.809+2.116}{3}\right) * \left(\frac{3+0.47*(3*(3)-4)}{3^{1.5}}\right)$$

$$= \left(\frac{6.96}{3}\right) * \left(\frac{5.35}{5.196152423}\right)$$

$$= 2.32 * 1.03$$

$$= 2.39$$

Step 3: Calculate estimate

- Determine appropriate formula for estimate type:

$$\hat{r}_{C,D} = 1000 * \left(\sum_{ijkl \in C,D} v_{ijkl} \right) \bigg/ \hat{N}_D$$

$$\hat{N}_D = \sum_{ijk \in D} w_{ijk}$$

Numerator:
File: Incident-level
Weight: newwgt
Subpopulation: year ge 2010 and year le 2012 and newcrime = 3 and notify = 1 and V4022 ne 1

Denominator:
File: Household-level
Weight: wgthhcy
Subpopulation: year ge 2010 and year le 2012

SAS code:
```
libname ncvs "<Location of Incident and Household Files>";
```

Numerator:
```
proc means data = ncvs.NCVS_Incident sum;
      where year ge 2010 and year le 2012 and newcrime = 3 and notify = 1 and V4022 ne
1;
      var newwgt;
run;
```

Denominator:
```
proc means data = ncvs.NCVS_Household sum;
      where year ge 2010 and year le 2012;
var wgthhcy;
run;
```

Numerator = 18,911,841
Denominator = 371,844,199

$$\frac{18,911,841}{371,844,199} * 1000 = 50.9$$

Step 4: Calculate GVF standard error:

$$\sqrt{V_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right)} =$$

$$= \sqrt{5{,}336 * \frac{50.9*(1{,}000-50.9)}{371{,}844{,}199} + 2.39 * \frac{50.9*(\sqrt{1{,}000*50.9}-50.9)}{\sqrt{371{,}844{,}199}}}$$

$$= \sqrt{0.693 + 1.10}$$

$$= \sqrt{1.79}$$

$$= 1.34$$

Example 12. Rate of Property Crimes Reported to Police: Northeast Vs. Midwest Regions, 2010–2012

First, calculate the rate for the Northeast region:

Step 1: Determine appropriate GVF formula:

$$V_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right) = b\frac{\hat{r}_{C,D}\left(1000 - \hat{r}_{C,D}\right)}{\widehat{N}_D} + c\frac{\hat{r}_{C,D}\left(\sqrt{1000\hat{r}_{C,D}} - \hat{r}_{C,D}\right)}{\sqrt{\widehat{N}_D}}$$

Step 2: Determine appropriate GVF parameters:

- Years of interest: 2010–2012
- Are aggregated parameters needed? Yes
- Personal or property crime? Property
- Overall or domain estimate? Domain
- $\hat{\rho} = 0.47$

Parameter	2010	2011	2012
b	3,297	2,981	2,810
c	1.687	3.852	2.057

Let Y = the number of pooled years. Then the aggregated, or pooled, parameters b_p and c_p are calculated as follows:

$$b_p = \left(\frac{\sum_{i=1}^{Y} b_i}{Y}\right) * \left(\frac{Y + \hat{\rho} * (3Y - 4)}{Y}\right)$$

$$= \left(\frac{3{,}297+2{,}981+2{,}810}{3}\right) * \left(\frac{3+0.47*(3*(3)-4)}{3}\right)$$

$$= 3{,}029.333*1.783$$

$$= 5{,}402$$

$$c_p = \left(\frac{\sum_{i=1}^{Y} c_i}{Y}\right) * \left(\frac{Y + \hat{\rho} * (3Y - 4)}{Y^{1.5}}\right)$$

$$= \left(\frac{1.687 + 3.852 + 2.057}{3}\right) * \left(\frac{3 + 0.47 * (3 * (3) - 4)}{3^{1.5}}\right)$$

$$= 2.532 * 1.03$$

$$= 2.61$$

Step 3: Calculate estimate

- Determine appropriate formula for estimate type:

$$\hat{r}_{C,D} = 1000 * \left(\sum_{ijkl \in C,D} v_{ijkl}\right) \Big/ \hat{N}_D$$

$$\hat{N}_D = \sum_{ijk \in D} w_{ijk}$$

Numerator:
File: Incident-level
Weight: newwgt
Subpopulation: year ge 2010 and year le 2012 and newcrime = 3 and notify = 1 and V4022 ne 1 and region = 1

Denominator:
File: Household-level
Weight: wgthhcy
Subpopulation: year ge 2010 and year le 2012 and region = 1

SAS code:
```
libname ncvs "<Location of Incident and Household Files>";
```

Numerator:
```
proc means data = ncvs.NCVS_Incident sum;
where year ge 2010 and year le 2012 and newcrime = 3 and notify = 1 and V4022 ne 1 and region = 1;
    var newwgt;
run;
```

Denominator:
```
proc means data = ncvs.NCVS_Household sum;
    where year ge 2010 and year le 2012 and region = 1;
    var wgthhcy;
run;
```

Numerator = 2,465,559
Denominator = 68,462,070

$$\frac{2,465,559}{68,462,070} * 1,000 = 36.0$$

Step 4: Calculate GVF standard error:

$$\sqrt{V_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right)} =$$

$$= \sqrt{5,402 * \frac{36.0*(1,000-36.0)}{68,462,070} + 2.61 * \frac{36.0*(\sqrt{1,000*36.0}-36.0)}{\sqrt{68,462,070}}}$$

$$= \sqrt{2.74 + 1.75}$$

$$= \sqrt{4.49}$$

$$= 2.12$$

Next, calculate the rate for the Midwest region:

Step 1: Determine appropriate GVF formula:

$$V_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right) = b\frac{\hat{r}_{C,D}\left(1000 - \hat{r}_{C,D}\right)}{\widehat{N}_D} + c\frac{\hat{r}_{C,D}\left(\sqrt{1000\hat{r}_{C,D}} - \hat{r}_{C,D}\right)}{\sqrt{\widehat{N}_D}}$$

Step 2: Determine appropriate GVF parameters:

- Years of interest: 2010–2012
- Are aggregated parameters needed? Yes
- Personal or property crime? Property
- Overall or domain estimate? Domain
- $\hat{\rho} = 0.47$

Note: aggregate parameters are the same as for the Northeast.

$$b_p = \left(\frac{\sum_{i=1}^{Y} b_i}{Y}\right) * \left(\frac{Y + \hat{\rho} * (3Y - 4)}{Y}\right)$$

$$= \left(\frac{3,297+2,981+2,810}{3}\right) * \left(\frac{3+0.47*(3*(3)-4)}{3}\right)$$

$$= 3029.333 * 1.783$$

$$= 5,402$$

$$c_p = \left(\frac{\sum_{i=1}^{Y} c_i}{Y}\right) * \left(\frac{Y + \hat{\rho} * (3Y - 4)}{Y^{1.5}}\right)$$

$$= \left(\frac{1.687 + 3.852 + 2.057}{3}\right) * \left(\frac{3 + 0.47 * (3*(3) - 4)}{3^{1.5}}\right)$$

$$= 2.532 * 1.03$$

$$= 2.61$$

Step 3: Calculate estimate

- Determine appropriate formula for estimate type:

$$\hat{r}_{C,D} = 1000 * \left(\sum_{ijkl \in C,D} v_{ijkl}\right) \Big/ \hat{N}_D$$

$$\hat{N}_D = \sum_{ijk \in D} w_{ijk}$$

Numerator:
File: Incident-level
Weight: newwgt
Subpopulation: year ge 2010 and year le 2012 and newcrime = 3 and notify = 1 and V4022 ne 1 and region = 2

Denominator:
File: Household-level
Weight: wgthhcy
Subpopulation: year ge 2010 and year le 2012 and region = 2

SAS code:
```
libname ncvs "<Location of Incident and Household Files>";
```

Numerator:
```
proc means data = ncvs.NCVS_Incident sum;
where year ge 2010 and year le 2012 and newcrime = 3 and notify = 1 and V4022 ne 1 and region = 2;
var newwgt;
run;
```

Denominator:
```
proc means data = ncvs.NCVS_Household sum;
    where year ge 2010 and year le 2012 and region = 2;
    var wgthhcy;
run;
```

Numerator: 4,538,627
Denominator: 85,397,321

$$\frac{4,538,627}{85,397,321} * 1000 = 53.1$$

Step 4: Calculate GVF standard error:

$$\sqrt{V_r\left(\hat{r}_{C,D}, \widehat{N}_D; b, c\right)} =$$

$$= \sqrt{5,402 * \frac{53.1 * (1,000 - 53.1)}{85,397,321} + 2.61 * \frac{53.1 * (\sqrt{1,000 * 53.1} - 53.1)}{\sqrt{85,397,321}}}$$

$$= \sqrt{3.18 + 2.66}$$

$$= \sqrt{5.84}$$

$$= 2.42$$

Finally, compare rates for Northeast and Midwest regions:

Step1: GVF comparison formulas:

$$z = \frac{|\hat{r}_1 - \hat{r}_2|}{\sqrt{\widehat{V}_{r_1} + \widehat{V}_{r_2} - 2\hat{\rho}\sqrt{\widehat{V}_{r_1}\widehat{V}_{r_2}}}}$$

Step 2: Specify $\hat{\rho}$

- Correlation years: n/a, $\hat{\rho} = 0$ for estimates within the same year and for aggregate years
- Crime type: n/a
- $\hat{\rho} = 0$

Step 3: Determine significance:

$$z = \frac{|\hat{r}_1 - \hat{r}_2|}{\sqrt{\widehat{V}_{r_1} + \widehat{V}_{r_2} - 2\hat{\rho}\sqrt{\widehat{V}_{r_1}\widehat{V}_{r_2}}}} = \frac{|36.0 - 53.1|}{\sqrt{2.12^2 + 2.42^2 - 2*(0)*\sqrt{2.12^2 * 2.42^2}}} = \frac{17.1}{\sqrt{10.4}} = \frac{17.1}{3.22} = 5.31$$

Conclusion: Because 5.31 > 1.96, the rates for the Northeast and Midwest regions are significantly different at the 5% level.

SECTION 5. REFERENCES

Ash, S., Hornick, D., McCune, J., Rottach, R., Sanduckchi, M., & Warner, T. (2008). *Variance methods of the Demographic Statistical Methods Division*. U.S. Census Bureau.

Cochran, W. G. (1977). *Sampling techniques*. New York, NY: Wiley.

Shook-Sa, B., Couzens, G. L, & Berzofsky, M. (2015). *National Crime Victimization Survey (NCVS) direct variance user's guide*. Prepared for the Bureau of Justice Statistics, Washington, DC.