Proceedings of Statistics Canada Symposium 2001
Achieving Data Quality in Statistical Agency: A Methodological Perspective

DATA ESTIMATION IN THE UNIFORM CRIME REPORTING (UCR) PROGRAM

Yoshio Akiyama[1] ,Samuel Berhanu[2]

ABSTRACT

In order to make up for unreported, missing, unreasonable, or unusable data, the UCR Program conducts data estimations/imputations using a variety of statistical methods. In this paper, the authors discuss and illustrate how offense and arrest data are estimated using a variety of approaches. The paper also points out the strength and the shortcomings of each approach.

KEY WORDS: Data Estimation; Offense and Arrest Data; Strata; UCR Program.

1. INTRODUCTION

Besides collecting, storing, processing, and disseminating crime data, the UCR Program performs a variety of simple and advanced statistical computations to provide insight into the national, regional, and local crime phenomena. The Program's major annual publications such as *Crime in the United States*, *Law Enforcement Officers Killed and Assaulted*, and Hate Crime Statistics present statistical information. The routine modes of data presentation of the national UCR Program include statistics such as percent change, crime rate, crime clock statistics, age-specific arrest rate, percent of crimes cleared, police employee rate, etc. These statistics are expressed in a simple nontechnical language that most people would understand.

For the above publications, estimation/imputation of crime data is conducted when the reported data from agencies are incomplete or unacceptable. The UCR Program's policy of estimating/imputing crime data has been in effect since 1960. The estimation of crime data for each agency is based on the size of the agency, the degree of urbanization, geography, and the agency type. The methodology of such data estimation is outlined in Appendix I of Crime in the United States. The shortcomings of this methodology are:

1. Large number of strata: Each state is stratified by eight UCR population groups and by Metropolitan Statistical Areas (MSA) designations. Formally, this gives rise to 11 strata for each state. However, this number of strata is too large, represents an over-refined stratification, and gives rise to a small number of agencies within a given stratum. The number of agencies for certain strata was often too small to estimate the missing data.

[1]Senior Statistician, Federal Bureau of Investigation, J. Edgar Hoover Building, Room 11194, 935 Pennsylvania Avenue, Washington, D.C., USA 20535.

[2] Statistician, Crime Analysis, Research, and Development Unit, Criminal Justice Information Services Division, Federal Bureau of Investigation, 1000 Custer Hollow Road, Clarksburg, WV, USA 26306.

2. Arbitrary cut-point of three usable months for Return A Reports: Incomplete agencies are classified based on the arbitrary threshold of 3 months of usable reports. Class I agencies are those that submitted complete (12 months) usable reports for all index crimes. Class II agencies contain

agencies that reported less than 3 usable monthly reports. Finally, Class III agencies encompass those that reported 3 to 11 usable monthly Return A reports.

Data for Class II agencies are estimated by the ratio estimation (i.e., proration by population). Data for Class III agencies are estimated by the proration by months. The ramification of agencies into Classes II and III is based on the arbitrary policy of 3 usable months.

3. The difficulty of computing standard errors: The use of different estimation procedures for Class II and Class III agencies makes the computations of standard errors tedious. It is particularly tedious to associate the precision/accuracy to the estimations for Class III agencies. Historically, precision estimates (standard errors) have not been associated with the UCR crime estimates.
4. The monthly proration for Class III agencies ignores seasonality and data trend: The monthly proration for Class III agencies ignores crime trends and crime seasonality. In order to generate viable annual estimates, data must be de-trended and de-seasoned prior to such prorations.

The main objective of this paper is, therefore, to present the new approach developed by the UCR Program to estimate UCR data.

2. STRATIFICATION OF AGENCIES

In detecting outliers, algorithms compare a given agency report to similar agencies reports. This is done through categorizing agencies into strata. Agencies belonging to the same stratum are called similar agencies. The new estimation procedure is based on the following factors:

- Jurisdictional size (i.e., agency population)
- Geographic region where an agency is located
- Degree of urbanization (i.e., in or outside MSA)
- Agency type (i.e., distinction between police departments and sheriff's offices).

The above four factors are used in the traditional UCR stratification (except the geographic breakdowns that were shown by states in the traditional stratification) and constitute a *minimum* set of variables that impinge upon the jurisdictional crime conditions. Geographic breakdowns are made in terms of the regions (instead of states) in order to increase the size of strata. Larger agencies with over 50,000 in population are not classified by regions. Except for these differences, the new stratification adopts the same variables and the same concept as the old method. One stratification scheme is used for both data quality control (outlier detections in data reasonableness review) and data estimations (crime and arrest estimations).

Most published UCR crime data are complete. Missing/incomplete data (represented by Classes II and III agencies) are exceptions rather than a rule. The UCR Program has actual data for the majority of agencies. Because the estimates are made for a relatively small number of agencies and the total number of agencies is finite, a stratification scheme would have a limited impact on the global estimates. Treated in the framework of sampling, the *finite adjustment factor* (a factor that is present in the variance of the UCR estimate) reduces the estimation errors close to zero. This should be compared to most regular sample surveys where researchers have to estimate the characteristics of the population by a small *sample fraction*.

3. NEW OFFENSE ESTIMATION PROCEDURES

3.1 Ratio Estimations of Offenses

The requirement (in the traditional offense estimation procedure) for 3 usable months is an arbitrary cut-point. Remember that 3 months of usable reports is the dividing line for Class II and III agencies, and monthly proration was performed only to Class III agencies. In order to alleviate the problem arising from this arbitrary cut-point, offense data estimation is made on a monthly basis.

Agencies with positive populations that report UCR data on a monthly basis can have their data estimated with little difficulty. For a given stratum defined as one of the 32 strata (except the one with zero population), let *n* be the number of agencies in the stratum that report UCR crime data on a monthly basis. Then assume that there are *j* agencies whose crime data are not usable (i.e., needs estimation), *k* agencies that received S (Self-Representation), and the remaining *m* = *n* - *j* - *k* agencies whose data are usable, i.e., require no estimation. In this context, a monthly report is usable if it requires no monthly estimation and is not S (Self-Representation).

Let X_i, $1 \le i \le m$, be the monthly usable report from the i^th agency, and P_i be the population of the i^th agency in the stratum. The assumption is that there are *m* such agencies. Let

$$X^* = \sum_{i=1}^{m} X_i, \quad P^* = \sum_{i=1}^{m} P_i, \quad \text{and} \quad r = \frac{X^*}{P^*}. \qquad [1]$$

Let Q_i ($m+1 \le i \le m+j = n-k$) be the population of the agency whose monthly crime data was not usable (i.e., needs estimation) and $Q = \sum Q_i$. The assumption is that there are *j* such agencies. The agency estimated number of crimes (in the whole number) for the month is defined by:

$$\tilde{X}_i = r\,Q_i, \quad m+1 \le i \le m+j = n-k. \qquad [2]$$

The number of crimes for each of the k agencies that received S (Self-Representation) is the actual number $X_i^{(S)}$, $n-k+1 \le i \le n$, that agencies reported. Let

$$\tilde{X} = \sum_{i=m+1}^{n-k} \tilde{X}_i \quad \text{and} \quad X^{(S)} = \sum_{i=n-k+1}^{n} X_i^{(S)}. \qquad [3]$$

Then, the total crime estimate for the month for the stratum is :

$$T = X^* + \tilde{X} + X^{(S)} + (\text{zero-population agencies prorated data}). \qquad [4]$$

For a given category of crime, the annual estimate for the stratum is obtained by adding 12 monthly estimates *T*. The national estimate for the month (and for the crime category) is obtained by adding the monthly estimates over the strata. Since for a given month each agency has the actual number X_i, the estimated number $\tilde{X}_i$ or the self-representing number $X_i^{(S)}$, crime estimate can be obtained for any area by adding appropriate individual agencies' monthly data.

3.2 Variances of the Ratio Estimates

For a given stratum, the above formula (4) defines the estimate of offenses $T = r\,P + X^{(S)} + (\text{others})$ for a crime category and a month, where $P = P^* + Q$ denotes the total population of the stratum, excluding those agencies that received an S (Self-Representation). The ratio *r* is a variable, since it depends on the class of *m* agencies that happen to submit "usable" data for the month. The variance of *r* is known to be approximated by the formula:

$$\sigma_r^2 \approx R^2 \left(1 - \frac{m}{n-k}\right) \frac{V_X^2 + V_P^2 - 2\rho V_X V_P}{m}, \qquad [5]$$

where *V* denotes the *relative precision* and ρ is the correlation coefficient between X_i and P_i. Since *j* (the number of agencies whose data are estimated) is generally small, the finite adjustment factor $1 - m/(n-k) = j/(n-k)$ in the above equation is small. It is noted that the values in *R*, V_X^2, V_P^2, and ρ can be estimated by data from the *m* agencies (considered as a sample) that gave "usable" reports. Then, (5) becomes as below:

$$\sigma_r^2 \approx r^2 \left[\frac{s_X^2}{(X')^2} + \frac{s_P^2}{(P')^2} - 2\,\hat{\rho}\,\frac{s_X s_P}{X' P'} \right], \qquad [6]$$

where s_X^2, s_P^2, and $\hat{\rho}$ denotes sample estimates of σ_X^2, σ_P^2, and ρ.

Since $T = r\,P + X^{(S)} + \text{(others)}$ and $X^{(S)}$ is a constant, the variance of T is given as:

$$\sigma_T^2 = \sigma_r^2\, P^2. \qquad [7]$$

Assuming the independence of months and strata in terms of data estimation, the combined variance is obtained by adding the appropriate variances.

3.3 Kalman Filter Approach to Data Estimation

The Kalman filter technique was initially developed and used by engineers. However, it has found a wide range of applications to other fields such as forecasting and quality control. Below, it is proposed to apply the Kalman filter technique to UCR data estimation in a simplified form (Meinhold, R. J. and Singpurwalla, 1983, 1989).

3.3.1 Kalman Filter Model

At time *t*, let Y_t denote the *observed* value of a variable (in UCR application, Y_t is the reported number of crimes for an agency) and U_t be the *unobserved* state of nature. U_t may be considered as the true but unobservable criminality (crime rate) of the agency. The relationship between Y_t and U_t is specified by the following *observation equation*:

$$Y_t = U_t + \eta_t, \qquad [8]$$

where η_t represents a normal variable $N(0, V_t)$. In the Kalman filter model, the state of nature is assumed to change in time *t*, being controlled by the following *system equation*:

$$U_t = G_t\, U_{t-1} + \xi_t, \qquad [9]$$

where G_t is a known quantity, and ξ_t is a normal variable $N(0, W_t)$ which is statistically independent of η_t.

At time $(t-1)$, knowledge on the state of nature can be expressed by the normal *posterior* distribution as:

$$U_{t-1} \mid \mathcal{U}_{t-1} \sim N(\hat{U}_{t-1}, \Sigma_{t-1}), \qquad [10]$$

where $\mathcal{U}_{t-1}$ represents the set of historical observations $\{Y_0, Y_1, \ldots, Y_{t-1}\}$, $\hat{U}_{t-1}$ denotes the expected value $E(U_{t-1} \mid \mathcal{U}_{t-1})$, and Σ_{t-1} denotes the variance $Var(U_{t-1} \mid \mathcal{U}_{t-1})$. The process is started off at time $t = 0$ by appropriately selecting $\hat{U}_0$ and Σ_0.

Before observing Y_t (which is not reported in our case), the *prior* distribution $U_t \mid \mathcal{U}_{t-1}$ is derived from the system equation. To be more specific we have the following mean and variance:

$$E(U_t \mid \mathcal{U}_{t-1}) = G_t\; E(U_{t-1} \mid \mathcal{U}_{t-1}) + E(\xi_t) = G_t\; \hat{U}_{t-1}. \qquad [11]$$

$$Var(U_t \mid \mathrm{U}_{t-1}) = G_t^2 \; Var(U_{t-1} \mid \mathrm{U}_{t-1}) + Var(\xi_t) \qquad [12]$$

$$= G_t^2 \; \Sigma_{t-1} + W_t .$$

Therefore, the variable $U_t \mid \mathrm{U}_{t-1}$ is normally distributed as below.

$$U_t \mid \mathrm{U}_{t-1} \sim N(G_t \hat{U}_{t-1} ,\; G_t^2 \; \Sigma_{t-1} + W_t) . \qquad [13]$$

If Y_t is observed or reported, the *posterior* distribution $U_t \mid \mathrm{U}_t$ is expressed as below:

$$U_t \mid \mathrm{U}_t \sim N(\hat{U}_t ,\; \Sigma_t) . \qquad [14]$$

The following theorem provides the algorithm for computing the mean $\hat{U}_t$ and the variance Σ_t:

Theorem:

$$\hat{U}_t = G_t \hat{U}_{t-1} + \frac{G_t^2 \; \Sigma_{t-1} + W_t}{V_t + G_t^2 \; \Sigma_{t-1} + W_t} (Y_t - G_t \hat{U}_{t-1}) . \qquad [15]$$

$$\Sigma_t = (G_t^2 \; \Sigma_{t-1} + W_t) - \frac{(G_t^2 \; \Sigma_{t-1} + W_t)^2}{V_t + G_t^2 \; \Sigma_{t-1} + W_t} . \qquad [16]$$

3.3.2 Application Of Kalman Filter Concept to UCR

The Kalman Filter is used when no data are reported throughout the year from a given state. In such situations, it is considered that historical (longitudinal) data of the state provide a better estimation than the cross-sectional ratio estimation. It is assumed that Y_t is not reported in the current context. First, we must define $\hat{U}_0$ and Σ_0. G_t is defined to be the rate of change (between the years $t-1$ and t) in the volume of crimes. This can be obtained from the agencies that submitted usable data for consecutive years in the stratum to which the agency under consideration belongs. It is assumed that $\eta_t \sim N(0 , \hat{U}_t)$, i.e., taking U_t as a Poisson variable with parameter $\lambda = \hat{U}_t$. Therefore, $V_0 = Y_0$ and $V_t = \hat{U}_t$.

W_t is defined as below. U_t was assumed to be Poisson distributed with parameter $\lambda = \hat{U}_t$. The variable $U_t' = G_t U_{t-1}$ is also assumed to be Poisson distributed with the same parameter (since it also pertains to the number of crimes for the year *t*). The difference $(U_t - U_t')$ of the two independent Poisson variables has variance $(2 \hat{U}_t)$. Therefore, it is defined that $W_t = 2 \hat{U}_t$. The prior distribution $U_t \mid \mathrm{U}_{t-1}$ of the state of nature is now computable.

Since we have no usable report Y_t, it is estimated by the formula $\hat{Y}_t = E(U_t \mid \mathrm{U}_{t-1}) = \hat{U}_t = G_t \hat{U}_{t-1}$, which is equivalent to $\hat{Y}_t = (G_t G_{t-1} \ldots G_1) Y_0$. This commonsense estimate $\hat{Y}_t = (G_t G_{t-1} \ldots G_1) Y_0$ is expected from the outset without appealing to the Kalman filter model. But, the Kalman filter model provides the variance to the estimates and utilizes the most recent report of the agency (instead of applying the average of similar agencies).

4. NEW ARREST ESTIMATION PROCEDURE

4.1 Estimating the Number of Arrests

The new arrest estimation is performed for each crime category. In terms of estimation algorithms, the ratio estimation is applied to usable arrest reports. The stratification of agencies is the same scheme defined earlier. For simplicity, a given offense category is assumed below in describing the new procedures involved.

In the i^{th} stratum and for a given offense category, let

n_{ij} = the annual total of arrests reported by the j^{th} usable agency,
p_{ij} = the UCR population of the j^{th} usable agency,

s_{ik} = the annual total of arrests reported by the k^{th} self-representing agency that received S,
q_{ik} = the population of the k^{th} self-representing agency, and
r_{im} = the population of the mth agency that requires arrest estimation.

Let, $n_i = \sum_j n_{ij}$, $p_i = \sum_k p_{ij}$, $s_i = \sum_k s_{ik}$, $q_i = \sum_m q_{ik}$, and $r_i = \sum_m r_{im}$.

The estimated number of arrests for the m^{th} agency (that requires arrest data estimation) is defined by

$$\tilde{n}_{im} = \frac{n_i}{p_i} r_{im}. \qquad [17]$$

It should be noted that [17] is based on the sample ratio estimation of the number of arrests and the population; the sample being the set of usable agencies for the crime category under consideration. The estimated portion of arrests for the ith stratum is

$$\tilde{n}_i = \sum_m \tilde{n}_{im}. \qquad [18]$$

The remaining numbers n_i and s_i are actual. Therefore, the total number of arrests for the i^{th} stratum, $\tilde{N}_i$, is the sum of n_i, s_i, and $\tilde{n}_i$, i.e.,

$$\tilde{N}_i = n_i + s_i + \tilde{n}_i, \qquad [19]$$

There is no estimation made for the zero-population agencies (Stratum 33) so that $\tilde{N}_{33}$ is the sum of numbers actually submitted by zero-population agencies. Therefore, the national arrest estimate $\tilde{N}$ for the prescribed offense category is the sm of N_i's:

$$\tilde{N} = \sum_{i=1}^{33} \tilde{N}_i \qquad [20]$$

4.2 Estimating Arrests by Age and Sex

For a given offense category, let $n_{ij}(\alpha, \beta)$ be the number of persons arrested for a given age α and sex β. The number $n_{ij}(\alpha, \beta)$ arises from the class of usable agencies in the i^{th} stratum. Using the index j for "usable" agencies as before, we have

$$n_{ij} = \sum_\alpha \sum_\beta n_{ij}(\alpha, \beta). \qquad [21]$$

Let, $n_i(\alpha, \beta) = \sum_j n_{ij}(\alpha, \beta)$. Then, $n_i = \sum_{\alpha, \beta} n_i(\alpha, \beta)$.

Similarly, let $s_{i,k}(\alpha, \beta)$ be the number of persons arrested with age α and sex β for self-representing agencies, and $s_i(\alpha, \beta) = \sum_k s_{ik}(\alpha, \beta)$. (self-representing agencies submit the age and sex data for persons arrested.) Then,

$$s_{ik} = \sum_\alpha \sum_\beta s_{ik}(\alpha, \beta). \qquad [22]$$

if we denote $s_i(\alpha, \beta) = \sum_k s_{ik}(\alpha, \beta)$, then $s_i = \sum_{\alpha, \beta} s_i(\alpha, \beta)$.

For the Agency m in the i^{th} stratum (whose arrest data was estimated), the estimated number of arrests for age α and sex β is defined by

$$\tilde{n}_{im}(\alpha, \beta) = \frac{n_i(\alpha, \beta)}{n_i} \tilde{n}_{im} = \frac{n_i(\alpha, \beta)}{p_i} r_{im}.$$

The estimate for the i^{th} stratum is therefore

$$\tilde{N}_i(\alpha, \beta) = n_i(\alpha, \beta) + s_i(\alpha, \beta) + \frac{n_i(\alpha, \beta)}{n_i} \tilde{n}_i, \qquad [23]$$

The above discussion does not include Stratum 33 of zero-population agencies. In order to include Stratum 33 the following definitions are made. Let the national estimate of arrests for age α and sex β (excluding Stratum 33) be defined by

$$\hat{N}(\alpha,\beta) = \sum_{i=1}^{32} \tilde{N}_i(\alpha,\beta) \text{ and} \qquad [24]$$

$$\hat{N} \equiv \sum_{\alpha}\sum_{\beta} \hat{N}(\alpha,\beta).$$

For the agency z in Stratum 33, let $n_{33,z}(\alpha,\beta)$ be the actual number of arrestees with age α and sex β. This data may not contain the complete arrest statistics of the year. The estimate for the ^{mth} agency is defined by $\tilde{n}_{33,z}(\alpha,\beta) = n_{33,z}(\alpha,\beta)$, if Agency z reported the age and sex data. Otherwise, the estimate is defined by:

$$\tilde{n}_{33,z}(\alpha,\beta) = \frac{\hat{N}(\alpha,\beta)}{\hat{N}}\, n_{33,z}.$$

Now, with the notation $\tilde{N}_{33}(\alpha,\beta) = \sum_{z} \tilde{n}_{33,z}(\alpha,\beta)$, we define the national estimate of arrests for age α and sex β by

$$\tilde{N}(\alpha,\beta) = \sum_{i=1}^{33} \tilde{N}_i(\alpha,\beta) = \hat{N}(\alpha,\beta) + \tilde{N}_{33}(\alpha,\beta).$$

It is noted that

$$\tilde{N}_i = \sum_{\alpha}\sum_{\beta} \tilde{N}_i(\alpha,\beta), \text{ and} \qquad [25]$$

$$\tilde{N} = \sum_{\alpha}\sum_{\beta} \tilde{N}(\alpha,\beta).$$

4.3 Estimating Arrests by Race

A slightly different method is used to estimate the number of arrests for a given race γ. The basis of arrestee race estimation is derived from usable agencies that also reported the race information on arrested persons.

Let the usable agency (in the i^{th} stratum) that reported the race information of the arrestees be indexed by $\hat{j}$. Let $n_{i\hat{j}}(\gamma)$ denote the number of arrestees with race classification γ, and $p_{i\hat{j}}$ be the agency population. Let $n_i(\gamma) = \sum_{\hat{j}} n_{i\hat{j}}(\gamma)$, $n_i^{(r)} = \sum_{\gamma} n_i(\gamma)$, and $p_i^{(r)} = \sum_{\hat{j}} p_{i\hat{j}}$. The superscript (r) is a reminder that the figure relates to race information. We have the following inequalities: $n_i^{(r)} \leq n_i$ and $p_i^{(r)} \leq p_i$.

Let the index $\hat{k}$ denote the self-representing agency that reported the race information of the arrestees. For the agency $\hat{k}$, let $s_{i\hat{k}}(\gamma)$ denote the number of arrested persons with race γ and $q_{i\hat{k}}$ be the corresponding population. Let $s_i(\gamma) = \sum_{\hat{k}} s_{i\hat{k}}(\gamma)$, $s_i^{(r)} = \sum_{\gamma} s_i(\gamma)$, and $q_i^{(r)} = \sum_{\hat{k}} q_{i\hat{k}}$. Clearly,

$s_i^{(r)} \leq s_i$ and $q_i^{(r)} \leq q_i$.

The population covered by the race information of arrestees is equal to $p_i^{(r)} + q_i^{(r)}$. Therefore, the population uncovered by the race information is given by $u_i^{(r)} = p_i + q_i + r_i - p_i^{(r)} - q_i^{(r)}$. The estimate of arrests for the race category γ for the agency t that did not report the race information of the arrestees is defined as follows:

$$\tilde{n}_{it}(\gamma) = \frac{n_i(\gamma)}{n_i^{(r)}}\, n_{it}, \qquad [26]$$

if $t\ (\neq j)$ represents an admissible agency that did not report the arrestees race.

$$\tilde{n}_{it}(\gamma) = \frac{n_i(\gamma)}{n_i^{(r)}}\, s_{it}, \qquad [27]$$

if $t\ (= k)$ represents a self-representing agency that did not report the arrestees race.

$$\tilde{n}_{it}(\gamma) = \frac{n_i(\gamma)}{n_i^{(r)}}\,\tilde{n}_{it}, \qquad [28]$$

if $t\ (= m)$ represents an agency that required arrest estimation.

Obviously, $\tilde{n}_{it}(\gamma) = n_{ij}(\gamma)$ if $t\ (= j)$ represents usable agency that reported the arrestees race. Likewise, $\tilde{n}_{it}(\gamma) = s_{it}(\gamma)$ if $t\ (= k)$ is a self-representing agency that reported the arrestees race information. When totaled over γ, $\tilde{n}_{it}(\gamma)$ add to the agency's actual or estimated number of arrests provided by the new arrest estimation procedure.For the i[th] stratum, the total estimated arrests for the race category γ is given by

$$\tilde{N}_i(\gamma) = \sum_t \tilde{n}_{it}(\gamma) = n_i(\gamma) + s_i(\gamma) + \frac{n_i(\gamma)}{n_i^{(r)}}\,[\tilde{N}_i - n_i^{(r)} - s_i^{(r)}], \qquad [29]$$

where $n_i^{(r)} = \sum_\gamma n_i(\gamma)$, $s_i^{(r)} = \sum_\gamma s_i(\gamma)$, and $i = 1, 2, \ldots, 32$. Therefore, we have

$$\sum_\gamma \tilde{N}_i(\gamma) = \tilde{N}_i\,. \qquad [30]$$

Up to this point, agencies in Stratum 33 (zero-population agencies) have not been addressed. In order to compute the race estimates for Stratum 33 agencies, the following definitions are made:

$$\hat{N}(\gamma) = \sum_{i=1}^{32} \tilde{N}_i(\gamma), \quad \hat{N} = \sum_\gamma \hat{N}(\gamma).$$

For Agency z in Stratum 33, we define $\tilde{n}_{33,z} = n_{33,z}$ if the agency reported the race information of the arrestees. Otherwise, we define

$$\tilde{n}_{33,z}(\gamma) = \frac{\hat{N}(\gamma)}{\hat{N}}\, n_{33,z}.$$

With the notation $\tilde{N}_{33}(\gamma) = \sum_z \tilde{n}_{33,z}(\gamma)$, we define $\tilde{N}(\gamma) = \sum_{i=1}^{33} \tilde{N}_i(\gamma)$. Then,

$$\tilde{N}_i = \sum_\gamma \tilde{N}_i(\gamma), \text{ for } i = 1, 2, \ldots, 33, \qquad [31]$$

and therefore,

$$\tilde{N} = \sum_\gamma \tilde{N}(\gamma).$$

The new arrest estimation procedure is different from the traditional arrest data estimation procedures. The new arrest data estimation is based on a more refined stratification and distinguishes self-representing agencies. The new age-specific and race-specific estimations are based on multiple-strata instead of one stratum, and defines compatible arrest estimations on age, sex, and race for each agency (compatible in the sense that at the agency level both the age and sex breakdown and the race breakdown add to the actual or estimated number of arrests).

5. CONCLUSION

The UCR program has undergone a dynamic process of improving the methodologies that are used in estimating offense and arrest data. The process not only gives the estimates but also provides the level of statistical precision. The new method represents an improvement over the traditional procedure in many respects. The result obtained by using this approach is far more superior to that obtained using the traditional approach. These improvements imply that the quality of UCR data is improved substantially.

REFERENCES

Meinhold, R. J. and Singpurwalla, (1983), AUnderstanding the Kalman Filter@, The American Statistician, Vol. 37, No.2., pp. 123-127.

Meinhold, R. J. and Singpurwalla, (1989), ARobustification of Kalman Filter Models@, Journal of the American Statistical Association, June 1989, Vol. 84, No. 406, pp. 479-486