Procedures for Estimating in the Uniform Crime Reporting (UCR) Program

James H. Noonan, PhD

The UCR Program will begin offense estimation in the 2013 *Crime in the United States* (*CIUS*) publication. The following information details the intended method for achieving a valid and reliable estimate and uses an example of Arson offense estimation to demonstrate the employed statistical model.

Currently, the UCR Program publishes Arson offense data by agency population group, shown in Arson Table 1, and by individual law enforcement agency (*CIUS* Tables 8-12, not shown). Offense totals are represented as a rate per 100,000 persons. Based on aggregated population group rates in Arson Table 1 and individual agency data, the UCR Program is capable of reporting agency imputations and national estimations for arson offenses in its publications.

Arson Table 1[1]

Arson Rate

by Population Group, 2011
[13,045 agencies; 2011 estimated population

263,872,098; rate per 100,000 inhabitants]

Population group	Rate
TOTAL ALL AGENCIES	**18.2**
TOTAL CITIES	20.7
Group I (250,000 and over)	32.8
1,000,000 and over (Group I subset)	28.2
500,000 to 999,999 (Group I subset)	31.8
250,000 to 499,999 (Group I subset)	38.6
Group II (100,000 to 249,999)	20.7
Group III (50,000 to 99,999)	16.9
Group IV (25,000 to 49,999)	14.9
Group V (10,000 to 24,999)	13.3
Group VI (under 10,000)	20.3
Metropolitan counties	13.2
Nonmetropolitan counties	12.2
Suburban areas[2]	12.8

[1] US Federal Bureau of Investigation (2012). *Arson Table 1*. Retrieved from http://www.fbi.gov/about us/cjis/ucr/crime-in-the-u.s/2011/crime-in-the-u.s.-2011/tables/arson-table-1

[2] Suburban areas include law enforcement agencies in cities with less than 50,000 inhabitants and county law enforcement agencies that are within a Metropolitan Statistical Area. Suburban areas exclude all metropolitan agencies associated with a principal city. The agencies associated with suburban areas also appear in other groups within this table.

National Estimation

For 2012, 17,010 agencies reported data to the FBI, and of those 15,387 reported a complete 12 months of data.[3] However, for arson, only 6,292 reported 12 months of data which means the missing information must be imputed. National estimates for offense data will be computed by summing agencies reporting a complete set of data with agencies who are estimated using one of two imputation methods based on the number of months the agency reported for the calendar year.

Number of Agencies by Number of Months of Competed UCR Reporting

Number of Months	Frequency	Percent	Valid Percent	Cumulative Percent
1	102	.6	.6	.6
2	94	.6	.6	1.2
3	141	.8	.8	2.0
4	70	.4	.4	2.4
5	87	.5	.5	2.9
6	91	.5	.5	3.4
7	108	.6	.6	4.1
8	116	.7	.7	4.8
9	152	.9	.9	5.6
10	226	1.3	1.3	7.0
11	436	2.6	2.6	9.5
12	15,387	90.5	90.5	100.0
Total	17,010	100.0	100.0	

Imputation for Agencies with 3-11 Months of Data

For agencies reporting 3-11 months of data to the FBI, it is assumed the agency best represents itself for imputing missing months. As such, the follow formula will be used to impute the agency's arson data:

$$Annual\ Reported \frac{Arson * 12}{Number\ of\ Reported\ Months} = Estimated\ Arson\ Count$$

For example if an agency reported 15 arsons in seven months of reporting there would be an estimated 26 arsons for the agency.

[3] FBI UCR Program. (2010). Crime by County Datafile. Processed May 13, 2013.

$$\frac{15*12}{7}=25.7\cong 26$$

Imputation for Agencies with Less Than Three Months of Data

For agencies reporting less than three months of data, the FBI UCR Program will impute the entire year of arson data based on the rate of arson per 100,000 persons within the agency's population group multiplied by the agency's population.

$$Population\ Group\ Arson\ \frac{Rate*Agency\ Population}{100{,}000}=Estimated\ Arson\ Count$$

For example, Philadelphia arson data was not reported in 2012. The arson rate for agencies over one million in population is 28.2 per 100,000 persons and Philadelphia has a population of 1,538,957 persons. Therefore its estimated arson count for 2012 is 434.

$$\frac{28.2*1{,}538{,}957\ persons}{100{,}000}=433.99\cong 434$$

Combining Estimates With Reported Totals

Once the agencies are imputed, their arson totals will be summed with the agencies that reported a complete 12 months of arson data to derive the national estimate. For 2012, there were 53,941 reported arsons by 12 month agencies and 12,397 estimated arsons. Therefore, the estimated number of arsons in the United States in 2012 is 66,338 arsons.

$$53{,}941+12{,}397=66{,}338$$

Limitations and Future Improvements

The imputation method demonstrated in this paper's estimation model is based on the assumption that agencies within population groups are most similar to one another. Further study needs to be done to determine if regional characteristics of crime would help better stratify agencies into more representative groups. Until other characteristics, such as poverty rates, employment rates, or other socioeconomic variables become attributable to specific agencies and statistically tested for validity, the best stratification variable the FBI UCR Program has is agency population group. The FBI UCR Program is also assuming that the best representation of an agency's missing data is the agency's data reported in other months, but there is no evidence which definitively shows that a partial report's own data qualifies as a representative sample.

Once new variables are tested and validated, other techniques for imputation may be used, such as regression analysis. Other types of imputation models, such as bootstrapping, are not currently available with the FBI UCR Program's current technological or staffing resources, however, the Program is working toward a partnership with the Bureau of Justice Statistics (BJS) which will allow for shared resources in the future which could handle more sophisticated and statistically valid and reliable methods. Arson may have seasonal variation that will need to be studied and identified for use in future estimations.

The 3,012 agencies with zero-population are not represented in the imputation model, however, 530 reported arson, and 962 did not report any type of crime offense. The proportion of the remaining 1,520 agencies which did report some type of crime offenses but simply did not have any arson to report is unknown. The FBI UCR Program will examine the value of longitudinal and ratio-based models for future estimation methods. For example, there may be a reliable ratio of burglaries to arson which can best represent missing data, including those with zero-population. Using an agency's previous reporting patterns over the last five years may also yield valuable estimation results, but these models are not yet tested. Discussions concerning these methods with the BJS showed there is value in examining new cross-sectional, longitudinal, and ratio models, but they will take some time to develop and validate.

A final potential future improvement for nationally estimating crime data is testing UCR data for representativeness. If the data are tested and found to support the idea that UCR data is a statistically representative sample of reported crime data, then sampling estimation models can be used to estimate national reported crime totals. Of course, should the tests show the data is not representative, the other aforementioned imputation models will still be available to the FBI UCR Program to meet the need for determining national crime estimates.

Given the limitations, this estimation proposal is proposed as the next step for the FBI UCR Program to enhance its value for estimating crime data. It is not intended as a once-and-for-all solution for estimation. It is also expected that a current National Academy of Sciences and future partnering opportunities with the BJS and other UCR stakeholders will provide recommendations for estimation that are not currently considered or anticipated in this document.

This proposed method also has other uses within the UCR Program, including providing estimations for other reported data collections, including Part I crimes, police employment data, arrest data, hate crime, cargo theft, human trafficking, and the Law Enforcement Officers Killed and Assaulted program. Specific methodological considerations need to be addressed for unique data concerns within each collection, but a method similar to the arson estimation example can be used as the base estimation model for those programs.