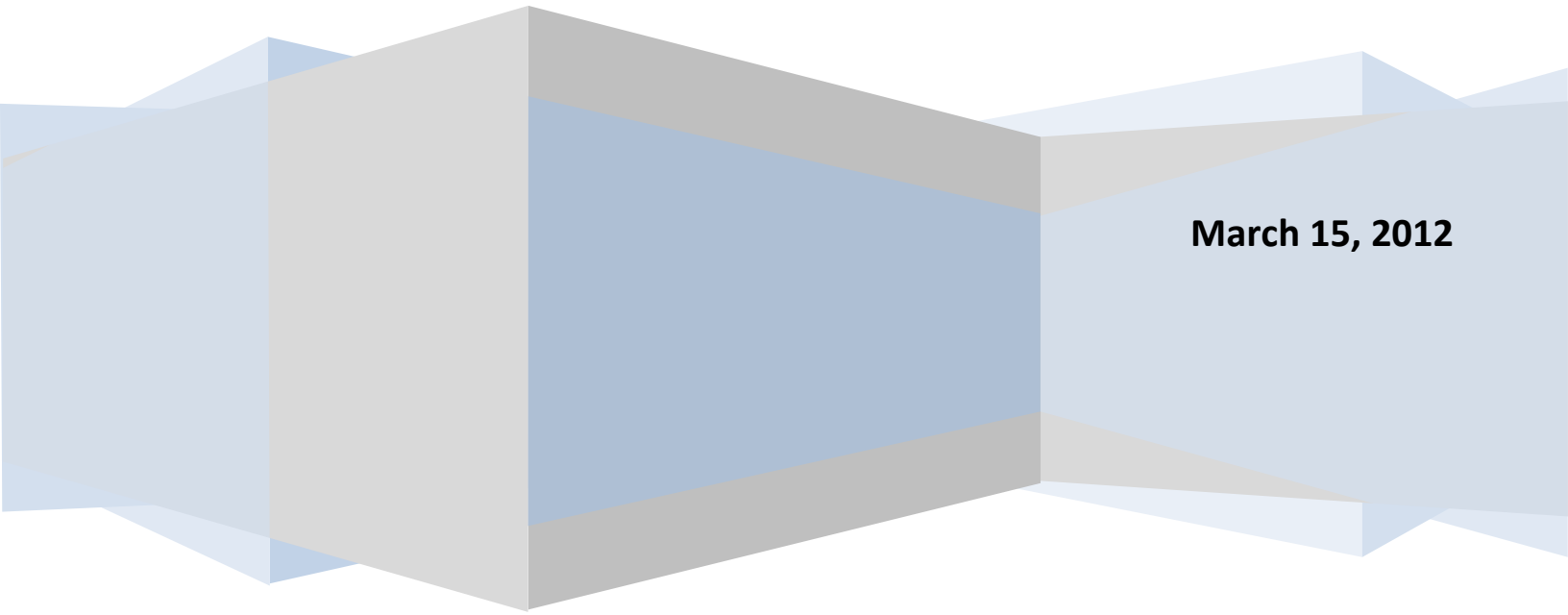
Censuses and Surveys of Governments: A Workshop on the Research and Methodology behind the Estimates

Progress on Presenting Derived Statistics and Coefficients of Variation

Carma R. Hogue



March 15, 2012

Progress on Presenting Derived Statistics and Coefficients of Variation

Carma R. Hogue
U.S. Census Bureau

Abstract

The Committee on National Statistics recommended that Governments Division add value to their data by presenting more derived statistics, such as estimates of change. In this paper we discuss possible derived statistics and the metadata that must accompany the release of such statistics. Plans for future releases as well as derived statistics that are currently being released are discussed along with the coefficients of variation of those statistics.

Introduction

In the 2007 release of the Committee on National Statistics (CNStat) report entitled State and Local Government Statistics at a Crossroads, the Committee recommended that the Census Bureau add value to the data that are released by "providing simple derived measures, such as per capita expenditures and taxes, more explanatory material, and comparative contextual analyses – for example, of trends by type of government and region."

Until the early 1990s, the Division wrote descriptive analytical and graphical reports that enabled the data user to compare data among governments or over time. The Division subsequently opted to release data files to the public using limited distribution news alerts rather than press releases that would highlight statistically significant changes, explain anomalies, and supply data visualizations. A limited number of data users are satisfied with the data files, but state and local government officials, the media, and the novice user benefit from the descriptive analytical reports. All data users benefit from greater transparency of methods, descriptions of the quality of the data, and explanations of concepts and anomalies.

The CNStat panel offered examples of analytical reports that could serve as a basis for future Governments Division products. The Division has been examining these suggestions and is looking towards publishing more analytical and graphical reports. As a part of this effort, the Division is studying possible derived statistics and measures of the variability of those estimates derived from sample surveys. Summary reports are currently being produced for most of the surveys with some derived statistics being added to the reports over time.

In this paper we discuss what has been done to date for our surveys as well as plans for future data product releases. The importance of transparency and education of the data users will be discussed.

Finally, we also discuss the statistics that have been released to date and those that may be released in the future as well as rankings, possible future rankings, currently released trends, and possible future releases of trend data.

Currently Released Reports and Statistics

Over the past three years, Governments Division has released all annual and census products with a press release. Quarterly reports are released with an abbreviated announcement, but all statistics (quarterly and annual) are now released with a summary report. Most of these reports are descriptive, often highlighting a few year-to-year or quarter-to-same quarter last year comparisons, but tables of these comparisons are sometimes provided without their associated coefficients of variation, thus making it difficult for the data user to determine if the changes are significant. Changes that are highlighted in the text have been tested at the 90-percent significance level using crude, conservative estimations of the coefficient of variation. With better estimates of the covariance terms, more changes may be found to be significantly different. For tables of state government (including state retirement systems) only estimates, school district estimates, and the quarterly panel of the 100 largest retirement systems, coefficients of variation are zero and are not shown. Unit response rates are usually in the 90-100 percent range also.

Some of the summary reports also offer graphical displays (pie charts) of the distributions of detailed subcomponents of a total. Some of the reports offer graphical displays of trends over recent years. Quarterly reports offer either trends of a selected quarter (example, third quarter) over several years, or quarter-to-quarter trends of the last several years.

To date, only one of these summary reports for sample surveys has offered descriptive statistics with the accompanying coefficients of variation. In February 2012, the report released with the revised Annual Survey of Public Employment and Payroll contained tables and a discussion of salaries paid per public employee, a ratio of an estimate of Total Salaries to an estimate of Number of Employees. The coefficient of variation of these ratios is included for each estimate. We used a Taylor Series approximation to estimate the coefficient of variation for these statistics (Wolter, 2007).

For surveys that have no sampling error (the Survey of State Government Finances, the State Tax Survey, State Public Employee Retirement Systems, Public Education Finances Survey, and the Quarterly Selected Government Employee Retirement Systems), there are no coefficients of variation provided in the tables. The unit response rates are usually 100 percent for the annual surveys and between 85 and 95 percent for the quarterly surveys. Unreported units are imputed prior to calculating the final estimates. The derived statistics for the surveys of state governments or state systems can be ranked or compared without testing for significance.

For the Public Education Finances Survey, State per Pupil Expenditure (SPPE) is a very important derived statistic that is calculated and displayed. This survey is actually an annual census of the finances of

school districts. The unit response rates are very high because of the system of State Data Coordinators that the survey uses to obtain local school district responses. Besides SPPE, tables of amounts per $1,000 of personal income are shown. State rankings of selected derived statistics and aggregations are displayed. Explanatory text is given in the publication to warn the user of misinterpretations of the rankings and derived statistics. This wording will serve as an example of the text that will be included in future releases of derived statistics in other Governments Division surveys. An excerpt from *Public Education Finances: 2009* is cited below:

> An analysis based on derived statistics can be misleading and misinterpreted because of differences between school systems in accounting methodology, governmental organization, and economic structure. For example, current spending or per pupil current spending as a measure of a school system's current expenses can be misleading because different school systems have different criteria on what they classify as current expenses.

Transparency and Education of Data Users

With the release of our data products, we started releasing methodology reports to describe the sample design, estimation, imputation, and editing for each survey. Quality indicators (unit, Total Quantity Response Rates, item rates, coefficients of variation) are also being released. This is just a beginning to educating the data users. For each survey, we plan to disclose our methodology and quality indicators and to educate the data users about anomalies and interpretations of the data.

There are times that although there are no concerns about sampling error, there are precautions that should be understood about comparisons of the data. For example, with tax data it is important to understand that a per capita tax estimate is not indicative of what an individual would pay for taxes in that state. The per capita estimate is an indication of the amount of tax revenue that is being collected per person counted as living in the state. This is important to note because for high tourist areas like Florida, a large contributor to the tax revenue would be the out-of-state visitors to the state. For Alaska, it is important to understand that most of the tax revenue is coming from severance taxes, i.e., taxes levied by a state on the extraction of a natural product like oil, coal, or gas that is sold outside of the state. Therefore, rankings on income tax or sales tax may be meaningless without such an understanding.

Likewise, differences in accounting from government to government may compromise the comparability of some estimates. These nuances must be understood in order to completely understand derived statistics. It is important for the Census Bureau to ensure that the appropriate clarifying documentation is available with all data releases.

Ancillary to this discussion, the Quarterly Selected Government Retirement Systems Survey is a panel survey of the hundred largest public employee retirement systems as determined by their cash and securities holdings reported in the 2007 Census of Governments. These systems accounted for 89.4 percent of the financial activity of public employee retirement systems in 2007. This panel is held

constant until two years after the Census when another panel will be determined by the most recent Census. There is no sampling, and hence no sampling error for statistics from these hundred largest systems. The data user must be mindful though that inferences cannot be made to all retirement systems. All change estimates, or other relationships between the variables, are only indicative of the change for the 100 largest systems. These data can be ranked and compared, but inferences beyond these 100 largest systems cannot be made. The limitations of the data must also be explained in the text.

Future Releases of Derived Statistics

Governments Division has made strides towards satisfying the data user need for more analytical reports, but the focus will now turn towards providing statistically significant period to period comparisons and providing tables of period to period ratios and their coefficients of variation. Likewise, tables of per capita expenditures and taxes and other such comparative statistics and their coefficients of variation will also be produced. The estimates of totals are very complicated, using small area estimation and Decision-based Estimation. A Taylor Series approximation will usually be tried first to estimate the coefficients of variation of the derived statistics. Tests of significance will be conducted and the results will be displayed in tables along with the coefficients of variation.

Perhaps the most important improvement will be an increase in the explanatory text. Using the text of the Public Education Finance Survey as a guide, the Division's analysts will turn their attention towards providing explanations of anomalies and helping data users understand the data better. Likewise, the Division's survey methodologists will continue to improve explanations of the methodology and quality.

SIMPLE DERIVED STATISTICS

More percentage changes from prior period to current period (annual, quarterly, or quinquennial) will be forthcoming with the appropriate coefficients of variation when there is sampling error. When sampling error is present, these changes will be marked to indicate which ones are significant at a 90 percent significance level, the Census Bureau's current standard.

Other simple derived statistics will be

per capita statistics: revenues, expenditures, taxes

inflation-adjusted dollars: adjusted to 1972 for revenue, expenditure, assets, debt, payroll

per $1,000 income: taxes, education expenditures, welfare, debt

expenditures per dollar taken in

RANKINGS & TRENDS

The Division's Data Visualization team will produce more graphical displays of trends (particularly quarterly) and more maps of comparative period to period changes, such as maps of changes in welfare expenditure, highways expenditure, education expenditure, tax revenue, intergovernmental revenue, etc.

The Division currently publishes rankings of per capita spending on education in the Public Education Finances publication, but the Division will carefully consider including other rankings as appropriate. As stated by the CNStat panel, these rankings are "often the subject of political rather than analytical interest." To date, there are no planned rankings.

Conclusion

The Division has made some steps towards meeting this recommendation by offering descriptive reports, but first increasing the number of period to period changes with corresponding measures of error for each survey along with additional explanatory material and better descriptions of our methodology will bring our programs much closer to satisfying the needs of our data users. The next phase would be to enhance our analytical reports even further, making them attractive to the novice user by offering an array of descriptive statistics and accompanying measures of error along with explanatory material and data visualizations.

References

National Research Council. State and Local Government Statistics at a Crossroads. Panel on Research and Development Priorities for the U.S. Census Bureau's State and Local Government Statistics Program, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press, 2007

U.S. Census Bureau. *Public Education Finances: 2009*, G09-ASPEF, U.S. Government Printing Office, Washington, DC, 2011

Wolter, Kirk. *Introduction to Variance Estimation*, Springer, 2007