

Reducing the Public Employment Survey Sample Size¹

Joseph James Barth, Yang Cheng, Carma R. Hogue
U.S. Census Bureau

Abstract

Like most establishment surveys the Public Employment Survey data are highly skewed. Our goal was to reduce the number of small units included in the sample since they collectively account for a small percentage of the total and currently account for a disproportionate number of sample units. A first stage sample of individual government units was selected via a probability proportional to size method within each state by type of government (city, county, township, special district, school district) using the sample sizes from the previous sample design. In the second stage, classification into size strata was determined using the Cumulative Square Root of the Frequency method [1] within selected states by type of government. After classification into size strata was determined, a subsample of individual government units in the stratum containing the small units was then taken to reduce the number of small units in the sample while maintaining comparability with the previous sample.

Keywords: Sample design, establishment surveys, probability proportional to size

1. Introduction²

The Annual Survey of Government Employment is used to collect full-time and part-time data on state and local government employment and payroll by governmental function (i.e., elementary and secondary education, higher education, police protection, fire protection, financial administration, judicial and legal). The data are collected from the state government as well as five types of local government: counties, municipalities, townships, special districts, and school districts. The first three types of government are referred to as general-purpose governments as they generally cover several governmental functions. School district governments cover only the education function. Special districts cover generally one, but sometimes two functions (e.g., sewer and water).

In 2007, the Committee on National Statistics (CNStat), National Research Council, released the findings of a two-year study of the U.S. Census Bureau's surveys of state and local governments. CNStat offered 21 recommendations on dissemination, data quality, strategic planning, and data uses. In response to the CNStat recommendations on examining new methodologies and in response to concerns the survey analysts have expressed about sample design, we decided to look into ways to modify the sample.

Currently, a stratified, modified probability proportional-to-size sample is used to obtain annual national and state estimates. The current sample design yields a large number of small townships and special districts. The response rate is poor for these units, and they account for a very small part of the final estimate. Within a geographic area, there is very little variability in the responses from units of the same type of government. The desire was to design a sample that would reduce the number of small units in areas of the country with many of these units.

¹ This report is released to inform interested parties of research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

² The methodology, questionnaires, full set of governmental functions, and classification documentation are available on www.census.gov/govs/index.html

In this paper, we discuss the background of the survey, the sample design methodology, the results, future research plans, and conclusions.

2. Survey Background

The Annual Survey of Government Employment is an annual survey of all state and local governments in the 50 states, plus Washington, D.C. The universe and frame are the same as those used in the Census of Governments, with updates made to reflect any births, deaths, or mergers that may have occurred. A unit is determined to be a government if it exists as an organized entity, has governmental character (such as the power to levy taxes), and displays substantial autonomy (i.e., considerable fiscal and administrative independence).

The Annual Survey of Government Employees collects data on five variables, and derives two additional variables from these. The five variables collected are full-time employees, full-time pay, part-time employees, part-time pay, and hours worked by part-time employees. The first derived variable is total pay, which is simply the sum of full-time and part-time pay. The second is full-time equivalent, which is calculated by dividing the number of part-time hours worked by the standard number of hours in a workweek for full-time employees in the particular government, added to the number of full-time employees in that government.

A new sample is selected two years after each census. The most recent sample prior to the 2009 sample was taken in 2004. The samples for 2005, 2006, and 2008 were the same as the sample taken in 2004, with the addition of any births that may have occurred since the 2004 sample was selected.

The data for each unit are subdivided into twenty-three different items, such as fire protection, sewerage, and hospitals. Not every unit has all twenty-three items. For instance special districts and school districts typically only have one or two items.

The sample taken in 2009 contains 10,489 units, not including births. These 10,489 units include units designated as initial certainties. For most types of governments these initial certainties are based on their size, either their population for counties, cities, and townships, or enrollment for schools. For special districts this inclusion is based on the items they contain.

Data are published at national and state levels for state-only, local-only, and state-and-local aggregates. For example, we can view just state government data for Alabama or all state governments combined with all local governments in Alabama. We can view a national total for all state governments combined, or we can view a national total for all local governments combined. If we do not consider data from Washington, D.C.³ this gives us 150 state level estimate tables, and 3 national level estimate tables.

3. Purpose

Our goal in decreasing the sample size was to reduce the time needed to collect and process data. As the population is skewed, analysts were spending a large amount of their time on units that collectively count for a relatively small part of the total. As such we wanted to reduce the number of small units included in the sample. This reduction in small units in the sample may also increase quality as less attention needs to be put on small cases that contribute very little, freeing up analyst time for units with greater impact.

One concern with reducing the sample in this way is that these were units which already had a small probability of selection. By further reducing that, we open ourselves to unforeseen

³ As Washington, D.C. is not a state it only has local estimates. Including Washington, D.C. yields a total of 152 estimate tables, with Washington D.C. not having a state-only estimate table.

jumps in variance; in particular we're concerned with the situation where a small unit grows considerably. Such cases have been rare historically, but we will need to watch for them.

4. Methodology

We used a two-stage approach to the sampling, first sampling using probability proportional to size sampling (PPS) and then using simple random sampling (SRS) to reduce the sample size of small units. Size was defined to be the variable Total Pay from the 2007 Census of Governments. Stratum boundaries, dividing units into small and large sizes, were determined using the cumulative square root frequency method. Birth units and units without any activity in the 2007 Census of Governments went through their own sampling procedures not discussed here, and all following statements in this section should be taken to exclude these units.

Sampling strata were defined by state and type of government with one modification. It was decided that cities and townships were similar enough that they could be combined into a single stratum. Since state governments are included with certainty, this left us with 4 strata per state in most cases: counties, special districts, independent schools, and cities and townships. Some states do not have all types of government and as such have fewer strata.

Sample sizes for strata were set equal to the number of units present in the 2006 sample for the given strata. This was done to aid in comparison between the two samples to see what effect the new sampling methodology may have had.

Only two strata were subsampled: special districts, and cities and townships. The strata containing counties and independent schools both contained too few units, for the majority of states, to warrant subsampling.

The strata selected for subsampling were then subdivided into substrata, small and large, using the cumulative square root frequency method. If this subdivision yielded two substrata that each contained at least 15 units and if the combined stratum had at least 40 units, subsampling was conducted in the "small" substratum. If one of the two substrata had fewer than 15 units, or if the combined stratum had fewer than 40 units, no subsampling was done.

The substrata containing small units were subsampled by SRS. The size for this subsample was determined by first selecting a total desired reduction in sample size as decided by the team lead. This number, eight hundred, was chosen arbitrarily. In the future we hope to use more rigorous methods for determining the size of reduction. This gave a proportion of units to be reduced, which was applied to each substratum selected for subsampling.

In addition to updated sampling methodology, research was done into the efficacy of new estimation methods. The final decision was to use a decision-based regression estimator. If both small and large unit strata yielded similar regression coefficients, then these two strata were combined. If the two strata yielded dissimilar results, then they each were estimated using their separate regression estimators. For further details see [2].

5. Results

The first two tables show sample sizes, not including state units, for the 2006, 2008, and 2009 survey years. For most states the changes were modest and resulted in a smaller sample. California, Illinois, Michigan, Missouri, Ohio, and Texas saw decreases of at least 50 units, while Minnesota saw an increase of similar size. Although cities and townships were joined in one stratum for the 2009 sample design, we leave the cities and townships separate in Tables 2 and 4 for comparison with prior years. From this table we see that the reduction in sample size for the combined city and township stratum is effected by an increase in the number of cities sampled and a larger decrease in the number of townships sampled.

Table 1: Local Government Sample Sizes for the ASGE
by State for 2008 and 2009

State	2006	2008	2009	State	2006	2008	2009
Alabama	240	244	249	Montana	186	204	181
Alaska	54	55	54	Nebraska	271	278	259
Arizona	102	106	108	Nevada	48	49	46
Arkansas	214	213	204	New Hampshire	146	148	132
California	609	645	587	New Jersey	233	227	208
Colorado	196	226	203	New Mexico	115	116	109
Connecticut	139	141	139	New York	267	274	292
Delaware	35	35	38	North Carolina	170	174	163
Washington D.C.	2	2	2	North Dakota	234	235	219
Florida	239	295	247	Ohio	391	420	353
Georgia	231	237	231	Oklahoma	211	212	196
Hawaii	19	19	15	Oregon	191	199	190
Idaho	155	157	146	Pennsylvania	368	374	312
Illinois	564	615	494	Rhode Island	44	44	48
Indiana	293	320	269	South Carolina	169	171	174
Iowa	278	278	256	South Dakota	202	215	197
Kansas	361	369	319	Tennessee	137	138	149
Kentucky	206	207	196	Texas	535	587	512
Louisiana	104	109	102	Utah	118	121	128
Maine	207	222	184	Vermont	168	174	154
Maryland	43	43	47	Virginia	99	99	100
Massachusetts	136	139	121	Washington	204	230	192
Michigan	341	394	320	West Virginia	163	162	150
Minnesota	370	379	431	Wisconsin	317	321	280
Mississippi	218	218	221	Wyoming	116	122	120
Missouri	397	443	367				

Source: U.S. Census Bureau

Table 2: Local Government Sample Sizes for the ASGE
by Type for 2008 and 2009

Type of Government	2006	2008	2009
Counties	1435	1436	1456
Cities	2549	2609	3022
Townships	1528	1534	624
Special Districts	3305	3772	3204
Independent Schools	2039	2054	2108

Source: U.S. Census Bureau

Tables 3 and 4 show the relative changes in sample size from the previous year to the current year for 2005 through 2009. Since 2007 was a census year, there is no information available for that year, and the ratio for 2008 shows the relative increase from the 2006 sample size.

Table 3: Relative Local Government Sample Size Change Rates
by State for the ASGE, 2005-2009

State	2005 /2004	2006 /2005	2008 /2006	2009 /2008	State	2005 /2004	2006 /2005	2008 /2006	2009 /2008
Alabama	1.00	1.02	1.02	1.02	Montana	1.00	1.02	1.10	0.89
Alaska	1.00	1.02	1.02	0.98	Nebraska	0.99	0.93	1.03	0.93
Arizona	1.01	1.02	1.04	1.02	Nevada	1.00	1.00	1.02	0.94
Arkansas	1.00	1.00	1.00	0.96	New Hampshire	0.99	1.00	1.01	0.89
California	1.00	1.00	1.06	0.91	New Jersey	1.00	1.00	0.97	0.92
Colorado	0.99	1.00	1.15	0.90	New Mexico	1.00	0.99	1.01	0.94
Connecticut	1.00	1.00	1.01	0.99	New York	1.00	1.00	1.03	1.07
Delaware	1.00	1.00	1.00	1.09	North Carolina	1.00	1.01	1.02	0.94
Washington D.C.	1.00	1.00	1.00	1.00	North Dakota	1.01	0.99	1.00	0.93
Florida	1.01	1.00	1.23	0.84	Ohio	1.00	1.00	1.07	0.84
Georgia	1.00	1.00	1.03	0.97	Oklahoma	1.00	1.01	1.00	0.92
Hawaii	1.00	1.00	1.00	0.79	Oregon	0.99	1.00	1.04	0.95
Idaho	1.00	0.99	1.01	0.93	Pennsylvania	0.99	0.99	1.02	0.83
Illinois	1.00	1.00	1.09	0.80	Rhode Island	1.00	1.00	1.00	1.09
Indiana	1.00	1.00	1.09	0.84	South Carolina	0.99	1.00	1.01	1.02
Iowa	1.00	0.99	1.00	0.92	South Dakota	1.00	1.00	1.06	0.92
Kansas	1.00	1.00	1.02	0.86	Tennessee	1.01	0.99	1.01	1.08
Kentucky	1.00	1.00	1.00	0.95	Texas	1.00	1.03	1.10	0.87
Louisiana	1.01	0.99	1.05	0.94	Utah	1.00	1.01	1.03	1.06
Maine	1.00	1.00	1.07	0.83	Vermont	0.99	0.99	1.04	0.89
Maryland	1.00	1.00	1.00	1.09	Virginia	1.00	1.00	1.00	1.01
Massachusetts	0.99	1.00	1.02	0.87	Washington	1.00	1.03	1.13	0.83
Michigan	0.98	1.00	1.16	0.81	West Virginia	1.01	1.00	0.99	0.93
Minnesota	1.00	1.01	1.02	1.14	Wisconsin	1.01	1.00	1.01	0.87
Mississippi	1.00	1.00	1.00	1.01	Wyoming	1.00	1.00	1.05	0.98
Missouri	1.00	1.00	1.12	0.83					

Source: U.S. Census Bureau

Table 4: Relative Local Government Sample Size Change Rates
by Type for the ASGE, 2005-2009

Type of Government	2005 /2004	2006 /2005	2008 /2006	2009 /2008
Counties	1.00	1.00	1.00	1.01
Cities	1.00	1.01	1.02	1.16
Townships	1.00	1.00	1.00	0.41
Special Districts	1.00	1.00	1.14	0.85
Independent Schools	1.00	0.99	1.01	1.03

Source: U.S. Census Bureau

Tables 5 and 6 show the coefficients of variation (CVs) for the 2004 sample design and for the 2009 sample design, both of which are years for which the sample is brand new, for four of the reported variables. Part-time hours is not presented as the data will not be shown separately in the published viewable tables. CVs are not yet available for 2008 as the data for 2008 are still going through processing at the time of this paper's writing. The CVs presented for 2009 are not actual CVs, as data collection is still underway as of the writing of this paper. Instead the CVs for 2009 are estimated by using a sampled unit's 2007 data as a surrogate for 2009, with any units that were not in existence for the 2007 census being excluded from the calculations. As such we expect the actual CVs to be somewhat higher. For further information on estimated 2009 CVs, see [2].

Table 5: Coefficient of Variation for Local Government Full-time Employees and Pay by State for the ASGE, for 2004 and 2009

State	Full-time Employees		Full-time Pay		State	Full-time Employees		Full-time Pay	
	2004	2009	2004	2009		2004	2009	2004	2009
Alabama	0.84	1.07	4.06	0.41	Montana	1.33	1.25	6.02	0.59
Alaska	1.04	0.44	3.06	0.29	Nebraska	2.96	1.31	3.03	0.43
Arizona	1.85	0.80	3.08	0.26	Nevada	4.48	0.25	6.87	0.12
Arkansas	1.11	1.86	7.39	0.54	New Hampshire	1.73	0.00	8.40	0.00
California	0.33	0.56	1.29	0.34	New Jersey	1.07	0.00	3.58	0.00
Colorado	0.55	1.50	2.29	0.51	New Mexico	1.19	2.48	3.79	0.45
Connecticut	0.71	2.54	2.77	1.91	New York	0.46	1.73	1.88	0.21
Delaware	27.64	0.00	28.18	0.00	North Carolina	0.49	0.44	1.97	0.24
Washington D.C.	0.00	0.00	0.00	0.00	North Dakota	0.77	0.84	4.24	0.68
Florida	0.39	0.53	1.08	0.37	Ohio	1.39	1.20	3.43	0.40
Georgia	0.81	0.84	5.20	0.31	Oklahoma	1.17	1.98	4.65	1.33
Hawaii	0.00	0.00	0.00	0.00	Oregon	1.00	1.32	3.90	0.61
Idaho	1.63	1.87	6.77	0.51	Pennsylvania	0.90	1.45	2.78	1.20
Illinois	0.57	1.13	3.23	0.62	Rhode Island	0.00	0.00	0.00	0.00
Indiana	1.25	1.95	3.74	0.50	South Carolina	1.33	1.35	5.68	0.26
Iowa	3.42	1.50	5.27	0.55	South Dakota	3.45	2.48	8.24	1.18
Kansas	1.13	2.18	2.07	0.80	Tennessee	0.96	0.72	6.05	0.14
Kentucky	1.18	1.19	5.10	0.32	Texas	0.32	0.59	1.62	0.31
Louisiana	0.92	1.74	7.31	0.33	Utah	1.58	0.49	1.71	0.53
Maine	1.72	0.00	8.19	0.00	Vermont	2.97	2.77	6.35	1.73
Maryland	2.34	0.20	2.41	0.09	Virginia	2.87	0.00	3.66	0.00
Massachusetts	2.14	0.00	7.27	0.00	Washington	0.55	0.78	1.91	0.58
Michigan	0.90	0.94	2.43	0.39	West Virginia	1.51	2.63	10.12	0.52
Minnesota	1.43	1.93	3.56	1.44	Wisconsin	1.43	1.65	4.40	0.59
Mississippi	1.36	1.74	5.17	0.28	Wyoming	0.81	1.94	5.28	0.86
Missouri	0.84	1.16	3.41	0.44					

Source: U.S. Census Bureau

Table 6: Coefficient of Variation for Local Government Part-time Employees and Pay by State for the ASGE, for 2004 and 2009

State	Part-time Employees		Part-time Pay		State	Part-time Employees		Part-time Pay	
	2004	2009	2004	2009		2004	2009	2004	2009
Alabama	0.84	8.20	2.04	4.65	Montana	1.23	5.87	4.92	4.66
Alaska	0.85	5.71	6.20	3.73	Nebraska	2.15	12.08	3.11	4.64
Arizona	1.79	4.32	2.02	2.70	Nevada	3.74	1.88	6.48	1.64
Arkansas	1.05	11.43	5.91	11.85	New Hampshire	2.21	0.00	10.48	0.00
California	0.35	2.27	1.00	1.64	New Jersey	1.12	0.00	4.34	0.00
Colorado	0.38	5.98	1.12	4.17	New Mexico	1.10	7.92	2.62	8.32
Connecticut	0.65	6.52	3.68	10.04	New York	0.38	4.17	1.93	2.85
Delaware	23.44	0.00	31.37	0.00	North Carolina	0.52	2.25	1.39	1.23
Washington D.C.	0.00	0.00	0.00	0.00	North Dakota	0.64	13.21	3.73	5.32
Florida	0.38	0.72	0.91	0.69	Ohio	0.88	5.09	3.26	4.77
Georgia	0.68	6.50	2.28	2.25	Oklahoma	1.00	13.06	5.37	8.73
Hawaii	0.00	0.00	0.00	0.00	Oregon	0.60	3.93	2.43	3.76
Idaho	1.48	8.78	4.91	5.96	Pennsylvania	0.89	5.65	3.77	4.86
Illinois	0.76	9.53	2.63	4.46	Rhode Island	0.00	0.00	0.00	0.00
Indiana	1.19	6.14	5.05	5.29	South Carolina	1.55	7.54	4.02	4.64
Iowa	3.38	5.82	4.87	4.57	South Dakota	2.05	9.46	6.18	10.89
Kansas	1.09	14.01	2.26	8.49	Tennessee	0.85	4.07	3.63	2.87
Kentucky	1.23	25.60	4.86	5.79	Texas	0.25	4.14	1.44	3.14
Louisiana	0.89	8.44	1.73	6.65	Utah	1.51	4.36	2.00	3.83
Maine	1.24	0.00	8.57	0.00	Vermont	2.36	7.64	8.25	7.99
Maryland	1.88	0.88	1.19	0.79	Virginia	2.30	0.00	3.38	0.00
Massachusetts	2.34	0.00	5.01	0.00	Washington	0.54	2.35	1.65	1.95
Michigan	0.87	5.84	3.06	3.48	West Virginia	2.66	8.07	8.03	7.58
Minnesota	0.75	6.60	3.51	4.88	Wisconsin	1.49	9.24	3.79	4.95
Mississippi	1.09	9.97	4.62	4.76	Wyoming	1.43	10.52	2.26	6.77
Missouri	0.63	4.70	2.60	4.42					

Source: U.S. Census Bureau

6. Conclusion and Future Research

For most states we see a reduction in overall sample size, or at least a retardation in sample size growth. This is entirely to be expected based on the construction of 2009 sample sizes. We began with 2006 sample sizes and then, although there were births and increases in the number of certainties, we removed sample. This is not a point of interest so much as a simple consequence of our methodology.

We do see a large shift of the sample from townships to cities as both were combined into a single sampling stratum and townships tend to be smaller than cities. As such the allocated

sample was pushed into the larger (city) governments. This result was expected but not guaranteed.

The estimated CVs for 2009 show a trend towards increasing CVs as compared with 2004. For full-time data the CVs are still generally under our target CV of 3 percent. For part-time data the CVs increased well into unacceptable levels and we will be relying on updated estimation procedures to correct for this.

In addition to the reduced sample size, we are also researching estimation methods to lower the variance of the estimates. Details of this research can be found in [2].

In the future we would like to spend more time determining the best way to determine the substratification into small and large units. In addition, further research into the optimal allocation of sample into the substrata would be beneficial. For both of these cases we decided to use standard methods, as there was not sufficient time to research optimal methods. For 2012, we hope to find better solutions.

References

1. Cochran, William G. *Sampling Techniques*, John Wiley & Sons, 1977
2. Cheng, Yang; Corcoran, Casey; Barth, Joseph; Hogue, Carma 2009 Estimation Procedure for New Public Employment Survey Design. In *JSM Proceedings*, Survey Research Methodology Section. Alexandria, VA: American Statistical Association.