Results of Monitoring the American Community Survey Using Statistical Process Control Methodologies

Jessica Schwan, Steven Hefter
U.S. Census Bureau, Washington, DC 20233

Abstract
To detect potential errors in response data from the American Community Survey (ACS) in near real-time, we have developed an automated statistical process control (SPC) system. For each year from 2005 to 2010, the ACS collected data about roughly two million housing units, 4.5 million people in the household population, and 150,000 people in group quarters facilities. Beginning in June of 2011, the ACS housing unit sample was increased by 22 percent. We expect to see approximately the same increase in the number of interviews. Several SPC methodologies are being used to investigate responses, using traditional Shewhart charts as well as basic statistical tests for differences between proportions. We are also using these techniques to monitor any impact the realignment of our Regional Office (RO) structure may have on data quality. This paper describes the details of the methodology, and a summary of results to date including preliminary results from the assessment of the realignment of the RO structure. We also discuss the inherent challenges and obstacles faced when applying traditional process control methods to a large-scale, multi-mode, demographic survey.

Key Words: American Community Survey, statistical process control

1. The American Community Survey

1.1. Introduction

The American Community Survey (ACS) fully implemented data collection starting in 2005. Data collected by the ACS is designed to replace the long-form data previously collected during Decennial Census operations. Using rolling monthly samples, the ACS collects data from approximately two million housing units, 4.5 million people in the household population, and 150,000 people in group quarters facilities. These monthly samples produce annually updated estimates for all geographic areas, regardless of their size. The ACS is also conducted in Puerto Rico, where it is referred to as the Puerto Rico Community Survey (PRCS). For the purposes of this paper, the term ACS will indicate the data collected in both the U.S. and PR.

Data collected from housing units (HUs) are collected using three different data collection modes. First, a paper questionnaire is sent to the mail addresses selected for sample for a particular month. If that questionnaire is not returned in a pre-determined period of time, a second questionnaire is mailed to the same address. If neither mail questionnaire is returned by the end of that month, then eligible HUs will move to the second phase of data collection: Computer Assisted Telephone Interview (CATI). If a completed telephone interview cannot be captured, or if the HU was not eligible for the CATI mode of data collection, then the HU may be selected for the third mode of data collection: Computer Assisted Personal Interview (CAPI). The personal visit is the final mode of data collection. This paper discusses all three modes of data collection.

Data collected from group quarters (GQs) facilities are collected by a combination of telephone interview and personal visit with the facility in question. GQs are places where people live or stay in a group living arrangement, where the management of the group living facility provides housing or other services for the residents. This definition differs from that of HUs, as they consist of houses, apartments, mobile homes, groups of rooms, or single rooms that are occupied as separate living quarters. Usually, the residents of HUs are related to one another, where as GQ residents are not. This paper discusses only HU data; however, GQ incorporation is forthcoming.

The data collected from the ACS are used to produce estimates of the demographic and socio-economic characteristics of the population of the U.S. and PR. Three sets of period estimates produced every year. The 1-year estimates are produced for areas with a population of at least 65,000. The 3-year estimates cover areas with a population of 20,000 or more. And lastly, the 5-year estimates cover all geographic areas, regardless of population size, down to the census tract and block group levels. The Population Estimates Program at the Census Bureau provides these official estimates of the population to the ACS annually.

1.2. Sample Selection and Data Collection

Each year, a sample of addresses is selected from a listing of all known living quarters. This listing is often called the sampling frame or sampling universe. The ACS sampling frame is produced from the Census Bureau's Master Address File, which is our official inventory of HUs and GQs for the U.S. and PR. Starting in June of 2011, the yearly ACS national sample was increased from 2.9 million addresses to 3.54 million addresses. This sample of addresses is randomly allocated to the twelve months of the calendar year. Each month (or panel) of HU sample undergoes the three modes of data collection in three sequential months. As depicted in Figure 1, the ACS data collection is conducted in a rolling nature.

Figure 1. The ACS Data Collection Strategy

	Calendar Month				
Sample Panel	**Feb-2012**	**Mar-2012**	**Apr-2012**	**May-2012**	**Jun-2012**
Dec-2011	CAPI				
Jan-2012	CATI	CAPI			
Feb-2012	Mail	CATI	CAPI		
Mar-2012		Mail	CATI	CAPI	
Apr-2012			Mail	CATI	CAPI

In this manner, the three data collection modes are always being conducted, but for different sample panels. Note that returned mail questionnaires are accepted during all three data collection periods.

2. Statistical Process Control

2.1. Introduction

Statistical process control (SPC) is a method of quality control which applies statistical methods and procedures to analyze the inherent variability of a process (Hefter and

Marquette, 2011). SPC methodology is traditionally utilized with industrial processes with little or no variation over time. For example, during the manufacture of chocolate bars, the length of the bars that are produced must be monitored. If the bars are too long, then the chocolate wrapper will not be large enough to contain the bar, whereas if the bars are too short, then consumers will complain that they are not getting enough chocolate. SPC techniques can be used to monitor the length of the chocolate bars, by alerting those operating the machine that the bars are outside a certain threshold.

2.2. Variation

There are two types of variation that must be considered where SPC is concerned. The first kind of variation is controlled variation, which is natural and expected in the process. The second kind of variation is uncontrolled variation, which is not normally present in the process and is triggered by an underlying special cause. While both kinds of variation can be detected using traditional SPC techniques, it can be difficult to determine which types of variation are controlled and which are uncontrolled.

While using SPC to measure controlled variation in the ACS is inherently interesting, we have been attempting to measure uncontrolled variation, in an effort to monitor data quality. In the past, errors have been found in the published ACS estimates. Once discovered, either by data users or by Census Bureau staff, these errors must be recalled, corrected, and re-published. This is inefficient and costly, not to mention damaging to our reputation as the foremost agency of data collection and dissemination. Therefore, using SPC to monitor the variation in the ACS will aid us in identifying potential errors in the data, and in enough time to correct them before publication.

Figure 1 shows that data collection for the February 2012 panel is finished after April of 2012 (with the exception of late mail returns, which are accepted up until the 15th of the following month). After minimal post-data collection processing, this data will not be addressed until pre-weighting, which occurs approximately a year later. This period of time between data collection and pre-weighting allows for the application SPC on a monthly basis, to monitor for potential issues. If a true data error is discovered, it is possible to make corrections in “near-real” time. This is a vast improvement over the current method of data review, error identification, and correction, which occurs after all data has been collected.

While there are two different kinds of variation, controlled and uncontrolled, there are also two types of uncontrolled variation. The first is the result of errors in the process, which is the kind in which we are most interested. The second is caused by verifiable changes to the process. The statisticians who work on the ACS are constantly researching ways to improve our survey design. These improvements can take form in content changes to the survey itself, formatting changes, modifications to the interviewer instructions, or processing changes. These internal changes often affect the process itself, and must be accounted for in our quest for early error detection. Additionally, as any content changes are experimentally tested before their inclusion in production, we often have an idea of the effect of such changes.

2.3. Challenges

SPC and the ACS, at first glance, seem to have conflicting goals. The major goal of SPC is to determine if a process is in statistical “control”. For something to be in control, we expect it to be consistently repeating the same actions. To return to our example with the chocolate factory; we expect the chocolate bars to be the same length. If they are not the same length consistently, then something must be adjusted or calibrated: the machinery, the ingredients, or the technicians. However, the major goal of the ACS is to *measure* change. Fluctuations in survey data are expected across time, as the population itself changes. This makes it difficult to distinguish between the natural variation in the changing population we are attempting to measure and unnatural variation that may identify a potential error.

Another challenge in developing SPC techniques for use with the ACS is the need to disseminate the results. For example, if the software identifies a potential issue, and no one is notified of the issue, there is no benefit from the software. Dissemination of the data is a crucial step in this process. Consequently, we spent a considerable amount of time developing an intranet application. This application is designed to allow the analysts and the subject matter experts easy access to the data from their desktop computers. This moves the data into their hands very quickly, and allows them the ability to investigate if outliers are identifying an actual error.

3. Methodology

3.1. Repeated Measures and the Time Window

Repeated measures over a relatively long period of time are required to use SPC techniques, and for them to be effective and accurate. As mentioned in the introduction, the ACS began full implementation of data collection procedures for the entire U.S. and PR in 2005. This gives us a time series of sufficient length, with relatively few major changes to the ACS questionnaire. Since January 2005, we have been collecting survey response data at not only the national level, but all the way down at the county level, and in some cases, tract level. We collect and test more than 120 questions and up to 17 categories per question, for a combination of almost 850 response categories for each of the three modes of data collection. We include non-response as a separate category to be tested. The non-response category includes all missing responses, invalid responses, “don’t know” responses, and refusals. This is a very large quantity of data.

One of the most important aspects of SPC methodology is the use of the process average. The process average is the average or mean of the process as it moves through time, and gives the best representation of the overall trend of the variable in which we are interested. To calculate a process average for each response category in the ACS, we had to begin by determining the appropriate length of time to use in its calculation. While we have data stretching all the way back to January 2005, the questionnaire itself has only been relatively stable since 2008. Additionally, we wanted to use the most recent survey data available, while keeping our estimates as reliable as possible. Thus, we decided to use data from the previous two years and the current year as our time window. This means that if we were interested in calculating the process average for July of 2011, we would use all the data from 2009, 2010, and 2011, up to our month of interest.

3.2. Data

The data we receive for each of the three modes of data collection comes from various locations. Returned mail forms are passed through the National Processing Center (NPC). The questionnaires are scanned and minimally processed at NPC, creating the unedited batch files. These files are then passed back to the Census Bureau’s Headquarters, where they are again processed to create the daily keying files. These daily keying files are not modified by any data quality processes, such as the Failed Edit Follow-Up operation, and are used as the mail data input to SPC. These files are the closest that we are able to review the raw data that are returned by the respondents, thus, negating possible biases and errors due to processing.

Data collected during the CATI and CAPI operations are collected using computer instruments. Telephone data is entered into a BLAISE instrument, which transmits data directly back to Headquarters daily. Personal visit data are collected by field representatives (FRs) using laptops and transmitted to Headquarters using an encrypted Internet connection. The Technologies Management Office (TMO) then collects the CATI and CAPI data into monthly files. These TMO-all files are the input to SPC.

One important fact to note is that data returned from NPC or through TMO are collected and reviewed by the original sample panel, not the actual calendar date that they were returned.

3.3. The Z-Score Filter

For each question in the ACS questionnaire, we began by determining the available universe of respondents for that question. This is because every question on the questionnaire does not necessarily apply to every respondent. For example, only those respondents who have served in the military are eligible to respond to the question concerning service-connected disabilities. These universes are primarily based on the edit universes. The edit universes define the respondents that are eligible to answer each series of question. For example, only those respondents who are 15 years or older are eligible to answer any questions concerning work or labor. Once we have the eligible universe for each question, the overall proportion of response for each legal value is reviewed, including “non-response”. For example, a respondent could answer a question such as “What is your sex?” in three ways: “Male”, “Female”, and without responding to the question at all. Choosing not to respond to a question, refusing to respond to a question, not answering a question due to misunderstanding the content, and accidentally missing a question all correspond to non-response.

Once we have the proportion of response for each response category of each question, we had to distinguish between those proportions that were normal and those proportions that were indicative of $a\ potential$ issue. To do so, we developed a conservative filter comprised of three z-score tests. This z-score filter compares the proportion of response from the current month of interest to the proportion from the previous month, the proportion from the same month from the previous year, and the proportion from the entire previous year. Only if all three comparison tests fail is the proportion flagged as an outlier. An individual test fails only if $z_{mt} > 2.5$, using the following formulas

$$z_{mt} = \frac{p_t - p_m}{SE_{mt}} \qquad SE_{mt} = \sqrt{p_{mT}\left(1 - p_{mT}\right)\left(\frac{1}{n_t} + \frac{1}{n_m}\right)} \qquad p_{mT} = \frac{p_t n_t + p_m n_m}{n_t + n_m}$$

where

m	Current month of interest (current month of analysis)
t	Comparison time period (either the previous month, the same month last year, or the entire previous year)
T	Total time period (the previous two years through the month of interest, m)

p_t	Sample proportion for comparison time period, t
n_t	Sample size for comparison time period, t
p_m	Sample proportion for current month of interest, m
n_m	Sample size for current month of interest, m
z_{mt}	Z-score, based on the month of interest, m, and comparison time period, t
SE_{mt}	Standard error, based on the month of interest, m, and comparison time period, t
p_{mT}	Overall process proportion (calculated using the total time period, T)

A negative z-score indicates that the proportion for the month of interest is higher than that of the comparison time period, while a positive z-score indicates that the proportion for the month of interest is lower than that of the comparison time period.

3.4. Shewhart Charts

An easy way to depict this time series of data and determine if it is in a state of statistical control is by using a control chart. Our control charts, or Shewhart Charts, illustrate a process as it moves through time by graphing the proportion of response for the question for each panel, displays the upper and lower confidence limits for each point, and gives the overall process average for the response category. The process average and the upper and lower control limits are calculated using the following formulas:

$$p_i = \frac{\sum_{j=1}^{N} n_{ij} p_{ij}}{\sum_{j=1}^{N} n_{ij}}$$

$$UCL_{ij} = \max\left(p_i + k\sqrt{p_i(1-p_i)/n_{ij}}, 1\right) \qquad LCL_{ij} = \min\left(p_i - k\sqrt{p_i\left(1-p_i\right)/n_{ij}}, 0\right)$$

where

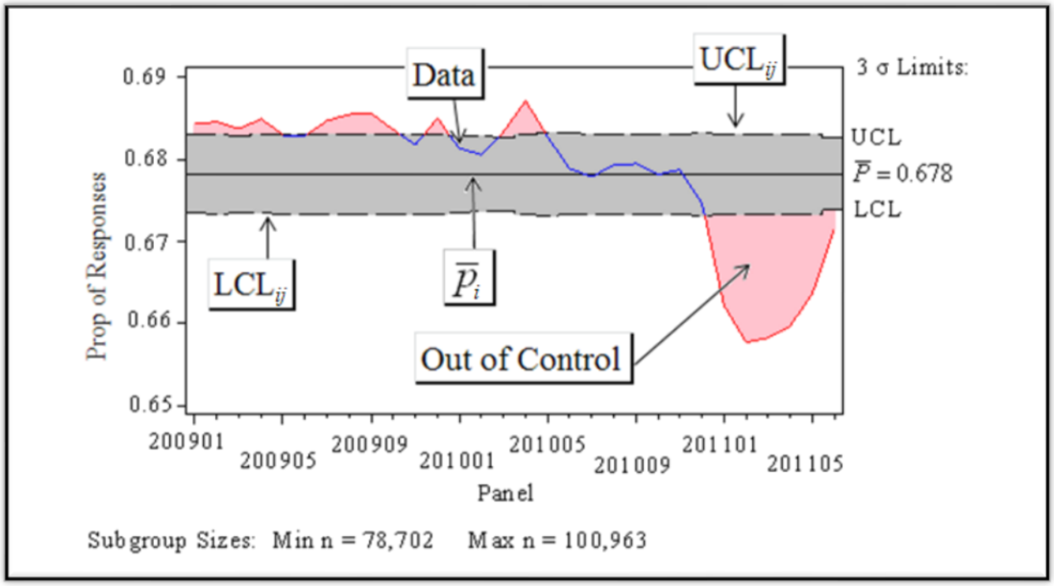
i	Response category for the question, for the current month of interest m
j	Panel
k	Sigma limit for control chart ($k = 3$)
N	Total number of panels in the time window ($25 \leq N \leq 36$)
n_{ij}	Sample size for the response category i for the panel j
p_{ij}	Proportion of response for the response category i for the panel j
p_i	Process average for the response category i (covers the entire time period)
UCL_{ij}	Upper confidence limit for the response category i for the panel j
LCL_{ij}	Lower confidence limit for the response category i for the panel j

Figure 2 gives an example of a control chart, with appropriate labels.

The grey area on the chart marks the section of the chart that is “in control”, that is, the points contained within that area are in statistical control. Notice that the data line is blue within that area. The red region on the chart is “out of control”, meaning points contained within that area are not in statistical control. Notice that the data line is red in that area.

One important fact to remember is that because the process average and control limits are based on the entire time series of data in the time window, the Shewhart chart will change as more data points are added. This means that the confidence limits and process average are dependent on the time window. For example, in Figure 2, the trend up through May 2005 is mostly illustrated in red, even though the points seem to be following a fairly linear pattern contained within a small area. If we had viewed this graph before the dramatic drop around January 2011, we may have seen these points in control. The January 2011 special cause has pulled the confidence limits lower, causing that section of the graph to appear as out of control.

Figure 2. Shewhart Chart Example



Now, while this method is very efficient and effective at displaying the data that we are producing, it does have some limitations. The sample size for the ACS is much larger than the sample size that is traditionally used with Shewhart charts. This results in a small scale, narrow control limits, and small variances. This causes a large number of false positives, meaning we see more response categories marked as out of control than we would usually expect. One possible solution that is being explored is to increase the control limits on the chart.

3.5. Seasonal and Trending Adjustments

Many questions on the ACS survey have predictable, nonrandom patterns that can be attributed to elements external to the survey. Changes in the population, changes in the economy, and changes in the weather can all have a dramatic effect on the proportion of certain response categories. For example, the overall population of the U.S. is gradually aging as the baby boomers approach retirement age. This has caused a marked trend in all response categories for any question that might be affected by age.

To account for this seasonality or trending, we are able to remove the seasonal or trend components and then generate Shewhart charts for the remaining portion of the data. We use PROC ARIMA for seasonal models and PROC ESM for trending models. We can use any combination of the seasonal or trending adjustments to create an appropriate model. Once we have the model, we subtract that model from the actual data to create residuals. These residuals are graphed using the same Shewhart chart methods as the non-adjusted data, and the graphs are reviewed with the rest of the out of control variables.

Currently, we apply the z-score filter to variables only *before* they undergo seasonal or trending adjustments, but have plans to apply the filter *after* the adjustments as well.

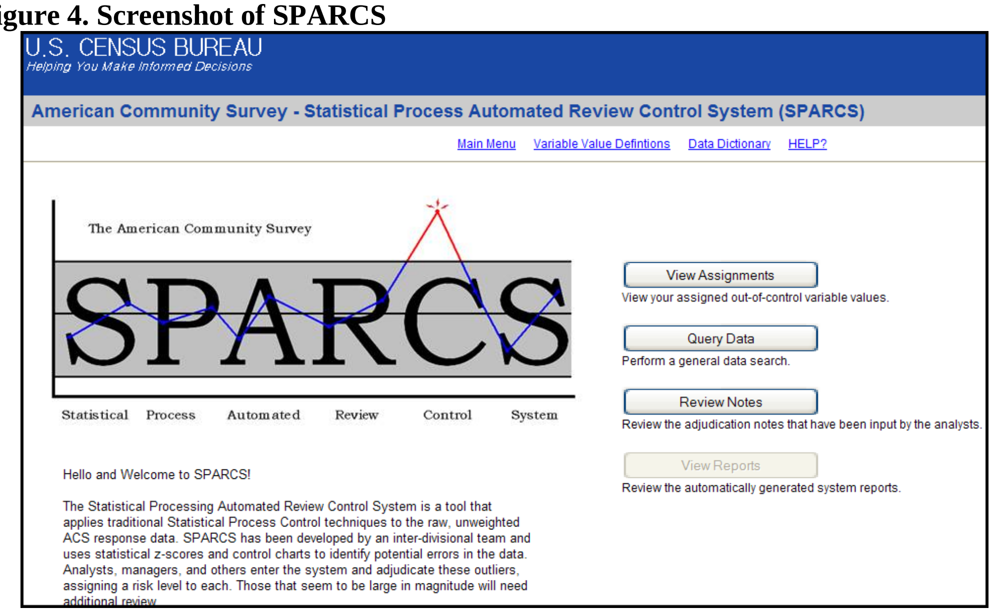
3.6. Geography

When we review the data for each response category for each question, we can aggregate the data to different geographic levels for review. We do this by separating the data that we receive based on the region and the data collection mode. Mail data can be aggregated and analyzed nationally, by state, by sampling substrata, and by Census segmentation group; CATI data can be aggregated and analyzed nationally and by telephone center; and CAPI data can be aggregated and analyzed nationally and by regional office (RO). Reviewing the data at various geographic levels helps in identifying specific regions of the country where an issue may be contained.

4. Statistical Process Automated Review Control System

As discussed earlier in the section on challenges, dissemination of the data is a vital step in any quality control and error correction process. As such, we developed the Statistical Process Automated Review Control System (SPARCS), to allow for easy access to the data and to automatically notify those individuals who are crucial in the quality control process. This system displays the Shewhart charts for all response categories, not only those that failed the z-score filter. This allows the analysts to see all response categories for a question, which is a crucial step in determining if an outlier is a potential error. The system also provides the associated data tables for each chart, containing the sample sizes, process averages, and control limits for each panel that is included in the time series. This allows the analysts to see the actual values that go into the creation of the chart, and allows them to download the information for further research.

Figure 4. Screenshot of SPARCS



For this system to be an effective tool from the error detection standpoint, we also incorporated some non-statistical functionality as well. First, a listing of the failing response categories by panel was necessary to keep track of the overall failures, and reduce information and data overload. This “view assignments” section of the web-tool allows the analysts to enter the system, immediately know their workload for the next month, and manage it effectively. View assignments also allows the users to quickly view only the outliers, without having to know anything about the internal workings of the system. This list can be tailored to show only specific response categories for a particular analyst (or group of analysts). Additionally, the system allows the analysts to input an adjudication, or risk level, for the outlier, and notes as to why they made that decision. This allows for collaboration with the other analysts and for documentation for future purposes. These notes can also be reviewed and sorted in the system, for easy cataloguing.

SPARCS also incorporates the functionality for any data user, with access to the system, to research the data. This “query data” section of the system allows a user to specify which data they would like to view. Users may select by panel (or panels), data collection mode, geography level, question, response category, z-score filter failure (or failures), and much more. This allows users and analysts to investigate potential problems, based on correlations between questions on the survey. This feature also has tremendous potential for research and evaluation.

Lastly, we have also included the capacity to generate an automatic report, but have not fully utilized this functionality.

Figure 4 shows the main page of SPARCS, through which you access the features detailed above.

5. Overall Results

Due to the amount of output from the system, we cover the overall number of outliers over time and only selected case studies for particularly interesting response categories.

Figure 5 shows the percentage of outliers that have been flagged at the national level for each mode of data collection. This table shows the total percentage of failures, or outliers, at the national level, for the U.S. English mail form, and CATI/CAPI instrument for July 2011 through April 2012. The total percentage of outliers range from zero in February 2012 CATI to 18.2% in January 2012 mail. We are currently reviewing the January 2012 mail run, to see why this percentage of failures is so high, compared to the other percentages. This may be in part due to the fact that this is the first month in 2012 we are comparing against data from the 2011 sample reallocation, which we will discuss in a later example. Notice that for the most recent runs, the percentage of outliers found in the mail data is much higher than that in the CATI or CAPI data. This partly because the U.S English mail data is separated from the other three mail forms (U.S. Spanish, PR English, PR Spanish), while the CATI and CAPI data is all collected in the same instrument.

The quantity of outliers is relatively small, considering that the system analyzes almost 850 response cateogires per mode for every panel of data collection. We are still attempting to discover if there is a certain threshold over which the quantity of outliers would be suspect. However, as the questions and response categories are so varied and subject to previous trends, a threshold has not yet been set.

Another important fact to consider is that we are testing each response category separately, using the basic z-tests described above. This leads to a large percentage of Type I errors, that is, response categories that are flagged as outliers, when they are actually not. And, in fact, we do see several graphs per month that were flagged as out of control, but were in control according to the limits of the Shewhart chart. In this manner, we are actually using two statistical tests to determine if the outliers are significant: the z-score filters reduce the amount of data we must sift through, and the Shewhart charts give us a visual reference of the magnitude of the outlier.

Figure 5. Percent of Failing Response Categories

	Mode		
Panel	**Mail Failures**	**CATI Failures**	**CAPI Failures**
July 2011	2.5%	3.2%	1.0%
August 2011	2.3%	3.1%	0.7%
September 2011	1.7%	0.7%	2.6%
October 2011	1.0%	0.8%	2.1%
November 2011	1.3%	2.3%	1.5%
December 2011	5.1%	1.1%	1.0%
January 2012	18.2%	0.9%	2.0%
February 2012	3.0%	0.0%	2.3%
March 2012	3.3%	1.1%	0.6%
April 2012	2.5%	0.9%	**

** April panel CAPI data was not yet available at the time this paper was prepared.

6. Case Studies

The first important fact to remember when reviewing results from SPARCS is that the Shewhart charts show only a single response category. For an analyst to understand what is actually happening, they must also review the other response categories for that particular question. Dramatic changes in trend may look odd if only a single graph is studied, but that same trend may have a reasonable explanation once the other response categories are reviewed. Another important fact is that the process average and control limits are strongly affected by outliers. As such, only the most recent data point can be truly defined as being in-control or out-of-control. Also, points may be statistically out-of-control on the chart, but that may have been caused by a large outlier sometime in the past. In this way, large outliers may mask other potential outliers.

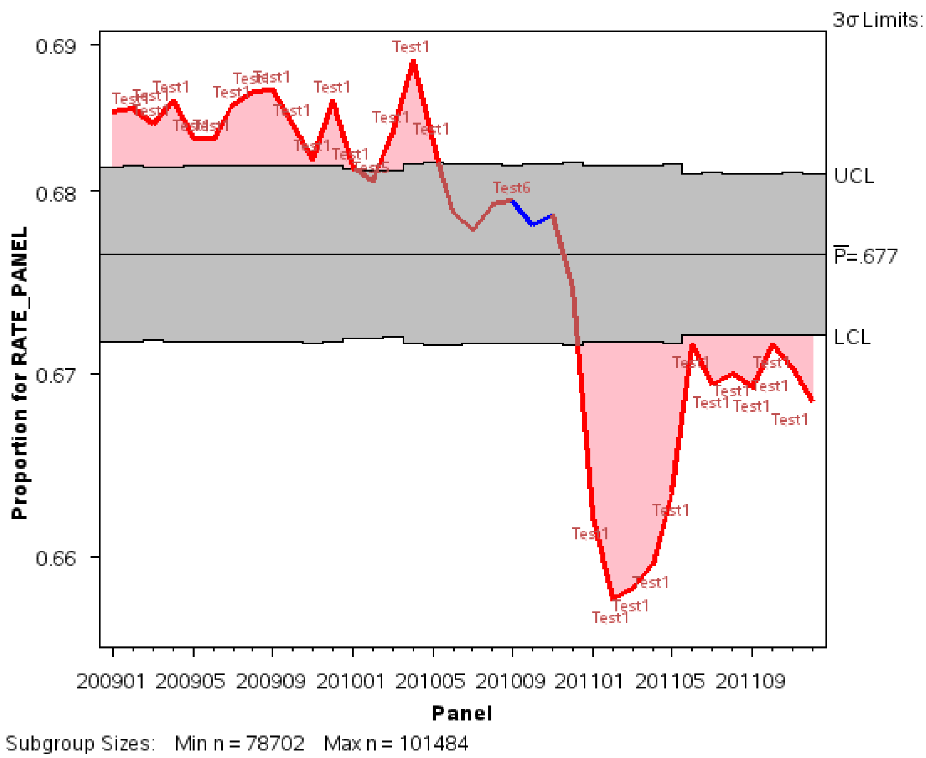
6.1. Case Study: Acreage

The question “How many acres is this house or mobile home on?” is broken up into four categories for analysis:

1. Less than one acre,
2. Between one and 9.9 acres,
3. Ten acres or greater, and
4. No response to acreage.

Note that respondents use checkboxes to answer this question. The response category "Less than 1 acre", also known as ACR1, was output as an outlier in the January 2011 panel for the U.S. English mail questionnaire. Figure 6 shows the December 2011 panel, which includes the entire time series from January 2009 through December 2011. Notice the large decrease below the lower confidence limit, starting in January 2011. Also, notice how the trend increases and approaches the lower confidence limit starting in June 2011. We believe both of these changes in trend are verification of known changes in the methodology for the ACS.

Figure 6. U.S. English Mail Questionnaire, December 2011 Panel, ACR1



Beginning with the January 2011 panel, there was a reallocation of the housing unit sample. This reallocation increased the number of sampling stratum from five to sixteen. It increased the sampling rates for the tracts with the smallest numbers of estimated occupied housing units. Instead of having one fixed sampling rate of 10% for only the smallest areas, we now have fixed rates of 15%, 10%, and 7% for the three smallest categories of governmental units. This reallocation effectively shifted sample to the smaller tracts. Additionally, the areas with the smallest population are usually the most rural, having most of their population spread across large land areas. This means that there were increased numbers of respondents with more than one acre of land, thus, the downward shift in the proportion.

Then, in June 2011, there was an increase to the sample size from 2.9 million addresses to 3.54 million addresses sampled per year. This increased the number of cases eligible to respond to this question from approximately 80,000 cases to 100,000 cases per panel. This sample increase was not uniform across the entire sample. The sampling rates for the smallest sampling stratum remain fixed at 15%, 10%, and 7%; this is so that sufficient sample is drawn from the smaller areas across the U.S.. This means that the sample increase was really only applied the larger areas, which have variable sampling rates. This implies that the overall proportion of respondents with large amounts of land decreased when the sample size was increased. This would explain why the proportion of people responding they have “less than one acre” of land would increase in June 2011.

One could potentially point out that if the data had been weighted, we might not have marked this ACR1 as an outlier, as all size areas would have been represented equally. However, viewing these proportions separated by the sampling substrata instead of aggregated to the national level gives the effect of weighting. After running the data separated by the sixteen substrata, there were no outliers. As expected, the proportions of respondents for ACR1 increased as the size of the tracts decreased. See Figure 7 for a table of the process averages for each substratum for the December 2011 panel.

Figure 7. Process Averages for the U.S. English Mail Questionnaire, December 2011 Panel, by Sampling Substrata, ACR1

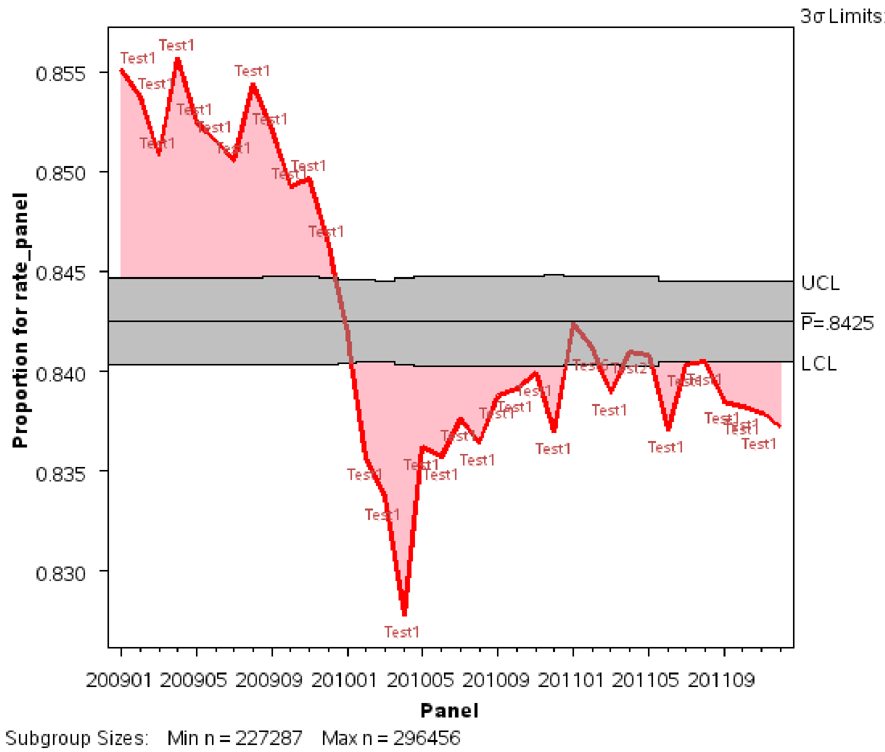
Sampling Substrata	Process Average	Sampling Substrata	Process Average
1 (Largest)	0.4637	**9**	0.7043
2	0.4546	**10**	0.7259
3	0.4654	**11**	0.7047
4	0.5271	**12**	0.7193
5	0.6854	**13**	0.7244
6	0.7259	**14**	0.7461
7	0.6849	**15**	0.8012
8	0.7219	**16 (Smallest)**	0.8161

6.2. Case Study: Citizenship

The question “Is this person a citizen of the United States?” is broken up into six categories for analysis:

1. Yes, this person was born in the U.S.,
2. Yes, this person was born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marinas,
3. Yes, this person was born abroad of U.S. citizen parent or parents,
4. Yes, this person is a U.S. citizen by naturalization,
5. No, this person is not a U.S. citizen, and
6. No response to citizenship.

Figure 8. U.S. English Mail Questionnaire, December 2011 Panel, CIT1



Note that respondents use checkboxes to answer this question, as well as a fill-in for the year of naturalization. The response category "Yes, this person was born in the U.S.", also known as CIT1, displays a very large decrease below the lower confidence limit, starting in approximately February 2010 and reaching its lowest point in April 2010. Figure 8 shows the December 2011 panel, which includes the entire time series from January 2009 through December 2011.

The lowest point happens to coincide with the 2010 Decennial Census day, April 1. Census advertising, which urges Americans to respond to the Decennial Census form, seems to have an effect on the response to the ACS. As demonstrated here, the overall proportion of response for "Yes, this person was born in the U.S." decreased around the time of the Census advertising campaigns were in effect. These advertisements attempt to increase the percentage of response for the demographics that tend to be less likely to respond. While no causal links have been proven, it appears that the overall increase in the number of respondents during Census time may have caused this decrease in proportion. Figure 9 shows the sample sizes for those eligible to respond to this question, before and after Census time.

This Census effect is visible in many other questions in the ACS for this same time period including: "How is the person related to person 1?", "Is the person of Hispanic, Latino, or Spanish Origin?", "What is the person's race?", "In the past 12 months, did anyone in the household receive Food Stamps or a food stamp benefit card?", among others.

Figure 9. Sample Sizes for Respondents Eligible for CIT1

Panel	Subgroup Sample Size
October 2009	243,486
November 2009	242,204
December 2009	246,493
January 2010	270,858
February 2010	282,506
March 2010	289,486
April 2010 (2010 Decennial Census)	248,919
May 2010	231,352
June 2010	234,512

7. Preliminary Results of Monitoring the Regional Office Realignment

The future of data collection at the Census Bureau features constrained and declining budgets, as well as increased challenges for data collection in the field. In response, the Census Bureau has decided to condense field activities to a six RO structure. This reduction, called the RO realignment, is the process by which six of the current twelve ROs are shifting data collection to other regions. This realignment is happening in seven waves, with the first wave having already occurred in January 2012. The goal is to close all six ROs by January 2013.

For us to monitor data quality during the RO Realignment, we began by developing and implementing a set of geography to use in monitoring the realigning regions. This geography is a composition of the unique combinations of the old RO, the new RO, and the wave in which the region is realigning. Combining the old RO, the new RO, and the wave number gave us 75 unique regions to monitor. Using this geography helps us pinpoint changes in individual regions across the country, as the steps of the realignment are carried out, and group areas that are encountering similar procedures.

The largest quantity of outliers for a single geographic region for the regions that shifted in January 2012 was 51 outliers in the geography that shifted from the Seattle RO to the Los Angeles RO. There were a total of 364 outliers across all 51,211 possible CAPI response category and geography combinations. 47 of the outliers were from the housing unit questions and 317 were from the person level questions.

Figure 10 shows the geographic regions that had the largest quantities of outliers. All other geographies had 19 or fewer outliers per region. 61 of the possible 75 regions had at least one outlier.

Of particular interest is the section of geography that transferred data collection from Seattle to Los Angeles. There were 51 outliers for January 2012, 17 of which were in a category of non-response. We are still investigating possible reasons for these outliers.

Figure 10. RO Realignment Geography: Largest Quantities of Outliers in January 2012

Old RO	New RO	Wave	Outliers
27 – Seattle	32 – Los Angeles	1	51
31 – Denver	31 – Denver	4	48
23 - Philadelphia	23 - Philadelphia	6	20

8. Conclusions and the Future

At this point, we are very excited and optimistic about our applications of SPC to the ACS data, including the adjudication system, SPARCS, which we have developed. We have a great deal of data to work with, and it has taken time to sift through all of it to find the important elements and areas on which we should focus. We have incorporated data from all three modes of data collection and we can analyze that data at many different geographic levels. To date, we have not found a truly significant data error, but we hope to catch any future data errors once our system has been fully implemented.

While we have made great strides in incorporating data, there is still much room for expansion. There are plans to implement new survey content and an additional mode of data collection, internet, which we plan to incorporate before they are implemented. We could also expand to cover the other three mail questionnaires (U.S. Spanish, PR English, and PR Spanish), group quarters data, and paradata collected along with the survey data. We could analyze more geography levels, including geographies as small as counties and tracts. As mentioned above in the acreage example, we could also switch to using weighted data instead of unweighted data, although that introduces the possibility for error in the weight calculations. Additionally, we could incorporate outside data, such as data from administrative records, other surveys, and more.

However, one of the largest untapped areas of expansion is comparing data across operations. Operations such as Failed Edit Follow-Up (FEFU) for the ACS have the overall goal of improving the data quality, but may also introduce errors into the data. If we could use the system to check across operations, we might be able to pick up some of these errors introduced by the operation itself. This would be a tremendous advancement in our current data quality measures, and we hope to be able to implement it soon.

References

Hefter, Steven and Marquette, Erica. *Using Statistical Process Control Techniques in the American Community Survey*. U.S. Census Bureau. 2011.