

Using Statistical Process Control to Understand Variation in Computer Assisted Personal Interviewing Data

Robyn Sirkis^{1*}

Matt Jans²

¹U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

²U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract

This paper discusses research using statistical process control with paradata obtained during data collection for the National Health Interview Survey (NHIS). Statistical process control (SPC) involves using statistical techniques to measure and analyze variation in operational processes. The goal with this approach is not to simply monitor, but to improve the quality of the process over time. For this paper, we group interviewers into statistical clusters based on census tract level housing unit and respondent demographic characteristics and produce control charts which examine the variation of the process over time for each cluster. We compare the means of interviewer performance indicators within each cluster to determine if they are significantly different from the overall mean of the process for the cluster, and examine some of the potential causes of process variation using selected control charts. We address advanced SPC techniques such as multivariate charting. Indicators of data quality used in the paper are item nonresponse and interview duration. The charts are intended to demonstrate how survey managers can use paradata to monitor the data collection process using SPC principles and techniques.

Key Words: Paradata, Statistical Process Control, NHIS

1. Introduction

There is an abundance of survey paradata or process data available to guide survey managers in their efforts to improve interviewer performance to maintain a certain level of data quality. Paradata are any information that describes a data collection process (Couper, 1998; Couper and Lyberg, 2005; Laflamme, 2008). Examples of paradata or process variables include interview pace, number of contact attempts, response rates, and item nonresponse rates. Examining paradata while the data are being collected provides an opportunity to make adjustments to procedures and may improve data quality in real time instead of waiting until the survey cycle is complete.

Paradata can be used to assess performance at the interviewer-level. Data quality can be compromised if an interviewer has few completed interviews, does not read the questions as worded, inappropriately skips over questions, or falsifies the responses. Because paradata are produced, in part, by characteristics of the sample units themselves (e.g., sociodemographic characteristics of households and geographic variability among

* Disclaimer: This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the authors and not necessarily those of the U. S. Census Bureau.

housing units) it is helpful to compare interviewers with similar workloads to one another. By controlling factors related to the sample units, we can assess more accurately whether a specific interviewer's performance is varying more than should be allowed.

There are many methods for analyzing survey paradata in a real-time production setting, of which statistical process control (SPC) is one. SPC approaches process management by analyzing and measuring the variation of an operational process over time. The fundamentals of SPC and the control chart technique were developed for manufacturing processes by Dr. Walter A Shewhart during the 1920's (Ryan, 1989). SPC tools include Pareto charts, scatter plots, fishbone diagrams, histograms, control charts, and analysis of means charts, but the focus of this paper is on control charts and analysis of means charts.

2. Background

The SPC techniques demonstrated in this paper were applied to paradata collected from the National Health Interview Survey (NHIS), and subsequently processed using the Performance and Data Analysis (PANDA) system. This section describes the NHIS and PANDA in more detail.

2.1 National Health Interview Survey (NHIS)

The National Health Interview Survey is an annual survey of the health of the civilian, noninstitutionalized household population of the United States, and is conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). The survey uses a multi-stage, clustered sample design, with oversampling of black, Hispanic, and Asian persons, and produces nationally representative data on health insurance coverage, health care access and utilization, health status, health behaviors, and other health-related topics. The survey produces nationally representative data on health insurance coverage, health care access and utilization, health status, health behaviors, and other health-related topics.

Roughly 650 trained interviewers with the U.S. Census Bureau, operating out of 12 regional offices (ROs), administer in-person interviews on laptop computers using computer-assisted personal interviewing (CAPI) with some telephone follow-up. Data collection is continuous throughout the calendar year. The survey instrument contains four main modules: household composition, family, sample child, and sample adult. In addition to the core survey modules, supplemental questions on special topics are added to the NHIS questionnaire each year.

The NHIS has embarked on extensive use of paradata. Paradata collected in conjunction with the NHIS are processed in a web-based tool called the Performance and Data Analysis (PANDA) system. Developed by the U.S. Census Bureau, PANDA tracks the performance of each interviewer during data collection to provide an early indication of the quality of specific data items. This tool provides timely data to the Census Bureau staff to determine whether interviewers are performing properly. Some of the performance measures tracked include interview durations, partial interview rates, and overnight completion rates. For more detail on PANDA, see Jans, Sirkis, Schultheis, Gindi, and Dahlhamer (2011) in the Proceedings of the Joint Statistical Meetings.

3. Statistical Analysis

The main objective of the control charts and tools demonstrated is to determine whether interviewers are collecting data in accordance with survey procedures. Applied to real-time data collection, these tools could facilitate prompt corrective action that would need to be taken if the data are not being collected properly.

3.1 Analysis Data

The data used in this paper come from the January 2008 through December 2010 NHIS. We used these data to create a demonstration situation in which we show how control charts can be used to improve interviewer performance. In this scenario we focus on December 2010 as our hypothetical “current month”. The quality indicators used in the analysis are sample adult interview pace, family interview pace, sample child interview pace and the item nonresponse on a question that asks for whom the respondent works. The interview pace measures were the estimated number of seconds per question for the respective interview modules. This was calculated by dividing the total number of seconds for the interview section by the number of questions asked in the interview section. We chose seconds per question rather than an total duration because there were different numbers of questions across years of data collection. The item nonresponse rate for the employer question is the number of don’t know and refused responses divided by the number of respondents who were asked the question.

3.2 Creation of the Clusters

One characteristic of sample units that influence the paradata they produce is where they are located. To compare interviewers who worked on cases that were similar to each other, independent of where those cases were located, we grouped census tracts into clusters based on housing unit characteristics, respondent demographic characteristics, and socioeconomic characteristics of the tracts. The characteristics of the tracts come from the U.S. Census Bureau 2000 Planning Database (PDB). The PDB contains 100-percent data from the Census 2000 and sample data from a sample of housing units that received the Census 2000 long form. See Bruce and Robinson, 2006 for a description on the variables on the database.

Clusters are created at the tract level. Therefore interviewers can be in more than one cluster if they had cases in more than one tract. A cluster with fewer than five interviewers, was combined with the nearest cluster with similar characteristics. New construction cases without a census tract identifier were placed into their own cluster.

To create the clusters, we first performed a variable reduction technique using the VARCLUS procedure in SAS to reduce the number of variables in the PDB to a manageable set. This was done to remove redundant statistical information, and simplify interpretation challenges present in any analyses with too many variables (Nelson, 2001). Similar to principal component analysis, the procedure divides variables into either disjoint or hierarchical clusters and finds groups of variables that are as correlated as possible among themselves and as uncorrelated as possible with variables in other clusters.

A variable is selected from each cluster that has a high correlation with its own cluster and a low correlation with other clusters by using the $1 - R^2$ ratio. R^2 is the proportion of variance accounted for by the clusters. The formula is below.

$$1 - R^2 \text{ Ratio} = \frac{1 - R^2_{\text{own cluster}}}{1 - R^2_{\text{nearest cluster}}}$$

After reducing the number of variables, a second clustering technique was used to group the census tracts into clusters using the *k means* model method. This involved using Euclidean distances so the cluster centers are based on least squares estimation. The cluster centers are the means of the observations assigned to each cluster when the algorithm is run to complete convergence. The FASTCLUS procedure was used to perform the analysis and involved selecting the number of clusters within RO. The procedure outputs statistics such as the approximate expected overall R^2 and the cubic clustering criterion (CCC).

The CCC is a comparative measure of the deviation of the clusters from the distribution expected if the data points were obtained from a uniform distribution (Sarle, 1983). The constant terms were chosen on the basis of extensive simulation results. A value greater than 2 indicates a good clustering algorithm.

$$CCC = \ln\left[\frac{1 - E(R^2)}{1 - R^2}\right] \times \frac{\left(\frac{np}{2}\right)^{0.5}}{(.001 + E(R^2))^{1.2}}$$

where $E(R^2)$ is the expected R^2 , and p is an estimate of the dimensionality of the between cluster variation.

The number of clusters was increased until the R^2 was greater than 0.70, at which point the clustering algorithm stopped. Many additional clusters would need to be added in order to obtain a much higher R^2 than 0.70. The decision was also based on what would be practical for implementation in a survey field environment.

Nine variable reduction clusters were chosen resulting in nine variables for grouping the tracts into interviewer comparison clusters. The nine variables were percent Hispanic origin, percent of units in structure containing ten or more housing units, percent of the population below poverty, percent Asian, percent vacant housing units, percent of the population under age 18, percent Native Hawaiian and Pacific Islander, percent White, and percent linguistically isolated other language households. Table 1 shows an example of how a variable was selected based on the $1 - R^2$ ratio for Cluster 1. Percent Hispanic origin had the smallest ratio.

Table 1: Results of Variable Reduction Analysis in Cluster 1

Variable	Next Closest Cluster R²	Own Cluster R²	1-R² Ratio
Percent Hispanic Origin	0.2558	0.8963	0.1394
Percent Linguistically Isolated Spanish Households	0.2875	0.9005	0.1397
Percent Linguistically Isolated Households	0.2601	0.8529	0.1988
Percent Other Race	0.2815	0.8312	0.2349
Percent Occupied Units With More Than 1.5 Persons Per Room	0.2896	0.7934	0.2908

The number of interviewer comparison clusters ranged from three to seven. An additional cluster was added within each RO for the new construction cases making it four to eight clusters. Table 2 shows an example of the clustering statistics for two select ROs. There was one cluster in an RO that had less than five interviewers and had to be collapsed with another cluster.

Table 2: Example Cluster-Level Statistics By Regional Office

Regional Office	Number of Clusters	Number of Clusters Including New Construction	Cubic Clustering Criterion	R²
1	4	5	21.13	0.73
2	3	4	9.45	0.70

3.3 Constructing Survey-Level and Interviewer-Level Control Charts within Clusters

The clustering described above provided the framework to apply control charts. Recall that the purpose of using the clustering technique to reduce geographic variation among sample units was so that interviewers and interviews plotted in control charts would be more similar to each other than if we had simply done national or RO-level control charts without clustering.

The application of control charts to the hypothetical “current month” (i.e., December 2010) proceeds in several steps. These steps aim to show how one would determine if the process reflected in the paradata collected by individual interviewers was not in control. The steps are as follows:

Step 1: We constructed the cluster-level control chart within RO including the data from January 2008 to November 2010. Mean and standard deviation charts were used for interviewer pace and proportion of nonconforming unit charts were used for item nonresponse.

Step 2: We constructed the trial control limits by removing observations in the chart that were outside the control limits. That is, we brought the process into control for demonstration purposes. This in-control system provides the process average and control limits going forward. In practice, out-of-control points due to special causes would need

to be individually identified and removed when constructing the trial control limits (Ryan, 1989).

Step 3: We constructed the final cluster-level control chart within RO including data from January 2008 to December 2010. That is, we added December, 2010's data to the chart.

Step 4: We constructed the analysis of means (ANOM) chart if the process in the final cluster-level control chart within RO was not in control.

Step 5: We constructed the interviewer-level control chart if the average interview pace or item nonresponse rate for the interviewer was outside the decision limits in the ANOM charts.

An example of these steps begins with Figure 1 using sample adult interview pace. This control chart includes data for Cluster 4 in RO 1. Months were used as the rational subgroup in order to have high between group variability and low within group variability when special cause variation is present (Montgomery, 1985). The rational subgroup size ranges from 17 to 194 interviews.

The center line or process average in Figure 1 is the average value of the sample adult interview pace and is bounded by upper and lower control limits. The process average in Figure 1 is 8.1.

The process is not in control if the sample points are outside the control limits indicated by the shading between the limit and the mean for the rational subgroup. One method of interpreting the charts is to first explore the standard deviation chart (the lower chart in Figure 1) and determine if it is in control. For the points where the standard deviation exceeds the limits the corresponding points in the average chart should not be interpreted regardless of whether it is above or below the control limits. Test one is signaled when the sample point is outside the limits. Test three is signaled when 6 neighboring points are increasing or decreasing (Western Electric Rules, 1956). Only tests one and three were used to identify special causes of variation.

The subgroup average was outside the control limits at the beginning of 2008 for several months. The subgroup average was steadily increasing for six months from May 2008 through October 2008 which is shown by the number 3 in the graph. Test 1 was signaled in the graph for five months since the average was outside the control limits. The subgroup standard deviation was outside the control limits for 19 months. The tests were not used in any of the standard deviation charts because there were many months where the standard deviation was outside the limits.

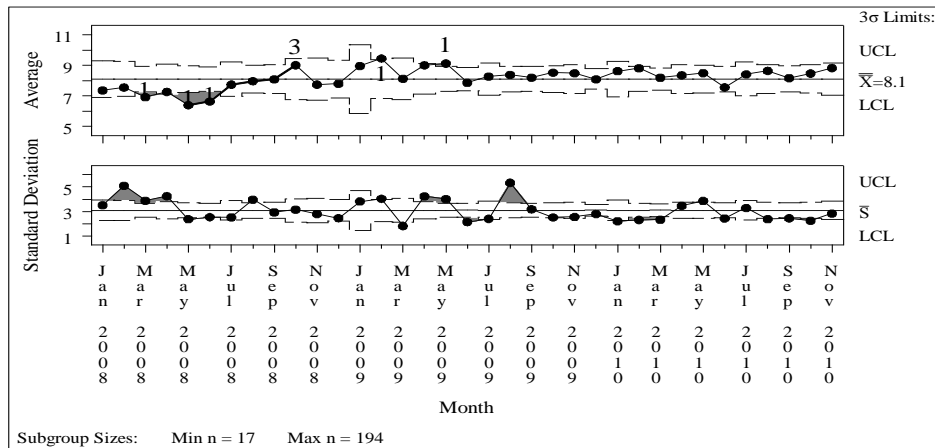


Figure 1: Sample Adult Interview Time Where Process Is Not In Control Excluding December 2010 (Regional Office = 1, Cluster=4)

The goal of statistical process control is to remove special cause variation that is nonrandom and unpredictable (Gitlow, 1989). In our example steps we do this by simply removing points from Figure 1 that are beyond the control limits in the mean and standard deviation charts. This was done iteratively until the process was in control as seen in Figure 2. In a real-world implementation of our process, a manager would have to assign causes to each of the points removed, thus validating that they were actually the result of special cause variation. (Ryan, 1989). This would be done in an initial research or test phase, but would result in a chart similar to Figure 2. If the process average and control limits from the test phase as seen in Figure 2 are carried forward into our hypothetical current month (December 2010) displayed in Figure 3, we can judge whether the current month is out of control. In this demonstration case it is, so we move forward to Step 4 and create ANOM charts for the interviewers represented by that data point.

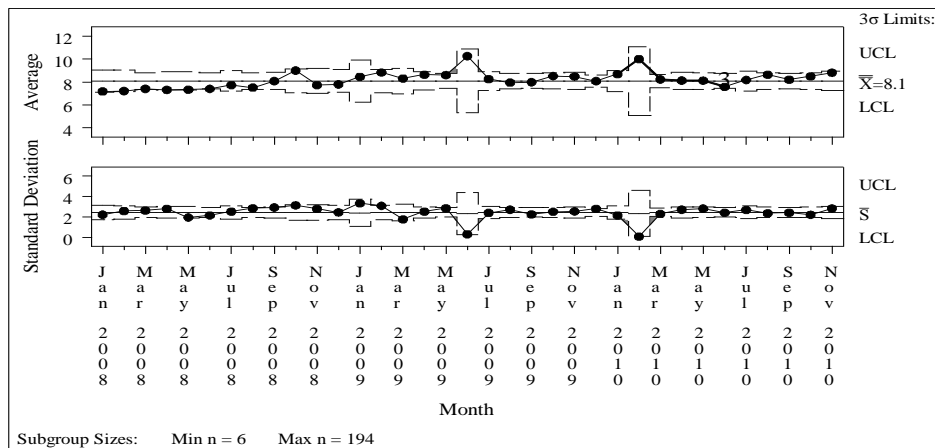


Figure 2: Sample Adult Interview Time Where Process Is In Control Excluding December 2010 (Regional Office = 1, Cluster=4)

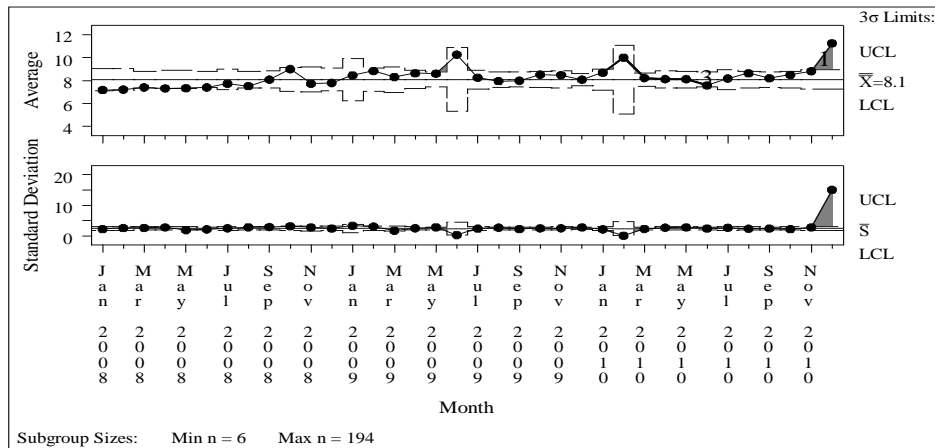


Figure 3: Sample Adult Interview Time Where Process Is Not In Control Including December 2010 (Regional Office =1, Cluster=4)

ANOM charts[†] were constructed for each cluster within RO for December 2010 because the associated cluster level control chart was not in statistical control. An example of this chart for Cluster 4 in RO 1 is shown in Figure 4. There were three interviewers where the average was outside the decision limits (1, 7, and 9) indicating that the average sample adult interview for that interviewer is significantly different from the overall average. An interviewer-level control chart was constructed for these three interviewers.

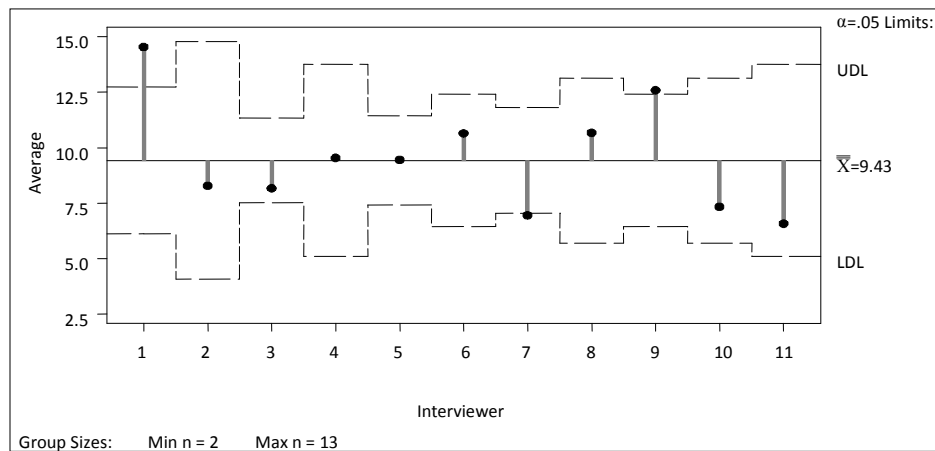


Figure 4: Sample Adult Interview Time ANOM Chart for December 2010 (Regional Office =1, Cluster=4)

Interviewer-level control charts were constructed for those interviewers within each cluster and RO that had average interview pace outside the decision limits in the ANOM charts (Step 5). Figure 5 shows Interviewer 7’s chart as an example.

[†] The ANOM procedure in SAS produces these types of charts.

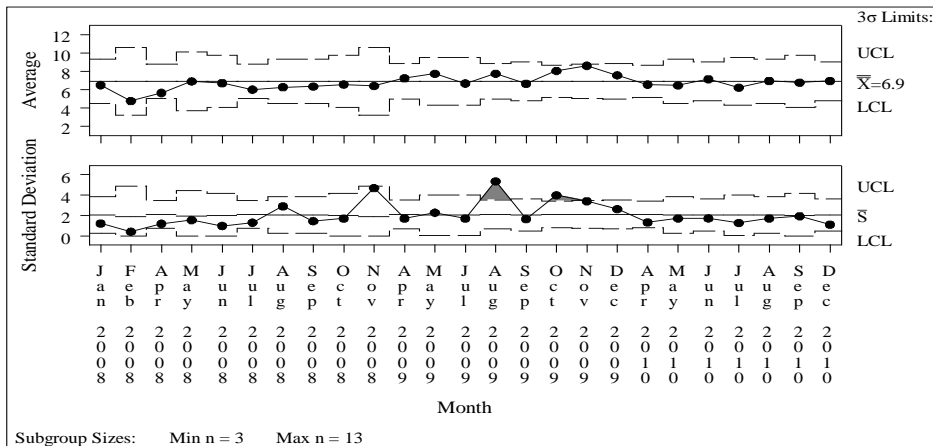


Figure 5: Sample Adult Interview Time Where Process Is Not In Control (Regional Office =1, Cluster=4 Interviewer=7)

The process was not brought into control for the previous historical data before including the current month. The observations that were removed in the cluster level control chart were also removed from the interviewer-level control chart. There can be more than one interviewer-level control chart for an interviewer that had cases in more than one cluster. Months with only one case were excluded from the interviewer-level control charts.

3.4 Item Nonresponse Process Control Charts

We do not demonstrate all 5 steps with the item nonresponse indicator, but they could be conducted in much the same manner as with the interview pace. Figure 6 is a cluster-level proportion of nonconforming units chart for the work item nonresponse rate. The process average and the rate on the y-axis are not percentages. The process average is .034 (3.4% missing data) denoted by \bar{p} in the chart. The important distinction between this chart and the charts used for interview pace is that in our chart the lower bound is zero.

The process was brought into a state of statistical control for January 2008 to November 2010 before December 2010 was included in the control chart. The process is in a state of statistical control when December 2010 was added to the cluster level control chart.

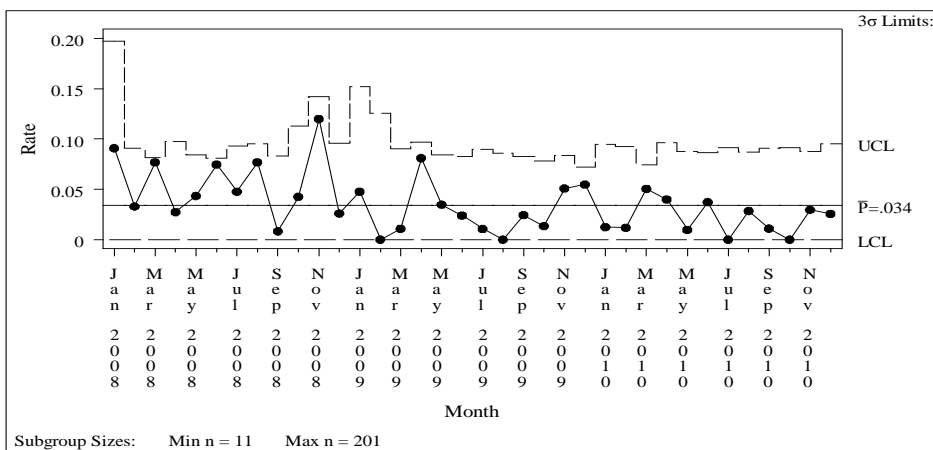


Figure 6: Who Do You Work For Item Nonresponse Rate Where Process Is In Control (Regional Office = 2, Cluster=3)

3.5 Moving Beyond One-Chart-Per-Indicator

Multivariate charts that combine two or more quality indicators into one measure using the Hotelling's T distribution (Ryan, 1989) would be applicable in situations where multiple correlated indicators need to be monitored. A multivariate chart was constructed using the sample adult interview pace and the family interview pace because these two measures are correlated with one another and have the same number of total observations. As a technical requirement for the multivariate chart in the JMP software, a simple random sample without replacement was conducted in each month to obtain the same number of observations in each month. The month that had the smallest sample size ($n=426$) determined the number of observations selected for each month. In application, individual control charts for each indicator would only need to be constructed if the process was not in control in the multivariate chart.

In our demonstration chart as displayed in Figure 7, the process was not in control since the points are above the upper control limits. Therefore univariate charts would be constructed for the sample adult interview pace and family interview pace measures and examined individually.

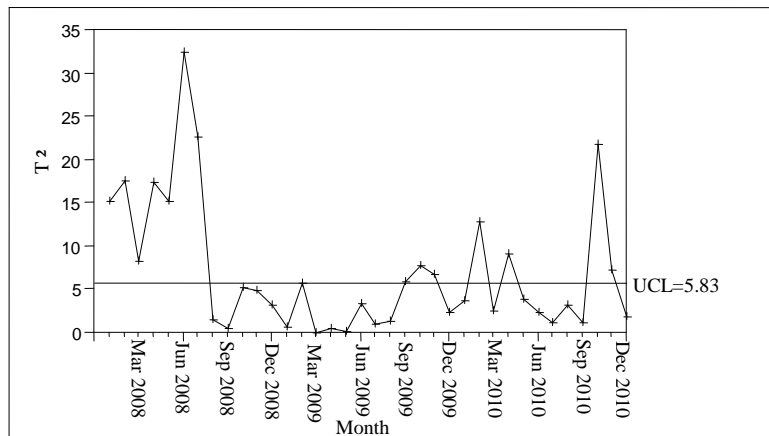


Figure 7: Multivariate Control Chart

4. Implementation

4.1 From Control Charts to Management Tools

SPC techniques can identify interviewers whose process is not in a state of statistical control but managers should examine other information before taking action. For interviewers for whom an interviewer-level control chart was produced, we constructed tables with information for every cluster in which the interviewer worked. Table 3 shows whether the measure exceeded the limits for the various charts for interviewer 7. The table also shows how many months were outside the limits for the cluster level and interviewer-level control charts. A 'Yes' entry in the control chart column indicates that the average or standard deviation for the sample adult interview pace was outside the control limits. A 'Yes' entry in the ANOM chart column indicates that the interviewer's average was significantly different from the overall average. The entry 'N/A' in Cluster 3 means "Not Applicable" indicating the interviewer did not have any cases in that cluster.

Cluster 4 was the only cluster where the average or standard deviation for the sample adult interview pace was outside the limits in the interviewer-level control charts.

Table 3: Sample Adult Interview Times – Summary of Charts

Cluster	Cluster Level			Interviewer Level		
	Jan 2008 to Nov 2010	Dec 2010	Number of Months Outside Limits	ANOM Chart	Control Chart	Number of Months Outside Limits
1	Yes	No	8	No	No	0
2	Yes	Yes	8	No	No	0
3	N/A	N/A	N/A	N/A	N/A	N/A
4	Yes	Yes	21	Yes	Yes	2
5	Yes	No	12	No	No	0

Table 4 shows the number of complete and total cases for the interviewer. The outliers were removed before the numbers of complete and total cases were calculated for the interviewer. The majority of the cases were in Cluster 4. It may be a bigger issue if the majority of the cases occur in a cluster where the interviewer's process was not in control.

Table 4: Sample Adult Interview Times - Number of Cases

Cluster	Number of Cases January 2008 to November 2010		Number of Cases December 2010	
	Complete	Total	Complete	Total
1	3	4	2	4
2	12	21	0	0
3	N/A	N/A	N/A	N/A
4	193	334	9	12
5	25	40	0	0

It is extremely helpful to examine whether the interviewer had an average or standard deviation for other quality indicators outside the limits. The SPC techniques were used on the sample adult, family respondent, and sample child interview times. Table 5 shows whether the measure was outside the limits for the cluster-level control charts, while Table 6 shows whether the measure was outside the limits for the interviewer-level control charts. Cluster 4 was the only cluster where the average or standard deviation for the sample adult interview pace and family respondent interview pace was outside the limits in the interviewer-level control charts. Viewing the data in each of these tables might help a supervisor determine what next steps to take with an interviewer who appeared to be out of control on the basis of control charts alone.

Table 5: Interview Times - Summary of Cluster Level Control Charts

Cluster	Sample Adult		Family Respondent		Sample Child	
	Jan 2008 to Nov 2010	Dec 2010	Jan 2008 to Nov 2010	Dec 2010	Jan 2008 to Nov 2010	Dec 2010
1	Yes	No	Yes	No	Yes	No
2	Yes	Yes	Yes	Yes	Yes	No
3	N/A	N/A	N/A	N/A	N/A	N/A
4	Yes	Yes	Yes	Yes	Yes	No
5	Yes	No	Yes	No	Yes	No

Table 6: Interview Times - Summary of Interviewer Level Control Charts

Cluster	Sample Adult		Family Respondent		Sample Child	
	ANOM Chart	Interviewer SPC Chart	ANOM Chart	Interviewer SPC Chart	ANOM Chart	Interviewer SPC Chart
1	No	No	No	No	No	No
2	No	No	No	No	No	No
3	N/A	N/A	N/A	N/A	N/A	N/A
4	Yes	Yes	Yes	Yes	No	No
5	No	No	No	No	No	No

5. Conclusion

There is a wealth of paradata that can be used to improve interviewer performance in real time. Data quality can be compromised if interviewers are not following survey procedures. SPC techniques allow managers to identify interviewers whose data need further examination. We have shown examples of how these charts can be used in practice by 1) displaying out-of-control processes, 2) demonstrating how to cluster to avoid confounding interviewer effort with characteristics of sample units, 3) suggesting techniques for monitoring individual interviewers, and 4) suggesting techniques for monitoring multiple indicators simultaneously. While the techniques are complex, we see them as implementable and likely useful for supervisory staff. We hope that readers will find these results interesting and our techniques useful.

Acknowledgements

James Dahlhamer and Renee Gindi at the National Center for Health Statistics were indispensable in the development of this work and writing of the paper. They deserve our deepest gratitude.

References

- Bruce, A., and Robinson, G. (2006). Tract-level planning database. *Internal Census Bureau Report*. U.S. Census Bureau, Suitland, MD.
- Centers for Disease Control and Prevention (2009). Retrieved May 20, 2011 from <http://www.cdc.gov/nchs/nhis.htm>.

- Couper, M. (1998). Measuring survey quality in a CASIC environment. *Proceedings of the Joint Statistical Meetings*.
- Couper, M. and Lyberg, L. (2005). The use of paradata in survey research. *Proceedings of the 55th Session of the International Statistical Institute, Sydney, Australia*.
- Gitlow, H., Gitlow., Oppenheim, A., Oppenheim, R. (1989). *Tools and Methods for the Improvement of Quality*. Illinois: Richard D. Irwin.
- Jans, M., Sirkis, R., Schultheis, C., Gindi R., and Dahlhamer J. (2011). Comparing CAPI trace file data and Quality Control Reinterview data as methods of maintaining data quality. *Proceedings of the 2011 Joint Statistical Meetings, Miami Beach, FL*.
- Laflamme, F. (2008). Understanding survey data collection through the analysis of paradata at Statistics Canada. *Paper presented at the Annual Meeting of the American Association for Public Opinion Research*.
- Montgomery, D., T. P. (1985). *Statistical Quality Control*. New York: John Wiley & Sons.
- Nelson, D. (2001). Variable reduction for modeling using PROC VARCLUS. *SAS Users Group International Proceedings*.
- Ryan, T. P. (1989). *Statistical Methods for Quality Improvement*, 2nd edition. New York: John Wiley & Sons.
- Sarle, W. S., (1983) *Cubic Clustering Criterion*, SAS Technical Report A-108, Cary, NC.
- Western Electric Co., Inc (1956). *Statistical Quality Control Handbook*, 2nd ed. Easton: Mack Printing Company.