Producing Control Charts to Monitor Response Rates for Business Surveys in the Economic Directorate of the U.S. Census Bureau[1]

Yarissa González, Broderick E. Oliver

U.S. Census Bureau
yarissa.gonzalez@census.gov, broderick.e.oliver@census.gov

Abstract

Control charts are powerful tools for monitoring ongoing processes, providing evidence of the process's stability and capability. We applied control chart techniques to the unit response rate (URR), a commonly used business survey quality indicator. At the U.S. Census Bureau, this measure is computed as part of survey processing and is subject to targeted benchmarks. Control charts will provide program managers a broader view of the unit response process and will help them to assess if their current procedures are capable of meeting a mandated benchmark. This paper presents our recommended methods for producing a *p*-chart for the URR. We employed simulation studies to address the challenges pertaining to this particular response indicator. Our objective was to find methods that could produce control limits that were reasonably precise and relatively easy to calculate.

1. Introduction

Business surveys in the Economic Directorate of the U.S. Census Bureau compute two types of response rates for usage as performance measures and quality indicators: a respondent level response rate and weighted-item level response rate. In this paper, we focus on the respondent level response rates**,** i.e., the unit response rates (URR). The URR is defined as the unweighted proportion of the sampled units that are eligible for data collection that "respond" to a survey. The URR is used as a performance measure (compared to predetermined benchmarks) and is an implied measure of data quality. It can also serve as a potential indicator of nonresponse bias. The Office of Management and Budget (OMB) offers the following guideline: "Given a survey with an overall unit response rate of less than 80 percent, conduct an analysis of nonresponse bias" (Federal Register Notice, 2006). The implication is that surveys with nonresponse rates greater than 20 percent may incur nonresponse bias. However, this is just a guideline as a survey may have a response rate lower than 80 percent and still be representative (Groves and Brick, 2008).

In this paper we move away from viewing the URR as simply a performance measure to be compared to a benchmark measure. Instead we seek to comprehend the variation of these response rates over time. By using control charts to monitor the URR, our focus switches from a single benchmark measure to a range of acceptable values that indicate if the response rate process is stable or not. By **process**[2], we mean a set of activities that follow some logical flow, where value is added and some output/result is expected.

All processes, including the response rate process that business surveys at the U.S. Census Bureau employs to get businesses to complete and return survey forms, have some form of variation. For example, the proportion of completed forms returned varies from one statistical period to the next. To understand this variation, one can track the unit response rates over time via a control chart. The control chart can help program managers to distinguish between variation that is inherent in the process and expected and variation that may not be. In addition to

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

[2] Source: Workbook on Leadership taught at a Six Sigma seminar conducted by **Six Sigma Qualtec** in Washington, D.C. on June 26, 2003.

monitoring variation, the control chart can be used to assess the response rate capability, particularly in light of mandated benchmark measures (Thompson and Oliver, 2010).

With response rates, the concerns in a production setting arise when a decrease occurs, especially when there appears to be a trend. Control charts can detect trends – downward or upward. In the case of the former, programs managers may make deliberate changes to their response rate process to increase their response rates. In the case of the latter, program managers can use the control charts to validate the effectiveness of their process intervention.

2. Analysis of the URR Process within a Statistical Process Control Framework

Regardless of how well a production process operates, there will always be an inherent or natural variability that exists. This natural variability or "background noise" is considered normal. In the framework of statistical quality control, this variability is often referred to as "a stable system of chance causes." A process that is operating with only chance causes of variation present is said to be in **statistical control (stable)**. In practice, most processes do not remain stable forever. In fact, control charts were first introduced to monitor manufacturing processes, where user error, equipment error, or differences due to the input material can lead the process to be unstable.

In a business survey environment, response processes could be affected by changes in reporting unit personnel, accounting practices, revised questionnaires, or the introduction of new collection methodology, for example. These sources of variability that are not part of the chance causes of variation are called **special causes** or **assignable causes**. A process that is operating in the presence of special causes is said to be **out of control** or **unstable**. If the special cause(s) is (are) found and corrected, the system returns to a state of control (Thompson & Oliver, 2010; Montgomery, 2005 pp. 148-149).

The statistical process control framework generally assumes a large number of measurements are available in a relatively short time. However, this is simply not true for the response rate process of an ongoing survey. At best, we can obtain twelve unit response rate values in a given year for a monthly survey and as few as four for a quarterly.

2.1 Formal Definition of the Unit Response Rate

In Appendix D3-B, to the Census Bureau Standards (2011), the URR computed for statistical period t is defined as:

$$URR_t = \frac{R_t}{E_t + U_t} = \frac{R_t}{n_t}$$

Where

R_t = count of reporting units that were eligible for data collection in the statistical period and classified as a response[3]

E_t = count of reporting units that were eligible for data collection in the statistical period

U_t = count of reporting units in the statistical period whose eligibility for reporting could not be determined

2.2 The Control Chart for the Unit Response Rate

We propose using a *p*-chart to monitor the URR process. Consider a process that consists of n_t independent trials conducted for statistical period (time) t where the outcome of each trial is dichotomous: "success" or "failure" (i.e., conforming or not conforming) and the probability of success on any trial (called Bernoulli trials) is p (Montgomery, 2005). With respect to the URR, we monitor the number reporting units that "respond" (i.e., conform) to a survey during the designated statistical period (R_t), recognizing that the response process follows a Binomial distribution in the ideal sense only. In practice, not every reporting unit has the same probability, p, of responding to the survey.

[3] A **respondent** is an eligible reporting unit for which (1) an attempt was made to collect data; (2) the unit belongs to the target population; and (3) the unit provided sufficient data to be classified as a response.

For example, larger businesses are often more likely to respond than smaller ones – particularly the businesses that are included in the sample with certainty.

A *p*-chart plots individual process measures (URR_t) against a centerline (a process average) and control limits. To develop these statistics, we assume the following:

- The production process is operating in a stable manner across a predetermined number of consecutive statistical periods, *T*.
- The URR value is approximately constant across consecutive statistical periods.
- Conditioning on the statistical period *t,* the URR is an exact value.
- Each reporting unit in the t^{th} statistical period has the same probability, *p* of "responding" to the survey.
- The response to the survey is binomially distributed: $R \sim \ bin\ (n\ , p)$

The true probability of responding, *p* is unknown and is estimated by a measure of central tendency obtained using the *T* most recent URR values: $\hat{p}_i, \hat{p}_{i+1}, \hat{p}_{i+2}, \cdots \hat{p}_I,$ that fall on the time interval [*i*, I]. In this paper, we consider two different estimators for *p* : a rolling process mean, $\hat{p}_{mean}$ and a rolling process median, $\hat{p}_{med}$ computed as follows.

Given = URR_t at time *t*,

$$\hat{p}_{mean} = \frac{1}{T}\sum_{t=i}^{I}\frac{R_t}{n_t} = \frac{1}{T}\sum_{t=i}^{I}\hat{p}_t$$ = **mean** value of URR over the studied time interval where T= I-i+1

$\hat{p}_{med}$ = median[$\hat{p}_t$] $_{t=i}^{I}$ = **median** value of URR over the studied time interval

$$\hat{\sigma}_m = \sqrt{\frac{\hat{p}_m(1-\hat{p}_m)}{\bar{n}}}$$ = estimated standard error of $\hat{p}_m$ (m = *mean* or *med*)

$$\bar{n} = \frac{\sum_{t=i}^{I} n_t}{T}$$ = average number of reporting units for the *T* consecutive statistical periods

The Upper Control Limit (UCL) is computed as $\hat{p}_m + 3\hat{\sigma}_m$, and the Lower Control Limit (LCL) is computed as $\hat{p}_m - 3\hat{\sigma}_m$.

The assumption is that the n_t are approximately the same across statistical periods.

The justification for development of the upper and lower control limits follows from the usage of a process mean to estimate the centerline, invoking the Central Limit Theorem and assuming a stable process. Thus, at $3\hat{\sigma}$ we would expect only 99.7 percent of the $\hat{p}_t$ to fall between $\hat{p} + 3\hat{\sigma}$ and $\hat{p} - 3\hat{\sigma}$. The 0.3 percent that fall outside of this range either do so by chance occurrence or can be attributed to an assignable cause. An investigation would be needed to determine the reason. These control limits are typically called "3-sigma" control limits. The larger the value of the sample size, the smaller the width of the control limits (Montgomery, 2005, pp. 152-153).

Although the above theory on *p*-charts calls for the mean as a measure of central tendency, we also incorporated the median due to the nature of our data. We recognize this as a limitation.

$$\hat{p}_t = \frac{R_t}{n_t}$$

Tague (2004) provides the following *general* guidelines to indicate when a stable process *may* have become unstable:

1. A single point plots outside the control limits.

2. Two out of three consecutive points that plot on the same side of the centerline exceed 2σ

3. Four out of five consecutive points plot on the same side of the centerline exceed 1σ

4. Eight consecutive points plot on the same side of the centerline

5. An obvious consistent, nonrandom pattern (e.g., an upwards or downwards monotone trend)

2.2.1 Related Issues in Developing Control Charts for URR

Upon first inspection, the derivation of the proposed formulae for the parameters for the URR control chart discussed in Section 2.2 appears to be relatively simple. All that is needed is an estimate of the process average to derive the control limits. However, the estimate of the response rate process average is affected by three separate factors: (1) the number of reporting units, n_t, in the t^{th} statistical period; (2) the number of consecutive statistical periods of estimates, T, used for the rolling average; and (3) the type of measure of central tendency used (i.e. mean or median).

If $\overline{n}$ is too small, then $\hat{\sigma}_m$ will be disproportionately large. The resulting control limits will be too wide, making it difficult to detect when the response rate process is out of control. If the number of consecutive statistical periods used for the rolling average, *T,* is too small, then a single outlier can overly affect the estimate of the process average. On the other hand, if T is too large, it will be more difficult for the control chart to indicate a potential shift in the process average when there is one.

We provide two illustrations below to indicate how the estimate of the process average is affected by the choice of averaging method. The first example illustrates a response rate process where a one-time outlier of 0.80 occurs in statistical period *t*. The second example illustrates a scenario where this outlier was an indication of an actual shift in the process average from 0.70 to 0.80.

Example 1: One Time Outlier with T = 5

Rolling Average *t*: (0.69, 0.70, 0.70, 0.71, **0.80**) $\hat{p}_{mean} = 0.72$; $\hat{p}_{med} = \mathbf{0.70}$
Rolling Average *t*+1: (0.70, 0.70, 0.71, **0.80**, 0.72) $\hat{p}_{mean} \approx 0.73$; $\hat{p}_{med} = 0.71$
Rolling Average *t*+2: (0.70, 0.71, **0.80,** 0.72, 0.70) $\hat{p}_{mean} \approx 0.73$; $\hat{p}_{med} = 0.71$
Rolling Average *t*+3: (0.71, **0.80,** 0.72, 0.70, 0.69) $\hat{p}_{mean} \approx 0.72$; $\hat{p}_{med} = 0.71$
Rolling Average *t*+4: (**0.80,** 0.72, 0.70, 0.69, 0.70) $\hat{p}_{mean} \approx 0.72$; $\hat{p}_{med} = 0.70$
Rolling Average *t*+5: (0.72, 0.70, 0.69, 0.70, 0.71) $\hat{p}_{mean} \approx \mathbf{0.70}$; $\hat{p}_{med} = 0.70$

In the above example, in statistical period *t*, the estimate of the process average includes a single outlier of 0.80. After first observing this outlier, one would question whether this particular value is an indication of a potential upward shift in the response rate process or is simply an aberration due to chance. Additional evidence would be needed to answer this question. In the meantime, $\hat{p}_{mean}$ is immediately affected by the one-time outlier. Consequently the values of the control limits will be affected until the one-time outlier value is no longer included in the computations. In contrast, the one-time outlier has very little effect on the control chart obtained using $\hat{p}_{med}$.

Example 2: Permanent Shift with T = 5

Rolling Average t: (0.69, 0.70, 0.70, 0.71, **0.80**) $\hat{p}_{mean} = 0.72$; $\hat{p}_{med} = \mathbf{0.70}$

Rolling Average t+1: (0.70, 0.70, 0.71, **0.80, 0.81**) $\hat{p}_{mean} \approx 0.74$; $\hat{p}_{med} = 0.71$

Rolling Average t+2: (0.70, 0.71, **0.80, 0.81, 0.80**) $\hat{p}_{mean} \approx 0.76$; $\hat{p}_{med} = \mathbf{0.80}$

Rolling Average t+3: (0.71, **0.80, 0.81, 0.80, 0.80**) $\hat{p}_{mean} \approx 0.78$; $\hat{p}_{med} = 0.80$

Rolling Average t+4: (**0.80, 0.81, 0.80, 0.80, 0.81**) $\hat{p}_{mean} \approx \mathbf{0.80}$; $\hat{p}_{med} = 0.80$

In this second example, the true process average has actually shifted to a higher value. Of course, additional evidence would be needed before one can conclude this, but as stated before, the mean provides an immediate indication of a potential shift. Via the mean, the estimate of the process average gradually increases over four subsequent time periods until the new process average is achieved. The median's shift to the new process average is not gradual in our example; it jumps to this new process average value after a two statistical period "burn-in".

Note: The results above are for illustrative purposes only. Had we used an outlier less than 0.80 for example, the conclusions pertaining to the mean would be different.

3. Methodology

The purpose of our study was to recommend URR control chart settings for the following parameters:

- n_t , the number of eligible reporting units used in computations for statistical period t
- T, size of the rolling average interval (number of statistical periods used to compute the process average)
- The measure of central tendency (mean or median) to use to calculate the estimate of the process average

Most business surveys will have a sufficiently large average number of eligible reporting units, n to produce a reliable estimate of the response rate process average, $\hat{p}_m$ and the associated measure of precision, $\hat{\sigma}_m$ at the program level. However, if the user wishes to produce the URR for a smaller subpopulation such as industry, the resulting control chart parameters may not be particularly useful. We used the following logic to develop a lower bound on n_t:

Assume that a program manager would want to inspect any URR value that is larger or smaller than $d = 0.02$ units of the process average. Then, in developing parameters for the control chart, we set the control limits at $\pm 3\sigma$, where $\sigma = \sqrt{\frac{p(1-p)}{n}}$. Let $3\sigma = d$, obtaining $n = \frac{9}{d^2} p(1-p)$. (See Montgomery, pp. 277-278).

With a fixed value of d and p, we can solve for n. However, p varies by survey. What do we do?

It can be shown that $\sigma = \sqrt{\frac{p(1-p)}{n}}$ is maximized for any n when $p = 0.5$. If $d = 0.02$ and $p = 0.5$, then $n = 5{,}625$. Thus, the conservative approach would require a minimum number of 5,625 eligible reporting units (on the average), allowing greater precision in the control limits for values of p other than 0.5 but enforcing the program-manager derived limits regardless.

Because our study involved examining the effects of varying parameter sizes on our control charts for our response rate process, we conducted a simulation study. For our simulation study, we generated a population of size 5,625. For each unit in the population, we created a time series of forty statistical periods, modeling the unit's binary response such that the probability of response ("1") for any unit was $p = 0.70$. Thus, the response propensity is independent both within and between statistical periods. Of course, in a survey setting, the unit response mechanism is not necessarily independent from one statistical period to the next, in the sense that a unit that responded in month t has a higher probability of responding in month t+1 than a unit that did not. Note that large population size ensures that any URR value that is greater or less than 0.02 from the centerline will be detected: the population size is much larger than actually needed.

In each statistical period, we calculated the URR by summing up the responding units. The unit response rates over the forty statistical periods varied but were all close to 0.70. To demonstrate that we had generated a stable response rate process, we plotted these forty URR values against the control chart constructed with centerline 0.70 and corresponding control limits of $0.70 \pm 3\sqrt{\frac{(0.70)(0.30)}{5625}}$. All values fell between the control limits. We used the simulated response rate population to study the effects of a one-time outlier and a process shift on the control chart parameters, examining eight different values of *T*, ranging from 5 to 12.

3.1 One-Time Outlier

We introduced an extreme outlier URR of 0.80 into our stable response rate process in statistical period 21. Thus, starting with statistical period 21 and ending with statistical period 21 + (*T*-1), this outlier could have an impact on the computed rolling averages. For example, if *T*=5, then the estimates of the process averages calculated from the following time intervals may be affected by this outlier: [17, 21], [18, 22], [19, 23], [20, 24], and [21, 25]. We examined the effect of this high outlier by varying the size of the rolling average interval, *T*=5 to 12, and by using two types of averaging methods: mean and median. We then repeated this experiment using a more realistic response rate outlier of 0.74.

3.2 Permanent Shift

This second scenario presents a permanent shift in the URR process beginning in the statistical period 21. The true URR process average is 0.70 in statistical periods 1 through 20; from statistical period 21 through 40, the true URR process average is $p = 0.80$. We repeated this experiment for a more realistic process average of $p = 0.74$.

4. Results

4.1 One Time Outlier

Figure 1 presents the URR control chart obtained using $\hat{p}_{mean}$ with a rolling average interval of size five (i.e., *T*=5) on the **stable** process data from the first twenty statistical periods. Hence, the process average was estimated using the mean of the five most recent URR values from statistical periods 16 through 20. The control chart obtained using the median instead of the mean to estimate the process average looks similar and is therefore not shown.

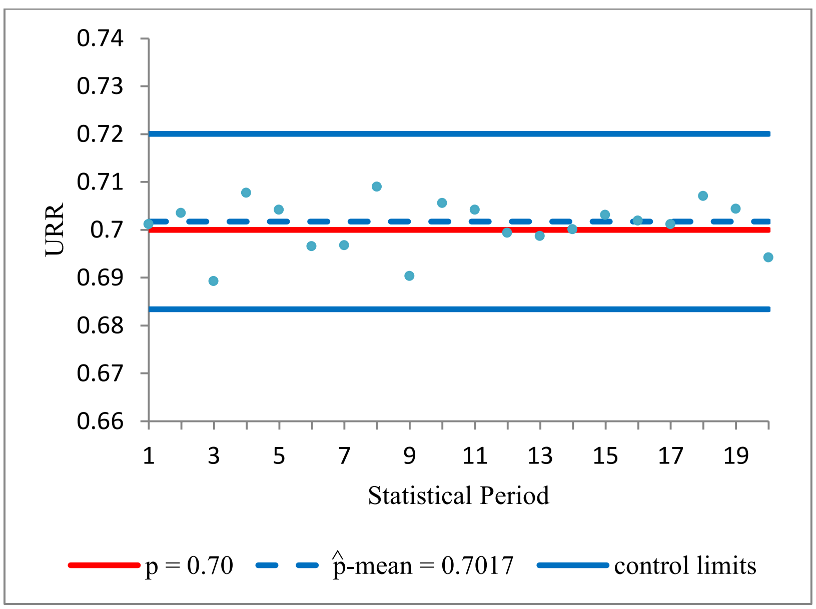

Figure 1 Control Chart Prior to Outlier using $\hat{p}_{mean}$ and interval **[16, 20]** (i.e. T=5).

In Figure 2, there is an outlier in statistical period 21. This outlier can be indicative of a process shift or a one-time outlier. However at this point we do not have sufficient evidence to draw any conclusion. Despite this, our control chart is indicating a potential shift in the process. For our simulation, we knew ahead of time that this outlier was a one-time occurrence; therefore, it is obvious that the mean estimator is too sensitive in such a scenario. Because of the complexities of the response rate process, program managers may not be able to identify the reason for the outlier, therefore we choose to keep all points in the calculation of the rolling average.

With an outlier of 0.80 present in statistical period 21, the computed estimate of the process average increases from the earlier value of 0.7017 (at t = 20) to 0.7214 (at t = 21). This increase in the process average shifts the control limits upwards, and as a result, 20 of 21 data points now fall below the estimated process average, with 14 of the 21 points being outside of the newly revised limits.

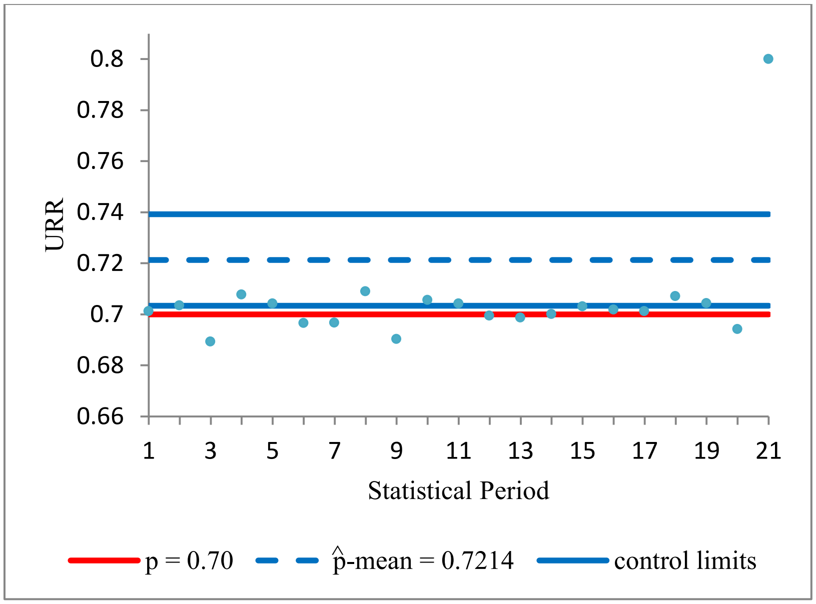

Figure 2 Control Chart with Outlier using $\hat{p}_{mean}$ and interval **[17, 21]**.

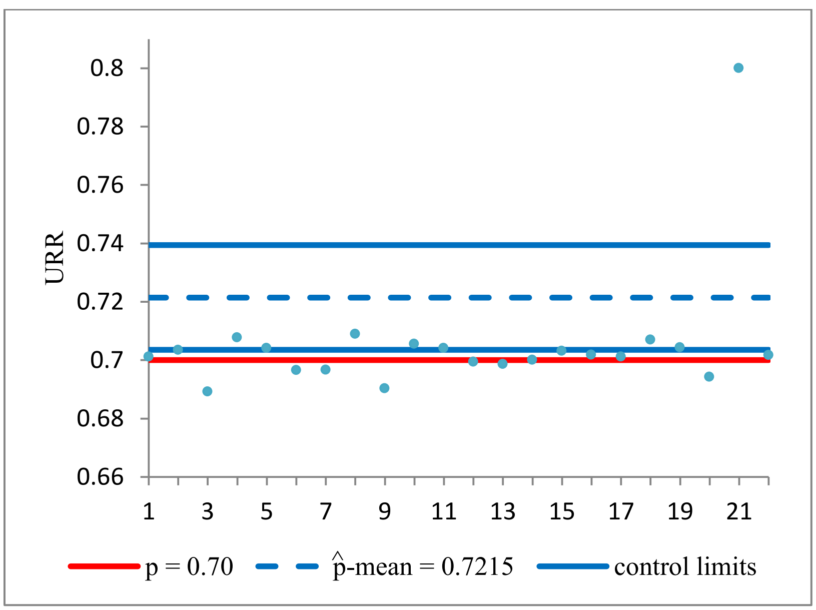

Figure 3 Control Chart with Outlier using $\hat{p}_{mean}$ and interval **[18, 22]**.

In Figure 3, the control limits were estimated using statistical period 18 through 22, which contain the outlying URR value. The effect of the outlier is still present despite the fact the URR value in statistical period 22 is much closer to the values observed before statistical period 21. This effect[4] continues through statistical period 25.

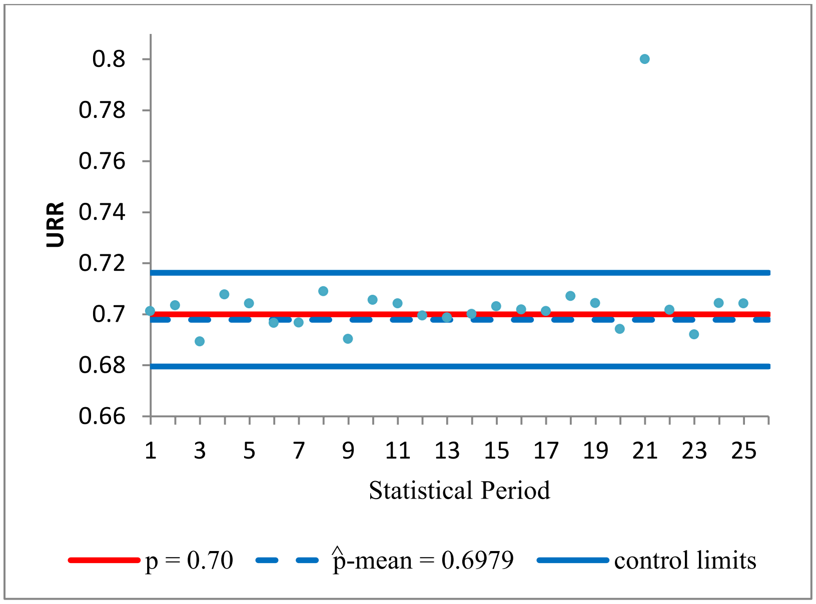

Figure 4 Control Chart with Outlier using $\hat{p}_{mean}$ and interval **[22, 26]**.

Finally, in Figure 4, we see that once the outlier is no longer used in the calculation of $\hat{p}_{mean}$, the control chart once again shows an in-control process for the entire time-period under consideration. Figures 1 through 4 demonstrate the sensitivity of the process average computed using a mean to a large one-time outlier. This is why we investigated the median as an alternative estimator to the mean.

In our graphical analysis of the effect of the mean estimator on the control chart, we concluded that a single outlier does adversely affect the estimates of the control chart parameters. To confirm our visual findings, we conducted a series of general linear hypothesis tests (Searle, 1971, pp. 188-203) to determine if the estimate is significantly different from the true process average of 0.70. We analyzed results for an extreme outlier URR of 0.80 and repeated the analysis using an outlier URR of 0.74.

The results confirmed our visual findings on the sensitivity of the mean process average to a one-time outlier. In general, $\hat{p}_{mean}$ is sensitive to both the magnitude of the outlier and the size of the rolling average interval, T. If the outlier is not extreme, using larger values of T (at least 9) greatly mitigates an outlier's effect on the estimated process average computed with a mean. This comes at the cost of increasing the amount of time needed for the process chart limits to reflect a true process change, such as a permanent shift, as shown in the following section.

In the discussion of the median that follows, we did not perform a series of general linear hypothesis tests. We were unable to find a comparable non-parametric test analog.

[4] With T=5, the estimates of the process average that are affected by the outlier of 0.80 are those estimated in the following time intervals (T=5): [17, 21], [18, 22], [19, 23], [20, 24], and [21, 25]. Figures for the other three affected intervals are not shown, as they are similar to Figures 2 and 3.

Figure 5 plots the same time series of URRs, but constructs the *p*-chart using a median to estimate the process average. The one-time outlier does not overly affect the control chart limits, and the control chart still depicts an in-control URR process through statistical period 20. This is true for every estimate of the process average that includes the single outlier, because the outlier will always be at the tail end of the distribution and will never attain the median value.

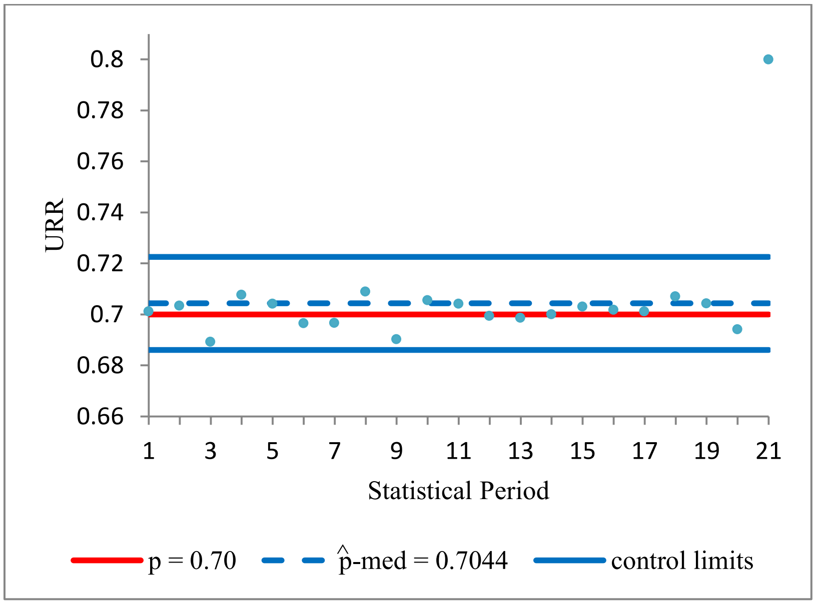

Figure 5 Control Chart with Outlier using $\hat{p}_{med}$ and interval **[17, 21]**.

4.2 Permanent Shift

It is important to ensure that the process limits are not overly sensitive to a rare event, but are adaptive enough to quickly reflect a true process shift after a sufficient "burn-in" period (at least three statistical periods).

In this scenario, the process average shifts from 0.70 to 0.80 in statistical period 21. In practice, neither estimator $\hat{p}_{mean}$ or $\hat{p}_{med}$ will yield an estimate of the new process average immediately. The time it takes to reflect the shift - the burn in period - is a function of the type of estimator used and the number of data points (statistical period) used in the estimator.

This burn-in period for the mean estimator is illustrated in Figure 6 where true process average of 0.70 (through statistical period 20) shifts to 0.80 in statistical period 21. However, our estimate of the process average (blue dashed lines) has a burn-in period of five. At this point, our estimate derives only from URR values from the new response rate process.

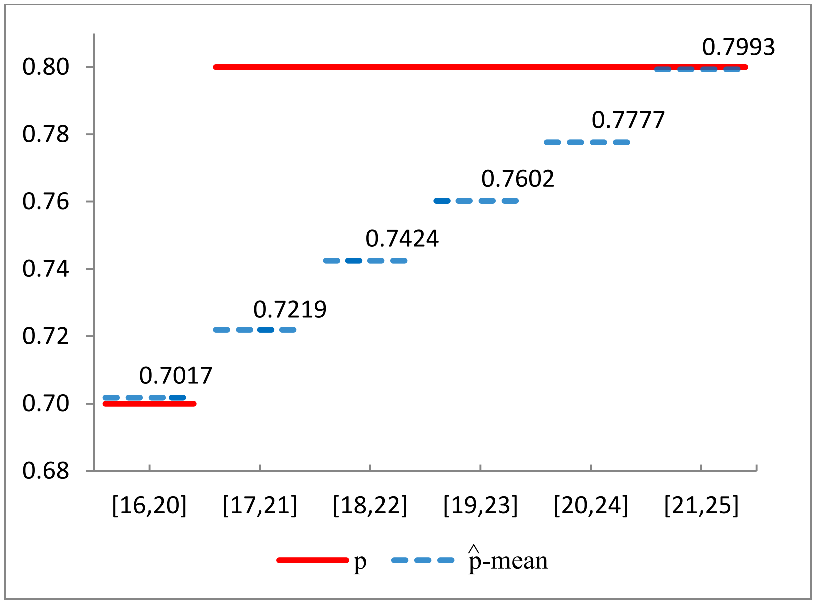

Figure 6 Estimates $\hat{p}_{mean}$ in the Presence of a Permanent Shift for a Rolling Average of Size 5

We once again conducted a series of general linear hypothesis tests to assess how estimates of the process average computed via the mean are affected when the process average shifts from 0.70 to 0.80 (extreme value) and from 0.70 to 0.74 (less-extreme value), beginning in statistical period 21. Our results show that in general it takes the mean approximately $\boldsymbol{T}$ -1statistical periods beyond the initial appearance of the outlier to assume the new process average regardless of the value (0.80 or 0.74).

Figure 7 shows that when using $\hat{p}_{med}$, the burn-in period to reflect the true new process average (p = 0.80) is three statistical periods. In Figure 8, T is an even number, the burn-in period increases to four, and, there is an intermediary value (0.7515) during the burn-in period.

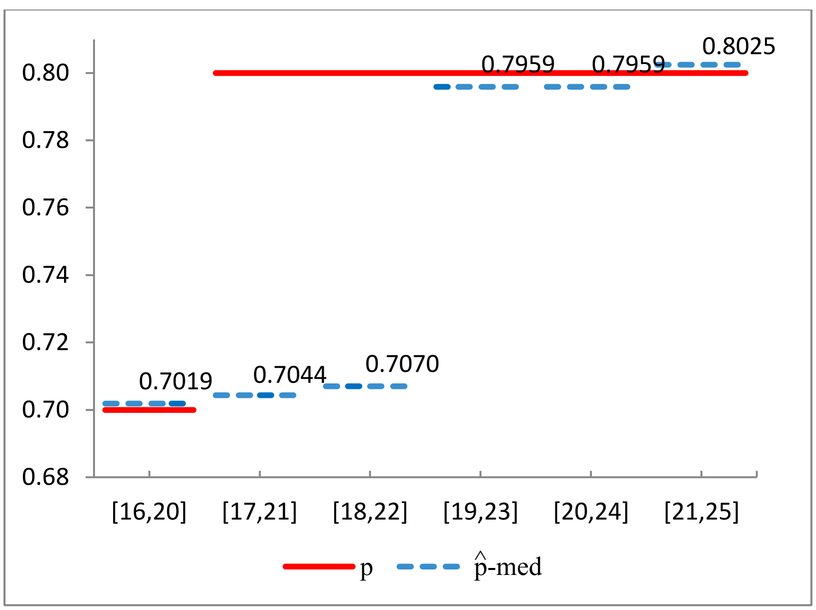

Figure 7 Estimates $\hat{p}_{med}$ in the Presence of a Permanent Shift for a Rolling Average of Size 5

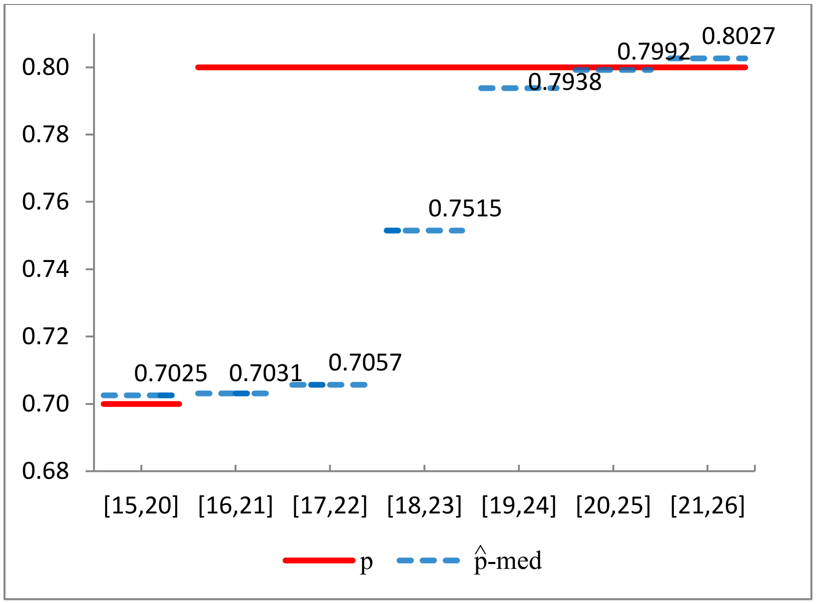

Figure 8 Estimates $\hat{p}_{med}$ in the Presence of a Permanent Shift for a Rolling average of Size 6

In general, when the shift in process average is permanent, the median assumes the new process in ***S*** subsequent statistical periods following the first appearance of the outlier, where ***S*** = **greatest integer function [*T*/2]**. The median achieves the new process average quicker than the mean for both values of *p* and occurs in a "jump" as opposed to changing gradually.

To summarize our findings, when there is a one-time outlier, using the median as process average maintains generally consistent control chart limits. In contrast, the process average obtained with the mean can be dramatically affected by a single outlier, which in turn affects the location and width of the control limits. Using a larger value of *T* reduces the effect of a one-time outlier on the process average estimate. However, using a larger value of *T* increases the number of statistical periods needed for both the mean and the median to achieve a new process average (an increase or a decrease). During this "burn-in" period, the control charts will depict an out-of-control process, which could in fact be stabilized much sooner than shown. In general, the median achieves the new process average quicker than the mean: **[*T*/2]** statistical periods following the first appearance of an outlier (low or high) versus approximately ***T*-1**.

5. Conclusion

The purpose of this study was to develop guidelines for producing control charts to monitor unit response rates produced by business surveys in the Economic Directorate of the U.S. Census Bureau. Our objective was to reframe the evaluation of unit response rates, considering the collective time-series of response rates as a process measure to be monitored instead of a performance measure to be evaluated. By studying the response process, the survey manager can determine whether fluctuations in the process are random and self-correcting, or whether a trend has occurred. More important, the survey manager can determine whether an intervention in the current procedures is necessary and whether the intervention effectively changes the process.

Our review of the literature suggests that a *p*-chart is appropriate for monitoring the unit response rate processes. However, translating the *p*-chart from the manufacturing setting for which it was designed to a survey setting poses several challenges. In a survey setting, there are fewer measurements and these measurements are generated over a longer time interval. Consequently, estimating the process average is more of a challenge particularly in the event of outliers. Our research concluded that these problems can be overcome under the following conditions:

- There are a sufficient number of response rates measures over the time period of interest to produce a reliable estimate of the process average.
- An appropriate measure of central tendency (mean or median) is used to mitigate the effect of outliers.

An assumption underlying the calculation of the response rate measures is that we have sufficient number of reporting units. This can be determined mathematically; for example, using a conservative hypothetical response rate of 0.50 and a tolerance level of 0.02 the number of reporting units needed is 5,625. For a hypothetical response rate value different from 0.50, the number of reporting units needed is fewer.

To determine the minimum number of response rate measures that can produce a reasonable estimate of the true process average we conducted simulation studies with the following scenarios that take into account the effect of potential outliers:

- A response rate process that is stable until it encounters a one time outlier.
- A response rate process that is stable until it encounters an outlier that marks a permanent shift in the process average.

When the response rate process is stable, both the mean and the median produce reliable estimates of the process average for all values of T (the length of the rolling average). However, a one-time outlier has a detrimental impact on the utility of the control limits when the mean is used as a measure of central tendency. Although this effect can be mitigated by a larger value of T, this larger value increases the number of statistical periods needed for both the mean and the median to reflect a shift in the process average (an increase or a decrease).

When the process shift is permanent, the rolling averages computed with the median approximate the new process change to the new level more slowly than those obtained with a mean, regardless of values of T. However, the shift is not gradual as it is with the mean. Using the median to compute the process average therefore provides a brief “burn in” period before shifting the limits, which again is desirable given the small number of point estimates. This "burn in" period is a function of the size of the rolling average. Based on our simulation study results, we recommend using the median to compute the estimate of the process average using T=5 or T=6 (the latter introduces an intermediary "burn in" value).

Our approach to studying response rates within a statistical process control framework is not common, and we were not able to strictly adhere to all of the assumptions that readily work for a manufacturing process. Despite these limitations, our analysis shows that when the number of data points used to estimate the process average is small, the median is capable of producing reliable control limits with a rolling average size of 5 or 6 data points.

Acknowledgment

Thanks to Laura T. Bechtel, Robyn Sirkis, and Katherine J. Thompson for their review and comments.

References

Groves, R.M. and Brick, J.M. (2008). *Practical Tools for Nonresponse Bias Studies*: JPSM Short Course Booklet.

Montgomery, D.C. (2005). *Introduction to Statistical Quality Control*. New York: John Wiley.

President’s Management Council Memorandum. (2006). Office of Management and Budget Standards and Guidelines for Statistical Surveys. Washington, DC.

Searle, S.R. (1971). *Linear Models*. New York: John Wiley.

Sigman, R. (2000). Estimation and Variance Estimation in a Standard Economic Processing System. *Proceedings of the International Conference on Establishment Surveys.*

Thompson, K.J. (2009). Computing Standard Response Measures for Economic Surveys in StEPS (For the Current Surveys). Internal Slide Presentation. Office of Statistical Methods and Research for Economic Programs, U.S. Census Bureau.

Thompson, Katherine J. and Oliver, B. (2010). Response Rates in Business Surveys: Going Beyond the Usual Performance Measures *(to appear)*. Journal of Official Statistics.