Title
A Note on the Measurement of Accuracy for Subnational Demographic Estimates

Permalink
https://escholarship.org/uc/item/4d07j047

Journal
Demography, 37(2)

Authors
Swanson, David A
Tayman, Jeff
Barr, Charles

Publication Date
2000-01-01

Peer reviewed
A Note on the Measurement of Accuracy for Subnational Demographic Estimates

David A. Swanson; Jeff Tayman; Charles F. Barr


Stable URL: http://links.jstor.org/sici?sici=0070-3370%28200005%2937%3A2%3C193%3AANOTMO%3E2.0.CO%3B2-9

*Demography* is currently published by Population Association of America.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/paa.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
A NOTE ON THE MEASUREMENT OF ACCURACY FOR SUBNATIONAL DEMOGRAPHIC ESTIMATES

DAVID A. SWANSON, JEFF TAYMAN, AND CHARLES F. BARR

Mean absolute percentage error (MAPE), the measure most often used for evaluating subnational demographic estimates, is not always valid. We describe guidelines for determining when MAPE is valid. Applying them to case study data, we find that MAPE understates accuracy because it is unduly influenced by outliers. To overcome this problem, we calculate a transformed MAPE (MAPE-T) using a modified Box-Cox method. Because MAPE-T is not in the same scale as the untransformed absolute percentage errors, we provide a procedure for calculating MAPE-R, a measure in the same scale as the original observations. We argue that MAPE-R is a more appropriate summary measure of average absolute percentage error when the guidelines indicate that MAPE is not valid.

The evaluation of measurement accuracy and the debate over the way to present summary measures of accuracy have a long history. As early as the eighteenth century, it was becoming commonplace for astronomers and navigators to take a set of measurements and then estimate the true value of what they were trying to measure through a simple arithmetic mean taken over the set. This practice was criticized by Roger Cotes; in a posthumous work published in 1722, he recommended the use of a weighted mean (Stigler 1986:16). In presenting his argument, Cotes clearly believed that observations should not have equal weights and that those subject to large errors should be assigned smaller weights than those subject to small errors.

Nearly 200 years after Cotes’s suggestion, Snow (1911) also struggled to find the best way to characterize and summarize the accuracy of a method of population estimation that he had developed. He decided that the best course of action in measuring the accuracy of his innovative method was to compare its estimates against census counts and to compute the percentage difference between the two figures. In developing a summary for these percentage errors, Snow decided to use the root mean square error (RMSE). This decision was not necessarily easy, as is evident in his discussion of the pros and cons of the different weighting schemes implicit in the RMSE and the mean absolute percentage error (MAPE).

The solutions offered by Snow in regard to the issues identified by Cotes (and many others) are still the subject of discussion. The National Research Council’s Panel on Small-Area Estimates of Population and Income devoted considerable thought to the characterization and summary presentation of errors (Kitagawa 1980). In regard to accuracy, the Panel defined four criteria that ideally should be met: (1) low average error, (2) low average relative error, (3) few extreme relative errors, and (4) absence of bias for subgroups (Kitagawa 1980:10). The Panel also acknowledged that it is generally not possible to produce a set of estimates that minimizes the four criteria simultaneously. Given the necessity to choose among the criteria, the Panel chose to focus on low average relative error and few extreme errors, with some additional attention to bias (Kitagawa 1980:11).

The path taken by the Panel is still followed by the U.S. Bureau of the Census. That is, MAPE is consistently discussed in evaluations of estimation accuracy. Davis (1994), for example, used seven measures of accuracy in an evaluation of county population estimates made by the Bureau in the 1980s. He discussed only MAPE, however, because “Except for (two), we found all the measures to be highly correlated for most of our tabulations. For this reason most of the discussion below will deal only with the familiar MAPE” (Davis 1994:2).

In addition to its correlation with other summary measures of error, what makes MAPE so desirable that an experienced analyst such as Davis labels it as “familiar”? First, MAPE has several highly desirable conceptual properties including reliability, ease of interpretation, and clarity of presentation (Tayman and Swanson 1996). Second, MAPE possesses a set of highly desirable statistical and mathematical properties: (1) It uses virtually all of the information about the error in its calculation; (2) each set of data has a unique MAPE; and (3) MAPE is the only measure of central tendency that provides the “center of gravity” for the errors.

MAPE also has a major drawback, however. Being an arithmetic mean, it is affected by extreme values, as noted in virtually every introductory statistics textbook in discussions of the relative merits of different measures of central tendency (Hamburg 1985:60; Norusis 1986:96). In the case of MAPE, extreme values occur only at the high end because it

*David A. Swanson, Helsinki School of Economics and Business Administration and Science Applications International Corporation, Las Vegas, NV; Jeff Tayman, San Diego Association of Governments; Charles F. Barr, Department of Economics, University of Nevada Las Vegas. Address correspondence to David A. Swanson, Helsinki School of Economics and Business Administration, International MBA and BBA Program, Lönrotinkatu 5, 50100 Mikkeli, Finland; E-mail: David.Swanson@mi.hkk.fi. An earlier version of this paper was presented at the Population Estimates Methods Conference held by the Population Estimates Branch, U.S. Bureau of the Census, Suitland, MD, June 8, 1999. The authors are grateful for comments and advice provided by Chuck Coleman, Tom Bryan, and Bill Schooling of the Population Estimates Branch, and for comments made by three anonymous reviewers and the editors of Demography.
is based on a right-skewed, asymmetrical distribution of absolute percentage errors (APEs), which is bounded on the left by zero and is unbounded on the right. Therefore MAPE, like any arithmetic average taken over such a set, is susceptible to being pulled upward, and thus to overstating the error represented by “most” of the observations (Emerson and Strenio 1983:61; Smith and Sincich 1988; Tayman and Swanson 1999). Yet in spite of its known shortcomings, MAPE remains the most commonly used summary measure of the accuracy of demographic estimates (Davis 1994; Lin 1999; Martindale 1999; Murdock and Hoque 1995; Shahidullah and Flotow 1999; Sink 1997; Smith and Cody 1999; Tedrow and Swanson 1990).

One might argue that being “pulled upward” is a desirable feature because if large errors are present, they should be represented in a summary measure. We, however, along with others (Kitagawa 1980:85; Morrison 1971), take a different view. Large, outlying errors can and should be examined separately from the central tendency of error. The arithmetic mean is most appropriate as a summary measure of central tendency for data that are distributed symmetrically. When this condition is present, the arithmetic mean provides not only the center of gravity but also the center of probability. Further, when such a symmetrical distribution is not “heavy-tailed,” the arithmetic mean aptly characterizes the bulk of the observations. When data are distributed asymmetrically, however, the arithmetic mean provides only a center of gravity.

The effects of outliers and asymmetrical distributions on the arithmetic mean are not new issues to statisticians (Barnett and Lewis 1984:1). Because of the shortcomings of this mean in characterizing an asymmetrical distribution, statisticians have devised alternatives. The median is one such alternative, as are the geometric mean, the weighted mean, the trimmed mean, and robust estimators such as the Tukey-M statistic (Goodall 1983). Other solutions include transformations of data, which we address in greater detail later in this paper.

Because of the round of estimation evaluations that will be triggered by the release of the 2000 census, we believe it is important not only to demonstrate MAPE’s shortcomings but also to provide a way to resolve them. We argue that in most instances MAPE will overstate the average error of a set of population estimates, and therefore is not valid in a criterion-related sense. As we use the term here, criterion-related validity is defined as the degree of correspondence between a given measure and some phenomenon of interest that is external to the measure itself; the latter is the criterion (Carmine and Zeller 1979:17). As an example of criterion-related validity, consider a 35-inch yardstick. Although reliable, such a yardstick does not produce valid measurements because it systematically overstates length. MAPE, like a 35-inch yardstick, is reliable but not valid: It systematically overstates average error by being pulled upward by outliers.

MAPE should not be the only criterion used in assessing the accuracy of population estimates. Like Kitagawa (1980), Davis (1994), and Smith and Cody (1999), we believe it is important to assess both bias and outliers as well as to use tools that are suitable for the purpose at hand. If it is important to select an estimation method that minimizes large errors, RMSE should be used. When one decides to use MAPE, however, the impact of outliers must be minimized.

Work with a direct bearing on the issue at hand has quantified the extent to which MAPE overstates demographic forecast error (Tayman, Schafer, and Carter 1998; Tayman and Swanson 1999; Tayman, Swanson, and Barr forthcoming). We extend this work by implementing a procedure that can be used to determine whether MAPE is valid for a given set of subnational demographic estimates, and by describing and empirically illustrating a process that leads to a more valid summary measure of accuracy. That is, we develop the capability to apply a summary measure of error that uses most of the information describing error, but neither overstates nor underestimates the error. In addition, this summary measure should be easy to interpret, clear in its presentation, and supportive of statistical analysis. In moving toward our objective, as a case study, we examine the errors in an illustrative set of estimates for the 39 counties of Washington State.

**TRANSFORMATION AND SYMMETRY**

The distribution of APEs may require a transformation to produce an average that reflects more accurately the error represented by most of the observations: that is, a symmetrical distribution. As noted earlier, in a right-skewed distribution—one with outliers in the upper tail—the average is vulnerable to being unduly dominated by these outliers. In determining whether a set of APEs should be transformed, Emerson and Stoto (1983:125) established the following guideline. If the (absolute) ratio of the largest value to the smallest value exceeds 20, a transformation may be useful; if the ratio is less than 2, transformation may not be useful. A ratio between 2 and 20 is indeterminate. We argue that a transformation is successful when the average of this new distribution neither overstates nor underestimates the level of error, yet makes use of all observations and remains intuitively interpretable and clear in its presentation.

To change the shape of a distribution efficiently and objectively, we apply a single nonlinear function. More formally, a transformation of the original observations \( x_1, x_2, \ldots, x_n \) is a nonlinear function \( N \) that replaces each \( x_i \) by a new value \( y_i = N(x_i) \), so the transformed values are \( y_1, y_2, \ldots, y_n \). For right-skewed distributions with only positive values (such as APEs), simple and familiar functions can be used, such as the natural logarithm, square root, or fourth root (Emerson and Strenio 1983:82). Each of these functions can transform the scale of the original observations in a nonlinear fashion while preserving rank order. They can achieve symmetry because they compress larger data values more than smaller values (Emerson and Stoto 1983:123), although

---

1. Most observations in a right-skewed distribution fall toward the lower values, and a relatively few large values or outliers form a tail that slopes to the right. Household and family incomes are probably the best-known examples of this kind of distribution (e.g., Fonseca and Tayman 1989).
a given value that is less than 1 will increase in size. Nonlinear transformations pull outliers in the upper tail more closely to the main body of the data set. Power transformations also preserve the order of the data; they are continuous functions in that points that are close together in the original data remain close together in the transformed distribution (Emerson and Stoto 1983:98).

The logarithmic transformation is directly related to one of the alternative summary measures of central tendency that may be useful when data are right-skewed, namely the geometric mean, which is calculated as $G = e^{\left[\frac{1}{n} \sum \ln(x_i)\right]}$. The geometric mean is easy to calculate. Equally important, it is easy to interpret and clear in its presentation because it is expressed in the original scale of the observations. Unfortunately it also has two drawbacks. First, if any observations are equal to zero, the geometric mean also is equal to zero. This disadvantage is not crippling, however, because a set of estimates containing no error is rare. The second drawback is more pernicious: The logarithmic transformation may not be the most appropriate.

Which nonlinear function is most appropriate? How can we find it? The answer depends on the characteristics of the original data. When we analyze these characteristics, however, it is not easy to determine the most appropriate function by looking at histograms, boxplots, and statistics of the distributional shape. One approach is to evaluate several simple functions and to select the one that leads to the most symmetrical distribution (e.g., to find a function for which the skewness coefficient is less than a preset value or to use a statistical test). This type of trial-and-discovery approach is time-consuming, however, and could involve many more functions than the examples just mentioned. It also is extremely vulnerable to interpretative differences, especially if the process is not standardized. Thus any two persons looking at the same data may reach different conclusions about the best transformation to use.²

Although the choice of transformation function is always partly a matter of judgment, our approach minimizes the degree of judgment required, is standardized, sets a well-defined criterion for determining the optimal function, and is easily implemented in an electronic spreadsheet. We use a modification of the power transformation developed by Box and Cox (1964), defined as $y = (x^\lambda - 1) / \lambda$, for $x \neq 0$; or $y = \ln(x)$, for $x = 0$, where $x$ is the untransformed absolute percentage error, $\bar{y}$ is the transformed absolute percentage error, and $\lambda$ is the power transformation constant.³

Our modified Box-Cox transformation not only compresses very large errors, but also increases errors greater than 1 in skewed distributions where $\lambda$ is relatively small—say, under 0.4. This property is the reason why this transformation may be more effective in achieving a symmetrical result than are simpler nonlinear functions that only increase untransformed errors less than 1. Because many, if not most, estimation errors are greater than 1%, the modified Box-Cox equations not only lower extremely high errors toward the body of the data but also can raise relatively low errors. These characteristics help to minimize skewness and increase symmetry. One determines $\lambda$ by finding its value that maximizes the function:

$$ml(\lambda) = -(n/2) \left[ \ln\left(\frac{1}{n} \sum (y_i - \bar{y})^2\right) + (\lambda - 1) \left( \sum \ln(x_i) \right) \right],$$

where $n$ is the sample size, $y$ is the transformed absolute percentage error, $\bar{y}$ is the mean of the transformed absolute percentage errors, $x$ is the untransformed absolute percentage error, and $\Sigma$ represents the sum over all observations.

According to Box and Cox (1964), $ml(\lambda)$ at a local maximum provides the power transformation ($\lambda$) for $x$ that optimizes the probability that the transformed distribution will be symmetrical. In other words, finding $\lambda$ does not guarantee symmetry, but it represents the transformation power most likely to yield a symmetrical distribution. In practice, the maximum value of $ml(\lambda)$ is found by solving for different values of $\lambda$ in the range from −2 to 2 and finding the largest value in that range (Draper and Smith 1981:225).

We developed a macro for Microsoft Excel (available from any of the authors) that finds the value of $\lambda$ that maximizes the maximum-likelihood function. This macro solves for values of $\lambda$ from −2 to 2 inclusive, using increments of .10. This is a “coarse grid” search, but it is sufficient for our purposes. The only input required is the untransformed distribution of APEs.

Figure 1 shows a graph of the maximum-likelihood value associated with each value of $\lambda$ for the APEs associated with our illustrative set of population estimates. As expected, the relationship between the maximum-likelihood

---

². Emerson and Stoto (1983:105–11) suggest one approach for finding the most appropriate exponent to use in simple nonlinear transformations. Their approach is appealing because it is systematic and contains clearly defined steps and criteria. It is more cumbersome to implement, however, than the approach shown here, and is more restrictive in its functional form.

³. The original Box-Cox formulation specified $y = (x^\lambda - 1) / \lambda$. According to information from a web site hosted by the Mathematics and Statistics Department at the University of Maryland, Baltimore County (http://www.math.umbc.edu/~imt9a/datatran.html), $y = (x^\lambda - \lambda) / \lambda$ is more useful for transforming right-skewed distributions.
value and $\lambda$ is nonlinear and the optimal value of $\lambda$ (0.3) corresponds to the largest maximum-likelihood value, which in this case is the smallest negative value. Figure 2 shows that the smallest skewness coefficient for these APEs also occurs at $\lambda = 0.3$; this validates our modified Box-Cox procedure for determining the most symmetrical transformed distribution of APEs.

**CASE STUDY DATA**

The illustrative data that we use from Washington State represent ex post facto comparisons of 1970 population estimates against 1970 census counts for counties based on the “unmodified” set of population estimates found in Swanson (1980: table 3). Our selection of these data was motivated by five considerations: (1) The estimates reflect an April 1 date; this allows a direct comparison with the decennial census, without any adjustments; (2) with 39 counties, the data set is large enough to illustrate our main points, but small enough to be shown in its entirety; (3) the methodology, the estimates, the census counts, and the errors are available in a single, readily accessible published source; (4) the estimates were produced the estimates by a standard method, the ratio-correlation model; and (5) our collective experience leads us to conclude that the data set contains errors representative of those found in a wide range of subnational demographic estimates. In addition to the criteria named above, these estimates meet the usual standard for acceptable accuracy.

As shown in Table 1, the model tends to be quite accurate. MAPE is 5.07; the highest APE is found for Pend Oreille County (14.16) and the lowest for Asotin County (0.36). Although this degree of accuracy would be judged acceptable, the guideline provided by Emerson and Stoto (1983:125) suggests that it would be useful to transform these APEs because the ratio of the APE for Pend Oreille County to the APE for Asotin County far exceeds 20: This ratio is 39.33 (14.16/0.36). Among other things, this finding suggests that some extreme observations are influencing MAPE unduly and pulling it upward so that it is not valid for this set of data: MAPE should characterize the bulk of the errors.

**ANALYTIC METHODS**

In addition to the Emerson-Stoto guideline, we employ graphic devices and statistics to determine whether a transformation is warranted. For the graphic devices, we use both a histogram and a boxplot to evaluate the symmetry of the untransformed and transformed APEs. The histogram is well known; the boxplot, less so.

A boxplot is a resistant graphic device that helps identify location, spread, skewness, tail length, and outliers (Emerson and Strenio 1983:58–61; Hintze 1996:273–80). Its main component is a box with endpoints representing the middle half of the distribution. A crossbar within the box shows the median, and the tails are represented by a line drawn from each end of the box to the most remote point that is not an outlier (outliers are usually represented by an asterisk). Spread or variability in the data is displayed by the length of the box. The relative position of the median in the box and the length and direction of the tails depict the distributional shape of the observations. A median closer to the lower end of the box with a long upper tail indicates a right-skewed distribution. A median in the middle of the box with lower and upper tails of equal length is characteristic of a symmetrical distribution.

In using statistics to determine whether a transformation is warranted, we first employ the skewness coefficient to quantify the degree of asymmetry (Snedecor and Cochran 1980:78). The skewness coefficient is zero for a symmetrical distribution, positive for a right-skewed distribution, and negative for a left-skewed distribution. Second, we use a statistical test of symmetry: the D’Agostino skewness test (D’Agostino, Belanger, and D’Agostino 1990; Hintze 1996:103).

**RESULTS**

**Transforming the APEs**

As we already know, the Emerson-Stoto procedure suggests that the APEs should be transformed. What do the graphic devices and statistics say? Figure 3 is a histogram of the absolute percentage error distribution taken from the last column in Table 1. It shows that the untransformed APEs are asymmetrical and right-skewed; this finding is confirmed by the boxplot shown in Figure 4. In addition, the skewness coefficient for these data is 0.802, and the D’Agostino skewness test suggests that the null hypothesis (the data are not skewed) should be rejected ($p = 0.03$).

---

4. Although there is no exact mathematical way to identify outliers with perfect certainty, the boxplot defines outlier cutoffs on the basis of the fourth spread or the width of the middle half of the data (Emerson and Strenio 1983:59–60). (Outliers are observations that fall above the upper cutoff or below the lower cutoff.) Outlier cutoff points are found by multiplying the fourth spread by 1.5, adding that result to the third quartile value, and subtracting the resulting sum from the first quartile value.
### Table 1. Percentage Errors for a Set of 1970 Population Estimates for Washington State Counties Made with the Ratio-Correlation Method

<table>
<thead>
<tr>
<th>County</th>
<th>Estimate</th>
<th>Census</th>
<th>Percentage Error</th>
<th>Absolute Percentage Error</th>
<th>Transformed Absolute Percentage Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams</td>
<td>11,387</td>
<td>11,102</td>
<td>2.57</td>
<td>2.57</td>
<td>3.42</td>
</tr>
<tr>
<td>Asotin</td>
<td>11,819</td>
<td>11,862</td>
<td>-0.36</td>
<td>0.36</td>
<td>1.45</td>
</tr>
<tr>
<td>Benton</td>
<td>67,823</td>
<td>63,144</td>
<td>7.41</td>
<td>7.41</td>
<td>5.08</td>
</tr>
<tr>
<td>Chelan</td>
<td>36,373</td>
<td>35,862</td>
<td>1.43</td>
<td>1.43</td>
<td>2.71</td>
</tr>
<tr>
<td>Clallam</td>
<td>31,371</td>
<td>30,023</td>
<td>4.49</td>
<td>4.49</td>
<td>4.23</td>
</tr>
<tr>
<td>Clark</td>
<td>111,312</td>
<td>116,663</td>
<td>-4.59</td>
<td>4.59</td>
<td>4.27</td>
</tr>
<tr>
<td>Columbia</td>
<td>4,222</td>
<td>3,771</td>
<td>11.96</td>
<td>11.96</td>
<td>6.02</td>
</tr>
<tr>
<td>Cowlitz</td>
<td>61,636</td>
<td>62,586</td>
<td>-1.52</td>
<td>1.52</td>
<td>2.78</td>
</tr>
<tr>
<td>Douglas</td>
<td>16,375</td>
<td>15,287</td>
<td>7.12</td>
<td>7.12</td>
<td>5.01</td>
</tr>
<tr>
<td>Ferry</td>
<td>3,408</td>
<td>3,336</td>
<td>2.16</td>
<td>2.16</td>
<td>3.20</td>
</tr>
<tr>
<td>Franklin</td>
<td>24,770</td>
<td>23,983</td>
<td>3.28</td>
<td>3.28</td>
<td>3.76</td>
</tr>
<tr>
<td>Garfield</td>
<td>2,770</td>
<td>2,546</td>
<td>8.80</td>
<td>8.80</td>
<td>5.40</td>
</tr>
<tr>
<td>Grant</td>
<td>42,750</td>
<td>38,921</td>
<td>9.84</td>
<td>9.84</td>
<td>5.62</td>
</tr>
<tr>
<td>Grays Harbor</td>
<td>52,173</td>
<td>52,583</td>
<td>-0.78</td>
<td>0.78</td>
<td>2.09</td>
</tr>
<tr>
<td>Island</td>
<td>22,215</td>
<td>20,589</td>
<td>7.90</td>
<td>7.90</td>
<td>5.20</td>
</tr>
<tr>
<td>Jefferson</td>
<td>9,551</td>
<td>9,235</td>
<td>3.42</td>
<td>3.42</td>
<td>3.82</td>
</tr>
<tr>
<td>King</td>
<td>1,035,704</td>
<td>1,054,271</td>
<td>-1.76</td>
<td>1.76</td>
<td>2.95</td>
</tr>
<tr>
<td>Kitsap</td>
<td>85,989</td>
<td>86,529</td>
<td>-0.62</td>
<td>0.62</td>
<td>1.89</td>
</tr>
<tr>
<td>Kittitas</td>
<td>19,972</td>
<td>22,764</td>
<td>-12.27</td>
<td>12.27</td>
<td>6.07</td>
</tr>
<tr>
<td>Klickitat</td>
<td>11,968</td>
<td>10,729</td>
<td>11.55</td>
<td>11.55</td>
<td>5.94</td>
</tr>
<tr>
<td>Lewis</td>
<td>40,124</td>
<td>39,265</td>
<td>2.19</td>
<td>2.19</td>
<td>3.22</td>
</tr>
<tr>
<td>Lincoln</td>
<td>9,185</td>
<td>8,168</td>
<td>12.45</td>
<td>12.45</td>
<td>6.10</td>
</tr>
<tr>
<td>Mason</td>
<td>17,867</td>
<td>18,411</td>
<td>-2.96</td>
<td>2.96</td>
<td>3.62</td>
</tr>
<tr>
<td>Okanogan</td>
<td>23,656</td>
<td>22,952</td>
<td>3.07</td>
<td>3.07</td>
<td>3.67</td>
</tr>
<tr>
<td>Pacific</td>
<td>12,834</td>
<td>13,310</td>
<td>-3.58</td>
<td>3.58</td>
<td>3.89</td>
</tr>
<tr>
<td>Pend Oreille</td>
<td>5,919</td>
<td>5,185</td>
<td>14.16</td>
<td>14.16</td>
<td>6.38</td>
</tr>
<tr>
<td>Pierce</td>
<td>346,430</td>
<td>339,048</td>
<td>2.18</td>
<td>2.18</td>
<td>3.21</td>
</tr>
<tr>
<td>San Juan</td>
<td>2,947</td>
<td>3,089</td>
<td>-4.60</td>
<td>4.60</td>
<td>4.27</td>
</tr>
<tr>
<td>Skagit</td>
<td>48,868</td>
<td>45,703</td>
<td>6.93</td>
<td>6.93</td>
<td>4.96</td>
</tr>
<tr>
<td>Skamania</td>
<td>5,360</td>
<td>5,330</td>
<td>0.56</td>
<td>0.56</td>
<td>1.80</td>
</tr>
<tr>
<td>Snohomish</td>
<td>231,238</td>
<td>245,193</td>
<td>-5.69</td>
<td>5.69</td>
<td>4.62</td>
</tr>
<tr>
<td>Spokane</td>
<td>256,882</td>
<td>251,057</td>
<td>2.32</td>
<td>2.32</td>
<td>3.29</td>
</tr>
<tr>
<td>Thurston</td>
<td>69,613</td>
<td>68,719</td>
<td>1.30</td>
<td>1.30</td>
<td>2.61</td>
</tr>
<tr>
<td>Wahkiakum</td>
<td>3,409</td>
<td>3,137</td>
<td>8.67</td>
<td>8.67</td>
<td>5.37</td>
</tr>
<tr>
<td>Walla Walla</td>
<td>38,323</td>
<td>36,608</td>
<td>4.69</td>
<td>4.69</td>
<td>4.30</td>
</tr>
<tr>
<td>Whatcom</td>
<td>70,801</td>
<td>72,111</td>
<td>-1.82</td>
<td>1.82</td>
<td>2.99</td>
</tr>
<tr>
<td>Whitman</td>
<td>32,510</td>
<td>34,843</td>
<td>-6.70</td>
<td>6.70</td>
<td>4.90</td>
</tr>
<tr>
<td>Yakima</td>
<td>136,455</td>
<td>128,960</td>
<td>5.81</td>
<td>5.81</td>
<td>4.65</td>
</tr>
</tbody>
</table>

**Notes:** Estimates and census counts are taken from Table 3 in Swanson (1980). Calculated values shown are subject to rounding, and in some cases they vary slightly from the corresponding calculated values shown in Table 3 in Swanson (1980). The calculation of summary statistics was not materially affected, however.

How well did the modified Box-Cox transformation work? The two rightmost columns of Table 1, along with Figures 5 and 6, show clearly that the transformed distribution is significantly more symmetrical than the original distribution of APEs. Moreover, the transformed APEs do not differ significantly from what one would expect under perfect symmetry. The skewness measure for the transformed APEs is -0.069. (Recall that perfect symmetry is indicated by a skewness score of zero.) The D'Agostino test resulted in a statistic of -0.200; this finding, with $p = 0.84$, leads us to accept
the null hypothesis that the transformed APEs are symmetrical. As expected, the modified Box-Cox transformation greatly reduces the largest values and does not change small values much: In the estimated population for Pend Oreille County (5,919) the APE was 14.16, which declined to 6.38 after the transformation. The original APE for Clark County was 4.59, but this decreased slightly, to 4.26, after the transformation.

Finally, using the Emerson-Stoto guideline, we find that the ratio of the highest transformed APE to the lowest transformed APE does not exceed 20. As shown in Table 1, the highest APE, 6.38, is (again) found for Pend Oreille County, while the lowest, 1.45, is (again) found for Asotin County. The ratio of these two transformed APEs is 4.4 (6.38/1.45).

Table 1 also shows that the average of the transformed error distribution (MAPE-T = 4.07) is lower than that of the
original error distribution \((\text{MAPE} = 5.07)\). In addition, the
standard deviation of the transformed APEs (not shown) is
1.3, while that of the untransformed APEs is 3.8. The trans-
formed APEs, in themselves, are not of interest in evaluat-
ing individual county estimates; for that purpose, the original
APEs should be used. For our present purposes, the interest-
ing point is the summary measure that the transformed APEs
produce, namely MAPE-T.

Reexpressing the Average to Match the Original Metric

Even if it is successful, the transformed estimation error dis-
tribution has a potential disadvantage: Transformation may
move the observations into a unit of measurement that is not
easily interpreted (Emerson and Stoto 1983:124). Conse-
quently we may lose some intuitive understanding and for-
feit the ability to make clear, appropriate interpretations re-
garding the magnitude of error in the estimates. This issue is
not trivial: In regard to the unit of measurement, most of us
find it easier to think of estimation error in terms of percent-
ages than (say) log percentages or square root percentages.
When the modified Box-Cox transformation is used, inter-
pretation also can be more difficult because the transforma-
tion is less intuitive than simpler transformations such as
the natural log and the square root.

Tayman, Swanson, and Barr (forthcoming) developed a
procedure for converting MAPE-T into the scale of the
untransformed observations; MAPE-R was the name given to
the reexpressed MAPE-T. To accomplish this reexpression,
we relate each observation from the untransformed error dis-
britution to its counterpart in the transformed error distribu-
tion by a nonlinear statistical equation, using regression to
estimate the coefficients that quantify that relationship. After
we complete this statistical mapping between the untrans-
formed and the transformed scales, we estimate MAPE-R by
solving the regression equation for MAPE-T, the average of
the transformed distribution.

Two nonlinear functions—quadratic and power—may be
useful for statistically mapping the scales of the transformed
and untransformed observations. We chose the power func-
tion because it fit our data best: \(x_i = a \times y_i^b\), where \(x\) is the
original absolute percentage error, \(y\) is the transformed abso-
lute percentage error, and \(a\) and \(b\) are estimated parameters.
The power function can be operationalized as a linear regres-
sion expressed in logarithm form: \(\ln(x) = \ln(a) + b[\ln(y)]\).
The estimated parameters from this equation are then used
in conjunction with MAPE-T to estimate

\[
\text{MAPE-R} = e^{(a + (b \times [\ln(MAPE-T)])}).
\]

Using the regression procedure from the NCSS package, we
found the coefficient of determination for this model to be
0.997; \(a = -2.146\) and \(b = 2.547\). Solving this equation with
MAPE-T yields a MAPE-R of 4.17; that is,

\[
4.17 = e^{(-2.146 + (2.547 \times [\ln(4.07)])}).
\]

For comparison we computed Tukey-M, a robust mea-
sure of location, from the untransformed errors. Both Tukey-
M and MAPE-R are designed to provide measures of the cen-
tral tendency of the error that are not influenced by the asym-
metry and outliers that characterize the untransformed abso-
lute percentage error distribution. We found that for the data
at hand, Tukey-M was equal to 4.21, quite close to the
MAPE-R value of 4.17. This correspondence provides em-
pirical support for our argument that outliers do not undue
fluence MAPE-R. Unlike Tukey-M, however, MAPE-R is
easier to interpret and present. Moreover, it has greater po-
tential for use among applied demographers, supports statis-
tical evaluation more effectively, and utilizes more of the in-
formation about estimation error.

With MAPE-R equal to 4.17 for our illustrative set of esti-
mates, we argue that we have found a measure of central
tendency that is more valid for this set of estimates than
MAPE because MAPE-R is more indicative of the bulk of
the APEs. For example, 19 of the 39 observations are within
2.5 points of MAPE-R, whereas only 12 of the 39 are within
2.5 points of MAPE.

If we were to summarize to a typical audience the perfor-
mance of this model in producing population estimates for
these counties, we would state that the average absolute per-
centage error is 4.17% and that the model tends to overesti-
mate the actual population (26 of 39 counties are overesti-
mated). Only five of the 39 counties (13%), however, have an
APE greater than 10.0; the highest is 14.2%. Using guide-
lines suggested by Goldberg, Rao, and Namboodiri (1964:
101–102) and by Zitter and Cavanaugh (1980:16–19), we
would conclude that this particular model provides an ac-
ceptable level of accuracy for counties in Washington State.

DISCUSSION AND CONCLUSION

MAPE is currently the standard that defines accuracy in
subnational demographic estimation. This position is not un-
derstood because MAPE has important, desirable features
including ease of interpretation, clarity of presentation, sup-
port of statistical evaluation, and the utilization of all the in-
formation about the error. We argue, however, that for
subnational demographic estimates, MAPE is not always
valid for any given set of data because APEs are right-
skewed and often impart an upward bias to MAPE.

Alternatives to the MAPE, such as M-estimators and the
median, provide summary measures that describe more ac-
curately most of the APEs found in a given set of popula-
tion estimates. These measures, unfortunately, have their
own disadvantages: M-estimators lack the intuitive, inter-
pretative qualities of an average, are not well known, and
require a specialized statistical package such as SPSS,
NCSS, or SAS. The median ignores most of the information
in the estimates. As a consequence, the median is influenced
by changes in the centermost observations resulting from

Other alternatives include the geometric mean and loss
functions. Unfortunately the geometric mean is problematic
because the logarithmic transformation does not always yield
a distribution that has optimal symmetry. Loss functions re-
main an intriguing possibility, although, as Bryan (1999) ac-
knowledges, standards for empirical weighting schemes must be developed if they are to be widely accepted.

Using as a point of departure the work of Tayman and Swanson (1999) on summary measures of population forecast accuracy, we pursued the idea of developing an average measure of population estimation accuracy that is not influenced unduly by extreme outlying observations. We tested a method based on the modified Box-Cox transformation; this turns the skewed, asymmetrical distribution of APEs into a symmetrical distribution, from which one can compute an average measure of error that we call MAPE-R.

Every new procedure raises questions that must be answered and issues that must be addressed before it can gain widespread use and acceptance. Our findings, however, suggest that it is possible to develop an average measure of estimation error that is demonstrably valid for the set of errors to which it is applied. These results seem sufficiently encouraging to warrant further research focused on several topics.

First, we suggest examining the Emerson-Stoto and other guidelines that we have presented for determining when MAPE is valid, particularly with an eye toward how they could be streamlined and strengthened.

Second, how should transformations be structured? How much would the results differ if (for example) one performed a single transformation for all counties across the United States versus state-by-state transformations? Further, is it necessary to apply these transformations for specific sizes, growth rates, or other analytical breakdowns, or would a more generalized application yield useful results?

Third, we suggest further examination of both the transformation and the “reexpression” procedure we have presented. Although our approach yielded the desired results, it is a cumbersome way to compute MAPE-R, and implementation requires skills in statistical modeling that may tax many applied demographers. Fortunately the computational formula for the geometric mean suggests that MAPE-R also can be computed in a single mathematical operation. In the case of the geometric mean, it is easy to overlook the fact that both a transformation and a reexpression occur in the same formula; the formula is “seamless” in the eyes of the user. With this point in mind, some informal tests and discussions with a Census Bureau mathematician have convinced us that it is possible to develop a “seamless” process for computing MAPE-R by moving directly from the Box-Cox transformation value of \( \lambda \) to MAPE-R by using an inverse deterministic function.

In conclusion, no single summary measure should be used for all situations. Like the median, the geometric mean, and any other summary measure of central tendency, MAPE-R imparts little, if any, information about the dispersion of estimation error, particularly in regard to outliers. Information on this aspect of the structure of estimation error requires at least a list of extreme errors, measures of skewness, and the use of graphic devices such as histograms and boxplots. When MAPE-R is supplemented with this type of information, and when the guidelines we have presented here show that MAPE is not valid, MAPE-R is a more appropriate summary measure of average absolute percentage error.

REFERENCES


Snow, E.C. 1911. “The Application of the Method of Multiple Cor-


