

Accuracy

STI: PopStats® — Unprecedented Accuracy

Understanding the accuracy of an estimate is fundamental to the use of that estimate. After all, a lack of faith in that estimate makes the use of it moot. STI: PopStats® is consistently rated by Synergos Technologies' (STI) clients as the best demographic data available. Here at STI, we are very proud of the fact that many of our clients have surpassed the status of being clients and have indeed become "fans." While all of this may be very interesting, most researchers are more pragmatic in their interest in the accuracy and demand unbiased proof. The best way to prove the accuracy of the estimate is, of course, to compare it to a known quantity. In our case that known quantity is the United States Census. Now that the April 2010 Census is available, we are able to provide irrefutable evidence that PopStats is the leading demographic product on the market.

STI's analysis was performed for population at five levels of geography: National, State, County, Tract, and Block Group using Census 2000 boundaries as the boundary definition standard. For national through county levels, virtually no data conversion was necessary to convert Census 2010 boundaries to 2000 (Four 2010 counties in Alaska had to be consolidated back into their original 2000 boundaries). However, extensive data conversion did take place for both Census Tract and Block Group boundary data.

Methodology

The methodology to derive an accuracy measure for populations was performed in two steps. First, we converted the Census 2010 Tract and Block Group boundaries into Census 2000 format. Then, we performed the variance calculation known as MAPE (Mean Absolute Percentage Error).

STI took the data for the analysis from commercially release CD's. We pulled the April 2010 estimate from the CD released in July 2010 and we took the Census 2010 data converted from 2010 to 2000 boundaries from the July 2011 CD. Accordingly, PopStats users who received both CD's can verify the results that follow.

Data Conversion

The Census Bureau provides relationship files so that a data user can translate data from 2000 to 2010 boundaries, and vice versa. STI used the relationship files that allowed us to convert the Census 2010 data at the block level to Census 2000 blocks. Once the conversion was



completed, the block data was aggregated to the block group level, the standard level of geography at which most data comparisons is performed. When this conversion is performed, a certain amount of rounding error is introduced. In our case, the original population changed from 308,745,538 to 308,745,517 at the national level or a difference of 21 people. The converted data was only used at the tract and block group levels; STI was able to use the actual figures for national through county levels.

MAPE

MAPE (Mean Absolute Percentage Error) is the standard that most data vendors use to measure the accuracy of their data. The calculation is straightforward: simply subtract the estimate from the actual, take the absolute value of the result, and divide that figure by the actual value. This yields the Absolute Percentage Error. Perform this calculation for every entity in a geographic layer, sum those results and then divide by the number of entities in that geographic level. This final value is the MAPE.

A newer method to judge accuracy was recently introduced and is referred to as MAPE-R or Mean Absolute Percentage Error - Rescaled. This method compensates for extreme outliers. This method accounts for the notion that if the method being judged gets 99.9% of the values accurately, is it "fair" to allow one extreme outlier to cloud the results? STI decided not to use MAPE-R for two reasons. First, the industry as a whole has traditionally used MAPE as the standard to judge all estimates, changing those rules now does not allow for comparisons of results to earlier years. Secondly, STI strongly believes that all entities in a geography should have a equal weight in the result, discounting extreme outliers does the researcher little good if his business happens to be located in those outlying areas. Most notable of which are high growth areas. The areas most likely to be inaccurately estimated by an estimating methodology.

If a researcher wants to compare a MAPE with a MAPE-R value, he must take into account that the MAPE-R method has potentially artificially lowered the methodology's true error rate, and therefore the difference between the two measures may actually be greater.

Our Results

STI results are extraordinary. At every level of geography, the PopStats methodology demonstrated an unparalleled level of accuracy.

National

The first level of concern is national. If the method fails to achieve a high degree of accuracy here, then by definition all lower levels of geography can not be anymore accurate then the level of above it. Our results show:

Geography	Census 2010	PopStats	MAPE
National	308,745,538	308,748,960	0.0011 %

The PopStats estimate differed from the actual national census by only 3,422 people.

State through Block Group

The extraordinary results at the national level followed through at every lower level of

geography:

Geography	MAPE
State	1.01
County	3.40
Tract	11.78
Block Group	18.99

For example, on average, the PopStats estimate varied at the state level by approximately 1.01%. In fact, a detail analysis showed that 33 states varied by less than 1%. Texas, one of the fastest growing states in the country, had differed by only 0.034%. Finally, over 1,750 counties, representing over 215 million people, differed by no more than 2.5% to the actual census.

Vendor Comparisons

Recently, other data vendors have released their results as well. Here at STI, we chose to compare our results to ESRI® which recently published its results on May 21, 2012, via a press release and supporting white paper. ESRI hired one of its partners (as indicated on the partner's web site), Cropper GIS, to analyze its results and 4 other unnamed vendors. As mentioned in the press release, four demographers were assembled (three of whom are part of the management team at Cropper GIS) and were given data obtained by ESRI. **Important: PopStats was NOT included in their analysis.**

Since the other vendors are unknown, the following table only shows our results as compared to those reported by ESRI in their white paper entitled "Vendor Accuracy Study - 2010 Estimates versus Census 2010". According to the press release, all the other vendors that were included in the study scored more poorly than ESRI on an overall basis.

Geography	PopStats	ESRI®
	MAPE	MAPE-R
State	1.01	7.2
County	3.40	19.1
Tract	11.78	47.1
Block Group	18.99	53.9
Total	35.18	127.3

The PopStats estimate consistently showed significantly greater accuracy than ESRI's own population estimate.

ESRI also performed an analysis at the household level. The following table compares those results*:

Geography	PopStats	ESRI®
	MAPE	MAPE-R
State	2.01	5.4
County	4.24	20.7
Tract	9.73	42.4
Block Group	14.39	51.9
Total	30.37	120.4

* Note: Households were ignored in the initial analysis because Census 2010 Household data

was not available to the STI user community in the CD's mentioned earlier. Census 2010 Household data was not released by the Census to the public until September 2011 and was incorporated in the October 2011 release of PopStats, in the 2010 boundary format.

PopStats Takes the Lead in Accuracy

Therefore in overall accuracy PopStats proved to be 3.77 times more accurate than ESRI. ESRI's overall MAPE-R aggregated score of 247.7 (127.3+120.4) versus PopStats' MAPE score of 65.55. Therefore, PopStats takes the lead in accuracy! If you have questions of how these results were achieved, we encourage you to attend our annual research conference where we discuss in detail the methodologies that stand behind PopStats.

Visit Our Web Site:

www.popstats.com