Understanding Medians

How a Simple Statistical Measure Responds to Distributional Change

Understanding variables that are reported as distributions—like income or age—can be challenging without summary measures like medians and averages. These statistics are commonly used to condense continuous economic measures (such as income, home value, and net worth) or demographic characteristics (such as age) into a single number. It's easier to determine whether a neighborhood is middle income by looking at the middle of the distribution—the median—than by assessing the proportion of households in nine different income intervals.

Medians represent the midpoint value of a distribution. Half of all data points fall below the median; the other half lie above.[1] The value of the median lies in its ability to be used to summarize a lot of information about any distribution into one simple measure. This also enables comparison across different geographic areas or populations. As a positional measure, the median's primary strength is its immunity from the influence of extreme values in the tails of the distribution.

The average is another statistical measure used to summarize a distribution. Its key advantage is that it captures the entire distribution, providing more stability than a median, but the average can be heavily skewed by a top- or bottom-heavy distribution.

The experienced analyst will refer to both the median and the average to characterize the underlying distribution. Their strengths are complementary in areas with relatively normal distributions. But small areas with small population bases can reveal the weaknesses in these summary measures. The simplicity of a positional measure like the median can also be its weakness. When analyzing demographic profiles in granular geographic areas such as census block groups, unexpected changes in positional measures like medians can occur. While such change is normal in areas that are undergoing rapid population growth or decline, this same outcome can be found in areas that have not experienced much demographic churn.

Shifts in the median commonly occur in areas with sparse base data (e.g., few households, small populations) and/or areas with bimodal distributions that represent two or more distinct population groups. While base totals can remain relatively unchanged, small changes *within* a distribution can produce large shifts in a median.

Consider the following example, which shows the impact of minor changes in the household income distribution of a block group (BG) in Hawaii:

[1] It is important to mention that Esri updates one- and five-year base distributions—not medians (or averages, for that matter). Both measures are computed directly from the updated distributions.

Hawaii Block Group				
Income Distribution	**Base**		**Update**	
<$15,000	5	7.5%	6	8.7%
$15,000–24,999	5	7.5%	6	8.7%
$25,000–$34,999	8	11.9%	9	13.0%
$35,000–$49,999	10	14.9%	12	17.4%
$50,000–$74,999	8	11.9%	7	10.1%
$75,000–$99,999	7	11.8%	6	8.7%
$100,000–$149,999	13	19.4%	12	17.4%
$150,000–$199,999	6	9.0%	6	8.7%
$200,000+	5	7.5%	5	7.2%
Total Households	67		69	
Median HH Income	$65,400		$54,200	
Average HH Income	$93,200		$88,800	

The household base is smaller than average. The typical number of households at the block group level exceeds 500. But this particular BG is also bimodal. The residential population falls primarily within two diverse groups in the low- and high-income ranges. More than 40 percent of the households earn less than $50,000; about 36 percent earn $100,000 or more. The middle of the distribution, $50,000 to $75,000, represents less than 12 percent of all households.

Now, assume a few minor changes in this area: the block group occupancy remains fairly stable with 69 households. The changes in the distribution of household income are also slight, with fewer upper-income households and a few more lower-income households—not exactly a compositional change in the distribution. But a few subtle changes decrease the median by 17 percent. The average, on the other hand, shows a decrease of less than 5 percent and still reflects the influence of households at the top end of the income distribution.

The next example demonstrates the effect of change in the age distribution of a small suburban neighborhood of single-family homes in Nashville, Tennessee.

Tennessee Block Group								
Age Distribution (years)	Previous Year				Current Year			
	Male	Female	Total	Total (%)	Male	Female	Total	Total (%)
0–4	5	5	10	18.5%	5	4	9	16.1%
5–9	1	2	3	5.6%	1	2	3	5.4%
10–14	7	0	7	13.0%	7	0	7	12.5%
15–19	2	2	4	7.4%	2	2	4	7.1%
20–24	0	0	0	0.0%	0	0	0	0.0%
25–29	0	3	3	5.6%	0	4	4	7.1%
30–34	0	1	1	1.9%	0	1	1	1.8%
35–39	5	6	11	20.4%	5	7	12	21.4%
40–44	1	0	1	1.9%	1	0	1	1.8%
45–49	5	3	8	14.8%	4	3	7	12.5%
50–54	0	1	1	1.9%	0	1	1	1.8%
55–59	1	1	2	3.7%	1	1	2	3.6%
60–64	0	1	1	1.9%	0	1	1	1.8%
65–69	0	1	1	1.9%	0	2	2	3.6%
70–74	1	0	1	1.9%	2	0	2	3.6%
75–79	0	0	0	0.0%	0	0	0	0.0%
80–84	0	0	0	0.0%	0	0	0	0.0%
85+	0	0	0	0.0%	0	0	0	0.0%
Total Population	28	26	54		28	28	56	
Median Age	17.5	35.0	30.0		17.5	35.7	35.0	

The population base is much smaller than average (about 1,500 persons per BG) and 3 years younger than the US, with a median age of 35 years. Most of the households are married-couple families with children and clearly skewed toward male children. As a result, the median age of all males is half that of females in this neighborhood.

Demographically, extreme change in median age is uncommon. Although the area's total population remains essentially stable, its distinct distribution by age and sex illustrates how the median age can increase from 30.0 to 35.0 years within a single year. This increase is effected by small changes in the older end of the distribution—specifically, an increase of one person in both the 65–69 and 70–74 age groups over this brief time period.

These examples are from very different residential markets. Aside from their stable base populations, they have one thing in common: skewed population distributions that illustrate how this can impact a positional measure like the median. The takeaway here is that when changes in summary measures look askew, peeking under the hood at the distribution can reveal the underlying cause.