Analysis Summarization Statistical Formulas

Formulas defined below are used in Business Analyst for summarizing values during many of the analyses, spatial overlay, and find similar and during report generation.

Consider BlockGroup (Bg) as a feature from a polygonal data layer. Instead of BlockGroup features from the Census Tract layer, ZIP Code layer, County layer, or State layer can be used.

Consider geometry "A" as a boundary geometry. For example, it could be a boundary feature from analysis results (i.e., drivetime, simple ring, other trade area).

Below is a discussion about computing the value for one field. The same field is used in "Bg" and (possibly) one additional weight field in "Bg" (only if WeightField is defined in metadata).

Terminology definition:

1. Value(Bg) is the value of the field in the Block group "Bg".
2. Value(A) is the summarized value of the corresponding field for specified geometry "A". This is an output value, and it will be computed using the formulas below.
3. Ratio(Bg,A) is the weight of the part of block group "Bg" that is inside geometry "A". The weight variable is determined by the user's choice during Analysis Layer Setup and is stored in the layer's metadata.

   For arbitrary geometry "C" ("C" can represent either "Bg" or "A") the following values will be determined:

   a. POP(C) is the number of population in geometry "C".
   b. HHD(C) is the number of households in geometry "C".
   c. HU(C) is the number of housing units in geometry "C".

   So Ratio of (Bg,A) could be:

- $Ratio(Bg, A) = \frac{Area(Bg \bigcap A)}{Area(Bg)}$
- $Ratio(Bg, A) = \frac{POP(Bg \bigcap A)}{POP(Bg)}$
- $Ratio(Bg, A) = \frac{HHD(Bg \bigcap A)}{HHD(Bg)}$
- $Ratio(Bg, A) = \frac{HU(Bg \bigcap A)}{HU(Bg)}$

- $Ratio(Bg, A) = 1$ if Bg is fully inside A

4. WeightFieldValue(Bg) is a value of the additional weight field of block group "Bg".
5. Weight(Bg,A) is a weight of the block group as the degree of influence of the value of this particular block group on the value calculated for the geometry A. It is not the same as the value of weight field for the block group. It depends on the basic layer's metadata and may be different for various fields.

   It could be equal to:
   - $Weight(Bg, A) = WeightFieldValue(Bg) * Ratio(Bg, A)$ (if weight field is used)
   - $Weight(Bg, A) = Ratio(Bg, A)$ (if weight field isn't used)

So the term Weight just takes into account both Ratio and value of weight field.

Now determine how to calculate Value(A). There are several ways to calculate it: summarization, summarization with weight, average, and average with weight.

Sum:

$$Value(A) = \sum_{\forall Bg, Bg \bigcap A \neq 0} Ratio(Bg, A) * Value(Bg)$$

Weighted sum:

$$Value(A) = \sum_{\forall Bg, Bg \bigcap A \neq 0} Ratio(Bg, A) * Value(Bg) * Weight(Bg, A)$$

Average:

$$Value(A) = \frac{\sum_{\forall Bg, Bg \bigcap A \neq 0} Ratio(Bg, A) * Value(Bg)}{\sum_{\forall Bg, Bg \bigcap A \neq 0} 1}$$

Average with weight:

$$Value(A) = \frac{\sum_{\forall Bg, Bg \bigcap A \neq 0} Ratio(Bg, A) * Value(Bg) * Weight(Bg, A)}{\sum_{\forall Bg, Bg \bigcap A \neq 0} Weight(Bg, A)}$$

The median is the number in the middle of a set of numbers; that is, half the numbers have values that are greater than the median, and half have values that are less.

Variance and standard deviation.

X is an equiprobable sample.
x is an element of X
n is a count of v in X,
$n = \sum_{x \in X} 1$

Variance:

$$Value(A) = \frac{n\sum_{x \in X} x^2 - \left(\sum_{x \in X} x\right)^2}{n(n-1)}$$

Standard deviation:

$$Value(A) = \sqrt{\frac{n\sum_{x \in X} x^2 - \left(\sum_{x \in X} x\right)^2}{n(n-1)}}$$

Density

Density is the most difficult type of field processing in Business Analyst. For each demography the demography area—for example, in square miles—needs to be calculated. To achieve precision it is required to create a spatial reference for each demography and then project it. Then, there are a number of methods that can be used to calculate the final value of Density depending on the case:

1. $$Density(A) = \sum_{Bg \cap A} \frac{\left(Ratio(Bg, A) * Value(Bg)\right)}{Area(Bg \cap A)} \approx \sum_{Bg \cap A} \frac{\left(\frac{Area(Bg \cap A)}{Area(Bg)} * Value(Bg)\right)}{Area(Bg \cap A)} \Leftrightarrow$$
   $$\Leftrightarrow \sum_{Bg \cap A} \frac{Value(Bg)}{Area(Bg)}$$

2. $$Density(A) = \frac{\sum_{Bg \cap A} \frac{\left(Ratio(Bg, A) * Value(Bg)\right)}{Area(Bg \cap A)}}{\sum_{Bg \cap A} Area(Bg \cap A)} \approx (see 1) \frac{\sum_{Bg \cap A} \frac{Value(Bg)}{Area(Bg)}}{\sum_{Bg \cap A} Area(Bg \cap A)}$$

3. $$Density(A) = \frac{\sum_{Bg \cap A} \left(Ratio(Bg, A) * Value(Bg)\right)}{\sum_{Bg \cap A} Area(Bg \cap A)}$$