Share this article:

Weight, Weight... Please Tell Me! (April 2013)

Weighting is a technique in survey research where the tabulation of results becomes more than a simple counting process. It can involve re-balancing the data in order to more accurately reflect the population and/or include a multiplier which projects the results to a larger universe. A demographic profile (based on known data such as a census age distribution) is often used as a target so that subsequent data analysis can match that profile. Multivariate weighting is a complex, iterative process that is used to achieve results which are as useful as possible in reflecting real-world results. The first step is to determine the goal of weighting.

Balancing Weights

As an example, if your research is based on a random national sample, it may be desirable to compare the demographic profile of your sample to the population profile in order to see how close your results match the overall population. For simplicity, we will restrict this analysis to a single variable: age.

Age Group:	Population Data	Sample (survey results)
18-34	24%	28%
35-44	36%	38%
45-54	28%	25%
55+	12%	9%

Based on the above table, the sample age profile turned out to be a little younger than the overall population. In other words, the sample results are biased towards younger respondents. This may have resulted from an issue with the sampling plan or execution, higher non-response from the older age groups, random variability or other factors. Sometimes, the need to weight may stem from an intentional oversampling of certain segments of the sample frame.

Whatever the reason, adjusting for this is relatively straightforward by calculating a multiplier, or weight, for each of the age groups. This process is called sample balancing, or sometimes "raking" the data. The formula to calculate the weights is W = T / A, where "T" represents the "Target" proportion, "A" represents the "Actual" sample proportions and "W" is the "Weight" value. The weights can be easily calculated using a spreadsheet or with a calculator.

Age Group:	Target (T)		Actual (A)		Weight (W)
18-34	24%	/	28%	=	0.857
35-44	36%	/	38%	=	0.947
45-54	28%	/	25%	=	1.120
55+	12%	/	9%	=	1.333

Before applying weights, each respondent in the sample is counted equally as one person. The weight values are used in lieu of this, so each respondent in the 18-34 or 35-44 age groups will count as slightly less than one individual, and those in the 45-54 or 55+ categories will count as more than one person. The weights will be applied to any subsequent tabulation or other analyses to adjust the sample to compensate for differences between the sample and the target profile.

Projection Weights

Sometimes we not only want to balance our sample data, but also desire to project the results to a larger population. Let's say our goal again was to reflect a national sample representative of the population age. We may want to know, for example, not just what percentage of people might be likely to buy our product, but also, *how many* of the roughly 315 million people in the population this represents. Again, a simple calculation can be used to determine this. Instead of using proportions, however, we will use counts for both the target population and the actual sample. The same formula is used (W = T / A)

Age Group:	Target (T)		Actual (A)		Weight (W)
18-34	75,709,877	/	280	=	270,392.4
35-44	113,564,816	/	380	=	298,854.8
45-54	88,328,190	/	250	=	353,312.8
55+	37,854,939	/	90	=	420,610.4
Total	315,457,821		1000		

Multivariate Weighting

In the above examples, a single weighting variable (age) was used for balancing. This worked well as an illustration, but in the real world, typically two or more variables are used in weighting plans. Adding one additional variable, such as gender, can still be handled similar to the above calculations using a spreadsheet. In this case, typically each unique combination of variables would be isolated to compute the final weights. Instead of 4 age groups, there would be 8 groups, each age by each of two genders.

Adding a third or fourth balancing variable makes this increasingly difficult and also leads to a situation where cells become too small for efficient calculations. Instead, a process called "rim weighting" is used. Targets are determined for each variable independently, and then a computerized program is used to iteratively create a single weight for each respondent which best fits the target variables.

Other Issues

One of the potential pitfalls of weighting is that it can result in extreme changes to the actual data. In the first example above, individuals in the older age groups were counted as slightly more than one person. Imagine the impact if those in a select group were counted as ten individuals each and others counted as only a small fraction. Care must be taken to watch the minimum and maximum balancing weights and the ratio between them to avoid creating a situation where data integrity suffers. Weighting is most successful as a technique where slight adjustments are made; it should not be used to attempt to salvage a poor sample design.

Special handling of significance testing and other statistical analyses may also be required when using weighted data. Always bear in mind that weighted data is like dealing with an altered sense of reality. A single individual no longer represents a single data entity.

The primary goal of weighting and sample balancing is to improve the quality and analytic strength of survey data after it has been collected. Results which better align with a broader population can be achieved, but must be done employing proper technique as well as caution to avoid unintended impacts.