

A dispersion model to describe market research panels

Few panel builders will release detailed source information illuminating the online portals from which they draw their respondent pool. Clients might seek such information as some means of measuring panel bias. Knowing the sources from which panel respondents are drawn, the predominance of a single source, future stability, and continuity are all important to those who bet their tracking studies on a panel provider. And yet, the information is likely to be impossible to obtain due to proprietary turf protection, and to be honest, once in hand might be very difficult to interpret. How then do we compare between panels and temporally within a single panel? We need some metrics.

One such metric proposed at the IIR in April of 2008 was a "panel diversity index" based on such indices employed in population ecology. The logic would be this, the greater the number of sources and the more equally they share in the burden of contributing respondents to a panel, the less each source would contribute to the bias inherent in the pool. Similarly if the number of sources were evenly apportioned, the loss or substitution of a few would have diminishing impact. If a panel had 100 respondents drawn from ten sources 91/1/1/1/1/1/1/1/1 and the first source were to be discontinued the constituency from which the panel had been sourced has a higher likelihood of suffering change, than if a second panel 10/10/10/10/10/10/10/10/10/10 were to lose its first source.

Research practitioners need to know something about panel sourcing. Panels that are diversely drawn and apportioned equally would provide more stable platforms for tracking studies. In addition, if there is any chance of achieving a representative sample it is not size alone that will get us there, diversity of sources is a true contributor.

Alas, completing such calculations has inherent problems. Panel managers would not have to provide the holy grail of proprietary information -- the names of sources -- but would have to release the number of sources each with the quantity of panelists drawn from that source to create the panel.

Here are two diversity indices that might be discussed:

$A = S - 1 / \log N$ where A is the first index, S is the number of discrete sources and N is the number of respondents in the panel.

$$B = S / \sqrt{N}$$

One index of dominance might be:

$$D = \sum (n/N)^2$$

Where n is the number of individuals from a discrete source.

N is the total number in the panel.

The benefit of calculating these indices is that at a given time we can obtain a scaled measure of how broad particular panels are based and thus venture assumptions about future predictability and stability given the threat of changes in sourcing.

Of course there are other and possibly better dispersion models to be found. Those presented are simple and easily understood. They get the ball rolling. I welcome others that come to mind.

One problem is how to discriminate between sources. It is a challenge to come up with language that draws the lines clearly. It is worth remembering that we are seeking portals that are likely to draw from different audiences thus broadening our scope in panel creation. No clearer I am afraid, but I thought I would throw it out there.

Imagine if we not only had the number of respondents drawn from each portal, but also the number of completes. Perhaps, one source is more often used than others. Our model would provide a different measure, equally as useful.

Steven H. Gittelman, Ph.D. - President CEO