

Risky business

The risk of bias in data collection is a threat for all market research. New software has been developed which is targeted at risks and biases within web data, but it can be used on any dataset.

The software called "Risk-E" was developed by CMR Group Ltd in association with Computable Functions Ltd. In the UK. It provides a quick and easy way to find out which variables have an effect on "key variables" within your dataset (go to www.risk-e.net for details).

The idea is that there are certain results, contained in key variables, which are important in business terms. The Risk-E software looks at which other variables exert an influence over these "key" variables. Tony Dent, Managing Director of the CMR Group explained: "The reason I thought the whole thing up was because of concern about web-based data collection. Regardless of whether the sample was 'representative' in terms of being 'balanced' by age, region, income etc, it might well be biased by 'speed of response' e.g. 50% responded within 2 hours. Or perhaps 'professional responders' - 50% having done 'every' invite they've been given." Risk-E can assess which variables "drive" other variables within your survey, variables which you may not have balanced for in your original sample, but have the potential to distort the results of your survey.

Re - sampling

Risk-E uses the concept of re-sampling, as Iain MacKay of Compactable Functions explained: "The key benefit of re-sampling is that it does not rely on assumptions about underlying data distributions or oblige a non-professional user to make decisions about applicability, e.g. paired sample versus independent sample." Re-sampling as the name implies works by repeatedly sampling the dataset (this is sampling "with replacement") and calculating statistics on each sample. The slight variations between the values of the calculated statistic for each sample provide the measure of the relationship between the "key" variables and predictors.

An important feature of Risk-E is that it is a vertical software package that does a couple of things well and makes the use of these statistical measurements accessible. Although you can perform these techniques in mainstream statistical software, they would probably be more difficult to apply, and as we all know, technical accessibility is as important as the statistic itself; if it's difficult to use it will be seldom used.

Concern and debate over web survey data quality is building to a peak. The biases that are inevitable with any data collection method (CATI was not immune) need to be quantified and monitored. The tools to do this need to be accessible, understandable to all users and reliable. By reliable I mean that it should be hard to misapply them or get false results. Risk-E seems to be easy to use, and the re-sampling technique is robust enough as it makes no assumptions about the data, unlike parametric tests for instance.

The re-sampling approach also seems more accessible when used in a teaching context. It appears (see: <http://www.statistics.com/resources/teachers/i-1reslt.php> that when used to teach students statistics "...Students handled more problems correctly, and liked statistics much better, with re-sampling". This is important for the future of market research. I frequently hear the cry of "the client made me do it" when there is a discussion on why some web surveys are badly designed. Being able to educate the practitioner and consuming populations about statistics would probably help in raising data quality, and with that web data quality.

Backend analysis

Being able to quickly parse out biases in web surveys with a tool is very useful. I used Risk-E to do this, and I quickly gained new insights into the data set I was working with. I could have done it with a mammoth statistical package, but I probably would not have run the same type of analyses and I would not have got it done so quickly. A lot of the emphasis on improving web data quality is based around front end checks, checking for respondent attention and response validity. But backend analysis can be just as powerful, and Risk-E can serve a very useful function in this regard. Maybe not so risky a business.... ■