Social Survey Division on NP3APP001

Created By: [redacted] on 14/06/2013 at 15:55

Title: Response Model for Social Data Collection Process from Scatter to Harvest-Proposal

Categorisation
Non Response / Response Rates\SVS Papers

As we discussed, I'm interested as my Summer project! to investigate the metrics we currently use, have or potentially have and don't use to monitor our measures of effectiveness on the LFS (or this survey to begin with).

The overall goal (perhaps ambitiously) would be to produce a mathematical model of collection metrics to predict measures of effectiveness. This is the sort of thing I did in the MoD over a decade ago now albeit looking at invasions of Kuwait rather than LFS collection. But the principle and the complexity of the problem is the same. Namely, you have some overall goal defined by your Measures of Effectiveness and then various variables that feed into the equations to predict the MoE. This could also involve a statistical simulation accompanied by a Visual Interactive component (another past expertise of mine).

Currently, our main MoE is our response rate. Are there any others we use currently? But metrics I would like to explore are for example:

Allocation per IA and stint area across the 4,4,5 quarterly periods. Number of eligible addresses per IA and stint area.
Absolute response by IA by stint area by TO by wave and by Face to face by wave including reissues.
Refusals and non-contacts by IA by stint area by TO by wave and by Face to face by wave including reissues.

Then the first basic model would be one where the absolute responses, refusals and non-contacts by IA and stint area for a given 4,4,5 period are weighted and aggregated to give an overall response and non-response rate. This would allow us to investigate absolute response versus response rate. How many cases actually contribute to a 1% fall or rise in response. Following this, absolute response, refusals and non-contacts could be modelled by:

Absolute response (face to face, wave 1,IA, stint area)= a function of calling pattern, calling days, interviewer experience (days in post), interviewer motivation, interview length, subject matter, interviewer

characteristics, persuasion methods employed minus a function of refusals (too busy, other) and non-contacts (always out, hard to reach) which may be a function of dwelling type etc.

And the same for the TO.

Obviously, this is only a brief idea of my thinking and I need to spend time in a dark room giving it more thought. But if this is OK can I:
a) be linked to all the available data or to someone who can provide me with it ([redacted]).
b) linked to other work in this area I can pull into the model eg I've heard [redacted] has done some work on interview time and response in the TO.
c) suggest who I should speak to [redacted]
d) make sure this fits with other work [redacted] or others in Titchfield may be doing.

This approach also compliments [redacted] LSS work as the MoEs could be considered Critical to Quality (CTQs) metrics. I can't promise anything but it could also lead to use understanding the important sub-metrics we need to monitor. I've done a web search and identified some interesting papers but none as yet that replicate this proposal. But any links to work that may be useful would also be appreciated.


LFSResponseRate.xlsx

Modelling the Process Leading to Cooperation or Refusal using Interviewer Call Record Data

[redacted]

[1]Southampton Statistical Sciences Research Institute (S3RI), University of Southampton, UK (j.darrigo@southampton.ac.uk)

[2]Centre for Multilevel Modelling, University of Bristol, UK

In recent years many surveys have seen a decline in response rates (de Heer, 1999). Survey agencies make great efforts to increase response rates and, at the same time, reduce the costs of the survey data collection process. Effective interviewer calling behaviours are critical in achieving contact and subsequent cooperation. Recent developments in the survey data collection process have led to the collection of so-called field process or paradata, which greatly extend the basic information on interviewer calls. This paper focuses on the process leading to cooperation or refusal, not just the final outcome per sampling unit. It jointly models the different types of outcomes at each call. The key research questions are:

1. Does the time of the day and the day of the week of the call play a role in gaining cooperation?
2. To what extent does cooperation depend on doorstep interviewer-householder interactions? To what extent does cooperation depend on doorstep negative/positive comments from householder?
3. Are different interviewer attributes important for the participation of householders with different socio-demographic attributes?

We use data from the UK Census Link Study which provides a unique opportunity to analyse the effectiveness of interviewer calls to establish contact and to gain cooperation, conditioning on individual, household and interviewer characteristics, in several face-to-face household surveys.

The data include process data, such as records of calls, interviewer observations about the household and information about the interviewer-household interaction, which is linked to census information on individuals, households and areas as well as to rich information about the interviewer. The data have a multilevel structure with individuals nested within households, which are nested within a cross-classification of interviewers and areas. This paper develops a multilevel multinomial logistic regression model based on interviewer call record data to predict the likelihood of interview or refusal at each call, conditioning on contact made, allowing for the hierarchical structure of the data. The project is part of a 3-year research programme funded by the UK Economic and Social Research Council.

Key Words: paradata, interviewer call data, multilevel multinomial logistic regression.