Sawtooth Software

RESEARCH PAPER SERIES

Consistency Cutoffs to Identify "Bad" Respondents in CBC, ACBC, and MaxDiff

Bryan Orme
Sawtooth Software, Inc.

Consistency Cutoffs to Identify "Bad" Respondents in CBC, ACBC, and MaxDiff

Bryan Orme, Sawtooth Software

December 2019

Introduction

Data quality is a big concern in survey research. Fortunately, conjoint analysis and MaxDiff (best-worst scaling) utilities (scores) include a fit statistic to help identify whether respondents seem to be consistent or seem to be acting randomly.

Over the last few years, the number of respondents (especially from panel sample) who are speeding, not paying attention, or who are bots is increasing. The game of cat-and-mouse continues between the cheaters and the panel providers along with the researchers who use their services. Fortunately, panel companies are getting better at screening out bots and cheaters. And, you as the researcher also have statistical tools at your disposal. This article gives you a straightforward approach to identify the vast majority of random responders (human or robotic) who still find their way into your CBC and MaxDiff surveys.

Cutoffs to Identify Random Responders

HB estimation for CBC, ACBC, and MaxDiff results in individual-level preference scores and an RLH (Root Likelihood) fit statistic for each respondent that is a proxy for within-respondent choice consistency. RLH is a probability value from 0 to 100% (0 to 1.0). The higher a respondent's RLH fit, the more consistently the respondent answered the choice questions; whereas random responders usually have RLH scores at or near the chance level (more about this below). For RLH to discriminate reliably between random and consistent responders, the conjoint or MaxDiff questionnaire must have a healthy amount of choice tasks per respondent such that we have repeated measures of the same levels or items. For MaxDiff, for example, this means showing each item about three times or more to each respondent.

To identify bad respondents, we recommend combining the RLH fit statistic with other data such as time to complete survey, straightlining behavior (say, from rating questions outside the choice experiment), quality of the open-ends, and other consistency checks. Many researchers I know regularly clean from 15% to 30% "bad" respondents from conjoint and MaxDiff data sets using a combination of these checks. Sadly, the proportion of "bad" respondents has risen over the last few years.

There's not a simple rule regarding what indicates a bad HB RLH value, because the RLH depends on the number of concepts per task, number of tasks, the number of levels or items in the study, the type of None alternative or anchoring method, HB settings, and the particular

choice method (CBC, ACBC, or MaxDiff). So, for each discrete choice survey project you create, you can follow the steps below in Sawtooth Software's Lighthouse Studio to find what RLH level points to a likely random responder.

Using the Data Generator to Create Random Responders

Before we can identify random responders among a real conjoint or MaxDiff data set, we generate hundreds of artificial (bot) respondents so we know what random respondents look like in terms of RLH. If using Sawtooth Software's Lighthouse Studio…

1. First, delete any existing data in your Test data area on your local device (in case you've already generated some test records) by clicking Test + Reset Data...
2. Next, click Test + Generate Data...
3. On the dialog that appears, specify the number of random-answering respondents you wish to generate (by default it is 100...but you could increase it to something like 300 and that would probably be sufficient).
4. Click Generate. This generates 300 (or as many as you request) random-responding robotic respondents locally on your device. This may take several minutes and you'll see a progress indicator.
5. After you have generated those random-responding robots, click Test + Download Data to move those respondents (as if they were real respondents) into your project's main data file.
6. Last, click Analysis + Analysis Manager. Then, click Add to add a new utility run, select HB as the "Analysis Type" and click Run.

Step 6 computes HB utilities for your random respondents for the CBC, ACBC, or MaxDiff project. In the HB report you will find a tab that shows the individual-level raw utility scores and their RLHs. Copy the respondent IDs and RLH scores to Excel (or your favorite analysis package) and sort the random responders from high to low RLH. Examine the median RLH for the random responders, but more importantly make note of the 95% percentile (the top 5% RLH that random responders can achieve). We recommend this 95% percentile value as your RLH cutoff point to discriminate between good and bad respondents for your real data set. You are 95% confident that a random responder who takes your survey will fall below this cutoff level.

After you're done creating and examining the random responders, click Test + Reset Survey again to clean them from your Test data file. And, prior to collecting real data, delete the random responders from your final data file by clicking File + Data Management...then by going to the View/Edit Data tab and deleting any random responders from your data set.

We should note that the above approach may occasionally misclassify as a bad respondent a respondent who is a well-intending person who has done a reasonable, yet humanly fallible job. (If your MaxDiff or conjoint survey is too short relative to the number of levels or items in the study, the likelihood of misclassifying good respondents as bad increases significantly.) My

colleague Keith Chrzan has looked into the rate of false positives when using the approach described in this article (with recommended-length MaxDiff surveys) and finds that only a very few real respondents answering with the expected rate of human error would be misclassified as random responders. Using the RLH cutoff approach described here, you'll end up throwing away a great many more bad respondents than good respondents.

MaxDiff vs. Conjoint for Detecting Random Responders

With traditional MaxDiff, it's very unlikely for a random responder to obtain a high RLH when you show each item about three times per respondent. This makes MaxDiff very robust for identifying random responders. On the other hand, sparse MaxDiffs where each item appears about one time per respondent will regularly produce high RLH scores and it will be difficult to distinguish between random and conscientious responders.

Turning to conjoint analysis, a random responder is unlikely to pass the RLH threshold test (as with MaxDiff). But with CBC or ACBC, respondents with very high RLH scores aren't necessarily good, conscientious respondents! A simplifying strategy such as always picking None or always picking the lowest priced alternative will result in a high RLH score (the respondent is very predictable). Thus, for conjoint analysis, it's especially important to combine RLH with additional survey quality information such as response times, straightlining behavior, or other consistency checks.

Respondents who fail the RLH cutoff test for random responders (the 95% cutoff) should probably be deleted based on that evidence alone. For the respondents that remain, fast response time, straightlining behavior, and relatively low RLH scores (e.g., in the bottom 25th percentile) also could be considered and respondents who fail two or more of those checks might be candidates for deletion.

Is Data Cleaning Necessary?

Some analysts argue that keeping random responders rather than deleting them should only add random noise that cancels out across respondents and thus the summary of relative preference scores (utility values) for the sample should be minimally affected. But there is typically more to conjoint analysis than reporting average utilities.

Random responders damp the differences in importance between most and least important conjoint attributes, leading to biased market simulator predictions and associated inferences regarding attribute tradeoffs (such as between a feature improvement vs. price). Another concern is that random responders can cause problems for price sensitivity simulations (derived demand curves) or especially for profit or revenue optimization searches: they can have reversed price utilities that make it appear that they prefer ever more expensive product offerings. These problems can lead to decisions that overprice products and overpredict profits.

If you are using a methodology such as latent class MNL for segmenting respondents via MaxDiff or CBC data, too many random responders in the data can break out as a segment—the group of respondents who are less predictable and uniformly have smaller magnitude utility scores. At the 2019 Sawtooth Software conference, Marco Hoogerbrugge presented a paper wherein he identified this problem and exploited it as a way to use latent class MNL to identify random responders.

Speeders who Simplify

We've focused most of this article on detecting and cleaning random responders. But, we also mentioned another kind of "bad" respondent we should be aware of: the speeder who simplifies in ways that lead to high fit, but is not choosing like she would in the real world. Unfortunately, it's difficult to know for certain whether simplifying respondents should be deleted. For example, a respondent could quickly answer a CBC survey by choosing the lowest priced concept, her favorite brand, or by mostly choosing None. Such a respondent would have a high RLH and be consistent and predictable. While it is possible that people in real life use such simplification strategies to choose and buy products, it's also likely that respondents who are simplifying bias the data because they are not choosing in ways that they would do in the real world. Fortunately, this problem of simplifiers rarely occurs with MaxDiff. For MaxDiff studies in which each item appears at least three times per respondent, it's very difficult to obtain a high RLH fit statistic for MaxDiff surveys via simplification strategies, unless the respondent does something perverse such as rank the items in terms of length of text or by alphabetical order.

Summary and Conclusion

Over the last few years, the incidence of bad respondents is increasing. Conjoint analysis and MaxDiff have a fit statistic called RLH when using HB estimation that helps identify bad respondents. As long as the conjoint or MaxDiff questionnaire has enough questions relative to the number of levels or items in the study, random responders can be identified with a high degree of accuracy. For well-powered conjoint and MaxDiff experiments, low RLH is sufficient evidence to delete a respondent. With conjoint analysis studies, one should make additional checks beyond RLH to identify bad respondents who are simplifying in unrealistic ways to answer the survey with little effort (e.g. always picking their favorite brand irrespective of all other attributes). Additional data regarding time to complete the survey, straightlining, and poor quality open-ends could be leveraged to delete respondents who are not conscientiously answering conjoint surveys.