



Using discrete choice conjoint to validate stated-choice data

With the explosive growth and ease of conducting online surveys, one of the questions that keeps coming up is the reliability and validity of self-administered stated-choice data. This is compounded by the fact that oftentimes the easiest way to increase the response rate is to make the survey anonymous. This further raises questions about the truthfulness of a participant's responses.

Now more than ever, survey administrators are faced with the question of how to accurately model (analytically) a respondent's state of mind and thoughts. A survey is essentially an effort to numerically quantify a respondent's opinion. In this article, we will explore a modeling technique that uses conjoint analysis as an auxiliary validating tool for other stated-choice questions asked to the same respondent.

Let us assume we would like to model importance or purchase intent related to different attributes of a service or product. A

good example is airline ticketing. If the problem is to measure the importance users place in the different attributes (price, leg room, on-time departure, airline miles, etc.) we have a few options for the stated choice, or asking the user directly what they think. There are a number of different question formats you can use:

Editor's note: Vivek Bhaskaran is CEO and co-founder of QuestionPro, a Seattle research firm. He can be reached at vivek@questionpro.com.

attributes they think are important to them. Some would argue that a scale of 1-7 or even 1-9 is most effective. An example is shown in the table.

Please rate the following in terms of importance when purchasing a ticket from Seattle to NYC				
	Extremely Important	Very Important	Neutral	Not Important
Price				
Bonus Airline Miles				
Extra Leg Room				
Preferred Security/Screening				
On-Time Departure/Arrival				
Assigned Seating				

• Likert scale: Ask users to rate the importance of each of the attributes on a Likert scale. Users can choose a score of 1-5 (or 1-7 etc.) to numerically tell us what

• Top box scoring: Users can be asked to choose the top two or top three, etc., factors affecting their purchasing decision. In this model, you list all the attributes and force

users to choose two and only two items.

Example:

Please choose the top two factors that affect your purchasing decision for a ticket from Seattle to NYC	
<input type="radio"/>	Bonus Airline Miles
<input type="radio"/>	Extra Leg Room
<input type="radio"/>	Preferred Security/Screening
<input type="radio"/>	On-Time Departure/Arrival
<input type="radio"/>	Assigned Seating

- Constant sum: Another way is to use a constant sum methodology, which forces people to rank their preferences by distributing a total of 100 points between all the different attributes.

Example:

Please distribute 100 points between the following attributes according to the influence each has on your decision to purchase	
	Points
Price	
Bonus Airline Miles	
Extra Leg Room	
Preferred Security/Screening	
On-Time Departure/Arrival	
Assigned Seating	
Total	100

So which should we use? Let us examine the differences between the three proposed models. While the Likert scale is very popular, you'll note that it is the only one of the three that does not have a trade-off component. What do we mean by trade-off? Well, survey respondents can say all the attributes are "very important," which, while it may be true, does not really help the surveyor find out what is most important or least important, which is, after all, the whole

point of asking the question.

Why is this ineffective? As consumers we like to think that everything is important, but the reality is some attributes (cost, brand, etc.) are definitely more important than others. Models that measure emotions like importance and purchase intent that do not have a trade-off component are much more vulnerable to random and inexplicable variations based on the sample population being surveyed. As a general rule, trade-off-based modeling resembles real-life scenarios more closely than other models.

At the risk of making top-box and constant sum models sound utopian, it should be pointed out that they have some shortcomings of their own.

Top-box scoring is typically best for consumer-oriented surveys given how easy they are to answer. All a respondent needs do is pick the one or two most-important attributes. This results in a low level of cognitive stress. Most people intuitively know immediately what's most important.

The problem with this approach is that the data captured is limited to high-level analysis only. For instance, you don't know which of the chosen attributes is "most" important, or the degree to which one or more is more important than the others. And for many surveys, you run the risk of confirming what you already knew to be true (i.e., the price of a ticket was the most important factor in the purchase decision).

With the constant sum model, this problem is solved. In this case, you know exactly how important some attributes are relative to others, i.e., price is five times more important than on-time arrival.

But this high degree of analysis will come at a price for your respondents. Frankly, they have to think more and take more time - two aspects that have been shown to dampen response rates. They need to think and distribute 100

points over multiple attributes. They may get frustrated in filling out the survey. This model is generally more effective when the sample population is relatively sophisticated and has a high level of interest in filling out the survey.

Overall issues

Now that we've established a couple of models that can work in capturing data, let us look at some of the other overall issues with self-administered and stated-choice data.

One of the criticisms of self-administered and rated models is the belief that intentions often do not match actions. Let's take our airline ticketing example. The purchasing decision that users make in real life is a fundamentally different experience than taking an online survey that asks them to distribute 100 points between the different attributes that impact the purchasing decision. In real life, they need to get from A to B, and a number of factors play into their decision. While cost might be an issue in general, if they have to get to New York for a wedding, cost might not be as important as the schedule. In a survey, however, these external stimuli are absent. We are relying heavily on the cognitive ability of the user to translate their purchasing factors into their numeric equivalents.

Conjoint analysis models attempt to solve this problem by creating actual hypothetical products that users choose between. One of the artifacts that conjoint analysis can generate is the "relative importance" scores for all the attributes. The relative importance table essentially gives you on a 100-point scale (percentage) the importance of each of the attributes. This is very much similar to asking the user to distribute 100 points between the different attributes.

We now have two distinct and different models giving us the same data. The constant sum stated-

Pros and cons of various standalone question types

- **Likert scale**

Pros: Most common survey question format. Easy for users to understand and respond to.

Cons: Does not require users to trade off one attribute for another, enabling them to say "it's all most-important" and skewing results.

- **Top-box scoring**

Pros: Low level of cognitive stress. Users simply have to identify and choose the most important attributes. Good for consumer-oriented surveys.

Cons: Analytical data captured is limited to high-level analysis only. No detailed data of degree of importance for each attribute can be attained.

- **Constant sum**

Pros: Analytical data captured includes a component for the degree of affinity to a particular attribute. For example, inferences like "cost is twice as important as brand" can be made.

Cons: Higher degree of cognitive stress. Users need to think and distribute 100 points over multiple attributes. Users may get frustrated in filling out the survey. This model is generally more effective when the sample population is relatively sophisticated and has a high level of interest in filling out the survey.

choice data asks users directly and the conjoint analysis model generates the same data.

This multi-mode approach of evaluating the relative importance using conjoint and also through stated choice gives us a mechanism to validate/compare/contrast both the models. If the distribution of the stated-choice (constant sum)

data is similar to the values obtained by conjoint modeling, then we can safely (with a high degree of reliability) say that the values actually represent the sample population that took the survey. However, if the values are distinctly different between the two models, further investigation will be required to reconcile them.

In general, users tend to be more extreme when stating the importance of the different attributes when asked directly (stated choice). The conjoint model allows the researcher to calibrate responses and scores based on two models that technically should arrive at a similar conclusion.

More confidence

We believe that this ability to calibrate and validate the responses using two different models allows researchers to have much more confidence in the data and make more reliable judgments.

I would like to specially thank Henry C. Eickelberg, staff vice president, human capital processes, at General Dynamics Corporation, who is largely responsible to bringing this validating model to our attention. While conjoint is traditionally and largely used in the market research space, Eickelberg has successfully used the same model in the human resource arena. His research initiatives have helped in determining optimal benefits packages and efficient benefits management across 18 business units within General Dynamics. | Q