

The logo for MACRO CONSULTING, INC. features the word "MACRO" in a stylized, teal, sans-serif font. A horizontal orange line is positioned directly beneath the letters "A" and "C".

CONSULTING, INC.

Marketing Research / Advanced Analytics

The background features a light gray grid of dots. A dashed gray line trends upwards from the bottom left towards the top right. A solid orange line trends downwards from the top right towards the bottom left. Several blue dots are scattered across the grid, with one notable dot at the intersection of the dashed and solid lines.

Bring Your Survey Design Out Of The Dark Ages

Modern marketing science offers us the chance to see a little more clearly, dig a little deeper, forecast a little more accurately. In some cases, it's not a little. It's a lot. We have to understand, however, how the data will be used prior to writing the questionnaire so we can collect data appropriate for the subsequent analysis.

Author: Paul Richard "Dick" McCullough

Published by *the AMA* in
MARKETING RESEARCH MAGAZINE
Column "*BEG TO DIFFER*"
Spring 2010

Bring Your Survey Design Out Of The Dark Ages

Take a questionnaire written last week and place it side by side with one written 20, 30 years ago. Chances are they will look identical. Same logic. Same skip patterns. Same batteries and scales. Same limitations. Even though today's questionnaire is most likely being programmed on the web, with all the new question formats and controls web surveys offer. Yet the resulting data are often appropriate for nothing more than cross-tabs, just like 30 years ago.

Back in the day, quantitative market research meant cross-tab decks with 20 point banners. Back in the day, that was rocket science, state-of-the-art, leading edge. I wrote those surveys (and analyzed their data) with suspender-snapping pride. Problem is, we are no longer back in the day. Back in the day, corporate main frames didn't have the computing power of today's smallest laptops. Marketing scientists and other brainiacs have had the last 30 years to develop new analytic techniques to take advantage of all this computing power. These new and not-so-new-anymore methodologies are designed to eliminate many of the biases and inaccuracies of traditional surveys. They deliver answers to questions we didn't even dare ask "back in the day".

But the analytics are just the engine. They need fuel to run. And they need high octane fuel to run at their optimum. Antiquated survey designs yield very low octane fuel. They keep these high-powered engines from blowing past the competition and hitting that checkered flag first. Bad survey design turns your Ferrari into a Model T. And it happens every day.

There are three main problem areas in old school surveys:

- Missing data
- Collinearity
- Direct questions

Missing Data

Missing data in survey data sets are epidemic. Don't knows and skip patterns are the primary culprits here. Generally speaking, both are entirely unnecessary. And both are devastating to advanced analytics.

Many advanced models do not handle missing data very well. Yes, we can attempt to do full-information data imputation and, yes, that is a better way than mean substitution to address missing data values. But no data imputation technique nor any other analytic fudge factor will be as accurate as simply asking everyone the question in the first place. Most questions can be reworded so that skip patterns and DON'T KNOWs are not necessary.

The only other alternative is to exclude large segments of your sample because you don't have data for them. This is fine (ok, perhaps tolerable) for cross-tabs but when using powerful statistical models to determine big questions like "why do they buy?", it's important to keep all the sample you can. Not only do you need sample for statistical precision, you want to answer the big questions for everybody, not just for the tiny fraction that accidentally qualified for every skip in the survey.

Collinearity

Any two questions that are highly correlated contain essentially the same information. That is, they are wasting survey real estate. Test virtually any survey data set and you'll find collinearity of epidemic proportions—100 questions with the information value of 10.

Item correlation is not inherently evil (like missing values, for example. That's always evil). Measurement theory tells us that if we ask a question four different ways and then construct a latent variable based on the four original questions, we will have a more stable, more accurate measure of the underlying theme than any one of the four original questions. So correlation itself is not necessarily bad.

What's bad is correlation that is an artifact of the survey design, rather than statement content. We want our results to reflect truth, not bad research.

Direct Questions

Did you buy that sports car because you want to attract women (Yes/No)? Did you buy my product because of the ad you just saw (Yes/No)? You can bury these types of questions in a check all that apply battery (or whatever else) but you're just putting a dress on a pig. Respondents will answer any question you ask them. But they won't necessarily answer truthfully. Sometimes they don't know. Sometimes they don't want you to know. Advanced analytics can ferret out the truth that respondents may not want or may not be able to share. But you have to ask the questions differently.

Ask a male respondent how important the Playboy channel is to his decision to buy the premium package from his cable company and you'll get very low importance scores. This was even more true when we did mall interviews with college coeds as interviewers.

But conduct a choice-based conjoint analysis and you might find a different answer entirely. Why? Choice-based conjoint derives the importance of the Playboy channel by analyzing the pattern of responses across a wide range of programming options. It's indirect. The respondent isn't aware (and neither is that coed administering the interview) that his answers will ultimately reveal his true motivations.

Summary

Modern marketing science offers us the chance to see a little more clearly, dig a little deeper, forecast a little more accurately. In some cases, it's not a little. It's a lot. We have to understand, however, how the data will be used prior to writing the questionnaire so we can collect data appropriate for the subsequent analysis. Even without fully understanding the analytic plan, following these simple guidelines will vastly improve the quality of your data and subsequent analysis:

- Avoid missing values by eliminating skip patterns and don't knows
- Prevent collinearity by mixing things up: item order, polarity
- Derive importances; don't ask directly

© 2010 / MACRO Consulting, Inc.

Published in *MARKETING RESEARCH MAGAZINE* - Column "BEG TO DIFFER", Spring 2010



We are an independent **marketing research consulting firm** dedicated to helping you make the most informed, insightful marketing decisions possible. We specialize in technology, consumer, and new product research, and are well recognized for our **State-of-the-Art Research** techniques.

Ultimately, we provide more than just technical expertise. We focus on developing **pragmatic solutions** that will have a positive impact on the profitability of our clients.

CONTACT US:

Telephone: 650-823-3042

General Inquiries:
info@macroinc.com

Advanced Analysis Inquiries:
analysis@macroinc.com

richard@macroinc.com

www.macroinc.com