Quantitative Analytics (Making Sense of Survey Data)

By Jerry W. Thomas

The analysis of survey data is a massive topic, and most of this exotic landscape is beyond the purview of this article. The purpose of this paper is to offer some suggestions for the novice researcher, but even those with experience might find one or two of the tips useful. It doesn't matter how the data is collected (in-person interviews, phone interviews, online surveys, smartphone surveys, mail surveys, etc.), so long as the data comes from some type of survey.

So let's pretend that your boss hands you a set of cross-tabulations and a study file and asks you to analyze the data and make some sense out of it. Where do you start? What do you look at? What's important and what's trivial? How can you make sense out of thousands of different numbers? What's the story hidden in the data?

Mush and Slop

Surveys are largely attempts to transform human feelings, memories, perceptions, and attitudes into numbers. That is, the goal of a survey is to convert the "mush and slop" of human emotions and memories into words, and then the words into numbers. It's magic! It's alchemy. One starts with emotional goo and jelly and ends up with statistical significance at the 99.7% level of confidence.

The mushy, squishy nature of the things we are trying to measure, and the limitations of survey research, should give us a large dose of humility and skepticism. Human memories are fast to fade, unreliable, and often distorted. People are known to tell lies, or twist and bend the truth, to serve selfish ends. Prejudices and biases often color or tarnish what respondents say. Ignorance and lack of awareness often lead to respondent guessing and errors. All of our survey questions assume that the consumer is self-aware and can answer our questions (a tenuous assumption at best). At times our survey questions can "lead the witness" and inject bias into the results, or poor wording of questions can confuse the respondents. Let's always remember that our survey data should be thought of as "soft" data (as contrasted to "hard" sales data).

Survey data are not the perfect measurements that our cross-tabs and precise numbers tend to imply. So the first rule of survey analytics is humility. Skepticism and doubt are your analytical friends. Don't get seduced by small differences in the cross-tabs, whether statistically

You are searching for the big picture, the big story. You are searching for elephants, not mosquitoes.

significant or not. Don't become a member of the "faith in numbers" cult. Significance testing rests atop a foundation of quicksand. Remember, you are searching for the big picture, the big story. You are searching for elephants, not mosquitoes.

Analysis Abhors a Vacuum

If you don't know why the survey was conducted, don't know the study objectives, and don't understand the business issues that led to the study, your chances of doing great analytical work are meager. So the first task is to talk to your boss and/or whoever commissioned the study to find out as much as possible about the "background." What business issues prompted the study? What decisions will be based on the research? What are the expectations and beliefs of senior executives related to the topics of the survey? For example, if senior executives believe that men are the primary consumers of the brand, the fact that women are the primary consumers becomes the "big story," but if you had not talked to senior executives before the analysis, you might easily have missed this important point. You cannot analyze survey data in a vacuum. You must understand the market, the issues, the context, the motives, and the executive expectations to help guide your analytical thinking as you study the cross-tabulations.

God's Random Sample

Before you dive into the cross-tabs and immerse yourself in the data, however, you must first understand the sample. When training employees, I always say, "God created the random sample; all other sampling schemes are the work of the devil." The random sample is usually representative of the total market. As you move away from a random sample, many dastardly sins lurk in the sampling and respondent-selection process. Exactly who was surveyed, and exactly how were those respondents chosen? What were the sampling and screening criteria? Who was included, and who was excluded, from the final sample? If looking at survey data on the U.S. automotive market, for example, it would be essential to know that respondents who favored foreign cars were excluded from the sample (we actually observed this recently). Often, marketing and advertising executives only want to survey the primary target audience (say 18- to 24-year-old men), when this group constitutes only a small part of the total market. In this case, you can only analyze what is happening in the 18- to 24-year-old male segment, and you cannot generalize these results to any other groups. Review the sampling plan and know it inside and out. Look in the cross-

tabulations to make sure that the sampling plan was followed, and that the final sample is what it's supposed to be. Does the final sample (those who completed the survey) appear to be representative of the total market? Or is the sample biased, distorted, or limited in some way? If the sample is not representative, you must state this fact at the beginning of your analysis, and identify the limitations of the survey imposed by the sample.

Inequality of Questions

The next thing you must master is the questionnaire itself. Exactly what questions were asked in what order? Are the questions precise and easy to understand, or nebulous and confusing? You will want to give extra weight to the questions that are clear and precise, and ignore or discount those of dubious value. Are questions worded in a fair and balanced way, or do some of the questions "lead the witness" or bias respondents toward one answer over another? Does the order of the questions create any sequence bias? Were questions and answer choices properly rotated? Were products, concepts, or ads properly rotated or randomized? For example, if respondents are shown a sequence of items (products, concepts, ads), those things shown first score higher (all other factors being equal). A questionnaire is made up of predictive questions (e.g., intent to buy), diagnostic questions (why did you prefer Coke over Pepsi?), descriptive questions (brand image ratings), and classification questions (demographics). By studying the questionnaire, you can identify the questions

most likely to yield the very best information relative to your analytic objectives, and your analysis should be slanted toward or based on these "best" questions. An analysis is not, I repeat, not a summary of the answers to all of the questions. An analysis is an examination of the answers important and relevant to the purposes of your survey. Your final report may only include the results from a fraction of the questions in the survey if that is all you need to tell the big, important story.

Data Accuracy Supreme

Before you start searching for the big story in the cross-tabulations, you must personally verify that the data are accurate. So spend some time going through the cross-tabs to make sure all of the numbers add up and make sense. If you have any doubts or suspicions, then request a thorough re-examination of the data to make certain it is accurate. If the study design, questionnaire, or sampling plan was complicated, be even more suspicious of the data. If anything looks amiss, raise your suspicion level and degree of scrutiny. If the data does not make logical sense or "looks funny" to you, you must force

Your final report may only include the results from a fraction of the questions in the survey if that is all you need to tell the big, important story.

You will want to pay extra attention to the "heavy users" of a product category (who often account for 70% to 80% of total sales).

a re-examination of the data. Don't take "no" for an answer, until you are fully satisfied. At Decision Analyst, for example, cross-tabs are calculated by two different and independent systems, and all the results are cross-checked between the two systems. Our Quality Assurance Department spends hours cross-checking tabulations before they are released for the analysis phase. Regardless of the level of quality-control checking, always check the cross-tabs yourself and be on the alert for any anomalies in the data. You are now ready to analyze the survey results.

Heavy Users and Other Oddities

Marketing research surveys are very democratic. Each respondent gets one vote. The purely democratic model, despite its political strengths, can lead you astray in data analysis. Some groups of consumers have more income than others. Some groups of consumers buy your brand more frequently. In fact, in almost every product category, there is a relatively small group (10% to 20% of category users) that consumes the bulk of the category (accounting for 70% to 80% of total category sales volume). Thus, always keep an eye on how these "heavy users" are answering the survey questions. Their answers are far more important than answers of the light users. Also, keep in mind that heavy users are typically buying the product more frequently, and typically are much more knowledgeable about the product category.

Percentage Points of Change

Researchers and writers of all sorts get confused and tangled up when trying to analyze changes in percentages. Let's suppose that you conducted a phone survey and measured Total Advertising Awareness for a brand at 20% in January. Six months later you repeated the survey, and Total Advertising Awareness increased to 30%. Some would report this as a 10% increase (30% minus 20% = 10%), while others would report it as a 50% increase (10% increase divided by 20% base = 50%). Which is correct, and what's the best way to report this change in Total Advertising Awareness? First, you should report exactly what happened: Total Advertising Awareness increased from 20% to 30%. You should describe this increase as a "10 percentage points" increase (not a 10% increase). Always use the term "percentage points" when reporting an increase or decrease in the level of a percentage, or comparing two percentages. If you also want to report the ratio of change (i.e., the 50% increase in our example), always show your math so that the reader understands exactly what you are doing. For example, say, "The increase in Total Advertising Awareness of 10 percentage

points represents a 50% increase from the initial level of 20% (10% divided by 20% = 50%)."

Top-Two Sins

The research industry loves answer scales. These scales typically range from three-point scales (e.g., Very true, Somewhat true, or Not true), up to five-point scales (e.g., the standard purchase-intent scale: Definitely buy, Probably buy, Might or might not buy, Probably not buy, Definitely not buy), up to 11-point scales (0 to 10) or even 100- (1 to 100) or 101-point scales (0 to 100). Scales allow us to more precisely turn words and phrases into numbers. Once you have the answers, however, you are faced with an immediate problem: how do you comprehend and present the results, since you have respondents choosing many different points on the scale? One way is to aggregate some of the points (or boxes) on the scale, and report a "top-two box" percentage, or "top-three box" percentage. This may be the greatest sin ever committed by the research industry, because it implies that all of the boxes are equal in value (that is, that the answers on the top end of the scale are all equal in value, when the whole premise of the scale is just the opposite). To illustrate, here are two sets of answers where the "top-two box" scores are equal, although it's obvious the results are quite different:

Purchase Intent	Product A	Product B	Answer Code
Definitely buy	10%	50%	5
Probably buy	50%	10%	4
Might or might not	20%	20%	3
Probably not buy	10%	10%	2
Definitely not buy	10%	10%	1
Total	100%	100%	

This is an extreme example, but it clearly illustrates the folly of the top-two-box percentage ("definitely buy" percentage plus "probably buy" percentage). It's obvious that Product B is much more likely to be purchased than Product A, but the top-two-box percentage equals 60%

for both, indicating that the two products are equal in purchase propensity. The top-two-box percentage (and similar aggregates) should be banished from marketing research forever.

What's a better way? Well, you could just use an average of the answer code values. In the above example, a "definitely buy" answer would have a value of 5, and a "probably buy" a value of 4, and so on. The weighted average (number of respondents choosing each answer multiplied by the corresponding answer code value, summed, and then divided by total number of respondents) would discriminate between Products A and B. So it's a better measure than top-two-box, but what's the rationale for assigning a value of 5 to "definitely buy" answers and a value of 4 to "probably buy"? Answer codes tend to be sequential numbers to help the computer understand the answer chosen, but generally answer codes do not accurately reflect the relative value of the words in the answer scale. Is a "definitely buy" only worth 25% more than a "probably buy" answer? That's what is suggested by the answer codes (5 divided by 4, less 1.00). A better approach is to assign values to the answer codes that best reflect the relative meaning of the points on the scale. So in the example above you might decide to count

Always explain to the reader what you are doing and why. Transparency is next to godliness.

100% of all "definitely buys" and only 60% of the "probably buys," because you think the "definitely buy" is quite a bit more important than a "probably buy." Whoa! Wait a minute, you say. You are advocating the intervention of human judgment to change the results. That is true. Remember, however, that a series of human judgments gave you the original numbers, so it's okay to make new human judgments to create new numbers (hopefully ones closer to the truth). Always explain to the reader what you are doing and why. Transparency is next to godliness.

Always look at the distribution of answers across an answer scale. If answers are concentrated in the top box, then you might only report the top-box percentages (the "definitely buy" percentages in the example given). But if few people choose top-box answers, then the top-box percentages could be nonrepresentative and misleading, and you may have to use the top-two boxes or even the top-three (but weight them properly).

Benchmarks and Norms

Survey results are easier to understand if you have benchmarks, standards, or normative data. Numbers and percentages only make sense in comparison to other numbers and percentages. Decision Analyst, for instance, maintains databases of answers to various types of surveys (product tests, advertising tests, and awareness, trail & usage studies, etc.) and conducts surveys on various industries specifically to create comparative data, which its analysts can access while studying survey data. Also, secondary and syndicated data are used as additional points of comparison and reference. In looking at the answers to questions, it's good policy to always ask the question, "Compared to what?"

Graphs and Tables

The purpose of a graph, chart, or table is to communicate; its purpose is not to dazzle the reader with how smart, creative, and complicated you are. Keep charts, graphs, and tables as simple and concise as possible. The current fashion in corporate America is to jam as much confusing information as possible onto every graph, so that everyone will understand just how clever the presenter is. And even if a complicated chart is clear to the presenter, he will never be able to explain it to senior-level executives. The gods created the whole universe out of electrons, protons, and neutrons (just three things). Keep your graphs and charts as simple as possible. Don't show sets of five color-coded bars. When you present a graph, explain the graph as though you are talking to third graders. I was halfway through a presentation recently, when a question by a high-level executive revealed that he did not understand the percentages in my charts. Those of us who slave away in the research industry often forget that all those graphs and numbers are not so simple to the high-level executives who rarely work with survey research data.

The Findings Found

As you assemble the graphs, charts, and tables that illustrate or support the most important findings, you will want to concisely state these findings atop the supporting chart or graph. Think of these "findings" as conclusions, as hypotheses. The finding

should be the core, most important meaning of the data in the chart or table. The finding at the top of your chart or table should not be a recitation of the numbers in the chart or table. That's the whole purpose of the table or chart, to present the foundational data in an easy-to-understand format. No need to repeat the data in the finding. Here is an example of a bad finding:

> *Total awareness of Skippy Peanut Butter is 58%. Total awareness of Peter Pan Peanut Butter is 35%, and Total awareness of Jif Peanut Butter is 21%, as shown in the table below.*

This finding is merely a repetition of the data in the supporting table or graph. It adds nothing. It offers no conclusions or interpretations of what the data means. You should almost never write findings of this ilk. Here is an example of a better finding:

> *Skippy Peanut Butter enjoys dominant brand awareness. Its total brand awareness is almost double the next brand, Peter Pan. This dominant brand awareness tends to create a positive halo around Skippy Peanut Butter, as demonstrated in later charts.*

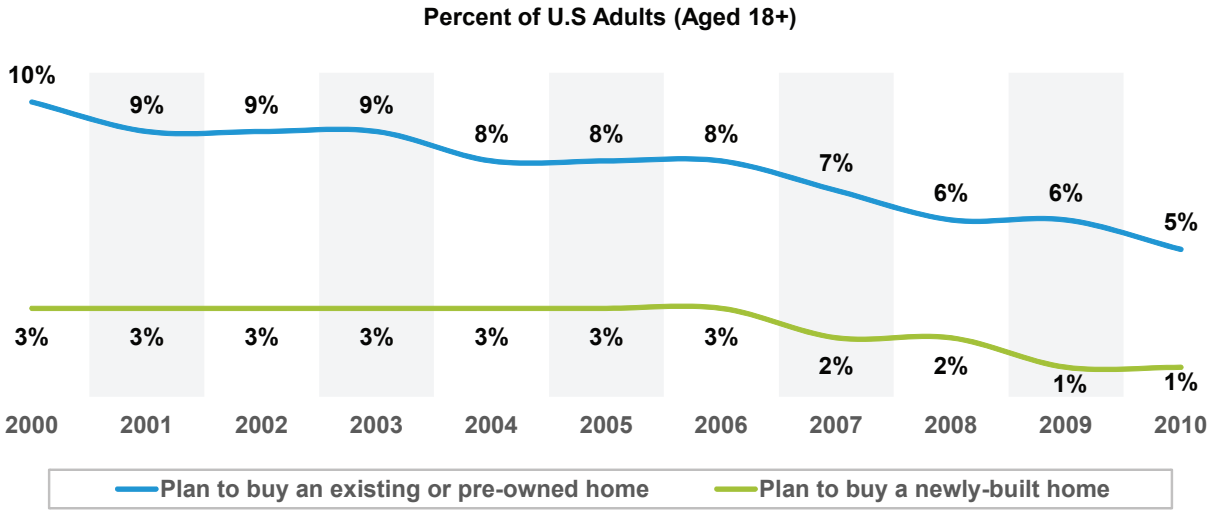
The second finding is better than the first. It reaches the conclusion that Skippy Peanut Butter dominates brand awareness, and

There are no shortcuts, no easy ways to analyze quantitative data.

Example of a Finding

Consumer Spending: Plan to Purchase a Home

The percentage of U.S. adults planning to buy a home in the next 12 months has gradually moved downward during the past decade, especially since 2006. The persistency of the decline indicates the profound weakness of the U.S. housing market.



points out an important implication of that dominance: a positive halo for the brand and its image.

The findings are the building blocks of your analytic report. Each finding should be concise and singular. It should focus upon a single, most important point or implication; it should not be a hodgepodge of 4 or 5 different ideas—all entangled with each other. Each finding should have some supporting role for your final conclusions and recommendations.

Conclusions and Recommendations

Once you have assembled the foundation of building blocks, the findings, think deeply about what the findings mean overall, and write down the most important conclusions. The conclusions should number no more than 5 to 10, typically. These conclusions are the findings, or are based on the findings, to help support and provide the rationale for the recommendations to follow. The recommendations must address and answer, if possible, the original objectives and goals for the study. Once the recommendations have addressed the study's goals, then you might offer other recommendations. What's the optimal target market? How should the brand be positioned? How can competitive threats be minimized? How can the product be improved? The recommendations section is where you pull all the pieces of the report together, and show your boss what a great analyst and writer you are.

The ultimate value of quantitative research is determined by the quality of the analysis.

Final Strokes

The triumphal analysis of survey data is not the result of short-term striving. You can master survey analytics in as few as 15 to 20 years of dedicated effort and attention. There are no shortcuts, just sweat, toil, and struggle. The prize at the end is an ability to derive truth and extract meaning from survey data, and translate all of the findings into a grand, poetic saga—the big story that leads to marketing and business success.

About the Author

Jerry W. Thomas is the President/CEO of Decision Analyst. The author may be reached by email at jthomas@decisionanalyst.com or by phone at *1-800-262-5974* or *1-817-640-6166*.

Decision Analyst is a global marketing research and analytical consulting firm. The company specializes in advertising testing, strategy research, new products research, and advanced modeling for marketing-decision optimization.

Decision Analyst
strategic research ■ analytics ■ modeling ■ optimization

604 Avenue H East ■ Arlington, TX 76011-3100, USA
1.817.640.6166 or 1.800. ANALYSIS ■ www.decisionanalyst.com