SURVEY ERROR

BY JERRY W. THOMAS

- SAMPLING
- UNIVERSE DEFINITION
- NONRESPONSE
- QUESTION WORDING
- SAMPLE SCREENING
- AGENDA
- CULTURAL
- OVERSTATEMENT
- UNCONSCIOUS
- ANSWER CHOICE
- TRANSITION
- ORDER
- ASSUMED KNOWLEDGE
- TABULATION
- SOCIAL DESIRABILITY
- TRANSLATION
- INTERPRETATION
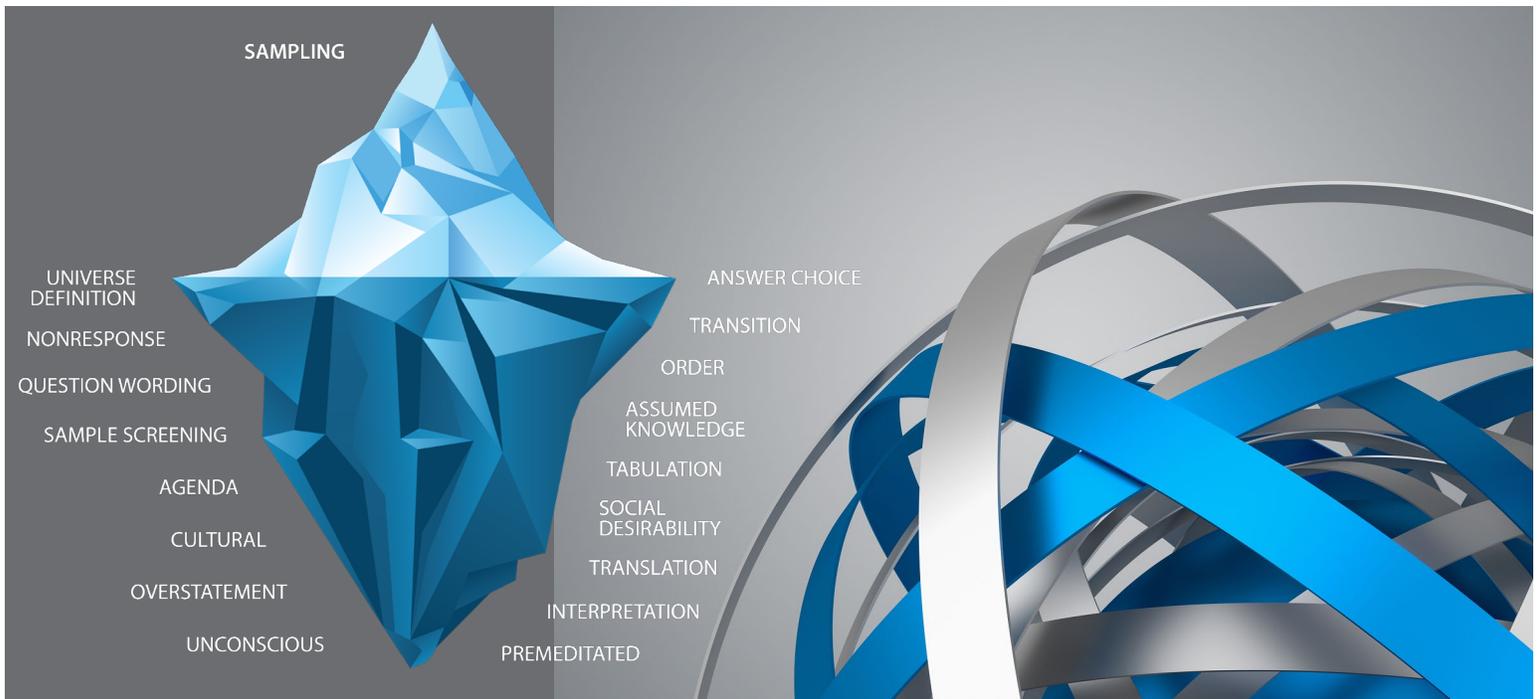- PREMEDITATED

The world is awash in data from surveys of all types. The rise of low-cost, do-it-yourself survey tools have added to the flood of survey data. We can scarcely buy a toothpick without a follow-on survey to measure how happy we are with the toothpick. All of these surveys and the data they generate (often based on relatively large samples, n > 1,000) tend to create a false sense of accuracy, based on the calculated Standard Error.

Decision Analyst
strategic research ■ analytics ■ modeling ■ optimization

Contents

STANDARD ERROR

The Standard Error is a widely accepted measure of Sampling Error, and it is typically the basis for the "accuracy of this survey is 5 percentage points, plus or minus, at a 95% level of confidence" footnote in research reports or survey results in newspapers, magazines, or websites. The Standard Error is the basis for Significance Testing. The Standard Error assumes…

> **SIGNIFICANCE TESTING:** helps determine if observed statistical differences are probably caused by chance variation.

- A sample is chosen by purely random methods from among all members of the target universe.
- All potential respondents do, in fact, respond to and participate in the survey (i.e., no response bias).
- The results of many identical surveys of the target universe (each based on a random sample) are normally distributed (i.e., the famous Bell Curve).

If these basic assumptions are met (and they rarely are), the Standard Error gives us a reasonably accurate measure of the Sampling Error in our survey data. However, the Sampling Error is only the tip of the iceberg. Many other types of Survey Error lie in wait for the innocent and the inexperienced.

UNIVERSE DEFINITION ERROR

If a mistake is made in defining the universe for a survey, the results can be very inaccurate. For example, if the universe for a liquor survey is defined as males aged 21 to 29, based on a belief that young adult males account for the bulk of liquor sales, the survey results could be completely misleading. The truth is that people aged 50 plus account for a large share of liquor sales. This is a good example of a potential Universe Definition Error.

Another example: For many years, a major U.S. automotive manufacturer defined its sampling universe as U.S. drivers who did not prefer Japanese or German cars over American produced cars. This peculiar definition of the sampling universe made all of the manufacturer's survey results appear to be much more positive (compared to a random sample). The auto manufacturer's executives could feel good about all the positive results coming in from their research department. It's easy to imagine the huge level of error introduced into this company's surveys by omitting consumers who preferred Japanese or German cars.

Another example of a Universe Definition Error is what happens in a typical customer satisfaction tracking survey. People who are unhappy with a company's services or products stop buying them, so they are no longer customers. The company's executives are happy, because satisfaction survey results are gradually getting better month by month, as unhappy customers drop out of the sample by becoming noncustomers. It is not uncommon for a company with declining customer counts to see its customer satisfaction ratings going up. The universe is changing over time, as unhappy customers leave. Universe Definition Errors, in most cases, will change survey results by significant amounts (usually much larger than the Sampling Error).

SAMPLE SCREENING ERROR

Sample Screening Errors are similar to Universe Definition Errors, in that we end up with a group that's not representative of the target universe. Most surveys consist of two parts: a screener to filter out consumers who don't qualify for the survey and a questionnaire for those who do qualify for the survey. Sample Screening Errors are common, and they can introduce large errors into survey results. For example, let's suppose a company wanted to survey people likely to buy a new refrigerator in the next three years. Such a survey might screen out consumers who purchased a new refrigerator in the past three years on the assumption that these individuals were "out of the market." Nothing could be further from the truth. That past-three-year refrigerator buyer might decide to buy a second refrigerator for the home, or they might buy a second home that needs a new refrigerator, or they might buy a refrigerator for their adult son/daughter. That buyer might also have a growing family that demands the purchase of a new, larger refrigerator, or momma might just get tired of the old refrigerator and want a new one. Survey results based on a final sample that excludes these past buyers of refrigerators would produce inaccurate results.

Another example of a Sample Screening Error: Let's suppose a candy company is completing 500 surveys a month to track its brand and advertising awareness. The company is screening for people who have purchased candy in the past 30 days, and only these individuals are accepted into the survey. The month-to-month survey results look stable, but suddenly in late October and early November brand awareness and advertising awareness decline. The brand manager is dismayed, because this is the peak selling season for candy. The problem is the past-30-day screening requirement. During the time around Halloween (the peak candy season), a much larger share of households buy candy. As the infrequent buyers of candy flood into the market at Halloween (and flood into the tracking survey), brand and advertising awareness fall because these infrequent candy buyers don't know as much about candy and aren't as interested in candy as the regulars who buy candy every month of the year.

NONRESPONSE ERROR

It's always possible that the people who do not respond to a survey are somehow different from those who decide to answer the survey. The U.S. government is especially concerned about Nonresponse Error, and often requires its research agencies to make a large number of attempts to complete a survey before giving up on a potential respondent. With the move to online surveys over the past 20 years and the increasing demands for instant survey results, the risks of Nonresponse Error are greater now than in the past. Ideally, the invitations to an online survey would be spread out over a seven-day period (to include a weekend) and would be emailed at a slow pace, with reminders on the third day to all those who had not responded. This process would tend to minimize Nonresponse Error, but it does not eliminate it completely.

AGENDA ERROR

A questionnaire is a set of questions. These questions create an agenda for the survey, and create an agenda in the survey-taker's mind. The subjects or topics of the questions (the agenda) can influence the results of the survey. If the first question asks about the dangers of knives in the kitchen, then a danger agenda (or bias) is created in the survey-taker's mind. Perhaps that individual never thinks much about the dangers of knives, but we've now planted an expanded awareness of knife-related dangers, and this might color the results of all the following survey questions.

Another example often happens in customer satisfaction tracking surveys. The inattentive customer may happily buy and consume a hamburger without ever thinking about how good (or bad) the hamburger is, and without thinking about how good or poor the service is. But then that customer gets the dreaded customer satisfaction survey and has to think about and rate his experiences at the hamburger chain. The satisfaction survey has now set the agenda and forced the customer to make judgments about the food and the service. By setting the agenda, the survey might actually cause the poor oblivious customer to come away unhappy. Without the survey, he might never have thought about his satisfaction or his happiness.

QUESTION WORDING ERROR

There are many types of Question Wording Errors, and these can be large sources of error in final survey results. As an example, suppose we asked the following question:

> **"Do you agree with leading scientists, college professors, and medical doctors that marijuana should be legalized in the U.S.?** (Choose One Answer)

> ○ Yes, I agree marijuana should be legalized.
> ○ No, I do not agree that marijuana should be legalized.

The above question is obviously leading and biased. A more neutral wording of the question might be:

> **"Do you think marijuana should be legalized for adults in the U.S., or do you think marijuana should not be legalized for adults in the U.S.?"** (Choose One Answer)

> ○ Marijuana should be legalized for adults in the U.S.
> ○ Marijuana should not be legalized for adults in the U.S.

As you might guess, the second question will give us a much more accurate measure of public attitudes toward the legalization of marijuana. The results to this question would be even more accurate if half of the respondents saw the question with "should be legalized" as the first part of the question, and the other half of the respondents saw "should not be legalized" as the first part of the question (and the answer choices were rotated to match). Poor composition of questions (that is, wording error) is one of the greatest sources of survey error.

ANSWER CHOICE ERROR

The vast majority of questions are closed-end; that is, respondents are presented with a set of predetermined answer choices. If you have not conducted some really good qualitative research before designing the questionnaire, you are at risk of coming up with answer choices that might not capture all of the answer possibilities. Do you include a "Don't Know" answer option? This can dramatically change the results to almost any question. Sometimes, knowing that 45% of your customers "don't know" the price they paid for your product is important information. When crafting survey answers, it's easy to leave out answer possibilities. You get results and they look reasonable, but you did not include the most important answer choice, so all of your results are meaningless. Pretesting or pilot testing every new questionnaire is absolutely essential to prevent Answer Choice Errors. Incomplete answer choices are a major source of survey error.

TRANSITION ERROR

If you are composing a questionnaire on the subject of peanut butter, and you suddenly switch to the topic of potato chips, respondents may overlook the change. It's not that respondents are not paying attention; they are thinking about peanut butter and concentrating on peanut butter, so everything starts looking like peanut butter—and they completely miss the change in topic to potato chips. Or if the survey is asking about advertising and suddenly switches to packaging, respondents might not notice the change in subject, and they might provide incorrect answers.

Transition Error can also creep into rating scales. If a rating scale suddenly changes during a survey, the respondent might not recognize the change. For example, if a rating scale goes from, "Excellent, Good, Fair, Poor" to "Poor, Fair, Good, Excellent" in the middle of a survey, even the most alert participant might not notice the change. Using many different types of answer scales within a questionnaire can lead to the same type of error. The respondent accidentally misreads the answer choices as the questionnaire bounces around from one scale to another. These sequencing and transition errors can be a significant source of survey error. It's especially a problem in omnibus surveys that contain many blocks of questions on different topics.

ORDER ERROR

If answer choices are not rotated or randomized, survey error can result. For example, in a paired-comparison product test, the product tasted first tends to be preferred by 55% of respondents (give or take), even if the two products are identical. If two new product concepts are compared, the one shown first (all things being equal) will be preferred over the one shown second. If a question is followed by a long list of brands, or a long list of possible answers, the brands toward the top of the list will be chosen more often, all other factors being equal. That means it's important to rotate or randomize the answer choices to prevent Order Error. Of course, not all answer choices are rotated or randomized, because some questions don't work if the answer choices are out of order (e.g., a purchase intent scale or a ratings scale). Order Error is typically not a huge source of error in most surveys, but in some instances it can be of major significance.

ASSUMED KNOWLEDGE ERROR

Survey creators often assume that survey participants possess more knowledge about a topic than is the case, and assume that survey respondents are familiar with the language and terms used in the survey—when often they aren't. Frequently, survey participants simply don't understand the words and terms in the questions, or they don't understand the answer choices. People will almost always give an answer, even if they have to guess. These guesses can be a significant source of survey error. That's why it's recommended that surveys be subjected to a pilot test, to make sure everyone understands the questions and answers and has enough topic knowledge to give meaningful answers. Assumed Knowledge Error is a potential problem in both consumer and B2B surveys.

TABULATION ERROR

Every company that tabulates survey answers makes assumptions and sets operating procedures that affect the reported results from a survey. For example, in calculating an average for grouped data, there is latitude for differences or error. In calculating average household income, as an example, how does one count the "Under $25,000" income group? Do you count an "Under $25,000" answer as $25,000? Or do you count it as $12,500? Or as $18,750? What about household income greater than $250,000? Do you count that household's income in computing an average as $250,000, or $300,000, or what? If a question asks how many times people have gone swimming in the past year, do you calculate an average number of times, or do you look only at the median? The base chosen to calculate a given percentage or average can change the results dramatically. These tabulation decisions can be a source of major differences or error.

SOCIAL DESIRABILITY OR SOCIAL PRESSURE ERROR

We humans are highly emotional, social creatures. We want others to like and admire us. This leads to something called "Social Desirability Bias," or Social Pressure Error, particularly in surveys conducted by a human interviewer (face to face and/or by telephone). The respondent gives answers intended to make the interviewer think better of the respondent. For example, the respondent might say he has a Master's degree, when he only has a Bachelor's degree; or the respondent might say that he goes to church every Sunday, when in truth he only goes once every six months. When asked how much he weighs, the respondent might say 165 pounds, when the truth is closer to 180. These social desirability biases are muted in online surveys or mail surveys, but they never disappear completely.

TRANSLATION ERROR

If the same survey is conducted in the U.S., France, Germany, and China, the translation of English into the other languages introduces differences (error) into survey results. That is, the results in English will be different from the results in French, German, and Chinese, purely because of language differences. No translation is a perfect mirror of the original language. Some words in English don't exist in German or Chinese, or they have no exact match in the other languages. Even if you are working with a highly skilled translator with marketing research experience, the questionnaires across different languages will never be the same.

CULTURAL ERROR

On top of the differences in language from one country to the next, cultures tend to be different. Some cultures are lively and festive, others drab and dull. Some cultures are happy and positive, while others are serious and dour. Some cultures like to give positive, glowing answers, while other cultures tend toward a negative worldview. These cultural differences lead to differences in survey results. We can think of culture as another source of bias or error.

OVERSTATEMENT ERROR

If you ask consumers how many cans of pinto beans they purchased in the past year (or the past month), they will overstate the actual number of cans purchased by a factor of 2 or 3 to 1. For high priced products, the overstatement factor might be 4 or 5 to 1 (or even higher). If the researcher accepts these reported purchase numbers at face value, the survey results will overstate the true numbers by huge margins. Likewise, if a consumer is asked how likely she is to buy a new peanut butter, she will overstate her likelihood to purchase by a factor of at least 2 or 3 to 1. If the manufacturer bases sales projections on these inflated purchase intentions, far too many jars of peanut butter would be produced and shipped. Overstatement Error is huge for certain types of questions and is another major source of error in survey results.

INTERPRETATION ERROR

Overstatement Error can lead to Interpretation Error. Let's suppose you have completed that new product concept survey and are ready to write the report—and impress your boss. Thirty-two percent (32%) of respondents said they would "definitely buy" the new product, and 23% said they would "probably buy?" Let's see, there are about 120 million households in the U.S., so 55% (32% plus 23%) means that 66 million households might buy this new product. And the respondents said they would buy the new product 6 times a year, and its price is $9.95. So 66 million times $9.95 times 6 equals market potential of just under $4 billion dollars. Wow! Your boss is going to be so happy when he hears the results. Your ascension to the corporate throne is only a matter of time. This tongue-in-cheek example illustrates the kinds of interpretation errors that human judgment can introduce into survey results. Yes, you used the results from the survey exactly as printed out in the cross-tabs. The numbers are correct. But your $4 billion market potential might only be $200 million once an experienced researcher discounts the survey results and adjusts for planned advertising spending, awareness, distribution, and competitive response. The interpretation of survey results is often a major source of error.

PREMEDITATED ERROR

You visit your favorite hamburger place, sit down to enjoy your lunch, and the store manager comes to your table to check on your happiness. Once assured that you are satisfied, the manager reminds you to respond to the satisfaction survey invitation printed on your receipt. Or, you buy that new car and, at the very end of the process, the salesman takes you aside and explains how his job and his future depend on getting good ratings on the satisfaction survey that you will shortly be asked to complete. He implores you to give him 10s on the survey, because anything less will doom him to poverty and failure. You feel sorry for the poor bloke and give him 10s. This type of error is commonplace in customer satisfaction surveys and raises questions about the true value of such surveys.
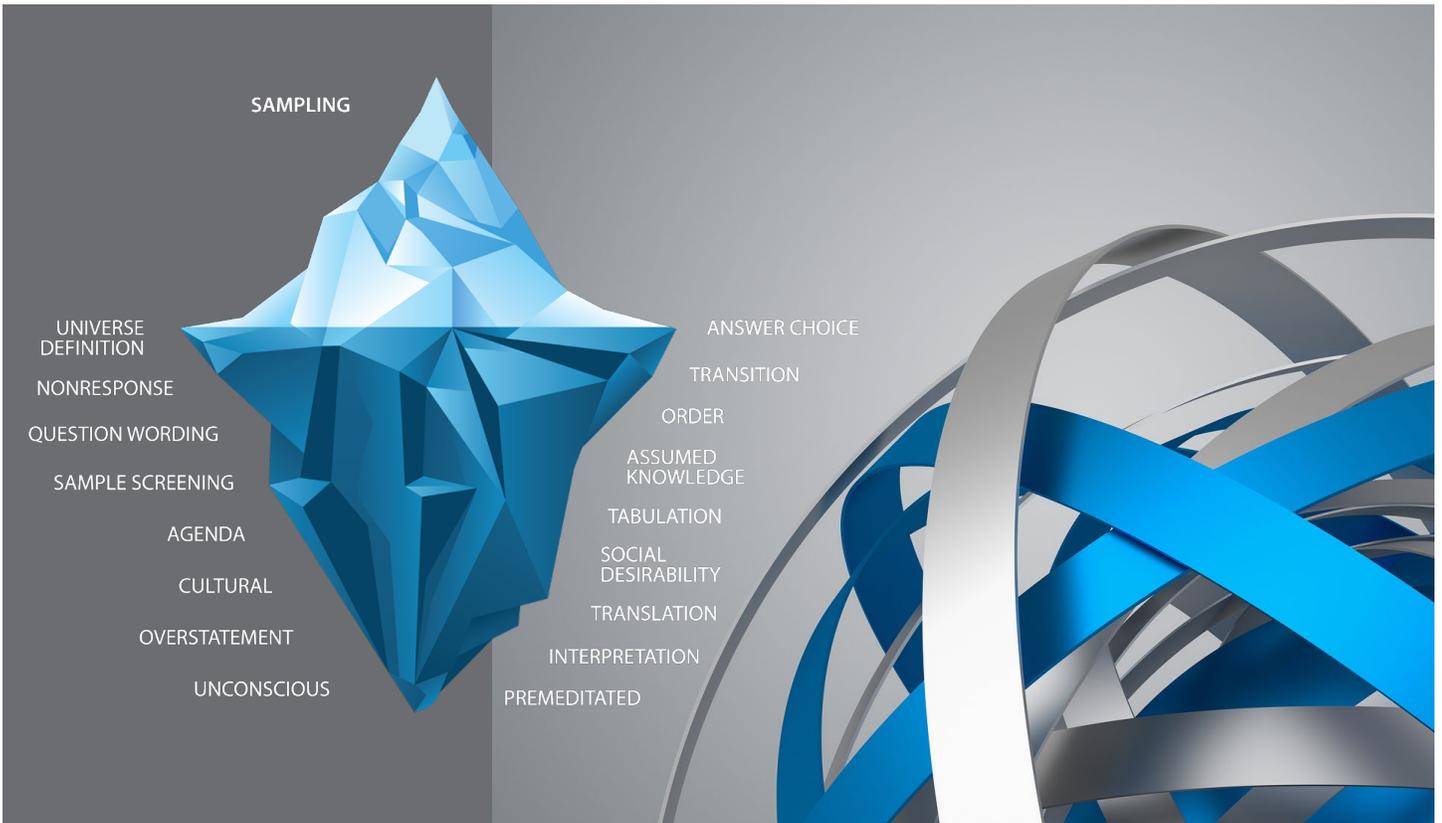
UNCONSCIOUS ERROR

If you are emotionally involved in a corporate project (say, the development of an exciting new product), you might just fall in love with the exciting new innovation that you are bringing to the world. When the higher-ups want some evidence to support your unbridled enthusiasm, you design a survey to provide the answers. Your enthusiasm, your emotional involvement, and your love of the new product causes you, without conscious awareness, to slant the definition of the sample and the wording of questions to provide the affirmation you so badly want. This is a major source of survey error, especially when corporate surveys are conducted directly by corporate employees.

CROSS-TABULATIONS AND SIGNIFICANCE TESTING

Many researchers demand that cross-tabulation tables be cluttered with significance tests based on the Standard Error (i.e., sampling error), or they spend considerable time running tests to determine if survey results are statistically significant (again based on the Standard Error). In the grand scheme of marketing research, sampling error is typically a minor source of survey error, compared to all of the other sources of error, yet it consumes 100% of the typical researcher's attention and time. It might be wise to skip the significance testing in cross-tabs and charts and instead focus attention on the other sources of potential error—where risks and degree of error are much greater than sampling error.

FINAL THOUGHTS

Significance Testing (based on Standard Error) is vastly overrated in importance, and it might actually cause us to overlook what's really important in research design, questionnaire design, and survey interpretation. Focus on minimizing nonsampling errors to dramatically improve the validity, reliability, and accuracy of your survey research. Always pretest or pilot test a totally new questionnaire design. Recuse yourself from a survey if you are emotionally involved in the subject of the survey or in love with the brand manager. Check your survey results against previous surveys and against secondary data to make sure the survey results are reasonable and within acceptable ranges. Be skeptical. If something seems amiss or out of kilter, keep searching for the explanation or the source of the error. Ultimate truth is elusive and shy, and it's our job to coax it out of the shadows.

SAMPLING

UNIVERSE DEFINITION

NONRESPONSE

QUESTION WORDING

SAMPLE SCREENING

AGENDA

CULTURAL

OVERSTATEMENT

UNCONSCIOUS

ANSWER CHOICE

TRANSITION

ORDER

ASSUMED KNOWLEDGE

TABULATION

SOCIAL DESIRABILITY

TRANSLATION

INTERPRETATION

PREMEDITATED

Jerry W. Thomas is President/CEO of Dallas/Fort Worth-based Decision Analyst (www.decisionanalyst.com), a global marketing research and analytical consulting firm.

The company specializes in strategy research, advertising testing, new product research, and advanced modeling for marketing-decision optimization. Mr. Thomas may be reached by email at jthomas@decisionanalyst.com or by phone at 1-817-640-6166.

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