Displaying the Results of an Employee Survey and a Balanced Scorecard Analysis to Non-Statisticians

Palmer Morrel-Samuels
University of Michigan, Ann Arbor, Michigan, and EDS Survey Services, Chelsea, Michigan

Abstract
This paper provides a brief outline that will assist those who present the results of employee surveys and attitude assessments to audiences that lack formal training in statistics. The analyst is urged

1) To provide proof of reliability and explain its importance;
2) To provide evidence of validity and explain its importance, and;
3) To provide evidence of direct business utility and to show how that information can be used to improve the organization. Guidelines are offered for

A) Accomplishing those three goals;
B) Responding to common criticisms;
C) Following conventional rules for legibility;
D) Accommodating the cognitive strengths and limitations of audience members;
E) Using graphics pulled from JMP® software written by SAS Institute, and;

Introduction
Everyday in businesses across the nation hundreds or thousands of unsuspecting and naive functionaries suggest that their company deploy a survey. It is fair, or even charitable, to describe these optimists as being unsuspecting and naive because they usually have little notion of the complexity that good survey design entails, and even less idea of the difficulties of proper statistical analysis. For example, on a recent inventory of managerial effectiveness I read the unintentionally impossible statement "I always think in terms of 'we' instead of 'I,'" a test item that may set a new standard for internal contradiction given the fact that respondents are all addressed as "I" in the first place. But the limits of survey work in the business world are even more starkly defined then that example suggests: On one recent engagement I was called in to analyze a relatively large dataset from a big company's customer survey; the company was delighted by the customer's response rate which was bringing in 31/2 thousand new respondents every month. Accordingly, when it was pointed out that the response rate was a mere 2% of the company's new customers, and that the norm for similar surveys in a study conducted by the National Academy of Sciences was 55%, representatives of the company were reluctant to conclude that their respondents could be anything other than entirely representative. In such a situation one can only politely respond that yes, the responses are probably very representative...representative at least of the opinions held by the two percent of the population that has plenty of time, and loves to fill out opinion surveys.

Compounding this problem is the fact that at every level in every bureaucracy there are folks who adore surveys, and think the best method for analysis is to designate the three highest averages as the best group, and the three lowest averages as the worst group, without ever pausing for a moment to imagine the possibility that nothing more than the flip of a coin has generated the illusion of differences between and within those groups.

However, the picture is not as bleak as one might suspect. It is almost invariably the case — or at least it has been in the institutions where I have been doing research during the last 15 or so years — that someone with administrative power, somewhere, in every organization is serious about collecting accurate, informative, and useful data... even if those data contradict some long-cherished stereotypes and require careful self-examination. In the brief space allowed here I will offer an outline of what those serious-minded people deserve to get when they pay someone to conduct a survey in their organizations, and, following that, will provide a few guidelines derived from both professional experience and published research that can assist you as you present survey results to non-statisticians.

What Survey Sponsors Deserve
When a business, school, hospital, or government agency sponsors a survey they deserve to get an assessment tool that has reliability, validity, and direct business utility. Lacking any of these three elements (each of which will be described in greater detail below) is the equivalent of turning a farmer's galvanized water pail into a high priced antique by emptying it out and moving it from a barn to a fancy gallery: the economic pressures of the situation may make some people want to treat a practical device as a decorative antique, but an ordinary steel pail from the local hardware store is still just an ordinary steel pail, and there are good reasons not to turn it into a decoration and rob it of all its functional value. Similarly, if a survey lacks reliability, validity and business utility, then it is devoid of all practical value, and becomes nothing more than an empty intellectual exercise.

Reliability
Survey sponsors deserve to be provided with clear evidence of reliability so that they can have proof that each respondent's answers are consistent (i.e., neither random nor self-contradictory) throughout the test. This can be done by including a reliability check in each section of the survey so that a correlation can be measured between overall responses in any given section and an individual item. For example, if you are asking students to evaluate their teachers, and the school administration claims there are four major attributes (e.g., being intelligent, personable, lucid, and well informed) that make their teachers uniquely effective and you have five questions to tap each of those four attributes, there should be four additional questions in the survey that ask respondents about each of the four attributes as a whole (e.g., my professor exemplifies what it means to be an intelligent teacher). If the questions are incomprehensible or if respondents are simply generating
answers by random then there will be no significant correlation between the average of the four specific intelligence questions and the one general intelligence question; similarly, if there is a negative correlation between that average and the general intelligence question, then reliability is so poor that respondents contradict themselves from moment to moment. Far from being a recondite and exotic concept of test design that survey sponsors have no interest in, when handled properly, evidence of reliability will convince skeptics and detractors that—at the very least—respondents were not just making up their answers by picking numbers out of a hat. This is an important component in presenting surveys to non-statisticians because many in the population believe Mark Twain's old saying that "there are lies, damn lies, and statistics". Proof of reliability will remind non-statisticians that letters are just as suited to lying as numbers, and that ethical folks use neither medium for anything other than honest purposes. Accordingly, proof of reliability is the first and most basic step in a credibility-enhancing process.

Validity

Even more important than evidence of reliability, it is imperative that survey sponsors get some evidence of the assessment's validity. Validity, of course, is the quality that allows you to know that a test, assessment, or survey tool does actually measure what it purports to measure. In fact, it is reasonable to argue that surveys in the business world are currently held in such low regard in those institutions where they are held in low regard precisely because previous statisticians, human resource specialists, and psychologists have failed (and failed rather miserably) to demonstrate that their tests and surveys actually measured what they claimed to measure. Unlike evidence of reliability where we can choose from a number of test statistics (e.g., the Spearman-Brown Rho, the product-moment correlation coefficient, etc.), validity can only be proven by looking at converging lines of evidence. But, as with reliability, the idea is not some rarefied piece of exotica: you can, for example, prove that your assessment of teaching ability has good validity by showing that teachers who score high on your assessment have fewer students dropping out of their non-required courses; have longer wait lists for enrollment, have greater numbers of graduate students requesting positions in their labs, etcetera. And, admittedly, while no single piece of evidence such as this could stand as irrefutable evidence (because, for example, graduate students may be basing their decisions on market driven factors that have little to do with teaching effectiveness), an entire constellation of consistent evidence can convince even the most hostile audience that your assessment, while not perfect, does an adequate and useful job of measuring the concept of interest. And that usefulness is a crucial part of the picture.

Business Utility

The gold standard of opinion surveys and assessment tools is something that, almost without exception, is never mentioned in the methodology text books. No survey or assessment is worth a solitary penny if it lacks business utility, even if it has great reliability and clear evidence of validity. That is, if the questions are irrelevant, have no bearing on the organizational culture where they were deployed, or merely address problems that people can neither improve nor alleviate, then there is no justifiable reason for conducting a survey. In the most basic manifestation, business utility is what makes the results of a survey actionable, so that remedial steps can be taken to enhance profits, increase staff motivation, improve quality, or whatever the needs of the moment require. And if business utility is in line, a good survey CAN be used as part of a program to diagnose problems, plan interventions, and document the effectiveness or ineffectuality of those remedial interventions.

The remainder of this paper will provide a set of guidelines for working with non-statisticians—more specifically, guidelines that pertain to data visualization using JMP software with such an audience. The work will draw heavily from Kaplan and Norton's fine articles in the Harvard Business Review on the Balanced Scorecard (Kaplan & Norton, 1992; 1993) Tufte's writings on data visualization (Tufte, 1983; 1991) and recent research on cognition and quantitative analysis from the field of experimental psychology.

General Guidelines

1) Prove reliability with a direct and straightforward statistic (e.g., a correlation coefficient) that has a set probability level attached and that can be displayed as a simple bivariate correlation. (Raw percents and Kappa are not especially helpful because the role of chance is either neglected from consideration, in the case of the former, or difficult to assess in the latter).

2) Prove validity by displaying bivariate plots or (preferably) leverage plots that show a clear relationship between the construct of interest and a hard and fast measure of performance common in the sponsor's own domain (e.g., between a leadership score and staff retention). JMP is especially helpful here for data visualization because leverage plots are generated by default in all multiple regressions.

3) Promise and deliver direct business utility by focusing on topics and domains of operation where remedial steps can be taken to improve actual performance in the sponsor's institution. Adopting the Balanced Scorecard approach of Kaplan and Norton can provide a sound general framework for such an analysis.

4) Prepare for two criticisms:

A) That the results are entirely implausible, and;

B) That the results are absolutely self-evident. The best approach here is to be sure that the presentation of results provides good grounds for both criticisms. That is, provide genuinely new information (e.g., from a factor analysis) that cannot be anticipated even by the skeptics who want to believe only what they already know, and provide some results that are nothing other than confirmation of the obvious. This counterintuitive approach does work, even for the hard-bitten critics who usually advance these charges. And being able to navigate through the east wind and the west wind of these contradictory claims is crucial because one half of this "implausible/obvious" quandary is invariably presented by influential non-statisticians who claim their (substantial) success is attributable only to gut instinct, real-world experience, and common sense. Such critics can only be converted into supporters by showing them that the self-evident results confirm the validity of the methodology, and that the surprising results justify the inconvenience and expense of the entire surveying process.
5) Follow generally accepted guidelines for clarity and legibility available from almost any company, but violate those guidelines with impunity whenever the content requires. After all, if we had followed the old saw that slides should never have more than 5 lines of text, the ten commandments could never have been written as a unified whole; Moses would instead have received the 8 plus 2 commandments, or the 5 & 5 Commandments. Some material just needs to be presented in its entirety so that audience members can represent the information in memory as a unified whole composed of equally weighted subsections. The best of these guidelines point out rules that are self-evident when you pore through Tufte’s books on data visualization. They include such straightforward advice as the recommendation to use color only to set off meaningful differences; to use graphic metaphors as a conscious vehicle for conveying a message rather than merely page filling gimmicks and polychromatic sight gags. Supplementing these are two guidelines from research journals in experimental psychology, where one almost never sees a truncated scale on a plot or graph, and where it is never ever the case that a subpopulation is separated out in a graph or plot unless data from that subpopulation require special attention.

6) It is good to bear in mind the fact that people have an easier time remembering information that is part of a set of interacting concrete images (Glass & Holyoak, 1986); that rote memory, in most cases, produces less accurate recall and recognition than memories based on a coherent story; That in most cases, recognition memory is much better than recall memory, and that, in general, the greater the frequency of exposure the better the recognition memory AND the higher the person’s positive feeling about the information being remembered. To round out this set of summary items, recall that people do better when called upon to remember the first few and the last few items in a list. It is also important to remember that recognition and recall will be most accurate and most efficient if information is clustered in sets where the number of elements is 7 + 2, a parameter that takes greatest advantage of the need to limit the number of sets in a big information base, but still keep the size of those sets within the range where recall and recognition are optimal.

7) Use photographs or videotapes to supplement your presentation. In no other class of endeavor is it so true that pictures can convey with succinct clarity what conceptual explanations can reveal only with wordy explanations. I cannot count the number of times I have tried to convey to undergraduate students the importance of not analyzing subtables that have been removed at will from a larger data matrix. But one glance at a picture of 20 or 30 assorted coins — some pennies, some nickels, some dimes — is all it takes to alert non-statisticians to the fact that themselves are nothing more than one of those one hundred marbles rolling into the only hole on the entire surface of that glass. Perhaps the mental exercise won’t help your non-statisticians appreciate the role of chance as it operates in their agency or business, but then again it just might.

8) And my final recommendation, the most important of any mentioned so far. If nothing else is remembered as you pass over the information sketched here in this brief outline, it is imperative to remember the one feature that allows non-statisticians to pay attention long enough to understand, appreciate and use the information you provide on your survey, questionnaire, or assessment tool. In all cases, build a multivariate set of dependent variables and predictor variables, so that part of those indices can be absolute, rock-hard, indisputable numbers drawn from the sponsors’ own organization. That is, follow the general guidelines of the Balanced Scorecard approach, so that performance can be examined across an entire enterprise: It makes no sense, for example, to talk about improving quality, unless you can show that that improvement is not going to be bought by sacrificing productivity. It makes no sense to improve employee motivation if that change is secured by sending the company’s profit margin down the slippery slope to bankruptcy. And — naturally enough — such an analysis will necessarily incorporate both perceptual measures (i.e., measures that quantify subjective opinions) and ratio scale measures based on discrete irreducible units (i.e., quantitative counts where there is an absolute zero point, where the steps between each incremental increase on the scale is equal in size to every other incremental increase throughout the scale’s range, and where the elements being counted are handled as discrete units like ten new cars in a sunny parking lot, three live pigs in a red barn, or a hundred black marbles rolling on a huge sheet of pure white glass. My closing advice is to have the non-statisticians in your audience imagine that huge sheet of white glass careening through empty back space, and to further imagine that they themselves are nothing more than one of those one hundred marbles rolling into the only hole on the entire surface of that glass. Perhaps the mental exercise won’t help your non-statisticians appreciate the role of chance as it operates in their agency or business, but then again it just might.

References

Acknowledgements
The author wishes to thank Paul Holland and Justin Ford of EDS Human Performance Services for their consistent encouragement, help, and support.
Palmer Morrel-Samuels
EDS Survey Services
210 Park St.
Chelsea, MI 48118
Tel: (313) 475-3145 palmer@umich.edu

JMP is a registered trademark of SAS Institute Inc. in the USA and other countries. © indicates USA registration.