

Title:

Should you trust your analyst? (Part I)

Word Count:

1111

Summary:

The first stage of most business decisions, such as marketing, hiring, and investing, is gathering data. In most cases, the information is captured in the form of words. Once the gathering of data is complete, the next step is analyzing the data. In many cases this analysis is performed by professional analysts, such as marketing researchers, human resource managers, and portfolio managers. In light of some recent scientific research, should we believe their analysis, and recommendations?

Keywords:

Focus Group, Interview, Survey, Qualitative Research, Qualitative Analysis, Investment Analysis, Open-ended questions, decision making, how to negotiate, conversation analysis, negotiations, text analysis

Article Body:

The first stage of most decision making in business is gathering data. In most cases the information is collected in the form of words (also called qualitative data, or unstructured data). For instance, marketing researchers conduct focus groups, perform in-dept interviews, or use open-ended questions in surveys to enable product managers and sales representatives to choose the best product design and the most effective message to convey to customers. Another example is human resource managers who conduct interviews with candidates to enable the company to choose the best candidate for the job. Once the gathering of data is complete, and the words are available, the professionals who gathered the data perform an analysis of these words.

A recent study (Craigie M, Loader B, Burrows R, Muncer S. Reliability of Health Information on the Internet: An Examination of Experts' Ratings. Journal of Medical Internet Research. 2002 Jan-Mar;4(1):e2) measured how consistent are experts when they analyze qualitative data. The data included the text from 18 threads (series of connected messages) posted on a message board by individuals suffering from a chronic disease. Each thread consisted of a start message, or question, and a number of responses, or answers. The experts processing the data were five doctors who worked together in the same specialist unit, and who had at least five years experience in treating the chosen disease. To process

the data, the doctors devised the following two scales. The start message or question was coded according to a 6-part scale: A = excellent; B = less good but with some details; C = poor with little detail; D = vague; E = misleading or irrelevant; F = incomprehensible. The responses or answers were coded according to another 6-part scale: A = evidence based, excellent; B = accepted wisdom; C = personal opinion; D = misleading, irrelevant; E = false; F = possibly dangerous.

After processing the data, the codes assigned by all five experts were compared using three statistical tests: kappa, gamma, and Kendall's W. The results showed **poor agreement** between the codes of all five experts in both the starting question and the responses. Moreover, two of the five experts showed a statistical significant **dis-agreement** between the codes they assigned to the question, and different pairing of experts showed **contradictions** between the codes they assigned to the responses. In simple terms, when one doctor labeled an answer with "A = evidence based, excellent," another doctor labeled the same answer with "E = false," or even "F = possibly dangerous."

Points to consider:

1. In this study, the analysts were doctors with at least five years of experience in treating the specific chronic disease. These analysts possess a much higher level of expertise in the research subject relative to even the most experienced market researchers analyzing qualitative customer data, or the most experienced human resource managers analyzing candidate data. So, if these highly trained experts failed to show consistent processing of qualitative data, what are the chances that the less trained professionals will show consistent analysis of their data?

2. The criterion in this study was whether an answer is "evidence based" (see code A) or not. This is an objective criterion. Unlike this study, the great majority of qualitative studies in business involve subjective criteria such as tastes, morals, values, or preferences. If the doctors failed to consistently apply a single objective criterion when coding the text, how can the less trained professionals be trusted to consistently apply a large set of subjective criteria when evaluating qualitative data?

3. How worried should you be when a market researcher is analyzing your focus groups? A typical focus group holds about 12,000 words. The data in this study included 18 threads. An average thread consists of about 5 postings with about 120 words each. These numbers suggest that the data in this study included 10,800 words; less than a single focus group. In contrast, a typical market research study consists of 4-8 focus groups, or 4 to 8 times more text. So, if the experts in this study failed to show consistency with a volume of data equivalent to a single focus group, what are the chances that a market

researcher will show consistency with a much larger dataset?

4. How worried should you be when a human resource manager is analyzing a pool of candidates? A transcript of a one hour interview holds about 6,000 words (when hiring middle and top managers, the interviews might take a whole day with an order of magnitude more words). When interviewing a few candidates, the total data may include 30,000 or more words (for 5 candidates). So, if the experts in this study failed to show consistency with a volume of data equivalent to a two interviews, what are the chances that a human resource manager will show consistency with a much larger dataset?

5. How worried should you be when an investment analyst is analyzing some companies for you? An annual report might include tens of thousand of words. For instance, the IBM 2004 annual report is 100 pages long and includes more than 65,000 words. So, if the experts in this study failed to show consistency when analyzing a dataset that holds less than 15% of the data included the annual report of a single company, what are the chances that an investment analyst will show consistency when analyzing a much larger dataset (such as the annual reports, financial statements, and press releases of a few companies)?

6. In this study pairs of doctors assigned different codes to the same question or answer. For instance, one doctor labeled an answer with "A = evidence based, excellent;" while another doctor labeled the same answer with "E = false," or even "F = possibly dangerous." Who is right? After all this is medicine and both cannot be right. Who should you believe? And what should you do as decision maker? If you believe that the first doctor is right, you should regard the response as great advice and follow its directives. If you believe that the second doctor is right, you should run for you life. Now, if such great experts failed to convince us that they can process a small dataset correctly, or at least consistently, how can we trust professionals when they say that they can?

Summary:

The first stage of most decision making in business is gathering data. In most cases the information is collected in the form of words. Once the words are available, the professionals who gather the data perform an analysis of these words, and present the results to the decision maker. As the study by Craigie, et. al., suggests, these professionals, most frequently, will fail in their analysis of qualitative data, and produce results which will prevent the decision maker from making the right decision.