QMETRICS™ - A NEW ONLINE QUALITY STANDARD

As online research has become the dominant mode of research, sample quality issues have finally reached their appropriate importance. We have developed an extensive database of 20 online U.S. and over 70 global panel sources using a standard questionnaire instrument. This has allowed us to explore techniques for monitoring and potentially controlling data-source quality. This paper outlines practical procedures to ferret out survey data that represents quality issues.

In online research where we function in a non probabilistic environment, it is at times difficult to identify and define problematic respondents. Much discussion has been centered on respondent behaviors that are considered undesirable; speeders, professionals, and satisficers have all been given considerable space in the market research literature. We propose here a method of blending these behaviors into a single index to assist researchers in rejecting the 'worst' respondents. These respondents provide us with information that does not reflect the balance of the data collected and thus represents a quality challenge. We must arrive at strategies that identify behaviors that bias our data.

Predicting "problem" respondents, as well as identifying them, is difficult. Most researchers approach projects with a sense for how the data will "fall out." However if the data always came out as expected we would not need research at all. It is the variability of data that makes life a challenge in the research profession. We are trained to detect change. It is this same variability that masks respondents who are providing us with spurious data because they are not inclined to take part in the process or are incapable of doing so.

Through digital fingerprinting we can screen out individuals that are duplicate survey takers prior to taking a survey. This type of screening is a growing industry procedure whose importance is on the rise. Respondent duplicates can be removed from the results using monitoring software in online surveys[1]or simply tagged for future consideration. For those who pass the test of de-duplication technology, finding a single metric that identifies other respondents who provide questionable data is not feasible. What we can do is identify different "troubling" responses and combine them into a single metric. If a sequence of questions targeting "troubling" behaviors is treated as screening questions, removing those who failed a single criterion would result in rejection of a large fraction of the total sample. As we will see later, this might be as much as 64% of the respondents[2]. This can eliminate almost all of the "problematic" respondents but in turn it will also remove far too many acceptable ones. More troubling, however, is that this process will most likely produce a biased sample. The integrity of the sample requires that its diversity be maintained. By being over critical, we could only expect to have reduced this diversity and produce results that would not reflect the true population.

In a recent white paper, Melanie Courtright and Denise Brien[3] have proposed the use of an exclusion decision rule based on a number of measures of "troubling" behavior. Their rule involved removal of respondents identified as having violated somewhat more than half of their

---

[1] The technique of fingerprinting is used to screen out duplicates. This involves measuring the characteristics of the respondent browser. Software then checks to see if this is a duplicate.

[2] MarketTools indicated in a recent white paper that almost 25% of their panel did not qualify under their screen criteria approach and were considered fraudulent, [Michael Conklin, *What Impact Do "Bad" Respondents have on Business Decisions*, published by MarketTools (2009)]. This is probably highly unlikely for all of them to be fraudulent.

[3] Courtright, M. and D. Brien *The Devil is in the Data* published by DMS Quirk's. April, 2009.

criteria. The procedure that is proposed here follows that line of thought of excluding respondents who have violated 3 or more out of six criteria.

QMETRICS™

A quality index was defined for each respondent by the number of errors and indicators of aberrant behavior. The index goes from zero for those without error to six for those who appear to get everything wrong. These can then be grouped into quality segments. Following the convention by Courtright and Brien[3] four segments are defined. Those having 3 or more errors are designated as the "Worst" segment. Those without error are considered "Ideal". Those with one error are designated as "Typical" and those with two errors are designated as "Imperfect". The results for a particular panel from our database are shown on Figure 1**Error! Reference source not found.**. For ease of presentation we will designate respondents rated by this Q-Metric™ standard as "QM" respondents.

These assignments could be merely random. That is, error and aberrant behavior could be just an event and not associated with other expected troubling behavior. Figure 1 also shows the expected distribution assuming these errors were purely random events. Note that for the second segment (typical) the frequencies are the same or at least within statistical precision. However, the first (ideal) and the last two segments (imperfect and worst) the frequencies are significantly different indicating that this is not simply a single independent random process, but that the likelihood of getting things repeatedly wrong is not random.

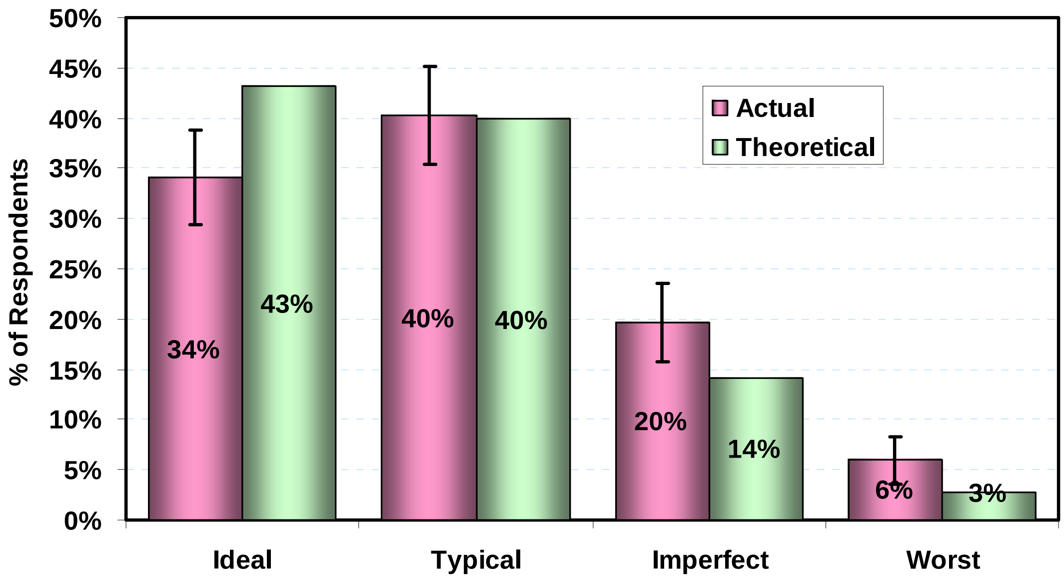

Figure 1, Frequency of Quality Segments

However, some panels may be random. Different panels appear to indicate different degrees of randomness in their quality structure. Below is the distribution of panels by the degree that the distributions of quality segments appear to be random[4], Figure 2. A value of 100% corresponds to a statistical agreement between the quality segment distribution and that which would be

---

4 A Chi Square (χ^2) test is used to compare distributions.

predicted for a completely random and independent process. On the reverse side, a value of zero corresponds to no likelihood that the two distributions are the same. Notice the broad range of values. Some panels appear to be described by a random process. But far more do not.

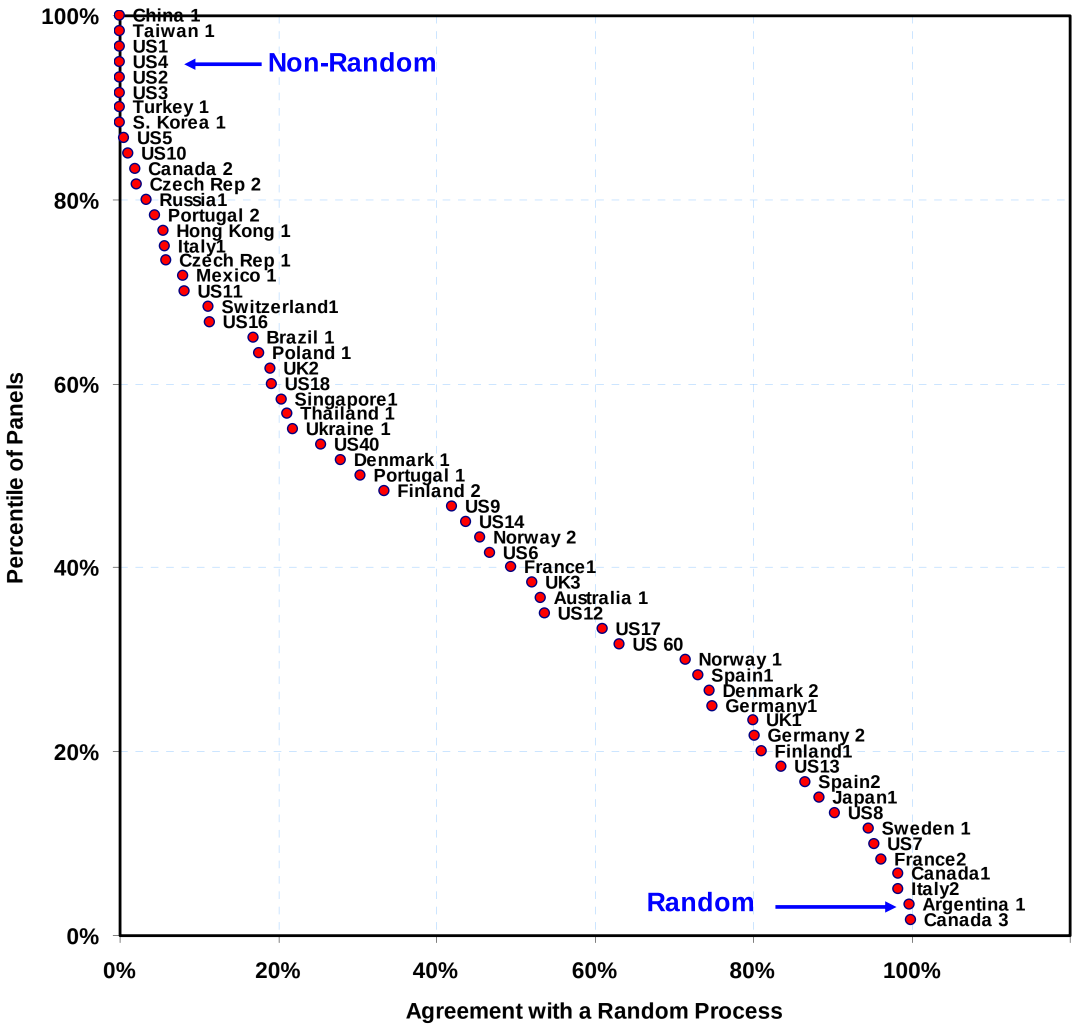

Figure 2, Agreement with a Random Process

We would expect particularly for non-random panels that the "worst" respondents would show distinctly different characteristics than the other segments. This is shown in terms of average values across questions in the survey. Figure 3, shows this result. Note that there is only a small, statistically insignificant difference among the three lower error segments while the "Worst" segment stands out, showing a significantly different value[5].

[5] Courtright and Brien[3] show this distinction among quality segments in terms of specific respondent behavior as well as attitudes.

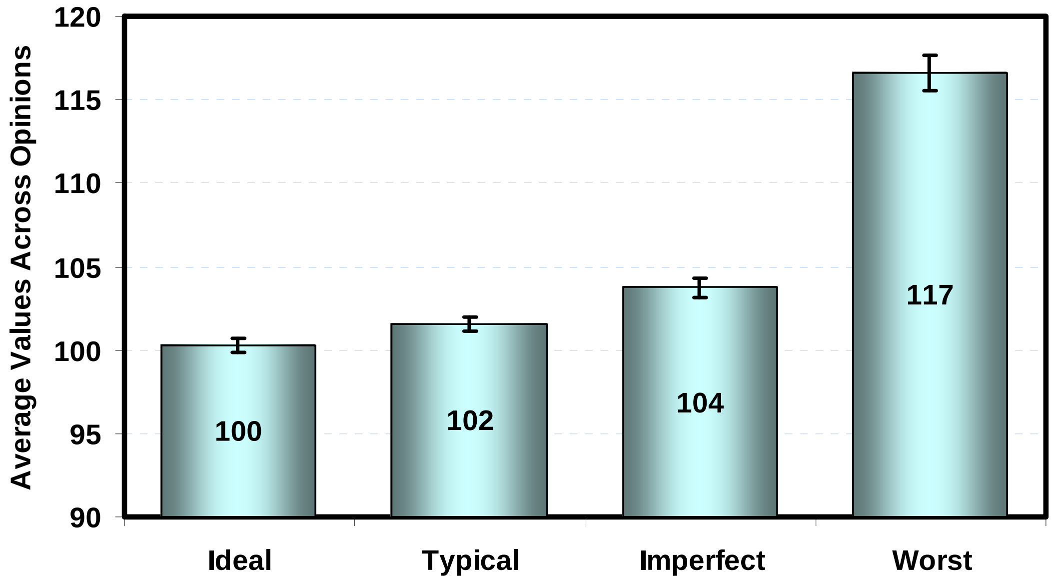

Figure 3, Average Values of Opinion by Quality Segments

THE METRICS BEHIND QMETRICS™

We identify six types of "troubling" behaviors. Three of these behaviors are considered 'question execution errors' while the other three are referred to as 'aberrant.'

Errors in execution are responses that we believe are wrong. They either reflect intent, confusion by the respondent or inattentiveness. They are captured as a metric by specific questions within the questionnaire. We refer to them as "failure to follow instructions" and "inconsistent responses".

Failure to follow instructions occurs when respondents ignore a directive to enter pre-determined answers. In our test questionnaire inconsistent responses are based on paired questions in reverse order. Those who rate them either both high or both low would be judged as having provided inconsistent responses. The standard questionnaire we have employed in our research on research contains one measure of failure to follow instructions and two measures of inconsistency (see Addendum).

Aberrant respondent behavior consists of a number of characteristics that while they are not consistently wrong, they are possible indicators of potential problems. These include professionalism, speeders, and straight-liners.

We refer to *professionals* as frequent survey takers or those individuals belonging to a large number of online panels. This measure is captured by specific questions in the questionnaire. Although the responses provided by frequent responders are different from the balance of the respondent population, difference is not in and of itself proof of problematic data.

Speeders are those respondents that complete the survey in a very short time. It is believed that this indicates inattentiveness in execution or at least the potential for less than fully considered answers. However, in these data there is an indication that cultural differences can speed up the process and that speeders are better educated than the average respondent.[6] The lowest ten percentile point[7] was chosen for the transition point defining speeders, as shown below, Figure 4. In more complex questionnaires, speeding measures will have to reflect skip out patterns.

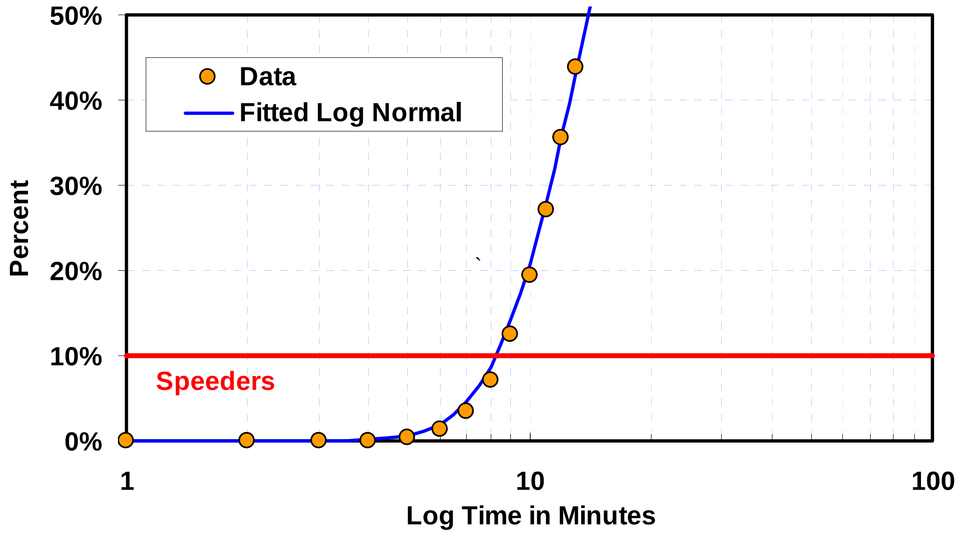

Figure 4, Defining Speeders, Distribution of Execution Times

Straight-liners are survey participants who have little variation in their responses. There are some respondents who have no variation. But many more, however, indicate only very little variation. That variation is computed as the standard error around a set of similar questions. In the case of the standard questionnaire, this was based on 31 opinion questions. The standard errors for respondents varied between zero to over three on a seven point scale. Having a standard error below 1 was used to define straight-liners[8], as shown in Figure 5.

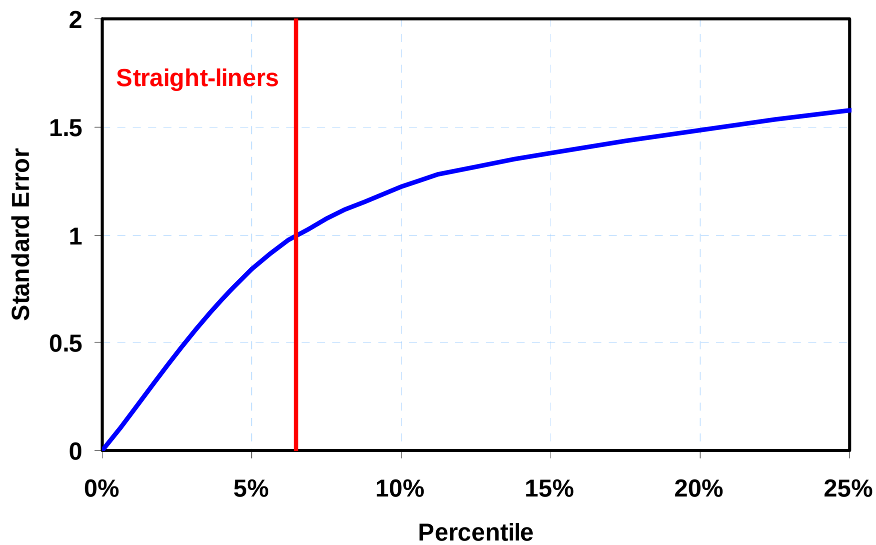

Figure 5, Defining Straight-liners, Distribution of Standard Errors

---

[6] Gittelman, Steven and Elaine Trimarchi. On the Road to Clarity: Differences between sample sources. Casro Panel. February, 2009.

[7] The ten percentile point also represents a deviation from the exponential portion of the distribution.

[8] The point appeared to separate two almost linear portions of the distribution.

VARIATION ACROSS DATA-SOURCES

The identification of the worst QM respondents can be used to improve results by excluding them from survey data. On average here they are a relatively small frequency and their exclusion should not affect the integrity of the rest of the sample. However, their frequency can differ widely among panels. In Figure 6, the distribution of QM segments is shown across 17 United States online panel data sources. Wide differences exist between the panels and the QM segments. It should be noted, that these panels have vastly different structures, management conditions, and respondent sources. It is not surprising that they would also differ in their incidence of errors.

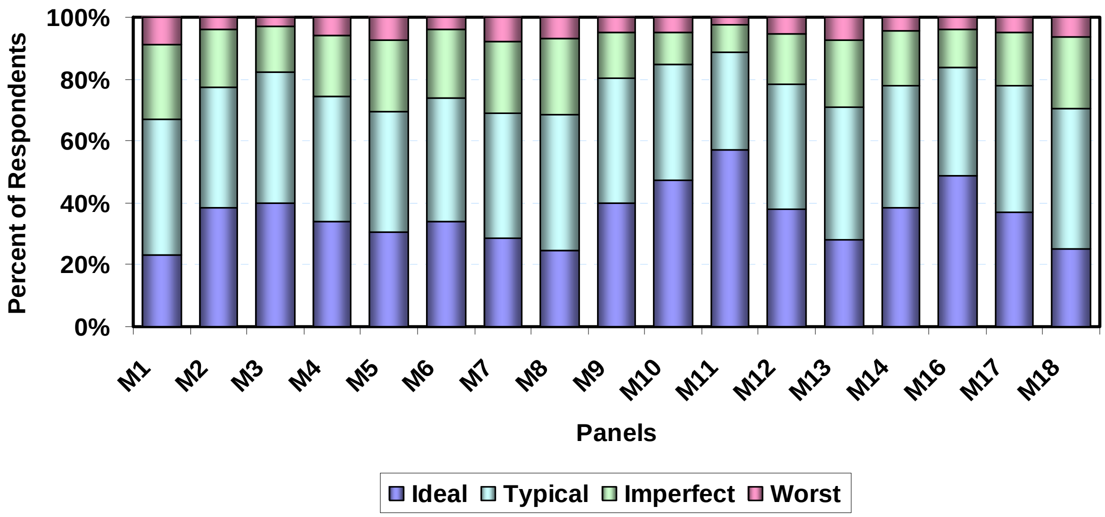

Figure 6, Distribution of Quality Segments by U.S. Data Source

While the distribution is interesting, the focus should be on the "Worst" segment. This is the QM segment containing the respondents that we might wish to exclude. In this respect, it is a composite measure of panel quality. This is shown on Figure 7. Note the variability between the panels – a low of 2.4% up to 8.7% with the average being 5.4%, a fourfold difference.

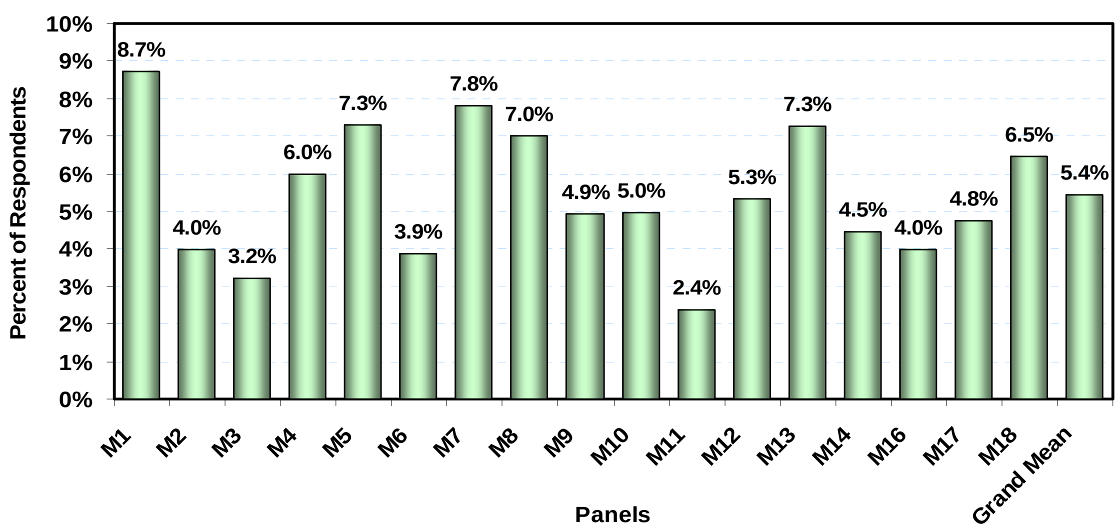

Figure 7, Frequency of Worst QM Segment Respondents by U.S. Data Source

PRACTICAL ISSUES

Market research practitioners have gravitated to the web for its obvious advantages of speed and cost. Quality issues lagged until about three years ago and have now become an important driver. Elimination of problem respondents has been a piecemeal practice usually relying upon the removal of one designated group or the next. Combinations of behaviors have rarely been entertained.

The cost of keeping respondents who contribute questionable data is incalculable. However, the cost of removing them is clearly measurable and consequential. The measures suggested here are in our opinion, minimal. We doubt that quality panel sources will penalize researchers who employ a method which on average forces replacement of 5.4% of all survey takers.

We are confronted by a cultural shift in our research. Those who use our data to make business decisions cannot afford spurious conclusions that are driven by dubious respondents.

There are direct costs that lie in the questionnaire real estate that the quality metric questions require. Measuring speeders and straight-liners is nominal in that it is a programming function that is not represented by questionnaire wording although there are some design considerations. However, the four questions that measure the two inconsistencies, one that captures the frequency of survey taking and another that serves as a “trap”, could amount to about ninety seconds.

CONCLUSION

The time has come to coordinate quality metrics. Here we provide a metric that requires six measures based on the analysis of a broad spectrum of sample sources. The differences between panel sources should alarm researchers sufficiently to employ techniques of this sort. Combined with digital fingerprinting to remove duplicate respondents, these behavioral measures are a beginning. One might think of them as a sliding scale. We have set the bar at three failures. For now, we hope that practices such as this become standard.

As we continue to grow our database of respondents around the world, we will be able to ground our quality metrics to a larger scope. For now the data in the United States has provided us with this launching point. Similar metrics can be applied in all global markets.

Please find the Q-Metrics™ Questions and Contact Information Below

We look forward to helping you make Q-Metrics™ your company standard.

Addendum:

The 6 quality metrics are as follows:

1. Inconsistency – Happy with standard of living / not happy with standard of living (see question below)
2. Inconsistency – price over brand / brand over price (see question below)

**one of the attributes should appear early in the survey and the opposite attribute later in the questionnaire (Eg – "price over brand" in the beginning of the survey and "brand over price" toward the end of the survey)

3. Speeders
4. Straight-liners
5. Failure to Follow Instructions
6. Participate in online surveys daily (see question below)

Q-METRICS™

The questions for the quality metrics are the following:

INCONSISTENCY #1

Please indicate your level of agreement or disagreement with the following statements.

I am perfectly happy with my standard of living	7	6	5	4	3	2	1

I am not really happy with my standard of living	7	6	5	4	3	2	1

INCONSISTENCY #2

How much do you agree or disagree with the following statements?

SCALE:
(7 Strongly Agree to 1 Strongly Disagree)

- Price is more important to me than brand names
- Brand names are more important to me than price

PARTICIPATING IN ONLINE SURVEYS DAILY

On average, how often do you participate in online surveys? Please consider only those with 6 or more questions.

Practically every day
Every few days
About once a week
About once a month
About every few months
About every six months or less

For additional information please contact:

Harvesting Quality Data...since 1979

Elaine Trimarchi | Executive Vice President – PRC Certified
631.277.7000 | Cell 631-664-1308 | etrimarchi@mktginc.com
200 Carleton Avenue, East Islip, NY 11730
www.mktginc.com

Steven Gittelman, Ph.D. | President
631.277.7000 | Cell 631-466-6604 | steve@mktginc.com
200 Carleton Avenue, East Islip, NY 11730
www.mktginc.com

