SPSS establishes India subsidiary, distributorship in Spain

Within six months of establishing a new French subsidiary, SPSS is once again expanding its international operations. SPSS India Private Ltd., formerly a branch of SPSS UK Ltd., is now a fully owned subsidiary, and SPSS has signed an agreement to establish a distributorship in Spain.

Under the agreement between SPSS and SPSS Hispanoportuguesa S.L., the Spanish distributorship has exclusive rights to sell SPSS software and support in Spain. SPSS is scheduled to release a Spanish version of SPSS for Windows later in the second quarter of 1994 in support of the Spanish operation and Latin America markets.

SPSS for Windows speaks Japanese

SPSS for Windows now supports the Kanji/Kana Japanese writing system. The Japanese version of SPSS for Windows has a graphical user interface with Japanese menus, dialog boxes, glossary of statistical terms, help and all online documentation. Since its recent release, initial orders for SPSS 6.0 for Windows in Japan have exceeded expectations.

"Regardless of where our customers are in the world, we continue to deliver on our promise of 'Real Stats. Real Easy,'" said Jack Noonan, SPSS president and CEO. "Local-language versions, such as the Japanese version, break down yet another barrier to ease of use - language. By offering local-language versions, we believe we get close to our customers, making it easier for them to offer suggestions for improvement of our products, statistical reporting, and graphical functionality."

1st quarter sets new records

SPSS Inc. reported record financial results for the first quarter of 1994, stating it was the best first quarter in the company's 13-year history. Management attributed the success during the period to a 39 percent increase in desktop software.

New income for the quarter totaled a record $358,000, an increase of 112 percent over the prior year. Earnings per share increased 25 percent to 15 cents from 12 cents in last year's first quarter. In March, SPSS Inc. agreed to buy Japan Systems Engineering Corp.'s 50 percent interest in SPSS Japan. SPSS will now report all revenues and expenses from Japan, previously treated as a joint venture. On a pro forma basis, earnings per share for the quarter increased 25 percent.

Revenue for the first quarter of 1994 were $11,463,000, an increase of 16 percent over first quarter 1993 revenues of $9,883,000. Large systems revenues were off as expected, but were supported by a 30 percent increase in UNIX sales. Operating income for the quarter increased 25 percent from the same period last year.

Send email to SPSS or Keywords

In the last issue of Keywords, we told how you can access technical support and SPSS company information through the Internet (support@spss.com) and through CompuServe (GO SPSS). Now, there are two more email addresses through which you can communicate with SPSS: sales@spss.com and keywords@spss.com.

To request the latest product information and pricing, send an email to sales@spss.com. If you want to submit a story idea or technical question to Keywords, or if you want to request copies of back issues, send email to keywords@spss.com. Both mailboxes are checked daily, so you should get a response within 24 hours.

How to determine appropriate survey sample size

by Pamela Narins
Manager of Market Research

Welcome to "Survey Samplings," a new Keywords column that will provide hints and tips for those of you doing survey research. This column will provide step-by-step instructions which should be of use to both the first-time researcher looking for practical answers and the seasoned survey veteran wanting to brush up on some of the basics.

In this first installment, we will talk about one way to determine the appropriate sample size for a survey. Many of you may have seen formulae for calculating the sample size or sampling error for a research project, but it can be difficult to see how to translate these equations into practical applications.

One warning: There are many considerations that come into play when designing a research study. Compromises are always being made on sample size, acceptable error levels, sources of bias and the like, based on the availability of resources (time, money, personnel). There is no one right way to decide what is acceptable. The following method for determining sample size is a conservative approach, and so should work for many people in many situations. It assumes a simple random sample, a large sample approximation, and that typical sources of error such as nonresponse, poor administration methods, and highly biased results are trivial.

The Equation

Here is one way to look at the equation needed for finding the appropriate number of cases to survey for a categorical variable.
\[
\frac{(Py)(Pn)}{\text{Std Error}^2} = N
\]

The Standard Error, when multiplied by the coefficient of 1.96, can be thought of as the error term for the entire sample for this item, and Py and Pn represent the proportion of respondents responding "Yes" or "No" to that question (more on that in a minute). What we need to do is assign values to both Py, Pn, and the standard error term, and then solve for N, which is the number of cases.

The Py and Pn terms
Py and Pn represent the proportion of people responding to each of the categories in a dichotomous variable (a dichotomous variable is one which has only two response choices, such as "Yes" and "No" or "Male" and "Female"). Even multiple-category or continuous variables can be thought of as dichotomous. You can look at any one response category (such as "Accounting" in a variable categorizing the department in which a respondent works), and count the number of people who are in Accounting and the number of people who are not in Accounting.

If you looked at data that had already been collected, you could see that 35 percent of your respondents said "Yes" to Accounting and 65 percent said "No" to Accounting. It would then be easy to calculate (Py)(Pn) for the Department variable: it simply would be (.35)(.65) or .2275.

The trick here, of course, is that we do not have real data to examine; we are trying to estimate the probable values of Py and Pn. Additionally, we are attempting to select values for Py and Pn that would work for every question in the survey, not just the Department question. Since both the distribution of responses and the number of missing cases will vary by question, this seems an almost unachievable task.

The easiest way to arrive at a number that would work for all the questions is to be as conservative as possible. If we were to use the analogy of an election between two candidates, it is easy to see that the outcome of a close election (where each candidate is expected to receive about 50 percent of the votes) would be harder to predict than a landslide, where the split in votes might be 80/20. Therefore, a close election might require a larger sample than a landslide.

If we know nothing about the types of responses we will be getting in our survey, it is always safest to maximize the variation, by assuming a 50/50 split in responses across questions. Thus, the computation of (Py)(Pn) becomes simple; it is (.5)(.5), or .25. So, our equation now looks like this:

\[
\frac{.25}{\text{Std Error}^2} = N
\]

So our equation now looks like this:

\[
\frac{.25}{.000657} = N
\]

And finally, we can solve for N:

\[
381 = N
\]

What this means
Given this calculation, we find that we would need 381 cases to be 95 percent confident of our overall results, within a range of plus or minus 5 percent. That is to say, N = 381 is the sample size to use if we want observed percentages to be within plus or minus 5 percent of the unknown population parameter. This will be true 19 out of 20 times, though in a given instance, it either is or is not true, but we don’t know which.

You can see that these numbers do not change. If you wanted to always have a sampling error of plus or minus 5 percent and always have the most conservative estimate of your response distribution (that is, assume a 50/50 split in responses to your questions), your sample would always be 381.

You could solve for more or less variation of Py and Pn, or you could solve for a different degree of certainty (let’s say plus or minus 1 percent), and know ahead of time how many people you would need to survey.

Remember that you can also use a variation of this formula for determining the sampling error for data you have already collected. In this case, you would solve for the standard error and plug in the values of Py, Pn and N. The equation would look like this:

\[
\sqrt{\frac{(Py)(Pn)}{N}} = \text{Std Error}
\]

Look for the next issue of Keywords for more hints! If there is a specific survey research topic you would like addressed, please send your comments to the Keywords address listed on the inside front cover.