Statistical Alternatives

Editorial Note:

William M. Bailey, Ph.D., an applied statistician, and market analyst, is principal of WMB & Associates. The company's Web site is http://home.earthlink.net/~statmanz

Published In Quirk's Marketing Research Review, May 2002

The Situation

"I can't afford the research plan you advise! Is there a way we can do fewer surveys but still get usable and as reliable results?" Does this sound familiar? In times like recent years, economic conditions affect strategies of consumer research, a simple fact, and one that can't be ignored. More and more clients are trying to find ways to cut cost while at the same time deliver to business objectives.

As market researchers, we tend to focus on cross tabulations that offer paired tests of proportions and generally take the results right to the portion of the paper that details the statistical results. This is not necessarily intended to be a criticism but the "way we 'typically' do consumer research." While this works in many cases, this author is finding that clients are asking somewhat different questions: "How do these two products differ in comparison to these other two?" "Is there a difference in opinion by product within gender or age or…?" And then there is the "Tell me how these product features rank as they apply to the respondent's overall opinion of my company." As can be seen, these questions begin to enter into the rim of data evaluation beyond the "basic" arena.

The preferred research plan is to interview a sufficient number of consumers to make the results "statistically reliable" at the 90% or 95% level of confidence with a certain margin of error, e.g., ±5 percentage point. Why? Because that is what we have always done! Depending on how one sets the constraint parameters, this works out to be from 250 to 350 completed interviews at the base level of analysis then we work up from there. With this base we can apply "standard analysis tools" such as paired t-tests, analysis of variance, and factor or regression analysis with "reasonable" comfort. Further, for this response base there usually is marginal violation of the implied assumptions: the data approaches a normal distribution and homogeneity of variance. But is this always the case? Depending on the response scales used, more likely not, there is some violation we could overlook. I am not suggesting that we have done a bad job just that we haven't done an appropriate job for the data's characteristics.

Back to the statement "I can't afford the research plan you advise! Is there a way we can do fewer surveys but still get usable and reliable results?" No worry; there are alternatives available that are often overlooked. These approaches fall into the general category of "sturdy or distribution-free" statistics or more specifically nonparametric statistics.

Sturdy Statistics

Most market researchers automatically use procedures that assume that the measurements are drawn from a normal distribution and then proceed to test hypotheses on parameters such as the mean or the variance [usually the standard deviation which is the square root of the variance]. Useful tests include but are not limited to the *Student's t* or the Z statistic, various forms of regression analysis, and/or analysis of variance to help understand a study's result and/or differences between product or control/treatment sets. These tools are a part of what is called *parametric* statistical tests.

While some of these statistical tests do work well in many cases even if the assumption of normality is violated, extreme violations of this assumption can affect the result's interpretation. There are technical reasons behind this, such as the affect of violating the assumption of normality is to decrease the **Type I error** [a conclusion is drawn that the null hypothesis is false when, in fact, it is true], but beyond the scope of this paper's intent.

Should a violation of an assumption be realized or as is often the case, the sample size desired for the analysis base is small, e.g., under 20 or 30 observations when "traditional" statistical tests become questionable, there is a collection of tests that do not depend that much on the precise shape of the distribution. This class of statistical tests base themselves on the *signs* of differences, *ranks* of measurements, and/or *counts* of objects falling into categories. Such methods may not rest heavily on the specific parameters of the distribution, and for this reason are called *nonparametric* or *distribution-free* tests that do not make any or as stringent assumptions about the distribution from which the numbers were sampled.

However, the term "nonparametric" is somewhat misleading, since these statistics do in fact deal with parameters such as the median of a distribution or the probability of success ***p*** in a binominal distribution. The main advantage to many of the methods described herein defend themselves against distribution outliers and "off normal distributions" and failures of assumptions. Statisticians use such adjectives as "robust," "resistant" and "sturdy."

Specifically, and more importantly, sturdy statistical techniques provide comparable test results to traditional tests when the samples are from asymmetric or skewed distributions. Here the term **power** is usually introduced. While there are transformations available such as taking logarithms or square roots of the data to bring them more inline with appropriate parametric assumptions, sturdy, or distribution-free tests are a worthwhile alternative.

Further, sturdy statistical methods are useful in cases when the researcher knows nothing about the parameters of the variable of interest in the population (hence the name *nonparametric*).

Comparative Grid

This section provides a comparison between tests in these two classifications [called parametric and nonparametric in the table] based on some popular study scenarios. It is not meant to be all-inclusive, further, there is always an affordable "interpretative debate" so the reader is reminded of this.

Most parametric tests have their nonparametric analogues. In other words, nonparametric tests exist for most situations a market analyst commonly uses: 2 independent groups, 2 matched groups, and multiple groups. The primary difference is that the data are no longer interval; instead it is ordinal (or is treated as ordinal). Below is a table that summarizes several 'crossover' tools. It offers a very simple comparison between several parametric tests with their analogues.

Parametric Test	**Nonparametric Tests**
Independent t-Test	Mann-Whitney Median
Matched Pairs t-Test	Wilcoxon Sign Test
One-Way ANOVA	Kruskal-Wallis
Pearson's Correlation	Spearman or Kendall

While nonparametric tests make fewer assumptions regarding the nature of distributions, they are usually less powerful than their parametric counterparts. However, in cases where assumptions are violated and interval data is treated as ordinal, not only is a nonparametric tests more proper, they can also be more powerful.

And What Are They

This section highlights the applicability of the nonparametric tests noted above. For more detail information the reader is directed to a statistical resource, the Internet, or software packages such as but certainty not limited to SPSS, SAS, and Prophet. [The author is not endorsing any of these packages, and no "rank" order is implied.]

The Mann-Whitney U test is the most popular of the two-independent-samples tests. It is equivalent to the Wilcoxon rank sum test and the Kruskal-Wallis test for two groups.

Mann-Whitney tests that two sampled populations are equivalent in location. The observations from both groups are combined and ranked, with the average rank assigned in the case of ties. The number of ties should be small relative to the total number of observations. If the populations are identical in location, the ranks should be randomly mixed between the two samples. The number of times a score from group 1 precedes a score from group 2 and the number of times a score from group 2 precedes a score from group 1 are calculated. The Mann-Whitney U statistic is the smaller of these two numbers.

The Median Test tests whether two or more independent samples are drawn from populations with the same median using the chi-square statistic. This test should not be used if any cell has an expected frequency less than 1, or if more than 20% of the cells have expected frequencies less than 5.

The Wilcoxon is used with two related variables to test the hypothesis that the two variables have the same distribution. It makes no assumptions about the shapes of the distributions of the two variables. This test takes into account information about the magnitude of differences within pairs and gives more weight to pairs that show large differences than to pairs that shows small differences. The test statistic is based on the ranks of the absolute values of the differences between the two variables.

The Sign Test is designed to test a hypothesis about the location of a population distribution. It is most often used to test the hypothesis about a population median, and often involves the use of matched pairs, for example, before and after data, in which case it tests for a median difference of zero. In many applications, this test is used in place of the one sample t-test when the normality assumption is questionable. It is a less powerful alternative to the Wilcoxon signed ranks test, but does not assume that the population probability distribution is symmetric. This test can also be applied when the observations in a sample of data are ranks, that is, ordinal data rather than direct measurements.

The Kruskal-Wallis test is used to test the null hypothesis that all populations have identical distribution functions against the alternative hypothesis that at least two of the samples differ only with respect to location (median), if at all. It is the analogue to the F-test used in analysis of variance. While analysis of variance tests depend on the assumption that all populations under comparison are normally distributed, the Kruskal-Wallis test places no such restriction on the comparison. It is a logical extension of the Wilcoxon-Mann-Whitney Test.

The Spearman Rank Correlation Coefficient bases itself on the rank ordering of each variable. It may also be a better indicator that a relationship exists between two variables when the relationship is non-linear.

Kendall's tau-b is a measure of association for ordinal or ranked variables that take ties into account. The sign of the coefficient indicates the direction of the relationship, and its absolute value indicates the strength, with larger absolute values indicating stronger relationships.

Conclusions

While, in most cases, we are able to be “traditional,” there are alternatives should the situation be applicable. Regardless, the analyst has a basic responsibility: Validate, validate, validate then analyze and interpret with confidence.