RESEARCH ● *Play by the rules*

One Chi Square beats two Z tests

By STEPHEN J. HELLEBUSCH

The statistical testing procedure built into software that produces marketing research data tables is widely used, and likely widely misused.

Some years ago, as a project director for a large research firm, I spent hours on data tables, conducting t tests, Z tests (approximation to the binomial), Chi Squares and ANOVAs (Analysis of Variance) by punching required summary data into a programmable calculator and recording the results by hand.

SPECIAL REPORT

Even then, I wondered why no one had found a way to conduct the tests automatically and print the results along with the tables.

Three simple rules help researchers determine which test to use and when:

- "Two groups" is not the same as "three or more groups";
- percentages are different from means; and
- matched samples are not the same as independent samples.

Back then, project directors knew that because of the third rule, the calculator method worked only for testing independent groups, those made up of totally different people—for example, men vs. women.

For dependent groups, we transferred the data to a statistical package to conduct the appropriate tests. Dependent groups are composed of the same people, such as when all respondents rate a product pre- and post-exposure to an advertisement, and the pre- and post-ratings are tested. For testing percentages, researchers use Chi Square and Z tests between independent groups, with Chi Square for three or more independent groups and the Z test for the two-group case. (I learned later that many refer to this Z test as a "t test" because, with the base sizes typical in marketing research, the Z and t distributions are the same.)

Similarly, researchers use t tests and ANOVAs (with appropriate post hoc follow-up) when means are tested. The t test is for the two-group case, and the ANOVA is used for three or more groups. These procedures were tedious but made sense based on statistical knowledge.

Measure	2 Groups	3+ Groups
Percentages	Z test	Chi Square
Means	t test	ANOVA

Years later, I learned that computer software programs could run data tables to include automatic statistical testing. I was thrilled, especially since I was still writing analyses from data tables near the way I had in those early days. But I also learned that the automatic testing was wrong in many cases. It still is, and it is not hard to explain why: Automatic statistical testing in survey programs regularly violates the rule that two groups is not the same as three or more groups. When percentages are tested and three or more groups are involved, multiple Z tests erroneously are used rather than the Chi Square test. And when three or more means are tested, multiple t tests are used instead of the appropriate ANOVA with follow-up.

Violating this rule has an interesting effect: The statistical reason for using a different test with three or more groups is to keep the confidence level at 95%, or whatever level you think you are using. With multiple Z tests, the confidence level is no longer what the computer says it is. I do not know what it becomes, but it is lower than the level desired.

The situation is the same when three or more means are tested using multiple t tests instead of the ANOVA. For example, say we have a purchase intent "top two box" percentage—the percent who said they "definitely" or "probably would buy." We want to learn whether the percentage of males saying they "definitely" or "probably" would buy differs from the percentage of females responding "definitely" or "probably." We interviewed 200 people—80 males (column B) and 120 females (column C). The percentage among males is 50%; among females it is 33%. The data tables have a "C" by the male percentage, indicating that 50% is significantly higher than 33% at the 95% confidence level. This is accurate and useful. The automatic Z test operated at the 95% level, and with only one Z test, these results are statistically sound.

Now, suppose the study includes a set of three age groups. We want to know if significant differences exist at the 90% confidence interval based on age, as represented by the three groups. With 63 18- to 29-year-olds (column D), 66 30- to 44-year-olds (column E), and 71 45-year-olds and over (column F), the percentages are 50%, 40% and 34%, respectively.

The automatic testing shows an "F" by the 50% in Column D, which indicates that 50% is statistically significantly higher than 34% at the 90% confidence level. But the test addresses a question we did not ask: We want to know whether the top two box purchase intent scores vary with age. With the three-group case, the program completed three different Z tests: D vs. E; D vs. F; and E vs. F, ignoring the rule that two groups is not the same as three or more. The appropriate test, Chi Square, would test across all three percentages at once to determine whether top two box purchase intent changes with age. In this admittedly contrived example, no statistically significant difference emerges in the top two box measure across the three age groups at the 90% confidence level. Think of it as using a test that takes all of the information available into account (Chi Square) vs. using a test that only looks at part of the information several times (Z test).

The consistent error of running multiple Z and t tests using the automatic option in survey research programs probably will not cause earth-shattering mistakes in marketing research decisions, and I doubt anyone could stop this standard misinterpretation at this point. But one can ask those producing the data tables if the appropriate test is available. In many cases, it is—though it is not the one typically used. Using tests that are automatically wrong is one small quality-control feature that can undermine credibility and should be avoided if at all possible. ■

Stephen J. Hellebusch is president of Cincinnati-based Hellebusch Research & Consulting, Inc.

Reprinted with permission from Marketing News, published by the American Marketing Association, June 4, 2001, Page 11.