SAMPLE SIZES FOR MONADIC AND PAIRED COMPARISONS

The Analytical Group Inc.

15300 N. 90th Street • Suite #500 • Scottsdale, AZ 85260
+1.480.483.2700 • www.analyticalgroup.com

Dr. Albert Madansky
Vice President, The Analytical Group, Inc.
and
H.G.B. Alexander Professor Emeritus of Business Administration Graduate School of Business
University of Chicago

Periodically I have been asked to determine sample sizes for monadic and paired comparison studies. Usually in market research studies the correlation between responses in paired comparison studies is positive, and this positive correlation reduces the standard error of the difference of the pair of means, and so for the same sample size one gets better results from paired comparison studies. But how much better are the results? And how can much smaller can the sample size for a paired comparison study be than that for a monadic comparison study to get the same degree of accuracy?

It is to answer these questions that I have written this note. This note is in three parts. In the first part I consider the case where one's goal is to estimate the magnitude of the difference of the two means. The metric used in such a study is the width of the confidence interval of the difference between the two means. In the second part I consider the case where one's goal is to test the hypothesis that the two means are different. The metric used in such a study is the power of the hypothesis test where the null hypothesis is that the difference of the means is zero. In the third part of this note I illustrate my conclusions with an example from a study.

And here is the simple punch line. It turns out that when the standard deviations of the two populations are equal that in both these cases the result is identical: the ratio of the size of the paired comparison sample to the size of the monadic comparison sample is $1-\rho$, where ρ is the correlation between responses in the paired comparison.

1. <u>Estimation of difference of means</u>

When one compares sample means from two monadic random samples drawn from their respective populations, each of size m, then the standard error of the difference of the two means (se_m) is given by

$$se_m = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{m}}$$

where σ_1 and σ_2 are, respectively, the standard deviations of the populations.

When one compares sample means from a paired comparison of n independent observations drawn from the two populations, then the standard error of the difference of the two means (se_p) is given by

$$se_p = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{n}}$$

where ρ is the correlation between the responses to the two items in the paired comparison.

The ratio r of these standard errors is

$$r = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}(\frac{m}{n})} = \sqrt{[1 - \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}](\frac{m}{n})}$$

When the standard deviations of the two populations are identical, i.e., when $\sigma_1 = \sigma_2 = \sigma$, this ratio reduces to

$$r = \sqrt{\frac{(1-\rho)m}{n}}$$

The two confidence intervals for the difference in means will be identical when r=1, that is, when $n=(1-\rho)m$

2. Testing significance of difference of means

In what follows I will assume that the null hypothesis being tested is that there is no difference between means and that the alternative hypothesis is that there is some difference between means, but that we do not know which mean is the larger. Consequently, the test of the null hypothesis will be a "two-tailed test." For ease of exposition I assume that the standard deviations of the populations are known, and that the level of significance for the comparison between means is set at 0.05. That is, the probability is 0.05 of rejecting the null hypothesis of "no difference between means" when in fact there is no difference.

The form of the test of the null hypothesis with two monadic samples of size m is: reject H_0 when

$$\frac{|\bar{x}_1 - \bar{x}_2|}{se_m} \geq 1.96$$

where $\bar{x}_1$ and $\bar{x}_2$ are the sample means from the two populations. The form of the test of the null hypothesis with one paired comparison sample of size n is: reject H_0 when

$$\frac{|\bar{x}_1 - \bar{x}_2|}{se_p} \geq 1.96$$

What is critical in comparing the two data elicitation methods is their relative "power." The power of a test is the probability of rejecting the null hypothesis of "no difference between means" when in fact there is a difference.

The power of these tests are functions of the magnitude of the difference between the population means. When the difference of the population means is δ, then the power of the test of the null hypothesis using two monadic samples of size m is:

$$\Pr\{\frac{|\bar{x}_1 - \bar{x}_2|}{se_m} \geq 1.96\} = \Pr\{\frac{|\bar{x}_1 - \bar{x}_2|}{se_m} - \frac{\delta}{se_m} \geq 1.96 - \frac{\delta}{se_m}\} = \Phi(1.96 - \frac{\delta}{se_m})$$

where $\Phi(y)$ is the two-tail area of the standard normal distribution to the right of y. Also, when the difference of the population means is δ, then the power of the test of the null hypothesis using one paired comparison sample of size n is:

$$\Pr\{\frac{|\bar{x}_1 - \bar{x}_2|}{se_p} \geq 1.96\} = \Pr\{\frac{|\bar{x}_1 - \bar{x}_2|}{se_p} - \frac{\delta}{se_p} \geq 1.96 - \frac{\delta}{se_p}\} = \Phi(1.96 - \frac{\delta}{se_p})$$

The power of the two tests will be identical when $se_m = se_p$. When the standard deviations of the two populations are identical, i.e., when $\sigma_1 = \sigma_2 = \sigma$, the power of the two tests will be equal when

$$r = \sqrt{\frac{(1-\rho)m}{n}} = 1$$

that is, when $n=(1-\rho)m$.

3. Example

Usually in paired comparisons studies the correlation ρ between the responses is positive. From the above analysis we see that regardless of whether we are estimating a difference of means or testing a hypothesis about the difference of means, if the standard deviations of the two populations are equal then one requires a smaller, by a factor of 1-ρ, sample for paired comparison than for monadic comparison studies. If ρ=0 then the paired comparison study will be equivalent to a monadic study. Also, if ρ=1 then the responses to each of the paired comparisons are exactly linearly related, so a paired comparison study isn't called for. Thus, when one is planning a paired comparison study it is worthwhile to determine beforehand the correlation between the pair of responses so as to determine the saving in sample size over that of monadic studies attainable by the paired comparison study.

I illustrate this result with the data from a market research study. The study presented four advertising messages about a product category and, after presenting each message, asked the following five questions about them:

1) What is your **overall** reaction to this statement?
2) How **unique** or different is this statement from others you are familiar with about these products?
3) How **believable** is this statement to you?
4) How does this statement make you **feel** about brand X?
5) Based on the idea communicated in the statement, how **likely** are you to purchase brand X in the future?

Following are the correlations between pairs of messages for each of the questions:

message	overall	unique	believable	feel	likely
1 vs 2	0.347	0.466	0.493	0.468	0.591
1 vs 3	0.415	0.535	0.547	0.466	0.659
1 vs 4	0.447	0.500	0.611	0.492	0.660
2 vs 3	0.362	0.419	0.462	0.410	0.573
2 vs 4	0.375	0.384	0.469	0.425	0.597
3 vs 4	0.473	0.551	0.610	0.535	0.653
min	0.347	0.384	0.462	0.410	0.573
max	0.473	0.551	0.611	0.535	0.660

Note the higher correlations for the **likely** question than for all the others. If this is universally true, than a paired comparison study with its primary goal of ascertaining "likelihood to purchase" needs a much smaller sample size than does a pair of monadic studies.

Suppose one were to use a sample of 1,000 respondents to study each of these four messages monadically, for a total sample size of 4,000. Based on the above correlations, here are the sample sizes needed for paired comparison studies of each pairs of messages to obtain results with equivalent statistical accuracy.

pair	overall	unique	believable	feel	likely	max
1 vs 2	653	534	507	532	409	653
1 vs 3	585	465	453	534	341	585
1 vs 4	553	500	389	508	340	553
2 vs 3	638	581	538	590	427	638
2 vs 4	625	616	531	575	403	625
3 vs 4	527	449	390	465	347	527

So even if one fielded all six paired comparison studies with separate samples, one would only need a total of 3,581 respondents for the studies, or roughly 600 per paired comparison.