

How Big Should My Test Be?

*By Jim Wheaton
Principal, Wheaton Group*

Original version of an article that appeared in the October 1, 2001 issue of "DM News"

Recently, a veteran list broker recommended test quantities of as few as 5,000, to a direct marketer with prospecting response rates as low as 0.25%. Unfortunately, the resulting thirteen responders would have been far from adequate to read the results of the tests.

At about the same time, a highly respected direct marketing consultant commented that test list quantities should be large enough to generate at least 50 responders. Unfortunately, this assumption is simplistic in its perspective.

A review of all the concepts behind good direct marketing testing is beyond the scope of this article – things such as confidence levels and intervals, one versus two-tail tests, stratified sampling, power testing, finite population correction factors, alpha versus beta "misreads," and the interpretation of dollar versus response rate performance. Nevertheless, we will focus on a single formula to provide some groundwork for answering a question that I have been asked countless times as a direct marketing consultant: "How big should my test be?"

Unfortunately, the short answer is that, "It depends!" (Bear with me, however, because things will become clearer.) For a given expected response rate, there is no one test panel quantity that will be optimal for every direct marketer. The appropriate quantity will depend on factors such as: 1) the amount of money available for testing, and 2) the level of risk, the direct marketer is willing to assume, that the rollout response rate will be significantly different from the test rate.

However, I will outline how you can intuitively arrive at your own well-considered conclusions. To do so requires a two-part statistical formula that every direct marketer should commit to heart:

Part 1: (Expected Response Rate * (1 - Expected Response Rate) * Z²) / Precision²

Part 2¹: Answer to Part 1 / (1 + (Answer to Part 1 / Rollout Universe Quantity))

First, a few sentences on "Precision" and "Z":

Precision describes the degree of "plus/minus" uncertainty around a test panel response rate. After all, we can never know for sure, by examining a test panel response rate, what the "true" rollout rate will be.

¹ This is what's known as a Finite Population Correction Factor.

Many direct marketers consider Precision of 10% to be acceptable; that is, the “true” rollout response rate will be within 10% of the test panel rate a certain percentage of the time. A 1.0% test panel rate, for example, translates into a rollout rate of between 0.9% and 1.1%.

Understanding “Z” would require a statistics lesson. All we need to know for our purposes, however, is that it corresponds to the degree of Confidence that we have in the accuracy of our test panel response rate. For example, a given test panel quantity will result in Confidence that – say – 80% of the time a test panel response rate of 1.0% will translate to a rollout rate of between 0.9% and 1.1%.

Direct marketers would love to be very Confident with very narrow Precision. Unfortunately, this generally requires a staggeringly high investment in very large test panel quantities. Therefore, they’re faced with the difficult decision of just how much of an investment to make.

While there is no one answer that is correct for every direct marketer, general guidelines can be posited. We’ll reference the table below as we continue to explore this issue:

Desired Confidence	Corresponding Z
95%	1.960
90%	1.645
80%	1.282
70%	1.04

Many direct marketers are unwilling to live with Confidence of less than 80%. So, let’s go with this for now, combine it with a Precision of +/- 10%, and see what that translates to in terms of test panel quantity.

The one thing that we are missing is an expected response rate. Because so much testing is done on rental lists, let’s focus on prospecting, where response rates are much lower than for customers. We’ll assume a response rate of 0.8%, take a list with a universe of 100,000, and use our two-part formula to calculate the corresponding test panel quantity:

Part 1:

The numerator is $0.8\% * (1 - 0.8\%) * (1.282 * 1.282)$, which equals 1.3043%.

The denominator is $(10\% \text{ of } 0.8\%) * (10\% \text{ of } 0.8\%)$, which equals 0.000064%.

Put them together – that is, $1.3043\% / 0.000064\%$, and the result is 20,380.

Part 2:

$20,380 / (1 + (20,380 / 100,000))$ equals 16,930.

Therefore, with a test panel response rate of 0.8%, and a universe size of 100,000, a test panel size of 16,930 will result in our being 80% Confident that the rollout response rate will be between 0.72% and 0.88%. In other words, 10% of the time our rollout rate will be less than 0.72%, or 10% less than expected. Conversely, 10% of the time it will be greater than 0.88%, or 10% more than expected.

Consider the problems that this uncertainty can create in circulation planning. All direct marketers have experienced what happens when a rollout response rate is significantly less than expected: a failed rollout!

Many do not realize it, but all have also experienced what happens when a rollout response rate is (or, more accurately, would have been) significantly greater than expected: based on poor test results, perfectly good rollouts that have not been exploited! This is because, frequently, the test panel rate is so artificially low that it dips below what's considered acceptable.

This hidden, second error of testing is particularly treacherous because it is magnified by the opportunity cost of not promoting a cost-effective rollout universe many times in the future. Considering how tough it is to find rental lists that work in today's competitive direct marketing environment, our industry is missing out on significant opportunities for expansion!

The problem is that a test panel quantity of 16,930 is a larger investment than most direct marketers are willing to make. As a point of reference, the 135 expected responders (i.e., $16,930 * 0.8\%$) is much more than the 50-responder rule-of-thumb that was referenced earlier.

In order to reduce the test panel quantity, we're going to have to either widen our Precision, decrease our level of Confidence, or both. So, let's run our formula under three additional scenarios, and see what we come up with. For each, you can decide for yourself if you're comfortable with the results:

- 1) With a test panel size of 11,826, and a Precision of +/- 10%, we can be 70% Confident that our rollout response rate will be between 0.72% and 0.88%. In other words, 15% of the time the rollout response rate will be less than 0.72%, and 15% of the time it will be greater than 0.88%. And, the resulting 95 responders is almost twice the 50 rule-of-thumb.
- 2) With a test panel size of 8,305, and a Precision of +/- 15%, we can be 80% Confident that our rollout response rate will be between 0.68% and 0.92%. In other words, 10% of the time the rollout response rate will be less than 0.68%, and 10% of the time it will be greater than 0.92%. And, the 66 responders is more than the 50 rule-of-thumb.
- 3) With a test panel size of 5,625, and a Precision of +/- 15%, we can be 70% Confident that our rollout response rate will be between 0.68% and 0.92%. In other words, 15% of the time the rollout

response rate will be less than 0.68%, and 15% of the time it will be greater than 0.92%. And, at 45 responders, we're pretty darn close to the 50 rule-of-thumb.

So, how big should your test be? If you enter the formula that I have given you into a spreadsheet, and run some test scenarios with response rates that are typical for your business, you'll have a basis for coming to your own conclusions.

Jim Wheaton is a Principal at Wheaton Group, and can be reached at 919-969-8859 or jim.wheaton@wheatongroup.com. The firm specializes in direct marketing consulting and data mining, data quality assessment and assurance, and the delivery of cost-effective data warehouses and marts. Jim is also a Co-Founder of Data University www.datauniversity.org.