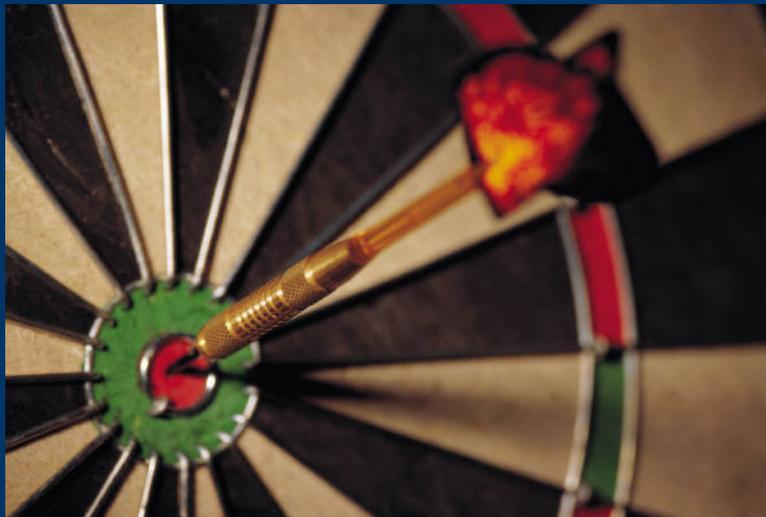


# The Secrets of Test Design Strategies

---

*A “how-to” for direct  
& interactive database marketers*



*By Perry D. Drake © 2005  
Vice President, Drake Direct & Professor, NYU*

*DMAW Annual Conference  
May 4, 2005*

# Presenter Background and Experience

## Perry D. Drake

### ➤ Current Responsibilities

- **Vice president & General Manager, Drake Direct, New York, NY**
- **Professor, Mercy College, Direct Marketing Program, Dobbs Ferry, NY.** Currently teaching test design and analysis techniques to future direct marketers.
- **Assistant Professor, New York University, Center for Direct and Interactive Marketing, New York, NY.** Currently teaching “Database Modeling and Analysis” and “Advanced Database Modeling.” 1999 Recipient of the “*Outstanding Master’s Faculty Award.*”
- **Instructor, Western Connecticut State University, Interactive Direct Marketing Certificate Program, Danbury Ct.** Currently lectures on testing and marketing financials.

### ➤ Prior Responsibilities

- **Director, Marketing Services, The Reader’s Digest Association, Pleasantville, NY**
- **Associate Director, Magazine Marketing, The Reader’s Digest Association, Pleasantville, NY**
- **Statistician, Quantitative Analysis, The Reader’s Digest Association, Pleasantville, NY**

### ➤ Education

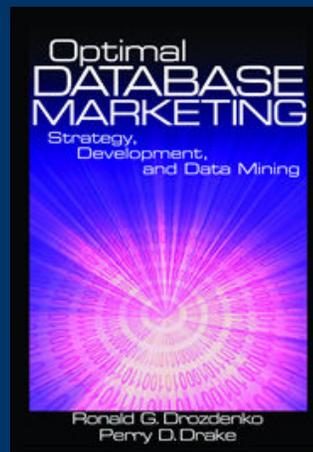
- **Master of Science, Applied Statistics, University of Iowa**
- **Bachelor of Science, Business Economics, University of Missouri**

# Presenter Background and Experience

## Perry D. Drake (Cont.)

### ➤ Publications and Presentations

- Over the years, Perry has presented at many industry conferences including the DMA, Circulation Days, and NCDM. He has also written numerous articles for the various trade journals including DM News, Direct, Target and Inside Direct Mail. To view his presentations and articles written visit his firm's web site at [www.DrakeDirect.com](http://www.DrakeDirect.com).
- In addition, Perry became a published author in April 2002. Perry's new book "Optimal Database Marketing" by Sage Publications delves into database marketing concepts and practices such as the process to evaluate database needs and then select a database vendor, segmenting the customer file, response modeling, lifetime value analysis, and test design and analysis. It is a "how-to" book geared solely for the marketer wanting to better understand the practices and principals of database marketing.



# Introduction – Why We Test

- Testing is the foundation upon which one builds and grows a direct marketing firm.
- With a database, names can be selected for certain treatments and comparisons on the customer's reaction to these treatments made.
- Based on these results, in conjunction with marketing cost and revenue figures, the most profitable decision can be made.
- Without knowledge of proper test planning and analysis, one therefore is not in the strongest position to help their company grow.

# Introduction – Why We Test

- The ability to easily test new marketing concepts, products or lists and read results is what sets direct marketers apart from other marketers.
- By testing:
  - we produce better front end response and better back end performance
  - we can determine the best creative approach
  - we can find the best offer
  - we find profitable mailing lists and other media

# Seven Rules of Test Design & Analysis

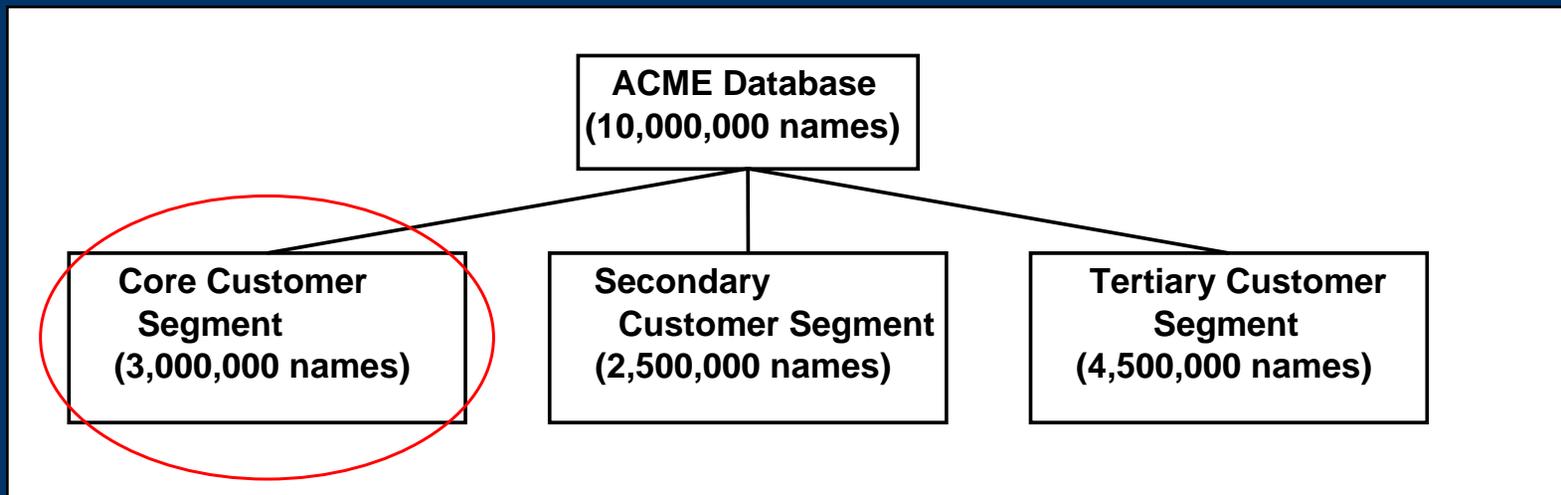
Below are seven rule of test design that must be followed in order to ensure the results of your tests are readable, reliable and projectable for roll-out.

1. Ensure the samples are random and representative
2. Block samples when appropriate
3. Include the control package in the test plan
4. Back test package changes to the control
5. Test each change in the package separately
6. Test for only meaningful interactions and not a full factorial
7. Ensure adequate sample sizes

# Seven Rules of Test Design & Analysis

## Rule 1: Ensure the Samples are Random and Representative

A sample is a subset of customer records and a random selection from the universe of interest on the direct marketer's database. For example, ACME Direct, a direct marketer of books, music, videos and magazines is interested in testing a new product offering. The universe of interest is the Core Customer Segment on their database; therefore, the sample will be comprised of names selected from this universe. A random and representative sample of names from this Core Segment will be test promoted as shown in the figure below.



# Seven Rules of Test Design & Analysis

Representative Samples - To be meaningful, the sample must be representative of the entire population of concern. A representative sample is a sample truly reflecting the population of interest from which the direct marketer draws inferences. For a sample to be representative, no members of the population of interest are purposely excluded from the sample.

To determine the effectiveness of a new format test sent to a specific segment of customers residing on the database, for example, the direct marketer cannot restrict the sample to only those names living in New York. If this new format is sent only to New Yorkers, the expected results will not be representative of the entire population, but only that of New Yorkers.

Some direct marketers overlook this very important concept and assume they can apply test results from one population to another. This may work in some cases but not always. Be careful!

# Seven Rules of Test Design & Analysis

Typically, the only names that should be eliminated from testing are names eliminated in roll-outs such as:

- DMA do-not-promotes
- Frauds
- Credit risk accounts

When testing new promotions, some direct marketers also consider eliminating:

- Names recently promoted for other marketing tests
- States or cities such as Washington, D. C. - known to have strict promotional restrictions. This is especially important if the new test promotion has not yet been fully reviewed by legal counsel.

# Seven Rules of Test Design & Analysis

Random Samples - A sample not taken randomly yields biased and misleading results. A random sample is one in which every member of the sample is equally likely to be chosen, ensuring a composition similar to that of the population. Pulling names one after another from the beginning of a geographically sequenced customer database will result in a geographically biased sample. In this case, depending on the size of the sample being drawn in comparison to the size of the population in total, some regions may not be represented in the sample.

To ensure random samples, many direct marketers utilize what is called “nth selects.” For example, ACME Direct maintains a database of 10,000,000 names. To test a new format to a random sample of 10,000 names from their entire database, the direct marketer will begin by selecting one name on the database, choosing every 1,000th ( $10,000,000/10,000$ ) name thereafter. The result is a random sample of 10,000 names representative of the entire population or database.

# Seven Rules of Test Design & Analysis

## Rule 2: Block Samples when Appropriate

Sometimes, when sampling there may be important variables you will want to control for to ensure a clean read of your tests versus the control.

A good example is in the telemarketing world. For example if you are about to conduct a new outbound telemarketing scripting test, you may be well advised to block based on the years of experience the telemarketing reps have. Doing this will ensure that you do not end up in a situation where the control test has reps with more years experience than the test. Or you might consider blocking on time of day the calls are made.

	Rep Experience	
Script	Less Than 1 Year	Over 1 Year
Control Script	25,000	25,000
Test Script	25,000	25,000

By blocking on years of experience you will ensure that the same proportion of experienced reps are assigned to both the test and control groups.

# Seven Rules of Test Design & Analysis

## Rule 3: Include the Control Package in the Test Plan

Marketing tests are typically conducted simultaneously within a major marketing campaign (the bulk mailing). To determine test winners we compare the response rates of the tests to that of the bulk. There are two main reasons to include a control panel in your test plans:

- 1) We as marketers typically are testing other segments within our bulk plans that we are not testing in the test plan. Therefore to compare the response rate of a test to that of the bulk mailing could lead to incorrect conclusions.
- 2) The lettershop will typically output and deliver to the post office the names associated with the test panels and the bulk (control) package in two separate streams. Larger bulk mail quantities will receive different handling by the post office -- the larger the quantity, the finer the sort and less expensive the postage. Larger quantities usually are dropped at the BMC directly. Smaller quantities are handled piece by piece at the local level.

Therefore names in the test mailing will be processed much more slowly than names in the bulk mailing. To compare the results of the tests which were handled differently to that of the bulk is risky. Therefore, to ensure that you properly read the test panel results against the control bulk package, you must include a control panel with the test mailing.

# Seven Rules of Test Design & Analysis

## Rule 4: Back Test Package Changes to the Control

As previously mentioned, when changing to a new promotional format one should always back test (re-test) the old promotional format to validate the lift in response you had forecasted. Without back tests you will not be able to determine if the cause for a campaign that is under forecast is due to the list selection, the promotion, or a general downturn in the business or any combination of these three.

Consider the following example....

# Seven Rules of Test Design & Analysis

Last year you conducted a new format test and received a .25% response rate versus a .225% for the control format (a +11% lift). You decided to roll-out with the new format to all eligible names (1,450,000 in size). You forecasted a .25% for this mailing. The results of the roll-out are as follows:

	<b>Number of Customers Mailed</b>	<b>Number of Customers who Responded</b>	<b>Response Rate</b>
<b>New Format</b>	<b>1,450,000</b>	<b>2,828</b>	<b>.195%</b>

Your boss is not happy with the results of the mailing and has questioned your decision to change to the new format. Do you have appropriate information to defend your decision?

# Seven Rules of Test Design & Analysis

Now, suppose I told you we conducted a reverse test of the old format within your promotion plan with results as shown below. Can you now defend your decision to your boss? What happened?

	<b>Number of Customers Mailed</b>	<b>Number of Customers who Responded</b>	<b>Response Rate</b>
<b>New Format</b>	<b>1,350,000</b>	<b>2,633</b>	<b>.195%</b>
<b>Reverse Test of Old Format</b>	<b>100,000</b>	<b>176</b>	<b>.176%</b>

# Seven Rules of Test Design & Analysis

## Rule 5: Test Each Change in a Package Separately

When applicable, test various changes to the control promotional package *separately*. Otherwise, you may be misled by the testing results and be left in a situation where no action can be taken.

For example, suppose you are planning to test a new format and new copy approach. Do not test the two changes combined into one test. Test each change separately. You will gain much more information from two separate test results than if they were combined into one test panel.

The same holds true if testing price and incentive tests.

# Seven Rules of Test Design & Analysis

In this example, it may be that the copy change increases response while the new format decreases response. This will not be apparent if tested together.

<b>Test Panel</b>	<b>Response Rate</b>	<b>Lift versus Control</b>
Control Package	1.10%	--
As Control with New Format and Copy Change	1.09%	-1%
As Control with New Format	0.99%	-10%
As Control with Copy Change	1.19%	+8%

Question: What would the decision have been had we not tested each element separately but only combined in the above test series?

# Seven Rules of Test Design & Analysis

Keep in mind that if you are making multiple changes to your champion package based on individual tests (for example, a new OE and new copy), unless you tested both together the pluses received on each will seldom be additive.

Most direct marketers will take half of the pluses and all of the minuses when forecasting the response for the champion package with the changes made.

# Seven Rules of Test Design & Analysis

## Rule 6: Test for Only Meaningful Package Element Interactions

Generally it is unnecessary to test every possible package element combination in your test plan. For example, the Marketing Manager may be interested in testing the following changes to the control package:

- Price/Rate increase
- Addition of a premium
- Color change to outer envelope
- New format

Testing every possible combination of price, premium, outer envelope color and format yields a total of 16 test panels. Testing all 16 is called a “full factorial test design.”

When should a direct marketer consider a full factorial test design?

# Seven Rules of Test Design & Analysis

The only reason a marketer would test a full factorial test design is if it was truly believed interactions will occur between all four elements with respect to response.

In this example, the only possible interaction to be concerned with would be one between price and premium. In other words, if testing a higher price, perhaps the minus in response (due solely to pricing) would be less for the package with a premium versus the package without the premium.

# Seven Rules of Test Design & Analysis

Assuming you are only interested in assessing a possible interaction between price and premium, the test series would appear as shown below:

Test Panel	Description
1	Control test package
2	Price test package - as control package with \$2 price increase
3	Premium test package - as control but with premium for order added
4	Format test package - as control but with a new format
5	New OE - as control but with a new OE
6	Price and premium test package - as test panel #3 with \$2 price increase

Question: Based on this test series, how would you determine if the addition of a premium to the control package offset any or all of the negative effect that a price increase might have on response?

# Seven Rules of Test Design & Analysis

## Rule 7: Ensure Adequate Sample Sizes

Without sampling enough names, your test results will have so much “error” associated with them that you will not be able to make a well informed and solid decision regarding roll-out.

Ensure you test enough names so that the results are readable, reliable and projectable.

# Seven Rules of Test Design & Analysis

What exactly do I mean by  
“error in test results?”

# Seven Rules of Test Design & Analysis

You conduct a new format test to a 10,000 random sampling of names drawn from your core universe of concern and receive a response rate of 1.19%.

Can you run to the bank with the 1.19% response rate?

# Seven Rules of Test Design & Analysis

## **Absolutely Not!**

Because you did not test the whole universe available, but only a sample, the response rate obtained is therefore only an estimate of what might be.

In fact, each time you conduct such a test you will get a different response rate.

# Seven Rules of Test Design & Analysis

Let's say you conducted not 1, but 10 tests of the same format to 10 unique samples of size 10,000 each drawn from the same universe.

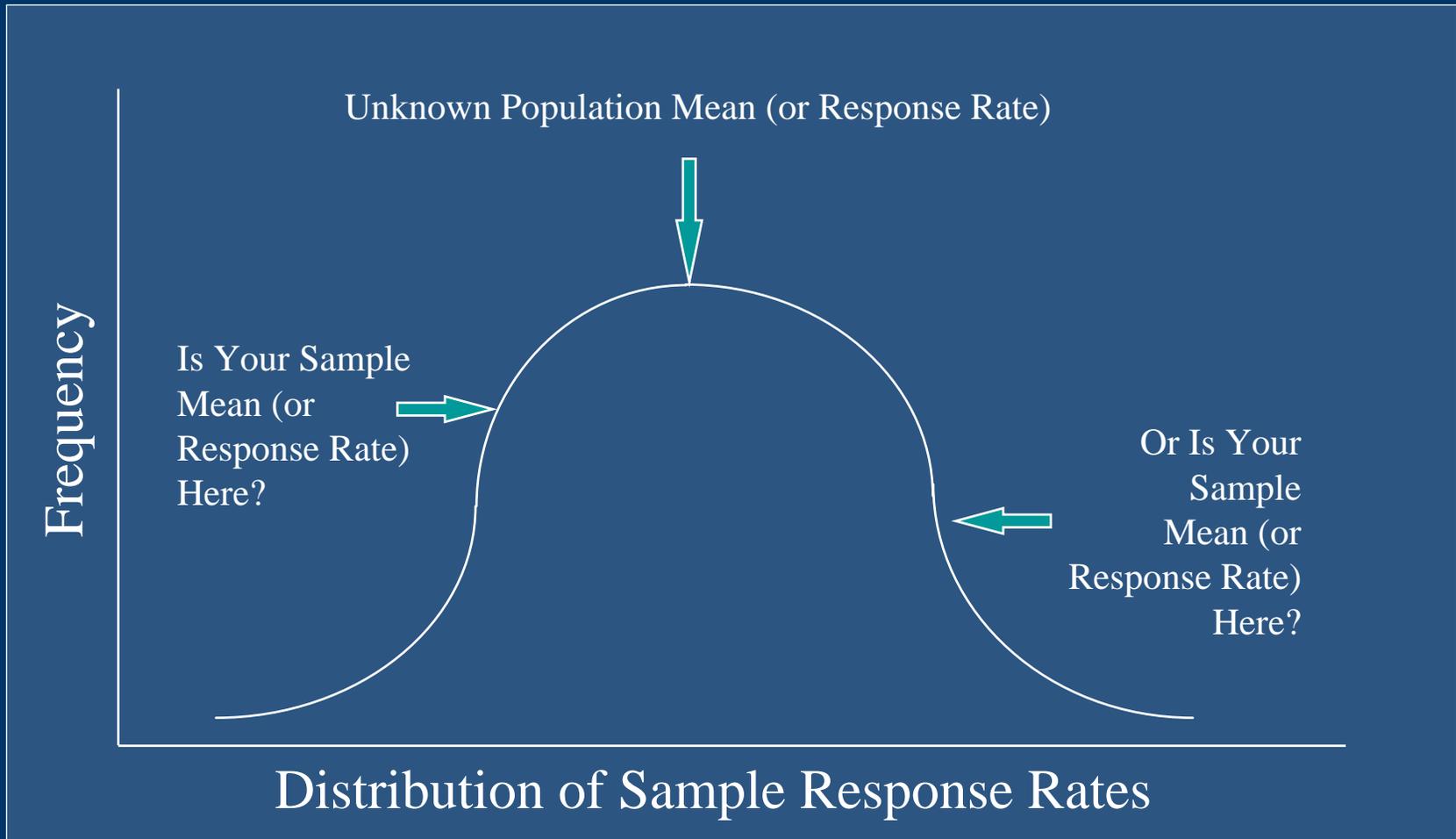
The results may look as follows:

<u>Test #</u>	<u>% Resp</u>	<u>Test #</u>	<u>% Resp</u>
1	1.23	6	1.17
2	0.99	7	1.22
3	1.06	8	0.97
4	1.15	9	1.08
5	1.19	10	0.95

*Your one test panel result of 1.19%*

# Seven Rules of Test Design & Analysis

Every time you test a list, you will get a different response rate. Some tests will yield results above the true response rate of the entire list and some below the response rate of the entire list.



# Seven Rules of Test Design & Analysis

Every test mailed has an error associated with it. The smaller the test quantity the more error associated with the test. A test of 5,000 names will have more error associated with the response rate received than the same test to 10,000 names.

So, the key is to understand just how much error you can tolerate in your test results and ensure that you sample enough names so that your error rate does not exceed that level.

Keep in mind that testing more names than is required for your level of precision is a waste of testing dollars. For more information on determining exactly how much error you can tolerate and hence the appropriate sample sizes see Perry's book "Optimal Database Marketing."

Once the amount of error is determined you can then use "*The Plan-analyzer*" software package created by Drake Direct to determine the exact sample size.



# Seven Rules of Test Design & Analysis

Assume you plan to test a new list estimated to yield a response rate of 1% and that you wish to control the error of the test so that it is no more than plus or minus 10% of the estimated response rate. How many names should you test?

Based on “The Plan-alyzer” you should test 38,000 names. Doing so will guarantee the error associated with the test will be no greater than plus or minus .10%.

The screenshot shows a software window titled "Plan-alyzer V3.0" with a green header bar. The header contains the text "Courtesy of Drake Direct Database Marketing Consulting Firm". The main content area is yellow and titled "Sample Size Estimation - A Single Estimate". It includes two buttons: "Return to previous screen" and "Recalculate". Below these is the instruction "Input the required information and press 'SUBMIT'". There are three input fields with the following labels and descriptions:

- Input: .01, Label: Test response rate estimate. Input in decimal format (e.g., input 4.00% as .04).
- Input: .001, Label: Allowable variance. Input in decimal format (e.g., input +/- .25% as .0025).
- Input: .95, Label: Confidence level. Input in decimal format (e.g., input 95% as .95).

A white box displays the results: "The sample size required for the conditions listed above is: 38,032. A sample of this size will guarantee with 95% confidence, the estimated response rate achieved in the test will be within +/- 0.1 percentage points of the true response rate to expect in a roll-out." Below this is a "Print" button. At the bottom, a green footer bar contains the text: "Any results and / or analysis of direct marketing tests and / or rollouts are not guaranteed or warranted by Drake Direct. Drake Direct. 225 East 46th Street, Penthouse D, New York, NY 10017 Telephone: 212-759-1225 Fax: 212-759-9756".

# Case Study #1 – Postage Test

A client of mine approximately two years ago decided to test changing a “trigger” promotional offer to their current customers from third class to first class.

The rationale for this promotional change was to see if a plus would be realized by getting an additional offer in the hand of the customer more quickly once the trigger occurred.

And, more importantly, would this plus cover the additional costs of changing from third to first class.

# Case Study #1 – Postage Test

The results of this mailing were as follows:

<b><u>Test Panel</u></b>	<b><u>MQ</u></b>	<b><u>Gross Resp</u></b>
Offer @ 3 <sup>rd</sup> Class	63,546	0.36%
Offer @ 1 <sup>st</sup> Class	63,669	0.44%

They asked for my help in making a decision.

# Case Study #1 – Postage Test

Analyzing the test results and assessing the amount of error associated with the test and control offers we came to the following conclusions:

- The test yielded a plus in response of +22%
- To break-even, the test needs to yield a plus of +18%
- Therefore, the test did meet break-even as observed
- So the question became how much error is there in reading this test given the sample sizes and can we trust the observed plus?
- Using “*The Plan-alyzer*” to assess the amount of error associated with this test, we did in fact determine the test was a significant winner with 95% confidence by anywhere from +42% in response (best case scenario) to only +2% in response (worst case scenario).
- With only a +2% in response being a possibility, we could end up in a situation where we are not even coming close to meeting our break-even plus required of +18%.

# Case Study #1 – Postage Test

## The Decision:

- Even though the test beat the control with confidence, the downside potential for the test (based on a full blown P&L) was deemed too risky.
- As a result, we decided to retest.
- But before re-testing, upper management wanted to ensure that we tested enough names this time to ensure a decision would be able to be made regarding roll-out.
- To determine exactly how many names to test was a complex process whereby we had to estimate the error we could tolerate in our results and still be able to make a decision.
- The resulting sample sizes jumped from 65,000 per panel to slightly over 300,000 per panel.

# Case Study #2 – Email Test

A publishing client of mine last year wished to gather as many email addresses of their current subscribers as they could. They decided to test asking for email addresses on each renewal effort coupon/return slip. The results are as shown below:

<b>Test Panel</b>	<b>MQ</b>	<b>Gross Resp</b>	<b>%Cash W/O</b>
Control	11,179	30.61%	61.43%
Email Test	11,161	31.54%	63.77%

# Case Study #2 – Email Test

Their decision: To roll-out and test on other magazine titles since the test yielded a response rate higher than the control (31.54% vs 30.61%).

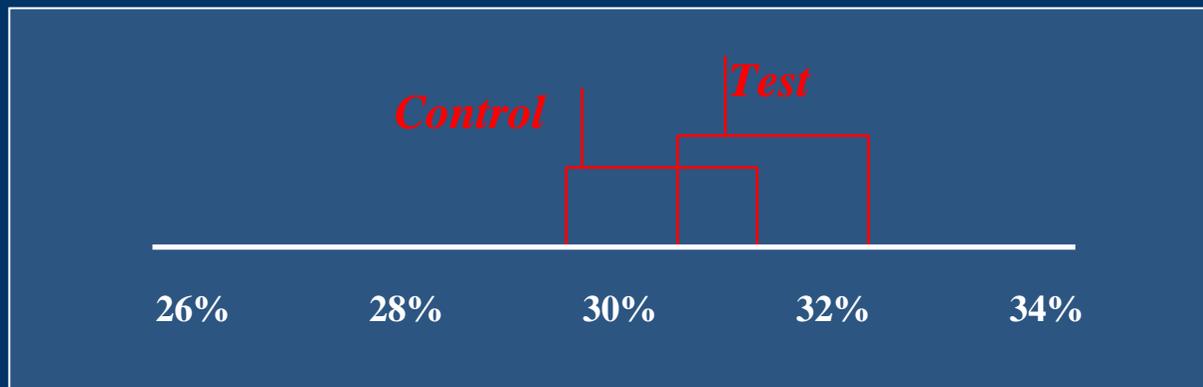
# Case Study #2 – Email Test

A Huge Risk. Why?

If you were to construct a confidence interval around the test and control response rates you would note that they overlap.

This basically implies no difference in the response rates of the control and test and it also implies the test could in fact do worse than the control in roll-out - not a good thing!

- The control really lies somewhere between 29.76% and 31.46%
- The test really lies somewhere between 30.68% and 32.40%



# Case Study #2 – Email Test

Other reasons not to roll-out include:

- A mistake in judgment here could cause the company a major loss in revenue.
- Typically, we as direct marketers see response negatively impacted when we ask a customer to find a pen and mark something down on the order form. So, in a way these test results should have been questioned from the start.
- Also, note the difference in cash with order. It too does not make sense. Sampling error must be present.
- As a result, I recommended they retest to a larger sample and perhaps on more than one title before making this decision.

# Questions?

## Perry's contact information is:

Drake Direct  
225 East 46<sup>th</sup> Street, Penthouse D  
New York, NY 10017

Phone: 212-759-1225  
Fax: 212-759-9756  
Mobile: 914-299-4879  
Email: [Perry@DrakeDirect.com](mailto:Perry@DrakeDirect.com)  
Website: [www.DrakeDirect.com](http://www.DrakeDirect.com)