

That Voodoo We Do - Marketers Are Embracing Statistical Design of Experiments

By
Richard Burnham

What had previously been a somewhat mystifying and obscure statistical technique took an abrupt turn to mainstream *marketing* acceptance eight years ago. In March of 1996, a Forbes article titled "*The New Mantra: MVT*" caught even statistical marketing experts by surprise. Mark Anderson, one of three principals at the privately held Stat-Ease, Inc. in Minneapolis recalls receiving a strange phone call: "Forbes magazine says you folks make the software for MVT - what can it do for me?" Anderson replied, "That's nice of them to say, but what's MVT?"

Listening intently to the caller's reply, Anderson immediately realized that MVT™ (multivariable testing) was simply a newly coined phrase for a 1920's invention called design of experiments (DOE) - a statistical tool, which Anderson's company has promoted through software and workshops since the early 1980's. Originally developed to increase agricultural yields, it began a slow diffusion into chemical and manufacturing sectors to improve product quality. It's turned out that when properly

applied to marketing, design of experiments is also a marketer's silver bullet. Look at some examples:

- Crayola® conducts an e-mail marketing DOE to attract parents and teachers to their new Internet site. The company discovers a combination of factors that makes their new e-mail pitch three-and-a-half times more effective than the control. (*Harvard Business Review*, October 2001, "Boost Your Marketing ROI with Experimental Design," Almquist, Wyner.)
- A travel club wants to increase response to their direct mail campaigns. It tests 17 factors at one time in 20 differing mail pieces, including: copy variations on the envelope and cover letter, the offer, graphics, and even fonts and logos. They find that short and simple works best, envelope text helps, and a free offer makes no difference. Extra inserts actually lower response. The DOE predicts an increased response from 0.3% to 0.5%, worth \$20 million to \$40 million in annual revenue. (*Multivariable Testing Methods in Marketing*, Gordon H. Bell.)
- A DOE investigates the effects of ad size versus product category on consumer recall. The effect is indeed a positive correlation, but the resulting sales increase does not outweigh the higher ad-space costs. Smaller and cheaper advertising turns out to be more profitable. (*Journal of Advertising*, Vol. XXIV, No. 4, pp. 1-12, "Ad Size as an

Indicator of Perceived Advertising Costs and Effort: The Effects on Memory and Perceptions," Homer.)

DOE's mysterious strength is the manner in which it tests combinations of factors instead of the traditional one-factor-at-a-time method of experimenting, such as A/B splits. As emphasized in the 1996 Forbes article, "If you test factors one at a time, there's a very low probability that you're going to hit the right one before everybody gets sick of it and quits." Not only that, you'll also miss out on interactions between factors, which often are key. (See sidebar.)

To prove this 80-year-old statistical technique's worth, let's review a marketing DOE conducted by Anderson's company. Every few months, Stat-Ease would mail two-color postcards announcing upcoming national DOE workshops. They wanted to know if four-color printing enhanced response. To test this, they created eight slightly different versions of their existing postcard. Each version contained two variations of each of three different factors:

- Factor A: Number of colors (two vs. four)
- Factor B: Postcard size (4" x 6" vs. 5.5" x 8.5")
- Factor C: Paper stock (index vs. heavyweight)

Table 1 shows all eight postcard versions and subsequent responses in this classic example of a powerful DOE design known as a two-level, full-factorial experiment.

TABLE 1 – Eight postcard versions test three factors at once. “High” and “low” factor levels are represented by plus (+) and minus (-) signs, respectively.

<u>VERSION</u> (Standard Order)	<u>FACTOR A:</u> NUMBER OF COLORS	<u>FACTOR B:</u> POSTCARD SIZE	<u>FACTOR C:</u> POSTCARD THICKNESS	<u>RESPONSE</u>
1	2-color (-)	4" x 6" (-)	Index (-)	34
2	4-color (+)	4" x 6" (-)	Index (-)	18
3	2-color (-)	5.5" x 8.5" (+)	Index (-)	44
4	4-color (+)	5.5" x 8.5" (+)	Index (-)	26
5	2-color (-)	4" x 6" (-)	Heavyweight (+)	26
6	4-color (+)	4" x 6" (-)	Heavyweight (+)	17
7	2-color (-)	5.5" x 8.5" (+)	Heavyweight (+)	29
8	4-color (+)	5.5" x 8.5" (+)	Heavyweight (+)	21

The eight versions were mailed out to eight different, randomly selected, sampling subsets of their mail list. For enhanced response, Stat-Ease offered a free report to those who faxed back the card. To track the incoming faxed responses, a code had been discretely added to the corner of each version that keyed to Table 1. For example: “- - -” was printed on only the two-color (-), 4" x 6" (-), index-stock (-) versions; and “+ + +” appeared solely on the four-color (+), 5.5" x 8.5" (+), heavyweight (+) postcards.

The results proved to be somewhat surprising. In a statistically significant “upset,” neither postcard size nor paperstock weight proved to be influential in response outcome. What *did* make a difference to Stat-Ease’s engineering-oriented audience was the number of printed colors. Two-color postcards

drew far better than four-color. As reflected in Figure 1, the two I-beam-shaped brackets shown in the software-generated graph don't overlap horizontally, indicating to the marketers that the response is statistically significant with over 95% confidence that the result is not a product of chance.

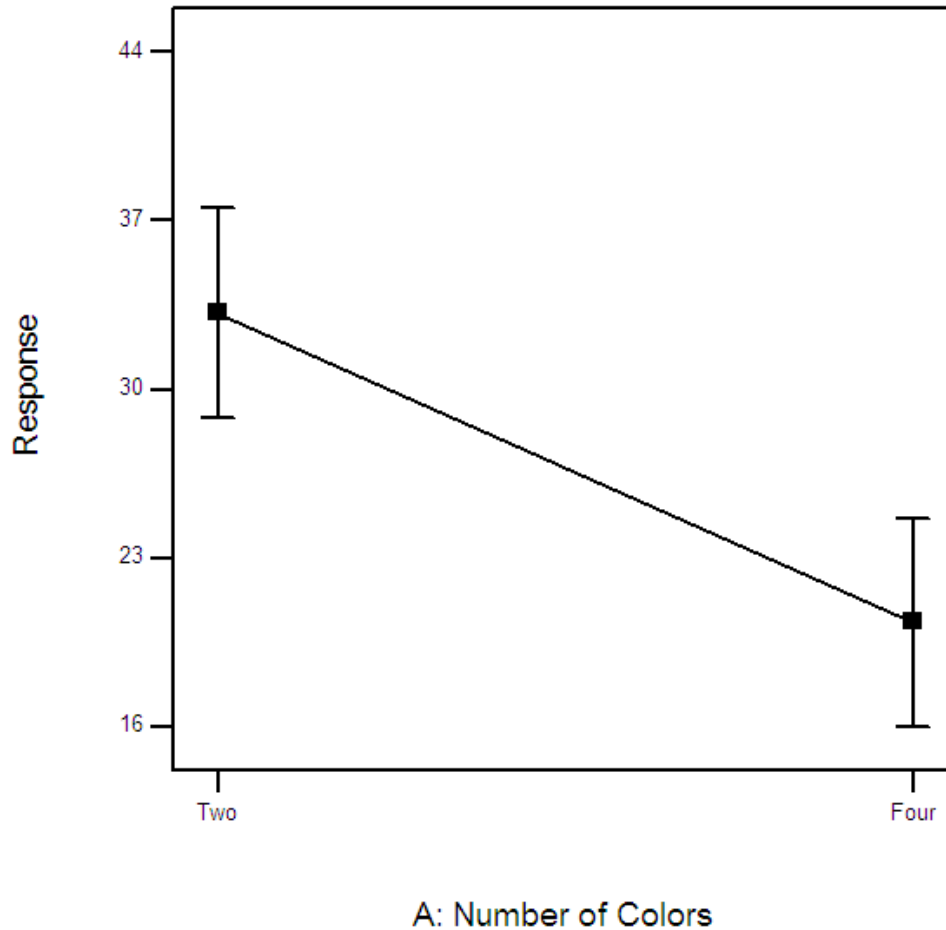


FIGURE 1 – INTUITIVELY DIFFICULT TO COMPREHEND, yet statistically significant: Stat-Ease workshop engineering prospects respond best to cheaper two-color postcards. (Graph courtesy Stat-Ease, Inc.)

Using the same DOE data, but viewing a different type of statistical graph, Figure 2 clarifies three important findings:

- Factor A (number of colors) stands alone in significance.
- Factor B (postcard size) and Factor C (card stock) hold no significance in influencing response.
- Interactions between factors are statistically insignificant.

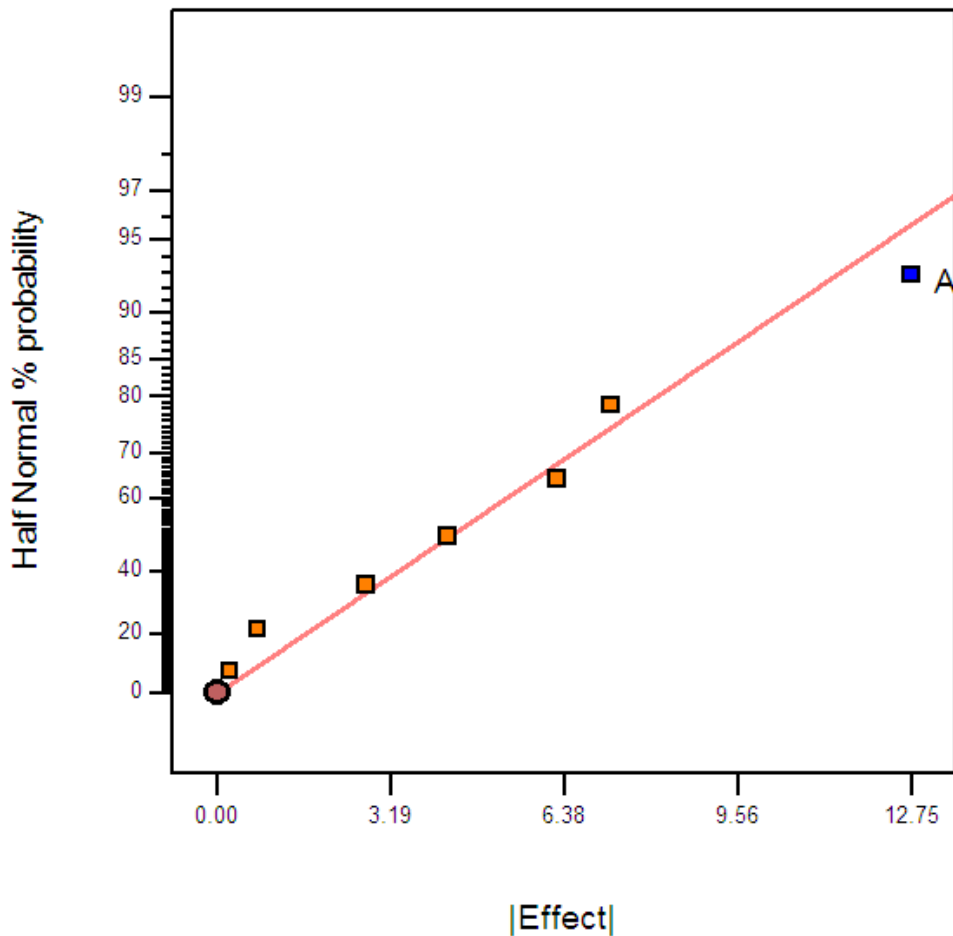


FIGURE 2 – FACTOR A (Number of colors in the postcard – two or four) stands alone, far to the upper right. For the marketers using this DOE data, the statistically isolated response point means that they don't need to pay extra for four-color printing. It's a breakthrough discovery worth thousands of dollars annually. (Graph courtesy Stat-Ease, Inc.)

Evidently, the statistically savvy engineering audiences whom Stat-Ease targets demographically are turned off by fancier postcards. They want the steak, not the sizzle. The results are now saving the software company thousands of dollars that would have been wasted with four-color printing. For their specific customers, simpler means better.

Marketers can't always be certain what triggers buyers to respond. In the past, we were always admonished to *test-test-test*, but only one factor at a time - relying on our gut feelings and uncertain hopes. With DOE, marketers have replaced voodoo with the science of statistics. Chanting and stickpins are optional.

Sidebar:

INTERACTIONS JUSTIFY MARKET SEGMENTATION

For decades, it's been common for marketers to test variables such as price, place, and product using A/B splits - only one factor changed for each test. Not any more, for this traditional method of testing one factor at a time may soon be going the way of the horse and buggy. Statistical terms such as interaction effects and design of experiments (DOE) are now part of the new terminology of modern marketing.

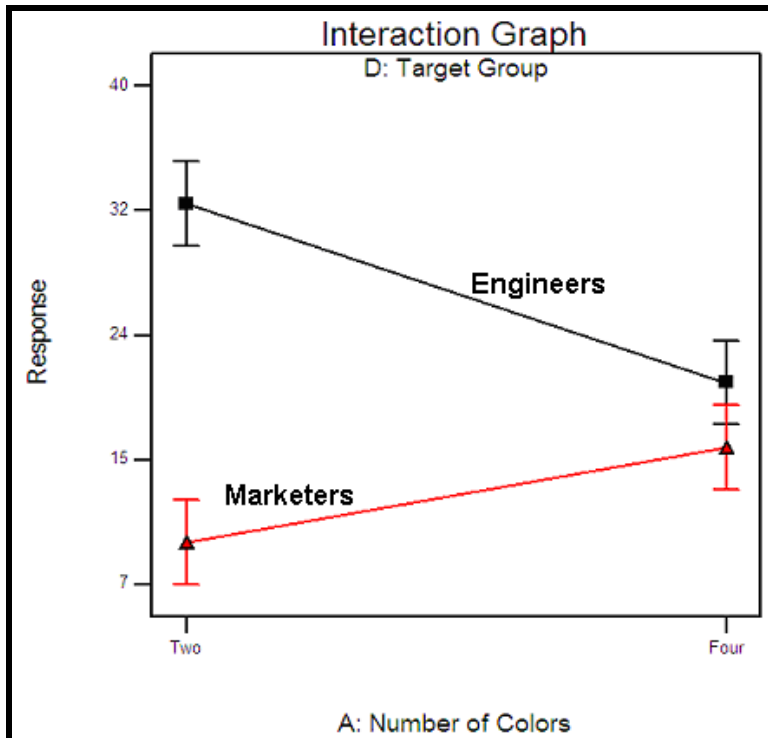
Interaction effects occur when the outcome of a factor's marketing response differs depending on the setting of another

factor being tested. DOE lets you change several variables at once, yet still isolates the impact of each. Testing multi-factor DOE interactions at an initial small extra cost, with an ultimately huge payoff, gives perceptive marketers a massive advantage over A/B split tests. Why? Because you can use the same sample size whether you're testing one factor or 31 factors.

Dr. Paul Berger, Professor of Marketing and Quantitative Methods at Boston University's School of Management, states, "Some marketing research firms still only change one factor at a time, out of ignorance." When not teaching university students or co-authoring DOE books, Berger offers his valuable lessons to marketing-consulting clients. He says of them, "Virtually always they began at the knowledge stage of 'knowing what they didn't know,' and had already been doing what they thought of as designing experiments. Of course, once in a while, I got a response of 'you guys and your interactions always making it more complicated!' When I'd hear that, I knew the company was near hopeless and felt bad for their clients, although I did try to explain to them that interactions are a naturally occurring phenomena, and what we DOE folks do is provide a way to measure and interpret it."

Sophisticated DOE software and the knowledge gained in workshops from firms like Stat-Ease (see accompanying story) today generate graphs that make it easy to interpret

interactions. When these interaction graphs display two non-parallel lines, the differing angles indicate that the effect of one factor depends on the level of the other. When an interaction occurs, the result very often is a breakthrough development in profits. The following DOE interaction graph was created from data that Stat-Ease marketers obtained. It shows lines that are far from parallel, indicating quite different response effects when changing the number of colors printed on a direct-mail piece from two to four. Note the interaction of Factor A (number of colors) with Factor D (target audience). It means that two-color printing appeals more to the just-the-facts-ma'am engineers, and four-color printing really sparks marketers' response. The obvious solution for enhanced sales is to design different direct mail for one audience versus the other. In other words - market segmentation.



When Factor A (number of colors) is set to two-color printing, the marketing target group (red line) generates fewer responses than when four-color printing is used. But the engineering target group (black line) at two-color printing generates more responses. As Factor A switches to four-color printing, its appeal to the engineering target group decreases, but appeals more to the marketing target group. (Graph courtesy Stat-Ease, Inc.)

ABOUT THE AUTHOR:

Richard Burnham's former positions include marketing and art directorships. He now resides in rural Wisconsin, where he writes trade publication articles in semi-retirement. He may be reached at RABURNHAM@PublicationCOORDINATION.com

CONTACT INFORMATION:

Design-Expert® software

Stat-Ease, Inc.

2021 East Hennepin Avenue, Suite 480

Minneapolis, MN 55413

Phone: (612)378-9449 Fax: (612)378-2152

E-mail: info@statease.com

Website: www.statease.com

R.A.BURNHAM Company

Publication Coordination

105 Aztalan Street, Suite Two

Johnson Creek, WI 53038

Phone: (920) 699-2904

E-mail: RABURNHAM@PublicationCOORDINATION.com
