



MarketDial

White Paper Series
Optimal Test Parameters

How to create a statistically valid test



Alen Delic
Data Scientist, MarketDial

How to create a statistically valid test

MarketDial White Paper Series:
Optimal Test Parameters

Alen Delic
Data Scientist, MarketDial

Content

- 1 Finding your optimal test parameters
- 2 Testing methodology without unlimited time or money
- 3 Achieving confidence
- 4 Statistical regularity
- 5 As easy as your favorite app, as powerful as your peace of mind

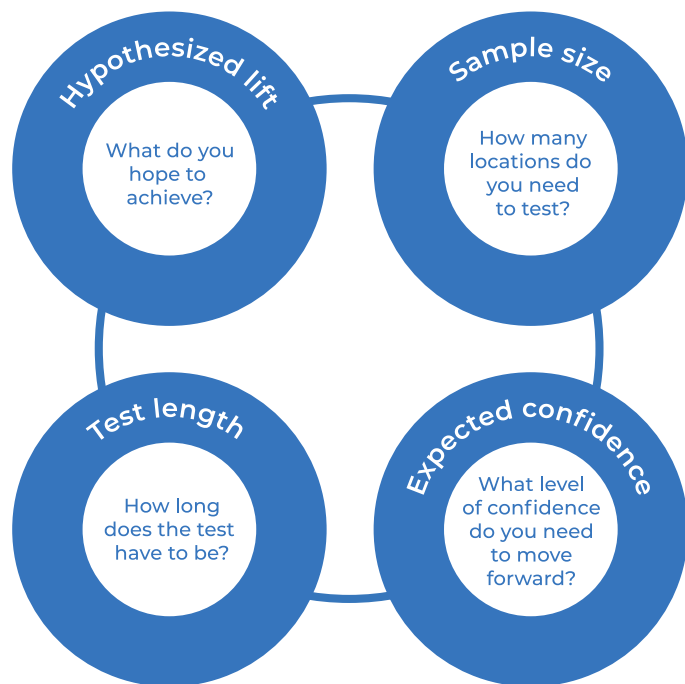
How to create a statistically valid test

Alen Delic, Data Scientist, MarketDial

Finding your optimal test parameters

“Four percent ... will this new idea lift revenue by that much?” The answer is simple when you know how to pull the right levers. Even the natural follow-up question, “Are you sure?” becomes a confident “yes” with proper test design. Here’s how.

Before you begin, you need to know a test’s four essential parameters:



Revealing reliable results

If you run a test in one location and see your desired lift, then test in another and get a negative result, you don’t know what’s accurate and what isn’t. Fortunately, any experiment can be designed to reveal the reliable results you need to determine if that next great idea is a true winner.

Testing methodology without unlimited time or money

The more constrained your timeline or testing budget is, the more you need to be confident your test is scalable across your footprint.

Consider a retailer with a fleet size of 500 that wants to attract more customers and believes that changing the layout of each store's outside display wall could be the answer. This change is expensive; it will not only take design and construction resources, it will also impact traffic during the renovation.

Here is the basic story problem:

The average revenue per store is \$5 million.

It costs \$200,000 to install the new display wall in each store.

So the retailer wants at least a 4% lift in revenue to pay for the change and begin to realize a net profit.

The \$200,000 cost drives a budget cap of \$2 million for the test.

That means they can transform 10 stores as test sites.

How can the retailer be confident that 10 stores will reveal the results needed to go all in across the other 500 stores?

Achieving confidence

Confidence is a statistical measure that ranges from 0% (no confidence) to 100% (as confident as possible). This measure describes the signal-to-noise ratio.

If we have a large lift (strong signal) and low variance (low noise), we'll have high confidence that the signal we're seeing is due to the test as implemented and not because of random noise.



Confidence increases as sample size gets larger, test length longer and hypothesized lift higher.



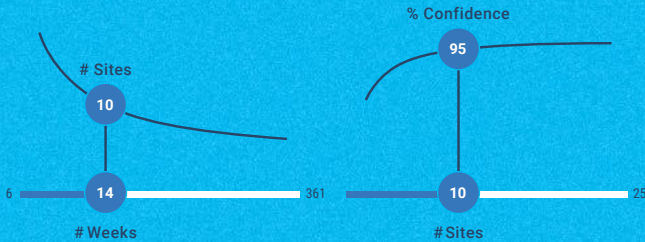
Confidence decreases with smaller sample size, short length of test and a small lift.

Cue the confidence simulator

MarketDial's software features a confidence simulator that uses the four parameters to inform clients of how many sites they need for what length of time to get the expected confidence at the hypothesized lift.

Clients often comment on how intuitive and easy to understand the simulator is. Here's an idea of what it offers for the client in our example:

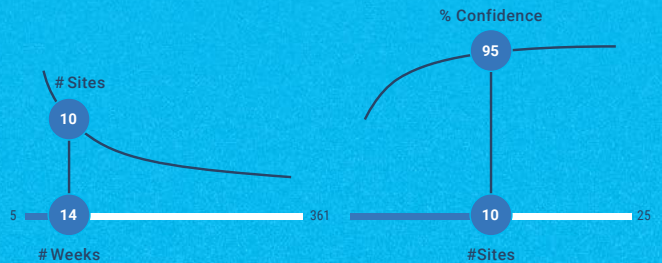
Length (weeks)	Number of Sites	Length (weeks)	Number of Sites
14	10	10	95% Best



▶ Factoring the hypothesized 4% increase in revenue with 10 sites, the confidence simulator says that to achieve 95% confidence, the test has to run for 14 weeks.

What if the retailer has to have a go/no go business decision in just 8 weeks?

Length (weeks)	Number of Sites	Length (weeks)	Number of Sites
8	13	13	95% Best



▶ The confidence simulator can refine the calculation to determine, for example, that 13 sites would be needed to maintain 95% confidence.

The experiment has to be powered correctly so that time and money aren't wasted with too little or too much testing. A tool like the MarketDial confidence simulator enables you to adjust the levers to get to the right balance of time and cost based on the hypothesized lift and expected confidence.

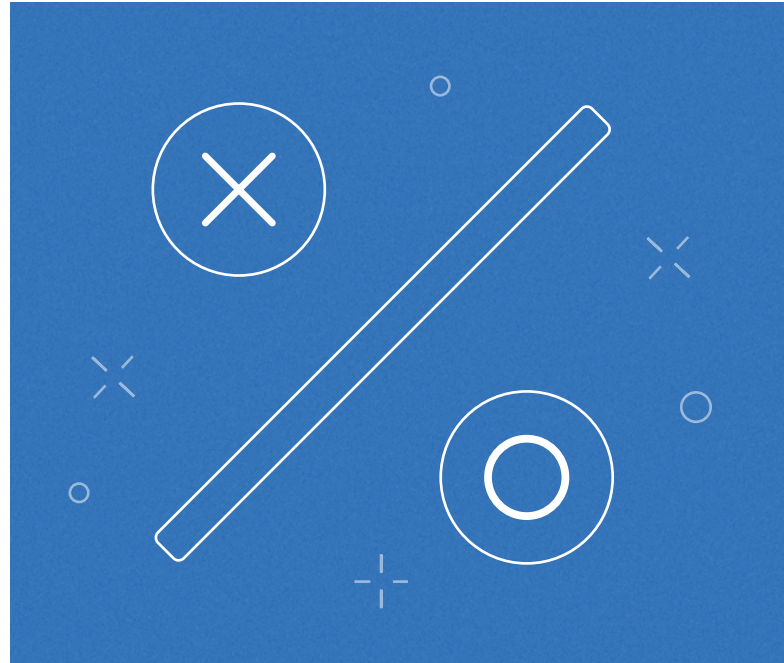
Statistical regulatory

True or false: the larger the lift you want, the fewer the data points you need

True. Yes, this statement is counterintuitive. But for larger lifts, we can use smaller sample sizes or shorter test periods because you don't need many data points to show a significant difference between the test and control group.

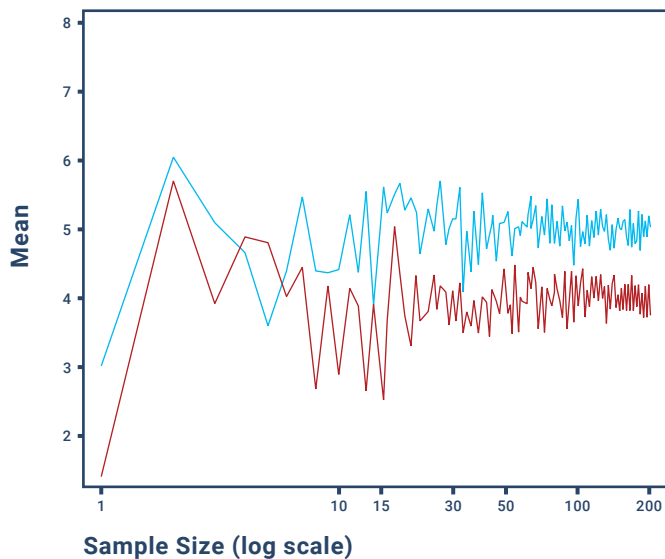
These graphs demonstrate this concept based on the law of large numbers and statistical regularity.

Each graph has two lines that represent random samples of two different distributions of data. Graph A contains two distributions with a smaller difference in means (similar to lift), compared to graph B, which compares a larger difference in means. As we see in graph B, with a larger lift, a distinct difference is observed with smaller sample sizes.

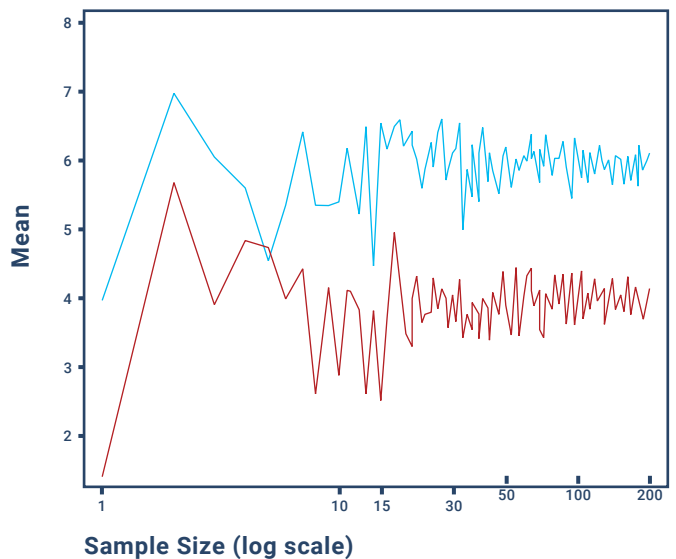


Finding your optimal test parameters

A N(5, 2) and N(4, 2) Variables



B N(5, 2) and N(4, 2) Variables



As easy as your favorite app, as powerful as your peace of mind



The MarketDial software makes it simple to be confident in your testing. Choose your desired lift, determine the balance between sample size and test length and reveal your expected confidence level ... all in a matter of minutes. MarketDial's platform and partnership helps you achieve more, faster.

Why MarketDial

More than 100 leading companies and global brands rely on MarketDial to help them create a culture of experimentation with intentionally easy A/B testing. MarketDial's simple interface makes asking "what if?" the cost-effective protocol for intelligent action. [Learn how now](#)

About Alen Delic, Data Scientist, MarketDial

With a Bachelor of Science in biochemistry and a Master of Science in biostatistics, you would expect Alen to be focused on data. Indeed, he was the lead biostatistician for the University of Utah's Neurology department before finding his home at MarketDial. But he also loves teaching, working often with Yale clinicians to create statistical lectures for Yale medical residents. True to his roots in Bosnia and Germany, he enjoys traveling. Next up: an African safari and a tour of Southeast Asia. When he isn't deep into data or flying internationally, he's restoring a 1993 Nissan 180sx and producing electronic music.

