

7

**A/B TESTING
SINS
THAT DESTROY YOUR
REVENUE**

Contrary to what the A/B testing tools want us to believe, A/B testing is not a simple game. We've set up and analyzed over 2,000 A/B tests on online stores in the past 7 years and have seen many companies make extremely costly mistakes.

When we were brainstorming for this list, we initially had a list of 33 costly A/B testing mistakes... 33! Obviously, that would go a bit overboard... Maybe we can use it as a topic for a new book, '33 A/B Testing Sins that Destroy Your Revenue' 😊

We've selected 7 of them based on 2 criteria:

1. how often we see people make this mistake; and
2. how big the impact can be on your revenue.

Happy reading!



Joris Bryon

Founder | @ dexter.agency

Joris Bryon



Marc Uitterhoeve

CEO | @ dexter.agency

M. Uitterhoeve



Sin
#1

Using frequentist stats instead of Bayesian

Let's get the most controversial one out of the way. Without getting too technical: you should use Bayesian stats to analyze your A/B tests, not frequentist stats.

This isn't a 'Statistics 101' course, so let's just stick to the essence of it. Frequentist stats are used in science. They're strict. Very strict. And that's OK. After all, if you're testing a drug, you want to make sure you're not killing anyone.

Bayesian stats on the other hand are a lot more pragmatic. Their main goal is to help you decide whether or not it's a good business decision to implement a test variation. If e.g. you have an 85% chance of making \$300,000 more per year with a test variation, you'd go for it, right? That's what Bayesian answers: is it a good business decision or not?

Why is this hurting your revenue?

If you use frequentist stats to analyze your tests, because they're so strict, you'll end up NOT implementing changes to your site that would actually increase your revenue. And obviously, that means you're leaving a lot of money on the table.





**Sin
#2**

Not applying segments

If you look at the results in your A/B testing tool, all you'll see is an average. And averages lie. You have to dig a little deeper. E.g. if it's a cross-device test, what do the results look like on desktop, mobile and tablet? We've seen it many times: overall there doesn't seem to be a difference between A and B.

But when you apply a device category segment, all of a sudden it turns out to be a winner on one of the devices.

It's not only about device categories. It can be any kind of segment, e.g. new vs returning visitors. Or US vs Canada vs UK. Just think of what makes sense for the particular test you're analyzing.

Why is this hurting your revenue?

If you just look at averages, you'll leave A LOT of money on the table, because you will overlook winners. Winners that often add hundreds of thousands in yearly revenue.





**Sin
#3**

Not calculating the statistical power upfront

If you don't calculate the statistical power before you run a test, you might as well not A/B test. I know, it's statistics, and most people try to stay away from statistics as far as humanly possible. So I'm not going to bore you with this.

Just know that you should calculate the power of a test upfront - a power of 80% is the rule. This will help you understand how long a test should run on e.g. your product pages before you can spot a 5% difference that is actually statistically significant.

Why is this hurting your revenue?

If you don't do this, you'll call a test a winner when it's not statistically significant. What's even worse: you'll think they're a winner, implement them, but actually they're a loser and they'll end up costing you a lot of revenue.





**Sin
#4**

Not having a clear prioritization system so you know which tests to run first

We see it all the time: teams have their favorites when it comes to test ideas. They pick the ones they feel are going to win and they work on those first.

This is a terrible idea! Because it's too subjective. We're all biased. So we need objective criteria to counter that bias. That way we can focus on building those tests with the biggest potential impact first. If you're familiar with the PIE framework, that's a good starting point.

However, we think that's not precise enough, so we have created The Dexter Score™, an A/B test scoring system with 9 criteria specific to online stores.

Why is this hurting your revenue?

If you don't run the tests with the highest potential impact on your revenue first, you're wasting a lot of time on tests that don't move the needle. It'll cost you a lot of resources with very little to show for.





**Sin
#5**

Not using server-side cookies

This one is a little technical, but ever so important. Since Safari's introduction of ITP2.1, client-side cookies are automatically removed from the browser after 7 days. That means that if a test runs for more than 7 days, your data from Safari can't be trusted anymore. That's why you need to use server-side cookies.

Why is this hurting your revenue?

If you still rely on client-side cookies, the Safari data can be heavily skewed. If Safari is responsible for a considerable portion of your total traffic (say 5 to 10%), this can mean the difference between calling a test a loser or calling it a winner. And obviously, you can miss out on a lot of extra revenue if you don't detect the winners or lose a lot of revenue if you think it's a winner but it's actually a loser.



Not doing (enough) QA

This is probably the most underrated part of A/B testing. All blogs, tutorials and courses talk about all the fun stuff of A/B testing. At best they mention Quality Assurance in passing. But it's an absolutely critical step in the A/B testing process.

You have to be sure your test variations are working properly on all browsers and devices before you push them live. Not just the browsers and devices you happen to use yourself.

Why is this hurting your revenue?

If you push a test live that is not working properly on one or several browsers or devices, it will skew your results: your test loses, but actually it could've won. So now you've invested all your time, resources and money in setting up the test, and you don't reap the rewards.

Focusing too much on the conversion rate

In the end, you want more revenue. Some changes can result in more conversions, but lower order values. It's important that you know what has more weight: the increase in conversion rate or the drop in AOV (or vice versa). The revenue per user will help you answer this question. It will tell you if people have a structurally higher AOV, or if one big order just skewed the results.

The difference in revenue per user cannot be calculated via the Bayesian method since it's not a binary metric. It requires some setup in the beginning with client IDs and some advanced reporting and analytical skills, but once you've set this up, you can calculate it for every single test you're running.

Why is this hurting your revenue?

You can either implement something that will actually hurt your revenue, or you discard a test because you haven't found a statistically significant result in terms of conversion rate. We have had numerous tests where there was no difference in conversion rate, but the B-version was responsible for a - statistically significant - higher revenue per user. If you are not able to calculate this, you're leaving hundreds of thousands of dollars on the table (if not more).



Do you have a high-7 or an 8-figure online store and do you want to avoid losing revenue because of these (or any of the other 26) costly mistakes?

Let's chat.

[SCHEDULE YOUR HIDDEN REVENUE OPPORTUNITIES™ SESSION](#)