

コンバージョン率が3倍になる A/B テストの7つのルール ラース・ロフグレン(2015年9月11日)

与えられた A/B テストがどういう結果になるか、ボクはまったく気にしてない。
もちろん、本心だよ。

だけど、待ってほしい。テストの結果を気にせずに、コンバージョン率を2倍または3倍にするには、どうすればいいんだろう？

ボクは実のところ、テストの**システム**全体に関心があるんだ。すべてのピースがぴったり合う必要がある。そうでなければ、A/B テストに多くの時間を浪費するわりに、何も得られないことになるだろうね。こういったことは、殆どのチームで起こるんだよ。

だけど、正しく行えば。正しいルールでプレーすれば。そうすれば、すべてのピースをぴったり合わせることができるし、ファネルのどの段階にいてもコンバージョン率を3倍にするのは、単に時間の問題になるんだ。

プレーを重ねればどんどん勝つように、ボクはシステムを設定した。上位に向かって十分な勝利を重ねていけば、コンバージョン率は3倍になる。そりゃ、与えられたテストが途中で失敗する可能性はあるよ。そんなこと、ボクは気にしない。

ボクの A/B テスト戦略はどのようなものかって？それは、とても簡単だ。

- ・ できるだけたくさんテストを繰り返して、10~40%の勝利となる1組のテストを見つける。
- ・ コンバージョン率を2倍か3倍にするために、上位に向かってこれらの勝利を重ねていく。
- ・ コンバージョン率を下げるような偽の勝者を出すことを避ける。

これをすべてやりきるには、7つの非常に具体的なルールに従う必要があるだろうね。どれも重要なものだ。ルールを1つスキップすると、システム全体がダメになる。ルールに従えば、ファネルはどんどん右肩上がりになるだろう。

ルール1:何はともあれ、コントロールが有効

ボクの A/B テストの見方は、殆どの人とまったく違う。

通常、誰かがテストを実行する場合、それぞれのバリエーションを等しいと見なす。コントロールとバリエーションはどちらも実行可能で、目標はどちらが優れているかを確認することだ。

ボクは、そのアプローチに我慢できない。

ボクたちは決定的な答えを求めているわけではないんだ。このステップでは、テストを繰り返して、どんどん上位に重ねていくことができる1組の偉大な勝者を見つけるんだよ。

もし、バリエーションとコントロールの違いが2%なら、どちらが**真**の勝者かなんてボクはまったく気にしない。ハイハイ、仮にこれらのテストで統計的に有意である(これについては詳述する)ことを示すのに十分なデータであれば、ボクだって2%の勝利を気にするだろう。

だけど、Facebook や Amazon でない限り、おそらくそれほどの母数はないよ。ボクは月に 100 万人以上の訪問者がいる複数のサイトで作業してきたけど、この手の変化が検出できるくらい十分なデータが 1 つのアセットにヒットしているなんてことは非常に稀だ。これをシステムで機能させるためには、バリエーションとコントロールのアプローチを変えなければならない。バリエーションそのものが明確な勝者であることを**証明**しない限り、コントロールは有効なんだ。言い換えれば、コントロールは**常に**勝者であるとみなされる。立証責任はバリエーションにあるんだ。バリエーションが勝たない限り、変更してはいけない。こうすれば、それ以降、確実に、アセットにプラスの変化のみをもたらし続ける。

ルール 2: 30 日以内に 2,000 人以上の人をテストに参加させる

トラフィックがまったくない? だったら、A/B テストはやめた方がいい。簡単なことだ。アセットをすっかり刷新してから、目測で検討を付ける。覚えておかなきゃならないのは、A/B テストの**システム**は連動している必要があるってことだ。そして、これは長期戦なんだ。つまり、適切な量のデータが必要で、だから様々なテスト・アイデアを繰り返す可能性があるってこと。1 つのテストを実行するのに 6 か月かかるなら、十分なテストを実行して少数の勝者を見つけることは絶対にできないだろうね。一般的に、ボクは、30 日以内にテスト中のアセットをヒットする 2,000 人以上の人を探す。だから、ホームページを A/B テストするなら、毎月 2,000 人のユニークな訪問者を獲得する方がいい。1 万~2 万人の人も好きだけど、ボクは月に 2,000 人くらいから始めるんだよ。それよりも少ないと、見合った価値はないだろうね。

ルール 3: 常に少なくとも 1 週間待つ

1 週間以内は、データは非常に不安定だ。テストを開始してから 24 時間以内に、99% の確実性で 240% の改善が見られた、なんてこともあったよ。これは勝者じゃない。そんなのはいつだって破綻する。最良のシナリオというのは、実際のところ 30~40% の勝利なんだ。最悪のシナリオは、事態が急展開して、実は 20% 減少した、という状況だ。あと、まるまる 1 週間のサイクル分のデータを取得することもできる。訪問者は、週末に平日と同じような行動をするとは限らない。だから、まるまる 1 週間分のデータがあれば、はるかに一貫性のあるサンプル・セットが手に入る。ボクがやったテストの 1 つで得た興味深い結果をお見せしよう。テスト開始直後は、10% のリフトがあるように見えた。テストを 1 週間実行したら、このテストに**完全な**急転直下が起こって、10% の敗者になってしまった(しかも、99% の確実性で)。



ボクが密かに疑っていることの 1 つは、ネットの周りで流れている 250%リフトっていうケース・スタディーの殆どが、最初の数日間で出現した極端な結果にすぎないということだよ。だから、もう少し長くテストを続けたら、控えめな上昇に落ち付いただろうってことだよ。急転直下で敗者に転落する場合もあるだろう。だけど、みんな勝ち名乗りをあげるが早すぎるから、Twitter で勝利宣言しまるくんだよ。

ルール 4:統計的有意性が 99%のバリエーションだけを取り上げる

待って、99%?95%はどうした?

もし A/B テストを実行したことがあるなら、95%の有意性に達するまで待つべきだ、という勧告に遭遇したことがあるかもしれないね。その場合、20 回のテストごとに 1 回の偽の勝者を選ぶことになるだろう。そして、誰も敗者を選びたいとは思わないから、ボクたちは通常このアドバイスに従うんだ。

君は、たくさんの A/B テストを実行した。たくさんの勝利を見つけた。その勝利を誇りに思っている。君は、巨大で幸せな A/B テストに泡のような誇りを感じているだろう。

じゃあ、ボクが君の A/B テストの泡のような誇りを壊して進ぜよう。

君の結果なんて何の意味もない。君は、20 人にたった 1 人、なんていうよりもはるかに多くの敗者を選んだんだ。ご愁傷様。

ちょっと振り返ってみよう。95%の統計的有意のルールはどこから来ていたっけ?

定量的な研究を行っている学術雑誌や科学雑誌を掘り下げみれば、どれにも 95%の統計的有意性が見つかるだろう。これは黄金律だ。

マーケターがテストを実行し始める場合、これと同じ標準を利用して、データが実際のところ

ろ何を語っているか確認するのは賢いやり方だ。でも、ボクたちは途中で重要なことを忘れちゃうんだよ。

ほら、テスト実行後にその統計的信頼度を評価することはできないじゃないか。君は最初にサンプル・サイズを決定する必要がある。これを行うには、ボクたちが検出を望んでいる最小限の改善率を決定するんだ。5%や10%という具合に。そうすれば、必要な統計的検出力を把握し、そこからサンプル・サイズを決めることができる。まだ混乱しているって? そうだよ、このようなことをするには、いくつかの統計学の知識が必要なんだ。テストのたびに、教科書を調べる必要がある。

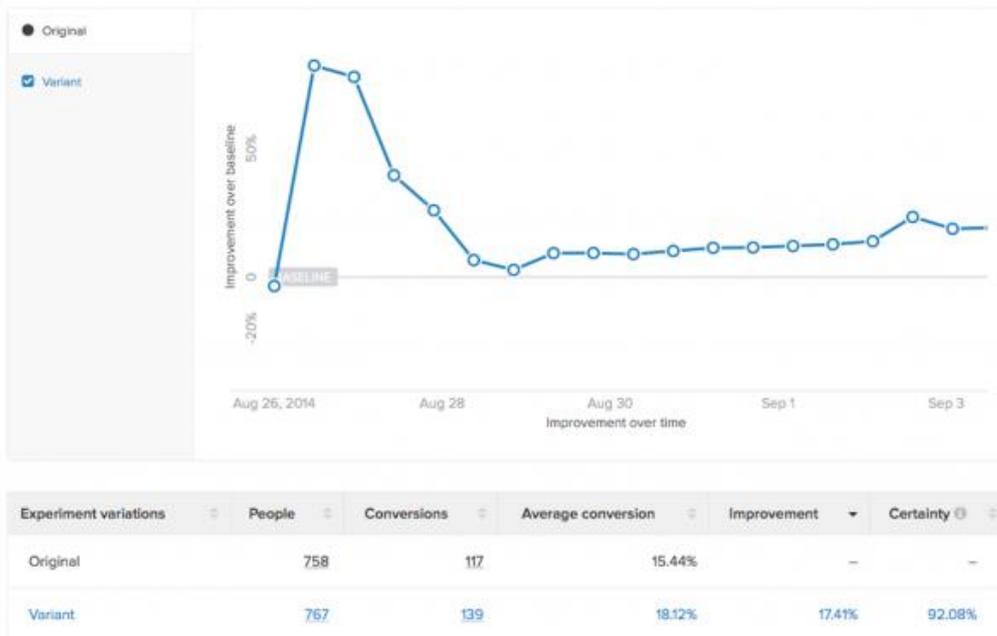
では、素敵な統計学のことはみんなすっ飛ばして、そんな心配をせずに95%の信頼度でテストを実行したらどうなるか?20回のテストのうち1回どころか、はるかに頻りに擬陽性検知が発生するね。

これはボクが実行したテストの例だ。最初の2日間で、97.7%の信頼度でコンバージョン率が58.7%増加した。



これは、殆どのマーケターにとって充分すぎるデータだ。ボクが知っている殆どの人は、これを勝者と呼び、発売して、先に進むことだろう。

それでは、1週間早送りしてみよう。偉大なる58.7%が勝ちだって?それは過去のこと。信頼度92%で17.4%になっているよ。



そして4週間後の結果は95.7%の確実性で11.7%の勝利にまで下落した。数週間で大勝利からわずかな勝利になってしまった。ここで安定するかもしれない。そうじゃないかもしれない。



データを収集するときに、テストは有意になったり、ならなかったりするんだよ。これが、必要なサンプル・サイズを決定することが非常に重要である理由だ。君だって、早い段階で

テストに騙されたいとは思わないだろう。

ところがどっこい!小さな勝者でも勝者のように見えるんだ!まだ発売すべきではないのかって?早まるには問題が2つある。

1. 長期的に勝者となった保証がない。もしテストを続けたら、更に確率が低下していたかもしれない。そして、たまには、急転直下、敗者になることもあるだろう。そうなると、それまでの勝者が苦労して獲得した勝利を失うことになる。
2. ビジネスへの予想される影響を大幅に盛っていたかもしれない。60%の勝利は山を動かす。そして、君の測定基準を押しつぶし、ボード・デッキを打ち砕く。一方で、11%の勝利は、君の成長に、はるかに穏やかな影響を与える。そして、君の測定基準に心地よいスパ・パッケージを与え、少しだけ正しい方向に誘う。早い段階で60%の勝利を呼び込むと、チーム全体が興奮しすぎてしまう。同じだけ膨らんだ希望と夢は、数週間後に成長はるかに穏やかになると押しつぶされる。それを何度も繰り返すと、人はA/Bテストの結果を信頼しなくなる。また、そこから間違った教訓を得て、ビジネスにまるでインパクトを与えない要素に集中し始めてしまう。

では、95%の統計的有意性が信頼できないとすると、ボクたちはどうすればいいんだ?

これをすっかり解決する簡単な方法があるよ。

ボクが Kissmetrics にいたとき、グロース・エンジニアの[ウィル・カート](#)と一緒に仕事をしていた。彼は今、[自身の統計ブログを運営している](#)最高に賢いやつだ。

ボクたちは、長期にわたる一連の A/B テスト戦略をモデル化した。すべての[データを網羅したブログ投稿](#)があり、[ウェビナーも行った](#)よ。抜群に統制のとれた学術研究戦略は、大雑把な 95%オンライン・マーケティング戦略と比較して、どうだろう?代わりに統計的有意性を 99%に引き上げたら、どうかな?

ボクたちは、99%の統計的有意性のルールを使ったら、長期的に非常に類似した結果が得られることを発見した。各テストで大変な統計作業を行う必要がなく、学術研究戦略と同じくらい信頼性があるんだよ。つまり、必要なサンプル・サイズもないのに 95%の統計的有意性を使うことは、殆どの人が考えるほど信頼性が高くないってことだ。

99%のルールは、ボクの A/B テスト戦略の基礎となるものだ。ボクは 99%の統計的有意性があったときのみ変更を加えている。それ以下だったら、コントロールから変えないんだ。こうして、偽の勝者を出す可能性を、もっと管理しやすいレベルにまで減少させ、悪いバリエーションで勝利を否定するというようなアクシデントを回避して、上位に向かって勝利を積み重ねることができるんだ。

ルール 5:テストが 10%のリフトを下まわったら、棄却する

よし、早速すべてのテストで 99%の確実性が得られるまで待つとしよう。

これだと、すべてのテストにかかる時間が劇的に増加しないかな?まさにそのとおり。

これがボクの最初の必殺技である理由は、そこにあるのさ。

もう一度言うけど、ボクはシステム全体に関心がある。ボクたちは勝者を見つけるためにテストを繰り返しているんだよ。だから、2~5%のテストに6か月もかけることなんてできないんだ。

それよりも君はどうしたいんだい？

- ・到達するのに6か月かかった確認済みの5%の勝者
- ・同じ6か月で6~12回のテストを繰り返した後の20%の勝者

5%の勝利なんてクソ喰らえだ、ボクに20%を寄越せ!

つまり、テストが長引けば長引くほど、機会費用がどんどん膨らんでいくんだ。待つのが長すぎると、他のテストを開始すれば見つかったはずの深刻な勝利を手放すことになるんだよ。

テストが10%のリフトを下まわったら、小さすぎて問題にならない。そんなのは却下。気を取り直して、次のテストへgoだ。

96%の確実性で8%の勝利が予測される場合はどうか? **とても近いね!** または、5%の勝利が素早く分かるだけの十分なデータがある場合はどうか?

その場合は、とても簡単な質問を1つ自分に問うてみる。このテストは、30日以内に確実な結果を出すだろうか? テストを開始して2週間で、99%の確実性に近づいているなら、もう少し粘ってみよう。ボク自身もそうしているよ。

30日で何が起こるかって?それが次の棄却ルールに繋がるんだ。

ルール 6:1 か月経っても勝者がいなかったら、棄却する

A/B テストの勝利を追いかけることは、中毒になる可能性がある。**データを、あと、もう一日、だけ。**

ボクたちは自分のアイデアに感情移入している。立ち上げたばかりの新しいページを愛している。そして、**勝利はとても近い。**もう少し粘ってみようか?**それで、いいだろう?**

分かるよ、これらのテストはいずれも個人的なペット・プロジェクトになるからね。そして、それを諦めることは悲痛を伴う。

勝利に向かっているテストがあれば、今のところはそれを続けよう。だけど、ボクたちは、ある時点で自分自身を断ち切る必要があるね。問題は、こういった“小さな勝利”のテストは、殆どが蜃気楼だってことなんだ。まず、15%勝ったように見える。それから10%。次に5%。そして2%。収集するデータが多くなれば、バリエントがコントロールに収束していくんだよ。

自分自身を断ち切る。ボクたちは感情を抑えるルールが必要だ。君はそうしないとイケない。大失敗したテストなんて忘れて、次のアイデアに進もう。

そういうわけで、ボクは30日棄却ルールを持っているんだ。バリエントが30日までに99%の確実性に達しない場合は、それを棄却する。98%でも、その場で中止して、先に進むわけ

だ。

ルール 7:データを待っている間に次のテストを仕込む

できるだけ速くテストを繰り返すことが、このゲームの鉄則だ。ボクたちは、テスト・パイプラインを**積み上げ**続ける必要がある。

テストとテストの間に停止時間なんかまったく**ない**方がいい。新しいバリエーションを仕込むのにどのくらい時間がかかるだろう?最初のアイデアから始めて、本番までどのくらいかかる?2週間?3週間?ひょっとして丸一ヶ月?

現在のテストが終了するまで次のテストを開始しないとすると、他の1~2個のテストに対して、結構な量のデータが無駄になるね。それは、他の勝利の上に積み重ねるべき20%の勝利を発見できた可能性のある別の1~2個のチャンスなんだ。

データを無駄にしてはいけない。これらのテストをフル・スピードで実行し続けないと。

1つのテストが終了したら、すぐに次のテストを始めるんだ。毎回。

そうさ、A/B テストに専念するチームを編成する必要があるだろうね。これは些細な作業ではないよ。君は、A/B テストをフル・タイムで開始することになるだろう。そして、君のチームは何の障壁もなくフル・スピードで動く必要がある。

それが簡単なことなら、みんなやっているだろう。

7つの A/B テスト・ルールすべてに従って、コンバージョン率を一貫して右肩上がりにする

このシステムに忠実に従えば、コンバージョン率を2倍または3倍にするのは時間の問題だ。長くプレーすればするほど、勝つ可能性が高くなるだろう。

すべてのルールをまとめると、次のとおりだ。

- 1.何はともあれ、コントロールが有効
- 2.30日以内に2,000人以上の人をテストに参加させる
- 3.常に少なくとも1週間待つ
- 4.有意性が99%のバリエーションだけを取り上げる
- 5.テストが10%のリフトを下まわったら、棄却する
- 6.1か月経っても勝者がいなかったら、棄却する
- 7.データを待っている間に次のテストを仕込む