



**Data Science  
Academy**

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

**Business Analytics**

**Análise de Testes A/B**



## **Elevação (Lift)**

Normalmente, a análise de Teste A/B mede a diferença entre a versão B e a versão A. Para uma métrica de resultado  $x$ , a diferença entre o teste e o controle é  $x_B - x_A$ . Essa diferença, especialmente para resultados cumulativos, pode aumentar com o tempo. Considere o exemplo de gasto por usuário exposto. Conforme o Teste A/B continua, os dois grupos continuam comprando e acumulando mais gastos. A versão B é um sucesso se os gastos do grupo de teste aumentarem mais rápido do que os de controle.

Em vez da diferença, medimos a elevação (Lift) de B sobre A. A elevação dimensiona a diferença pelo valor da linha de base. Para uma métrica de resultado  $x$ , a elevação do teste sobre o controle é:

$$(x_B - x_A) / x_A * 100\%$$

Descobrimos que o aumento das métricas cumulativas tende a ser estável ao longo do tempo. Isso é o que chamamos de Lift.

## **Análise de Potência**

Antes de iniciar um Teste A/B, é bom fazer duas perguntas:

- 1- Que porcentagem de usuários deve fazer o teste versus o controle?
- 2- Quanto tempo o teste precisa ser executado?

A maneira estatística formal de responder a essas perguntas é uma análise de potência. Primeiro, precisamos saber qual é a menor diferença (ou aumento) que seria significativo para o negócio. Isso é chamado de tamanho do efeito. Em segundo lugar, precisamos saber o quanto a métrica de resultado normalmente flutua. A análise de potência calcula o tamanho da amostra, o número de usuários necessários para detectar esse tamanho de efeito com significância estatística.

Existem dois componentes para usar o tamanho da amostra. A divisão é a fração de usuários em teste versus controle e isso afeta o tamanho da amostra necessário. O tempo para o teste é o tempo total que levará para expor tantos usuários. Como os usuários podem voltar e serem expostos novamente, o número cumulativo exposto aumentará mais lentamente com o passar do tempo. Matematicamente, quanto mais desequilibrada a divisão (quanto mais distante de 50-50 em qualquer direção), mais longo será o teste. Da mesma forma, quanto menor o tamanho do efeito, mais longo será o teste.

Muitas vezes, a análise de potência não conta toda a história. Por exemplo, se em um portal online temos um forte ciclo semanal - as pessoas compram de maneira diferente nos fins de semana dos dias de semana.



Sempre recomendamos a execução de Testes A/B por pelo menos uma semana e, de preferência, em múltiplos de sete dias. Obviamente, se os resultados parecerem drasticamente negativos após o primeiro ou dois dias, é bom desligar o teste mais cedo.

O equilíbrio da divisão afeta a duração da execução do teste, mas também consideramos o nível de risco. Se tivermos uma grande quantidade de produtos à venda, podemos começar com 90% de controle, 10% de teste. Por outro lado, se quisermos ter certeza de que um recurso importante no site continua fornecendo sustentação, podemos manter um controle de 5% e teste de 95%. Mas, se tivermos um teste de baixo risco, como uma pequena mudança na interface do site, uma divisão em 50% do controle, o teste de 50% significará um tempo de teste mais curto.