

Eltecon Data Science Course by Emarsys

Measuring effect through experimenting

András Bérczi

October 9, 2019

About me

- Background in Economics
- Works as Data Scientist @ Emarsys

What is an effect?

What is an effect?

Added value of a treatment

Why do we want to measure the effect?

To decide if our treatment works

What to measure?

What is worth measuring?

- **Based on customer's need**
- Validated by data based research
 - Will the algo work?
 - Does it scale?
 - Cost of the feature?
- Make sure you understand your feature/algorithm!

Measure the effect of what?

- Adding a new feature to the software
- Change in the algorithm
- Change on the website/UI
- etc.

How can we measure the effect?

- Simulation
- Based on historical data
- **Experimenting**

How do we experiment?

Setup

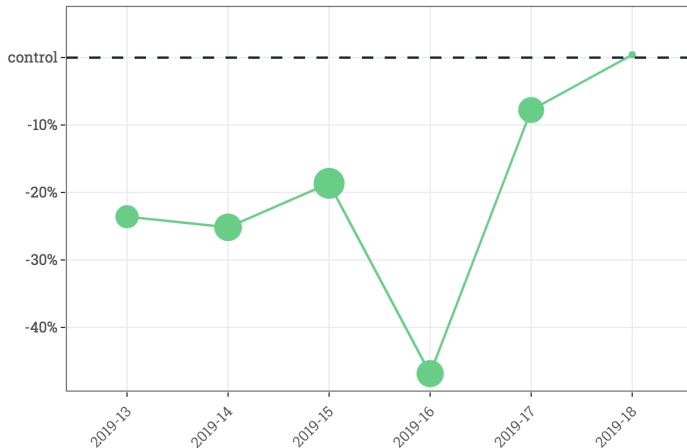
- Define the goal
- Measure one feature at a time
- (Or control for other effects)
- Split contacts *randomly* into control and treatment group(s)
- Do not change parameters during the experiment (Simpson-paradox)

There is always an effect. . .

- We can always measure something.
- Is there really an effect?

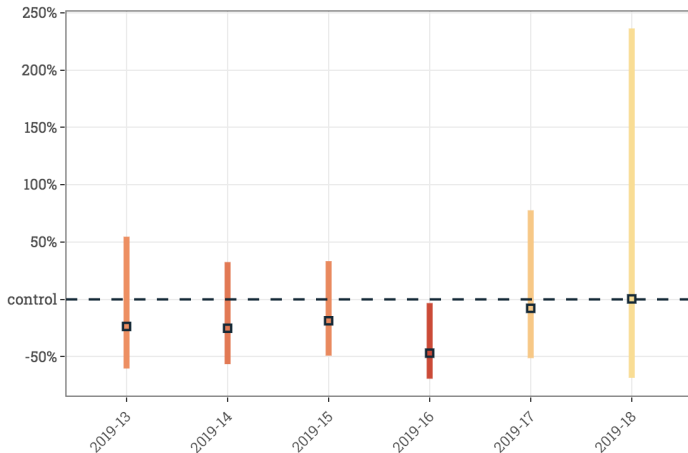
There is always an effect. . .

STO's effect on open rate



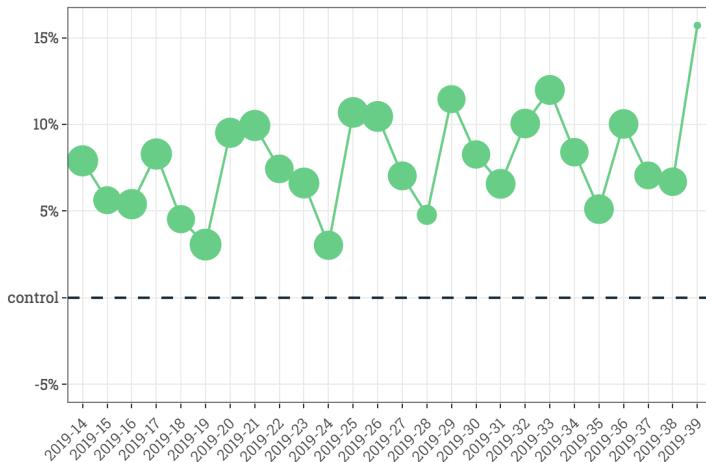
But not necessarily significant!

STO's effect on open rate



Know your data!

STO's effect on click rate



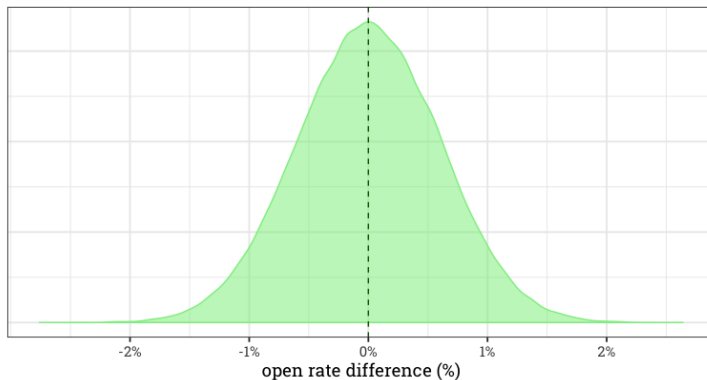
Minimum Detectable Effect

A great blogpost by a great guy

Detectable Effects for Useless Feature

Distribution of Detectable Effects

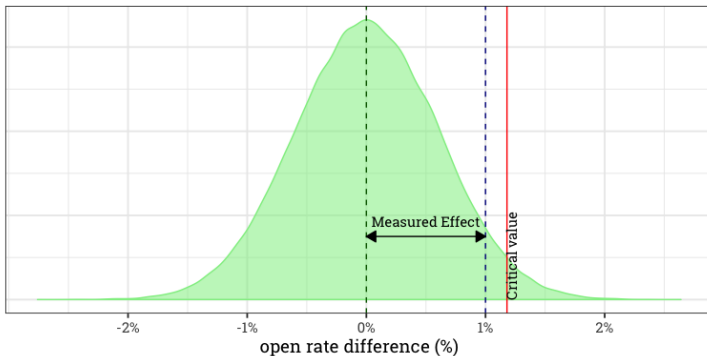
$n = 10,000$



Hypothesis testing

Distribution of Detectable Effects when there is actually no difference in open rates

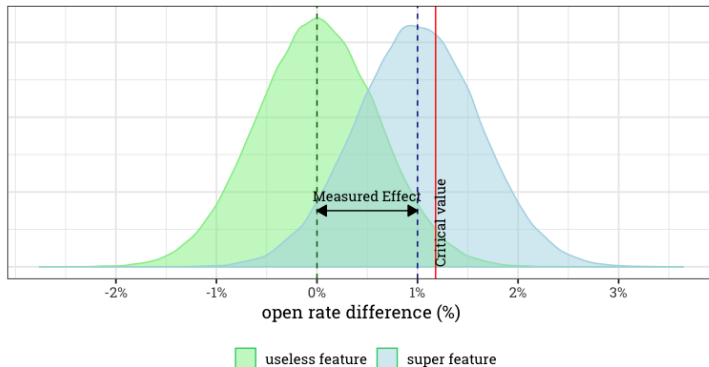
$n = 10,000$



But what if we really have an effect?

Distribution of Detectable Effects

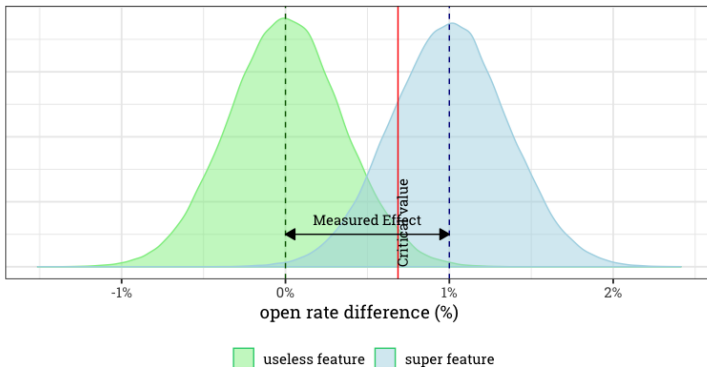
n = 10,000



Use more data points!

Distribution of Detectable Effects

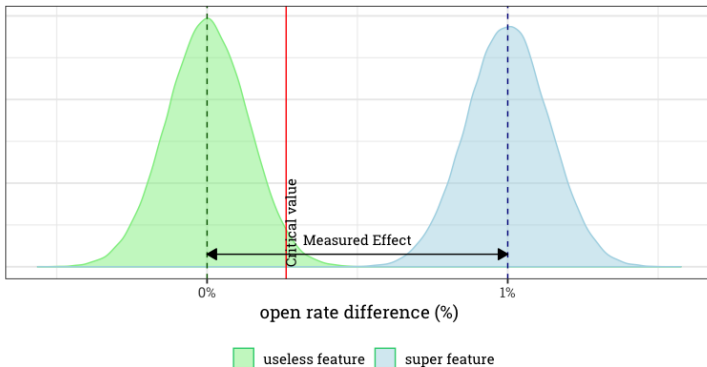
n = 30,000



Or even more!

Distribution of Detectable Effects

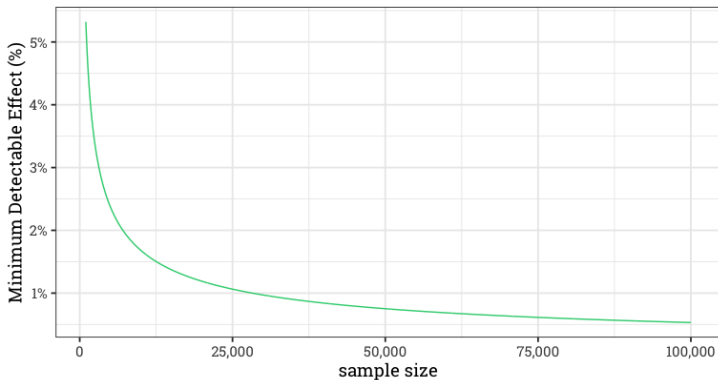
n = 200,000



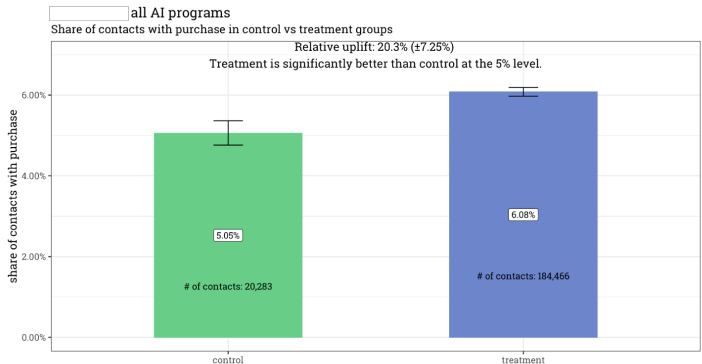
We can calculate this in advance!

Distribution of MDE given different sample sizes

When base rate is 10%, with 95% significance level and 80% power



How to present



Contact behaviour is measured for 7 days from entering the program (currently until May 22, 2019)

How to present: Shiny app from Emarsys