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

We want to understand the effect of a new feature

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 experiment on?

What is worth experimenting?

- **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)

How to calculate the effect

$$Uplift_{KPI} = \frac{KPI_{treatment}}{KPI_{control}} - 1$$

How to aggregate the effect

$$KPI_{treatment} = \frac{\sum_{i=1}^{n} KPI_{treatment,i} * SampleSize_i}{\sum_{i=1}^{n} SampleSize_i}$$

$$KPI_{control} = \frac{\sum_{i=1}^{n} KPI_{control,i} * SampleSize_i}{\sum_{i=1}^{n} SampleSize_i}$$

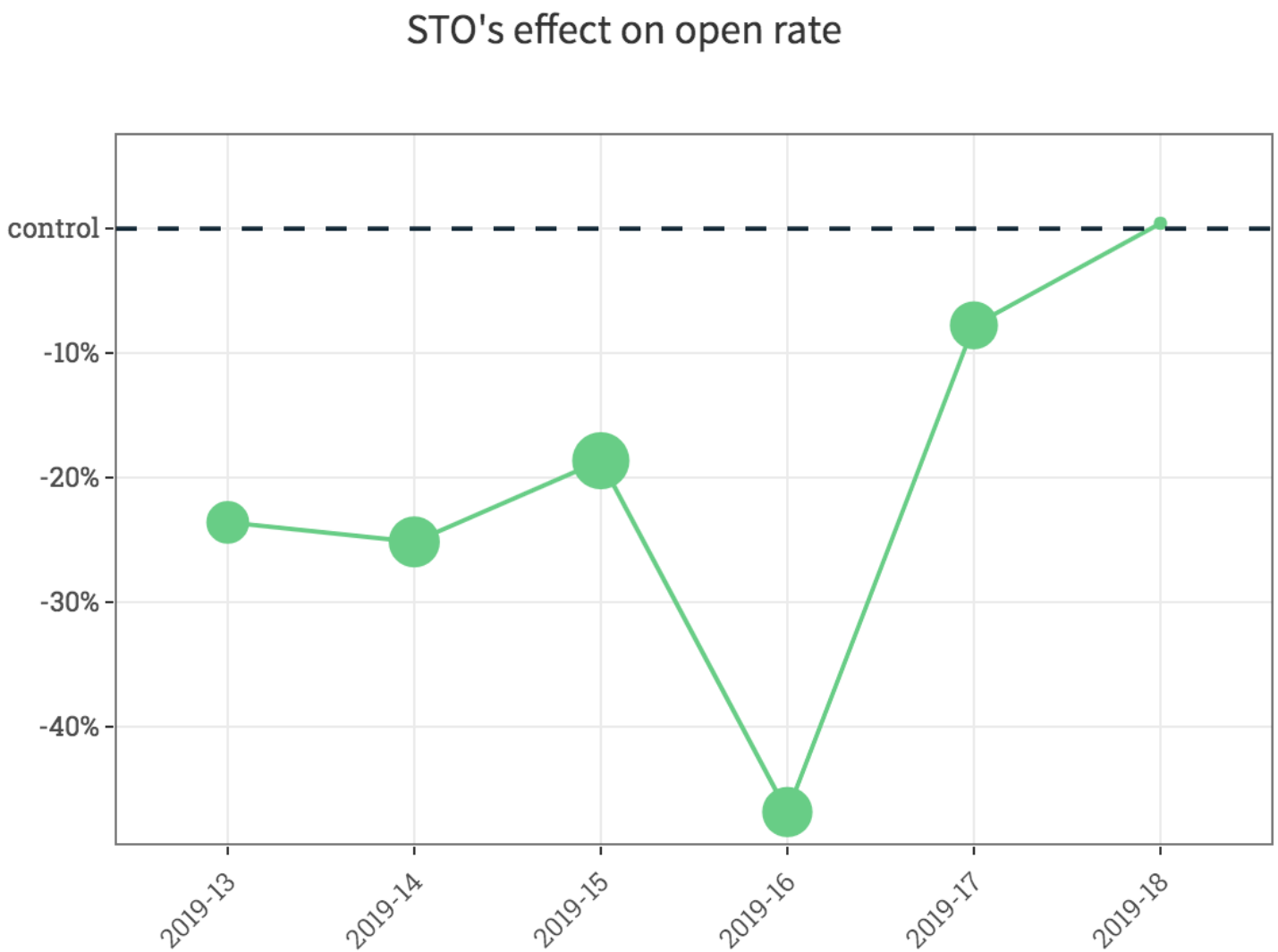
$$Uplift_{KPI} = \frac{KPI_{treatment}}{KPI_{control}} - 1$$

, where n is the number of unit levels.

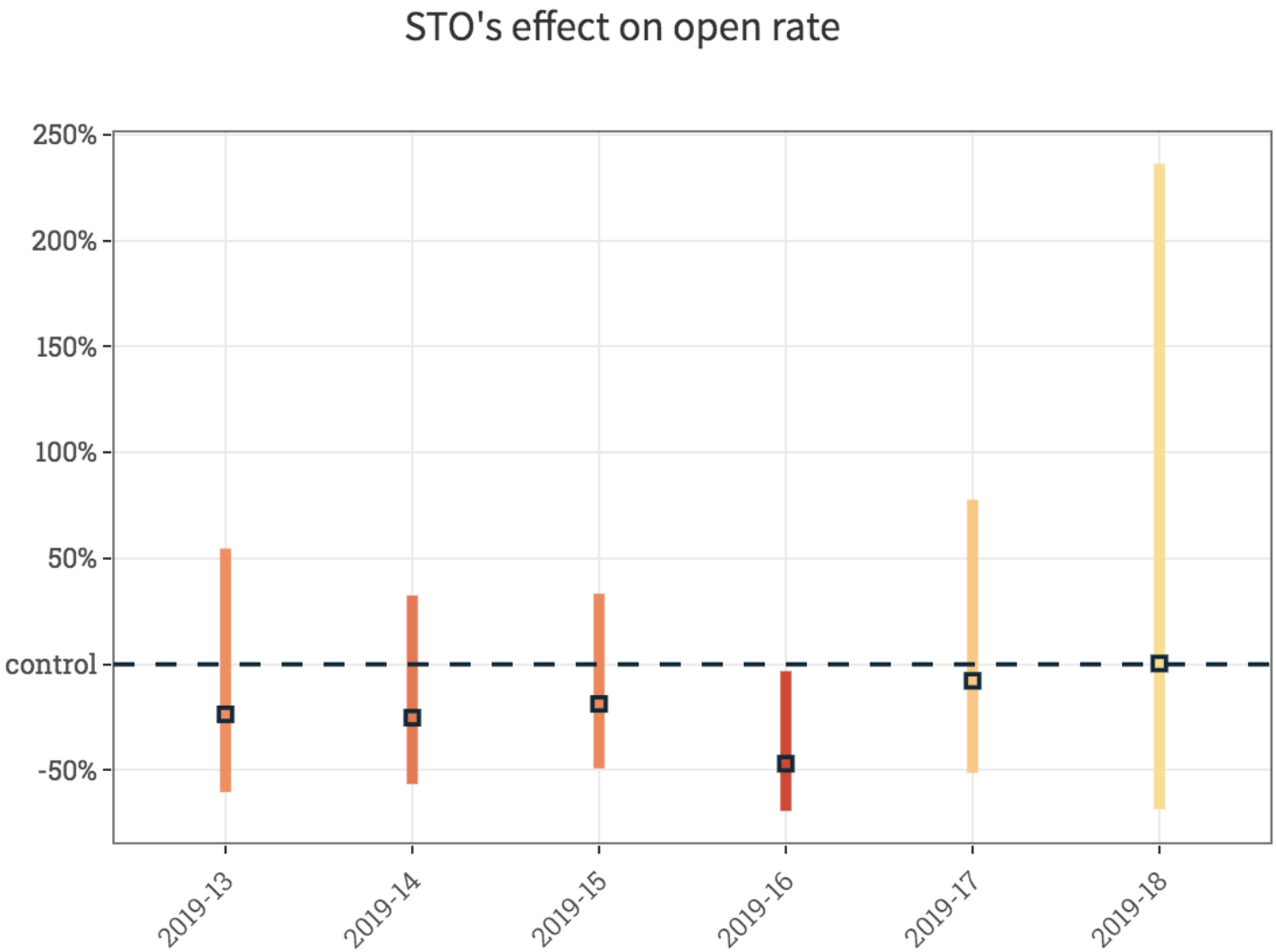
There is always an effect...

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

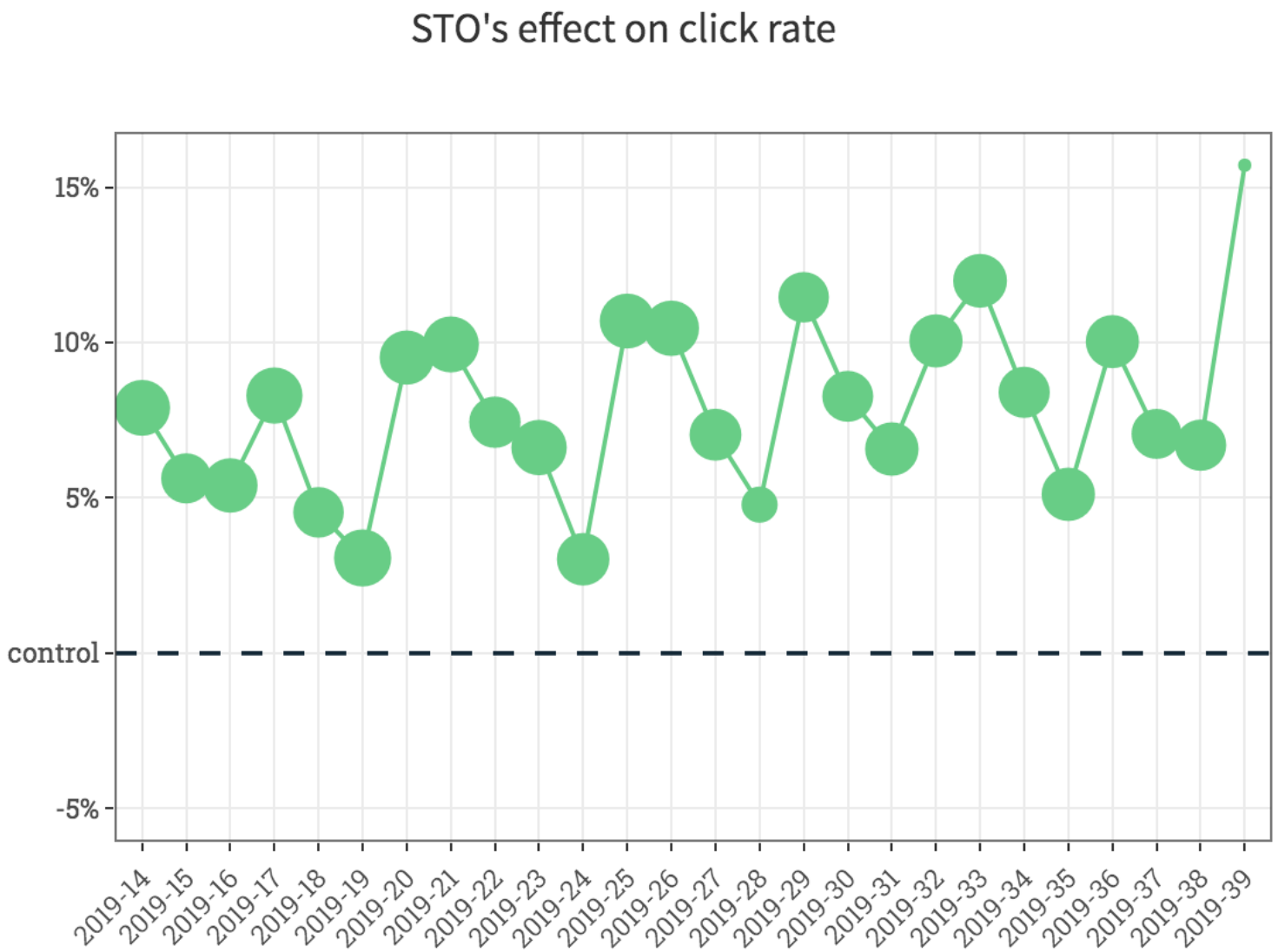
There is always an effect...



But not necessarily significant!



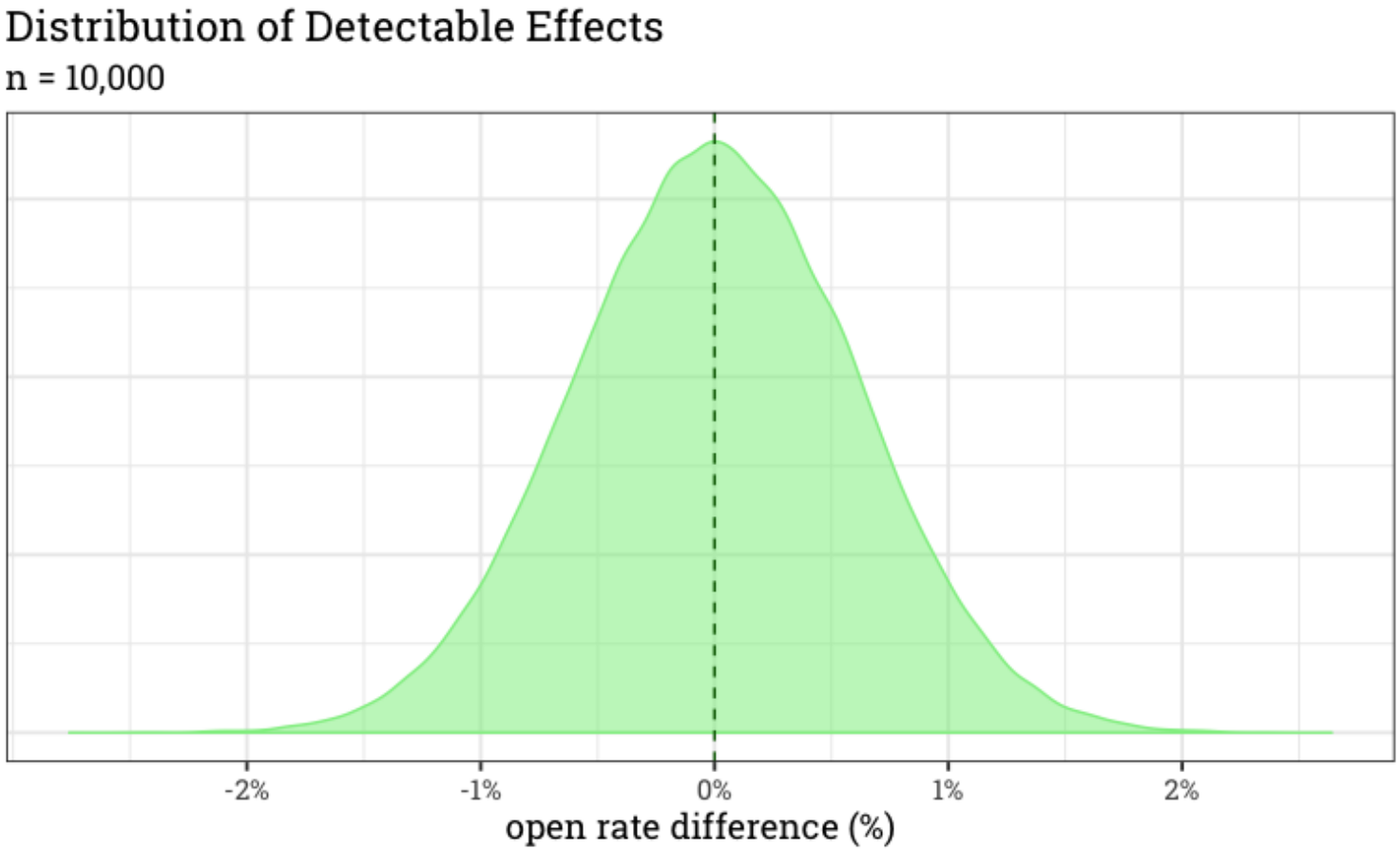
Know your data!



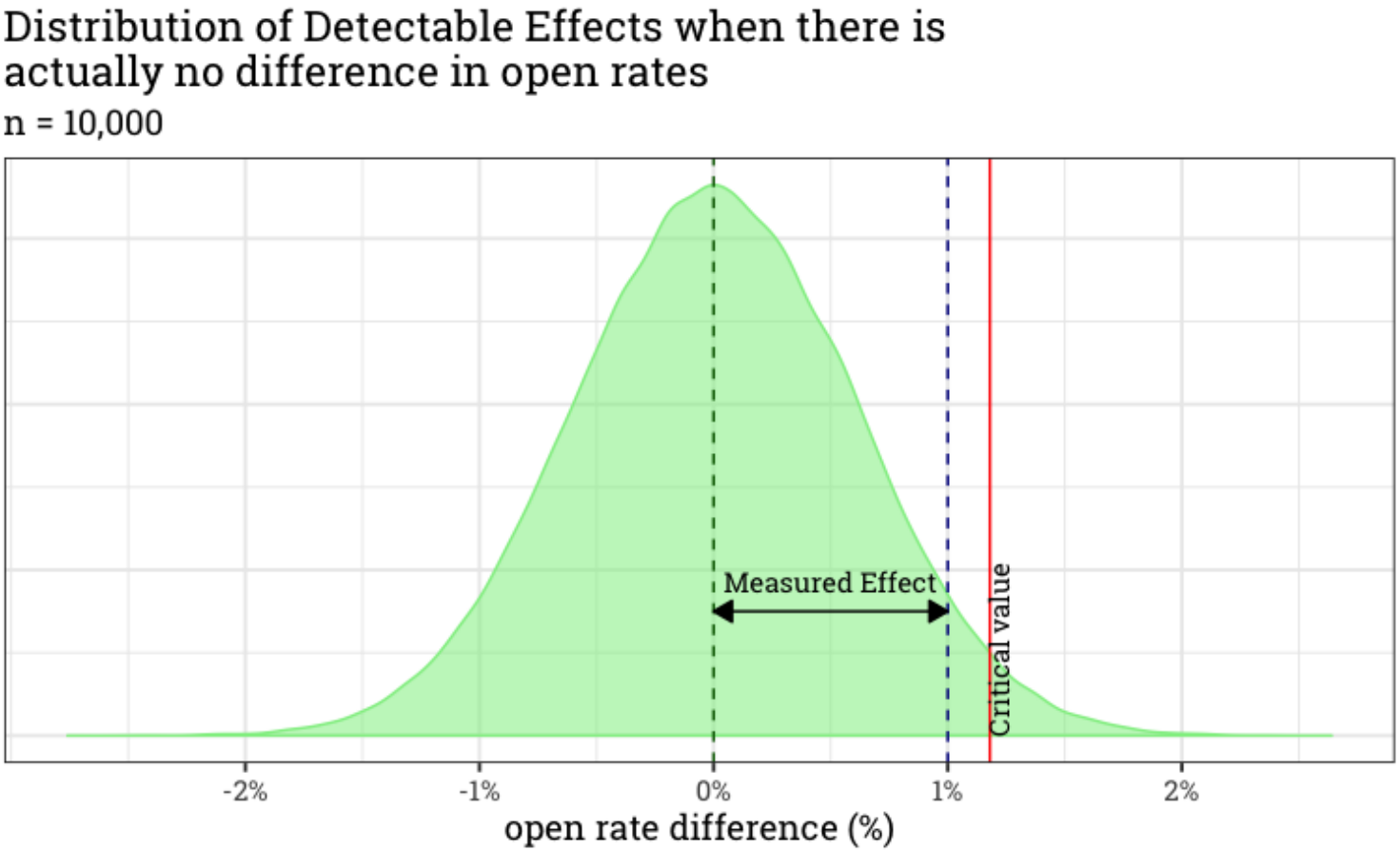
Minimum Detectable Effect

A great blogpost by a great guy

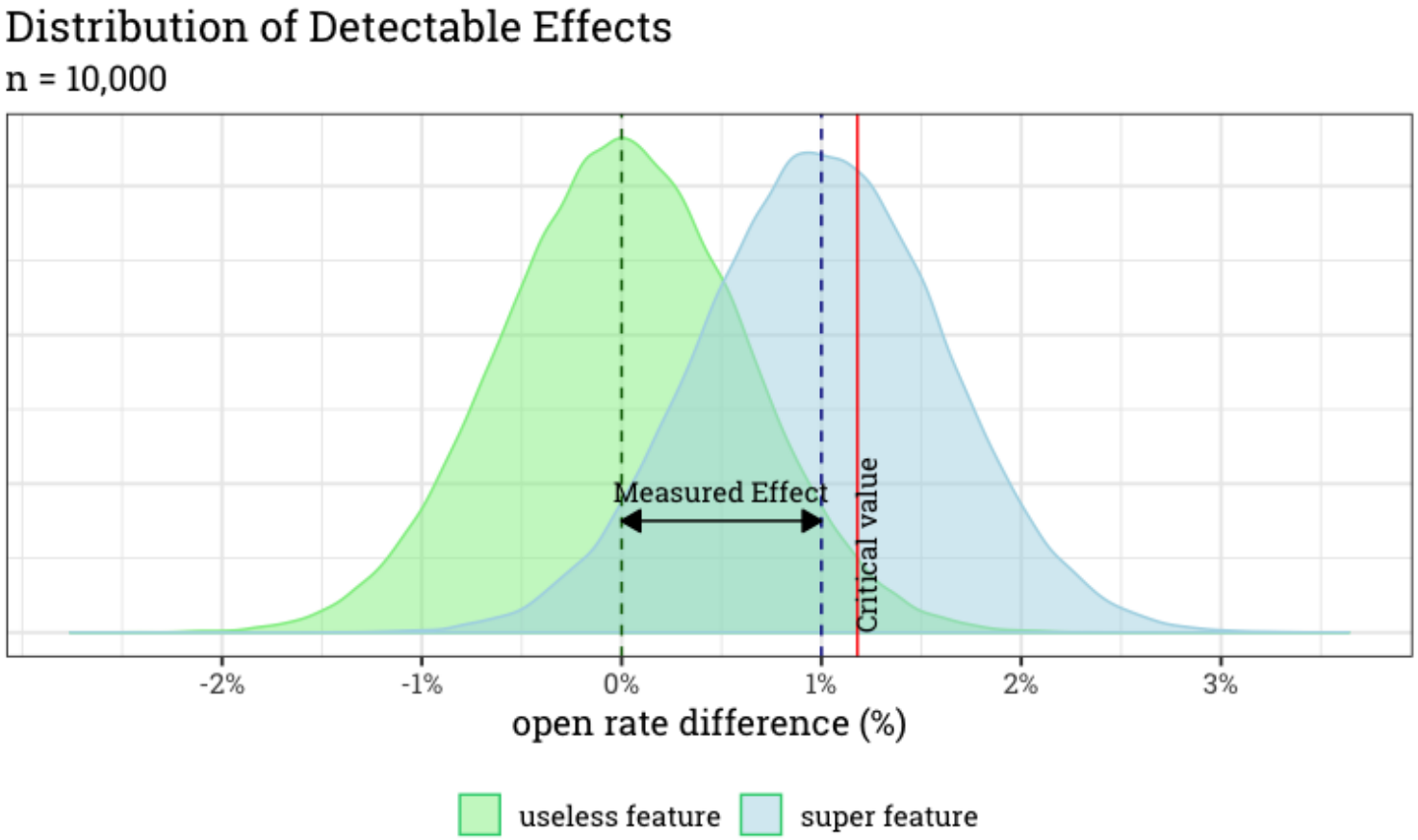
Detectable Effects for Useless Feature



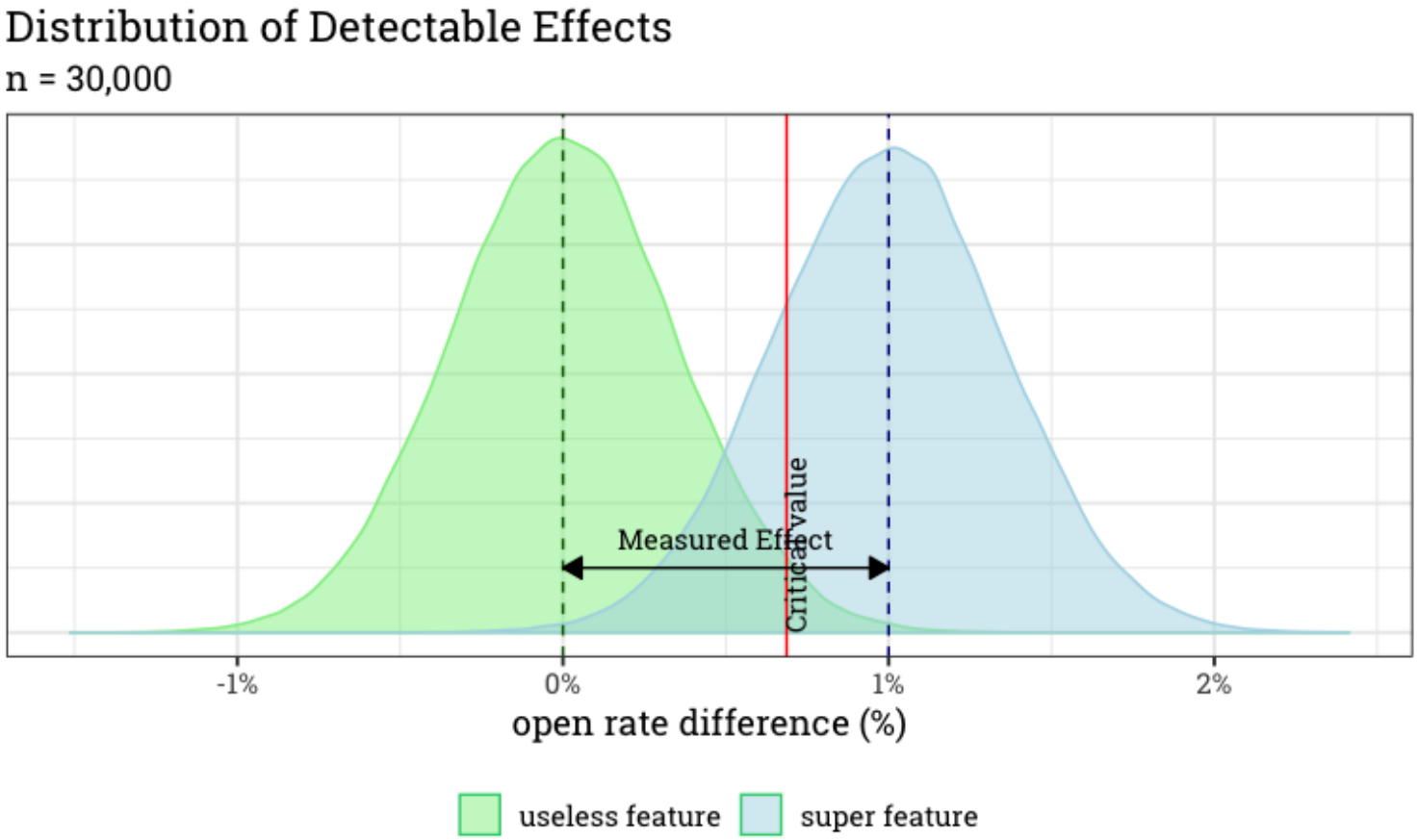
Hypothesis testing



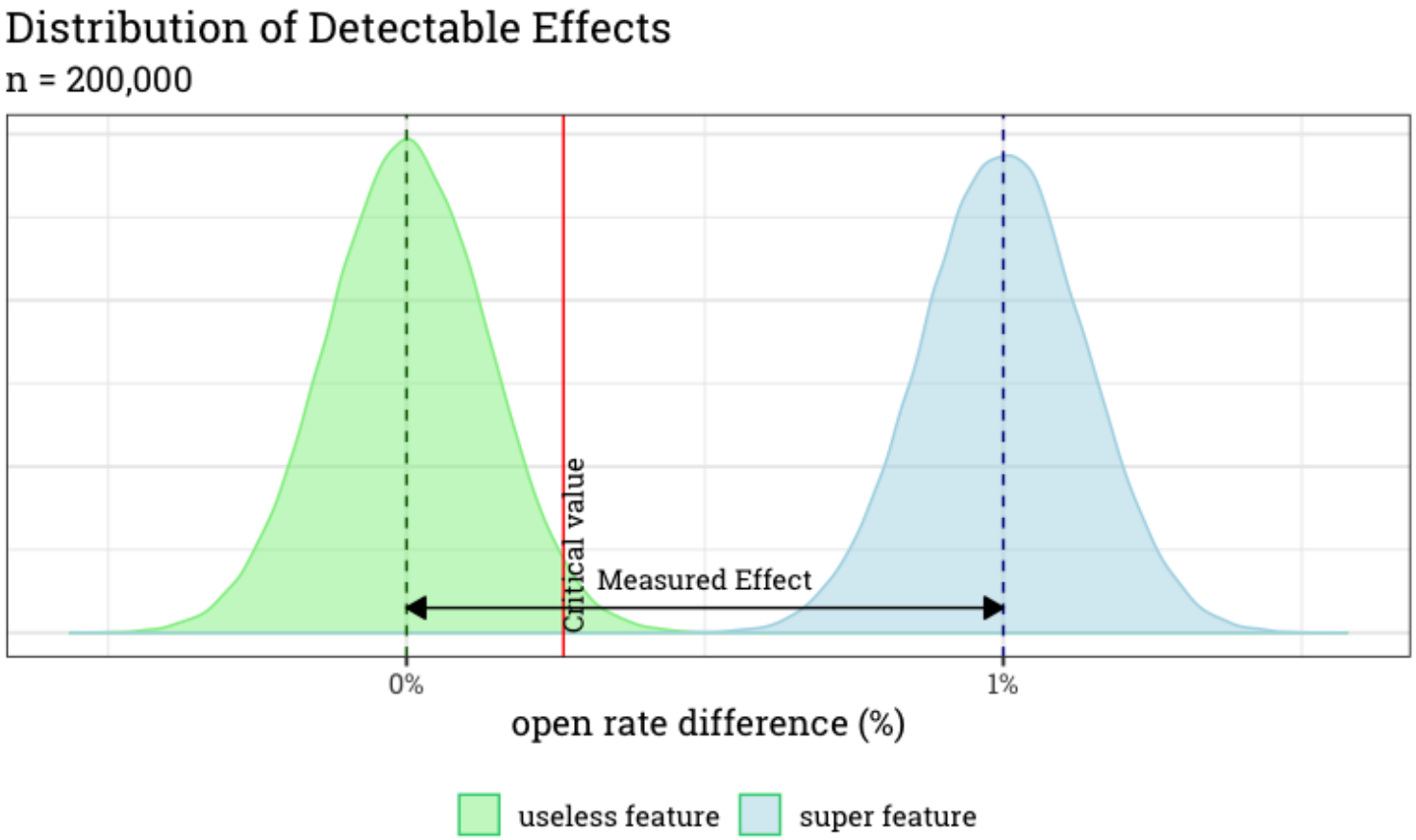
But what if we really have an effect?



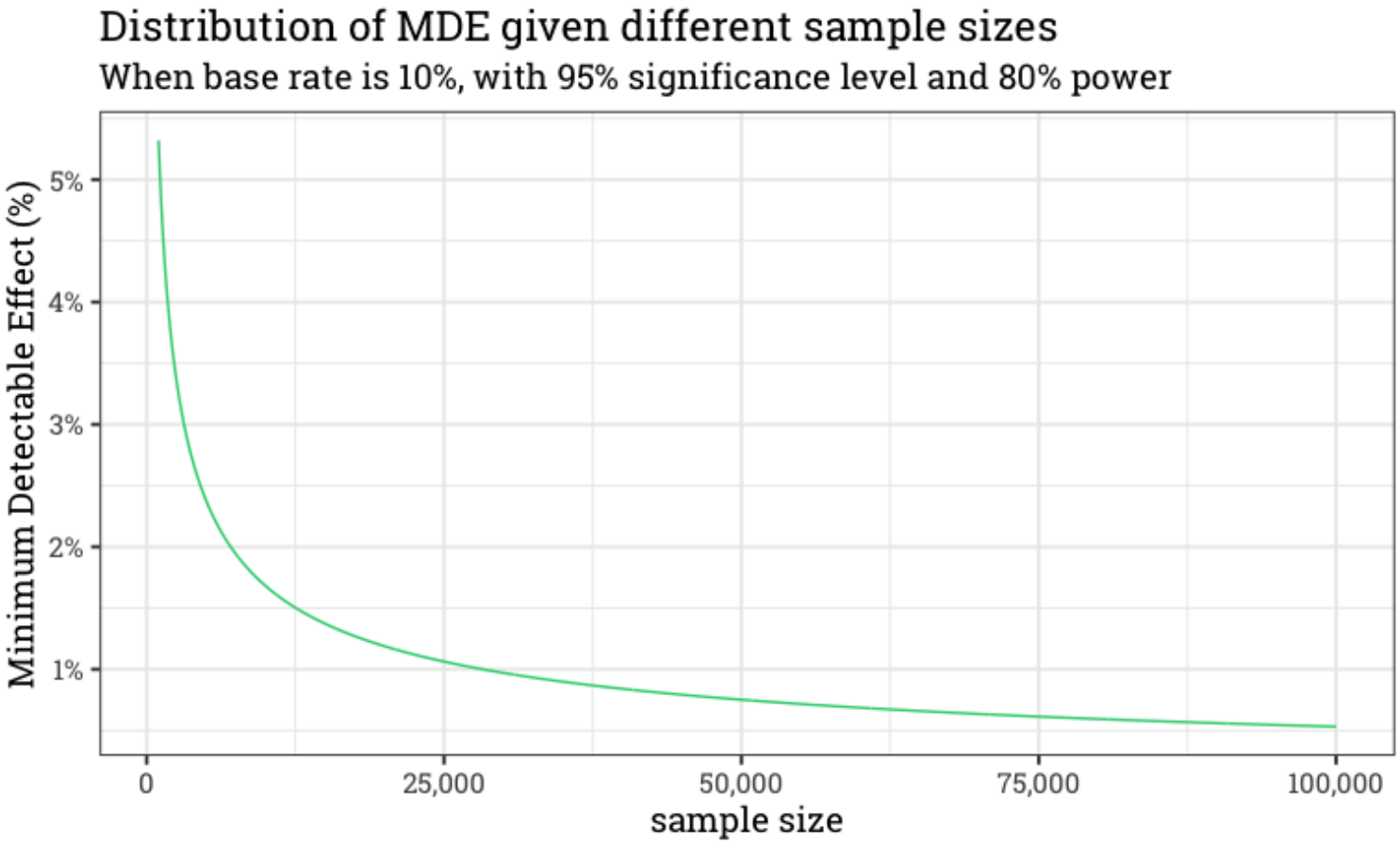
Use more data points!



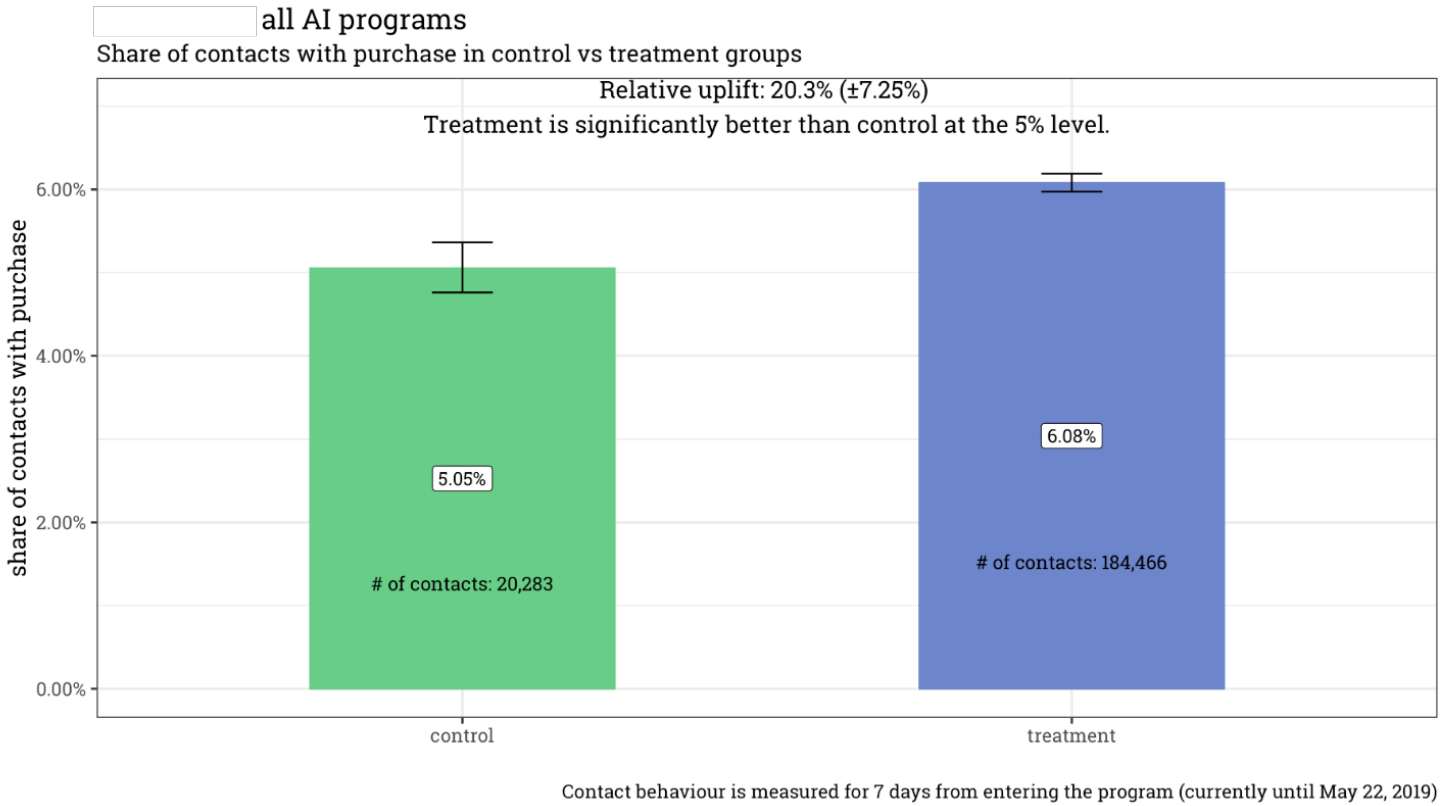
Or even more!


Distribution of Detectable Effects
n = 200,000
Critical value
Measured Effect
0%
1%
open rate difference (%)
useless feature
super feature

We can calculate this in advance!



How to present



How to present: Shiny app from Emarsys