Eltecon Data Science Course by Emarsys

Measuring effect through experimenting

Sándor Sóvágó

November 24, 2021

Section 1

Why talk about effect measurement in Data Science?

What is an effect?

Added value of a treatment (the new feature)

How can we measure the effect?

- Simulation
- Based on historical data
- **Experimenting** (A/B test or something like that...)

Section 2

How do we experiment?

What to experiment on?

- **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!

Setup

- Together with customer or darklaunch?
- Define the goal
  - Measure one feature at a time
- Definte KPIs
  - How much added value does it bring to the user?
- Make sure you have enough data to measure significant results
  - check the minimum effect you can measure beforehand!
- Split contacts *randomly* into control and treatment groups
- Do not change parameters during the experiment

What are good KPIs?

Treatment receives personalized email (80% of contacs), control receives "standard" email (20% of contacs).

id	group	did_open	did_click	sales_amount
1	control	0	0	0
2	treatment	0	0	0
3	control	1	1	30
4	treatment	0	0	0
5	control	1	0	50
6	treatment	1	1	12
...	...	...	...	...
100	treatment	1	0	0

How to calculate the effect

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

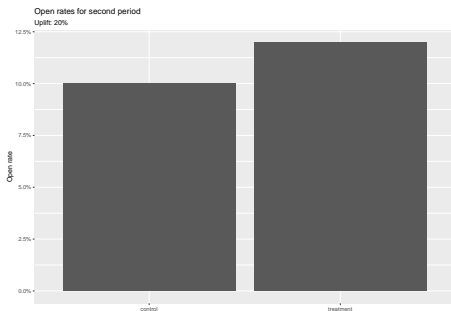# Calculate the effect for the first period!

Use `experiment_results.csv`

Calculate the effect for the second period and plot the results!

Use `experiment_results.csv`

When you are done with both the calculation and the plot, enter your results in the chat!

The *design* should look something like this:

Calculate the effect for the whole period and plot the results!

When you are done with both the calculation and the plot, enter your results in the chat!

Do you notice anything weird?

What's happening here?

We have too much 'bad' control! (Let's check the code!)

One way 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.
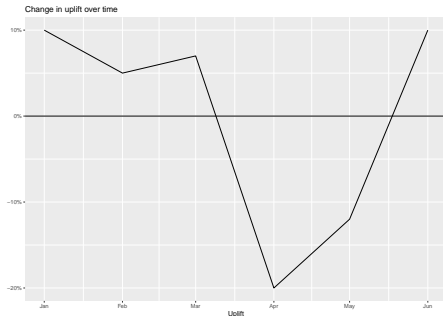
Calculate the effect for the whole period and plot the results!

Weight the open rates by the number of contacts present in the period!

When you are done with both the calculation and the plot, enter your results in the chat!

Plot the effect (uplift) over time!

Use `experiment_results_over_time.csv`
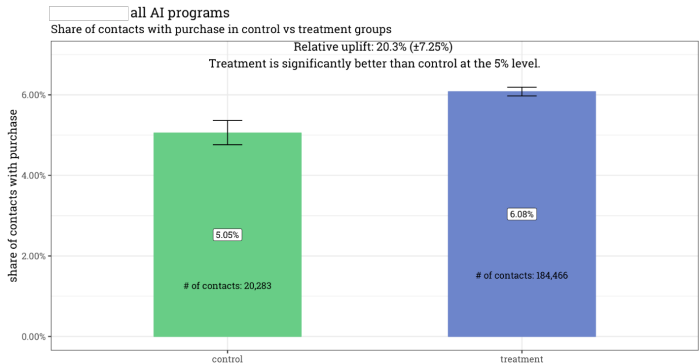
The *design* should look something like this:

Section 3

How to present

How to present

What is missing from the previous plots?

How we do it in Emarsys

How we do it in Emarsys

Shiny app from Emarsys

Section 4

Homework for next week

Homework for next week

Use `experiment_result_HW.csv` for your homework. Define KPIs which describe the best the succes of the experiment. Calculate and plot the uplift for the *whole* period for each KPI.

You should send the plots along with a short explanation how you assess the results based on the KPIs.

Section 5
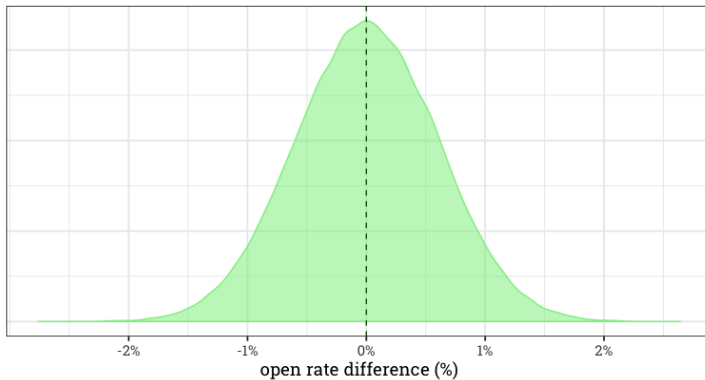
How to design experiments?

Minimum Detectable Effect

A great blogpost by a great guy

Detectable Effects for Useless Feature
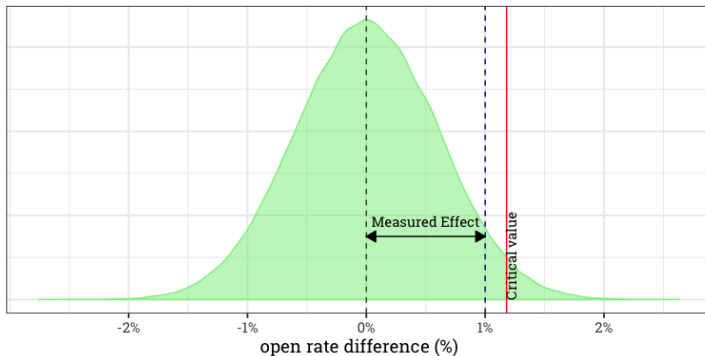
Distribution of Detectable Effects

n = 10,000

open rate difference (%)

Hypothesis testing

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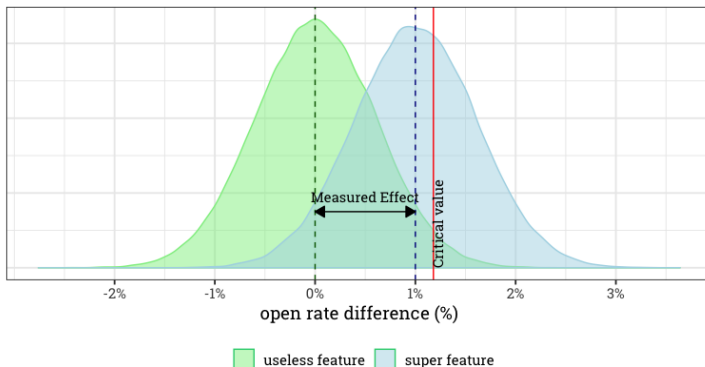actually no difference in open rates

n = 10,000

But what if we really have an effect?

Distribution of Detectable Effects
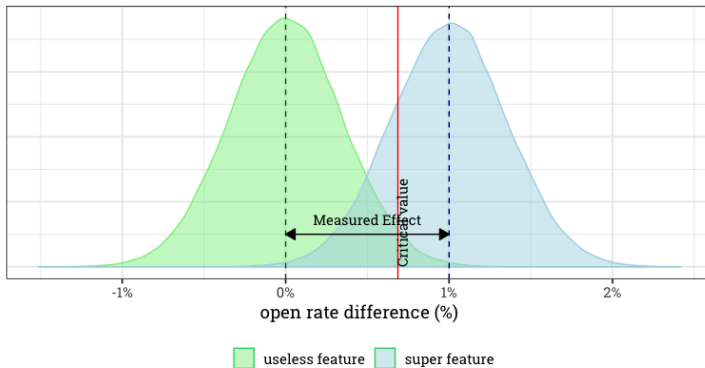n = 10,000

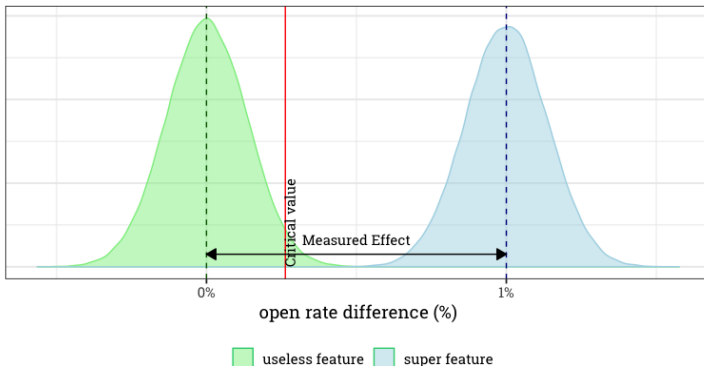# Use more data points!

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