

# Resp\_takehome

Martin Gascon

May 22, 2015

## Question 1

With the available data, what do you believe the results suggest? Do you have more confidence in Acme's data or the third party data? Why?

The available data suggest that there is an improvement but is not statistically significant. In detail, if we consider the raw data only, the completion rates for the old and new pages are 10.08% and 10.14% respectively, which means there is a lift of only 0.6%. Now, if we remove repeated users to have only independent cases, the lift goes up to 1.009% and that's why the experiment was stopped. The question is if this lift is statistically significant and the evidence (t-test) suggest that we need more data points to conclude that.

Regarding the second question, I have more confidence on Acme's data because, third party data stopped the experiment without taking into account the statistical fluctuation.

## Question 2

Knowing that the project manager is very smart, but that she is not a statistician, how would you explain it? What suggestions would you make for a future test?

I would say the following: The experiment stopped because we reached a difference of 1% after removing repeated users. These users were removed because the experiment has to take into account only independent cases (users have a different attitude after having responded the first time). The main problem is that the data suggest an improvement that is not statistically significant, which means the variation of 1% is within the limits of fluctuation when using such a low amount of cases (real difference is  $<100$  cases). That is why we need to continue the experiment.

In other words, if we want to stop an hypothetical experiment of flipping an unbalanced coin when the difference is 55%, we can't stop the experiment after 10 flips because most likely the rate won't be

50/50 but 60/40 or 30/70 and even though we reached the desired difference the number of needed flips to have a real (statistical) difference is hundreds of thousand of times.

In the future I would suggest to have the IP address (or Time zone) so I can have a completion rate for the different times of the day (see other possible questions).

## **Other possible questions**

- Using the time stamp (we would also need Time Zone) we could answer the following questions:
  - 1) Do we have a different completion rate if the user responded during the weekend rather during weekdays? Presumably users have more time and are willing to spend some time responding an quality survey.
  - 2) For the same reason, if we consider only weekdays, do we have a different completion rate if the user responded during night hours rather than working hours?