AB testing report

1. The first approach is to use conversion rate as a gauge for the significant test. The conversion rates are binary(Binomial distributed) and the normal distribution assumption for the t-test is not fulfilled. Therefore, we use the χ²-test which is appropriate for the binomially distributed data.

The sample size of the control group and test group are almost equal. So, there is no need for normalization to avoid bias in our significant test.

The hypotheses to test are:

H_0: "the conversion rate is the same for the two versions"

H_a: "the conversion rate is higher for the test version "

Result:

The p-value is, p= 0.0895. This is $p \simeq 9$ %.

we observe that the p-value is larger than the significance level i.e alpha=0.05. So, we cautiously accept the null hypothesis. However, one needs more information, other user experience features, to be sure that the Type-II error is not committed here. Also, further power analysis for this significant test is suggested.

2 . We can also observe if the change in the website environment, changes the number of characters translated, Thus, the second approach is to use the z-test for the number of characters translated. We use the z-test as the sample size is sufficiently big. Alternatively, one can use a t-test which gives a similar result.

Result:

The p-value is, p= 0.00000023.

The p-value is significantly smaller than the significance level i.e alpha=0.05. So we reject the null hypothesis. We should also carefully interpret this result as the causal relationship is not very clear for this metric. Additional metrics can help for the interpretation of this result.

Comments:

I. From the significant test for the number of characters translated, we can cautiously suggest that users like the change in the website, but from the significant test for conversion we can suggest that the change is not enough to lead to (significant) an increase in conversion rate, hence profit.

II. The conversion rate and the number of translated characters are the available metrics in this data frame. The importance of conversion rate is obvious, as it is related to the profit of subscription for the company. This is a metric which is called directional, as it directly influence the revenue. It is the main metric to measure the effectiveness of the website improvements and the first metric to use for the significant test.

The number of translated characters is also important to measure for the user experience of the new version of the website though not sufficient. Such metrics are called sensitivity metrics, as opposed to directional,and are important in the long term performance. Besides applying the significant tests for this metric, there is another implication here. Together with other sensitivity metrics, which are absent in this data frame, one can filter data for the conversion rate significant test, mentioned above. This filtering extracts data for the cases in which the users paid sufficient attention to the website.

III. We need other sensitivity features to be able to detect how much the users pay attention to when visiting the website. The number of characters translated is not sufficient. Following, there are some suggestions for other necessary features for better understanding of the user's experience:

i. Time spent on the website: This is a better metric than the number of characters translated. It shows how much time and also attention is spent on the website by users.

ii. user-id: to detect the users separately rather than using the season-id.

iii. Number of visits for each user: it is also related to the overall attention given to the website.

iv. Click-through rate for each user: more sensitive metric than conversion.

V. Users return rate: important metric that indicates that the content of the website has improved.