WEEK 4 INTERIM SUBMISSION

NAME: KEN MBELENZI MBAYA

Which online users belong to the control and exposed groups?
- Controlled users are those who have been showed a dummy ad
- Exposed users have been shown a creative and interactive online ad

How are the users targeted?
- Users are targeted at random since the website server sends them to different versions at random

Could we use the counts of yes and no answers to make a judgement on which experiment is performing better? For example, if #yes > #no for the exposed group than the control group, could we declare that the ad had a significant impact Why or why not?
- No, we cannot use the counts of yes and no because probably the sample size being used is not large enough and does not represent the whole population
- The sample may have been split unequally making the number of people in the controlled group larger that the number of people in the exposed group and vice versa
- This might be a false positive
- Only a portion of the visitors have seen the ad this does not fully represent how the other people who have not seen it will react

What is the statistical process that generates the data? Which kind of statistical model will you use if you were to simulate the data?
- The statistical process used to generate this data is a binomial process since users are answering a yes or no question
- A binomial distribution will be used to simulate the data

Assessment of the statistical significance of an A/B test is dependent on what kind of probability distribution the experimental data follows. Given your answer above, which statistical tests (z-test, t-test, etc.) are appropriate to use for this project?
- Paired t test

In classical (frequentist) A/B testing, we use p-values to measure the significance of the experimental feature (being exposed to an ad in our case) over the null hypothesis (the hypothesis that there is no difference in brand awareness between the exposed and control groups in the current case). How are p-values computed? What information do p-values provide? What is the type-I and type-II errors you may have in the analysis? Can you comment to which error types p-values are related?
- A p value is the probability of finding the observed or more extreme results when the null hypothesis is true i.e. (h0 = h1).
- P values are computed by first calculating the probability a random chance generated the data then adding everything that has the same equal probability and the probability of the rare outcome happening
- P values quantify how confident we should be that the ad influenced the brand awareness. The closer it is to zero the more confident we are that there is a difference in brand awareness between the two groups
- Type 1 error occur when we incorrectly reject the null hypothesis. In the challenge this may occur when we say that there is a difference between brand awareness while there isn't

- Type 2 error occur when we fail to reject the null hypothesis when we should have rejected it. In the challenge this may occur when we say there is no difference between the two groups in brand awareness while infect there is a difference
- P values are related to type 1 errors since a p value calculates the probability of obtaining a result at least as extreme as the current one while assuming the null is true

How does the classical A/B testing (using z-test, f-test, etc.) framework work?
- We first pick a feature to test
- We frame hypothesis statements as per what we are measuring
- Split the users into two groups and determine our sample size
- Calculate the z scores, p value, critical value and the standard deviation
- Make a conclusion whether to accept or reject h0 based on the p value

How does sequential A/B testing work?
- At the beginning of the experiment, choose a sample size N.
- Assign subjects randomly to the treatment and control, with 50% probability each.
- Track the number of incoming successes from the treatment group. Call this number T.
- Track the number of incoming successes from the control group. Call this number C.
- If $T-C$ reaches $\sqrt{2N}$, stop the test. Declare the treatment to be the winner.
- If $T+C$ reaches N, stop the test. Declare no winner.
-

What are some of the advantages of sequential A/B testing?
- It takes less time compared to classical testing
- Saves more traffic
- Less errors

How is A/B testing done using machine learning? What is the core idea behind this approach? In other words, what part of the machine learning analysis provides the insight regarding the high or no significance of the experimental feature?

What are the pros and cons of using Machine learning to perform A/B testing?
- Using machine learning can help you understand a complex system by providing insights that statistical inference would not have identified
- Provides a direction and magnitude of the experiment
  Cons
- Machine learning might overfit the data