Which online users belong to the control and exposed groups?

Control: users who have been shown a dummy ad
Exposed: users who have been shown a creative (ad) that was designed by SmartAd for the client.

Targeted ads are meant to reach certain customers based on demographics, psychographics, behavior and other second-order activities that are learned usually through data exhaust produced by users themselves.

“Is the world different from usual, is it surprising?” Because the world is usually not surprising and because in statistics you are never 100 percent sure about what a sample tells you

“You cannot say that the world is surprising, that the population is unusual, unless the evidence is very strong. This means that when you arrange your tests, you have to do it in a manner that makes it difficult for the unusual, surprising world to win support.”

No, Because we do not know if the result is due to a random chance. Before jumping to conclusions, we need to keep in mind, the raw results we have are only samples of bigger populations. Their statistical properties vary around the ones of the populations they come from. Therefore, statistically modeling these outputs is necessary. We need Statistical significance to give us a degree of confidence that their findings are real, reliable, and not due to chance.

The binomial distribution is the right process to generate the data since it satisfies all 4 requirements of the binomial distribution.
Each observation falls into one of two categories called success or failure.
There is a fixed number of observations.
The observations are all independent.
The probability of success (p) for each observation is the same - equally likely.

A test statistic is a random variable that is calculated from sample data and used in a hypothesis test. We can use test statistics to determine whether to reject the null hypothesis. The test statistic compares our data with what is expected under the null hypothesis. The test statistic is used to calculate the p-value.
A test statistic measures the degree of agreement between a sample of data and the null hypothesis. Z-test is the Z-statistic, which has the standard normal distribution under the null hypothesis. We can approximate the binomial distribution by the normal distribution for a large enough sample size.

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 are the type-I and type-II errors you may have in the analysis? Can you comment on which error types p-values are related?

- The p-value is the probability of finding the observed, or more extreme, results when the null hypothesis (H0) of a study question is true. It's the measure of the strength of evidence in support of a null hypothesis.
- Type I error (false positive) is the rejection of a true null hypothesis.
- Type II error (false negative) is the non-rejection of a false null hypothesis.
- A p-value is a probability associated with your critical value. The critical value depends on the probability you are allowing for a Type I error. It measures the chance of getting results at least as strong as yours if the claim (H0) were true.
- The probability of type 1 error, (alpha) is the p-value
- The p-value is calculated using the sampling distribution of the test statistic under the null hypothesis, the sample data, and the type of test being done (lower-tailed test, upper-tailed test, or two-sided test).
- The p-value for:
    - a lower-tailed test is specified by: p-value = cdf(ts)
    - an upper-tailed test is specified by: p-value = 1 - cdf(ts)
    - two-sided test is specified by: p-value = 2 * (1 - cdf(|ts|))
    - Where:
        - ts: the observed value of the test statistic calculated from our sample
        - cdf(): Cumulative distribution function of the distribution of the test statistic under the null hypothesis

How does the classical A/B testing (using z-test, f-test, etc.) framework work?

1. Make our Hypothesis:
    a. H0: there’s no difference in brand awareness between the 2 groups
    b. H1:There's a difference in brand awareness.
2. Set significant level (alpha)
3. Sample grouped into control and exposed groups randomly with equal probability.
4. Perform hypothesis testing to acquire p-value
5. Rejecting or failing to reject the null hypothesis. Rejection happens when p value is less than α (level of significance) else fails to reject the null hypothesis.

How does sequential A/B testing work?

Sequential testing is the practice of making decisions during an A/B test by sequentially monitoring the data as it accrues.

- In sequential sampling, instead of a fixed sample size, we choose one item (or a few) at a time and then test our hypothesis. We can either:
    - Reject the null hypothesis (H0) in favor of the alternate hypothesis (H1) and stop
    - Keep the null hypothesis and stop
    - Unable to reach either conclusion with current observation and continue sampling

What are some of the advantages of sequential A/B testing?

- Optimize necessary observation(sample size)
- Reduce the likelihood of error
- Finish experiment earlier without increasing the possibility of false results

A major issue with traditional, statistical-inference approaches to A/B testing is that it only compares 2 variables - an experiment/control to an outcome. The problem is that customer behavior is vastly more complex than this. Customers take different paths, spend different amounts of time on the site, come from different backgrounds (age, gender, interests) and more. This is where machine learning excels generating insights from complex systems. Unlike statistical inference, machine learning algorithms enable us to model complex systems that include all of the ongoing events, users features, and more. There are a number of algorithms each with strengths and weaknesses.
An attractive benefit to machine learning is that we can combine multiple approaches to gain insights.