**Homework assignment for analysts**

Background:

SFLY has batch emails that go out to all existing users in the database that are emailable (for the purpose of this assignment, emailable = users that have not opted out of marketing email). SFLY periodically tests to improve performance of these marketing campaigns. There were 2 test that were run by the team. There is a 14-day pre period that is used to check whether the group has any bias. There is also a 14 day post period that is used to check if there is demand being pulled forward (which means that customers that would have bought later but bought in the campaign time frame due to the offer- in these cases, we might see that the control group does better in the post period). We use 95% significance or greater as an acceptable standard.

When comparing test and control groups, we typically compare them in the following periods for conversions, orders and revenue related numbers:

1. Pre- to test for bias
2. Campaign
3. Post
4. Campaign through post combined

Email engagement numbers are not split by the above periods- they are measured for the entire time starting campaign launch date.

Following 2 tests were run by SFLY

1. Test 1 was run to test the hypothesis that SL2 (test) was better than SL1 (control)
2. Test 2 was run to test the hypothesis that content 2 (test) was better than content 1 (control)

For each test give the following descriptive answers. Try to be exhaustive and include as many points as possible. Don’t restrict yourself by data availability. This is to see how you think through a problem and approach it

1. What are the success metrics that you will use for the test?
2. You have a full population base of emailable customers available to you (size = 10M)
   1. How do you split your audience for testing?

The data for test 2 has been provided in the excel embedded. Do an analysis and share your results.   
The analysis should answer the following questions

1. What was the outcome of the test?
2. What is the confidence you have in these results?

This analysis you do will be shared with the CMO of the company. Think about how you’d present your findings. Use any format (a PowerPoint deck or pdf) you deem appropriate to share results.

**Formulas to use:**  
  
**Response metric, t statistic**= [SQROOT(mean1-mean2)2]/[(std dev1)2 + (std dev2)2]

Response metrics are calculated by counting if a customer qualifies or not. Orders and conversions are examples. Mean in these cases are percentages or rates.

**Continuous metric, t statistic** = A/(B\*C)

A = ABSOLUTE[(mean1-mean2)

B = SQROOT{[(aud size1)\*(variance1) + (aud size2)\*(variance2)] / (aud size1 + aud size2 -2)}

C = SQROOT[(Aud size1 + Aud size2)/(Aud size1)\*(Aud size2)]

Continuous metrics are calculated by summing them up. Examples are revenue and photo uploads. Mean in these cases is an average

**T statistic value and corresponding threshold values**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 99% | 95% | 90% | 85% | 80% |
| 2.5758 | 1.96 | 1.6449 | 1.4395 | 1.2815 |

Spreadsheet with data:

