

Experiment Design: Email Campaign Testing Plan

28th October 2019

Cici Chen

CASE BACKGROUND

The Engagement Marketing Team is designing an email campaign to 1 million customers. Their goals are to increase customer retention and decrease subscription cancellation. The insights team is helping them to achieve the goal and succeed in the test.

Propose an outline of testing plan. Please be specific about the details of your design. Details can include but are not limited to primary KPIs to measure, sample selection method, test cadence to achieve a statistical significant result.

PART I: OUTLINE OF TESTING PLAN

This proposal could be implemented through several methods, but for our specific case, I would recommend an email marketing strategy with different A/B Testing to get better results from the testing plans.

The following shows a simple framework:

1. Set up a SMART goal

It is difficult to track the whole process of clients' conversions, which makes it even more important to know what is the goal of the testing and how to track our campaigns' performance effectively.

- **SMART** stands for Specific, Measurable, Achievable, Relevant, and Time-bound. A SMART goal will inform us what we want to achieve during the whole process.
- **Target Audience** is 1 million customers who have already subscribed Audible and do not have a high frequency of activities.
- **Our goal** is to increase customer retention and decrease subscription cancellation

2. Pick up primary KPIs

KPIs provide us with quantitative methods to evaluate the performance of the engagement activities within the campaigns. We need to decide which criteria matter before our testing, which also leads to the KPIs of email marketing. There are some primary KPIs that we need to consider:

- **Inbox Placement Rate/Delivery Rate (IPR)** = $\frac{\text{Number of emails delivered to inbox}}{\text{Total number of emails sent}}$

The number of emails delivered is of great importance during email marketing.

It is impossible to have a 100% delivery rate of emails. There is always a difference between an email being “delivered” and actually making it into our customers' email inbox. It is very likely that the email campaign is

delivered while it goes straight to the junk folder for some reason. As such, how many emails truly delivered matters in our testing plan.

- **Open Rate** = $\frac{\text{Open the email campaign}}{\text{Total impressions}}$

Due to the convenience of panes of email apps, some people may just preview panes to get the information from Audible, and they might not open the email actually. As a result, it is not enough to make the email campaigns to the inbox of customers, we also have to measure the number of emails that are opened.

This rate could be heavily influenced by many factors of email campaigns, and this part needs help with the marketing and UX/UI team together to make the wordings of the subject lines more relevant to our customers' demand. The average open rate is usually around 17.92%^[1]

- **Click-Through Rate (CTR)** = $\frac{\text{Total clicks}}{\text{Total impressions}}$

The percentage of customers who clicked on one or more links contained in a given email campaign. CTR indicates how many people truly clicked on a link to our email campaign. We need to measure which link they are interested in and where is the link located in our emails. Even our customers click the unsubscribe link, it still counts as a click, while it will be counted as a negative sign. The average click-through rate is usually around 2.69%^[1].

- **Click-To-Open Rate (CTOR)** = $\frac{\text{Total clicks}}{\text{Total opens}}$

CTO is the number of opens compared to the number of click-throughs. If we have a lot of clicks but few opens, it means that we have a great offer but we need to nail our subject lines better. The average click-to-open rate: 14.10%^[1].

- **Unsubscribe Rate** = $\frac{\text{Unsubscribed number}}{\text{Emails delivered}}$

This is of great significance because we are able to see which part goes wrong during the whole promotion process. Also, we need to monitor the spam complaints as some people just go ahead to report spam instead of going ahead to unsubscribe. The average unsubscribe rate: 0.17%^[1].

- **Bounce Rate** = $\frac{\text{Total visitors to a webpage linked to an email who left without any interactions}}{\text{Total visitors to a webpage}}$

It represents the percentage of visitors who enter the site and then leave rather than continuing to view other pages within the same site. The average bounce rate: 1.06%^[1].

- **Conversions (CR)** = $\frac{\text{Total attributed conversions}}{\text{Total clicks}}$

Finally, the KPI that matters the most: conversions. Different goals may contain different conversions. A conversion could be a single click, registration, subscription, downloading a blog or anything else, and it depends on our goal.

- **Social Shares & Forwards**

This KPI indicates engagement on social media and is a good indicator of the quality of our content. So make it easy for your readers to share your content by adding social buttons to your emails.

There are more KPIs could be measured during the testing based on different goals of the experiment designs, and those listed above are the primary ones when we design a testing plan. Generally speaking, 3-5 KPIs will be enough to track for one testing plan, and knowing what why we are measuring is as crucial as knowing the KPIs.

3. Hypothesis Testing

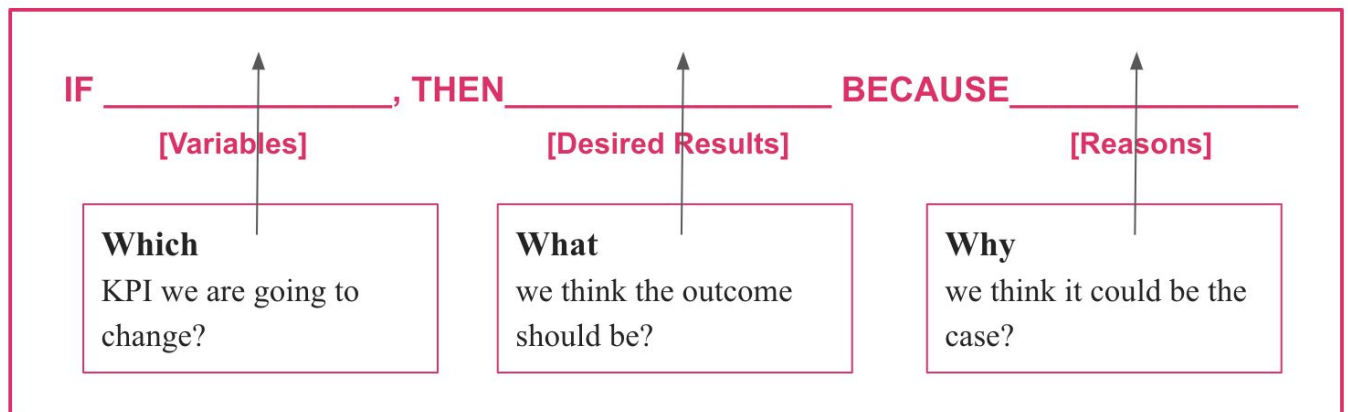
The American Heritage Dictionary defines a **hypothesis** as, “a tentative explanation for an observation, phenomenon, or scientific problem that can be tested by further investigation.” [2]

A Hypothesis is a prediction we create prior to running an experiment test. It states clearly what is being changed, what we believe the outcome will be, and why we think it is the case. Testing an experiment will either agree or disagree with our hypothesis.

We will have two common kinds of hypotheses: **Null hypothesis** and **alternative hypothesis**, which are two statements about our testing samples. A **hypothesis test** uses sample data to determine whether to reject the **null hypothesis** or no.

- **Null Hypothesis (H_0):** what we want to reject, and it indicates “No difference”
- **Alternative Hypothesis (H_A):** what we believe to be true or hope to prove true, usually “There is a difference”

A good hypothesis should contain three parts:



Once we have a hypothesis, we can then decide what to test to meet the variable conditions the hypothesis states for different testing questions.

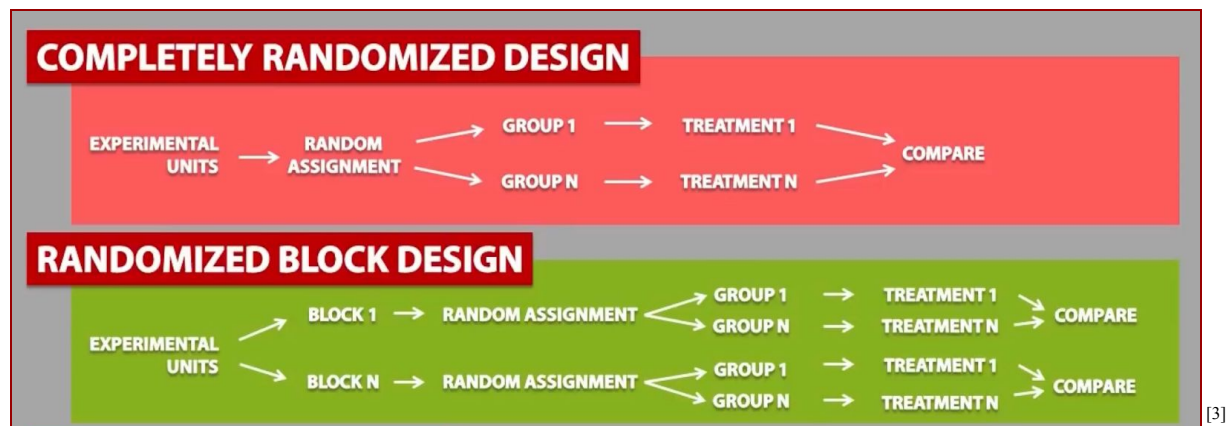
4. Select Samples

Sample mechanism

There are three things we need to consider when we select samples:

- **Control:** In order to avoid experimental bias, the favoring of certain outcomes over others. We need to control the variable, and hold the other factors fixed.

- **Randomization:** It is generally extremely difficult for experiments to eliminate bias, so we will use randomization in practice. In our email campaign, we need to randomly send the testing emails to our sample customers, and we can use the two most popular randomized experimental designs



Completely randomized design: objects or subjects are assigned to groups completely at random.

Randomly blocked design: If we know the specific differences among groups of subjects within an experimental group, we can use this kind of design and we first divide customers based on homogeneous blocks, such as gender, age, and other known factors.

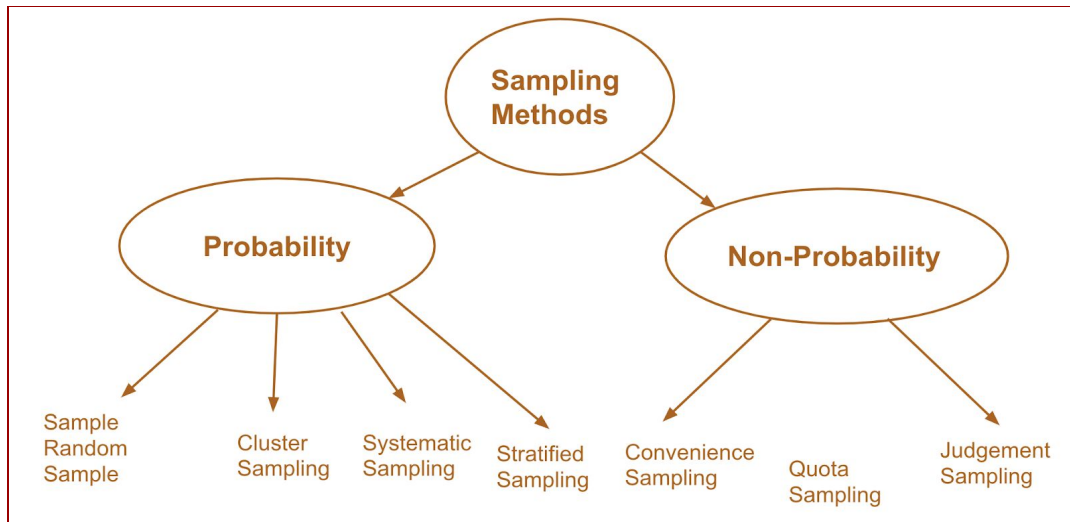
- **Replication:** Although randomization has ensured the results of our testing as reliable as possible. In order to improve the significance of our email experimental results, replication will make a contribution. It will reduce the variability during the testing, and it will increase the significance and confidence level, which leads to a better conclusion in statistical area.

Sampling Methods

As we are conducting a testing plan, we do not want to use up all of the customers we have in the market.

In the meantime, we do not want to have a small sample size as it won't help with producing meaningful results. As such, we will use some sampling methods to improve the testing efficiency and effectiveness.

Sampling methods could be divided into two categories: **Probability Sampling** and **Non-Probability Sampling**^[4].



Sample Random Sample: Each individual is chosen entirely by chance and each member of the population has an equal probability of being selected.

Cluster Sampling: Individuals are selected at regular intervals

Stratified Sampling: First divide the population into subgroups (or strata) who all share a similar characteristic

Cluster Sampling: Subgroups of the population are used as the sampling unit, rather than individuals

Convenience Sampling: The participants are selected based on availability and willingness to take part

Quota Sampling: Interviewers are given a quota of subjects of a specified type to attempt to recruit

Judgement Sampling: Relies on the judgement of the researcher when choosing who to ask to participate

Sample size

$$n = \frac{Z_{\frac{\alpha}{2}}^2 S^2}{e^2 + \frac{Z_{\frac{\alpha}{2}}^2 S^2}{N}}$$

- n : sample size
- $Z_{\frac{\alpha}{2}}$: critical value in statistics, usually it is $Z_{\frac{0.05}{2}} = 1.96$ when $\alpha = 0.05$
- S^2 : population variance, it could be estimated as $S^2 = \frac{N}{N-1} p(1-p) \approx p(1-p) \leq \frac{1}{4}$ when N is large
- e : Margin of error, which determines how precise we want our estimation of proportion of conversion. Suppose we are interested in estimation of p : the proportion of our conversion rate. $P(|\hat{p} - p| \leq e) = 1 - \alpha$ shows an illustration of how e works here.
- α : the confidence level we set, usually it is 0.05
- N : population size

We can use the above formulas to determine our sample size, and it could be explained later in an example.

5. Build a Testing Calendar

Calendar: Creating a testing calendar will keep the tracking process organized. It helps us to outline the testing procedure clearly. We are likely to run multiple tests of the same hypothesis to find a meaningful result, and we also need to test several times to ensure the results are not a coincidence. A calendar helps with the priority and the list of the order to test them.

Set a live test: Now we are ready to test for the first time. Set a pre-test of our testing plan means setting a live test to ensure everything is onsite.

6. Collect, Analyze and Report Data

During the testing, we will collect data and analyze the data to get useful information. We will transform the data into the results of which KPI perform better, and which KPI could be declared by statistical significance. This part is profound and should be treated very carefully, even with the help with many advanced tools such as Google Analytics, Adwords, and other email marketing tracking platforms.

7. Retesting to Optimize

The same idea as randomization, we want our results to be reliable, which could be implemented into the final email campaign sent to Audible's customers successfully. We can do retesting to draw meaningful conclusions within significant confidence.

Here we can do A/A Testing, which is exactly the same procedure as A/B Testing. The only difference is that we want to test two identical versions of an email campaign. We want to use A/A Testing to check that our results are fair enough. As such, the results of A/A Testing should not report a big difference. Since the two groups of users experience are treated with the same environment, we expect that the KPIs of the two variations to be roughly the same.

Notice that people have complex emotional feelings about what they are doing, and sometimes even the color change of a click bottom or the change of the bottom position will lead to a different testing result. No testing plan will be perfectly conducted for the first time, and that's the reason why we need to optimize our testing plan accordingly. Even if we do not get the desired results, we still learn something from the data.

PART II: INSIGHTS TO PRESENT & RECOMMENDATIONS

Suppose you concluded that the test was successful. With the test result, you are about to communicate to the marketing team. What kind of insights would you present and what would be your recommendations?

Being a storyteller rather than a repeater

The testing process is complicated and might be boring to people who are not interested in the numbers and data. We do not need to communicate the whole procedure to the marketing team, but we need to present the most important results as well as the data behind to support the whole story.

Insights present with one simple and valid example

1. Set up a SMART goal

We want to increase the conversion rate of Audible by conducting experimental designs to measure KPIs within a specific testing period and achieve a statistically significant result eventually.

2. Pick up Primary KPIs

Conversion Rate

3. Hypothesis Testing

H_0 : If Audible change the sending times to evenings, then the conversion rate will **NOT** increase, because people are **NOT** likely to convert when they finished working.

H_A : If Audible change the sending times to evenings, then the conversion rate will increase, because people are more likely to convert when they finished working.

4. Select Samples

$$n = \frac{Z_{\frac{\alpha}{2}}^2 S^2}{e^2 + \frac{Z_{\frac{\alpha}{2}}^2 S^2}{N}} = \frac{1.96^2 \times \frac{1}{4}}{0.03^2 + \frac{1.96^2 \times \frac{1}{4}}{10^6}} = 1065.974 \approx 1066 \Rightarrow \text{Sample Size} = 1100$$

In our case, we set $e=3\%$, and we want a 95% confidence significance conclusion, and the customers population is 1 million, so $N=10^6$, so our sample size in this case will be $n=1100$.

5. Build a Testing Calendar

6. Collect, Analyze and Report Data

7. Retesting To Optimize

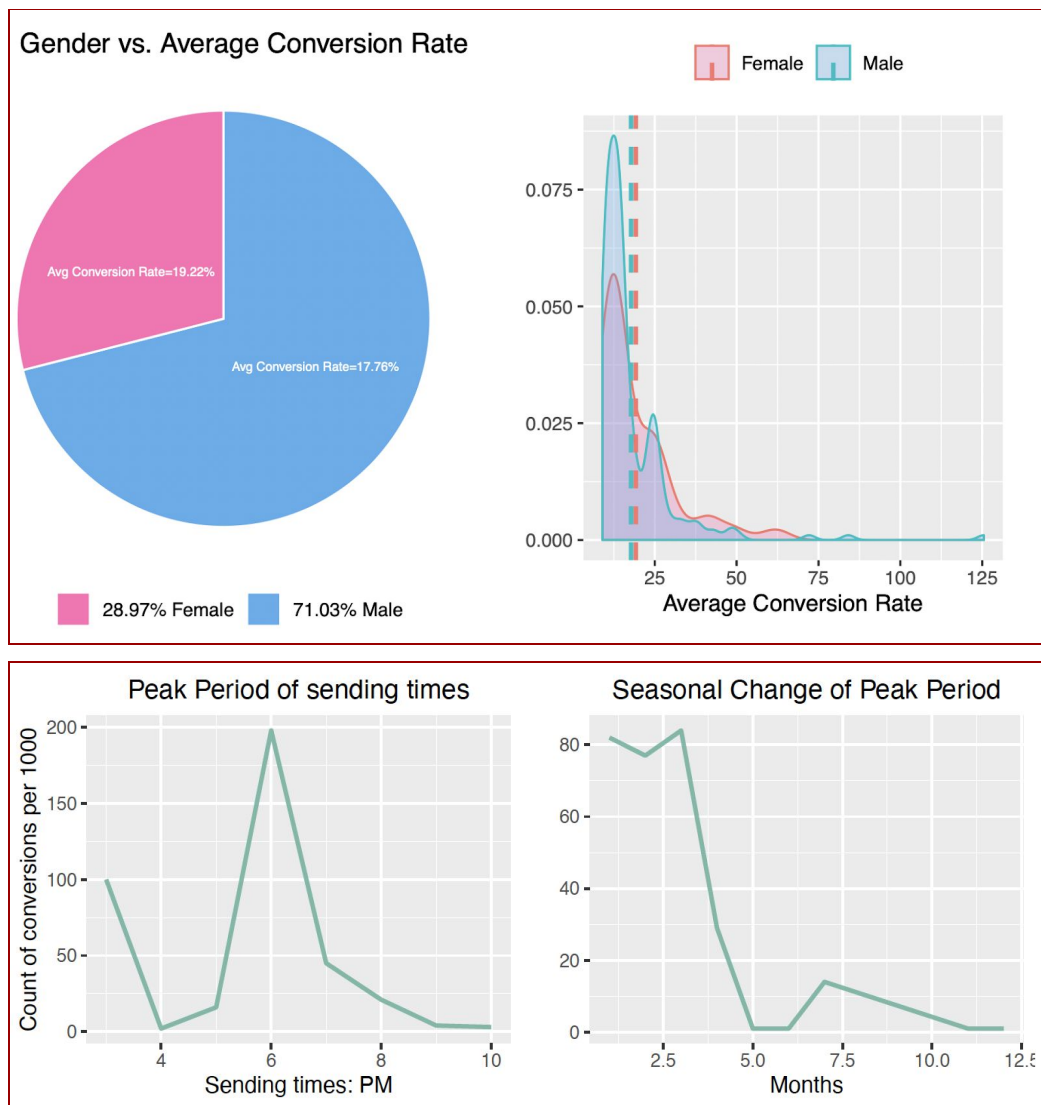
Other Recommendations

- **Data visualization**

“Visualization works from a human perspective because we respond to and process visual data better than any other type of data. In fact, the human brain processes images 60,000 times faster than text, and 90 percent of information transmitted to the brain is visual. Since we are visual by nature, we can use this skill to enhance data processing and organizational effectiveness.”^[5]

The testing plan will help us to know what we are doing, and data visualization is a smart way to continually communicate the results of the testing plan efficiently. It shows our values on making a difference for the marketing team, so the data visualization part must be done and data-driven strategies will contribute a lot to the final plan.

I’ve made **two sample plots** for the data visualization for our example before, and I also combined it with the factor of Gender as block design. I simulated some data to generate the graphs below as a reference, and hope it will give you some insights if I am going to present.



- **Effective Comparison**

Technology has changed our life. Here is an example about why we need to make comparison when we display why the testing plan will work.



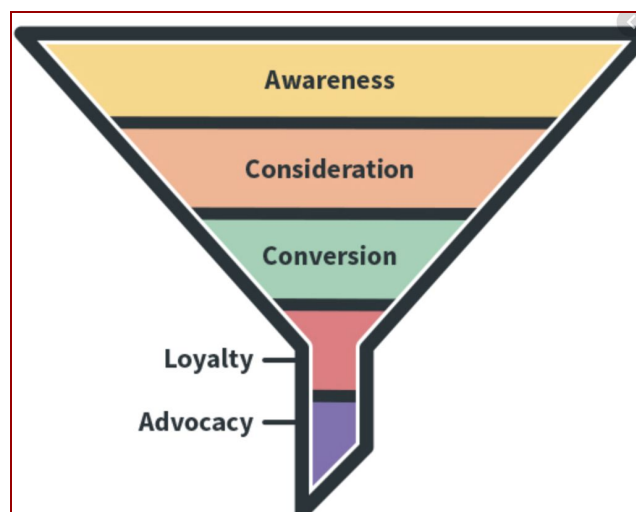
As you can see, the old logo is less attractive compared to the new logo for now, and the same idea when we communicate our insights to the marketing team, the easy and simple way to present our testing results is to make an effective comparison like this.

PART III: ATTRACT POTENTIAL CUSTOMERS

Now the Acquisition Marketing Team will be launching a campaign to acquire customer through targeting people that have shown interest on Audible but have not converted yet. The insights team is brought into the conversation to help the business understand what would be a good campaign.

Before the campaign, you are asked to design what would be considered “measures of success” for the campaign. What KPIs would you define to measure its success? Consider that the goal will be acquiring as many customers as possible (short-term), but making sure we create sustained value for the company by acquiring customer that would remain with Audible for a long time (long-term).

Here comes a marketing funnel, which will help us to determine the KPIs we want to measure the success.



In order to attract potential customers who have already shown the interest on Audible, besides the KPIs that I mentioned earlier, there are some other KPIs that we can consider as measurements of success.

- **Conversion Rate** (As it is the key)
- **Landing Page:** Even it is not part of the email, if our customers leave the page when they click the links in the email, we may need to see whether we need to change to the landing page. Usually, the conversion rate will decrease if the landing page does not reinforce the email campaigns' words.
- **Subject Lines:** If the subject line of the email campaign does not appealing to our potential customers, they may change the status from consideration to awareness
- **Call-to-Action Rate:** If one of our potential customers click a link and then make the action right after, it will be a positive sign, which should be tracked carefully as we want to see how this happens and how could we reproduce the same procedures.
- **Personalization Rate:** We need to measure the personalization level of our email campaigns. "Hello John" is always better than "Hello my customer ", and it will also be impressive if we can conclude locations, interests, and other related information.
- **Product image sizes, Video in email, etc**

REFERENCE

[1]The Ultimate Email Marketing Benchmarks (2019): By Day, Industry. (n.d.). Retrieved from <https://www.campaignmonitor.com/resources/guides/email-marketing-benchmarks/>.

[2]YourDictionary. (2017, July 20). Examples of Hypothesis. Retrieved from <https://examples.yourdictionary.com/examples-of-hypothesis.html>.

[3](n.d.). Retrieved from <https://www.youtube.com/watch?v=10ikXret7Lk>.

[4]Admin. (2018, August 20). Methods of sampling from a population. Retrieved from <https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/methods-of-sampling-population>.

[5]Humans Process Visual Data Better. (2018, October 8). Retrieved from <http://www.t-sciences.com/news/humans-process-visual-data-better>.