**Scenario Description**

Let's say that you're working for a fictional productivity software company that is looking for ways to increase the number of people who pay for their software. The way that the software is currently set up, users can download and use the software free of charge, for a 7-day trial. After the end of the trial, users are required to pay for a license to continue using the software.

One idea that the company wants to try is to change the layout of the homepage to emphasize more prominently and higher up on the page that there is a 7-day trial available for the company's software. The current fear is that some potential users are missing out on using the software because of a lack of awareness of the trial period. If more people download the software and use it in the trial period, the hope is that this entices more people to make a purchase after seeing what the software can do.

In this case study, you'll go through steps for planning out an experiment to test the new homepage. You will start by constructing a user funnel and deciding on metrics to track. You'll also perform experiment sizing to see how long it should be run. Afterwards, you'll be given some data collected for the experiment, perform statistical tests to analyze the results, and come to conclusions regarding how effective the new homepage changes were for bringing in more users.

## Building a Funnel

Before we do anything else, the first thing we should do is specify the objective or goal of our study:

*Revising the structure of the homepage will increase the number of people that download the software and, ultimately, the number of people that purchase a license.*

Now, we should think about the activities that a user will take on the site that are relevant to measuring our objective. This path or funnel will help us figure out how we will create experimental condition groups and which metrics we'll need to track to measure the experiment's effect. To help you construct the funnel, here's some information about the way the company's website is structured, and how the software induces users to purchase a license.

The company's website has five main sections:

1. the homepage;
2. a section with additional information, gallery, and examples;
3. a page for users to download the software;
4. a page for users to purchase a license; and
5. a support sub-site with documentation and FAQs for the software.

For the software itself, the website requires that users create an account in order to download the software program. The program is usable freely for seven days after download. When the trial period is hit, the program will bring up a dialog box that takes the user to the license page. After purchasing a license, the user will receive a unique code associated with their site account. This code can then be used with the program to register it with that user, and the program can be used thereafter without issue.

Using the information above, fill in your responses to the questions below regarding the construction of a user funnel, then check on the next page for my thoughts.

1. What steps do you expect typical visitors to take from their initial visit to the webpage through purchasing a license for continued use of the program? Are there any 'typical' steps that certain visitors might not take?
2. Consider the webpage as a whole. What kinds of events might occur outside of the expected flow for the experiment that might interfere with measuring the effects of our manipulation?

**Re: An expected flow**

A straightforward flow might include the following steps:

* Visit homepage
* Visit download page
* Sign up for an account
* Download software
* After 7-day trial, software takes user to license-purchase page
* Purchase license

Note that it is possible for the visitor to drop from the flow after each step, forming a funnel. There might be additional steps that a user might take between visiting the homepage and visiting the download page that aren't accounted for in the above flow. For example, someone might want to check out the additional informational pages before visiting the download page, or even visit the license purchase page to check the license price before even deciding to download. Considering the amount of browsing that a visitor could perform on the page, it might be simplest just to track whether or not a user gets to the download page at some point, without worrying about the many paths that they could have taken to get there.

**Re: Atypical events**

There are a few events in the expected flow that might not correspond with the visitors we want to target. For example, there might be users on the homepage who aren't new users. Users who already have a license might just be visiting the homepage as a way to access the support sub-site. A user who wants to buy a license might also come in to the license page through the homepage, rather than directly from the software.

When it comes to license purchasing, it's possible that users don't come back after exactly seven days. Some users might come back early and make their purchase during their trial period. Alternatively, a user might end up taking more than seven days to decide to make their purchase, coming back days after the end of the trial. Anticipating scenarios like this can be useful for planning the design, and coming up with metrics that come as close as possible to measuring desired effects.

## Deciding on Metrics

From our user funnel, we should consider two things: where and how we should split users into experiment groups, and what metrics we will use to track the success or failure of the experimental manipulation. The choice of unit of diversion (the point at which we divide observations into groups) may affect what metrics we can use, and whether the metrics we record should be considered invariant or evaluation metrics. To start, decide on a unit of diversion and brainstorm some ideas for metrics to capture.

To be clear, the overall plan is to test the effect of the new homepage using a true experiment; in particular, we'll be using an A/B testing framework. This means that prospective users should be split into two groups. The control, or 'A' group, will see the old homepage, while the experimental, or 'B' group, will see the new homepage that emphasizes the 7-day trial.

1. By which means should we divide visitors into our experimental and control groups?
2. Write down some potential metrics or ways of objectively measuring things related to evaluating the success of the experiment. You don't need to decide on invariant or evaluation metrics here: that'll be discussed on the next page.

## Re: Unit of Diversion

Three main categories of diversion were presented in the course: event-based diversion, cookie-based diversion, and account-based diversion.

An event-based diversion (like a pageview) can provide many observations to draw conclusions from, but doesn't quite hit the mark for this case. If the condition changes on each pageview, then a visitor might get a different experience on each homepage visit. Event-based diversion is much better when the changes aren't as easily visible to users, to avoid disruption of experience. In addition, pageview-based diversion would let us know how many times the download page was accessed from each condition, but can't go any further in tracking how many actual downloads were generated from each condition.

Diverting based on account or user ID can be stable, but it's not the right choice in this case. Since visitors only register after getting to the download page, this is too late to introduce the new homepage to people who should be assigned to the experimental condition.

So this leaves the consideration of cookie-based diversion, which feels like the right choice. We can assign a cookie to each visitor upon their first page hit, which allows them to be separated into the control and experimental groups. Cookies also allow tracking of each visitor hitting each page, recording whether or not they eventually hit the download page and then whether or not they actually register an account and perform the download.

That's not to say that the cookie-based diversion is perfect. The usual cookie-based diversion issues apply: we can get some inconsistency in counts if users enter the site via incognito window, different browsers, or cookies that expire or get deleted before they make a download. This kind of assignment 'dilution' could dampen the true effect of our experimental manipulation. As a simplification, however, we'll assume that this kind of assignment dilution will be small, and ignore its potential effects.

## Re: Brainstorm Potential Metrics

In terms of metrics, we might want to keep track of the number of cookies that are recorded in different parts of the website. In particular, the number of cookies on the homepage, download page, and account registration page (in order to actually make the download) could prove useful. We can track the number of licenses purchased through the user accounts, each of which can be linked back to a particular condition. Though it hasn't been specified, it's also possible that the software includes usage statistics that we could track.

The above metrics are all based on absolute counts. We could instead perform our analysis on ratios of those counts. For example, we could be interested in the proportion of downloads out of all homepage visits. License purchases could be stated as a ratio against the number of registered users (downloads) or the original number of cookies.

Below, you will decide for each of the proposed metrics whether or not you would want to use them as an invariant metric or an evaluation metric. To recap, an invariant metric is an objective measure that you should expect will not vary between conditions and that indicate equivalence between groups. Evaluation metrics, on the other hand, represent measures where you expect there will be differences between groups, and whose differences should say something meaningful about your experimental manipulation.

## Re: Selecting Invariant and Evaluation Metrics

There's one invariant metric that really stands out here, and that's the number of cookies that hit the homepage. If we've done things correctly, each visitor should have an equal chance of seeing each homepage, and that means that the number of cookies assigned to each group should be about the same. Since visitors come in without any additional information (e.g. account info) and the change effected by the experimental manipulation comes in right at the start, there aren't other invariant metrics we should worry about.

Selecting evaluation metrics is a trickier proposition. Count-based metrics at other parts of the process seem like natural choices: the number of times the software was downloaded and the number of licenses purchased are exactly what we want to change with the new homepage. The issue is that even though we expect the number of cookies assigned to each group to be about the same, it's much more likely than not they they won't be exactly the same. Instead, we should prefer using the download rate (# downloads / # cookies) and purchase rate (# licenses / # cookies) relative to the number of cookies as evaluation metrics. Using these ratios allows us to account for slight imbalances between groups.

As for the other proposed metrics, the ratio between the number of licenses and number of downloads is potentially interesting, but not as direct as the other two ratios discussed above. It's possible that the manipulation increases both the number of downloads and number of licenses, but increases the former to a much higher rate. In this case, the licenses-to-downloads ratio might be worse off for the new homepage compared to the old, even though the new homepage has our desired effects. There's no such inconsistency issue with the ratios that use the number of cookies in the denominator.

Product usage statistics like the average time the software was used in the trial period are potentially interesting features, but aren't directly related to our experiment. We might not have a strong feeling about what kind of effect the homepage will have on people that actually download the software. Stated differently, product usage isn't a direct target of the homepage manipulation. Certainly, these statistics might help us dig deeper into the reasons for observed effects after an experiment is complete. They might even point toward future changes and experiments to conduct. But in terms of experiment success, product usage shouldn't be considered an invariant or evaluation metric.

## Experiment Sizing

Now that we have our main metrics selected: number of cookies as an invariant metric, and the download rate and license purchase rate (relative to number of cookies) as evaluation metrics, we should take a look at the feasibility of the experiment in terms of the amount of time it will take to run. We can use historical data as a baseline to see what it might take to detect our desired levels of change.

Recent history shows that there are about 3250 unique visitors per day, with slightly more visitors on Friday through Monday, than the rest of the week. There are about 520 software downloads per day (a .16 rate) and about 65 licenses purchased each day (a .02 rate). In an ideal case, both the download rate and license purchase rate should increase with the new homepage; a statistically significant negative change should be a sign to not deploy the homepage change. However, if only one of our metrics shows a statistically significant positive change we should be happy enough to deploy the new homepage.

1. Consider that we want to preserve a maximum 5% Type I error rate for falsely deploying the homepage without any actual effect. Should we apply the Bonferroni correction in this case?
2. Let's say that we want to detect an increase of 50 downloads per day (up to 570 per day, or a .175 rate). How many days of data would we need to collect in order to get enough visitors to detect this new rate at an overall 5% Type I error rate and at 80% power?
3. What if we wanted to detect an increase of 10 license purchases per day (up to 75 per day, or a .023 rate). How many days of data would we need to collect in order to get enough visitors to detect this new rate at an overall 5% Type I error rate and at 80% power?

## Re: Experiment Sizing

Because we have two evaluation metrics of interest, we should make sure that we are choosing an appropriate significance level to conduct each test, in order to preserve a maximum overall Type I error rate of .05. Since we would be happy to deploy the new homepage if either download rate or license purchase rate showed a statistically significant increase, performing both individual tests at a .05 error rate carries the risk of making too many Type I errors. As such, we'll apply the Bonferroni correction to run each test at a .025 error rate so as to protect against making too many errors. If it were the case that we needed to see both metrics with a statistically significant increase, then we wouldn't need to include the correction on the individual tests.

For an overall 5% Type I error rate with Bonferroni correction and 80% power, we should require 6 days (rounded up from 5.55) to reliably detect a 50 download increase per day and 21 days (rounded up from 20.44) to detect an increase of 10 license purchases per day. 21 days is actually a convenient number since the three-week timespan helps to account for weekly cycles. In addition, the 21-day data collection period is a short enough timeframe that running the experiment is a reasonable proposition. If the required experiment length were a few weeks longer, then we might have needed to forego measuring the license purchasing rate as a critical metric.

One thing that isn't accounted for in the base experiment length calculations is that there is going to be a delay between when users download the software and when they actually purchase a license. That is, when we start the experiment, there could be about seven days before a user account associated with a cookie actually comes back to make their purchase. Any purchases observed within the first week might not be attributable to either experimental condition. As a way of accounting for this, we'll run the experiment for about one week longer to allow those users who come in during the third week a chance to come back and be counted in the license purchases tally.

## Validity, Bias, and Ethics

Before getting to the data and its analysis, let's review a few of the conceptual points that go into creation of an experiment: validity, bias, and ethics.

We probably don't have too much to worry about in terms of validity. For conceptual validity, the evaluation metrics are directly aligned with the experimental goals, no abstraction needed. Internal validity is maintained by performing an experiment with properly-handled randomization and controls. We don't really need to answer to external validity since we're drawing from the full site population and there's no other population we're looking to generalize to.

As for biases, we might think of novelty bias as being a potential issue. However, we don't expect users to come back to the homepage regularly. Downloading and license purchasing are actions we expect to only occur once per user, so there's no real 'return rate' to worry about. One possibility, however, is that if more people download the software under the new homepage, the expanded user base is qualitatively different from the people who came to the page under the original homepage. This might cause more homepage hits from people looking for the support pages on the site, causing the number of unique cookies under each condition to differ. If we do see something wrong or out of place in the invariant metric (number of cookies), then this might be an area to explore in further investigations.

Finally, for ethical issues, the changes to the homepage should be benign and present no risk to users. Our experiment objectives are also clearly stated. Considering the low risks of the experiment, informed consent is at worst a minor concern; a standard popup to let visitors know that cookies are used to track user experience on the site will likely suffice. The largest ethics principle we should be concerned about is data sensitivity. We shouldn't get any sensitive data out of the cookie assignment and collection, though some information will be collected from the user when they go to download the software. No sensitive data is required for the metrics we've laid out, so what we should do is just aggregate daily visits, downloads, and purchase counts without looking at any individual outcomes.

## Analyze Data

Let's assume that the experiment was given the green light to go ahead, and data was collected for 29 days. As a reminder of the discussion on experiment sizing, it was found that a three-week period was needed to collect enough visitors to achieve our desired power level. Eight additional days of collection were added to allow visitors in the last week to complete their trials and come back to make a purchase – if you look at the data linked in the next paragraph, you will see that it takes about eight days before the license purchases reaches its steady level.

The collected data can be found [here](https://s3.amazonaws.com/video.udacity-data.com/topher/2018/October/5bbbcebf_homepage-experiment-data/homepage-experiment-data.csv). The data file reports the daily counts for the number of unique cookies, number of downloads, and number of license purchases attributed to each group: the experimental group with the new homepage, or the control group with the old homepage. The number of license purchases only includes purchases by users who joined after the start of the experiment, so there will be some time before the counts reach their steady state. As noted earlier, we'll assume that the potentially muddying effects of visits across multiple days, established user visits, and 'lost' cookie tracking will be ignorable, at least unless we find reason to doubt our findings.

## Invariant Metric

First, we should check our invariant metric, the number of cookies assigned to each group. If there is a statistically significant difference detected, then we shouldn't move on to the evaluation metrics right away. We'd need to first dig deeper to see if there was an issue with the group-assignment procedure, or if there is something about the manipulation that affected the number of cookies observed, before we feel secure about analyzing and interpreting the evaluation metrics.

1. What is the p-value for the test on the number of cookies assigned to each group?

## Evaluation Metrics

Assuming that the invariant metric passed inspection, we can move on to the evaluation metrics: download rate and license purchasing rate. For a refresher, the download rate is the total number of downloads divided by the number of cookies, and the license purchasing rate the number of licenses divided by the number of cookies.

One tricky point to consider is that there is a seven or eight day delay between when most people download the software and when they make a purchase. There's no direct way of attributing cookies all the way through license purchases due to the daily aggregation of results, so the best we can do is to make a justified argument for handling the data. To answer the question below about the license purchasing rate, you should only take the cookies observed through day 21 as the denominator of the ratio as being responsible for all of the license purchases observed. (A more informed model of license purchasing could come up with a different handling of the data, such as including part of the day 22 cookies in the denominator.) (Note that we don't need to perform this kind of correction for the download rate, since the link between homepage visits and downloads is much closer.)

1. What is the p-value for the test on the download rate between groups?
2. What is the p-value for the test on the license purchasing rate between groups?

## Re: Analyze Data

For the test of the invariant metric, number of cookies, there were a larger number of cookies recorded in the experiment group, 47 346 vs. 46 851. This ends up generating a p-value of 0.107 (z = -1.61), which is within a reasonable range under the null hypothesis. Since we lack sufficient reason to reject the null, we can continue on to evaluating the evaluation metrics. (Note that this doesn't mean that there wasn'tsomething actually different about the cookie counts between groups, only that we couldn't detect it if such a difference existed.)

For the first evaluation metric, download rate, there was an extremely convincing effect. An absolute increase from 0.1612 to 0.1805 results in a z-score of 7.87, well beyond any standard significance bound. However, the second evaluation metric, license purchasing rate, only shows a small increase from 0.0210 to 0.0213 (following the assumption that only the first 21 days of cookies account for all purchases). This results in a p-value of 0.398 (z = 0.26).

## Draw Conclusions

Despite the fact that statistical significance wasn't obtained for the number of licenses purchased, the new homepage appeared to have a strong effect on the number of downloads made. Based on our goals, this seems enough to suggest replacing the old homepage with the new homepage. Establishing whether there was a significant increase in the number of license purchases, either through the rate or the increase in the number of homepage visits, will need to wait for further experiments or data collection.

One inference we might like to make is that the new homepage attracted new users who would not normally try out the program, but that these new users didn't convert to purchases at the same rate as the existing user base. This is a nice story to tell, but we can't actually say that with the data as given. In order to make this inference, we would need more detailed information about individual visitors that isn't available. However, if the software did have the capability of reporting usage statistics, that might be a way of seeing if certain profiles are more likely to purchase a license. This might then open additional ideas for improving revenue.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

You can calculate as below.

1. First part

In every AB test, we formulate the null hypothesis which is that the two conversion rates for the control design () and the new tested design () are equal:

The null hypothesis is tested against the alternative hypothesis which is that the two conversion rates are not equal:

1. **The significance level for the experiment:** A 5% significance but we will take this as 2.5 after Bernoulli correction
2. **Minimum detectable effect**: The desired relevant difference between the rates you would like to discover(0.175-0.16)
3. **The test power**: the probability of detecting that difference between the original rate and the variant conversion rates. This is 80%

1

Formula: n = ((z\_null\*sd\_null - z\_alt\*sd\_alt) / p\_diff) \*\* 2

*z\_null=1.95996398454 (Z score for 97.5 interval)*

*z\_alt=-0.841621233573 (Z score for B=20)*

*sd\_null=0.518459255873*

*sd\_alt=0.527991477204*

The above formula n=9481, now number of visitors is 3250 per day, dividing it into two groups 1625=

Total number of Days will be 9481/1625 =6(approx)

2.Second Part

Follow the steps to calculate if the invariant metric( statistically significant difference)

* Sum of the total cookies in control group= 46851
* Total number of cookies=(46851+47346 = 94197
* Ratios of total number of cookies in control to total cookies is .4973 now we are sure if this statistical significant or not., Therefore we check for p values
* Calculate standard deviation for total cookies sd = (np.sqrt(0.5 \* (1-0.5) \* 94197 ) =153.457648881
* Calculate Zscore =z = ((46851 + 0.5) - 0.5 \* 94197 ) / sd = **-1.6**
* Then calculate the p value from [Zscore](http://www.z-table.com/" \t "_blank), which is 2\*0.503=0.10 for two tailed test and this is not statistically significant(<0.05)

Also, you can refer to this wonderful free course on [A/B Testing](https://classroom.udacity.com/courses/ud257/lessons/4085798776/concepts/40713087720923)