We’re launching a redesigned home page for our site and want to study how it affects visitors. What are some of the key things you’d want to study? What metrics would you measure? Provide an explanation for every metric.

Our question is whether the redesigned home page affects visitors. The question doesn’t define the type of website that we’re dealing with, but let’s say that you can buy products from this website, and that there are links from ads to products as well. That being said, we’re going to measure three metrics. The first metric will measure the click-through rate. The second metric is measuring the amount of purchases made. The third metric is measuring the ratio of clicking on links to products and then purchasing those products. The first metric of clicks will measure overall activity of users who view the site due to the changes with the design. The second metric measure will tell us if the change in design has any effect on purchases or sales of products. The third metric will measure if the change in design affects the direct translation of clicks to purchases by measuring their ratio.

Using those metrics, how would you statistically evaluate if the new site was a success?

We need to sample our users for experimentation. We’ll randomly select a representative sample and control for region by splitting the study between different areas like New York, LA and Miami. 50% of users in each region are introduced to the site as treatment, and the rest continue using the same design (in each region). We’ll also divide the sample into age and gender groups to avoid the Simpsons paradox.

We’ll define a two-tailed null hypothesis to measure if the new design has any affect at all on our metrics, for example, “The change in website has no difference on mean click-through rates, purchases, or click-to-purchase translations”.

After a period of a month of collecting date, we’ll compare the difference between the clicks, purchases and ratios and calculate their significance between our samples as well as their demographics to reject our null hypothesis and see if the new design has a significant effect. This will happen separately in each region.

We can then conduct a one-tailed hypothesis test to see if the new designed increased or decreased our click and purchase rates. Based on a number we define as successful for our metric, we can decide if we want to implement the design completely, make changes, or scrap the idea.

How would you approach vary if we could offer our redesign to a subset of customers in perpetuity?

 \*

I would increase the time period of data collection from a month to 5 or 6 months in which a subset of customers will have access to the redesigned website as a treatment group. This way, the treatment sample has time to adjust to the new design.

After 6 months, we’d run a two-tailed t-test to see if there are any significant differences in the metrics between our treatment group and a sample that uses the original design. Then, we’d follow up with a one-tailed t-test to measure if there are significantly higher (or lower) measures of our metrics in the treatment group. We can decide to implement the new design indefinitely if the treatment group’s metrics (click-through rates, purchases, click-to-purchase ratios) are significantly higher.