Thinkful Data Science Boot Camp

Unit 1.4 Drill 3
Am I Biased?

For each of the following scenarios, call out the potential biases in the proposed experiment. Do your best to try to discover not only the bias, but the initial design. There is plenty of room for interpretation here, so make sure to state what assumptions you're making.

- **You're testing advertising emails for a bathing suit company and you test one version of the email in February and the other in May.**

  We can say that the sampling procedure is biased because of the span between versions as well as the months that each version was released. If the intent was to see what advertising email would work better, they should have been sent in the same period to more accurately measure response instead of having a two month span between the two.

  Also, the two months in which the advertisements were sent can be problematic and be considered biased, depending what region this is being advertised in. If the company is advertising in a place like Miami, where the temperature is relatively the same all year long, them there might not be a problem as people shop for bathing suits all year round. However, if they are advertising in a place like Chicago, then there would be a problem. We can assume that in February, people in Chicago would be least likely to buy swimsuits since it is still winter. We can then assume that in May, people in Chicago would be more likely to buy swimsuits because it is almost summer and would be much warmer than in February. This is seasonality bias.

- **You open a clinic to treat anxiety and find that the people who visit show a higher rate of anxiety than the general population.**

  In this case, sampling bias would be an issue. If the purpose of the study were to determine what percentage of the general population suffer from anxiety, using the patients from the anxiety clinic to make an assumption about the anxiety of the general population would not be ideal. Since this is an anxiety clinic, we can assume that those who come in will have anxiety problems or at least think that they have anxiety. This means that the percentage of people who will be coming to the clinic will always show a higher rate of anxiety than the general population.

- **You launch a new ad billboard based campaign and see an increase in website visits in the first week.**

  If the initial design was to see if a new billboard ad would increase website visits, then this ad can be considered a success. There are many factors that we do not know: product, location of ad, target audience, and time of the year. These are all factors that can affect the outcome of the initial design and can give us bias. For example, if we were put up a new billboard ad for swimsuits in May, in downtown Chicago, there would be an increase ifn website visits but it might not be because of the billboard. It might be because its May, it's warm and now people can enjoy Lake Michigan. There are too many unknowns in this example to determine if the

increased web traffic was a direct result of the billboard ad. This study would suffer from contextual bias as well as possible more, like seasonality in the swimsuit example.

- **You launch a loyalty program but see no change in visits in the first week.**

  A one week measurement to see if a loyalty program is effective may be to short of a time to measure any significant change. It is also unclear how the loyalty program was launched. What kind of marketing campaign was set? In what type of media was this loyalty program launched in? We can expect the conversion time to be longer for more traditional media like mailers or newspaper ads than Twitter, Instagram, or other social media. Another factor to consider is the program itself and if there is incentive enough for people to join?