Galvanize Tech Challenge Q3

Christopher Stewart

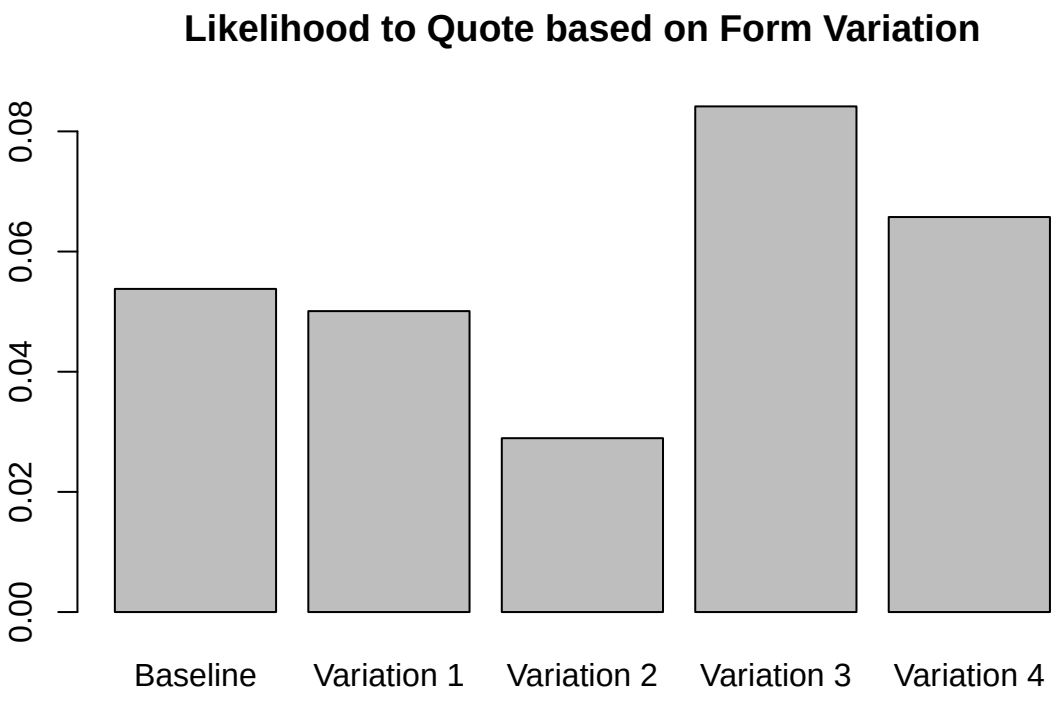
June 27, 2015

This document describes a split test analysis of quote form data. The goal is to ascertain how the 4 form variations affects service providers' likelihood to make a quote.

To begin, we read the data into R and make a simple barplot.

```
baseline = 32/595
variation_1 = 30/599
variation_2 = 18/622
variation_3 = 51/606
variation_4 = 38/578

data <- c(baseline, variation_1, variation_2, variation_3, variation_4)
barplot(data, main = "Likelihood to Quote based on Form Variation", names.arg=c("Baseline","Variation 1
```



What is the goal of this analysis?

If the goal is to determine which form is most effective, the barplot suggests that Variation 3 could be a good candidate. However, it is difficult to make a solid conclusion without knowing more about how the data were collected. Optimally each form would have the same number of viewers. It is also important to establish whether or not viewers were randomly distributed across the 5 variations.

Looking forwardly

I would begin by asking what the goal of the analysis is. Assuming that you want to find the most successful form variation in terms of likelihood to lead to a quote, I would begin by asking how viewers accessed the forms. This would me to establish why there are not a constant number of viewer numbers over the 5 form variations and, hopefully, the conditions under which the form variations were accessed.

If the goal was to model quote likelihood more broadly, it may be desirable to include category and location, for example, in a model in which they function as independent variables (possibly along with form variation), which porportion of quotes / viewers as the dependent variable.