**Maru Batting Center II**

**Using Monte Carlo Simulation of Individual Customers to Estimate CLV**

This assignment is a follow up to Maru Batting Center I where you calculated CLV for various segments and performed sensitivity analyses to assess how variation in assumptions affects conclusions about CLV at the aggregate level. In this case you will use an individual-level simulation approach to try to estimate CLV and compare the analysis to the conclusions you reached with respect to the aggregate level analysis.

Again, we will concentrate on the elite ballplayers (party) segment. Your objective is to run a simulation that randomly samples values for the annual margin, retention rate, and acquisition cost from a probability distribution of your choosing and computes CLV for a large number of “virtual customers.” You will analyze and visualize the CLV distribution of virtual customers and compare it to your conclusions based on the aggregate analysis.

Create a PowerPoint presentation that describes your modeling choices (i.e. which distributions did you sample from and why?), your results and interpretation.

The variables to simulate:

You will simulate annual margins and retention rates for your virtual customers. For all analyses use .1 for the interest rate and 50,000 for the acquisition cost. Note that in reality, both of these numbers have uncertainty, but for simplicity use the fixed values.

Annual Margin: To help you determine the right distribution to sample from look at the file “customers.rdata.” This shows annual margins for a subset of Maru’s past customers in the Elite Ballplayer segment. Load the data into R, visualize the data in a histogram or other format and calculate summary statistics. Based on this past data, decide on an appropriate distribution to sample annual margins for your virtual customer.

Retention Rate: The only thing you know about retention rate is that the average is 60%. You have no additional data about retention so you will have to use your intuitions and common sense to decide what kind of distribution makes sense. In order to decide on parameters for the distributions, play around and try out a bunch of different parameterizations. Pay attention to how variations in the distribution type and parameters you choose affect your conclusions.

Simulation: Write R code to generate a sample of virtual customers. You can choose how large of a sample to generate and you may try different sample sizes to see how this choice affects your results. Two notes: 1. Vector-based calculations are much more computationally efficient in R than For Loops. When possible, use vector computations. 2. If you want to be able to recreate your results exactly generate a random seed at the beginning of your code using the set.seed() function.

Analyses and Visualization: Create graphs and compute summary statistics based on your simulations. Your analyses should focus on two issues:

1. Value concentration. Value concentration refers to how much of your aggregate CLV is concentrated in a subset of customers. If value concentration is high, it means a small number of customers is responsible for a large percentage of your margins. Assess how value concentration depends on the distribution types and parameters you choose for the margin and retention rate distributions. Does value concentration depend on assumptions about how these variables are distributed? If so, how?

2. Comparison to aggregate level analysis. In Maru Batting Center I (last week’s case) you calculated an average CLV for customers in the elite ballplayer segment. Does the average CLV based on the simulation agree with the number you calculated in part 1? Do assumptions about the distribution of margins appear to affect the extent of agreement or disagreement? How about assumptions about retention rates? Speculate on why the individual analysis does or does not agree with the aggregate level analysis from part 1.

Finally, do you have any high-level recommendations about the kind of data that it would be useful to track or collect in order to improve the simulation and get more precise estimates of CLV?

Bonus: Repeat the simulation process for the little leaguers segment. Based on this simulation and all your analysis so far, which segment is more important to Maru, little leaguers or elite ballplayers?