Using R: Bootstrap and Monte Carlo Simulation

Definition of Bootstrap and Monte Carlo Simulation

The goal of analysis is to compute the 95% confidence intervals for values derived from Willingness to Pay (WTP), which is calculated by beta coefficients associated with product attributes such as Screen85Inch, Resolution 4K, and Sony. Willingness to Pay differs by coefficients and customers' utility per dollar which indicates this is a variable without knowing the exact probability distribution. However, if we want to expect the range within which the real WTP values are located by calculating CI, there are three methods we can use such as Residuals Bootstrap, Data Bootstrap and Monte Carlo Simulation. Residual Bootstrap resamples residuals from the original model with replacement to generate new datasets where new coefficients will be extracted used for calculating WTP values. Data Bootstrap resamples the rows of the original dataset directly with replacement to create new datasets which are also used for calculating beta coefficients used for the calculation of WTP values. Monte Carlo Simulation does not repeat the process of linear regression to calculate coefficients for each new dataset but it uses the estimated coefficients and their variance and covariance matrix from the original dataset and multivariate normal distribution is used to generate new coefficients.

Goals and Dataset Review

The goal of this analysis uses three different methods for calculating the confidence intervals of WTP when the distribution of WTP is not defined. We cannot use the normal distribution to calculate the confidence interval of WTP unless there is an assumption that the distribution of WTP is a normal distribution. However, if we want to find a specific range where the true values of WTP are located, three different kinds of methods such as Residuals Bootstrap, Data Bootstrap and Monte Carlo Simulation are useful because they do not need to assume the probability distribution of the target variable. The dataset is the same as that we have used for the conjoin analysis. Utilizing 'Customer Preference' dataset, WTP (Willingness to Pay) is calculated based on beta coefficients extracted from the original dataset but these values differ when resampling with replacement, indicating WTP is the variable. All independent variables such as screen size, brand, resolution are dummy variables. 'Screen 75 inch', 'Screen 85 inch' are dummy variables with the baseline variable of 'Screen 65 inch'. The 'Resolution 4K = 1' is a dummy variable with the baseline variable of 'Resolution 1K = 0'. Sony is encoded as 1 with the baseline variable of 'Sharp = 0'. Resampling is executed 1000 times and the 95% confidence interval is calculated for each method.

Mechanism of Residual Bootstrap, Data Bootstrap and Monte Carlo Simulation and Outcomes

(1) Residual Bootstrap

The Residual Bootstrap resamples residuals from the original regression model with replacement. This process repeated 1000 times. The new predicted y value is calculated by adding the original predicted value to these new resampled residuals. The new predicted y regresses on the original X independent variables to extract new beta coefficients used for

calculating the WTP. The confidence interval is calculated for the WTP values derived from each beta coefficient associated with features of the linear regression, such as Screen size, Resolution and Brand and the quantile function is used to determine the 2.5% and 97.5% points of the WTP distribution for each feature. Here is the result of the confidence intervals for the WTP values of each beta coefficient.

(2) Data Bootstrap

The data bootstrap analysis resamples the rows of the original matrix of dataset with replacement to generate new datasets. This method fits a linear regression using each new dataset and calculate new beta coefficient estimates used for calculating WTP values. When calculating the confidence intervals, each confidence interval for beta coefficients associated with WTP is respectively calculated. Here is the result of the confidence intervals for the WTP values of each beta coefficient.

(3) Monte Carlo Simulation

The Monte Carlo Simulation does not fit a linear regression repeatedly to extract different beta coefficient estimates for the calculation of WTP. This method fits a linear regression using the original dataset to extract beta coefficient estimates with intercepts as well as a covariance matrix. New beta coefficients follow the multivariate normal distribution with the mean of beta coefficients matrix from the original dataset and the covariance matrix. This multivariate resamples 1000 times and extracts new beta coefficient estimates. Compared to bootstrap methods, the Monte Carlo Simulation utilizes the normal distribution to resample datasets. Here is the result of the confidence intervals for the WTP values of each beta coefficient.

Conclusion

The two different bootstrap methods are different in the perspective of uncertainty. The Residual Bootstrap uses the original independent variable when extracting new beta coefficients by regressing Y on the original X independent variables. On the other hand, the Data Bootstrap uses both new dependent and independent variables, which increase overall uncertainty compared to the Residual Bootstrap. The Monte Carlo Simulation uses a multivariate normal distribution when resampling new beta coefficients based on the original beta coefficient estimates and their covariance matrix, which is different from the bootstrap methods only relying on the linear regression repeatedly to extract new beta coefficients for calculating WTP.