



Mathematical Statistics
Stockholm University
Bachelor Thesis **2015:27**
<http://www.math.su.se>

Generating Predictive Behavior Models from Web Analytics Data

Kiowa Jackson*

December 2015

Abstract

In this work we seek to evaluate three different datasets from different scenarios to see ways in which behavior of web visitors can be evaluated statistically, for purposes of predicting future behavior. Our aim is to create predictive models that can be used as a starting point for modeling the potential value of a web visitor. These models, then, could be used for predicting acquisition costs and inform overall on-line marketing strategy. The data was collected with Google Analytics (both standard and premium versions), and includes features such as age, gender, location and the number of visitors who took a particular action. The multiple regression model and Pearson's Chi-Square test of independence have been used to analyze the data. The three different datasets examined respectively the difference in performance for two variations of a homepage, the relationship between age and estimated value of a visitor, and the relationship between age, gender and the likelihood of the visitor making a purchase on the website. The results show statistically significant relationships in all three datasets, finding that the new version of the homepage has significant impact on conversion rate; estimated value of a web visitor drops off with age; and women age 25-34 from the West region have the highest predicted likelihood of making a purchase on the website.

As Google Analytics, and especially Google Analytics Premium, makes available a large amount of data more advanced models can be built utilizing a larger set of features. The models developed in this work can serve as initial proof of concepts to benchmark more complex models against.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: kiowa.jackson@yahoo.com. Supervisor: Mathias Lindholm.