DATA DRIVEN DEVELOPMENT

Matthew Newman, Sonia Xu, Alex Zrenner
Project Manager: Sheng Jiang

Summary

- Duke Alumni Affairs and Development Office provided a dataset containing 163,060 observations of Duke alumni from 1972 – 2015
- We analyzed characteristics of Duke alumni and giving habits to statistically identify potential large donors
  - Segmented alumni into 11 groups based on total lifetime giving (TLG)
  - Fitted predictive models in response to TLG
  - Clustered alumni into 11 groups based on similarities

EDA & METHODS

Exploratory Data Analysis

- Identified 11 segments of alumni from the local minima of the distribution of TLG

Total Lifetime Giving Not Including Gifts of Zero, Logged
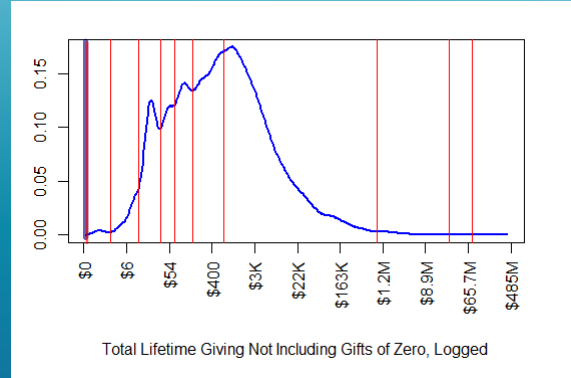
Methods and Equations

- Logistic Regression and LASSO variable selection
  - $\min - [\frac{1}{N} \sum_{i=1}^{N}([y_i(\beta_0 + x_i^T \beta) - \log(1 + e^{\beta_0 + x_i^T \beta}))] + \lambda ||\beta||_1$
  - $P(is\ donor) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
- Multinomial Logistic Regression and LASSO variable selection
  - $\min - [\frac{1}{N} \sum_{i=1}^{N}(\sum_{k=1}^{K}([y_i(\beta_0 + x_i^T \beta_k) - \log(1 + e^{\beta_{0k} + x_i^T \beta_k}))] + \lambda ||\beta_k||_1$
  - $P(Alumni\ belongs\ to\ group\ i) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
- Tobit Model
  - $Total\ Lifetime\ Giving = \begin{cases} \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots if\ Total\ Lifetime\ Giving > 0 \\ 0\ if\ Total\ Lifetime\ Giving \leq 0 \end{cases}$
- Clustering with Divisive Analysis Clustering algorithm, diana()

GRP	# of Alumni	Mean TLG
0	54,113	$0
1	301	$1
2	2,756	$9
3	10,670	$23
4	7,457	$50
5	13,060	$109
6	24,350	$386
7	49,915	$18,449
8	423	$3 m
9	10	$41 m
10	5	$140 m

CONCLUSIONS

Key Assumption:

Similar alumni share similar giving habits

- Density plots of clusters and groups segmented by TLG should be similar

Clustering Results *20% sample of dataset

- Shown is the distribution of TLG segmented by the diana() clusters and a table summarizing clusters
- Two plots are NOT similar
- There exist alumni who don't give to Duke in the same capacity that their "similar" peers do

By identifying these alumni, the Duke Alumni Affairs and Development Office can target and cultivate them to further Duke Forward's fundraising.

Density Plots of "True" Clusters

Density Plots of Clusters Based on Characteristics