MA478 TEE

Cooper Klein

May 2024

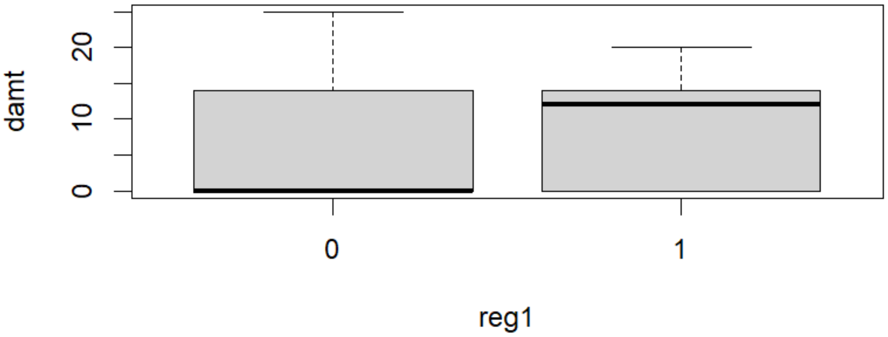
1 Data

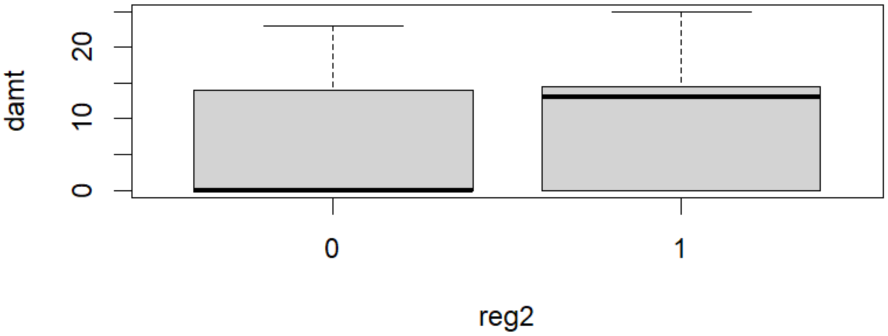
1.1 Numeric Predictors

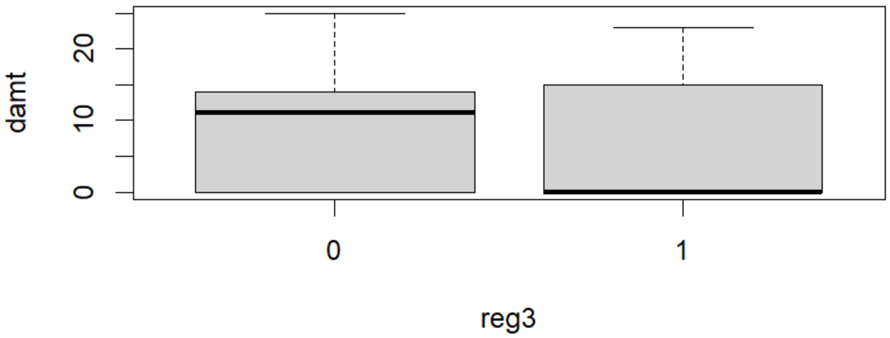
Description	**Non-Donor Mean**	**Donor Mean**
Number of Children	2.33	0.830
Household Income (7 categories)	3.91	3.98
Wealth rating (scale of 0-9)	6.47	7.63
Average Home Value in Neighborhood	176	194
Average Family Income in Neighborhood	53.6	60.7
Percent low-income in Neighborhood	15.6	11.9
Number of lifetime promotions	57.5	65.7
Dollar amount of lifetime gifts	107	127
Dollar amount of largest gift	22.4	24.0
Number of months between 1st and 2nd gift	6.81	5.79

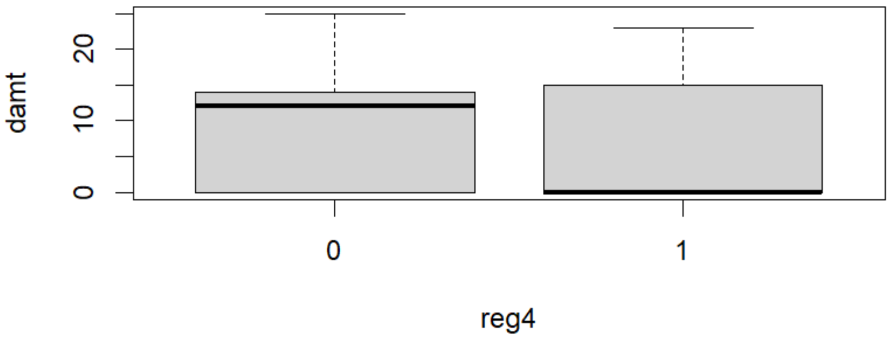
Table 1: Comparison of numeric statistics for Donors and Non-Donors

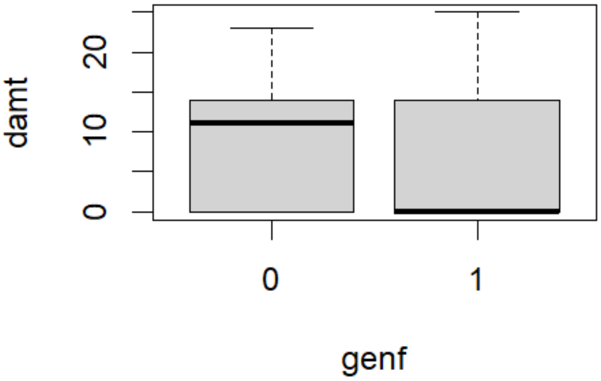
1.2 Region Comparison

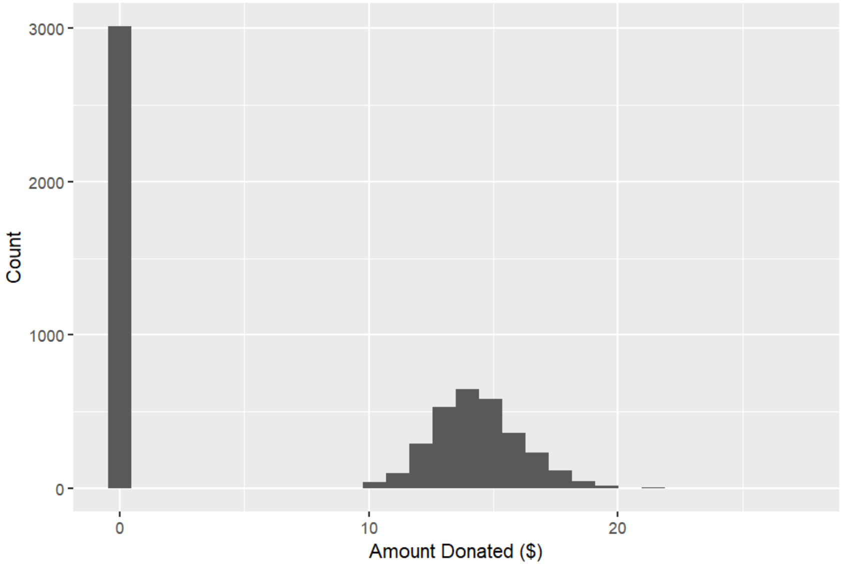





damt
0
10
20
0
1
reg3


damt
0
10
20
0
1
reg4





2 Models

2.1 Model 1: Full Binary Classification and Linear Regression

Our first model predicts whether an individual will be a donor with a binary classification, and then predicts the amount donated with a linear regression. This model uses all available predictors for both the logistic and linear regression models.

$$\begin{aligned}\text{logit}(p_i) = &-0.25 + 0.56Region1_i + 1.21Region2_i + 0.02Region3_i - 0.01Region4_i \\ &- 1.96Child_i + 1.11Homeowner_i + 0.82WRAT_i - 0.03Female_i \\ &+ 0.37INCM_i + 0.01INCA_i - 0.19PLOW_i - 0.44TLAG_i \\ &+ 0.032AGIF_i - 0.46TLAG_i - 0.23TDON_i - 0.03LGIF_i + 0.14TGIF_i\end{aligned} \tag{1}$$

where:

- $z_i = \begin{cases} 1 & \text{if Donor} \\ 0 & \text{if non-donor} \end{cases}$
- $z_i \sim Ber(p_i)$

The linear regression was fit as follows:

$$\begin{aligned}\text{DAMT} = &14.2 - 0.04Region1_i - 0.07Region2_i + 0.32Region3_i + 0.64Region4_i \\ &- 0.60Child_i + 0.24Homeowner_i - 0.01WRAT_i - 0.03Female_i \\ &+ 0.37INCM_i + 0.01INCA_i - 0.19PLOW_i - 0.44TLAG_i + 0.03AGIF_i \\ &- 0.003TLAG_i + 0.07TDON_i - 0.03LGIF_i + 0.14TGIF_i + \epsilon\end{aligned} \tag{2}$$

2.2 Model 2: Subset Binary Classification and Linear Regression

Our second model similarly predicts whether an individual will be a donor with a binary classification, and then predicts the amount donated with a linear regression. This model uses a smaller subset of available predictors, that are believed to hold more significance from our data exploration.

The logistic regression was fit as follows:

$$\begin{aligned}\text{logit}(p_i) = &-0.22 + 0.56Region1_i + 1.18Region2_i + 0.03Region3_i + 0.01Region4_i \\ &- 1.91Child_i + 1.08Homeowner_i + 0.79WRAT_i - 0.02Female_i \\ &+ 0.35INCM_i + 0.06INCA_i - 0.19PLOW_i - 0.44TLAG_i + 0.032AGIF_i\end{aligned} \tag{3}$$

where:

- $z_i = \begin{cases} 1 & \text{if Donor} \\ 0 & \text{if non-donor} \end{cases}$
- $z_i \sim Ber(p_i)$

The linear regression was fit as follows:

$$\begin{aligned}\text{DAMT} = &14.28 - 0.04Region1_i - 0.09Region2_i + 0.34Region3_i \\ &+ 0.66Region4_i - 0.53Child_i + 0.26Homeowner_i + \epsilon\end{aligned} \tag{4}$$

2.3 Model 3: Zero-Inflated Poisson

$$\log \lambda_i = 2.66 - 0.01 Region1_i - 0.02 Region2_i + 0.07 Region3_i + 0.01 Region4_i - 0.02 Child_i + 0.002 TFIG_i \tag{5}$$

$$\text{logit}(\phi_i) = -0.29 - 1.13 Region1_i - 2.11 Region2_i - 0.04 Region3_i + 0.07 Region4_i + 1.15 Child_i - 0.01 TGIF \tag{6}$$

where:

- $y_i \sim \begin{cases} 0 & \text{with prob}\phi_i \\ & \text{Pois}(\lambda_i) w/probability(1-\phi_i) \end{cases}$
- $Z_i \sim \text{Bern}(\phi_i)$
- $Y_i \sim \text{Po}(\lambda_i)$

3 Results

We will evaluate our 3 models on their AIC and their MSE results on our withheld test set. Additionally, we will present the profit curve for our best model.

Model	**AIC**	**MSE**
Model 1 (binary classification)	2804.3	1.76
Model 1 (linear regression)	6646.78	1.76
Model 2 (binary classification)	2916.71	2.12
Model 2 (linear regression)	7880.144	2.12
Model 3 (zero-inflated Poisson)	13119.15	1.77

Table 2: AIC Performance of Models

From Table 2, we will recommend our model combining logistic regression and linear regression with all available predictors. We deduce this is our preferable model because it yields the lowest MSE and lowest AIC. We believe, however, that our zero-inflated poisson model offers the most potential, we just did not have enough time to fully develop the model with the optimal subset of predictors.

F

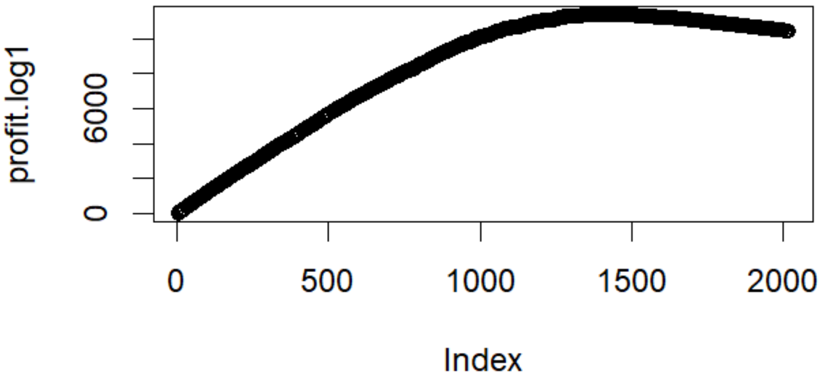


Figure 1: Expected Profit Curve of Model 1