Title: Increased loan interest rate associated with increased amount requested

Introduction:

When a borrower applies for a loan with a lending institution information -- such as FICO credit score, the amount requested, and the borrower's loan history -- are provided to aid the lender in determining the interest rate of the loan.

The criteria that influences this decision most strongly is the borrower's FICO credit score [1]. Beyond this score many other variables are considered, some influencing the interest rate more than others. This analysis was performed to determine which, if any, of these secondary variables were significantly associated with the resulting loan interest rate.

Using exploratory analysis and linear regression models this analysis shows that there is a positive, highly significant association between interest rate and the amount of money requested by the borrower. Even when taking into account the borrower's FICO score this association remains.

Methods:

Data Collection

The data for this analysis was provided by Prof. Leek from a central AWS repository [2] and was collected from The Lending Club [3]. The data was downloaded for this analysis on Wednesday, February 6, 2013. The data consists of 2500 peer-to-peer loans and provides information about the loan, such as interest rate, and the borrower, such as FICO credit score.

Exploratory Analysis

Examination of the data for missing values revealed only two samples with missing data. Because this represented such a small percentage of the overall data they were omitted. Examination of the structure of the data led to three data transformations (in each case converting strings to numerics) all made to ease data visualization and statistical analyses.

Exploratory analysis was then performed on this cleaned data set in the form of basic plots and examination of distribution within variables to assess the quality of the data and to determine which variables were likely to have the most important associations with loan interest rate and should be analyzed further.

Statistical Modeling

To evaluate associations with the interest rate variable, linear regression models were fit to determine potential covariates [4]. For each linear model generated 1) the p-value was noted to

determine the probability of the variable being predictive for the interest rate and 2) the confidence intervals were evaluated for the coefficients to determine the reliability.

Results:

The sample data included information about loans and loan borrowers. Exploratory analysis of the data found a small amount of missing data, 2 of 2500 samples included missing data, and these samples were omitted from further analysis. Analysis of the distribution of values for each variable did not reveal any values out of the assumed normal ranges.

Several variables, when evaluated with a linear model, were determined to be correlated with interest rate. Those which were deemed to be more likely to be a result of the evaluation of the borrower and therefore not predictive, but rather outcomes, were omitted. These variables included loan length and amount funded by investors.

Three variables were found to be highly correlated with interest rate, and each were evaluated as possible confounding variables: amount of money requested, loan inquiries made in the last 6 months, and number of open lines of credit. Of these, both amount requested and inquiries made in the last 6 months had the smallest p-value the analyzing machine could evaluate (< 2.2e-16), while open lines of credit had a larger p-value (6.169e-06) indicating of the three it was the weakest predictor of interest rate.

When taking into account FICO credit score the variable amount requested outperformed inquiries in the last 6 months. FICO credit scores were binned into five groups and a linear model was fit for each variable with interest rate as the response. Amount requested produced a smaller p-value for 4 of the 5 FICO credit score tiers, indicating it was the stronger predictor of the two variables when considering FICO credit score.

The linear model produced by the R code lm(data$Rate~data$Amount.Requested) -- where Rate represents the resulting loan interest rate and Amount.Requested represents the amount of money requested by the borrower -- produced a p-value of <2.2e-16, indicating a highly statistically significant association between the loan interest rate and the amount of money the borrower requested. The standard error was found to be 1.011e-05. When evaluating the confidence intervals for this same linear model it was observed that a $1 increase in amount requested is associated with a 0.00017% interest rate increase (95% CI: 0.00016% - 0.00020%) (Figure 1).

Conclusions:

This analysis suggests that there exists a positive association between the interest rate of a loan and the amount of money requested by the borrower. When taking into account the strongest predictor for interest rate, the FICO credit score of the borrower, the amount requested remains a strong predictor. It can be assumed based on this analysis that it is likely when two borrowers

with the same FICO credit score apply for a loan at the Lending Club the borrower requesting the larger amount will receive the loan with the higher interest rate.

It should be noted that relatively few samples, only 2500, were included in this data set. A larger sample, or a sample selected with a different method, may produce different results. Further, this data was collected from the Lending Club, a peer-to-peer loan network which may use different criteria for assigning loan interest rates than a traditional bank or even a different peer-to-peer lending system. It should not be assumed that the results of this analysis would apply to these.

References:

[1] FICO credit score http://en.wikipedia.org/wiki/FICO_score#FICO_score
[2] Coursera data download URL
https://spark-public.s3.amazonaws.com/dataanalysis/loansData.rda
[3] The Lending Club https://www.lendingclub.com/home.action
[4] R: Fitting Linear Models http://stat.ethz.ch/R-manual/R-patched/library/stats/html/lm.html