Submission Phase

1. Do assignment ☐ (/dataanalysis-001/human_grading/view/courses/294/assessments/4/submissions)

Evaluation Phase

2. Evaluate peers 🔒 (/dataanalysis-001/human_grading/view/courses/294/assessments/4/peerGradingSets)
3. Self-evaluate 🔒 (/dataanalysis-001/human_grading/view/courses/294/assessments/4/selfGradingSets)

Results Phase

4. See results 🔒 (/dataanalysis-001/human_grading/view/courses/294/assessments/4/results/mine)

No work was submitted before the submission deadline. You will not be able to evaluate the work of your peers or receive an evaluation.

Data

For this analysis you will use the loans data available from here:

https://spark-public.s3.amazonaws.com/dataanalysis/loansData.csv (https://spark-public.s3.amazonaws.com/dataanalysis/loansData.csv)
https://spark-public.s3.amazonaws.com/dataanalysis/loansData.rda (https://spark-public.s3.amazonaws.com/dataanalysis/loansData.rda)

There is a code book for the variables in the data set available here:

https://spark-public.s3.amazonaws.com/dataanalysis/loansCodebook.pdf (https://spark-public.s3.amazonaws.com/dataanalysis/loansCodebook.pdf)

Prompt

The data above consist of a sample of 2,500 peer-to-peer loans issued through the Lending Club (https://www.lendingclub.com/home.action (https://www.lendingclub.com/home.action)). The interest rate of these loans is determined by the Lending Club on the basis of characteristics of the person asking for the loan such as their employment history, credit history, and creditworthiness scores.

The purpose of your analysis is to identify and quantify associations between the interest rate of the loan and the other variables in the data set. In particular, you should consider whether any of these variables have an important association with interest rate after taking into account the applicant's FICO score. For example, if two people have the same FICO score, can the other variables explain a difference in interest rate between them?

What you should submit

Your data analysis submission will consist of the following components:

1. The main text of your document including a numbered list of references. This can be uploaded either as a pdf document or typed into the text box (not both!). The limit for the text and references is 2000 words. Your main text should be written in the form of an essay with an introduction, methods, results, and conclusions section.
2. One figure for your data analysis uploaded as a .png, .jpg, or .pdf file, along with a figure caption of up to 500 words.

Reproducibility

Due to security concerns with the exchange of R code, you will no longer be asked to submit code to reproduce your analyses. I still believe reproducibility is a key component of data analysis and I encourage you to create reproducible code for your data analysis.

Submission Deadline

You must submit your data analysis by February 18th, 2013 at 7:00AM UTC-5:00 (Baltimore time). No late days may be applied to the data analysis. Note that this is an extension of the original date posted on the class website.

Please either enter the body of your data analysis in the text box or upload a pdf file with your analysis. This file should both contain the main text of your analysis and the numbered list of references. It may be no more than 2000 words.

Introduction

There are many alternate options for lending and borrowing money without using banking system. One of them is available on Lending Club (https://www.lendingclub.com (https://www.lendingclub.com)). By allowing their members to directly invest in and borrow from each other, they bypass the complexity of the banking system and provide better rates to borrowers and better returns to investors.

In order to determine the interest rate of loans, Lending Club use a system of characteristics of the person asking for the loan such as their employment history, credit history, and creditworthiness scores. FICO score is the key component of this system. It provides a snapshot of risk that banks and other institutions use to help make lending decisions. Applicants with higher FICO scores might be offered better interest rates on mortgages or automobile loans as well as higher credit limit amounts [1].

Understanding the relationship between interest rate and FICO score and/or other related variables will give us information whether the FICO score alone or any other variables have an important association with final interest rate at Lending Club.

Methods

Data Collection
For our analysis we used the data on sample of 2,500 peer-to-peer loans issued through the Lending Club https://www.lendingclub.com/home.action (https://www.lendingclub.com/home.action)).. The data were downloaded from data.gov

on January 31, 2013 using the R programming language [3].

Exploratory Analysis

Exploratory analysis was performed by examining tables and plots of the observed data. We identified transformations to perform on the raw data on the basis of plots and knowledge of the scale of measured variables. Exploratory analysis was used to (1) identify missing values, (2) verify the quality of the data, and (3) determine the terms used in the regression model relating earthquake magnitude to earthquake depth.

Statistical Modeling

To relate earthquake magnitude to earthquake depth we performed a standard multivariate linear regression model [4]. Model selection was performed on the basis of our exploratory analysis and prior knowledge of the relationship between geography and earthquake

magnitude. Coefficients were estimated with ordinary least squares and standard errors were calculated using standard asymptotic approximations [5].

Reproducibility

All analyses performed in this manuscript are reproduced in the R markdown file earthquakesFinal.Rmd [6]. To reproduce the exact results presented in this manuscript the cached version of the analysis must be performed, as the data available from data.gov changes based on the date.

Results

Conclusions

FICO wiki http://en.wikipedia.org/wiki/Credit_score_in_the_United_States (http://en.wikipedia.org/wiki/Credit_score_in_the_United_States)

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Does the analysis have an introduction, methods, results, and conclusions section?

Are figures labeled and referred to by number in the text?

Is the analysis written in clear and understandable English?

Are the names of variables reported in plain language, rather than in coded names?

Does the analysis report the number of observations/samples?

Does the analysis report any missing data or other unusual features?

Does the analysis include description and justification for data transformations?

Does the analysis include a discussion of potential confounders?

Are the statistical models correctly applied?

Are estimates reported with appropriate units and measures of uncertainty?

Are estimators/predictions appropriately interpreted?

Does the analysis make concrete conclusions?

Please upload the figure for your data analysis here. It must be in .png, .jpg, or .pdf format. Type in a corresponding figure caption (no more than 500 words).

h

Evaluation/feedback on the above work

Note: this section can only be filled out during the evaluation phase.

Is the figure caption descriptive enough to stand alone?

Does the figure focus on a key issue in the processing/modeling of the data?


PDF-XChange
Click to buy NOW!
www.docu-track.com


PDF-XChange
Click to buy NOW!
www.docu-track.com