Mortgage Approval Analysis Project

This project will follow a paper published in the *American Economic Review*. The purpose of the paper was to test for the presence of racial discrimination in the mortgage approval process. The dataset provided contains records of the 2,380 applicants for mortgages in Boston.

Instructions

1. Create a folder called `Project`.

2. Read the paper as it provides a detailed discussion of the data and insight on the determinants of the probability of being approved.

3. Bring in the dataset provided within my repository. The variables in the dataset do not have intuitive names (e.g., the meaning of S3 is unclear). Referencing the data description and the AER paper, identify the qualitative dependent variable that you will be modeling and the set of co-variates: debt-to-income ratio, race, self-employed, marital status, and education indicator variables.

4. Generate summary statistics on the set of variables selected, and explain the composition of the sample and of the characteristics of an average (representative) applicant. In the process, you should also generate scatterplots, histograms, and frequency counts on particular variables of interest, which can be referenced in your explanation of the composition of the sample and of a representative applicant.

5. What is the baseline probability of an individual being approved for a mortgage?

6. Based on the data you read in, create a table with the following structure (values will not be the same as the example):

Applicant Race	Approved	Not Approved	Total
Black	23	32	55
White	75	25	100
Total	98	57	155

Table 1: Example Mortgage Approval Table

7. From the table you create in step 6, calculate the following:

- $P(\text{Approved}|\text{White})$
- $P(\text{Approved}|\text{Black})$