Interest Rate on Loan Analysis

Introduction

The objective of this analysis is to utilize the random sample with 10000 subjects for analyzing the factors related to the borrowers' interest rate on their loan in the United States. Data in this random sample comes from the Lending Club platform, which is a platform that allows individuals to lend to other individuals. This analysis will focus on the association between the interest rate on a borrower's loan in the United States and the home-ownership status of a borrower. Thus, this analysis will answer the following question: Is there a difference in the average interest rate on a loan based on a borrower's home-ownership status?

Methods

This analysis used Analysis of Variance (ANOVA) to compare the variability of interest rate among the different borrowers' groups with different home-ownership status because the two variables studied in this analysis were one numerical variable and one categorical variable with 3 categories.

Exploratory data analysis

First, we looked at the average interest rate, the median of interest rate, the standard deviation of interest rate as well as the number of cases in each group of different home-ownership status. (See Table 1)

Interest Rate in different groups of home-ownership status

Home-ownership Status	Average interest rate	Median of interest rate	Standard deviation of interest rate	Number of cases
MORTGAGE	12.1	10.9	4.96	4789
OWN	12.3	12.0	4.92	1353
RENT	12.9	12.0	5.04	3858

Table 1

Hypotheses

We would like to study whether there is a real difference in the average interest rate regarding to the borrowers' home-ownership status. So, we are interested in testing if the average interest rates in group of "mortgage", "own" and "rent" are the same.

So, we set up the following hypotheses:

H_0: $\mu_{mortgage} = \mu_{own} = \mu_{rent}$

H_A: At least one of these means is not equal to the others.

Where $\mu_{mortgage} = \mu_{own} = \mu_{rent}$ are the true average interest rate for the borrowers in each home-ownership status group.

Check for Independence

Since the data used in this analysis is from a random sample, the independence can assume to be hold.

Check for Normality

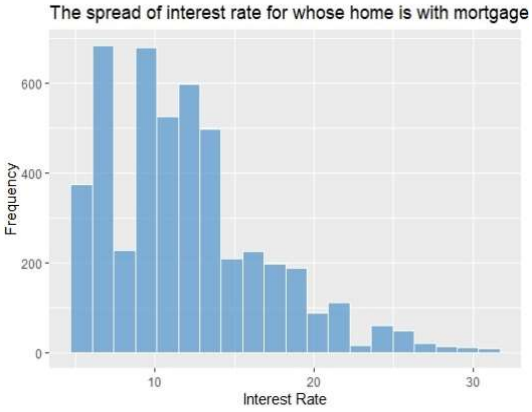

Figure1

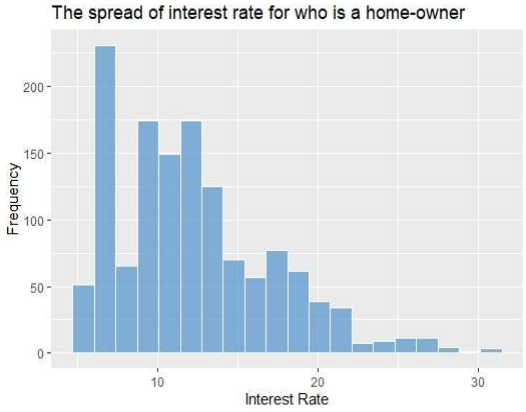

Figure 2

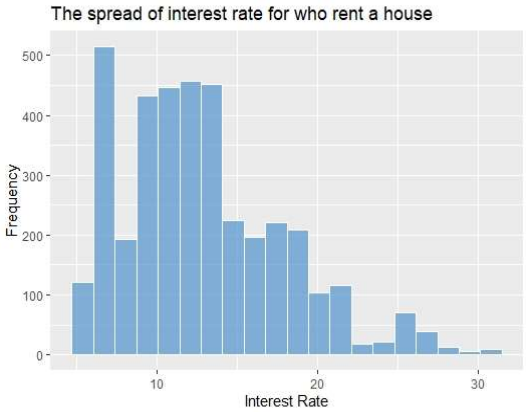

Figure3

Mortgage: The interest rate distribution of mortgage group looks like bimodal and right skewed. However, there are 4789 observations in this group, normality can be assumed to hold. (See figure 1)

Own: The interest rate distribution of own group looks like bimodal and right skewed. However, there are 1353 observations in this group, normality can be assumed to hold. (See figure 2)

Rent: The interest rate distribution of rent group looks like bimodal and right skewed. However, there are 3858 observations in this group, normality can be assumed to hold. (See figure 3)

Check for Constant Variance

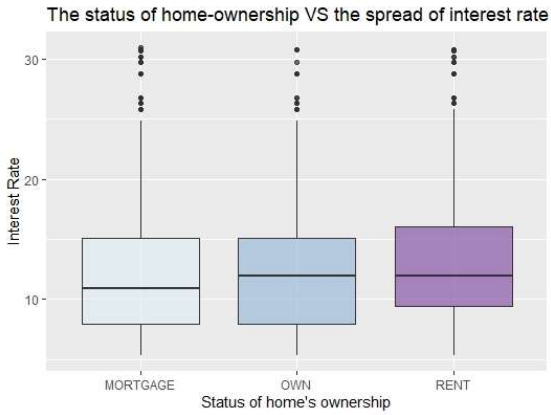

Figure4

The constant variance assumption seems to hold well. The boxplots are of the equal length and the IQR of these three boxplots looks similar. The IQR of rent boxplot is a little bit greater than another two groups. But the difference is so small that we can assume the constant variance to be hold.

Since all the assumptions needed to be check for Analysis of Variance are met, we can process the ANOVA test.

Results

Conduct Test for Analysis of Variance (ANOVA)

The result of ANOVA test

ANOVA test	Degree of Freedom	Sum of Squares	Mean Sum of Squares	F-Statistic	P-value
Home-ownership status	2	1623	811.5	32.65	7.35e-15
Residuals	9997	248463	24.9		

Table 2

The significance level was set at 0.05. The p-value from this ANOVA test is only 7.35e-15 which is much smaller than 0.05 so that we can reject the null hypothesis and conclude that there is some difference in the average interest rate on a loan based on the home-ownership status.

Pairwise t-tests with a Bonferroni correction

The result from Pairwise t-tests is as following:

```
        Pairwise comparisons using t tests with pooled SD

data:  loans$interest_rate and loans$homeownership

     MORTGAGE OWN
OWN  0.28954  -
RENT 3.2e-15  0.00032
```

Figure5

The p-value comparing the average interest rate in home-ownership status as mortgage to own is 0.28, which is greater than 0.05. Thus, we cannot posit there is significant difference in interest rate between these two groups. But the p-value comparing mortgage to rent is 3.2e-15 as well as the p-value comparing own to rent is 0.00032 are both much smaller than 0.05, so, we can conclude that there is significant difference in the average interest rate for whose home is with mortgage and for who rents the house. Also, we can conclude that there is significant difference in the average interest rate for who owns the house and for who rents the house.

Discussion

Regarding to the random sample from the Landing Club Platform, we can conclude that there is significant difference in the average interest rate on a loan comparing the borrower's home-ownership status as rent to the other two, mortgage and own. However, when comparing the borrower's home-ownership status as mortgage to own, we cannot posit there is significant difference in the average interest rate between them.

When we checked the normality of the distributions of interest rate in each category, three distributions are all right skewed. It might impact the result of the ANOVA test.

I am still considering whether the loan information from the Landing Club Platform can be representative of all the loans in the U.S. because the Landing Club Platform is an individual-to-individual platform. We cannot avert the data bias based on the data only from one channel, which is only focused on a specific market. Therefore, it is possible that the sample is not representative of all the loan borrowers in the U.S.

Before our research team publishes the final analysis report, we might need to search for more data from different channels and manipulate data much more precisely for the normality.