MKT 6337

Homework 1

1. Read the big\_loan data

**Please see CODE**

1. What variables are in the data?

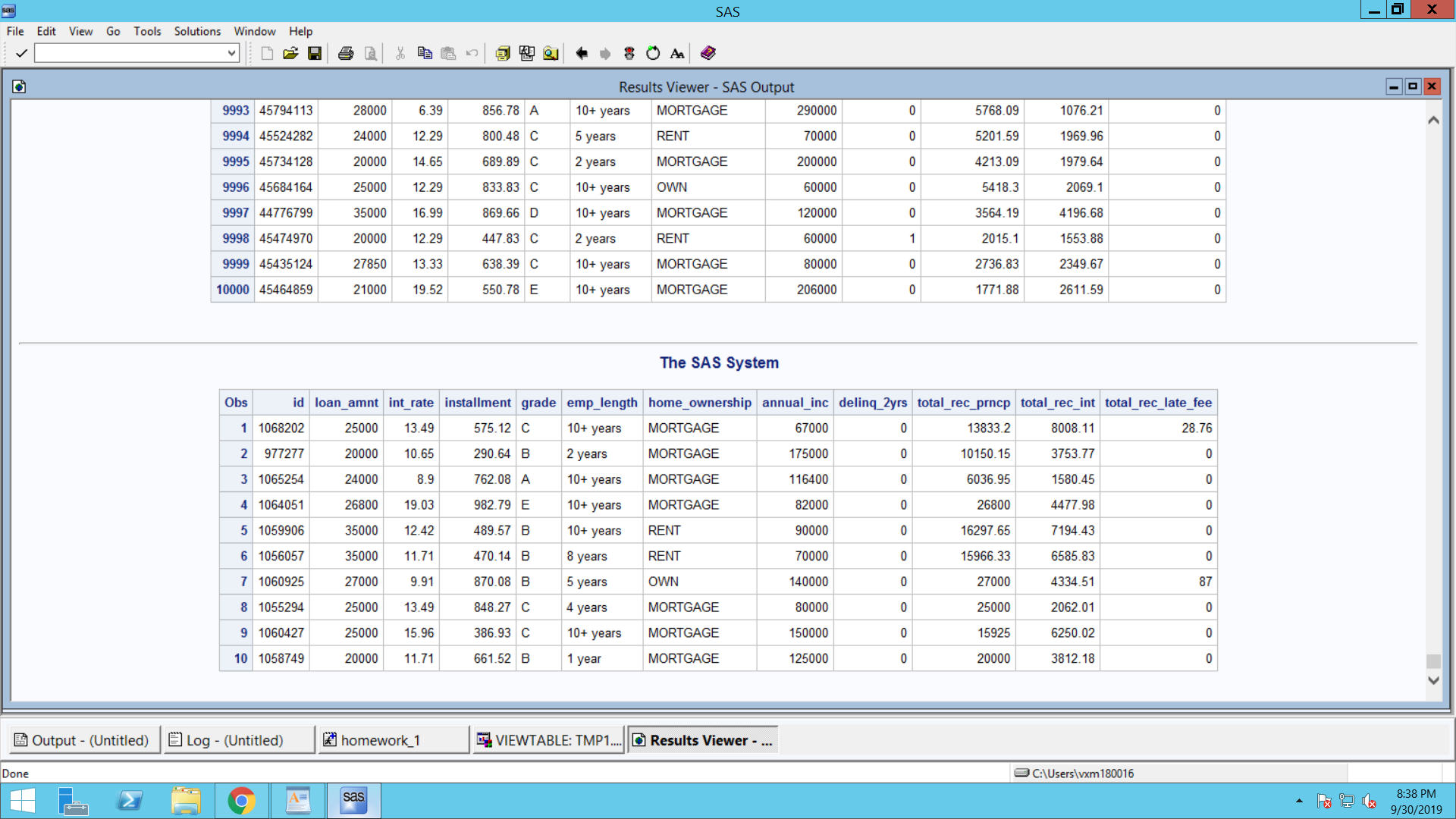
annual\_inc, delinq\_2yrs, emp\_length, grade, home\_ownership, id, installment, int\_rate, loan\_amnt, total\_rec\_int, total\_rec\_late\_fee, total\_rec\_prncp

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| annual\_inc | Num |  | 8 | BEST12. | BEST32. |
|  | delinq\_2yrs |  | Num | 8 | BEST12. | BEST32. |
|  | emp\_length |  | Char | 9 | $9. | $9. |
|  | grade |  | Char | 1 | $1. | $1. |
|  | home\_ownership |  | Char | 8 | $8. | $8. |
|  | id |  | Num | 8 | BEST12. | BEST32. |
|  | installment |  | Num | 8 | BEST12. | BEST32. |
|  | int\_rate |  | Num | 8 | BEST12. | BEST32. |
|  | loan\_amnt |  | Num | 8 | BEST12. | BEST32. |
|  | total\_rec\_int |  | Num | 8 | BEST12. | BEST32. |
|  | total\_rec\_late\_fee |  | Num | 8 | BEST12. | BEST32. |
|  | total\_rec\_prncp |  | Num | 8 | BEST12. | BEST32. |

1. Write SAS code to print 10 observations?

**PROC** **print** Data = data(obs= **10**);

**RUN**;



1. What is the average loan amount?

$25935.64

1. How many home-owners are in the data?

909

1. What is the average loan amount across different categories of home-ownership?

**MORTGAGE** 26259.51

**OTHER** 22475.00

**OWN** 25769.31

**RENT** 25219.74

1. Do a ttest to check if the average loan amount is different between mortgage owners and renters.

* Who has higher average loan amount?  **MORTGAGE**
* What is the difference in average loan amount between the two groups? **1039.8**
* What does the F Value indicate? **Probability is less than 0.05 therefore variance are unequal**
* Which is the correct t-value to use and why? **Satterthwaite, since variance are unequal so 2 population are unequal**

1. Do a ttest to check if the delinquency in 2 years is different between mortgage owners and renters.

* Which group has a higher delinquency rate? **MORTGAGE**
* What is the difference in average delinquency rate between the two groups? **0.0532**
* What does the F Value indicate? **The similarity of Variance between 2 groups**
* Which is the correct t-value to use and why? **Satterthwaite, since variance are unequal so 2 population are unequal**

1. Check for significant differences in principal, interest, and late fee between the two groups. What do you conclude?

**2 Groups are unequal or not similar in case of principal and late fees**

1. Use PROC ANOVA to test if there is a difference in mean loan amount between the three groups in home-ownership? First delete “Other” since there are only 2 observations in this group.

**p-value is small, we can conclude that there is evidence that there is a statistically significant difference in home-ownership.**

1. Create a new variable INCOME that divides all consumers into three groups based on whether they are high income, medium income or low income.

* How will you decide the cutoff?

**Using percentile 0-25 th percentile as low and 25-75 as medium and 75-100 as high income.**

* Use PROC UNIVARIATE to get more information on the distribution of the variable annual\_inc

Please refer SAS code

* What is the median income? **90000**
* What is the 25th percentile income? **70000**
* What is the 75th percentile income? **120000**
* Now use 25th percentile income to create INCOME that takes the value 1
* Use the 75th percentile to create INCOME that takes the value 3
* Anyone in between can be coded as income 2.

**PLEASE SEE CODE**

1. Check whether there is a relationship between income levels and home\_ownership using Chi-square test.

* What do you conclude?

**There is an association between income levels and home ownership.**

* Which income group are more likely to be renters? **MEDIUM**
* Which income group are more likely to be Mortgage owners? **MEDIUM**

1. Check if there is a relationship between home\_ownership and empl\_length using chi-square test. **yes**
2. Run a regression model to explain the loan amount. Use as independent variables annual\_inc, interest rate, emp\_length, home\_ownership.

* What is the meaning of the F-value? **Model is significant**
* Which coefficients are significant? **ALL**
* What is the meaning of the coefficient of annual\_inc? Is it significant?

**1 unit ($) increase in annual income will lead to 0.02431 times increase in loan amount, It is significant**

* What is the meaning of the coefficient of MORTGAGE?

**If home ownership type is Mortgage then the loan amount will increase by 919.52266 (reg) or 901.97989 (GLM)**

1. Run a regression model to explain the delinq\_2yrs. Use as independent variables annual\_inc, interest rate, emp\_length, home\_ownership and loan amount.

* What is the meaning of the F-value? **Model is significant**
* Which coefficients are significant?

annual\_inc, int\_rate, mortgage, loan\_amnt

* What is the meaning of the coefficient of loan amount? Is it significant?

-.0000060881 (significant)

* What is the meaning of the coefficient of “renter”?

**If the person is a renter then it doesn’t affect the delinq\_2yrs.**

1. Save your SAS code
2. Log out of your server

Submit answers to questions and SAS code online