\* 1. Read the big\_loan data \*;

LIBNAME HW1 'C:\Users\yxp120230\Desktop\BUAN 6337\HW1'; **RUN**;

**DATA** loan; SET HW1.big\_loan; **run**;

\* 3. Write SAS code to print 10 observations \*;

**PROC** **PRINT** DATA=loan (obs=**10**); **RUN**;

\* 4. What is the average loan amount? \*;

**PROC** **MEANS** DATA=loan; **RUN**;

\* 5. What is the average loan amount across different categories of home-ownership? \*;

**PROC** **SORT** DATA=loan OUT=ownership;BY home\_ownership; **RUN**;

**PROC** **MEANS**; VAR loan\_amnt; BY home\_ownership; **RUN**;

\* 7. Do a ttest to check if the average loan amount is different between mortgage owners and renters. \*;

**DATA** q7; SET loan; IF home\_ownership='MORTGAGE' or home\_ownership='RENT' THEN output; **RUN**;

**PROC** **TTEST** DATA=q7 sides=**2** alpha = **0.05** h0 = **0**;

CLASS home\_ownership; VAR loan\_amnt;

title "Two sample t-test"; **RUN**;

\* 8.Do a ttest to check if the delinquency in 2 years is different between mortgage owners and renters. \*;

**PROC** **TTEST** DATA=q7 sides=**2** alpha = **0.05** h0 = **0**;

CLASS home\_ownership; VAR delinq\_2yrs;

title "Two sample t-test Q8"; **RUN**;

\* 9. Check for significant differences in principal, interest, and late fee between the two groups. What do you conclude? \*;

**PROC** **TTEST** DATA=q7 sides=**2** alpha = **0.05** h0 = **0**;

CLASS home\_ownership; VAR total\_rec\_prncp;

title "Two sample t-test - principal"; **RUN**;

**PROC** **TTEST** DATA=q7 sides=**2** alpha = **0.05** h0 = **0**;

CLASS home\_ownership; VAR total\_rec\_int;

title "Two sample t-test - interest"; **RUN**;

**PROC** **TTEST** DATA=q7 sides=**2** alpha = **0.05** h0 = **0**;

CLASS home\_ownership; VAR total\_rec\_late\_fee;

title "Two sample t-test - late fee"; **RUN**;

\* 10. 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.

\*;

**DATA** q10; SET loan; IF home\_ownership='OTHER' THEN delete; **RUN**;

**PROC** **ANOVA** DATA=q10;

CLASS home\_ownership; MODEL loan\_amnt = home\_ownership;

MEANS home\_ownership / tukey;

TITLE 'Compare loan\_amnt across home\_ownership'; **RUN**;

\* 11. Create a new variable INCOME that divides all consumers into three groups based on whether they are high income, medium income or low income. \*;

**PROC** **UNIVARIATE** DATA=q10; VAR annual\_inc; TITLE 'Annual Income'; **RUN**;

**DATA** income; SET q10;

IF annual\_inc <= **70000** THEN income = **1**;

IF annual\_inc > **70000** AND annual\_inc <= **120000** THEN income = **2**;

IF annual\_inc > **120000** THEN income = **3**;

**RUN**;

\* 12. Check whether there is a relationship between income levels and home\_ownership using Chi-square test. \*;

**PROC** **FREQ** DATA = income; TABLES home\_ownership\*income / expected chisq; TITLE 'Chi-Square Test for Income and Ownership'; **RUN**;

\* 13. Check if there is a relationship between home\_ownership and empl\_length using chi-square test. \*;

**PROC** **FREQ** DATA = income; TABLES home\_ownership\*emp\_length / expected chisq; TITLE 'Chi-Square Test for Income and Tenure'; **RUN**;

\* 14. Run a regression model to explain the loan amount. Use as independent variables annual\_inc, interest rate, emp\_length, home\_ownership. \*;

**DATA** q14; SET loan;

KEEP loan\_amnt annual\_inc int\_rate emp\_length home\_ownership;

IF home\_ownership='OTHER' THEN DELETE;

**RUN**;

**PROC** **GLM** DATA = q14;

CLASS emp\_length home\_ownership; /\* generates dummy variables internally \*/

MODEL loan\_amnt = annual\_inc int\_rate emp\_length home\_ownership / SOLUTION;

**RUN**;

/\*

proc glmmod data=q14 outdesign=GLMDesign outparm=GLMParm;

class emp\_length home\_ownership;

model loan\_amnt = annual\_inc int\_rate emp\_length home\_ownership;

run;

proc reg data=GLMDesign;

DummyVars: model loan\_amnt = COL2-COL18;

\*/

\* 15. Run a regression model to explain the delinq\_2yrs. Use as independent variables annual\_inc, interest rate, emp\_length, home\_ownership and loan amount. \*;

**DATA** q15; SET loan;

KEEP delinq\_2yrs loan\_amnt annual\_inc int\_rate emp\_length home\_ownership;

IF home\_ownership='OTHER' THEN DELETE;

**RUN**;

**PROC** **GLM** DATA = q15;

CLASS emp\_length home\_ownership; /\* generates dummy variables internally \*/

MODEL delinq\_2yrs = annual\_inc int\_rate emp\_length home\_ownership loan\_amnt / solution;

**RUN**;

/\* Poisson Regression

proc freq data=q15;

tables delinq\_2yrs / plots=freqplot;

run;

PROC COUNTREG DATA = q15;

CLASS emp\_length home\_ownership;

MODEL delinq\_2yrs = annual\_inc int\_rate emp\_length home\_ownership loan\_amnt / dist=poisson;

RUN;

\*/