ASSIGNMENT

ASS02

DATA PREPARATION – STA6714

DONE BY:

SUSHMITHA MANI

**ASSIGNMENT –ASS02**

**DATA PREPARATION**

**NAME: SUSHMITHA MANI**

**UCF ID: 5016977**

***Problem 1(A).*** *Calculate the summary statistics including mean, median, standard deviation, skewness, and the percentage of missing values of the following numerical variables: LOAN, MORTDUE, VALUE, YOJ, CLAGE, and DEBTINC.*

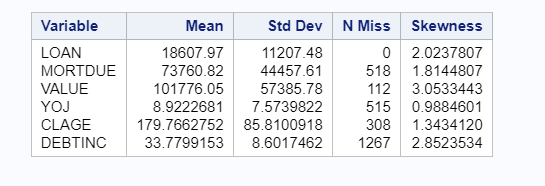
**Answer 1 (A):**

**CODE:**

proc means data=WORK.IMPORT2 chartype mean std nmiss vardef=df skewness;

var LOAN MORTDUE VALUE YOJ CLAGE DEBTINC;

run;



***Problem 1(B).*** *Calculate the summary statistics including mean, median, standard deviation, skewness and the percentage of missing values of the following numerical variables when BAD = 1: LOAN, MORTDUE, VALUE, YOJ, CLAGE, and DEBTINC.*

**ANSWER 1(B)**

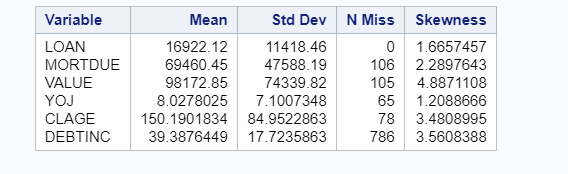
**The following code is executed when BAD=1:**

proc means data=WORK.IMPORT2 chartype mean std nmiss vardef=df skewness;

var LOAN MORTDUE VALUE YOJ CLAGE DEBTINC;

where Bad=1;

run;



***Problem 1(C).*** *Calculate the summary statistics including mean, median, standard deviation, skewness and the percentage of missing values of the following numerical variables when BAD = 0: LOAN, MORTDUE, VALUE, YOJ, CLAGE, and DEBTINC.*

**ANSWER 1 (C)**

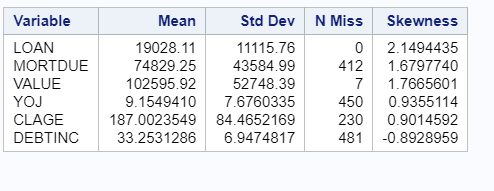
**The following code is executed when BAD=0:**

proc means data=WORK.IMPORT2 chartype mean std nmiss vardef=df skewness;

var LOAN MORTDUE VALUE YOJ CLAGE DEBTINC;

where Bad=0;

run;



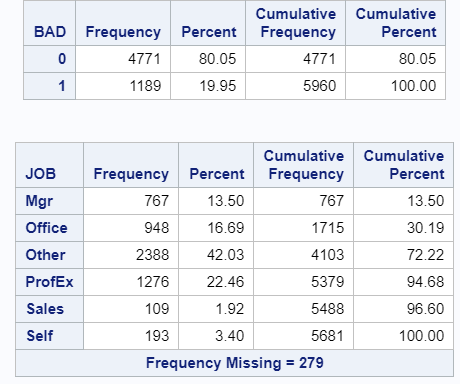
***Problem 2(A).*** Produce the frequency table of the following categorical variables: “BAD”, “JOB”, “DELINQ”, “MINQ” and “CLN”. (Note: including missing values)

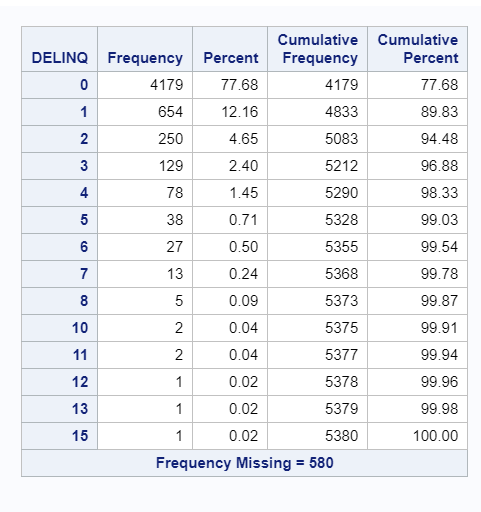
**ANSWER 2(A)**

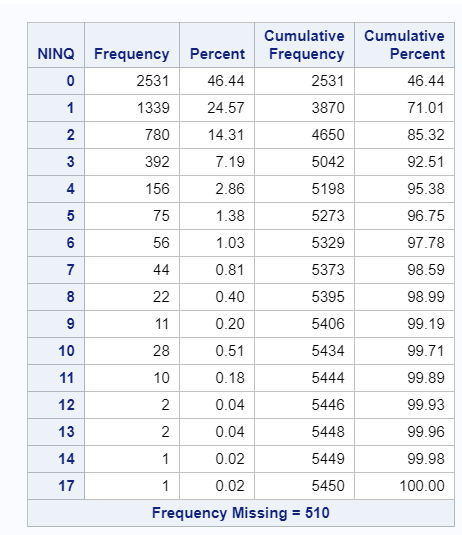
proc freq data=work.import2;

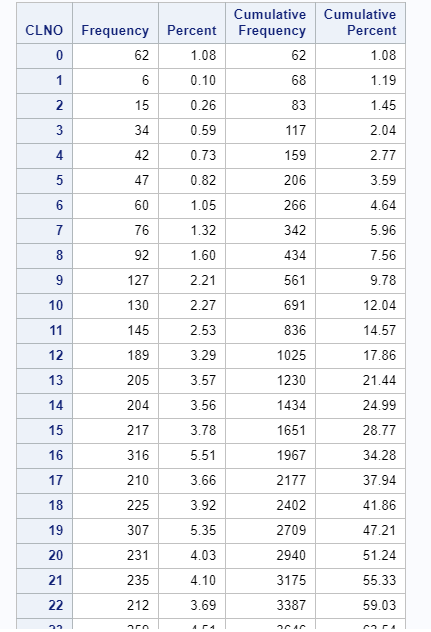
Tables BAD JOB DELINQ NINQ CLNO;

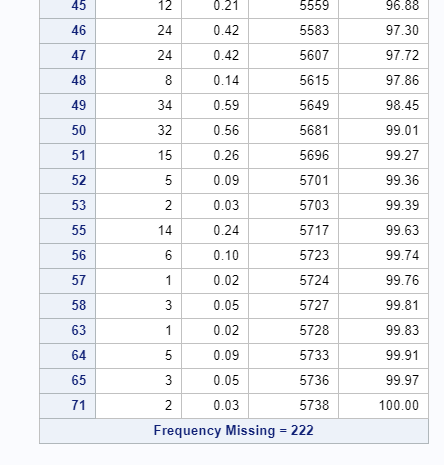
run;











***Problem 2(B).*** *Produce the two-way contingency table of the variable “BAD” with each of the following variables: “JOB”, “DELINQ”, “MINQ” and “CLN”.*

**ANSWER 2(B)**

**The following code is executed:**

proc freq data=work.import2;

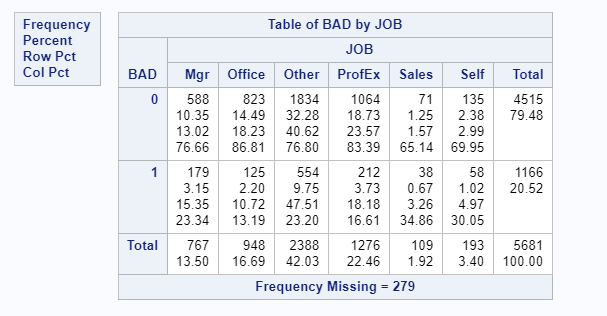
Tables BAD \* JOB;

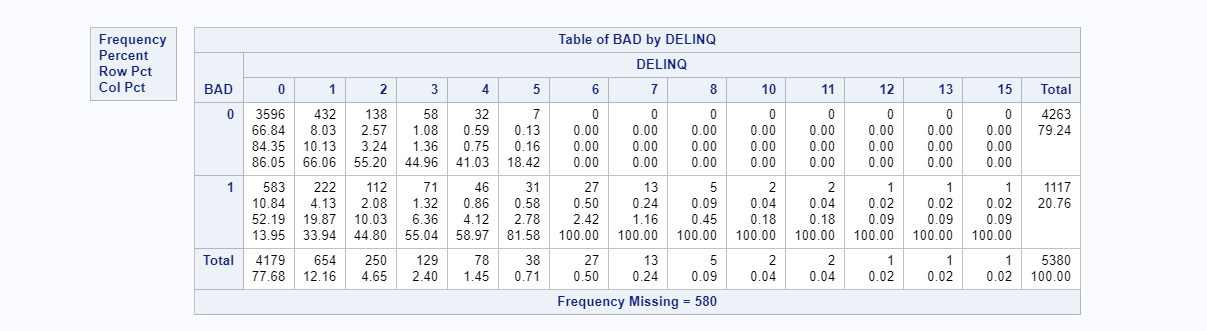
Tables BAD \* DELINQ;

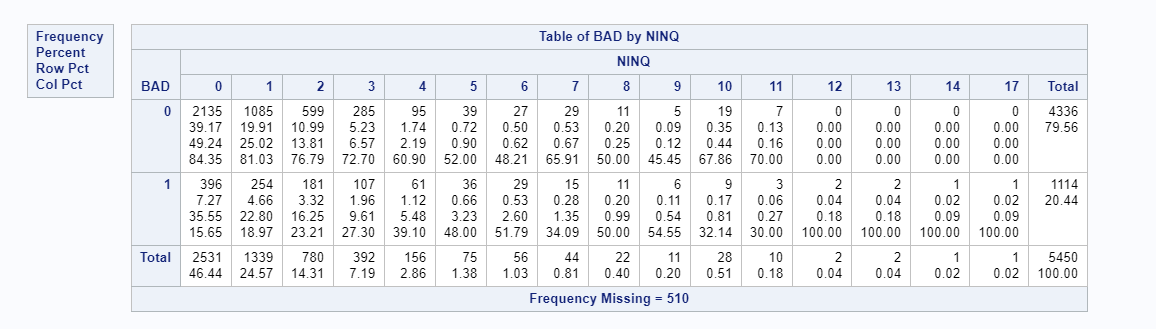
Tables BAD \* NINQ;

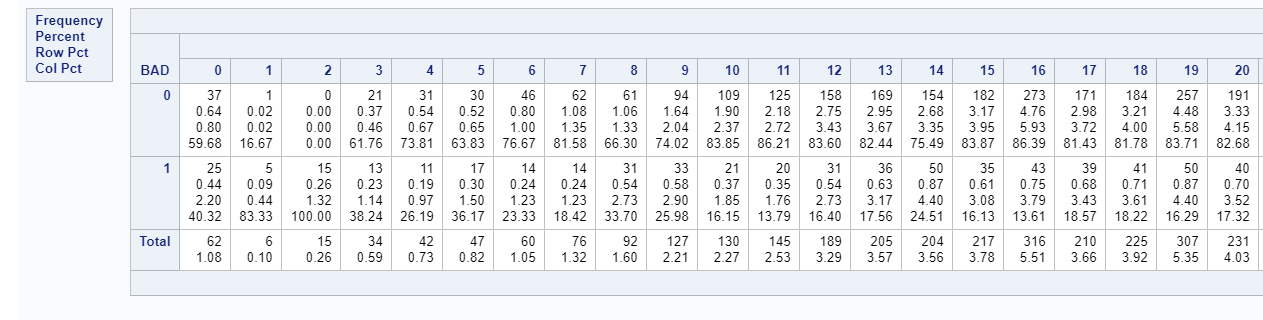
Tables BAD \* CLNO;

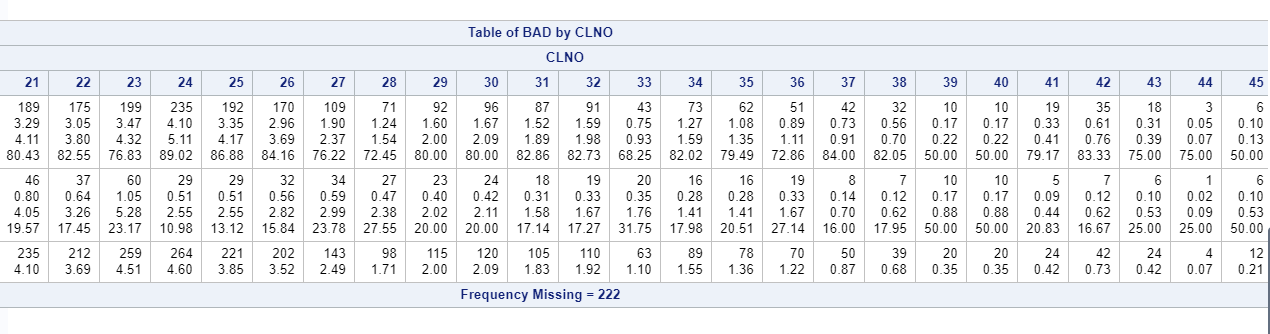
run;

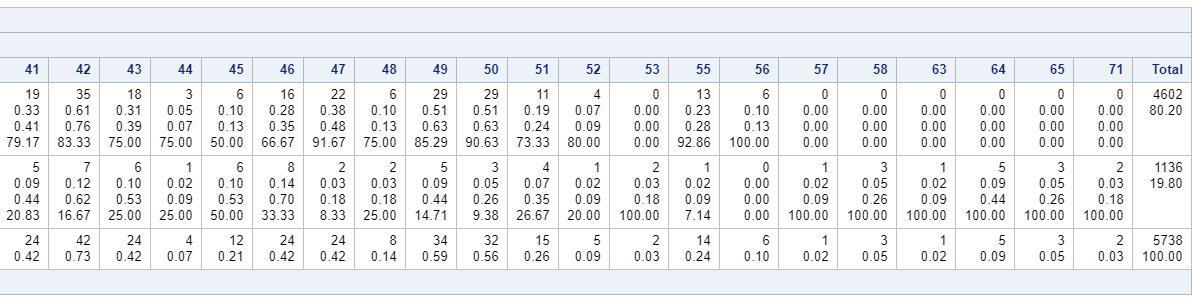












***Problem 3(A).*** *Produce a histogram of the numerical variable “LOAN”with kernel density line and normal density imposed*

**ANSWER 3(A):**

proc sgplot data=WORK.IMPORT2;

histogram LOAN /;

density LOAN;

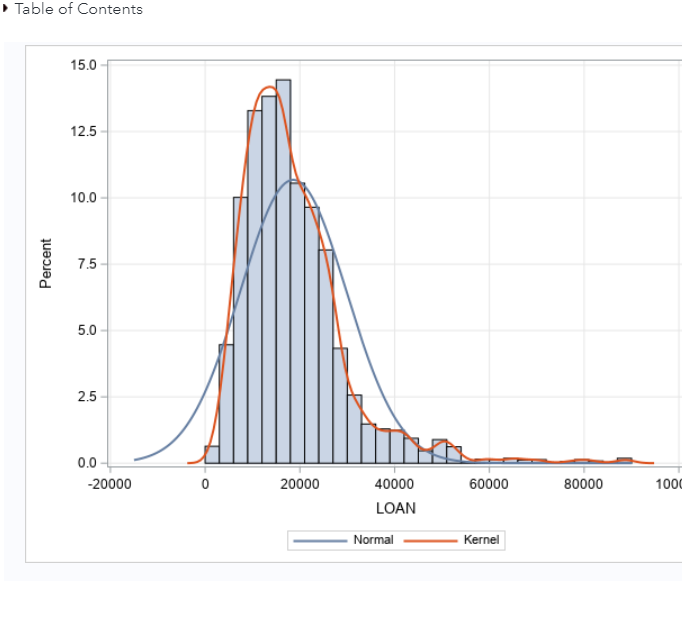
density LOAN / type=Kernel;

xaxis grid;

yaxis grid;

run;

ods graphics / reset;



***Problem 3(B).*** Produce a histogram of the numerical variable “MORTDUE” with kernel density line and normal density imposed.

**ANSWER 3(B)**

proc sgplot data=WORK.IMPORT2;

histogram MORTDUE /;

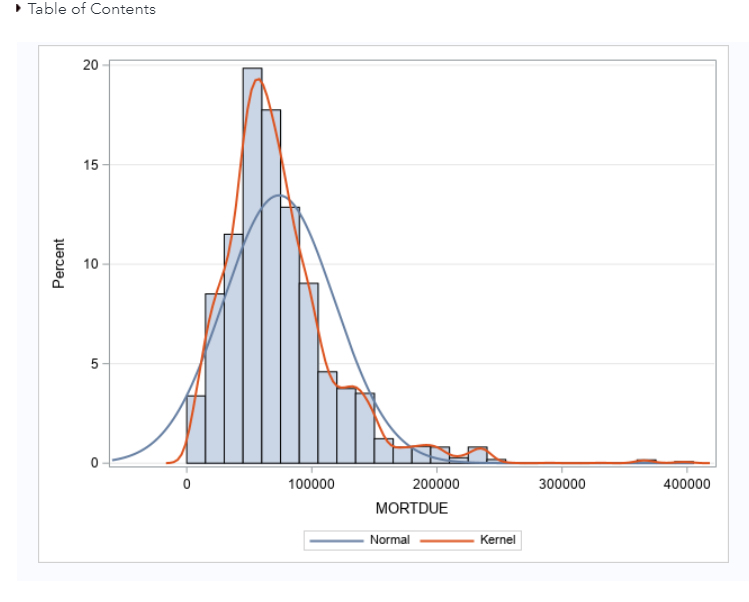
density MORTDUE;

density MORTDUE / type=Kernel;

yaxis grid;

run;

ods graphics / reset;



***Problem 3 (C).*** *Produce a histogram of the numerical variable “VALUE” with kernel density line and normal density imposed.*

**Answer 3 (C):**

proc sgplot data=WORK.IMPORT2;

histogram VALUE /;

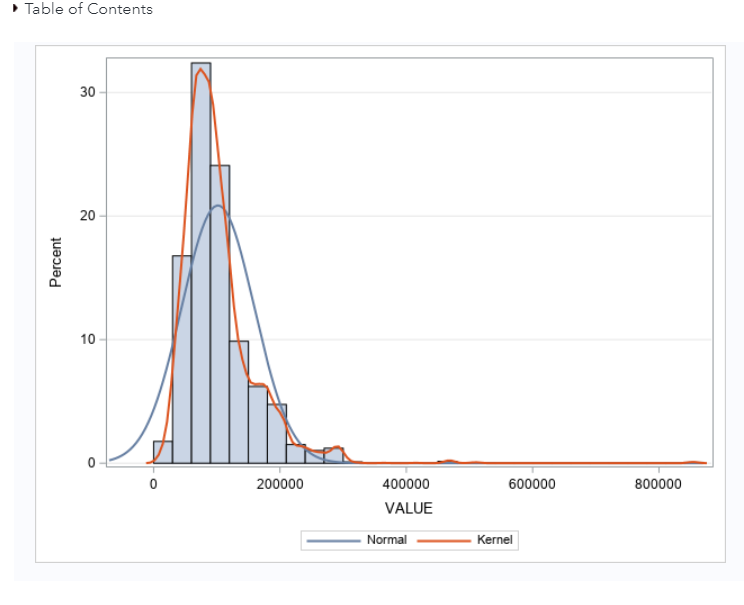
density VALUE;

density VALUE / type=Kernel;

yaxis grid;

run;

ods graphics / reset;



***Problem 3(D).*** *Produce a histogram of the numerical variable “CLAGE” with kernel density line and normal density imposed.*

**ANSWER 3(D)**

proc sgplot data=WORK.IMPORT2;

histogram CLAGE /;

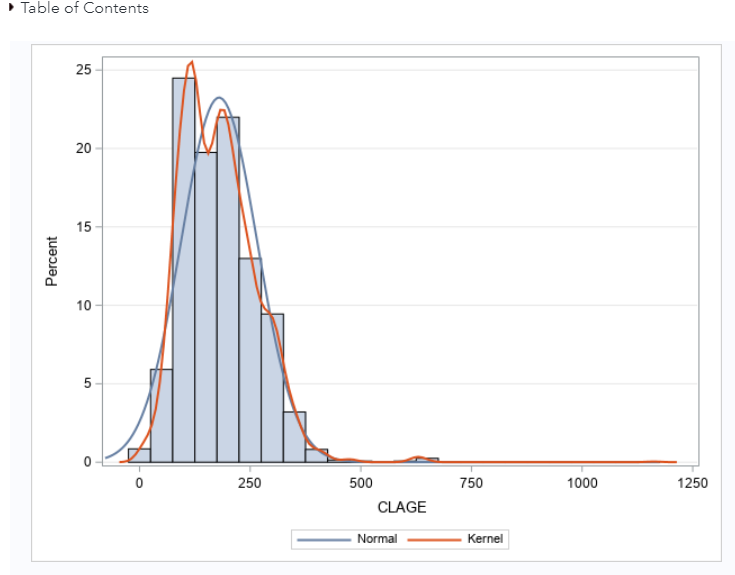
density CLAGE;

density CLAGE / type=Kernel;

yaxis grid;

run;

ods graphics / reset;



***Problem 3(E).*** *Summarized your findings in a table that including distribution property for each variable examined by you.*

**ANSWER 3(E):**

**The code:**

ods noproctitle;

ods graphics / imagemap=on;

proc means data=WORK.IMPORT2 chartype mean std median n mode vardef=df skewness

qmethod=os;

var LOAN MORTDUE VALUE CLAGE;

run;

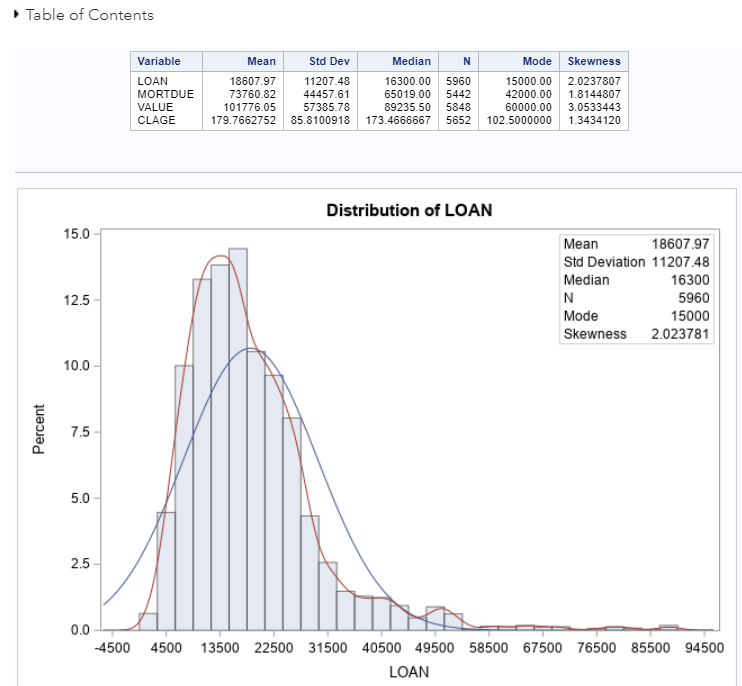
proc univariate data=WORK.IMPORT2 vardef=df noprint;

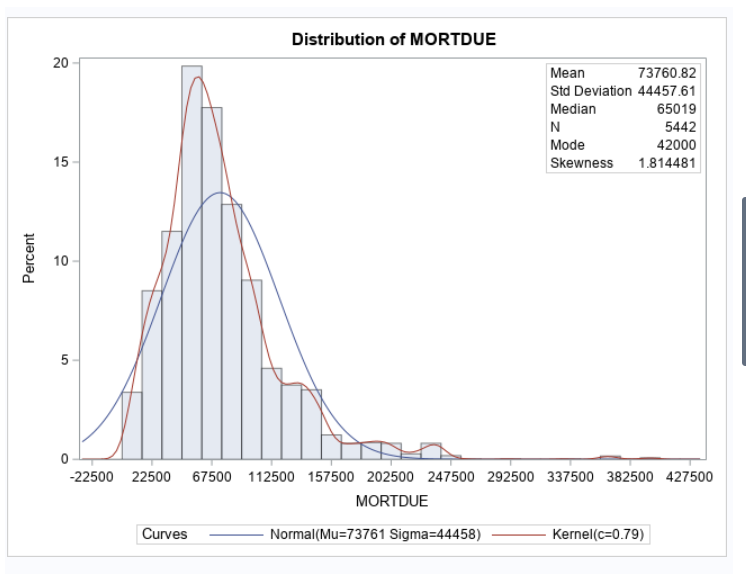
var LOAN MORTDUE VALUE CLAGE;

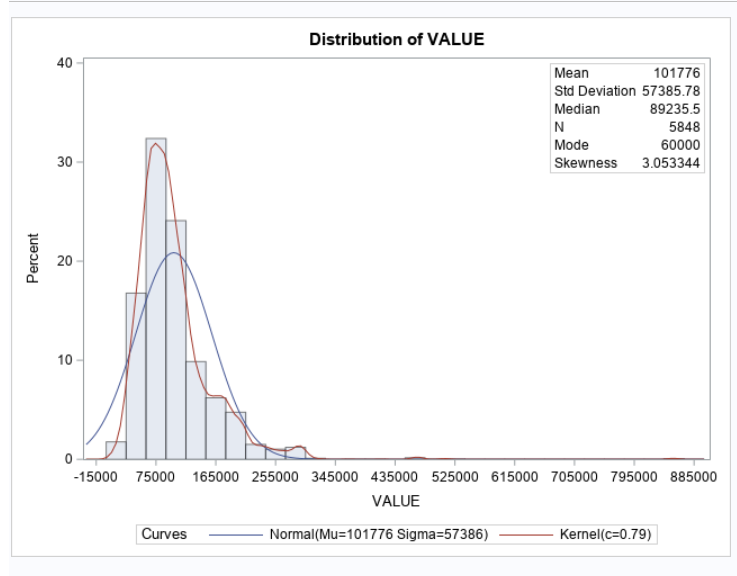
histogram LOAN MORTDUE VALUE CLAGE / normal(noprint) kernel;

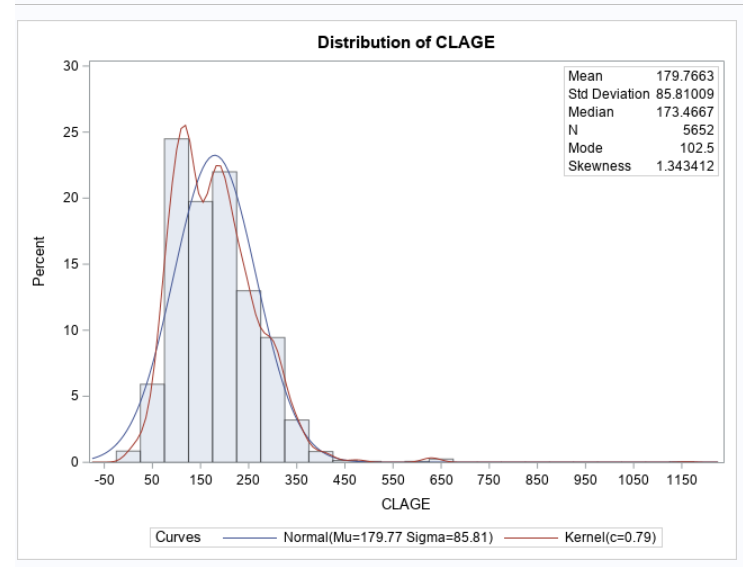
inset mean std median n mode skewness / position=ne;

run;









***Problem 4(A).*** *A scatter plots of LOAN and MORTDUE with a linear regression line imposed*.

**ANSWER 4 (A)**

proc sgplot data=WORK.IMPORT2;

reg x=LOAN y=MORTDUE / nomarkers;

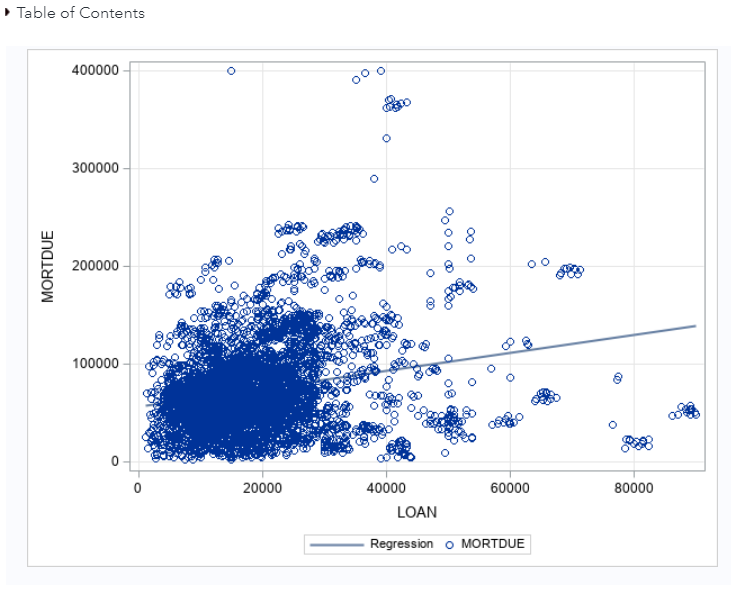
scatter x=LOAN y=MORTDUE /;

xaxis grid;

yaxis grid;

run;

ods graphics / reset;



***Problem 4(B).*** *A scatter plots of LOAN and VALUE with a linear regression line imposed.*

**ANSWER 4 (B)**

proc sgplot data=WORK.IMPORT2;

reg x=LOAN y=VALUE / nomarkers;

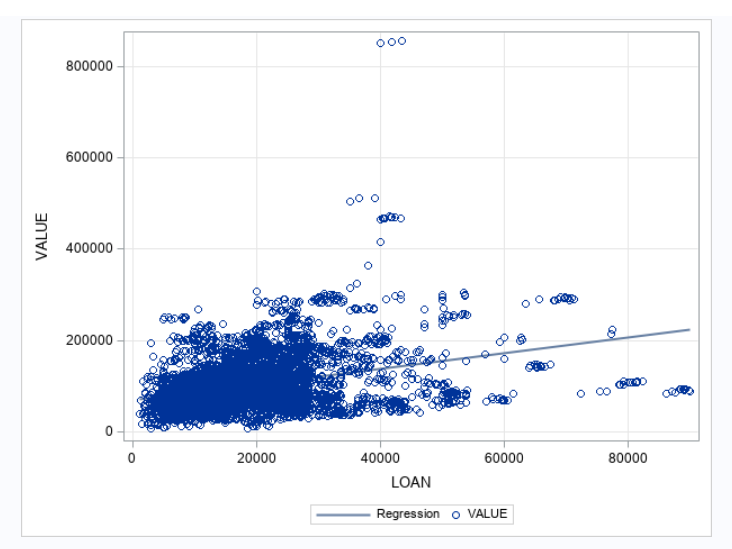
scatter x=LOAN y=VALUE /;

xaxis grid;

yaxis grid;

run;

ods graphics / reset;



***Problem 4(c):*** *A scatter plots of LOAN and YOJ with a linear regression line imposed*

**ANSWER 4(C):**

proc sgplot data=WORK.IMPORT2;

reg x=LOAN y=YOJ / nomarkers;

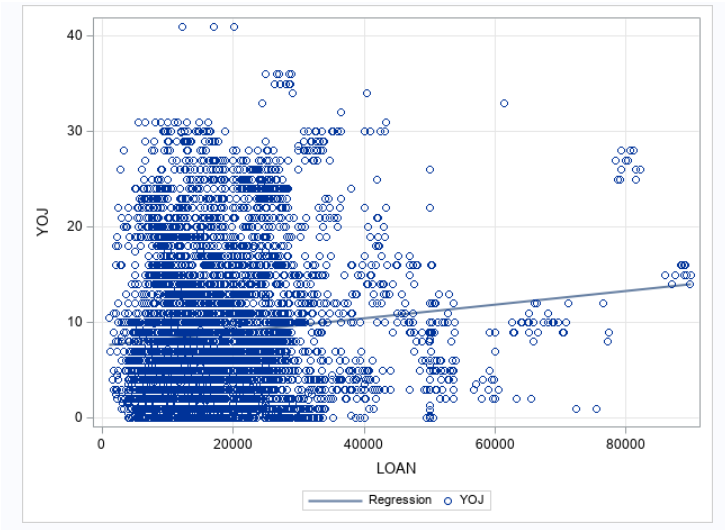
scatter x=LOAN y=YOJ /;

xaxis grid;

yaxis grid;

run;

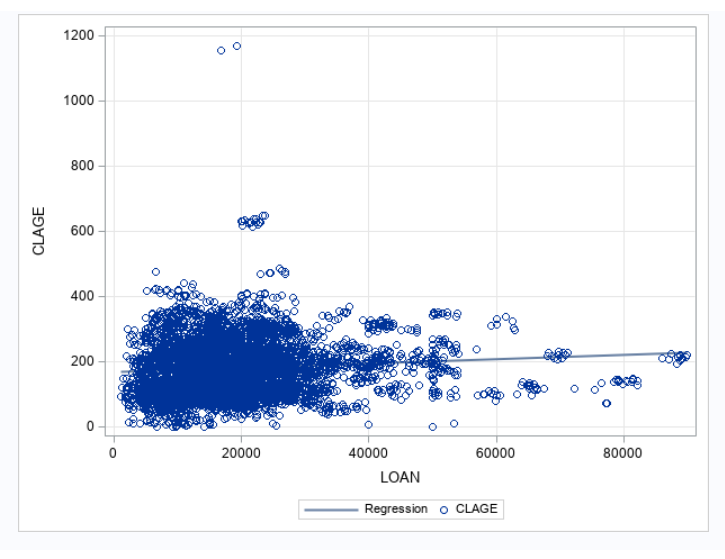
ods graphics / reset;



***PROBLEM 4(D):*** *A scatter plots of LOAN and CLAGE with a linear regression line imposed*

**ANSWERS 4 (D):**

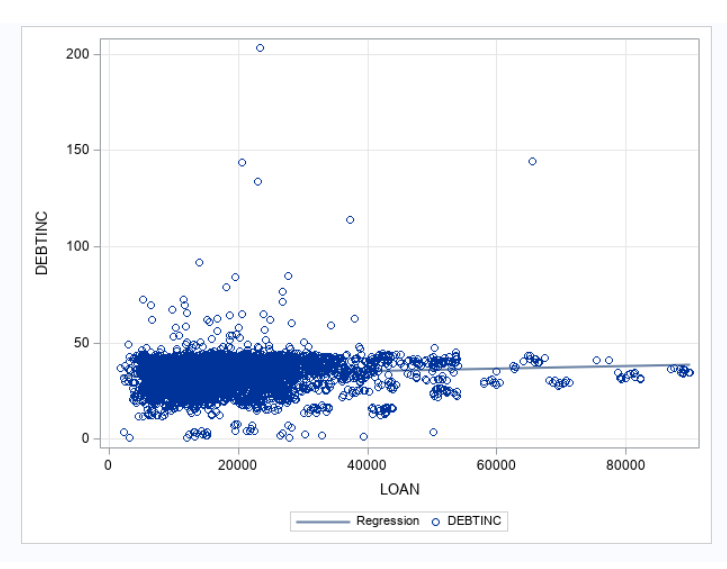
proc sgplot data=WORK.IMPORT2;  
 reg x=LOAN y=CLAGE / nomarkers;  
 scatter x=LOAN y=CLAGE /;  
 xaxis grid;  
 yaxis grid;  
run;  
  
ods graphics / reset;



***Problem 4(E):*** A scatter plots of LOAN and DEBTINC with a linear regression line imposed

**Answer 4(E):**

ods graphics / reset width=6.4in height=4.8in imagemap;  
  
proc sgplot data=WORK.IMPORT2;  
 reg x=LOAN y=DEBTINC / nomarkers;  
 scatter x=LOAN y=DEBTINC /;  
 xaxis grid;  
 yaxis grid;  
run;  
  
ods graphics / reset;



***Problem 5****: Produce a logit plot for the numerical predictor “LOAN” .*

**ANSWER 5:**

**The following code is executed when BAD =0:**

ods noproctitle;

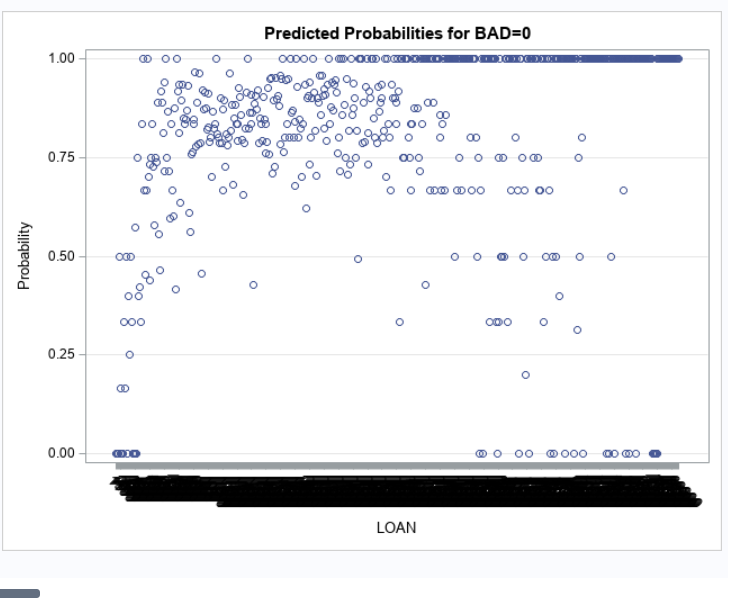
ods graphics / imagemap=on;

proc logistic data=WORK.IMPORT2 plots=(effect);

class LOAN / param=glm;

model BAD(event='0')=LOAN / link=logit technique=fisher;

run;



**The following code is executed when BAD=1:**

ods noproctitle;

ods graphics / imagemap=on;

proc logistic data=WORK.IMPORT2 plots=(effect);

class LOAN / param=glm;

model BAD(event='1')=LOAN / link=logit technique=fisher;

run;

