**Logistic Regression Graded Assignment by Avijit Mallick(jig154684)**

**Path: D:\Users\Jig15684\Practice\Avijit Mallick Graded assignment**

LIBNAME creditr "Y:\Practice\Graded assignment";

**Run**;

/\* Import Data \*/

**PROC** **IMPORT** DATAFILE = "Z:\Assignments\Graded Assignment\Topic 10 - Regression Models\Credit.csv"

OUT = creditr.credit1

DBMS = CSV

REPLACE;

GUESSINGROWS=150000;

**RUN**;

/\* exploring the data \*/

**proc** **contents** data = creditr.credit1;

**run**;

**Output**

|  |
| --- |
| The WPS System |

The CONTENTS Procedure

| Data Set Name | CREDIT1 |
| --- | --- |
| Member Type | DATA |
| Engine | WPD |
| Created | 16OCT2017:11:02:37 |
| Last Modified | 16OCT2017:11:02:37 |
| Observations | 150002 |
| Variables | 17 |
| Indexes | 0 |
| Observation Length | 143 |
| Deleted Observations | 0 |
| Data Set Type |  |
| Label |  |
| Compressed | NO |
| Sorted | NO |
| Data Representation | Little endian, IEEE Windows |
| Encoding | wlatin1 Windows-1252 Western |

| **Engine/Host Dependent Information** | |
| --- | --- |
| Data Set Page Size | 4096 |
| Number of Data Set Pages | 5359 |
| First Data Page | 1 |
| Max Obs Per Page | 28 |
| Obs In First Data Page | 28 |
| Data Set Diagnostic Code | 0009 |
| File Name | Y:\Practice\Graded assignment\CREDIT1.wpd |
| WPD Engine Version | 3 |
| Large Data Set Support | no |

| **Alphabetic List of Variables and Attributes** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Number** | **Variable** | **Type** | **Len** | **Pos** | **Format** | **Informat** | **Label** |
| 11 | DebtRatio | Num | 8 | 32 | BEST12. | BEST32. | DebtRatio |
| 9 | Education | Char | 15 | 116 | $15. | $15. | Education |
| 4 | Gender | Char | 7 | 72 | $7. | $7. | Gender |
| 6 | MonthlyIncome | Char | 9 | 87 | $9. | $9. | MonthlyIncome |
| 12 | MonthlyIncome1 | Char | 9 | 131 | $9. | $9. | MonthlyIncome |
| 1 | NPA\_Status | Num | 8 | 0 | BEST12. | BEST32. | NPA Status |
| 17 | NumberOfDependents | Char | 3 | 140 | $3. | $3. | NumberOfDependents |
| 13 | NumberOfOpenCreditLinesAndLoans | Num | 8 | 40 | BEST12. | BEST32. | NumberOfOpenCreditLinesAndLoans |
| 10 | NumberOfTime30\_59DaysPastDueNotW | Num | 8 | 24 | BEST12. | BEST32. | NumberOfTime30-59DaysPastDueNotWorse |
| 16 | NumberOfTime60\_89DaysPastDueNotW | Num | 8 | 64 | BEST12. | BEST32. | NumberOfTime60-89DaysPastDueNotWorse |
| 14 | NumberOfTimes90DaysLate | Num | 8 | 48 | BEST12. | BEST32. | NumberOfTimes90DaysLate |
| 15 | NumberRealEstateLoansOrLines | Num | 8 | 56 | BEST12. | BEST32. | NumberRealEstateLoansOrLines |
| 8 | Occupation | Char | 12 | 104 | $12. | $12. | Occupation |
| 5 | Region | Char | 8 | 79 | $8. | $8. | Region |
| 7 | Rented\_OwnHouse | Char | 8 | 96 | $8. | $8. | Rented\_OwnHouse |
| 2 | RevolvingUtilizationOfUnsecuredL | Num | 8 | 8 | BEST12. | BEST32. | RevolvingUtilizationOfUnsecuredLines |
| 3 | age | Num | 8 | 16 | BEST12. | BEST32. | age |

/\*monthly income showing 2 times\*/

**data** creditr.credit1(drop = monthlyincome1);

set creditr.credit1;

**run**;

/\* proc means of the data \*/

**proc** **means** data = creditr.credit1 mean stddev n nmiss min max;

**run**;

/\* convert character variable to numeric variable of the data \*/

**data** creditr.credit2;

set creditr.credit1;

if gender = "Male" then gender\_M =1;

else gender\_M = 0;

if gender = "Male" then gender=1;

else if gender = "Female" then gender = 0;

if Education = "Matric" then EMatric = 1;

else EMatric = 0;

if Education = "Graduate" then EGraduate = 1;

else EGraduate = 0;

if Education = "PhD" then EPhD = 1;

else EPhD = 0;

if Education = "Professional" then EProfessional = 1;

else Eprofessional = 0;

if Education = "Post-Grad" then Epostgrad = 1;

else Epostgrad = 0;

if Occupation = "Self\_Emp" then occself = 1;

else occself = 0;

if Occupation = "Officer1" then occoffic1 = 1;

else occoffic1 = 0;

if Occupation = "Officer2" then occoffic2 = 1;

else occoffic2 = 0;

if Occupation = "Officer3" then occoffic3 = 1;

else occoffic3 = 0;

if Occupation = "Non-officer" then occnonoffic = 1;

else occnonoffic = 0;

if Region = "South" then Regsouth = 1;

else Regsouth = 0;

if Region = "Central" then Regcent = 1;

else Regcent = 0;

if Region = "North" then Regnorth = 1;

else Regnorth = 0;

if Region = "East" then Regeast = 1;

else Regeast = 0;

if Region = "West" then Regwest = 1;

else Regwest = 0;

if Rented\_OwnHouse = "Ownhouse" then OwnHouse = 1;

else OwnHouse = 0;

monthlyincome1 = monthlyincome\*1;

drop monthlyincome Education Occupation Region Gender rented\_Ownhouse;

Dependent = numberofdependents\*1;

drop numberofdependents;

**run**;

/\* again check for missing data of the data \*/

/\* Missing NPA to be delete \*/

**data** creditr.credit2;

set creditr.credit2;

if NPA\_status = . then delete;

**run**;

**proc** **means** data = creditr.credit2 mean stddev n nmiss min max;

**run**;

/\* checking for monthlyincome1 outlier\*/

**data** creditr.credit3;

set creditr.credit2;;

if 0 = monthlyincome1> 1000000 then delete;

**run**;

**proc** **means** data = creditr.credit3 mean stddev n nmiss min max;

**run**;

|  |
| --- |
| The WPS System |

The MEANS Procedure

| **Summary statistics** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **Mean** | **Std Dev** | **N** | **N Miss** | **Minimum** | **Maximum** |
| NPA\_Status | NPA Status | 0.0668400 | 0.2497455 | 150000 | 0 | 0 | 1.0000000 |
| RevolvingUtilizationOfUnsecuredL | RevolvingUtilizationOfUnsecuredLines | 6.0484381 | 249.7553706 | 150000 | 0 | 0 | 50708.00 |
| age | age | 52.2952067 | 14.7718659 | 150000 | 0 | 0 | 109.0000000 |
| NumberOfTime30\_59DaysPastDueNotW | NumberOfTime30-59DaysPastDueNotWorse | 0.4210333 | 4.1927813 | 150000 | 0 | 0 | 98.0000000 |
| DebtRatio | DebtRatio | 353.0050758 | 2037.82 | 150000 | 0 | 0 | 329664.00 |
| NumberOfOpenCreditLinesAndLoans | NumberOfOpenCreditLinesAndLoans | 8.4527600 | 5.1459510 | 150000 | 0 | 0 | 58.0000000 |
| NumberOfTimes90DaysLate | NumberOfTimes90DaysLate | 0.2659733 | 4.1693038 | 150000 | 0 | 0 | 98.0000000 |
| NumberRealEstateLoansOrLines | NumberRealEstateLoansOrLines | 1.0182400 | 1.1297710 | 150000 | 0 | 0 | 54.0000000 |
| NumberOfTime60\_89DaysPastDueNotW | NumberOfTime60-89DaysPastDueNotWorse | 0.2403867 | 4.1551794 | 150000 | 0 | 0 | 98.0000000 |
| gender\_M |  | 0.6153600 | 0.4865117 | 150000 | 0 | 0 | 1.0000000 |
| EMatric |  | 0.1054000 | 0.3070692 | 150000 | 0 | 0 | 1.0000000 |
| EGraduate |  | 0.2650333 | 0.4413524 | 150000 | 0 | 0 | 1.0000000 |
| EPhD |  | 0.0420067 | 0.2006050 | 150000 | 0 | 0 | 1.0000000 |
| EProfessional |  | 0.3394667 | 0.4735299 | 150000 | 0 | 0 | 1.0000000 |
| Epostgrad |  | 0.2480933 | 0.4319077 | 150000 | 0 | 0 | 1.0000000 |
| occself |  | 0.4274400 | 0.4947087 | 150000 | 0 | 0 | 1.0000000 |
| occoffic1 |  | 0.1010933 | 0.3014533 | 150000 | 0 | 0 | 1.0000000 |
| occoffic2 |  | 0.0888867 | 0.2845810 | 150000 | 0 | 0 | 1.0000000 |
| occoffic3 |  | 0.1084933 | 0.3110035 | 150000 | 0 | 0 | 1.0000000 |
| occnonoffic |  | 0.2740867 | 0.4460544 | 150000 | 0 | 0 | 1.0000000 |
| Regsouth |  | 0.1566333 | 0.3634559 | 150000 | 0 | 0 | 1.0000000 |
| Regcent |  | 0.2930400 | 0.4551581 | 150000 | 0 | 0 | 1.0000000 |
| Regnorth |  | 0.2273267 | 0.4191067 | 150000 | 0 | 0 | 1.0000000 |
| Regeast |  | 0.1370067 | 0.3438555 | 150000 | 0 | 0 | 1.0000000 |
| Regwest |  | 0.1859933 | 0.3891026 | 150000 | 0 | 0 | 1.0000000 |
| OwnHouse |  | 0.5730200 | 0.4946410 | 150000 | 0 | 0 | 1.0000000 |
| monthlyincome1 |  | 6670.22 | 14384.67 | 120269 | 29731 | 0 | 3008750.00 |
| Dependent |  | 0.7572223 | 1.1150861 | 146076 | 3924 | 0 | 20.0000000 |

/\* 29731 miss value for monthlyincome1, replace the missing value \*/

**proc** **means** mean data = creditr.credit3;

var monthlyincome1 dependent;

output out = creditr.credit4 (drop = \_type\_ \_freq\_) *mean*(monthlyincome1) = meanmoninc *mean*(dependent) = meandepend;

**run**;

/\* Since Monthly income is fixed the Debt Ratio needs to be fixed and number of dependents is also fixed \*/

**data** creditr.credit5;

set creditr.credit4;

common = 1;

**run**;

**data** creditr.credit6;

set creditr.credit3;

common = 1;

**run**;

**data** creditr.credit7;

merge creditr.credit5 creditr.credit6;

by common;

if monthlyincome1 = . then Debtratio = *round*((debtratio/meanmoninc),0.01);

if monthlyincome1 = . then monthlyincome1 = meanmoninc;

if 0 < meandepend <1 then meandepend = 1;

if 1< meandepend <2 then meandepend = 2;

if dependent = . then dependent = meandepend;

drop meanmoninc meandepend;

**run**;

|  |
| --- |
| The WPS System |

The MEANS Procedure

| **Summary statistics** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **Mean** | **Std Dev** | **N** | **N Miss** | **Minimum** | **Maximum** |
| common |  | 1.0000000 | 0 | 150000 | 0 | 1.0000000 | 1.0000000 |
| NPA\_Status | NPA Status | 0.0668400 | 0.2497455 | 150000 | 0 | 0 | 1.0000000 |
| RevolvingUtilizationOfUnsecuredL | RevolvingUtilizationOfUnsecuredLines | 6.0484381 | 249.7553706 | 150000 | 0 | 0 | 50708.00 |
| age | age | 52.2952067 | 14.7718659 | 150000 | 0 | 0 | 109.0000000 |
| NumberOfTime30\_59DaysPastDueNotW | NumberOfTime30-59DaysPastDueNotWorse | 0.4210333 | 4.1927813 | 150000 | 0 | 0 | 98.0000000 |
| DebtRatio | DebtRatio | 21.3764251 | 380.2066313 | 150000 | 0 | 0 | 61106.50 |
| NumberOfOpenCreditLinesAndLoans | NumberOfOpenCreditLinesAndLoans | 8.4527600 | 5.1459510 | 150000 | 0 | 0 | 58.0000000 |
| NumberOfTimes90DaysLate | NumberOfTimes90DaysLate | 0.2659733 | 4.1693038 | 150000 | 0 | 0 | 98.0000000 |
| NumberRealEstateLoansOrLines | NumberRealEstateLoansOrLines | 1.0182400 | 1.1297710 | 150000 | 0 | 0 | 54.0000000 |
| NumberOfTime60\_89DaysPastDueNotW | NumberOfTime60-89DaysPastDueNotWorse | 0.2403867 | 4.1551794 | 150000 | 0 | 0 | 98.0000000 |
| gender\_M |  | 0.6153600 | 0.4865117 | 150000 | 0 | 0 | 1.0000000 |
| EMatric |  | 0.1054000 | 0.3070692 | 150000 | 0 | 0 | 1.0000000 |
| EGraduate |  | 0.2650333 | 0.4413524 | 150000 | 0 | 0 | 1.0000000 |
| EPhD |  | 0.0420067 | 0.2006050 | 150000 | 0 | 0 | 1.0000000 |
| EProfessional |  | 0.3394667 | 0.4735299 | 150000 | 0 | 0 | 1.0000000 |
| Epostgrad |  | 0.2480933 | 0.4319077 | 150000 | 0 | 0 | 1.0000000 |
| occself |  | 0.4274400 | 0.4947087 | 150000 | 0 | 0 | 1.0000000 |
| occoffic1 |  | 0.1010933 | 0.3014533 | 150000 | 0 | 0 | 1.0000000 |
| occoffic2 |  | 0.0888867 | 0.2845810 | 150000 | 0 | 0 | 1.0000000 |
| occoffic3 |  | 0.1084933 | 0.3110035 | 150000 | 0 | 0 | 1.0000000 |
| occnonoffic |  | 0.2740867 | 0.4460544 | 150000 | 0 | 0 | 1.0000000 |
| Regsouth |  | 0.1566333 | 0.3634559 | 150000 | 0 | 0 | 1.0000000 |
| Regcent |  | 0.2930400 | 0.4551581 | 150000 | 0 | 0 | 1.0000000 |
| Regnorth |  | 0.2273267 | 0.4191067 | 150000 | 0 | 0 | 1.0000000 |
| Regeast |  | 0.1370067 | 0.3438555 | 150000 | 0 | 0 | 1.0000000 |
| Regwest |  | 0.1859933 | 0.3891026 | 150000 | 0 | 0 | 1.0000000 |
| OwnHouse |  | 0.5730200 | 0.4946410 | 150000 | 0 | 0 | 1.0000000 |
| monthlyincome1 |  | 6670.22 | 12880.45 | 150000 | 0 | 0 | 3008750.00 |
| Dependent |  | 0.7635733 | 1.1010861 | 150000 | 0 | 0 | 20.0000000 |

/\* proc mean suggest age min = 0 need to check with univariant\*/

**data** creditr.credit7;

set creditr.credit7;

if age = 0 then delete;

**run**;

/\*Looking at final time for data\*/

**proc** **contents** data = creditr.credit7;

**run**;

**proc** **univariate** data = creditr.credit7;

**run**;

**proc** **means** data = creditr.credit7 mean stddev n nmiss min max;

**run**;

/\* Split data set in 65:35 for Train and Validation data set \*/

**proc** **surveyselect** data = creditr.credit7

method= srs out = creditr.credit8 samprate = .065 outall;

**run**;

**data** creditr.credit8\_tr creditr.credit8\_val;

set creditr.credit8;

if selected=0 then output creditr.credit8\_tr;

else if selected = 1 then output creditr.credit8\_val;

**run**;

**proc** **freq** data = creditr.credit8\_val;

table NPA\_status;

**run**;

|  |
| --- |
| The WPS System |

The FREQ Procedure

| **NPA Status** | | | | |
| --- | --- | --- | --- | --- |
| **NPA\_Status** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 0 | 9070 | 93.03 | 9070 | 93.03 |
| 1 | 680 | 6.97 | 9750 | 100.00 |

**proc** **freq** data = creditr.credit8\_tr;

table NPA\_status;

**run**;

|  |
| --- |
| The WPS System |

The FREQ Procedure

| **NPA Status** | | | | |
| --- | --- | --- | --- | --- |
| **NPA\_Status** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| 0 | 130904 | 93.34 | 130904 | 93.34 |
| 1 | 9346 | 6.66 | 140250 | 100.00 |

/\*multicollinearity\*/

**proc** **corr** data = creditr.credit8\_tr;

var RevolvingUtilizationOfUnsecuredL age gender\_M NumberOfTime30\_59DaysPastDueNotW DebtRatio

NumberOfOpenCreditLinesAndLoans NumberOfTime60\_89DaysPastDueNotW NumberOfTimes90DaysLate

NumberRealEstateLoansOrLines OwnHouse Regcent Regeast Regnorth Regsouth Regwest occnonoffic

occoffic1 occoffic2 occoffic3 occself Dependent EGraduate EMatric EPhD EProfessional Epostgrad

NPA\_Status;

**run**;

|  |
| --- |
| The WPS System |

The CORR Procedure

| 27 | Variables: | RevolvingUtilizationOfUnsecuredL age gender\_M NumberOfTime30\_59DaysPastDueNotW DebtRatio NumberOfOpenCreditLinesAndLoans NumberOfTime60\_89DaysPastDueNotW NumberOfTimes90DaysLate NumberRealEstateLoansOrLines OwnHouse Regcent Regeast Regnorth Regsouth Regwest occnonoffic occoffic1 occoffic2 occoffic3 occself Dependent EGraduate EMatric EPhD EProfessional Epostgrad NPA\_Status |
| --- | --- | --- |

| **Simple Statistics** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** | **Label** |
| RevolvingUtilizationOfUnsecuredL | 140250 | 6.18029 | 254.32726 | 866786 | 0 | 50708 | RevolvingUtilizationOfUnsecuredLines |
| age | 140250 | 52.29548 | 14.76744 | 7334441 | 0 | 109.00000 | age |
| gender\_M | 140250 | 0.61540 | 0.48650 | 86310 | 0 | 1.00000 |  |
| NumberOfTime30\_59DaysPastDueNotW | 140250 | 0.42469 | 4.23677 | 59563 | 0 | 98.00000 | NumberOfTime30-59DaysPastDueNotWorse |
| DebtRatio | 140250 | 21.52373 | 379.41068 | 3018703 | 0 | 61107 | DebtRatio |
| NumberOfOpenCreditLinesAndLoans | 140250 | 8.44873 | 5.14553 | 1184934 | 0 | 58.00000 | NumberOfOpenCreditLinesAndLoans |
| NumberOfTime60\_89DaysPastDueNotW | 140250 | 0.24409 | 4.19977 | 34234 | 0 | 98.00000 | NumberOfTime60-89DaysPastDueNotWorse |
| NumberOfTimes90DaysLate | 140250 | 0.26992 | 4.21382 | 37856 | 0 | 98.00000 | NumberOfTimes90DaysLate |
| NumberRealEstateLoansOrLines | 140250 | 1.01834 | 1.12987 | 142822 | 0 | 54.00000 | NumberRealEstateLoansOrLines |
| OwnHouse | 140250 | 0.57377 | 0.49453 | 80471 | 0 | 1.00000 |  |
| Regcent | 140250 | 0.29278 | 0.45504 | 41062 | 0 | 1.00000 |  |
| Regeast | 140250 | 0.13676 | 0.34360 | 19181 | 0 | 1.00000 |  |
| Regnorth | 140250 | 0.22785 | 0.41945 | 31956 | 0 | 1.00000 |  |
| Regsouth | 140250 | 0.15678 | 0.36359 | 21988 | 0 | 1.00000 |  |
| Regwest | 140250 | 0.18583 | 0.38897 | 26063 | 0 | 1.00000 |  |
| occnonoffic | 140250 | 0.27422 | 0.44612 | 38459 | 0 | 1.00000 |  |
| occoffic1 | 140250 | 0.10118 | 0.30157 | 14191 | 0 | 1.00000 |  |
| occoffic2 | 140250 | 0.08870 | 0.28431 | 12440 | 0 | 1.00000 |  |
| occoffic3 | 140250 | 0.10870 | 0.31126 | 15245 | 0 | 1.00000 |  |
| occself | 140250 | 0.42720 | 0.49467 | 59915 | 0 | 1.00000 |  |
| Dependent | 140250 | 0.76379 | 1.10247 | 107121 | 0 | 20.00000 |  |
| EGraduate | 140250 | 0.26531 | 0.44150 | 37210 | 0 | 1.00000 |  |
| EMatric | 140250 | 0.10544 | 0.30712 | 14788 | 0 | 1.00000 |  |
| EPhD | 140250 | 0.04200 | 0.20058 | 5890 | 0 | 1.00000 |  |
| EProfessional | 140250 | 0.33873 | 0.47328 | 47507 | 0 | 1.00000 |  |
| Epostgrad | 140250 | 0.24852 | 0.43216 | 34855 | 0 | 1.00000 |  |
| NPA\_Status | 140250 | 0.06664 | 0.24940 | 9346 | 0 | 1.00000 | NPA Status |

/\* Running the model iteration 1 without redundant varibales Tried with Age group p value are high\*/

**proc** **logistic** data = creditr.credit8\_tr descending outmodel = creditr.dmm;

model NPA\_status = /\*RevolvingUtilizationOfUnsecured\*/age gender\_M NumberOfTime30\_59DaysPastDueNotW DebtRatio

NumberOfOpenCreditLinesAndLoans NumberOfTime60\_89DaysPastDueNotW NumberOfTimes90DaysLate

NumberRealEstateLoansOrLines OwnHouse Regeast Regnorth Regsouth Regwest /\*occnonoffic\*/occoffic1

occoffic2 occoffic3 occself Dependent EGraduate EMatric EProfessional Epostgrad /\*EPhD\*/ /\*Regcent\*/;

score out = creditr.dmp;

**run**;

/\* Running the model iteration 1 without redundant varibales Tried with Age value only\*/

**proc** **logistic** data = creditr.credit8\_tr descending outmodel = creditr.dmm;

model NPA\_status =RevolvingUtilizationOfUnsecuredl age gender\_M NumberOfTime30\_59DaysPastDueNotW DebtRatio

NumberOfOpenCreditLinesAndLoans NumberOfTime60\_89DaysPastDueNotW NumberOfTimes90DaysLate

NumberRealEstateLoansOrLines OwnHouse Regeast Regnorth Regsouth Regwest /\*occnonoffic\*/occoffic1

occoffic3 occself Dependent EGraduate EMatric EProfessional Epostgrad /\*EPhD\*/ /\*Regcent\*/ /\*occoffic2\*/;

score out = creditr.dmp;

**run**;

/\* Running the model iteration 1 without redundant varibales Tried with Age value only\*/

**proc** **logistic** data = creditr.credit8\_tr descending outmodel = creditr.dmm;

model NPA\_status =RevolvingUtilizationOfUnsecuredl age gender\_M NumberOfTime30\_59DaysPastDueNotW DebtRatio

NumberOfOpenCreditLinesAndLoans NumberOfTime60\_89DaysPastDueNotW NumberOfTimes90DaysLate

NumberRealEstateLoansOrLines OwnHouse Regeast Regnorth Regsouth Regwest /\*occnonoffic\*/occoffic1

occoffic3 occself Dependent EGraduate EMatric EProfessional Epostgrad /\*EPhD\*/ /\*Regcent\*/ /\*occoffic2\*/

/ctable lackfit maxiter=100;

score out = creditr.dmp;

**run**;

/\*\*RAN MULTIPLE MODELS TO SEE IF ANY SIGNIFICANT BETTER MODEL \*\*\*//

/\* Running the model iteration 2 without insignificant RevolvingUtilizationOfUnsecuredL and debt ratio variable\*/

**proc** **logistic** data = creditr.credit8\_tr descending outmodel = creditr.dmm;

model NPA\_status = /\*RevolvingUtilizationOfUnsecuredl\*/ /\*DebtRatio\*/ age gender\_M NumberOfTime30\_59DaysPastDueNotW

NumberOfOpenCreditLinesAndLoans NumberOfTime60\_89DaysPastDueNotW NumberOfTimes90DaysLate

NumberRealEstateLoansOrLines OwnHouse Regeast Regnorth Regsouth Regwest /\*occnonoffic\*/occoffic1

occoffic3 occself Dependent EGraduate EMatric EProfessional Epostgrad /\*EPhD\*/ /\*Regcent\*/ /\*occoffic2\*/

/ctable lackfit maxiter=100;

score out = creditr.dmp;

**run**;

**proc** **logistic** data = creditr.credit8\_tr descending outmodel = creditr.dmm;

model NPA\_status = RevolvingUtilizationOfUnsecuredl /\*DebtRatio\*/ age gender\_M NumberOfTime30\_59DaysPastDueNotW

NumberOfOpenCreditLinesAndLoans NumberOfTime60\_89DaysPastDueNotW NumberOfTimes90DaysLate

NumberRealEstateLoansOrLines OwnHouse Regeast Regnorth Regsouth Regwest /\*occnonoffic\*/occoffic1

occoffic3 occself Dependent EGraduate EMatric EProfessional Epostgrad /\*EPhD\*/ Regcent occoffic2

/ctable lackfit maxiter=100;

score out = creditr.dmp;

**run**;

/\*\*Using stepwise to run the test model validation\*/

**proc** **logistic** data = creditr.credit8\_val descending outmodel = creditr.dmm;

model NPA\_status = RevolvingUtilizationOfUnsecuredl /\*DebtRatio\*/ age gender\_M NumberOfTime30\_59DaysPastDueNotW

NumberOfOpenCreditLinesAndLoans NumberOfTime60\_89DaysPastDueNotW NumberOfTimes90DaysLate

NumberRealEstateLoansOrLines OwnHouse Regeast Regnorth Regsouth Regwest /\*occnonoffic\*/occoffic1

occoffic3 occself Dependent EGraduate EMatric EProfessional Epostgrad /\*EPhD\*/ Regcent occoffic2

/ctable lackfit maxiter=100;

score out = creditr.dmp;

**run**;

/\*before that,run training dataset once again\*/

**proc** **logistic** data = creditr.credit8\_tr descending outmodel = creditr.dmm;

model NPA\_status = RevolvingUtilizationOfUnsecuredl /\*DebtRatio\*/ age gender\_M NumberOfTime30\_59DaysPastDueNotW

NumberOfOpenCreditLinesAndLoans NumberOfTime60\_89DaysPastDueNotW NumberOfTimes90DaysLate

NumberRealEstateLoansOrLines OwnHouse Regeast Regnorth Regsouth Regwest /\*occnonoffic\*/occoffic1

occoffic3 occself Dependent EGraduate EMatric EProfessional Epostgrad /\*EPhD\*/ Regcent occoffic2

/ctable lackfit maxiter=100;

score out = creditr.dmp;

**run**;

|  |
| --- |
| The WPS System |

The LOGISTIC Procedure

| **Model Information** | |
| --- | --- |
| Data Set | CREDITR.credit8\_tr |
| Response Variable | NPA\_Status |
| Number of Response Levels | 2 |
| Model | Binary logit |
| Optimisation Technique | Fisher's scoring |

| Number of Observations Read | 140250 |
| --- | --- |
| Number of Observations Used | 140250 |

| **Response Profile** | | |
| --- | --- | --- |
| **Ordered Value** | **NPA\_Status** | **Total Frequency** |
| 1 | 1 | 9346 |
| 2 | 0 | 130904 |
| **Probability modeled is NPA\_Status='1'.** | | |

| **Model Convergence Status** |
| --- |
| Convergence criterion (GCONV=1e-008) satisfied. |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept only** | **Intercept and Covariates** |
| AIC | 68683.758 | 53872.388 |
| SC | 68693.609 | 54098.966 |
| -2 Log L | 68681.758 | 53826.388 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > Chi-Square** |
| Likelihood Ratio | 14855.3698 | 22 | <.0001 |
| Score | 16836.7973 | 22 | <.0001 |
| Wald | 9834.1318 | 22 | <.0001 |

NOTE: The following parameters have been set to 0, since the variables are a linear combination of other variables as shown.

| Regcent = | 226.88\*Intercept + 988.331\*RevolvingUtilizationOfUnsecuredL + 10297.3\*age + 116.894\*gender\_M + 894.379\*NumberOfTime30\_59DaysPastDueNotW + 1769.52\*NumberOfOpenCreditLinesAndLoans + 772.858\*NumberOfTime60\_89DaysPastDueNotW + 816.027\*NumberOfTimes90DaysLate + 220.918\*NumberRealEstateLoansOrLines + 83.2104\*OwnHouse + 6.36623\*occoffic1 + 6.79845\*occoffic3 + 50.36\*occself + 218.115\*Dependent + 83.4444\*EGraduate + 131.286\*EProfessional + 12.1496\*Epostgrad |
| --- | --- |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| Intercept | 1 | -2.0702 | 0.1016 | 415.5384 | <.0001 |
| RevolvingUtilizationOfUnsecuredL | 1 | -0.00004 | 0.000065 | 0.4464 | 0.5041 |
| age | 1 | -0.0305 | 0.000900 | 1146.3221 | <.0001 |
| gender\_M | 1 | -0.4470 | 0.0304 | 215.6001 | <.0001 |
| NumberOfTime30\_59DaysPastDueNotW | 1 | 0.4991 | 0.0125 | 1599.0141 | <.0001 |
| NumberOfOpenCreditLinesAndLoans | 1 | -0.0104 | 0.00270 | 14.8883 | 0.0001 |
| NumberOfTime60\_89DaysPastDueNotW | 1 | -0.8742 | 0.0195 | 2020.3005 | <.0001 |
| NumberOfTimes90DaysLate | 1 | 0.4105 | 0.0164 | 626.0871 | <.0001 |
| NumberRealEstateLoansOrLines | 1 | 0.0294 | 0.0111 | 6.9616 | 0.0083 |
| OwnHouse | 1 | -0.4839 | 0.0283 | 291.2974 | <.0001 |
| Regeast | 1 | 1.4960 | 0.0839 | 317.5625 | <.0001 |
| Regnorth | 1 | 2.6365 | 0.0738 | 1277.2267 | <.0001 |
| Regsouth | 1 | 2.0757 | 0.0779 | 710.7677 | <.0001 |
| Regwest | 1 | 3.6744 | 0.0701 | 2743.7579 | <.0001 |
| occoffic1 | 1 | -0.4830 | 0.0577 | 70.0686 | <.0001 |
| occoffic3 | 1 | -0.4777 | 0.0555 | 73.9633 | <.0001 |
| occself | 1 | -0.5435 | 0.0374 | 210.7826 | <.0001 |
| Dependent | 1 | 0.0832 | 0.00980 | 72.1204 | <.0001 |
| EGraduate | 1 | -1.3419 | 0.0590 | 517.2210 | <.0001 |
| EMatric | 1 | 0.2167 | 0.0524 | 17.0677 | <.0001 |
| EProfessional | 1 | -1.2462 | 0.0585 | 453.2451 | <.0001 |
| Epostgrad | 1 | -0.7730 | 0.0606 | 162.7562 | <.0001 |
| Regcent | 0 | 0 | . | . | . |
| occoffic2 | 1 | 0.2328 | 0.0534 | 18.9834 | <.0001 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **Lower 95% Wald Confidence Limit** | **Upper 95% Wald Confidence Limit** |
| RevolvingUtilizationOfUnsecuredL | 1.000 | 1.000 | 1.000 |
| age | 0.970 | 0.968 | 0.972 |
| gender\_M | 0.640 | 0.602 | 0.679 |
| NumberOfTime30\_59DaysPastDueNotW | 1.647 | 1.607 | 1.688 |
| NumberOfOpenCreditLinesAndLoans | 0.990 | 0.984 | 0.995 |
| NumberOfTime60\_89DaysPastDueNotW | 0.417 | 0.402 | 0.433 |
| NumberOfTimes90DaysLate | 1.508 | 1.460 | 1.557 |
| NumberRealEstateLoansOrLines | 1.030 | 1.008 | 1.053 |
| OwnHouse | 0.616 | 0.583 | 0.652 |
| Regeast | 4.464 | 3.787 | 5.262 |
| Regnorth | 13.965 | 12.085 | 16.137 |
| Regsouth | 7.970 | 6.842 | 9.284 |
| Regwest | 39.424 | 34.360 | 45.235 |
| occoffic1 | 0.617 | 0.551 | 0.691 |
| occoffic3 | 0.620 | 0.556 | 0.692 |
| occself | 0.581 | 0.540 | 0.625 |
| Dependent | 1.087 | 1.066 | 1.108 |
| EGraduate | 0.261 | 0.233 | 0.293 |
| EMatric | 1.242 | 1.121 | 1.376 |
| EProfessional | 0.288 | 0.256 | 0.323 |
| Epostgrad | 0.462 | 0.410 | 0.520 |
| occoffic2 | 1.262 | 1.137 | 1.402 |

| **Association of Predicted Probabilities and Observed Responses** | | | |
| --- | --- | --- | --- |
| Percent Concordant | 81.6 | Somer's D | 0.639 |
| Percent Discordant | 17.7 | Gamma | 0.644 |
| Percent Tied | 0.7 | Tau-a | 0.08 |
| Pairs | 1223428784 | c | 0.82 |

| **Partition for the Hosmer and Lemeshow Test** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Group** | **Total** | **Observed Events** | **Expected Events** | **Observed Nonevents** | **Expected Nonevents** |
| 1 | 14034 | 62 | 26.25 | 13972 | 14007.75 |
| 2 | 14030 | 84 | 55.91 | 13946 | 13974.09 |
| 3 | 14028 | 165 | 109.32 | 13863 | 13918.68 |
| 4 | 14025 | 378 | 241.57 | 13647 | 13783.43 |
| 5 | 14025 | 479 | 414.10 | 13546 | 13610.90 |
| 6 | 14025 | 604 | 604.51 | 13421 | 13420.49 |
| 7 | 14025 | 709 | 834.80 | 13316 | 13190.20 |
| 8 | 14025 | 1003 | 1166.82 | 13022 | 12858.18 |
| 9 | 14025 | 1461 | 1754.19 | 12564 | 12270.81 |
| 10 | 14008 | 4401 | 4138.62 | 9607 | 9869.38 |

| **Hosmer and Lemeshow Goodness-of-Fit Test** | | |
| --- | --- | --- |
| **Chi-Square** | **DF** | **Pr > Chi-Square** |
| 305.2785 | 8 | <.0001 |

| **Classification Table** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Prob Level** | **Correct Events** | **Correct Non- events** | **Incorrect Events** | **Incorrect Non- events** | **Percentage Correct** | **Sensi- tivity** | **Speci- ficity** | **False POS** | **False NEG** |
| 0.000 | 9346 | 0 | 131E3 | 0 | 6.7 | 100.0 | 0.0 | 93.3 | . |
| 0.020 | 8771 | 51726 | 79178 | 575 | 43.1 | 93.8 | 39.5 | 90.0 | 1.1 |
| 0.040 | 8006 | 72782 | 58122 | 1340 | 57.6 | 85.7 | 55.6 | 87.9 | 1.8 |
| 0.060 | 7205 | 89541 | 41363 | 2141 | 69.0 | 77.1 | 68.4 | 85.2 | 2.3 |
| 0.080 | 6490 | 101E3 | 29829 | 2856 | 76.7 | 69.4 | 77.2 | 82.1 | 2.7 |
| 0.100 | 5846 | 109E3 | 21951 | 3500 | 81.9 | 62.6 | 83.2 | 79.0 | 3.1 |
| 0.120 | 5277 | 115E3 | 16391 | 4069 | 85.4 | 56.5 | 87.5 | 75.6 | 3.4 |
| 0.140 | 4830 | 118E3 | 12514 | 4516 | 87.9 | 51.7 | 90.4 | 72.2 | 3.7 |
| 0.160 | 4413 | 121E3 | 9733 | 4933 | 89.5 | 47.2 | 92.6 | 68.8 | 3.9 |
| 0.180 | 4045 | 123E3 | 7582 | 5301 | 90.8 | 43.3 | 94.2 | 65.2 | 4.1 |
| 0.200 | 3734 | 125E3 | 5930 | 5612 | 91.8 | 40.0 | 95.5 | 61.4 | 4.3 |
| 0.220 | 3447 | 126E3 | 4709 | 5899 | 92.4 | 36.9 | 96.4 | 57.7 | 4.5 |
| 0.240 | 3164 | 127E3 | 3783 | 6182 | 92.9 | 33.9 | 97.1 | 54.5 | 4.6 |
| 0.260 | 2935 | 128E3 | 2982 | 6411 | 93.3 | 31.4 | 97.7 | 50.4 | 4.8 |
| 0.280 | 2720 | 129E3 | 2355 | 6626 | 93.6 | 29.1 | 98.2 | 46.4 | 4.9 |
| 0.300 | 2523 | 129E3 | 1888 | 6823 | 93.8 | 27.0 | 98.6 | 42.8 | 5.0 |
| 0.320 | 2316 | 129E3 | 1520 | 7030 | 93.9 | 24.8 | 98.8 | 39.6 | 5.2 |
| 0.340 | 2130 | 13E4 | 1211 | 7216 | 94.0 | 22.8 | 99.1 | 36.2 | 5.3 |
| 0.360 | 1950 | 13E4 | 981 | 7396 | 94.0 | 20.9 | 99.3 | 33.5 | 5.4 |
| 0.380 | 1801 | 13E4 | 799 | 7545 | 94.1 | 19.3 | 99.4 | 30.7 | 5.5 |
| 0.400 | 1676 | 13E4 | 660 | 7670 | 94.1 | 17.9 | 99.5 | 28.3 | 5.6 |
| 0.420 | 1562 | 13E4 | 561 | 7784 | 94.0 | 16.7 | 99.6 | 26.4 | 5.6 |
| 0.440 | 1448 | 13E4 | 476 | 7898 | 94.0 | 15.5 | 99.6 | 24.7 | 5.7 |
| 0.460 | 1330 | 13E4 | 415 | 8016 | 94.0 | 14.2 | 99.7 | 23.8 | 5.8 |
| 0.480 | 1224 | 131E3 | 366 | 8122 | 93.9 | 13.1 | 99.7 | 23.0 | 5.9 |
| 0.500 | 1122 | 131E3 | 334 | 8224 | 93.9 | 12.0 | 99.7 | 22.9 | 5.9 |
| 0.520 | 1028 | 131E3 | 297 | 8318 | 93.9 | 11.0 | 99.8 | 22.4 | 6.0 |
| 0.540 | 944 | 131E3 | 260 | 8402 | 93.8 | 10.1 | 99.8 | 21.6 | 6.0 |
| 0.560 | 883 | 131E3 | 240 | 8463 | 93.8 | 9.4 | 99.8 | 21.4 | 6.1 |
| 0.580 | 815 | 131E3 | 219 | 8531 | 93.8 | 8.7 | 99.8 | 21.2 | 6.1 |
| 0.600 | 742 | 131E3 | 199 | 8604 | 93.7 | 7.9 | 99.8 | 21.1 | 6.2 |
| 0.620 | 677 | 131E3 | 178 | 8669 | 93.7 | 7.2 | 99.9 | 20.8 | 6.2 |
| 0.640 | 627 | 131E3 | 158 | 8719 | 93.7 | 6.7 | 99.9 | 20.1 | 6.3 |
| 0.660 | 566 | 131E3 | 140 | 8780 | 93.6 | 6.1 | 99.9 | 19.8 | 6.3 |
| 0.680 | 515 | 131E3 | 126 | 8831 | 93.6 | 5.5 | 99.9 | 19.7 | 6.3 |
| 0.700 | 466 | 131E3 | 110 | 8880 | 93.6 | 5.0 | 99.9 | 19.1 | 6.4 |
| 0.720 | 416 | 131E3 | 103 | 8930 | 93.6 | 4.5 | 99.9 | 19.8 | 6.4 |
| 0.740 | 368 | 131E3 | 84 | 8978 | 93.5 | 3.9 | 99.9 | 18.6 | 6.4 |
| 0.760 | 336 | 131E3 | 72 | 9010 | 93.5 | 3.6 | 99.9 | 17.6 | 6.4 |
| 0.780 | 297 | 131E3 | 64 | 9049 | 93.5 | 3.2 | 100.0 | 17.7 | 6.5 |
| 0.800 | 259 | 131E3 | 58 | 9087 | 93.5 | 2.8 | 100.0 | 18.3 | 6.5 |
| 0.820 | 224 | 131E3 | 51 | 9122 | 93.5 | 2.4 | 100.0 | 18.5 | 6.5 |
| 0.840 | 187 | 131E3 | 43 | 9159 | 93.4 | 2.0 | 100.0 | 18.7 | 6.5 |
| 0.860 | 150 | 131E3 | 32 | 9196 | 93.4 | 1.6 | 100.0 | 17.6 | 6.6 |
| 0.880 | 113 | 131E3 | 22 | 9233 | 93.4 | 1.2 | 100.0 | 16.3 | 6.6 |
| 0.900 | 82 | 131E3 | 14 | 9264 | 93.4 | 0.9 | 100.0 | 14.6 | 6.6 |
| 0.920 | 67 | 131E3 | 11 | 9279 | 93.4 | 0.7 | 100.0 | 14.1 | 6.6 |
| 0.940 | 41 | 131E3 | 8 | 9305 | 93.4 | 0.4 | 100.0 | 16.3 | 6.6 |
| 0.960 | 16 | 131E3 | 5 | 9330 | 93.3 | 0.2 | 100.0 | 23.8 | 6.7 |
| 0.980 | 2 | 131E3 | 3 | 9344 | 93.3 | 0.0 | 100.0 | 60.0 | 6.7 |
| 1.000 | 0 | 131E3 | 0 | 9346 | 93.3 | 0.0 | 100.0 | . | 6.7 |

/\* Rank Chart \*/

**proc** **rank** data = creditr.dmp out =creditr.decile\_data groups = 10 ties = mean;

ranks decile;

**run**;

**proc** **sort** data = creditr.decile\_data;

by descending p\_1;

**run**;

/\* export this file to creat lift chart\*/

**proc** **export** data = creditr.decile\_data

outfile = "Y:\Practice\Graded assignment\Regression\credit.csv"

dbms = csv replace;

**run**;

/\* Trying with complete data for creating lift chart by importing the original file again\*/

/\*Scoring is done to creat lift chart\*/

**PROC** **IMPORT** DATAFILE = "Z:\Assignments\Graded Assignment\Topic 10 - Regression Models\Credit.csv"

OUT = creditr.scoring1

DBMS = CSV

REPLACE;

**RUN**;

/\*Some data preparation is again required\*/

/\*monthly income showing 2 times\*/

**data** creditr.scoring2(drop = monthlyincome1);

set creditr.scoring1;

**run**;

/\* Missing NPA to be delete and the change of char var to num var\*/

**data** creditr.scoring2;

set creditr.scoring1;

if NPA\_status = . then delete;

if age = 0 then delete;

if gender = "Male" then gender\_M =1;

else gender\_M = 0;

if Education = "Matric" then EMatric = 1;

else EMatric = 0;

if Education = "Graduate" then EGraduate = 1;

else EGraduate = 0;

if Education = "PhD" then EPhD = 1;

else EPhD = 0;

if Education = "Professional" then EProfessional = 1;

else Eprofessional = 0;

if Education = "Post-Grad" then Epostgrad = 1;

else Epostgrad = 0;

if Occupation = "Self\_Emp" then occself = 1;

else occself = 0;

if Occupation = "Officer1" then occoffic1 = 1;

else occoffic1 = 0;

if Occupation = "Officer2" then occoffic2 = 1;

else occoffic2 = 0;

if Occupation = "Officer3" then occoffic3 = 1;

else occoffic3 = 0;

if Occupation = "Non-officer" then occnonoffic = 1;

else occnonoffic = 0;

if Region = "South" then Regsouth = 1;

else Regsouth = 0;

if Region = "Central" then Regcent = 1;

else Regcent = 0;

if Region = "North" then Regnorth = 1;

else Regnorth = 0;

if Region = "East" then Regeast = 1;

else Regeast = 0;

if Region = "West" then Regwest = 1;

else Regwest = 0;

if Rented\_OwnHouse = "Ownhouse" then OwnHouse = 1;

else OwnHouse = 0;

monthlyincome1 = monthlyincome\*1;

if 0 = monthlyincome1> 1000000 then delete;

if monthlyincome1 = . then Debtratio = *round*((debtratio/6670.22),0.01);

if monthlyincome1 = . then monthlyincome1 = 6670.22;

drop monthlyincome Education Occupation Region Gender rented\_Ownhouse;

Dependent = numberofdependents\*1;

if dependent = . then dependent = 1;

drop numberofdependents;

**run**;

**proc** **contents** data = creditr.scoring2;

**run**;

|  |
| --- |
| The WPS System |

The CONTENTS Procedure

| Data Set Name | SCORING2 |
| --- | --- |
| Member Type | DATA |
| Engine | WPD |
| Created | 14OCT2017:13:23:05 |
| Last Modified | 14OCT2017:13:23:05 |
| Observations | 149999 |
| Variables | 28 |
| Indexes | 0 |
| Observation Length | 223 |
| Deleted Observations | 0 |
| Data Set Type |  |
| Label |  |
| Compressed | NO |
| Sorted | NO |
| Data Representation | Little endian, IEEE Windows |
| Encoding | wlatin1 Windows-1252 Western |

| **Engine/Host Dependent Information** | |
| --- | --- |
| Data Set Page Size | 4096 |
| Number of Data Set Pages | 8335 |
| First Data Page | 1 |
| Max Obs Per Page | 18 |
| Obs In First Data Page | 18 |
| Data Set Diagnostic Code | 0009 |
| File Name | Y:\Practice\Graded assignment\SCORING2.wpd |
| WPD Engine Version | 3 |
| Large Data Set Support | no |

| **Alphabetic List of Variables and Attributes** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Number** | **Variable** | **Type** | **Len** | **Pos** | **Format** | **Informat** | **Label** |
| 5 | DebtRatio | Num | 8 | 32 | BEST12. | BEST32. | DebtRatio |
| 28 | Dependent | Num | 8 | 208 |  |  |  |
| 13 | EGraduate | Num | 8 | 88 |  |  |  |
| 12 | EMatric | Num | 8 | 80 |  |  |  |
| 14 | EPhD | Num | 8 | 96 |  |  |  |
| 15 | EProfessional | Num | 8 | 104 |  |  |  |
| 16 | Epostgrad | Num | 8 | 112 |  |  |  |
| 6 | MonthlyIncome1 | Char | 7 | 216 | $7. | $7. | MonthlyIncome |
| 1 | NPA\_Status | Num | 8 | 0 | BEST12. | BEST32. | NPA Status |
| 7 | NumberOfOpenCreditLinesAndLoans | Num | 8 | 40 | BEST12. | BEST32. | NumberOfOpenCreditLinesAndLoans |
| 4 | NumberOfTime30\_59DaysPastDueNotW | Num | 8 | 24 | BEST12. | BEST32. | NumberOfTime30-59DaysPastDueNotWorse |
| 10 | NumberOfTime60\_89DaysPastDueNotW | Num | 8 | 64 | BEST12. | BEST32. | NumberOfTime60-89DaysPastDueNotWorse |
| 8 | NumberOfTimes90DaysLate | Num | 8 | 48 | BEST12. | BEST32. | NumberOfTimes90DaysLate |
| 9 | NumberRealEstateLoansOrLines | Num | 8 | 56 | BEST12. | BEST32. | NumberRealEstateLoansOrLines |
| 27 | OwnHouse | Num | 8 | 200 |  |  |  |
| 23 | Regcent | Num | 8 | 168 |  |  |  |
| 25 | Regeast | Num | 8 | 184 |  |  |  |
| 24 | Regnorth | Num | 8 | 176 |  |  |  |
| 22 | Regsouth | Num | 8 | 160 |  |  |  |
| 26 | Regwest | Num | 8 | 192 |  |  |  |
| 2 | RevolvingUtilizationOfUnsecuredL | Num | 8 | 8 | BEST12. | BEST32. | RevolvingUtilizationOfUnsecuredLines |
| 3 | age | Num | 8 | 16 | BEST12. | BEST32. | age |
| 11 | gender\_M | Num | 8 | 72 |  |  |  |
| 21 | occnonoffic | Num | 8 | 152 |  |  |  |
| 18 | occoffic1 | Num | 8 | 128 |  |  |  |
| 19 | occoffic2 | Num | 8 | 136 |  |  |  |
| 20 | occoffic3 | Num | 8 | 144 |  |  |  |
| 17 | occself | Num | 8 | 120 |  |  |  |

**proc** **means** data = creditr.scoring2 mean stddev n nmiss min max;

**run**;

|  |
| --- |
| The WPS System |

The MEANS Procedure

| **Summary statistics** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **Mean** | **Std Dev** | **N** | **N Miss** | **Minimum** | **Maximum** |
| NPA\_Status | NPA Status | 0.0668404 | 0.2497463 | 149999 | 0 | 0 | 1.0000000 |
| RevolvingUtilizationOfUnsecuredL | RevolvingUtilizationOfUnsecuredLines | 6.0484717 | 249.7562028 | 149999 | 0 | 0 | 50708.00 |
| age | age | 52.2955553 | 14.7712980 | 149999 | 0 | 21.0000000 | 109.0000000 |
| NumberOfTime30\_59DaysPastDueNotW | NumberOfTime30-59DaysPastDueNotWorse | 0.4210295 | 4.1927950 | 149999 | 0 | 0 | 98.0000000 |
| DebtRatio | DebtRatio | 21.3765647 | 380.2078948 | 149999 | 0 | 0 | 61106.50 |
| NumberOfOpenCreditLinesAndLoans | NumberOfOpenCreditLinesAndLoans | 8.4527764 | 5.1459642 | 149999 | 0 | 0 | 58.0000000 |
| NumberOfTimes90DaysLate | NumberOfTimes90DaysLate | 0.2659751 | 4.1693176 | 149999 | 0 | 0 | 98.0000000 |
| NumberRealEstateLoansOrLines | NumberRealEstateLoansOrLines | 1.0182335 | 1.1297719 | 149999 | 0 | 0 | 54.0000000 |
| NumberOfTime60\_89DaysPastDueNotW | NumberOfTime60-89DaysPastDueNotWorse | 0.2403883 | 4.1551932 | 149999 | 0 | 0 | 98.0000000 |
| gender\_M |  | 0.6153574 | 0.4865123 | 149999 | 0 | 0 | 1.0000000 |
| EMatric |  | 0.1054007 | 0.3070701 | 149999 | 0 | 0 | 1.0000000 |
| EGraduate |  | 0.2650284 | 0.4413498 | 149999 | 0 | 0 | 1.0000000 |
| EPhD |  | 0.0420069 | 0.2006057 | 149999 | 0 | 0 | 1.0000000 |
| EProfessional |  | 0.3394689 | 0.4735306 | 149999 | 0 | 0 | 1.0000000 |
| Epostgrad |  | 0.2480950 | 0.4319087 | 149999 | 0 | 0 | 1.0000000 |
| occself |  | 0.4274362 | 0.4947081 | 149999 | 0 | 0 | 1.0000000 |
| occoffic1 |  | 0.1010940 | 0.3014542 | 149999 | 0 | 0 | 1.0000000 |
| occoffic2 |  | 0.0888873 | 0.2845819 | 149999 | 0 | 0 | 1.0000000 |
| occoffic3 |  | 0.1084941 | 0.3110044 | 149999 | 0 | 0 | 1.0000000 |
| occnonoffic |  | 0 | 0 | 149999 | 0 | 0 | 0 |
| Regsouth |  | 0.1566344 | 0.3634569 | 149999 | 0 | 0 | 1.0000000 |
| Regcent |  | 0 | 0 | 149999 | 0 | 0 | 0 |
| Regnorth |  | 0.2273282 | 0.4191077 | 149999 | 0 | 0 | 1.0000000 |
| Regeast |  | 0.1370076 | 0.3438565 | 149999 | 0 | 0 | 1.0000000 |
| Regwest |  | 0.1859946 | 0.3891036 | 149999 | 0 | 0 | 1.0000000 |
| OwnHouse |  | 0.5730238 | 0.4946404 | 149999 | 0 | 0 | 1.0000000 |
| Dependent |  | 0.7635651 | 1.1010851 | 149999 | 0 | 0 | 20.0000000 |

/\*validating with the above export file with the previously created creditr.dmm file\*/

**proc** **logistic** inmodel = creditr.dmm;

score data = creditr.scoring2 out = creditr.scoring3;

**run**;

**proc** **rank** data = creditr.dmp out =creditr.decile\_data1 groups = 10 ties = mean;

var p\_1;

ranks decile;

**run**;

**proc** **sort** data = creditr.decile\_data1;

by descending p\_1;

**run**;

/\* Exporting the scored dataset to excel for chart creation and further model validation \*/

**proc** **export** data = creditr.decile\_data1 outfile ="Y:\Practice\Graded assignment\Regression\credit\_varify.csv"

dbms = csv replace;

**run**;

**Creating lift chart**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Decile** | **Predicted** | **Actual** | **Cum Pred** | **Cum Actual** | **Cum Pred %** | **Cum Actual%** | **Fitting Value** |  |
| 9 | 4153.6 | 1000.2 | 4153.6 | 1000.2 | 41.53 | 10 | 4.153 |  |
| 8 | 1766.9 | 1000.2 | 5920.5 | 2000.4 | 59.19 | 20 | 2.96 |  |
| 7 | 1775.4 | 1000.2 | 7695.9 | 3000.6 | 76.94 | 30 | 2.565 |  |
| 6 | 841.1 | 1000.2 | 8537 | 4000.8 | 85.35 | 40 | 2.134 |  |
| 5 | 608.95 | 1000.2 | 9145.95 | 5001 | 91.44 | 50 | 1.829 |  |
| 4 | 417.13 | 1000.2 | 9563.08 | 6001.2 | 95.61 | 60 | 1.593 |  |
| 3 | 243.76 | 1000.2 | 9806.84 | 7001.4 | 98.04 | 70 | 1.401 |  |
| 2 | 111.95 | 1000.2 | 9918.79 | 8001.6 | 99.16 | 80 | 1.24 |  |
| 1 | 56.6 | 1000.2 | 9975.39 | 9001.8 | 99.73 | 90 | 1.108 |  |
| 0 | 27.1 | 1000.2 | 10002.49 | 10002 | 100 | 100 | 1 |  |
|  | **10002.49** | **10002** |  |  |  |  |  |  |

**Summery**

**Data expolaration and preparation is required:**

1. I observed that there is a duplicate column of “monthlyincome1”, so I removed the duplicate column.
2. I found that 2 missing value of NPA\_status so I have removed NPA\_status rows, with the missing value.
3. Created dummy independent variable and with that I have also change the charracter variable to numeric variable for Gender region eduction occupation numberofdependents monthlyincome1 ownhouse.
4. I have replaced missing value of monthlyincome1 with its mean value.
5. I have replaced missing value of numberofdependents with the mean rounds to figure i.e. 1.
6. Tried iteration process and found best concordant of 81.6%. Later I form the Gain or lift chart as a model scoring.