**Prediction of Loan Status using Logistic Regression Model**

Gayatri Pant

Youngstown State University

Dr. Andy Chang

SAS Programming and Data Analytics

1. **Introduction**

The term project is about analyzing loan data. The objective of the term project is to explore various factors related to customers’ loan status and find a good predictive model for customers’ loan status, such as Fully Paid (not delinquent) or Not Fully Paid (delinquent). In this project, we will build a model to predict customers’ loan status, delinquent or not.

This essay discusses loan data predictive modeling. Logistic regression is the model used in this study. A customer borrows money from the bank for a predetermined period of time. Before approving a loan, the banks must examine if the borrower has the ability to repay it. We have numerous distinct credit scores. The FICO score is the most popular. This approach is great for learning about the client because someone with a good FICO score may make a good borrower. But a borrower with a low FICO score might not necessarily be a bad borrower, and a borrower with a high FICO score would not always be a good borrower either. Being a successful borrower and a bad borrower are influenced by more things than just a FICO score. Banks should make wise financial judgments rather than depending just on FICO scores. In these situations, predictive modeling is useful. Using predictive modeling, it is possible to determine the borrower's capacity to pay the amount. Whether the account will eventually become delinquent or not can be determined using the predictive modeling utilized in the article.

1. **Methods/Data Preparation**

The data is obtained from blackboard provided by the Dr. Andy Chang. The data contains 141 columns and 20,000 rows and was in the excel format. After that we imported the data in SAS studio. Since the 'Loan status' variable was the response variable, we used Binary Logistic Regression to predict our model. The model must be able to forecast whether or not the account will be past due at the end. Therefore, the only variable that needs to be predicted is "Loan status," which might either be "Delinquent" or "Not Delinquent."

The Data Dictionary file provided by the Dr. Andy Chang, helped us in understanding the meaning of all the variables in the data. We studied the dictionary and chose the predictors variables whose meaning was closely related with loan status. The variables that we found to the best predictors with their meanings were as follows:

* loan\_amnt: The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
* int\_rate: Interest rate on the loan.
* total\_rec\_last\_fee: Late fees received to date.
* term: The number of payments on the loan. Values are in months and can be either 36 or 60.
* purpose: A category provided by the borrower for the loan request.
* Int rate: Interest Rate on the loan.
* fundedAmnt: The total amount committed to that loan at that point in time.
* delinq2Yrs: The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
* delinqAmnt: The past-due amount owed for the accounts on which the borrower is now delinquent.
* accNowDelinq: The number of accounts on which the borrower is now delinquent.

1. **Analysis/Results**

**3.1. The Logistic Regression Model.**

Binary logistic regression was used to find the best predictors for predicting the response variable and to find the good model. Looked at significant factors to choose the model. Since term, funded amount, total payment, interest rate, total rec late fee, were significant we chose these as predictor variables. Since purpose, loan amount, delinq\_2yrs were not significant we didn’t use these variables as predictor variables.

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  | 1 | 4.2621 | 0.3394 | 157.6758 | <.0001 |
| **purpose** | **car** | 1 | -0.0910 | 0.3229 | 0.0794 | 0.7782 |
| **purpose** | **credit\_card** | 1 | -0.5094 | 0.2982 | 2.9183 | 0.0876 |
| **purpose** | **debt\_consolidation** | 1 | -0.3801 | 0.2823 | 1.8130 | 0.1781 |
| **purpose** | **home\_improvement** | 1 | -0.3536 | 0.3088 | 1.3113 | 0.2522 |
| **purpose** | **house** | 1 | -0.1827 | 0.4846 | 0.1422 | 0.7061 |
| **purpose** | **major\_purchase** | 1 | 0.1547 | 0.3341 | 0.2144 | 0.6434 |
| **purpose** | **medical** | 1 | -0.1400 | 0.3845 | 0.1327 | 0.7157 |
| **purpose** | **moving** | 1 | -0.7844 | 0.3454 | 5.1565 | 0.0232 |
| **purpose** | **other** | 1 | -0.5271 | 0.2948 | 3.1957 | 0.0738 |
| **purpose** | **renewable\_energy** | 1 | -0.5766 | 0.5961 | 0.9356 | 0.3334 |
| **purpose** | **small\_business** | 1 | -0.8026 | 0.3271 | 6.0220 | 0.0141 |
| **purpose** | **vacation** | 1 | -0.2753 | 0.4037 | 0.4650 | 0.4953 |
| **purpose** | **wedding** | 0 | 0 | . | . | . |
| **term** | **36 months** | 1 | 1.0268 | 0.0998 | 105.9540 | <.0001 |
| **term** | **60 months** | 0 | 0 | . | . | . |
| **delinq\_2yrs** |  | 1 | -0.0452 | 0.0624 | 0.5242 | 0.4690 |
| **delinq\_amnt** |  | 0 | 0 | . | . | . |
| **acc\_now\_delinq** |  | 0 | 0 | . | . | . |
| **funded\_amnt** |  | 1 | -0.00160 | 0.000051 | 964.8926 | <.0001 |
| **loan\_amnt** |  | 1 | 0.000030 | 0.000034 | 0.7560 | 0.3846 |
| **int\_rate** |  | 1 | -21.4701 | 1.1199 | 367.5135 | <.0001 |
| **total\_pymnt** |  | 1 | 0.00149 | 0.000034 | 1882.1072 | <.0001 |
| **total\_rec\_late\_fee** |  | 1 | -0.0887 | 0.00472 | 353.9050 | <.0001 |

Table 1: Analysis of the maximum likelihood estimates

Observed at only significant variables to predict the model. Since term, funded amount, total payment, interest rate, total rec late fee, were significant we chose these as predictor variables and find the logistic regression equation.

The logistic regression equation is:

P (Fully Paid) =

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  | 1 | 3.9763 | 0.1908 | 434.5329 | <.0001 |
| **term** | **36 months** | 1 | 0.9705 | 0.0955 | 103.2821 | <.0001 |
| **term** | **60 months** | 0 | 0 | . | . | . |
| **funded\_amnt** |  | 1 | -0.00157 | 0.000037 | 1828.8098 | <.0001 |
| **total\_pymnt** |  | 1 | 0.00149 | 0.000034 | 1898.5082 | <.0001 |
| **int\_rate** |  | 1 | -21.9283 | 1.0924 | 402.9324 | <.0001 |
| **total\_rec\_late\_fee** |  | 1 | -0.0891 | 0.00470 | 359.5925 | <.0001 |

Table 2: Analysis of the maximum likelihood estimates

**3.2. Interpretation of Odds Ratio**

Odds ratio are an important tool for analysis since odds ratio greater than 1 indicates a positive association and odds ratio less than 1 indicates a negative association. We have positive association between loan status and total payment, positive association in term 36 months vs 60 months. Since funded amount point estimate is closely to 1, so we have closely positive association between loan status and funded amount. We have negative association between loan status and interest rate.

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **term 36 months vs 60 months** | 2.639 | 2.189 | 3.183 |
| **funded\_amnt** | 0.998 | 0.998 | 0.999 |
| **total\_pymnt** | 1.001 | 1.001 | 1.002 |
| **int\_rate** | <0.001 | <0.001 | <0.001 |
| **total\_rec\_late\_fee** | 0.915 | 0.906 | 0.923 |

Table 3: Odds Ratio Estimates

* 1. **Model Performance**

We applied many statistical tests to evaluate how well our model worked. First, we utilized the concordance statistic to evaluate how well our statistical model predicted future events. The higher the c-statistic, the better the model can discriminate between subjects who do experience the outcome of interest and subjects who do not. As we can see from the table, our model's c-statistic value is 0.963, which is quite near to 1, making it appear better. To assess the prediction efficacy of our statistical model, the concordance statistic is insufficient.

| **Association of Predicted Probabilities and Observed Responses** | | | |
| --- | --- | --- | --- |
| **Percent Concordant** | 96.3 | **Somers' D** | 0.925 |
| **Percent Discordant** | 3.7 | **Gamma** | 0.925 |
| **Percent Tied** | 0.0 | **Tau-a** | 0.240 |
| **Pairs** | 51780864 | **c** | 0.963 |

Table 4: Association of predicted probabilities and observed responses.

We must establish a cutoff value in order to use a logistic regression model's prediction, which is a logit or probability, to categorize positive and negative responses. The tradeoff we're making between the two measures was shown and quantified using the ROC curve to establish a cutoff point. We choose to utilize 0.9 as a cutoff value to distinguish between the two responses of the loan status variable after visualizing the ROC curve.

Chart, line chart

Description automatically generated

Figure1: ROC curve.

We also tested the predictive accuracy of our logistic regression model using a classification table. The dependent outcome's observed values are cross-classified in this table with the predicted values (at a user-defined cut-off value, in this case p=0.9). Table 5 demonstrates that the correct categorization rate is 92.2. This test shows that our model can generalize effectively.

| **Classification Table** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Prob Level** | **Correct** | | **Incorrect** | | **Percentages** | | | | |
| **Event** | **Non- Event** | **Event** | **Non- Event** | **Correct** | **Sensi- tivity** | **Speci- ficity** | **Pos Pred** | **Neg Pred** |
| **0.900** | 15619 | 2824 | 232 | 1325 | 92.2 | 92.2 | 92.4 | 98.5 | 68.1 |

Table 5: Classification Table

We can see from figure 2 that the probability of loan status charged off is higher than the probability of fully paid.

Chart, line chart

Description automatically generated

Figure2: Empirical Distribution for probability.

From the Hosmer and Lemeshow Goodness of fit test p-value is less than 0.05 indicating that our model is not fit well. So, we can improvise our model by adding more variables.

| **Hosmer and Lemeshow Goodness-of-Fit Test** | | |
| --- | --- | --- |
| **Chi-Square** | **DF** | **Pr > ChiSq** |
| 13646.3924 | 8 | <.0001 |

Table 6: Goodness of fit test

We calculated the predicted probability of case number 20000 to see if we can approve the loan for case like 20000 or not. We got the predicted probability 0.790473. That means we have nearly 79% chance that the case number 20000 will fully pay the loan.

| **Case\_No** | **Predicted Probability: loan\_status=Fully Paid** |
| --- | --- |
| 20000 | 0.790473 |
| 20000 |  |

Table 7: Predicted Probability in case no. 20000

In the table 8, we can see that the fully paid number are more than the charged off number in both cases term 36 months and term 60 months. The mean of funded amount and total amount is higher in fully paid than charge off. The mean of interest rate and total rec late fee is lower in fully paid than charged off, in the term 36 months. But in term 36 months, the mean of funded amount is higher in charged off, mean of total payment is lower in charged off, mean of interest and mean of total rec late fee is higher in charged off than the fully paid.

| **term** | **loan\_status** | **N Obs** | **Variable** | **Label** | **Mean** | **Std Dev** | **Minimum** | **Maximum** | **N** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 36 months | Charged Off | 1377 | funded\_amnt  total\_pymnt  int\_rate  total\_rec\_late\_fee | funded\_amnt  total\_pymnt  int\_rate  total\_rec\_late\_fee | 8369.05  4966.74  0.1214211  2.6058208 | 6125.25  5127.49  0.0348243  8.6969594 | 1000.00  0  0.0542000  0 | 35000.00  37527.48  0.2248000  120.8100000 | 1377  1377  1377  1377 |
|  | Fully Paid | 11546 | funded\_amnt  total\_pymnt  int\_rate  total\_rec\_late\_fee | funded\_amnt  total\_pymnt  int\_rate  total\_rec\_late\_fee | 9449.58  10827.99  0.1042662  0.5759417 | 6493.71  7581.99  0.0340634  4.1433016 | 1000.00  1011.43  0.0542000  0 | 35000.00  46205.60  0.2322000  87.0000001 | 11546  11546  11546  11546 |
| 60 months | Charged Off | 1679 | funded\_amnt  total\_pymnt  int\_rate  total\_rec\_late\_fee | funded\_amnt  total\_pymnt  int\_rate  total\_rec\_late\_fee | 16776.97  10101.24  0.1637310  3.8587436 | 8465.54  8713.46  0.0337756  12.7280146 | 1000.00  0  0.0729000  0 | 35000.00  55836.73  0.2440000  180.2000000 | 1679  1679  1679  1679 |
|  | Fully Paid | 5398 | funded\_amnt  total\_pymnt  int\_rate  total\_rec\_late\_fee | funded\_amnt  total\_pymnt  int\_rate  total\_rec\_late\_fee | 16634.13  22158.30  0.1509934  0.8977010 | 8350.42  11658.38  0.0351878  5.9175708 | 1000.00  1151.17  0.0600000  0 | 35000.00  58886.47  0.2459000  170.7600004 | 5398  5398  5398  5398 |

Table 8: Summery Statistics

We observe the boxplot because Boxplot is best to compare the one categorical and one quantitative variable. Also, we can see the relationship clearly from boxplot. We can see from figure 3 that the mean of funded amount, mean of interest rate and mean of total rec late fee has higher mean in the charged off, than the fully paid. The mean of total payment is lower in charged off than the fully paid.

Chart, box and whisker chart

Description automatically generated Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated Chart, box and whisker chart

Description automatically generated

Figure3: Boxplots by loan status.

We can see from figure 4 that the mean of funded amount, mean of total payment, mean of interest rate and mean of total rec late fee has higher mean in 60 months term than the 36 months term.

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated Chart, box and whisker chart

Description automatically generated

Figure4: Boxplots by term.

1. **Conclusion**

The objective of the term project was to explore various factors related to customers’ loan status and find a good predictive model for customers’ loan status, such as Fully Paid (not delinquent) or Not Fully Paid (delinquent).

For this project, we examined the data, found the list of potentially good predictor variables for predicting customers’ loan status, and performed some preliminary analysis about the data such as descriptive statistics, correlation analysis for examining correlation between the potential predictors and the customers’ loan status, calculated concordance statistic to evaluate how well our statistical model predicted future events, observed ROC curve to establish a cutoff point, utilized classification table to demonstrate the correct categorization rate etc.

We examined 5 potential predictor variables term, funded amount, total payment, interest rate, total rec late fee for predicting loan status. We chose these variables because these were significant variables.

Using data with various variables, we constructed a binary logistic regression model to address the issue of predicting loan delinquency status. Our logistic regression equation is:

P (Fully Paid) =

In our model, the correct rate of classification is 92.2.

We found the probability of non-delinquency for a customer with case like 20000 of our data is 0.7904731829. Since the p-value is less than 0.9, we do not approve the loan for cases like 20000.

**SAS Codes**

/\* Term project-STAT 5811

Gayatri Pant

\*/

ods noproctitle;

title "Term Project from Gayatri Pant";

libname Gp "~/termpaper/";

FILENAME REFFILE '~/termpaper/LoanStats-20000.xlsx';

PROC IMPORT DATAFILE=REFFILE

DBMS=XLSX

OUT=Gp.IMPORT;

GETNAMES=YES;

RUN;

PROC CONTENTS DATA=Gp.IMPORT; RUN;

/\*Interested to look at the predictor variables for loan status

that why people took loan, how much loan they took, how much loan they paid, with interest and late fee\*/

proc logistic data=GP.IMPORT;

class purpose term / param=glm;

model loan\_status(event='Fully Paid')=purpose term delinq\_2yrs delinq\_amnt acc\_now\_delinq

funded\_amnt loan\_amnt int\_rate total\_pymnt total\_rec\_late\_fee delinq\_amnt acc\_now\_delinq delinq\_2yrs / link=logit

lackfit technique=fisher;

run;

/\*Looked at significant factors to choose the model. since purpose,

loan amount, delinq\_2yrs are not significant I don't want to study

those variables as predictor variables\*/

proc logistic data=GP.IMPORT;

class term / param=glm;

model loan\_status(event='Fully Paid')= term

funded\_amnt total\_pymnt int\_rate total\_rec\_late\_fee / link=logit lackfit technique=fisher;

run;

/\*linear regression\*/

ods noproctitle;

ods graphics / imagemap=on;

proc glmselect data=GP.IMPORT outdesign(addinputvars)=Work.reg\_design

plots=(criterionpanel);

model loan\_amnt=funded\_amnt total\_pymnt delinq\_2yrs delinq\_amnt / showpvalues

selection=stepwise

(select=sbc);

run;

proc reg data=Work.reg\_design alpha=0.05 plots(only)=(diagnostics residuals

partial rstudentbypredicted observedbypredicted);

ods select ParameterEstimates CollinDiag DiagnosticsPanel ResidualPlot PartialPlot RStudentByPredicted

ObservedByPredicted;

model loan\_amnt=&\_GLSMOD / partial;

output out=work.Reg\_stats h=h\_ p=p\_ lcl=lcl\_ ucl=ucl\_ lclm=lclm\_ uclm=uclm\_

r=r\_ student=student\_;

run;

quit;

proc reg data=Work.reg\_design alpha=0.05 plots(only)=(diagnostics residuals

observedbypredicted);

ods select ParameterEstimates CollinDiag DiagnosticsPanel ResidualPlot

ObservedByPredicted;

model loan\_amnt=&\_GLSMOD / collin tol vif;

output out=work.Reg\_stats0002 p=p\_ lcl=lcl\_ ucl=ucl\_ lclm=lclm\_ uclm=uclm\_

r=r\_;

run;

quit;

proc delete data=Work.reg\_design;

run;

/\*summary statistics\*/

ods noproctitle;

ods graphics / imagemap=on;

proc means data=GP.IMPORT chartype mean std min max n vardef=df;

var funded\_amnt total\_pymnt int\_rate total\_rec\_late\_fee;

class term loan\_status;

output out=work.Means\_stats mean=std=min=max=n= / autoname;

run;

proc means data=GP.IMPORT chartype mean std min max n vardef=df;

var funded\_amnt total\_pymnt int\_rate total\_rec\_late\_fee;

class term;

output out=work.Means\_stats mean=std=min=max=n= / autoname;

run;

proc univariate data=GP.IMPORT vardef=df noprint;

var funded\_amnt total\_pymnt int\_rate total\_rec\_late\_fee;

class loan\_status term;

histogram funded\_amnt total\_pymnt int\_rate total\_rec\_late\_fee/ normal(noprint);

run;

/\*Looking at boxplot because, Boxplot is best to compare the one caterogical and one quantative variables\*/

%DEHisto(data=GP.IMPORT, avar=funded\_amnt total\_pymnt, classVar=loan\_status delinq\_amnt delinq\_2yrs acc\_now\_delinq);

proc sort data=GP.IMPORT out=WORK.TempSorted4877;

by loan\_status;

run;

proc boxplot data=WORK.TempSorted4877;

plot (funded\_amnt)\*loan\_status / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (total\_rec\_late\_fee)\*loan\_status / boxstyle=schematic;

run;

proc sort data=GP.IMPORT out=WORK.TempSorted4877;

by term;

run;

proc boxplot data=WORK.TempSorted4877;

plot (funded\_amnt)\*term / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (total\_pymnt)\*term / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (int\_rate)\*term / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (total\_rec\_late\_fee)\*term / boxstyle=schematic;

run;

proc sort data=GP.IMPORT out=WORK.TempSorted4877;

by delinq\_amnt;

run;

proc boxplot data=WORK.TempSorted4877;

plot (funded\_amnt)\*delinq\_amnt / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (total\_pymnt)\*delinq\_amnt / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (int\_rate)\*delinq\_amnt / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (total\_rec\_late\_fee)\*delinq\_amnt / boxstyle=schematic;

run;

proc sort data=GP.IMPORT out=WORK.TempSorted4877;

by delinq\_2yrs;

run;

proc boxplot data=WORK.TempSorted4877;

plot (funded\_amnt)\*delinq\_2yrs / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (total\_pymnt)\*delinq\_2yrs / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (int\_rate)\*delinq\_2yrs / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (total\_rec\_late\_fee)\*delinq\_2yrs / boxstyle=schematic;

run;

proc sort data=GP.IMPORT out=WORK.TempSorted4877;

by acc\_now\_delinq;

run;

proc boxplot data=WORK.TempSorted4877;

plot (funded\_amnt)\*acc\_now\_delinq / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (total\_pymnt)\*acc\_now\_delinq / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (int\_rate)\*acc\_now\_delinq / boxstyle=schematic;

run;

proc boxplot data=WORK.TempSorted4877;

plot (total\_rec\_late\_fee)\*acc\_now\_delinq / boxstyle=schematic;

run;

data predict1;

set GP.IMPORT;

run;

data lregp1;

input term Case\_No;

cards;

1 20000

;

run;

proc append base=predict1 data=lregp1 force;

run;

proc logistic data=predict1;

class term / param=glm;

model loan\_status(event='Fully Paid')=term Case\_No / link=logit lackfit

technique=fisher;

output out=work.Logistic\_stats predicted=pred\_ lower=lcl\_ upper=ucl\_ /

alpha=0.05;

score out=work.Logistic\_scores;

run;

proc logistic data=gp.import;

class term / param=glm;

model loan\_status(event='Fully Paid')=term funded\_amnt total\_pymnt int\_rate

total\_rec\_late\_fee / link=logit lackfit rsquare

ctable clodds=pl pprob=0.9 selection=backward slstay=0.05 hierarchy=single details

technique=fisher;

output out=work.Logistic\_stats1 xbeta=xbeta\_ predicted=pred\_ reslik=reslik\_;

score out=work.Logistic\_scores1 fitstat outroc=vroc;

code;

run;

proc npar1way edf data=work.Logistic\_scores;

class loan\_status;

var 'P\_Fully Paid'n;

run;

proc sql;

select case\_no, 'P\_Fully Paid'n

from work.Logistic\_scores

where case\_no=20000;

quit;

proc sort data=GP.IMPORT out=WORK.TempSorted4877;

by loan\_status;

run;

data WORK.TempSorted4877;

set WORK.TempSorted4877;

loan\_status\_new=0;

if loan\_status="Fully Paid" then

loan\_status\_new=1;

run;

proc means data=WORK.TempSorted4877 noprint nway;

class purpose;

var loan\_status\_new;

output out=level mean=prop;

run;

proc print data=level;

run;

%global sl;

proc sql;

select 1-probchi(log(sum(loan\_status\_new ge 0)), 1) into: sl from

WORK.TempSorted4877;

quit;

proc logistic data=predict1

plots(maxpoints=none);

class term / param=ref;

model loan\_status(event="Fully Paid")=term

funded\_amnt total\_pymnt int\_rate total\_rec\_late\_fee

/selection=backward fast slstay=&sl lackfit ctable clodds=pl pprob=0.7;

output out=test p=ppred;

units funded\_amnt=1000 total\_pymnt=1000 /default=1;

score data=WORK.TempSorted4877 out=logit\_File fitstat outroc=vroc;

run;