

## SAS assignment 5.2

### Variables

revol\_util(+). The amount of credit the borrower is using relative to all available revolving credit. Higher utility means the borrower is using more of his credit and is less likely to pay it back, resulting in high default probability.

int\_rate(+). Interest rate. Higher interest rate indicates borrowers need to pay more money back to the club. This will cause higher default rate.

total\_pymnt(-). Payment received for total amount. The more money already paid back, the less likely the loan would default.

ins\_to\_inc(+). A ratio defined as installment/annual\_inc. It describes the paying ability of the borrower. Higher ins\_to\_inc indicates either high installment or low annual income. Both have a negative influence on avoiding default.

last\_pymnt\_amnt(-). Last total payment received. It is negatively correlated with the probability of default.

inq\_last\_6mths(+). The number of inquiries by creditors during the past 6 months. This is a quite subjective because it depends on the action of creditors. The assumption is the creditors would inquiry a loan more often if they have doubt whether it would default or not.

*The SAS System**The LOGISTIC Procedure*

Model Information	
Data Set	WORK.ORIGDATA
Response Variable	loan_status
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	189640
Number of Observations Used	189640

Response Profile		
Ordered Value	loan_status	Total Frequency
1	1	32992
2	0	156648

*Probability modeled is loan\_status=1.*

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	175278.04	65946.294
SC	175288.19	66017.364
-2 Log L	175276.04	65932.294

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	109343.746	6	<.0001
Score	57661.0429	6	<.0001
Wald	23426.8989	6	<.0001

*The SAS System**The LOGISTIC Procedure*

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.2735	0.0409	6403.5498	<.0001
revol_util	1	0.00280	0.000421	44.2253	<.0001
int_rate	1	0.2259	0.00285	6287.2603	<.0001
total_pymnt	1	-0.00025	2.402E-6	10782.1770	<.0001
ins_to_inc	1	373.6	3.6865	10272.4193	<.0001
last_pymnt_amnt	1	-0.00139	0.000015	8193.5328	<.0001
inq_last_6mths	1	0.0969	0.00952	103.6554	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
revol_util	1.003	1.002	1.004
int_rate	1.253	1.246	1.260
total_pymnt	1.000	1.000	1.000
ins_to_inc	>999.999	>999.999	>999.999
last_pymnt_amnt	0.999	0.999	0.999
inq_last_6mths	1.102	1.081	1.122

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	96.6	Somers' D	0.935
Percent Discordant	3.1	Gamma	0.937
Percent Tied	0.2	Tau-a	0.269
Pairs	5168130816	c	0.968

*The SAS System*

Obs	rank	count	percentage
1	0	5357	0.61265
2	1	3295	0.37683
3	2	83	0.00949
4	3	1	0.00011
5	4	0	0.00000
6	5	1	0.00011
7	6	0	0.00000
8	7	2	0.00023
9	8	3	0.00034
10	9	2	0.00023