INTERPRETATION OF CREDIT CARD DATA

DETAILS

Interpreting the credit card data to understand the behavior of dependent variable using Tobit model, Selection model and Survival Analysis

Course

Predictive Analytics with SAS

About the Dataset

Dataset consists of 7401 credit card customers. Of these credit card customers some are inactive (i.e., have never used the card) and the rest are active. We have the following variables.

1. The mode of acquisition (whether they were acquired through direct mail (DM), direct selling (DS), telephone sales (TS) or through internet (NET))
2. Whether they have a Reward card (i.e., a card that gives points for every dollar purchased)
3. Whether they have an affinity card and the type of affinity card they have.
4. The type of card that they were given (that is, whether they have a standard, gold, platinum or quantum card). Note: Quantum > Platinum > Gold > Standard card in terms of credit worthiness.
5. Note that profit = totfc + 1.6%*TotalTrans (approximately)

	HID	ID of the account
	Rewards	whether the customer has a reward card (=1) or not (=0)
	Limit	credit limit of the customer
	numcard	number of cards that the customer has from this bank
Mode of acquisition	**DM**	whether the customer was acquired through direct mail (1=Yes, 0=No)
	DS	whether the customer was acquired through direct selling (1=Yes, 0=No)
	TS	whether the customer was acquired though telephone selling (1=Yes, 0=No)
	NET	whether the customer was acquired though internet (1=Yes, 0=No)
Type of card	**Gold**	whether the customer has a GOLD card (1=Yes, 0=No)
	Platinum	whether the customer has a PLATINUM card (1=Yes, 0=No)
	Quantum	whether the customer has a QUANTUM card (1=Yes, 0=No)
	Standard	whether the customer has a STANDARD card (1=Yes, 0=No)
	Totaltrans	Total transaction amount (money spent) by the customer over a 3 year period
	Totfc	Total finance charges paid by the customer over a 3 year period
	Age	Age in years
	Dur	Duration: Number of months a customer has stayed with the firm
Types of Affinity cards	**sectorA**	No affinity – card is not associated with affinity to an organization
	SectorB	Affinity card affiliated with Professional organization (e.g. Am. Medical. Assoc) if a customer has an affinity card of this type value =1 else 0.
	SectorC	Affinity card affiliated with Sports
	SectorD	Affinity card affiliated with Financial institution
	SectorE	Affinity card affiliated with University (e.g. UTD card)
	SectorF	Affinity card affiliated with Commercial (e.g. Macy's card)

1. If profit is negative, set it to 0, since profit cannot be negative.
2. IF TOTTRANS=0 THEN CREATE A NEW VARIABLE CALLED ACTIVE THAT TAKES THE VALUE=0, ELSE IT TAKES THE VALUE =1. Only active customers have positive transactions.
3. Create a new variable climit = limit/10000

4. Create a new variable ttrans=tottrans/10000;
5. Create a new variable profit that is = totfc + 1.6%*TotalTrans

1. Tobit model

Model profit = age ttrans rewards climit numcard, modes of acquisition, type of card, types of affinity

a. Write a summary of the results. Focus on important effects, interpretation, model fit etc.

Tobit Model:

The SAS System

The QLIM Procedure

Summary Statistics of Continuous Responses							
Variable	Mean	Standard Error	Type	Lower Bound	Upper Bound	N Obs Lower Bound	N Obs Upper Bound
profit	1068.474	1686.534704	Censored	0		2424	

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	profit
Number of Observations	7401
Log Likelihood	-45918
Maximum Absolute Gradient	0.00361
Number of Iterations	100
Optimization Method	Quasi-Newton
AIC	91872
Schwarz Criterion	91997

dm =	Intercept – ds – ts – net
platinum =	Intercept – standard – gold – quantum
sectorA =	Intercept – sectorB – sectorC – sectorD – sectorE – sectorF

Algorithm converged.

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	1752.347177	131.582172	13.32	<.0001
age	1	-18.598025	1.845423	-10.08	<.0001
ttrans	1	475.310306	12.492710	38.05	<.0001
rewards	1	-315.862853	70.881843	-4.46	<.0001
climit	1	-93.069731	34.204368	-2.72	0.0065
numcard	1	56.901968	51.638740	1.10	0.2705
dm	0	0	.	.	.
ds	1	-1026.593099	99.628843	-10.30	<.0001
ts	1	-1047.398116	56.258325	-18.62	<.0001
net	1	-180.050098	106.342454	-1.69	0.0904
standard	1	-651.927976	74.239845	-8.78	<.0001
gold	1	-640.588025	185.542294	-3.45	0.0006
platinum	0	0	.	.	.
quantum	1	-1134.879125	108.718217	-10.44	<.0001
sectorA	0	0	.	.	.
sectorB	1	229.052900	87.433330	2.62	0.0088
sectorC	1	-173.630081	103.587380	-1.68	0.0937
sectorD	1	-94.496050	92.653464	-1.02	0.3078
sectorE	1	-283.807156	92.379496	-3.07	0.0021
sectorF	1	-390.186354	82.890590	-4.71	<.0001
_Sigma	1	1848.645741	19.119082	96.69	<.0001

Interpretation:

- In the “Parameter Estimates” table there are twenty one rows. The first twenty of these rows correspond to the vector estimate of the regression coefficients. The last one is called _Sigma, which corresponds to the estimate of the error variance.

- A one unit increase in age is associated with a 18.6 point decrease in the predicted value of profit.
- A one unit increase in total transaction amount is associated with a 475.3 point increase in the predicted value of profit.
- A one unit increase in total transaction amount is associated with a 475.3 point increase in the predicted value of profit.
- When the customer has a reward card, then there is a 315.9 point decrease in the predicted value of profit when compared to a customer without a reward card.

- A one unit increase in credit limit is associated with a 93 point decrease in the predicted value of profit.
- Number of cards that a customer has from this bank does not have a significant relationship with predicted value of profit.

- The terms for modes of acquisition, Type of cards and types of affinity cards have a slightly different interpretation.
- The predicted value of profit is 1026.6 units lower for customers who were acquired through direct selling than for customers who were acquired through direct mail.
- The predicted value of profit is 1047.4 units lower for customers who were acquired through telephone selling than for customers who were acquired through direct mail.
- Customers acquired through internet were not significant against predicted value of profit when compared to customers acquired through direct mail.

- The predicted value of profit is 652 units lower for customers who have a standard card than for customers who have a platinum card.
- The predicted value of profit is 640 units lower for customers who have a gold card than for customers who have a platinum card.
- The predicted value of profit is 1134.9 units lower for customers who have a quantum card than for customers who have a platinum card.

- The predicted value of profit is 229 units higher for customers who have an affinity card affiliated with Professional organization than for customers who have no affinity card.
- Customers who have an affinity card affiliated with Sports were not significant against predicted value of profit when compared to customers who have no affinity card.
- Customers who have an affinity card affiliated with financial institution were not significant against predicted value of profit when compared to customers who have no affinity card.
- The predicted value of profit is 283.8 units lower for customers who have an affinity card affiliated with university than for customers who have no affinity card.
- The predicted value of profit is 173 units lower for customers who have an affinity card affiliated with commercial than for customers who have no affinity card.

b. Which mode of acquisition generates the highest profit?
Direct mail and internet are the modes of acquisition that generates the highest profit.

c. Order the modes of acquisition from high to low in terms of profit.
Direct mail > Internet > Direct selling > Telephone selling. This can be determined using the coefficient estimates.

2. Selection model

Model active = age, rewards, climit, numcard, modes of acquisition, type of card, types of affinity

Model totfc = age, ttrans, rewards, climit, numcard, modes of acquisition, type of card, types of affinity

The SAS System

The QLIM Procedure

Summary Statistics of Continuous Responses

Variable	N	Mean	Standard Error	Type	Lower Bound	Upper Bound	N Obs Lower Bound	N Obs Upper Bound
totfc	4977	1307.693	1743.560707	Regular				

Discrete Response Profile of active

Index	Value	Total Frequency
1	0	2424
2	1	4977

Model Fit Summary

Number of Endogenous Variables	2
Endogenous Variable	active totfc
Number of Observations	7401
Log Likelihood	-47500
Maximum Absolute Gradient	0.97633
Number of Iterations	105
Optimization Method	Quasi-Newton
AIC	95070
Schwarz Criterion	95311

sible values), since the variables are a linear combination of other variables as shown.

active.dm =	active.Intercept - active.ds - active.ts - active.net
active.platinum =	active.Intercept - active.standard - active.gold - active.quantum
active.sectorA =	active.Intercept - active.sectorB - active.sectorC - active.sectorD - active.sectorE - active.sectorF
totfc.dm =	totfc.Intercept - totfc.ds - totfc.ts - totfc.net
totfc.platinum =	totfc.Intercept - totfc.standard - totfc.gold - totfc.quantum
totfc.sectorA =	totfc.Intercept - totfc.sectorB - totfc.sectorC - totfc.sectorD - totfc.sectorE - totfc.sectorF

Algorithm converged.

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
totfc.Intercept	1	1786.598564	131.846888	13.55	<.0001
totfc.age	1	-3.975666	1.810259	-2.20	0.0281
totfc.ttrans	1	152.613418	11.743011	13.00	<.0001
totfc.rewards	1	-116.943719	70.005120	-1.67	0.0948
totfc.climit	1	-156.502287	30.953522	-5.06	<.0001
totfc.numcard	1	-62.878969	50.401153	-1.25	0.2122
totfc.dm	0	0	.	.	.
totfc.ds	1	15.551083	93.540064	0.17	0.8680
totfc.ts	1	-1.791589	24.851538	-0.07	0.9425
totfc.net	1	-117.424392	100.285628	-1.17	0.2416
totfc.standard	1	-774.079205	78.072156	-9.91	<.0001
totfc.gold	1	-338.464097	191.306547	-1.77	0.0769
totfc.platinum	0	0	.	.	.
totfc.quantum	1	-828.583913	116.034481	-7.14	<.0001
totfc.sectorA	0	0	.	.	.
totfc.sectorB	1	273.039065	87.196323	3.13	0.0017
totfc.sectorC	1	-59.277429	103.808940	-0.57	0.5680
totfc.sectorD	1	-163.999818	93.237326	-1.76	0.0786
totfc.sectorE	1	-168.219435	93.681649	-1.80	0.0726
totfc.sectorF	1	-455.416901	81.491081	-5.59	<.0001
_Sigma.totfc	1	1669.747170	16.734676	99.78	<.0001
active.Intercept	1	1.673375	0.102023	16.40	<.0001
active.age	1	-0.020175	0	.	.
active.rewards	1	-0.345130	0.055480	-6.22	<.0001
active.climit	1	0.373093	0.028233	13.21	<.0001
active.numcard	1	0.230368	0.040320	5.71	<.0001
active.dm	0	0	.	.	.
active.ds	1	-1.292270	0.069631	-18.56	<.0001
active.ts	1	-1.342287	0.042207	-31.80	<.0001
active.net	1	-0.080809	0.097463	-0.83	0.4070
active.standard	1	-0.035809	0.052850	-0.68	0.4981
active.gold	1	-0.420344	0.135673	-3.10	0.0019
active.platinum	0	0	.	.	.
active.quantum	1	-0.765891	0.071981	-10.64	<.0001
active.sectorA	0	0	.	.	.
active.sectorB	1	-0.028011	0.068327	-0.41	0.6818
active.sectorC	1	-0.217259	0.077942	-2.79	0.0053
active.sectorD	1	-0.004220	0.070458	-0.06	0.9522
active.sectorE	1	-0.215705	0.069547	-3.10	0.0019
active.sectorF	1	0.065724	0.066340	0.99	0.3218
_Rho	1	0.017705	.	.	.

a. Write a summary of the results. Focus on important effects, interpretation, model fit etc.

The Customer is active, total transaction is above zero.

- A one unit increase in age is associated with 3.97 units decrease in the predicted value of total financial charge.
- A one unit increase in total transaction is associated with 152.61 unit increase in the predicted value of total financial charge.
- Customer with a reward card is insignificant at 5% level in determining the total financial charge.
- A one unit increase in credit limit is associated with 156.52 unit decrease in the predicted value of total financial charge.
- Number of cards is not significant at 5% level in determining the total financial charge.
- Each mode of acquisition is not significant at 5% level in determining the total financial charge.
- The predicted value of total financial charge is 774.07 dollars lower for customer who has standard card than for customer who has platinum card.
- The predicted value of total financial charge is 828.58 dollars lower for customer who has a quantum card than for customer who has a platinum card.
- Customer who has gold card is not significant against customer who has platinum card.
- The predicted value of total financial charge is 273.03 dollars higher for customer who has Affinity card affiliated with Professional organization than for customer who has no affinity card. The predicted value of total financial charge is 455.41 lower for customer who has Affinity card affiliated with Commercials than for customer who has no affinity card. Customer who has Affinity card affiliated with Sports, financial intuition and university are not significant at 5% level against customer who has no affinity card.
- The value of rho is small and not significant, so the selection bias is not a big problem in the estimation of total financial charge.

b. Which mode of acquisition generates the total financial charge?

\card affiliated with Professional organization, Sports, financial intuition and university have highest total financial charge.

c. Order the Affinity card from high to low in terms of total financial charge:

Professional organization =Sports =financial intuition = university> No card > Commercials.

3. Survival analysis

Note that duration is censored if its value is 37 as we have only 37 months of data.
Create a new variable Censor which takes the value=1 if dur=37 (the maximum value)
and value=0 otherwise. Use this as a censoring variable.

1. **Run a proportional hazards model (PROC PHREG)**

Duration = age, ttrans, rewards, climit, numcard, modes of acquition, type of card, types of affinity

Write a summary of the results. Focus on important effects, interpretation, model fit etc.

Model Information	
Data Set	WORK.A1
Dependent Variable	dur
Censoring Variable	censor
Censoring Value(s)	1
Ties Handling	BRESLOW

Number of Observations Read	7401
Number of Observations Used	7401

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
7401	4308	3093	41.79

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

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	74842.297	73604.214
AIC	74842.297	73636.214
SBC	74842.297	73738.106

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1238.0827	16	<.0001
Score	725.9541	16	<.0001
Wald	780.8872	16	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
age	1	0.0004637	0.00113	0.1684	0.6815	1.000
ttrans	1	-0.44892	0.02069	470.7059	<.0001	0.638
rewards	1	0.04670	0.04839	0.9315	0.3345	1.048
climit	1	-0.14929	0.02689	30.8133	<.0001	0.861
numcard	1	-0.03644	0.03686	0.9775	0.3228	0.964
dm	1	-0.06645	0.08320	0.6380	0.4244	0.936
ds	1	0.10547	0.09089	1.3465	0.2459	1.111
ts	1	0.06354	0.08362	0.5774	0.4473	1.066
net	0	0	.	.	.	.
standard	1	-0.34545	0.07464	21.4171	<.0001	0.708
gold	1	-0.28489	0.13430	4.5003	0.0339	0.752
platinum	1	-0.10218	0.06034	2.8683	0.0903	0.903
quantum	0	0	.	.	.	.
sectorA	1	0.29082	0.05819	24.9768	<.0001	1.338
sectorB	1	0.09376	0.06258	2.2442	0.1341	1.098
sectorC	1	0.12288	0.06728	3.3358	0.0678	1.131
sectorD	1	0.05986	0.06261	0.9142	0.3390	1.062
sectorE	1	0.08807	0.06144	2.0547	0.1517	1.092
sectorF	0	0	.	.	.	.

Interpretation:

- If total transaction is increased by one unit, the hazard decreases by 36.2%
- If credit limit is increased by one unit, the hazard decreases by 13.9%
- If the customer has the standard card, the hazard decreases by 29.2% than the customer has quantum card.
- If the customer has no affinity card, the hazard increases by 33.8 % than the customer has affinity card affiliated with Commercial.
- All other variables are not significant at 5 % level.

  R-square= 1.6%

4. Run the same model as above using PROC LIFEREG with Weibull distribution.

Write a summary of the results. Focus on important effects, interpretation, model fit etc.

Survival Analysis: Lifereg Model:

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.5278	0.0290	3.4711	3.5846	14830.9	<.0001
age	1	-0.0003	0.0003	-0.0008	0.0002	1.17	0.2799
ttrans	1	0.1200	0.0051	0.1101	0.1299	563.21	<.0001
rewards	1	-0.0152	0.0111	-0.0368	0.0065	1.88	0.1706
climit	1	0.0329	0.0061	0.0209	0.0448	29.17	<.0001
numcard	1	0.0088	0.0084	-0.0076	0.0253	1.11	0.2916
dm	1	0.0151	0.0190	-0.0222	0.0523	0.63	0.4275
ds	1	-0.0435	0.0207	-0.0840	-0.0029	4.42	0.0356
ts	1	-0.0333	0.0191	-0.0707	0.0040	3.06	0.0804
net	0	0.0000	.	.	.	.	.
standard	1	0.0874	0.0171	0.0540	0.1209	26.25	<.0001
gold	1	0.0641	0.0307	0.0040	0.1242	4.37	0.0367
platinum	1	0.0297	0.0138	0.0026	0.0567	4.63	0.0314
quantum	0	0.0000	.	.	.	.	.
sectorA	1	-0.0689	0.0133	-0.0950	-0.0428	26.80	<.0001
sectorB	1	-0.0229	0.0143	-0.0510	0.0052	2.54	0.1108
sectorC	1	-0.0294	0.0154	-0.0595	0.0007	3.66	0.0557
sectorD	1	-0.0128	0.0143	-0.0409	0.0153	0.79	0.3733
sectorE	1	-0.0233	0.0141	-0.0509	0.0043	2.75	0.0973
sectorF	0	0.0000	.	.	.	.	.
Scale	1	0.2284	0.0033	0.2220	0.2350		
Weibull Shape	1	4.3778	0.0638	4.2544	4.5047		

Interpretation:

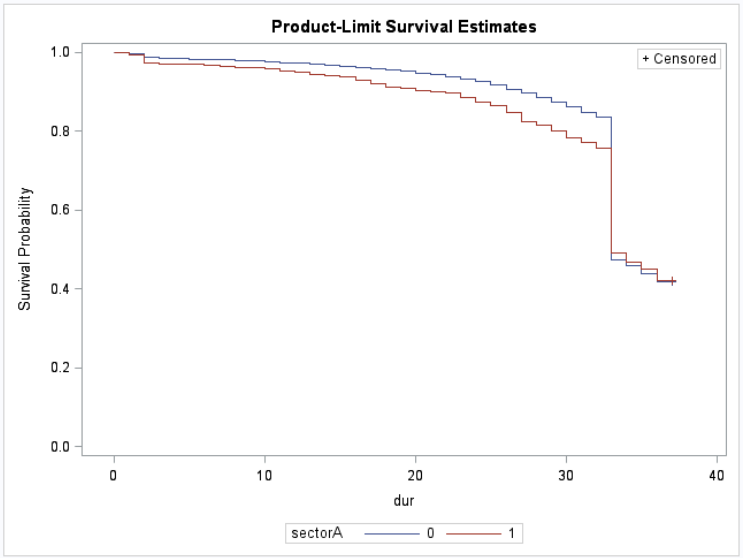
- If total transaction is increased by one unit, the survival time increases by 12%.
- If credit limit is increased by one unit, the survival time increases by 3.29%.
- If the customer has the standard card, the survival time increases by 8.74 % than the customer has quantum card.

- If the customer has no affinity card, the survival time decreases by 6.89 % than the customer has affinity card affiliated with Commercial.
- All other variables are not significant at 5 % level.

5. **Use PROC LIFETEST to test whether survivor function of affinity groups are significantly different from that of non-affinity groups. (that is compare sectorA with other sectors)**

What do you conclude?

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	3.5093	1	0.0610
Wilcoxon	5.0690	1	0.0244
-2Log(LR)	0.4856	1	0.4859



Log-Rank test shows that the survivor function of affinity groups are not significantly different from that of non-affinity groups.