Sixth Homework- Data Analysis Course of Professor Murthi on Logit, Probit, Tobit

Meisam Hejazinia

04/08/2013

Rewards	Freq.	Percent
0	7172	79.69
1	1828	20.31

Table 1: Reward card owner percentage

Affinity	Freq.	Percent
0	1500	16.67
1	7500	83.33

Table 2: Affinity card owner

Question 1
As is shown in table 1, aroud 80% of customers do not have any reward card, and only around 20% have reward card.

Question 2
As shown in table 2 around 83% of customers have affinity card.

Question 3
Average credit limit of customers was $12,255, and their average profit was $1,108 during last three years. Table 3 is showing this.

Question 4
Most popular mode of acquisition is direct mail. Also telephone sales has only small distance from it. Table 4 is showing this.

Variable	N	Mean	Std
limit	9000	12,255.58	8082.87
profit	9000	1,108.57	1773.82

Table 3: Average profit and average credit limit

Acquis.	Feq.	Percent
direct mail (DM)	3745	41.6
direct selling (DS)	1187	13.19
telephone sales (TS)	3609	40.1
internet (NET)	459	5.1

Table 4: Mode of acquisition

Question 5
Model of profit regression is significant. Table 5 shows around 60% of the variation of profit is explained by total fee, affinity, reward, limit, number of cards, mode of acquisitions, and type of card. Among them, as table 6 is showing, almost every variable is significant in 95% confidence interval. Total fee followed by credit limit is the most effective factor on profit. It is kind of obvious, since usually profitable customers pay the highest fee, and are given with the higher credit limit. Also I think credit limit is indogeneous variable. Total fee, credit limit, number of cards, direct mail acquisition, gold and plantinum cards have positive effect on profit, and reward cards, affinity cards, direct selling, telephone selling, and quantum cards have negative effect on profit. Affinity card has the highest magnitude of negative effect on the profit.

Question 6
Total fee has positive effect on profit. The magnitude is high, and it is significant, according to table 6. It is saying that company is getting higher profit from customers who have paid higher fee during last three years, which sounds reasonable.

Variable	Estimate	Mean	Effect	St. Err.	$Pr > \|t\|$
Intercept	-261.77081			81.80736	0.0014
totalfee	4.64973	157.40	731.8457771	0.04565	$< .0001$
Affinity	-103.91844	0.83	-86.59869654	38.14113	0.0065
Rewards	-192.01504	0.20	-39.00038599	34.77634	$< .0001$
c. limit	0.0395	12255.58	484.09541	0.00169	$< .0001$
numcard	54.5348	1.25	68.37452157	27.6296	0.0484
dm	189.86385	0.42	79.00445547	59.45313	0.0014
ds	-146.52387	0.13	-19.32487204	65.5146	0.0253
ts	-122.37192	0.40	-49.07113992	59.52196	0.0398
gold	246.73063	0.02	4.002513626	101.74829	0.0153
platinum	275.79347	0.75	207.4886114	37.26704	$< .0001$
quantum	-134.73906	0.06	-8.518499799	62.07365	0.03

Table 6: Effect of regressors on profit

F value	$< .0001$
Root MSE	1160.58209
Dependent Mean	1110.4068
Coeff Var	104.51864
R-Square	0.5727
Adj R-sq	0.5722

Table 5: Profit regression on credit card variabels

Criterion	Intercept Only	Inter. & Cov.
AIC	10571.544	8852.148
SC	10578.649	8930.302
-2 Log L	10569.544	8830.148

Table 7: Logit model fit statistics

Effect	W. Chi-Sq.	$Pr > ChiSq$
Affinity	0.09	0.7699
Rewards	36.32	$< .0001$
limit	175.45	$< .0001$
numcard	16.74	$< .0001$
dm	0.25	0.6176
ds	122.85	$< .0001$
ts	160.63	$< .0001$
gold	5.45	0.0195
platinum	10.88	0.001
quantum	130.33	$< .0001$

Table 8: Analysis of affects on choice of being active customer

Question 7

As table 7 shows model fit of logit based on Akaike information criteria, Shwarts bayesian criteria, and $-2log(L)$, intercept and covariates model does better job in explaining the choice to be active customers than intercept only model, since its AIC, and SC are lower.

Question 8

Wald chi square shows that effects of Rewards, credit limit, number of credit cards, direct selling and telephone selling acquisition method, and gold, platinum, and quantum card on the odds of being active is significant. In other word out of the model only direct mail, and affinity card did not have significant effect on odds of being active customer. The confidence interval was 95%. The result is shown in table 8.

Question 9

As table 9 is showing direct mail which turned out not to be significant in the whole model had estimated magnitude of .0436 indicating positive effect on log odds of being active.

Question 10

In your proposed model you have not put the internet channel of acquistion, so I assumed that in this question you meant which out of three channels of direct

Parameter	Estimate	$Pr > Chi.Sq$
Intercept	-1.6681	$< .0001$
Affinity	-0.0128	0.7699
Rewards	0.2303	$< .0001$
limit	0.000058	$< .0001$
numcard	0.2566	$< .0001$
dm	0.0436	0.6176
ds	0.9737	$< .0001$
ts	1.0685	$< .0001$
gold	0.2538	0.0195
platinum	0.1293	0.001
quantum	0.7345	$< .000$

Table 9: Maximum likelihood estimates

parameter	Estimate	mean	effect
dm	1.091	0.42	0.45397721
ds	7.011	0.13	0.924673078
ts	8.475	0.40	3.398475

Table 10: Effect of mode of acuisition on being active odds

mail, direct selling, and telephone selling. According to table 10 telephone selling has highest effect on odds of being active. I think it is indogeneous, since the company has acquired this customer with the most costly channel means telephone selling, so they have expected them to be more active. On the other hand it also could have been the result of special treatment of these customers, yet both of these are confounde, and needed to be teased apart.

Question 11

Since you have the general comment that Q7-Q16 should be responded based on logit model, I assumed this question means which type of card has greater effect on the odds of being active. According to table 11 quantum card has greates effect on the odds of being active in comparison with the other cards. If in the question you meant which has the highest frequency the answer is platinum card, but again this one could be indogeneous, due to the decision of managers of company to make the cards available to different people, based on company's decided policy.

Question 12

parameter	Estimate	mean	effect
gold	0.2538	0.02	0.004
platinum	0.1293	0.75	0.097
quantum	0.7345	0.06	0.046

Table 11: Effect of type of card on log odds of being active

Intercept only (log)	10569.544
Int. and covar (log)	8830.148
Ratio(LL/LL0)	0.835433203
R-sq	0.164566797

Table 13: Logit model of activeness goodness of fit

According table 12 having reward card compared to not having it will affect the odds of having active card to not having it positively with magnitude of 1.585.

Question 13

According to table 12 having affinity card versus not having it will affect negatively on log of odds, mean the odds of being active will decrease by $(1 - .975) * 100\%$, mean 2.5 percent.

Question 14

As table 13 shows model is doing poor job in explaining whether the customer is active or not, since only 16% of the variation in making the choice is explained according to the $R^2 = 1 - \frac{LL}{LL_0}$ that is calculated from the model.

Question 15

As table 14 shows using $ln(odd) = \alpha + X.\beta$, and $p = \frac{o}{o+1}$ the probability of the customer being active was calculated .62.

Question 16

Elasticity of probability of being active with respect to affinity is $(1-pr(active))*c_{af}*af$, and with respect to reward is $(1-pr(active))*c_{rw}*rw$, where c_{af}, c_{rw} are estimated values for coefficients of affinity and reward, and rw, af are affinity and reward at specific point. To calculated this I used the mean as data point, yet I am not sure whether mean here makes sense, since the probability of having affinity card and reward card is used as their mean. Table 15 is showing this calculation.

Effect	estimate	Upper C. I	Lower C. I.
Rewards 0 vs 1	1.585	1.365	1.841
Affinity	0.975	0.822	1.157

Table 12: Odds ratio of reward car.d

	coeff	point value	prob.	elasticity
affinity	-0.0128	0.83	0.522	-0.0051
reward	0.2303	0.20	0.522	0.0224

Table 15: Elasticity at the mean

Parameter	Estimate	cust	affect
Intercept	-1.6681	1	-1.6681
Affinity	-0.0128	0	0
Rewards	0.2303	0	0
limit	0.000058	10000	0.58
numcard	0.2566	1	0.2566
dm	0.0436	0	0
ds	0.9737	0	0
ts	1.0685	1	1.0685
gold	0.2538	1	0.2538
platinum	0.1293	0	0
quantum	0.7345	0	0
$\ln(\text{odd}) = \alpha + \beta.X$	0.4908	odds	1.6336
$p = \frac{o}{o+1}$	0.62029		

Table 14: Calculate probability for single customer

Number of end var.	1
Endogenous	profit
Number of obs	8987
Missing	13
Log	-76164
Max absol. Gradient	3.02E-10
Number of iteration	0
Optimization method	Quasi-Newton
AIC	152355
Schwarz Criterion	152447

Table 17: Tobit model fit

Question 17

The probit estimate is shown in table 16. As could be seen still affinity and direct mail are not significant, consistent with logit model. I caclulated $\frac{c_{probit}}{c_{logit}}$, and for most cases it was around .58, but for insignificant ones it is different. Difference on insignificant ones could be attributed to not converging.

Question 18

The result of running tobit model is shown in table 17. The data does not show up to be censored here, since we also have negative profit, and we do not have high density on extremes. I compared the data item by item with result of normal regression, and since the table was completely the same I did not bring it.

Question 19

Table 18 shows result of running selection model of heckman on data. AIC has decreased, showing better fit of the model, in comparison to tobit model. Table 19 shows the estimated effects. Due to self selection almost everything in the probit and logit model was biased. The positive effects after applying the selection model converted to negative for example for number of cards, credit limit, direct mail, direct selling, and telephone selling acquiring channel. The same effect could also be seen for type of card, mean they had negative effect rather than positive.

Parameter	Estimate	Std Error	$Pr > Chi - sq$	$Probit/logit$
Intercept	-0.95	0.1042	$< .0001$	0.57
Affinity	0.00018	0.0249	0.9942	-0.01
Rewards	0.1346	0.0221	$< .0001$	0.58
limit	0.000034	2.49E-06	$< .0001$	0.59
numcard	0.1472	0.0362	$< .0001$	0.57
dm	0.0354	0.0453	0.4351	0.81
ds	0.5616	0.0466	$< .0001$	0.58
ts	0.621	0.0441	$< .0001$	0.58
gold	0.1359	0.063	0.0309	0.54
platinum	0.0681	0.0231	0.0032	0.53
quantum	0.4272	0.0378	$< .0001$	0.58

Table 16: Probit estimates and logit ratio

Parameter	Estimate	STD err.	$Pr > \|t\|$	$Prev.Mdl$
profit.Intercept	-320.115889	103.430455	0.002	-261.77081
profit.totalfee	4.486566	0.056341	$< .0001$	4.64973
profit.Affinity	-116.771477	49.908109	0.0193	-103.91844
profit.Rewards	-232.056135	45.761686	$< .0001$	-192.01504
profit.limit	0.040859	0	.	0.0395
profit.numcard	42.122587	35.769686	0.239	54.5348
profit.dm	169.935355	72.332577	0.0188	189.86385
profit.ds	-104.227907	80.30452	0.1943	-146.52387
profit.ts	-67.772401	68.02506	0.3191	-122.37192
profit.gold	345.916617	141.363014	0.0144	246.73063
profit.platinum	469.617034	51.798037	$< .0001$	275.79347
profit.quantum	8.987919	4:57	0.9215	-134.73906
Sigma.profit	1337.504775	11.704989	$< .0001$	
Active.Intercept	1.030208	0.113832	$< .0001$	-0.95
Active.Affinity	-0.001638	0.050255	0.974	0.00018
Active.Rewards	-0.270283	0.044698	$< .0001$	0.1346
Active.limit	0.000033921	0	.	0.000034
Active.numcard	0.147986	0.036253	$< .0001$	0.1472
Active.dm	-0.067236	0.09096	0.4598	0.0354
Active.ds	-1.118828	0.093701	$< .0001$	0.5616
Active.ts	-1.238611	0.088689	$< .0001$	0.621
Active.gold	-0.28191	0.127503	0.027	0.1359
Active.platinum	-0.136494	0.04605	0.003	0.0681
Active.quantum	-0.854464	0.075873	$< .0001$	0.4272
Rho	0.024559	.	.	

Table 19: Selection model

Number of end var.	2
Endogenous	active profit
Number of obs	8988
Missing	12
Log	-60615
Max absol. Gradient	25.01239
Number of iteration	52
Optimization method	Quasi-Newton
AIC	121281
Schwarz Criterion	121459

Table 18: Selection model fit