

New Age Marketing: Past Life Regression versus Logistic Regression

C. Olivia Rud, Providian Direct Insurance, Frazer, PA

ABSTRACT

We don't need to examine your past lives to determine your likelihood to buy an insurance policy from us. The ability to identify and measure certain characteristics about your current incarnation will allow us to measure your propensity to purchase our products. Demographic features as well as financial and lifestyle characteristics can easily be modeled using PROC LOGISTIC to calculate your individual probability of purchase behavior.

INTRODUCTION

Increasing competition in the field of direct marketing has forced companies to adopt methods that improve their efficiency. While some companies may hire astrologers or fortune tellers to guide their strategies, many are embracing the 'Engineering Approach' to direct marketing. This involves the development and implementation of sophisticated segmentation and predictive models which allow a company to calculate a profit measure for each perspective customer.

The purpose of this paper is to detail the steps involved in building a simple logistic model and interpreting the results for decision making in a direct marketing application. It begins with a definition of the logistic model and a comparison to other types of models. The next steps describe the model building process. This involves defining the objective function, preparing the independent variables and processing the model. The final steps will explain model evaluation and validation.

CONCEPTS AND DEFINITIONS

The distinguishing feature of a logistic model is its ability to use continuous variables to predict the probability of a discrete response, i.e. the dependent variable is categorical. It is often confused with a logit model which also predicts the probability of a discrete response. However, the logit model uses only categorical independent variables.

The log-linear model is another type that is often confused with logistic and logit models. Technically, a log-linear model does not distinguish between the dependent and independent variables. Since all variables are categorical, it is often a preliminary step to logit model building.

The most common use of logistic regression is one in which the dependent variable is binary or dichotomous. The logistic model calculates a probability of an

occurrence, which is a continuous value. The difference between logistic regression and linear regression is the selection of a parametric model and in the assumptions. Once these differences are handled, the methods employed in the analysis using logistic regression are very similar to those used in linear regression.

The major characteristics which differentiate logistic regression from linear regression are as follows:

- 1) The conditional mean of the logistic regression model must be bounded between 0 and 1.
- 2) The distribution of the errors is binomial.
- 3) The estimation is based on an iterative method called maximum likelihood.

THE DATA

To predict the performance of future insurance promotions, data is selected from a previous campaign consisting of about 200,000 offers. To create a validation dataset, the file is split in half using the following SAS® code:

```
DATA LIB.MODEL LIB.VALID;
  SET LIB.DATA;
  IF RANUNI(0) < .5 THEN OUTPUT LIB.MODEL;
  ELSE OUTPUT LIB.VALID;
RUN;
```

The descriptive variables are as follows:

GENDER: Male, Female, Unknown
 AGE: Numeric value
 STATE: Of residence
 SUNSIGN: Sign of the Zodiac

The behavioral variables are as follows:

MODE: Frequency of payment on current policy.
 METHOD: Method of payment (Mail, Credit Card)
 POLCYAGE: Age of current policy in months.
 PREM: Annual Premium

OBJECTIVE FUNCTION

Determining what you want to predict is the single most important step in the model building process. Consider the following choices:

- 1) Probability of response.
- 2) Probability of approval.
- 3) Probability of a payment received.
- 4) Probability of a continued relationship
- 5) Probability of a claim.

Within these possibilities is the choice of selecting them singly or in sequence.

Modeling the probability of a 'Paid Sale' can be accomplished by treating all the unpaid responders like non-responders. This is the most efficient method and will usually produce a strong model. However, an alternative method is recommended when the non-paying responders look more like responders than the non responders or look different from both groups. This method involves two steps. 1) Calculate the probability of response. 2) Using only the responders, calculate the probability that a responder will be accepted and become a 'Paid Sale'. This method takes advantage of Bayes Theorem:

$P(A)$ = Probability of response.

$P(B)$ = Probability of payment received.

So we can write:

$P(A \& B) = P(A) \times P(B|A)$

This says that the probability of two sequential events occurring is the probability of the first event occurring, times the probability of the second event occurring, given the first one has occurred.

To determine which technique will work best, create a variable called OUTCGRP (Outcome Group). Define it to have three values: NRSP (Non Responder), RESP (Responder who did not buy), and PDSL (Paid Sale). Perform analysis on OUTCGRP by the categorical variables with limited levels using the following code:

```
PROC FREQ DATA=LIB.MODEL;
WHERE OUTCGRP IN ('NRSP','RESP');
TABLE OUTCGRP*(GENDER MODE METHOD)
/MISSING NOPERCENT NOROW CHISQ;
TITLE "Non Responder Versus Non Paid Responder";
RUN;
```

```
PROC FREQ DATA=LIB.MODEL;
WHERE OUTCGRP IN ('PDSL','RESP');
TABLE OUTCGRP*(GENDER MODE METHOD)
/MISSING NOPERCENT NOROW CHISQ;
TITLE 'Non Paid Responder Versus Paid Sale';
RUN;
```

If the chi square is significant for all variables in both frequencies, two models should be considered. If only some of the differences are significant, examination of the continuous variables can assist in your decision.

A univariate logistic regression on the continuous variables can determine which groups are most different. The first step is to code the OUTCGRP variable in a numeric form:

```
DATA LIB.MODEL;
SET LIB.MODEL;
IF OUTCGRP = 'PDSL' THEN LEVEL1 = 1;
ELSE IF OUTCGRP = 'RESP' THEN LEVEL1 = 0;
ELSE LEVEL1 = .;
IF OUTCGRP = 'RESP' THEN LEVEL0 = 1;
ELSE IF OUTCGRP = 'NRSP' THEN LEVEL0 = 0;
ELSE LEVEL0 = .;
RUN;
```

```
PROC LOGISTIC DATA=LIB.MODEL DESCENDING;
MODEL LEVEL1 = AGE;
RUN;
```

```
PROC LOGISTIC DATA=LIB.MODEL DESCENDING;
MODEL LEVEL0 = AGE;
RUN;
```

NOTE: The descending option forces the procedure to predict the probability of LEVE1 = 1;

Run these models for each continuous variable. The significance of the -2 log likelihood for the two models will assist you in deciding which groups are most similar. More detail on this decision will be provided in the presentation.

Since the alternative method involves building two logistic models using similar techniques, this paper will focus on the first method. The dependent variable is coded as follows:

```
DATA LIB.MODEL;
SET LIB.MODEL;
IF OUTCOME = 'PDSL' THEN ACCEPT = 1;
ELSE ACCEPT = 0;
RUN;
```

VARIABLE PREPARATION

Once the objective function has been established, the variables must be examined for suitability as predictors. Continuous variables may need to be transformed to achieve linearity and/or segmented to improve fit. Categorical variables may need to be smoothed or grouped and defined as indicator variables.

An effective technique for linearizing continuous variables is to break the continuous variable into 4-5 groups. Before you can determine logical groups you should perform a PROC UNIVARIATE on each variable to find the distribution. Once you have determined the best groupings you calculate the logit of each group. Then plot the logit versus the group mean. To further clarify, consider the following code:

```
DATA LIB.MODEL;
SET LIB.MODEL;
IF AGE <= 50 THEN GRP1 = 1; ELSE GRP1 = 0;
IF 50 < AGE <= 55 THEN GRP2 = 1; ELSE GRP2 = 0;
IF 55 < AGE <= 60 THEN GRP3 = 1; ELSE GRP3 = 0;
IF 60 < AGE <= 65 THEN GRP4 = 1; ELSE GRP4 = 0;
IF 65 < AGE <= 70 THEN GRP5 = 1; ELSE GRP5 = 0;
RUN;
```

```
PROC LOGISTIC DATA=LIB.MODEL DESCENDING;
MODEL ACCEPT = GRP1 GRP2 GRP3 GRP4 GRP5;
TITLE 'Age Groupings to Assess Linearity';
RUN;
```

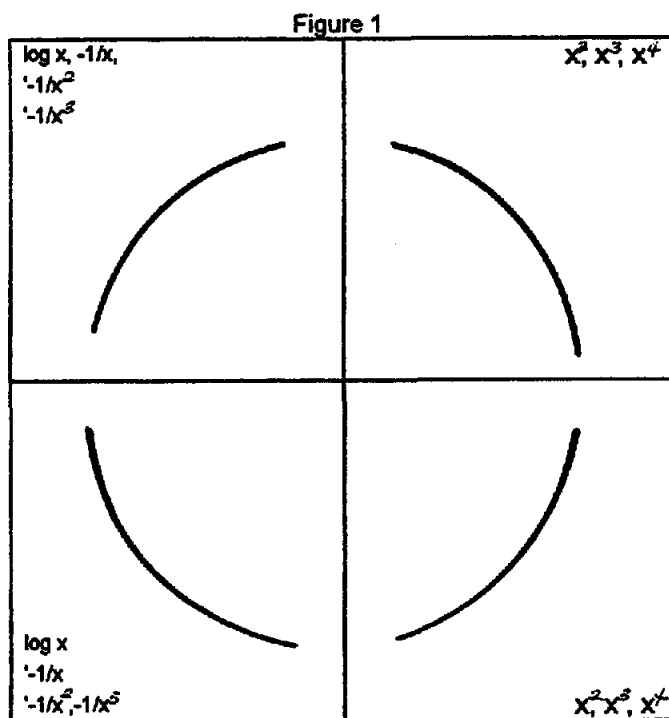
Note: Ages 70+ are treated as the referent group and are not needed in the model.

Next, run the logistic regression and input the regression coefficients into a dataset for the plot procedure as follows:

```
DATA PLOT;
INPUT MIDPOINT REGCOEF;
CARDS;
37.5 -0.2286
52.5 -0.2292
57.5 -0.1926
62.5 -0.1700
67.5 -0.1235
;
RUN;

PROC PLOT;
PLOT REGCOEF*MIDPOINT;
TITLE 'Age Curve';
RUN;
```

The plot is then examined for the shape of the curve (See Appendix A). To determine what transformation is best suited for your data consider the following Figure 1:



There are basically four types of curves. Select the shape that best resembles the curve of your data. Then follow the ladders of power (some are shown in the corners of the diagram).

Since this age appears to be linear except for the first group, create an indicator variable which separates the first group from the rest. After creating these alternative forms of the variables, you can run a PROC LOGISTIC

with the STEPWISE option to determine the form or forms of the variable to use:

```
DATA TEST;
SET LIB.MODEL;
TAGE = AGE*AGE;
TTAGE = AGE*AGE*AGE;
IF AGE < 50 THEN AGE50 = 0;
ELSE AGE50 = 1;
RUN;

PROC LOGISTIC DATA=TEST DESCENDING;
MODEL ACCEPT = AGE AGE50 TAGE TTAGE
/SELECTION=STEPWISE;
RUN;
```

The logistic regression selects three forms of in the following order: TTAGE, AGE, AGE50 (See Appendix B). Because the different transformations compliment each other, we will use all significant forms in the final model.

Again, by definition, logistic regression sees all independent variables as continuous. For categorical variables to work in the model, they must be in a form which is interpreted as continuous by the model. The best solution is to create indicator variables. This establishes a new variable for each level. If the categorical variable has more than two levels, use PROC FREQ to determine which levels have similar behavior with respect to the dependent variable.

```
PROC FREQ DATA=LIB.MODEL;
TABLE ACCEPT*(GENDER MODE METHOD STATE)
/NOPERCENT NOROW MISSING;
RUN;
```

Examination of the column percentages will allow you to see which categories have similar accept rates. Once you have determined the appropriate groupings, create indicator variables for your categorical variables or groupings. Transform your continuous variables in a data step as follows:

```
DATA LIB.MODEL;
SET LIB.MODEL;
IF GENDER = 'F' THEN IGENDER = 1;
ELSE IGENDER = 0;
IF MODE = '01' THEN HMODE = 1;
ELSE HMODE = 0;
IF MODE IN ('02','03') THEN MMODE = 1;
ELSE MMODE = 0;
IF METHOD IN ('06','12') THEN IMETHOD = 1;
ELSE IMETHOD = 0;
IF STATE IN ('AZ','CA','ID','MN','NJ','NY','OH','TX','WY')
THEN HSTATE = 1;
ELSE HSTATE = 0;
RUN;
```

NOTE: The variables names were developed using a first letter 'I' for indicator or 'H', 'M', 'L' for high, medium and low performing groups.

INTERACTIONS

If available, a CHAID (Chi Square Automatic Interaction Detection) analysis is the best method for detecting interactions. It is a decision tree methodology which based on your dependent variable, splits the population on the independent variable with the strongest difference.

In lieu of a decision tree software, a brute force method using PROC MEANS can uncover many first and second degree interactions. To uncover interaction between categorical and/or continuous variables, perform PROC MEANS on the continuous variables with a CLASS statement on the categorical variables. The following code demonstrates this:

```
PROC MEANS DATA=LIB.MODEL;
CLASS IGENDER HMODE;
VAR ACCEPT;
RUN;
```

The output allows you to compare the 'Paid Sales' rate for each combination of GENDER and MODE (See Figure 3). If this rate changes in a different direction or intensity when comparing males and females for HMODE=1 versus HMODE=0, there is a possible interaction present.

```
PROC MEANS DATA=LIB.MODEL;
CLASS GENDER ACCEPT;
VAR TTAGE LPREM LPAGE;
RUN;
```

This output allows you to look for different average values for continuous values within subgroups of the categorical variable (See Appendix C). Again you are looking for changes in direction or intensity of the continuous variables while comparing males and females among the ACCEPT=1 versus ACCEPT=0;

For interactions among continuous variables, you must create the various combinations and test them in PROC LOGISTIC. The following code demonstrates coding for all types of interactions:

```
DATA LIB.MODEL;
SET LIB.MODEL;
HMOD_GEN=HMODE*GENDER;
AGE_PREM=AGE*PREM;
RUN;
```

To test for significance, use the following code:

```
PROC LOGISTIC DATA=LIB.MODEL DESCENDING;
MODEL = HMOD_GEN;
RUN;
```

Variables can become more significant when used in combination with other variables. Therefore, unless your number of available variables is prohibitively large (> 25), keep all variables which have a -2 Log Likelihood p-value of < .50.

THE LOGISTIC PROCEDURE

If the total number of independent variables is reasonable (less than 25) you can allow the stepwise procedure to provide automatic data reduction. For example, if two independent variables are highly correlated, once the variable with the higher predictive power (with respect to the dependent variable) enters the model, the power of the other variable is greatly reduced.

The following program will create your model and an output dataset with your predicted probabilities:

```
PROC LOGISTIC DATA=LIB.MODEL;
MODEL ACCEPT = IGENDER HMODE MMODE AGE
TTAGE AGE50 HSUNSIGN LSUNSIGN TPAGE
TTPAGE LPAGE PREM LPREM IMETHOD HSTATE
LSUNPAG MMODLPAG MMODTTAG PREMTTAG
TTAGLPAG PREMLPAG / SELECTION=STEPWISE
SLE=.001 SLS=.001;
OUTPUT OUT=LIB.MODELX PRED=PREDPROB;
RUN;
```

The final output shows the order in which the variables entered the model as well as the regression coefficients (See Appendix D). The output data set is used to evaluate the model.

MODEL EVALUATION

The best method of testing your model is to create a 'Gains Table' which calculates the 'lift' achieved by selecting only the best scoring names. This involves sorting the names by the predicted probability (PREDPROB). The 'Gains Table' is created by dividing the data into deciles with the highest scoring names in the lowest deciles. The following code will create a 'Gains Table':

```
PROC SORT DATA=LIB.MODELX;
BY DESCENDING PREDPROB;
RUN;
```

```
DATA LIB.MODELX;
SET LIB.MODELX NOBS=COUNT;
TOTREC=COUNT;
RECORDS=1;
LABEL ACCEPT1='Paid Sale Rate';
LABEL RECORDS='Total Contacts';
DECILE=INT((_N_-1)/(1*TOTREC));
```

```
PROC TABULATE DATA=LIB.MODELX;
CLASS DECILE;
VAR ACCEPT1 PREDPROB RECORDS;
TABLE DECILE ALL,
RECORDS*SUM*F=COMMA9.
ACCEPT1*(SUM*F=COMMA5. MEAN*F=5.3)
PREDPROB*(MEAN MIN MAX)*F=5.3/RTS=9;
TITLE 'Stepwise Logistic on Model Data';
RUN;
```

Table 1 displays the performance by decile for the model data:

Table 1

DECILE	Total Con- tacts	Paid Sale Rate		Estimated Probability		
	SUM	SUM	MEAN	MEAN	MIN	MAX
0	9,069	3,701	0.408	0.400	0.350	0.662
1	9,068	3,093	0.341	0.330	0.312	0.350
2	9,068	2,826	0.312	0.299	0.286	0.312
3	9,069	2,657	0.293	0.275	0.264	0.286
4	9,068	2,371	0.261	0.254	0.243	0.264
5	9,068	2,243	0.247	0.232	0.221	0.243
6	9,069	2,026	0.223	0.210	0.199	0.221
7	9,068	1,720	0.190	0.185	0.171	0.199
8	9,068	1,431	0.158	0.156	0.139	0.171
9	9,068	996	0.110	0.104	0.000	0.139
ALL	90,683	23,064	0.254	0.245	0.000	0.662

From this 'Gains Table' table you can calculate the 'lift'. For example, if you chose to mail only the best performing 30%, you would capture 9,620 of the 'Paid Sales' or 41.7% of the total buyers. This provides a 'lift' of 139 $((41.7/30)*100)$. Without the model you would have mailed 41.7% of your population to capture 41.7% of the buyers. Therefore you have reduced your mailing expense by 28% $((41.7 - 30)/41.7)$.

If you chose to mail only the best performing 70%, you would capture 18,917 of the 'Paid Sales' or 82% of the total buyers. This provides a 'lift' of 117 $((82/70)*100)$. Without the model you would have mailed 82% of your population to capture 82% of the buyers. Therefore you have reduced your mailing expense by 14.6% $((82 - 70)/82)$.

With many direct mail programs generating millions of pieces annually, these savings can substantially improve profits.

VALIDATION

To insure that the model is not biased by the data, the validation data must be scored with the model parameters. Since the validation data does not have any transformed or interaction variables, these must first be created. The final steps of the program calculate the predicted probability for each record. This involves

scoring each record with the parameter estimates and using the link function

$$\frac{\exp(b_0 + b_1X + b_2X + \dots)}{1 + \exp(b_0 + b_1X + b_2X + \dots)}$$

to calculate the predicted probability. These tasks are both completed in the following program:

```
DATA LIB.VALID;
  SET LIB.VALID NOBS=NUMBER;
  TOTREC=NUMBER;
  TTAGE=AGE*AGE*AGE;
  LPAGE=LOG(PAGE);
  LPREM=LOG(PREM);
  IF STATE IN ('GA','MN','NJ','TX') THEN HSTATE=1;
  ELSE HSTATE=0;
  IF GENDER='F' THEN IGENDER=1;
  ELSE IGENDER=0;
  IF MODE IN ('03') THEN MMODE = 1;
  ELSE MMODE = 0;
  IF MODE IN ('01') THEN HMODE = 1;
  ELSE HMODE = 0;
  IF METHOD IN ('01') THEN IMETHOD = 1;
  ELSE IMETHOD = 0;
  TTAGLPAG=TTAGE*LPAGE;
  VBETA = .1526 + .1999*IGENDER + .7758*HMODE +
    .4562*MMODE - .0305*AGE - .4479*LPAGE -
    .00198*PREM + .1569*LPREM + .6263*IMETHOD +
    .5693*HSTATE + .0000007042*TTAGLPAG;
  VPRDPOB = (EXP(VBETA))/(1+ EXP(VBETA));
  RUN;
```

Next, sort the validation data by the predicted probability and create the deciles:

```
PROC SORT DATA=LIB.VALIDX;
  BY DESCENDING VPRDPROB;
  RUN;

DATA LIB.VALIDX;
  SET LIB.VALIDX NOBS=COUNT;
  TOTREC=COUNT;
  RECORDS=1;
  LABEL ACCEPT1='Paid Sale Rate';
  LABEL RECORDS='Total Contacts';
  DECILE=INT((_N_-1)/(.1*TOTREC));

PROC TABULATE DATA=LIB.MODELX;
  CLASS DECILE;
  VAR ACCEPT1 VPRDPROB RECORDS;
  TABLE DECILE ALL,
    RECORDS*SUM*F=COMMA9.
    ACCEPT1*(SUM*F=COMMA5. MEAN*F=5.3)
    VPRDPROB*(MEAN MIN MAX)*F=5.3/RTS=9;
  TITLE 'Stepwise Logistic on Validation Data';
  RUN;
```

Table 2 displays the performance by decile for the validation data:

Table 2

DECILE	Total Conta- cts	Paid Sale Rate		Estimated Probability		
	SUM	SUM	MEAN	MEAN	MIN	MAX
0	9,030	3,729	0.413	0.400	0.350	0.670
1	9,030	3,069	0.340	0.330	0.312	0.350
2	9,029	2,843	0.315	0.299	0.286	0.312
3	9,030	2,529	0.280	0.275	0.263	0.286
4	9,029	2,355	0.261	0.253	0.242	0.263
5	9,030	2,215	0.245	0.231	0.220	0.242
6	9,030	1,964	0.217	0.209	0.198	0.220
7	9,029	1,676	0.186	0.185	0.171	0.198
8	9,030	1,493	0.165	0.155	0.138	0.171
9	9,029	955	0.106	0.104	0.000	0.138
ALL	90,296	22,828	0.253	0.244	0.000	0.670

The validation data 'Gains Table' shows a similar 'lift' to the model data. At 30% of the file, the model captures 42.4% of the 'Paid Sales', or a 'lift' of 141. At 70% of the file, the model captures 81.9% of the 'Paid Sales'. This implies the model will be stable across other data sets.

CONCLUSION

Predicting behavior patterns using Past Life Regression can be enlightening and entertaining. However, to produce a statistically significant probability of behavior, logistic regression is one of the most powerful tools. The procedures detailed in this paper can provide a useful guide for attaining this goal.

In many instances, a better model is possible through the introduction of more predictive information. Attend the presentation to see if information on wealth, marital status, population density and custom clustering can improve your model.

REFERENCES

David Shepard Associates, Inc. (1995), *The New Direct Marketing*, New York: Irwin Professional Publishing.

Kass, G.V., "Significance Testing in, and Some Extension of, Automatic Interaction Detection" (doctoral dissertation, University of Witwatersrand, Johannesburg, South Africa, 1976).

Hosmer, D.W., Jr. and Lemeshow, S. (1989), *Applied Logistic Regression*, New York: John Wiley & Sons, Inc.

Mallozzi, J., (1995), "A Cosmic Consulting Firm," *New Age Journal*, (June).

SAS Institute Inc. (1989) *SAS/STAT User's Guide, Vol. 2, Version 6, Fourth Edition*, Cary NC: SAS Institute Inc.

Tukey, John W. (1977), *Exploratory Data Analysis*, Philippines: Addison-Wesley Publishing Company, Inc.

AUTHOR CONTACT

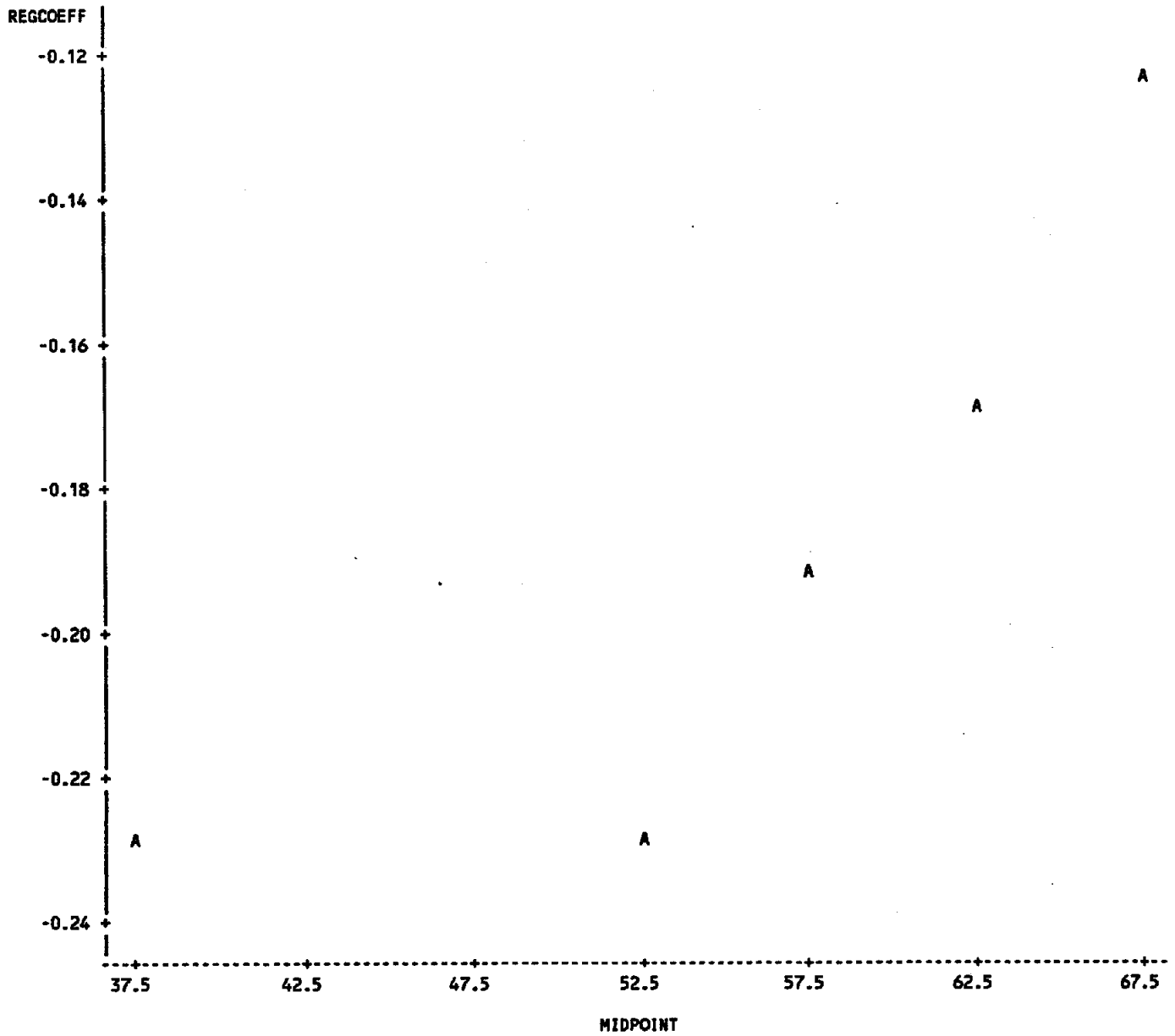
Providian Direct Insurance
20 Moores Road I-3
Frazer, PA 19355

Voice: (610) 648-4957
Fax: (610) 648-5348
Internet: orud@ccantares.wcupa.edu

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Age Curve

Plot of REGCOEFF*MIDPOINT. Legend: A = 1 obs, B = 2 obs, etc.



Test for Age Variable

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	0.00489	0.1541	0.0010	0.9747	.	1.005
AGE	1	-0.0358	0.00430	69.1751	0.0001	-0.222890	0.965
AGE50	1	0.1870	0.0460	16.5022	0.0001	0.035391	1.206
TTAGE	1	3.667E-6	3.399E-7	116.3955	0.0001	0.243031	1.000

Association of Predicted Probabilities and Observed Responses

Concordant = 50.7%	Somers' D = 0.063
Discordant = 44.3%	Gamma = 0.066
Tied = 5.0%	Tau-a = 0.024
(1559564616 pairs)	c = 0.532

Appendix C

Test for Interaction

Analysis Variable : ACCEPT1

IGENDER	HMODE	N Obs	N	Mean	Std Dev	Minimum	Maximum
0	0	21370	21370	0.1905475	0.3927421	0	1.0000000
	1	27599	27599	0.2460234	0.4307001	0	1.0000000
1	0	17349	17349	0.2519454	0.4341426	0	1.0000000
	1	24365	24365	0.3214037	0.4670249	0	1.0000000

IGENDER	ACCEPT1	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
0	0	38107	TTAGE	38107	272199.78	121917.02	8000.00	704969.00
			LAARP	38107	5.6687211	0.6492368	2.1587147	8.5713177
			LPAGE	38106	4.4437858	0.8206147	1.6094379	7.5678626
	1	10862	TTAGE	10862	278718.22	124961.97	8000.00	804357.00
			LAARP	10862	5.5618856	0.6354204	2.1610215	8.0119917
			LPAGE	10862	4.2754767	0.8706691	1.7917595	7.2730926
1	0	29512	TTAGE	29512	295957.26	115761.46	8000.00	778688.00
			LAARP	29512	5.4197567	0.6033516	2.1690537	8.3305479
			LPAGE	29512	4.2781809	0.8914011	1.6094379	7.5363639
	1	12202	TTAGE	12202	306812.70	115519.97	8000.00	704969.00
			LAARP	12202	5.4371454	0.5987212	2.8622009	7.8971403
			LPAGE	12202	4.2013860	0.9158529	1.6094379	6.8710913

Stepwise on Model Data

The LOGISTIC Procedure

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	100886.12	97296.587	.
SC	100895.53	97400.153	.
-2 LOG L	100884.12	97274.587	3609.529 with 10 DF (p=0.0001)
Score	.	.	3347.202 with 10 DF (p=0.0001)

Residual Chi-Square = 28.2103 with 10 DF (p=0.0017)

NOTE: No (additional) variables met the 0.001 significance level for entry into the model.

Summary of Stepwise Procedure

Step	Variable Entered	Variable Removed	Number In	Score Chi-Square	Wald Chi-Square	Pr > Chi-Square
1	AARPLPAG		1	1035.5	.	0.0001
2	HMODE		2	681.6	.	0.0001
3	HSTATE		3	474.0	.	0.0001
4	IMETHOD		4	394.8	.	0.0001
5	IGENDER		5	179.4	.	0.0001
6	HMODE		6	177.3	.	0.0001
7	LPAGE		7	126.4	.	0.0001
8	TTAGLPAG		8	58.8204	.	0.0001
9	AGE		9	145.6	.	0.0001
10	LAARP		10	31.3762	.	0.0001
11	AARP		11	50.1069	.	0.0001
12		AARPLPAG	10	.	1.6631	0.1972

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	0.1526	0.1856	0.6764	0.4108	.	1.165
IGENDER	1	0.1999	0.0162	152.3548	0.0001	0.054941	1.221
HMODE	1	0.7758	0.0363	456.2743	0.0001	0.211571	2.172
HMODE	1	0.4562	0.0368	153.4113	0.0001	0.120222	1.578
AGE	1	-0.0305	0.00243	156.7456	0.0001	-0.189819	0.970
LPAGE	1	-0.4479	0.0195	527.4658	0.0001	-0.214426	0.639
AARP	1	-0.00198	0.00009	481.5701	0.0001	-0.190552	0.998
LAARP	1	0.1569	0.0202	60.1640	0.0001	0.055068	1.170
IMETHOD	1	0.6263	0.0326	368.0549	0.0001	0.097233	1.871
HSTATE	1	0.5693	0.0270	443.3525	0.0001	0.082403	1.767
TTAGLPAG	1	7.042E-7	5.058E-8	193.8437	0.0001	0.241040	1.000

Association of Predicted Probabilities and Observed Responses

Concordant = 62.6%	Somers' D = 0.259
Discordant = 36.7%	Gamma = 0.261
Tied = 0.7%	Tau-a = 0.096
(1519142725 pairs)	c = 0.630