Logistic Regression

BAN 100: ASSIGNMENT 6

PROBLEM 2

Data information

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	2038
Obs in First Data Page	100
Number of Data Set Repairs	0
Filename	/tmp/SAS_work508F0000098B_localhost.localdomain/SAS_work553B0000098B_localhost.localdomain/import.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	281107
Access Permission	rw-rw-r--
Owner Name	sasdemo
File Size	128KB
File Size (bytes)	131072

Alphabetic List of Variables and Attributes					
#	**Variable**	**Type**	**Len**	**Format**	**Label**
2	GPA	Num	8	15.2	GPA
3	Program	Num	8	BEST.	Program
4	Return	Num	8	BEST.	Return
1	Student	Num	8	BEST.	Student

Question A

Write the logistic regression equation relating X1 and X2 to y.

$$P(\text{Return}=1 \mid \text{GPA Program}) = \frac{e^{\beta 0+\beta 1*GPA+\beta 2*Program}}{1+ e^{\beta 0+\beta 1*GPA+\beta 2*Program}}$$

- Here, GPA and program are independent variables it means our X1 and X2 are GPA and Program respectively
- Return is our target variable it means return is our Y

Question B

What is the interpretation of E(y) when x2 = 0? .

$$E(y) = P(\text{Return}=1 \mid GPA) = \frac{e^{\beta_0+\beta_1*GPA}}{1+e^{\beta_0+\beta_1*GPA}}$$

- X2 = program
- If It becomes zero we left with above formula

Question C

For the Lakeland data, use SAS to compute the estimated logistic regression equation.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	130.207	86.338
SC	132.812	94.153
-2 Log L	128.207	80.338

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	47.8694	2	<.0001
Score	40.3936	2	<.0001
Wald	24.4151	2	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.8926	1.7472	15.5615	<.0001
GPA	1	2.5388	0.6729	14.2362	0.0002
Program	1	1.5608	0.5631	7.6813	0.0056

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
GPA	12.664	3.387	47.351
Program	4.762	1.579	14.361

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	88.3	Somers' D	0.768
Percent Discordant	11.5	Gamma	0.770
Percent Tied	0.2	Tau-a	0.348
Pairs	2244	c	0.884

$$\hat{y} = p(y=1|x_1\ x_2) = \frac{e^{-6.8926+2.5388x_1+1.56x_2}}{1+e^{-6.8926+2.5388x_1+1.56x_2}}$$

Question D

Use the estimated logit computed in c) to estimate the probability that students with a 2.5 grade point average who did not attend the orientation program will return to Lakeland for their sophomore year. What is the estimated probability for students with a 2.5 grade point average who attended the orientation program?

Here, value of our variables are
X1: GPA = 2.5
X2: Program = 0
Y : Return = 1
Using formula of question B because X2=0

$$E(y) = P(\text{Return}=1 \mid \text{GPA}) = \frac{e^{\beta_0+\beta_1*GPA}}{1+ e^{\beta_0+\beta_1*GPA}}$$

We get 0.366 means 36.6% Probability

Question E

What is the estimated odds ratio for the orientation program? Interpret it.

Odds of return = 1 are 12.664% higher than zero for students with low GPA

And Odds of return = 1 are 4.76% higher while program = 1

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
GPA	12.664	3.387	47.351
Program	4.762	1.579	14.361

Question F

Would you recommend making the orientation program a required activity? Why or why not?

- Probability of student return to the campus who has attended the Orientation program is 0.733 which is more than 70%
- So, in my opinion Orientation program is great initiative and it should be held every intakes for new students
- Orientation makes good impression of place and make them aware of the culture which is really important for a student and also in marketing perspective for college