LOGISTIC REGRESSION ANALYSIS

Seth Howells
Concepts of Statistics II
Week# 6 Project – Logistic Regression
08/16/20

TABLE OF CONTENTS

OVERVIEW

Logistic regression analysis was performed on 3 variables on the HBAT.xls file with x_4 (region) as the nonmetric criterion, or dependent variable, and all other variables listed below as he predictors, or explanatory/independent variables. The tables provide visual understanding of the multivariate relationships.

ID	Variable	Measurement	Description	Type
x_4	Region	nonmetric	Classification	Dependent
x_{13}	Competitive Pricing	metric	Performance	Independent
x_{17}	Price Flexibility	metric	Performance	Independent

-2 Log Likelihood (-2LL): comparable to the *F test* in multiple regression – assessing model fit.

Score Statistics: measure used for selective variables in stepwise approach

Significance Level: determines whether the independent variable is statistically significant (<0.05) or not (>0.05). Reject null hypothesis if p-value is (<0.05) and conclude that at least one of the Betas is significantly different than 0.

Logistic Model: $p = \frac{1}{1+e^{-b_0+b_1x}}$

Due to the binary nature of logistic regression, x_4 (region) was split into two groups: customers in the USA/North America and customers that are outside the USA/North America. The sample of 100 included 60 observations and 40 observations for validation. The goal of logistic regression analysis is to estimate and understand the difference between the two groups in x_4 (region).

A stepwise approach was used in the analysis for a greater understanding of the variables impacting the binary dependent variable. Variables will be entered in each step and assessed before the variable added into selection. There are a few criteria that can be used to assess the entry selection: greatest reduction in -2LL, greatest Wald coefficient, or highest conditional probability, with the significance level taken into consideration.

TABLE 6-2

As part of the stepwise approach, TABLE 6-2 assesses the overall fit and it found x_{13} (Competitive Pricing) and x_{17} (Price Flexibility) to have the highest Score Statistic. Both significance levels are (<0.05). The following tables, TABLE 6-3 and TABLE 6-4, examine these two variables in the step of estimation process.

FIG 6-2

TABLE 2 Logistic Regression Base Model Results

Overall Model Fit: Goodness-of-Fit Measures		
	Value	
–2 Log Likelihood (–2*LL*)	82.108	
Variables Not in the Equation		
Independent Variables	**Score Statistic**	**Significance**
X_6 Product Quality	11.925	.001
X_7 E-Commerce Activities	2.052	.152
X_8 Technical Support	1.609	.205
X_9 Complaint Resolution	.866	352
X_{10} Advertising	.791	7
X_{11} Product Line	18.323	000
X_{12} Salesforce Image	8.622	.003
X_{13} Competitive Pricing	21.330	.000
X_{14} Warranty & Claims	.465	.495
X_{15} New Products	.614	.433
X_{16} Order & Billing	.090	.764
X_{17} Price Flexibility	21.204	.000
X_{18} Delivery Speed	.157	.692

TABLE 6-3

In the next step, x_{13} (Competitive Pricing) was the first variable in entry to the stepwise approach because Competitive Pricing has a slightly higher Score Statistic than the next highest, Price Flexibility. The p-value shows that this variable is statistically significant, and thus would indicate that the researcher should reject the null hypothesis.

TABLE 3 Logistic Regression Stepwise Estimation: Adding X_{13} (Competitive Pricing)

Overall Model Fit: Goodness-of-Fit Measures

		CHANGE IN –2LL			
		From Base Model		*From Prior Step*	
	Value	**Change**	**Significance**	**Change**	**Significance**
–2 Log Likelihood (–2LL)	56.971	25.136	.000	25.136	.000
Cox and Snell R^2	.342				
Nagelkerke R^2	.459				
Pseudo R^2	.306				

	Value	**Significance**
Hosmer and Lemeshow χ^2	17.329	.027

Variables in the Equation

Independent Variable	**B**	**Std. Error**	**Wald**	***df***	**Sig.**	**Exp(B)**
X_{13} Competitive Pricing	1.129	.287	15.471	1	.000	3.092
Constant	–7.008	1.836	14.57	1	.000	.001

B = logistic coefficient, Exp(B) = exponentiated coefficient

Variables Not in the Equation

Independent Variables	**Score Statistic**	**Significance**
X_6 Product Quality	4.859	.028
X_7 E-Commerce Activities	.132	.716
X_8 Technical Support	.007	.932
X_9 Complaint Resolution	1.379	.240
X_{10} Advertising	.129	.719
X_{11} Product Line	6 154	.013
X_{12} Salesforce Image	2.745	.098
X_{14} Warranty & Claims	.640	.424
X_{15} New Products	.344	.557
X_{16} Order & Billing	2.529	.112
X_{17} Price Flexibility	13.723	.000
X_{18} Delivery Speed	1.206	.272

Classification Mat ix

	Predicted Group Membership[c]					
	ANALYSIS SAMPLE[a]			HOLDOUT SAMPLE[b]		
	X_4 *Region*			X_4 *Region*		
A tual Group Membership	**USA/ North America**	**Outside North America**	**Total**	**USA/ North America**	**Outside North America**	**Total**
USA/ North America	19 (73.1)	7	26	4 (30.8)	9	13
Outside North America	9	25 (73.5)	34	34	26 (96.3)	27

TABLE 6-4

x_{17} (Price Flexibility) was entered in the next step. The -2LL value was reduced, which indicates a better model fit because it is a lower percent of unexplained information after the variable was selected. Significance levels were maintained below 0.05 which indicates that the variables are statistically significant.

FIG 6-4

TABLE 4 Logistic Regression Stepwise Estimation: Adding X_{17} (Price Flexibility)

Overall Model Fit: Goodness-of-Fit Measures

		CHANGE IN –2LL			
		From Base Model		*From Prior Step*	
	Value	**Change**	**Significance**	**Change**	**Significance**
–2 Log Likelihood (–2LL)	39.960	42.148	.000	17.011	.000
Cox and Snell R^2	.505				
Nagelkerke R^2	.677				
Pseudo R^2	.513				

	Value	**Significance**
Hosmer and Lemeshow χ^2	5.326	.722

Variables in the Equation

Independent Variable	**B**	**Std. Error**	**Wald**	***df***	**Sig.**	**Exp(B)**
X_{13} Competitive Pricing	1.079	.357	9.115	1	.003	2.942
X_{17} Price Flexibility	1.844	.639	8.331	1	.004	6.321
Constant	–14.192	3.712	14.614	1	.000	.000

B = logistic coefficient, Exp(B) = exponentiated coefficient

Variables Not in the Equation

Independent Variables	**Score Statistic**	**Significance**
X_6 Product Quality	.656	.418
X_7 E-Commerce Activities	3.501	.061
X_8 Technical Support	.006	937
X_9 Complaint Resolution	.693	.405
X_{10} Advertising	.091	.762
X_{11} Product Line	3.409	.065
X_{12} Salesforce Image	.849	.357
X_{14} Warranty & Claims	2.327	.127
X_{15} New Products	.026	.873
X_{16} Order & Billing	.0 0	.919
X_{18} Delivery Speed	2.907	.088

Classification Matrix

	Predicted Group Membership[c]					
	ANALYSIS SAMPLE[a]			**HOLDOUT SAMPLE[b]**		
	X_4 Region			*X_4 Region*		
Actual Group Membership	**USA/ North America**	**Outside North America**	**Total**	**USA/ North America**	**Outside North America**	**Total**
USA/ North America	25 (96.2)	1	26	9 (69.2)	4	13
Outside North America	6	28 (82.4)	34	2	25 (92.6)	27

TABLE 6-2

-2 Log L	=	82.108

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x6	1	11.925	0.001
x7	1	2.052	0.152
x8	1	1.609	0.205
x9	1	0.866	0.352
x10	1	0.791	0.9214
x11	1	18.323	0.000
x12	1	8.622	0.003
x13	1	21.330	0.000
x14	1	0.464	0.495
x15	1	0.614	0.433
x16	1	5.090	0.764
x17	1	21.204	0.000
x18	1	13.157	0.692

TABLE 6-3

-2 Log L	=	56.971

Parameter	Estimate	Standard Error	Wald Chi-Square	Standardized Estimate	Partial Correlation	Exp(Est)
Intercept	-7.008	1.836	15.471			0.001
x13	1.129	0.287	14.574			3.092

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x6	1	4.859	0.028
x7	1	0.132	0.716
x8	1	0.007	0.932
x9	1	1.379	0.240
x10	1	0.129	0.719
x11	1	6.154	0.013

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x12	1	2.745	0.098
x14	1	0.640	0.424
x15	1	0.344	0.557
x16	1	2.529	0.112
x17	1	13.723	0.000
x18	1	1.206	0.272

TABLE 6-4

-2 Log L	=	39.960

Parameter	Estimate	Standard Error	Wald Chi-Square	Standardized Estimate	Partial Correlation	Exp(Est)
Intercept	-14.192	3.712	14.614			0.000
x13	1.079	0.357	9.115			2.942
x17	1.844	0.639	8.331			6.321

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x6	1	0.656	0.418
x7	1	3.501	0.061
x8	1	0.006	0.937
x9	1	0.693	0.405
x10	1	0.091	0.762
x11	1	3.409	0.065
x12	1	0.849	0.357
x14	1	2.327	0.127
x15	1	0.026	0.873
x16	1	0.070	0.919
x18	1	2.907	0.088