Factor Analysis on HBAT Data
Check MSA, partial corr, and eigenvalues

The REG Procedure
Model: MODEL1
Dependent Variable: Firm_size

Number of Observations Read	100
Number of Observations Used	100

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	13	9.32846	0.71757	3.94	<.0001
Error	86	15.66154	0.18211		
Corrected Total	99	24.99000			

Root MSE	0.42674	**R-Square**	0.3733
Dependent Mean	0.51000	**Adj R-Sq**	0.2786
Coeff Var	83.67541		

Parameter Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	1	0.18205	0.84417	0.22	0.8298	0
Prod_Qual	1	0.02983	0.03972	0.75	0.4547	1.67169
Ecommerce	1	-0.37888	0.10286	-3.68	0.0004	2.82256
Tech_support	1	0.10314	0.04892	2.11	0.0379	3.04740
Complaint	1	-0.05692	0.07807	-0.73	0.4679	4.83774
Adv	1	0.00839	0.04734	0.18	0.8598	1.54742
Prod_Line	1	0.18750	0.20096	0.93	0.3534	37.97842
Sales_Image	1	0.31611	0.07645	4.13	<.0001	3.65361
Pricing	1	0.06206	0.03632	1.71	0.0911	1.71202
Warranty	1	-0.33537	0.09459	-3.55	0.0006	3.26841
New_Prod	1	-0.01768	0.02979	-0.59	0.5544	1.07545
Ordering	1	0.07631	0.07876	0.97	0.3353	2.90906
Price_Flex	1	0.23514	0.20532	1.15	0.2553	33.33234
Del_Speed	1	-0.31509	0.38738	-0.81	0.4183	44.00376

Logistic Regression

Research Objective:

I will use the HBAT data set to explore a set of 13 independent metric variables to predict or estimate the size of the value of the variable Firm_size. The non-metric variable Firm_size indicates a smaller firm of less than 500 employees when the value is zero, and a value of one indicates a larger firm above 500 employees. The objective is to find discriminating variables to predict or group the response variable firm_size. To create a probability model that predicts or estimates the probability of *Firm_size* = 0.

Regression Equation:

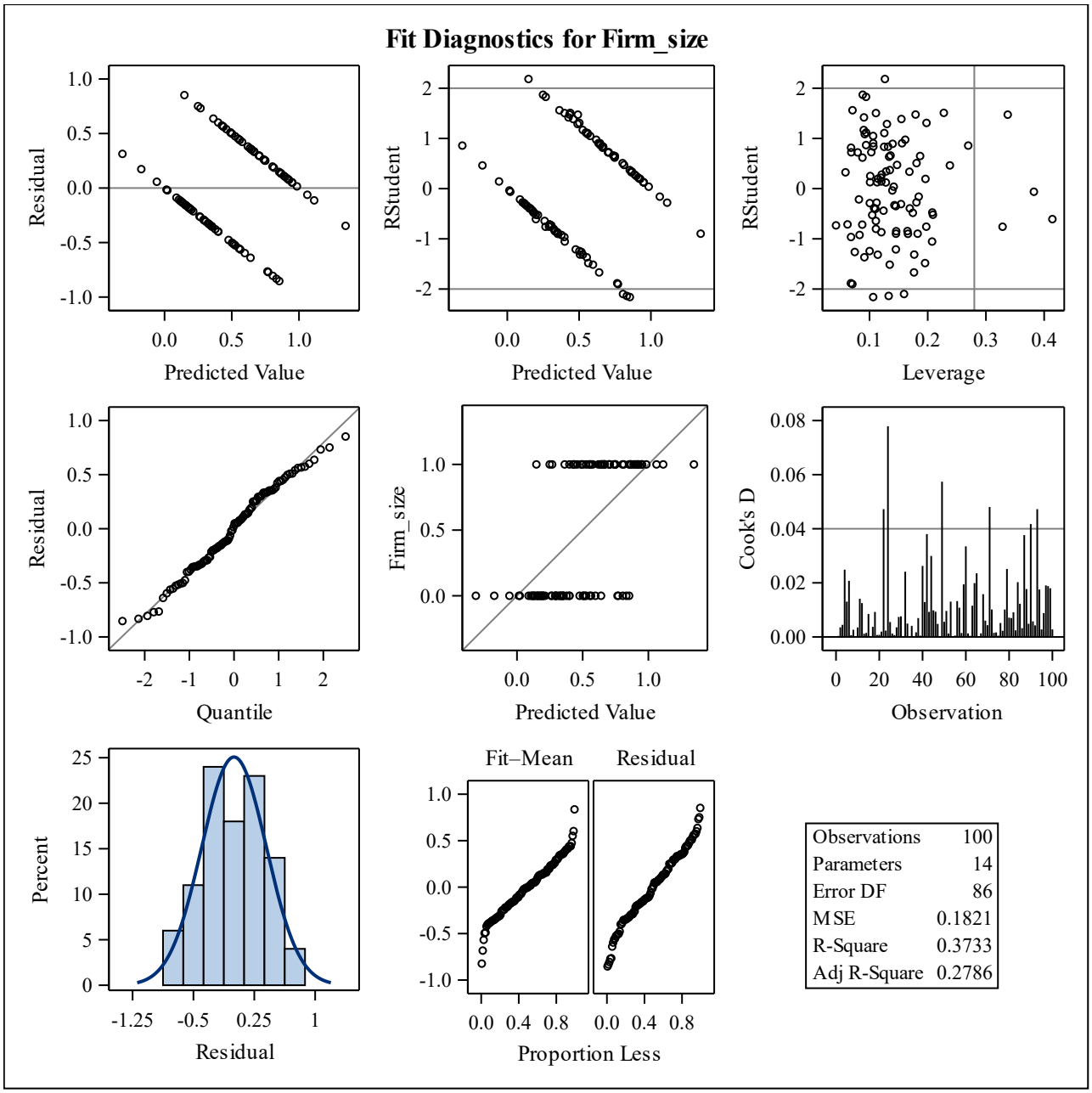
$$logodds = -3.8298 + 2.3763Ecommerce - 0.6159Tech\_support - 1.9895\text{Sales\_Image} - 0.4711\text{Pricing} + 1.9644\text{Warranty}$$

VIF

Prod_Line, Price_Flex, and Del_Speed have VIF values greater than 10 indicating high multicollinearity and are highly correlated with each other. They may be problematic in the model.

Factor Analysis on HBAT Data
Check MSA, partial corr, and eigenvalues

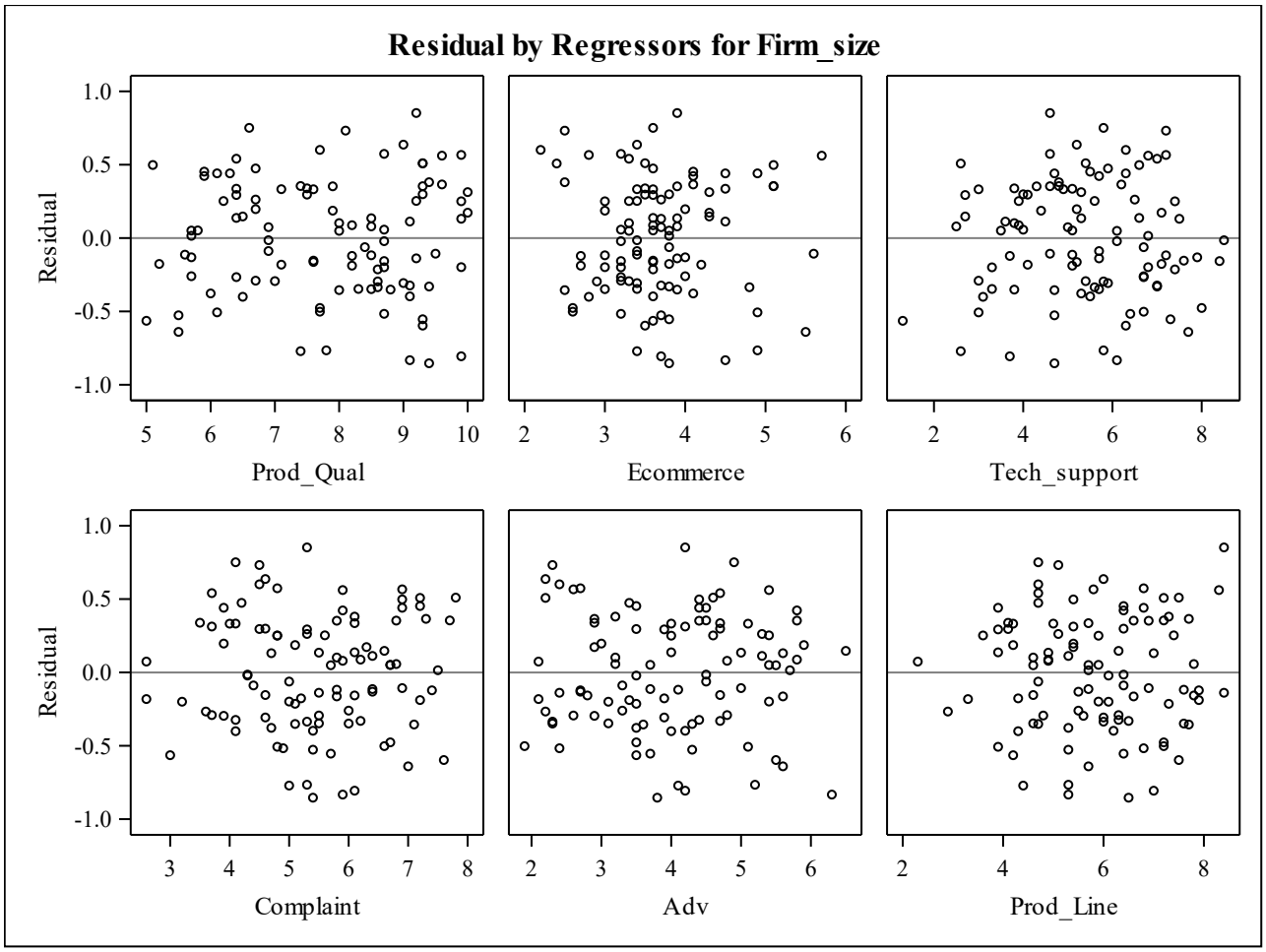
The REG Procedure
Model: MODEL1
Dependent Variable: Firm_size



The graph of RStudent vs Leverage has four observations with a RStudent value just outside of -2, but no outliers are present. From the plot of Cook's D vs Observation, no observations have values > 1, thus no influential observations detected.

Factor Analysis on HBAT Data
Check MSA, partial corr, and eigenvalues

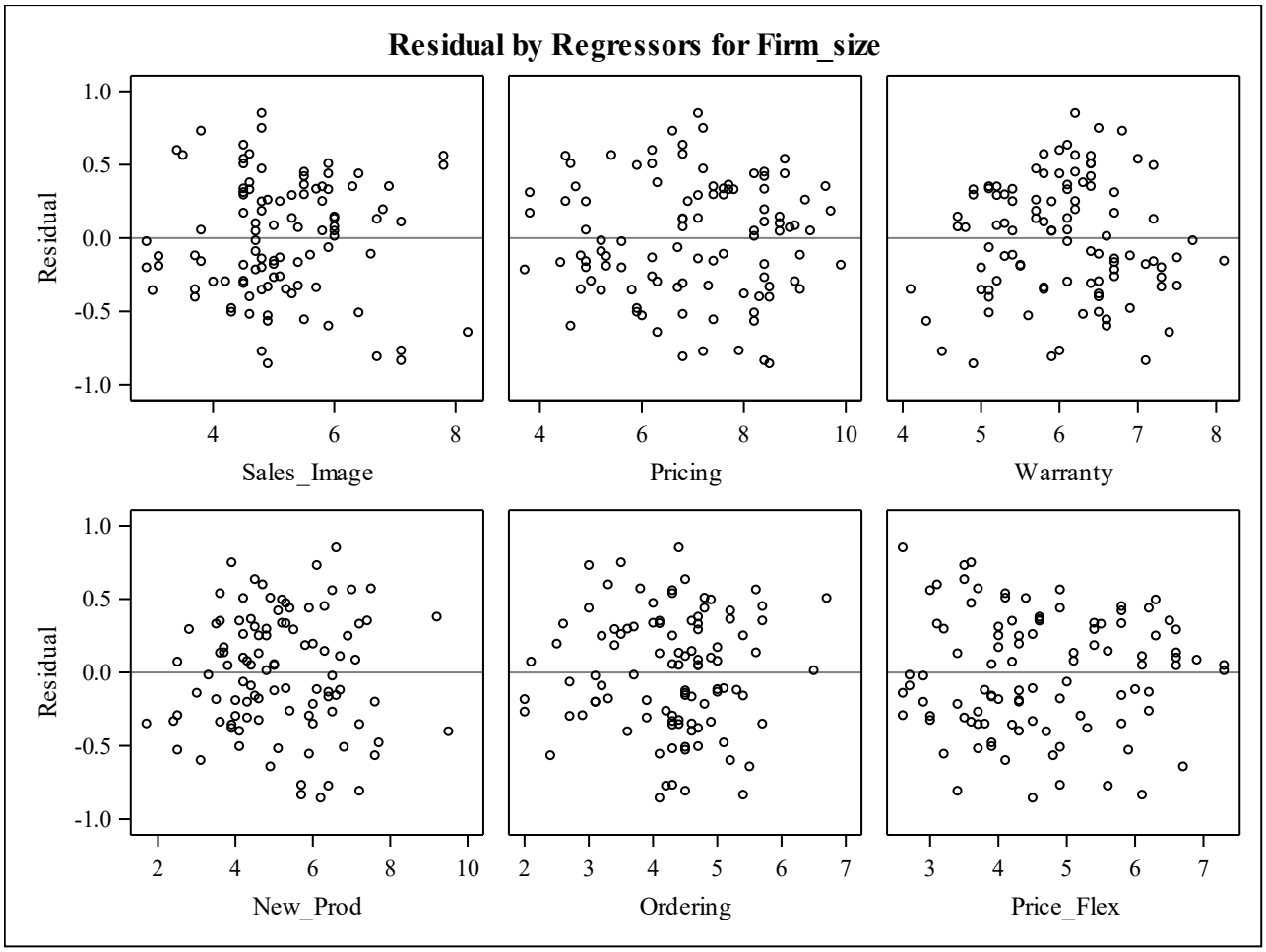
The REG Procedure
Model: MODEL1
Dependent Variable: Firm_size



There are no obvious bivariate outliers.

Factor Analysis on HBAT Data
Check MSA, partial corr, and eigenvalues

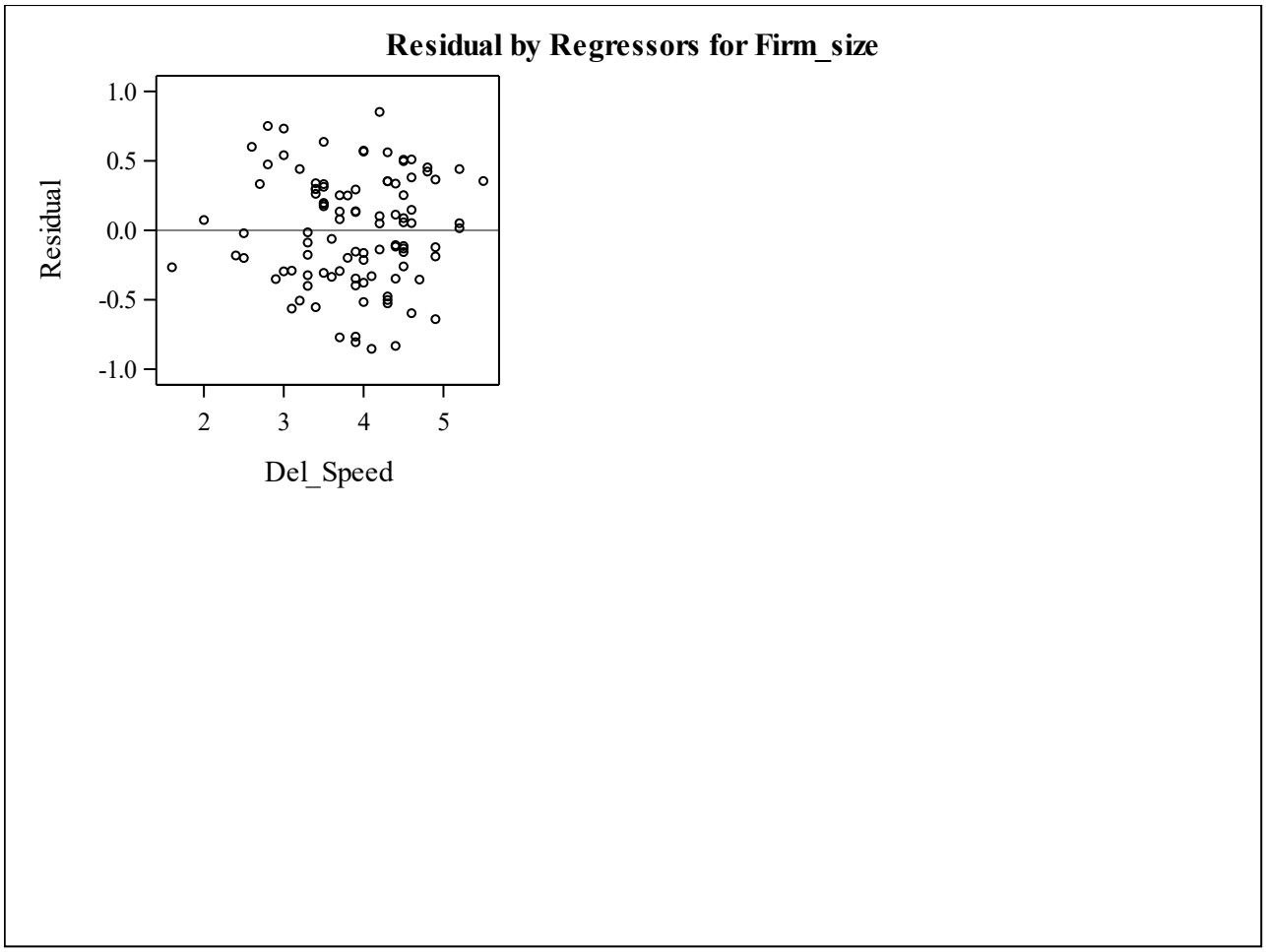
The REG Procedure
Model: MODEL1
Dependent Variable: Firm_size



There are no obvious bivariate outliers.

Factor Analysis on HBAT Data
Check MSA, partial corr, and eigenvalues

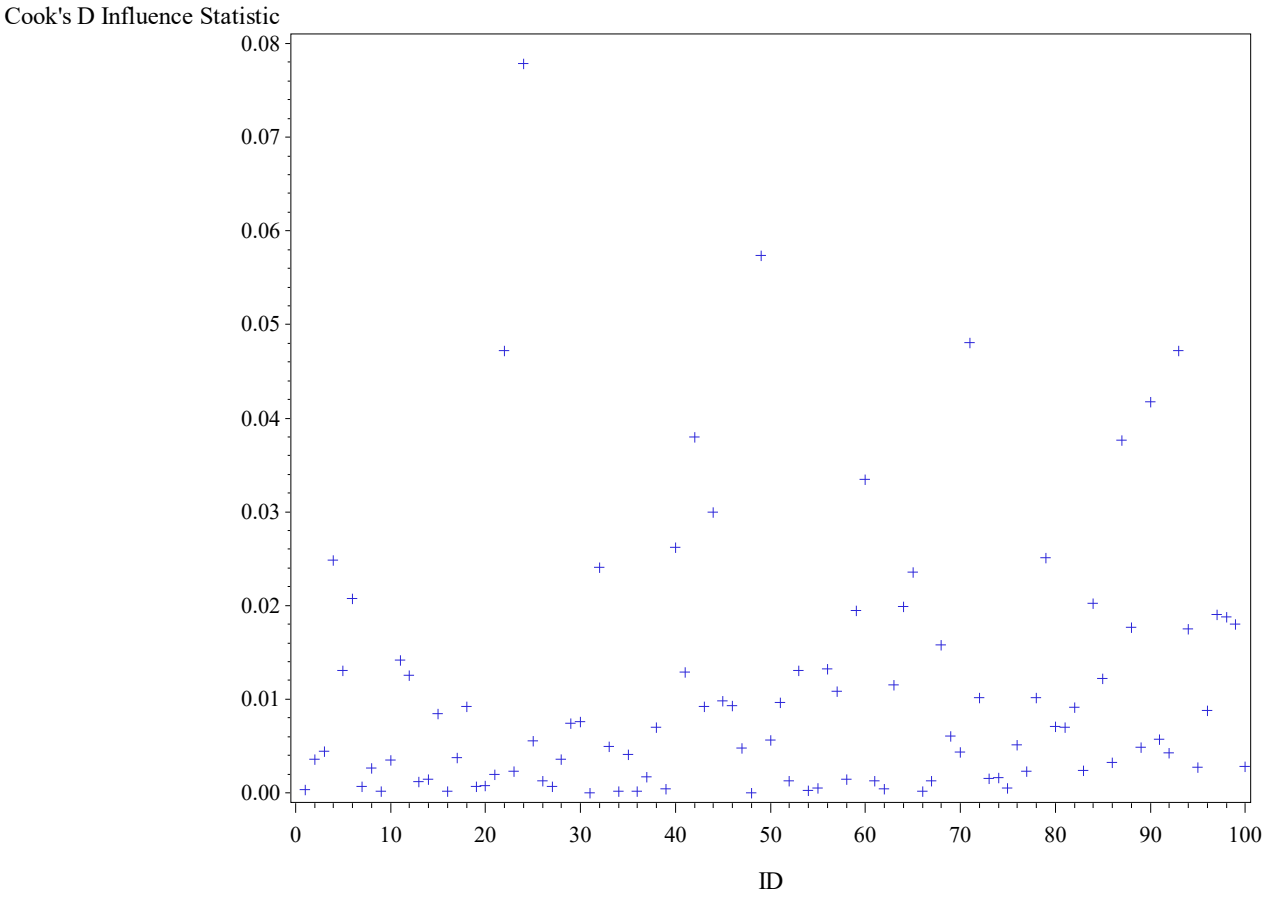
The REG Procedure
Model: MODEL1
Dependent Variable: Firm_size



There are no obvious bivariate outliers.

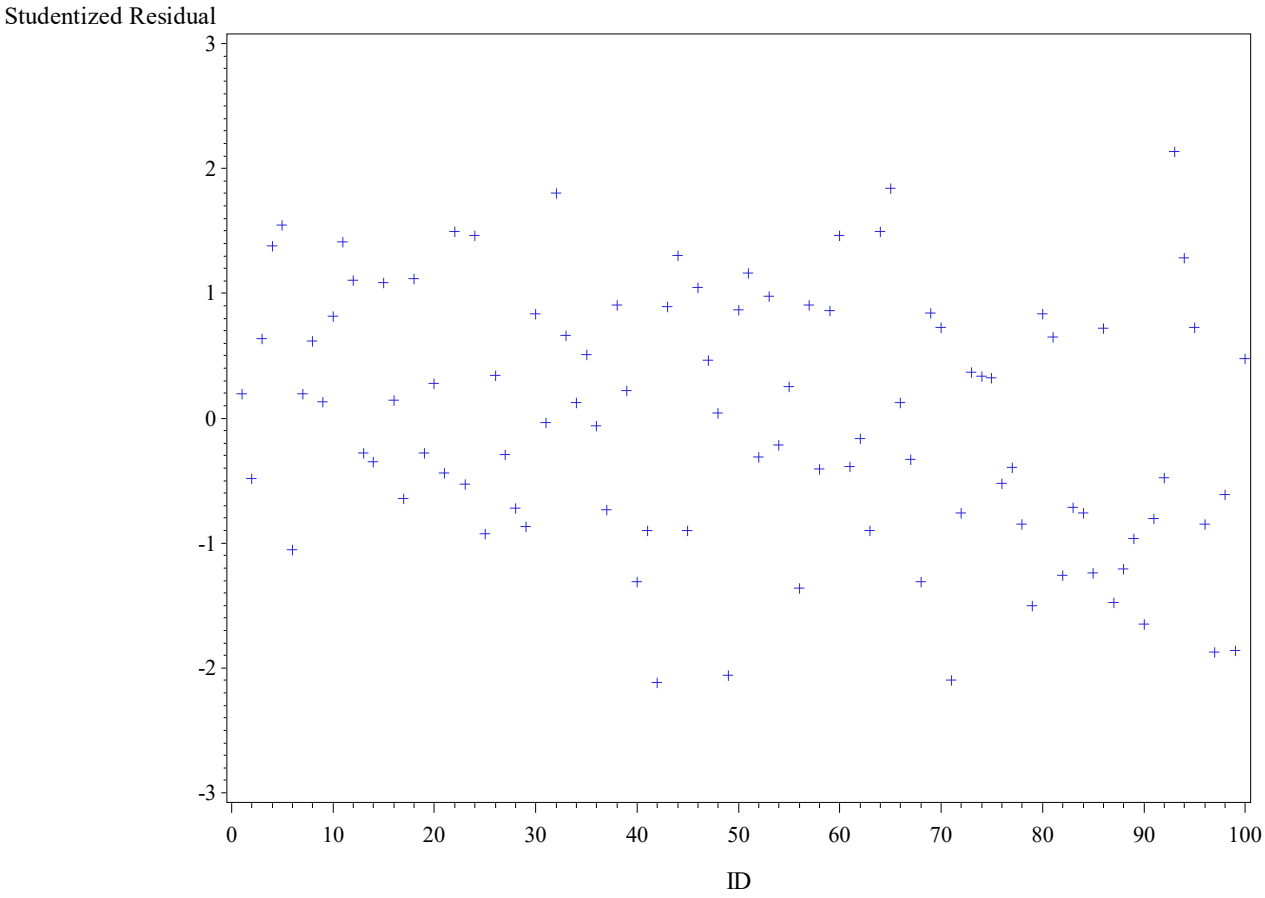
Factor Analysis on HBAT Data

Check MSA, partial corr, and eigenvalues



None of the Cook's D values are > 1, it is concluded that none of the observations are influential.

Factor Analysis on HBAT Data
Check MSA, partial corr, and eigenvalues



No bivariate outliers are observed.

Preliminary Analysis

BK-2

Variable	t Value	df	p-value
Adv	-2.05	98	0.0426
Sales image	-2.99	98	0.0035
Pricing	-3.38	98	0.0010
Warranty	3.06	98	0.0029
Price flexibility	-3.50	98	0.0007

Five variables, Adv, Sales_image, Pricing, Warranty, and Price_Flex, have significantly different means for *Firm_Size* at alpha = 0.05. For each of these 5 variables the mean for *Firm_Size* group 1 is higher than the mean for *Firm_Size* group 0. Thus, smaller firms of less than 500 employees that purchase from HBAT have a lower average perception of HBAT for those 5 variables when compared to larger firms above 500 employees.

These five variables are potential predictor or discriminating variables to predict or group the response variable *firm_size*.

BK-3

Model Information	
Data Set	SASUSER.HBAT
Response Variable	Firm_size
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	100
Number of Observations Used	100

Response Profile		
Ordered Value	**Firm_size**	**Total Frequency**
1	0	49
2	1	51

Probability modeled is Firm_size='0'.

Descriptive Statistics for Continuous Variables					
Variable	**Firm_size**	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Prod_Qual	0	7.953061	1.404176	5.000000	10.000000
	1	7.672549	1.388536	5.100000	9.900000
	Total	7.810000	1.396279	5.000000	10.000000
Ecommerce	0	3.628571	0.705632	2.500000	5.600000
	1	3.713725	0.700006	2.200000	5.700000
	Total	3.672000	0.700516	2.200000	5.700000
Tech_support	0	5.638776	1.655397	1.300000	8.500000
	1	5.101961	1.364916	2.500000	7.500000
	Total	5.365000	1.530457	1.300000	8.500000
Complaint	0	5.332653	1.198330	2.600000	7.600000
	1	5.547059	1.220550	2.600000	7.800000
	Total	5.442000	1.208403	2.600000	7.800000

The group frequencies are 49 for group 0 and 51 for group 1.

Group 0 shows 5 means that appear higher than group 1 about 38% of the variables.

Group 1 shows 8 means that appear higher than group 0 about 62% of the variables.

Descriptive Statistics for Continuous Variables					
Variable	**Firm_size**	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Adv	0	3.777551	1.059730	1.900000	6.300000
	1	4.233333	1.154412	2.100000	6.500000
	Total	4.010000	1.126943	1.900000	6.500000
Prod_Line	0	6.040816	1.288235	2.900000	8.400000
	1	5.578431	1.313516	2.300000	8.400000
	Total	5.805000	1.315285	2.300000	8.400000
Sales_Image	0	4.808163	1.111462	2.900000	8.200000
	1	5.425490	0.948861	3.400000	7.800000
	Total	5.123000	1.072320	2.900000	8.200000
Pricing	0	6.467347	1.553811	3.700000	9.900000
	1	7.460784	1.384063	4.500000	9.700000
	Total	6.974000	1.545055	3.700000	9.900000
Warranty	0	6.289796	0.892152	4.300000	8.100000
	1	5.805882	0.670347	4.100000	7.200000
	Total	6.043000	0.819738	4.100000	8.100000
New_Prod	0	5.183673	1.570078	2.400000	9.500000
	1	5.117647	1.430064	1.700000	9.200000
	Total	5.150000	1.493048	1.700000	9.500000
Ordering	0	4.200000	0.885532	2.000000	5.700000
	1	4.352941	0.971463	2.100000	6.700000
	Total	4.278000	0.928840	2.000000	6.700000
Price_Flex	0	4.202041	1.026014	2.600000	6.700000
	1	5.001961	1.244747	2.600000	7.300000
	Total	4.610000	1.206004	2.600000	7.300000
Del_Speed	0	3.795918	0.714131	1.600000	4.900000
	1	3.972549	0.750221	2.000000	5.500000
	Total	3.886000	0.734437	1.600000	5.500000

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	140.589	120.065
SC	143.195	156.537
-2 Log L	138.589	92.065

R-Square	0.3720	**Max-rescaled R-Square**	0.4961

Overall Model Test Likelihood ratio

H_0: none of the variables help to explain Firm_size
H_a: at least one variable helps to explain Firm_size.

$\chi^2(13) = 46.5248$ with p-value < .0001. Reject **H_0** at least one of the variables helps to explain firm_size.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	46.5248	13	<.0001
Score	37.3288	13	0.0004
Wald	25.2958	13	0.0211

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.2374	5.2263	0.0561	0.8128
Prod_Qual	1	-0.0592	0.2650	0.0499	0.8233
Ecommerce	1	2.6456	0.7664	11.9176	0.0006
Tech_support	1	-0.7616	0.3295	5.3417	0.0208
Complaint	1	0.3124	0.4647	0.4519	0.5014
Adv	1	0.0216	0.2958	0.0053	0.9418
Prod_Line	1	-1.5398	1.4213	1.1737	0.2786
Sales_Image	1	-2.0836	0.5690	13.4076	0.0003
Pricing	1	-0.5147	0.2450	4.4139	0.0356
Warranty	1	2.3338	0.7124	10.7327	0.0011
New_Prod	1	0.1410	0.1750	0.6492	0.4204
Ordering	1	-0.5784	0.4973	1.3526	0.2448
Price_Flex	1	-1.6222	1.3962	1.3499	0.2453
Del_Speed	1	2.5976	2.7205	0.9117	0.3397

Wald Chi-Square Test

Due to the presence of multicollinearity the Wald Chi-Square tests cannot be used to determine which variables are significant.

A stepwise selection will be used since it controls for multicollinearity.

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
Prod_Qual	0.943	0.561	1.584
Ecommerce	14.092	3.138	63.282
Tech_support	0.467	0.245	0.891
Complaint	1.367	0.550	3.398
Adv	1.022	0.572	1.824
Prod_Line	0.214	0.013	3.476
Sales_Image	0.124	0.041	0.380
Pricing	0.598	0.370	0.966
Warranty	10.317	2.554	41.682
New_Prod	1.151	0.817	1.622
Ordering	0.561	0.212	1.486
Price_Flex	0.197	0.013	3.048
Del_Speed	13.432	0.065	>999.999

Exponentiated Coefficients

Due to multicollinearity, these odds ratios are not subject to interpretation.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	86.1	**Somers' D**	0.721
Percent Discordant	13.9	**Gamma**	0.721
Percent Tied	0.0	**Tau-a**	0.364
Pairs	2499	**c**	0.861

Predicted Probabilities and Observed Responses

Percent Concordant: Relatively high, indicating good predictive performance.

Percent Discordant: Relatively low, a desirable result.

Somers' D: 0.721, Good strength/magnitude.

Goodman-Kruskal Gamma: 0.721, Good strength/magnitude.

Kendall's Tau-a: 0.721, Moderate strength/magnitude.

C (Concordance Index): 0.861 indicates a high level of predictive discrimination.

Partition for the Hosmer and Lemeshow Test					
		Firm_size = 0		**Firm_size = 1**	
Group	**Total**	**Observed**	**Expected**	**Observed**	**Expected**
1	10	0	0.43	10	9.57
2	10	1	1.06	9	8.94
3	10	4	1.97	6	8.03
4	10	0	3.04	10	6.96
5	10	7	3.86	3	6.14
6	10	2	5.33	8	4.67
7	10	8	6.95	2	3.05
8	10	8	7.93	2	2.07
9	10	9	8.90	1	1.10
10	10	10	9.54	0	0.46

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
17.0266	8	0.0298

The Hosmer and Lemeshow

H_0: The model fits the data
H_a: The model does not fit the data.

$\chi^2(8) = 17.0266$, p-value = 0.0298. At alpha = 0.05, reject H_0 and conclude the logistic model does not fit the data well. This is likely due to having non-significant independent variables along with multicollinearity present in this model. Stepwise logistic regression will be used to address and control for both these issues.

BK-4

Stepwise Selection

Model Information	
Data Set	SASUSER.HBAT
Response Variable	Firm_size
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	100
Number of Observations Used	100

Response Profile		
Ordered Value	**Firm_size**	**Total Frequency**
1	0	49
2	1	51

Probability modeled is Firm_size='0'.

Stepwise Selection Procedure

Step 0. Intercept entered:

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

-2 Log L	=	138.589

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
37.3288	13	0.0004

Step 1. Effect Price_Flex entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	140.589	130.983
SC	143.195	136.193
-2 Log L	138.589	126.983

R-Square	0.1096	Max-rescaled R-Square	0.1461

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11.6068	1	0.0007
Score	11.1052	1	0.0009
Wald	10.1818	1	0.0014

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
28.8267	12	0.0042

Note: No effects for the model in Step 1 are removed.

Step 2. Effect Warranty entered:

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	140.589	126.328
SC	143.195	134.144
-2 Log L	138.589	120.328

R-Square	0.1669	**Max-rescaled R-Square**	0.2226

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	18.2611	2	0.0001
Score	17.0300	2	0.0002
Wald	14.8524	2	0.0006

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
24.0778	11	0.0124

Note: No effects for the model in Step 2 are removed.

Step 3. Effect Sales_Image entered:

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	140.589	122.360
SC	143.195	132.781
-2 Log L	138.589	114.360

R-Square	0.2152	Max-rescaled R-Square	0.2869

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	24.2290	3	<.0001
Score	22.3968	3	<.0001
Wald	19.0627	3	0.0003

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
18.7476	10	0.0436

Step 4. Effect Price_Flex is removed:

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	140.589	123.910
SC	143.195	131.726
-2 Log L	138.589	117.910

R-Square	0.1868	Max-rescaled R-Square	0.2491

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	20.6790	2	<.0001
Score	19.2282	2	<.0001
Wald	16.5544	2	0.0003

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
21.7457	11	0.0264

Note: No effects for the model in Step 4 are removed.

Step 5. Effect Ecommerce entered:

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	140.589	114.928
SC	143.195	125.349
-2 Log L	138.589	106.928

R-Square	0.2714	**Max-rescaled R-Square**	0.3619

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	31.6610	3	<.0001
Score	27.8202	3	<.0001
Wald	21.4124	3	<.0001

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
13.1993	10	0.2127

Note: No effects for the model in Step 5 are removed.

Step 6. Effect Pricing entered:

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	140.589	111.641
SC	143.195	124.667
-2 Log L	138.589	101.641

R-Square	0.3089	**Max-rescaled R-Square**	0.4119

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	36.9486	4	<.0001
Score	30.8586	4	<.0001
Wald	21.9249	4	0.0002

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
9.0518	9	0.4325

Note: No effects for the model in Step 6 are removed.

Step 7. Effect Tech_support entered:

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	140.589	109.112
SC	143.195	124.743
-2 Log L	138.589	97.112

R-Square	0.3395	Max-rescaled R-Square	0.4527

The R-squared value of 0.3395 indicates that my model explains approximately 33.95% of the variation in the dependent variable, a moderate level of explanation. The max-rescaled R-squared value of 0.4527 suggests that my model's explanatory power is about 45.27%. The next step is to examine the likelihood ration to confirm if the model is a good fit.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	41.4772	5	<.0001
Score	33.6591	5	<.0001
Wald	23.2752	5	0.0003

Likelihood Ratio Test

H_0: none of the variables help to explain Firm_size
H_a: at least one variable helps to explain Firm_size.

$\chi^2(5) = 41.4772$ with p-value < .0001.
Reject **H_0** At least one of the 5 variables helps to explain firm_size.

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
4.9528	8	0.7626

Residual Chi-Square Test

H_0: The model provides a good fit to the data.
H_a: The model is not a good fit for the data.

$\chi^2(8) = 4.9528$ with p-value > $\alpha = 0.5$.
Fail to reject **H_0** the model provides a good fit for the data.

Note: No effects for the model in Step 7 are removed.

Note: No (additional) effects met the 0.05 significance level for entry into the model.

Summary of Stepwise Selection							
	Effect			Number	Score	Wald	
Step	**Entered**	**Removed**	**DF**	**In**	**Chi-Square**	**Chi-Square**	**Pr > ChiSq**
1	**Price_Flex**		1	1	11.1052		0.0009
2	**Warranty**		1	2	6.5508		0.0105
3	**Sales_Image**		1	3	5.9377		0.0148
4		**Price_Flex**	1	2		3.4218	0.0643
	Ecommerce		1	3	10.5396		0.0012
6	**Pricing**		1	4	5.2433		0.0220
7	**Tech_support**		1	5	4.3857		0.0362

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-3.8298	2.7863	1.8892	0.1693
Ecommerce	1	2.3763	0.7034	11.4135	0.0007
Tech_support	1	-0.6159	0.3070	4.0245	0.0448
Sales_Image	1	-1.9895	0.4969	16.0280	<.0001
Pricing	1	-0.4711	0.1907	6.1010	0.0135
Warranty	1	1.9644	0.6101	10.3662	0.0013

Logistic Regression Model

All five variables are significant, but the intercept may not be significantly different than zero.

The Model: log(odds of Firm_Size = 0) = -3.8298 2.3763**Ecommerce* – -0.6159**Tech_support* - 1.9895**sales_image* -0.4711**Pricing* +1.9644**Warranty*.

There are two positive coefficients 2.3763**Ecommerce**, and 1.9644**Warranty** that *increase* the probability of being in Firm_size group 0. There are three negative coefficients - 0.6159**Tech_support**, - 1.9895**Sales_Image**, and - 0.4711**Pricing** that *decrease* the probability of being in Firm_size group 0.

Increase or Decrease in Odds of being in Group 0

For a one-unit change in each predictor.

Odds of being in Group 0

*(odds ratio -1) *100*

Ecommerce	976.5%
Tech_support	-46%
Sales_Image	-86.3%
Pricing	-37.6%
Warranty	613%

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
Ecommerce	10.765	2.712	42.733
Tech_support	0.540	0.296	0.986
Sales_Image	0.137	0.052	0.362
Pricing	0.624	0.430	0.907
Warranty	7.130	2.157	23.574

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	83.9	**Somers' D**	0.678
Percent Discordant	16.1	**Gamma**	0.678
Percent Tied	0.0	**Tau-a**	0.342
Pairs	2499	**c**	0.839

Partition for the Hosmer and Lemeshow Test					
		Firm_size = 0		**Firm_size = 1**	
Group	**Total**	**Observed**	**Expected**	**Observed**	**Expected**
1	10	1	0.59	9	9.41
2	10	2	1.31	8	8.69
3	10	3	2.12	7	7.88
4	10	2	3.30	8	6.70
5	10	5	4.11	5	5.89
6	10	2	5.07	8	4.93
7	10	8	6.72	2	3.28
8	10	7	7.56	3	2.44

Partition for the Hosmer and Lemeshow Test					
		Firm_size = 0		Firm_size = 1	
Group	Total	Observed	Expected	Observed	Expected
9	10	9	8.66	1	1.34
10	10	10	9.56	0	0.44

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
7.5209	8	0.4816

The Hosmer and Lemeshow

H_0: The model fits the data
H_a: The model does not fit the data.

$\chi^2(8) = 7.5209$, p-value > $\alpha = 0.5$. Fail to reject **H_0** the model fits the data.

Final Model Summary

Research Objective and Analysis Plan

I will use the HBAT data set to explore a set of 13 independent metric variables to predict or estimate the size of the value of the variable Firm_size. The non-metric variable Firm_size indicates a smaller firm of less than 500 employees when the value is zero, and a value of one indicates a larger firm above 500 employees. The objective is to find discriminating variables to predict or group the response variable firm_size. To create a probability model that predicts or estimates the probability of Firm_size = 0. First, I will screen my data by checking for multicollinearity and bivariate outliers. A two-sample t-test will be used to determine if any of the means are significantly different by Firm_size, and an examination of descriptive statistics to determine which of the two groups is preferred with a higher mean for each independent variable. Logical regression will be performed on the full model and tested to see if at least one variable explains Firm_size. A Wald Chi-square test will be used to test individual parameters, exponentiated coefficients, and predicted probabilities will be interpreted unless multicollinearity is present in the model. If multicollinearity is present, a stepwise selection will be used to control multicollinearity. R-squared and Max-rescaled R-squared and a Residual Chi-squared test will be used to determine the model fitness. The selected model will be tested with the likelihood ratio to see if at least one variable explains Firm_size, After the model is selected, the final logistic regression will be run and examined, and a regression equation will be created from the output. Finally, Summarization and validation of the findings.

Statistacal Assumptions

All relevant variables are included. There are no bivariate outliers observed.
Multicollinearity is present in the model, the variables Prod_Line, Price_Flex, and Del_Speed having VIF values > than 10.

Model Fit & Stepwise Estimation

Overall Model Fit

The Hosmer and Lemeshow After Stepwise Selection.

H_0: The model fits the data
H_a: The model does not fit the data.

$\chi^2(8) = 7.5209$, p-value = 0.4816. Fail to reject **H_0** the model fits the data.

Significance of Coefficients

Exponentiated Coefficients

Due to multicollinearity, these odds ratios are not subject to interpretation.

Case Wise Diagnostics

R-squared explains approximately 33.95% and max-rescaled R-squared suggests that my model's explanatory power is about 45.27%.

Likelihood Ratio Test
H_0: none of the variables help to explain Firm_size
H_a: at least one variable helps to explain Firm_size.
$\chi^2(5) = 41.4772$ with p-value < .0001. Reject.

Residual Chi-squared Test
H_0: The model provides a good fit to the data.
H_a: The model is not a good fit for the data.
$\chi^2(8) = 4.9528$ with p-value > $\alpha = 0.5$. FTR.

Interpretation

Probability Model for *Firm_size* = 0. The model is predicting or estimating the probability of occurring in group 0.

Regression Equation: $\log(odds\ of\ Firm\_size = 0) = -3.8298 + 2.3763Ecommerce - 0.6159Tech\_support - 1.9895\text{Sales\_Image} - 0.4711\text{Pricing} + 1.9644\text{Warranty}$

➔ **Intercept** (-3.8298): Representing the log odds of the outcome when all independent variables are zero.

1. **Ecommerce** (2.3763): The coefficient is positive, suggesting an increase in Ecommerce is associated with **higher** odds of Firm_Size = 0.
2. **Tech_support** (-0.6159): The coefficient is negative, suggesting an increase in Tech_support is associated with **lower** odds of Firm_Size = 0.
3. **Sales_Image** (-1.9895): The coefficient is negative, suggesting an increase in Sales_Image is associated with **lower** odds of Firm_Size = 0.
4. **Pricing** (-0.4711): The coefficient is negative, suggesting an increase in Pricing is associated with **lower** odds of Firm_Size = 0.
5. **Warranty** (1.9644): The coefficient is positive, suggesting an increase in Warranty is associated with **higher** odds of Firm_Size = 0.

Variable Strengths: The two positive coefficients 2.3763**Ecommerce** and 1.9644**Warranty** that *increase* the probability of being in Firm_size group 0. The three negative coefficients -0.6159**Tech_support**, -1.9895**Sales_Image**, and -0.4711**Pricing** *decrease* the probability of being in Firm_size group 0. Ecommerce has the most positive value and Sales_Images has the most negative value of the five coefficients.

Analysis of the Variate: A Firm with 500 or less employees, group 0 in the Firm_size variable, will increase the probability of being in group 0 with each one-unit increase in the predictor variables by the following percentages for each of the predictor variables included in the final model. Ecommerce by 976.5% and Warranty by 613%, both variables have a positive coefficient. Respectively, the variables with negative coefficients will decrease for a one-unit increase in the predictor variables. Tech_support by -46%, Sales_Image by -86.3%, and Pricing by -37.6% probability of being in group 0. Ecommerce, the overall friendliness of HBAT's website is viewed more favorably by smaller firms with 500 or fewer employees. Warranty, the extent to which HBAT stands behind its product and service claims is also appreciated by smaller firms. Investing in Ecommerce and improving website functionality, enhancing user experience, and expanding online marketing efforts to attract small businesses. Introducing an extended warranty can potentially increase customer satisfaction, loyalty, customer retention, brand value, and may increase sales volume. Human resources could work to provide improved training programs to current employs so that they can provide better technical support. Marketing analysis could be used to address sales image and pricing, consider a reputable data analytics firm with a good track record. Stay vigilant and follow the data.

Validation

The probability values versus predicted had the following results, showing that the model predicts the correct results 76% of the time.

Number of Errors Found	24
Number of Correct Predictions	76