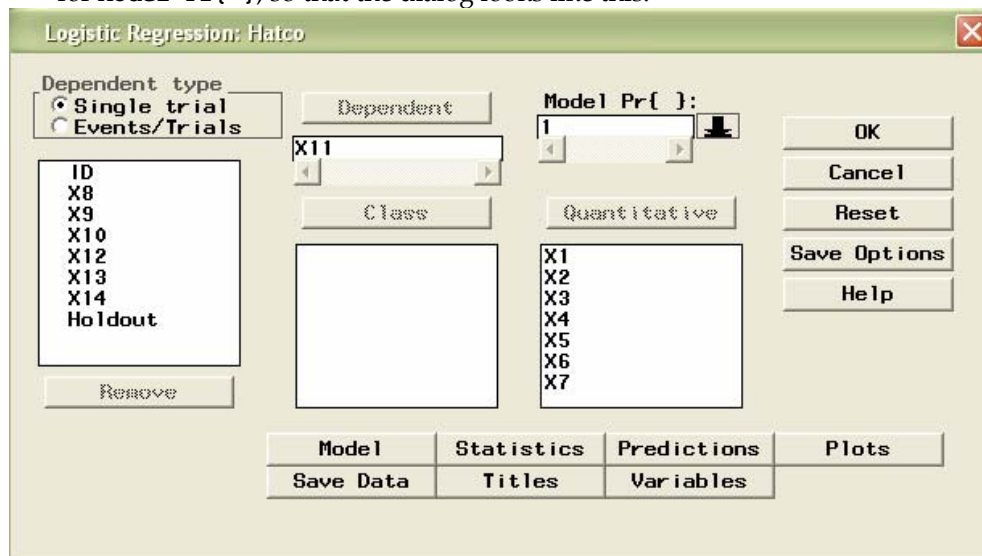


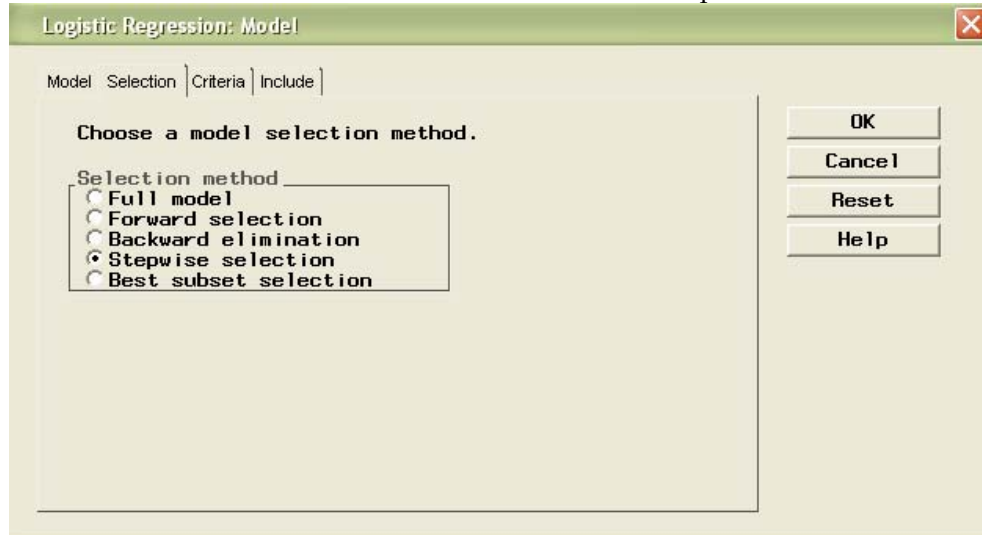
DSC 410/510 – Logistic Regression

Illustrative Example from p314-321

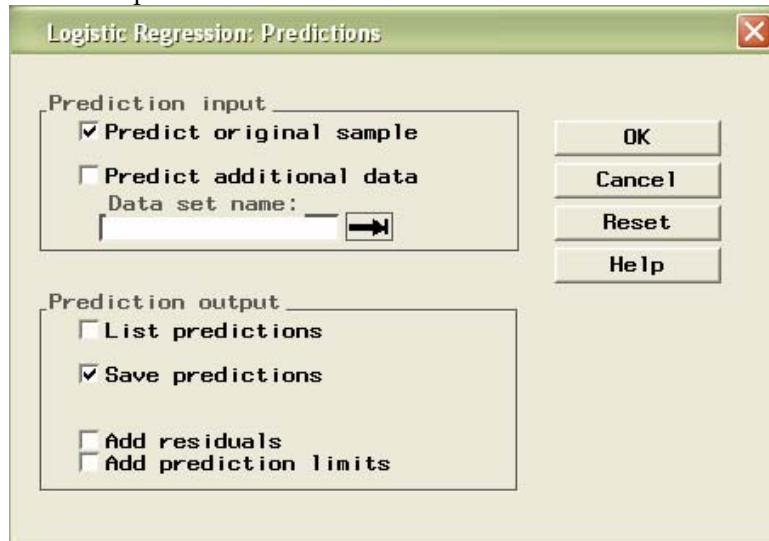
1. This document will help you get started with logistic regression using SAS software. Follow the instructions below to begin the illustrative example on p314-321 of the textbook.
2. Import the data from the Excel spreadsheet “hatco.xls” (which can be downloaded from the class website) as usual by selecting **File > Import Data**.
3. Start the “Analyst” application by selecting **Solutions > Analysis > Analyst**
 - a. Select **File > Open By SAS Name** and select Hatco in the “Work” library; the Hatco data table should appear in the Analyst spreadsheet window.
 - b. Select **Statistics > Regression > Logistic**; the “Logistic Regression dialog” should appear.
 - c. Move “X11” to the **Dependent** box, “X1, X2, ..., X7” to the **Quantitative** box, and select “1” for **Model Pr{ }**, so that the dialog looks like this:



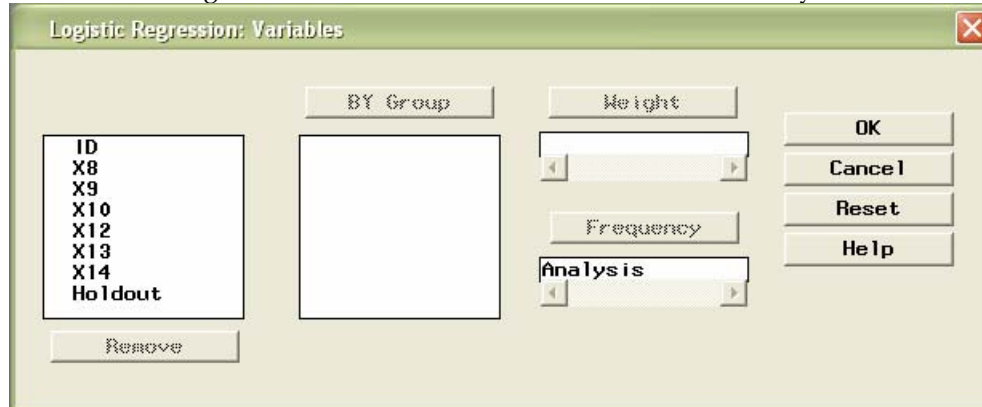
- d. Next click on “Model” then “Selection” and select “Stepwise selection.”



- e. After clicking "OK," next click on "Predictions" and check "Predict original sample" and "Save predictions."



- f. After clicking "OK," next click on "Variables" and move "Analysis" to the **Frequency** box.



4. Hit "OK" (for the Variables dialog) and then "OK" (for the main dialog) to run the analysis - you should get a window showing the main results of the model fit (we'll go over what this all means in class):

The LOGISTIC Procedure
Model Information

Data Set	WORK.HATCO	
Response Variable	X11	X11
Number of Response Levels	2	
Number of Observations	60	
Frequency Variable	Analysis	Analysis
Sum of Frequencies	60	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Response Profile

Ordered Value	X11	Total Frequency
1	1	38
2	0	22

Probability modeled is X11=1.

NOTE: 40 observations having zero frequencies or weights were excluded since they do not contribute to the analysis.

Stepwise Selection Procedure

Step 0. Intercept entered:

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

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
41.4133	7	<.0001

Step 1. Effect X7 entered:

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

Model Fit Statistics

Criterion	Intercept and Covariates	
	Intercept Only	Intercept and Covariates
AIC	80.859	41.524
SC	82.953	45.713
-2 Log L	78.859	37.524

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	41.3346	1	<.0001
Score	30.0455	1	<.0001
Wald	14.6788	1	0.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
19.5275	6	0.0034

Step 2. Effect X3 entered:

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

Model Fit Statistics

Criterion	Intercept and Covariates	
	Intercept Only	Intercept and Covariates
AIC	80.859	26.258
SC	82.953	32.541
-2 Log L	78.859	20.258

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	58.6006	2	<.0001
Score	37.2940	2	<.0001
Wald	7.2119	2	0.0272

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
11.4094	5	0.0438

Step 3. Effect X5 entered:

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

Model Fit Statistics

Criterion	Intercept and Covariates	
	Intercept Only	Intercept and Covariates
AIC	80.859	14.254
SC	82.953	22.631
-2 Log L	78.859	6.254

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	72.6053	3	<.0001
Score	40.0604	3	<.0001
Wald	2.4552	3	0.4834

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
4.3184	4	0.3646

Step 4. Effect X5 is removed:

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

Model Fit Statistics

Criterion	Intercept and Covariates	
	Intercept Only	Intercept and Covariates
AIC	80.859	26.258
SC	82.953	32.541
-2 Log L	78.859	20.258

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	58.6006	2	<.0001
Score	37.2940	2	<.0001
Wald	7.2119	2	0.0272

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
11.4094	5	0.0438

NOTE: Model building terminates because the last effect entered is removed by the Wald statistic criterion.

Summary of Stepwise Selection

Step	Effect		DF	Number		Score	Wald	Pr > ChiSq	Variable Label
	Entered	Removed		In	Out				
1	X7		1	1		30.0455	.	<.0001	X7
2	X3		1	2		15.6104	.	<.0001	X3
3	X5		1	3		8.8247	.	0.0030	X5
4		X5	1	2		.	1.7329	0.1880	X5

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	8.3295	5.1098	2.6572	0.1031
X3	1	1.8306	0.7171	6.5168	0.0107
X7	1	-2.9126	1.1354	6.5809	0.0103

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X3	6.237	1.530 25.433	
X7	0.054	0.006 0.503	

Association of Predicted Probabilities and Observed Responses

Percent Concordant	98.6	Somers' D	0.972
Percent Discordant	1.3	Gamma	0.974
Percent Tied	0.1	Tau-a	0.459
Pairs	836	c	0.986

To find the predicted probabilities, click on **Analyst** at the bottom of the main SAS Window, then double-click on **Predictions Table** in the left pane of the Analyst Window. Next select **File > Save as By SAS Name**, navigate to the **Work Library**, and give the data table a "member name" like "results." Then in the Explorer pane (on the left), double-click on **Libraries**, then **Work**, then **Results** (or whatever you called the data table you just saved). Finally, select **Edit > Edit Mode** to allow sorting on variables (using **Data > Sort**) and other spreadsheet-style operations, so that you can determine correct/incorrect classifications, etc.