```
*;
*;
* HATCO - Logistic Regression Analysis;
*;
*;
ods graphics on;
*;
options ls=80 ps=50 nodate pageno=1;
*;
Title 'Logistic Regression Homework';
*;
* Input HATCO ;
*;
Data HATCO;
Infile '/folders/myfolders/HATCO_Split60.txt' DLM = '09'X TRUNCOVER;
Input ID Split60 X1 X2 X3 X4 X5 X6 X7 X11;
*;

Data HATCO;
Set HATCO (Keep = ID Split60 X4 X6 X7 X8 X9 X10 X11 X12 X13 X14 X15 X16 X17 X18);
Label ID = 'ID - Identification Number'
Split60 = 'Split60'
X1 = 'X1 - Delivery Speed'
X2 = 'X2 - Price level'
X3 = 'X3 - Price flexibility'
X4 = 'X4 - Manufacturer"s image'
X5 = 'X5 - Service'
X6 = 'X6 - Salesforce"s image'
X7 = 'X7 - Product quality'
X11 = 'X11 - Specification buying'

*;

* Create HBAT Split 60 (Original/Initial) and Split 40 (Validation/Holdout) Datasets ;

*;

Data HATCO60;
Set HATCO;
If Split60 = 1;
*;

Data HATCO40;
Set HATCO;
If Split60 = 0;
*;

Proc Print Data = HATCO60;
*;

Proc Print Data = HATCO40;
*;
*;
* Stepwise Logistic Regression Analysis - X11 = X1 X2 X3 X4 X5 X6 X7;
*;
* EVENT='category' | keyword
* specifies the event category for the binary response model.
*;
* SELECTION = option specifies the method used to select the explanatory variables in the
model.
* STEPWISE requests stepwise selection;
*;
* SLENTRY = option specifies the significance level for entry into the model
* SLSTAY = option specifies the significance level for staying in the model
*;
* DETAILS option produces detailed printout at each step of the model-building process
*;
* LACKFIT requests Hosmer and Lemeshow goodness-of-fit test
*;
* RSQUARE displays generalized R^2
*;
* CTABLE option requests the printing of a classification table for the final model
produced by the procedure.
*;
* PPROB = option specifies possibly multiple cutpoints used to classify observations for
the CTABLE option.
* The values must be between 0 and 1. If the PPROB= option is not specified, the
* default is to print the classification for a range of probabilities from the
smallest estimated
* probability (rounded below to the nearest .02) to the highest estimated
probability (rounded above
```

```
* to the nearest .02) with 0.02 increments. Note that the PPROB= option has no
effect unless the
* CTABLE option is also specified.
*;
*;

Proc Logistic Data = HATCO60;
Model X11(event='0') = X1 X2 X3 X4 X5 X6 X7
/ Selection=Stepwise SLEntry=0.05 SLStay=0.05 Details
LackFit RSquare CTable PProb =(0 to 1 by .10);
*;
* Final Resultant Model and Output Model;
*;

Proc Logistic Data = HATCO60 OutModel=Logistic60;
Model X11(event='0') = X3 X7
/ LackFit RSquare CTable PProb =(0.40 to 0.60 by
.01);
*;
* Original Split60 Logistic Model Fitted to Split40 validation Data;
*;

Proc Logistic InModel=Logistic60;
Score Data = HATCO60 (Keep = X11 X3 X7) Out = HATCO60Score;
*;
* Proc Freq Crosstabulations Original and Holdout Validation Datasets;
*;

Proc Print Data = HATCO60Score;

Proc Freq Data = HATCO60Score;
Table F_X11 * I_X11;
*;

Proc Logistic InModel=Logistic60;
Score Data = HATCO40 (Keep = X11 X3 X7) Out = HATCO40Score;

Proc Print Data = HATCO40Score;

Proc Freq Data = HATCO40Score;

Table F_X11 * I_X11;
*;
*;
* ods graphics off;
*;
*;
Run;
Quit;
```