Assignment 6

Sai Kumar Murarishetti
Lewis ID: L30079224

1. **Logistic Regression, Linear Regression or Linear Discriminant Analysis? Write your thoughts in one page summary**

Logistic Regression**:**

- Logistic regression is used to predict the binary or categorical dependent variable using a given set of independent variables
- The Logistic Regression is an analysis that is used to predict the missing data by using another related data values.
- It doesn't follow linearity between the Dependent variables and independent variables.
- It helps weather the model is good or not.

Linear Regression**:**

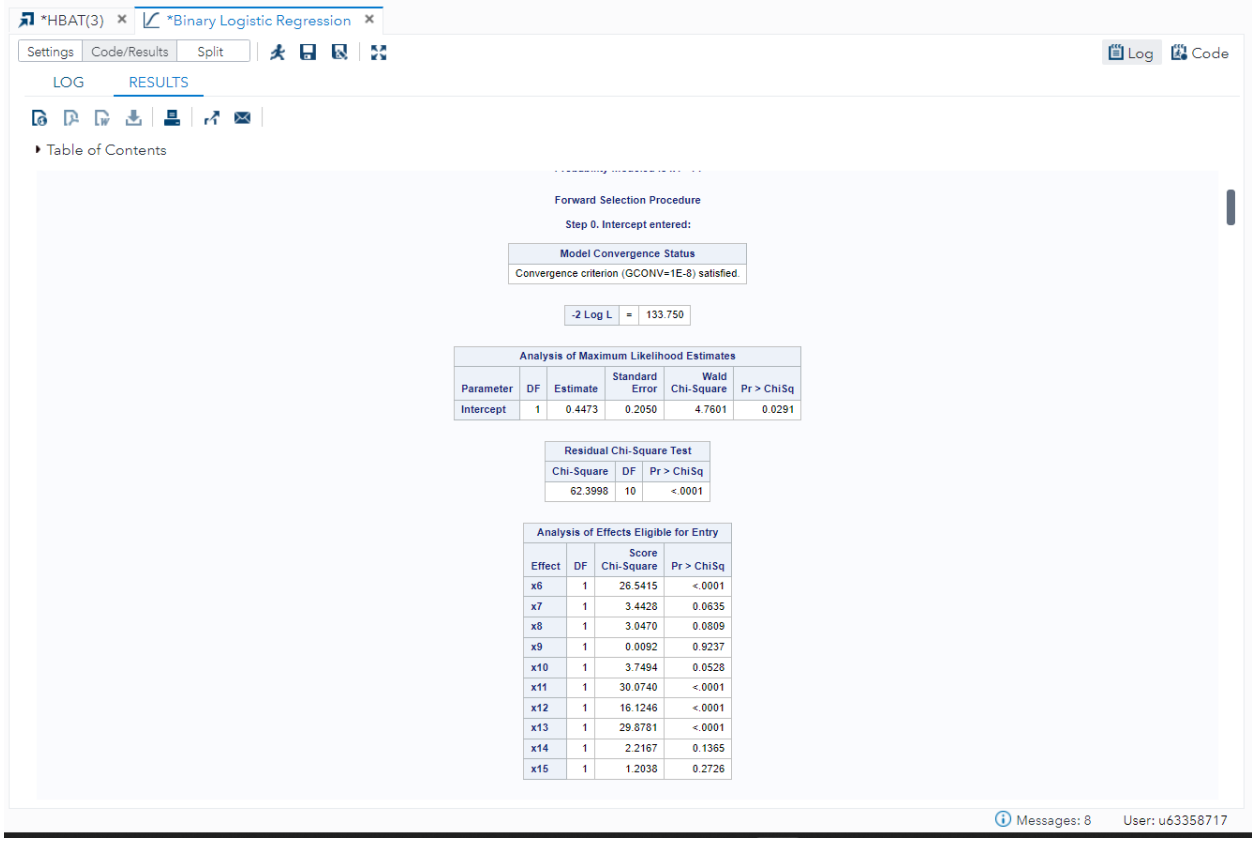
- Linear regression is a method that is used to build the relationship between dependent variable and independent variables. Linear regression assumes that there is a linear relationship between the dependent variable and the independent variables.
- Linear regression analysis is relatively simple an easy-to-interpret in various predictions.
- In these dependent missing data is predicted by using the independent variables.
- It is mostly used for continuous statistical variable prediction.

Linear Discriminant:

- Discriminant analysis is a multivariate analysis used to separate two or more groups of observations.
- In the linear discriminant of dependent variable is predicted by several independent variables.
- This follows the linearity.
- It is used for categorical independent variables and a continuous dependent variable.

2. **For the data set associated with this homework (HBAT) Using X4 as the non-metric response variableand (X6 up to X15) as the metric variables:**

   a. **Apply forward selection binary logistic regression (1 is the level of interest with single non-crosseffects) and report what variable is entered into the model after each step. (Use 0.05 significancelevel). Report the final summary of the regression model and the ROC curve and the area underthe ROC curve after each step.**



Step 1. Effect x11 entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	102.699
SC	138.355	107.910
-2 Log L	133.750	98.699

R-Square	0.2957	Max-rescaled R-Square	0.4009

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x6	1	9.8161	0.0017
x7	1	2.8646	0.0905
x8	1	1.0181	0.3130
x9	1	16.0663	<.0001
x10	1	3.5431	0.0598
x12	1	17.9196	<.0001
x13	1	12.8206	0.0003
x14	1	0.0177	0.8943
x15	1	2.7846	0.0952

Step 2. Effect x12 entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	83.119
SC	138.355	90.934
-2 Log L	133.750	77.119

R-Square	0.4324	Max-rescaled R-Square	0.5863

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x6	1	6.9973	0.0082
x7	1	10.9080	0.0010
x8	1	0.4477	0.5034
x9	1	8.6547	0.0033
x10	1	1.1844	0.2765
x13	1	4.4374	0.0352
x14	1	0.0203	0.8868
x15	1	2.8288	0.0926

Step 3. Effect x7 entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	73.494
SC	138.355	83.915
-2 Log L	133.750	65.494

R-Square	0.4947	Max-rescaled R-Square	0.6707

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x6	1	7.7382	0.0054
x8	1	0.3925	0.5310
x9	1	10.5702	0.0011
x10	1	2.9553	0.0856
x13	1	6.6448	0.0099
x14	1	0.0169	0.8967
x15	1	1.5048	0.2199

Step 4. Effect x9 entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	64.172
SC	138.355	77.197
-2 Log L	133.750	54.172

R-Square	0.5488	Max-rescaled R-Square	0.7441

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x6	1	5.7531	0.0165
x8	1	0.2646	0.6070
x10	1	2.8011	0.0942
x13	1	5.3643	0.0206
x14	1	0.0079	0.9290
x15	1	0.5725	0.4493

Step 5. Effect x6 entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	60.136
SC	138.355	75.767
-2 Log L	133.750	48.136

R-Square	0.5752	Max-rescaled R-Square	0.7799

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x8	1	0.0434	0.8349
x10	1	3.5158	0.0608
x13	1	3.8538	0.0496
x14	1	0.0034	0.9532
x15	1	1.1257	0.2887

Step 6. Effect x13 entered:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	58.094
SC	138.355	76.331
-2 Log L	133.750	44.094

R-Square	0.5920	Max-rescaled R-Square	0.8027

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
x8	1	0.0090	0.9244
x10	1	2.7967	0.0945
x14	1	0.0512	0.8210
x15	1	1.3770	0.2406

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

Summary of Forward Selection						
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Variable Label
1	x11	1	1	30.0740	<.0001	x11
2	x12	1	2	17.9196	<.0001	x12
3	x7	1	3	10.9080	0.0010	x7
4	x9	1	4	10.5702	0.0011	x9
5	x6	1	5	5.7531	0.0165	x6
6	x13	1	6	3.8538	0.0496	x13

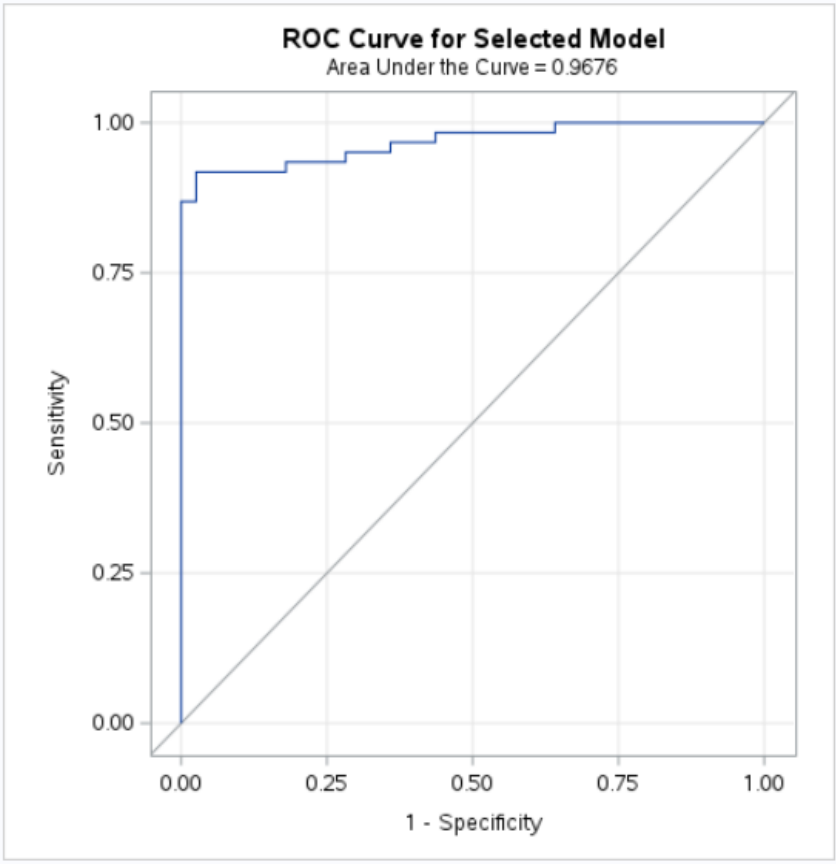
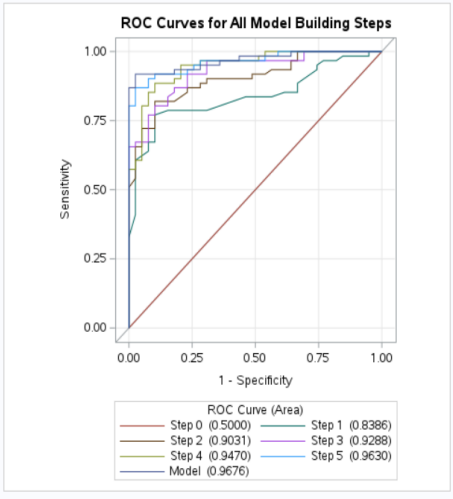


Table of Contents



Messages: 8 User: u63358717

b. Apply backward selection binary logistic regression (1 is the level of interest with single non-cross effects) and report what variable is eliminated from the model after each step. (Use 0.05significance level). Report the final summary of the regression model and the ROC curve and thearea under the ROC curve after each step.

Backward Elimination :

Backward Elimination Procedure

Step 0. The following effects were entered:

Intercept x6 x7 x8 x9 x10 x11 x12 x13 x14 x15

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	61.041
SC	138.355	89.698
-2 Log L	133.750	39.041

R-Square	0.6121	Max-rescaled R-Square	0.8300

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
x6	1	3.9809	0.0460
x7	1	8.1623	0.0043
x8	1	0.0038	0.9510
x9	1	5.3421	0.0208
x10	1	3.2318	0.0722
x11	1	8.8432	0.0029
x12	1	8.6225	0.0033
x13	1	2.7117	0.0996
x14	1	0.0013	0.9710
x15	1	2.0937	0.1479

Step 1. Effect x14 is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	59.042
SC	138.355	85.094
-2 Log L	133.750	39.042

R-Square	0.6121	Max-rescaled R-Square	0.8300

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
x6	1	4.0236	0.0449
x7	1	8.5561	0.0034
x8	1	0.0125	0.9109
x9	1	5.3478	0.0207
x10	1	3.2535	0.0713
x11	1	9.0240	0.0027
x12	1	8.7058	0.0032
x13	1	2.8027	0.0941
x15	1	2.0906	0.1482

Step 2. Effect x8 is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	57.055
SC	138.355	80.501
-2 Log L	133.750	39.055

R-Square	0.6121	Max-rescaled R-Square	0.8299

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
x6	1	4.0153	0.0451
x7	1	8.5996	0.0034
x9	1	5.4533	0.0195
x10	1	3.2408	0.0718
x11	1	9.4403	0.0021
x12	1	8.7440	0.0031
x13	1	2.8018	0.0942
x15	1	2.0636	0.1509

Step 3. Effect x15 is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	57.364
SC	138.355	78.205
-2 Log L	133.750	41.364

R-Square	0.6030	Max-rescaled R-Square	0.8177

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
x6	1	3.1636	0.0753
x7	1	9.3813	0.0022
x9	1	5.3541	0.0207
x10	1	2.5759	0.1085
x11	1	9.9298	0.0016
x12	1	9.4485	0.0021
x13	1	2.8394	0.0920

Step 4. Effect x10 is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	58.094
SC	138.355	76.331
-2 Log L	133.750	44.094

R-Square	0.5920	Max-rescaled R-Square	0.8027

Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
x6	1	3.3634	0.0667
x7	1	9.3911	0.0022
x9	1	5.6419	0.0175
x11	1	9.4554	0.0021
x12	1	10.0191	0.0015
x13	1	3.5327	0.0602

Step 5. Effect x6 is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	135.750	60.097
SC	138.355	75.728
-2 Log L	133.750	48.097

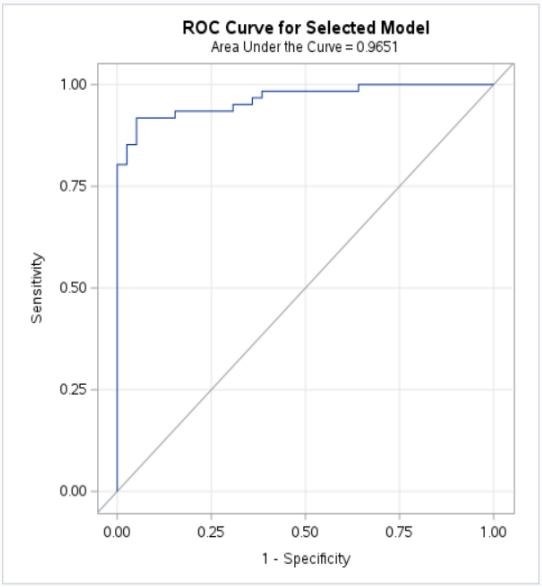
R-Square	0.5754	Max-rescaled R-Square	0.7802

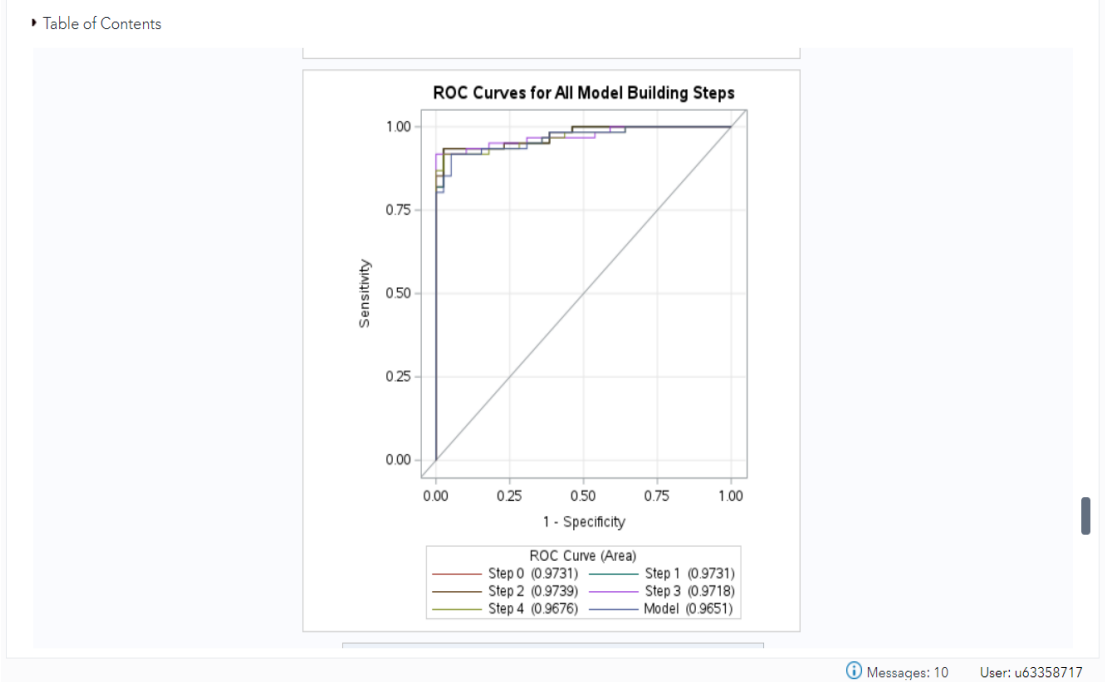
Analysis of Effects Eligible for Removal			
Effect	DF	Wald Chi-Square	Pr > ChiSq
x7	1	9.5863	0.0020
x9	1	7.4104	0.0065
x11	1	14.9434	0.0001
x12	1	10.2752	0.0013
x13	1	4.7979	0.0285

Note: No (additional) effects met the 0.05 significance level for removal from the model.

Summary of Backward Elimination						
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq	Variable Label
1	x14	1	9	0.0013	0.9710	x14
2	x8	1	8	0.0125	0.9109	x8
3	x15	1	7	2.0636	0.1509	x15
4	x10	1	6	2.5759	0.1085	x10
5	x6	1	5	3.3634	0.0667	x6

Table of Contents





c. Which selection method from (a) or (b) provides better model? Explain.

- In Forward Selection method, the area under ROC curve is 0.9676.
- In Backward Elimination the area under ROC curve is 0.9651.

If each model value is around 0.9 but that model is good.

Forward selection is better model compared to Backward elimination due to high area under ROC compared to backward elimination.