# Appendix

## Exploratory Data Figures

## SAS Code – EDA

## Objective 1 – SAS Output

## Assumptions

## Model Selection

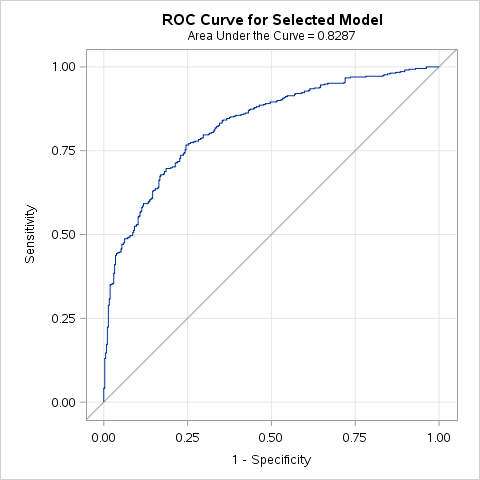
## SAS Code for Regressions

## Objective 2

## Logistic Regression 2

| **Hosmer and Lemeshow Goodness-of-Fit Test** | | |
| --- | --- | --- |
| **Chi-Square** | **DF** | **Pr > ChiSq** |
| 4.8489 | 8 | 0.7736 |
| **Model Fit Statistics** | | |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 1106.827 | 825.142 |
| **SC** | 1111.511 | 862.619 |
| **-2 Log L** | 1104.827 | 809.142 |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | -11.6910 | 3.0092 | 15.0941 | 0.0001 |
| **log.chlorides** | 1 | -1.0685 | 0.3215 | 11.0442 | 0.0009 |
| **log.total.sulfur.dio** | 1 | -0.3460 | 0.1332 | 6.7514 | 0.0094 |
| **log.sulphates** | 1 | 3.5737 | 0.4936 | 52.4218 | <.0001 |
| **log.alcohol** | 1 | 9.0324 | 1.1476 | 61.9431 | <.0001 |
| **volatile.acidity** | 1 | -3.2189 | 0.6516 | 24.4033 | <.0001 |
| **citric.acid** | 1 | -2.1847 | 0.6547 | 11.1336 | 0.0008 |
| **pH** | 1 | -2.0224 | 0.7030 | 8.2758 | 0.0040 |



## SAS Code for Regressions

proc import datafile="/home/anhainguyen820/sasuser.v94/wine\_train.csv"

dbms=dlm out=wine replace;

delimiter=',';

getnames=yes;

run;

data wine;

set wine;

"log.residual.sugar"N = log("residual.sugar"N);

"log.chlorides"N = log(chlorides);

"log.free.sulfur.dioxide"N = log("free.sulfur.dioxide"N);

"log.total.sulfur.dioxide"N = log("total.sulfur.dioxide"N);

"log.sulphates"N = log(sulphates);

"log.alcohol"N = log(alcohol);

run;

/\*Logistic regression\*/

proc logistic data=wine;

class Outcome / param=ref;

model Outcome(event='fine') = "log.residual.sugar"N

"log.chlorides"N

"log.free.sulfur.dioxide"N

"log.total.sulfur.dioxide"N

"log.sulphates"N

"log.alcohol"N

"fixed.acidity"N

"volatile.acidity"N

"citric.acid"N

density

pH/ selection=forward scale=none lackfit;

run;

## QDA

**Test of Homogeneity of Within Covariance Matrices**

| **Chi-Square** | **DF** | **Pr > ChiSq** |
| --- | --- | --- |
| 264.243869 | 66 | <.0001 |

| **Number of Observations and Percent Classified into Outcome** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **From Outcome** | | | **fine** | | **poor** | **Total** |
| **fine** | | | 312  72.73 | | 117  27.27 | 429  100.00 |
| **poor** | | | 76  20.49 | | 295  79.51 | 371  100.00 |
| **Total** | | | 388  48.50 | | 412  51.50 | 800  100.00 |
| **Priors** | | | 0.5 | | 0.5 |  |
| **Error Count Estimates for Outcome** | | | | |
|  | **fine** | **poor** | | **Total** |
| **Rate** | 0.2727 | 0.2049 | | 0.2388 |
| **Priors** | 0.5000 | 0.5000 | |  |

| **Number of Observations and Percent Classified into Outcome** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **From Outcome** | | | **fine** | | **poor** | **Total** |
| **fine** | | | 313  73.47 | | 113  26.53 | 426  100.00 |
| **poor** | | | 106  28.42 | | 267  71.58 | 373  100.00 |
| **Total** | | | 419  52.44 | | 380  47.56 | 799  100.00 |
| **Priors** | | | 0.5 | | 0.5 |  |
| **Error Count Estimates for Outcome** | | | | |
|  | **fine** | **poor** | | **Total** |
| **Rate** | 0.2653 | 0.2842 | | 0.2747 |
| **Priors** | 0.5000 | 0.5000 | |  |

## SAS Code for QDA

proc import datafile="/home/anhainguyen820/sasuser.v94/wine\_train.csv"

dbms=dlm out=wine replace;

delimiter=',';

getnames=yes;

run;

proc import datafile="/home/anhainguyen820/sasuser.v94/wine\_test.csv"

dbms=dlm out=test replace;

delimiter=',';

getnames=yes;

run;

data wine; set wine;

"log.residual.sugar"N = log("residual.sugar"N);

"log.chlorides"N = log(chlorides);

"log.free.sulfur.dioxide"N = log("free.sulfur.dioxide"N);

"log.total.sulfur.dioxide"N = log("total.sulfur.dioxide"N);

"log.sulphates"N = log(sulphates);

"log.alcohol"N = log(alcohol);

run;

data test; set test;

"log.residual.sugar"N = log("residual.sugar"N);

"log.chlorides"N = log(chlorides);

"log.free.sulfur.dioxide"N = log("free.sulfur.dioxide"N);

"log.total.sulfur.dioxide"N = log("total.sulfur.dioxide"N);

"log.sulphates"N = log(sulphates);

"log.alcohol"N = log(alcohol);

run;

proc discrim data=wine pool=test testdata=test;

class Outcome;

var "log.residual.sugar"N

"log.chlorides"N

"log.free.sulfur.dioxide"N

"log.total.sulfur.dioxide"N

"log.sulphates"N

"log.alcohol"N

"fixed.acidity"N

"volatile.acidity"N

"citric.acid"N

density

pH;

run;