

```

data wine;
  infile '/home/yleung2/6372 HW/6372Project1/winequality-red.csv' dlm=';' firstobs=2;
  input fixed_acidity volatile_acidity citric_acid residual_sugar chlorides free_sulfur_dioxide total_sulfur_dioxide den:
run;
/*proc print;run;*/

proc univariate data=wine plot ;
var fixed_acidity volatile_acidity citric_acid residual_sugar chlorides free_sulfur_dioxide total_sulfur_dioxide density pl
run;

proc glm data = wine PLOTS=(DIAGNOSTICS RESIDUALS);
class fixed_acidity volatile_acidity citric_acid residual_sugar chlorides free_sulfur_dioxide total_sulfur_dioxide density
model quality= fixed_acidity volatile_acidity citric_acid residual_sugar chlorides free_sulfur_dioxide total_sulfur_dioxide
run;

/*log fixed_acidity, free_sulfur_dioxide and total_sulfur_dioxide */
data LogWine;
set wine;
log_fixed_acidity=log10(fixed_acidity);
log_free_sulfur_dioxide=log10(free_sulfur_dioxide);
log_total_sulfur_dioxide=log10(total_sulfur_dioxide);

proc univariate data=LogWine plot ;
var log_fixed_acidity volatile_acidity citric_acid residual_sugar chlorides log_free_sulfur_dioxide log_total_sulfur_dioxide
run;

proc corr data=LogWine plots(maxpoints = none)=matrix(histogram);
run;

proc glm data = LogWine PLOTS=(DIAGNOSTICS RESIDUALS);
class fixed_acidity volatile_acidity citric_acid residual_sugar chlorides free_sulfur_dioxide total_sulfur_dioxide density
model quality= fixed_acidity volatile_acidity citric_acid residual_sugar chlorides free_sulfur_dioxide total_sulfur_dioxide
run;

data wine2;
set LogWine;
RandNumber = ranuni(11);
quality = quality
;

data train;
set wine2;
if RandNumber <= 1/4 then delete;
run;

data test;
set wine2;
if RandNumber > 1/4 then delete;
run;

/*
proc print data = train;
run;

proc print data = test;
run;
*/

/*OLS model, standard stepwise model to determine the predictors in a linear model*/
ods graphics on;
Title "LASSO steps set at 120";
proc glmselect data=train testdata = test
  seed=1 plots=all /*(stepAxis=number)=(criterionPanel ASEPlot CRITERIONPANEL)*/;
model quality= log_fixed_acidity volatile_acidity citric_acid residual_sugar chlorides log_free_sulfur_dioxide log_total_s

run;
quit;
ods graphics off;

/*Select model from the OLS and then apply Lasso and Elastic Net*/
/*Lasso model that set to end the selection at 120 steps, 10 fold cross validation*/
ods graphics on;
Title "LASSO steps set at 120";
proc glmselect data=train testdata = test seed=3 plots=all;

```

```
model quality= log_fixed_acidity volatile_acidity citric_acid residual_sugar chlorides log_free_sulfur_dioxide log_total_sulfur_dioxide;
*modelAverage tables=(EffectSelectPct(all) ParmEst(all)) alpha=0.05;

run;
quit;
ods graphics off;

/*Elastic Net */
/*Elastic net model that set to end the selection at 120 steps, 10 fold cross validation. This model supposed to be a middle ground between Ridge and Lasso.
/*Ridge sets coeff close to 0 but not equal to 0 and Lasso will set coeff to exactly 0. Elastic Net can be 0, not equal to 0.

ods graphics on;
Title "Elastic Net";
proc glmselect data=train testdata = test plots= all;
    model quality= log_fixed_acidity volatile_acidity citric_acid residual_sugar chlorides log_free_sulfur_dioxide log_total_sulfur_dioxide;
    store out=glmselectStore;
run;
quit;
ods graphics off;
```