

```
/** Rohit Bhardwaj **
```

```
Multivariate project on alcohol consumption
```

```
***/
```

```
/**Import complete data**/
```

```
PROC IMPORT OUT= WORK.All_data
```

```
    DATAFILE= "E:\MS IS studies\3rd Sem\Multivariate\project\data1.csv"
```

```
    DBMS=CSV REPLACE;
```

```
    GETNAMES=YES;
```

```
    DATAROW=2;
```

```
RUN;
```

```
title"Logistic Regression model for Other";
```

```
proc logistic data=alldata_z descending;
```

```
    class FatherIH(ref='0') MotherIH(ref='0') Sex(ref='0')
```

```
        Work(ref='0') Retired(ref='0') work_12_months(ref='0')
```

```
    food_stamps(ref='0') medicaid(ref='0') private_health(ref='0') / param =ref;
```

```
    model Drink = FatherIH MotherIH Sex Work Retired work_12_months food_stamps
```

```
    medicaid private_health;
```

```
quit;
```

```
title"Logistic Regression model for marital status";
```

```
proc logistic data=All_data descending;
```

```
    class marital_status_divoced(ref='0') marital_status_live_in(ref='0')
```

```
        marital_satatus_married(ref='0') marital_satatus_never_married(ref='0')
```

```
    marital_satatus_seperated(ref='0')
```

```
        marital_satatus_widowed(ref='0')/ param =ref;
```

```
    model Drink = marital_satatus_widowed marital_satatus_seperated
```

```
    marital_satatus_never_married marital_satatus_married marital_status_live_in
```

```
    marital_status_divoced;
```

```
quit;
```

```

title"Logistic Regression model for drinking based on region";
proc logistic data=All_data descending;
    class Division_East_North_central(ref='0') Division_East_North_central (ref='0')
        Division_Middle_Atlantic(ref='0') division_mountain(ref='0')
division_new_england(ref='0') division_pacific(ref='0')
        Division_south_Atlantic(ref='0') Division_west_north_central(ref='0')
Division_west_south_central(ref='0')/ param =ref;
    model Drink = Division_East_North_central Division_East_North_central
Division_Middle_Atlantic division_mountain division_new_england
        division_pacific Division_south_Atlantic Division_west_north_central
Division_west_south_central;
quit;

```

```

title"Logistic Regression model for Ethnicity";
proc logistic data=All_data descending;
    class Ethnicity_Asian(ref='0') Ethnicity_Black (ref='0') Ethnicity_Hispanic(ref='0')
Ethnicity_Native(ref='0') Ethnicity_white(ref='0')/ param =ref;
    model Drink = Ethnicity_Asian Ethnicity_Black Ethnicity_Hispanic Ethnicity_Native
Ethnicity_white
    /SS1 SS2 STB dwProb VIF selection= backward; *stb means standarized beta, gives
values for standarized data;

                                *ss1 : sequential sum of the square gives individual
sum of the square of regression of that variable;
                                *ss2 : partial sum of the square, that mean what will
be the difference between sum of the square if that bvariable is not there
                                in the equation but others variables are;
                                *goes through list, keep highest ssr, and goes
sequential, and keeps highest ssr;
                                *forward: adding variables one after the other
starting from 1;
                                *backward: starting from all, removes the most
insignificant;
                                *MaxR: picks up the most significat with the best
combinations;
    OUTPUT OUT=reg_cerealOUT PREDICTED= predict_cereal RESIDUAL=Res
L95M=C_I95m U95M=C_u95m L95=C_I95 U95=C_u95
    rstudent=C_rstudent h=lev cookd=Cookd dffits=dffit
    STDP=C_spredicted STDR=C_s_residual STUDENT=C_student ;
quit;

```

title "Multi Regression for Drinking";

```
proc reg data=All_data outest=multi_reg_out ;
```

```
model Drink = schooling number_of_persons income weight age
```

```
FatherIH MotherIH Sex Work Retired
```

```
work_12_months food_stamps medicaid private_health
```

```
marital_satatus_widowed marital_satatus_seperated
```

```
marital_satatus_never_married
```

```
marital_satatus_married marital_status_live_in
```

```
marital_status_divoced
```

```
Ethnicity_Asian Ethnicity_Black Ethnicity_Hispanic
```

```
Ethnicity_Native Ethnicity_white
```

```
Division_East_North_central Division_East_North_central
```

```
Division_Middle_Atlantic division_mountain division_new_england
```

```
division_pacific Division_south_Atlantic
```

```
Division_west_north_central Division_west_south_central
```

```
 /SS1 SS2 STB dwProb VIF selection= backward; *stb means standarized beta,  
gives values for standarized data;
```

```
 *ss1 : sequential sum of the square gives individual  
sum of the square of regression of that variable;
```

```
 *ss2 : partial sum of the square, that mean what will  
be the difference between sum of the square if that bvariable is not there  
in the equation but others variables are;
```

```
 *goes through list, keep highest ssr, and goes  
sequential, and keeps highest ssr;
```

```
 *forward: adding variables one after the other  
starting from 1;
```

```
 *backward: starting from all, removes the most  
insignificant;
```

```
 *MaxR: picks up the most significat with the best  
combinations;
```

```
 OUTPUT OUT=reg_cerealOUT PREDICTED= predict_cereal RESIDUAL=Res
```

```
 L95M=C_l95m U95M=C_u95m L95=C_l95 U95=C_u95
```

```
 rstudent=C_rstudent h=lev cookd=Cookd dffits=dffit
```

```
 STDP=C_spredicted STDR=C_s_residual STUDENT=C_student ;
```

```
quit;
```

title "Multi Regression for Drinking";

```
proc reg data=All_data outest=multi_reg_out ;
  model Drink = schooling number_of_persons income weight age
              FatherIH MotherIH Sex Work Retired height_inches
              work_12_months food_stamps medicaid private_health
              /SS1 SS2 STB dwProb VIF selection= backward; *stb means standarized beta,
gives values for standarized data;

                                *ss1 : sequential sum of the square gives individual
sum of the square of regression of that variable;
                                *ss2 : partial sum of the square, that mean what will
be the difference between sum of the square if that bvariable is not there
                                in the equation but others variables are;
                                *goes through list, keep highest SSR, and goes
sequential, and keeps highest SSR;
                                *forward: adding variables one after the other
starting from 1;
                                *backward: starting from all, removes the most
insignificant;
                                *MaxR: picks up the most significant with the best
combinations;
  OUTPUT OUT=reg_cereal OUT_PREDICTED=predict_cereal RESIDUAL=Res
L95M=C_l95m U95M=C_u95m L95=C_l95 U95=C_u95
  rstudent=C_rstudent h=lev cookd=Cookd dffits=dffit
  STDP=C_spredicted STDR=C_s_residual STUDENT=C_student ;

quit;
```

```
proc reg data=All_data outest=multi_reg_out ;
  model Drink = income age MotherIH Sex Retired work_12_months food_stamps
              medicaid private_health schooling
              number_of_persons height_inches
              /SS1 SS2 STB dwProb VIF selection= Setpwise; *stb means standarized beta,
gives values for standarized data;

                                *ss1 : sequential sum of the square gives individual
sum of the square of regression of that variable;
                                *ss2 : partial sum of the square, that mean what will
be the difference between sum of the square if that bvariable is not there
                                in the equation but others variables are;
```

sequential, and keeps highest ssr;  
 starting from 1;  
 insignificant;  
 combinations;

\*goes through list, keep highest ssr, and goes  
 \*forward: adding variables one after the other  
 \*backward: starting from all, removes the most  
 \*MaxR: picks up the most significant with the best

```

    OUTPUT OUT=reg_cereal OUT_PREDICTED= predict_cereal RESIDUAL=Res
L95M=C_l95m U95M=C_u95m L95=C_l95 U95=C_u95
    rstudent=C_rstudent h=lev cookd=Cookd dffits=dffit
    STDP=C_s_predicted STDR=C_s_residual STUDENT=C_student ;

quit;

```

```

** Normalize the data **;
PROC STANDARD DATA=All_data
    MEAN=0.5 STD=0.5
    OUT=alldata_z;
var income weight age schooling number_of_persons health;
run;

```

/\*\*corelation analysis \*\*/

```

proc corr data=all_data cov;
var income weight age fatherIH MotherIH Sex work Retired work_12_months food_stamps
medicaid private_health
    deaths schooling number_of_persons drink height_inches health
    Division_East_North_central Division_East_South_central Division_Middle_Atlantic
division_mountain division_new_england
    division_pacific Division_south_Atlantic Division_west_north_central
Division_west_south_central
    ms_divorced ms_live_in ms_married ms_never_married ms_seperated ms_widowed
    Ethnicity_Asian Ethnicity_Black Ethnicity_Hispanic Ethnicity_Native Ethnicity_white;
run;

```

```
/** logistic regression **/
```

```
title "Logistic Regression model for Other";
proc logistic data=alldata_z ;
    model Drink = income weight age fatherIH MotherIH Sex Retired work_12_months
    food_stamps medicaid private_health
    deaths schooling number_of_persons height_inches health
    Division_East_North_central Division_East_South_central Division_Middle_Atlantic
    division_mountain division_new_england
    division_pacific Division_south_Atlantic Division_west_north_central
    Division_west_south_central
    ms_divorced ms_live_in ms_married ms_never_married ms_seperated ms_widowed
    Ethnicity_Asian Ethnicity_Black Ethnicity_Hispanic Ethnicity_Native Ethnicity_white
    /selection= backward;
quit;
```

```
/** multiple regression **/
```

```
title "Logistic Regression for Drinking";
```

```
proc reg data=All_data outest=multi_reg_out ;
    model Drink = income weight age fatherIH MotherIH Sex Retired work_12_months
    food_stamps medicaid private_health
    deaths schooling number_of_persons height_inches health
    Division_East_North_central Division_East_South_central Division_Middle_Atlantic
    division_mountain division_new_england
    division_pacific Division_south_Atlantic Division_west_north_central
    Division_west_south_central
    ms_divorced ms_live_in ms_married ms_never_married ms_seperated ms_widowed
    Ethnicity_Asian Ethnicity_Black Ethnicity_Hispanic Ethnicity_Native Ethnicity_white
    /SS1 SS2 STB dwProb VIF selection= backward; *stb means standarized beta,
    gives values for standarized data;
```

\*ss1 : sequential sum of the square gives individual  
sum of the square of regression of that variable;

\*ss2 : partial sum of the square, that mean what will  
be the difference between sum of the square if that bvariable is not there

in the equation but others variables are;

\*goes through list, keep highest ssr, and goes

sequential, and keeps highest ssr;

\*forward: adding variables one after the other

starting from 1;

insignificant;  
 combinations;  
 OUTPUT OUT=reg\_cereal OUT PREDICTED=predict\_cereal RESIDUAL=Res  
 L95M=C\_l95m U95M=C\_u95m L95=C\_l95 U95=C\_u95  
 rstudent=C\_rstudent h=lev cookd=Cookd dffits=dffit  
 STDP=C\_spredicted STDR=C\_s\_residual STUDENT=C\_student ;  
 quit;

OUTPUT OUT=reg\_cerealOUT PREDICTED= predict\_cereal RESIDUAL=Res  
L95M=C\_l95m U95M=C\_u95m L95=C\_l95 U95=C\_u95  
rstudent=C\_rstudent h=lev cookd=Cookd dffits=dffit  
STDP=C\_s\_predicted STDR=C\_s\_residual STUDENT=C\_student ;

```
/** principal component */
```

```
class FatherIH(ref='0') MotherIH(ref='0') Sex(ref='0') Work(ref='0') Retired(ref='0')
    work_12_months(ref='0') food_stamps(ref='0') medicaid(ref='0')
private_health(ref='0')
    Division_East_North_central(ref='0') Division_East_South_central (ref='0')
    Division_Middle_Atlantic(ref='0') division_mountain(ref='0')
division_new_england(ref='0') division_pacific(ref='0')
    Division_south_Atlantic(ref='0') Division_west_north_central(ref='0')
Division_west_south_central(ref='0')
    ms_divorced(ref='0') ms_live_in(ref='0') ms_married(ref='0')
ms_never_married(ref='0') ms_seperated(ref='0') ms_widowed(ref='0')
    Ethnicity_Asian(ref='0') Ethnicity_Black(ref='0') Ethnicity_Hispanic(ref='0')
Ethnicity_Native(ref='0') Ethnicity_white(ref='0')
    / param =ref;
```

model Drink = income weight age fatherIH MotherIH Sex Retired work\_12\_months food\_stamps  
medicaid private\_health  
deaths schooling number\_of\_persons height\_inches health  
Division\_East\_North\_central Division\_East\_South\_central Division\_Middle\_Atlantic  
division\_mountain division\_new\_england  
Division\_south\_Atlantic Division\_west\_north\_central Division\_west\_south\_central  
ms\_divorced ms\_live\_in ms\_seperated ms\_widowed  
Ethnicity\_Asian Ethnicity\_Black Ethnicity\_Hispanic Ethnicity\_white