ods pdf file="C:\Users\HP\OneDrive\Desktop\BS 805\project\_output.pdf";

libname bs805 "C:\Users\HP\OneDrive\Desktop\BS 805\";

/\*\*linking baseline and clinical data for cohort 1\*\*/

/\*\*used match-merging technique to link the datasets\*\*/

**proc** **sort** data=bs805.baseline\_cohort1;

by id;

**run**;

**proc** **sort** data=bs805.clinical\_cohort1;

by id;

**run**;

**data** cohort1;

merge bs805.baseline\_cohort1 bs805.clinical\_cohort1;

by id;

**run**;

/\*\*linking baseline and clinical data for cohort 2\*\*/

/\*\*used match-merging technique to link the datasets\*\*/

**proc** **sort** data=bs805.baseline\_cohort2;

by id;

**run**;

**proc** **sort** data=bs805.clinical\_cohort2;

by id;

**run**;

**data** cohort2;

merge bs805.baseline\_cohort2 bs805.clinical\_cohort2;

by id;

**run**;

/\*\*combining the two cohort datasets into a temp dataset\*\*/

**data** combined\_cohort;

set cohort1 cohort2 ;

by id;

**run**;

/\*\*creating a new combined dataset for analyses\*\*/

**proc** **format**;

value BMI\_catf **1**='Normal'

**2**='Overweight'

**3**='Obese';

value MALEf **0**='female'

**1**='male';

**run**;

**data** new\_combined\_cohort;

set combined\_cohort;

weight\_kg = weight\*(**0.453592**);/\*converting weight in lbs to kg\*/

height\_m = height \* (**0.0254**);/\*converting height in inchies to meters\*/

BMI = (weight\_kg/(height\_m\*height\_m) ); /\*creating a BMI variable\*/

if BMI < **19** then delete;/\*excluding BMI values less than 19\*/

if BMI=**.** then BMI\_cat=**.**;/\*grouping BMI into 3 categories\*/

else if BMI < **25** then BMI\_cat = **1**;

else if **25** <= BMI < **30** then BMI\_cat =**2**;

else if BMI >= **30** then BMI\_cat = **3**;

if (**19** <= BMI < **25**) then BMI1=BMI;/\*created 3 piecewise variables for each BMI interval as defined in the BMI cat variable\*/

else if BMI >=**25** then BMI1=**25**;

if(**19** <= BMI < **25**) then BMI2=**25**;

else if (**25** <= BMI < **30**) then BMI2=BMI;

else if BMI >= **30** then BMI2=**30**;

if (**19** <= BMI < **30**) then BMI3=**30**;

else if BMI >=**30** then BMI3=BMI;

waist\_hip\_ratio = ((waist)/(hip));/\*created a waist to hip ratio variables in meters\*/

Mean\_SBP = mean(of BP\_1S BP\_2S);/\*mean systolic blood pressure\*/

glyhb\_ln = log(glyhb);/\*natural log of HbA1c\*/

if gender='male' then MALE=**1**; else MALE = **0**;/\*dummy variable with male as 1 and female as 0\*/

label MALE="Sex"

chol="Total Cholestrol"

weight\_kg="Weight in kilograms"

height\_m ="Height in meters"

stab\_glu="Blood Glucose Level"

glyhb="hemoglobin A1c"

waist="waist circumference in inches"

hip="hip circumference in inches"

BMI="Body Mass Index"

BMI\_cat="Categorical BMI variable "

waist\_hip\_ratio="Waist to Hip Ratio"

glyhb\_ln="Natural logarithm of HbA1c"

Mean\_sbp="Mean systolic blood pressure";

format BMI\_cat BMI\_catf. MALE MALEf.;

**run**;

**proc** **means** data=new\_combined\_cohort;

var MEAN\_SBP;

**run**;

/\*\*Descriptive Statistics for all study variables\*\*/

/\*\*table displays distribution of variables by BMI categories\*\*/

/\*Checking for normality of continuous variables within BMI categories to know which measures of central tendency and dispersion to report\*/

**proc** **univariate** data=new\_combined\_cohort noprint;

histogram BMI waist\_hip\_ratio age waist hip Mean\_SBP glyhb chol stab\_glu weight\_kg height\_m;

class BMI\_cat;

title'checking for normality and skewness of continuous variables';

**run**;

title;

/\*reporting means and standard deviation for these variables because data distribution looks normal\*/

**proc** **means** data=new\_combined\_cohort maxdec=**2** mean std;

var waist\_hip\_ratio age waist hip chol weight\_kg height\_m;

class BMI\_cat;

title 'descriptive statistics for continuous variables with normal distribution';

**run**;

title;

/\*reporting median and IQR for these variables because data distribtuion looks skewed\*/

**proc** **means** data=new\_combined\_cohort maxdec=**2** median Q1 Q3;

var stab\_glu glyhb BMI Mean\_SBP;

class BMI\_cat;

title 'descriptive statistics for continuous variables with skewed distribution';

**run**;

title;

/\*for categorical variables\*/

**proc** **freq** data=new\_combined\_cohort;

tables MALE \* BMI\_cat / norow nocol nocum ;

title 'descriptive statistics for categorical variable (sex)';

**run**;

title;

/\*\*One-way ANOVA to to test a common mean of HbA1C across BMI categories using log vs non-log version of HbA1C variable\*\*/

/\*one-way anova for non log version\*/

**proc** **glm** data=new\_combined\_cohort;

class BMI\_cat (ref='Normal');

model glyhb= BMI\_cat/solution;

title'one-way ANOVA across BMI categories for non log version of HbA1C';

**run**;

title;

/\*one-way anova for log version\*/

**proc** **glm** data=new\_combined\_cohort;

class BMI\_cat (ref='Normal');

model glyhb\_ln= BMI\_cat/solution;

lsmeans BMI\_cat/adjust=tukey;

title'one-way ANOVA across BMI categories for log version of HbA1C';

**run**;

title;

/\*\*best approach modeling approach to assess relationship between BMI and HbA1C\*\*/

/\*\*a. creating dummy variables for BMI\*\*/

**data** new\_combined\_cohort;

set new\_combined\_cohort;

\*\*creating dummy variables for BMI categories;

\*\* Normal (1) is the reference group;

if BMI\_cat=**2** then over\_weight=**1**; else over\_weight=**0**;

if BMI\_cat =**3** then obese=**1**; else obese=**0**;

**run**;

**proc** **glm** data=new\_combined\_cohort;

model glyhb\_ln=over\_weight obese/solution;

title'Linear regression model for log transformed HbA1C using dummy variables for BMI';

**run**;

title;

/\*\*b. BMI categories as an ordinal variable\*\*/

**proc** **glm** data=new\_combined\_cohort;

model glyhb\_ln=BMI\_cat/solution;

title'Linear regression model for log transformed HbA1C using BMI categories as an ordinal variable';

**run**;

title;

/\*\*c. BMI as a continuous variable\*\*/

**proc** **reg** data=new\_combined\_cohort;

title1'Linear regression model for log transformed HbA1C using BMI as a continuous variable';

model glyhb\_ln=BMI;

**run**;

title;

/\*\*d. Piecewise linear model using piecewise variables\*/

**proc** **reg** data=new\_combined\_cohort;

model glyhb\_ln= BMI1 BMI2 BMI3/stb;

test BMI1 = BMI2;

test BMI2 = BMI3;

output out=piece pred=pred\_glyhb\_ln;

title'Piecewise linear model for log transformed HbA1C using piecewise variables for BMI';

**run**;

title;

**proc** **sort** data=piece;

by BMI;

**run**;

symbol1 value =dot color=black;

symbol2 line=**1** interpol =join color=black;

**proc** **gplot** data=piece;

plot glyhb\_ln\*BMI pred\_glyhb\_ln\*BMI / overlay;

title 'A plot of the slopes of the piecewise model with observed values of log transformed HbA1C and BMI';

**run**;

title;

/\*checking for linearity assumption between log transformed HbA1C and BMI\*/

**proc** **gplot** data=new\_combined\_cohort;

plot glyhb\_ln \* BMI;

title' Log transformed HbA1C with BMI plot';

**run**;

title;

/\*\*Assessing effect modifiers on relationship between log transformed HbA1C and BMI\*\*/

/\*\*Age as an effect modifier using BMI as a categorical variable\*\*/

**proc** **glm** data=new\_combined\_cohort;

class BMI\_cat (ref='Normal');

model glyhb\_ln = BMI\_cat age BMI\_cat\*age/solution;

title' Assessing age as an effect modifier between log transformed HbA1C and BMI';

**run**;

title;

/\*\*Sex as an effect modifier using BMI as a categorical variable\*\*/

**proc** **glm** data=new\_combined\_cohort;

class BMI\_cat (ref='Normal') MALE (ref='female');

model glyhb\_ln = BMI\_cat MALE BMI\_cat\*MALE/solution;

title' Assessing age as an effect modifier between log transformed HbA1C and BMI';

**run**;

title;

/\*\*A. Best predictor out of waist, hip, waist to hip ratio\*\*/

/\*\*waist\*\*/

**proc** **reg** data=new\_combined\_cohort;

model glyhb\_ln = waist;

title'Linear regression of log-transformed HbA1C and waist';

**run**;

title;

/\*\*waist to hip ratio\*\*/

**proc** **reg** data=new\_combined\_cohort;

model glyhb\_ln = waist\_hip\_ratio;

title'Linear regression of log-transformed HbA1C and waist to hip ratio';

**run**;

title;

/\*\*hip\*\*/

**proc** **reg** data=new\_combined\_cohort;

model glyhb\_ln = hip;

title'Linear regression of log-transformed HbA1C and hip';

**run**;

title;

/\* Means Plots for each of the three variables to assess linearity with log transformed HbA1C\*/

**proc** **rank** groups=**10** data=new\_combined\_cohort out=three;

var waist waist\_hip\_ratio hip;

ranks rwaist rwaist\_hip\_ratio rhip;

**run**;

/\*\*waist\*\*/

**proc** **sort** data=three;

by rwaist;

**run**;

**proc** **means** data=three noprint;

by rwaist;

var glyhb\_ln waist;

output out=waiststr mean=glyhb\_lnmean waistmean;

**run**;

**proc** **plot** data=waiststr;

plot glyhb\_lnmean\*waistmean='W';

title 'Plot of Decile Means for Waist';

**run**;

title;

/\*\*hip\*\*/

**proc** **sort** data=three;

by rhip;

**run**;

**proc** **means** data=three noprint;

by rhip;

var glyhb\_ln hip;

output out=hipstr mean=glyhb\_lnmean hipmean;

**run**;

**proc** **plot** data=hipstr;

plot glyhb\_lnmean\*hipmean='H';

title1 'Plot of Decile Means for Hip';

**run**;

/\*\*waist to hip ratio\*\*/

**proc** **sort** data=three;

by rwaist\_hip\_ratio;

**run**;

**proc** **means** data=three noprint;

by rwaist\_hip\_ratio;

var glyhb\_ln waist\_hip\_ratio;

output out=waist\_hip\_ratiostr mean=glyhb\_lnmean waist\_hip\_ratiomean;

**run**;

**proc** **plot** data=waist\_hip\_ratiostr;

plot glyhb\_lnmean\*waist\_hip\_ratiomean='W';

title 'Plot of Decile Means for Waist to Hip ratio';

**run**;

title;

/\*\*multi-collinearity analysis, outliers and influence points\*\*/

**proc** **reg** data=new\_combined\_cohort;

model glyhb\_ln=waist waist\_hip\_ratio hip/ tol vif collinoint r;

id;

title'Multicollinearity analysis of waist, hip, and waist to hip ratio';

output out=two pred=p\_glyhb\_ln student=str\_glyhb\_ln residual=resid\_glyhb\_ln cookd=cooksd;

**run**;

title;

**proc** **print** data=two;

id id;

where abs(str\_glyhb\_ln)>=**3**;

var str\_glyhb\_ln p\_glyhb\_ln resid\_glyhb\_ln waist waist\_hip\_ratio hip;

title'outliers among waist, hip and waist to hip ratio';

**run**;

title;

**proc** **print** data=two;

id id;

where cooksd >= **4**/**386**;

var str\_glyhb\_ln p\_glyhb\_ln resid\_glyhb\_ln cooksd waist waist\_hip\_ratio hip;

title'influence points among waist, hip and waist to hip ratio';

**run**;

title;

/\*\*b\*\*/

/\*\*assessing for confounding\*\*/

/\*Crude model\*/

**proc** **glm** data=new\_combined\_cohort;

class BMI\_cat(ref='Normal');

model glyhb\_ln = BMI\_cat/solution;

title'Crude model of log transformed HbA1C and BMI categories';

**run**;

title;

/\*Adjusted model\*/

**proc** **glmselect** data=new\_combined\_cohort;

class BMI\_cat(ref='Normal') MALE(ref='female');

model glyhb\_ln = BMI\_CAT age chol waist MALE Mean\_SBP/selection=none stb;

title'Adjusted model of log transformed HbA1C and BMI categories with other predictors';

**run**;

title;

/\*\*Assessing whether the multivariate regression model is over-fit or under-fit\*\*/

/\*\*LASSO based linear regression\*\*/

**proc** **glmselect** data=new\_combined\_cohort;

class BMI\_cat(ref='Normal') MALE(ref='female');

model glyhb\_ln=BMI\_cat age chol waist MALE Mean\_SBP / selection=lasso (stop=none choose=aic);

title'LASSO based linear regression for the multivariate linear regression model';

**run**;

title;

/\*\*Backward selection based linear regression\*\*/

**proc** **glmselect** data=new\_combined\_cohort;

class BMI\_cat(ref='Normal') MALE(ref='female');

model glyhb\_ln=BMI\_cat age chol waist MALE Mean\_SBP / selection=backward (choose=aic stop=sl) sle=**0.05** sls=**0.05** select=aic;

title'Backward selection linear regression for the multivariate linear regression model';

**run**;

title;

ods pdf close;