```
Libname p1 '/home/u63581499/Biostat/data/';

options fmtsearch=(p1.ps24format);
proc contents data= p1.ps24;
run;

 data p1.working;
 set p1.ps24;
 keep  CHCOCNCR AVE_DRINK  BMI SEX DIABETES AGEGROUP SMOKDAY2;
 run;

proc print data=p1.working (obs=10);
 run;

 /*task 2 comapare distributions and summarize*/
/*numeric*/

PROC FREQ DATA=p1.working;
Tables CHCOCNCR;
run;

PROC UNIVARIATE DATA=p1.working normal;
   VAR AVE_DRINK BMI;
   CLASS CHCOCNCR;
   TITLE " normal, Iqr";
RUN;

/*Since variables are not normally distributed I used wilcoxon test.*/

Proc NPAR1WAY DATA=p1.working Wilcoxon;
  VAR AVE_DRINK BMI;
   CLASS CHCOCNCR;
   TITLE "p value";
run;

/*categorical*/
/*I used chi square test and reported adjusted chi square for those with two levels and chi square for those with more than two levels.*/

PROC FREQ DATA=p1.working;
 Tables DIABETES*CHCOCNCR SEX*CHCOCNCR AGEGROUP*CHCOCNCR SMOKDAY2*CHCOCNCR/ CHISQ NOPERCENT NOROW EXPECTED;
RUN;

  /*task 3 log reg*/

 PROC LOGISTIC DATA=p1.working;
   MODEL CHCOCNCR (event='Yes') = AVE_DRINK;
   TITLE 'Logistic Regression';
RUN;

 PROC LOGISTIC DATA=p1.working;
 class sex (ref= 'Male')/param=ref;
   MODEL CHCOCNCR (event='Yes') = SEX;
   TITLE 'Logistic Regression';
RUN;

 PROC LOGISTIC DATA=p1.working;
 class agegroup (ref= 'Age 45 - 54')/param=ref;
   MODEL CHCOCNCR (event='Yes') = AGEGROUP;
   TITLE 'Logistic Regression';
RUN;

 PROC LOGISTIC DATA=p1.working;
 class SMOKDAY2 (ref= 'Every day')/param=ref;
   MODEL CHCOCNCR (event='Yes') = SMOKDAY2;
   TITLE 'Logistic Regression';
RUN;

 PROC LOGISTIC DATA=p1.working;
   MODEL CHCOCNCR (event='Yes') = BMI;
   TITLE 'Logistic Regression';
RUN;

 PROC LOGISTIC DATA=p1.working;
 class Diabetes (ref= 'Yes')/param=ref;
   MODEL CHCOCNCR (event='Yes') = Diabetes;
   TITLE 'Logistic Regression';
RUN;

/*task 4*/

 PROC LOGISTIC DATA=p1.working;
 class SMOKDAY2 (ref= 'Every day') Diabetes (ref= 'Yes') agegroup (ref= 'Age 45 - 54') sex (ref= 'Male')/param=ref;
  MODEL CHCOCNCR (event='Yes') = AVE_DRINK SMOKDAY2 Diabetes BMI SEX AGEGROUP /SELECTION=BACKWARD DETAILS INCLUDE=3;
 RUN;

 PROC LOGISTIC DATA=p1.working;
class SMOKDAY2 (ref= 'Every day') Diabetes (ref= 'Yes') agegroup (ref= 'Age 45 - 54') sex (ref= 'Male')/param=ref;
 MODEL CHCOCNCR (event='Yes') = AVE_DRINK SMOKDAY2 Diabetes BMI SEX AGEGROUP /SELECTION=FORWARD DETAILS INCLUDE=3;
RUN;

 PROC LOGISTIC DATA=p1.working;
class SMOKDAY2 (ref= 'Every day') Diabetes (ref= 'Yes') agegroup (ref= 'Age 45 - 54') sex (ref= 'Male')/param=ref;
 MODEL CHCOCNCR (event='Yes') = AVE_DRINK SMOKDAY2 Diabetes BMI SEX AGEGROUP /SELECTION=STEPWISE DETAILS INCLUDE=3;
RUN;

 PROC LOGISTIC DATA=p1.working;
class SMOKDAY2 (ref= 'Every day') Diabetes (ref= 'Yes') agegroup (ref= 'Age 45 - 54') sex (ref= 'Male')/param=ref;
 MODEL CHCOCNCR (event='Yes') = AVE_DRINK SMOKDAY2 Diabetes BMI SEX AGEGROUP /RSQUARE;
RUN;

 PROC LOGISTIC DATA=p1.working;
class SMOKDAY2 (ref= 'Every day') Diabetes (ref= 'Yes') agegroup (ref= 'Age 45 - 54') sex (ref= 'Male')/param=ref;
 MODEL CHCOCNCR (event='Yes') = AVE_DRINK SMOKDAY2 Diabetes BMI SEX AGEGROUP /LACKFIT;
RUN;

 PROC LOGISTIC DATA=p1.working PLOTS (ONLY)= ROC;
class SMOKDAY2 (ref= 'Every day') Diabetes (ref= 'Yes') agegroup (ref= 'Age 45 - 54') sex (ref= 'Male')/param=ref;
 MODEL CHCOCNCR (event='Yes') = AVE_DRINK SMOKDAY2 Diabetes BMI SEX AGEGROUP;
RUN;
```