```
/*read in the raw dental data*/
FILENAME Dental '/home/lingdizhang68/6624/Project0/Data/Project0_dental_data.csv';

PROC IMPORT DATAFILE=Dental
    DBMS=CSV replace
    OUT=WORK.IMPORT;
    GETNAMES=YES;
RUN;

/* generate new variables for the change from baseline to one year*/
DATA dentalclean;
set WORK.IMPORT;
attachchange=attach1year-attachbase;
pdchange=pd1year-pdbase;
RUN;


***Export the clean dataset for use in other analysis programs****;
PROC EXPORT DATA= WORK.dentalclean
            OUTFILE= '/home/lingdizhang68/6624/Project0/dentalclean.csv'
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;

PROC PRINT data=Dentalclean;
run;

/***open ods (output delivery system) to write all my figures       ***/
/***and tables to a directory and file of my choosing for editing   ***/
/***later                                                           ***/

ods listing close;
ods html path='/home/lingdizhang68/6624/Project0/Reports' file = 'DDescriptives.html';

/*import the new clean dataset*/
PROC IMPORT OUT= WORK.dentalclean
            DATAFILE= '/home/lingdizhang68/6624/Project0/dentalclean.csv'
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;


/*check the missing data and mean for numerical variables*/
PROC MEANS data=Dentalclean N NMISS;
    var pdchange attachchange age sites pdbase pd1year attachbase attach1year;
    title 'Missing data description for Continuous Variables in Dental-Project0';
RUN;title;

ods select MissPattern;
PROC MI data=Dentalclean nimpute=0;
var pdchange attachchange age sites pdbase pd1year attachbase attach1year;
run;

/*Descriptive Statistics for the dataset*/
PROC MEANS data=Dentalclean;
    var pdchange attachchange age sites pdbase pd1year attachbase attach1year;
    title 'Descriptive Statistics for Continuous Variables in Dental-Project0';
RUN;title;

PROC FREQ data=Dentalclean;
    tables smoker sex trtgroup race;
    title 'Descriptive Statistics for the Categorical Variables in Dental-Project0';
RUN;title;

PROC FREQ data=Dentalclean;
    tables trtgroup*race;
```

```
    title 'Descriptive Statistics for treatment and race';
RUN;title;

PROC CORR data=Dentalclean;
    var pdchange attachchange age sites pdbase pd1year attachbase attach1year;
    title 'Correlations between all variables: After data cleaning';
RUN;title;

PROC GPLOT data=Dentalclean;
  plot attachchange*(trtgroup race age sites smoker sex);
  symbol I=r1 value=dot color=black;
RUN;

PROC GPLOT data=Dentalclean;
  plot pdchange*(trtgroup race age sites smoker sex);
  symbol I=r1 value=dot color=red;
RUN;

QUIT

ods graphics on;

/*check if the treatment group has important effect on the outcomes*/

PROC GLM data=Dentalclean;
class trtgroup;
model attachchange=trtgroup;
run;
/*0.045*/
/* below t-test still find there is no clinical meaning for the treatment group*/


PROC GLM data=Dentalclean;
class trtgroup;
model attachchange=trtgroup attachbase race sex age smoker sites;
run;
/*0.1063*/

PROC GLM data=Dentalclean;
class trtgroup sex race;
model attachchange=trtgroup attachbase race sex age smoker sites;
run;
/*0.0907*/


quit

/*t-test to check if two treatment groups are significantly diffrent */

ods graphics on;
proc sql;
create table trt12 as
select *
from Dentalclean
where trtgroup=1 or trtgroup=2;
quit;

proc sql;
create table trt13 as
select *
from Dentalclean
where trtgroup=1 or trtgroup=3;
quit;

proc sql;
create table trt14 as
select *
from Dentalclean
```

```
where trtgroup=1 or trtgroup=4;
quit;

proc sql;
create table trt15 as
select *
from Dentalclean
where trtgroup=1 or trtgroup=5;
quit;


proc sql;
create table trt23 as
select *
from Dentalclean
where trtgroup=2 or trtgroup=3;
quit;

proc sql;
create table trt24 as
select *
from Dentalclean
where trtgroup=2 or trtgroup=4;
quit;

proc sql;
create table trt25 as
select *
from Dentalclean
where trtgroup=2 or trtgroup=5;
quit;


proc ttest data=trt12;
class trtgroup;
var attachchange;
run;

/*0.0883*/

proc ttest data=trt13;
class trtgroup;
var attachchange;
run;
/*0.3706*/

proc ttest data=trt14;
class trtgroup;
var attachchange;
run;
/*0.2733*/

proc ttest data=trt15;
class trtgroup;
var attachchange;
run;

/0.3081*/

proc ttest data=trt23;
class trtgroup;
var attachchange;
run;
/*0.0176*/

proc ttest data=trt24;
class trtgroup;
var attachchange;
```

```
run;
/*0.0096*/

proc ttest data=trt25;
class trtgroup;
var attachchange;
run;
/*0.5635*/


proc ttest data=trt12;
class trtgroup;
var pdchange;
run;

/*0.8792*/

proc ttest data=trt23;
class trtgroup;
var  pdchange;
run;
/*0.0944*/

proc ttest data=trt24;
class trtgroup;
var  pdchange;
run;
/*0.0852*/

proc ttest data=trt25;
class trtgroup;
var  pdchange;
run;
/*0.5717*/

PROC GLM data=Dentalclean;
class trtgroup;
model pdchange=trtgroup;
run;
/*0.089*/

PROC GLM data=Dentalclean;
class trtgroup;
model pdchange=trtgroup attachbase sex age race smoker sites;
run;
/0.1208*/

PROC GLM data=Dentalclean;
class trtgroup sex race;
model pdchange=trtgroup attachbase race sex age smoker sites;
run;
/*0.1276*/


/*model selection for the regression linear model*/
proc glmselect data=Dentalclean;
    class trtgroup sex race smoker;
    model attachchange=attachbase trtgroup age sex race smoker sites/selection=stepwise(select=SL SLS=0.05);
run;

/*attachbase is significant p<0.001*/

proc glmselect data=Dentalclean;
    class trtgroup sex race smoker;
    model pdchange=pdbase trtgroup age sex race smoker sites/selection=stepwise(select=SL SLS=0.05);
run;

/*sex is significant p=0.0330*/
```

```
/*please forget the interaction term and imputation*/

/*linear regression for model selection*/
ods graphics on;
proc glmselect data=Dentalclean;
    class trtgroup sex race smoker;
    model attachchange=attachbase|trtgroup|age|sex|race|smoker|sites/selection=stepwise(select=SL SLS=0.05);
run;

proc glmselect data=Dentalclean;
    class trtgroup sex race smoker;
    model attachchange=attachbase trtgroup age sex race smoker sites/selection=stepwise(select=SL SLS=0.05);
run;

/*attachbase is significant p<0.001*/

proc glmselect data=Dentalclean;
    class trtgroup sex race smoker;
    model pdchange=pdbase trtgroup age sex race smoker sites/selection=stepwise(select=SL SLS=0.05);
run;

/*sex is significant p=0.0330*/

proc glm data=Dentalclean PLOTS=(DIAGNOSTICS RESIDUALS);
    class trtgroup sex race smoker;
    model attachchange=attachbase trtgroup smoker attachbase*smoker/solution;
    output out=d_attach COOKD=COOKD STUDENT=STUDENT;
    /*R=0.3492*/

proc glm data=Dentalclean PLOTS=(DIAGNOSTICS RESIDUALS);
    class trtgroup sex race smoker;
    model attachchange=attachbase trtgroup/solution;
    output out=d_attach COOKD=COOKD STUDENT=STUDENT;
    /*R=0.237*/


    quit;

ods html close;
ods listing;



/*impute the missing data*/

PROC MI DATA = Dentalclean;
EM OUT = Dentalimpute;
VAR attach1year pd1year age;
RUN;

PROC PRINT  data=Dentalimpute;
run;

/*clean data by generating the change scores*/
DATA dental1;
set Dentalimpute;
dattach=attach1year-attachbase;
```

```
dchange=pd1year-pdbase;
RUN;

PROC PRINT  data=dental1;
run;

/*select the interaction and run linear regression model*/
proc glmselect data=dental1;
    class trtgroup sex race smoker;
    model dattach=attachbase|trtgroup|age|sex|race|smoker|sites/selection=stepwise(select=SL SLS=0.05);
run;

proc glmselect data=dental1;
    class trtgroup sex race smoker;
    model dchange=pdbase|trtgroup|age|sex|race|smoker|sites/selection=stepwise(select=SL SLS=0.05);
run;



proc glm data=dental1 PLOTS=(DIAGNOSTICS RESIDUALS);
    class trtgroup sex race smoker;
    model dattach=attachbase trtgroup sex smoker attachbase*smoker/solution;
    output out=d_attach COOKD=COOKD STUDENT=STUDENT;
    quit;

proc glm data=dental1 PLOTS=(DIAGNOSTICS RESIDUALS);
    class trtgroup sex race smoker;
    model dchange=pdbase trtgroup pdbase*sites*sex/solution;
    output out=attach1year COOKD=COOKD STUDENT=STUDENT;
```