

```

/*****
NAME: C:\Users\legof\Desktop\M403\SAS Final\Task 1\Task_1_Program.sas
DATE: 12.2.2019
CREATED BY: Stephen Li
PURPOSE: Code for the SAS Final Task 1
*****/

Title1 "Stephen Li (Final)";

libname task1 "C:\Users\legof\Desktop\M403\SAS Final\Task 1";

/* Proc Import Code */
/* Survey Sheet 1 */
PROC IMPORT OUT= TASK1.survey1
            DATAFILE= "C:\Users\legof\Desktop\M403\SAS Final\STD scores.
xls"
            DBMS=EXCEL REPLACE;
            RANGE="STDquiz1$";
            GETNAMES=YES;
            MIXED=NO;
            SCANTEXT=YES;
            USEDATE=YES;
            SCANTIME=YES;
RUN;

/* Survey Sheet 2 */
PROC IMPORT OUT= TASK1.SURVEY2
            DATAFILE= "C:\Users\legof\Desktop\M403\SAS Final\STD scores.
xls"
            DBMS=EXCEL REPLACE;
            RANGE="STDquiz2$";
            GETNAMES=YES;
            MIXED=NO;
            SCANTEXT=YES;
            USEDATE=YES;
            SCANTIME=YES;
RUN;

/* Survey Sheet 3 */
PROC IMPORT OUT= TASK1.SURVEY3
            DATAFILE= "C:\Users\legof\Desktop\M403\SAS Final\STD scores.
xls"
            DBMS=EXCEL REPLACE;
            RANGE="STDquiz3$";
            GETNAMES=YES;
            MIXED=NO;
            SCANTEXT=YES;
            USEDATE=YES;
            SCANTIME=YES;
RUN;

/* Options */
Option pageno=1;
Option formdlm=' - ';
Option fmtsearch=(task1);

```

```

/** Proc Format **/
Proc Format library=task1;
    value answer      0 = 'No'
                    1 = 'Yes';

    /* Use for step 6 */
    value newanswer 1 = 'Yes'
                    2 = 'No';

    value quizft      low - 79    = 'Fail'
                    80 - high    = 'Pass';

Run;

/* survey1 labels, sorting, remove duplicate values, proc contents */
Data task1.survey1permanent;
    set task1.survey1;
    label    id          = 'ID Number'
            city         = 'City of Residence'
            age          = 'Age Group'
            gender       = 'Gender'
            zipcode      = 'Zipcode'
            reuse        = 'Reuses Needles'
            crack        = 'Smokes crack during sex'
            sex4drugs    = 'Commercial sex workers'
            anonymous    = 'Had at least one anonymous sex partner in the past year'
            marijuana    = 'Has used marijuana during sex in the past year'
            knows        = 'Knows when someone has STD/HIV'
            syp          = 'Had syphilis in the past year'
            gc           = 'Had gonorrhea in the past year'
            chl          = 'Had chlamydia in the past year'
            hiv          = 'Has tested positive for HIV'
            quizscore    = 'Score on STD prevention quiz'
            survdate     = 'Date student was surveyed'
    ;
Run;

Proc Sort data=task1.survey1permanent dupout=survey1dup nodupkey;
    By id;
Run;

Proc Print data=survey1dup;
    Title2 "Survey1 duplicates";
Run;

Proc Contents data=task1.survey1permanent varnum;
    Title2 "Proc Contents of STDquiz1";
Run;

/* survey2 labels, sorting, remove duplicate values, proc contents */
Data task1.survey2permanent;
    set task1.survey2;
    label    id          = 'ID Number'
            city         = 'City of Residence'
            age          = 'Age Group'
            gender       = 'Gender'
            zipcode      = 'Zipcode'

```

```

reuse      = 'Reuses Needles'
crack      = 'Smokes crack during sex'
sex4drugs  = 'Commercial sex workers'
anonymous  = 'Had at least one anonymous sex partner in the past year'
marijuana  = 'Has used marijuana during sex in the past year'
knows      = 'Knows when someone has STD/HIV'
syp        = 'Had syphilis in the past year'
gc         = 'Had gonorrhea in the past year'
chl        = 'Had chlamydia in the past year'
hiv        = 'Has tested positive for HIV'
quizscore  = 'Score on STD prevention quiz'
survdate   = 'Date student was surveyed'
;
Run;

Proc Sort data=task1.survey2permanent dupout=survey2dup nodupkey;
  By id;
Run;

Proc Print data=survey2dup;
  Title2 "Survey2 Duplicates";
Run;

Proc Contents data=task1.survey2permanent varnum;
  Title2 "Proc Contents of STDquiz2";
Run;

/* survey3 labels, sorting, remove duplicate values, proc contents */
Data task1.survey3permanent;
  set task1.survey3;
  label    id          = 'ID Number'
         city          = 'City of Residence'
         age           = 'Age Group'
         gender        = 'Gender'
         zipcode       = 'Zipcode'
         reuse         = 'Reuses Needles'
         crack         = 'Smokes crack during sex'
         sex4drugs     = 'Commercial sex workers'
         anonymous     = 'Had at least one anonymous sex partner in the past year'
         marijuana     = 'Has used marijuana during sex in the past year'
         knows        = 'Knows when someone has STD/HIV'
         syp          = 'Had syphilis in the past year'
         gc           = 'Had gonorrhea in the past year'
         chl          = 'Had chlamydia in the past year'
         hiv          = 'Has tested positive for HIV'
         quizscore    = 'Score on STD prevention quiz'
         survdate     = 'Date student was surveyed'
;
Run;

Proc Sort data=task1.survey3permanent dupout=survey3dup nodupkey;
  By id;
Run;

```

```

Proc Print data=survey3dup;
  Title2 "Survey3 Duplicates";
Run;

Proc Contents data=task1.survey3permanent varnum;
  Title2 "Proc Contents of STDquiz3";
Run;

/**** Concatenate Datasets ****/

Data task1.surveyCombined;
  set task1.survey1permanent task1.survey2permanent task1.survey3permanent;

  /* Create a new date variable */
  new_date = intck('days',survdate,'13DEC2019'D);

  /* convert city to lower case */
  city = lowercase(city);

  /* Use if then statements for city */
  if city      = ' ' then city = 'Missing';
  else if city = 'outside los angeles county' then city = 'Outside Los Angeles County';
  else if city = 'out los angeles county' then city = 'Outside Los Angeles County';
  else if city = 'otside los angeles county' then city = 'Outside Los Angeles County';
  else if city = 'otside la county' then city = 'Outside Los Angeles County';
  else if city = 'otherside la county' then city = 'Outside Los Angeles County';
  else if city ^= 'Outside Los Angeles County'
and city ^= 'Missing' then city = 'Los Angeles County';

  /* Get rid of extra characters in age */
  age = compress(age, 'xz');
  if age = '40-49r' then age = '40-49';
  if age = '60-69r' then age = '60-69';

  /* Gender use if then statements */
  /* Note that Women is counted as Female */
  gender = substr(gender, 1, 1);
  if gender = 'M' then gender = 'Male';
  if gender = 'F' then gender = 'Female';
  if gender = 'W' then gender = 'Female';
  if gender = 'K' then gender = '';

  /* Zipcode - first 5 characters */
  zipcode = substr(zipcode,1,5);

  /* Get rid of extra characters in quizscore */
  quizscore = compress(quizscore, 'abcdefghijklmnopqrstuvwxyz#*$-');

  /* Get rid of extra characters in reuse--hiv (no numbers for now)*/
  reuse      = compress(reuse, 'mp');
  crack      = compress(crack, '*p463');
  sex4drugs  = compress(sex4drugs, 'pq');
  anonymous   = compress(anonymous, 'lop4');
  marijuana  = compress(marijuana, 'sr4');

```

```

knows          = compress(knows, 'p');
syp            = compress(syp, 'p5');
gc            = compress(gc, 'p5');
chl           = compress(chl, 'p');
hiv           = compress(hiv, 'p');

temp1 = input(id, 4.);
temp2 = input(zipcode, 5.);
temp3 = input(reuse, 2.);
temp4 = input(crack, 2.);
temp5 = input(sex4drugs, 2.);
temp6 = input(anonymous, 2.);
temp7 = input(marijuana, 2.);
temp8 = input(knows, 2.);
temp9 = input(syp, 2.);
temp10 = input(gc, 2.);
temp11 = input(chl, 2.);
temp12 = input(hiv, 2.);
temp13 = input(quizscore, 3.);

/* Use arrays to recode 7, 8, 9 to missing*/

Array temp[13] temp1-temp13;
Do I = 1 to 13;
    If temp[I] = 7 then temp[I] = .;
    If temp[I] = 8 then temp[I] = .;
    If temp[I] = 9 then temp[I] = .;
End;

/* Recode 0=no to 2=no */
If temp3 = 0 then temp3 = 2;
If temp12 = 0 then temp12 = 2;

/* Rename variables */

id1          = temp1;
zipcode1     = temp2;
reuse1       = temp3;
crack1       = temp4;
sex4drugs1   = temp5;
anonymous1   = temp6;
marijuana1   = temp7;
knows1       = temp8;
syp1         = temp9;
gc1          = temp10;
chl1         = temp11;
hiv1         = temp12;
quizscore1   = temp13;

/* Label the new variables */
Label  new_date    = 'Days between survey date and Dec. 13, 2019'
      id1          = 'ID Number'
      zipcode1     = 'Zipcode'
      reuse1       = 'Reuses needles'
      crack1       = 'Smokes crack during sex'
      sex4drugs1   = 'Commercial sex worker'

```

```

        anonymous1 = 'Had at least one anonymous sex partner in the past year'
        marijuana1 = 'Has used marijuana during sex in the past year'
        knows1     = 'Knows when someone has STD/HIV'
        syp1       = 'Had syphilis in the past year'
        gc1        = 'Had gonorrhea in the past year'
        chl1       = 'Had chlamydia in the past year'
        hiv1       = 'Has tested positive for HIV'
        quizscore1 = 'Score on STD prevention quiz'
;

/* Format and drop variables */
Format crack1 -- chl1 answer. reuse1 hiv1 newanswer. survdate WEEKDATE. quizscore1 quizft.;
Drop I;
Drop temp1 -- temp13;
Drop id zipcode--quizscore;
Run;

Proc Contents data=task1.surveyCombined varnum;
    Title2 "Proc Contents of task1.surveyCombined";
Run;

/**/ Proc Tabulate Step for City /**/
Proc Tabulate data=task1.surveyCombined Format=12.0;
    Title2 "Proc Tabulate for City of Residence";
    Class city;
    Table city;
Run;

/**/ Proc Tabulate Step for Age /**/
Proc Tabulate data=task1.surveyCombined Format=12.0;
    Title2 "Proc Tabulate for Age Groups";
    Class age;
    Table age;
Run;

/**/ Proc Tabulate Step for Anonymous /**/
Proc Tabulate data=task1.surveyCombined Format=12.0;
    Title2 "Proc Tabulate for anonymous sex partners";
    Class anonymous1;
    Table anonymous1;
Run;

/**/ Proc Tabulate Step for Quizscore /**/
Proc Tabulate data=task1.surveyCombined Format=12.0;
    Title2 "Proc Tabulate for Quiz Score";
    Class quizscore1;
    Table quizscore1;
Run;

/**/ Logistic Regression Models using Parameterization /**/

/* Crude Analysis */
Proc Logistic data=task1.surveyCombined;
    Title2 'Crude logistic regression of gender
on quizscore (OR=1.282, 95% CL 1.064, 1.544)';

```

```

class gender (PARAM=REF REF='Male') quizscore1 (PARAM=REF REF='Pass');
model quizscore1=gender;
Run;

/*

Odds Ratio Estimates

Effect                                Point          95% Wald
Estimate          Confidence Limits

gender Female vs Male                1.282          1.064          1.544

Interpretation:
The odds of women failing the STD prevention quiz was 1.282 times that of men
(95% CL 1.064, 1.544)
*/

/* Adjusted Predictive Analysis*/
Proc Logistic data=task1.surveyCombined;
  Title2 'Effect of gender (AOR=1.272, 95% CL 1.056, 1.533) and
reuse AOR=1.202, 95% CL 0.976, 1.481) on quizscore';

  class gender (PARAM=REF REF='Male') reuse1 (PARAM=REF REF='Yes')
quizscore1 (PARAM=REF REF='Pass');

  model quizscore1=gender reuse1/lackfit;
Run;

/*

Odds Ratio Estimates

Effect                                Point          95% Wald
Estimate          Confidence Limits

gender Female vs Male                1.272          1.056          1.533
reuse1 No vs Yes                    1.202          0.976          1.481

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square      DF      Pr > ChiSq

2.8886           2           0.2359

Interpretation:
Females were less likely to pass the STD prevention quiz than men (AOR=1.272, 95% CL 1.056, 1.533).
Whereas, individuals who said they did not reuse needles were less likely to pass the
STD prevention quiz than those who did (AOR=1.202, 95% CL 0.976, 1.481).
*/

/* ODS Step */
ods pdf file= "C:\Users\legof\Desktop\M403\SAS Final\Task 1\Task1Results.pdf";

Proc Contents data=task1.surveyCombined varnum;
  Title2 "Proc Contents of task1.surveyCombined";
Run;

```

```

Proc Freq data=task1.surveyCombined;
  Title2 "Proc Freq of task1.surveyCombined";
  /* Everything except id and zipcode */
  Tables city--gender survdate new_date reuse1--quizscore1;
Run;

ods pdf close;

/*
TODO:
  When Printing - make sure titles fit within page
*/

/**** Proc Export Step ****/

PROC EXPORT DATA= TASK1.SURVEYCOMBINED
  OUTFILE= "C:\Users\legof\Desktop\M403\SAS Final\Task 1\surveycombined.xlsx"
  DBMS=EXCEL REPLACE;
  SHEET="surveycombined";
RUN;

```