```
/*****************************************************************************/

/* Program Name: STAT 604 HW#15
                                */
/* Date Created: 12/03/2021
                                */
/* Author: Jack Rodoni
                                */
/* Purpose: STAT 604 HW#15
                                */
/* Date Modified: 12/07/2021
                         */
/* Location: /home/u59649056/Homeworks/JRodoni_Homework15.sas
*/
/*****************************************************************************/


/* This assignment will use three separate sources of data as input.  One source will be the Master  */

/* Location Pop Table.txt file that was used in an R assignment earlier in the semester.  Another will be the  */

/* permanent data set of all Monthly Stats which was created in step 7 of Homework 13.  If you had  */

/* difficulty creating this data set, the professor's version (monthly_stats) is included on the weekly  */

/* module for your use.  This assignment will also use the county_jobs data set that is provided on the  */

/* Weekly module in Canvas.  This data set is a modified version of the professor's final data set from  */

/* Homework 14.  Since there was some confusion regarding column names in the instructions for that  */

/* assignment, use the data set provided for consistency.  The FIPS column has also been converted for  */

/* convenience in merging with the other data.  Familiarize yourself with this file and its contents after  */

/* downloading it to a folder on your computer that is accessible to SAS.  Programming efficiency should be  */

/* incorporated throughout the program.  Unneeded information should be eliminated as early as possible.  */
```

```
/* Make sure the lines in your program do not get too long to fit on the PDF output page when you
convert  */

/* your program for submission.  Please read the entire assignment instructions before beginning.  */


/* 1.) Add a header comment section to the beginning of a new program in your SAS session.  Be sure  */

/*          to include a comment line above each section of the program that identifies the associated  */

/*          assignment step and a brief description of what the section is doing.  Include housekeeping  */

/*          statements to clear titles and footnotes and suppress the printing of procedure titles.  Use a
*/

/*          system option to prevent an error message when SAS cannot locate a permanent format and
*/

/*          another option to allow SAS to locate your permanent formats.  */


title;

footnote;

ods noproctitle;

options nofmterr fmtsearch=(mylib);


/* 2.) Assign librefs to the downloaded data folder (set to readonly) and the mylib folder containing  */

/*          your permanent data sets.  Create a fileref to the pdf file for output.  Create a fileref to the
text  */

/*          file.  */



libname mylib "/home/u59649056/Homeworks/mylib";

filename HW15pdf "/home/u59649056/Homeworks/mylib/JRodoni_HW15_Output.pdf";

libname  HWDATA "/home/u59649056/Homeworks/Homework Data";

filename locpop "/home/u59649056/Homeworks/Homework Data/Master Location Pop Table.txt";
```

/* 3.) Open the PDF destination to receive your output. */

ods pdf file=HW15pdf;

/* 4.) Write a single proc step that converts the text file to a temporary data set. Ensure the program */

/*   will overwrite the data set if it already exists. You will need to have SAS evaluate all rows of the */

/*   text file to determine the attributes of the data. */

```
proc import datafile=locpop

dbms=dlm

out=locpop_temp

replace;

delimiter = ':';

guessingrows=max;

run;
```

/* 5.) Print the descriptor portion of the new data set. Supply an appropriate title. */

```
proc contents data=locpop_temp;

title1 "Descriptor of Location Pop Table";

run;
```

/* 6.) Write a single SAS step that will use the imported data as input and create in mylib a new */

/*   permanent data set of Texas county populations that is suitable for combining with the Covid */

/*   data using the FIPS number as the common value: The output data set will contain three */

/*   columns: the county name, the FIPS number, and a column of population values renamed */

/*   Population. The rows will be only Texas counties based on a FIPS number that begins with 48. */

```
data mylib.TxPop;

length COUNTY_NAME $ 17;

set locpop_temp(rename=(geo_region_population_count=Population));

where province_state_name = "Texas";

keep county_name county_fips_number Population;

run;
```

/* 7.) Use the Monthly Stats data set as the source for a FREQ procedure that will show the number in */

/*   each fatality category by month. The month name will be the rows in the output. Show only */

/*   the Frequency and Row Percent statistics. Since some students may use their own data set and */

/*   others use the professor's for this step, apply your permanent format to the fatality rate column */

/*   in this PROC step so everyone's code will be consistent. If you were unable to create the */

/*   permanent format correctly, you will need to view the Solution Review video for Homework 13 */

/*   and run the code shown in the solution to get the format created on your system */


```
proc freq data=HWDATA.MONTHLY_STATS;

tables report_month*fatality_rate / nocum nocol nopercent;

format fatality_rate pct.;

run;
```

/* 8.) In a single step create a temporary copy of the monthly stats data set that is sorted by FIPS, Year */

/*   and Month number. Subset the data so that it only contains rows from 2020 months 3 and 6 */

/*   and from 2021 months 1, 2, and 3. */

```
proc sort data=HWDATA.Monthly_stats

 out=monthly_stats_temp;

  where (report_year=2020 and (report_month=3 or report_month=6)) or (report_year=2021 and
report_month between 1 and 3);

by county_fips_number report_year report_month;

run;
```

/* 9.) Without using a DATA step, create a new temporary data set that is a "wide" version of the */

/*   monthly stats data created in the previous step. There will be one row per county FIPS. The */
*/

/*   value for the "Cases" columns will be the number of monthly cases for that specific year and */

/*   month time period as identified in the column name. The first two rows from that data set is */

/*   shown below as a sample. */

```
proc transpose data=monthly_stats_temp

out = monthly_stats_transpose(drop=_:)

prefix=Cases;

var monthly_cases;

by county_fips_number;

id report_year report_month;

run;
```

/* 10.) Use a single DATA step to combine the columns from the permanent data set of Texas county */

/*   populations, the downloaded county_jobs data set, and the wide data set created in the */

/*   previous step. Keep only those rows for which there is a match in county_jobs. Create the */

/*   following new variables for analysis. Begin the variable names with Pct so they can be
accessed */

/*   with a variable list: */

```
data mergedHW15;

        merge HWDATA.county_jobs(IN=CountyJobs)

                mylib.txpop

                work.monthly_stats_transpose;

        by county_fips_number;

        if CountyJobs = 1;


/* (a) Percentage of population employed before the pandemic by dividing jobs20m1 by the  */

/* county population.  */



        Pct_pop_prior = jobs20m1/Population;



/* (b) Percentage of population employed early in the pandemic by dividing jobs20m4 by the  */

/* county population.  */



        Pct_pop_early = jobs20m4/Population;



/* (c) Percentage of population employed one year into the pandemic by dividing jobs21m3  */

/* by the county population.  */



        Pct_pop_1year = jobs21m3/Population;



/* (d) Percentage of change in monthly cases from the end of 2020 Q2 to the beginning of  */

/* 2021 Q1.  Subtract cases20206 from cases20211 and divide the result by cases20206.   */

/* Use conditional logic on this statement to prevent a divide by 0 message from occurring  */

/* in the log.  */



        if cases20206 ^= 0 then Pct_Change_cases = (cases20211-cases20206)/cases20206;
```

```
/* (e) Percentage of change in monthly employment from the end of 2020 Q2 to the beginning  */
/* of 2021 Q1.  Subtract jobs20m6 from jobs21m1 and divide the result by jobs20m6.  Use  */
/* conditional logic on this statement to prevent a divide by 0 message from occurring in  */
/* the log.  */


        if jobs20m6 ^= 0 then Pct_change_emp = (jobs21m1-jobs20m6)/jobs20m6;


/* f. Format these 5 new variables as a percent with 1 decimal place.  */


        format Pct: percent10.1;


run;




/* 11)  Use a single PROC step to show the extreme observations of employment percentage before the
*/
/* pandemic and one year into the pandemic using two of the variables created above.  Show the  */
/* county name and population value in the tables of extreme observations.  Supply an appropriate  */
/* title.  */


title "Unappropriate Title";
proc univariate data=work.mergedhw15;
        var Pct_pop_early Pct_pop_1year;
        id County_Name Population;
run;


/* 12)  Use the data set from step 10 with the means procedure to create an analysis of all the "Cases"
*/
/* variables by using a variable list.  Show the default statistics to one decimal place.  At the same  */
```

```
/* time create a temporary data set that contains only the mean and range statistics.  For training  */
/* purposes, only supply one variable for each of the statistics.  Supply an appropriate title. */


title "Insert *Appropriate Title Here";
proc means data=work.mergedhw15 maxdec=1 mean min max range;
        var case: ;
        output out=means_rangeHw15

                MEAN=Mean1 /*Mean2 Mean3 Mean4 Mean5 */

                RANGE=Range1 /*Range2 Range3 Range4 Range5;*/
run;


proc print data=work.means_rangeHW15;
run;




/* 13.)  Use the TABULATE procedure and the sorted data set created in step 8 to show the mean and  */
/* range of monthly cases by year and month to one decimal place.  Include the mean and range  */
/* for all values at the bottom of the report.  Supply an appropriate title.  The layout of the report  */
/* is shown below:  */


title "I am the Smartest Man Alive!";
proc tabulate data = work.monthly_stats_temp;
class report_year report_month;
var monthly_cases;
table report_year*report_month all, monthly_cases*(mean range) *f=7.1;
run;


/* 14.  Close the PDF destination.  */
```

ods pdf close;

/* 15. Examine the data sets and use report information contained in your PDF output document to */

/* find the answers to the questions below and include the answers in a comment section at the */

/* bottom of your program file: */

/* (a) Which month had the highest row percent value for the Extreme fatality rate? */

/*   March
                                                                                       */

/* (b) Which county had the lowest percentage employment before the pandemic and 1 year */

/* into the pandemic? */

/*   Before pandemic: San Jacinto
                                                                   */
/*   1 Year into pandemic: San Jacinto
                                                                   */

/* (c) Which county had the highest percentage employment before the pandemic and 1 year */

/* into the pandemic? */

/*   Before pandemic: Kenedy
                                                                 */
/*   1 Year into pandemic: Loving
                                                             */

/* (d) Which of the "Cases" variables had the highest mean value? What was the value */

/*   Cases20211-2371.8
                                                             */

/* (e) Compare the data in the data set created by the MEANS procedure with the output data */

/* it created in the PDF. What data is represented in the data set? */

/*   The data is the mean and range of cases20203
                              */

/* (f) What is the overall mean and range of the Monthly Cases? */

/*   882.20
                    */