

```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable some
output features.
69
70
/*****
*****/
71      /* Program Name: STAT 604 HW#15    */
72      /* Date Created: 12/03/2021    */
73      /* Author: Jack Rodoni    */
74      /* Purpose: STAT 604 HW#15    */
75      /* Date Modified: 12/07/2021    */
76      /* Location:
/home/u59649056/Homeworks/JRodoni_Homework15.sas    */
77
/*****
*****/
78
79      /* This assignment will use three separate sources of data
as input. One source will be the Master */
80      /* Location Pop Table.txt file that was used in an R
assignment earlier in the semester. Another will be the */
81      /* permanent data set of all Monthly Stats which was
created in step 7 of Homework 13. If you had */
82      /* difficulty creating this data set, the professor's
version (monthly_stats) is included on the weekly */
83      /* module for your use. This assignment will also use the
county_jobs data set that is provided on the */
84      /* Weekly module in Canvas. This data set is a modified
version of the professor's final data set from */
85      /* Homework 14. Since there was some confusion regarding
column names in the instructions for that */
86      /* assignment, use the data set provided for consistency.
The FIPS column has also been converted for */
87      /* convenience in merging with the other data.
Familiarize yourself with this file and its contents after */
88      /* downloading it to a folder on your computer that is
accessible to SAS. Programming efficiency should be */
89      /* incorporated throughout the program. Unneeded
information should be eliminated as early as possible. */
90      /* Make sure the lines in your program do not get too long
to fit on the PDF output page when you convert */
91      /* your program for submission. Please read the entire
assignment instructions before beginning. */

```

```

92
93      /* 1.) Add a header comment section to the beginning of a
new program in your SAS session. Be sure */
94      /*      to include a comment line above each section of the
program that identifies the associated */
95      /*      assignment step and a brief description of what the
section is doing. Include housekeeping */
96      /*      statements to clear titles and footnotes and
suppress the printing of procedure titles. Use a */
97      /*      system option to prevent an error message when SAS
cannot locate a permanent format and */
98      /*      another option to allow SAS to locate your permanent
formats. */
99
100     title;
101     footnote;
102     ods noproctitle;
103     options nofmterr fmtsearch=(mylib);
104
105     /* 2.) Assign librefs to the downloaded data folder (set
to readonly) and the mylib folder containing */
106     /*      your permanent data sets. Create a fileref to the
pdf file for output. Create a fileref to the text */
107     /*      file. */
108
109
110     libname mylib "/home/u59649056/Homeworks/mylib";
NOTE: Libref MYLIB refers to the same physical library as _TEMP1.
NOTE: Libref MYLIB was successfully assigned as follows:
      Engine:          V9
      Physical Name: /home/u59649056/Homeworks/mylib
111     filename HW15pdf
"/home/u59649056/Homeworks/mylib/JRodoni_HW15_Output.pdf";
112     libname HWDATA "/home/u59649056/Homeworks/Homework Data";
NOTE: Libref HWDATA refers to the same physical library as _TEMP2.
NOTE: Libref HWDATA was successfully assigned as follows:
      Engine:          V9
      Physical Name: /home/u59649056/Homeworks/Homework Data
113     filename locpop "/home/u59649056/Homeworks/Homework
Data/Master Location Pop Table.txt";
114
115
116     /* 3.) Open the PDF destination to receive your output.
*/
117
118     ods pdf file=HW15pdf;

```

NOTE: Writing ODS PDF output to DISK destination "HW15PDF", printer "PDF".

```
119
120      /* 4.) Write a single proc step that converts the text
file to a temporary data set. Ensure the program */
121      /* will overwrite the data set if it already exists.
You will need to have SAS evaluate all rows of the */
122      /* text file to determine the attributes of the data.
*/
123
124      proc import datafile=locpop
125      dbms=dlm
126      out=locpop_temp
127      replace;
128      delimiter = ':';
129      guessingrows=max;
130      run;
```

NOTE: Unable to open parameter catalog: SASUSER.PARMS.PARMS.SLIST in update mode. Temporary parameter values will be saved to WORK.PARMS.PARMS.SLIST.

```
131
/*****
*
132      *   PRODUCT:    SAS
133      *   VERSION:    9.4
134      *   CREATOR:    External File Interface
135      *   DATE:       07DEC21
136      *   DESC:       Generated SAS Datastep Code
137      *   TEMPLATE SOURCE: (None Specified.)
138
*****/
*/
139      data WORK.LOCPOP_TEMP      ;
140      %let _EFIERR_ = 0; /* set the ERROR detection macro
variable */
141      infile LOCPop delimiter = ':' MISSOVER DSD  firstobs=2
;
142      informat COUNTRY_SHORT_NAME $74. ;
143      informat COUNTRY_ALPHA_3_CODE $3. ;
144      informat COUNTRY_ALPHA_2_CODE $2. ;
145      informat PROVINCE_STATE_NAME $28. ;
146      informat COUNTY_NAME $33. ;
147      informat COUNTY_FIPS_NUMBER best32. ;
148      informat GEO_LATITUDE best32. ;
149      informat GEO_LONGITUDE best32. ;
```

```

150          informat GEO_REGION_POPULATION_COUNT best32. ;
151          informat DATA_SOURCE_NAME $29. ;
152          format COUNTRY_SHORT_NAME $74. ;
153          format COUNTRY_ALPHA_3_CODE $3. ;
154          format COUNTRY_ALPHA_2_CODE $2. ;
155          format PROVINCE_STATE_NAME $28. ;
156          format COUNTY_NAME $33. ;
157          format COUNTY_FIPS_NUMBER best12. ;
158          format GEO_LATITUDE best12. ;
159          format GEO_LONGITUDE best12. ;
160          format GEO_REGION_POPULATION_COUNT best12. ;
161          format DATA_SOURCE_NAME $29. ;
162      input
163          COUNTRY_SHORT_NAME $
164          COUNTRY_ALPHA_3_CODE $
165          COUNTRY_ALPHA_2_CODE $
166          PROVINCE_STATE_NAME $
167          COUNTY_NAME $
168          COUNTY_FIPS_NUMBER
169          GEO_LATITUDE
170          GEO_LONGITUDE
171          GEO_REGION_POPULATION_COUNT
172          DATA_SOURCE_NAME $
173      ;
174      if _ERROR_ then call symputx('_EFIERR_',1); /* set
ERROR detection macro variable */
175      run;

```

NOTE: The infile LOCPop is:  
 Filename=/home/u59649056/Homeworks/Homework Data/Master  
 Location Pop Table.txt,  
 Owner Name=u59649056,Group Name=oda,  
 Access Permission=-rw-r--r--,  
 Last Modified=04Dec2021:12:41:30,  
 File Size (bytes)=335452

NOTE: 3483 records were read from the infile LOCPop.  
 The minimum record length was 61.  
 The maximum record length was 121.

NOTE: The data set WORK.LOCPOP\_TEMP has 3483 observations and 10  
 variables.

NOTE: DATA statement used (Total process time):  
 real time 0.00 seconds  
 user cpu time 0.01 seconds  
 system cpu time 0.00 seconds  
 memory 10183.75k

OS Memory	40996.00k
Timestamp	12/07/2021 04:53:54 PM
Step Count	53 Switch Count 2
Page Faults	0
Page Reclaims	155
Page Swaps	0
Voluntary Context Switches	13
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	1544

3483 rows created in WORK.LOCPOP\_TEMP from LOCPop.

NOTE: WORK.LOCPOP\_TEMP data set was successfully created.

NOTE: The data set WORK.LOCPOP\_TEMP has 3483 observations and 10 variables.

NOTE: PROCEDURE IMPORT used (Total process time):

real time	1.20 seconds
user cpu time	1.19 seconds
system cpu time	0.01 seconds
memory	10183.75k
OS Memory	41508.00k
Timestamp	12/07/2021 04:53:54 PM
Step Count	53 Switch Count 10
Page Faults	0
Page Reclaims	5318
Page Swaps	0
Voluntary Context Switches	79
Involuntary Context Switches	1
Block Input Operations	0
Block Output Operations	1600

176

177 /\* 5.) Print the descriptor portion of the new data set.

Supply an appropriate title. \*/

178 proc contents data=locpop\_temp;

179 title1 "Descriptor of Location Pop Table";

180 run;

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.05 seconds
user cpu time	0.05 seconds

```

system cpu time      0.00 seconds
memory              3412.65k
OS Memory           36528.00k
Timestamp            12/07/2021 04:53:54 PM
Step Count           54   Switch Count   1
Page Faults          0
Page Reclaims        266
Page Swaps           0
Voluntary Context Switches 6
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 24

```

```

181
182      /* 6.) Write a single SAS step that will use the imported
data as input and create in mylib a new */
183      /*      permanent data set of Texas county populations that
is suitable for combining with the Covid */
184      /*      data using the FIPS number as the common value: The
output data set will contain three */
185      /*      columns: the county name, the FIPS number, and a
column of population values renamed */
186      /*      Population. The rows will be only Texas counties
based on a FIPS number that begins with 48. */
187
188      data mylib.TxPop;
189      length COUNTY_NAME $ 17;
190      set
locpop_temp(rename=(geo_region_population_count=Population));
191      where province_state_name = "Texas";
192      keep county_name county_fips_number Population;
193      run;

```

WARNING: Multiple lengths were specified for the variable COUNTY\_NAME by input data set(s). This can cause truncation of data.

NOTE: There were 254 observations read from the data set WORK.LOCPOP\_TEMP.

```
WHERE province_state_name='Texas';
```

NOTE: The data set MYLIB.TXPOP has 254 observations and 3 variables.

NOTE: DATA statement used (Total process time):

```

real time          0.02 seconds
user cpu time       0.00 seconds
system cpu time     0.00 seconds
memory             1624.03k
OS Memory           36528.00k

```

Timestamp	12/07/2021 04:53:54 PM
Step Count	55
Page Faults	0
Page Reclaims	253
Page Swaps	0
Voluntary Context Switches	66
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```

194
195
196      /* 7.) Use the Monthly Stats data set as the source for a
FREQ procedure that will show the number in */
197      /*      each fatality category by month. The month name
will be the rows in the output. Show only */
198      /*      the Frequency and Row Percent statistics. Since
some students may use their own data set and */
199      /*      others use the professor's for this step, apply your
permanent format to the fatality rate column */
200      /*      in this PROC step so everyone's code will be
consistent. If you were unable to create the */
201      /*      permanent format correctly, you will need to view
the Solution Review video for Homework 13 */
202      /*      and run the code shown in the solution to get the
format created on your system */
203
204      proc freq data=HWDATA.MONTHLY_STATS;
NOTE: Data file HWDATA.MONTHLY_STATS.DATA is in a format that is
native to another host, or the file encoding does not match the
      session encoding. Cross Environment Data Access will be used,
which might require additional CPU resources and might reduce
      performance.
205      tables report_month*fatality_rate / nocum nocol nopercnt;
206      format fatality_rate pct.;
207      run;

```

NOTE: There were 5355 observations read from the data set  
HWDATA.MONTHLY\_STATS.

NOTE: PROCEDURE FREQ used (Total process time):

real time	0.11 seconds
user cpu time	0.11 seconds
system cpu time	0.01 seconds
memory	2026.00k
OS Memory	36468.00k

Timestamp	12/07/2021 04:53:55 PM
Step Count	56 Switch Count 4
Page Faults	0
Page Reclaims	273
Page Swaps	0
Voluntary Context Switches	33
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	560

```

208
209      /* 8.) In a single step create a temporary copy of the
monthly stats data set that is sorted by FIPS, Year */
210      /*      and Month number. Subset the data so that it only
contains rows from 2020 months 3 and 6 */
211      /*      and from 2021 months 1, 2, and 3. */
212
213      proc sort data=HWDATA.Monthly_stats
214          out=monthly_stats_temp;
NOTE: Data file HWDATA.MONTHLY_STATS.DATA is in a format that is
native to another host, or the file encoding does not match the
session encoding. Cross Environment Data Access will be used,
which might require additional CPU resources and might reduce
performance.
215          where (report_year=2020 and (report_month=3 or
report_month=6)) or (report_year=2021 and report_month between 1 and
3);
216          by county_fips_number report_year report_month;
217      run;

```

NOTE: There were 1275 observations read from the data set  
HWDATA.MONTHLY\_STATS.

WHERE ((report\_year=2020) and report\_month in (3, 6)) or  
((report\_year=2021) and (report\_month>=1 and report\_month<=3));

NOTE: The data set WORK.MONTHLY\_STATS\_TEMP has 1275 observations and  
8 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time	0.01 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	2300.34k
OS Memory	36988.00k
Timestamp	12/07/2021 04:53:55 PM
Step Count	57 Switch Count 2
Page Faults	0



Page Reclaims	195
Page Swaps	0
Voluntary Context Switches	20
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	272

```

218
219      /* 9.) Without using a DATA step, create a new temporary
data set that is a "wide" version of the */
220      /*      monthly stats data created in the previous step.
There will be one row per county FIPS. The */
221      /*      value for the "Cases" columns will be the number of
monthly cases for that specific year and */
222      /*      month time period as identified in the column name.
The first two rows from that data set is */
223      /*      shown below as a sample. */
224
225      proc transpose data=monthly_stats_temp
226      out = monthly_stats_transpose(drop=_)
227      prefix=Cases;
228      var monthly_cases;
229      by county_fips_number;
230      id report_year report_month;
231      run;

```

NOTE: There were 1275 observations read from the data set  
WORK.MONTHLY\_STATS\_TEMP.

NOTE: The data set WORK.MONTHLY\_STATS\_TRANSPOSE has 255 observations  
and 6 variables.

NOTE: PROCEDURE TRANSPOSE used (Total process time):

real time	0.00 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	2801.71k
OS Memory	38068.00k
Timestamp	12/07/2021 04:53:55 PM
Step Count	58 Switch Count 6
Page Faults	0
Page Reclaims	191
Page Swaps	0
Voluntary Context Switches	27
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	544

```

232
233
234      /* 10.) Use a single DATA step to combine the columns from
the permanent data set of Texas county */
235      /*      populations, the downloaded county_jobs data set,
and the wide data set created in the */
236      /*      previous step. Keep only those rows for which
there is a match in county_jobs. Create the */
237      /*      following new variables for analysis. Begin the
variable names with Pct so they can be accessed */
238      /*      with a variable list: */
239

245
246      data mergedHW15;
247      merge HWDATA.county_jobs(IN=CountyJobs)
248            mylib.txpop
249            work.monthly_stats_transpose;
NOTE: Data file HWDATA.COUNTY_JOBS.DATA is in a format that is native
to another host, or the file encoding does not match the
      session encoding. Cross Environment Data Access will be used,
which might require additional CPU resources and might reduce
      performance.
250      by county_fips_number;
251      if CountyJobs = 1;
252
253      /* (a) Percentage of population employed before the
pandemic by dividing jobs20m1 by the */
254      /* county population. */
255
256      Pct_pop_prior = jobs20m1/Population;
257
258      /* (b) Percentage of population employed early in the
pandemic by dividing jobs20m4 by the */
259      /* county population. */
260
261      Pct_pop_early = jobs20m4/Population;
262
263      /* (c) Percentage of population employed one year into the
pandemic by dividing jobs21m3 */
264      /* by the county population. */
265
266      Pct_pop_1year = jobs21m3/Population;
267

```

```

268      /* (d) Percentage of change in monthly cases from the end
of 2020 Q2 to the beginning of */
269      /* 2021 Q1. Subtract cases20206 from cases20211 and
divide the result by cases20206. */
270      /* Use conditional logic on this statement to prevent a
divide by 0 message from occurring */
271      /* in the log. */
272
273      if cases20206 ^= 0 then Pct_Change_cases = (cases20211-
cases20206)/cases20206;
274
275      /* (e) Percentage of change in monthly employment from the
end of 2020 Q2 to the beginning */
276      /* of 2021 Q1. Subtract jobs20m6 from jobs21m1 and divide
the result by jobs20m6. Use */
277      /* conditional logic on this statement to prevent a divide
by 0 message from occurring in */
278      /* the log. */
279
280      if jobs20m6 ^= 0 then Pct_change_emp = (jobs21m1-
jobs20m6)/jobs20m6;
281
282      /* f. Format these 5 new variables as a percent with 1
decimal place. */
283
284      format Pct: percent10.1;
285
286      run;

```

NOTE: There were 254 observations read from the data set  
HWDATA.COUNTY\_JOBS.

NOTE: There were 254 observations read from the data set MYLIB.TXPOP.

NOTE: There were 255 observations read from the data set  
WORK.MONTHLY\_STATS\_TRANSPOSE.

NOTE: The data set WORK.MERGEDHW15 has 254 observations and 37  
variables.

NOTE: DATA statement used (Total process time):

real time	0.01 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	1975.93k
OS Memory	36532.00k
Timestamp	12/07/2021 04:53:55 PM
Step Count	60 Switch Count 2
Page Faults	0
Page Reclaims	216

Page Swaps	0
Voluntary Context Switches	24
Involuntary Context Switches	0
Block Input Operations	288
Block Output Operations	272

```

287
288
289      /* 11) Use a single PROC step to show the extreme
observations of employment percentage before the */
290      /* pandemic and one year into the pandemic using two of
the variables created above. Show the */
291      /* county name and population value in the tables of
extreme observations. Supply an appropriate */
292      /* title. */
293
294      title "Unappropriate Title";
295      proc univariate data=work.mergedhw15;
296      var Pct_pop_early Pct_pop_1year;
297      id County_Name Population;
298      run;

```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

real time	0.12 seconds	
user cpu time	0.12 seconds	
system cpu time	0.00 seconds	
memory	1151.56k	
OS Memory	36012.00k	
Timestamp	12/07/2021 04:53:55 PM	
Step Count	61	Switch Count 0
Page Faults	0	
Page Reclaims	74	
Page Swaps	0	
Voluntary Context Switches	0	
Involuntary Context Switches	0	
Block Input Operations	0	
Block Output Operations	56	

```

299
300      /* 12) Use the data set from step 10 with the means
procedure to create an analysis of all the "Cases" */
301      /* variables by using a variable list. Show the default
statistics to one decimal place. At the same */

```

```

302      /* time create a temporary data set that contains only the
mean and range statistics.  For training */
303      /* purposes, only supply one variable for each of the
statistics.  Supply an appropriate title. */
304
305      title "Insert *Appropriate Title Here";
306      proc means data=work.mergedhw15 maxdec=1 mean min max
range;
307      var case: ;
308      output out=means_rangeHw15
309      MEAN=Mean1 /*Mean2 Mean3 Mean4 Mean5 */
310      RANGE=Range1 /*Range2 Range3 Range4 Range5;*/
311      run;
312

```

NOTE: There were 254 observations read from the data set WORK.MERGEDHW15.

NOTE: The data set WORK.MEANS\_RANGEHW15 has 1 observations and 5 variables.

NOTE: PROCEDURE MEANS used (Total process time):

real time	0.03 seconds
user cpu time	0.03 seconds
system cpu time	0.01 seconds
memory	6649.40k
OS Memory	41412.00k
Timestamp	12/07/2021 04:53:55 PM
Step Count	62 Switch Count 3
Page Faults	0
Page Reclaims	1424
Page Swaps	0
Voluntary Context Switches	33
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```

313      proc print data=work.means_rangeHW15;
314      run;

```

NOTE: There were 1 observations read from the data set WORK.MEANS\_RANGEHW15.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.01 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	672.93k

OS Memory	36012.00k	
Timestamp	12/07/2021 04:53:55 PM	
Step Count	63	Switch Count 0
Page Faults	0	
Page Reclaims	61	
Page Swaps	0	
Voluntary Context Switches	0	
Involuntary Context Switches	0	
Block Input Operations	0	
Block Output Operations	8	

```

315
316
317      /* 13.) Use the TABULATE procedure and the sorted data
set created in step 8 to show the mean and */
318      /* range of monthly cases by year and month to one decimal
place. Include the mean and range */
319      /* for all values at the bottom of the report. Supply an
appropriate title. The layout of the report */
320      /* is shown below: */
321
322      title "I am the Smartest Man Alive!";
323      proc tabulate data = work.monthly_stats_temp;
324      class report_year report_month;
325      var monthly_cases;
326      table report_year*report_month all, monthly_cases*(mean
range) *f=7.1;
327      run;

```

NOTE: There were 1275 observations read from the data set WORK.MONTHLY\_STATS\_TEMP.

NOTE: PROCEDURE TABULATE used (Total process time):

real time	0.03 seconds
user cpu time	0.01 seconds
system cpu time	0.01 seconds
memory	9673.68k
OS Memory	46288.00k
Timestamp	12/07/2021 04:53:55 PM
Step Count	64
Page Faults	1
Page Reclaims	2239
Page Swaps	0
Voluntary Context Switches	80
Involuntary Context Switches	0
Block Input Operations	8

```
328
329      /* 14. Close the PDF destination. */
330
331      ods pdf close;
NOTE: ODS PDF printed 10 pages to
/home/u59649056/Homeworks/mylib/JRodoni_HW15_Output.pdf.
332
333      /* 15. Examine the data sets and use report information
contained in your PDF output document to */
334      /* find the answers to the questions below and include the
answers in a comment section at the */
335      /* bottom of your program file: */
336
337      /* (a) Which month had the highest row percent value for
the Extreme fatality rate? */
338
339      /* March */
340
341      /* (b) Which county had the lowest percentage employment
before the pandemic and 1 year */
342      /* into the pandemic? */
343
344      /* Before pandemic: San Jacinto */
345      /* 1 Year into pandemic: San Jacinto */
346
347      /* (c) Which county had the highest percentage employment
before the pandemic and 1 year */
348      /* into the pandemic? */
349
350      /* Before pandemic: Kenedy */
351      /* 1 Year into pandemic: Loving */
352
353      /* (d) Which of the "Cases" variables had the highest mean
value? What was the value */
354
355      /* Cases20211-2371.8 */
356
357      /* (e) Compare the data in the data set created by the
MEANS procedure with the output data */
358      /* it created in the PDF. What data is represented in the
data set? */
359
360      /* The data is the mean and range of cases20203 */
```

```
361
362      /* (f) What is the overall mean and range of the Monthly
Cases? */
363
364      /* 882.20 */
365
366      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
376
```