```
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#14   */
72         /* Date Created: 11/28/2021   */
73         /* Author: Jack Rodoni   */
74         /* Purpose: STAT 604 HW#14   */
75         /* Date Modified: 12/02/2021    */
76         /* Location: /home/u59649056/Homeworks/JRodoni_Homework14.sas   */
77         /*********************************************************************************/
78
79
80         /* This assignment will use the “All Texas” - permanent data set that was created in Homework 10 and */
81         /* used in Homework 11. If you had difficulty creating this data set, the professor’s version, named */
82         /* alltx.sas7bdat, is available on the Week 9 module in Canvas and in the Fall2021 folder on SoDA. You will */
83         /* also be using three quarterly employment data sets downloaded from the U.S. Bureau of Labor */
84         /* Statistics. Download these three data sets to your homework data folder for PC SAS or download them */
85         /* then upload them to your homework data folder on SoDA. Familiarize yourself with these data sets */
86         /* before you start writing your program code. */
87
88
89         /* 1.) Add a header comment section to the beginning of a new program in your SAS session. Be sure */
90         /*    to include a comment line above each section of the program that identifies the associated */
91         /*    assignment step and a brief description of what the section is doing. Include housekeeping */
92         /*    statements to clear titles and footnotes and suppress the printing of procedure titles. */
93
94         title;
95         footnote;
96         ods noProctitle;
97
98         /* 2.) If you are using the professor’s data set, assign a libref to the folder where it is located and add */
99         /*    access=readonly at the end of the libname statement, before the semicolon, to protect data */
100        /*    sets in this folder from being accidentally overwritten. Assign a libref to the mylib folder */
101        /*    containing your permanent data sets. Create a fileref to the pdf file for output. Ensure that */
102        /*    your SAS session can locate any permanent user defined formats that you create. */
103
104        libname mylib "/home/u59649056/Homeworks/mylib";
NOTE: Libref MYLIB was successfully assigned as follows:
      Engine:        V9
      Physical Name: /home/u59649056/Homeworks/mylib
105        filename HW14pdf "/home/u59649056/Homeworks/mylib/JRodoni_HW14_Output.pdf";
106        libname  HWDATA "/home/u59649056/Homeworks/Homework Data" access = readonly;
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
107
108        /* 3.) Open a PDF destination to receive your output. */
109
110        ods pdf file=HW14pdf;
NOTE: Writing ODS PDF output to DISK destination "HW14PDF", printer "PDF".
111
112        /* 4.) The FIPS code is the common value between the COVID data we have been working with and */
113        /* the Employment data. The employment data sets are already ordered by the column containing */
114        /* the FIPS code and do not need any other modifications prior to merging. However, they do not */
115        /* contain county names and the FIPS code is a character value. Use one PROC SORT and one */
116        /* DATA step to create an unduplicated list of county names and FIPS codes from the permanent */
117        /* “All Texas” data set without altering the original data set. The final result of these two SAS steps */
118        /* will be a permanent data set with two columns ready to be merged with the employment data */
119        /* sets. NOTE: You must deal with any extra blanks that are created in the conversion from */
120        /* numeric to character or you will not get any results from your match merge process. */
121
122        PROC SORT data=HWDATA.alltx out=alltx_sort; /*out = so we dont alter the orignial data set*/
123        by COUNTY_FIPS_NUMBER;
124        run;

NOTE: There were 153255 observations read from the data set HWDATA.ALLTX.
NOTE: The data set WORK.ALLTX_SORT has 153255 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.06 seconds
      user cpu time       0.04 seconds
      system cpu time     0.04 seconds
      memory              19162.78k
      OS Memory           51800.00k
      Timestamp           11/30/2021 08:07:24 PM
      Step Count                        356  Switch Count  3
      Page Faults                       0
      Page Reclaims                     4538
```

```
      Page Swaps                     0
      Voluntary Context Switches     21
      Involuntary Context Switches   0
      Block Input Operations         0
      Block Output Operations        24080


125
126        data mylib.alltx_sort;
127        set alltx_sort;
128        by COUNTY_FIPS_NUMBER;
129        if first.COUNTY_FIPS_NUMBER;
130        AREA_FIPS = PUT(COUNTY_FIPS_NUMBER,7. -L);
131        keep area_fips COUNTY_NAME;
132        run;

NOTE: There were 153255 observations read from the data set WORK.ALLTX_SORT.
NOTE: The data set MYLIB.ALLTX_SORT has 255 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              2449.75k
      OS Memory           35076.00k
      Timestamp           11/30/2021 08:07:24 PM
      Step Count                     357  Switch Count  2
      Page Faults                    0
      Page Reclaims                  433
      Page Swaps                     0
      Voluntary Context Switches     41
      Involuntary Context Switches   0
      Block Input Operations         0
      Block Output Operations        264


133
134
135        /* 5.) Use the match merge process in a single DATA step to combine the three employment data sets */
136        /* with the list of counties to create a new permanent data set. Start with the county list then add */
137        /* the employment data sets from the oldest to the newest. The data sets are named to indicate */
138        /* the year and quarter of the data they contain. The resulting data set should have 254 */
139        /* observations and 26 variables. NOTE: A significant amount of the code for this step will be in */
140        /* data set options. */
141
142        /* a. The output data set must only contain employment data for FIPS codes that are in the */
143        /* Texas county list. */
144
145        data mylib.merged_data;
146        merge mylib.alltx_sort(IN=Texas)
147          HWDATA.employ2020q1(RENAME=(qtrly_estabs = qtrly_estabs20q1
148           month1_emplvl = emplvl20m1
149           month2_emplvl = emplvl20m2
150           month3_emplvl = emplvl20m3
151           avg_wkly_wage = avg_wkly_wage20q1))
152          HWDATA.employ2020q2(RENAME=(qtrly_estabs = qtrly_estabs20q2
153           month1_emplvl = emplvl20m4
154           month2_emplvl = emplvl20m5
155           month3_emplvl = emplvl20m6
156           avg_wkly_wage = avg_wkly_wage20q2))
157          HWDATA.employ2021q1(RENAME=(qtrly_estabs = qtrly_estabs21q1
158           month1_emplvl = emplvl21m1
159           month2_emplvl = emplvl21m2
160           month3_emplvl = emplvl21m3
161           avg_wkly_wage = avg_wkly_wage21q1));
NOTE: Data file HWDATA.EMPLOY2020Q1.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.
NOTE: Data file HWDATA.EMPLOY2020Q2.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.
NOTE: Data file HWDATA.EMPLOY2021Q1.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.
162           by area_fips;
163           if Texas = 1;
164
165
166
167        /* b. There could be up to 7 rows per FIPS depending on the types of business owners in the */
168        /* county as indicated by the own_code column. The row with an own_code value of 0 is a  */
169        /* summary of all owner types in the county. This is the only row we want to use for each */
```

```
170        /* county in our data step. The own_code column is not to be included in the output data */
171        /* set. */
172
173        if own_code='0';
174
175        /* c. The only employment statistics we want are qtrly_estabs, the three columns that begin */
176        /* with month (monthly jobs), and the avg_weekly_wage. However, we do not want any */
177        /* of the new data to overwrite the older data due to same named columns. Use naming */
178        /* patterns so that the qtrly_estabs can be accessed as a group using a variable list. */
179        /* Include the two-digit year and quarter number at the end of the column name. */
180        /* Similarly, all of the “month” columns should be named as a group with the year and */
181        /* month at the end of the name. However, use the true month number instead of the */
182        /* month within the quarter. For example, Month1 in quarter 2 is June so its name should */
183        /* end with the number 6. Finally, the avg_weekly_wage columns should be named as a */
184        /* third distinct group with the year and quarter number at the end of each name. */
185
186        /* renaming done with the rename statements*/
187
188        /* keep statement done at the end*/
189
190        /* d. Define an array that can be used to access the monthly jobs columns incrementally. */
191        /* Make the array definition dynamic such that it would not need to be changed should we */
192        /* add another quarterly data set to the merge list. */
193
194        array monthly{*} emplvl:;
195
196        /* e. Use a second array definition that will create 8 numeric variables to store the difference */
197        /* in numbers of jobs from one month to the next. The variable should get their name */
198        /* from the array name, and it should be distinct enough to not be confused with any of */
199        /* the other variable lists we have used so far. */
200
201        array diffs{8};
202
203        /* f. Use a loop to populate the monthly difference variables. Base the stop value of the loop */
204        /* on the size of the array instead of hard coding the number. The difference will be */
205        /* calculated by subtracting the month1 value from month2 and so on. NOTE: Even */
206        /* though there is a 6-month gap in the reported data, we still want to compute the */
207        /* difference between January 2021 and June of 2020 in sequence. The index variable */
208        /* must not be in the output data set. */
209
210        n = dim(diffs);
211        do i=1 to n;
212        diffs{i}=monthly{i+1}-monthly{i};
213        /* output; */
214        end;
215
216        /* g. Use the array in the argument of a function to compute the mean of all the monthly */
217        /* differences for the county. */
218
219        MEAN_DIFS = mean(of diffs{*});
220         keep qtrly_estabs: emplvl: avg_wkly_wage: COUNTY_NAME area_fips diffs: MEAN_DIFS;
221
222        run;

NOTE: MERGE statement has more than one data set with repeats of BY values.
NOTE: There were 255 observations read from the data set MYLIB.ALLTX_SORT.
NOTE: There were 19248 observations read from the data set HWDATA.EMPLOY2020Q1.
NOTE: There were 19248 observations read from the data set HWDATA.EMPLOY2020Q2.
NOTE: There were 19248 observations read from the data set HWDATA.EMPLOY2021Q1.
NOTE: The data set MYLIB.MERGED_DATA has 254 observations and 26 variables.
NOTE: DATA statement used (Total process time):
      real time           0.27 seconds
      user cpu time       0.26 seconds
      system cpu time     0.01 seconds
      memory              4709.75k
      OS Memory           37136.00k
      Timestamp           11/30/2021 08:07:24 PM
      Step Count                        358  Switch Count  2
      Page Faults                       0
      Page Reclaims                     862
      Page Swaps                        0
      Voluntary Context Switches        66
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           280



223
224
225        /* 6. Report the descriptor portion of the permanent data set of county names. Supply an */
226        /* appropriate title. */
```

```
227
228         title "My titles never make sense anyways";
229         proc contents data = mylib.alltx_sort;
230         run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.06 seconds
      user cpu time       0.05 seconds
      system cpu time     0.01 seconds
      memory              2751.00k
      OS Memory           36100.00k
      Timestamp           11/30/2021 08:07:24 PM
      Step Count                        359  Switch Count  0
      Page Faults                       0
      Page Reclaims                     441
      Page Swaps                        0
      Voluntary Context Switches        4
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           16


231
232
233         /* 7. Report the descriptor portion of the permanent data set of merged data. The variables must be */
234         /* listed in creation order instead of alphabetically. Supply an appropriate title. */
235
236         title "Do you even check this?";
237         proc contents data = mylib.merged_data varnum;
238         run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.07 seconds
      user cpu time       0.07 seconds
      system cpu time     0.00 seconds
      memory              1235.87k
      OS Memory           36356.00k
      Timestamp           11/30/2021 08:07:24 PM
      Step Count                        360  Switch Count  0
      Page Faults                       0
      Page Reclaims                     160
      Page Swaps                        0
      Voluntary Context Switches        9
      Involuntary Context Switches      0
      Block Input Operations            288
      Block Output Operations           32


239
240         /* 8. Print the changes in monthly jobs for each county from the last data set created. Show the */
241         /* county name, the 8 monthly difference variables, and the mean of the monthly differences. Use */
242         /* a variable list to specify the variables when appropriate. Do not show column labels or */
243         /* observation numbers in the output. Supply an appropriate title. */
244
245         title "Seriously, do you?";
246         proc print data=mylib.merged_data NOOBS;
247         var COUNTY_NAME diff: MEAN_DIFS;
248         run;

NOTE: There were 254 observations read from the data set MYLIB.MERGED_DATA.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.58 seconds
      user cpu time       0.59 seconds
      system cpu time     0.00 seconds
      memory              1710.53k
      OS Memory           37120.00k
      Timestamp           11/30/2021 08:07:25 PM
      Step Count                        361  Switch Count  0
      Page Faults                       0
      Page Reclaims                     277
      Page Swaps                        0
      Voluntary Context Switches        4
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           312


249
250         /* 9. Close the PDF destination. */
251
252         ods pdf close;
```

```
NOTE: ODS PDF printed 10 pages to /home/u59649056/Homeworks/mylib/JRodoni_HW14_Output.pdf.
253
254         /* 10. View the data sets, log, and report information contained in your PDF output document to find */
255         /* the answers to the questions below and include the answers in a comment section at the */
256         /* bottom of your program file: */
257
258         /* (a) What county name is in the data set of counties but not in the match merge output data */
259         /* set? Why is it not included? */
260
261         /* Unknown, b/c there is no FIPS number to work with */
262
263         /* (b) How many observations were read from each of the employment data sets? */
264
265         /* 19248 */
266
267         /* (c) How does the average monthly difference from Brazos County compare to McLennan */
268         /* County? */
269
270         /* Brazos County: -157.5 McLennan: -299.5, brazos is losing less jobs on average */
271
272         /* (d) How do the extreme monthly difference values in Brazos County compare to McLennan */
273         /* County? */
274
275         /* -14113 is the largest absolute monthly difference in Brazos, 5542 is the largest monthly gain in brazos */
276         /* -9820 is the largest absolute monthly differnce in McLennan, 4212 is the largest monthly gain in McLennan */
277
278
279         /* 11.) Save the final version of the program and convert it to a PDF file. Convert the log to PDF */
280
281
282
283         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
293
```