```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         ************************************************************
70         *  Study:             Impact of BMI on Diabetes and Cardiovascular Disease Risk
71         *  Program:           Midterm Project.sas
72         *  Created by:        Emma Mo, Emily Chiang
73         *  Date Created:      10/9/24
74         *  Description:       Merge dataset, clean data, create Tables and Graphs, perform hypothesis testing and logistic
74       ! regression
75         *
76         *
77         *  Input files:       diabetes_binary_5050split_health_indicators_BRFSS2015.csv,
77       ! heart_disease_health_indicators_BRFSS2015.csv
78         *  Output files:
79         *
80         *  Other Comments:    None
81         *
82         *
83         ************************************************************
84         *  Modification History:
85         *
86         *  Modification Number:  x
87         *  Modified by:          Shengge Cheng, Junlian Cao
88         *  Date Modified:        31OCT2024
89         *  Reason Modified:      describe changes made
90         *
91         ************************************************************
92         * Copyright  .
93         * All rights reserved.
94         ************************************************************;
95
96
97         ************************************************************;
98         ** APPLY CENTRAL CONFIGURATIONS                          **;
99         ************************************************************;
100
101        proc import datafile="/home/u64041732/Midterm/diabetes_binary_5050split_health_indicators_BRFSS2015.csv"
102            out=diabetes
103            dbms=csv
104            replace;
105            getnames=yes;
106        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.
107         /*************************************************************************
108         *   PRODUCT:   SAS
109         *   VERSION:   9.4
110         *   CREATOR:   External File Interface
111         *   DATE:      01NOV24
112         *   DESC:      Generated SAS Datastep Code
113         *   TEMPLATE SOURCE:  (None Specified.)
114         *************************************************************************/
115            data WORK.DIABETES    ;
116            %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
117            infile '/home/u64041732/Midterm/diabetes_binary_5050split_health_indicators_BRFSS2015.csv' delimiter = ',' MISSOVER
117       ! DSD lrecl=32767 firstobs=2 ;
118               informat Diabetes_binary best32. ;
119               informat HighBP best32. ;
120               informat HighChol best32. ;
121               informat CholCheck best32. ;
122               informat BMI best32. ;
123               informat Smoker best32. ;
124               informat Stroke best32. ;
125               informat HeartDiseaseorAttack best32. ;
126               informat PhysActivity best32. ;
127               informat Fruits best32. ;
128               informat Veggies best32. ;
129               informat HvyAlcoholConsump best32. ;
130               informat AnyHealthcare best32. ;
131               informat NoDocbcCost best32. ;
132               informat GenHlth best32. ;
133               informat MentHlth best32. ;
134               informat PhysHlth best32. ;
135               informat DiffWalk best32. ;
136               informat Sex best32. ;
137               informat Age best32. ;
138               informat Education best32. ;
139               informat Income best32. ;
140               format Diabetes_binary best12. ;
141               format HighBP best12. ;
142               format HighChol best12. ;
143               format CholCheck best12. ;
144               format BMI best12. ;
145               format Smoker best12. ;
146               format Stroke best12. ;
147               format HeartDiseaseorAttack best12. ;
148               format PhysActivity best12. ;
149               format Fruits best12. ;
```

```
150               format Veggies best12. ;
151               format HvyAlcoholConsump best12. ;
152               format AnyHealthcare best12. ;
153               format NoDocbcCost best12. ;
154               format GenHlth best12. ;
155               format MentHlth best12. ;
156               format PhysHlth best12. ;
157               format DiffWalk best12. ;
158               format Sex best12. ;
159               format Age best12. ;
160               format Education best12. ;
161               format Income best12. ;
162            input
163                        Diabetes_binary
164                        HighBP
165                        HighChol
166                        CholCheck
167                        BMI
168                        Smoker
169                        Stroke
170                        HeartDiseaseorAttack
171                        PhysActivity
172                        Fruits
173                        Veggies
174                        HvyAlcoholConsump
175                        AnyHealthcare
176                        NoDocbcCost
177                        GenHlth
178                        MentHlth
179                        PhysHlth
180                        DiffWalk
181                        Sex
182                        Age
183                        Education
184                        Income
185            ;
186            if _ERROR_ then call symputx('_EFIERR_',1);  /* set ERROR detection macro variable */
187            run;

NOTE: The infile '/home/u64041732/Midterm/diabetes_binary_5050split_health_indicators_BRFSS2015.csv' is:
      Filename=/home/u64041732/Midterm/diabetes_binary_5050split_health_indicators_BRFSS2015.csv,
      Owner Name=u64041732,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=22Oct2024:23:02:29,
      File Size (bytes)=6347570

NOTE: 70692 records were read from the infile '/home/u64041732/Midterm/diabetes_binary_5050split_health_indicators_BRFSS2015.csv'.
      The minimum record length was 88.
      The maximum record length was 91.
NOTE: The data set WORK.DIABETES has 70692 observations and 22 variables.
NOTE: DATA statement used (Total process time):
      real time           0.09 seconds
      user cpu time       0.09 seconds
      system cpu time     0.00 seconds
      memory              10679.37k
      OS Memory           37408.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        189  Switch Count  2
      Page Faults                       0
      Page Reclaims                     311
      Page Swaps                        0
      Voluntary Context Switches        15
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           24592


70692 rows created in WORK.DIABETES from /home/u64041732/Midterm/diabetes_binary_5050split_health_indicators_BRFSS2015.csv.



NOTE: WORK.DIABETES data set was successfully created.
NOTE: The data set WORK.DIABETES has 70692 observations and 22 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.13 seconds
      user cpu time       0.12 seconds
      system cpu time     0.02 seconds
      memory              10679.37k
      OS Memory           37924.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        189  Switch Count  10
      Page Faults                       0
      Page Reclaims                     1598
      Page Swaps                        0
      Voluntary Context Switches        88
      Involuntary Context Switches      5
      Block Input Operations            0
      Block Output Operations           24648


188
```

```
189        proc import datafile="/home/u64041732/Midterm/heart_disease_health_indicators_BRFSS2015.csv"
190            out=heartdisease
191            dbms=csv
192            replace;
193            getnames=yes;
194        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.
195         /*************************************************************************
196         *   PRODUCT:   SAS
197         *   VERSION:   9.4
198         *   CREATOR:   External File Interface
199         *   DATE:      01NOV24
200         *   DESC:      Generated SAS Datastep Code
201         *   TEMPLATE SOURCE:  (None Specified.)
202         *************************************************************************/
203            data WORK.HEARTDISEASE    ;
204            %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
205            infile '/home/u64041732/Midterm/heart_disease_health_indicators_BRFSS2015.csv' delimiter = ',' MISSOVER DSD
205      ! lrecl=32767 firstobs=2 ;
206               informat HeartDiseaseorAttack best32. ;
207               informat HighBP best32. ;
208               informat HighChol best32. ;
209               informat CholCheck best32. ;
210               informat BMI best32. ;
211               informat Smoker best32. ;
212               informat Stroke best32. ;
213               informat Diabetes best32. ;
214               informat PhysActivity best32. ;
215               informat Fruits best32. ;
216               informat Veggies best32. ;
217               informat HvyAlcoholConsump best32. ;
218               informat AnyHealthcare best32. ;
219               informat NoDocbcCost best32. ;
220               informat GenHlth best32. ;
221               informat MentHlth best32. ;
222               informat PhysHlth best32. ;
223               informat DiffWalk best32. ;
224               informat Sex best32. ;
225               informat Age best32. ;
226               informat Education best32. ;
227               informat Income best32. ;
228               format HeartDiseaseorAttack best12. ;
229               format HighBP best12. ;
230               format HighChol best12. ;
231               format CholCheck best12. ;
232               format BMI best12. ;
233               format Smoker best12. ;
234               format Stroke best12. ;
235               format Diabetes best12. ;
236               format PhysActivity best12. ;
237               format Fruits best12. ;
238               format Veggies best12. ;
239               format HvyAlcoholConsump best12. ;
240               format AnyHealthcare best12. ;
241               format NoDocbcCost best12. ;
242               format GenHlth best12. ;
243               format MentHlth best12. ;
244               format PhysHlth best12. ;
245               format DiffWalk best12. ;
246               format Sex best12. ;
247               format Age best12. ;
248               format Education best12. ;
249               format Income best12. ;
250            input
251                        HeartDiseaseorAttack
252                        HighBP
253                        HighChol
254                        CholCheck
255                        BMI
256                        Smoker
257                        Stroke
258                        Diabetes
259                        PhysActivity
260                        Fruits
261                        Veggies
262                        HvyAlcoholConsump
263                        AnyHealthcare
264                        NoDocbcCost
265                        GenHlth
266                        MentHlth
267                        PhysHlth
268                        DiffWalk
269                        Sex
270                        Age
271                        Education
272                        Income
273            ;
274            if _ERROR_ then call symputx('_EFIERR_',1);  /* set ERROR detection macro variable */
275            run;
```

```
NOTE: The infile '/home/u64041732/Midterm/heart_disease_health_indicators_BRFSS2015.csv' is:
      Filename=/home/u64041732/Midterm/heart_disease_health_indicators_BRFSS2015.csv,
      Owner Name=u64041732,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=09Oct2024:17:43:47,
      File Size (bytes)=22738147

NOTE: 253680 records were read from the infile '/home/u64041732/Midterm/heart_disease_health_indicators_BRFSS2015.csv'.
      The minimum record length was 88.
      The maximum record length was 91.
NOTE: The data set WORK.HEARTDISEASE has 253680 observations and 22 variables.
NOTE: DATA statement used (Total process time):
      real time           0.31 seconds
      user cpu time       0.29 seconds
      system cpu time     0.03 seconds
      memory              10807.78k
      OS Memory           37408.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        190  Switch Count  2
      Page Faults                       0
      Page Reclaims                     273
      Page Swaps                        0
      Voluntary Context Switches        16
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           87568


253680 rows created in WORK.HEARTDISEASE from /home/u64041732/Midterm/heart_disease_health_indicators_BRFSS2015.csv.



NOTE: WORK.HEARTDISEASE data set was successfully created.
NOTE: The data set WORK.HEARTDISEASE has 253680 observations and 22 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.36 seconds
      user cpu time       0.34 seconds
      system cpu time     0.03 seconds
      memory              10807.78k
      OS Memory           37924.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        190  Switch Count  10
      Page Faults                       0
      Page Reclaims                     1797
      Page Swaps                        0
      Voluntary Context Switches        90
      Involuntary Context Switches      4
      Block Input Operations            0
      Block Output Operations           87584


276
277        /* Sort the diabetes dataset by Age */
278        proc sort data=diabetes;
279            by age;
280        run;

NOTE: There were 70692 observations read from the data set WORK.DIABETES.
NOTE: The data set WORK.DIABETES has 70692 observations and 22 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.03 seconds
      user cpu time       0.02 seconds
      system cpu time     0.01 seconds
      memory              17535.15k
      OS Memory           47268.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        191  Switch Count  3
      Page Faults                       0
      Page Reclaims                     3776
      Page Swaps                        0
      Voluntary Context Switches        15
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           24584


281
282        /* Sort the heartdisease dataset by Age */
283        proc sort data=heartdisease;
284            by age;
285        run;

NOTE: There were 253680 observations read from the data set WORK.HEARTDISEASE.
NOTE: The data set WORK.HEARTDISEASE has 253680 observations and 22 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.13 seconds
      user cpu time       0.07 seconds
      system cpu time     0.06 seconds
      memory              53981.90k
      OS Memory           83532.00k
```

```
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        192  Switch Count  5
      Page Faults                       0
      Page Reclaims                     12759
      Page Swaps                        0
      Voluntary Context Switches        19
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           87568


286
287         /* Now merge the two datasets */
288         data merged_data;
289             merge diabetes heartdisease;
290             by age;
291         run;

NOTE: MERGE statement has more than one data set with repeats of BY values.
NOTE: There were 70692 observations read from the data set WORK.DIABETES.
NOTE: There were 253680 observations read from the data set WORK.HEARTDISEASE.
NOTE: The data set WORK.MERGED_DATA has 253680 observations and 23 variables.
NOTE: DATA statement used (Total process time):
      real time           0.07 seconds
      user cpu time       0.04 seconds
      system cpu time     0.04 seconds
      memory              5492.53k
      OS Memory           36276.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        193  Switch Count  5
      Page Faults                       0
      Page Reclaims                     927
      Page Swaps                        0
      Voluntary Context Switches        22
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           91400


292
293
294         proc contents data = diabetes;
295         run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.04 seconds
      user cpu time       0.04 seconds
      system cpu time     0.00 seconds
      memory              3319.15k
      OS Memory           32176.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        194  Switch Count  0
      Page Faults                       0
      Page Reclaims                     283
      Page Swaps                        0
      Voluntary Context Switches        3
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           24


296
297         proc contents data = heartdisease;
298         run;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              2440.68k
      OS Memory           32176.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        195  Switch Count  0
      Page Faults                       0
      Page Reclaims                     283
      Page Swaps                        0
      Voluntary Context Switches        3
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           24


299
300
301         /* Step 1: Assign descriptive labels to each age group */
302
303         data categorized_data;
304             set merged_data;
305
306             /* Categorize BMI */
307             if bmi < 18.5 then nbmi = 'Underweight';
```

```
308           else if bmi <= 24.9 and bmi >= 18.5 then nbmi = 'Healthy';
309           else if bmi <= 29.9 and bmi >= 25 then nbmi = 'Overweight';
310           else if bmi <= 39.9 and bmi >= 30 then nbmi = 'Obese';
311           else if bmi >= 40 then nbmi = 'Severely obese';
312
313       run;

NOTE: There were 253680 observations read from the data set WORK.MERGED_DATA.
NOTE: The data set WORK.CATEGORIZED_DATA has 253680 observations and 24 variables.
NOTE: DATA statement used (Total process time):
      real time           0.06 seconds
      user cpu time       0.03 seconds
      system cpu time     0.04 seconds
      memory              3599.18k
      OS Memory           34224.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        196  Switch Count  9
      Page Faults                       0
      Page Reclaims                     561
      Page Swaps                        0
      Voluntary Context Switches        38
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           99344


314
315       /* Step 2: Calculate the percentage of Diabetes and Heart Disease for each BMI category */
316
317       proc freq data=categorized_data;
318           tables nbmi*Diabetes_binary / nocol nopercent norow;
319           tables nbmi*HeartDiseaseorAttack / nocol nopercent norow;
320       run;

NOTE: There were 253680 observations read from the data set WORK.CATEGORIZED_DATA.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.07 seconds
      user cpu time       0.07 seconds
      system cpu time     0.01 seconds
      memory              2666.53k
      OS Memory           32436.00k
      Timestamp           11/01/2024 10:18:25 PM
      Step Count                        197  Switch Count  11
      Page Faults                       0
      Page Reclaims                     376
      Page Swaps                        0
      Voluntary Context Switches        49
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           544


321
322       /* Step 3: Use PROC TABULATE to generate a summary table with age ranges */
323
324       proc tabulate data=categorized_data;
325           class nbmi;
326           var Diabetes_binary HeartDiseaseorAttack;
327           table nbmi, (Diabetes_binary HeartDiseaseorAttack)*(mean='Percentage') / misstext='0.0%';
328       run;

NOTE: There were 253680 observations read from the data set WORK.CATEGORIZED_DATA.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.03 seconds
      user cpu time       0.04 seconds
      system cpu time     0.02 seconds
      memory              9537.46k
      OS Memory           38344.00k
      Timestamp           11/01/2024 10:18:26 PM
      Step Count                        198  Switch Count  12
      Page Faults                       0
      Page Reclaims                     1846
      Page Swaps                        0
      Voluntary Context Switches        134
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           544


329
330       /* Pie Chart for Diabetes */
331       proc gchart data=categorized_data;
332           where Diabetes_binary = 1; /* Select individuals with diabetes */
333           pie nbmi / coutline=black
334                      percent=inside
335                      noheading
336                      value=arrow;
337           title "Percentage of Individuals with Diabetes by BMI Category";
338       run;

339
```

```
340        /* Pie Chart for Heart Disease */

NOTE: There were 218334 observations read from the data set WORK.CATEGORIZED_DATA.
      WHERE Diabetes_binary=1;
NOTE: PROCEDURE GCHART used (Total process time):
      real time           0.23 seconds
      user cpu time       0.20 seconds
      system cpu time     0.03 seconds
      memory              18221.90k
      OS Memory           48304.00k
      Timestamp           11/01/2024 10:18:26 PM
      Step Count                        199  Switch Count  36
      Page Faults                       0
      Page Reclaims                     4411
      Page Swaps                        0
      Voluntary Context Switches        102
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           5216


341        proc gchart data=categorized_data;
342            where HeartDiseaseorAttack = 1; /* Select individuals with heart disease */
343            pie nbmi / coutline=black
344                       percent=inside
345                       noheading
346                       value=arrow;
347            title "Percentage of Individuals with Heart Disease by BMI Category";
348        run;

349
350        /* Reshape data for plotting */

NOTE: There were 23893 observations read from the data set WORK.CATEGORIZED_DATA.
      WHERE HeartDiseaseorAttack=1;
NOTE: PROCEDURE GCHART used (Total process time):
      real time           0.16 seconds
      user cpu time       0.14 seconds
      system cpu time     0.02 seconds
      memory              7652.40k
      OS Memory           37152.00k
      Timestamp           11/01/2024 10:18:26 PM
      Step Count                        200  Switch Count  5
      Page Faults                       0
      Page Reclaims                     1199
      Page Swaps                        0
      Voluntary Context Switches        25
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           800


351        data plot_data;
352            set categorized_data;
353            Category = "Diabetes";
354            Count = Diabetes_binary;
355            output;
356
357            Category = "Heart Disease";
358            Count = HeartDiseaseorAttack;
359            output;
360
361            keep nbmi Category Count;
362        run;

NOTE: There were 253680 observations read from the data set WORK.CATEGORIZED_DATA.
NOTE: The data set WORK.PLOT_DATA has 507360 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.05 seconds
      user cpu time       0.04 seconds
      system cpu time     0.02 seconds
      memory              3589.53k
      OS Memory           36016.00k
      Timestamp           11/01/2024 10:18:26 PM
      Step Count                        201  Switch Count  7
      Page Faults                       0
      Page Reclaims                     561
      Page Swaps                        0
      Voluntary Context Switches        26
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           32008


363
364        /* Stacked Bar Chart */
365        proc sgplot data=plot_data;
366            vbar Category / response=Count group=nbmi groupdisplay=stack datalabel;
367            keylegend / title="BMI Category";
368            xaxis label="Condition";
369            yaxis label="Count";
```

```
370           title "Stacked Bar Chart of Diabetes and Heart Disease by BMI Category";
371       run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.17 seconds
      user cpu time       0.12 seconds
      system cpu time     0.01 seconds
      memory              9457.42k
      OS Memory           37940.00k
      Timestamp           11/01/2024 10:18:26 PM
      Step Count                        202  Switch Count  5
      Page Faults                       0
      Page Reclaims                     1595
      Page Swaps                        0
      Voluntary Context Switches        227
      Involuntary Context Switches      4
      Block Input Operations            0
      Block Output Operations           480

NOTE: There were 507360 observations read from the data set WORK.PLOT_DATA.

372
373       /* hypothesis testing - Chi-square */
374
375       proc freq data=categorized_data;
376           tables nbmi*Diabetes_binary / chisq;
377       run;

NOTE: There were 253680 observations read from the data set WORK.CATEGORIZED_DATA.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.07 seconds
      user cpu time       0.06 seconds
      system cpu time     0.01 seconds
      memory              2692.71k
      OS Memory           37812.00k
      Timestamp           11/01/2024 10:18:26 PM
      Step Count                        203  Switch Count  4
      Page Faults                       0
      Page Reclaims                     458
      Page Swaps                        0
      Voluntary Context Switches        27
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           528


378
379       proc freq data=categorized_data;
380           tables nbmi*HeartDiseaseorAttack / chisq;
381       run;

NOTE: There were 253680 observations read from the data set WORK.CATEGORIZED_DATA.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.07 seconds
      user cpu time       0.07 seconds
      system cpu time     0.01 seconds
      memory              2469.34k
      OS Memory           38068.00k
      Timestamp           11/01/2024 10:18:26 PM
      Step Count                        204  Switch Count  11
      Page Faults                       0
      Page Reclaims                     385
      Page Swaps                        0
      Voluntary Context Switches        44
      Involuntary Context Switches      2
      Block Input Operations            0
      Block Output Operations           544


382
383       /* hypothesis testing - t-test */
384
385       proc ttest data=categorized_data;
386       var BMI;
387       class Diabetes_binary;
388       RUN;

NOTE: PROCEDURE TTEST used (Total process time):
      real time           6.71 seconds
      user cpu time       5.24 seconds
      system cpu time     0.59 seconds
      memory              29906.00k
      OS Memory           64488.00k
      Timestamp           11/01/2024 10:18:33 PM
      Step Count                        205  Switch Count  149
      Page Faults                       0
      Page Reclaims                     62800
      Page Swaps                        0
      Voluntary Context Switches        60937
      Involuntary Context Switches      74
      Block Input Operations            0
```

```
      Block Output Operations           603512


389
390
391       proc ttest data=categorized_data;
392       var BMI;
393       class heartdiseaseorattack;
394       RUN;

NOTE: PROCEDURE TTEST used (Total process time):
      real time           6.82 seconds
      user cpu time       5.36 seconds
      system cpu time     0.62 seconds
      memory              31150.68k
      OS Memory           66580.00k
      Timestamp           11/01/2024 10:18:40 PM
      Step Count                        206  Switch Count  138
      Page Faults                       0
      Page Reclaims                     63841
      Page Swaps                        0
      Voluntary Context Switches        57073
      Involuntary Context Switches      75
      Block Input Operations            0
      Block Output Operations           625480


395
396
397       proc tabulate data=categorized_data;
398       class age nbmi;
399       var diabetes_binary heartdiseaseorattack;
400       table age*nbmi,(diabetes_binary heartdiseaseorattack)*(mean='Percentage')/misstext='0.0%';
401       run;

NOTE: There were 253680 observations read from the data set WORK.CATEGORIZED_DATA.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.06 seconds
      user cpu time       0.07 seconds
      system cpu time     0.02 seconds
      memory              9764.71k
      OS Memory           45512.00k
      Timestamp           11/01/2024 10:18:40 PM
      Step Count                        207  Switch Count  7
      Page Faults                       0
      Page Reclaims                     1902
      Page Swaps                        0
      Voluntary Context Switches        217
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           568


402
403       /*perform logistic regression because of binary data*/
404       proc logistic data=categorized_data;
405           model Diabetes_binary(event='1') = BMI; /* Replace with actual variable names */
406           title "Logistic Regression of Diabetes on BMI";
407       run;

NOTE: PROC LOGISTIC is modeling the probability that Diabetes_binary='1'.
NOTE: Convergence criterion (GCONV=1E-8) satisfied.
NOTE: There were 253680 observations read from the data set WORK.CATEGORIZED_DATA.
NOTE: PROCEDURE LOGISTIC used (Total process time):
      real time           0.45 seconds
      user cpu time       0.41 seconds
      system cpu time     0.05 seconds
      memory              30073.18k
      OS Memory           68436.00k
      Timestamp           11/01/2024 10:18:40 PM
      Step Count                        208  Switch Count  13
      Page Faults                       0
      Page Reclaims                     5082
      Page Swaps                        0
      Voluntary Context Switches        51
      Involuntary Context Switches      4
      Block Input Operations            0
      Block Output Operations           46128


408
409       proc logistic data=categorized_data;
410           model Diabetes_binary(event='1') = BMI age; /* Replace with actual variable names */
411           title "Logistic Regression of Diabetes with Multiple Predictors";
412       run;

NOTE: PROC LOGISTIC is modeling the probability that Diabetes_binary='1'.
NOTE: Convergence criterion (GCONV=1E-8) satisfied.
NOTE: There were 253680 observations read from the data set WORK.CATEGORIZED_DATA.
NOTE: PROCEDURE LOGISTIC used (Total process time):
      real time           0.47 seconds
```

```
      user cpu time       0.42 seconds
      system cpu time     0.05 seconds
      memory              32058.25k
      OS Memory           70164.00k
      Timestamp           11/01/2024 10:18:41 PM
      Step Count                        209  Switch Count  24
      Page Faults                       0
      Page Reclaims                     5067
      Page Swaps                        0
      Voluntary Context Switches        89
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           48688


413
414        proc logistic data=categorized_data;
415            model heartdiseaseorattack(event='1') = BMI age; /* Replace with actual variable names */
416            title "Logistic Regression of Diabetes with Multiple Predictors";
417        run;

NOTE: PROC LOGISTIC is modeling the probability that HeartDiseaseorAttack='1'.
NOTE: Convergence criterion (GCONV=1E-8) satisfied.
NOTE: There were 253680 observations read from the data set WORK.CATEGORIZED_DATA.
NOTE: PROCEDURE LOGISTIC used (Total process time):
      real time           0.51 seconds
      user cpu time       0.46 seconds
      system cpu time     0.05 seconds
      memory              31946.00k
      OS Memory           70164.00k
      Timestamp           11/01/2024 10:18:41 PM
      Step Count                        210  Switch Count  24
      Page Faults                       0
      Page Reclaims                     5067
      Page Swaps                        0
      Voluntary Context Switches        78
      Involuntary Context Switches      5
      Block Input Operations            0
      Block Output Operations           48688


418
419
420        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
430
```