```
 1 *------------------------------------------------------------
   -*
 2 User:                   u63452984
 3 Date:                   07 January 2024
 4 Time:                   09:25:48
 5 Site:                   70094220
 6 Platform:               Linux
 7 Maintenance Release: 9.04.01M7P080620
 8 EM Version:             15.2
 9 *
10 *------------------------------------------------------------
   -*
11 * Training Log
12 Date:                   07 January 2024
13 Time:                   09:25:45
14 *------------------------------------------------------------
   -*
15 11450  proc freq data=EMWS3.Stat2_VariableSet noprint;
16 11451  table ROLE*LEVEL/out=WORK.Stat2META;
17 11452  run;
18 11453  proc print data=WORK.Stat2META label noobs;
19 11454  var ROLE LEVEL COUNT;
20 11455  label ROLE = "%sysfunc(sasmsg(sashelp.dmine, meta_rol
   e_vlabel, NOQUOTE))" LEVEL = "%sysfunc(sasmsg(sashelp.dmine,
    meta_level_vlabel, NOQUOTE))" COUNT = "%sysfunc(sasmsg(sash
   elp.dmine, rpt_count_vlabel, NOQUOTE))";
21 11456  title9 ' ';
22 11457  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_varSummar
   y_title  , NOQUOTE))";
23 11458  run;
24 11459  title10;
25 11460  data WORK.M2F5TXB6;
26 11461  set WORK.M2F5TXB6;
27 11462  where((role in('TARGET' 'FREQ' 'INPUT') and use in('D
   ','Y')) or (role = 'REJECTED' and use = 'Y'));
28 11463  if ROLE = 'REJECTED' then role = 'INPUT';
```

```
29 11464  run;
30 11465  *-------------------------------------------------
   --------*;
31 11466  * Stat2: Determining Analysis Variables;
32 11467  *-------------------------------------------------
   --------*;
33 11468  *-------------------------------------------------
   --------*;
34 11469  * Stat2: Creating Macros for Variable Selection;
35 11470  *-------------------------------------------------
   --------*;
36 11471  *-------------------------------------------------
   --------* ;
37 11472  * Stat2: Interval Input Variables Macro ;
38 11473  *-------------------------------------------------
   --------* ;
39 11474  %macro INTINPUTS;
40 11475      Age IMP_TotalSpent TotalPurchases
41 11476  %mend INTINPUTS;
42 11477  *-------------------------------------------------
   --------*;
43 11478  * Stat2: Creating Macros for Variable Selection;
44 11479  *-------------------------------------------------
   --------*;
45 11480  *-------------------------------------------------
   --------* ;
46 11481  * Stat2: Binary and Nominal Input Variables Macro ;
47 11482  *-------------------------------------------------
   --------* ;
48 11483  %macro NOMINPUTS;
49 11484      M_Variable MembershipLevel PaymentMethod
50 11485  %mend NOMINPUTS;
51 11486  *-------------------------------------------------
   --------*;
52 11487  * Stat2: Creating Macros for Variable Selection;
53 11488  *-------------------------------------------------
```

```
--------*;
54 11489  *-------------------------------------------------
--------* ;
55 11490  * Stat2: Ordinal Input Variables Macro ;
56 11491  *-------------------------------------------------
--------* ;
57 11492  %macro ORDINPUTS;
58 11493
59 11494  %mend ORDINPUTS;
60 11495  data work.EM_Stat2_tree / view=work.EM_Stat2_tree;
61 11496  set EMWS3.Impt_TRAIN(obs=100000 keep=%INTINPUTS %ORDI
   NPUTS %NOMINPUTS IMP_Churn);
62 11497  run;
63 11498  proc arbor data=EMWS3.Impt_TRAIN(obs=100000 keep=%INT
   INPUTS %ORDINPUTS %NOMINPUTS IMP_Churn) Criterion=GINI Leafs
   ize=5 Mincatsize = 5 Maxbranch=5 Maxdepth=1 Padjust= NONE NO
   RULELIMIT MAXRULES=6 MAXSURRS=0 Missing=USEINSEARCH Exhausti
   ve=5000;
64 WARNING: PADJUST and PVAR options are ignored with this spli
   tting criterion.
65 11499  input %INTINPUTS / level = interval;
66 11500  input %NOMINPUTS / level=nominal;
67 11501  target IMP_Churn / level=BINARY;
68 11502  Performance Disk NodeSize=10000;
69 11503  Assess NoValidata measure=MISC;
70 11504  SUBTREE BEST;
71 11505  save RULES=WORK.Stat2_RULE;
72 11506  run;
73 11507  quit;
74 11508  data WORK.Stat2_RULE(keep=Target Name Rank Numeric_Va
   lue StatVar rename=(numeric_value=Worth));
75 11509  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
   rget_vlabel,      NOQUOTE))" Name = "%sysfunc(sasmsg(sashelp
   .dmine, rpt_variable_vlabel,    NOQUOTE))" Rank = "%sysfunc(
   sasmsg(sashelp.dmine, rpt_importance_vlabel,  NOQUOTE))" Num
   eric_Value =
```

```
 76 11510     "%sysfunc(sasmsg(sashelp.dmine, rpt_worth_vlabel,
          NOQUOTE))" StatVar = "%sysfunc(sasmsg(sashelp.dmine, r
    pt_analysisVar_vlabel, NOQUOTE))";
 77 11511  length Target $32 Name $32;
 78 11512  retain TARGET "IMP_Churn" Name;
 79 11513  format STATVAR 6.0;
 80 11514  set WORK.Stat2_RULE;
 81 11515  where stat in('VARIABLE','WORTH');
 82 11516  if stat = 'VARIABLE' then Name = character_value;
 83 11517  else do;
 84 11518  if _N_<= 2*1000 then STATVAR=1;
 85 11519  else STATVAR=0;
 86 11520  output;
 87 11521  end;
 88 11522  run;
 89 11523  proc append base=EMWS3.Stat2_WORTH data=WORK.Stat2_RU
    LE force;
 90 11524  run;
 91 11525  *-------------------------------------------------
    --------*;
 92 11526  * Stat2: Counting Levels;
 93 11527  *-------------------------------------------------
    --------*;
 94 11528  proc sql;
 95 11529  create view WORK.Stat2_distinct as select distinct IM
    P_Churn from EMWS3.Impt_TRAIN(obs=100000);
 96 11530  quit;
 97 11531  proc sql;
 98 11532  reset noprint;
 99 11533  select count(*) into :_tmpcount from WORK.Stat2_disti
    nct;
100 11534  quit;
101 11535  proc sort data=WORK.Stat2_count NOTHREADS;
102 11536  by NAME ROLE DATA;
103 11537  run;
104 11538  proc print data=WORK.Stat2_count(obs=500) label noobs
```

```
    ;
105 11539  label NAME = "%sysfunc(sasmsg(sashelp.dmine, rpt_vari
    able_vlabel,    NOQUOTE))" COUNT = "%sysfunc(sasmsg(sashelp.
    dmine, rpt_count_vlabel,       NOQUOTE))" ROLE = "%sysfunc(s
    asmsg(sashelp.dmine, meta_role_vlabel,       NOQUOTE))";
106 11540  var NAME ROLE COUNT;
107 11541  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_varLevels_
    title, NOQUOTE))";
108 11542  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title, NOQUOTE))";
109 11543  run;
110 11544  title9;
111 11545  title10;
112 11546  proc sort data=EMWS3.Stat2_WORTH nodupkey out=WORK.St
    at2_SELVAR(keep=Name StatVar);
113 11547  by NAME;
114 11548  where statvar=1;
115 11549  run;
116 11550  proc sort data=EMWS3.Stat2_WORTH;
117 11551  by Rank;
118 11552  where statvar=1;
119 11553  run;
120 11554  proc sort data=WORK.M1L5A6IE;
121 11555  by name;
122 11556  run;
123 11557  data WORK.Stat2_analysisMeta;
124 11558  merge WORK.M1L5A6IE WORK.Stat2_SELVAR(in=_a);
125 11559  by name;
126 11560  if _a then STATVAR = 1;
127 11561  else if ROLE in('INPUT', 'REJECTED') and REPORT ne 'Y
    ' then delete;
128 11562  if REPORT eq 'Y' then STATVAR = 1;
129 11563  run;
130 11564  data WORK.M00AWJ9_;
131 11565  set WORK.M00AWJ9_;
132 11566  where(use = 'Y' or Report='Y' or (role in('TARGET' 'F
```

```
    REQ' 'PREDICT' 'RESIDUAL' 'INPUT') and use = 'D'));
133 11567  if ROLE = 'SEGMENT' then delete;
134 11568  else if ROLE ^in('FREQ','TARGET') then role = 'INPUT'
    ;
135 11569  run;
136 11570  data WORK.M3KM_NHK;
137 11571  set WORK.M3KM_NHK;
138 11572  where(use = 'Y' or Report='Y' or (role in('SEGMENT',
    'TARGET' 'FREQ' 'PREDICT' 'RESIDUAL' 'INPUT') and use = 'D')
    );
139 11573  if ROLE ne 'FREQ' then role = 'INPUT';
140 11574  run;
141 11575  *-------------------------------------------------
    --------*;
142 11576  * Stat2: Computing Statistics for Interval Variables;
143 11577  *-------------------------------------------------
    --------*;
144 11578  proc dmdb data=EMWS3.Impt_TRAIN(obs=100000) nonorm ma
    xlevel=513
145 11579  varout=work._DMDBVAR(RENAME=(NAME=VARIABLE))
146 11580  classout=WORK.Stat2CLASS(drop=NMISSPERCENT rename=(NA
    ME=VARIABLE FREQUENCY=COUNT FREQPERCENT=PERCENT))
147 11581  ;
148 11582  var
149 11583  Age IMP_TotalSpent TotalPurchases
150 11584  ;
151 11585  class
152 11586  IMP_Churn M_Variable MembershipLevel PaymentMethod
153 11587  ;
154 11588  ;
155 11589  run;
156 11590  proc stdize data = EMWS3.Impt_TRAIN(obs=100000) out=_
    null_ outstat=work._STDIZE add=0 fuzz=1E-14 initial=MAD vard
    ef=df method=STD mult=1 pctlmtd=ORD_STAT pctldef=2 pctlpts=(
    0 5 50 95 100);
157 11591  var
```

```
158 11592  Age IMP_TotalSpent TotalPurchases
159 11593  ;
160 11594  run;
161 11595  proc transpose data=work._STDIZE out=work._TRANSSTDIZ
    E(drop=_LABEL_ rename=(_NAME_=VARIABLE));
162 11596  id _type_;
163 11597  where _TYPE_='P50';
164 11598  run;
165 11599  proc sort data=work._TRANSSTDIZE;
166 11600  by VARIABLE;
167 11601  run;
168 11602  proc sort data=work._DMDBVAR;
169 11603  by VARIABLE;
170 11604  run;
171 11605  data WORK.Stat2INTERVAL;
172 11606  merge work._TRANSSTDIZE work._DMDBVAR;
173 11607  format N NMISS 8. MIN MAX P50 MEAN STD SKEWNESS KURTO
    SIS BEST8.3;
174 11608  by VARIABLE;
175 11609  run;
176 WARNING: Multiple lengths were specified for the BY variable
     VARIABLE by input data sets. This might cause unexpected re
    sults.
177 11610  proc sort data=WORK.Stat2CLASS;
178 11611  by Variable Level;
179 11612  run;
180 11613  data colorindex;
181 11614  retain LevelIndex 0;
182 11615  set WORK.Stat2CLASS(keep=Variable Level);
183 11616  by variable level;
184 11617  if first.variable then LevelIndex = 0;
185 11618  if first.level then LevelIndex + 1;
186 11619  run;
187 11620  data WORK.Stat2CLASS;
188 11621  merge WORK.Stat2CLASS(in=_a) colorindex;
189 11622  by variable level;
```

```
190 11623  if _a then output;
191 11624  run;
192 11625  data WORK.Stat2INTERVAL;
193 11626  length DATAROLE $20;
194 11627  retain DATAROLE "TRAIN";
195 11628  length Target $32 TargetValue $32;
196 11629  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
197 11630  retain Target '_OVERALL_';
198 11631  set WORK.Stat2INTERVAL;
199 11632  run;
200 11633  proc append base=EMWS3.Stat2_INTERVAL data=WORK.Stat2
    INTERVAL force;
201 11634  run;
202 11635  data WORK.Stat2CLASS;
203 11636  length DATAROLE $20;
204 11637  retain DATAROLE "TRAIN";
205 11638  length Target $32 TargetValue $32;
206 11639  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
207 11640  retain Target '_OVERALL_';
208 11641  set WORK.Stat2CLASS;
209 11642  run;
210 11643  proc append base=EMWS3.Stat2_CLASS data=WORK.Stat2CLA
    SS force;
211 11644  run;
212 11645  data WORK.M31M6GJD;
213 11646  set WORK.M31M6GJD;
214 11647  where(use = 'Y' or Report='Y' or (role in('TARGET' 'F
    REQ' 'PREDICT' 'RESIDUAL' 'INPUT') and use = 'D'));
215 11648  if ROLE = 'SEGMENT' or (ROLE = 'TARGET' and LEVEL ne
    'INTERVAL') then delete;
216 11649  else if ROLE ne 'FREQ' then role = 'INPUT';
217 11650  run;
```

```
218 11651  proc freq data=EMWS3.Impt_TRAIN noprint;
219 11652  table
220 11653  IMP_Churn / out=EMWS3.Stat2_BYVAR(drop=COUNT PERCENT)
     missing;
221 11654  run;
222 11655  data WORK.STATEXPLORE_DATA / view=WORK.STATEXPLORE_DA
    TA;
223 11656  set EMWS3.Impt_TRAIN;
224 11657  where left(trim(put(IMP_Churn,BEST12.))) ="0";
225 11658  run;
226 11659  *-------------------------------------------------
    --------*;
227 11660  * Stat2: Computing Statistics for Interval Variables;
228 11661  *-------------------------------------------------
    --------*;
229 11662  proc dmdb data=WORK.STATEXPLORE_DATA(obs=100000) nono
    rm maxlevel=513
230 11663  varout=work._DMDBVAR(RENAME=(NAME=VARIABLE))
231 11664  classout=WORK.Stat2CLASS(drop=NMISSPERCENT rename=(NA
    ME=VARIABLE FREQUENCY=COUNT FREQPERCENT=PERCENT))
232 11665  ;
233 11666  var
234 11667  Age IMP_TotalSpent TotalPurchases
235 11668  ;
236 11669  class
237 11670  M_Variable MembershipLevel PaymentMethod
238 11671  ;
239 11672  ;
240 11673  run;
241 11674  proc stdize data = WORK.STATEXPLORE_DATA(obs=100000)
    out=_null_ outstat=work._STDIZE add=0 fuzz=1E-14 initial=MAD
     vardef=df method=STD mult=1 pctlmtd=ORD_STAT pctldef=2 pctl
    pts=(0 5 50 95 100);
242 11675  var
243 11676  Age IMP_TotalSpent TotalPurchases
244 11677  ;
```

```
245 11678  run;
246 11679  proc transpose data=work._STDIZE out=work._TRANSSTDIZ
    E(drop=_LABEL_ rename=(_NAME_=VARIABLE));
247 11680  id _type_;
248 11681  where _TYPE_='P50';
249 11682  run;
250 11683  proc sort data=work._TRANSSTDIZE;
251 11684  by VARIABLE;
252 11685  run;
253 11686  proc sort data=work._DMDBVAR;
254 11687  by VARIABLE;
255 11688  run;
256 11689  data WORK.Stat2INTERVAL;
257 11690  merge work._TRANSSTDIZE work._DMDBVAR;
258 11691  format N NMISS 8. MIN MAX P50 MEAN STD SKEWNESS KURTO
    SIS BEST8.3;
259 11692  by VARIABLE;
260 11693  run;
261 WARNING: Multiple lengths were specified for the BY variable
     VARIABLE by input data sets. This might cause unexpected re
    sults.
262 11694  proc sort data=WORK.Stat2CLASS;
263 11695  by Variable Level;
264 11696  run;
265 11697  data WORK.Stat2CLASS;
266 11698  merge WORK.Stat2CLASS(in=_a) colorindex;
267 11699  by variable level;
268 11700  if _a then output;
269 11701  run;
270 11702  data WORK.Stat2INTERVAL;
271 11703  length DATAROLE $20;
272 11704  retain DATAROLE "TRAIN";
273 11705  length Target $32 TargetValue $32;
274 11706  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
```

```
275 11707  Target = "IMP_Churn";
276 11708  TargetValue = '0';
277 11709  set WORK.Stat2INTERVAL;
278 11710  run;
279 11711  proc append base=EMWS3.Stat2_INTERVAL data=WORK.Stat2
    INTERVAL force;
280 11712  run;
281 11713  data WORK.Stat2CLASS;
282 11714  length DATAROLE $20;
283 11715  retain DATAROLE "TRAIN";
284 11716  length Target $32 TargetValue $32;
285 11717  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
286 11718  Target = "IMP_Churn";
287 11719  TargetValue = '0';
288 11720  set WORK.Stat2CLASS;
289 11721  run;
290 11722  proc append base=EMWS3.Stat2_CLASS data=WORK.Stat2CLA
    SS force;
291 11723  run;
292 11724  data WORK.STATEXPLORE_DATA / view=WORK.STATEXPLORE_DA
    TA;
293 11725  set EMWS3.Impt_TRAIN;
294 11726  where left(trim(put(IMP_Churn,BEST12.))) ="1";
295 11727  run;
296 11728  *------------------------------------------------
    --------*;
297 11729  * Stat2: Computing Statistics for Interval Variables;
298 11730  *------------------------------------------------
    --------*;
299 11731  proc dmdb data=WORK.STATEXPLORE_DATA(obs=100000) nono
    rm maxlevel=513
300 11732  varout=work._DMDBVAR(RENAME=(NAME=VARIABLE))
301 11733  classout=WORK.Stat2CLASS(drop=NMISSPERCENT rename=(NA
    ME=VARIABLE FREQUENCY=COUNT FREQPERCENT=PERCENT))
```

```
302  11734  ;
303  11735  var
304  11736  Age IMP_TotalSpent TotalPurchases
305  11737  ;
306  11738  class
307  11739  M_Variable MembershipLevel PaymentMethod
308  11740  ;
309  11741  ;
310  11742  run;
311  11743  proc stdize data = WORK.STATEXPLORE_DATA(obs=100000)
     out=_null_ outstat=work._STDIZE add=0 fuzz=1E-14 initial=MAD
      vardef=df method=STD mult=1 pctlmtd=ORD_STAT pctldef=2 pctl
     pts=(0 5 50 95 100);
312  11744  var
313  11745  Age IMP_TotalSpent TotalPurchases
314  11746  ;
315  11747  run;
316  11748  proc transpose data=work._STDIZE out=work._TRANSSTDIZ
     E(drop=_LABEL_ rename=(_NAME_=VARIABLE));
317  11749  id _type_;
318  11750  where _TYPE_='P50';
319  11751  run;
320  11752  proc sort data=work._TRANSSTDIZE;
321  11753  by VARIABLE;
322  11754  run;
323  11755  proc sort data=work._DMDBVAR;
324  11756  by VARIABLE;
325  11757  run;
326  11758  data WORK.Stat2INTERVAL;
327  11759  merge work._TRANSSTDIZE work._DMDBVAR;
328  11760  format N NMISS 8. MIN MAX P50 MEAN STD SKEWNESS KURTO
     SIS BEST8.3;
329  11761  by VARIABLE;
330  11762  run;
331  WARNING: Multiple lengths were specified for the BY variable
      VARIABLE by input data sets. This might cause unexpected re
```

```
    sults.
332 11763  proc sort data=WORK.Stat2CLASS;
333 11764  by Variable Level;
334 11765  run;
335 11766  data WORK.Stat2CLASS;
336 11767  merge WORK.Stat2CLASS(in=_a) colorindex;
337 11768  by variable level;
338 11769  if _a then output;
339 11770  run;
340 11771  data WORK.Stat2INTERVAL;
341 11772  length DATAROLE $20;
342 11773  retain DATAROLE "TRAIN";
343 11774  length Target $32 TargetValue $32;
344 11775  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
345 11776  Target = "IMP_Churn";
346 11777  TargetValue = '1';
347 11778  set WORK.Stat2INTERVAL;
348 11779  run;
349 11780  proc append base=EMWS3.Stat2_INTERVAL data=WORK.Stat2
    INTERVAL force;
350 11781  run;
351 11782  data WORK.Stat2CLASS;
352 11783  length DATAROLE $20;
353 11784  retain DATAROLE "TRAIN";
354 11785  length Target $32 TargetValue $32;
355 11786  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
356 11787  Target = "IMP_Churn";
357 11788  TargetValue = '1';
358 11789  set WORK.Stat2CLASS;
359 11790  run;
360 11791  proc append base=EMWS3.Stat2_CLASS data=WORK.Stat2CLA
    SS force;
```

```
361 11792  run;
362 11793  proc sort data=EMWS3.Stat2_VariableSet out=tempVariab
    le(keep=name label role rename=(name=variable));
363 11794  by name;
364 11795  run;
365 11796  proc sort data=EMWS3.Stat2_CLASS;
366 11797  by Variable;
367 11798  run;
368 11799  data EMWS3.Stat2_CLASS;
369 11800  merge EMWS3.Stat2_CLASS(in=_a) tempVariable;
370 11801  by variable;
371 11802  if label = '' then label=Variable;
372 11803  if _a then output;
373 11804  run;
374 WARNING: Multiple lengths were specified for the BY variable
     VARIABLE by input data sets. This might cause unexpected re
    sults.
375 11805  proc sort data=EMWS3.Stat2_INTERVAL NOTHREADS;
376 11806  by DATAROLE
377 11807  Target TargetValue
378 11808  ;
379 11809  run;
380 11810  run;
381 11811  proc sort data=EMWS3.Stat2_VariableSet out=tempVariab
    le(keep=name label role rename=(name=variable));
382 11812  by name;
383 11813  run;
384 11814  proc sort data=EMWS3.Stat2_INTERVAL;
385 11815  by Variable;
386 11816  run;
387 11817  data EMWS3.Stat2_INTERVAL;
388 11818  merge EMWS3.Stat2_INTERVAL(in=_a) tempVariable;
389 11819  by variable;
390 11820  if label = '' then label=Variable;
391 11821  if _a then output;
392 11822  run;
```

```
393 WARNING: Multiple lengths were specified for the BY variable
     VARIABLE by input data sets. This might cause unexpected re
    sults.
394 11823  proc datasets library=EMWS3 nolist;
395 11824  modify Stat2_INTERVAL;
396 11825  label DATAROLE = "%sysfunc(sasmsg(sashelp.dmine, rpt_
    datarole_vlabel, NOQUOTE))";
397 11826  label Variable = "%sysfunc(sasmsg(sashelp.dmine, rpt_
    variable_vlabel, NOQUOTE))";
398 11827  label MEAN = "%sysfunc(sasmsg(sashelp.dmine, rpt_mean
    _vlabel,    NOQUOTE))";
399 11828  label N = "%sysfunc(sasmsg(sashelp.dmine, rpt_nonMiss
    _vlabel, NOQUOTE))";
400 11829  label STD = "%sysfunc(sasmsg(sashelp.dmine, rpt_std_v
    label,     NOQUOTE))";
401 11830  label NMISS = "%sysfunc(sasmsg(sashelp.dmine, rpt_mis
    sing_vlabel, NOQUOTE))";
402 11831  label p50 = "%sysfunc(sasmsg(sashelp.dmine, rpt_media
    n_vlabel, NOQUOTE))";
403 11832  label MIN = "%sysfunc(sasmsg(sashelp.dmine, rpt_minim
    um_vlabel, NOQUOTE))";
404 11833  label MAX = "%sysfunc(sasmsg(sashelp.dmine, rpt_maxim
    um_vlabel, NOQUOTE))";
405 11834  label SKEWNESS = "%sysfunc(sasmsg(sashelp.dmine, rpt_
    skewness_vlabel, NOQUOTE))";
406 11835  label KURTOSIS = "%sysfunc(sasmsg(sashelp.dmine, rpt_
    kurtosis_vlabel, NOQUOTE))";
407 11836  label TARGET = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel,   NOQUOTE))" TARGETVALUE = "%sysfunc(sasmsg(sas
    help.dmine, rpt_targetLevel_vlabel,   NOQUOTE))";
408 11837  run;
409 11838  proc sort data=EMWS3.Stat2_CLASS NOTHREADS;
410 11839  by DATAROLE ROLE Variable descending COUNT;
411 11840  run;
412 11841  data WORK.Stat2CLASS(keep=DATAROLE Variable ROLE Numc
    at Nmiss Mode ModePct Mode2 Mode2Pct);
```

```
413 11842  set EMWS3.Stat2_CLASS;
414 11843  length Mode Mode2 $32;
415 11844  retain NMiss 0 ModePct 0 Mode2Pct 0 Mode '' Mode2 '';
416 11845  by DATAROLE ROLE Variable;
417 11846  if first.DATAROLE or first.variable then do;
418 11847  Numcat = 1;
419 11848  NMiss = 0;
420 11849  Mode = strip(LEVEL);
421 11850  ModePct = PERCENT;
422 11851  Mode2 = '';
423 11852  Mode2Pct = 0;
424 11853  end;
425 11854  else Numcat +1;
426 11855  if Numcat =2 then do;
427 11856  Mode2 = strip(LEVEL);
428 11857  Mode2Pct = PERCENT;
429 11858  end;
430 11859  if NRAW=. and CRAW='' then NMiss= COUNT;
431 11860  if last.DATAROLE or last.variable then output;
432 11861  where TARGET='_OVERALL_';
433 11862  run;
434 11863  %let _cn = %sysfunc(getoption(CENTER));
435 11864  %let _nb = %sysfunc(getoption(NUMBER));
436 11865  options nonumber nocenter;
437 11866  title;
438 11867  title9 ' ';
439 11868  proc print data=WORK.Stat2CLASS(obs=500) label noobs;
440 11869  var DATAROLE VARIABLE ROLE NUMCAT NMISS Mode ModePct
    Mode2 Mode2Pct;
441 11870  label DATAROLE = "%sysfunc(sasmsg(sashelp.dmine, rpt_
    datarole_vlabel, NOQUOTE))" ROLE = "%sysfunc(sasmsg(sashelp.
    dmine, meta_role_vlabel, NOQUOTE))" NUMCAT = "%sysfunc(sasms
    g(sashelp.dmine, rpt_numcat_vlabel,   NOQUOTE))" NMISS =
442 11871     "%sysfunc(sasmsg(sashelp.dmine, rpt_missing_vlabel
    ,  NOQUOTE))" MODE = "%sysfunc(sasmsg(sashelp.dmine, rpt_mod
    e_vlabel,     NOQUOTE))" MODEPCT = "%sysfunc(sasmsg(sashelp.
```

```
    dmine, rpt_modepct_vlabel,  NOQUOTE))" MODE2 =
443 11872     "%sysfunc(sasmsg(sashelp.dmine, rpt_mode2_vlabel,
        NOQUOTE))" MODE2PCT = "%sysfunc(sasmsg(sashelp.dmine, rp
    t_mode2pct_vlabel, NOQUOTE))";
444 11873  format ModePct Mode2Pct 5.2;
445 11874  by DATAROLE;
446 11875  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_classStats
    _title,     NOQUOTE))";
447 11876  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title,  NOQUOTE))";
448 11877  run;
449 11878  proc print data=EMWS3.Stat2_CLASS(obs=500) noobs labe
    l;
450 11879  var DATAROLE VARIABLE ROLE LEVEL COUNT PERCENT;
451 11880  where ROLE in('TARGET', 'SEGMENT');
452 11881  by DATAROLE;
453 11882  label DATAROLE = "%sysfunc(sasmsg(sashelp.dmine,  rpt
    _datarole_vlabel, NOQUOTE))" ROLE = "%sysfunc(sasmsg(sashelp
    .dmine, meta_role_vlabel,          NOQUOTE))" LEVEL = "%sysf
    unc(sasmsg(sashelp.dmine, rpt_level_vlabel,          NOQUOTE
    ))" COUNT =
454 11883     "%sysfunc(sasmsg(sashelp.dmine, rpt_count_vlabel,
            NOQUOTE))" PERCENT = "%sysfunc(sasmsg(sashelp.dmine
    , rpt_percent_vlabel,        NOQUOTE))";
455 11884  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_classTarge
    tSegmentDist_title, NOQUOTE))";
456 11885  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title,          NOQUOTE))";
457 11886  run;
458 11887  title9;
459 11888  title10;
460 11889  options &_cn &_nb;
461 11890  data WORK.Stat2INTERVALSUMMARY;
462 11891  set EMWS3.Stat2_INTERVAL;
463 11892  length ROLE $32;
464 11893  where TARGET='_OVERALL_';
```

```
465 11894  if Variable = "Age" then ROLE="INPUT";
466 11895  else
467 11896  if Variable = "IMP_TotalSpent" then ROLE="INPUT";
468 11897  else
469 11898  if Variable = "TotalPurchases" then ROLE="INPUT";
470 11899  run;
471 11900  proc sort data=WORK.Stat2INTERVALSUMMARY NOTHREADS;
472 11901  by DATAROLE ROLE Variable;
473 11902  run;
474 11903  %let _cn = %sysfunc(getoption(CENTER));
475 11904  %let _nb = %sysfunc(getoption(NUMBER));
476 11905  options nonumber nocenter;
477 11906  title;
478 11907  proc print data=WORK.Stat2INTERVALSUMMARY(obs=500) la
    bel noobs;
479 11908  var Variable ROLE MEAN STD N NMISS MIN P50 MAX SKEWNE
    SS KURTOSIS;
480 11909  by DATAROLE;
481 11910  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_intervalSt
    ats_title, NOQUOTE))";
482 11911  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title, NOQUOTE))";
483 11912  run;
484 11913  title9;
485 11914  title10;
486 11915  options &_cn &_nb;
487 11916  proc sort data=EMWS3.Stat2_CLASS NOTHREADS;
488 11917  by DATAROLE DATAROLE Target TargetValue Variable desc
    ending COUNT;
489 11918  run;
490 11919  data WORK.Stat2CLASS(keep=DATAROLE Target TargetValue
     DATAROLE Variable Numcat NMiss Mode ModePct Mode2 Mode2Pct
    );
491 11920  set EMWS3.Stat2_CLASS;
492 11921  length Mode Mode2 $32;
493 11922  retain NMiss 0 ModePct 0 Mode2Pct 0 Mode '' Mode2 '';
```

```
494  11923  by DATAROLE DATAROLE Target TargetValue Variable;
495  11924  if first.datarole or first.variable then do;
496  11925  Numcat = 1;
497  11926  NMiss = 0;
498  11927  Mode = LEVEL;
499  11928  ModePct = PERCENT;
500  11929  Mode2 = '';
501  11930  Mode2Pct = 0;
502  11931  end;
503  11932  else Numcat +1;
504  11933  if Numcat =2 then do;
505  11934  Mode2 = LEVEL;
506  11935  Mode2Pct = PERCENT;
507  11936  end;
508  11937  if NRAW=. and CRAW='' then NMiss= COUNT;
509  11938  if last.datarole or last.variable then output;
510  11939  where ROLE ^in('TARGET', 'SEGMENT');
511  11940  run;
512  11941  proc sort data=WORK.Stat2CLASS NOTHREADS;
513  11942  by DATAROLE
514  11943  VARIABLE Target TargetValue;
515  11944  run;
516  11945  %let _cn = %sysfunc(getoption(CENTER));
517  11946  %let _nb = %sysfunc(getoption(NUMBER));
518  11947  options nonumber nocenter;
519  11948  title;
520  11949  title9 ' ';
521  11950  proc print data=WORK.Stat2CLASS(obs=500) label noobs;
522  11951  label DATAROLE = "%sysfunc(sasmsg(sashelp.dmine, rpt_
     datarole_vlabel, NOQUOTE))" NUMCAT = "%sysfunc(sasmsg(sashel
     p.dmine, rpt_numcat_vlabel,   NOQUOTE))" NMISS = "%sysfunc(s
     asmsg(sashelp.dmine, rpt_missing_vlabel,  NOQUOTE))" MODE =
523  11952     "%sysfunc(sasmsg(sashelp.dmine, rpt_mode_vlabel,
        NOQUOTE))" MODEPCT = "%sysfunc(sasmsg(sashelp.dmine, rpt_
     modepct_vlabel,  NOQUOTE))" MODE2 = "%sysfunc(sasmsg(sashelp
     .dmine, rpt_mode2_vlabel,     NOQUOTE))" MODE2PCT =
```

```
524 11953     "%sysfunc(sasmsg(sashelp.dmine, rpt_mode2pct_vlabe
    l, NOQUOTE))";
525 11954  format ModePct Mode2Pct 5.2;
526 11955  by DATAROLE;
527 11956  label TARGET = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel,   NOQUOTE))" TARGETVALUE = "%sysfunc(sasmsg(sas
    help.dmine, rpt_targetLevel_vlabel,   NOQUOTE))";
528 11957  var Target TargetValue NumCat NMiss Mode ModePct Mode
    2 Mode2Pct;
529 11958  by DATAROLE VARIABLE;
530 11959  title9 "Class Variable Summary Statistics by Class Ta
    rget";
531 11960  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title, NOQUOTE))";
532 11961  run;
533 11962  title9;
534 11963  title10;
535 11964  options &_cn &_nb;
536 11965  proc sort data=EMWS3.Stat2_INTERVAL NOTHREADS;
537 11966  by DATAROLE Variable
538 11967  Target
539 11968  TargetValue
540 11969  ;
541 11970  run;
542 11971  %let _cn = %sysfunc(getoption(CENTER));
543 11972  %let _nb = %sysfunc(getoption(NUMBER));
544 11973  options nonumber nocenter;
545 11974  title;
546 11975  proc print data=EMWS3.Stat2_INTERVAL(obs=500) label n
    oobs;
547 11976  by DATAROLE Variable
548 11977  ;
549 11978  title9 "Interval Variable Summary Statistics by Class
     Target";
550 11979  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title, NOQUOTE))";
```

```
551 11980  run;
552 11981  title9;
553 11982  title10;
554 11983  options &_cn &_nb;
555 11984  *-------------------------------------------------
    --------*;
556 11985  * Stat2: Computing Chi-Square Statistics;
557 11986  *-------------------------------------------------
    --------*;
558 11987  ods listing close;
559 11988  ods output chisq=WORK.TEMPCHISQUARE;
560 11989  ods output crosstabFreqs=WORK.TEMPCROSSTAB;
561 11990  proc freq data=EMWS3.Impt_TRAIN(obs=100000) order=for
    matted addnames;
562 11991  table IMP_Churn*M_Variable /missing chisq outexpect s
    parse;
563 11992  table IMP_Churn*MembershipLevel /missing chisq outexp
    ect sparse;
564 11993  table IMP_Churn*PaymentMethod /missing chisq outexpec
    t sparse;
565 11994  ;
566 11995  run;
567 11996  quit;
568 11997  ods listing;
569 11998  data WORK.Stat2CHI2(keep=TARGET INPUT CRAMERV CHIMEAS
    URE PROBCHI CHIDF rename=(PROBCHI=PROB CHIDF=DF));
570 11999  length TARGET INPUT $32;
571 12000  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel ,      NOQUOTE))" INPUT = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_input_vlabel ,       NOQUOTE))" CRAMERV = "%sy
    sfunc(sasmsg(sashelp.dmine, rpt_cramerv_vlabel ,     NOQUOTE
    ))" CORRVALUE =
572 12001     "%sysfunc(sasmsg(sashelp.dmine, rpt_correlation_vl
    abel , NOQUOTE))" PROBCHI = "%sysfunc(sasmsg(sashelp.dmine,
    rpt_probchi_vlabel ,     NOQUOTE))" CHIMEASURE = "%sysfunc(s
    asmsg(sashelp.dmine, rpt_chisquare_vlabel,    NOQUOTE))" CHI
```

```
    DF =
573 12002     "%sysfunc(sasmsg(sashelp.dmine, rpt_df_vlabel ,
          NOQUOTE))";
574 12003  retain CHIMEASURE CHIDF PROBCHI;
575 12004  format PROBCHI PVALUE6.4 CHIMEASURE 10.4;
576 12005  set WORK.TEMPCHISQUARE;
577 12006  Target = strip(RowVariable);
578 12007  Input = strip(ColVariable);
579 12008  where StatisticId in('_CRAMV_','_PCHI_');
580 12009  if StatisticId = '_PCHI_' then do;
581 12010  PROBCHI = Prob;
582 12011  CHIMEASURE = VALUE;
583 12012  CHIDF = DF;
584 12013  end;
585 12014  else do;
586 12015  CRAMERV = abs(VALUE);
587 12016  OUTPUT;
588 12017  end;
589 12018  run;
590 12019  proc sort data=WORK.TEMPCROSSTAB;
591 12020  by RowVariable ColVariable _Type_;
592 12021  where _Type_ in('11','00');
593 12022  run;
594 12023  data WORK.Stat2CHI(keep=YVAR XVAR X Y XTEXT YTEXT COU
    NT CHISQUARE LOGCHISQUARE);
595 12024  retain N;
596 12025  dsid = open("WORK.TEMPCROSSTAB");
597 12026  rowvarnum = varnum(dsid, 'RowVariable');
598 12027  colvarnum = varnum(dsid, 'ColVariable');
599 12028  freqnum = varnum(dsid, 'FREQUENCY');
600 12029  rownum = varnum(dsid, 'RowPercent');
601 12030  colnum = varnum(dsid, 'ColPercent');
602 12031  typenum = varnum(dsid, '_TYPE_');
603 12032  length YVAR XVAR $32 YTEXT $32 XTEXT $32;
604 12033  do while(^fetch(dsid));
605 12034  COUNT = getvarn(dsid, freqnum);
```

```
606 12035  YVAR = (getvarc(dsid, rowvarnum));
607 12036  XVAR = strip(getvarc(dsid, colvarnum));
608 12037  ynum = varnum(dsid, YVAR);
609 12038  ytype = vartype(dsid, ynum);
610 12039  yformat = varfmt(dsid, ynum);
611 12040  if yformat eq '' then yformat = 'best12.';
612 12041  Y=.;
613 12042  if ytype = 'N' then do;
614 12043  Y = getvarn(dsid, ynum);
615 12044  YTEXT = left(putn(Y, yformat));
616 12045  end;
617 12046  else YTEXT = getvarc(dsid, ynum);
618 12047  xnum = varnum(dsid, XVAR);
619 12048  xtype = vartype(dsid, xnum);
620 12049  xformat = varfmt(dsid, xnum);
621 12050  if xformat eq '' then xformat = 'best12.';
622 12051  X=.;
623 12052  if xtype = 'N' then do;
624 12053  X = getvarn(dsid, xnum);
625 12054  XTEXT = left(putn(X, xformat));
626 12055  end;
627 12056  else XTEXT = getvarc(dsid, xnum);
628 12057  if getvarc(dsid, typenum)= '00' then N= count;
629 12058  else do;
630 12059  EXP = (getvarn(dsid,rownum)/100)*(getvarn(dsid,colnum
    )/100)*N;
631 12060  if exp >0 then do;
632 12061  CHISQUARE=(EXP-count)**2/(EXP);
633 12062  if chisquare>0 then LOGCHISQUARE = log(CHISQUARE);
634 12063  OUTPUT;
635 12064  end;
636 12065  end;
637 12066  end;
638 12067  dsid = close(dsid);
639 12068  run;
640 12069  data WORK.Stat2CHI;
```

```
641  12070  length DATAROLE $20;
642  12071  retain DATAROLE "TRAIN";
643  12072  SegmentVar = "";
644  12073  SegmentValue = "";
645  12074  SegmentId = "_OVERALL_";
646  12075  set WORK.Stat2CHI;
647  12076  run;
648  12077  proc append base=EMWS3.Stat2_CHI2 data=WORK.Stat2CHI
     force;
649  12078  run;
650  12079  data WORK.Stat2CHI2;
651  12080  length DATAROLE $20;
652  12081  retain DATAROLE "TRAIN";
653  12082  SegmentVar = "";
654  12083  SegmentValue = "";
655  12084  SegmentId = "_OVERALL_";
656  12085  set WORK.Stat2CHI2;
657  12086  run;
658  12087  proc append base=EMWS3.Stat2_CHIMEASURE data=WORK.Sta
     t2CHI2 force;
659  12088  run;
660  12089  proc sort data=EMWS3.Stat2_VariableSet out=tempVariab
     le(keep=name label role rename=(label=XLabel name=XVAR));
661  12090  by name;
662  12091  run;
663  12092  proc sort data=EMWS3.Stat2_CHI2;
664  12093  by XVAR;
665  12094  run;
666  12095  data EMWS3.Stat2_CHI2;
667  12096  merge EMWS3.Stat2_CHI2(in=_a) tempVariable;
668  12097  by XVAR;
669  12098  if XLabel = '' then XLabel=XVAR;
670  12099  if _a then output;
671  12100  run;
672  WARNING: Multiple lengths were specified for the BY variable
      XVAR by input data sets. This might cause unexpected result
```

```
    s.
673 12101  proc datasets library=EMWS3 nolist;
674 12102  modify Stat2_CHI2;
675 12103  label DATAROLE = "%sysfunc(sasmsg(sashelp.dmine,  rpt
    _datarole_vlabel, NOQUOTE))";
676 12104  label SEGMENTVAR = "%sysfunc(sasmsg(sashelp.dmine, rp
    t_segment_vlabel,      NOQUOTE))";
677 12105  label SEGMENTID = "%sysfunc(sasmsg(sashelp.dmine, rpt
    _segmentnamevalue_vlabel,      NOQUOTE))";
678 12106  label SegmentValue= "%sysfunc(sasmsg(sashelp.dmine, r
    pt_segmentid_vlabel, NOQUOTE))";
679 12107  label YVAR = "%sysfunc(sasmsg(sashelp.dmine, rpt_targ
    et_vlabel,         NOQUOTE))";
680 12108  label XVAR = "%sysfunc(sasmsg(sashelp.dmine, rpt_inpu
    t_vlabel,          NOQUOTE))";
681 12109  label X = "%sysfunc(sasmsg(sashelp.dmine, rpt_inputNu
    mValue_vlabel,  NOQUOTE))";
682 12110  label Y = "%sysfunc(sasmsg(sashelp.dmine, rpt_targetN
    umValue_vlabel, NOQUOTE))";
683 12111  label XTEXT = "%sysfunc(sasmsg(sashelp.dmine, rpt_inp
    utFmtValue_vlabel,  NOQUOTE))";
684 12112  label YTEXT = "%sysfunc(sasmsg(sashelp.dmine, rpt_tar
    getFmtValue_vlabel, NOQUOTE))";
685 12113  label COUNT = "%sysfunc(sasmsg(sashelp.dmine, rpt_cou
    nt_vlabel,          NOQUOTE))";
686 12114  label CHISQUARE = "%sysfunc(sasmsg(sashelp.dmine, rpt
    _chiSquare_vlabel,      NOQUOTE))";
687 12115  label LOGCHISQUARE = "%sysfunc(sasmsg(sashelp.dmine,
    rpt_logChiSquare_vlabel,   NOQUOTE))";
688 12116  run;
689 12117  proc sort data=EMWS3.Stat2_VariableSet out=tempVariab
    le(keep=name label role rename=(name=INPUT));
690 12118  by name;
691 12119  run;
692 12120  proc sort data=EMWS3.Stat2_CHIMEASURE;
693 12121  by INPUT;
```

```
694  12122  run;
695  12123  data EMWS3.Stat2_CHIMEASURE;
696  12124  merge EMWS3.Stat2_CHIMEASURE(in=_a) tempVariable;
697  12125  by INPUT;
698  12126  if label = '' then label=INPUT;
699  12127  if _a then output;
700  12128  run;
701  WARNING: Multiple lengths were specified for the BY variable
      INPUT by input data sets. This might cause unexpected resul
     ts.
702  12129  proc sort data=EMWS3.Stat2_CHIMEASURE NOTHREADS;
703  12130  by DATAROLE TARGET descending CRAMERV;
704  12131  run;
705  12132  data EMWS3.Stat2_CHIMEASURE;
706  12133  set EMWS3.Stat2_CHIMEASURE;
707  12134  label ORDEREDINPUTS = "%sysfunc(sasmsg(sashelp.dmine,
      rpt_orderedInput_vlabel  , NOQUOTE))" DATAROLE = "%sysfunc(
     sasmsg(sashelp.dmine,  rpt_datarole_vlabel, NOQUOTE))" SEGME
     NTID =
708  12135     "%sysfunc(sasmsg(sashelp.dmine, rpt_segmentnameval
     ue_vlabel,       NOQUOTE))" SegmentValue = "%sysfunc(sasmsg(
     sashelp.dmine, rpt_segmentid_vlabel, NOQUOTE))" SEGMENTVAR =
      "%sysfunc(sasmsg(sashelp.dmine, rpt_segment_vlabel,      NO
     QUOTE))";
709  12136  format ORDEREDINPUTS 5.0;
710  12137  retain ORDEREDINPUTS 0;
711  12138  if first.DATAROLE or first.SEGMENTID or first.TARGET
     then ORDEREDINPUTS=1;
712  12139  else ORDEREDINPUTS+1;
713  12140  by DATAROLE SEGMENTID TARGET;
714  12141  run;
715  12142  %let _cn = %sysfunc(getoption(CENTER));
716  12143  %let _nb = %sysfunc(getoption(NUMBER));
717  12144  options nonumber nocenter;
718  12145  title;
719  12146  proc print data=EMWS3.Stat2_CHIMEASURE(obs=500) label
```

```
 noobs;
720 12147  var INPUT CHIMEASURE DF PROB;
721 12148  label INPUT = "%sysfunc(sasmsg(sashelp.dmine, rpt_inp
    ut_vlabel ,       NOQUOTE))" CHIMEASURE = "%sysfunc(sasmsg(s
    ashelp.dmine, rpt_chisquare_vlabel,    NOQUOTE))" PROB = "%s
    ysfunc(sasmsg(sashelp.dmine, rpt_probchi_vlabel ,     NOQUOT
    E))" DF =
722 12149     "%sysfunc(sasmsg(sashelp.dmine, rpt_df_vlabel ,
           NOQUOTE))";
723 12150  by DATAROLE TARGET;
724 12151  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_chisquareS
    tats_title, NOQUOTE))";
725 12152  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title,  NOQUOTE))";
726 12153  run;
727 12154  title10;
728 12155  options &_cn &_nb;
729 12156  proc sort data=EMWS3.Stat2_CHIMEASURE out=trainChimea
    sure;
730 12157  by TARGET descending CramerV;
731 12158  where DATAROLE='TRAIN';
732 12159  run;
733 12160  data trainChimeasure;
734 12161  set trainchimeasure;
735 12162  by TARGET;
736 12163  retain groupcount 0;
737 12164  if first.Target then groupcount =1;
738 12165  else groupcount + 1;
739 12166  if groupcount<21 then Plot=1;
740 12167  else Plot = 0;
741 12168  keep Target Input GroupCount Plot;
742 12169  run;
743 12170  proc sort data=trainChiMeasure;
744 12171  by TARGET INPUT;
745 12172  run;
746 12173  proc sort data=EMWS3.Stat2_CHIMEASURE;
```

```
747 12174  by TARGET INPUT;
748 12175  run;
749 12176  data EMWS3.Stat2_CHIMEASURE;
750 12177  merge EMWS3.Stat2_CHIMEASURE trainchimeasure;
751 12178  by TARGET INPUT;
752 12179  label DATAROLE = "%sysfunc(sasmsg(sashelp.dmine,  rpt
    _datarole_vlabel, NOQUOTE))" Plot = "%sysfunc(sasmsg(sashelp
    .dmine, rpt_plot_vlabel,  NOQUOTE))" groupcount = "%sysfunc(
    sasmsg(sashelp.dmine, rpt_group_vlabel, NOQUOTE))";
753 12180  run;
754 12181  proc sort data=EMWS3.Stat2_CLASS(drop=CRAW NRAW) NOTH
    READS;
755 12182  by DATAROLE TARGET VARIABLE LEVEL;
756 12183  where TARGET ne '_OVERALL_';
757 12184  run;
758 12185  data WORK.Stat2tempDs(keep=DATAROLE target VARIABLE n
    obs);
759 12186  retain nobs;
760 12187  set EMWS3.Stat2_CLASS;
761 12188  by DATAROLE target VARIABLE;
762 12189  if first.DATAROLE or first.variable then nobs=count;
763 12190  else nobs = nobs + count;
764 12191  if last.DATAROLE or last.variable then output;
765 12192  run;
766 12193  data EMWS3.Stat2_CLASS(drop=NOBS);
767 12194  merge EMWS3.Stat2_CLASS(where=(TARGET ne '_OVERALL_')
     rename=(PERCENT=WITHINPCT)) WORK.Stat2tempDs;
768 12195  label percent = "%sysfunc(sasmsg(sashelp.dmine, rpt_p
    ercent_vlabel,         NOQUOTE))" withinpct = "%sysfunc(sasm
    sg(sashelp.dmine, rpt_percentwithin_vlabel, NOQUOTE))";
769 12196  by DATAROLE TARGET VARIABLE;
770 12197  if NOBS>0 then PERCENT=COUNT/NOBS;
771 12198  run;
772 12199  proc sort data=EMWS3.Stat2_CHIMEASURE out=WORK.Stat2c
    utoff(RENAME=(INPUT=VARIABLE) keep=DATAROLE TARGET INPUT PLO
    T);
```

```
773 12200  by DATAROLE TARGET INPUT;
774 12201  run;
775 12202  data EMWS3.Stat2_CLASS;
776 12203  drop groupcount;
777 12204  retain groupcount 0;
778 12205  merge EMWS3.Stat2_CLASS(in=_a) WORK.Stat2cutoff;
779 12206  if _a then do;
780 12207  plot=1;
781 12208  groupcount+1;
782 12209  end;
783 12210  if groupcount>21 then plot=0;
784 12211  by DATAROLE Target Variable;
785 12212  if PERCENT eq . or plot = . then delete;
786 12213  label DATAROLE = "%sysfunc(sasmsg(sashelp.dmine,  rpt
    _datarole_vlabel, NOQUOTE))" LEVEL = "%sysfunc(sasmsg(sashel
    p.dmine, rpt_level_vlabel,          NOQUOTE))" TYPE = "%sysf
    unc(sasmsg(sashelp.dmine, rpt_type_vlabel,           NOQUOTE
    ))" COUNT =
787 12214     "%sysfunc(sasmsg(sashelp.dmine, rpt_count_vlabel,
             NOQUOTE))" ROLE = "%sysfunc(sasmsg(sashelp.dmine, m
    eta_role_vlabel,          NOQUOTE))" LEVELINDEX = "%sysfunc(
    sasmsg(sashelp.dmine, rpt_levelIndex_vlabel,     NOQUOTE))";
788 12215  run;
789 12216  proc sort data=EMWS3.Stat2_CHIMEASURE;
790 12217  by DATAROLE ORDEREDINPUTS;
791 12218  run;
792 12219  data WORK.Stat2INTERVALPLOT(drop=TARGET N MIN NMISS P
    50 MAX STD SKEWNESS KURTOSIS TARGETVALUE rename=(Mean=OMean)
    );
793 12220  set EMWS3.Stat2_INTERVAL;
794 12221  where TARGET='_OVERALL_';
795 12222  run;
796 12223  data EMWS3.Stat2_INTERVAL(drop=OMean);
797 12224  merge EMWS3.Stat2_INTERVAL(where=(TARGET ne '_OVERALL
    _')) WORK.Stat2INTERVALPLOT;
798 12225  by DATAROLE Variable;
```

```
799  12226  if OMean ne 0 then ScaleDevMean = (Mean - OMean)/OMea
     n;
800  12227  run;
801  12228  proc sort data=EMWS3.Stat2_INTERVAL NOTHREADS;
802  12229  by DATAROLE Variable Target;
803  12230  run;
804  12231  data WORK.Stat2MAXINTERVAL(keep=DATAROLE Target Varia
     ble MaxDev);
805  12232  set EMWS3.Stat2_INTERVAL;
806  12233  if first.DATAROLE or first.Variable then MaxDev = abs
     (ScaleDevMean);
807  12234  else if abs(ScaleDevMean)>MaxDev then MaxDev = abs(Sc
     aleDevMean);
808  12235  if last.datarole or last.variable or last.target then
      output;
809  12236  by DATAROLE Variable Target;
810  12237  run;
811  12238  data EMWS3.Stat2_INTERVAL;
812  12239  merge EMWS3.Stat2_INTERVAL WORK.Stat2MAXINTERVAL;
813  12240  label LevelId = "%sysfunc(sasmsg(sashelp.dmine, rpt_l
     evelId_vlabel,             NOQUOTE))" Variable = "%sysfunc(s
     asmsg(sashelp.dmine, rpt_variable_vlabel,            NOQUOTE
     ))" MaxDev =
814  12241      "%sysfunc(sasmsg(sashelp.dmine, rpt_maxdev_vlabel,
                    NOQUOTE))" ScaleDevMean = "%sysfunc(sasmsg(sas
     help.dmine, rpt_scaledMeanDeviation_vlabel, NOQUOTE))";
815  12242  format LevelId best3.;
816  12243  retain LevelId 0;
817  12244  by DATAROLE Variable Target;
818  12245  if first.DATAROLE or first.Target then LevelId=1;
819  12246  else LevelId + 1;
820  12247  run;
821  12248  proc sort data=EMWS3.Stat2_INTERVAL NOTHREADS;
822  12249  by DATAROLE Target descending MaxDev Variable LevelId
     ;
823  12250  run;
```

```
824 12251  proc sort data=WORK.M1L5A6IE;
825 12252  by name;
826 12253  run;
827 12254  proc sort data=EMWS3.Stat2_WORTH;
828 12255  by NAME;
829 12256  run;
830 12257  data EMWS3.Stat2_WORTH;
831 12258  merge EMWS3.Stat2_WORTH(in=_a) WORK.M1L5A6IE(Keep=NAM
    E LABEL REPORT);
832 12259  by NAME;
833 12260  if REPORT = 'Y' then plot = 1;
834 12261  if LABEL = "" then LABEL = NAME;
835 12262  if _a then output;
836 12263  drop report;
837 12264  run;
838 WARNING: Multiple lengths were specified for the BY variable
     Name by input data sets. This might cause unexpected result
    s.
839 12265  proc sort data=EMWS3.Stat2_WORTH;
840 12266  by rank;
841 12267  run;
842 *------------------------------------------------------------
    -*
843 * Score Log
844 Date:                07 January 2024
845 Time:                09:25:47
846 *------------------------------------------------------------
    -*
847 12369  *------------------------------------------------------
    --------*;
848 12370  * Stat2: Scoring DATA data;
849 12371  *------------------------------------------------------
    --------*;
850 12372  data EMWS3.Stat2_TRAIN
851 12373  / view=EMWS3.Stat2_TRAIN
852 12374  ;
```

```
853 12375  set EMWS3.Impt_TRAIN
854 12376  ;
855 12377  run;
856 12378  quit;
857 12379  *------------------------------------------------------
    --------*;
858 12380  * Stat2: Computing metadata for TRAIN data;
859 12381  *------------------------------------------------------
    --------*;
860 *------------------------------------------------------------
    -*
861 * Report Log
862 Date:                 07 January 2024
863 Time:                 09:25:48
864 *------------------------------------------------------------
    -*
```