```
 1 *------------------------------------------------------------
   -*
 2 User:                   u63452984
 3 Date:                   07 January 2024
 4 Time:                   09:10:39
 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:                   09:10:37
14 *------------------------------------------------------------
   -*
15 11449  proc freq data=EMWS3.Stat_VariableSet noprint;
16 11450  table ROLE*LEVEL/out=WORK.StatMETA;
17 11451  run;
18 11452  proc print data=WORK.StatMETA label noobs;
19 11453  var ROLE LEVEL COUNT;
20 11454  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 11455  title9 ' ';
22 11456  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_varSummar
   y_title  , NOQUOTE))";
23 11457  run;
24 11458  title10;
25 11459  data WORK.M39AYFW7;
26 11460  set WORK.M39AYFW7;
27 11461  where((role in('TARGET' 'FREQ' 'INPUT') and use in('D
   ','Y')) or (role = 'REJECTED' and use = 'Y'));
28 11462  if ROLE = 'REJECTED' then role = 'INPUT';
```

```
29 11463  run;
30 11464  *-------------------------------------------------
   --------*;
31 11465  * Stat: Determining Analysis Variables;
32 11466  *-------------------------------------------------
   --------*;
33 11467  *-------------------------------------------------
   --------*;
34 11468  * Stat: Creating Macros for Variable Selection;
35 11469  *-------------------------------------------------
   --------*;
36 11470  *-------------------------------------------------
   --------* ;
37 11471  * Stat: Interval Input Variables Macro ;
38 11472  *-------------------------------------------------
   --------* ;
39 11473  %macro INTINPUTS;
40 11474      Age TotalPurchases TotalSpent
41 11475  %mend INTINPUTS;
42 11476  *-------------------------------------------------
   --------*;
43 11477  * Stat: Creating Macros for Variable Selection;
44 11478  *-------------------------------------------------
   --------*;
45 11479  *-------------------------------------------------
   --------* ;
46 11480  * Stat: Binary and Nominal Input Variables Macro ;
47 11481  *-------------------------------------------------
   --------* ;
48 11482  %macro NOMINPUTS;
49 11483      MembershipLevel PaymentMethod
50 11484  %mend NOMINPUTS;
51 11485  *-------------------------------------------------
   --------*;
52 11486  * Stat: Creating Macros for Variable Selection;
53 11487  *-------------------------------------------------
```

```
--------*;
54 11488  *-------------------------------------------------
--------* ;
55 11489  * Stat: Ordinal Input Variables Macro ;
56 11490  *-------------------------------------------------
--------* ;
57 11491  %macro ORDINPUTS;
58 11492
59 11493  %mend ORDINPUTS;
60 11494  data work.EM_Stat_tree / view=work.EM_Stat_tree;
61 11495  set EMWS3.Ids_DATA(obs=100000 keep=%INTINPUTS %ORDINP
   UTS %NOMINPUTS Churn);
62 11496  run;
63 11497  proc arbor data=EMWS3.Ids_DATA(obs=100000 keep=%INTIN
   PUTS %ORDINPUTS %NOMINPUTS Churn) Criterion=GINI Leafsize=5
   Mincatsize = 5 Maxbranch=5 Maxdepth=1 Padjust= NONE NORULELI
   MIT MAXRULES=5 MAXSURRS=0 Missing=USEINSEARCH Exhaustive=500
   0;
64 WARNING: PADJUST and PVAR options are ignored with this spli
   tting criterion.
65 11498  input %INTINPUTS / level = interval;
66 11499  input %NOMINPUTS / level=nominal;
67 11500  target Churn / level=BINARY;
68 11501  Performance Disk NodeSize=10000;
69 11502  Assess NoValidata measure=MISC;
70 11503  SUBTREE BEST;
71 11504  save RULES=WORK.Stat_RULE;
72 11505  run;
73 11506  quit;
74 11507  data WORK.Stat_RULE(keep=Target Name Rank Numeric_Val
   ue StatVar rename=(numeric_value=Worth));
75 11508  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 11509     "%sysfunc(sasmsg(sashelp.dmine, rpt_worth_vlabel,
          NOQUOTE))" StatVar = "%sysfunc(sasmsg(sashelp.dmine, r
    pt_analysisVar_vlabel, NOQUOTE))";
 77 11510  length Target $32 Name $32;
 78 11511  retain TARGET "Churn" Name;
 79 11512  format STATVAR 6.0;
 80 11513  set WORK.Stat_RULE;
 81 11514  where stat in('VARIABLE','WORTH');
 82 11515  if stat = 'VARIABLE' then Name = character_value;
 83 11516  else do;
 84 11517  if _N_<= 2*1000 then STATVAR=1;
 85 11518  else STATVAR=0;
 86 11519  output;
 87 11520  end;
 88 11521  run;
 89 11522  proc append base=EMWS3.Stat_WORTH data=WORK.Stat_RULE
     force;
 90 11523  run;
 91 11524  *-------------------------------------------------
    --------*;
 92 11525  * Stat: Counting Levels;
 93 11526  *-------------------------------------------------
    --------*;
 94 11527  proc sql;
 95 11528  create view WORK.Stat_distinct as select distinct Chu
    rn from EMWS3.Ids_DATA(obs=100000);
 96 11529  quit;
 97 11530  proc sql;
 98 11531  reset noprint;
 99 11532  select count(*) into :_tmpcount from WORK.Stat_distin
    ct;
100 11533  quit;
101 11534  proc sort data=WORK.Stat_count NOTHREADS;
102 11535  by NAME ROLE DATA;
103 11536  run;
104 11537  proc print data=WORK.Stat_count(obs=500) label noobs;
```

```
105 11538  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 11539  var NAME ROLE COUNT;
107 11540  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_varLevels_
    title, NOQUOTE))";
108 11541  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title, NOQUOTE))";
109 11542  run;
110 11543  title9;
111 11544  title10;
112 11545  proc sort data=EMWS3.Stat_WORTH nodupkey out=WORK.Sta
    t_SELVAR(keep=Name StatVar);
113 11546  by NAME;
114 11547  where statvar=1;
115 11548  run;
116 11549  proc sort data=EMWS3.Stat_WORTH;
117 11550  by Rank;
118 11551  where statvar=1;
119 11552  run;
120 11553  proc sort data=WORK.M1RNLGO9;
121 11554  by name;
122 11555  run;
123 11556  data WORK.Stat_analysisMeta;
124 11557  merge WORK.M1RNLGO9 WORK.Stat_SELVAR(in=_a);
125 11558  by name;
126 11559  if _a then STATVAR = 1;
127 11560  else if ROLE in('INPUT', 'REJECTED') and REPORT ne 'Y
    ' then delete;
128 11561  if REPORT eq 'Y' then STATVAR = 1;
129 11562  run;
130 11563  data WORK.M1M0O2R1;
131 11564  set WORK.M1M0O2R1;
132 11565  where(use = 'Y' or Report='Y' or (role in('TARGET' 'F
    REQ' 'PREDICT' 'RESIDUAL' 'INPUT') and use = 'D'));
```

```
133 11566  if ROLE = 'SEGMENT' then delete;
134 11567  else if ROLE ^in('FREQ','TARGET') then role = 'INPUT'
    ;
135 11568  run;
136 11569  data WORK.M338F2ZO;
137 11570  set WORK.M338F2ZO;
138 11571  where(use = 'Y' or Report='Y' or (role in('SEGMENT',
    'TARGET' 'FREQ' 'PREDICT' 'RESIDUAL' 'INPUT') and use = 'D')
    );
139 11572  if ROLE ne 'FREQ' then role = 'INPUT';
140 11573  run;
141 11574  *-------------------------------------------------
    --------*;
142 11575  * Stat: Computing Statistics for Interval Variables;
143 11576  *-------------------------------------------------
    --------*;
144 11577  proc dmdb data=EMWS3.Ids_DATA(obs=100000) nonorm maxl
    evel=513
145 11578  varout=work._DMDBVAR(RENAME=(NAME=VARIABLE))
146 11579  classout=WORK.StatCLASS(drop=NMISSPERCENT rename=(NAM
    E=VARIABLE FREQUENCY=COUNT FREQPERCENT=PERCENT))
147 11580  ;
148 11581  var
149 11582  Age TotalPurchases TotalSpent
150 11583  ;
151 11584  class
152 11585  Churn MembershipLevel PaymentMethod
153 11586  ;
154 11587  ;
155 11588  run;
156 11589  proc stdize data = EMWS3.Ids_DATA(obs=100000) out=_nu
    ll_ outstat=work._STDIZE add=0 fuzz=1E-14 initial=MAD vardef
    =df method=STD mult=1 pctlmtd=ORD_STAT pctldef=2 pctlpts=(0
    5 50 95 100);
157 11590  var
158 11591  Age TotalPurchases TotalSpent
```

```
159 11592  ;
160 11593  run;
161 11594  proc transpose data=work._STDIZE out=work._TRANSSTDIZ
    E(drop=_LABEL_ rename=(_NAME_=VARIABLE));
162 11595  id _type_;
163 11596  where _TYPE_='P50';
164 11597  run;
165 WARNING: The variable _LABEL_ in the DROP, KEEP, or RENAME l
    ist has never been referenced.
166 11598  proc sort data=work._TRANSSTDIZE;
167 11599  by VARIABLE;
168 11600  run;
169 11601  proc sort data=work._DMDBVAR;
170 11602  by VARIABLE;
171 11603  run;
172 11604  data WORK.StatINTERVAL;
173 11605  merge work._TRANSSTDIZE work._DMDBVAR;
174 11606  format N NMISS 8. MIN MAX P50 MEAN STD SKEWNESS KURTO
    SIS BEST8.3;
175 11607  by VARIABLE;
176 11608  run;
177 WARNING: Multiple lengths were specified for the BY variable
     VARIABLE by input data sets. This might cause unexpected re
    sults.
178 11609  proc sort data=WORK.StatCLASS;
179 11610  by Variable Level;
180 11611  run;
181 11612  data colorindex;
182 11613  retain LevelIndex 0;
183 11614  set WORK.StatCLASS(keep=Variable Level);
184 11615  by variable level;
185 11616  if first.variable then LevelIndex = 0;
186 11617  if first.level then LevelIndex + 1;
187 11618  run;
188 11619  data WORK.StatCLASS;
189 11620  merge WORK.StatCLASS(in=_a) colorindex;
```

```
190  11621  by variable level;
191  11622  if _a then output;
192  11623  run;
193  11624  data WORK.StatINTERVAL;
194  11625  length DATAROLE $20;
195  11626  retain DATAROLE "TRAIN";
196  11627  length Target $32 TargetValue $32;
197  11628  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
     rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
     lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
198  11629  retain Target '_OVERALL_';
199  11630  set WORK.StatINTERVAL;
200  11631  run;
201  11632  proc append base=EMWS3.Stat_INTERVAL data=WORK.StatIN
     TERVAL force;
202  11633  run;
203  11634  data WORK.StatCLASS;
204  11635  length DATAROLE $20;
205  11636  retain DATAROLE "TRAIN";
206  11637  length Target $32 TargetValue $32;
207  11638  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
     rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
     lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
208  11639  retain Target '_OVERALL_';
209  11640  set WORK.StatCLASS;
210  11641  run;
211  11642  proc append base=EMWS3.Stat_CLASS data=WORK.StatCLASS
      force;
212  11643  run;
213  11644  data WORK.M1OK8TNV;
214  11645  set WORK.M1OK8TNV;
215  11646  where(use = 'Y' or Report='Y' or (role in('TARGET' 'F
     REQ' 'PREDICT' 'RESIDUAL' 'INPUT') and use = 'D'));
216  11647  if ROLE = 'SEGMENT' or (ROLE = 'TARGET' and LEVEL ne
     'INTERVAL') then delete;
217  11648  else if ROLE ne 'FREQ' then role = 'INPUT';
```

```
218 11649  run;
219 11650  proc freq data=EMWS3.Ids_DATA noprint;
220 11651  table
221 11652  Churn / out=EMWS3.Stat_BYVAR(drop=COUNT PERCENT) miss
    ing;
222 11653  run;
223 11654  data WORK.STATEXPLORE_DATA / view=WORK.STATEXPLORE_DA
    TA;
224 11655  set EMWS3.Ids_DATA;
225 11656  where left(trim(put(Churn,BEST12.))) =".";
226 11657  run;
227 11658  *-------------------------------------------------
    --------*;
228 11659  * Stat: Computing Statistics for Interval Variables;
229 11660  *-------------------------------------------------
    --------*;
230 11661  proc dmdb data=WORK.STATEXPLORE_DATA(obs=100000) nono
    rm maxlevel=513
231 11662  varout=work._DMDBVAR(RENAME=(NAME=VARIABLE))
232 11663  classout=WORK.StatCLASS(drop=NMISSPERCENT rename=(NAM
    E=VARIABLE FREQUENCY=COUNT FREQPERCENT=PERCENT))
233 11664  ;
234 11665  var
235 11666  Age TotalPurchases TotalSpent
236 11667  ;
237 11668  class
238 11669  MembershipLevel PaymentMethod
239 11670  ;
240 11671  ;
241 11672  run;
242 11673  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);
243 11674  var
244 11675  Age TotalPurchases TotalSpent
```

```
245 11676  ;
246 11677  run;
247 11678  proc transpose data=work._STDIZE out=work._TRANSSTDIZ
    E(drop=_LABEL_ rename=(_NAME_=VARIABLE));
248 11679  id _type_;
249 11680  where _TYPE_='P50';
250 11681  run;
251 WARNING: The variable _LABEL_ in the DROP, KEEP, or RENAME l
    ist has never been referenced.
252 11682  proc sort data=work._TRANSSTDIZE;
253 11683  by VARIABLE;
254 11684  run;
255 11685  proc sort data=work._DMDBVAR;
256 11686  by VARIABLE;
257 11687  run;
258 11688  data WORK.StatINTERVAL;
259 11689  merge work._TRANSSTDIZE work._DMDBVAR;
260 11690  format N NMISS 8. MIN MAX P50 MEAN STD SKEWNESS KURTO
    SIS BEST8.3;
261 11691  by VARIABLE;
262 11692  run;
263 WARNING: Multiple lengths were specified for the BY variable
     VARIABLE by input data sets. This might cause unexpected re
    sults.
264 11693  proc sort data=WORK.StatCLASS;
265 11694  by Variable Level;
266 11695  run;
267 11696  data WORK.StatCLASS;
268 11697  merge WORK.StatCLASS(in=_a) colorindex;
269 11698  by variable level;
270 11699  if _a then output;
271 11700  run;
272 WARNING: Multiple lengths were specified for the BY variable
     LEVEL by input data sets. This might cause unexpected resul
    ts.
273 11701  data WORK.StatINTERVAL;
```

```
274 11702  length DATAROLE $20;
275 11703  retain DATAROLE "TRAIN";
276 11704  length Target $32 TargetValue $32;
277 11705  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
278 11706  Target = "Churn";
279 11707  TargetValue = '.';
280 11708  set WORK.StatINTERVAL;
281 11709  run;
282 11710  proc append base=EMWS3.Stat_INTERVAL data=WORK.StatIN
    TERVAL force;
283 11711  run;
284 11712  data WORK.StatCLASS;
285 11713  length DATAROLE $20;
286 11714  retain DATAROLE "TRAIN";
287 11715  length Target $32 TargetValue $32;
288 11716  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
289 11717  Target = "Churn";
290 11718  TargetValue = '.';
291 11719  set WORK.StatCLASS;
292 11720  run;
293 11721  proc append base=EMWS3.Stat_CLASS data=WORK.StatCLASS
     force;
294 11722  run;
295 WARNING: Variable LEVEL has different lengths on BASE and DA
    TA files (BASE 12 DATA 11).
296 11723  data WORK.STATEXPLORE_DATA / view=WORK.STATEXPLORE_DA
    TA;
297 11724  set EMWS3.Ids_DATA;
298 11725  where left(trim(put(Churn,BEST12.))) ="0";
299 11726  run;
300 11727  *-------------------------------------------------
    --------*;
```

```
301 11728  * Stat: Computing Statistics for Interval Variables;
302 11729  *-------------------------------------------------
    --------*;
303 11730  proc dmdb data=WORK.STATEXPLORE_DATA(obs=100000) nono
    rm maxlevel=513
304 11731  varout=work._DMDBVAR(RENAME=(NAME=VARIABLE))
305 11732  classout=WORK.StatCLASS(drop=NMISSPERCENT rename=(NAM
    E=VARIABLE FREQUENCY=COUNT FREQPERCENT=PERCENT))
306 11733  ;
307 11734  var
308 11735  Age TotalPurchases TotalSpent
309 11736  ;
310 11737  class
311 11738  MembershipLevel PaymentMethod
312 11739  ;
313 11740  ;
314 11741  run;
315 11742  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);
316 11743  var
317 11744  Age TotalPurchases TotalSpent
318 11745  ;
319 11746  run;
320 11747  proc transpose data=work._STDIZE out=work._TRANSSTDIZ
    E(drop=_LABEL_ rename=(_NAME_=VARIABLE));
321 11748  id _type_;
322 11749  where _TYPE_='P50';
323 11750  run;
324 WARNING: The variable _LABEL_ in the DROP, KEEP, or RENAME l
    ist has never been referenced.
325 11751  proc sort data=work._TRANSSTDIZE;
326 11752  by VARIABLE;
327 11753  run;
328 11754  proc sort data=work._DMDBVAR;
```

```
329 11755  by VARIABLE;
330 11756  run;
331 11757  data WORK.StatINTERVAL;
332 11758  merge work._TRANSSTDIZE work._DMDBVAR;
333 11759  format N NMISS 8. MIN MAX P50 MEAN STD SKEWNESS KURTO
    SIS BEST8.3;
334 11760  by VARIABLE;
335 11761  run;
336 WARNING: Multiple lengths were specified for the BY variable
     VARIABLE by input data sets. This might cause unexpected re
    sults.
337 11762  proc sort data=WORK.StatCLASS;
338 11763  by Variable Level;
339 11764  run;
340 11765  data WORK.StatCLASS;
341 11766  merge WORK.StatCLASS(in=_a) colorindex;
342 11767  by variable level;
343 11768  if _a then output;
344 11769  run;
345 WARNING: Multiple lengths were specified for the BY variable
     LEVEL by input data sets. This might cause unexpected resul
    ts.
346 11770  data WORK.StatINTERVAL;
347 11771  length DATAROLE $20;
348 11772  retain DATAROLE "TRAIN";
349 11773  length Target $32 TargetValue $32;
350 11774  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
351 11775  Target = "Churn";
352 11776  TargetValue = '0';
353 11777  set WORK.StatINTERVAL;
354 11778  run;
355 11779  proc append base=EMWS3.Stat_INTERVAL data=WORK.StatIN
    TERVAL force;
356 11780  run;
```

```
357 11781  data WORK.StatCLASS;
358 11782  length DATAROLE $20;
359 11783  retain DATAROLE "TRAIN";
360 11784  length Target $32 TargetValue $32;
361 11785  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
    lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
362 11786  Target = "Churn";
363 11787  TargetValue = '0';
364 11788  set WORK.StatCLASS;
365 11789  run;
366 11790  proc append base=EMWS3.Stat_CLASS data=WORK.StatCLASS
     force;
367 11791  run;
368 WARNING: Variable LEVEL has different lengths on BASE and DA
    TA files (BASE 12 DATA 11).
369 11792  data WORK.STATEXPLORE_DATA / view=WORK.STATEXPLORE_DA
    TA;
370 11793  set EMWS3.Ids_DATA;
371 11794  where left(trim(put(Churn,BEST12.))) ="1";
372 11795  run;
373 11796  *-------------------------------------------------
    --------*;
374 11797  * Stat: Computing Statistics for Interval Variables;
375 11798  *-------------------------------------------------
    --------*;
376 11799  proc dmdb data=WORK.STATEXPLORE_DATA(obs=100000) nono
    rm maxlevel=513
377 11800  varout=work._DMDBVAR(RENAME=(NAME=VARIABLE))
378 11801  classout=WORK.StatCLASS(drop=NMISSPERCENT rename=(NAM
    E=VARIABLE FREQUENCY=COUNT FREQPERCENT=PERCENT))
379 11802  ;
380 11803  var
381 11804  Age TotalPurchases TotalSpent
382 11805  ;
383 11806  class
```

```
384  11807  MembershipLevel PaymentMethod
385  11808  ;
386  11809  ;
387  11810  run;
388  11811  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);
389  11812  var
390  11813  Age TotalPurchases TotalSpent
391  11814  ;
392  11815  run;
393  11816  proc transpose data=work._STDIZE out=work._TRANSSTDIZ
     E(drop=_LABEL_ rename=(_NAME_=VARIABLE));
394  11817  id _type_;
395  11818  where _TYPE_='P50';
396  11819  run;
397  WARNING: The variable _LABEL_ in the DROP, KEEP, or RENAME l
     ist has never been referenced.
398  11820  proc sort data=work._TRANSSTDIZE;
399  11821  by VARIABLE;
400  11822  run;
401  11823  proc sort data=work._DMDBVAR;
402  11824  by VARIABLE;
403  11825  run;
404  11826  data WORK.StatINTERVAL;
405  11827  merge work._TRANSSTDIZE work._DMDBVAR;
406  11828  format N NMISS 8. MIN MAX P50 MEAN STD SKEWNESS KURTO
     SIS BEST8.3;
407  11829  by VARIABLE;
408  11830  run;
409  WARNING: Multiple lengths were specified for the BY variable
      VARIABLE by input data sets. This might cause unexpected re
     sults.
410  11831  proc sort data=WORK.StatCLASS;
411  11832  by Variable Level;
```

```
412  11833  run;
413  11834  data WORK.StatCLASS;
414  11835  merge WORK.StatCLASS(in=_a) colorindex;
415  11836  by variable level;
416  11837  if _a then output;
417  11838  run;
418  WARNING: Multiple lengths were specified for the BY variable
      LEVEL by input data sets. This might cause unexpected resul
     ts.
419  11839  data WORK.StatINTERVAL;
420  11840  length DATAROLE $20;
421  11841  retain DATAROLE "TRAIN";
422  11842  length Target $32 TargetValue $32;
423  11843  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
     rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
     lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
424  11844  Target = "Churn";
425  11845  TargetValue = '1';
426  11846  set WORK.StatINTERVAL;
427  11847  run;
428  11848  proc append base=EMWS3.Stat_INTERVAL data=WORK.StatIN
     TERVAL force;
429  11849  run;
430  11850  data WORK.StatCLASS;
431  11851  length DATAROLE $20;
432  11852  retain DATAROLE "TRAIN";
433  11853  length Target $32 TargetValue $32;
434  11854  label Target = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
     rget_vlabel, NOQUOTE))" TargetValue = "%sysfunc(sasmsg(sashe
     lp.dmine, rpt_targetlevel_vlabel, NOQUOTE))";
435  11855  Target = "Churn";
436  11856  TargetValue = '1';
437  11857  set WORK.StatCLASS;
438  11858  run;
439  11859  proc append base=EMWS3.Stat_CLASS data=WORK.StatCLASS
      force;
```

```
440 11860  run;
441 WARNING: Variable LEVEL has different lengths on BASE and DA
    TA files (BASE 12 DATA 11).
442 11861  proc sort data=EMWS3.Stat_VariableSet out=tempVariabl
    e(keep=name label role rename=(name=variable));
443 11862  by name;
444 11863  run;
445 11864  proc sort data=EMWS3.Stat_CLASS;
446 11865  by Variable;
447 11866  run;
448 11867  data EMWS3.Stat_CLASS;
449 11868  merge EMWS3.Stat_CLASS(in=_a) tempVariable;
450 11869  by variable;
451 11870  if label = '' then label=Variable;
452 11871  if _a then output;
453 11872  run;
454 WARNING: Multiple lengths were specified for the BY variable
     VARIABLE by input data sets. This might cause unexpected re
    sults.
455 11873  proc sort data=EMWS3.Stat_INTERVAL NOTHREADS;
456 11874  by DATAROLE
457 11875  Target TargetValue
458 11876  ;
459 11877  run;
460 11878  run;
461 11879  proc sort data=EMWS3.Stat_VariableSet out=tempVariabl
    e(keep=name label role rename=(name=variable));
462 11880  by name;
463 11881  run;
464 11882  proc sort data=EMWS3.Stat_INTERVAL;
465 11883  by Variable;
466 11884  run;
467 11885  data EMWS3.Stat_INTERVAL;
468 11886  merge EMWS3.Stat_INTERVAL(in=_a) tempVariable;
469 11887  by variable;
470 11888  if label = '' then label=Variable;
```

```
471  11889  if _a then output;
472  11890  run;
473  WARNING: Multiple lengths were specified for the BY variable
      VARIABLE by input data sets. This might cause unexpected re
     sults.
474  11891  proc datasets library=EMWS3 nolist;
475  11892  modify Stat_INTERVAL;
476  11893  label DATAROLE = "%sysfunc(sasmsg(sashelp.dmine, rpt_
     datarole_vlabel, NOQUOTE))";
477  11894  label Variable = "%sysfunc(sasmsg(sashelp.dmine, rpt_
     variable_vlabel, NOQUOTE))";
478  11895  label MEAN = "%sysfunc(sasmsg(sashelp.dmine, rpt_mean
     _vlabel,    NOQUOTE))";
479  11896  label N = "%sysfunc(sasmsg(sashelp.dmine, rpt_nonMiss
     _vlabel, NOQUOTE))";
480  11897  label STD = "%sysfunc(sasmsg(sashelp.dmine, rpt_std_v
     label,     NOQUOTE))";
481  11898  label NMISS = "%sysfunc(sasmsg(sashelp.dmine, rpt_mis
     sing_vlabel, NOQUOTE))";
482  11899  label p50 = "%sysfunc(sasmsg(sashelp.dmine, rpt_media
     n_vlabel, NOQUOTE))";
483  11900  label MIN = "%sysfunc(sasmsg(sashelp.dmine, rpt_minim
     um_vlabel, NOQUOTE))";
484  11901  label MAX = "%sysfunc(sasmsg(sashelp.dmine, rpt_maxim
     um_vlabel, NOQUOTE))";
485  11902  label SKEWNESS = "%sysfunc(sasmsg(sashelp.dmine, rpt_
     skewness_vlabel, NOQUOTE))";
486  11903  label KURTOSIS = "%sysfunc(sasmsg(sashelp.dmine, rpt_
     kurtosis_vlabel, NOQUOTE))";
487  11904  label TARGET = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
     rget_vlabel,   NOQUOTE))" TARGETVALUE = "%sysfunc(sasmsg(sas
     help.dmine, rpt_targetLevel_vlabel,   NOQUOTE))";
488  11905  run;
489  11906  proc sort data=EMWS3.Stat_CLASS NOTHREADS;
490  11907  by DATAROLE ROLE Variable descending COUNT;
491  11908  run;
```

```
492 11909  data WORK.StatCLASS(keep=DATAROLE Variable ROLE Numca
    t Nmiss Mode ModePct Mode2 Mode2Pct);
493 11910  set EMWS3.Stat_CLASS;
494 11911  length Mode Mode2 $32;
495 11912  retain NMiss 0 ModePct 0 Mode2Pct 0 Mode '' Mode2 '';
496 11913  by DATAROLE ROLE Variable;
497 11914  if first.DATAROLE or first.variable then do;
498 11915  Numcat = 1;
499 11916  NMiss = 0;
500 11917  Mode = strip(LEVEL);
501 11918  ModePct = PERCENT;
502 11919  Mode2 = '';
503 11920  Mode2Pct = 0;
504 11921  end;
505 11922  else Numcat +1;
506 11923  if Numcat =2 then do;
507 11924  Mode2 = strip(LEVEL);
508 11925  Mode2Pct = PERCENT;
509 11926  end;
510 11927  if NRAW=. and CRAW='' then NMiss= COUNT;
511 11928  if last.DATAROLE or last.variable then output;
512 11929  where TARGET='_OVERALL_';
513 11930  run;
514 11931  %let _cn = %sysfunc(getoption(CENTER));
515 11932  %let _nb = %sysfunc(getoption(NUMBER));
516 11933  options nonumber nocenter;
517 11934  title;
518 11935  title9 ' ';
519 11936  proc print data=WORK.StatCLASS(obs=500) label noobs;
520 11937  var DATAROLE VARIABLE ROLE NUMCAT NMISS Mode ModePct
    Mode2 Mode2Pct;
521 11938  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 =
522 11939     "%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 =
523 11940     "%sysfunc(sasmsg(sashelp.dmine, rpt_mode2_vlabel,
        NOQUOTE))" MODE2PCT = "%sysfunc(sasmsg(sashelp.dmine, rp
    t_mode2pct_vlabel, NOQUOTE))";
524 11941  format ModePct Mode2Pct 5.2;
525 11942  by DATAROLE;
526 11943  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_classStats
    _title,     NOQUOTE))";
527 11944  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title,  NOQUOTE))";
528 11945  run;
529 11946  proc print data=EMWS3.Stat_CLASS(obs=500) noobs label
    ;
530 11947  var DATAROLE VARIABLE ROLE LEVEL COUNT PERCENT;
531 11948  where ROLE in('TARGET', 'SEGMENT');
532 11949  by DATAROLE;
533 11950  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 =
534 11951     "%sysfunc(sasmsg(sashelp.dmine, rpt_count_vlabel,
            NOQUOTE))" PERCENT = "%sysfunc(sasmsg(sashelp.dmine
    , rpt_percent_vlabel,        NOQUOTE))";
535 11952  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_classTarge
    tSegmentDist_title, NOQUOTE))";
536 11953  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title,          NOQUOTE))";
537 11954  run;
538 11955  title9;
539 11956  title10;
540 11957  options &_cn &_nb;
541 11958  data WORK.StatINTERVALSUMMARY;
542 11959  set EMWS3.Stat_INTERVAL;
```

```
543 11960  length ROLE $32;
544 11961  where TARGET='_OVERALL_';
545 11962  if Variable = "Age" then ROLE="INPUT";
546 11963  else
547 11964  if Variable = "TotalPurchases" then ROLE="INPUT";
548 11965  else
549 11966  if Variable = "TotalSpent" then ROLE="INPUT";
550 11967  run;
551 11968  proc sort data=WORK.StatINTERVALSUMMARY NOTHREADS;
552 11969  by DATAROLE ROLE Variable;
553 11970  run;
554 11971  %let _cn = %sysfunc(getoption(CENTER));
555 11972  %let _nb = %sysfunc(getoption(NUMBER));
556 11973  options nonumber nocenter;
557 11974  title;
558 11975  proc print data=WORK.StatINTERVALSUMMARY(obs=500) lab
    el noobs;
559 11976  var Variable ROLE MEAN STD N NMISS MIN P50 MAX SKEWNE
    SS KURTOSIS;
560 11977  by DATAROLE;
561 11978  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_intervalSt
    ats_title, NOQUOTE))";
562 11979  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title, NOQUOTE))";
563 11980  run;
564 11981  title9;
565 11982  title10;
566 11983  options &_cn &_nb;
567 11984  proc sort data=EMWS3.Stat_CLASS NOTHREADS;
568 11985  by DATAROLE DATAROLE Target TargetValue Variable desc
    ending COUNT;
569 11986  run;
570 11987  data WORK.StatCLASS(keep=DATAROLE Target TargetValue
    DATAROLE Variable Numcat NMiss Mode ModePct Mode2 Mode2Pct )
    ;
571 11988  set EMWS3.Stat_CLASS;
```

```
572  11989  length Mode Mode2 $32;
573  11990  retain NMiss 0 ModePct 0 Mode2Pct 0 Mode '' Mode2 '';
574  11991  by DATAROLE DATAROLE Target TargetValue Variable;
575  11992  if first.datarole or first.variable then do;
576  11993  Numcat = 1;
577  11994  NMiss = 0;
578  11995  Mode = LEVEL;
579  11996  ModePct = PERCENT;
580  11997  Mode2 = '';
581  11998  Mode2Pct = 0;
582  11999  end;
583  12000  else Numcat +1;
584  12001  if Numcat =2 then do;
585  12002  Mode2 = LEVEL;
586  12003  Mode2Pct = PERCENT;
587  12004  end;
588  12005  if NRAW=. and CRAW='' then NMiss= COUNT;
589  12006  if last.datarole or last.variable then output;
590  12007  where ROLE ^in('TARGET', 'SEGMENT');
591  12008  run;
592  12009  proc sort data=WORK.StatCLASS NOTHREADS;
593  12010  by DATAROLE
594  12011  VARIABLE Target TargetValue;
595  12012  run;
596  12013  %let _cn = %sysfunc(getoption(CENTER));
597  12014  %let _nb = %sysfunc(getoption(NUMBER));
598  12015  options nonumber nocenter;
599  12016  title;
600  12017  title9 ' ';
601  12018  proc print data=WORK.StatCLASS(obs=500) label noobs;
602  12019  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 =
603  12020     "%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 =
604 12021     "%sysfunc(sasmsg(sashelp.dmine, rpt_mode2pct_vlabe
    l, NOQUOTE))";
605 12022  format ModePct Mode2Pct 5.2;
606 12023  by DATAROLE;
607 12024  label TARGET = "%sysfunc(sasmsg(sashelp.dmine, rpt_ta
    rget_vlabel,   NOQUOTE))" TARGETVALUE = "%sysfunc(sasmsg(sas
    help.dmine, rpt_targetLevel_vlabel,   NOQUOTE))";
608 12025  var Target TargetValue NumCat NMiss Mode ModePct Mode
    2 Mode2Pct;
609 12026  by DATAROLE VARIABLE;
610 12027  title9 "Class Variable Summary Statistics by Class Ta
    rget";
611 12028  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title, NOQUOTE))";
612 12029  run;
613 12030  title9;
614 12031  title10;
615 12032  options &_cn &_nb;
616 12033  proc sort data=EMWS3.Stat_INTERVAL NOTHREADS;
617 12034  by DATAROLE Variable
618 12035  Target
619 12036  TargetValue
620 12037  ;
621 12038  run;
622 12039  %let _cn = %sysfunc(getoption(CENTER));
623 12040  %let _nb = %sysfunc(getoption(NUMBER));
624 12041  options nonumber nocenter;
625 12042  title;
626 12043  proc print data=EMWS3.Stat_INTERVAL(obs=500) label no
    obs;
627 12044  by DATAROLE Variable
628 12045  ;
629 12046  title9 "Interval Variable Summary Statistics by Class
     Target";
```

```
630 12047  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title, NOQUOTE))";
631 12048  run;
632 12049  title9;
633 12050  title10;
634 12051  options &_cn &_nb;
635 12052  *-------------------------------------------------
    --------*;
636 12053  * Stat: Computing Chi-Square Statistics;
637 12054  *-------------------------------------------------
    --------*;
638 12055  ods listing close;
639 12056  ods output chisq=WORK.TEMPCHISQUARE;
640 12057  ods output crosstabFreqs=WORK.TEMPCROSSTAB;
641 12058  proc freq data=EMWS3.Ids_DATA(obs=100000) order=forma
    tted addnames;
642 12059  table Churn*MembershipLevel /missing chisq outexpect
    sparse;
643 12060  table Churn*PaymentMethod /missing chisq outexpect sp
    arse;
644 12061  ;
645 12062  run;
646 WARNING: Some statistics are not computed because there is n
    o valid numeric table score for the missing values.
647 WARNING: Some statistics are not computed because there is n
    o valid numeric table score for the missing values.
648 12063  quit;
649 12064  ods listing;
650 12065  data WORK.StatCHI2(keep=TARGET INPUT CRAMERV CHIMEASU
    RE PROBCHI CHIDF rename=(PROBCHI=PROB CHIDF=DF));
651 12066  length TARGET INPUT $32;
652 12067  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 =
```

```
653 12068     "%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 =
654 12069     "%sysfunc(sasmsg(sashelp.dmine, rpt_df_vlabel ,
           NOQUOTE))";
655 12070  retain CHIMEASURE CHIDF PROBCHI;
656 12071  format PROBCHI PVALUE6.4 CHIMEASURE 10.4;
657 12072  set WORK.TEMPCHISQUARE;
658 12073  Target = strip(RowVariable);
659 12074  Input = strip(ColVariable);
660 12075  where StatisticId in('_CRAMV_','_PCHI_');
661 12076  if StatisticId = '_PCHI_' then do;
662 12077  PROBCHI = Prob;
663 12078  CHIMEASURE = VALUE;
664 12079  CHIDF = DF;
665 12080  end;
666 12081  else do;
667 12082  CRAMERV = abs(VALUE);
668 12083  OUTPUT;
669 12084  end;
670 12085  run;
671 12086  proc sort data=WORK.TEMPCROSSTAB;
672 12087  by RowVariable ColVariable _Type_;
673 12088  where _Type_ in('11','00');
674 12089  run;
675 12090  data WORK.StatCHI(keep=YVAR XVAR X Y XTEXT YTEXT COUN
    T CHISQUARE LOGCHISQUARE);
676 12091  retain N;
677 12092  dsid = open("WORK.TEMPCROSSTAB");
678 12093  rowvarnum = varnum(dsid, 'RowVariable');
679 12094  colvarnum = varnum(dsid, 'ColVariable');
680 12095  freqnum = varnum(dsid, 'FREQUENCY');
681 12096  rownum = varnum(dsid, 'RowPercent');
682 12097  colnum = varnum(dsid, 'ColPercent');
```

```
683 12098  typenum = varnum(dsid, '_TYPE_');
684 12099  length YVAR XVAR $32 YTEXT $32 XTEXT $32;
685 12100  do while(^fetch(dsid));
686 12101  COUNT = getvarn(dsid, freqnum);
687 12102  YVAR = (getvarc(dsid, rowvarnum));
688 12103  XVAR = strip(getvarc(dsid, colvarnum));
689 12104  ynum = varnum(dsid, YVAR);
690 12105  ytype = vartype(dsid, ynum);
691 12106  yformat = varfmt(dsid, ynum);
692 12107  if yformat eq '' then yformat = 'best12.';
693 12108  Y=.;
694 12109  if ytype = 'N' then do;
695 12110  Y = getvarn(dsid, ynum);
696 12111  YTEXT = left(putn(Y, yformat));
697 12112  end;
698 12113  else YTEXT = getvarc(dsid, ynum);
699 12114  xnum = varnum(dsid, XVAR);
700 12115  xtype = vartype(dsid, xnum);
701 12116  xformat = varfmt(dsid, xnum);
702 12117  if xformat eq '' then xformat = 'best12.';
703 12118  X=.;
704 12119  if xtype = 'N' then do;
705 12120  X = getvarn(dsid, xnum);
706 12121  XTEXT = left(putn(X, xformat));
707 12122  end;
708 12123  else XTEXT = getvarc(dsid, xnum);
709 12124  if getvarc(dsid, typenum)= '00' then N= count;
710 12125  else do;
711 12126  EXP = (getvarn(dsid,rownum)/100)*(getvarn(dsid,colnum
    )/100)*N;
712 12127  if exp >0 then do;
713 12128  CHISQUARE=(EXP-count)**2/(EXP);
714 12129  if chisquare>0 then LOGCHISQUARE = log(CHISQUARE);
715 12130  OUTPUT;
716 12131  end;
717 12132  end;
```

```
718 12133  end;
719 12134  dsid = close(dsid);
720 12135  run;
721 12136  data WORK.StatCHI;
722 12137  length DATAROLE $20;
723 12138  retain DATAROLE "TRAIN";
724 12139  SegmentVar = "";
725 12140  SegmentValue = "";
726 12141  SegmentId = "_OVERALL_";
727 12142  set WORK.StatCHI;
728 12143  run;
729 12144  proc append base=EMWS3.Stat_CHI2 data=WORK.StatCHI fo
    rce;
730 12145  run;
731 12146  data WORK.StatCHI2;
732 12147  length DATAROLE $20;
733 12148  retain DATAROLE "TRAIN";
734 12149  SegmentVar = "";
735 12150  SegmentValue = "";
736 12151  SegmentId = "_OVERALL_";
737 12152  set WORK.StatCHI2;
738 12153  run;
739 12154  proc append base=EMWS3.Stat_CHIMEASURE data=WORK.Stat
    CHI2 force;
740 12155  run;
741 12156  proc sort data=EMWS3.Stat_VariableSet out=tempVariabl
    e(keep=name label role rename=(label=XLabel name=XVAR));
742 12157  by name;
743 12158  run;
744 12159  proc sort data=EMWS3.Stat_CHI2;
745 12160  by XVAR;
746 12161  run;
747 12162  data EMWS3.Stat_CHI2;
748 12163  merge EMWS3.Stat_CHI2(in=_a) tempVariable;
749 12164  by XVAR;
750 12165  if XLabel = '' then XLabel=XVAR;
```

```
751 12166  if _a then output;
752 12167  run;
753 WARNING: Multiple lengths were specified for the BY variable
     XVAR by input data sets. This might cause unexpected result
    s.
754 12168  proc datasets library=EMWS3 nolist;
755 12169  modify Stat_CHI2;
756 12170  label DATAROLE = "%sysfunc(sasmsg(sashelp.dmine,  rpt
    _datarole_vlabel, NOQUOTE))";
757 12171  label SEGMENTVAR = "%sysfunc(sasmsg(sashelp.dmine, rp
    t_segment_vlabel,      NOQUOTE))";
758 12172  label SEGMENTID = "%sysfunc(sasmsg(sashelp.dmine, rpt
    _segmentnamevalue_vlabel,      NOQUOTE))";
759 12173  label SegmentValue= "%sysfunc(sasmsg(sashelp.dmine, r
    pt_segmentid_vlabel, NOQUOTE))";
760 12174  label YVAR = "%sysfunc(sasmsg(sashelp.dmine, rpt_targ
    et_vlabel,         NOQUOTE))";
761 12175  label XVAR = "%sysfunc(sasmsg(sashelp.dmine, rpt_inpu
    t_vlabel,          NOQUOTE))";
762 12176  label X = "%sysfunc(sasmsg(sashelp.dmine, rpt_inputNu
    mValue_vlabel,  NOQUOTE))";
763 12177  label Y = "%sysfunc(sasmsg(sashelp.dmine, rpt_targetN
    umValue_vlabel, NOQUOTE))";
764 12178  label XTEXT = "%sysfunc(sasmsg(sashelp.dmine, rpt_inp
    utFmtValue_vlabel,  NOQUOTE))";
765 12179  label YTEXT = "%sysfunc(sasmsg(sashelp.dmine, rpt_tar
    getFmtValue_vlabel, NOQUOTE))";
766 12180  label COUNT = "%sysfunc(sasmsg(sashelp.dmine, rpt_cou
    nt_vlabel,          NOQUOTE))";
767 12181  label CHISQUARE = "%sysfunc(sasmsg(sashelp.dmine, rpt
    _chiSquare_vlabel,      NOQUOTE))";
768 12182  label LOGCHISQUARE = "%sysfunc(sasmsg(sashelp.dmine,
    rpt_logChiSquare_vlabel,   NOQUOTE))";
769 12183  run;
770 12184  proc sort data=EMWS3.Stat_VariableSet out=tempVariabl
    e(keep=name label role rename=(name=INPUT));
```

```
771  12185  by name;
772  12186  run;
773  12187  proc sort data=EMWS3.Stat_CHIMEASURE;
774  12188  by INPUT;
775  12189  run;
776  12190  data EMWS3.Stat_CHIMEASURE;
777  12191  merge EMWS3.Stat_CHIMEASURE(in=_a) tempVariable;
778  12192  by INPUT;
779  12193  if label = '' then label=INPUT;
780  12194  if _a then output;
781  12195  run;
782  WARNING: Multiple lengths were specified for the BY variable
      INPUT by input data sets. This might cause unexpected resul
     ts.
783  12196  proc sort data=EMWS3.Stat_CHIMEASURE NOTHREADS;
784  12197  by DATAROLE TARGET descending CRAMERV;
785  12198  run;
786  12199  data EMWS3.Stat_CHIMEASURE;
787  12200  set EMWS3.Stat_CHIMEASURE;
788  12201  label ORDEREDINPUTS = "%sysfunc(sasmsg(sashelp.dmine,
      rpt_orderedInput_vlabel  , NOQUOTE))" DATAROLE = "%sysfunc(
     sasmsg(sashelp.dmine,  rpt_datarole_vlabel, NOQUOTE))" SEGME
     NTID =
789  12202     "%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))";
790  12203  format ORDEREDINPUTS 5.0;
791  12204  retain ORDEREDINPUTS 0;
792  12205  if first.DATAROLE or first.SEGMENTID or first.TARGET
     then ORDEREDINPUTS=1;
793  12206  else ORDEREDINPUTS+1;
794  12207  by DATAROLE SEGMENTID TARGET;
795  12208  run;
796  12209  %let _cn = %sysfunc(getoption(CENTER));
```

```
797 12210  %let _nb = %sysfunc(getoption(NUMBER));
798 12211  options nonumber nocenter;
799 12212  title;
800 12213  proc print data=EMWS3.Stat_CHIMEASURE(obs=500) label
    noobs;
801 12214  var INPUT CHIMEASURE DF PROB;
802 12215  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 =
803 12216     "%sysfunc(sasmsg(sashelp.dmine, rpt_df_vlabel ,
          NOQUOTE))";
804 12217  by DATAROLE TARGET;
805 12218  title9 "%sysfunc(sasmsg(sashelp.dmine, rpt_chisquareS
    tats_title, NOQUOTE))";
806 12219  title10 "%sysfunc(sasmsg(sashelp.dmine, rpt_maxObsPri
    nted_title,  NOQUOTE))";
807 12220  run;
808 12221  title10;
809 12222  options &_cn &_nb;
810 12223  proc sort data=EMWS3.Stat_CHIMEASURE out=trainChimeas
    ure;
811 12224  by TARGET descending CramerV;
812 12225  where DATAROLE='TRAIN';
813 12226  run;
814 12227  data trainChimeasure;
815 12228  set trainchimeasure;
816 12229  by TARGET;
817 12230  retain groupcount 0;
818 12231  if first.Target then groupcount =1;
819 12232  else groupcount + 1;
820 12233  if groupcount<21 then Plot=1;
821 12234  else Plot = 0;
822 12235  keep Target Input GroupCount Plot;
823 12236  run;
```

```
824 12237  proc sort data=trainChiMeasure;
825 12238  by TARGET INPUT;
826 12239  run;
827 12240  proc sort data=EMWS3.Stat_CHIMEASURE;
828 12241  by TARGET INPUT;
829 12242  run;
830 12243  data EMWS3.Stat_CHIMEASURE;
831 12244  merge EMWS3.Stat_CHIMEASURE trainchimeasure;
832 12245  by TARGET INPUT;
833 12246  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))";
834 12247  run;
835 12248  proc sort data=EMWS3.Stat_CLASS(drop=CRAW NRAW) NOTHR
    EADS;
836 12249  by DATAROLE TARGET VARIABLE LEVEL;
837 12250  where TARGET ne '_OVERALL_';
838 12251  run;
839 12252  data WORK.StattempDs(keep=DATAROLE target VARIABLE no
    bs);
840 12253  retain nobs;
841 12254  set EMWS3.Stat_CLASS;
842 12255  by DATAROLE target VARIABLE;
843 12256  if first.DATAROLE or first.variable then nobs=count;
844 12257  else nobs = nobs + count;
845 12258  if last.DATAROLE or last.variable then output;
846 12259  run;
847 12260  data EMWS3.Stat_CLASS(drop=NOBS);
848 12261  merge EMWS3.Stat_CLASS(where=(TARGET ne '_OVERALL_')
    rename=(PERCENT=WITHINPCT)) WORK.StattempDs;
849 12262  label percent = "%sysfunc(sasmsg(sashelp.dmine, rpt_p
    ercent_vlabel,         NOQUOTE))" withinpct = "%sysfunc(sasm
    sg(sashelp.dmine, rpt_percentwithin_vlabel, NOQUOTE))";
850 12263  by DATAROLE TARGET VARIABLE;
851 12264  if NOBS>0 then PERCENT=COUNT/NOBS;
```

```
852 12265  run;
853 12266  proc sort data=EMWS3.Stat_CHIMEASURE out=WORK.Statcut
    off(RENAME=(INPUT=VARIABLE) keep=DATAROLE TARGET INPUT PLOT)
    ;
854 12267  by DATAROLE TARGET INPUT;
855 12268  run;
856 12269  data EMWS3.Stat_CLASS;
857 12270  drop groupcount;
858 12271  retain groupcount 0;
859 12272  merge EMWS3.Stat_CLASS(in=_a) WORK.Statcutoff;
860 12273  if _a then do;
861 12274  plot=1;
862 12275  groupcount+1;
863 12276  end;
864 12277  if groupcount>21 then plot=0;
865 12278  by DATAROLE Target Variable;
866 12279  if PERCENT eq . or plot = . then delete;
867 12280  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 =
868 12281     "%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))";
869 12282  run;
870 12283  proc sort data=EMWS3.Stat_CHIMEASURE;
871 12284  by DATAROLE ORDEREDINPUTS;
872 12285  run;
873 12286  data WORK.StatINTERVALPLOT(drop=TARGET N MIN NMISS P5
    0 MAX STD SKEWNESS KURTOSIS TARGETVALUE rename=(Mean=OMean))
    ;
874 12287  set EMWS3.Stat_INTERVAL;
875 12288  where TARGET='_OVERALL_';
876 12289  run;
```

```
877 12290  data EMWS3.Stat_INTERVAL(drop=OMean);
878 12291  merge EMWS3.Stat_INTERVAL(where=(TARGET ne '_OVERALL_
    ')) WORK.StatINTERVALPLOT;
879 12292  by DATAROLE Variable;
880 12293  if OMean ne 0 then ScaleDevMean = (Mean - OMean)/OMea
    n;
881 12294  run;
882 12295  proc sort data=EMWS3.Stat_INTERVAL NOTHREADS;
883 12296  by DATAROLE Variable Target;
884 12297  run;
885 12298  data WORK.StatMAXINTERVAL(keep=DATAROLE Target Variab
    le MaxDev);
886 12299  set EMWS3.Stat_INTERVAL;
887 12300  if first.DATAROLE or first.Variable then MaxDev = abs
    (ScaleDevMean);
888 12301  else if abs(ScaleDevMean)>MaxDev then MaxDev = abs(Sc
    aleDevMean);
889 12302  if last.datarole or last.variable or last.target then
     output;
890 12303  by DATAROLE Variable Target;
891 12304  run;
892 12305  data EMWS3.Stat_INTERVAL;
893 12306  merge EMWS3.Stat_INTERVAL WORK.StatMAXINTERVAL;
894 12307  label LevelId = "%sysfunc(sasmsg(sashelp.dmine, rpt_l
    evelId_vlabel,             NOQUOTE))" Variable = "%sysfunc(s
    asmsg(sashelp.dmine, rpt_variable_vlabel,            NOQUOTE
    ))" MaxDev =
895 12308      "%sysfunc(sasmsg(sashelp.dmine, rpt_maxdev_vlabel,
                  NOQUOTE))" ScaleDevMean = "%sysfunc(sasmsg(sas
    help.dmine, rpt_scaledMeanDeviation_vlabel, NOQUOTE))";
896 12309  format LevelId best3.;
897 12310  retain LevelId 0;
898 12311  by DATAROLE Variable Target;
899 12312  if first.DATAROLE or first.Target then LevelId=1;
900 12313  else LevelId + 1;
901 12314  run;
```

```
902 12315  proc sort data=EMWS3.Stat_INTERVAL NOTHREADS;
903 12316  by DATAROLE Target descending MaxDev Variable LevelId
    ;
904 12317  run;
905 12318  proc sort data=WORK.M1RNLGO9;
906 12319  by name;
907 12320  run;
908 12321  proc sort data=EMWS3.Stat_WORTH;
909 12322  by NAME;
910 12323  run;
911 12324  data EMWS3.Stat_WORTH;
912 12325  merge EMWS3.Stat_WORTH(in=_a) WORK.M1RNLGO9(Keep=NAME
     LABEL REPORT);
913 12326  by NAME;
914 12327  if REPORT = 'Y' then plot = 1;
915 12328  if LABEL = "" then LABEL = NAME;
916 12329  if _a then output;
917 12330  drop report;
918 12331  run;
919 WARNING: Multiple lengths were specified for the BY variable
     Name by input data sets. This might cause unexpected result
    s.
920 12332  proc sort data=EMWS3.Stat_WORTH;
921 12333  by rank;
922 12334  run;
923 *------------------------------------------------------------
    -*
924 * Score Log
925 Date:                 07 January 2024
926 Time:                 09:10:39
927 *------------------------------------------------------------
    -*
928 12436  *------------------------------------------------------
    --------*;
929 12437  * Stat: Scoring DATA data;
930 12438  *------------------------------------------------------
```

```
    --------*;
931 12439  data EMWS3.Stat_TRAIN
932 12440  / view=EMWS3.Stat_TRAIN
933 12441  ;
934 12442  set EMWS3.Ids_DATA
935 12443  ;
936 12444  run;
937 12445  quit;
938 12446  *-------------------------------------------------
    --------*;
939 12447  * Stat: Computing metadata for TRAIN data;
940 12448  *-------------------------------------------------
    --------*;
941 *-----------------------------------------------------
    -*
942 * Report Log
943 Date:                 07 January 2024
944 Time:                 09:10:39
945 *-----------------------------------------------------
    -*
```