Execution Environment

Author:	di00222@surrey.ac.uk
File:	File: /export/viya/homes/di00222@surrey.ac.uk/casuser/Cleaned_Road_Accident_2021_MergedwithDescription_Data_Balancing.sas
SAS Context:	SAS Studio compute context
SAS Version:	V.04.00M0P091624
SAS Client:	SAS® Studio 6.0
SAS Locale:	
Submission Time:	May 7, 2025, 2:18:35 PM
Time Zone:	GMT+01:00
User Agent:	Chrome 135.0.0.0

Code:Cleaned_Road_Accident_2021_MergedwithDescription_Data_Balancing.sas

```
%let NumSamples1 = 70;
%let NumSamples2 = 3;
%let NumSamples3 = 1;
/* Sort the dataset by acci_severity */
proc sort data=WORK.IMPORT;
by acci_severity;
run;
/* Use PROC SURVEYSELECT to make the dataset balanced */
proc surveyselect data=WORK.IMPORT NOPRINT out=BalancedData1
method=urs
seed=12345
samprate=(1 0 0); /* Adjust the samprate to balance the strata */
strata acci_severity;
run;
/* Create a variable with a constant value for each observation */
data BalancedData_with_constant1;
set BalancedData1;
constant = 1;
run;
/* Use PROC SURVEYSELECT to perform bootstrap sampling based on
acci_severity */
proc surveyselect data=BalancedData_with_constant1 NOPRINT seed=1
method=urs
samprate=1
OUTHITS
reps=&NumSamples1(repname=row)
out=BootSamp1;
strata acci_severity;
run;
/* Overall, how often was each observation selected? */
proc freq data=BootSamp1;
tables acci_severity;
run;
/* Use PROC SURVEYSELECT to make the dataset balanced */
proc surveyselect data=WORK.IMPORT NOPRINT out=BalancedData2
method=urs
seed=12345
samprate=(0 1 0); /* Adjust the samprate to balance the strata */
strata acci_severity;
run;
/* Create a variable with a constant value for each observation */
data BalancedData_with_constant2;
set BalancedData2;
constant = 1;
run;
/* Use PROC SURVEYSELECT to perform bootstrap sampling based on
acci_severity */
proc surveyselect data=BalancedData_with_constant2 NOPRINT seed=1
method=urs
samprate=1
OUTHITS
reps=&NumSamples2(repname=row)
out=BootSamp2;
strata acci_severity;
run;
/* Use PROC SURVEYSELECT to make the dataset balanced */
proc surveyselect data=WORK.IMPORT NOPRINT out=BalancedData3
method=urs
seed=12345
samprate=(0 0 0.9); /* Adjust the samprate to balance the strata */
strata acci_severity;
run;
/* Create a variable with a constant value for each observation */
data BalancedData_with_constant3;
set BalancedData3;
constant = 1;
run;
/* Use PROC SURVEYSELECT to perform bootstrap sampling based on
acci_severity */
proc surveyselect data=BalancedData_with_constant3 NOPRINT seed=1
method=urs
samprate=1
OUTHITS
reps=&NumSamples3(repname=row)
```

```
out=BootSamp3;
strata acci_severity;
run;
proc freq data=BootSamp3;
tables acci_severity;
run;
data BootSamp;
set BootSamp1 BootSamp2 BootSamp3;
run;
proc freq data=BootSamp;
tables acci_severity;
run;
```

Log:Cleaned_Road_Accident_2021_MergedwithDescription_Data_Balancing.sas

```
1    /* region: Generated preamble */
2    /* Make sure the current directory is writable */
3    data _null_;
4        length rc 4;
5        %let tworkloc="%sysfunc(getoption(work))";
6        rc=dlgcdir(&tworkloc);
7    run;
NOTE: The current working directory is now
      "/opt/sas/viya/config/var/tmp/compsrv/default/02b0b9be-accf-49c4-975b-2b75911438f5/SAS_workC95300000211_sas-compute-server-0d5
      a343e-620c-4cd1-8f0a-d5f23fc7bbc0-20030".
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

8
9    /* Setup options */
10   title;
11   footnote;
12   options validvarname=any;
13   options validmemname=extend;
14   options dtreset date number;
15   options device=png;
16
17   /* Setup macro variables */
18   %let syscc=0;
19   %let _clientapp = %nrquote(%nrstr(SAS Studio));
20   %let _clientappabbrev = %nrquote(%nrstr(Studio));
21   %let _clientappversion=2024.09;
22   %let _clientversion=;
23   %let _sasservername=&SYSHOSTNAME;
24   %let _sashostname=&SYSHOSTNAME;
25   %let _sasprogramfilehost=&SYSHOSTNAME;
26   %let _clientuserid = %nrquote(%nrstr(di00222@surrey.ac.uk));
27   %let _clientusername = %nrquote(%nrstr(di00222@surrey.ac.uk));
28   %let clientmachine = %nrquote(%nrstr());
29   %let _clientmachine = %nrquote(%nrstr());
30   %let _clientmode = %nrquote(%nrstr(viya));
31   %let sasworklocation="%sysfunc(getoption(work))/";
32   filename _cwd &sasworklocation;
33   data _null_;
34       call symput('_sasworkingdir',pathname('_cwd'));
35   run;
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

36   filename _cwd;
NOTE: Fileref _CWD has been deassigned.
37   %let _sasprogramfile = %nrquote(%nrstr());
38   %let _baseurl = %nrquote(%nrstr(https://vfl-040.engage.sas.com/SASStudio/));
39   %let _execenv = %nrquote(%nrstr(SASStudio));
40   %symdel _dataout_mime_type _dataout_name _dataout_url _dataout_table / nowarn;
41   %let _sasws_ = %bquote(%sysfunc(getoption(work)));
42   %let _saswstemp_ = %bquote(%sysfunc(getoption(work)));
43
44   /* Detect SAS/Graph and setup graph options */
45   data _null_;
46       length rc $255;
47       call symput("graphinit","");
48       call symput("graphterm","");
49       rc=tslvl('sasxgopt','n');
50       _error_=0;
51       if (rc^=' ') then do;
52           call symput("graphinit","goptions reset=all gsfname=_gsfname;");
53           call symput("graphterm","goptions noaccessible;");
54       end;
55   run;
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds

56   data _null_;
57       length rc 4;
58       rc=sysprod("PRODNUM002");
59       if (rc^=1) then do;
60           call symput("graphinit","");
```

```
61          call symput("graphterm","");
62       end;
63    run;
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

64
65    /* Setup ODS destinations */
66    ods _all_ close;
67    %studio_results_directory;
68    filename _htmlout "&_results_prefix_..html";
69    filename _listout "&_results_prefix_..lst";
70    filename _gsfname temp;
71    filename _dataout "&_results_prefix_..dat";
72    ods autonavigate off;
73    ods graphics on;
74    ods html5 (id=web) METATEXT='http-equiv="Content-Security-Policy" content="default-src ''none''; style-src ''unsafe-inline''; 
74 ! img-src data: ;"' device=png gpath="&_saswstemp_" path="&_saswstemp_" encoding=utf8 file=_htmlout (title='Results:SAS 
74 ! Program.sas') style=Ignite options(bitmap_mode='inline' outline='on' svg_mode='inline' css_prefix=".ods_&SYS_COMPUTE_JOB_ID" 
74 ! body_id="div_&SYS_COMPUTE_JOB_ID" );
NOTE: Writing HTML5(WEB) Body file: _HTMLOUT
75    ods listing file=_listout;
76    &graphinit;
77    %studio_initialize_custom_output;
78    /* endregion */
79
80    %let NumSamples1 = 70;
81    %let NumSamples2 = 3;
82    %let NumSamples3 = 1;
83    /* Sort the dataset by acci_severity */
84    proc sort data=WORK.IMPORT;
85    by acci_severity;
86    run;
NOTE: There were 2476 observations read from the data set WORK.IMPORT.
NOTE: The data set WORK.IMPORT has 2476 observations and 42 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds

87    /* Use PROC SURVEYSELECT to make the dataset balanced */
88    proc surveyselect data=WORK.IMPORT NOPRINT out=BalancedData1
89    method=urs
90    seed=12345
91    samprate=(1 0 0); /* Adjust the samprate to balance the strata */
92    strata acci_severity;
93    run;
NOTE: The SAMPRATE= value is 0. No sample is selected from this stratum.
NOTE: The above message was for the following stratum:
      acci_severity=2.
NOTE: The SAMPRATE= value is 0. No sample is selected from this stratum.
NOTE: The above message was for the following stratum:
      acci_severity=3.
NOTE: The data set WORK.BALANCEDDATA1 has 18 observations and 45 variables.
NOTE: PROCEDURE SURVEYSELECT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds

94    /* Create a variable with a constant value for each observation */
95    data BalancedData_with_constant1;
96    set BalancedData1;
97    constant = 1;
98    run;
NOTE: There were 18 observations read from the data set WORK.BALANCEDDATA1.
NOTE: The data set WORK.BALANCEDDATA_WITH_CONSTANT1 has 18 observations and 46 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

99    /* Use PROC SURVEYSELECT to perform bootstrap sampling based on
100   acci_severity */
101   proc surveyselect data=BalancedData_with_constant1 NOPRINT seed=1
102   method=urs
103   samprate=1
104   OUTHITS
105   reps=&NumSamples1(repname=row)
106   out=BootSamp1;
107   strata acci_severity;
108   run;
NOTE: Variable NumberHits already exists on file WORK.BOOTSAMP1, using NumberHits2 instead.
NOTE: Variable ExpectedHits already exists on file WORK.BOOTSAMP1, using ExpectedHits2 instead.
NOTE: Variable SamplingWeight already exists on file WORK.BOOTSAMP1, using SamplingWeight2 instead.
NOTE: The data set WORK.BOOTSAMP1 has 1260 observations and 50 variables.
NOTE: PROCEDURE SURVEYSELECT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

109   /* Overall, how often was each observation selected? */
110   proc freq data=BootSamp1;
111   tables acci_severity;
112   run;
NOTE: There were 1260 observations read from the data set WORK.BOOTSAMP1.
```

```
NOTE: The PROCEDURE FREQ printed page 3.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds

113  /* Use PROC SURVEYSELECT to make the dataset balanced */
114  proc surveyselect data=WORK.IMPORT NOPRINT out=BalancedData2
115  method=urs
116  seed=12345
117  samprate=(0 1 0); /* Adjust the samprate to balance the strata */
118  strata acci_severity;
119  run;
NOTE: The SAMPRATE= value is 0. No sample is selected from this stratum.
NOTE: The above message was for the following stratum:
      acci_severity=1.
NOTE: The SAMPRATE= value is 0. No sample is selected from this stratum.
NOTE: The above message was for the following stratum:
      acci_severity=3.
NOTE: The data set WORK.BALANCEDDATA2 has 386 observations and 45 variables.
NOTE: PROCEDURE SURVEYSELECT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds

120  /* Create a variable with a constant value for each observation */
121  data BalancedData_with_constant2;
122  set BalancedData2;
123  constant = 1;
124  run;
NOTE: There were 386 observations read from the data set WORK.BALANCEDDATA2.
NOTE: The data set WORK.BALANCEDDATA_WITH_CONSTANT2 has 386 observations and 46 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

125  /* Use PROC SURVEYSELECT to perform bootstrap sampling based on
126  acci_severity */
127  proc surveyselect data=BalancedData_with_constant2 NOPRINT seed=1
128  method=urs
129  samprate=1
130  OUTHITS
131  reps=&NumSamples2(repname=row)
132  out=BootSamp2;
133  strata acci_severity;
134  run;
NOTE: Variable NumberHits already exists on file WORK.BOOTSAMP2, using NumberHits2 instead.
NOTE: Variable ExpectedHits already exists on file WORK.BOOTSAMP2, using ExpectedHits2 instead.
NOTE: Variable SamplingWeight already exists on file WORK.BOOTSAMP2, using SamplingWeight2 instead.
NOTE: The data set WORK.BOOTSAMP2 has 1158 observations and 50 variables.
NOTE: PROCEDURE SURVEYSELECT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

135  /* Use PROC SURVEYSELECT to make the dataset balanced */
136  proc surveyselect data=WORK.IMPORT NOPRINT out=BalancedData3
137  method=urs
138  seed=12345
139  samprate=(0 0 0.9); /* Adjust the samprate to balance the strata */
140  strata acci_severity;
141  run;
NOTE: The SAMPRATE= value is 0. No sample is selected from this stratum.
NOTE: The above message was for the following stratum:
      acci_severity=1.
NOTE: The SAMPRATE= value is 0. No sample is selected from this stratum.
NOTE: The above message was for the following stratum:
      acci_severity=2.
NOTE: The data set WORK.BALANCEDDATA3 has 1101 observations and 45 variables.
NOTE: PROCEDURE SURVEYSELECT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

142  /* Create a variable with a constant value for each observation */
143  data BalancedData_with_constant3;
144  set BalancedData3;
145  constant = 1;
146  run;
NOTE: There were 1101 observations read from the data set WORK.BALANCEDDATA3.
NOTE: The data set WORK.BALANCEDDATA_WITH_CONSTANT3 has 1101 observations and 46 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

147  /* Use PROC SURVEYSELECT to perform bootstrap sampling based on
148  acci_severity */
149  proc surveyselect data=BalancedData_with_constant3 NOPRINT seed=1
150  method=urs
151  samprate=1
152  OUTHITS
153  reps=&NumSamples3(repname=row)
154  out=BootSamp3;
155  strata acci_severity;
156  run;
NOTE: Variable NumberHits already exists on file WORK.BOOTSAMP3, using NumberHits2 instead.
```

```
NOTE: Variable ExpectedHits already exists on file WORK.BOOTSAMP3, using ExpectedHits2 instead.
NOTE: Variable SamplingWeight already exists on file WORK.BOOTSAMP3, using SamplingWeight2 instead.
NOTE: The data set WORK.BOOTSAMP3 has 1101 observations and 50 variables.
NOTE: PROCEDURE SURVEYSELECT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds

157  proc freq data=BootSamp3;
158  tables acci_severity;
159  run;
NOTE: There were 1101 observations read from the data set WORK.BOOTSAMP3.
NOTE: The PROCEDURE FREQ printed page 4.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

160  data BootSamp;
161  set BootSamp1 BootSamp2 BootSamp3;
162  run;
NOTE: There were 1260 observations read from the data set WORK.BOOTSAMP1.
NOTE: There were 1158 observations read from the data set WORK.BOOTSAMP2.
NOTE: There were 1101 observations read from the data set WORK.BOOTSAMP3.
NOTE: The data set WORK.BOOTSAMP has 3519 observations and 50 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds

163  proc freq data=BootSamp;
164  tables acci_severity;
165  run;
NOTE: There were 3519 observations read from the data set WORK.BOOTSAMP.
NOTE: The PROCEDURE FREQ printed page 5.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds

166
167  /* region: Generated postamble */
168  /* Close ODS destinations */
169  &graphterm; ;*';*";*/;run;quit;
170  quit;run;
171  ods html5 (id=web) close;
172  ods listing close;
173  %if %sysfunc(fileref(_gsfname)) lt 0 %then %do;
174      filename _gsfname clear;
NOTE: Fileref _GSFNAME has been deassigned.
175  %end;
176  %studio_capture_custom_output;
177  /* endregion */
178
```

Results:Cleaned_Road_Accident_2021_MergedwithDescription_Data_Balancing.sas

The FREQ Procedure

acci_severity	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1260	100.00	1260	100.00

The FREQ Procedure

acci_severity	Frequency	Percent	Cumulative Frequency	Cumulative Percent
3	1101	100.00	1101	100.00

The FREQ Procedure

acci_severity	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1260	35.81	1260	35.81
2	1158	32.91	2418	68.71
3	1101	31.29	3519	100.00