

SAS-AF APPLICATION IN DETERMINING SAMPLE SIZE

Zhongmin Li*

for Poster Session of Second Annual South-Central Regional
SAS Users Group Conference
November 1992

Sampling survey designers are often facing a challenge of determining a proper sample size. SAS-AF facility makes it easier to accomplish this task. The program presented below provides an alternative for dealing with either nominal scale or interval scale variable, and calculating the minimum sample size required from a finite population with a variety of options. In addition, if the samples are drawn from a given SAS data set, applying the number of the sample size will generate a set of Simple Random Sample (SRS) from the given data.

There are two major advantages in using this AF application: (1) it is a user-friendly interactive application, and no program coding required for end-users. (2) by changing values one or more required fields, an user is able to compare a variety of combinations and select the most adequate sample size. However, this AF application is not complicate enough to determine sample size for any multi-stage or stratified sampling. I hope this application will serve as a start point for such an effort.

The attached examples provide a self-explanatory instruction for developing this AF application.

References:

Hubert M. Blalock, Jr. (1979). Social Statistics, Revised Second Edition. Chapter 12.4 Determining the Sample Size and Chapter 21. Sampling. McDraw-Hill, Inc.

SAS Institute Inc. (1989). SAS/AF Software: Usage and Reference, Version 6, First Edition. SAS Institute Inc., Cary, NC.

SAS Institute Inc. (19??). Technical Report ?, Chapter 5.6, Additional Feature (Optional) Random Sampling without Replacement. p298.

* Research Associate at Texas Workers' Compensation Research Center.

```

ELOG#####N
e Command ==>
e
e NOTE: Running on IBM Model 3090 Serial Number 174186,
e             IBM Model 3090 Serial Number 274186,
e             IBM Model 3090 Serial Number 374186,
e             IBM Model 3090 Serial Number 474186.
e
e
e Welcome to the SAS System, Release 6.07.
e
e SAS Release 6.07 is now in production mode.
e
e
e NOTE: No options specified.
e
e NOTE: The initialization phase used 0.19 CPU seconds and 3545K.
D#####SM
EPROGRAM EDITOR#####N
e Command ==>
e
e 00001 LIBNAME AFAPP 'CT0.ZML.SASLIB1';RUN;
e 00002 PROC BUILD C=AFAPP.SAMPLE;RUN;
e 00003
e 00004
e 00005
e 00006
e 00007
e 00008
e 00009
D#####SM

```

[illegible]

[illegible]

```

EBUILD: DISPLAY NOMINAL.PROGRAM (E)#####N
e Command ==>e
e
eEBUILD: SOURCE NOMINAL.PROGRAM (E)#####Ne
ee Command ==>ee
ee
ee 00001 init:ee
ee 00002     return;ee
ee 00003ee
ee 00004 main:ee
ee 00005     if ci=95 then z=1.96;ee
ee 00006     if ci=95 then z=1.96;ee
ee 00007     if ci=99 then z=2.33;ee
ee 00008 n=pop;ee
ee 00009 p=pro/100;ee
ee 00010 m=me/100;ee
ee 00011 if (n=_blank_ or ci=_blank_ or me=_blank_) then do;ee
ee 00012     _msg_='Note: This field must be entered.';ee
ee 00013     return;ee
ee 00014 end;ee
ee 00015ee
ee 00016 srs=round(((z**2)*p*(1-p)*n)/(n*m**2+(z**2)*p*(1-p)),1);ee
ee 00017ee
ee 00018 term:ee
ee 00019     return;ee
ee 00020ee
ee 00021ee
ee 00022ee
ee 00023ee
ee 00024ee
eD#####Me
D#####R#####M

```



```

EBUILD: DISPLAY INTERVAL.PROGRAM (E)ssssssssssssssssssssssssssssssssssssssssssssssN
e Command ==>
e
eEBUILD: SOURCE INTERVAL.PROGRAM (E)ssssssssssssssssssssssssssssssssssssssssssssNe
ee Command ==> COMPILE
ee
ee 00001 init:
ee 00002 return;
ee 00003
ee 00004 main:
ee 00005     n=pop;
ee 00006     z=1.96;
ee 00007     v=(sd)**2;
ee 00008 if (n=_blank_ or sd=_blank_ or me=_blank_) then do;
ee 00009     _msg_='Note: This field must be entered.';
ee 00010     return;
ee 00011 end;
ee 00012
ee 00013 srs=round(((z**2)*v*n)/(n*me**2+(z**2)*v),1);
ee 00014
ee 00015 term:
ee 00016 return;
ee 00017
ee 00018
ee 00019
ee 00020
ee 00021
ee 00022
ee 00023
ee 00024
eDssssssssssssssssssssssssssssssssssssssssssssssMe
DssssssssssssssssssssssssssssssssssssssssssssssRssssssssssM

```

```

EBUILD: DISPLAY INTERVAL.PROGRAM (E)ssssssssssssssssssssssssssssssssssssssssssssN
e Command ==>
e
eEBUILD: SOURCE INTERVAL.PROGRAM (E)ssssssssssssssssssssssssssssssssssssssssssssNe
ee Command ==> TESTAF
ee NOTE: Code generated for INTERVAL. Code size=824.
ee 00001 init:
ee 00002 return;
ee 00003
ee 00004 main:
ee 00005     n=pop;
ee 00006     z=1.96;
ee 00007     v=(sd)**2;
ee 00008 if (n=_blank_ or sd=_blank_ or me=_blank_) then do;
ee 00009     _msg_='Note: This field must be entered.';
ee 00010     return;
ee 00011 end;
ee 00012
ee 00013 srs=round(((z**2)*v*n)/(n*me**2+(z**2)*v),1);
ee 00014
ee 00015 term:
ee 00016 return;
ee 00017
ee 00018
ee 00019
ee 00020
ee 00021
ee 00022
ee 00023
ee 00024
eDssssssssssssssssssssssssssssssssssssssssssssssMe
DssssssssssssssssssssssssssssssssssssssssssssssRssssssssssM

```

```

#####N
e Command ==->
e
e
e What is the population size? 25000
e
e What is the standard deviation for
e the variable you are interested in? 1050
e
e What is the acceptable marginal error? 100
e
e *****
e
e The minimum sample size required at
e 95% confident interval is 416
e
e
e
e
e #####R#####M

```

```

EDIT ----- CTO.ZML.PROG(SAMPLE) - 01.03 ----- COLUMNS 001 072
COMMAND ==> SCROLL ==> PAGE
000001 /*IN BATCH MODE:*/
000002 /*APPLYING THE CALCULATED SAMPLE SIZE AND SELECTING RANDOM*/
000003 /*SAMPLES FROM A GIVEN SAS DATASET.*/
000004
000005 LIBNAME NEWLIB 'SASLIB1';
000006 DATA SAMPLE(DROP=SAMPLESIZE OBSLEFT);
000007     SAMPsize=?????; /*MODIFY HERE*/
000008     OBSLEFT=TOTOBS;
000009     DO WHILE(SAMPsize>0);
000010         COUNTER+1;
000011         IF RANUNI(0)<SAMPsize/OBSLEFT THEN DO;
000012             SET NEWLIB.SASDATA POINT=COUNTER NOBS=TOTOBS;
000013             OUTPUT;
000014             SAMPsize=SAMPsize-1;
000015             END;
000016             OBSLEFT=OBSLEFT-1;
000017         END;
000018     STOP;
000019 RUN;
000020 PROC PRINT DATA=SAMPLE SPLIT='*' D N NOOBS;
000021 .....
***** BOTTOM OF DATA *****

```

```

F1=HELP      F2=SPLIT    F3=END       F4=RETURN    F5=RFIND     F6=RCHANGE
F7=UP        F8=DOWN      F9=SWAP      F10=LEFT     F11=RIGHT    F12=RETRIEVE

```