Paper 3820-2016

Using SURVEYSELECT to Draw Stratified Cluster Samples with Unequally Sized Clusters

Chuanjie "George" Liao, Pearson, Iowa City, IA

Brian F. Patterson, Questar Assessment, Inc., Minneapolis, MN

ABSTRACT

PROC SURVEYSELECT is a useful procedure for sample selection for a wide variety of applications. This paper presents the application of PROC SURVEYSELECT to a complex scenario involving a state-wide educational testing program, namely drawing inter-dependent stratified cluster samples of schools for the field-testing of test questions, which we call "items". These stand-alone field tests are given to only small portions of the testing population and as such, a stratified procedure was used to ensure representativeness of the field-test samples. As these items' field test statistics are evaluated for use in future operational tests, an efficient procedure is needed to sample schools, while satisfying pre-defined sampling criteria and targets. This paper provides an adaptive sampling application and then generalizes the methodology as much as possible for potential use in more industries.

INTRODUCTION

The SAS Institute 9.2 User's Guide, Second Edition provides a thorough introduction about PROC SURVEYSELECT. That along with past papers presented at the SAS Global Forum (e.g., Lewis, 2013) provide examples of replicated sampling, stratified sampling, etc, the latter of which is the focus of our work. A template of PROC SURVEYSELECT is presented below and some of the keywords and options that are most relevant for the purposes of stratified sampling are:

- METHOD specifies the sampling method. SRS (simple random sampling) is used in our example.
- N sample size, indicating the total number of primary sampling units (PSUs). In our case, it refers to school-grade combinations.
- STRATA specifies the strata variable for the entire sampling frame.
- ALLOC determines the sampling weight by either a value or dataset. A user-defined number value can be used to proportionally allocate the sample size. A system value such as PROP or OPTIMAL is also often used in practice. In our example, we use a dataset which specifically includes the allocation variable and related weights to allocate the total sample across the strata.
- ID specifies the key identifying variables, which will be included in the output dataset.

```
PROC SURVEYSELECT DATA= DataFrame
            METHOD= SRS
            N= n_Schools
            OUT= Sample;
      STRATA School_Geog_Score_Quant /
            ALLOC= Target_OP;
      ID Subject Grade Form_Type School_ID School_Geog;
RUN;
```

To extract cluster samples from a large sampling frame, this paper first introduces the preparation of strata and allocation datasets and then focuses on PROC SURVEYSELECT which is iteratively executed until meeting the stopping criteria, namely total sampled students and the representativeness criteria. The total number of sampled schools is adjusted adaptively across iterations. The actual implementation is more complicated due to state specific requirements, so this paper will instead simplify codes and details to enable readers to grasp the thrust of the issues at play and the relevant SAS code - some data steps and procedures are in open code but some macros are presented along to enhance the readability. While our application was limited to an educational setting, the potential for applications to other research areas are vast. Any reader needing to draw a cluster sample with unequally sized clusters with varying sample weights could benefit from these methods.

A large state testing program had newly developed test items that needed empirical validation and as such, we conducted a stand-alone field test – that is a test administration separate from a regular large-scale, high-stakes setting in which all students are expected to participate. Because only a small subset of students participate in the stand-alone field test, it is important to have samples that are representative both in terms of school geography and urbanicity (rural, suburban and urban) and in terms of prior academic achievement. Therefore, we used a stratified sampling approach. For simplicity, the variable that we'll use in this paper to characterize school geography and urbanicity will be called "School_Geog". In a similar way, we'll characterize the prior achievement of schools' students by that school's prior year's mean test scale score, using the variable name "School_SS". Scale score is the score appearing on students' regular test score reports. An additional constraint that complicated our work was the fact that a given school can be selected at most two grades and the grades must be adjacent.

A flowchart is included here to illustrate the whole procedure, which can be generalized as (1) creating the sampling frame with only eligible schools; (2) preparing sampling targets via the population; (3) drawing independent samples, evaluating and (4) as needed, adjusting sampling targets. In our paper, we will elaborate on each step in detail and at the same time, we will simplify our discussion of our work to reach the widest possible audience of readers.

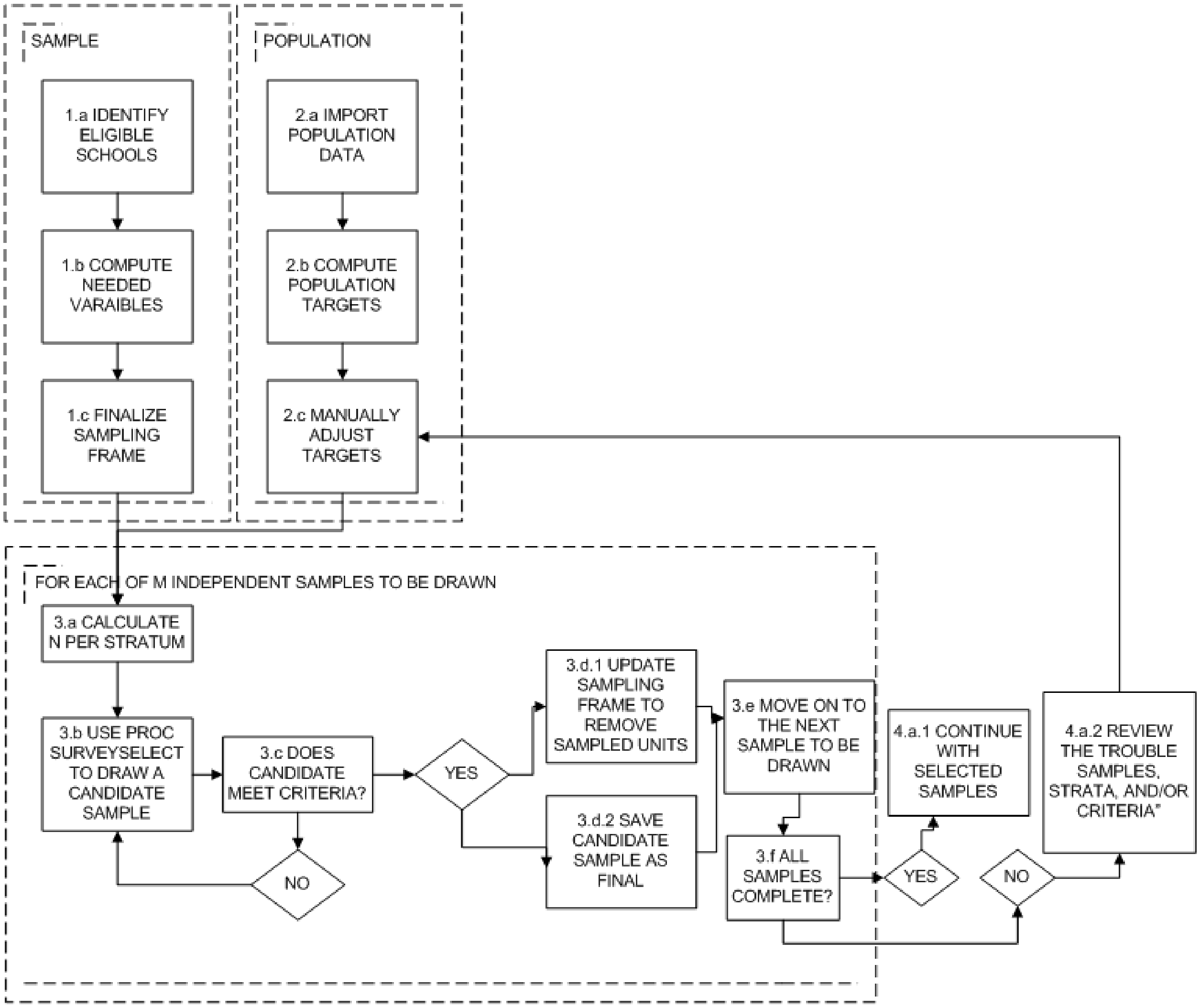

Figure 1. Flowchart of Iterative Sampling Procedure

I. PREPARING TO DRAW THE SAMPLES

CREATE SAMPLING FRAME

As shown in the block of Figure 1 with the header "SAMPLE", the first step is to prepare the sampling frame. From all schools in one state, only eligible schools can be selected as a sampling frame. In order to get a sampling frame with eligible schools only, the state had a variety of rules related to sample size, school type, and participation in other testing programs' field tests. Suffice it to say that approximate school enrollment, the number of high-stakes (i.e., operational) tests ordered by grade and subject, and prior years' testing data all went into the construction of the sampling frame and the determining of school eligibility for our stand-alone field test. Note that the way we identify one particular school is to use a unique School_ID. In terms of school eligibility, it is also important to note that not all schools share the same subjects and grades. It is possible that a given school may be eligible for field testing in a few grades (e.g., elementary schools may only have grades 3 through 5) or one subject, but here we want to focus on the sampling from a more general perspective.

CONSTRUCT THE STRATIFYING VARIABLE

The general rule for sampling in our example is to draw schools so that the whole drawn samples roughly match the prior year's population test score means for the relevant grade and subject, i.e. the entire state operational data population. However, the challenge is we do not have targets per se that could be used to construct sampling weights.

We create a strata variable School_Geog_SS_Qnt at school level, which is the combination of School_Geog and the quantile for the school's prior scale score means. This is the strata variable we will use in the later PROC SURVEYSELECT for sampling. School_Geog values are obtained from the sampling frame but scale scores are from prior year's population-wide test data.

```
* Calculate scale score means by School ID;
PROC MEANS NOPRINT DATA= OP_Prior_Year;
      BY Subject Grade School_ID School_Geog;
      VAR SCALE_SCORE;
      OUTPUT OUT= OP_Year_SS_Means
            N(SCALE_SCORE)= OP_Year_Examinees
            MEAN(SCALE_SCORE)= OP_Year_SS_Mean;
RUN;
```

In order to better match the state-wide scale score means, we calculate school-level percentile ranks for each subject and grade. By incorporating those percentile ranks into the development of sampling weights, we will ensure not only that the school geographic/urbanicity distribution for each selected sample matches that of the state, but also that the distribution of school scale score means matches the target distribution. In the code to follow, we calculate the scale score mean percentile ranks for each subject and grade from the 0.01 to the 99.99 percentile. In our case, we only stratified schools into 5 quantiles (i.e., using 20th, 40th, etc. percentile ranks) and that number can vary by setting. We were able to balance computation time, precision of meeting sampling criteria, and other factors using five quantiles. For the sake of convenience, we then concatenated the scale score mean quantile with school geographic/urbanicity information to construct a single strata variable. This helped simplify the manner in which we assigned sampling weights while still achieving the sampling representativeness requirement.

```
*Calculate percentiles;
PROC UNIVARIATE NOPRINT DATA= OP_Year_SS_Means;
      BY Subject Grade;
      VAR OP_Year_SS_Mean;
      OUTPUT OUT= OP_Year_SSM_Pctl
            PCTLPTS= 0.01 TO 99.99 BY 0.01
            PCTLPRE= OP_Year_SS_Mean_P_;
RUN;

%LET SS_MEAN_QNT= 5; *User-defined number of quantiles;
```

```
%DO Qnt_Ind= 1 %TO %EVAL(&SS_Mean_Qnt-1);
      %LET SSM_Year_Qnt_&Qnt_Ind=
            %SYSFUNC(TRANSLATE(%SYSFUNC(
            ROUND(%SYSEVALF(100*(&Qnt_Ind/&SS_Mean_Qnt)),0.01)),_,.));
%END;

DATA OP_Year_SSM_Qnt;
      SET OP_Year_SSM_Pctl
            (KEEP= Subject Grade
                  %DO Quantile= 1 %TO %EVAL(&SS_Mean_Qnt-1);
                        OP_SS_Mean_P_&&SSM_Year_Qnt_&Qnt_Ind
                  %END;);
RUN;
```

From the above procedure, we get school scale score quantile from population and School_Geog values are already directly obtained from the sampling frame. The strata variable School_Geog_SS_Qnt is constructed as shown below. An example strata level could be "1|03", which means school geographic code of 1 + 3rd quantile. Note variable OP_Year_SSM_Qnt is a new variable sharing the same meaning of same scale score mean quantile as the naming of data set OP_Year_SSM_Qnt, and is just obtained from merging data set OP_Year_SSM_Qnt back to sampling frame.

```
* Construct the strata variable;
School_Geog_SS_Qnt= CATX('|', School_Geog, OP_Year_SSM_Qnt);
```

After the strata variable is constructed for each school in the sampling frame, we merge it back to update main dataset OP_Year_SS_Means, which has information for Subject, Grade, School_Geog_SS_Qnt, School_ID, expected number of examinees, and SS means. Using dataset OP_Year_SS_Means, the next step is to assign different sampling weights to all unique levels of strata variable School_Geog_SS_Qnt, so that we can allocate the total sample size.

DEVELOP THE SAMPLING WEIGHTS

In the logic of stratified sampling and sample size allocating, it is intuitive to think of assigning related sampling weights after the corresponding strata variable is constructed. For instance, if one subject and one grade have 8 school geographic codes and 5 quantiles, the maximum number of possible School_Geog_SS_Qnt level values are 8*5 = 40 and percentages would need to sum up to 100 percent.

As mentioned previously, the dataset OP_Year_SS_Means has all information such as scale score quantile at school level. Aside from school level percentage School_Geog_SS_Qnt_Pct_Sch, we can also get percentage School_Geog_SS_Qnt_Pct_Stu, which is weighted by total number of students expected to test in the current year's operational test. These weighted frequencies on the strata variable themselves may be directly used as sampling weights and PROC SURVEYSELECT requires that we have a single scalar quantity for the number of sampled schools, and so using these frequencies tend to give rise to samples with too many large schools and too few small schools. To resolve this, we adjust the sampling weights using the mean expected test-takers by subject, grade, and school geographic/urbanicity variable. In particular, we simply divide the weighted frequency by the expected number of examinees in that particular category and then normalize that measure to ensure that it sums to one. We name this weighted measure as School_Geog_SS_Qnt_Wgt, which is another important adjustable factor in our sampling process. As for all the preparation steps by far, we hope this could shed some insights for readers about how to construct a strata and related weight from stratum layers of information.

```
* Calculate school-weighted frequencies;
PROC FREQ NOPRINT DATA= OP_Year_SS_Means;
      BY Subject Grade;
      TABLE School_Geog_SS_Qnt / MISSING
            OUT= OP_Year_Alloc_Sch
                  (RENAME= (PERCENT= School_Geog_SS_Qnt_Pct_Sch));
```

```
RUN;

* Calculate student-weighted frequencies;
PROC FREQ NOPRINT DATA= OP_Year_SS_Means;
      BY Subject Grade;
      WEIGHT OP_Year_Examinees;
      TABLE School_Geog_SS_Qnt / MISSING
            OUT= OP_Year_Alloc_Stu
                  (RENAME= (PERCENT= School_Geog_SS_Qnt_Pct_Stu));
RUN;

* Calculate mean expected examinees by School_Geog value;
PROC MEANS NOPRINT DATA= Frame;
      BY Subject Grade;
      CLASS School_Geog_SS_Qnt School_Geog;
      VAR Expected_Examinees;
      OUTPUT OUT= Frame_Mean_Exp
            MEAN(Expected_Examinees)= Mean_Expected_Examinees;
RUN;

DATA _NULL_;
      SET Frame_Mean_Exp;
      CALL SYMPUTX(CATX('_', 'Mean_Exp', Subject, Gr, School_Geog),
            Mean_Expected_Examinees);
RUN;

DATA Target_OP_Year_Alloc;
      MERGE OP_Year_Alloc_Sch
            OP_Year_Alloc_Stu
            Frame_Mean_Exp
            /* Frame_Mean_Exp is a dataset containing
                  mean enrollment by subject, grade, and School_Geog */;
      BY Subject Grade School_Geog_SS_Qnt;

      * Update simple frequency to get sampling weight;
      School_Geog_SS_Qnt_Wgt= School_Geog_SS_Qnt_Pct_Stu /
                              Mean_Exp_&Subject._&Gr._&School_Geog;
RUN;
```

At last, we create a dataset Target_OP_Year_Alloc which contains the sampling weight (i.e., School_Geog_SS_Qnt_Wgt) by subject, grade and the stratifying variable. This dataset in addition to the previously constructed sampling frame datasets are the main inputs to the PROC SURVEYSELECT instances in the “II. Drawing the samples” section later.

The last input to the sampling procedure is determining the number of primary sampling units (i.e., school-grade combinations) to draw. In our practice, we were constrained by a total sample size (i.e., number of students) for each of the inter-dependent, roughly 40 samples, that we needed to draw. Because of the manner in which we specified the sample size (i.e., indicating the number of PSUs or schools), it was non-trivial to calculate an initial number of schools to sample in our first sampling pass. We ran a number of replications of this procedure and in doing so had a deterministic component to the initial cluster count and to that we added a random factor to ensure that the replicates covered a reasonable range of sample sizes and would give rise to better guesses at future cluster counts. The deterministic component was the target sampled individuals for the relevant subject, grade, and form (i.e., item) type divided by the mean expected number of examinees in that subject and grade. The details of this procedure are less important than understanding that we passed the procedure an initial target number of clusters which would be adjusted via the iterative procedure described below in the section “REPEATING THE PROCESS FOR ALL INDEPENDENT SAMPLES”.

II. DRAWING THE SAMPLES

DRAW AN INITIAL SAMPLE USING PROC SURVEYSELECT

We eventually have the dataset with stratification information ready for the sampling. The sampling method we chose was simple random sampling (SRS). Recall that variable School_Geog_SS_Qnt is defined as the strata variable. We also specify n_Schools observations we would want to sample. We specify the dataset containing sampling targets with our allocation proportions via the ALLOC option. The sampling data frame Frame_&Sample_Num._&Subject._&Gr._&Form is classified by subject, grade and form in the ith drawn sample, and we have two forms usually – multiple-choice form and open-ended question form. So the allocation datasets Target_OP_Year_Alloc_&Subject_&Gr._&Form. are children datasets from the Target_OP_Year_Alloc dataset, by Subject, Grade and Form.

```
PROC SURVEYSELECT NOPRINT DATA= Frame_&Sample_Num._&Subject._&Gr._&Form
            METHOD= SRS
            N= &n_Schools
            OUT= Sample_&Sample;
      STRATA School_Geog_SS_Qnt /
            ALLOC= Target_OP_Year_Alloc_&Subject._&Gr._&Form;
      ID Subject Grade Form_Type School_ID School_Geog;
RUN;
```

The n_Schools MACRO variable, i.e. number of clusters, is the total number of primary sampling units. However, we still need to make sure it leads to reasonably good sample draws, otherwise, we should adjust it. It is possible that, due to random adjustment, the number of schools that we request be sampled exceeds the number of possible schools in this strata named N_Strata. The value of N_Strata can be derived from the number of records in the children allocation datasets Target_OP_Year_Alloc_&Subject._&Gr._&Form., for instance, how many schools in Mathematics Grade 3 Multiple-Choice Form. If N_school is less than N_Strata, it is re-set to be equal to N_Strata. If we ever get down this low, would probably be in trouble in terms of meeting our sampling criteria, but this bit of code helps avoid an error in our SAS log file.

```
%IF &n_Schools < &n_Strata %THEN %LET n_Schools= &n_Strata;
```

Thus far, we have the structure of the data and SAS code in place to get initial samples drawn using a reasonable guess at the number of school clusters and implementing SURVEYSELECT procedure, before evaluating the samples against our strata and sample size targets.

EVALUATE SAMPLE DRAW FOR SAMPLE SIZE AND MATCH TO STRATA TARGET

We evaluate the drawn samples from three criteria – total number of students, grand mean of scale scores, and school geographic targets. These three targets are translated into three variables referred in the checking codes below - total target students needed Target_n_&Subject._&Gr._&Form, grand scale score mean Grand_Mean_&Subject_&Gr and school geographic percentage Target_School_Geog_Percent. These three variable values can be derived from dataset Target_OP_Year_Alloc_&Subject._&Gr._&Form., which is the allocation dataset from the previous PROC SURVEYSELECT procedure.

It is worth noting that the evaluation for each individual sample is by subject, grade and form, but not by school. In the evaluation procedure, a binary MACRO variable Sample_Criteria_Met is first initialized as 1. If any of the three criteria is not met, this MACRO variable is updated as 0, which means we have drawn a bad sample and must iterate again.

```
%LET Sample_Criteria_Met= 1;
```

A. Evaluate sampled students

This is our first checking criteria. We have a target number of sampled examinees and a target absolute deviation proportion.

```
Target_Examinees= Target_n_&Subject._&Gr._&Form;

Sampled_Examinees_Dev= Sampled_Examinees - Target_Examinees;

Sampled_Examinees_Abs_Prop_Dev==
ABS(Sampled_Examinees_Dev)/ Target_Examinees;
Target_to_Sampled_Examinees= Target_Examinees / Sampled_Examinees;

IF Sampled_Examinees_Abs_Prop_Dev > &Target_Sample_Tolerance_Prop OR
Sampled_Examinees_Dev < 0 THEN CALL SYMPUTX('Sample_Criteria_Met', 0);
```

B. Evaluate grand scale score means

In our practice, we pre-set the target scale score mean tolerance as 3 points, which was the state sampling criteria. If absolute deviance in the sample was within this range, it was acceptable. It is worth noting that as we worked to refine our code, we were able to make adjustments to do better than simply meeting the stated sampling criteria. For example, having been able to successfully draw a sample within the required three (+/- 3) test score points on the prior year's assessment, we reduced that parameter and were able to draw a sample within two (+/- 2) points of the prior year's population mean test score. This is just an example of the sort of tinkering that we found necessary to balance meeting the competing targets and trying to do so in a reasonably efficient manner.

```
%LET Target_SS_Mean_Tolerance= 3;

Target_Year_SS_Mean= Grand_Mean_&Subject_&Gr;

Samp_Year_SS_Mean_Abs_Dev= ABS(Samp_Year_SS_Mean - Target_Year_SS_Mean);

IF Samp_Year_SS_Mean_Abs_Dev > &Target_SS_Mean_Tolerance
THEN CALL SYMPUTX('Sample_Criteria_Met', 0);
```

C. Evaluate geography and urbanicity representativeness

For each School_Geog category, we calculate the absolute School_Geog proportion deviation. If the largest absolute deviation is within the pre-determined 5% range, it is acceptable.

```
%LET Target_School_Geog_Pct_Tolerance= 5;

%DO School_Geog_Ind= 1 %TO 8;
        Sampled_School_Geog_Percent_Abs_Dev=
                ABS(Sampled_School_Geog_Percent - Target_School_Geog_Percent);
        IF Sampled_School_Geog_Percent_Abs_Dev >
                &Target_School_Geog_Pct_Tolerance THEN
                CALL SYMPUTX('Sample_Criteria_Met', 0);
%END;
```

REPEATING THE PROCESS FOR ALL INDEPENDENT SAMPLES

The whole sample drawing process was implemented in a repetitive loop and did not stop until either the sampling criteria were all met or we hit a pre-defined number of maximum iterations. If the sample passes the evaluation procedure, we update the sampling frame to remove those selected PSUs. We did this by a data-step merge between the original sampling frame and selected sample dataset. If we do meet the sampling criteria, we may update the sampling targets, number of PSUs to sample, etc. in order to draw better candidate samples in subsequent iterations. Like many statistical applications, PROC SURVEYSELECT seems relatively straightforward to implement for sampling, but does require us to make reasonable adaptive adjustments during the procedure to more effectively find good samples. Next, we will mainly focus on how to make adjustment during the iterations from three perspectives. These are meant to be examples for researchers, rather than a comprehensive accounting of all possible adjustments. Subject-matter expertise will guide which of these – if any – approaches would help readers accomplish their particular goals in sampling.

We rely on PROC SURVEYSELECT, and the main options, number of clusters and sampling weight, could be tweaked to aid the adjustment as we cannot change the strata variable. We will show the sampling order matters too.

A. ADJUST NUMBER OF CLUSTERS

The number of clusters, i.e. the number of schools, which is initially equal to total target students needed divided by average number of students per school but no more than the number of strata, nevertheless needs to be adjusted to find more feasible samples rather than restricting itself to a deterministic local optima deadlock. We adaptively adjust the number of clusters only when all sampling criteria are not met, that is to say, if bad samples are drawn. To conquer this, a parameter n_Schools_Factor with initial value 1.00 is introduced first but adjusted randomly a bit, every 10 iterations or so, to help search around the target number of schools while still following the general determination of the number of schools.

The extent of adjustment is determined by the mean of a proportion of target examinees to total expected examinees in the current sample. The mean of target to sampled examinees is attained by a standard PROC MEANS with a class statement for the current sample number, Subject Grade and Form_Type. In other words, all these are still for each individual iteration and limited to a sample level.

```
%LET n_Schools_Factor= 1.00;

Sampled_Examinees = SUM(Expected_Examinees);
Target_to_Sampled_Examinees= Target_Examinees / Sampled_Examinees;
M_Target_to_Sampled_Examinees= MEAN(Target_to_Sampled_Examinees);

n_Schools_Factor_Rev= &n_Schools_Factor * M_Target_to_Sampled_Examinees;

*Update n_Schools_Factor;
CALL SYMPUTX('n_Schools_Factor', n_Schools_Factor_Rev);

*Adjust the number of school;
N_schools= n_Schools_Factor * N_schools;
```

B. ADJUST SAMPLING WEIGHT

We have elaborated in detail in Section “DEVELOP THE SAMPLING WEIGHTS” about the sampling weight. Besides the cluster size N_school, we also adjust sample weights. The reason is the sampling weight plays an essential role in the repeating process adaptively. The sampling allocations is finalized according to school size which is determined by the mean expected examinees, and as a result, the fewer mean expected examinees there are, like in smaller sized schools, the more weight is assigned and the more likely a school-grade combination is to be sampled.

```
%DO Grade= 1 %TO 5;
      %DO Subject_Ind= 1 %TO 4;
            %DO School_Geog= 1 %TO 8;
                  School_Geog_SS_Qnt_Wgt= School_Geog_SS_Qnt_Pct_Stu /
                        Mean_Exp_&Subject._&Gr._&School_Geog;
            %END;
      %END;
%END;
```

Another obvious consequence is that the adjusted weights are not theoretically bound to sum to one due to the number of different mean expected examinees. Hence, we wrote a simple post hoc MACRO to normalize the weights but still obey the weight adjusting rule per se.

C. ADJUST SAMPLING ORDER

Just as with many other types of constrained optimization, the order in which we select samples (e.g., do we start with the lower grades first, or the upper grades? Do we select math or English Language Arts samples first?) does affect our ability to meet our sampling targets. After the sampling process, we only output samples which have sampled all possible combinations, i.e. subject x grade x form type in our setting. For instance, we could track how many presented form types are sampled, output the error and append the failed samples into a dataset for identification. If not all combinations have been sampled, we identified where the process failed and consider adjusting the sampling order preference, if there was a systematic tendency to fail at a given subject or grade level. Based on our experience, we present some recommendations for adjusting sampling order at the end of this paper.

```
%LET All_Form_Types_Present= 0;
DATA _NULL_;
      SET Sample_&Sample_Num._Complete_QC END= Last;
      IF Last THEN DO;
            CALL SYMPUTX('Form_Types_Present', _N_);
            *Check if all possible 40 combinations have been sampled;
            IF _N_ = 40 THEN CALL SYMPUTX('All_Form_Types_Present', 1);
      END;
RUN;

*If fail to sample all combinations;
%IF &All_Form_Types_Present = 0 %THEN %DO;
      PROC APPEND BASE= SAFT.Sample_Complete_QC
            DATA= Sample_&Sample_Num._Complete_QC FORCE;
      RUN;

      %PUT ERROR: Only &Form_Types_Present form types sampled.;
%END;
```

III. CONCLUSION AND RECOMMENDATIONS

In this paper, we presented the application of PROC SURVEYSELECT in an educational testing program to select sample schools based on a combination of different sources of information. We incorporated the general methodology of the SURVEYSELECT procedure and hands-on adjustment to meet strict state testing requirements. We have demonstrated the flexibility and applicability of the procedure and would like to close with some more general guidance for interested readers.

- Consider drawing the largest samples before drawing the smaller. Based on additional sampling constraints, the ideal order of sampling may vary. For example, the higher grades tended to have larger average class sizes and this coarseness led us to sample the middle school grades before the lower grades. Generally, we drew the samples from the higher grades to the lower grades, but we also could manipulate any specific grade for the sampling order after inspecting the sampling frame.

- Inspect the sampling frame and consider drawing samples first that may have fewer clusters available in the sampling frame. It is always a good practice to look into the original data before processing. We inspected the geographic information and number of schools per grade, and realized that due to the nature of our constraints to sample only consecutive grades when we sample two grades from one school, it was easier to meet the sampling criteria for certain grades, above and beyond the observation about average class size.

- Similar to different sampling order preference for each grade, we adjusted the allocation proportions for non-public schools downward, so we didn't run out of non-public schools by the time we reach the last subject, grade, and item type. In other words, we biased the non-public school targets downward but only so far down as would still enable us to meet our +/-%5 sampling criterion. As we needed to allocate both public and non-public schools to different clusters by subject, grade and item type, we knew there were fewer non-public schools beforehand.

- Constructing meaningful strata is also critical. As we discussed above, in our combination of School_Geog and the school mean scale score quantile, we stratified schools into 5 quantiles (i.e., using 20^{th}, 40^{th}, etc. percentile ranks) rather than investigating all 100 quantiles.

REFERENCES

Lewis, T. (2013). "PROC SURVEYSELECT as a Tool for Drawing Random Samples," Paper presented at the annual conference of the Midwest SAS Users Group. Columbus, OH, September 22-24. Available online at: http://www.mwsug.org/proceedings/2013/AA/MWSUG-2013-AA02.pdf.

Lohr, S. (1999). Sampling: Design and Analysis. Pacific Grove, CA: Duxbury Press.

ACKNOWLEDGMENTS

We would like to sincerely thank Brian Wrobel for his advice and support on this paper.

CONTACT INFORMATION

Your comments and questions are most welcome. Please feel free to contact the author at:

Chuanjie "George" Liao
Pearson
2510 N. Dodge Street
Iowa City, IA 52245
Office: (319) 354-9200 X 215781
E-mail: george.liao@pearson.com