Designing a Sampling Method for a Survey of Iowa High School Seniors

Kyle A. Hewitt and Michael D. Larsen, Iowa State University
Department of Statistics, Snedecor Hall, Ames, Iowa 50011-1210, larsen@iastate.edu

Keywords: Cluster sampling, education research, margin of error, probability sampling, simple random sampling, stratified random sampling.

1. Introduction

Are the distributions of courses taken by high school students in Iowa substantially different in small and large school districts? What factors affect student plans after high school graduation? Representatives of Iowa's State Board of Education brought this question and many others to Iowa State University's Center for Survey Statistics and Methodology (CSSM) in the fall of 2004. Their research questions primarily concern the question of equality of opportunity for students across the state. It was decided to conduct a series of surveys to gather needed information. This article describes the research project and a simulation study. Section 2 provides background on the research questions and surveys. Sections 3 and 4 discuss sampling designs. Section 5 reports on a simulation study. The actual survey design used in the study is described in Section 6.

2. Background

Public school districts in Iowa range in size from a couple hundred to several thousand students in grades K-12. The number of students per high school ranges from less than 100 to more than 2000. Figure 1 is a histogram of school sizes. The large schools tend to be grouped together in districts, thereby making the district size distribution even more right skew. The districts design their own curricula and set graduation requirements independently. There is no standard reporting requirement to the State Board that would have enabled the school representatives to answer their questions based on existing data. As a result, it was decided to conduct some surveys in Iowa to gather the needed information. The 2005 Iowa legislature debated, but did not pass a bill, setting a minimum district size (Higgins 2005; see also Des Moines Register 2005a). A poll at that time suggested that Iowans are split evenly on whether minimum school sizes should be enforced (Des Moines Register 2005b). Survey data on course distributions and after graduation plans, therefore, could prove very important in future education policy debates.

The first survey is a survey of high school seniors about their experiences in high school and plans after graduation. The Department of Education wants to learn if there are any predictors that might indicate what a high school student will do after graduation. They want to know what influences whether a graduate plans to enter the work force, join the military, enroll in a two or four year college, or pursue one of the many other options available to high school graduates. The best way of obtaining this information, the Department of Education decided, is to conduct a survey of high school seniors across the state of Iowa, asking seniors what their plans are for the following year. They will compare a student's future plans with the student's GPA, the size of the student's high school, and other information. A follow-up survey will gather information a year later from these same students to ask what they actually did after graduation from high school. Staff members at CSSM helped with wording of individual questions and the organization of the survey instrument (the whole layout of survey questions) and will be involved with data entry.

This survey will be administered to school districts based on a pre-set schedule of reporting to the state. As such, it is not a probability sample, because a random device was not used to randomly select the districts and schools to participate. Instead it is a convenience sample based on an established administrative schedule. As such, the year to year samples could be affected by statistical bias. The districts chosen in the first year of the study, for example, could be systematically different in terms of student experiences and plans from the ones in the fifth and last year of the study. Random sampling would not eliminate all differences across years and samples, but on average there would not have been any bias. The information likely will still be useful to the Board of Education, because each AEA and districts of small, medium, and large sizes are represented in the schedule each year. AEA location and district size could affect opportunities in high school and typical post graduation plans. Reports that make general statements about the entire student high school population, however, should be somewhat cautious in their claims due to the fact that there are known to be differences in sample composition across years.

The second survey is a survey of student transcripts and is the focus of the rest of this article (see also Hewitt and Larsen 2005). In this survey, students will not provide information directly. Instead, the data gatherers will examine student transcripts and record the types of courses taken in grades 9 through 12. Since districts can give their courses unique names and organize their course offerings as they wish, recording transcript information is not as easy as it might seem. First, for each district in the sample, a data collector has to examine the courses in the course catalog for the district and note which of the courses are of interest to the study. Second, for each student in the sample, the courses of interest on the transcript are recorded. Although the transcript study will not rely on students giving honest answers to questions, carefully coding the courses and thoroughly examining transcripts will present challenges for data collection.

One of the main statistical tasks for CSSM was to help Iowa's Department of Education select which high school seniors to survey and which transcripts to examine. That is, CSSM was responsible for developing the sampling methodology for the surveys. For the transcript survey it was possible to implement a sophisticated probability sampling scheme. The rest of this paper will report on the process of selecting the sampling method for the Iowa Public Schools Transcript (IPST) survey. Simulation will be used to illustrate the impact of different sample design choices.

3. Studying the Population by Sampling

There are countless different sampling methods and no one sampling method is best for all surveys. In this sense, choosing a sampling method is like buying a car —a small two-seat sport convertible is a great automobile, but to a family of seven it is not very useful. Such a large family would be better off purchasing a less stylish station wagon since it would better fit its needs. Likewise, when designing a survey one needs to select a sampling method that will best fit the population that the survey intends to study. Thus the first step in selecting the sampling method for the IPST survey is to learn more about the public school system in Iowa and the needs of Iowa's Department of Education.

The Iowa school districts are divided by geographic location into twelve Area Education Agencies (AEAs). The AEAs divide the state's 370 school districts into more manageable groups: see the map at http://www.state.ia.us/educate/aea/map.html. So, instead of overseeing each school district independently, the Department of Education oversees the AEAs, and the AEAs oversee the individual school districts.

No two AEAs are exactly the same. Not only do the AEAs have their own separate administrators and policies, they also vary drastically in the number of schools and students served. For example, there are 3,436 students enrolled in AEA 4 in the Northwest corner of the state and almost ten-times that many, 32,819, enrolled in AEA 11, which includes the capitol Des Moines. The median number of students per school also varies across AEAs ranging from 130 students to 328. Yet there is as much variation in school size within each AEA as there is between the AEAs. The standard deviation of school size within an AEA can reach as high as 500 students. Table 1 summarizes the number of high schools and students by AEA in Iowa for 2003-2004. Within each AEA are several districts. There are 364 public high schools in Iowa used in this study. Most districts in Iowa have only one high school, but the eleven largest districts have multiple schools. Some geographic areas house primarily rural populations at great distances from cities, whereas others include urban centers.

For each student, the transcript study will record the number of classes of certain types that are taken by the student in grades 9 through 12. The actual transcript study will focus on employment preparation classes taken by general education students and by students with disabilities following individualized education programs (IEPs; see http://www.state.ia.us/educate/ecese/cfcs/iep/index.html). Members of Iowa's State Board of Education were responsible for identifying the courses in selected districts that qualify according to their interest. Of course, surveys such as this could concentrate on a variety of other topics. An alternate interest could have been the number of college preparation or AP classes. College preparation classes could include, for example, advanced literature and language classes, trigonometry and calculus, and science courses beyond the initial high school courses. Another interest could have been job and life preparation classes, such as classes in personal finances and accounting, citizenship issues, workshop and automotive courses, and business computing. Although it would have been interesting to study a variety of topics, doing so would have added a lot of complexity, time, and cost to the data collection process.

4. Sampling Methods

The second step in buying a car, once one has identified what type of car is needed, is to shop around in order to compare the pros and con of the vehicles that are available. Likewise, the second step that one can take when selecting a sampling method for the IPST survey is to compare a variety of different common sampling methods. Three common sampling methods are simple random sampling, stratified random sampling, and cluster sampling.

Simple Random Sampling

There are two types of simple random sampling: simple random sampling with replacement (SRSWR) and simple random sampling without replacement (SRS). SRSWR allows for individual units from the target population to be included in the sample more than once, whereas SRS ensures that the sample contains no duplicates. Because one generally does not gain any additional information by interviewing the same student twice, SRS is preferable to SRSWR for the Iowa schools surveys.

SRS is like picking slips of paper out of a hat: every unit in the population is listed a piece of paper and each piece of paper is just as likely to be selected as the next. Once a piece of paper is removed from the hat, it is not returned to the hat, therefore it cannot be pulled out again. Every possible combination of drawing n slips of paper is equally likely. The method of SRS is designed to protect a survey from selection bias by randomly selecting the sample with equal probability to any other possible sample (Scheuren 2004).

Even though SRS is designed to avoid selection bias, it is possible (although unlikely) to get a very unrepresentative sample. In the case of the IPST survey, under SRS it would be possible to select every unit from AEA 1, and neglect to sample any other AEA. Even more likely is the possibility that a sample selected by SRS would completely neglect a small AEA.

Neglecting a subset of the population due to random non-selection may not be a problem, depending on what the statistician wants to learn about the population and on the characteristics of the neglected sub-population. For example, we could safely exclude every student with green eyes from our sample of high school seniors *if* we knew that eye color is completely unrelated to a person's plans after high school. In this case we assume that the characteristic traits the statistician is interested in are the same within the neglected sub-population and within the sample. If the neglected sub-population is different from the sampled population on a characteristic of interest, however, then the statistician wants to ensure that the sub-population is accurately represented in the sample. To ensure that main sub-populations are represented in a survey, one can use a sampling technique called stratification.

Stratification

Stratification breaks the population of size N into H distinct sub-populations called strata. Suppose that the size of stratum h is N_h and $N_1 + N_2 + \cdots + N_H = N$. A statistician will then (often) take a separate SRS of each individual stratum. Let n_h be the sample size in stratum h ($n_1 + n_2 + \cdots + n_H = n$). So, going back to the drawing slips of paper out of a hat analogy, with stratification a statistician sorts the slips of paper into meaningful groups—say, by AEA—and puts each group of papers into a separate hat. Then the statistician draws some slips of paper out of each hat.

A statistician can choose how many units to select from each stratum. Some common allocations include sampling each stratum equally or determining what percentage of the population is within each stratum and sampling proportionally. There also are optimal allocations if additional information about the subpopulations is known (Lohr 1999). The population total estimate is simply the sum of the stratum population total estimates. Because the strata are sampled independently of each other, the variance of the population total estimate is the sum of the individual stratum variances (Lohr 1999).

Stratification's obvious advantage over SRS is that it ensures that every sub-population of interest (every stratum created) is sampled. For the IPST survey, this means that we can ensure that every AEA is represented within our sample. Because stratification ensures that every stratum is sampled, sampling by stratification also allows the statistician to compare strata to see if there are any differences across sub-populations. So, not only would we be assured that every AEA is represented in our sample, stratification would allow us to compare characteristics across AEAs—which is one of the goals the Department of Education has for this survey.

The other big advantage of stratification is that it often produces more precise estimates of population characteristics than does SRS. If the stratification breaks the target population into sub-populations in which the characteristic of interest is more homogenous than over the entire target population, stratification will produce a more precise estimate of the characteristic than SRS. Taking the means of the samples from more homogenous sub-populations reduces the variance the statistician encounters since the total variance within each stratum is lower than the variance over the entire target population. Lower total variance yields more precise estimation. The more homogenous the subpopulations are, the lower the total variance (using stratification).

Stratification seems to be an obvious sampling design to use in the IPST survey — if not for all surveys. Stratification, however, does have at least one disadvantage compared with a convenience sample. Suppose one collected data from students in Des Moines (a large district), Ames (a medium district), and Zearing (a small district), all located in the center of the state not far from Interstate 35. Would you believe these results are representative of the state? Iowans probably would not be satisfied. But it would be much less expensive than traveling all over the state. Stratification requires one to sample from every stratum and in practice it can be very difficult and costly to gather information from every stratum. It may increase necessary travel time, the operational cost of locating the units sampled, or the start-up costs (permission slips, negotiating with school officials, etc.) in many different locations. A sampling method that is often used to reduce cost is cluster sampling.

Cluster Sampling

A statistician taking a cluster sample will first select sets of units, such as households or schools, rather than selecting individual units. The elements selected in this first stage of the sampling process are called the primary sampling units (PSUs). Then the statistician typically will sample units (students) within each PSU. The statistician could sample every unit within the PSUs (this would be a one stage cluster sample). The units selected in the second stage of the sampling process are called secondary sampling units (SSUs). The statistician could continue to use cluster sampling, and take sampling units beyond SSUs if the population is organized in such a hierarchically structured manner.

Cluster sampling has two big advantages over SRS and stratification. First, cluster sampling is often less expensive and easier to conduct, because it does not require that every sub-population be sampled, like stratification does. This reduces travel time and can reduce start-up costs by sampling more units within a smaller geographical location. In the IPST survey, for example, if we take schools to be our PSUs and students to be SSUs, then we would need to travel to fewer schools to obtain the same size sample. We would likely also reduce start up costs since it would probably be less expensive to meet all the regulations (such as sending out permission slips, etc.) at a few schools than at many schools.

Cluster sampling's other advantage is that it does not require that the statistician have a list of the entire target population. The statistician can make do with a list of all the PSUs, select from this sampling frame, and then obtain a list of all the possible SSUs within the PSUs selected. In other words, we do not need to list all the students in the state before choosing school districts; we can list the students of only the selected schools.

What one gains in cost savings with cluster sampling, however, one might lose in precision. Sampling 100 students from one school does not provide as much information about the target population as if we were to question 100 students across the state. Students from the same school would likely give more similar responses than the randomly chosen students on topics such as distributions of courses and graduation plans. This explains the loss in precision when using cluster sampling over SRS or typical stratification sampling methods. See Flanagan-Hyde's (2005) article in *Stats* for another example of this result. See Richardson and Gajewski (2003) for an activity on this topic.

5. A Simulation Study

Since preliminary data from a pilot study or past survey are not available, population data were simulated in order to compare sampling strategies. In four years of high school, suppose that a student takes fifty-six classes: seven classes per semester for eight semesters. For the purposes of the study, we will think about the number of AP/college preparation classes a student takes in four years of high school. A Poisson distribution is a discrete distribution that takes on non-negative integer values. Table 2 gives Poisson probabilities when the mean number of classes is 5. Four college prep courses during high school, for example, could be accomplished by taking two advanced courses in fall and spring semesters of the senior year. For each school, a Poisson mean value was generated using a linear model. Schools with larger sizes were assumed to have more AP/college preparation classes available, so the means were generated proportional to the school size. The largest and smallest means were close to seven and three, respectively. The number of classes for each student was generated according to the school's Poisson distribution. Values of school means and students' number of classes were generated independently.

Several sampling methods were applied to the simulated population. The methods were simple random sampling (SRS), stratified random sampling with equal and with proportional allocation to strata, single stage cluster sampling with simple random sampling of clusters and with probability proportional to size (PPS) sampling of clusters, and two-stage cluster sampling with PPS sampling of clusters and SRS selection of students within clusters. Formulas for estimating means and standard errors of means are presented in Lohr (1999). Sample sizes of students for each sampling method were controlled to be approximately the same. Samples of size 1200 and 4800 were taken from each population. One hundred populations were randomly generated. Sampling was implemented using each method on each population. Replication was used in the simulation study so that the performance of the methods on average could be compared. All methods that were used produce unbiased estimates of the mean number of college preparation/AP courses, so comparisons are based on margins of error. Margins of error are approximately two times the standard error of the estimates. The average margin of error for each sampling method is depicted in Figure 2.

Simple random sampling (SRS) with sample size 1200 produced an average margin of error of approximately 0.18. When sample size was 4800 (4 times as large), the average margin of error was approximately $1/2 = \sqrt{1/4}$ as much (0.09). Stratified sampling with proportional allocation does a little better on average than SRS, but stratified sampling with equal allocation actually is worse when estimating the state average. This result occurs because sampling an equal number of students within each AEA means that large AEAs have a proportionally smaller representation per student and AEAs with larger variability are sampled the same amount as those with much smaller variability. In other words, resources are used in ways that are not very efficient for estimation at the state level. One advantage of stratification with equal allocation is that estimates of individual AEA averages have more uniform precision. Proportional allocation gave some small AEAs very small sample sizes and consequently large margins of error. The School Board would have to weigh the competing interests, estimates at the state level and estimates for each AEA, if choosing between these two designs.

Of course, SRS and stratified random sampling are not very practical in this study due to the need to review the course catalogs in depth for each selected district, arrange cooperation with each selected district, and travel to the schools. Cluster sampling therefore has significant practical advantages. As can be seen in Figure 2, single stage cluster sampling of schools, however, produces a very large margin of error on average. This method can result in the selection of very few schools, because some high schools in Iowa are attended by a few hundred seniors. Valley West high school, for example, has more than 400 seniors each year. Collecting data at only a few schools greatly increases the uncertainty when estimating the state average. Two stage cluster sampling allows one to sample more schools, but still reduce the workload from that of SRS or stratified random sampling. In simulations, taking up to 100 students per school, two-stage cluster sampling produced average margins of error slightly larger than those of the best two methods. The gain in convenience and decrease in cost for two-stage cluster sampling might be worth the slight loss in precision.

6. Actual Design for the Transcript Survey

The transcript study was implemented in the spring and summer of 2005. It used a stratified, two-stage cluster design. The strata were defined by two factors: district size (small, medium, and large) and AEA. Five AEAs did not have any large school districts, so there were 31 strata. The plan was to select two school districts per stratum. At least two observations are needed to directly estimate a variance. Within each selected district a sample of student transcript are being coded and reviewed. A selection of students in grades nine and twelve from both general education and special education populations are being studied. In small and most of the medium districts, data are being recorded for all students. Due to the start-up cost of coding the district course catalog and the heterogeneity of students within most schools, it was decided to review several student records per school. In some medium and the large districts, a sample of transcripts is included in the study. It was required by the Board that all schools in selected districts be included in the study, so it would not appear that any school in a selected district was being treated differently than any other. If schools within districts had been selected, the stratified two-stage survey design would have turned into a three-stage design.

Data should be available for analysis in the fall of 2005 and winter of 2006. At the conclusion of the transcript survey and the senior exit interview survey, the State of Iowa should have available to it concrete and precise information from a representative sample of schools and students in Iowa. The Board of Education should be able to address more rigorously questions of equality of opportunity and factors influencing plans after high school graduation for students in Iowa.

Acknowledgments

Kyle Hewitt graduated from Lawrence University in the spring of 2005 and began graduate school in statistics at Iowa State University (ISU) in the fall. He participated in ISU's summer research experience program for undergraduate students, which was supported by National Science Foundation grant DMS 0091953 in the VIGRE program. Michael Larsen is an assistant professor in ISU's Department of Statistics and Center for Survey Statistics and Methodology. This work is supported in part by Iowa's State Board of Education.

References

Des Moines Register. (2005a). Iowa must set graduation requirements. June 12, 2005.

Des Moines Register. (2005b). School size opinions. April 25, 2005.

Flanagan-Hyde, Peter. (2005). Clustering sampling: high quality information at a bargain basement price. *Stats*, 42, 20-23.

Hewitt, K., and Larsen, M.D. (2005). Design of a Survey of High School Students in Iowa: Studying Design Choices Using Simulation. Submitted to *Stats* magazine.

Higgins, T. (2005). School size minimums dropped from bill. *Des Moines Register*, April 29, 2005.

Lohr, Sharon. (1999). *Sampling: Design and Analysis.* New York: Duxbury Press.

Richardson, Mary, and Gajewski, Byron. (2003). Archaeological sampling strategies. *Journal of Statistics Education* Volume 11, Number 1, [http://www.amstat.org/publications/jse/v11n1/richardson.html]

Scheuren, Fritz. (2004). *What is a survey?* [http://www.whatisasurvey.info]

Figure 1: Number of public high school students per school in Iowa, 2003-2004.

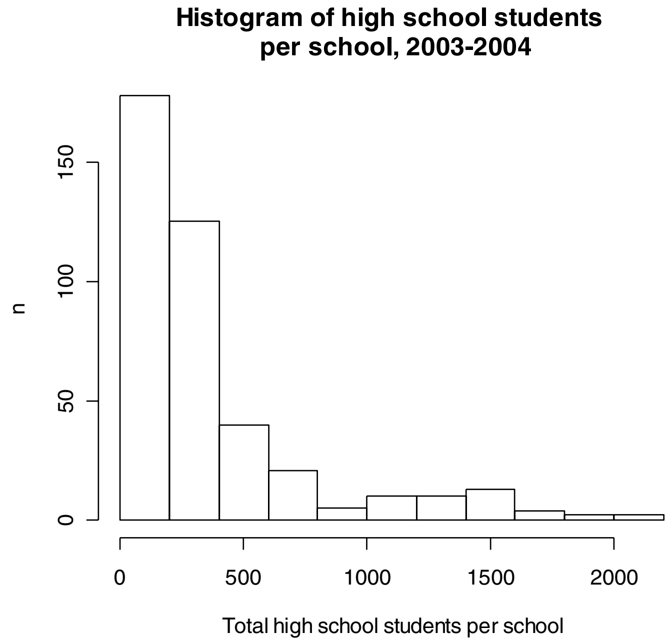

Figure 2: Mean margin of error for sampling methods applied to simulated data, 100 replications.

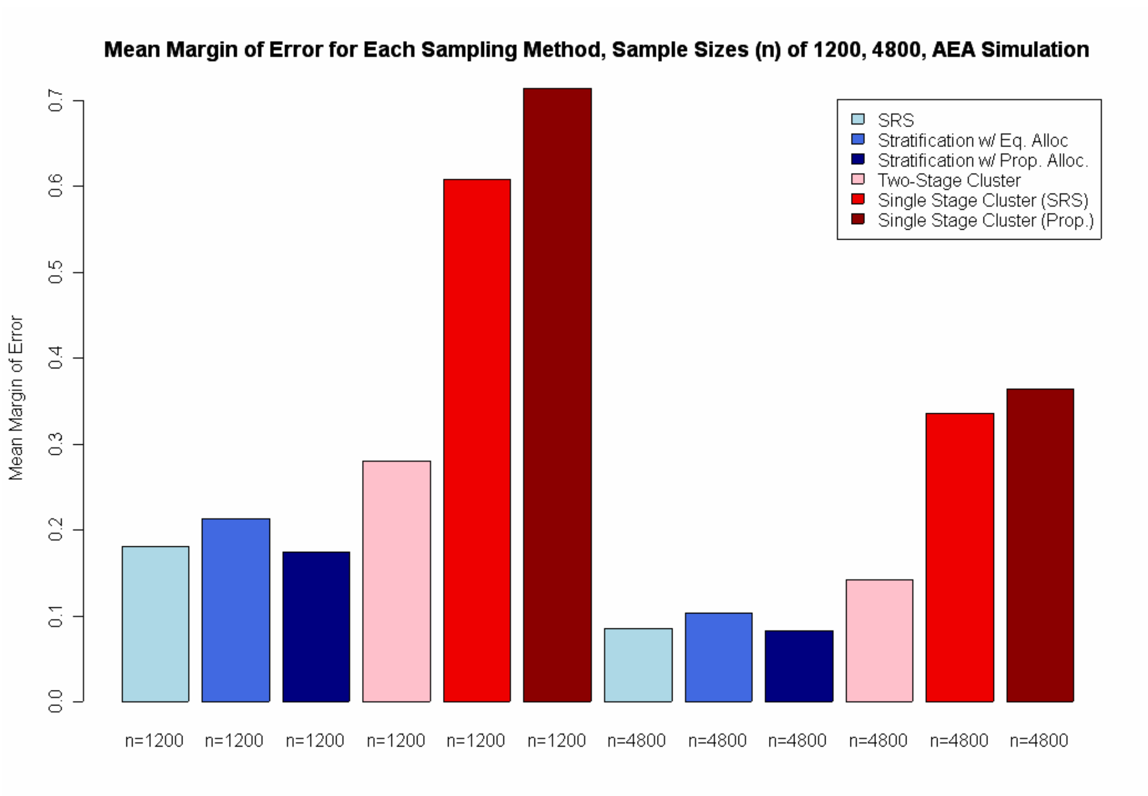

Table 1: Number of high schools and students by AEA in Iowa, 2003-2004.

AEA	Number of high schools	Number of students					
		Total	Mean	Standard Deviation	1st quartile	Median	3rd quartile
1	27	11110	412	377	182	289	450
4	13	3436	264	86	198	247	309
8	47	11350	242	234	110	167	290
9	26	15581	599	582	177	305	1057
10	38	19311	508	501	190	328	545
11	70	34599	494	506	162	293	642
12	26	9429	363	397	157	202	336
13	39	10298	264	305	102	166	267
14	20	3558	178	138	101	130	237
15	25	7143	286	298	109	202	374
16	15	5353	357	337	156	232	518
267	64	21397	334	346	159	232	322

Table 2: Poisson probabilities for λ equal to 5. Outcomes above 11 have probabilities less than 0.01 total.

Value	0	1	2	3	4	5	6	7	8	9	10	11
Probability	.01	.03	.08	.14	.18	.18	.15	.10	.07	.04	.02	.01