Sampling Design in Education

STAT 726 Fall 2021
Lin Rong Shao, Reina Li

Contents

(click on link to quickly jump to section)

1. Introduction

Researchers in the industry of education have different research goals and questions that require different sampling techniques, which also brings sampling difficulties. Censuses are not very common in educational research because it is often difficult to study large groups and is often expensive in cost and time. Therefore, a subset of the population is commonly used for the study; however, there is not a single, correct method to conduct the sampling. Different sampling techniques used in educational research include purposeful sampling, stratified sampling, and convenience sampling.

In the industry of education, schools have always been the popular setting for researchers to conduct their studies. Students and educators also are excellent participants in their educational research. However, researchers would also encounter difficulties when sampling in the field of education, including general challenges, psychological and social risks, as well as numerical difficulties.

2. Main Sampling Techniques

It can be difficult to conduct a census, where all members of a population are treated as subjects, not only in the field of education, but also in other fields. Censuses are not very common in educational research because it is often difficult to study large groups and is often expensive in cost and time. Also, research goals and questions, and contextual feasibility varies. Therefore, a subset of the population is commonly used for the study; however, there is not a single, correct method to conduct the sampling. Different sampling techniques discussed in this paper, such as purposeful sampling, stratified sampling, and convenience sampling are used in educational research.

2.1. Purposeful sampling

A purposeful sample is conducted by selecting specific subjects from the population that will help answer or be beneficial to the research question and goal the most. This type of sampling is common in qualitative research. In qualitative research, it is more important to select participants who can best understand the aim of the study, and not those who represent some larger population. On the other hand, generally different, smaller, and less representative samples are selected for quantitative research because the two approaches have different goals. Four forms of purposeful sampling include informant sampling, extreme case and intensity sampling, quota sampling, and snowball sampling.

An informant sample is conducted by selecting specific "informants" from the population based on their experience and skills that will help with answering the research questions and goals. Conducting an informant sample will be most useful when the researcher wants to develop a deep understanding of the topic from experienced and skilled people. However, results from an informant sample will not be representative to the whole population and will depend on the criteria used to select subjects. An example of informant sampling is when interviewing students representing minority ethnic groups, you should sample 10 students from the target ethnic group.

An extreme case sample is conducted by selecting subjects from the population that are from opposite ends or contradictory sides in order to develop an understanding of the differences in experiences, opinions, and characteristics. This type of sampling requires the researcher to know what the extremities are before selecting the subjects. Extreme case sampling allows the researcher to show a range in their sampling. In an intensity sample, the researcher is more interested in identifying the diversity of the sampled subjects. This type of sampling is less focused on the extremities. An example of extreme case and intensity sampling is when interviewing students representing both majority and minority groups, you should sample 5 students from the minority group and 5 students from the majority group.

A quota sample is conducted by selecting a specific number of subjects from predetermined groups in order to account for the differences between the groups. This type of sampling requires the researcher to make assumptions about what types of differences between subjects are important, and the researcher will treat each group of subjects like a separate population. The overall representation of the subgroup in the larger population is not important. The goal of a quota sample is to make sure that each group is sampled sufficiently. Quota sampling is useful if the researchers have categories of analysis they want to focus on for their research questions and if researchers are able to identify subjects based on those categories they selected. If the categories specified by the researchers are poorly defined or unnecessary to the research questions and goals, then quota sampling is less useful. The key to quota sampling is to ensure categories are important, accurate, and clear to the research questions and goals. An example of quota sampling is when interviewing the same number of students from each target ethnic group, you should sample 2 students from each of the 5 different ethnic groups.

A snowball sample is conducted by selecting a small number of initial subjects and then slowly growing the sample by additional sampling. This type of sampling is useful if researchers have difficulty identifying or recruiting subjects in the target population. Snowball sampling allows the researcher to use the subjects' knowledge and social networks as sampling tools. However, snowball sampling may not be representative of the whole population. An example of snowball sampling is when interviewing a student representing a target ethnic group, and then asking the interviewed student to help identify other students in the target group and so on, you should sample 1 student from the target group, 2 of that student's friends, and 7 of their friends.

2.2. Random sampling

A random sample is conducted by selecting subjects randomly to make sure the results are representative of the population. If the research goal is to generalize results from the sample for the population, random sampling methods must be used; otherwise, there may be bias in the study. Studying a random subset of the population can generalize results that represent the population, and makes research studies possible that would otherwise be impossible due to high costs, etc. Three forms of random sampling include simple random sampling, stratified sampling, and cluster sampling.

A simple random sample is conducted by selecting subjects from the target population where each subject has an equal probability of being selected. Simple randomization is great for researchers to show that the results are representative of the population. However, this type of sampling often does not provide enough data to make specific analyses. Simple random sampling is an easy method to quickly and efficiently collect generalizable data. An example of simple random sampling is when surveying random teachers from a state-level list of teachers, you should sample 1,000 teachers.

A stratified sample is conducted by selecting subjects from different strata in the population. In a stratified sample, results will account for important variation in the population. This type of sampling requires the researchers to identify the stratification variables and then they have to sample an appropriate amount of subjects from each stratum. This process allows the sample to be more diverse and representative of the larger population. Stratified sampling is more useful than simple random sampling when the researchers want to be able to compare results between groups. The sample size for each stratum needs to be sufficient in order to generalize results. However, stratum-based results, especially for strata with smaller populations, may not be representative to the entire population. An example of stratified sampling is when surveying teachers from each type of school, you should sample 334 teachers from traditional public schools, 333 teachers from private schools, and 333 teachers from charter schools.

A cluster sample is conducted by selecting subjects from “clusters” of the population. The clusters should be similar to the larger population in terms of characteristics and internal variance within the cluster. This type of sampling requires the researcher to identify clusters and then to randomly select which clusters to study. A major benefit of using cluster sampling is the ability to study subjects in clusters in finite settings. However, clusters must be representative of the population; otherwise, there will be problems in the results if clusters are different in size and have different variance from other clusters. An example of cluster sampling is when surveying all teachers from randomly selected schools, you should sample 300 teachers from school A, 300 teachers from school B, and 400 teachers from school C.

2.3. Convenience sampling

A convenience sample is conducted by selecting subjects that are “convenient”, easy to study, or close in location. In convenience sampling, data collection is easy. This type of sampling is common in user testing, design settings, and other situations that use less-formal research methods. Convenience sampling is the easiest type of sampling to do when compared to purposeful sampling, random sampling, and census, and is most commonly used in teaching and design scenarios. While this type of sampling is great for gathering data quickly, convenience sampling is the least accurate and most problematic when compared to other sampling techniques. The results are also not generalizable and representative of the population and also does not account for the variations in the population. Convenience sampling is useful for designing the research questions, as well as iterative designs of instruments, tools, and products.

3. Main Sampling Difficulties

In the industry of education, schools have always been the popular setting for researchers to conduct their studies. Students and educators are excellent participants in educational research. However, researchers would also encounter difficulties when sampling in the field of education, including general challenges, psychological and social risks, as well as numerical difficulties.

3.1. General Challenges

One of the most challenging issues to consider when conducting school-based research is accessing the field. One of the important duties for researchers is to maintain contact with the school and organize the data collection there. Sometimes it can be problematic when recruiting and accessing schools, especially when it is difficult to gain the contact with the relevant person to pass on the message.

It is crucial to gain informed consent from all relevant authorities when taking educational research. When the sampling is taking place in the school setting and children get involved, it often requires the approval from the school principal and class-head teacher. Additionally, it also requires the consent from children themselves, also their parents or guardians. Besides that, there is an Opt-out Right for children, which means parents could withdraw their children from the study at any point and if children themselves don’t want to participate anymore they could quit at any time. Such opt-out procedures often cause the failure of the research.

As researchers, it is also their responsibility to ensure the participant is not experiencing any kind of suffering or loss in the process of the study. Sometimes sampling is required to collect the data around the sensitive topics such as mental health or wellbeing, so there may be a chance of the sensitive or potentially harmful information being exposed to the participants. Once that situation happened, participants could choose to not complete or withdraw from any data collection around the sensitive topics at any time.

3.2. Psychological and Social Risks

When conducting research in school settings, it is common for participants facing psychological and social difficulties. Emotional upset or destabilization, emotional fatigue, and emotional dependency are the potential psychological risks. Among these risks, emotional upset would be the most significant risk and is often caused by the following situation: 1. Confrontation with particular emotionally evocative material. 2. Confrontation with age-inappropriate material. 3. Confrontation with sensitive topics. 4. Confrontation with topics related to personal difficulties or difficult life experiences. 5. Triggering traumatic memories. Basically, all the above situations highlight the fact that sometimes the topic of the study may not be suitable for the children participants (such as underage smoking, alcohol drinking), or the context of the studies (for example, school violence) may remain the bad experiences in their personal life which cause the negative emotional impacts to the participants and make them suffer from psychological difficulties.

Social risks arise when studies take place within an established peer group or the method of the research relies heavily on social interaction within the peer group. Social risks may include: 1. Being singled out. 2. Embarrassment in front of other students. 3. Change of image within the peer group. 4. Loss of status within the peer group.

Unfortunately, psychological and social risks are complicated and impossible to avoid, as a consequence the result of studies may lose some degree of accuracy and precision.

3.3. Numerical Difficulties

Walter S. Monroe had argued in his paper that in the view of the numerical text on statistical methods applied to educational research, researchers would find difficulties to apply appropriate formulae in treating their data, which may result in an unjustified interpretation or imprecise analysis of the statistical data. For example, one study using the stratified sampling for data collection process, however when doing the data calculation, the researcher uses the formula for simple random sampling not the one for stratified sampling, which would cause an error in research results. Moreover, numerical difficulties are also encountered when interpreting the coefficient of correlation in the industry of education. Coefficients of correlation are more likely to be used in the field of biology, economics and sociology, later it also gets involved in the field of education, but the required correlation technique for educational research is different from the technique used in other fields. In those industries, researchers focus more on the degree of the relationship that exists between two sets of data, while in education, research should focus more on the degree of the perfection of a relationship that is already known to exist.

4. Examples

4.1. PISA 2015 Technical Report

The Programme for International Student Assessment (PISA) in 2015 measured how well 15-year-old students in school, in grades 7 and higher, can use their knowledge and skills to meet real-life challenges, instead of mastering a specific school curriculum. The 2-hour assessment was conducted in 35 OECD countries and 37 partner countries and economies, and approximately 540,000 students completed the assessment, representing about 29 million 15-year-old students in the 72 participating countries and economies. The main mode of assessment was computer-based; however, paper-based alternatives were provided for countries without resources for computer-based testing. PISA 2015 covered reading, math, science, collaborative problem solving and financial literacy with a primary focus on science, through multiple choice and open-ended questions.

The target population was 15-year-old students in school in grades 7 and higher, which included full-time and part-time students, students in vocational training or related educational programs, and students attending foreign schools within the country. Schools as a whole were used as the primary sampling unit.

A two-stage stratified sampling design was used. The first-stage consisted of individual schools that had 15-year-old students. Schools were sampled systematically using the international contractor sampling software, KeyQuest, from a national list of all PISA-eligible schools sorted by grade and gender, which was the sampling frame, and the probabilities were proportional to the estimated number of PISA-eligible 15-year-old students enrolled in the school (ENR). This is called systematic probability proportional to size (PPS) sampling. The second-stage consisted of students within sampled schools. After schools were selected to be in the sample, a list of each sampled school's 15-year-old students was created. Each country had a target cluster size (TCS) of 42 students for computer-based countries and 35 students for paper-based countries. From each list of students that contained more than the target cluster size, a sample of around 42 students were selected with equal probability, and for lists that contained less than the target cluster size, all students on the list were selected.

A minimum of 150 schools were selected in each country. All schools were selected if a participating country had fewer than 150 schools.

Initially-selected schools had a required response rate of 85%. A school with student participation rate between 25% and 50% was not considered as a participating school, but the data were included in the database and contributed to the report. Non-respondents consisted of schools with student participation rate less than 25%.

Before sampling, schools were stratified in the sampling frame using explicit and implicit stratification. Explicit stratification variable examples include sorting by states or regions within a country. Implicit stratification variable examples include type of school, school size (very small, moderately small, or large), urbanization, or minority composition.

A measure of size (MOS) was assigned to each school: $MOS = max(ENR, TCS)$. MOS was equal to the enrollment estimate (ENR) if ENR was greater than the target cluster size. MOS was equal to the target cluster size (TCS) if ENR was less than TCS. MOS of small schools was set to 42 students for computer-based, and 35 students for paper-based, so small schools would have an equally likely chance of being selected to participate.

PPS-systematic sampling method was used to determine which schools to sample. Calculations include: total measure of size, S, for all schools in the sampling frame for each explicit stratum; number of schools, D, to be sampled from the explicit stratum; sampling interval, $I = S/D$; identifying all schools in the sample where the school size exceeded I and removing them from the frame and then recalculating S, D, and I. Then a random number (RN) from a uniform distribution between 0 and 1 was generated for each explicit stratum. Then, selection numbers were calculated for each explicit stratum, one for each of the D schools to be selected in the explicit stratum. The first selection number was used to identify the first sampled school in the explicit stratum $= I * RN$. The next selection number was used to identify the next sampled school $previous\ selection\ number + I$, and so on, until all had been assigned a selection number. Selection numbers were generated independently for each explicit stratum, with a new RN generated for each explicit stratum.

Let Z be the first selection number for a particular explicit stratum. Find the first school in the sampling frame where the cumulative MOS (C_s) was equal to or greater than Z. This was the first sampled school. This rule was applied to each selection number to generate the original sample of schools for that stratum.

For example, suppose that in an explicit stratum in a participant country, the PISA-eligible student population is 105,000, then:

$S = 105000$

$D = 150$

$I = S/D = 105000/150 = 700$

Generated random number, $RN = 0.3230$

First selection number, $I * RN = 700 * 0.3230 = 226$

Second selection number, $(I * RN) + I = 226 + 700 = 926$

...

and so on until all schools on the list have a selection number, resulting in a school sample size of 150 schools.

School	MOS	Cumulative MOS (C_s)	Selection number	School selection
1	550	550	226	Selected
2	364	914		
3	60	974	926	Selected
4	93	1067		
5	88	1155		
6	200	1355		
7	750	2105	1626	Selected
...	...	...	...	...

The small school analysis procedure was carried out to manage small schools in the sampling frame, for each individual explicit stratum. Assume an initial target school sample size of 150 and a target student sample size of 6,300. First, from the complete sampling frame, find the proportions of total ENR that come from very small schools with ENR of 0, 1, or 2 ($P1$); from very small schools with ENR greater than 2 but less than TCS/2 ($P2$); from moderately small schools (Q); and from large schools (R). $P1 + P2 + Q + R = 1$. Calculate $L = 1 + 3(P1)/4 + (P2)/2$. The minimum sample size for large schools is $= 150 * R * L$, rounded up to the nearest integer. Calculate the mean value of ENR for moderately small schools (MENR), and for very small schools (V1ENR and V2ENR). MENR is a number between TCS/2 and TCS. V1ENR is a number between 0 and 2. V2ENR is a number larger than

2 but no greater than TCS/2. The number of schools that must be sampled from moderately small schools is $= (6300 * Q * L)/MENR$. The number of schools that must be sampled from very small schools (type P2) is $= (3150 * P2 * L)/V2ENR$. The number of schools that must be sampled from very small schools (type P1) is $= (1575 * P1 * L)/V1ENR$.

For example, suppose that in a participant country, $TCS = 42$ students, 10% of total enrollment of 15-year-old students in moderately small schools, 5% in P1 type of very small schools, and 5% in P2 type of very small schools. Suppose average student enrollment in moderately small schools is 25, type P2 very small schools is 12 students, and type P1 very small schools is 1.5 students.

$P1 = 0.05$

$P2 = 0.05$

$Q = 0.10$

$R = 0.80$

$L = 1 + 3(P1)/4 + (P2)/2 = 1 + 3(0.05)/4 + 0.05/2 = 1.0625$

Minimum sample size for large schools, $150 * R * L = 150 * 0.80 * 1.0625 = 127.5$

Rounded up to the nearest integer, 128 of the large schools must be sampled.

$MENR = 25$

$V2ENR = 12$

$V1ENR = 1.5$

Number of schools that must be sampled from moderately small schools,
$(6300 * Q * L)/MENR = (6300 * 0.10 * 1.0625)/25 = 26.8$

Rounded up to the nearest integer, at least 27 moderately small schools must be sampled.

Number of schools that must be sampled from type P2 very small schools,
$(3150 * P2 * L)/V2ENR = (3150 * 0.05 * 1.0625)/12 = 13.9$

Rounded up to the nearest integer, at least 14 type P2 very small schools must be sampled.

Number of schools that must be sampled from type P1 very small schools,
$(1575 * P1 * L)/V1ENR = (1575 * 0.05 * 1.0625)/1.5 = 55.8$

Rounded up to the nearest integer, at least 56 type P1 very small schools must be sampled.

Combining them gives a total sample size, $128 + 27 + 14 + 56 = 225$ schools. Before considering school and student non-response, larger schools will give an initial sample of approximately $128 * 42 = 5376$ students. Moderately small schools will give an initial sample of approximately $27 * 25 = 675$ students. Type P2 very small schools will give an initial sample of approximately $14 * 12 = 168$ students. Type P1 very small schools will give an initial sample of approximately $56 * 1.5 = 84$ students. The total expected sample size of students is $5376 + 675 + 168 + 84 = 6303$ students.

4.2. The Importance of Teachers' Emotions and Instructional Behavior for their Students' Emotions – An Experience Sampling Analysis

This study is focused on the relationship between teachers' emotion, instructional behavior and students' emotion in class. There are three hypotheses investigated in this study. Hypothesis 1 is Teachers' discrete emotions relate to students' discrete emotions in class. Second hypothesis is Teachers' instructional behavior relates to students' discrete emotions in class. Last hypothesis is Teachers' emotions are significantly related to students' emotions in class, above and beyond teachers' instructional behavior.

The study is conducted with a convenience sample drawn from eight different schools from the German speaking part of Switzerland. Data was collected from 44 different 9th grade classes and three to four students per class were selected in the study. There are a total of 149 students (55% female) with a mean age 15.63 years old.

Each participant was required to use the handheld device to record their immediate emotional experience and their perception of their teacher's emotion and behavior over a period of 10 school days. Data collection process contains both event-based sampling and random sampling. Students were instructed to activate their device in the beginning of their lesson in four different subject classes (German, English, French, and mathematics) with each lesson lasting for 45 mins. In the next 10-35 mins, the device would randomly pop up the questionnaire that asked them to report their emotion at that moment as well as the perceived teacher's emotion and behavior. In total, students activated their devices in 2890 lessons and completed a total of 2668 questionnaires. In 431 of these questionnaires contain no data for teacher's emotion and instructional behavior, resulting in a final sample of 2230 questionnaires.

There were three types of emotions accessed, angry, anxious and happy, which also rate on a 5 point scale, 1 represents not at all, and 5 represents very strongly. In order to avoid the mood influence the way students perceive their teachers' emotions. They used the control variable "moods before class" and students asked to indicate the mood when they activated their devices before a lesson started. In addition to that, they also distinguished the analysis for each of the subjects. Students indicated the subject when they activated the device (1=German, 2=English, 3=French, 4=Mathematics).

Across all four subjects, enjoyment was the emotion experienced most intensely by study participants during class ($M = 2.85$, $SD = 1.19$), followed by anger ($M = 1.83$, $SD = 1.16$) and anxiety ($M = 1.45$, $SD = 0.93$). The same pattern was found for perceived teachers' enjoyment ($M = 3.01$, $SD = 1.10$), anger ($M = 1.82$, $SD = 1.06$) and for anxiety ($M = 1.76$, $SD = 0.98$). The mean student mood before entering the class was 6.46 ($SD = 1.19$), indicating a slightly positive mood (1 = negative mood, 5 = neutral, 9 = positive mood).

Test of Hypothesis: The result indicates that students and teacher's emotions were substantially related, the strongest relationship was found for enjoyment, followed by anger, then anxiety. The second hypothesis stated that teachers' instructional behavior is related to students' emotions within a lesson. Results indicate that teachers' instructional behavior was related to students' anger and students' enjoyment but unrelated to students' anxiety. Last hypothesis test if teachers' emotions are significantly related to students' emotions in class, above and beyond teachers' instructional behavior. And the study results provide support for the third hypothesis.

Overall, the finding of the study indicates that teacher and student's emotions are closely related.and that teachers' emotions explain incremental variance in students' emotions above and beyond their instructional behavior.

5. Sources

1. Becker, E. S., Goetz, T., Morger, V., & Ranellucci, J. (2014). The importance of teachers' emotions and instructional behavior for their students' emotions – an experience sampling analysis. *Teaching and Teacher Education*, *43*, 15–26. https://doi.org/10.1016/j.tate.2014.05.002
2. Felzmann, H. (2009). Ethical issues in school-based research. *Research Ethics*, *5*(3), 104–109. https://doi.org/10.1177/174701610900500304
3. Kimmons, R. (1970, January 1). *Sampling*. Education Research. Retrieved November 4, 2021, from https://edtechbooks.org/education_research/sampling.
4. Monroe, W. S. (1930). Difficulties in Educational Research. *Educational Research Bulletin*, *9*(1), 1–6. http://www.jstor.org/stable/1470187
5. OECD. (2017). (tech.). PISA 2015 Technical Report (pp. 65–89).
6. Ross, K. N. (n.d.). Sample design for educational survey research. In Quantitative research methods in educational planning (pp. 1–82). essay, UNESCO.
7. Simpson, A. (2021, September 21). *The challenges of conducting school-based research*. Methods Blog. Retrieved November 16, 2021, from https://walesdtp.ac.uk/methodsblog/2019/02/27/the-challenges-of-conducting-school-based-research/.