Exercise 1.1

A large school district plans to survey parents to measure their opinions on various issues. The survey will be done as follow. A random sample of 20 elementary schools, 10 middle schools, and 5 high schools will be drawn: 5 homerooms per elementary school, 10 homerooms per middle school, and 20 homerooms per high school. A self-administered questionnaire and envelope will be placed in the "home communication" folder of each student in selected classes. Parents will be asked to complete the questionnaire, seal it in the envelope, and send it back to school with their child. Is this an EPSEM sampling procedure? What is the probability of any given student being selected?

It depends on some of the unknowns involved in the survey. If all homerooms contain the same number of students across all schools at all levels and there are an equal number of schools at each level, then the sampling procedure is EPSEM. This is proven in the following equations:

$$probability\ of\ selections = \left(\frac{n_{schools}}{N_{level}}\right)\left(\frac{n_{selected}}{N_{students}}\right) where$$

- $n_{schools} = number\ of\ schools\ selected\ at\ particular\ level$
- $n_{selected} = number\ of\ students\ selected\ in\ a\ selected\ homeroom\ at\ a\ particular\ level$
- $N_{level} = total\ number\ of\ schools\ at\ a\ particular\ level$
- $N_{students} = total\ number\ of\ students\ in\ a\ selected\ homeroom$

Elementary: $\left(\frac{20}{N_{level}}\right)\left(\frac{5}{N_{students}}\right) = \frac{100}{N_{level}\ N_{students}}$

Middle: $\left(\frac{10}{N_{level}}\right)\left(\frac{10}{N_{students}}\right) = \frac{100}{N_{level}\ N_{students}}$

High: $\left(\frac{5}{N_{level}}\right)\left(\frac{20}{N_{students}}\right) = \frac{100}{N_{level}\ N_{students}}$

Certain assumptions can be made to assume an EPSEM sampling procedure depending on the demographics for each particular subpopulation, even if N_{level} and $N_{students}$ are different across schooling levels. For instance, if there are a smaller pool of high schools to sample compared to middle and elementary schools but the number of students per homeroom is higher at the high school level, these deviations would offset each other resulting in an equal probability. An example is shown below:

Elementary: $\left(\frac{20}{40}\right)\left(\frac{5}{10}\right) = \frac{100}{400} = \frac{1}{4}$

Middle: $\left(\frac{10}{20}\right)\left(\frac{10}{20}\right) = \frac{100}{400} = \frac{1}{4}$

High: $\left(\frac{5}{10}\right)\left(\frac{20}{40}\right) = \frac{100}{400} = \frac{1}{4}$

Again, this offsetting phenomenon only applies if students per homeroom are held constant across all homerooms at that particular level. Varying homeroom sizes at the same school level will result in varying probabilities across homerooms at that school level.

Exercise 1.2

A university wants to learn about the problems its freshman experience in making the transition to college, so it can design and prioritize programs to keep these students in school. The university plans to gather this information from freshman who sign up for interviews in exchange for extra credit in an Introduction to Psychology class. How do you evaluate this sample in terms of potential coverage bias, selections bias, and nonresponse bias? Overall, is this sample acceptable for the research purpose? Would the university do better if it sent a request to freshmen's e-mail addresses, asking them to participate in the research, and collected the information online?

       The sampling procedure contains some inherent sample bias based on the design stated above, but this bias may actually work in the university's favor. The purpose is to design and prioritize programs to keep students in school so the university should want to target students that may be struggling. Offering extra credit in exchange for an interview is 2-fold is what it accomplishes. It properly introduces selection bias by targeting students struggling because those with poor grades are more inclined to participate and it combats nonresponse bias by giving the incentive of extra credit. There is still potential for students who are not struggling to sign up simply because they would like extra credit, say to go from a B to an A. Therefore, a small degree of unwanted selection bias could occur.

       Other sources of bias come from those who are taking an introductory psychology course. It does not specify is only freshman are allowed to take the course. This would give potential to upper classmen being in the sampling pool. The bias may not be a bad thing, although not as likely there may be upper classmen that are also struggling which can benefit from transition programs. In addition, coverage bias is made around the fact that only those, which are required to take an introductory psychology class, have a chance of being sampled. This may exclude students struggling in other majors like engineering.

       Overall, the sample may need slight modifications before it is acceptable. The overall procedure of offering extra credit in exchange for an interview should remain unchanged, however, I would suggest adding an additional level of selection by picking introductory courses that span across a wider spectrum of majors offered by the university, in turn covering more of the struggling student population.

       I believe an email sent to every freshman at the university would actually perform worse. It does not address the aspect of nonresponse bias. Although it covers a bigger population, there is no incentive to respond to the email and spend time out of a day for an interview. Struggling student would be more concerns with courses that they are struggling in rather than giving answers to an interview.