

SM 625: Week 5 Sampling Project Notes

From a previous study, you obtain estimates of the following **design effects** for each of these three estimates:

- proportion ever smoked one cigarette = 2.5;
- proportion ever smoked marijuana = 2.0; and
- mean age when first asked to smoke = 1.7.

This previous study featured a sample of size $n = 7,500$ students between the ages of 13 and 19, selected from a total of $a = 150$ clusters. Using this information, compute a synthetic estimate of ρ_h for each of the three variables. These synthetic estimates of ρ_h will be used to consider alternative cluster sample designs as you continue with your project work.

Finally, budget and cost information is now available. The total budget for data collection for this project will be \$500,000. The client and the data collection organization estimate that the data collection will cost \$3,000 per primary stage cluster (school), and \$50 per completed questionnaire within a cluster. We will use this cost information moving forward for optimal subsample size calculations.