

# Cluster Sampling

## Bullying Behavior

Consider a (one-stage) cluster sampling design using a population of 2000 second grade students in a school district. The researchers want to estimate the number of incidents of bullying behavior among these students during a school year. But because special training and protocols are necessary to help teachers recognize bullying behavior, the researchers decide to select a sample of students by selecting a sample 5 classrooms from the 100 classrooms of second grade students in the district (these are some lazy researchers). The teachers for the sampled classroom keep a record of the number of times each student exhibited bullying behavior. The table below shows the total number of incidents of bullying behavior for each classroom as well as the number of students in the classroom.

Classroom	Incidents	Students
1	0	16
2	4	20
3	2	19
4	3	18
5	7	22

Use these data for the following concerning the estimation of the total number of bullying incidents among second grade students in the school district for the school year.

1. Assume that the researchers selected classrooms using *simple random sampling*. Confirm that estimate of the total number of incidents using the *unbiased* estimator is 320, and that the bound on the error of estimation of this estimator is approximately 226.
2. Assume that the researchers selected classrooms using *simple random sampling*. Confirm that estimate of the total number of incidents using the *ratio* estimator is approximately 337, and that the bound on the error of estimation of this estimator is approximately 195.
3. Assume that the researchers selected classrooms using *sampling with replacement with selection probabilities proportional to classroom size*. Confirm that the estimate of the total number of incidents using the Hansen-Hurwitz estimator is approximately 316, and that the bound on the error of estimation of this estimator is approximately 210.

Note: The bounds on the error of estimation are quite large because the sample size is (probably unrealistically) small.

## Educational Testing

Consider a (one-stage) cluster sampling design using a population of 2000 second grade students in a school district. Researchers would like to estimate the average score of these students on a particular test based on a sample of students since the testing is expensive. They decide to use a cluster sampling design where they select a sample of 5 classrooms of students out of the 100 classrooms in the district (again, these are some lazy researchers). The table below shows the average test score and the number of students in each of the sampled classrooms.

Use these data to answer the following questions concerning the estimation of the mean test score for all second grade students in the district. (Hint: Many calculations for cluster sampling are expressed in terms

Classroom	Mean Score	Students
1	70	16
2	80	20
3	75	19
4	72	18
5	88	22

of cluster *totals* rather than cluster *means*, but cluster totals can be found from cluster means by simply multiplying the mean by the number of elements in the cluster.)

1. Assume that the researchers selected the classrooms using *simple random sampling*. Confirm that the estimate of the mean test score for all the second grade student in the district is approximately 77.65, with a bound on the error of estimation of approximately 6.2376.
2. Assume that the researchers selected the classrooms using *sampling with replacement* with *selection probabilities proportional to classroom size*. Confirm that the estimate of the mean test score using the Hansen-Hurwitz estimator is 77, with a bound on the error of estimation of approximately 6.4498.