1. In the current population survey, an observational unit will be part of the monthly survey for 2 consecutive months, out of the survey for 4 months, and then return for 2 more months. Suppose each month 1,000 new units enter the sample. How many units are in the sample in any month (once the sample has reached steady state)? Of those units how many will still be in the sample in the following month? Explain carefully.

2. Consider the following description of the Gambia survey. How many stages are there? What is the sampled unit in each stage?

   The sampling frame consisted of all rural villages of fewer than 3000 people in the Gambia. The villages were stratified by three geographic regions (eastern, central, and western) and by whether the village had a public health clinic (PHC) or not. In each region five districts were chosen with probability proportional to the district population estimated in the 1983 national census. In each district four villages were chosen, again with probability proportional to census population: two PHC villages and two non-PHC villages. Finally, six compounds were chosen more or less randomly from each village, and a researcher recorded the number of beds and nets, along with other information, for each compound.

3. In a water pollution study for the region below, the following sampling design is used. Pairs of random numbers are generated uniformly from the unit square. If the corresponding point falls in one of the lakes in the region then the lake is included in the sample. This procedure is repeated until \( n \) pairs of points have landed in a lake. For each lake sampled, the pollution level is measured.

   a. What is the sampling unit?
   b. Name the sample design.
   c. Provide an unbiased estimator for the average pollution level of the lakes in the region.
   d. Prove that your estimator is unbiased.

4. In a two stage cluster sample of birds, 382 birds were known to be banded. A geographic region was divided up into \( N \) regions (non-overlapping). Then 8 randomly selected regions were surveyed and the number of birds were observed
(total, banded) in each region. The data are (33,20), (13,2), (1,0), (45,15), (21,5),
(82,29), (14,4), (0,0). Provide an estimate of the total number of birds in the
region.

5. Suppose you want to estimate the percentage of persons who have been
immunized against polio in Gilbert (population 59,000) and can take a SRS of
persons. What should be your sample size if you want the estimate to have margin
of error of 4 percentage points for a 95% confidence interval?

6. A researcher wants to study the prevalence of smoking and other high-risk
behaviors among female high school students in a region with 35 high schools.

<table>
<thead>
<tr>
<th>Number of Students</th>
<th>Number of Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-499</td>
<td>3</td>
</tr>
<tr>
<td>500-999</td>
<td>7</td>
</tr>
<tr>
<td>1000-1499</td>
<td>18</td>
</tr>
<tr>
<td>1500-2000</td>
<td>5</td>
</tr>
</tbody>
</table>

She intends to drive to $n$ of the schools and then interview some or all female
students in the selected schools. She has conducted a similar study with 4 schools
out of 29 in another region. The results were as follows:

<table>
<thead>
<tr>
<th>School</th>
<th>Number of Students</th>
<th>Number of Female Students</th>
<th>Number of Females Interviewed</th>
<th>Number of Smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1471</td>
<td>792</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>890</td>
<td>447</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1021</td>
<td>511</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>1587</td>
<td>800</td>
<td>40</td>
<td>27</td>
</tr>
</tbody>
</table>

a. Estimate the percentage of female students who smoke, from the study of the
4 schools.

b. Using information from the previous study, propose a design for the new one.
Suppose it takes about 50 hours per school to make contact with school
officials, obtain permission, obtain a list of female students, and travel back
and forth. Although interviews themselves are only about 10 minutes, it takes
about 30 minutes per interview obtained to allow for additional scheduling for
no-shows, obtaining parental permission, and other administrative tasks. The
investigator would like to spend 300 hours or less on the data collection.