ST 432 Spring 2008
Cluster and Multi-Stage Sampling Summary Comments

1. Cluster and Multi-stage sampling are used for practical reasons because natural populations often have a nested structure. For example, a survey of a city could have structure:
   Blocks (primary units)
   Households (secondary units)
   Individuals (tertiary units)
   At each level the units are nested within the level above. In class we have focused on two-stage sampling because it is the simplest, but in practice there may be three or more stages as shown.

2. If one exploits the nested structure of the population, then one may be able to sample a larger total number of units for a fixed cost than under simple random sampling of all the units (ignoring the nesting). Typically a simple random sample would be more efficient than a cluster sample if they were of the same size, because usually members of the same primary unit (cluster) are positively correlated with each other.

3. Systematic random sampling is a special form of cluster sampling and we will consider that next in lectures.

4. The equations for two-stage sampling in Chapter 13 are simple generalizations of the results in Chapter 12 for cluster sampling. The generalization is necessary because only some of the secondary units are sampled.

5. The variance equations in Chapter 13 involve components of variance for each stage of sampling. They can be quite complex if there are not equal numbers of units at each stage. Variance estimation typically requires that at least two units be sampled at each level.

6. We emphasized simple random sampling of the primary units and simple random sampling of the secondary units. It is also possible to use probability proportional to size sampling or stratified random sampling of the primary units and systematic random sampling of the primary or secondary units (although the variance estimates may be very complex).