Sampling Designs

• 1. **Simple random sampling (SRS)**

  Steps:
  
  – (1) Assign a single number to each element in the sampling frame.
  
  – (2) Use random numbers to select elements into the sample until the desired number of cases is obtained.

• The method is not very different from winning a lottery.
2. Systematic Sampling

• Steps:
  – (1) Calculate the sampling interval as the ratio between population size and sample size, \( I = \frac{N}{n} \).
  – (2) Arrange all elements in the population in an order.
  – (3) Select a case in the first interval randomly.
  – (4) Select every \( i^{th} \) case from this point.
2. Systematic Sampling (continued)

- Systematic Sampling is easier and simpler than SRS
- The text warns of a danger to this method. What is it?
3. Stratified Sampling

- Stratified sampling is more complicated than SRS. The advantage is the guaranteed representativeness in some important characteristics.

- For example, say that we want to select a sample of 100 individuals. Sex ratio in the sample is up to chance if we do SRS. We can guarantee the 50-50 split if we do stratified sampling:
Stratified sampling is often used to reduce the variability of a sample.
Oversampling, Graphic Representation

- Increasing the representation of a group in a sample. This is often done when groups are very different in size – e.g., race
4. Cluster Sampling

• Cluster sampling is desirable from an economic point of view:

• It saves money, at the expense of lowering the quality of data.

• e.g.: we are interested in studying students' experiences at the University of Wisconsin, the unit of analysis is the student. An economic way is to sample classes. Once a class is selected, everyone in the class is selected.
Cluster Sampling, Graphic Representation

Collection of selected elements
Cluster Sampling, Loss of Efficiency

• For cluster sampling to work well, we assume clusters do not differ radically from each other.

• If this assumption is not true, the sample has more variability than a sample obtained by SRS, resulting in inefficiency.

• In general, we can only lose efficiency with cluster sampling.
Sources of Variability in a Sample Statistic

• 1. Population Variability
   All elements in a population are inherently variant.

• 2. Random Selection
   Precisely because elements in the population have different values on a variable, random selection is meaningful and necessary.
Decomposing Variability

• 1. ANOVA: Analysis of Variance
  Total variation =
  between-group variation + within-group variation.

• Internal homogeneity => external heterogeneity => between-group variance is large (e.g., gender and height).

• Internal heterogeneity => external homogeneity => between-group variance is small (e.g., gender and GPA).
Effects of Stratification

• Stratification reduces sampling variation.
• Total variation - between-strata variation = within-strata variation.
• The more heterogeneous are the strata externally (or equivalently, the more homogeneous internally), the greater the gain in precision arising from stratification.
• Example of gun control law and region.
Design Effect

• The ratio of the variance of the estimator based on the complex design to the variance of the estimator based on simple random sampling of the same size.

• \( D^2(z) = \frac{V(z)}{V(z_0)} \)

• For stratified samples, \( D^2 \leq 1 \). That is, stratified samples cannot be less efficient than simple random samples. \( D^2 = 1 \) if strata do not differ from each other.
Effects of Clustering

• Clustering increases sampling variation.
• For a cluster sample, the Design Effect \((D^2)\) \(\geq 1\). That is, cluster samples cannot be more efficient than simple random samples. \(D^2 = 1\) if clusters do not differ from each other.
• Example of cluster sampling of individuals based on state of residence.
Cluster sampling vs. Stratification.

- Since strata are all represented in the sample, it is advantageous if they are internally homogeneous (i.e., externally heterogeneous).
- With cluster sampling, it is best when the clusters are internally heterogeneous in the characteristics being studied.
- The cluster size also affects the design effect. The larger the cluster size, the less efficient is the sample (i.e., the higher the variance).
Practical Matters

• **1. Implicit Stratified Sampling**
  Combining systematic sampling with stratified sampling.

• **2. Multi-stage Cluster Sampling**
  Combining cluster sampling at one level with cluster sampling at another level.
Combination of Clustering and Stratification

• You may have cluster sampling at a higher level and have stratification at a lower level. e.g., cluster sampling of counties first, and then stratification by race and sex second.

• Or, you may have stratified sampling at a higher level and have cluster sampling at a lower level. e.g., stratify regions, and then cluster sampling of counties.
Probability Proportionate to Size Sampling

• A type of cluster sampling where a cluster's probability of being selected is proportional to its size. That is, the larger a cluster, the higher its probability of being selected.

• Within each cluster, a fixed number of cases are selected. For individual $i$ in cluster $j$ (by the rule of conditional probability):

$$P(i|j) = P(i|j) \times p(j) = \frac{1}{N_j} \times \frac{N_j}{N} = \frac{1}{N},$$
equal probability for the element.
Disproportionate Sampling and Weighting

• Disproportionate sampling usually is a kind of stratified sampling where strata have different sampling ratios. Example: oversampling of American Indians.

• Weighting is necessary. Weight is usually the inverse of the sampling ratio (e.g., if blacks sampled at three times population representation we would weight each black respondent by 1/3).
Alternative Ways of Sampling

• 1. Convenience Sampling
  Soc. 357 class in 2009.

• 2. Quota Sampling
  Matching overt characteristics. No random selection.
  – Problem: bias in other characteristics.

• 3. Purposive sampling

• 4. Snowball Sampling