**EPSY 5244 Survey Design, Sampling, and Implementation**

**Coverage Error**

Ability to include (capture) all communities/groups within the intended population – which influences our ability to generalize. This is primarily a concern because we don’t know if the missing community would respond differently than those that are included.

* Issues related to the sampling frame (the list of the population)
  + Does it contain everyone in the survey population?
  + Does it contain individuals NOT in the survey population?
  + Is the list maintained and current?
  + Repeated entries need to be removed.

**Probability Sampling & Sampling Error**

Error introduced by using a portion of a population – variation in results due to sampling

Sample size determination (relate to precision)

* How much error are you willing to tolerate?
* How large is the population? (to a point)
* How variable is your population?
* How much confidence in your estimate is needed?

**Probabilistic Sampling Methods**

(assumes we have a sampling frame, from which we can estimate probability of selection)

1. Simple random sampling
   1. Every individual in the population has an equal opportunity (probability) to be selected into the sample
   2. Every sample has the equal probability of being selected for participation in the survey

We use sampling without replacement in practice, but rely on the statistics of sampling with replacement, since they are much easier – the probabilities don’t change with each selection.

1. Stratified random sampling
   1. Subdivide the population into known (naturally occurring) groups (grouping variable), called strata
   2. This requires that the stratification information is available in the sampling frame.
   3. Draw a sample from each strata
   4. Choose strata based on a characteristic that would be helpful to achieve the purpose of the survey
   5. Typically we select sample sizes within strata as a function of *n* or proportion, to secure representation
2. Simple random cluster sampling
   1. We define clusters as naturally organized individuals – predefined clusters
   2. We sample, randomly, clusters, not individuals
   3. Done to address administration barriers
   4. One-stage or two-stage (or more) cluster sampling

**Non-Probabilistic Sampling Methods**

1. Systematic sampling
   1. Every *n*th individual is selected
   2. This is most obviously a problem when there is a systematic characteristic that varies in the order of individuals
2. Convenience sampling
   1. Accessible, available, and willing
   2. May be a problem if timing and location are related to the characteristics being studied
3. Snowball sampling
   1. Start with some participants, then ask them to recommend others to participate
   2. Only participants know others from their community – or those with similar characteristics
4. Focus group
5. Expert panels
6. Purposive sampling