Sampling and Statistical Terminology Lunch & Learn

Daniel Guzman

Statistics and Methods Unit

March 2017

- On going effort
- No need to cover all today
- Please ask question/interrupt any moment
- References:
  - Encyclopedia of Survey Research Methods
  - Cross-cultural Survey Guidelines

1) Define Target Population

2) Determine Sampling Frame

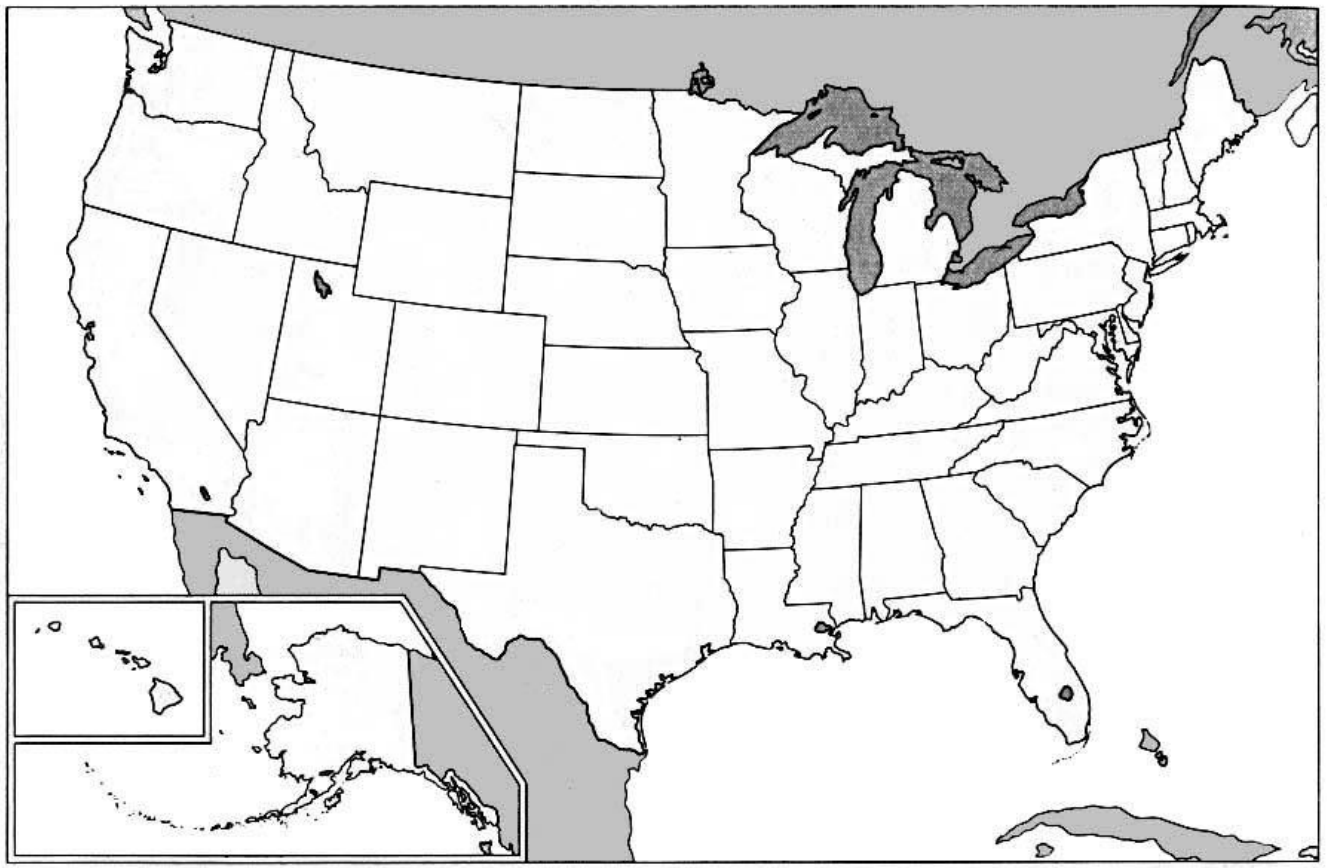
3) Select Sampling technique

4) Execute sampling plan (data collection)

5) Post-Survey adjustments

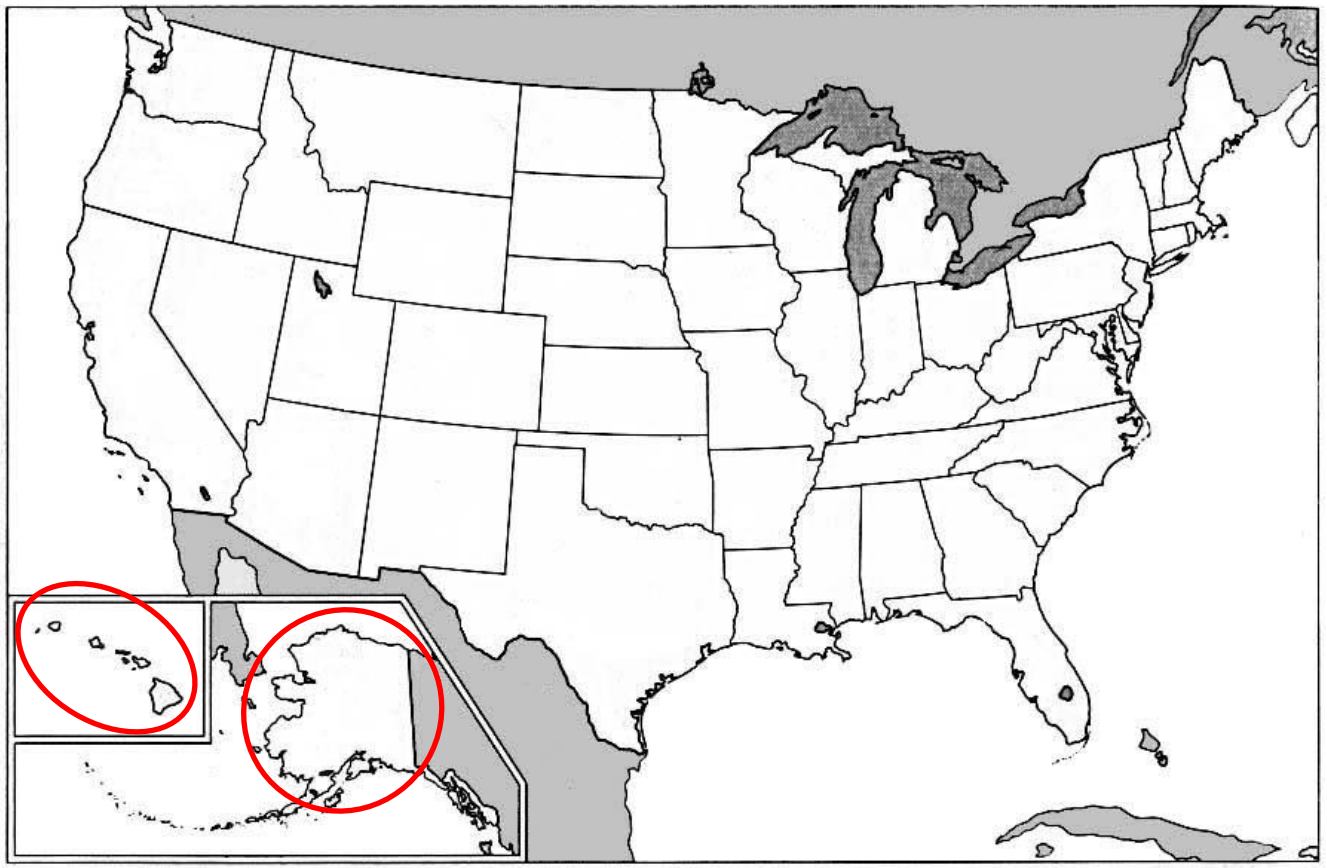
- Target Population:

The finite population for which the survey sponsor wants to make inferences using the sample statistics.



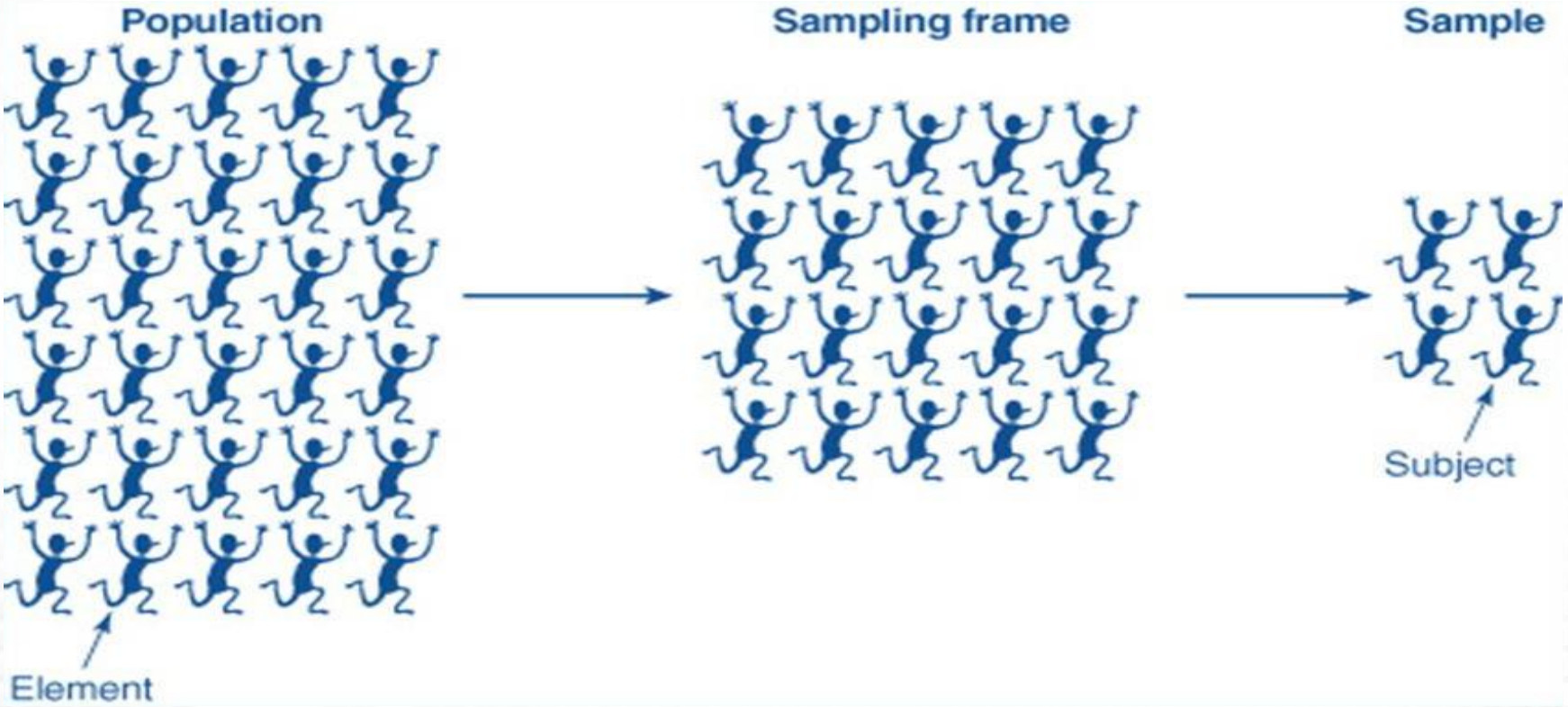
- Survey Population:

The actual population from which the survey data are collected, given the restrictions from data collection operations.



- # Sampling Frame (1)

A list or group of materials used to identify all elements of a survey population from which the sample will be selected. This list or group of materials can include maps of areas in which the elements can be found, lists of members of a professional association, and registries of addresses or persons.

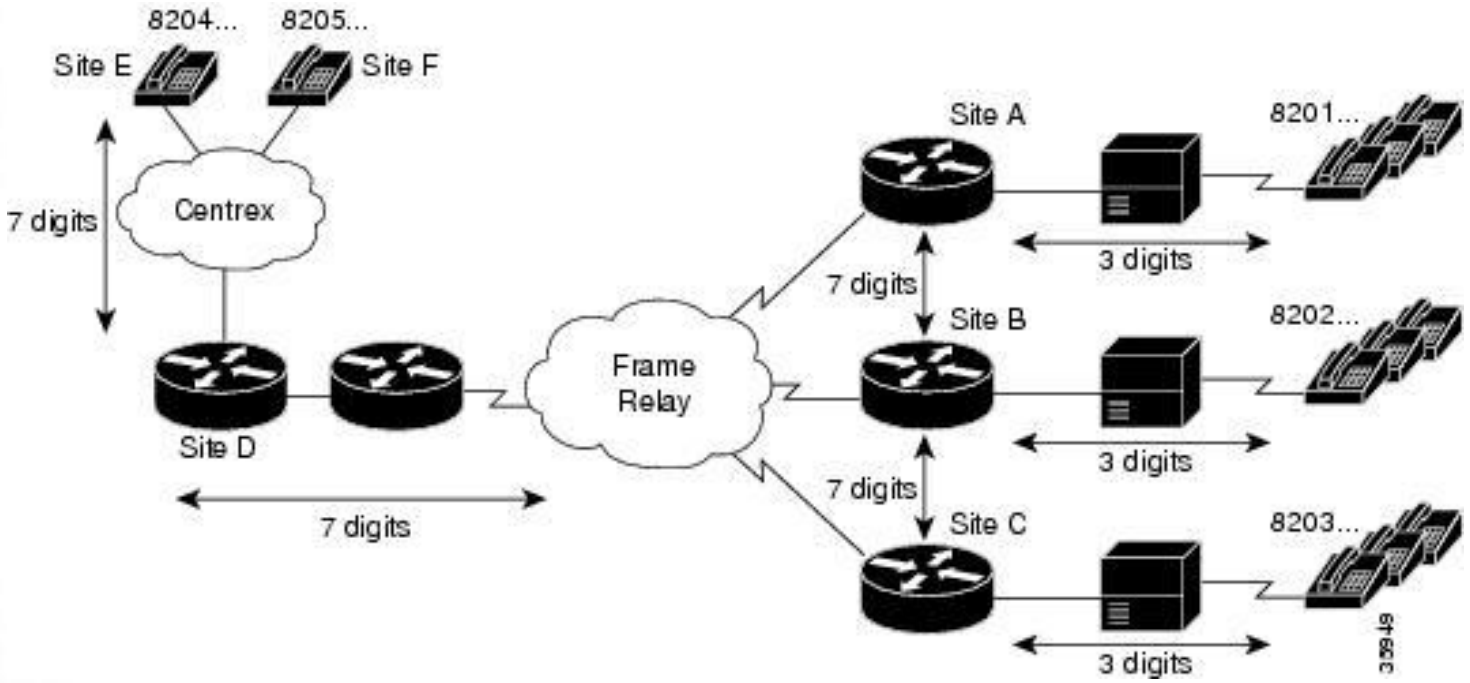


- Sampling Frame (2)

VIEWTABLE: TMP1.hrs2016_2010psu_cnty_xwalk

	psu_name	psu_2016_code	Concatenated 2010 FIPS State and County code	Census 2010 County Name	State Post Letter Abbreviat
157	Sevier County, AR	3477	05133	SEVIER	AR
158	Sharp County, AR	3478	05135	SHARP	AR
159	Stone County, AR	3479	05137	STONE	AR
160	Union County, AR	3480	05139	UNION	AR
161	Van Buren County, AR	3481	05141	VAN BURE	AR
162	White County, AR	3482	05145	WHITE	AR
163	Woodruff County, AR	3483	05147	WOODRUFF	AR
164	Yell County, AR	3484	05149	YELL	AR
165	Atlanta, GA SMSA	92901	13151	HENRY	GA
166	Atlanta, GA SMSA	92901	13247	ROCKDALE	GA
167	Atlanta, GA SMSA	92901	13097	DOUGLAS	GA
168	Atlanta, GA SMSA	92901	13135	GWINNETT	GA
169	Atlanta, GA SMSA	92901	13223	PAULDING	GA
170	Atlanta, GA SMSA	92901	13113	FAYETTE	GA
171	Atlanta, GA SMSA	92901	13035	BUTTS	GA
172	Atlanta, GA SMSA	92901	13057	CHEROKEE	GA
173	Atlanta, GA SMSA	92901	13067	COBB	GA
174	Atlanta, GA SMSA	92901	13117	FORSYTH	GA
175	Atlanta, GA SMSA	92901	13089	DE KALB	GA
176	Atlanta, GA SMSA	92901	13217	NEWTON	GA
177	Atlanta, GA SMSA	92901	13121	FULTON	GA
178	Atlanta, GA SMSA	92901	13297	WALTON	GA
179	Atlanta, GA SMSA	92901	13063	CLAYTON	GA
180	Atlantic City, NJ SMSA, NJ	90838	34001	ATLANTIC	NJ
181	Augusta, GA-SC SMSA, SC	92810	13245	RICHMOND	GA
182	Augusta, GA-SC SMSA, SC	92810	13073	COLUMBIA	GA
183	Augusta, GA-SC SMSA, SC	92810	45003	AIKEN	SC
184	Alpine-El Dorado Counties, CA	4254	06003	ALPINE	CA

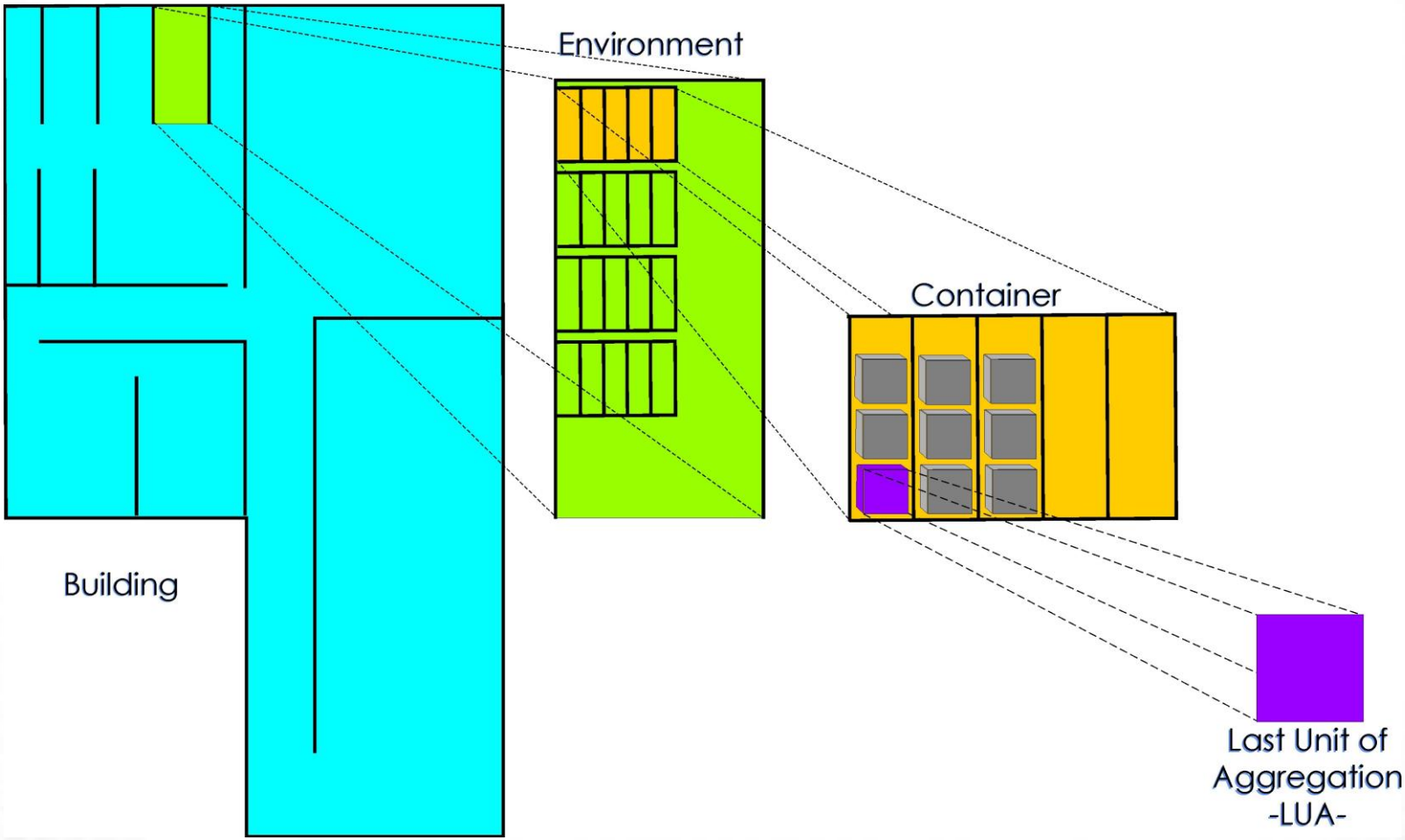
- Sampling Frame (3)



Multi Stage Sampling (1)

- A multi-stage sample: is one in which sampling is done sequentially across two or more hierarchical level
- Sampling Unit: Elements or group of elements considered for selection in some stage of sampling.

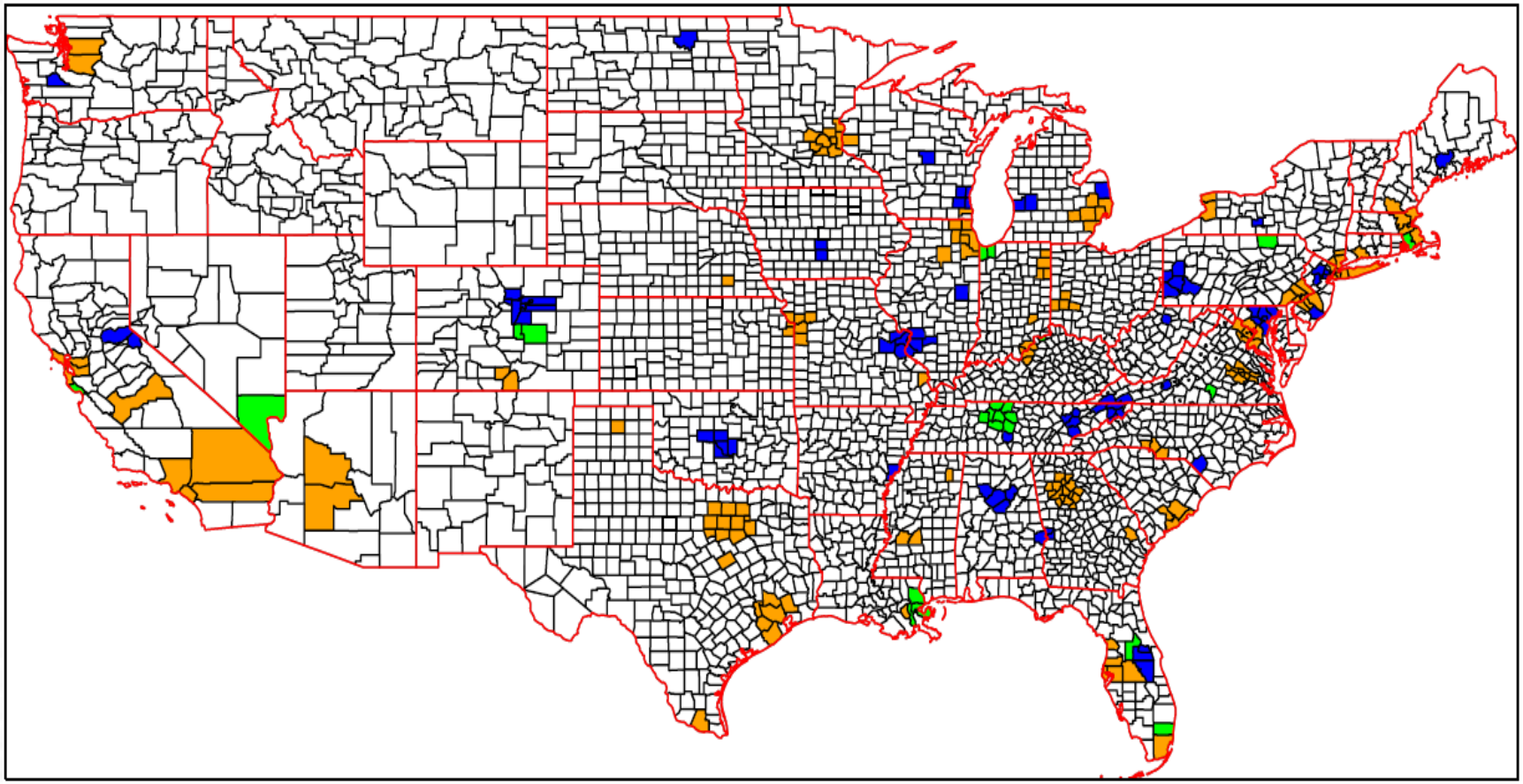
Multi Stage Sampling (2)



Multi Stage Sampling (3)

- PSU: *Primary sampling unit* refers to sampling units that are selected in the first (primary) stage of a multi-stage sample ultimately aimed at selecting individual elements.
- Examples: Counties, School districts, Schools

Multi Stage Sampling (4)



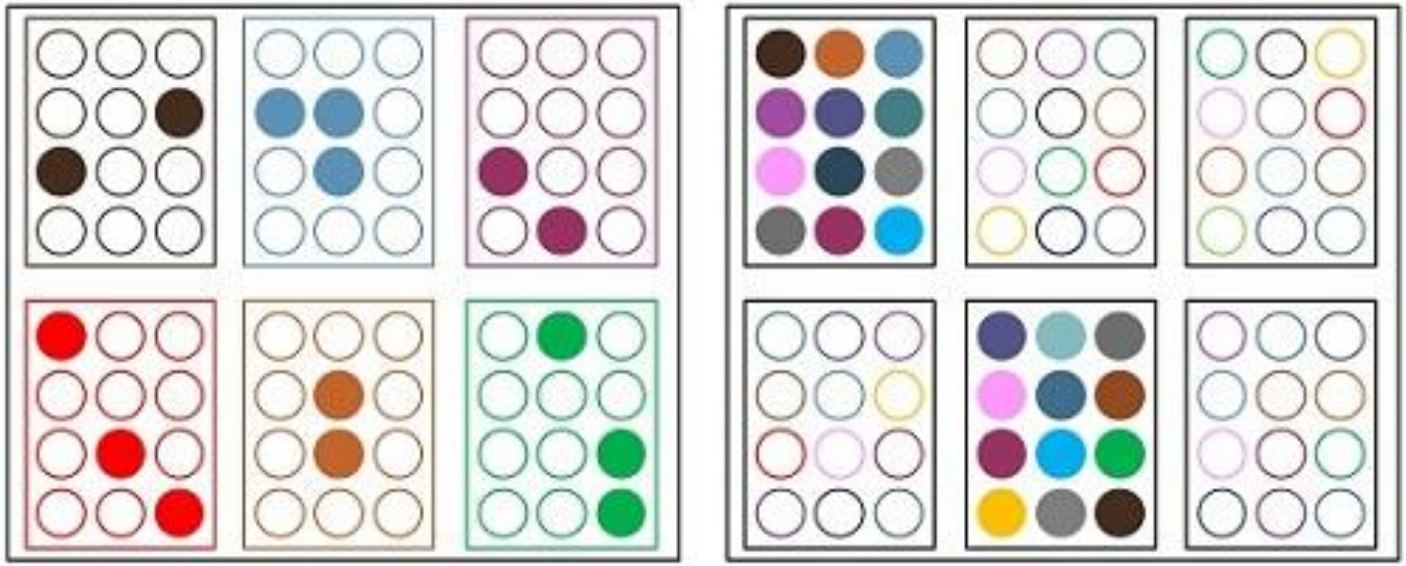
Multi Stage Sampling (5)

- SSU: *Secondary sampling unit* refers to sampling units that are selected in the second stage of a multi-stage sample. Often, segments are sample units in the second stage of area probability sampling
- Examples: Area segments, Schools, class rooms

Clusters vs Strata (1)

- Cluster: A grouping of units on the sampling frame that is similar on one or more variables, typically geographic.
- Strata: In stratified sampling are distinct subsets of all entries on a sampling frame. These subsets are strategically defined by one or more stratification variables to improve the statistical quality of findings for the population as a whole or for important population subgroups

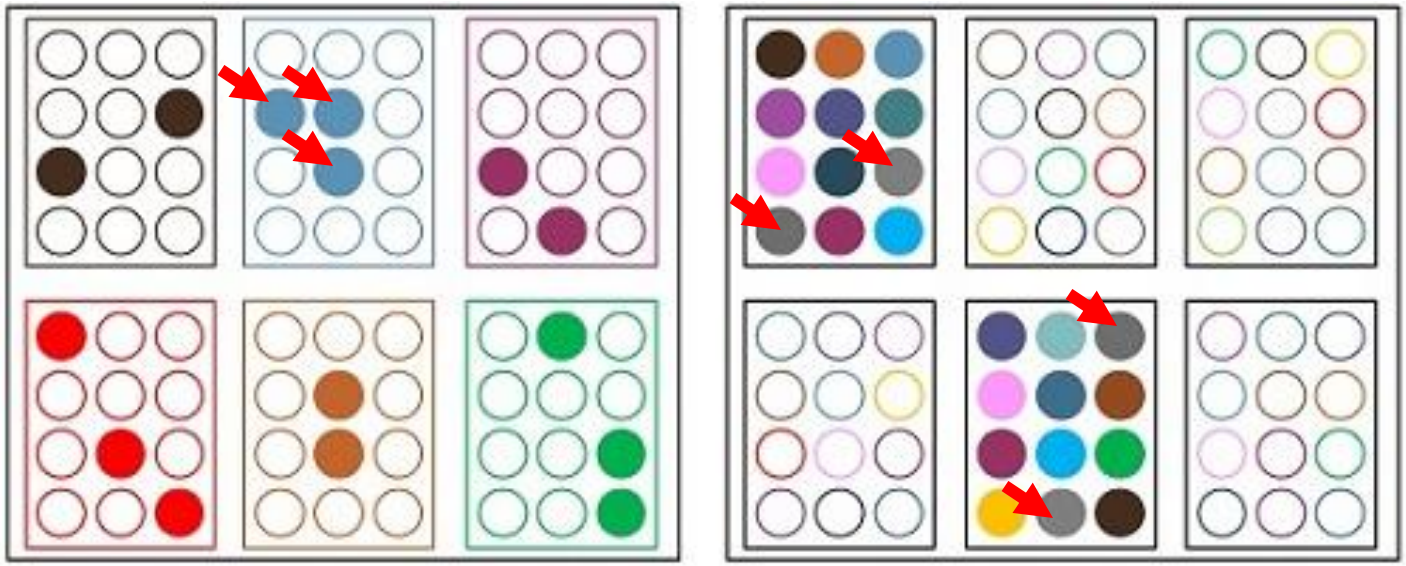
Clusters vs Strata (2)



Domain/Subpopulation/Subclass

Surveys are seldom designed to yield information only for the total population. Subclasses are subpopulations within the target population for which separate estimates are prepared. A subclass is thus a portion of the sample for which inferences are to be made to the totality of subclass elements in the population.

Domain/Subpopulation/Subclass

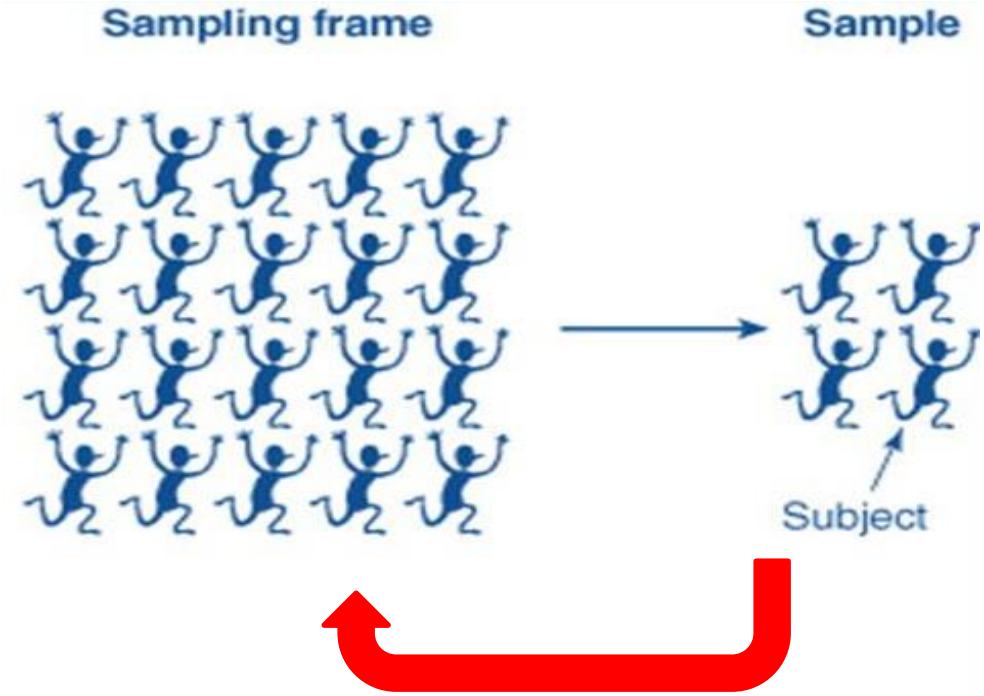


Probability of Selection

- The term probability of selection refers to the chance that a member of a population can be chosen for a given survey. When using a probability sample, the term also means that every member of the sampling frame that is used to represent the population has a known nonzero chance of being selected

EPSEM

- Equal Probability of Selection Method



Oversampling

- For some surveys, it is important to ensure that there are enough members of a certain subgroup in the population so that more reliable estimates can be reported for that group. To do this, more people from this group are selected than would typically be done if everyone in the sample had an equal chance of being selected.

Weights/Weighting

- Weighting is a correction technique to improve the accuracy of the survey estimates.

There are two basic reasons:

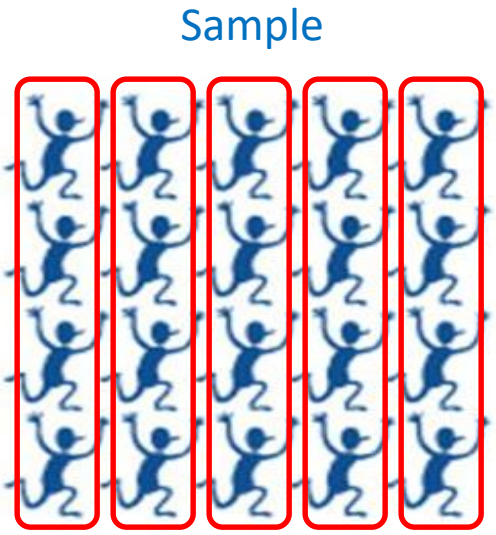
- to correct for unequal probabilities of selection.
- To compensate for survey nonresponse.

Design Effect

- The effect of the complex survey design on sampling variance measured as the ratio of the sampling variance under the complex design to the sampling variance computed as a simple random sample of the same sample size.

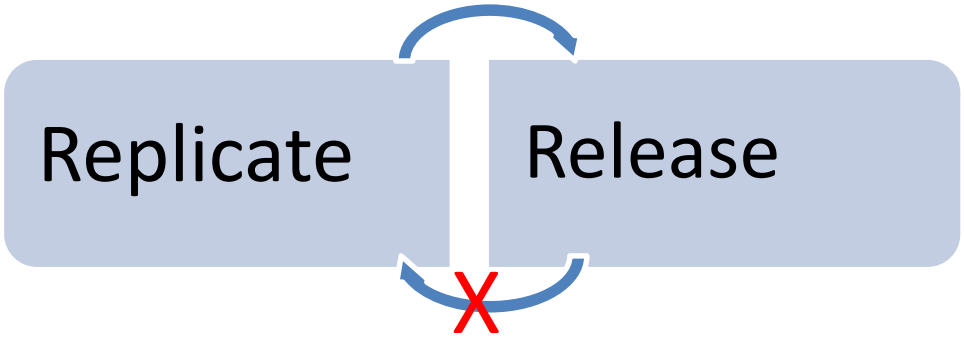
Sample Replicate

- A sample replicate is a random subset of the entire available sample (i.e. *sampling pool*) that has been drawn for a particular survey. Sample replicates help survey managers coordinate the progress that is made on data collection during the survey's field period.



Sample Release

In survey research, sample release refers to the process of how sample cases are deliver to the field. The best practice is to deliver the sample in replicates.

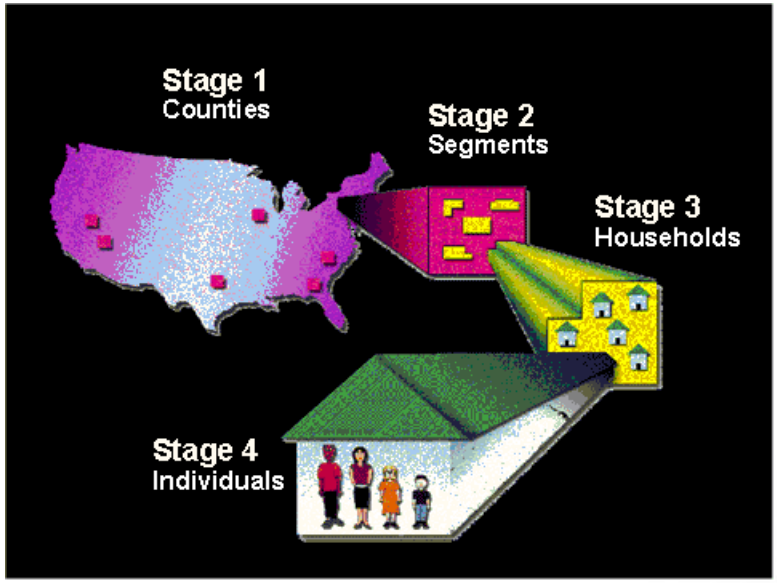


Double Sample

- Designs in which initially a sample of units is selected for obtaining auxiliary information only, and then a second sample is selected in which the variable of interest is observed in addition to the auxiliary information.

Sub-Selection

- It is a generic term referring to selecting elements from a sample.
- Examples:
  - Multi-Stage sampling
  - Double sampling



Responsive Design

- 1. Preidentify a set of design features that affect cost and error tradeoffs.
- 2. Identify indicators for these costs and errors. Monitor these during data collection.
- 3. Alter the design features based on pre-identified decision rules based on the indicators in step 2.

Adaptive Design

The sampling design is modified in real time as data collection continues—based on what has been learned from previous sampling that has been completed. Its purpose is to improve the selection of elements during the remainder of the sampling.

“Adaptive" design as a subset of responsive design

Post-survey adjustments

- Post-survey adjustments refer to a series of statistical adjustments applied to survey data prior to data analysis and dissemination. Although no universal definition exits, post-survey adjustments typically include data editing, missing data imputation, weighting adjustments, and disclosure limitation procedures.

Post-stratification

- A statistical adjustment that assures that sample estimates of totals equal population totals. The adjustment cells for poststratification are formed in a similar way as strata in sample selection, but variables can be used that were not on the original sampling frame at the time of selection.

Non-Response Adjustment

- Non-response adjusted weights are based on achieving the alignment between the responding sample and the original sample.

Imputation

- A computation method that, using some protocol, assigns one or more replacement answers for each missing, incomplete, or implausible data item.

Weight smoothing

- Highly disproportional sample designs have large weights, which can introduce unnecessary variability in statistics such as the population mean estimate
  - Trim weights
  - Apply a model to smooth weight

References

- http://methods.sagepub.com/Reference/encyclopedia-of-survey-research-methods/n339.xml
- http://methods.sagepub.com/Reference/encyclopedia-of-survey-research-methods/n339.xml
- http://jameswagnersurv.blogspot.com/2010/09/responsive-design-and-adaptive-design.html
- http://www.pewresearch.org/methodology/u-s-survey-research/sampling/#oversampling

Thank you!