Survey Sampling And Political Polls

- In the 2000 Presidential election, pollsters came within a couple of percentage points of estimating the votes of 100 million people.

- To gather this information, they interviewed fewer than 2,000 people.

Election Eve Polls - Voting for U.S.Presidential Candidates, 2000

Date	Agency	Gore	Bush	Nader	Buchanan
11/6	IDB/CSM	47	**49**	4	0
11/6	CBS	**48**	47	4	1
11/6	CNN/USA Today]	46	**48**	4	1
11/6	Reuters/ MSNBC	**48**	46	5	1
11/6	Voter.com	45	**51**	4	0
11/7	**Results**	**48**	**48**	**3**	**1**

Probability Sampling

- Creating a sample in accord with or using **probability theory**. Typically involving **random selection** of elements.
- All elements must have an equal chance of being selected.
- Is the use of a telephone directory a good way of allowing for random sampling?

Probability Sampling

- **Why Probabilility Sampling**:
  - (1) findings can be generalized to the larger population
  - (2) it allows one to estimate the amount of error in one's generalization to the larger population and
  - (3) it allows one to know how much confidence one can have in this estimate of error

- **Example**: We might be 95% confident (confidence level) that the percentage of voters for President Bush is between 48% and 52% (plus or minus 2% or confidence interval).
- **Example**: Or, we might be 99% confident that the percentage of voters for Bush is between 46% and 54% (plus or minus 4%)

Nonprobability Sampling

Generally a reliance on available subjects

- Only justified if less risky sampling methods are not possible. Typically no inclusive list of elements to sample from (e.g., the homeless).

- Researchers must exercise great caution in generalizing from their data when this method is used.

Nonprobability Sampling

Purposive or Judgmental sampling

- Units are selected on the basis of one's judgment about which ones will be the most useful or representative

- For example, to get an understanding of how conservative and liberal students think, one might pick students from two political groups within the university who reflect these ideologies.

Nonprobability Sampling

Snowball sampling

- Appropriate when members of a population are difficult to locate (homeless, migrant workers, or undocumented immigrants.)

- Researcher collects data on members of the target population she can locate, then asks those individuals to help locate other members of that population.

Nonprobability Sampling

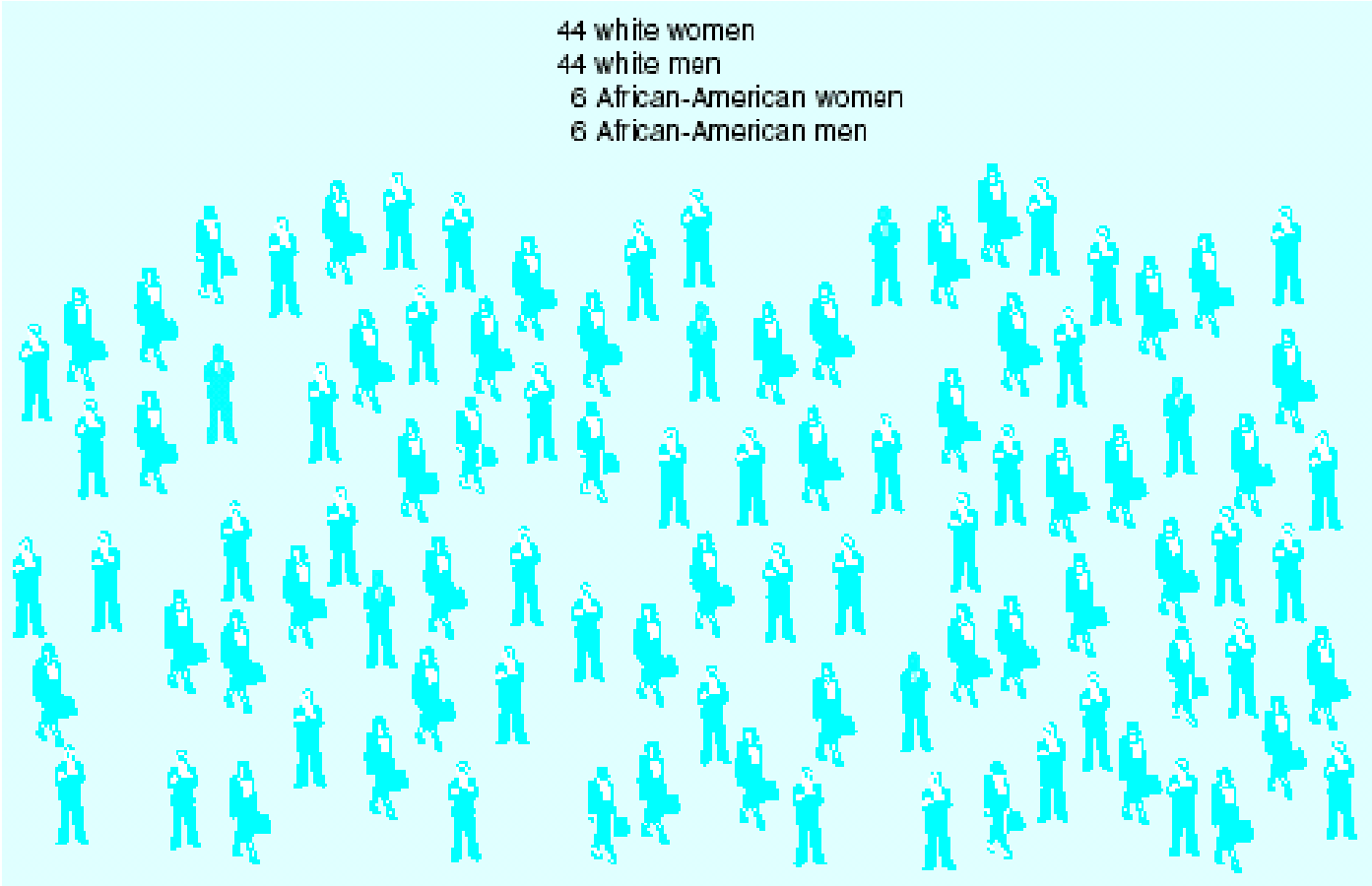
Quota sampling

- Begins with a matrix of the target population.
- Data is collected from people with the characteristics of a given cell.
- Each group in a cell is assigned a weight appropriate to their portion of the total population.
- When the elements are properly weighted, data should provide a representation of the total population.

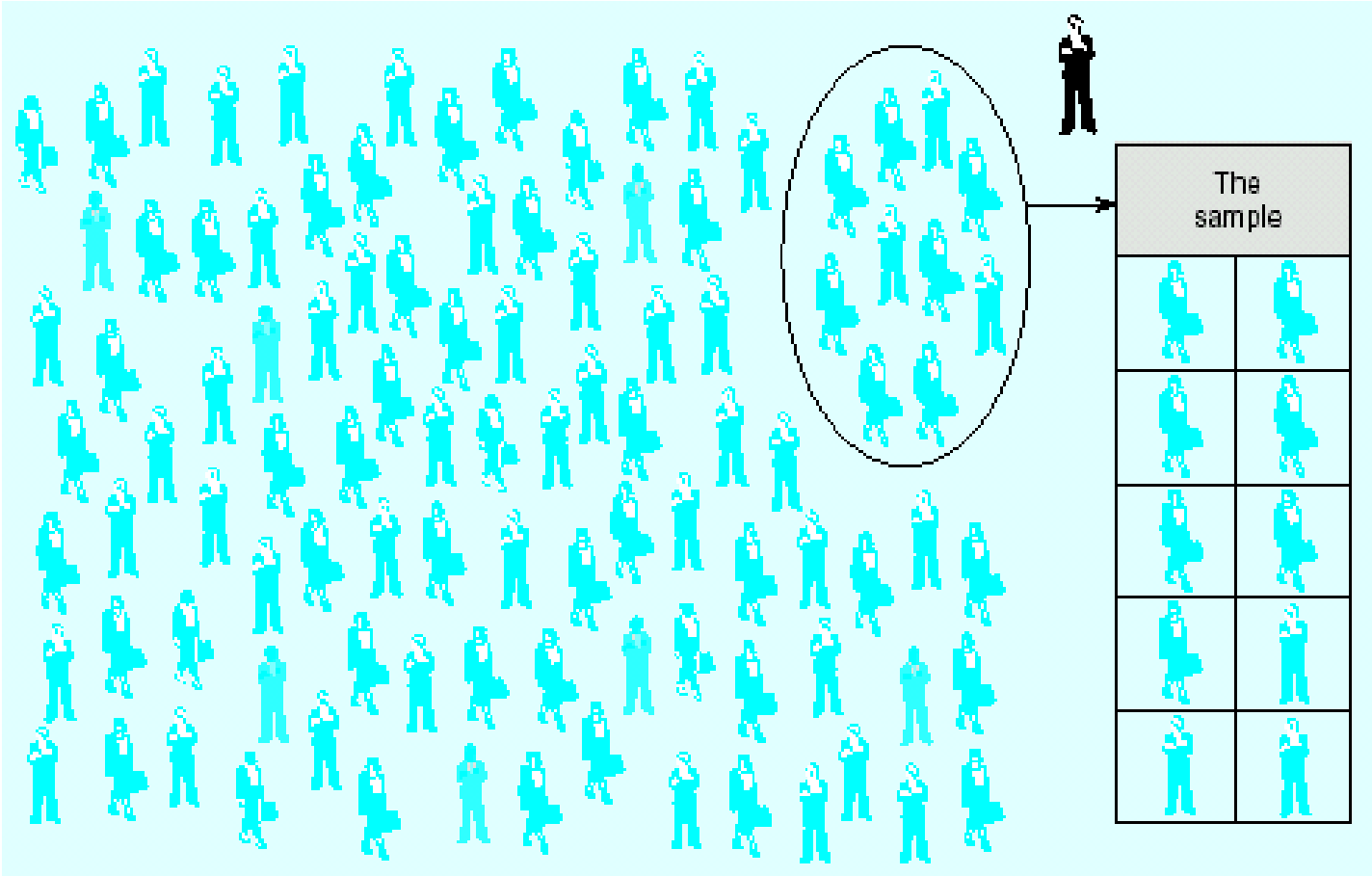
Probability Sampling

- **Why Probabilility Sampling**:
  - (1) findings **can be generalized** to the larger population
  - (2) it allows one to **estimate the amount of error** in one's generalization to the larger population and
  - (3) it allows one **to know how much confidence** one can have in this estimate of error
- **Example**: We might be 95% confident that the percentage of voters for President Bush is between 48% and 52% (plus or minus 2%).
- **Example**: Or, we might be 99% confident that the percentage of voters for Bush is between 47% and 53% (plus or minus 3%)

A Population of 100 Folks



A Sample of Convenience: Easy, but Not Representative



What Makes Probability Theory Work?

- (1) The **larger the sample** the smaller the samples error. In other words, the larger the sample, the closer the sample statistic will reflect the true population statistic.

- And, (2) **the more homogeneous** the total population the smaller the sample's error.

Probability Theory and A Sample's Error

- Probability theory provides a formula for estimating a sample's error for a specific variable (that is, how different the sample statistic is from the population statistic that we are trying to predict).
- When using a dichotomous variable (the simplest data), the formula is:

  "Standard" Error = $\sqrt{(P \times Q)/N}$
- Dichotomous variables have two statistics such as (1) % voted for Bush (P above), and (2) % voted against Bush (Q above)
- Mathematically the sample's error is referred to as the sample's **"standard error"** since it is, mathematically, the standard deviation of the error.

Probability Theory and Sampling Error

$$\sqrt{(P \times Q)/N} = \text{standard error}$$

- If we were to apply the formula to a dichotomous variable such as: Did you vote for OR vote against Bush, the formula would be carried out like this:

- P = % voted for
- Q =% voted against
- N = number of cases

(see next slide for example)

What Makes Probability Theory Work?

- Example #1: If 60% are for Bush and 40% are against Bush and we determine this after sampling 400 people, the sample's "standard" error would be:

$$\sqrt{(.60 \times .40)/400} = .024 = \text{standard error}$$

- That is, those in favor of Bush would be between 57.6% and 62.4% (.60 - .024 and .60 + .024) or those opposed 37.6% & 42.4%

  - In this example the population is NOT very homogeneous since almost half are for and the other half against Bush.

  - How confident are we that the sample statistic falls between 57.6% and 62.4%?

What Makes Probability Theory Work?

- How confident are we that the sample statistic falls between 57.6% and 62.4%?

- **Probability Theory's** mathematical formula provides an estimated range of error with 68% confidence that the range is correct.

- However, doubling the standard error provides 95% confidence in the range and tripling the SE provides 99% confidence.

What Makes Probability Theory Work?

Example #1 (continued):

Therefore:

1. if the standard error is .024 and
2. In our sample 60% were in favor of Bush and
3. we want to be 95% confident in our prediction of the total population
4. We would report that we are 95% confident that the population statistic ranges between:

   .552 and .648

   (.024 x 2 = .048, and then subtracted from and added to the statistic 60%)

What Makes Probability Theory Work: Example 2?

- Example #2: If 90% are for Bush and 10% are against Bush and we determine this after sampling 400 people, the sampling error would be:

$$\sqrt{(.90 \times .10)/400} = .015$$

  - That is, those in favor of Bush would be between 88.5% and 91.5% with 68% confidence and between 87% and 93% with 95 % confidence

  - In this example the population is very homogeneous since almost all are in favor of Bush

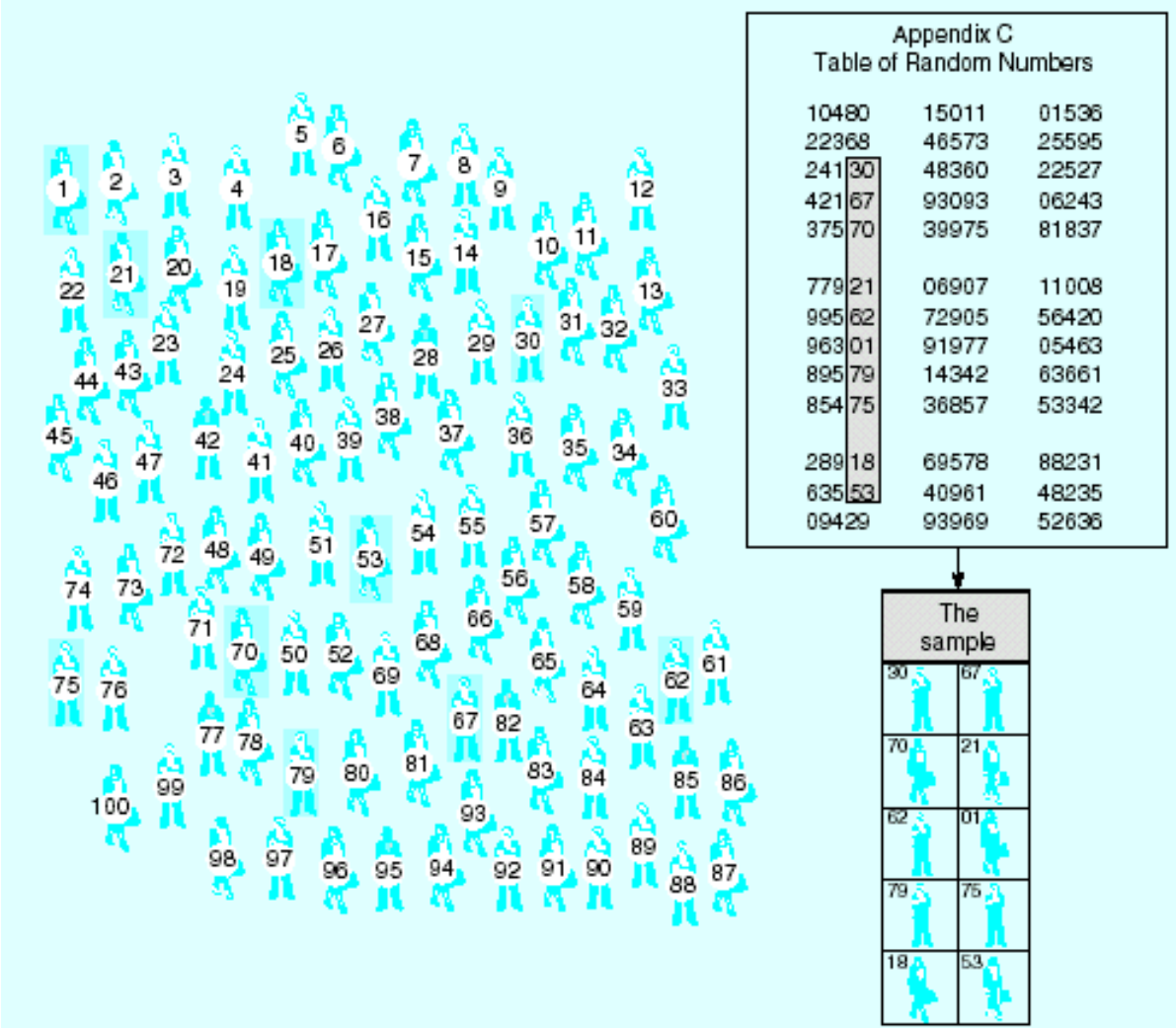
Requirements Associated with Probability Theory

- Elements or units must be randomly selected.
- Each unit must have an equal chance of being selected.

Types of Sampling Designs

- Simple random sampling (SRS)
- Systematic sampling
- Stratified sampling
- Cluster sampling

A Simple Random Sample



Simple Random Sampling

- Feasible only with the simplest sampling frame.
- Not the most accurate method available.

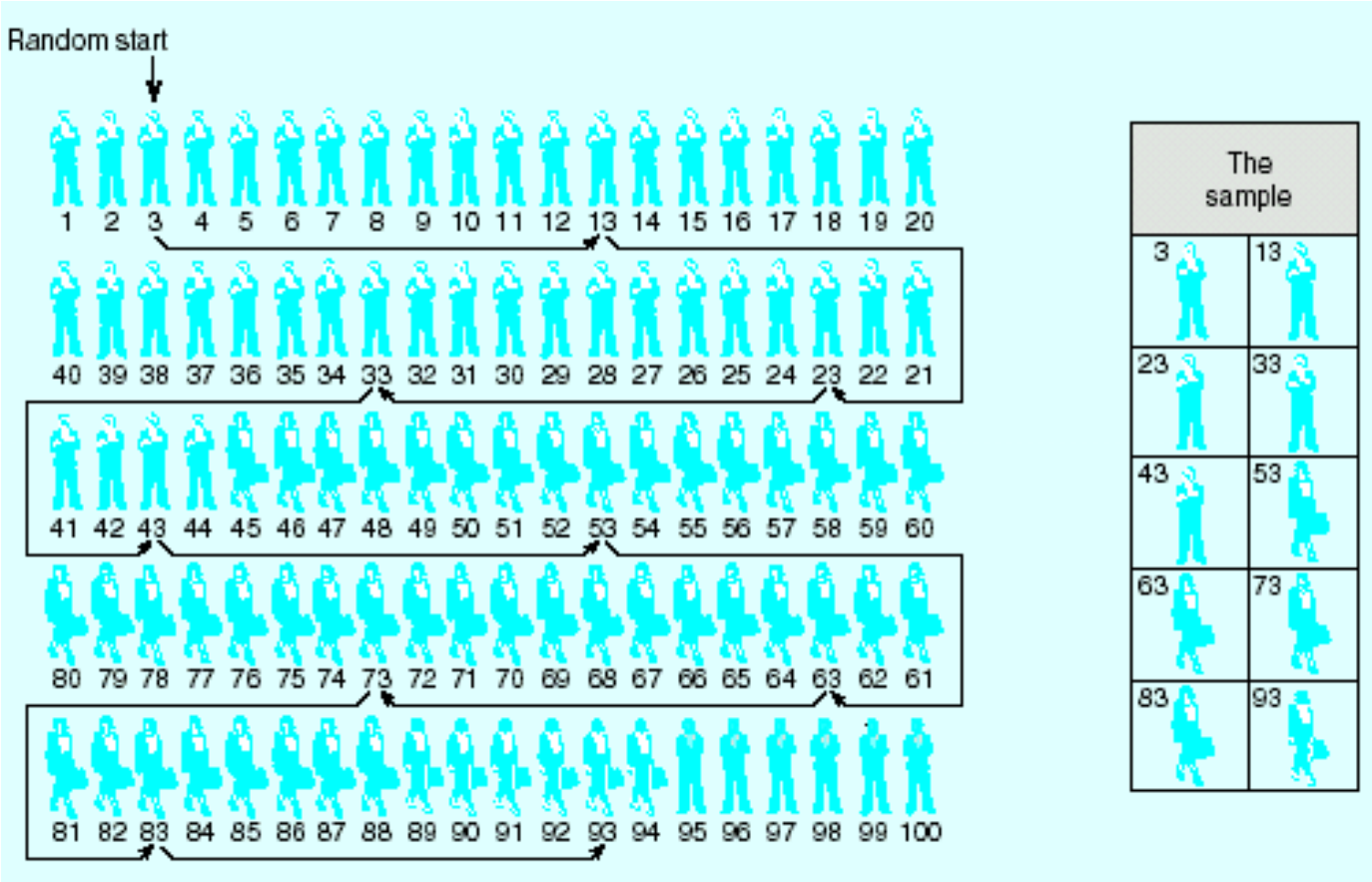
Systematic Sampling

- List all the elements in the sampling frame and then select every x^{th} element
- Easier to use than simple random sampling.
- However, arrangement of elements in the list can result in a biased sample.

Stratified Sampling

- Researcher organizes the elements into homogeneous subsets of the population that she is interested in (e.g., classes of students)
- For example, to survey undergraduate university students, one would create a list of students (or elements) from each of four groups (or strata) of students with each group representing one of the classes (freshman, sophomore, etc.). Then students would be randomly drawn from the list.
- This allows the researcher to ensure the representativeness of those subsets of interest. It can result in a more homogeneous sample and consequently a lower sampling error than simple random sampling.
- However, the sample drawn will not necessarily be representative of the entire population (e.g, all university students) but instead only of those subsets drawn (e.g., would not include graduate students).

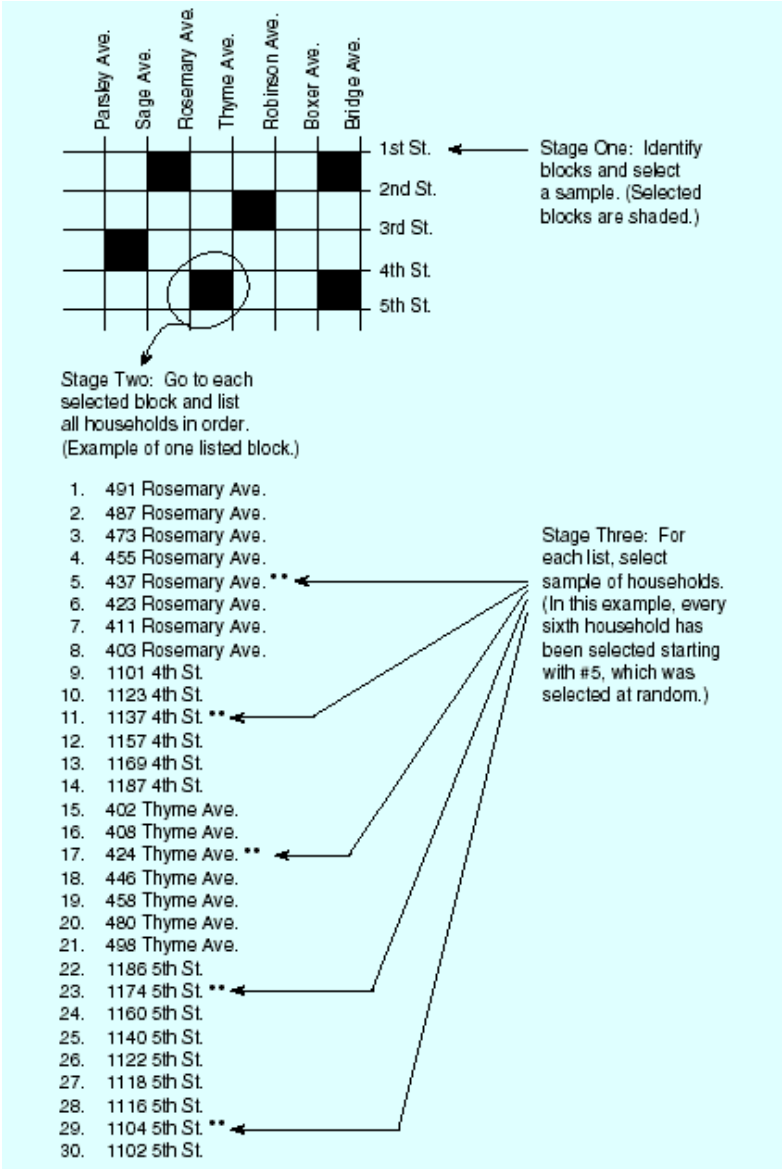
A Stratified, Systematic Sample with a Random Start



Multistage Cluster Sampling

- Used when it's not possible or practical to create a list of all the elements that compose the target population.
- Involves repetition of two basic steps: listing and sampling.
- Highly efficient but less accurate. For example, a telephone survey might first randomly pick a page and then randomly pick a person on the page. Standard error occurs during each random selection.

Multistage Cluster Sampling



Probability Proportionate to Size (PPS) Sampling

- Sophisticated form of cluster sampling.
- Used in many large scale survey sampling projects.
- The object is to make sure that all elements have an equal chance of getting selected.

Summary: Probability Sampling

- Most effective method for selection of study elements.
- Avoids researchers biases in element selection.
- Permits estimates of sampling error and level of confidence in those estimates.