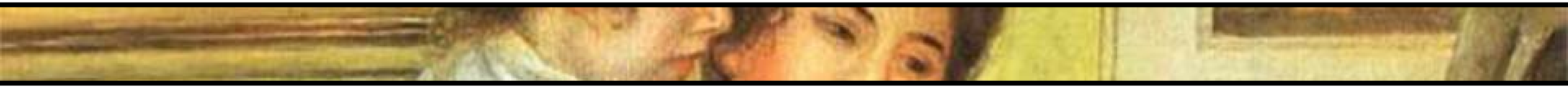
The Logic of Sampling

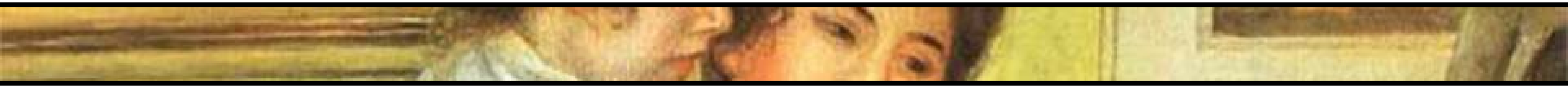
Pierre-Auguste Renoir: Yvonne & Christine Lerolle Playing the Piano, 1897.



A Brief History of Sampling

1. Some Failures

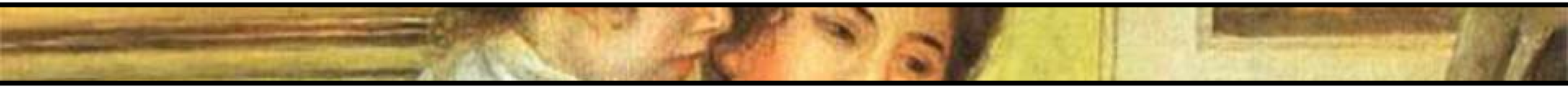
- <u>President Alf Landon</u>: In 1936, the *Literary Digest* predicted that Landon would defeat President Franklin Roosevelt. It based this opinion on a poll it conducted, wherein it selected its sample from a list of telephone numbers and automobile registrations.
- The flaw in this sampling procedure was that this sample frame was biased toward the educated and affluent, who tended to vote for Landon (Roosevelt won in a landslide!).



A Brief History of Sampling

1. Some Failures (Continued)

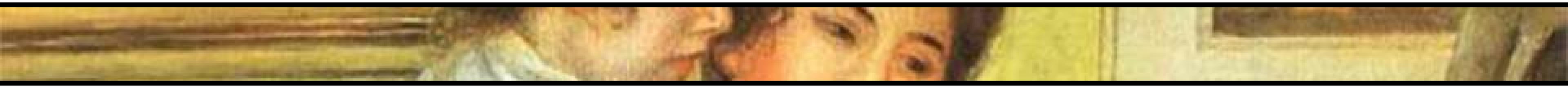
- President Thomas Dewey: In 1948, the George Gallop agency predicted that Dewey would defeat Harry S. Truman. It based this opinion on a poll it conducted, wherein the sample was selected by quota using the 1940 Census figures.
- The sample was biased because the 1940 Census did not reflect the rapid move to urban areas following WWII. The many new unaccounted for urban dwellers tended to vote for Truman.



Nonprobability Sampling

1. Definition

- A sample that relies upon available subjects.
- Researchers sometimes rely upon available subjects rather than draw samples using probability sampling.
- Available subjects are selected because:
  1. Lack of access all members of the population,
  2. Reduction in costs and time,
  3. Lack of need for a probability sample.



Nonprobability Sampling

2. Purposive or Judgmental Sampling

- Selection of individuals with specific characteristics.
- One might:
  1. Request certain individuals within a population (e.g., ask for an adult male in the household in a telephone survey),
  2. Restrict the sampling to certain audiences (e.g., select a sample from readers of *Popular Mechanics*),
  3. Specify a need for individuals with certain characteristics (e.g., solicit with ads).

Nonprobability Sampling

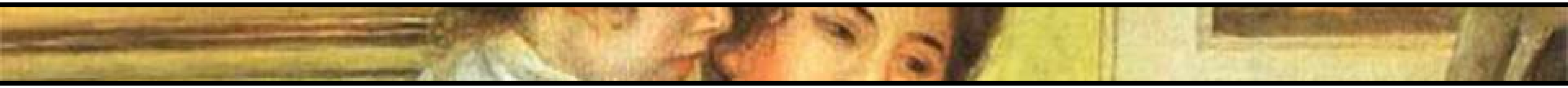
2. Purposive or Judgmental Sampling

- This type of sampling has the advantage of collecting information from a targeted element. For example, one might place an advertisement in the newspaper to solicit “all current or former members of the armed forces who have served in Iraq” to join your sample.
- Thus, one can request that elements with specific characteristics join the sample.

Nonprobability Sampling

2. Purposive or Judgmental Sampling

- At the same time, through the broad appeal, the researcher might encourage some to join the sample that are not deemed appropriate by the researcher.
- For example, in the previous example, the researcher might be interested in interviewing combat veterans, but the request for respondents might encourage those without combat experience in Iraq to join the sample.



Nonprobability Sampling

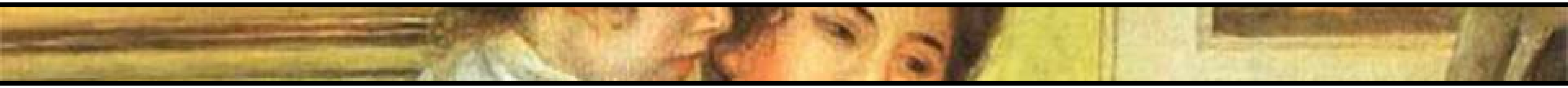
3. Snowball Sampling

- Selection of individuals who are recommended by others already selected.
- This procedure is appropriate for difficult to locate populations or persons with specific characteristics:
    - Vietnam veterans who fought in a specific area of the country.
    - Influential leaders in a community.
    - Persons who wish to remain anonymous, but who will respond to introductions from their associates.

Nonprobability Sampling

3. Snowball Sampling

- Snowball sampling allows the researcher to screen potential members of the sample, thereby building a sample of only those whom the researcher wants to study.
- Snowball sampling takes more time and money to implement than purposive sampling.



Nonprobability Sampling

4. Quota Sampling

- Selection of individuals to fill a quota for a certain characteristic.
- This procedure is appropriate for building a representative sample.
- One must have an accurate depiction of the total sample.
- Filling out some cells in the quota might require an unreasonable amount of resources.

Nonprobability Sampling

4. Quota Sampling

- Quota sampling has the advantage of collecting information from elements of interest. For example, the researcher might want to survey 100 males and 100 females. So, the researcher continues to contact individuals until the sample has 100 males and 100 females.

Nonprobability Sampling

4. Quota Sampling

- If the characteristics of interest become too complicated, however, it can be difficult to fill all the cells of a quota sample.

- For example, the researcher might have to contact many persons before finding 100 white females, aged 65+, with a college education to join the sample.

Nonprobability Sampling

5. Key Informants

- Selection of individuals who know information about other individuals or events.
- Assurances of confidentiality can become important in this type of sampling.

Nonprobability Sampling

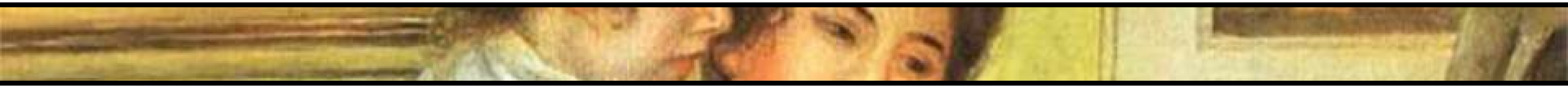
6. Summary

- Nonprobability sampling sometimes is the only reasonable procedure for building a sample.
- Probability samples are unnecessary for studies aimed at theory building or testing, wherein the researcher is not attempting to generalize the findings to a population.

Nonprobability Sampling

6. Summary

- Nonprobability sampling typically is less expensive than probability sampling, but not in all cases.
- The theoretical assumptions necessary for inferential statistics requires a probability sample. Therefore, non-probability samples should not be used to make inferences to a population.



Probability Sampling

1. Definition

- A sample that selects subjects with a known probability.
- Probability samples are important when one wishes to generalize to the larger population because one knows how to weight the responses to fit the characteristics of the population.

Probability Sampling

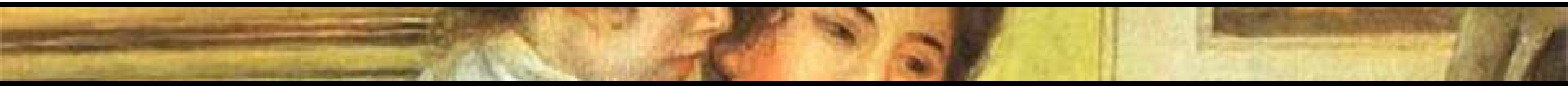
2. Conscious and Unconscious Sample Bias

- A biased sample is one whose characteristics do not match those of the population.
- If the sample is biased, and the responses are not weighted to reflect this bias, then generalizations to the population will be flawed.
- A randomly selected sample is not necessarily an unbiased one. A minority subpopulation, for example, might be missed or underrepresented when selecting a sample at random.

Probability Sampling

3. Representativeness

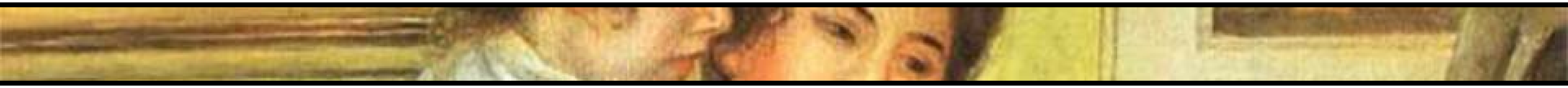
- Representativeness: The extent to which a sample has the same characteristics as the population.
  - Representativeness is judged by comparing selected characteristics.
  - Representativeness is not needed for accurate generalization to a population:
    - some characteristics are not important.
    - weighting can adjust differences between sample and population.



Probability Sampling

3. EPSEM

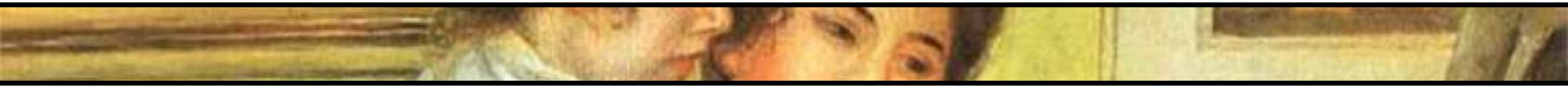
- Equal Probability of Selection Method: All members of a population have an equal probability of selection in the sample.
  - This is the basic principle of probability sampling.
  - Perfect representation still might not be achieved.
  - EPS is not always desirable. Sometimes, one wants to oversample some segments of a population.



Probability Sampling

3. EPSEM (Continued)

- Element: The unit from which information is collected.
  - An ISU student.
- Population: The aggregate of elements.
  - All ISU students.
- Study Population: The part of the population that is known or available to be sampled.
  - All ISU students properly registered.



Probability Sampling

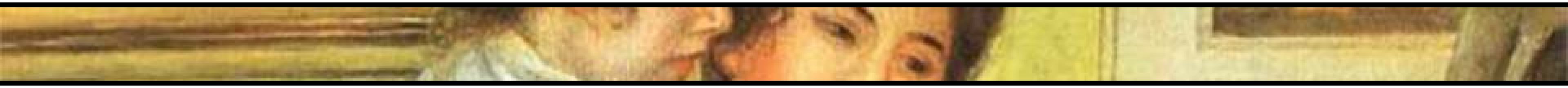
3. EPSEM (Continued)

- Sample Frame: The list from which the sample is drawn.
  - ISU students listed in the telephone directory.
- Sampling Unit: The element or set of elements considered for selection into the sample.
  - ISU students taking 12 or more hours this semester.

Probability Sampling

3. EPSEM (Continued)

- Initial Sample: Sampling units selected from the sample frame.
  - ISU students taking 12 or more hours this semester who were selected at random from the telephone directory.
- Final Sample: The elements who complete the survey.
  - ISU students in the initial sample who completed the survey.



Probability Sampling

4. Sampling Distributions

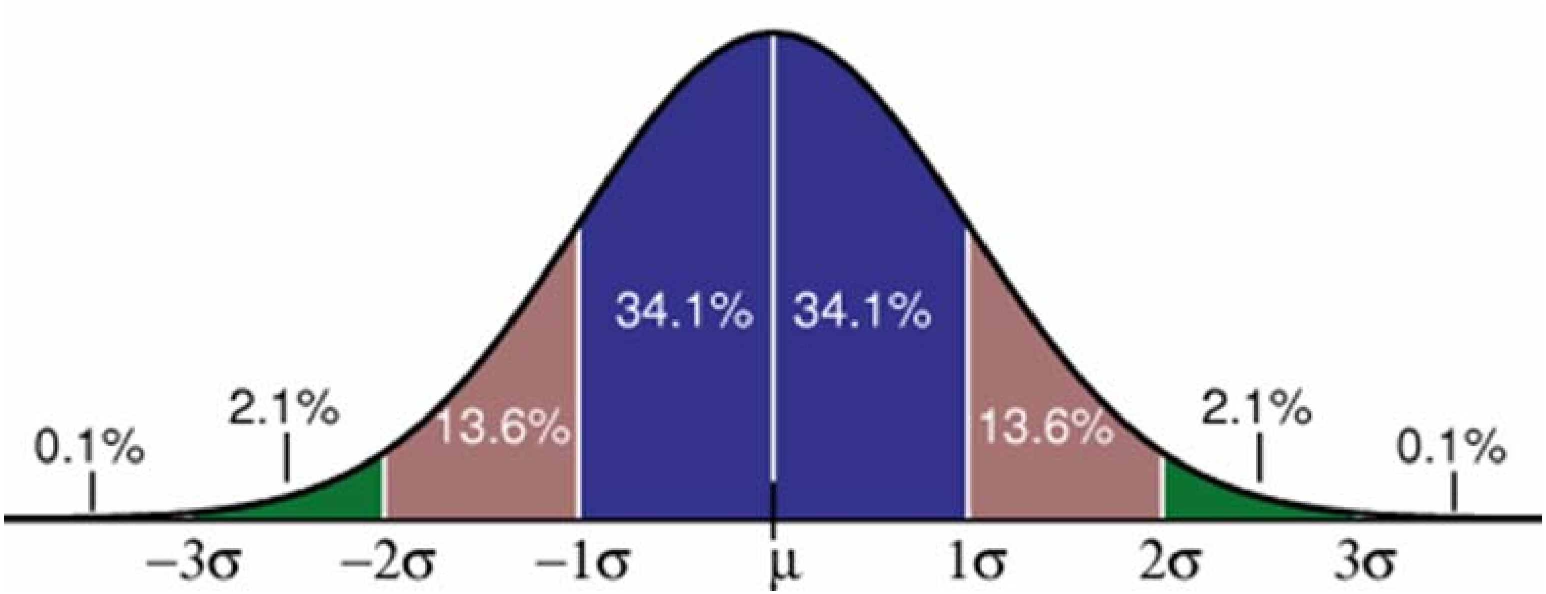
- Probability Theory: A branch of mathematics that provides the tools for estimating the representativeness of a sample.
- A key aspect of probability theory is the *Central Limit Theorem*: If the sum of variables has a finite variance (i.e., set end points), then it will be approximately normally distributed (i.e., have a bell-shaped curve).
    - A normal distribution sometimes is called a Gaussian distribution.

Probability Sampling

4. Sampling Distributions (Continued)

- The normal distribution is very useful because it defines boundaries by which to judge the representativeness of a sample.
- The key boundary of interest is the standard deviation, which is a range of values from the mean that includes a certain percentage of area beneath the bell-shaped curve.
- For example, one standard deviation accounts for all values from the mean included within ≈34.1% of the bell-shaped curve.

The Normal Distribution

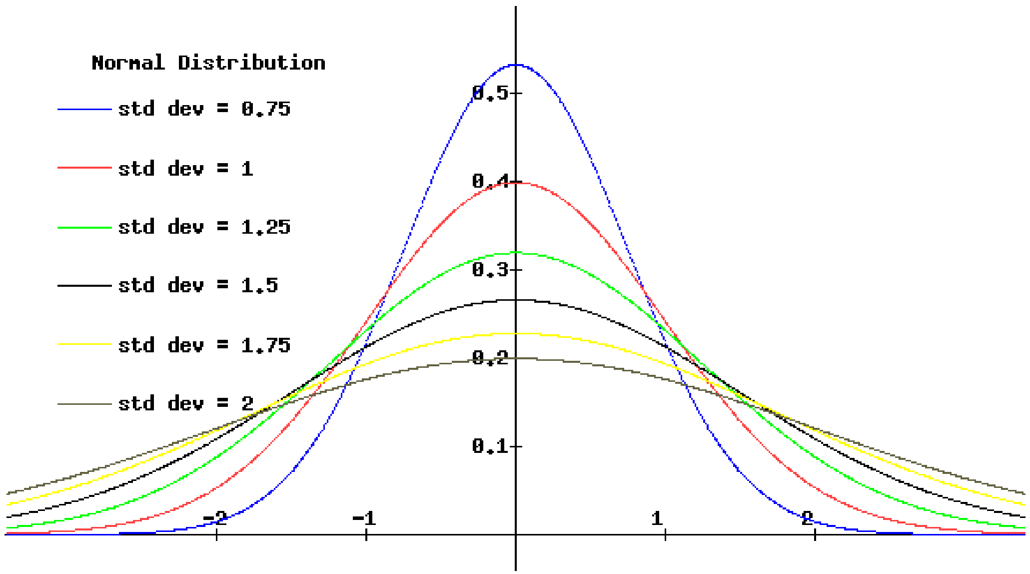


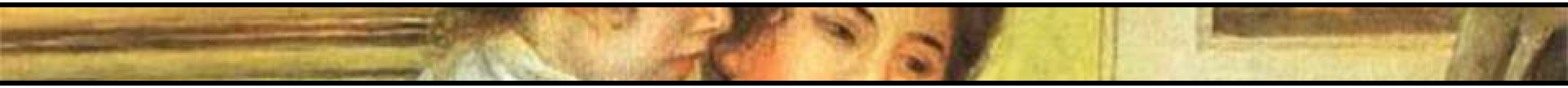
Each standard deviation (σ) represents a defined area from the mean (μ) beneath the curve.

Probability Sampling

4. Sampling Distributions (Continued)

- *Variance* is a statistic that represents how spread out the observations are from the mean.
- The standard deviation is the square root of the variance.

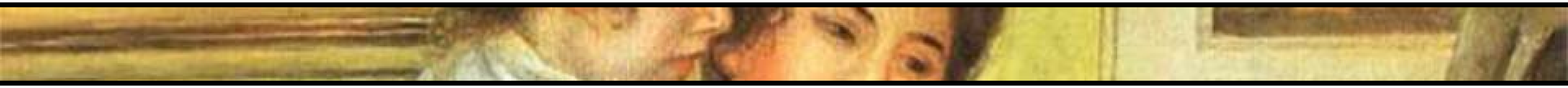




Probability Sampling

4. Sampling Distributions (Continued)

- Therefore, if one knows the mean and variance of the population and the mean and variance of the sample, one can estimate how closely the sample characteristics match the population characteristics using a standardized criterion of judgment: the bell-shaped, normal distribution.



Probability Sampling

4. Sampling Distributions (Continued)

- Parameter: A summary description of a variable in the *population* (e.g., the mean and standard deviation are parameters).
- Statistic: A summary description of a variable in the *sample*.
- Confidence Level: The amount of error the researcher is willing to tolerate (e.g., 5%)
- Confidence Interval: The range of values about a statistic where the parameter might be located for a given confidence level.

Types of Sampling Designs

1. Simple Random Sampling

1. Number the elements of the sample frame.
2. Generate **n** unique random numbers within the range of numbers assigned to the sample frame, where **n** = the size of the initial sample.
   - This is the simplest procedure for drawing a probability sample.
   - Might not capture minority elements of a sample frame.
   - Cannot draw independent samples for specific sub-populations.

Types of Sampling Designs

2. Systematic Sampling

1. Number the elements of the sample frame.
2. Determine the ratio (**k**) of **n** (sample size) to **N** (size of sample frame).
3. Generate one random number within the range of this ratio.
4. Select each **k**th element beginning with this random number.

Types of Sampling Designs

2. Systematic Sampling: Example

- Sample frame (**N**) = 10,000.
- Desired initial sample (**n**) = 1,000.
- Ratio (**k**) = 10.
- Generate a random number between 1 and 10.
- Suppose this number is 8.
- Begin with element 8 and pick each **k**th element (18, 28, 38...).

Types of Sampling Designs

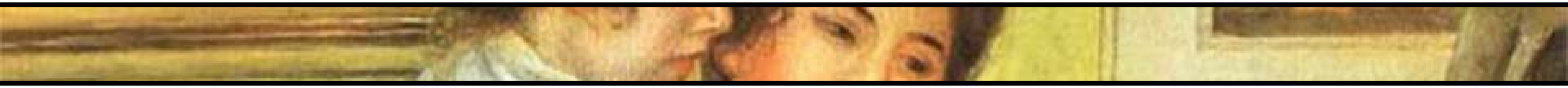
2. Systematic Sampling (Continued)

- This is a very simple procedure for drawing a probability sample.
- Certain **k** elements might be excluded (i.e., *periodicity*):
    - Elements might be ordered in certain ways to exclude or include specific elements.
        - The corner apartments might all end in number 9. If k = 9 then only corner apartments are selected. If k ≠ 9, then all corner apartments are excluded.

Types of Sampling Designs

3. Stratified Sampling

- In its simplest form, a stratified sample is a set of simple random samples selected from sub-segments of the sample frame.
- Example: One might select a simple random sample of males and a simple random sample of females.



Types of Sampling Designs

3. Stratified Sampling (Continued)

- Allows one to control the number of elements selected from each sub-segment of the sample frame.
- Homogeneous sub-samples will have smaller standard errors on parameter estimates than will more heterogeneous samples of the entire sample frame.
- More expensive and time consuming.
- Must know the size of each segment in the sample frame.

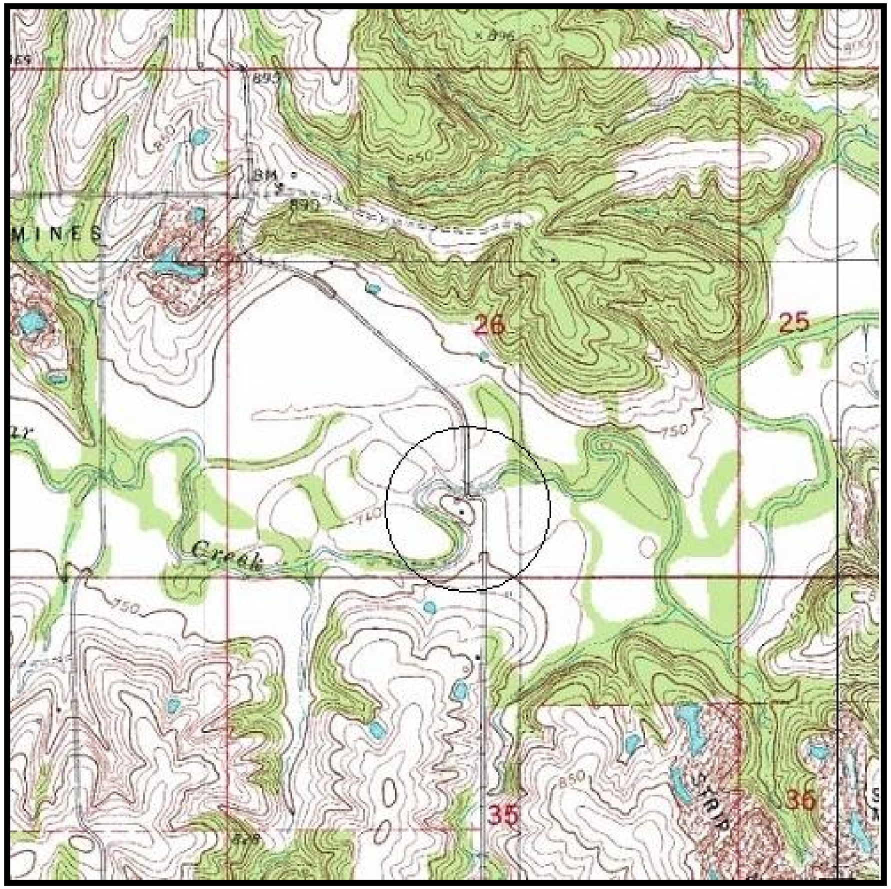
Types of Sampling Designs

4. Cluster Sampling

- Elements are divided into groups of equal number of elements (i.e., clusters).
- The clusters are selected at random.
- All elements within a cluster are included in the initial sample.
- ❑ Saves time and money for personal interviews.
- ❑ Do not have to know the exact size of the sample frame.
- ❑ Increased sampling error because of the clustering procedure.

Types of Sampling Designs

4. Cluster Sampling (Continued)



Types of Sampling Designs

5. Multistage Cluster Sampling

- Clusters are selected at random.
- Elements are selected at random within each cluster.

- ❑ Saves time and money for personal interviews.
- ❑ Do not have to know the exact size of the sample frame.
- ❑ Increased sampling error because of the clustering procedure.
- ❑ Increased sampling error because of the selection of elements within a cluster.

Types of Sampling Designs

5. Multistage Cluster Sampling: Example

- The city has 10,000 households.
- The city has 1,000 blocks of 10 hh each.
- We want an initial sample of 500.
- We want to select 1/20 households.
- In Stage 1, we select 100 blocks.
- In Stage 2, we select 5 hh per block.
- Probability of selection for each hh:
  - 1/10 (block) x ½ (hh in block) = 1/20.

Types of Sampling Designs

6. Probability Proportionate to Size (PPS)

- What if a few city blocks contain many more households than others, and we anticipate that density of housing is an important characteristic that will affect our study?
- Then, we want to select clusters and households proportionate to the number of households in each city block to insure that we select households from the large city blocks.

Types of Sampling Designs

6. PPS Sampling: Example

- The city has 10,000 households.
- The city has 110 blocks.
- 10 blocks contain 500 hh each.
- 100 blocks contain 50 hh each.
- We want an initial sample of 500 hh.
- If we selected blocks at random, we might miss all 10 of the very large blocks.
- So, we use PPS sampling.

Types of Sampling Designs

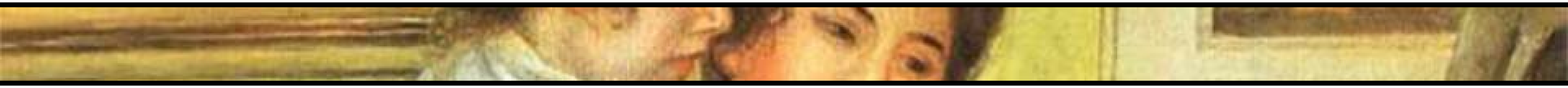
6. PPS Sampling: Example

- The 10 large blocks contain 1/2 of the hh.
- So, we want to select blocks and hh so that we obtain 1/2 of our sample from the large blocks and 1/2 from the small blocks.
- We decide to select 50 hh from each block.
- Therefore, we need to select 5 of the 10 large blocks and 5 of the small blocks.
- Probabilities:
  - Large blocks: 1/2 (block) x 1/10 (hh) = 1/20.
  - Small blocks: 1/20 (block) x 1/1 (hh) = 1/20.

Types of Sampling Designs

7. Weighting

- Suppose we want to survey a city with a population of 4,500 whites and 500 blacks.
- We want a sample of 1/10 = 500.
- If we selected at random, we would obtain only 50 blacks in our initial sample.
- We might want to over sample blacks to improve the validity and reliability of our estimates of their opinions.
- If we do so, we need to adjust the weights of their opinions when we generalize to the total population.



Types of Sampling Designs

7. Weighting (Continued)

	Whites	*Blacks*
Number in population……………………	4,500	500
Percentage of population……………….	90	10
Sampling fraction (oversample blacks)..	1/10	1/5
Number in initial sample……………….	450	100
Unweighted percentage of sample……..	81.8	18.2
Weight (to adjust for oversampling)…….	1	1/2
Weighted number in initial sample………	450	50
Weighted percentage of initial sample….	90	10

Questions?