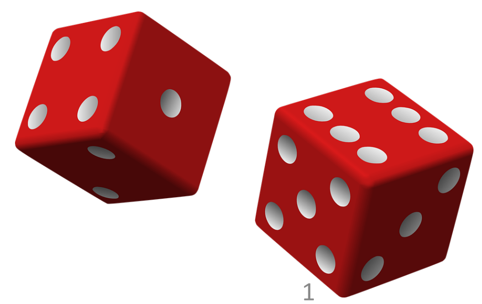
Sampling People, Records, & Networks

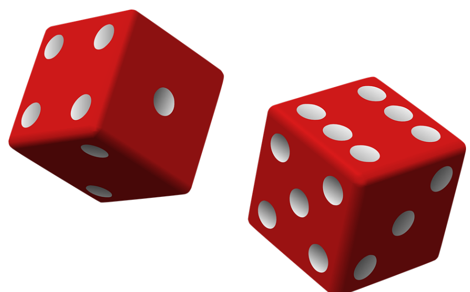
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



Unit 6

1. Stratified multistage sampling
2. Weights for over/under sampling
3. Nonresponse & noncoverage weighting
4. Variance estimation and software
5. Statistical software for sample selection
6. Sampling networks: multiplicity weighting

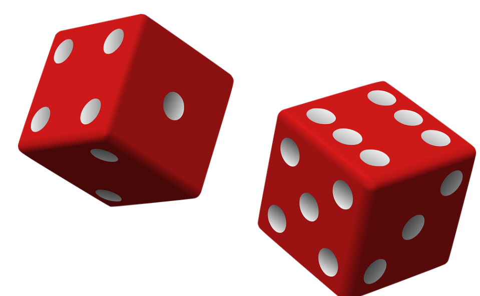
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



Unit 6

1. Stratified multistage sampling
2. Weights for over/under sampling
3. Nonresponse & noncoverage weighting
4. Variance estimation and software
5. Statistical software for sample selection
6. Sampling networks: multiplicity weighting

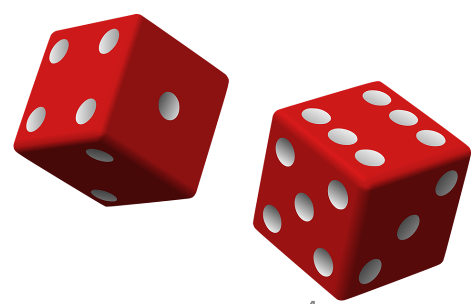
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- **Unit 6: Some extensions & applications**
  1. Statistical software for sample selection
  2. Stratified multistage sampling
  3. Weights for over/under sampling
  4. Nonresponse & noncoverage weighting
  5. Sampling networks: multiplicity weighting
  6. Non-probability sampling



Unit 6

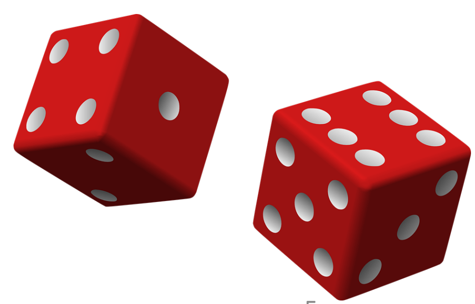
1. Statistical software for sample selection
2. Stratified multistage sampling
3. **Weights for over/under sampling**
4. Nonresponse & noncoverage weighting
5. Sampling networks: multiplicity weighting
6. Non-probability sampling

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- **Unit 6: Some extensions & applications**
  1. Statistical software for sample selection
  2. Stratified multistage sampling
  3. **Weights for over/under sampling**
  4. Nonresponse & noncoverage weighting
  5. Sampling networks: multiplicity weighting
  6. Non-probability sampling

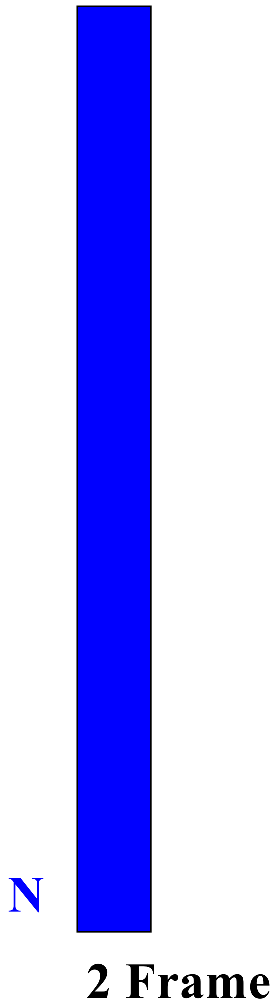


- **Weighting framework**
- **Weighted estimation**
- **Weighting and over-sampling**

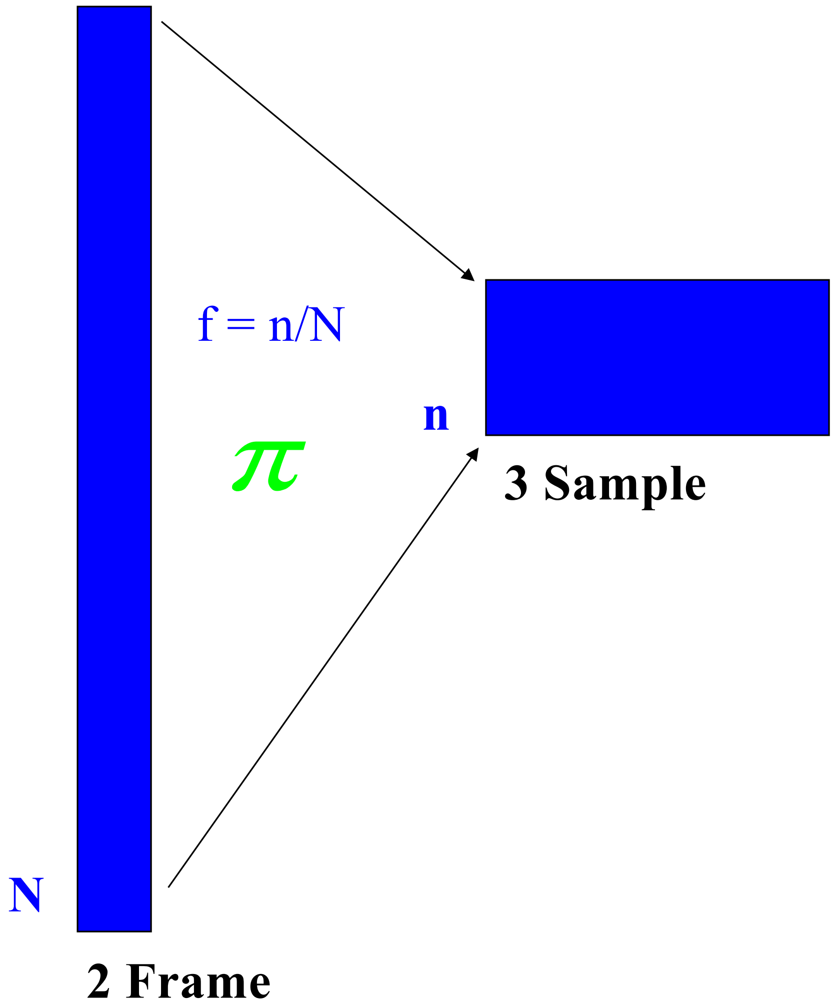
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications
    1. Statistical software for sample selection
    2. Stratified multistage sampling
    3. **Weights for over/under sampling**
    4. Nonresponse & noncoverage weighting
    5. Sampling networks: multiplicity weighting
    6. Non-probability sampling



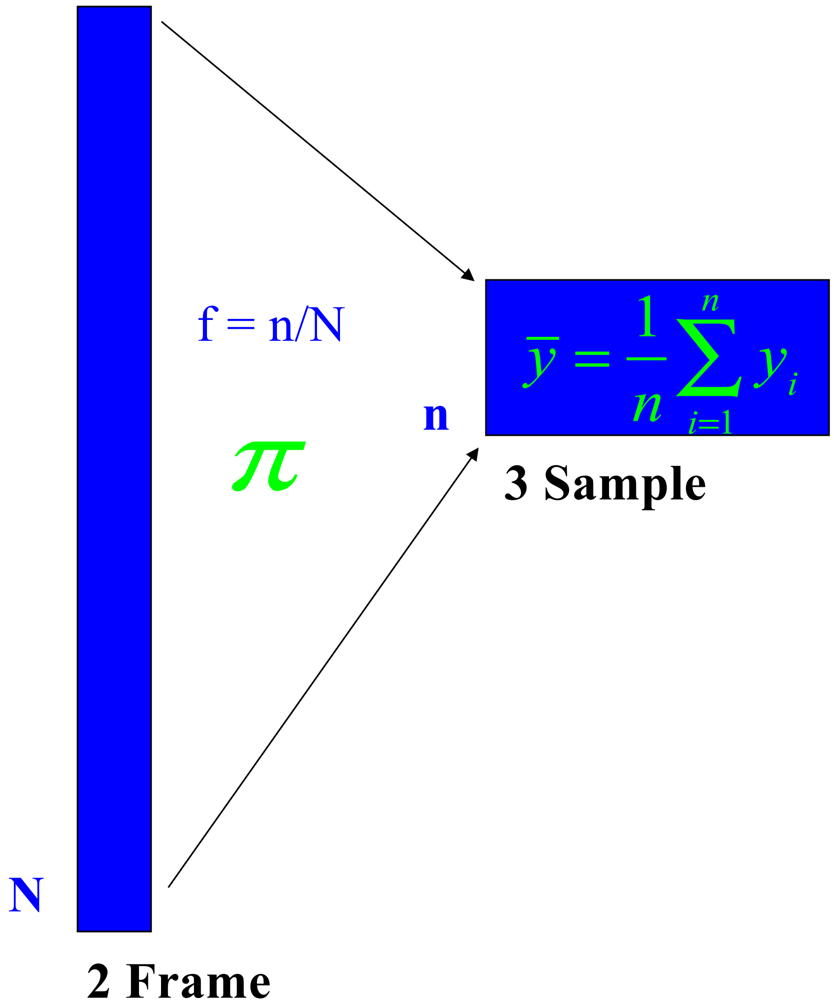
- **Weighting framework**
- Weighted estimation
- Weighting and over-sampling



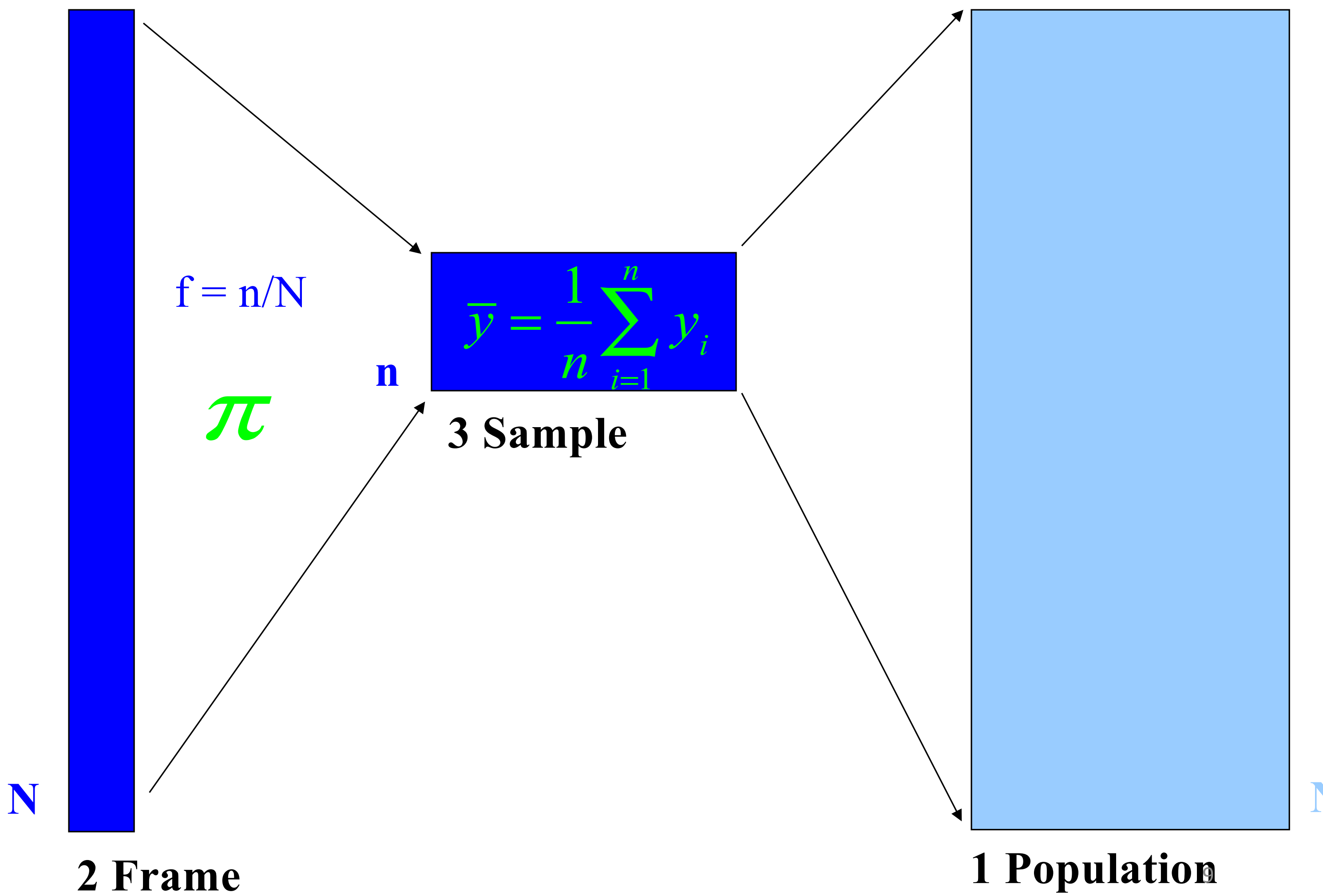
- **Weighting framework**
- Weighted estimation
- Weighting and over-sampling



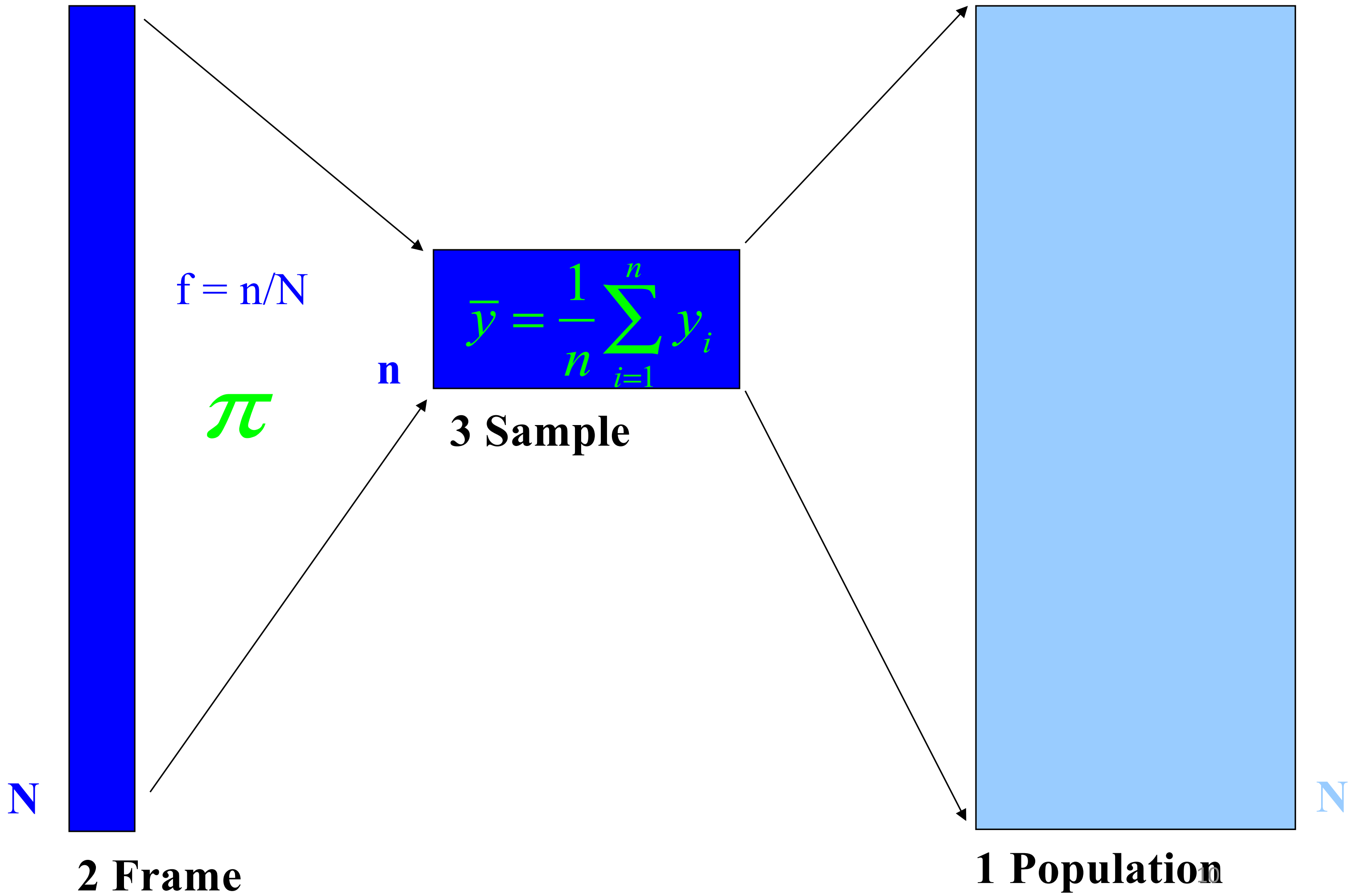
- **Weighting framework**
- Weighted estimation
- Weighting and over-sampling



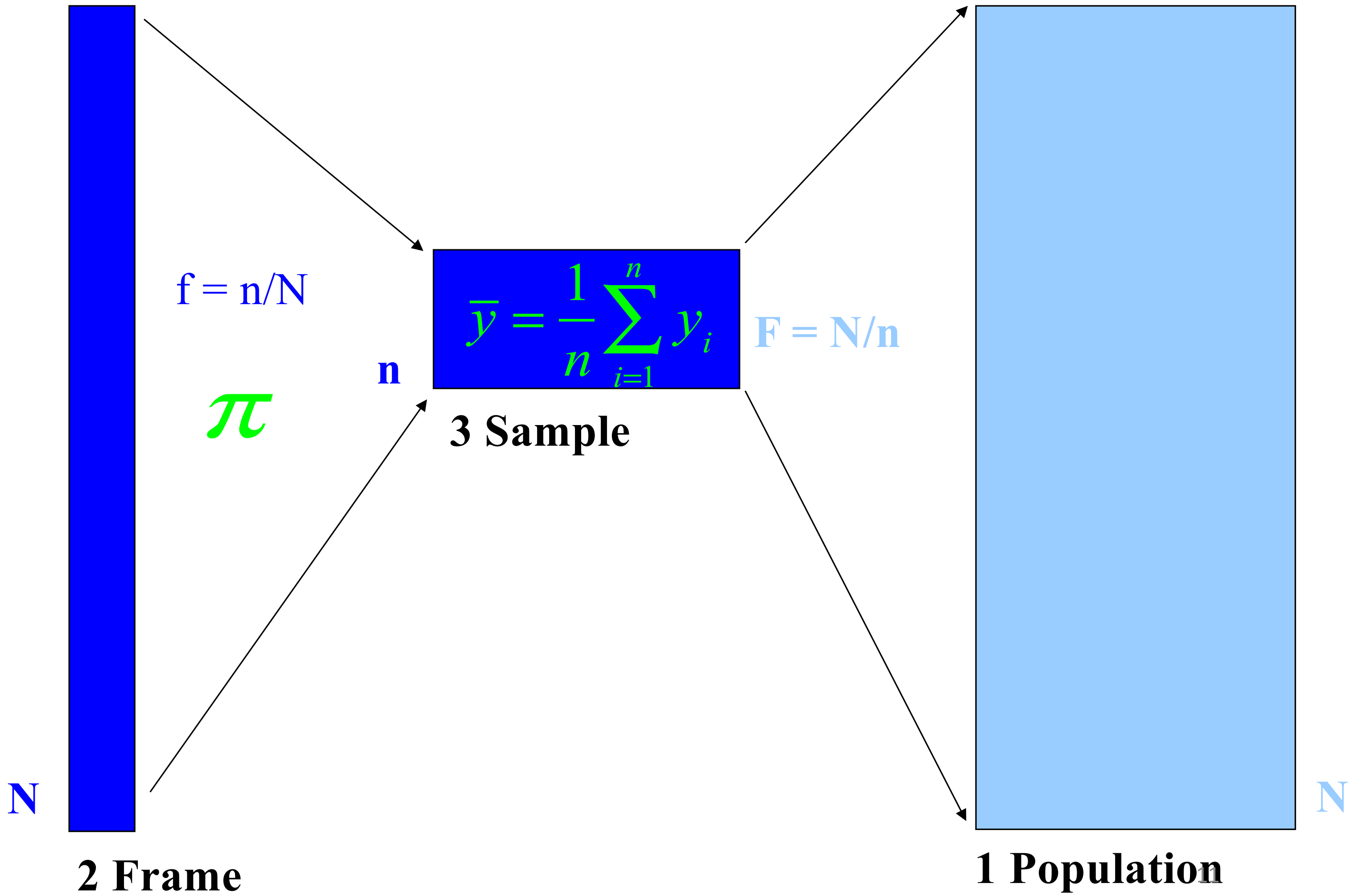
- **Weighting framework**
- Weighted estimation
- Weighting and over-sampling



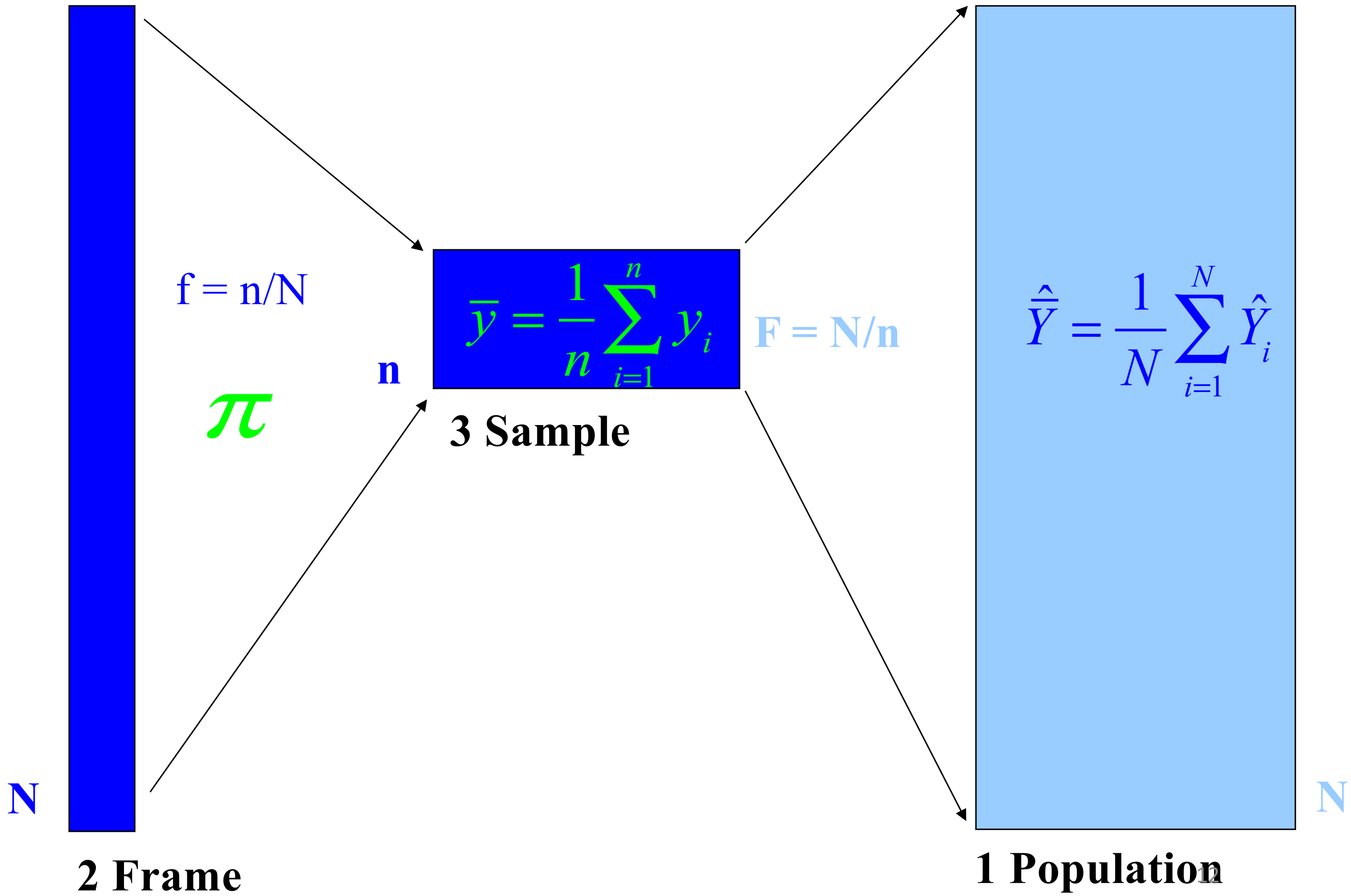
- **Weighting framework**
- Weighted estimation
- Weighting and over-sampling



- **Weighting framework**
- Weighted estimation
- Weighting and over-sampling



- **Weighting framework**
- Weighted estimation
- Weighting and over-sampling



- Weighting framework
- **Weighted estimation**
- Weighting and over-sampling

- **An *epsem* sampling system…**

$$\pi_i = \pi = f = n/N$$

- Weighting framework
- **Weighted estimation**
- Weighting and over-sampling

- **An *epsem* sampling system…**

$$\pi_i = \pi = f = n/N$$

- **Then an *epsem* estimation system …**

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{y_1 + y_2 + \cdots + y_n}{1 + 1 + \cdots 1}$$

- Weighting framework
- **Weighted estimation**
- Weighting and over-sampling

- **An *epsem* sampling system…**

$$\pi_i = \pi = f = n/N$$

- **Then an *epsem* estimation system …**

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{y_1 + y_2 + \cdots + y_n}{1+1+\cdots 1}$$

- **For example, from *N* = 2,000 people, select *n* = 20 with *epsem***

$$\pi_i = \frac{20}{2000} = \frac{1}{100} \quad and \quad w_i = 100$$

- **Each person represents themself and 99 others**

- Weighting framework
- **Weighted estimation**
- Weighting and over-sampling

- **But the mapping may not be equal for every element – a non-*epsem* design:**

$$\pi_i \neq \pi = f = n/N$$

- Weighting framework
- **Weighted estimation**
- Weighting and over-sampling

- **But the mapping may not be equal for every element – a non-*epsem* design:**

$$\pi_i \neq \pi = f = n/N$$

- **A weighted estimator is required …**

$$\bar{y}_w = \frac{\sum_i w_i y_i}{\sum_i w_i} = \frac{w_1 \cdot y_1 + w_2 \cdot y_2 + \cdots + w_n \cdot y_n}{w_1 \cdot 1 + w_2 \cdot 1 + \cdots + w_n \cdot 1}$$

- Weighting framework
- **Weighted estimation**
- Weighting and over-sampling

- **But the mapping may not be equal for every element – a non-*epsem* design:**

$$\pi_i \neq \pi = f = n/N$$

- **A weighted estimator is required …**

$$\bar{y}_w = \frac{\sum_i w_i y_i}{\sum_i w_i} = \frac{w_1 \cdot y_1 + w_2 \cdot y_2 + \cdots + w_n \cdot y_n}{w_1 \cdot 1 + w_2 \cdot 1 + \cdots + w_n \cdot 1}$$

- **The unweighted mean is a special case of the weighted – constant weights cancel**

- Weighting framework
- **Weighted estimation**
- Weighting and over-sampling

- **The basic approach: weight by** $1/\pi_i$
  - **Counting a sample person** $1/\pi_i$ **times**

- Weighting framework
- **Weighted estimation**
- Weighting and over-sampling

- **The basic approach: weight by $1/\pi_i$**
  - **Counting a sample person $1/\pi_i$ times**
- **Consider the following population distribution for 10th grade students in the U.S.**

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	N	n	Sampling rate	Weight A	Weight B
High	800,000	2,400	1/333.33	333.33	1
Low	3,200,000	9,600	1/333.33	333.33	1
Total	**4,000,000**	**12,000**	1/333.33	333.33	1

- Weighting framework
- **Weighted estimation**
- Weighting and over-sampling

- **The basic approach: weight by $1/\pi_i$**
  - **Counting a sample person $1/\pi_i$ times**
- **Consider the following population distribution for 10th grade students in the U.S.**
- **Divided into two groups, 10th graders in schools with a high proportion receiving Free or Reduced Price Lunches (High) and those in low proportion schools (Low)**

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	N	n	Sampling rate	Weight A	Weight B
High	800,000	2,400	1/333.33	333.33	1
Low	3,200,000	9,600	1/333.33	333.33	1
Total	**4,000,000**	**12,000**	1/333.33	333.33	1

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	N	n	Sampling rate	Weight A	Weight B
High	800,000	2,400	1/333.33	333.33	1
Low	3,200,000	9,600	1/333.33	333.33	1
Total	**4,000,000**	**12,000**	1/333.33	333.33	1

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	N	n	Sampling rate	Weight A	Weight B
High	800,000	2,400	1/333.33	333.33	1
Low	3,200,000	9,600	1/333.33	333.33	1
Total	**4,000,000**	**12,000**	1/333.33	333.33	1

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

- **This is an allocation of sample across the strata that is proportionate.**

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

- **This is an allocation of sample across the strata that is proportionate.**
- **Proportionate allocation has equal probabilities in each group**

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

- This is an allocation of sample across the strata that is proportionate.
- Proportionate allocation has equal probabilities in each group
- Some investigators might prefer that the distribution in the sample be an equal sample size across the two groups:

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	N	n	Sampling rate	Weight A	Weight B
High	800,000	6,000	1/133.33	133.33	1
Low	3,200,000	6,000	1/533.33	533.33	4
Total	**4,000,000**	**12,000**	1/333.33	--	--

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	N	n	Sampling rate	Weight A	Weight B
High	800,000	6,000	1/133.33	133.33	1
Low	3,200,000	6,000	1/533.33	533.33	4
Total	**4.000,000**	**12,000**	1/333.33	--	--

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

- This equal allocation would be used for comparing the two groups
- The proportionate allocation would be used to represent the population

- Weighting framework
- Weighted estimation
- Weighting and over-sampling

- This equal allocation would be used for comparing the two groups
- The proportionate allocation would be used to represent the population
- Consider the consequences of the equal allocation when estimating a mean test score among 10th graders, averaging across samples from the two groups:

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	Mean test score	Proportionate allocation		Weights	
		n	Mean test score	A	B
High	72	2,400	72	333.33	1
Low	92	9,600	92	333.33	1
Total	**88**	**12,000**	**88**	**333.33**	1

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	Mean test score	Proportionate allocation		Weights	
		n	Mean test score	A	B
High	72	2,400	72	333.33	1
Low	92	9,600	92	333.33	1
Total	**88**	**12,000**	**88**	**333.33**	1

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	Mean test score	Dispro-portionate allocation		Weights		Weighted estimate
		n	Mean test score	A	B	
High	72	6,000	72	133.33	1	(6,000)(1)(72)
Low	92	6,000	92	533.33	4	(6,000)(4)(92)
Total	**88**	**12,000**	**82**	**--**	**--**	**88**

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

Group	Mean test score	Dispro-portionate allocation		Weights		Weighted estimate
		n	Mean test score	A	B	
High	72	6,000	72	133.33	1	(6,000)(1)(72)
Low	92	6,000	92	533.33	4	(6,000)(4)(92)
Total	**88**	**12,000**	**82**	**--**	**--**	**88**

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

- **Weights will restore the population distribution:**

$$\bar{y} = \frac{\sum y_i}{n} = \frac{6{,}000 \times 72 + 6{,}000 \times 92}{6{,}000 + 6{,}000} = 82$$

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

- **Weights will restore the population distribution:**

$$\bar{y} = \frac{\sum y_i}{n} = \frac{6{,}000 \times 72 + 6{,}000 \times 92}{6{,}000 + 6{,}000} = 82$$

$$y_{w(B)} = \frac{\sum w_{i(B)} y_i}{\sum w_{i(B)}} = \frac{6{,}000 \times 4 \times 92 + 6{,}000 \times 1 \times 72}{6{,}000 \times 4 + 6{,}000 \times 1} = 88$$

- Weighting framework
- Weighted estimation
- **Weighting and over-sampling**

- **Weights will restore the population distribution:**

$$\bar{y} = \frac{\sum y_i}{n} = \frac{6{,}000 \times 72 + 6{,}000 \times 92}{6{,}000 + 6{,}000} = 82$$

$$y_{w(B)} = \frac{\sum w_{i(B)} y_i}{\sum w_{i(B)}} = \frac{6{,}000 \times 4 \times 92 + 6{,}000 \times 1 \times 72}{6{,}000 \times 4 + 6{,}000 \times 1} = 88$$

$$y_{w(A)} = \frac{\sum w_{i(A)} y_i}{\sum w_{i(A)}} = \frac{6{,}000 \times 533.33 \times 92 + 6{,}000 \times 133.33 \times 72}{6{,}000 \times 533.33 + 6{,}000 \times 133.33} = 88$$

Unit 6

1. Statistical software for sample selection
2. Stratified multistage sampling
3. Weights for over/under sampling
4. **Nonresponse & noncoverage weighting**
5. Sampling networks: multiplicity weighting
6. Non-probability sampling

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- **Unit 6: Some extensions & applications**
  1. Statistical software for sample selection
  2. Stratified multistage sampling
  3. Weights for over/under sampling
  4. **Nonresponse & noncoverage weighting**
  5. Sampling networks: multiplicity weighting
  6. Non-probability sampling