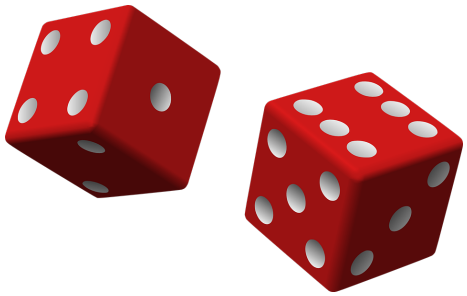
Sampling People, Records, & Networks

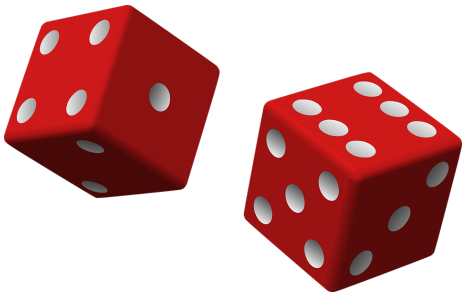
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



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

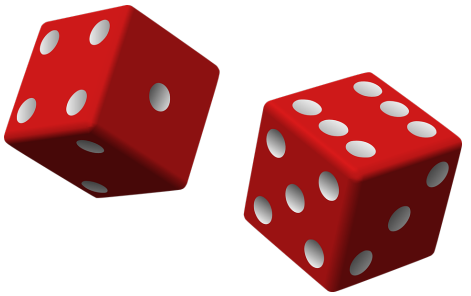
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



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

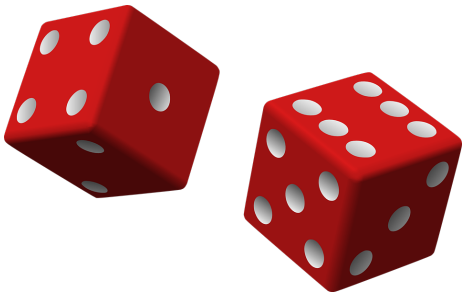
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- **Unit 6: Some extensions & applications**
  1. Statistical software for sample selection
  2. Stratified multistage sampling
  3. Weights for over/under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



Unit 6

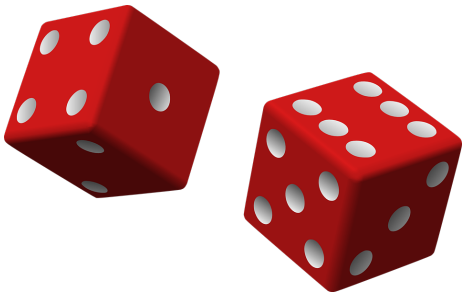
1. Statistical software for sample selection
2. Stratified multistage sampling
3. Weights for over/under sampling
4. **Nonresponse & noncoverage weighting**
5. Sampling networks: multiplicity weighting
6. Non-probability sampling

- 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



- Nonresponse
- Post-stratification & non-coverage
- A final weight

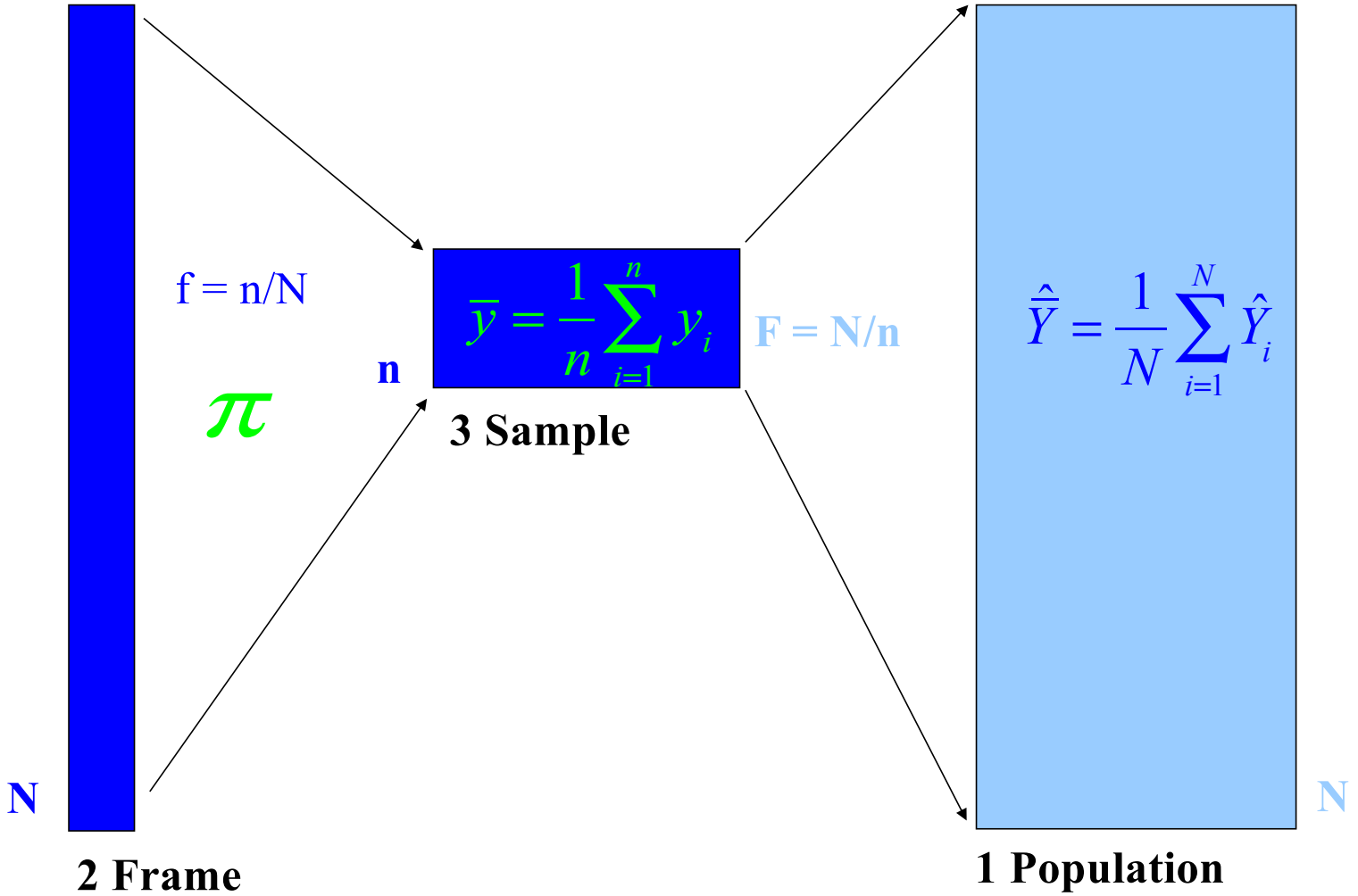
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications
  1. Statistical software for sample selection
  2. Stratified multistage sampling
  3. Weights for over/under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



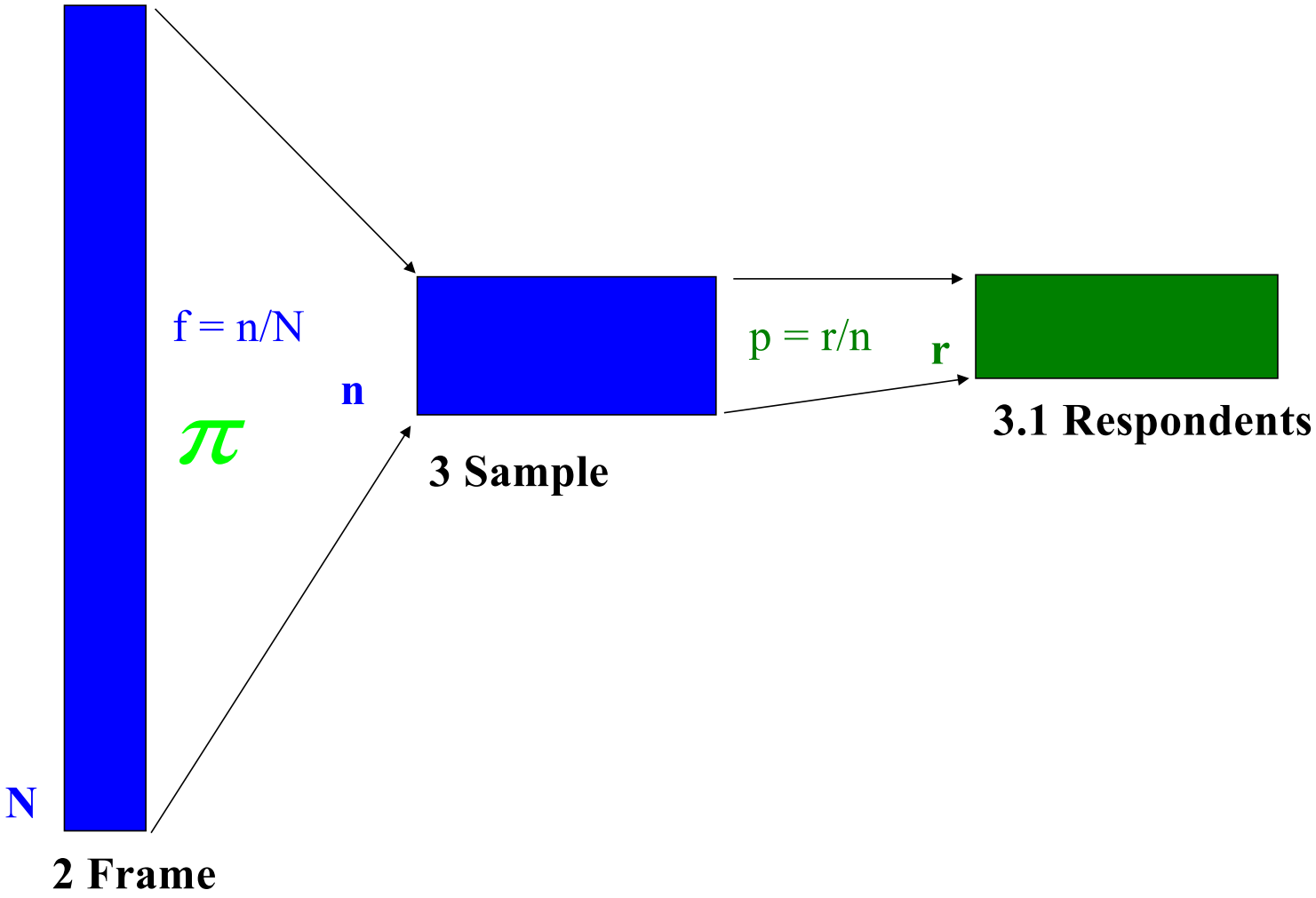
- Nonresponse
- Post-stratification & non-coverage
- A final weight

- Suppose that not everyone in the sample of 12,000 drawn from the two groups respond
- Ignoring nonresponse may produce biased estimates

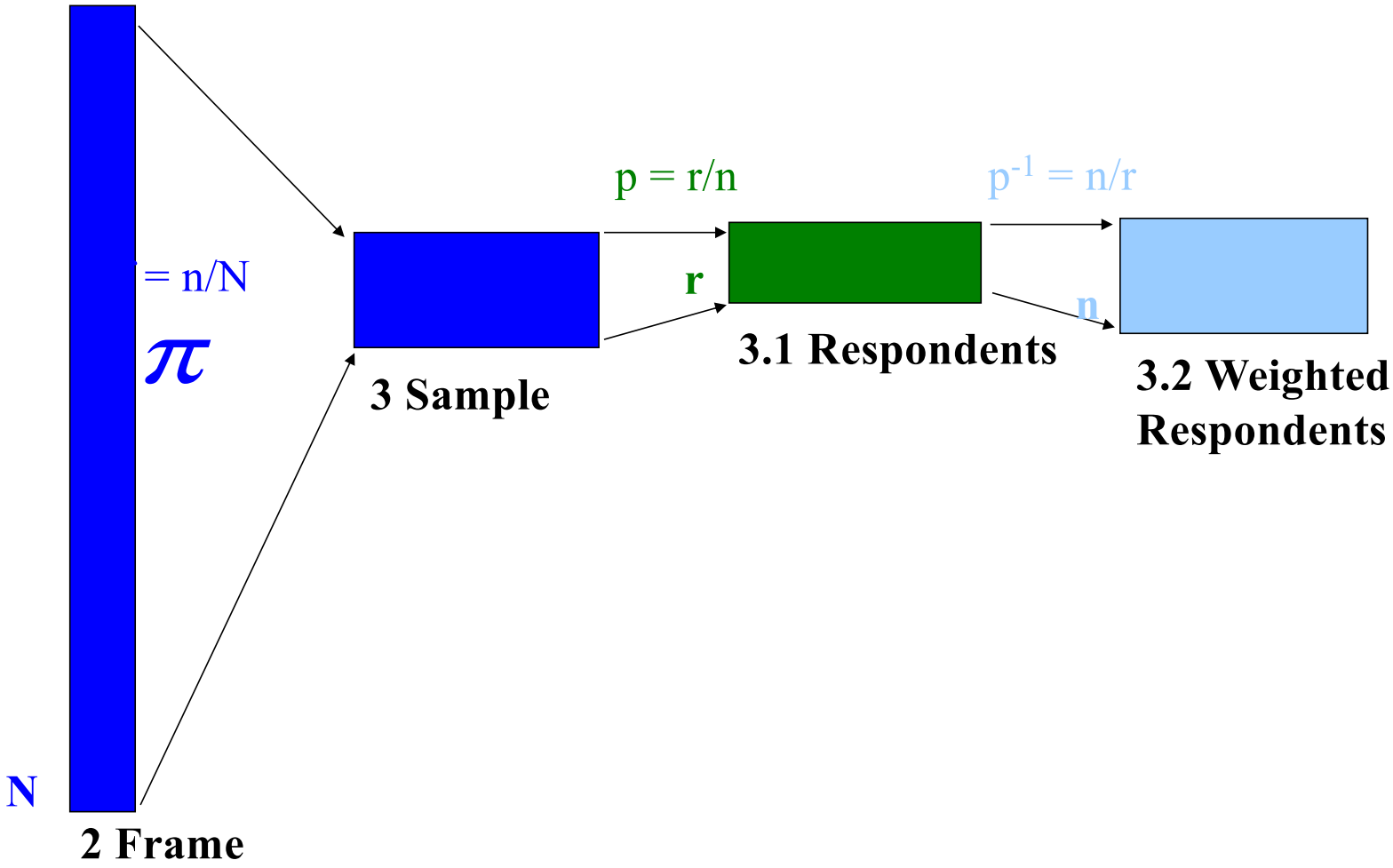
- **Nonresponse**
- Post-stratification & non-coverage
- A final weight



- **Nonresponse**
- Post-stratification & non-coverage
- A final weight



- **Nonresponse**
- Post-stratification & non-coverage
- A final weight



- Nonresponse
- Post-stratification & non-coverage
- A final weight

- **Biased estimates may be produced when averaging across potentially disproportionately-distributed groups**
- **Consider the disproportionate equal sample size allocation for 10th grade students**
- **Suppose, that the response rate across 10th grade student location (urban, rural school) differs:**

- **Nonresponse**
- Post-stratification & non-coverage
- A final weight

Group	n	r	Mean score	Respondents
Metro	8,000	5,600	60	(60)
Non-metro	4,000	3,400	75	(75)
Total	**12,000**	**9,000**	**65**	**65.67**

- Nonresponse
- Post-stratification & non-coverage
- A final weight

- Compute response rates in each group

- **Nonresponse**
- Post-stratification & non-coverage
- A final weight

Location	n_h	r_h	$\frac{r_h}{n_h}$	$w_{2i} = \frac{n_h}{r_h}$
Metro	8,000	5,600	0.7	1.43
Non-metro	4,000	3,400	0.85	1.18
Total	**12,000**	**9,000**	**0.75**	**--**

- Nonresponse
- Post-stratification & non-coverage
- A final weight

- Compute response rates in each group
- Adjust the base weights (those computed to compensate for unequal probabilities of selection) for nonresponse – a product of weights

- **Nonresponse**
- Post-stratification & non-coverage
- A final weight

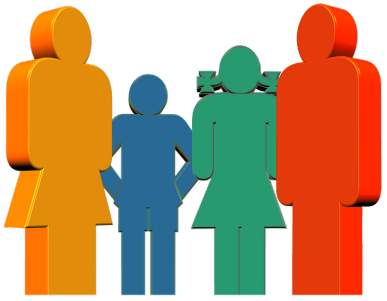
FRPL	Location	w_{1i}	r_h	$\frac{r_h}{n_h}$	$\frac{n_h}{r_h}$	$w_{2i} = w_{1i}\frac{n_h}{r_h}$
High	Metro	1	2,800	0.70	1.43	1.43
	Non-metro	1	1,700	0.85	1.18	1.18
Low	Metro	4	2,800	0.70	1.43	5.72
	Non-metro	4	1,700	0.85	1.18	4.72
	Total		**9,000**	**0.70**		

- Nonresponse
- Post-stratification & non-coverage
- A final weight

- **Assumption: data is missing at random (MAR)**
- **Response rate in each group is a "sampling rate" under the MAR assumption**

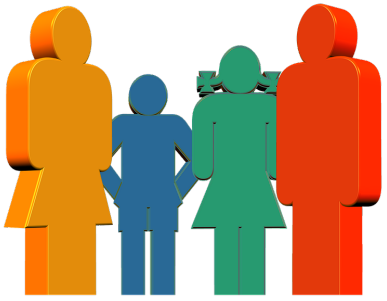
- Nonresponse
- Post-stratification & non-coverage
- A final weight

- **Weighting classes: cross-classification of multiple variables**
  - Choice of variables: stepwise regression, 'effect sizes'
  - Choose variable related both the "propensity" and the variables ("prediction")
- **Logistic regression**
  - Using good propensity/prediction variables, estimate logistic regression model for response
  - Inverse of predicted probabilities as the weight



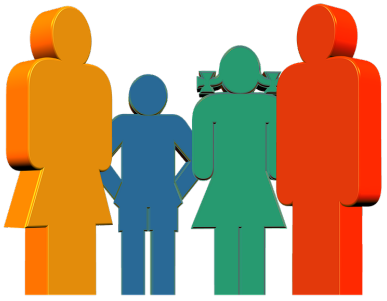
- Nonresponse
- Post-stratification & non-coverage
- A final weight

- Poststratification is used to make the weighted sample distribution conform to a known population distribution
- Typically poststratification adjusts the nonresponse adjusted weights
- Suppose that family type (single parent, other) is not known in advance for each sample 10th grade student, but is only obtained in data collection

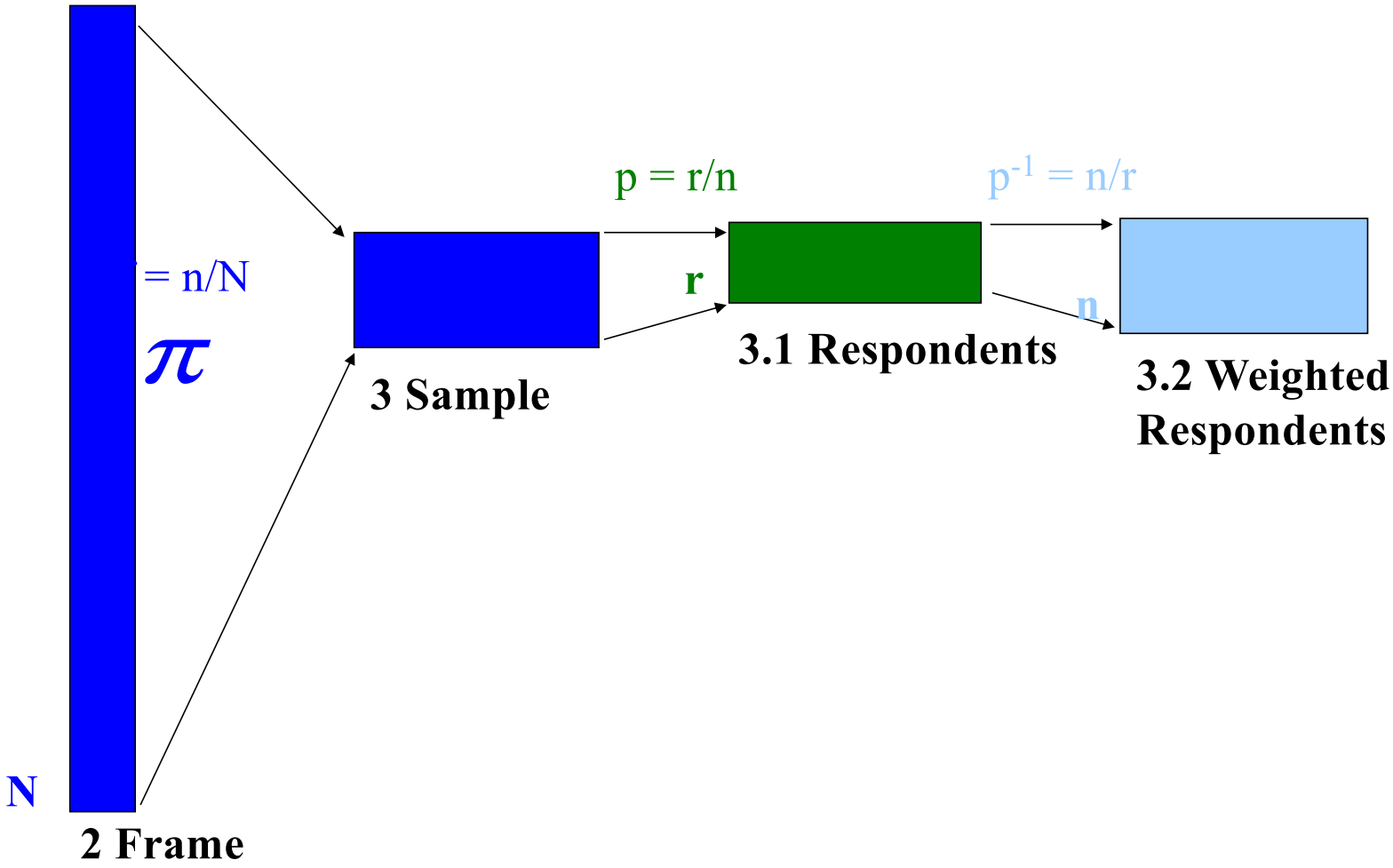


- Nonresponse
- Post-stratification & non-coverage
- A final weight

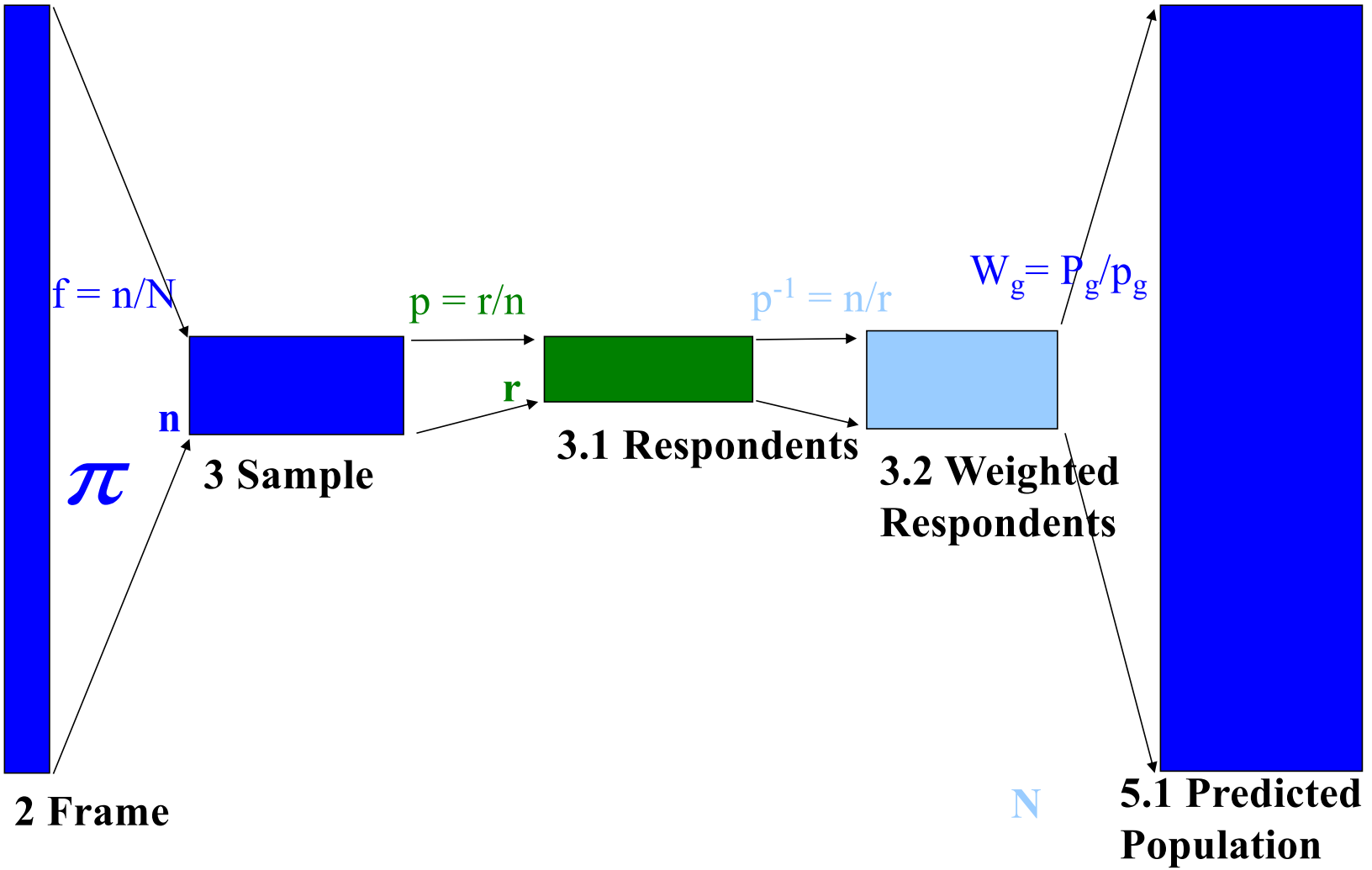
- Suppose that family type (single parent, other) is not known in advance for each sample 10th grade student, but is only obtained in data collection
- Suppose also that from recent Census data the proportion of 10th grade students' living with a single parent was tabulated



- Nonresponse
- **Post-stratification & non-coverage**
- A final weight



- Nonresponse
- **Post-stratification & non-coverage**
- A final weight

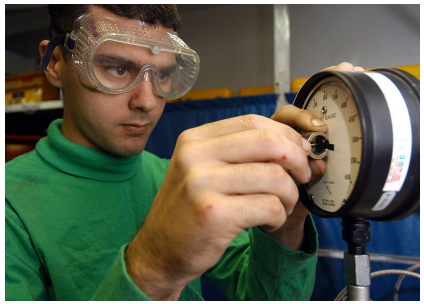


- Nonresponse
- **Post-stratification & non-coverage**
- A final weight

Family Type	Wtd n_g	p_g	N_g	P_g	$w_g = P_g / p_g$
Single parent	1,800	0.2	1,200,000	0.3	1.500
Other	7,200	0.8	2,800,000	0.7	0.875
Total	**9,000**	**1.0**	**4,000,000**	**1.000**	**--**

- Nonresponse
- **Post-stratification & non-coverage**
- A final weight

- **In poststratification, the weights for the individuals in groups are adjusted up or down to obtain the distribution of the sum of weights that corresponds to the population distribution**
- **The final weight is an adjustment of the baseline weight for nonresponse and poststratification:**



- Nonresponse
- Post-stratification & non-coverage
- **A final weight**

Group			n_{hcg}	$w_{3i} = w_{1i} \times w_{2i} \times w_{gi}$
High	Metro	Single parent	560	1 x 1.43 x 1.500 = 2.145
		Other	2,240	1 x 1.43 x 0.875 = 1.125
	Non-metro	Single parent	340	1 x 1.18 x 1.500 = 1.770
		Other	1,360	1 x 1.18 x 0.875 = 1.251
Low	Metro	Single parent	560	4 x 1.43 x 1.500 = 8.580
		Other	2,240	4 x 1.43 x 0.875 = 5.005
	Non-metro	Single parent	340	4 x 1.18 x 1.500 = 7.080
		Other	1,360	4 x 1.18 x 0.875 = 4.130
Total			**9,000**	

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