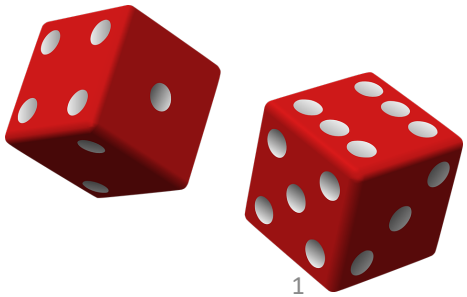
Sampling People, Records, & Networks

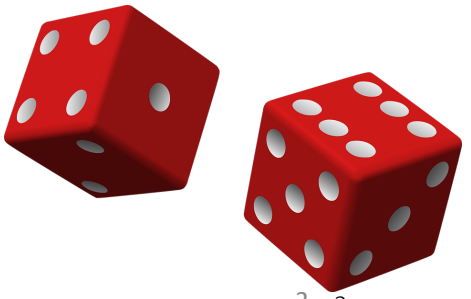
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



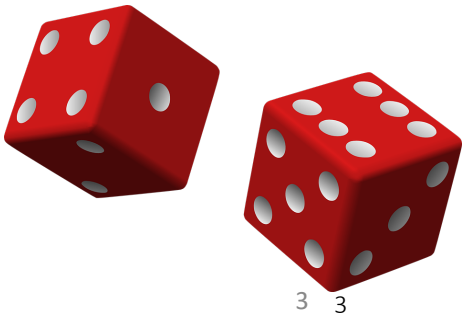
Unit 3

- 1 Simple complex
- 2 deff & roh
- 3 2-stage sampling
- 4 Designing 2-stage samples
- 5 Unequal sized clusters
- 6 Subsampling

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
  - Lecture 1: Simple complex sampling – choosing entire clusters
  - Lecture 2: Design effects & intraclass correlation
  - Lecture 3: Two-stage sampling
  - Lecture 4: Designing for two-stage samples
  - Lecture 5: Dealing with the real world – unequal sized clusters
  - Lecture 6: Subsampling
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications

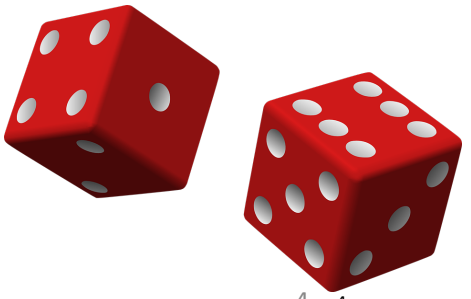


Unit 3

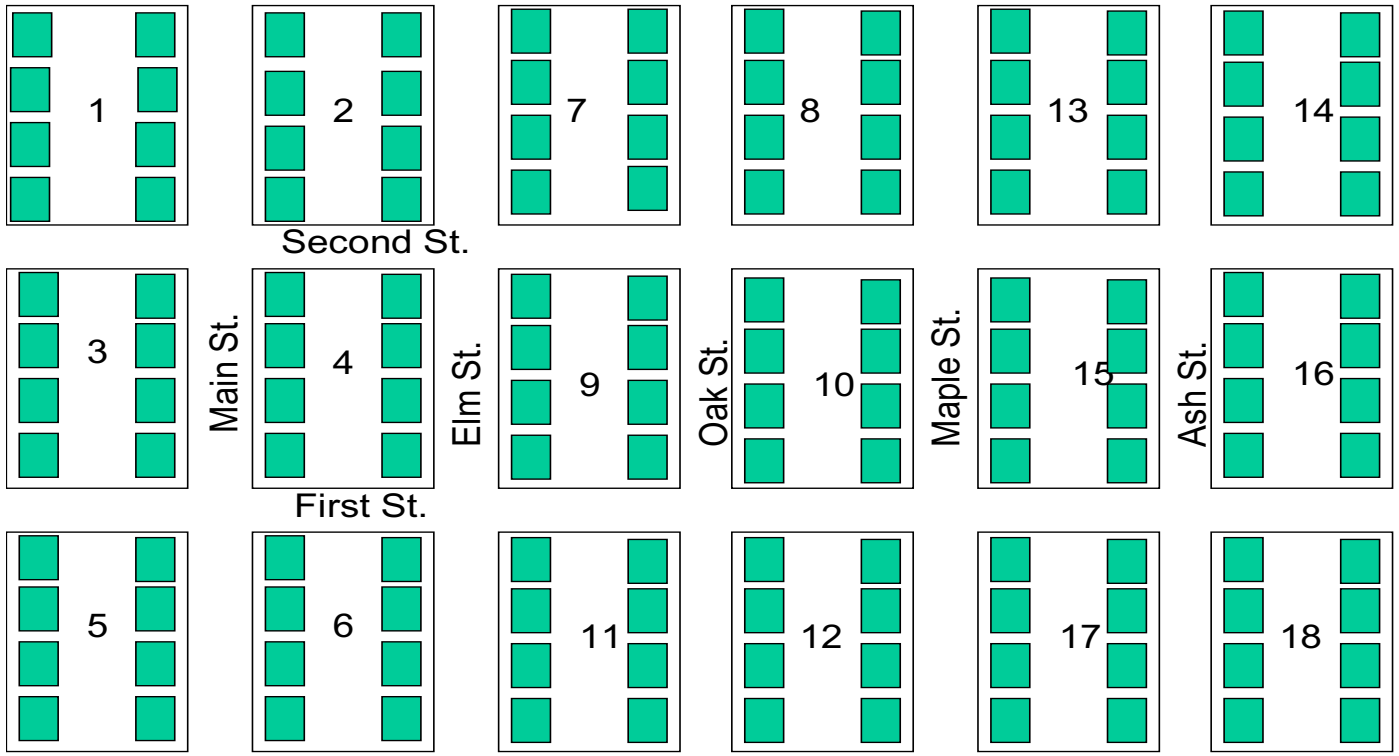


- Sample selection
- Sampling variance
- Design effect
- Change in design effect

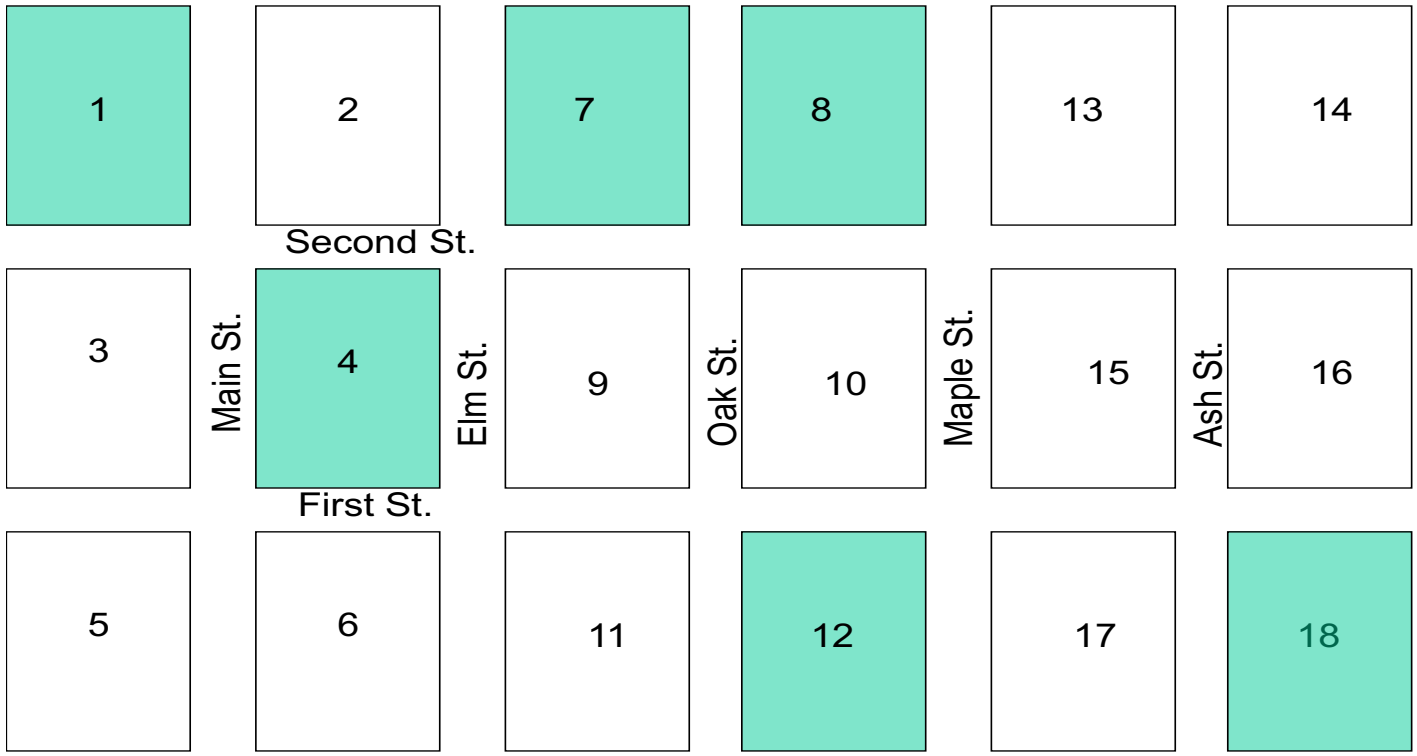
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
  - Lecture 1: Simple complex sampling – choosing entire clusters
  - Lecture 2: Design effects & intraclass correlation
  - **Lecture 3: Two-stage sampling**
  - Lecture 4: Designing for two-stage samples
  - Lecture 5: Dealing with the real world – unequal sized clusters
  - Lecture 6: Subsampling
- Unit 4: Being more efficient
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



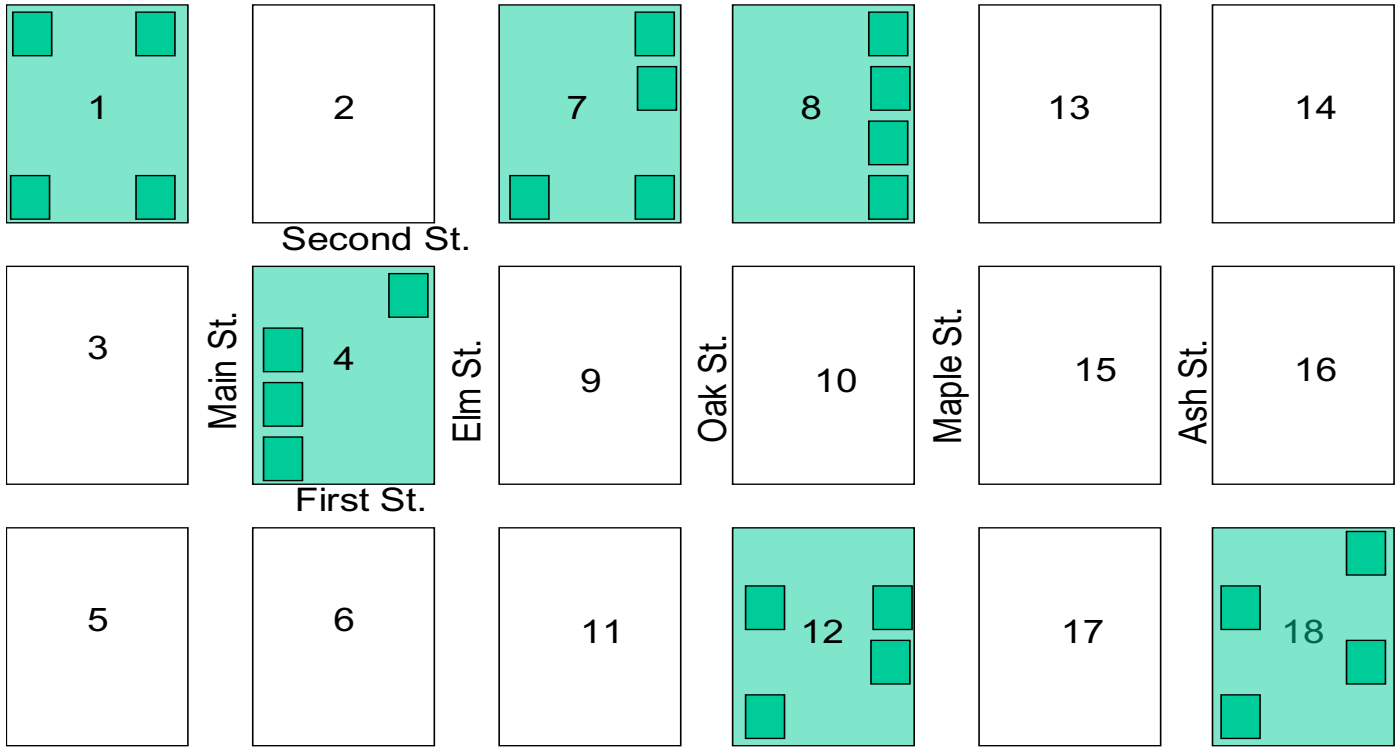
- **Sample selection**
- Sampling variance
- Design effect
- Change in design effect



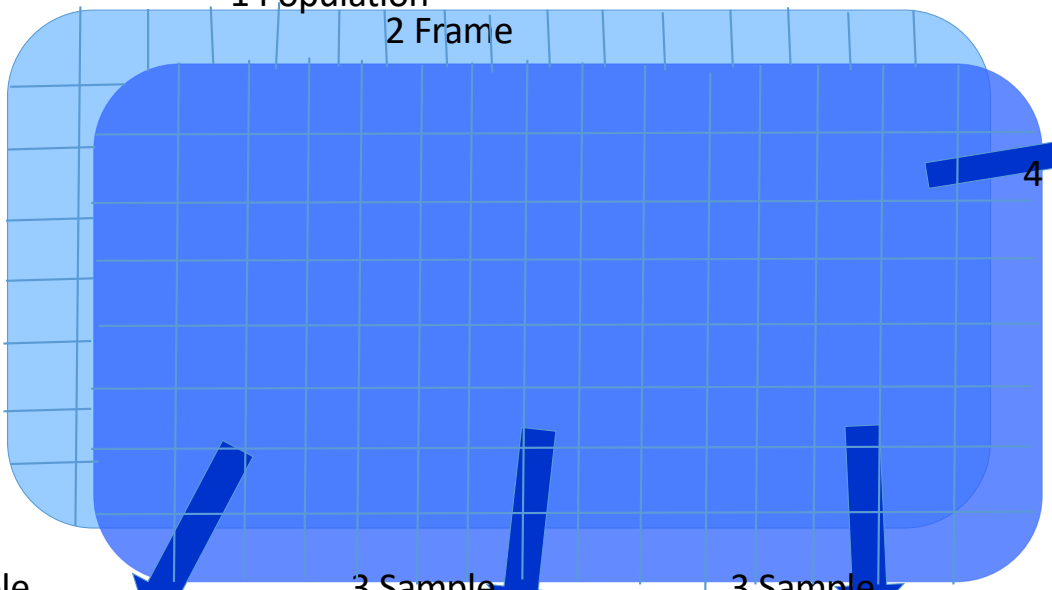
- **Sample selection**
- Sampling variance
- Design effect
- Change in design effect



- **Sample selection**
- Sampling variance
- Design effect
- Change in design effect



$$Var(\bar{y}) = \frac{(1-f)}{a} S_a^2$$



3 Sample

4 Estimate

$$\bar{y}_{\binom{N}{n}} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

5 Sampling distribution

$$Var(\bar{y}) = \frac{1}{\binom{N}{n}} \sum_{s=1}^{\binom{N}{n}} \left(\bar{y}_s - \bar{\bar{y}}\right)^2$$

6 Standard error

$$se(\bar{y}) = \sqrt{\frac{1-f}{a} s_a^2}$$

3 Sample

4 Estimate

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

3 Sample

4 Estimate

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

3 Sample

4 Estimate

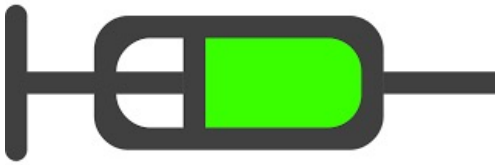
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- **Suppose we select an SRS of *a* = 20 classrooms from *A* = 1,000, and examine the immunization history of only *b* = 12 children in selected classrooms**
- **Here again**

$$N = A \times B = 1{,}000 \times 24 \; and \; n = a \times b = 240$$

- **For each of the a = 20 selected PSU's, we record the number of children immunized:**

$$\frac{4}{12},\frac{5}{12},\frac{5}{12},\frac{6}{12},\frac{6}{12},\frac{6}{12},\frac{7}{12},\frac{8}{12},\frac{8}{12},\frac{8}{12},$$
$$\frac{8}{12},\frac{8}{12},\frac{8}{12},\frac{9}{12},\frac{9}{12},\frac{10}{12},\frac{10}{12},\frac{11}{12},\frac{12}{12},\frac{12}{12}$$

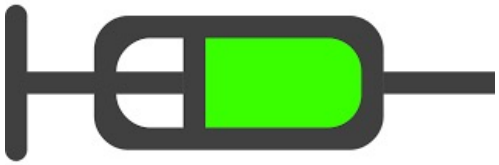


- **For each of the *a* = 20 selected PSU's, we record the number of children immunized:**

$$\frac{4}{12},\frac{5}{12},\frac{5}{12},\frac{6}{12},\frac{6}{12},\frac{6}{12},\frac{7}{12},\frac{8}{12},\frac{8}{12},\frac{8}{12},$$
$$\frac{8}{12},\frac{8}{12},\frac{8}{12},\frac{9}{12},\frac{9}{12},\frac{10}{12},\frac{10}{12},\frac{11}{12},\frac{12}{12},\frac{12}{12}$$

- **Again, the overall proportion immunized is**

$$p = 160/240 = 0.67$$



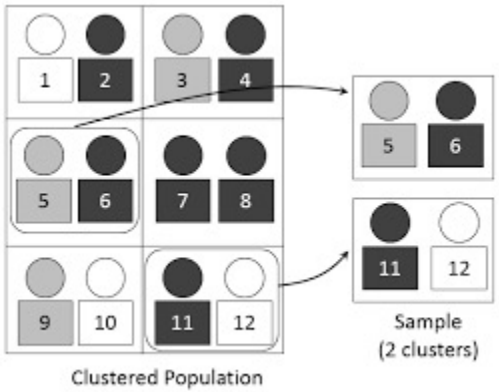
- Sample selection
- **Sampling variance**
- Design effect
- Change in design effect

- **Again, as in cluster sampling, treat the sample as an SRS of *a* = 20 units from *A* = 240:**

$$\text{var}(p) = \frac{(1-f)}{a} s_a^2$$

  - **Where**

$$s_a^2 = \sum_{\alpha=1}^{a} (p_\alpha - p)^2 \Big/ (a-1)$$

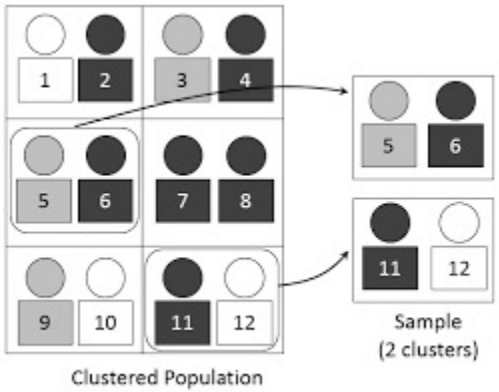


- **Again, as in cluster sampling, treat the sample as an SRS of *a* = 20 units from *A* = 240:**

$$\text{var}(p) = \frac{(1-f)}{a} s_a^2$$

  - **Where**

$$s_a^2 = \sum_{\alpha=1}^{a} (p_\alpha - p)^2 \Big/ (a-1) \qquad f = \frac{a}{A} \times \frac{b}{B}$$



- **Again, as in cluster sampling, treat the sample as an SRS of *a* = 20 units from *A* = 240:**

$$\mathrm{var}(p) = \frac{(1-f)}{a} s_a^2$$

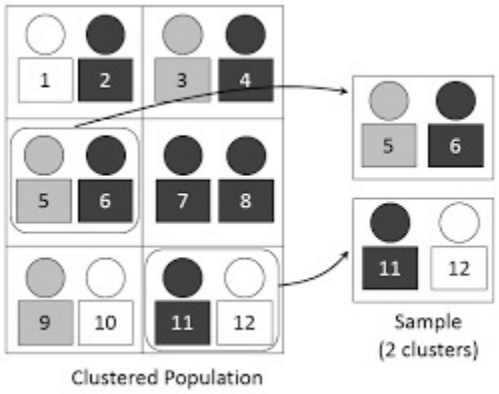
  - **Where**

$$s_a^2 = \sum_{\alpha=1}^{a} (p_\alpha - p)^2 \Big/ (a-1) \qquad f = \frac{a}{A} \times \frac{b}{B}$$

- **That is,**

$$\mathrm{var}(p) = \frac{(1-f)}{a} \frac{\sum_{\alpha=1}^{a} (p_\alpha - p)^2}{a-1}$$

$$se(p) = \sqrt{\mathrm{var}(p)}$$



- **The design effect for two-stage sampling is the same as for simple cluster sampling:**

$$deff(p) = \frac{\text{var}(p)}{\text{var}_{SRS}(p)}$$

- **The design effect for two-stage sampling is the same as for simple cluster sampling:**

$$deff(p) = \frac{\text{var}(p)}{\text{var}_{SRS}(p)}$$

- **Selecting many elements per cluster increases variances**

- **The design effect for two-stage sampling is the same as for simple cluster sampling:**

$$deff\left(p\right)=\frac{\mathrm{var}\left(p\right)}{\mathrm{var}_{SRS}\left(p\right)}$$

- **Selecting many elements per cluster increases variances**
- **As noted before, even small values of *roh* can be magnified by large *b* since**

$$deff\left(p\right)=1+\left(b-1\right)roh$$

- **One way to think about the design effect now is to see how it affects potentially the sampling variance**
- **Remember**

$$\text{var}(p) = deff(p) \times \text{var}_{SRS}(p)$$

- **One way to think about the design effect now is to see how it affects potentially the sampling variance**
- **Remember**

$$\text{var}(p) = deff(p) \times \text{var}_{SRS}(p)$$

- **If we keep the same sample size, then the SRS sampling variance does not change**
- **Then any change to the design effect is a change to the sampling variance**

- Sample selection
- Sampling variance
- Design effect
- **Change in design effect**

- **Manipulation of sampling fractions between first and second stages, maintaining the overall sample size, reveals the nature of the design effect, and the effective sample size**

- **Manipulation of sampling fractions between first and second stages, maintaining the overall sample size, reveals the nature of the design effect, and the effective sample size**
- **Sample *a* = 20 classrooms and *b* = 12:**

$$deff\left(p\right)=1+\left(12-1\right)\times 0.088=1.97 \qquad n_{eff}=122$$

- Sample selection
- Sampling variance
- Design effect
- **Change in design effect**

- **Manipulation of sampling fractions between first and second stages, maintaining the overall sample size, reveals the nature of the design effect, and the effective sample size**
- **Sample *a* = 20 classrooms and *b* = 12:**

$$deff(p) = 1 + (12 - 1) \times 0.088 = 1.97 \quad n_{eff} = 122$$

- **Sample *a* = 30 classrooms and *b* = 8:**

$$deff(p) = 1 + (8 - 1) \times 0.088 = 1.62 \quad n_{eff} = 148$$

- Sample selection
- Sampling variance
- Design effect
- Change in design effect

- **Manipulation of sampling fractions between first and second stages, maintaining the overall sample size, reveals the nature of the design effect, and the effective sample size**
- **Sample *a* = 20 classrooms and *b* = 12:**

$$deff\left(p\right)=1+\left(12-1\right)\times 0.088=1.97 \quad n_{eff}=122$$

- **Sample *a* = 30 classrooms and *b* = 8:**

$$deff\left(p\right)=1+\left(8-1\right)\times 0.088=1.62 \quad n_{eff}=148$$

- **Sample *a* = 80 classrooms and *b* = 3:**

$$deff\left(p\right)=1+\left(3-1\right)\times 0.088=1.18 \quad n_{eff}=204$$

Unit 3

- 1 Simple complex
- 2 deff & roh
- 3 2-stage sampling
- **4 Designing 2-stage samples**
- 5 Unequal sized clusters
- 6 Subsampling