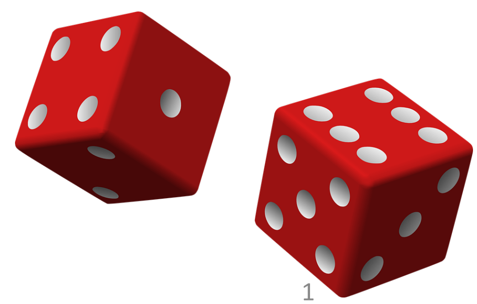
Sampling People, Records, & Networks

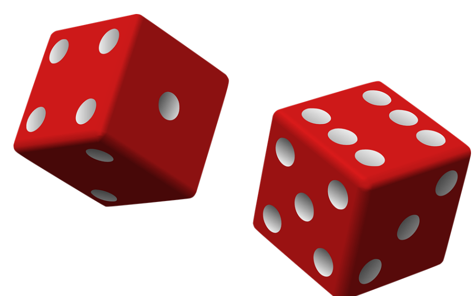
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



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

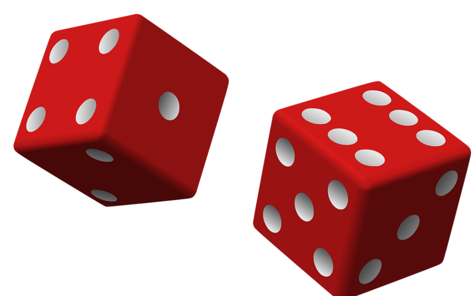


 2

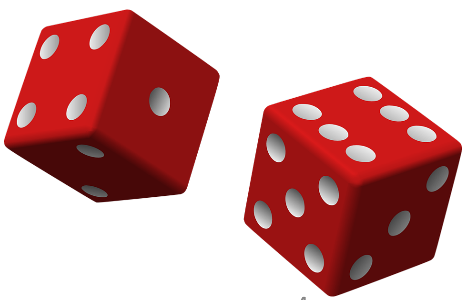
Unit 3

- 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



- Projecting design effects
- Design effects and sample size
- Projecting standard errors and confidence intervals



 4

- **Estimation**

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

- **Estimation**

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

$$\mathrm{var}_{(1),SRS}(p) = (1-f)\frac{p(1-p)}{n_{(1)}-1}$$

- **Estimation**

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

$$\text{var}_{(1),SRS}(p) = (1-f)\frac{p(1-p)}{n_{(1)}-1}$$

$$deff_{(1)} = \frac{\text{var}_{(1)}(p)}{\text{var}_{(1),SRS}(p)}$$

- **Estimation**

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

$$\mathrm{var}_{(1),SRS}(p) = (1-f)\frac{p(1-p)}{n_{(1)}-1}$$

$$deff_{(1)} = \frac{\mathrm{var}_{(1)}(p)}{\mathrm{var}_{(1),SRS}(p)}$$

$$roh = \frac{deff_{(1)}-1}{b_{(1)}-1}$$

- **Estimation**

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

$$\mathrm{var}_{(1),SRS}(p) = (1-f)\frac{p(1-p)}{n_{(1)}-1}$$

$$deff_{(1)} = \frac{\mathrm{var}_{(1)}(p)}{\mathrm{var}_{(1),SRS}(p)}$$

$$roh = \frac{deff_{(1)}-1}{b_{(1)}-1}$$

- **Design**

- **Estimation**

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

$$\mathrm{var}_{(1),SRS}(p) = (1-f)\frac{p(1-p)}{n_{(1)}-1}$$

$$deff_{(1)} = \frac{\mathrm{var}_{(1)}(p)}{\mathrm{var}_{(1),SRS}(p)}$$

$$roh = \frac{deff_{(1)}-1}{b_{(1)}-1}$$

- **Design**

$$\boxed{roh}$$

- **Estimation**

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

$$\mathrm{var}_{(1),SRS}(p) = (1-f)\frac{p(1-p)}{n_{(1)}-1}$$

$$deff_{(1)} = \frac{\mathrm{var}_{(1)}(p)}{\mathrm{var}_{(1),SRS}(p)}$$

$$roh = \frac{deff_{(1)}-1}{b_{(1)}-1}$$

- **Design**

$$deff_{(2)} = 1 + (b_{(2)}-1)roh$$

$$\boxed{roh}$$

- **Estimation**

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

$$\mathrm{var}_{(1),SRS}(p) = (1-f)\frac{p(1-p)}{n_{(1)}-1}$$

$$deff_{(1)} = \frac{\mathrm{var}_{(1)}(p)}{\mathrm{var}_{(1),SRS}(p)}$$

$$roh = \frac{deff_{(1)}-1}{b_{(1)}-1}$$

- **Design**

$$\mathrm{var}_{(2),SRS}(p) = \frac{p(1-p)}{n_{(2)}}$$

$$deff_{(2)} = 1 + (b_{(2)}-1)roh$$

$$\boxed{roh}$$

- **Estimation**

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

$$\mathrm{var}_{(1),SRS}(p) = (1-f)\frac{p(1-p)}{n_{(1)}-1}$$

$$deff_{(1)} = \frac{\mathrm{var}_{(1)}(p)}{\mathrm{var}_{(1),SRS}(p)}$$

$$roh = \frac{deff_{(1)}-1}{b_{(1)}-1}$$

- **Design**

$$\mathrm{var}_{(2)}(p) = deff_{(2)} \times \mathrm{var}_{(2),SRS}(p)$$

$$\mathrm{var}_{(2),SRS}(p) = \frac{p(1-p)}{n_{(2)}}$$

$$deff_{(2)} = 1 + (b_{(2)}-1)roh$$

$$\boxed{roh}$$

- **Estimation**

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

$$\mathrm{var}_{(1),SRS}(p) = (1-f)\frac{p(1-p)}{n_{(1)}-1}$$

$$deff_{(1)} = \frac{\mathrm{var}_{(1)}(p)}{\mathrm{var}_{(1),SRS}(p)}$$

$$roh = \frac{deff_{(1)}-1}{b_{(1)}-1}$$

- **Design**

$$\mathrm{var}_{(2)}(p) = deff_{(2)} \times \mathrm{var}_{(2),SRS}(p)$$

$$\mathrm{var}_{(2),SRS}(p) = \frac{p(1-p)}{n_{(2)}}$$

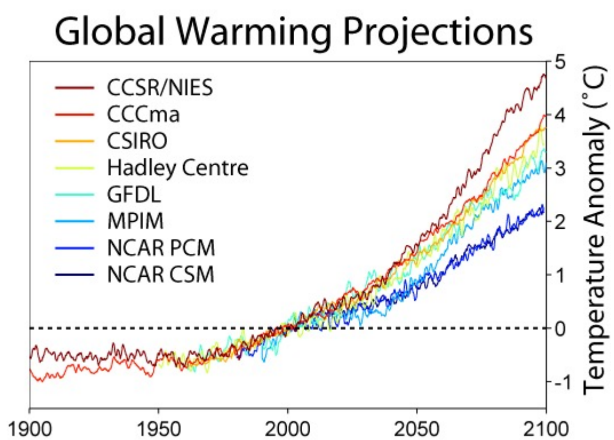
$$deff_{(2)} = 1 + (b_{(2)}-1)roh$$

$$\boxed{roh}$$

- **Projecting design effects**
- Design effects and sample size
- Projecting standard errors and confidence intervals

A. Suppose the sample described in Exercise 4 (with n = 2,400 and a = 60) is to be repeated with a smaller sample of n =1,200 and in only a = 30 equal-sized clusters

Project (say what is expected) how large the sampling variance of p will be under this new design.

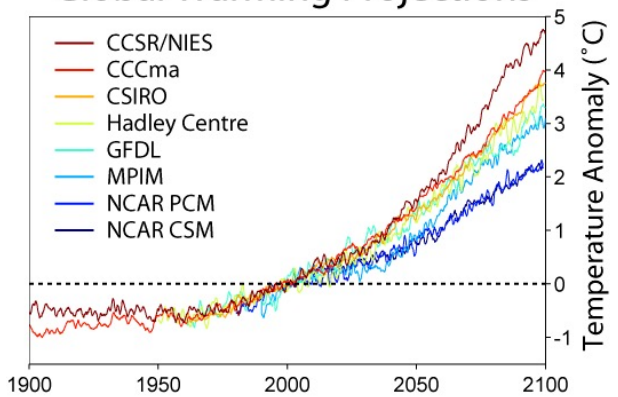


- **Projecting design effects**
- Design effects and sample size
- Projecting standard errors and confidence intervals

B. Now suppose the reduced size of *n* = 1,200 is retained, but we want to consider *a* = 60 equal-sized clusters.

Project (say what is expected) how large the sampling variance of *p* will be under this new design.

Global Warming Projections



- **Projecting design effects**
- Design effects and sample size
- Projecting standard errors and confidence intervals

1. **For A), compute the simple random sampling variance**

$$\text{var}_{SRS}(p) = \frac{p(1-p)}{n-1}$$

2. **Computed the design effect**

$$deff(p) = 1 + (b-1)roh = 1 + (40-1)roh$$

3. **Compute the projected sampling variance**

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

4. **For B), repeat steps 1-3 for the second design, replacing b = 40 with b = 20 in Step 2 above.**

- **Projecting design effects**
- Design effects and sample size
- Projecting standard errors and confidence intervals

For A), n = 1200 for a = 30 and b = 40.

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

Here,

$$deff(p) = 2.1795$$

Thus, using the design effect and new SRS variance, we can obtain *var*(*p*) under the new design.

- **Projecting design effects**
- Design effects and sample size
- Projecting standard errors and confidence intervals

And, ignoring the *fpc*,

$$\mathrm{var}_{SRS}\left(p\right)=\frac{p\left(1-p\right)}{n}=\frac{0.4\times 0.6}{1,200}=0.0002$$

Then,

$$\mathrm{var}\left(p\right)=2.1795\times\boxed{0.0002}=0.0004358$$

- **Projecting design effects**
- Design effects and sample size
- Projecting standard errors and confidence intervals

For B), when a = 60 and b = 20 for n = 1200,

$$deff\,(p) = 1 + (20 - 1)(0.03024) = 1.575$$

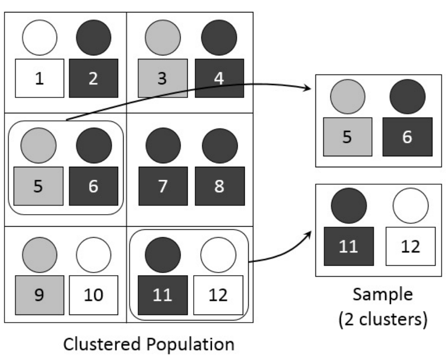
$$\text{var}\left(p\right) = 1.575 \times \boxed{0.0002} = 0.0003150$$

- **Projecting design effects**
- Design effects and sample size
- Projecting standard errors and confidence intervals

n	*a*	*b*	*deff*	*var(p)*
2400	60	40	2.1795	.000218
1200	30	40	2.1795	.000436
1200	60	20	1.5750	.000315

- Projecting design effects
- **Design effects and sample size**
- Projecting standard errors and confidence intervals

- **Design effects, when projected, can also help us determine sample size in cluster sampling**

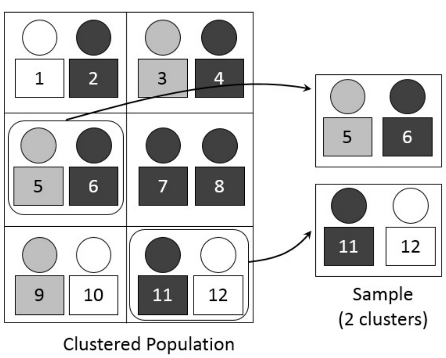


- Projecting design effects
- **Design effects and sample size**
- Projecting standard errors and confidence intervals

- **Design effects, when projected, can also help us determine sample size in cluster sampling**
- **Cluster sampling increases variances by a factor**

$$deff(p) = 1 + (b-1)roh$$

compared to SRS

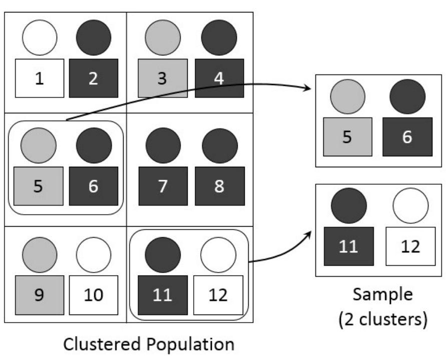


- Projecting design effects
- **Design effects and sample size**
- Projecting standard errors and confidence intervals

- **Design effects, when projected, can also help us determine sample size in cluster sampling**
- **Cluster sampling increases variances by a factor**

$$deff(p) = 1 + (b-1)roh$$

compared to SRS



- Projecting design effects
- **Design effects and sample size**
- Projecting standard errors and confidence intervals

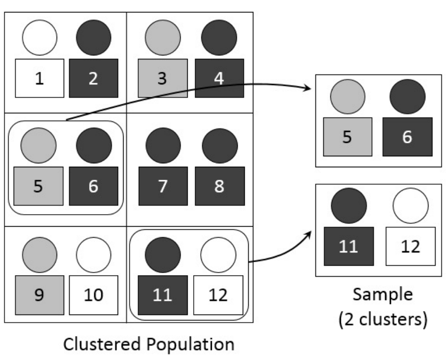
- **Design effects, when projected, can also help us determine sample size in cluster sampling**
- **Cluster sampling increases variances by a factor**

$$deff(p) = 1 + (b-1)roh$$

compared to SRS

- **Let's 'offset' this increase by increasing sample size by**

$$deff(p) = 1 + (b-1)roh$$



- Projecting design effects
- **Design effects and sample size**
- Projecting standard errors and confidence intervals

- **Design effects, when projected, can also help us determine sample size in cluster sampling**
- **Cluster sampling increases variances by a factor**

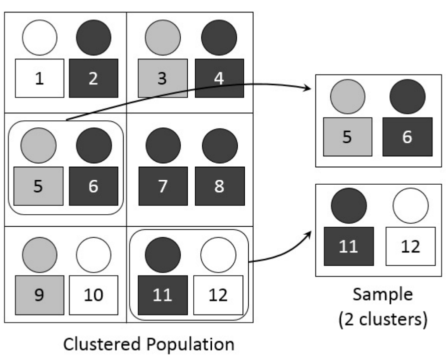
$$deff(p) = 1 + (b-1)roh$$

compared to SRS

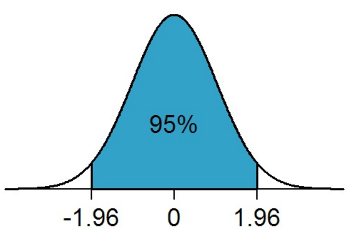
- **Let's 'offset' this increase by increasing sample size by**

$$deff(p) = 1 + (b-1)roh$$

- **That is, compute an SRS sample size and inflate it by a design effect …**

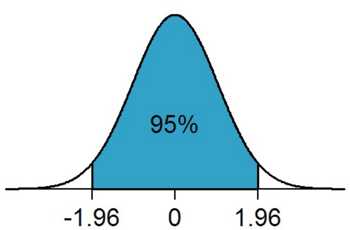


- **For example, suppose, for our proportion p = 0.4 we want a 95% confidence interval (0.37,0.43)**



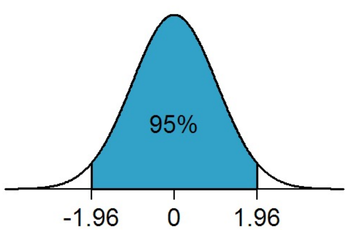
- Projecting design effects
- Design effects and sample size
- Projecting standard errors and confidence intervals

- For example, suppose, for our proportion $p = 0.4$ we want a 95% confidence interval (0.37,0.43)
- This is a "margin of error" of 0.03 …



- Projecting design effects
- Design effects and sample size
- Projecting standard errors and confidence intervals

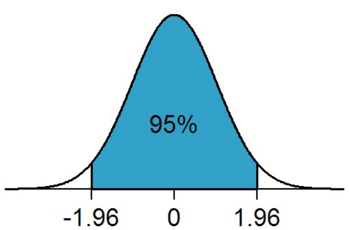
- For example, suppose, for our proportion $p = 0.4$ we want a 95% confidence interval (0.37,0.43)
- This is a "margin of error" of 0.03 …
- Or a standard error of 0.015 ...



- Projecting design effects
- **Design effects and sample size**
- Projecting standard errors and confidence intervals

- For example, suppose, for our proportion $p = 0.4$ we want a 95% confidence interval (0.37,0.43)
- This is a "margin of error" of 0.03 …
- Or a standard error of 0.015 ...
- Which for a proportion yields an SRS sample size

$$n_{SRS} = \frac{S^2}{\left[se(p)\right]^2} = \frac{(0.4)(1-0.4)}{\left[0.015\right]^2} = 1066.67$$



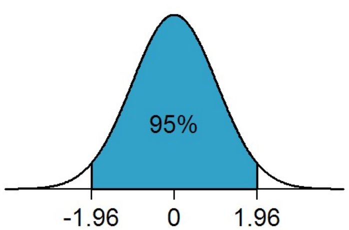
- Projecting design effects
- Design effects and sample size
- Projecting standard errors and confidence intervals

- For example, suppose, for our proportion p = 0.4 we want a 95% confidence interval (0.37,0.43)
- This is a "margin of error" of 0.03 …
- Or a standard error of 0.015 ...
- Which for a proportion yields an SRS sample size

$$n_{SRS} = \frac{S^2}{\left[se\left(p\right)\right]^2} = \frac{\left(0.4\right)\left(1-0.4\right)}{\left[0.015\right]^2} = 1066.67$$

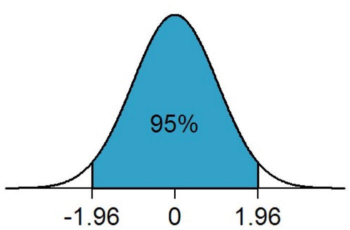
- If the cluster sample has *deff* = 2.1795, the sample size for the cluster sample would be

$$n = n_{SRS} \times deff\left(p\right) = 1066.67 \times 2.1795 \approx 2{,}325$$



- Projecting design effects
- Design effects and sample size
- **Projecting standard errors and confidence intervals**

- **We can take the variance projection one step further, and project what a 95% confidence interval would look like …**



- **Projecting design effects**
- Design effects and sample size
- Projecting standard errors and confidence intervals

For B), when *a* = 60 and *b* = 20 for *n* = 1200,

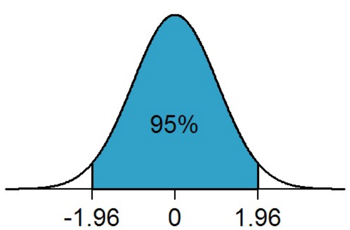
$$deff(p) = 1 + (20-1)(0.03024) = 1.575$$

$$\text{var}\left(p\right) = 1.575 \times \boxed{0.0002} = 0.0003150$$

… the 95% confidence interval, using the 'Normal' distribution multiplier, is

$$\left(0.4 - 1.96 \times \sqrt{0.000315}, 0.4 + 1.96 \times \sqrt{0.000315}\right)$$

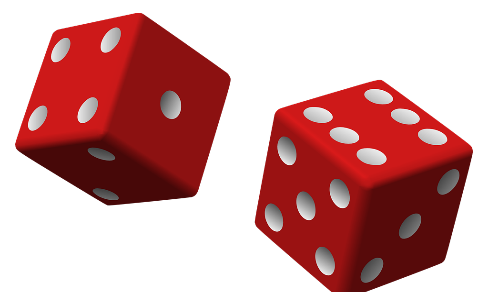
$$\left(0.365, 0.435\right)$$



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



 34