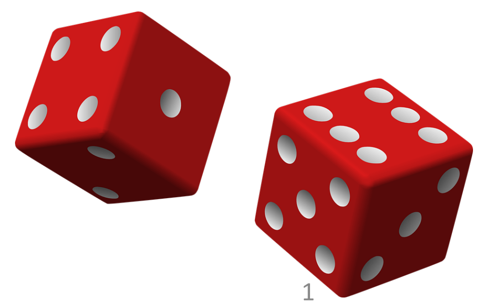
Sampling People, Records, & Networks

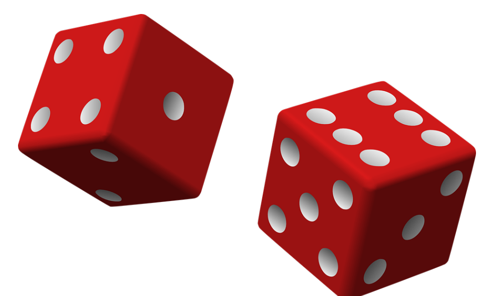
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



Unit 4

- 1 Forming groups
- 2 Sampling variance
- 3 More on grouping
- 4 Allocate sample
- 5 Other allocations
- 6 Weights

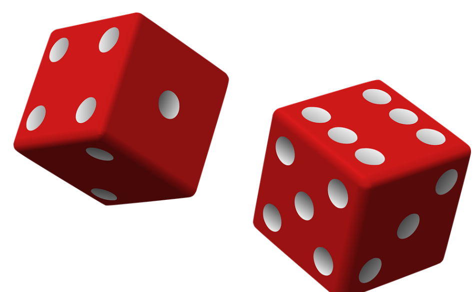
- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- **Unit 4: Being more efficient**
  - Forming groups
  - Sampling variance
  - More on grouping
  - Allocate sample
  - Other allocations
  - Weights to combine across strata
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



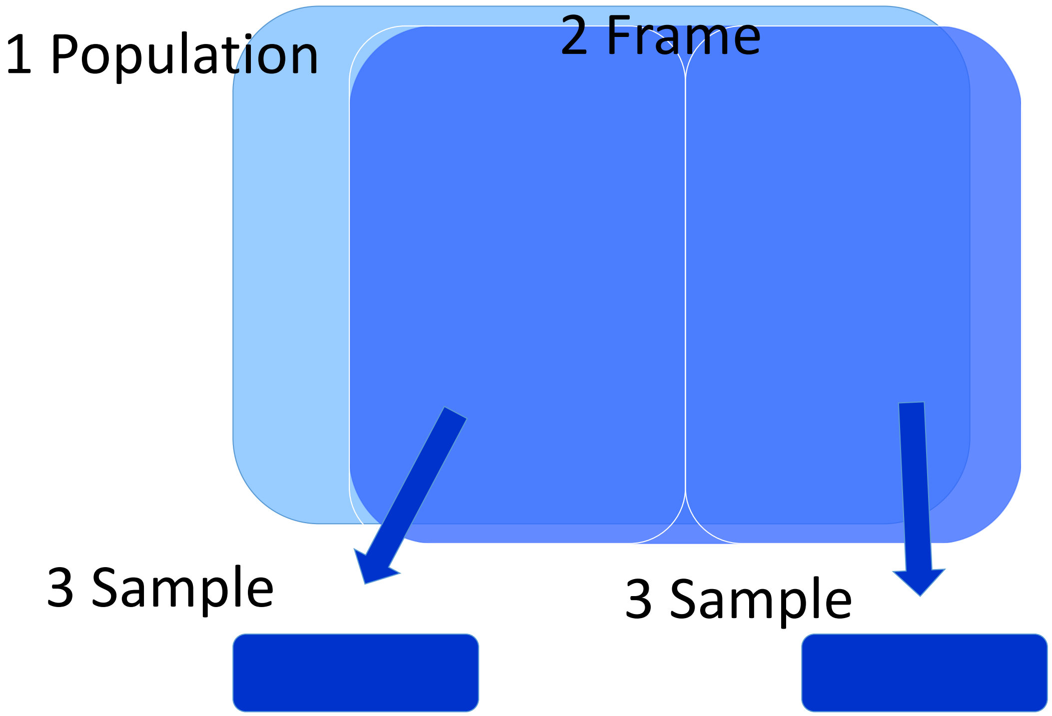
2

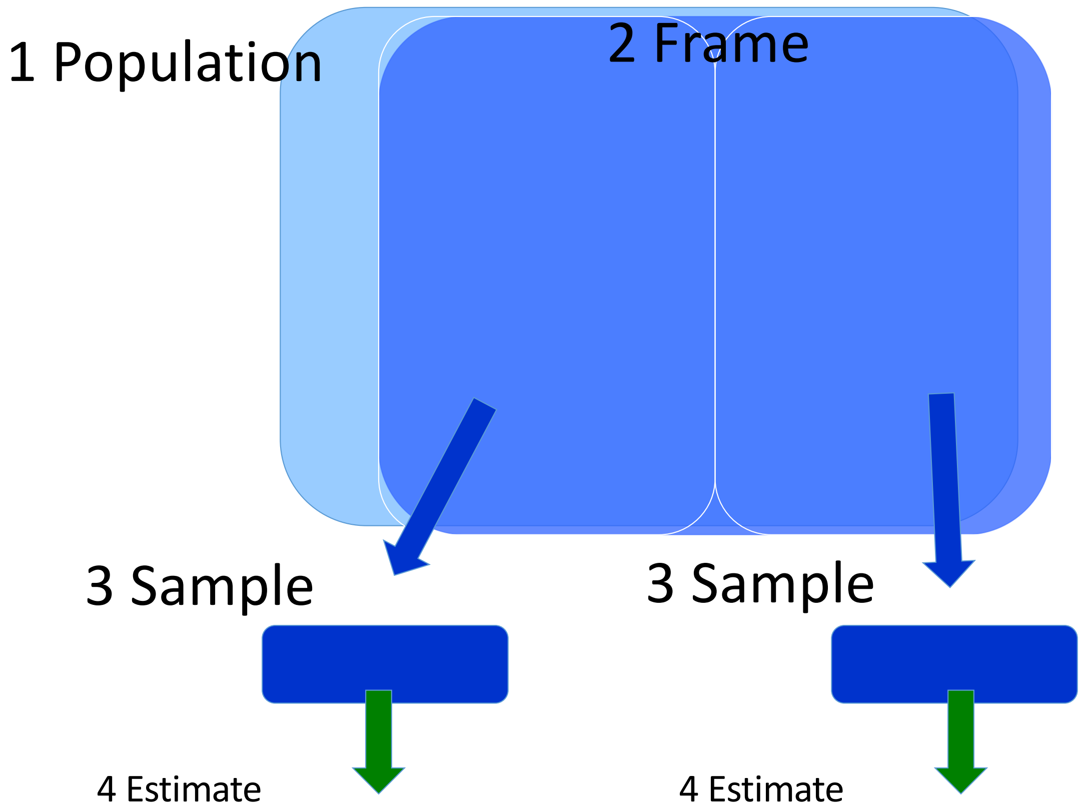
- Principles
- Example
- Confidence interval
- Design effect
- Effective sample size

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
  - Forming groups
  - **Sampling variance**
  - More on grouping
  - Allocate sample
  - Other allocations
  - Weights to combine across strata
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



3


1 Population
2 Frame
3 Sample
3 Sample



$$\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i} \qquad \bar{y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i} \;\Rightarrow\; \bar{y}_w = \sum_{h=1}^{2} W_h \bar{y}_h$$

- **Principles**
- Example
- Confidence interval
- Design effect
- Effective sample size

- **In theory,**

$$Var\left(\bar{y}\right) = \sum_{h=1}^{H} W_h^2 Var\left(\bar{y}_h\right)$$

- **Principles**
- Example
- Confidence interval
- Design effect
- Effective sample size

- **In theory,**

$$Var\left(\bar{y}\right)=\sum_{h=1}^{H}W_h^2 Var\left(\bar{y}_h\right)$$

- **Estimate this variance by**

$$var\left(\bar{y}\right)=\sum_{h=1}^{H}W_h^2 var\left(\bar{y}_h\right)$$

- **Principles**
- Example
- Confidence interval
- Design effect
- Effective sample size

- **In theory,**

$$Var\left(\bar{y}\right)=\sum_{h=1}^{H}W_h^2 Var\left(\bar{y}_h\right)$$

- **Estimate this variance by**

$$var\left(\bar{y}\right)=\sum_{h=1}^{H}W_h^2 var\left(\bar{y}_h\right)$$

- **And what is** $var(\bar{y}_h)$ **?**
- **For SRS within strata,** $var(\bar{y}_h) = \frac{(1-f_h)}{n_h} s_h^2$

- **Principles**
- Example
- Confidence interval
- Design effect
- Effective sample size

- **In theory,**

$$Var\left(\bar{y}\right) = \sum_{h=1}^{H} W_h^2 Var\left(\bar{y}_h\right)$$

- **Estimate this variance by**

$$var\left(\bar{y}\right) = \sum_{h=1}^{H} W_h^2 var\left(\bar{y}_h\right)$$

- **And what is** $var(\bar{y}_h)$ **?**
- **For SRS within strata,** $var(\bar{y}_h) = \frac{(1-f_h)}{n_h} s_h^2$
- **We thus need the within stratum variances:**

- Principles
- **Example**
- Confidence interval
- Design effect
- Effective sample size

h	Stratum	N_h	W_h	n_h	f_h	$\bar{y}_h$	s_h^2
1	Assistant	115	0.2875	23	0.2	50	125
2	Associate	75	0.1875	15	0.2	70	250
3	Full	210	0.5250	42	0.2	90	500
Total		400	1.0000	80	0.2	$74.75	

- **This simplifies a little, since here $1\text{-}f_h = 0.8$**
- **The calculation is really beyond the scope of this course, but since we've come this far …**

- Principles
- **Example**
- Confidence interval
- Design effect
- Effective sample size

- **This simplifies a little, since here $1\text{-}f_h = 0.8$**
- **The calculation is really beyond the scope of this course, but since we've come this far …**
- $var(\bar{y}) = \frac{(0.2875)^2(0.8)\,(125)}{23} + \frac{(0.1875)^2\,(0.8)250}{15} + \frac{(0.5250)^2\,(0.8)500}{42} = 3.453$

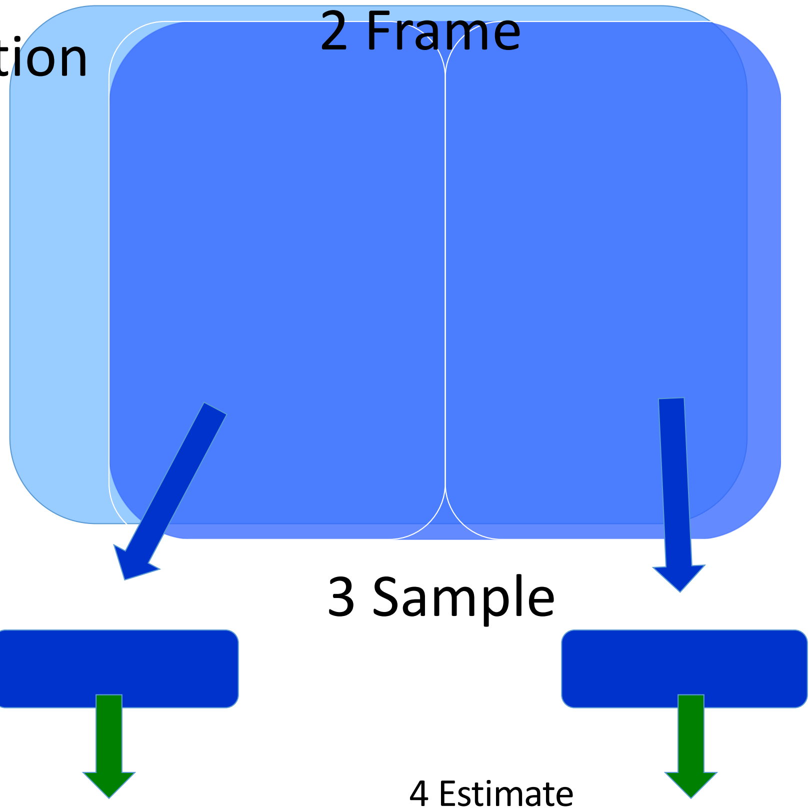
- **This simplifies a little, since here *1-f_h = 0.8***
- **The calculation is really beyond the scope of this course, but since we've come this far …**
- $var(\bar{y}) = \frac{(0.2875)^2(0.8)\,(125)}{23} + \frac{(0.1875)^2\,(0.8)250}{15} + \frac{(0.5250)^2\,(0.8)500}{42} = 3.453$
- **And** $se(\bar{y}) = 1.858$

- Principles
- **Example**
- Confidence interval
- Design effect
- Effective sample size

- **This simplifies a little, since here *1-f_h = 0.8***
- **The calculation is really beyond the scope of this course, but since we've come this far …**
- $var(\bar{y}) = \frac{(0.2875)^2(0.8)\,(125)}{23} + \frac{(0.1875)^2\,(0.8)250}{15} + \frac{(0.5250)^2\,(0.8)500}{42} = 3.453$
- **And** $se(\bar{y}) = 1.858$
- **Thus, we have completed step 6a & 6b – within stratum sampling variances and combining across strata**

1 Population

2 Frame



3 Sample

3 Sample

4 Estimate

4 Estimate

$$\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i}$$

$$\bar{y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i} \Rightarrow \bar{y}_w = \sum_{h=1}^{2} W_h \bar{y}_h$$

6 Standard error

$$se\left(\bar{y}\right) = \sqrt{\sum_{h=1}^{2} W_h^2 \frac{\left(1-f_h\right)}{n_h} s_h^2}$$

7 Confidence interval

- Principles
- Example
- **Confidence interval**
- Design effect
- Effective sample size

- **The last step, step 7, is the confidence interval**
- **Here, let's use the *t*-distribution**
- **We have $n_h - 1$ degrees of freedom for each stratum and $n - H$ for combining across strata**
- **For a 95% confidence interval, then, $t_{(0.975,80-3)} = 1.991$**

- Principles
- Example
- **Confidence interval**
- Design effect
- Effective sample size

- **The last step, step 7, is the confidence interval**
- **Here, let's use the *t*-distribution**
- **We have $n_h - 1$ degrees of freedom for each stratum and $n - H$ for combining across strata**
- **For a 95% confidence interval, then, $t_{(0.975,80-3)} = 1.991$**
- **The 95% confidence interval is then**
- $\left(\bar{y} - t_{(0.975,77)} se(\bar{y}), \bar{y} + t_{(0.975,77)} se(\bar{y})\right)$
- $(74.75 - 1.991 \times 1.858, 74.75 + 1.991 \times 1.858)$
- $(71.05, 78.45)$

- Principles
- Example
- Confidence interval
- Design effect
- Effective sample size

- Here, as in cluster sampling, there is another issue to be addressed – how does the sampling variance from stratified sampling compare to simple random sampling?
- That is, what is the design effect, deff?

- Principles
- Example
- Confidence interval
- Design effect
- Effective sample size

- **Here, as in cluster sampling, there is another issue to be addressed – how does the sampling variance from stratified sampling compare to simple random sampling?**
- **That is, what is the design effect, deff?**
- **For SRS, we need to calculate** $var(\bar{y}) = (1-f)\frac{s^2}{n}$
- **From a separate calculation,** $s^2 = 647.8$

RATIO®

- Principles
- Example
- Confidence interval
- **Design effect**
- Effective sample size

- **For our sample of size n = 80, from a population of size N = 400, or a sampling fraction f = 0.2,**
  $var(\bar{y}) = (1 - \mathbf{0.2})\frac{\mathbf{647.8}}{\mathbf{80}}$ **=6.478**

RATIO®

- Principles
- Example
- Confidence interval
- **Design effect**
- Effective sample size

- **For our sample of size n = 80, from a population of size N = 400, or a sampling fraction f = 0.2,** $var(\bar{y}) = (1 - \mathbf{0.2})\frac{\mathbf{647.8}}{\mathbf{80}}$ **=6.478**
- **Compared to the stratified sample,** $deff(\bar{y}) = \frac{var(\bar{y})}{var_{SRS}(\bar{y})} = \frac{3.453}{6.478} = 0.5331$

RATIO®

- Principles
- Example
- Confidence interval
- **Design effect**
- Effective sample size

- **For our sample of size *n* = 80, from a population of size *N* = 400, or a sampling fraction *f* = 0.2,** $var(\bar{y}) = (1 - \mathbf{0.2})\frac{\mathbf{647.8}}{\mathbf{80}}$ **=6.478**
- **Compared to the stratified sample,** $deff(\bar{y}) = \frac{var(\bar{y})}{var_{SRS}(\bar{y})} = \frac{3.453}{6.478} = 0.5331$
- **A 47% reduction in sampling variance**

RATIO®

- Principles
- Example
- Confidence interval
- Design effect
- **Effective sample size**

- **As we did for cluster samples, this design effect can be thought of in several ways**

- Principles
- Example
- Confidence interval
- Design effect
- **Effective sample size**

- **As we did for cluster samples, this design effect can be thought of in several ways**
- **For example, the actual sample size of 400 in the stratified proportionately allocated sample is the equivalent of having a simple random sample of**

$$n_{eff} = \frac{400}{0.5331} = 750$$

- Principles
- Example
- Confidence interval
- Design effect
- **Effective sample size**

- **As we did for cluster samples, this design effect can be thought of in several ways**
- **For example, the actual sample size of 400 in the stratified proportionately allocated sample is the equivalent of having a simple random sample of**

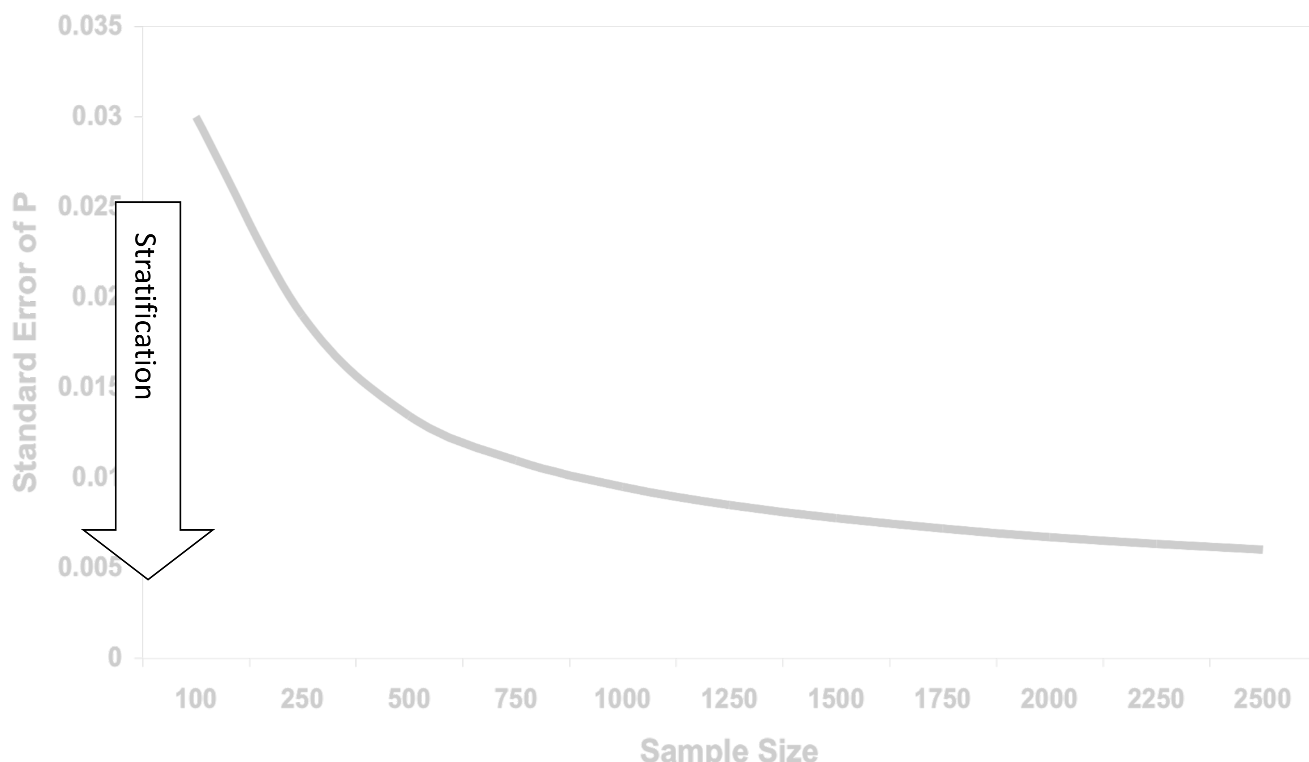
$$n_{eff} = \frac{400}{0.5331} = 750$$

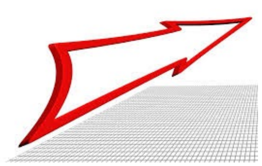
- **That is, the gains in precision are the equivalent of adding 350 cases to the sample**

- Principles
- Example
- Confidence interval
- Design effect
- **Effective sample size**

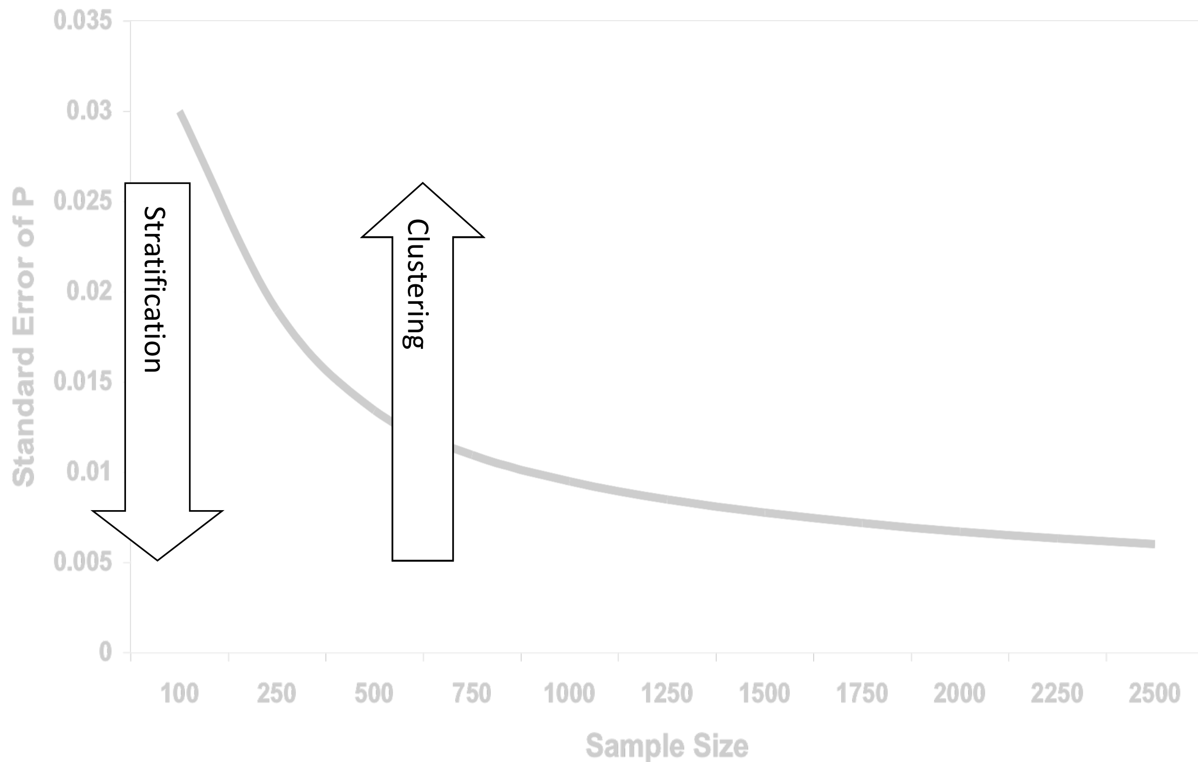
- **As we did for cluster samples, this design effect can be thought of in several ways**
- **For example, the actual sample size of 400 in the stratified proportionately allocated sample is the equivalent of having a simple random sample of** $n_{eff} = \frac{400}{0.5331} = 750$
- **That is, the gains in precision are the equivalent of adding 350 cases to the sample**
- **Alternatively, the standard errors are smaller and the confidence intervals are narrower by a factor of** $1 - \sqrt{0.5331} = 1 - 0.7301 = 0.2699$, **27%**

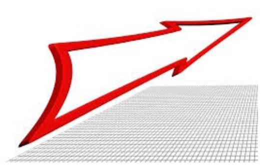
- Principles
- Example
- Confidence interval
- Design effect
- **Effective sample size**



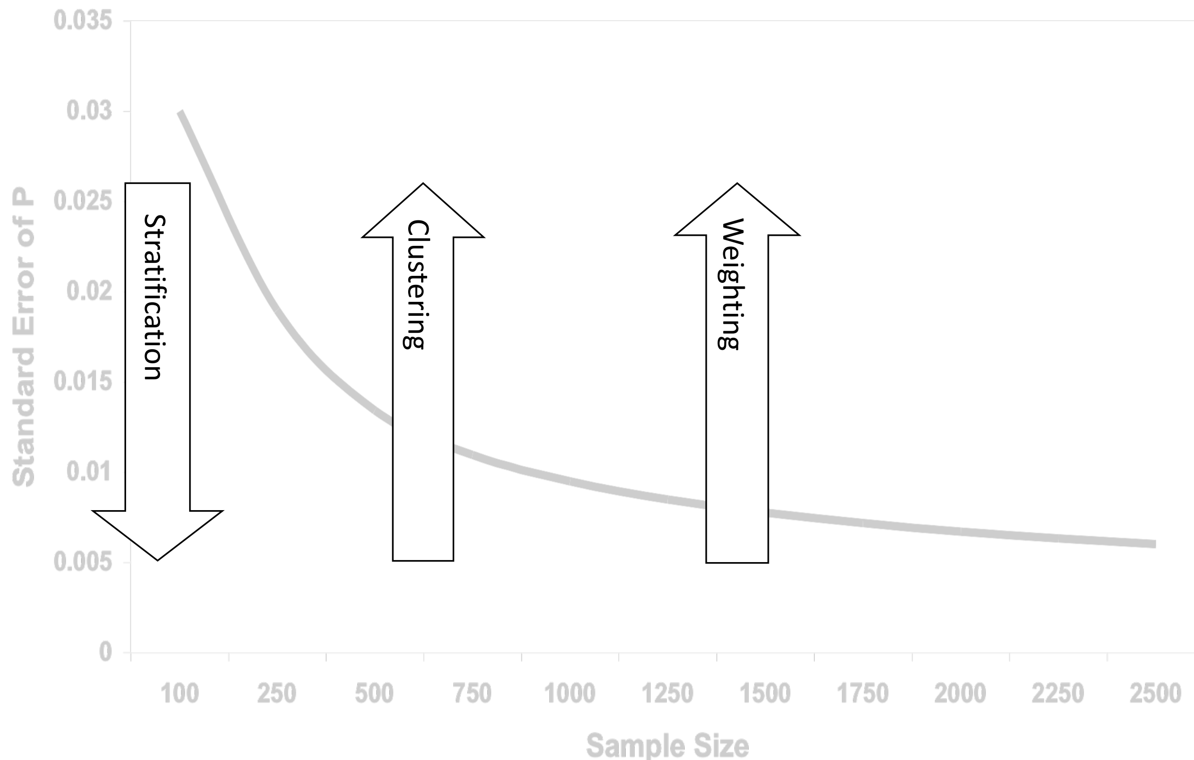


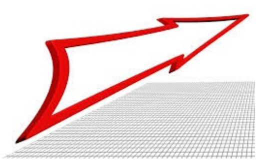
- Principles
- Example
- Confidence interval
- Design effect
- **Effective sample size**





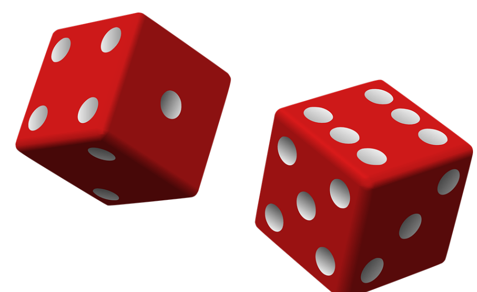
- Principles
- Example
- Confidence interval
- Design effect
- **Effective sample size**





Unit 4

- 1 Forming groups
- 2 Sampling variance
- 3 More on grouping
- 4 Allocate sample
- 5 Other allocations
- 6 Weights



30